Statistical Consistency of Distributionally Robust Convex Optimization

Qiaoyi Wang*
Chongqing Normal University
National Center for Applied Mathematics in Chongqing
Chongqing, China

April 29, 2026

Abstract

We study the statistical consistency of distributionally robust optimization (DRO) with general ambiguity sets. While convergence of optimal values is well understood, a unified set-valued analysis of feasible regions and solution sets remains largely missing, especially for constrained DRO. We develop a unified framework for statistical consistency based on an abstract reference consistency and a collapse principle. We establish uniform convergence of robust objective and constraint functionals, and combine it with Painlevé–Kuratowski (PK) set convergence to derive consistency of optimal values and solution sets. In particular, we prove upper convergence of minimizers and convergence of feasible regions, providing a full set-valued characterization of statistical consistency. The framework applies to general discrepancy-based ambiguity sets and extends naturally to constrained and multi-constraint DRO problems. All results hold almost surely.

Keywords: Distributionally robust optimization, Statistical consistency, Set convergence, PK convergence

1 Introduction

We consider a class of data-driven distributionally robust optimization (DRO) problems of the form

$$\min_{x\in\mathcal{X}} \sup_{P\in\mathcal{P}_N} \int_{\Xi} f(x,\xi)\, dP(\xi), \tag{1}$$

where $\mathcal{X}$ is a compact decision set, ξ takes values in an uncertainty space Ξ (which we assume to be a Polish space), and $\mathcal{P}_N$ is an ambiguity set constructed from data. In

*Corresponding author. Email: 2024110518099@stu.cqnu.edu.cn

contrast to classical stochastic programming, which assumes full knowledge of the data-generating distribution P^* [?], DRO evaluates decisions against a family of plausible distributions and thus provides robustness against distributional uncertainty.

A standard approach in data-driven DRO is to construct $\mathcal{P}_N$ as a neighborhood of a reference distribution P_N^{ref} (e.g., the empirical distribution). A prominent example is the Wasserstein ambiguity set, which has been extensively studied due to its favorable statistical and computational properties [?, ?]. Under suitable regularity conditions, such as Lipschitz continuity of $f(x, \cdot)$ and appropriate decay of the ambiguity radius, it is known that the DRO problem enjoys both finite-sample guarantees and asymptotic consistency.

Despite recent advances in distributionally robust optimization (DRO), a systematic variational treatment of solution and feasibility convergence remains largely underdeveloped, especially for constrained problems. Compared with specific DRO frameworks that rely on empirical reference measures and exact metric structures, our approach abstracts away from the specific construction of the reference distribution and the exact metric used.

The purpose of this paper is to develop such a unified framework. Our approach is based on a simple but fundamental observation: the core statistical properties of DRO depend mainly on three modular components: (i) the reference measure consistency, which centers the ambiguity set around the true distribution; (ii) the sequential weak collapse of the ambiguity set, which controls the radius and shape of the worst-case expectations; and (iii) the variational stability of the optimization problem itself.

Building on this collapse principle, we develop a variational analysis framework for DRO. Under suitable regularity conditions, we establish uniform convergence of the empirical robust objective and constraint functionals, and combine this with Painlevé–Kuratowski (PK) set convergence to derive consistency of optimal values and solution sets. In particular, we show that

$$\hat{v}_N \to v^*, \qquad \operatorname*{Limsup}_{N\to\infty} \hat{X}_N \subseteq X^*, \quad \text{a.s.}$$

which provides a full set-valued characterization of solution convergence.

Furthermore, we extend the analysis to constrained DRO problems. We prove that the feasible regions converge in the PK sense, and establish consistency of constrained optimal solutions. The framework naturally extends to problems with multiple constraints, without requiring additional structural assumptions. All convergence results hold almost surely.

2 Problem Formulation and Statistical Framework

2.1 Data-generating process and Test Functions

Let the uncertainty set be a closed set $\Xi \subset \mathbb{R}^n$, equipped with its Borel σ-algebra. Depending on the setting, we may further assume that Ξ is compact to avoid tail-related technicalities. If Ξ is not compact, appropriate light-tail or moment conditions must be imposed, as in the standard Wasserstein DRO framework [?].

Let $P^* \in \mathscr{P}(\Xi)$ denote the unknown data-generating distribution. We consider an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which all random objects are defined. In canonical data-driven settings, this may be taken as $(\Xi^\infty, \mathcal{B}(\Xi)^{\otimes\infty}, (P^*)^\infty)$.

Let $\mathcal{F}$ be a class of real-valued measurable functions on Ξ, serving as the test functions for our analysis.

We consider a sequence of reference measures $\{P_N^{\mathrm{ref}}\} \subset \mathscr{P}(\Xi)$ and an associated sequence of ambiguity sets $\mathcal{P}_N \subset \mathscr{P}(\Xi)$. The notation P_N^{ref} emphasizes that this sequence represents a general reference distribution. It is not restricted to the empirical measure and may arise from weighted empirical constructions, conditional models, bootstrap procedures, or other estimators.

We denote the target distribution by $P^\circ \in \mathscr{P}(\Xi)$. All asymptotic statements in this paper are understood to hold almost surely with respect to $\mathbb{P}$.

2.2 Optimization model

Assumption 2.1 (Regular integrand)**.** Let $\mathcal{X} \subset \mathbb{R}^d$ be nonempty and compact, equipped with the Euclidean metric. Let $f : \mathcal{X} \times \Xi \to \mathbb{R}$ satisfy:

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

(ii) There exists a nondecreasing continuous function $\rho_f : [0, \infty) \to [0, \infty)$, with $\rho_f(0) = 0$, such that

$$|f(x, \xi) - f(x', \xi)| \le \rho_f(d_{\mathcal{X}}(x, x')), \qquad \forall x, x' \in \mathcal{X},\ \forall \xi \in \Xi.$$

2.3 Reference Consistency and Collapse Assumption

Our consistency analysis decouples the statistical properties of the reference measure from the geometric properties of the ambiguity set. The reference consistency controls the center P_N^{ref}, while the collapse assumption controls the radius and shape of $\mathcal{P}_N$.

Assumption 2.2 (Pointwise reference consistency)**.** For every $\psi \in \mathcal{F}$, there exists an event A_ψ^R with $\mathbb{P}(A_\psi^R) = 1$ such that, on A_ψ^R,

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

We now introduce the sequential collapse assumption, which dictates that any sequence chosen from the ambiguity set must become indistinguishable from the reference measure.

Assumption 2.3 (Pointwise sequential weak collapse)**.** For every $\psi \in \mathcal{F}$, there exists an event A_ψ^C with $\mathbb{P}(A_\psi^C) = 1$ such that, on A_ψ^C, for every sequence $P_N \in \mathcal{P}_N$,

$$\int_\Xi \psi \, dP_N - \int_\Xi \psi \, dP_N^{\text{ref}} \to 0.$$

Assumption 2.3 requires that we first arbitrarily pick a test function ψ, and then arbitrarily pick a sequence $P_N \in \mathcal{P}_N$. We do not require uniform convergence via $\sup_{\psi \in \mathcal{F}}$ in the theoretical body. If a specific discrepancy metric (like Wasserstein or MMD) admits a supremum representation, that structure is utilized only in the verification examples.

Assumption 2.4 (Nonemptiness of ambiguity sets)**.** On the full-probability event under consideration, $\mathcal{P}_N \neq \emptyset$ for every N.

3 Statistical Consistency

We first formalize the probabilistic collapse result for ambiguity sequences on the abstract test class $\mathcal{F}$, and then analyze the deterministic convergence of the associated robust objective and solution sets.

Lemma 3.1 (Pointwise collapse to the target)**.** *Suppose Assumptions 2.2 and 2.3 hold. For every $\psi \in \mathcal{F}$, there exists an event*

$$A_\psi := A_\psi^R \cap A_\psi^C$$

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. Fix $\psi \in \mathcal{F}$ and define $A_\psi := A_\psi^R \cap A_\psi^C$. Then $\mathbb{P}(A_\psi) = 1$. For every $\omega \in A_\psi$ and every sequence $P_N \in \mathcal{P}_N(\omega)$,

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

The first term converges to zero by Assumption 2.3, and the second term converges to zero by Assumption 2.2. □

Under Assumptions 2.1, 2.2, 2.3, and 2.4, we define the empirical robust objective for a given sample path ω as

$$\mathcal{R}_N(f)(x,\omega) := \sup_{P\in\mathcal{P}_N(\omega)} \int_\Xi f(x,\xi)\, dP(\xi).$$

For brevity, when the sample path is fixed, we write

$$\mathcal{R}_N(f)(x) := \sup_{P\in\mathcal{P}_N} \int_\Xi f(x,\xi)\, dP(\xi), \qquad \mathcal{R}_\infty(f)(x) := \int_\Xi f(x,\xi)\, dP^\circ(\xi).$$

3.1 Modulus Continuity and Uniform Convergence

Lemma 3.2 (Inherited modulus continuity of robust objectives)**.** *Under Assumption 2.1, for every N,*

$$|\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x')| \le \rho_f(d_\mathcal{X}(x,x')), \qquad x,x' \in \mathcal{X},$$

and

$$|\mathcal{R}_\infty(f)(x) - \mathcal{R}_\infty(f)(x')| \le \rho_f(d_\mathcal{X}(x,x')), \qquad x,x' \in \mathcal{X}.$$

In particular, $\mathcal{R}_N(f)$ and $\mathcal{R}_\infty(f)$ are equicontinuous on $\mathcal{X}$.

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

$$\int_\Xi f(x,\xi)\, dP(\xi) \le \int_\Xi f(x',\xi)\, dP(\xi) + \rho_f(d_\mathcal{X}(x,x')).$$

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

$$\mathcal{R}_N(f)(x) \le \mathcal{R}_N(f)(x') + \rho_f(d_\mathcal{X}(x,x')).$$

Interchanging x and x' yields

$$\mathcal{R}_N(f)(x') \le \mathcal{R}_N(f)(x) + \rho_f(d_\mathcal{X}(x,x')).$$

Hence

$$|\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x')| \le \rho_f(d_\mathcal{X}(x,x')).$$

The same argument with $\mathcal{P}_N$ replaced by the singleton $\{P^\circ\}$ gives the bound for $\mathcal{R}_\infty(f)$. □

Lemma 3.3 (Pointwise convergence at fixed decisions)**.** *Under Assumptions 2.1, 2.2, 2.3, and 2.4, 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. Fix $x \in \mathcal{X}$ and set $\psi_x(\xi) := f(x,\xi)$. By Assumption 2.1, $\psi_x \in \mathcal{F}$. Let $A_x := A_{\psi_x}$ be the full-measure event obtained from Lemma 3.1.

Fix $\omega \in A_x$. Define

$$g_x(P) := \int_\Xi f(x,\xi)\, dP(\xi).$$

For any sequence $P_N \in \mathcal{P}_N(\omega)$, Lemma 3.1 gives

$$g_x(P_N) \to g_x(P^\circ).$$

For each N, choose $P_N^\varepsilon \in \mathcal{P}_N(\omega)$ such that

$$g_x(P_N^\varepsilon) \geq \sup_{P \in \mathcal{P}_N(\omega)} g_x(P) - \frac{1}{N}.$$

Then

$$\limsup_{N\to\infty} \mathcal{R}_N(f)(x) \leq g_x(P^\circ).$$

For the lower bound, take any sequence $Q_N \in \mathcal{P}_N(\omega)$, which exists by Assumption 2.4. Then

$$\mathcal{R}_N(f)(x) \geq g_x(Q_N),$$

and hence

$$\liminf_{N\to\infty} \mathcal{R}_N(f)(x) \geq g_x(P^\circ).$$

Therefore

$$\mathcal{R}_N(f)(x) \to g_x(P^\circ) = \mathcal{R}_\infty(f)(x).$$

□

Theorem 3.4 (Uniform convergence of empirical robust objectives)**.** *Under Assumptions 2.1, 2.2, 2.3, and 2.4, there exists an event A_f with $\mathbb{P}(A_f) = 1$ such that, on A_f,*

$$\sup_{x\in\mathcal{X}} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \to 0. \tag{2}$$

Proof. Since $\mathcal{X}$ is a compact metric space, it is separable. Fix once and for all a countable dense subset

$$D = \{x_m : m \in \mathbb{N}\} \subset \mathcal{X}.$$

For each $m \in \mathbb{N}$, by Lemma 3.3, 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

$$A_f := \bigcap_{m=1}^{\infty} A_{x_m}.$$

Then $\mathbb{P}(A_f) = 1$. Fix $\omega \in A_f$.

By Lemma 3.2, both $\mathcal{R}_N(f)$ and $\mathcal{R}_\infty(f)$ admit the common modulus ρ_f.

Let $\epsilon > 0$. Since ρ_f is continuous and $\rho_f(0) = 0$, there exists $\delta > 0$ such that

$$\rho_f(\delta) < \frac{\epsilon}{3}.$$

Since D is dense in the compact set $\mathcal{X}$, there exist indices $m_1, \dots, m_M$ such that

$$\mathcal{X} \subset \bigcup_{\ell=1}^{M} B(x_{m_\ell}, \delta).$$

Since $\omega \in A_f$, for each $\ell = 1, \dots, M$,

$$\mathcal{R}_N(f)(x_{m_\ell}) \to \mathcal{R}_\infty(f)(x_{m_\ell}).$$

Hence there exists $N_0 = N_0(\omega, \epsilon)$ such that for all $N \geq N_0$,

$$\max_{1 \leq \ell \leq M} |\mathcal{R}_N(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x_{m_\ell})| < \frac{\epsilon}{3}.$$

Now take arbitrary $x \in \mathcal{X}$. Choose $\ell \in \{1, \dots, M\}$ such that

$$d_{\mathcal{X}}(x, x_{m_\ell}) < \delta.$$

Then for all $N \geq N_0$,

$$\begin{aligned}
|\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| &\leq |\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x_{m_\ell})| \\
&\quad + |\mathcal{R}_N(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x_{m_\ell})| \\
&\quad + |\mathcal{R}_\infty(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x)| \\
&\leq \rho_f(\delta) + \frac{\epsilon}{3} + \rho_f(\delta) < \epsilon.
\end{aligned}$$

Therefore,

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

on A_f. □

3.2 Preliminaries on Set Convergence

To analyze the asymptotic behavior of feasible regions and minimizer sets, we use set convergence in the sense of Painlevé–Kuratowski (PK); see [**?**, Chapter 4].

For any set $C \subseteq \mathcal{X}$ and any point $x \in \mathcal{X}$, define $\operatorname{dist}(x, C) := \inf_{y \in C} d_{\mathcal{X}}(x, y)$. By convention, if $C = \emptyset$, then $\operatorname{dist}(x, C) = \infty$.

Definition 3.5 (PK set convergence)**.** The inner (lower) limit of $\{C_N\}$ is

$$\operatorname*{Liminf}_{N \to \infty} C_N = \left\{ x \in \mathcal{X} : \limsup_{N \to \infty} \operatorname{dist}(x, C_N) = 0 \right\},$$

and the outer (upper) limit is

$$\operatorname*{Limsup}_{N\to\infty} C_N = \left\{x \in \mathcal{X} : \liminf_{N\to\infty} \operatorname{dist}(x, C_N) = 0\right\}.$$

We write $C_N \xrightarrow{\text{PK}} C$ if $\operatorname{Limsup}_{N\to\infty} C_N \subseteq C \subseteq \operatorname{Liminf}_{N\to\infty} C_N$.

Corollary 3.6 (Epigraphical convergence of robust objectives)**.** *Under the assumptions of Theorem 3.4, there exists an event A_f with $\mathbb{P}(A_f) = 1$ such that, on A_f,*

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

Equivalently,

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

where

$$\operatorname{epi}_{\mathcal{X}} h := \{(x, \alpha) \in \mathcal{X} \times \mathbb{R} : \ h(x) \le \alpha\}.$$

Proof. By Theorem 3.4, on A_f,

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

Thus $\mathcal{R}_N(f)$ converges uniformly, hence continuously, to $\mathcal{R}_\infty(f)$ on $\mathcal{X}$. By [**?**, Theorem 7.11], continuous convergence is equivalent to simultaneous epi- and hypo-convergence. 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:

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

□

3.3 Unconstrained Robust Consistency

Define the unconstrained optimal values and minimizer sets by

$$\hat{v}_N := \inf_{x\in\mathcal{X}} \mathcal{R}_N(f)(x), \qquad v^* := \inf_{x\in\mathcal{X}} \mathcal{R}_\infty(f)(x),$$

$$\hat{X}_N := \operatorname*{arg\,min}_{x\in\mathcal{X}} \mathcal{R}_N(f)(x), \qquad X^* := \operatorname*{arg\,min}_{x\in\mathcal{X}} \mathcal{R}_\infty(f)(x).$$

Theorem 3.7 (Unconstrained robust consistency)**.** *Under the assumptions of Theorem 3.4, on the event A_f,*

$$\hat{v}_N \to v^*, \tag{3}$$

and in the Painlevé–Kuratowski sense,

$$\operatorname*{Limsup}_{N\to\infty} \hat{X}_N \subseteq X^*. \tag{4}$$

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

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

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

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

By Theorem 3.4, the right-hand side converges to 0, and therefore $\hat{v}_N \to v^*$.

We next prove the outer inclusion for minimizer sets. By compactness of $\mathcal{X}$ and continuity of $\mathcal{R}_N(f)$ and $\mathcal{R}_\infty(f)$ (from Lemma 3.2), the sets $\hat{X}_N$ and X^* are nonempty. 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 $\mathcal{R}_{N_k}(f)(x_{N_k}) = \hat{v}_{N_k} \le \mathcal{R}_{N_k}(f)(x^*)$. Hence

$$\begin{aligned}
\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_\infty(f)(x^*) &= \big(\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_{N_k}(f)(x_{N_k})\big) + \big(\mathcal{R}_{N_k}(f)(x_{N_k}) - \mathcal{R}_{N_k}(f)(x^*)\big) \\
&\quad + \big(\mathcal{R}_{N_k}(f)(x^*) - \mathcal{R}_\infty(f)(x^*)\big) \\
&\le |\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_{N_k}(f)(x_{N_k})| + |\mathcal{R}_{N_k}(f)(x^*) - \mathcal{R}_\infty(f)(x^*)| \\
&\le 2 \sup_{z\in\mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)|.
\end{aligned}$$

Taking the limit superior gives $\limsup_{k\to\infty} \big(\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_\infty(f)(x^*)\big) \le 0$. By continuity of $\mathcal{R}_\infty(f)$, passing to the limit along $x_{N_k} \to \bar{x}$ yields $\mathcal{R}_\infty(f)(\bar{x}) \le \mathcal{R}_\infty(f)(x^*) = v^*$. 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 $\operatorname{Limsup}_{N\to\infty} \hat{X}_N \subseteq X^*$. ☐

Remark 3.8 (Failure of reverse PK inclusion)**.** Even when the ambiguity sets collapse to a singleton, the reverse inclusion $X^* \subseteq \operatorname{Limsup}_{N\to\infty} \hat{X}_N$ fails in general.

Construction. Let $\Xi = [0,1]$, $\mathcal{X} = [-1,1]$, and $P^\circ = \delta_0$. Assume $\xi_i \equiv 0$ almost surely, so that $P_N^{\text{ref}} = \delta_0$. Take the probability metric d to be the 1-Wasserstein distance W_1. Define $\mathcal{P}_N := \{P \in \mathscr{P}(\Xi) : W_1(P, \delta_0) \le 1/N\}$, so that $\mathcal{P}_N \to \{P^\circ\}$.

Let $f(x,\xi) = \max\{x, 0\} - \xi x$. Then $\mathcal{R}_\infty(x) = \mathbb{E}_{\delta_0}[f(x,\xi)] = \max\{x, 0\}$, hence $X^* = [-1, 0]$.

Computation of $\mathcal{R}_N$. For any $P \in \mathcal{P}_N$, every coupling $\pi \in \Pi(P, \delta_0)$ satisfies $\pi = P \otimes \delta_0$, and hence $W_1(P, \delta_0) = \int_0^1 |\xi|\, P(d\xi) = \int_\Xi \xi \, dP(\xi) \le 1/N$. Therefore, $\mathcal{R}_N(x) = \sup_{P\in\mathcal{P}_N} \big(\max\{x,0\} - x \int_\Xi \xi \, dP(\xi)\big)$. If $x \le 0$, then $-x \ge 0$, and the supremum is attained at $\sup_{0\le m\le 1/N} m = 1/N$, so $\mathcal{R}_N(x) = -x/N$. If $x \ge 0$, then $-x \le 0$, and the supremum is attained at $\inf_{0\le m\le 1/N} m = 0$, so $\mathcal{R}_N(x) = x$. Therefore, $\mathcal{R}_N(x) = (-x/N)$ for $x \le 0$ and $\mathcal{R}_N(x) = x$ for $x \ge 0$.

Conclusion. It follows that $\hat{X}_N = \{0\}$ for all N, hence $\operatorname{Limsup}_{N\to\infty} \hat{X}_N = \{0\} \subsetneq [-1, 0] = X^*$. Moreover, $\sup_{x\in\mathcal{X}} |\mathcal{R}_N(x) - \mathcal{R}_\infty(x)| \le 1/N \to 0$, so all assumptions of Theorem 3.7 are satisfied, yet equality fails.

3.4 Constrained DRO and PK Convergence

We now extend the variational consistency analysis to structurally constrained DRO problems. Let $g_j(x, \xi)$, $j = 1, \ldots, J$, be a finite family of constraint functions. In the constrained analysis, we work on the event

$$A_* := A_f \cap A_g,$$

which has probability one (where A_f is from Theorem 3.4 and A_g will be defined in Lemma 3.10). All subsequent arguments are deterministic on this event. Define the population and empirical robust constraint functionals by

$$\Phi_{\infty,j}(x) := \mathcal{R}_\infty(g_j)(x), \qquad \Phi_{N,j}(x) := \mathcal{R}_N(g_j)(x), \qquad j = 1, \ldots, J.$$

Assumption 3.9 (Constraint regularity and Slater condition)**.** Let $\mathcal{X} \subset \mathbb{R}^d$ be nonempty, compact, and convex. For each $j = 1, \ldots, J$, let $g_j : \mathcal{X} \times \Xi \to \mathbb{R}$ 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

$$|g_j(x, \xi) - g_j(x', \xi)| \le \rho_{g_j}(d_{\mathcal{X}}(x, x')), \qquad \forall x, x' \in \mathcal{X},\ \forall \xi \in \Xi.$$

Furthermore, there exists $x^\circ \in \mathcal{X}$ such that

$$\Phi_{\infty,j}(x^\circ) < 0, \qquad j = 1, \ldots, J,$$

where $\Phi_{\infty,j}(x) := \mathcal{R}_\infty(g_j)(x)$.

Define the population and empirical feasible regions by

$$\Gamma_\infty := \bigcap_{j=1}^{J} \{x \in \mathcal{X} : \Phi_{\infty,j}(x) \le 0\}, \qquad \Gamma_N := \bigcap_{j=1}^{J} \{x \in \mathcal{X} : \Phi_{N,j}(x) \le 0\}. \tag{5}$$

For finite N with $\Gamma_N = \emptyset$, we set $\hat{v}_N^c = +\infty$ and $\hat{X}_N^c = \emptyset$. Under the Slater condition and uniform convergence, this convention is immaterial asymptotically because $\Gamma_N \ne \emptyset$

eventually. Define the corresponding constrained optimal values and minimizer sets:

$$v_\infty^c := \inf_{x\in\Gamma_\infty} \mathcal{R}_\infty(f)(x), \qquad \hat{v}_N^c := \inf_{x\in\Gamma_N} \mathcal{R}_N(f)(x),$$

$$X_c^* := \arg\min_{x\in\Gamma_\infty} \mathcal{R}_\infty(f)(x), \qquad \hat{X}_N^c := \arg\min_{x\in\Gamma_N} \mathcal{R}_N(f)(x).$$

Lemma 3.10 (Uniform convergence of empirical robust constraint functionals)**.** *Under Assumptions 2.2, 2.3, 2.4, and 3.9, there exists an event A_g with $\mathbb{P}(A_g) = 1$ such that, on A_g,*

$$\max_{1\le j\le J} \sup_{x\in\mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \to 0. \tag{6}$$

Proof. By Assumption 3.9, each constraint function g_j satisfies the regularity conditions of Assumption 2.1. Thus, applying Theorem 3.4 to each g_j yields a full-measure event A_{g_j} on which

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

Letting $A_g := \bigcap_{j=1}^J A_{g_j}$, we have $\mathbb{P}(A_g) = 1$ since J is finite. On this intersection, the convergence holds simultaneously for all $j = 1, \dots, J$, which immediately gives the desired limit.

Theorem 3.11 (Painlevé–Kuratowski convergence of empirical feasible sets)**.** *Under Assumption 3.9 and the assumptions of Lemma 3.10, on the event A_*,*

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

Proof. Define $\varepsilon_N := \max_{1\le j\le J} \sup_{x\in\mathcal{X}} \left|\Phi_{N,j}(x) - \Phi_{\infty,j}(x)\right|$. By Lemma 3.10, $\varepsilon_N \to 0$.

Step 1: Outer inclusion ($\mathrm{Limsup}_{N\to\infty} \Gamma_N \subseteq \Gamma_\infty$). 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 $\Phi_{N_k,j}(x_{N_k}) \le 0$ for $j = 1, \dots, J$. Hence, for each j,

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

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

Step 2: Inner inclusion ($\Gamma_\infty \subseteq \mathrm{Liminf}_{N\to\infty} \Gamma_N$). Let $x \in \Gamma_\infty$. We distinguish two cases.

Case 1: $\Phi_{\infty,j}(x) < 0$ for all $j = 1, \dots, J$. Set $\eta := -\max_{1\le j\le J} \Phi_{\infty,j}(x) > 0$. Since $\varepsilon_N \to 0$, there exists N_0 such that for all $N \ge N_0$, $\varepsilon_N < \frac{\eta}{2}$. Then, for every $j = 1, \dots, J$ and every $N \ge N_0$,

$$\Phi_{N,j}(x) \le \Phi_{\infty,j}(x) + \left|\Phi_{N,j}(x) - \Phi_{\infty,j}(x)\right| \le -\eta + \varepsilon_N < -\frac{\eta}{2} < 0.$$

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

Case 2: $\max_{1\le j\le J}\Phi_{\infty,j}(x)=0$. Let x° be the point from Assumption 3.9 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 $\Phi_{\infty,j}$,

$$\Phi_{\infty,j}(x^\lambda)\le(1-\lambda)\Phi_{\infty,j}(x)+\lambda\Phi_{\infty,j}(x^\circ)<0,\qquad j=1,\dots,J.$$

Fix $\lambda\in(0,1)$. By Case 1, there exists N_λ such that $x^\lambda\in\Gamma_N$ for all $N\ge N_\lambda$. Consequently, $\operatorname{dist}(x,\Gamma_N)\le d_{\mathcal{X}}(x,x^\lambda)$ for all $N\ge N_\lambda$. Taking the limit superior as $N\to\infty$ yields $\limsup_{N\to\infty}\operatorname{dist}(x,\Gamma_N)\le d_{\mathcal{X}}(x,x^\lambda)$. Since $x^\lambda\to x$ as $\lambda\downarrow 0$, we obtain $\limsup_{N\to\infty}\operatorname{dist}(x,\Gamma_N)=0$. By the distance characterization of the inner limit, $x\in\operatorname{Liminf}_{N\to\infty}\Gamma_N$.

Combining Step 1 and Step 2 yields $\Gamma_N\xrightarrow{\text{PK}}\Gamma_\infty$. □

Theorem 3.12 (Constrained robust consistency)**.** *Under the assumptions of Theorem 3.11, on the event A_*,*

$$\hat{v}_N^c\to v_\infty^c,\tag{7}$$

and

$$\operatorname*{Limsup}_{N\to\infty}\hat{X}_N^c\subseteq X_c^*.\tag{8}$$

Proof. **Step 1: Upper bound ($\limsup_{N\to\infty}\hat{v}_N^c\le v_\infty^c$).** Let $x^*\in X_c^*\subset\Gamma_\infty$ (which is valid since $X_c^*\neq\emptyset$). By Theorem 3.11, $\Gamma_\infty\subseteq\operatorname{Liminf}_{N\to\infty}\Gamma_N$, 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 $\hat{v}_N^c\le\mathcal{R}_N(f)(x_N)$. Therefore,

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

Step 2: Lower bound ($\liminf_{N\to\infty}\hat{v}_N^c\ge v_\infty^c$). 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 $v_\infty^c\le\mathcal{R}_\infty(f)(\bar{x})$. Moreover,

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

The right-hand side converges to 0. Therefore, $\hat{v}_{N_m}^c\to\mathcal{R}_\infty(f)(\bar{x})$, and thus $v_\infty^c\le\liminf_{N\to\infty}\hat{v}_N^c$.

Combining Step 1 and Step 2 yields $\hat{v}_N^c\to v_\infty^c$.

Step 3: Outer inclusion for constrained minimizers. Let $\{N_k\}$ be an arbitrary subsequence and let $\hat{x}_{N_k}\in\hat{X}_{N_k}^c$. 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 3.11 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 $\hat{v}^c_{N_k} \to v^c_\infty$,

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

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, $\operatorname{Limsup}_{N\to\infty} \hat{X}^c_N \subseteq X^*_c$. □

4 Verification of Assumptions

4.1 Reference distribution consistency

We discuss statistical conditions under which the reference measures $\{P^{\text{ref}}_N\}$ satisfy the consistency requirement in Assumption 2.2.

Throughout this section, let $P^\circ \in \mathscr{P}(\Xi)$ denote the target distribution. We present several representative constructions of P^{ref}_N and establish their convergence.

4.2 Reference consistency via deviation bounds

In the DRO literature, finite-sample guarantees are often established through deviation inequalities for the distance between a reference distribution Q_N and the true distribution P°. A commonly used form is an exponential concentration inequality of the form

$$\mathbb{P}\big(W_1(Q_N, P^\circ) > \epsilon\big) \leq c_1 \exp(-c_2 N \epsilon^m), \qquad \forall \epsilon > 0, \tag{9}$$

for some constants $c_1, c_2, m > 0$.

While such exponential bounds are convenient, they are not necessary for establishing reference consistency. The following result shows that a much weaker summability condition suffices.

Proposition 4.1 (Summable deviation implies reference consistency)**.** *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 $\epsilon > 0$, $\sum_{N=1}^\infty \mathbb{P}\big(\mathbb{D}(P^{\text{ref}}_N, P^\circ) > \epsilon\big) < \infty$. Then $\mathbb{D}(P^{\text{ref}}_N, P^\circ) \to 0$ almost surely, and hence Assumption **??** *holds.*

Proof. Fix $\epsilon > 0$ and define the sequence of error events $E_N(\epsilon) := \{\omega : \mathbb{D}(P^{\text{ref}}_N(\omega), P^\circ) > \epsilon\}$. By assumption, $\sum_{N=1}^\infty \mathbb{P}(E_N(\epsilon)) < \infty$. According to the first Borel–Cantelli lemma see Theorem 4.3 in [**?**], the probability that infinitely many events $E_N(\epsilon)$ occur is zero: $\mathbb{P}\big(\limsup_{N\to\infty} E_N(\epsilon)\big) = 0$, 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 $\mathbb{D}(P^{\text{ref}}_N(\omega), P^\circ) \leq \epsilon$ 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.

We obtain $\int_\Xi \psi \, dP_N^{\text{ref}} \to \int_\Xi \psi \, dP^\circ$ almost surely for every $\psi \in \mathcal{F}$. □

4.3 Examples of admissible reference distributions

These examples illustrate that exponential concentration is only one sufficient condition among many; the essential requirement for reference consistency is the summability of deviation probabilities.

Wasserstein distance. Let $\Xi = \mathbb{R}^d$. For $p > 0$, define $T_p(\mu, \nu) := \inf_{\xi \in H(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, \xi(dx, dy)$, where $H(\mu, \nu)$ denotes the set of couplings of μ and ν. The Wasserstein distance is defined by $W_p(\mu, \nu) = T_p(\mu, \nu)$ if $p \in (0, 1]$, $\quad W_p(\mu, \nu) = T_p(\mu, \nu)^{1/p}$ if $p > 1$. See [?] for details.

Example 4.2 (Empirical reference via Fournier–Guillin inequality)**.** [?] Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ and let μ_N be the empirical measure associated with an i.i.d. sample from μ.

For $q > 0$, $\alpha > 0$, $\gamma > 0$, define $M_q(\mu) := \int_{\mathbb{R}^d} |x|^q \, \mu(dx)$, $\quad E_{\alpha,\gamma}(\mu) := \int_{\mathbb{R}^d} e^{\gamma |x|^\alpha} \, \mu(dx)$.

Assume one of the following conditions:

1. $\exists \alpha > p, \ \exists \gamma > 0, \ E_{\alpha,\gamma}(\mu) < \infty$,
2. or $\exists \alpha \in (0, p), \ \exists \gamma > 0, \ E_{\alpha,\gamma}(\mu) < \infty$,
3. or $\exists q > 2p, \ M_q(\mu) < \infty$.

Then for all $N \geq 1$ and all $x \in (0, \infty)$,

$$\mathbb{P}\big(T_p(\mu_N, \mu) \geq x\big) \leq a(N, x) \mathbf{1}_{\{x \leq 1\}} + b(N, x),$$

where

$$a(N, x) = C \begin{cases} \exp(-cNx^2), & \text{if } p > d/2, \\ \exp(-cN(x/\log(2 + 1/x))^2), & \text{if } p = d/2, \\ \exp(-cNx^{d/p}), & \text{if } p \in [1, d/2), \end{cases}$$

and

$$b(N, x) = C \begin{cases} \exp(-cNx^{\alpha/p}) \mathbf{1}_{\{x > 1\}}, & \text{under (1)}, \\ \exp(-c(Nx)^{(\alpha-\varepsilon)/p}) \mathbf{1}_{\{x \leq 1\}} + \exp(-c(Nx)^{\alpha/p}) \mathbf{1}_{\{x > 1\}}, & \forall \varepsilon \in (0, \alpha) \text{ under (2)}, \\ N(Nx)^{-(q-\varepsilon)/p}, & \forall \varepsilon \in (0, q) \text{ under (3)}. \end{cases}$$

The positive constants C, c depend only on p, d and either on $\alpha, \gamma, E_{\alpha,\gamma}(\mu)$ (under (1)), or on $\alpha, \gamma, E_{\alpha,\gamma}(\mu), \varepsilon$ (under (2)), or on $q, M_q(\mu), \varepsilon$ (under (3)).

We could also treat the critical case $E_{\alpha,\gamma}(\mu) < \infty$ with $\alpha = p$, but the resulting bound is more intricate and less satisfactory for small values of x.

Verification of Proposition 4.1. Define $r_N(\epsilon) := a(N,\epsilon)\mathbf{1}_{\{\epsilon\le 1\}} + b(N,\epsilon), \quad \epsilon > 0$. By Theorem 2 of Fournier and Guillin [?], $\mathbb{P}\big(T_p(\mu_N,\mu) \ge \epsilon\big) \le r_N(\epsilon)$ for every $N \ge 1$ and every $\epsilon > 0$.

It remains to check that $\sum_{N=1}^{\infty} r_N(\epsilon) < \infty$ for each fixed $\epsilon > 0$. The term $a(N,\epsilon)\mathbf{1}_{\{\epsilon\le 1\}}$ is exponentially summable in N. For $b(N,\epsilon)$, under conditions (1) and (2), the bounds are exponential or stretched-exponential in N, hence summable. Under condition (3), choose $\eta \in (0, q-2p)$. Then $N(N\epsilon)^{-(q-\eta)/p} = \epsilon^{-(q-\eta)/p} N^{1-(q-\eta)/p}$, and since $(q-\eta)/p > 2$, the exponent $1-(q-\eta)/p < -1$. Hence this term is also summable.

Therefore, $\sum_{N=1}^{\infty} r_N(\epsilon) < \infty$ for every $\epsilon > 0$. Proposition 4.1 applies with $\mathbb{D} = T_p$, $P_N^{\mathrm{ref}} = \mu_N$, and $P^\circ = \mu$, yielding $T_p(\mu_N,\mu) \to 0$ almost surely.

(ii) Contextual and conditional reference. In contextual DRO, the reference distribution may depend on side information. Fix a covariate value z, and consider the kernel-based conditional estimator introduced by Stute [?]:

$$m_N(z,y) = \frac{\sum_{i=1}^{N} \mathbf{1}_{\{\xi_i \le y\}} K\left(\frac{z-Z_i}{h_N}\right)}{\sum_{i=1}^{N} K\left(\frac{z-Z_i}{h_N}\right)},$$

where K is a kernel function and h_N is a bandwidth parameter.

This defines a probability measure $P_{N,z}^{\mathrm{ref}}$ on Ξ via its distribution function $m_N(z,\cdot)$.

Under standard bandwidth conditions,

$$h_N \to 0, \qquad N h_N^d \to \infty,$$

Stute [?] proves that

$$\sup_y |m_N(z,y) - F(y \mid z)| \to 0 \quad \text{almost surely},$$

where $F(\cdot \mid z)$ denotes the true conditional distribution function.

Consequently, for each fixed z,

$$P_{N,z}^{\mathrm{ref}} \Rightarrow P^*(\cdot \mid z) \quad \text{almost surely.}$$

Thus, the limiting reference distribution is

$$P_z^\circ = P^*(\cdot \mid z),$$

which satisfies the reference consistency condition.

The above examples demonstrate that the reference consistency assumption is satisfied by a broad class of data-driven constructions, including empirical, weighted, conditional, and concentration-based reference measures.

4.4 Sequential weak collapse of ambiguity sets

These discrepancies provide sufficient mechanisms for Assumption 2.3 on suitable test classes. For the following examples, we verify the required conditions.

Example 1: c_V-Wasserstein ambiguity sets. (i) We define the discrepancy $D = W_{c_V}(P,Q) = \inf_{\pi\in\Pi(P,Q)} \int c_V(u,v)d\pi(u,v)$, and the ambiguity set $\mathcal{P}_N = \{P : W_{c_V}(P, P_N^{\mathrm{ref}}) \le \delta_N\}$. Provided $\delta_N \ge 0$, $\mathcal{P}_N \neq \emptyset$ because $W_{c_V}(P_N^{\mathrm{ref}}, P_N^{\mathrm{ref}}) = 0 \le \delta_N$, ensuring $P_N^{\mathrm{ref}} \in \mathcal{P}_N$.

(ii) The admissible test class is specified as $\mathcal{F} = \mathrm{Lip}_{c_V}(\Xi)$, where functions satisfy $|\psi(u) - \psi(v)| \le |\psi|_{c_V} c_V(u,v)$.

(iii) For every $\psi \in \mathcal{F}$ and any sequence $P_N \in \mathcal{P}_N$, the expectation difference is bounded by $|\int \psi\, dP_N - \int \psi\, dP_N^{\mathrm{ref}}| \le |\psi|_{c_V} W_{c_V}(P_N, P_N^{\mathrm{ref}}) \le |\psi|_{c_V}\delta_N$. Since $\delta_N \to 0$, this vanishes, verifying Assumption 2.3.

(iv) The condition $P_N^{\mathrm{ref}} \Rightarrow_{\mathcal{F}} P^\circ$ holds, for instance, when $P_N^{\mathrm{ref}} = \hat{P}_N$ and the target distribution P° admits suitable moments, via Wasserstein convergence of empirical measures.

Example 2: Bounded Lipschitz ambiguity sets. (i) Let $D = \beta(P,Q) = \sup_{\|\psi\|_{BL}\le 1} \left|\int \psi\, dP - \int \psi\right.$ and $\mathcal{P}_N = \{P \in \mathscr{P}(\Xi) : \beta(P, P_N^{\mathrm{ref}}) \le \delta_N\}$. The set is nonempty since $P_N^{\mathrm{ref}} \in \mathcal{P}_N$.

(ii) We specify $\mathcal{F} = BL(\Xi)$.

(iii) For every $\psi \in \mathcal{F}$ and any sequence $P_N \in \mathcal{P}_N$, $|\int \psi dP_N - \int \psi dP_N^{\mathrm{ref}}| \le \|\psi\|_{BL}\beta(P_N, P_N^{\mathrm{ref}}) \le \|\psi\|_{BL}\delta_N \to 0$, verifying Assumption 2.3.

(iv) Varadarajan's theorem gives sufficient conditions for $P_N^{\mathrm{ref}} \Rightarrow_{\mathcal{F}} P^\circ$ when P_N^{ref} is the empirical measure on a Polish space.

Example 3: MMD ambiguity sets. (i) Let $D = \mathrm{MMD}_{\mathcal{H}}(P,Q)$ and $\mathcal{P}_N = \{P : \mathrm{MMD}_{\mathcal{H}}(P, P_N^{\mathrm{ref}}) \le \delta_N\}$. The set is nonempty as $P_N^{\mathrm{ref}} \in \mathcal{P}_N$.

(ii) The test class is $\mathcal{F} = \mathcal{H}$, the reproducing kernel Hilbert space.

(iii) For every $\psi \in \mathcal{F}$ and $P_N \in \mathcal{P}_N$, $|\int \psi dP_N - \int \psi dP_N^{\mathrm{ref}}| \le \|\psi\|_{\mathcal{H}}\mathrm{MMD}_{\mathcal{H}}(P_N, P_N^{\mathrm{ref}}) \le \|\psi\|_{\mathcal{H}}\delta_N \to 0$, verifying Assumption 2.3.

(iv) $P_N^{\mathrm{ref}} \Rightarrow_{\mathcal{F}} P^\circ$ is satisfied under standard RKHS laws of large numbers.

Example 4: ϕ-divergence ambiguity sets. (i) Let $D = D_\phi(P\|Q)$ and $\mathcal{P}_N = \{P : D_\phi(P\|P_N^{\mathrm{ref}}) \le \delta_N\}$. The set is nonempty because $D_\phi(P_N^{\mathrm{ref}}\|P_N^{\mathrm{ref}}) = 0 \le \delta_N$.

(ii) We specify $\mathcal{F}$: for bounded ψ, collapse is governed by the total variation distance via Pinsker-type bounds. For unbounded ψ, $\mathcal{F}$ requires specific Orlicz integrability or uniform integrability conditions corresponding to the dual structure of ϕ.

(iii) Under these integrability conditions on $\psi \in \mathcal{F}$, $D_\phi(P_N \| P_N^{\text{ref}}) \leq \delta_N \to 0$ implies $\int \psi \, dP_N - \int \psi \, dP_N^{\text{ref}} \to 0$, satisfying Assumption 2.3.

(iv) Sufficient conditions for $P_N^{\text{ref}} \Rightarrow_{\mathcal{F}} P^\circ$ rely on corresponding tail bounds and moment assumptions of the target distribution.

Example 5: Mixture ambiguity sets. (i) Suppose $P_N = \sum_{i=1}^k \alpha_{N,i} P_{N,i}$, where $\alpha_{N,i}$ are weights.

(ii) The class $\mathcal{F}$ is given by the discrepancy of the underlying components.

(iii) If for every $\psi \in \mathcal{F}$, $\sum_{i=1}^k \alpha_{N,i} \left| \int \psi \, dP_{N,i} - \int \psi \, dP_N^{\text{ref}} \right| \to 0$, then $\left| \int \psi \, dP_N - \int \psi \, dP_N^{\text{ref}} \right| \to 0$. Thus mixture ambiguity sets satisfy collapse whenever the weighted discrepancy contributions vanish.

(iv) $P_N^{\text{ref}} \Rightarrow_{\mathcal{F}} P^\circ$ follows standard consistency arguments for the underlying reference.

5 Example / Extension: Mixed Aggregation Ambiguity Sets

5.1 Abstract Model

We construct a single-layer mixed ambiguity set by aggregating componentwise ambiguity sets through uncertain mixing weights. Let K be the number of components. We define the mixed aggregation ambiguity set as:

$$\mathcal{P}_N^{MA} := \left\{ \sum_{k=1}^K w_{N,k} P_{N,k} : P_{N,k} \in \mathcal{P}_{N,k}, \ \boldsymbol{w}_N \in \mathcal{W}_N \right\}, \tag{10}$$

where $\mathcal{P}_{N,k} \subset \mathscr{P}(\Xi)$ represents the component ambiguity sets for each $k = 1, \ldots, K$, and $\mathcal{W}_N \subset \{\boldsymbol{w} \in \mathbb{R}_+^K : \sum_{k=1}^K w_k = 1\}$ represents the weight uncertainty set.

This construction generalizes mixture-type ambiguity sets and includes multi-ambiguity DRO as a special case. By formulating $\mathcal{P}_N^{MA}$ directly in the space of probability measures, we seamlessly embed it into the proposed collapse framework.

5.2 Two-Layer Equivalence

The single-layer robust evaluation under $\mathcal{P}_N^{MA}$ naturally admits a two-layer representation. Consider the inner worst-case expectation for a given decision $x \in \mathcal{X}$:

$$\inf_{x \in \mathcal{X}} \sup_{P \in \mathcal{P}_N^{MA}} \int_\Xi f(x, \xi) \, dP(\xi). \tag{11}$$

By the definition of $\mathcal{P}_N^{MA}$ and the linearity of expectation over mixture distributions, the supremum can be perfectly decoupled. Since the choice of $P_{N,k}$ within each $\mathcal{P}_{N,k}$ is independent across components, we obtain the exact equivalence:

$$\inf_{x\in\mathcal{X}} \sup_{\boldsymbol{w}_N\in\mathcal{W}_N} \sum_{k=1}^{K} w_{N,k} \sup_{P_{N,k}\in\mathcal{P}_{N,k}} \mathbb{E}_{P_{N,k}}[f(x,\xi)]. \tag{12}$$

This two-layer structure serves as the core formulation for Multi-Ambiguity DRO (MADRO), acting here as a direct consequence of the mixed ambiguity definition.

5.3 Collapse Stability

The following theorem is the main result of this section. It establishes that if a subset of the component ambiguity sets satisfies the collapse assumption, and the weights assigned to the remaining non-collapsing components vanish uniformly, then the entire mixed ambiguity set $\mathcal{P}_N^{MA}$ preserves the collapse property. Let Q_N denote the global reference measure.

Theorem 5.1 (Mixed collapse stability)**.** *Let $I_c \subset \{1,\ldots,K\}$ be an index set of collapsing components. Suppose the following conditions hold:*

(i) ***Component collapse:*** *For every $k \in I_c$, and any sequence $P_{N,k} \in \mathcal{P}_{N,k}$:*

$$\left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right| \to 0, \quad \forall \phi \in \mathcal{T}.$$

(ii) ***Vanishing non-collapse weights:*** *The uncertainty weights associated with the non-collapsing components vanish uniformly:*

$$\sup_{\boldsymbol{w}_N\in\mathcal{W}_N} \sum_{k\notin I_c} w_{N,k} \to 0.$$

(iii) ***Bounded test class:*** *The test functions are globally bounded, i.e., $\phi \in \mathcal{T} \subset B_b(\Xi)$. Let $|\phi|_\infty := \sup_{\xi\in\Xi} |\phi(\xi)|$.*

Then, the mixed ambiguity set $\mathcal{P}_N^{MA}$ satisfies the sequential collapse assumption.

Proof. Take any sequence $P_N \in \mathcal{P}_N^{MA}$. By definition, $P_N = \sum_{k=1}^K w_{N,k} P_{N,k}$ for some $\boldsymbol{w}_N \in \mathcal{W}_N$ and $P_{N,k} \in \mathcal{P}_{N,k}$. Since $\sum_{k=1}^K w_{N,k} = 1$, we can decompose the difference as:

$$\begin{aligned}
|\mathbb{E}_{P_N}[\phi] - \mathbb{E}_{Q_N}[\phi]| &= \left|\sum_{k=1}^{K} w_{N,k}\left(\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right)\right| \\
&\le \sum_{k\in I_c} w_{N,k}\left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right| + \sum_{k\notin I_c} w_{N,k}\left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right|.
\end{aligned}$$

For the collapsing part ($k \in I_c$), we have $\left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right| = o(1)$. Since $w_{N,k} \le 1$, this sum vanishes. For the non-collapsing part ($k \notin I_c$), since $\phi \in B_b(\Xi)$, the difference in expectations is bounded by $2|\phi|_\infty$. Thus, this part is bounded by:

$$2|\phi|_\infty \sum_{k\notin I_c} w_{N,k} \le 2|\phi|_\infty \sup_{\boldsymbol{w}_N\in\mathcal{W}_N} \sum_{k\notin I_c} w_{N,k}.$$

By the uniform vanishing weights condition, this term also converges to 0. Therefore, $|\mathbb{E}_{P_N}[\phi] - \mathbb{E}_{Q_N}[\phi]| \to 0$, concluding the proof. □

5.4 Extension to Non-compact Spaces

If the uncertainty space Ξ is unbounded (non-compact), test functions in $\mathcal{T}$ may not be globally bounded, breaking condition (iii) in Theorem 5.1. To close the theoretical loop, we impose an envelope control assumption to regulate the non-collapsing components.

Assumption 5.2 (Envelope control)**.** There exists an envelope function $V \ge 1$ such that for any $\phi \in \mathcal{T}$, there is a constant $C_\phi > 0$ satisfying $|\phi(\xi)| \le C_\phi V(\xi)$ for all $\xi \in \Xi$. Furthermore, the expected value of V is uniformly bounded across both the component ambiguity sets and the reference measures:

$$\sup_N \sup_{P\in\mathcal{P}_{N,k}} \int_\Xi V\, dP < \infty, \quad \text{and} \quad \sup_N \int_\Xi V\, dQ_N < \infty, \quad \forall k \notin I_c.$$

Under Assumption 5.2, there exists a global constant $C > 0$ such that for any $P_{N,k} \in \mathcal{P}_{N,k}$, we have

$$\left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right| \le \int_\Xi |\phi|\, dP_{N,k} + \int_\Xi |\phi|\, dQ_N \le C_\phi \cdot C.$$

Consequently, the remainder term in the proof of Theorem 5.1 is controlled by:

$$\sum_{k\notin I_c} w_{N,k} \left|\mathbb{E}_{P_{N,k}}[\phi] - \mathbb{E}_{Q_N}[\phi]\right| \le C_\phi C \sup_{\boldsymbol{w}_N\in\mathcal{W}_N} \sum_{k\notin I_c} w_{N,k} \to 0.$$

This technical patch ensures the collapse stability holds strictly even in non-compact spaces with unbounded test functions.

5.5 MADRO as a Special Case

We now demonstrate how the original Multi-Ambiguity DRO (MADRO) model can be recovered precisely as a special case within this framework.

Corollary 5.3 (MADRO as a special case)**.** *Consider the Wasserstein component ambiguity sets and the ℓ_1-norm weight ambiguity set:*

$$\mathcal{P}_{N,k} = \left\{P \in \mathscr{P}(\Xi_k) : W_1(P, \hat{P}_{N,k}) \le \delta_{N,k}\right\}, \quad \forall k = 1, \dots, K,$$

$$\mathcal{W}_N = \left\{\boldsymbol{w} \in \mathbb{R}_+^K : \|\boldsymbol{w} - \hat{\boldsymbol{w}}_N\|_1 \le \delta_w(N)\right\}.$$

Suppose that as $N \to \infty$, the empirical centers converge to the true underlying distribution P^*. If the radii satisfy:

- $\delta_{N,k} \to 0 \implies$ *Component collapse* ($\mathcal{P}_{N,k} \to \{P_k^*\}$),
- $\delta_w(N) \to 0 \implies$ *Weight collapse* ($\mathcal{W}_N \to \{\boldsymbol{w}^*\}$),

then the mixed ambiguity set collapses to the target measure:

$$\mathcal{P}_N^{MA} \to \{P^*\}.$$

This naturally recovers the statistical consistency of the MADRO model without requiring separate proofs for the aggregated distance metric.

5.6 Interpretation and Applications

The mixed aggregation ambiguity structure provides a powerful tool for modeling complex stochastic environments. It can be interpreted in several ways:

- **Mixture Distributions:** Capturing multi-modal data distributions where each component $\mathcal{P}_{N,k}$ models a local peak, and $w_{N,k}$ models the proportion.
- **Heterogeneous Populations:** Addressing fairness or heterogeneity where different sub-populations have distinct local distribution uncertainty ($\mathcal{P}_{N,k}$) and relative sizes ($w_{N,k}$).
- **Regime-Switching:** Representing different temporal or spatial states (e.g., market regimes), where $w_k(z)$ could act as the conditional mixing weights based on feature z.

In these applications, the component sets $\mathcal{P}_{N,k}$ capture the *local distributional uncertainty*, while the weight sets $\mathcal{W}_N$ govern the *macroscopic aggregation logic*. The proposed framework thus allows the model to enforce robust self-adaptive weighting across distinct distributions while fundamentally preserving the heterogeneous local structures.