On Approximations of Data-Driven Chance Constrained Programs over Wasserstein Balls

Zhi Chen
Department of Management Sciences, College of Business, City University of Hong Kong, Kowloon Tong, Hong Kong, zhi.chen@cityu.edu.hk

Daniel Kuhn
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, daniel.kuhn@epfl.ch

Wolfram Wiesemann
Imperial College Business School, Imperial College London, London, United Kingdom, ww@imperial.ac.uk

Distributionally robust chance constrained programs minimize a deterministic cost function subject to the satisfaction of one or more safety conditions with high probability, given that the probability distribution of the uncertain problem parameters affecting the safety condition(s) is only known to belong to some ambiguity set. We study two popular approximation schemes for distributionally robust chance constrained programs over Wasserstein balls, where the ambiguity set contains all probability distributions within a certain Wasserstein distance to a reference distribution. The first approximation replaces the chance constraint with a bound on the conditional value-at-risk, whereas the second approximation decouples different safety conditions via Bonferroni's inequality. We show that the conditional value-at-risk approximation can be characterized as a tight convex approximation, which complements earlier findings on classical (non-robust) chance constraints, and we offer a novel interpretation in terms of transportation savings. We also show that the two approximation schemes can both perform arbitrarily poorly in data-driven settings, and that they are generally incomparable with each other—in contrast to earlier results for moment-based ambiguity sets.

Key words: Distributionally robust optimization; ambiguous chance constraints; Wasserstein distance; conditional value-at-risk; Bonferroni's inequality.

History: May 31, 2022

---

1. Introduction

In this paper we study distributionally robust chance constrained programs of the form

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}\in\mathcal{X}} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \mathcal{S}(\boldsymbol{x})] \geq 1-\varepsilon \quad \forall \mathbb{P} \in \mathcal{F}(\theta). \end{array} \tag{1}$$

The goal is to find a decision $\boldsymbol{x}$ from within a compact polyhedron $\mathcal{X} \subseteq \mathbb{R}^L$ that minimizes a linear cost function $\boldsymbol{c}^\top \boldsymbol{x}$ and ensures that the exogenous random vector $\tilde{\boldsymbol{\xi}}$ falls within a decision-dependent safety set $\mathcal{S}(\boldsymbol{x}) \subseteq \mathbb{R}^K$ with high probability $1-\varepsilon$ under every distribution $\mathbb{P}$ that resides in the Wasserstein ball $\mathcal{F}(\theta)$ of radius $\theta \geq 0$:

$$\mathcal{F}(\theta) = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^K) \mid d_{\mathrm{W}}(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta\}. \tag{2}$$

Here, $\hat{\mathbb{P}} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\hat{\boldsymbol{\xi}}_i}$ is the empirical distribution over the historical samples $\{\hat{\boldsymbol{\xi}}_i\}_{i\in[N]}$ of $\tilde{\boldsymbol{\xi}}$, and the (type-1) Wasserstein distance $d_{\mathrm{W}}(\mathbb{P}_1, \mathbb{P}_2)$ between two distributions $\tilde{\boldsymbol{\xi}}_1 \sim \mathbb{P}_1$ and $\tilde{\boldsymbol{\xi}}_2 \sim \mathbb{P}_2$ on $\mathbb{R}^K$, equipped with a general norm $\|\cdot\|$, is defined as

$$d_{\mathrm{W}}(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{P}\in\mathcal{P}(\mathbb{P}_1,\mathbb{P}_2)} \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{\xi}}_1 - \tilde{\boldsymbol{\xi}}_2\|],$$

where $\mathcal{P}(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all joint distributions on $\mathbb{R}^K \times \mathbb{R}^K$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$. Problem (1) generalizes both individual chance constrained programs, where $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{a}(\boldsymbol{\xi})^\top \boldsymbol{x} < b(\boldsymbol{\xi})\}$ for affine functions $\boldsymbol{a}(\cdot): \mathbb{R}^K \to \mathbb{R}^L$ and $b(\cdot): \mathbb{R}^K \to \mathbb{R}$, and joint chance constrained programs with right-hand side uncertainty, where $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{A}\boldsymbol{x} < \boldsymbol{b}(\boldsymbol{\xi})\}$ for $\boldsymbol{A} \in \mathbb{R}^{M\times L}$ and an affine function $\boldsymbol{b}: \mathbb{R}^K \to \mathbb{R}^M$.

It has been shown that a fixed decision $\boldsymbol{x}$ satisfies the ambiguous chance constraint in (1) if and only if the partial sum of the εN smallest transportation distances to the unsafe set $\bar{\mathcal{S}}(\boldsymbol{x}) = \mathbb{R}^K \setminus \mathcal{S}(\boldsymbol{x})$, multiplied by the mass $1/N$ of a training sample, exceeds θ.

Theorem 1 (Chen et al. (2022)). *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$, the ambiguous chance constraint in (1) is satisfied if and only if*

$$\frac{1}{N}\sum_{i=1}^{\varepsilon N} \mathbf{dist}(\hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})}, \bar{\mathcal{S}}(\boldsymbol{x})) \geq \theta.$$

Here, $\boldsymbol{\pi}(\boldsymbol{x}) : [N] \to [N]$ is a decision-dependent permutation that orders the training samples $\{\hat{\boldsymbol{\xi}}_i\}_{i\in[N]}$ in order of non-decreasing distance to the unsafe set $\bar{\mathcal{S}}(\boldsymbol{x})$, and the distance with respect to a norm $\|\cdot\|$ is defined as $\mathbf{dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x})) = \min\{\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_i\| \mid \boldsymbol{\xi} \in \bar{\mathcal{S}}(\boldsymbol{x})\}$.

(The sum in Theorem 1 is defined even if $\varepsilon N \notin \mathbb{N}$; please refer to the notation at the end of this section.) Theorem 1 allows us to reformulate individual and joint chance constrained programs as deterministic mixed-integer conic programs (Chen et al. 2022, Xie 2021).

Proposition 1 (Chen et al. (2022), Xie (2021)). *Assume that $\boldsymbol{A}^\top \boldsymbol{x} \neq \boldsymbol{b}$ for all $\boldsymbol{x} \in \mathcal{X}$. For the safety set $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid (\boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{a})^\top \boldsymbol{x} < \boldsymbol{b}^\top \boldsymbol{\xi} + b\}$, problem (1) is equivalent to the mixed-integer conic program*

$$\begin{aligned}
Z^\star_{\mathrm{ICC}} = \min_{\boldsymbol{q},\boldsymbol{s},t,\boldsymbol{x}} \;& \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \;& \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\
& (\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x} + \mathrm{M} q_i \geq t - s_i \quad \forall i \in [N] \\
& \mathrm{M}(1 - q_i) \geq t - s_i \quad \forall i \in [N] \\
& \boldsymbol{q} \in \{0,1\}^N, \;\; \boldsymbol{s} \geq \mathbf{0}, \;\; \boldsymbol{x} \in \mathcal{X},
\end{aligned} \tag{3}$$

where M *is a suitably large (but finite) positive constant.*

The condition that $\boldsymbol{A}^\top \boldsymbol{x} \neq \boldsymbol{b}$ for all $\boldsymbol{x} \in \mathcal{X}$ in Proposition 1 is non-restrictive. In fact, the weaker condition that $\boldsymbol{A}^\top \boldsymbol{x}^\star \neq \boldsymbol{b}$ for any *optimal* solution $\boldsymbol{x}^\star \in \mathcal{X}$ is sufficient, and if an optimal solution $\boldsymbol{x}^\star$ satisfies $\boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{b}$, then an alternative optimal solution $\boldsymbol{x}'$ satisfying $\boldsymbol{A}^\top \boldsymbol{x}' \neq \boldsymbol{b}$ can be identified from the solution of auxiliary optimization problems. We refer to Chen et al. (2022) for the details.

Proposition 2 (Chen et al. (2022), Xie (2021)). *For the safety set* $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{a}_m^\top \boldsymbol{x} < \boldsymbol{b}_m^\top \boldsymbol{\xi} + b_m \ \forall m \in [M]\}$, *where* $\boldsymbol{b}_m \neq \boldsymbol{0}$ *for all* $m \in [M]$, *the chance constrained program* (1) *is equivalent to the mixed-integer conic program*

$$\begin{aligned} Z^\star_{\text{JCC}} = \min_{\boldsymbol{q},\boldsymbol{s},t,\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \\ & \frac{\boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i + b_m - \boldsymbol{a}_m^\top \boldsymbol{x}}{\|\boldsymbol{b}_m\|_*} + \mathrm{M} q_i \geq t - s_i \quad \forall m \in [M], \ i \in [N] \\ & \mathrm{M}(1 - q_i) \geq t - s_i \quad \forall i \in [N] \\ & \boldsymbol{q} \in \{0,1\}^N, \ \boldsymbol{s} \geq \boldsymbol{0}, \ \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{4}$$

where M *is a suitably large (but finite) positive constant.*

The assumption that $\boldsymbol{b}_m \neq \boldsymbol{0}$ for all $m \in [M]$ in Proposition 2 is non-restrictive since any safety condition with $\boldsymbol{b}_m = \boldsymbol{0}$ is deterministic and can thus be absorbed in the definition of the set $\mathcal{X}$.

In recent years, the mixed-integer conic programming reformulations developed in Propositions 1 and 2 have been strengthened so as to scale better to larger problem sizes (Ho-Nguyen et al. 2020, 2021, Ji and Lejeune 2021, Shen and Jiang 2022). Nevertheless, similar to classical chance constraints as well as distributionally robust chance constraints over moment ambiguity sets, exact reformulations of the distributionally robust chance constrained program (1) quickly become computationally prohibitive for large problems. As a result, there has been significant interest in safe (*i.e.*, conservative) tractable approximations to problem (1) that scale gracefully with problem size.

Distributionally robust chance constrained programs are most commonly approximated by the Bonferroni approximation or the worst-case conditional value-at-risk (CVaR) approximation. The quality of the Bonferroni approximation crucially depends on the choice of the associated Bonferroni weights. While Xie et al. (2019) show that these Bonferroni weights can be optimized efficiently under specific conditions, Chen et al. (2010) show that the quality of the Bonferroni approximation can be poor even if the Bonferroni weights are chosen optimally. Chen et al. (2010) also show that the worst-case CVaR approximation can outperform the Bonferroni approximation with optimally chosen Bonferroni weights for Chebyshev (*i.e.*, second-order moment) ambiguity sets, provided that certain scaling factors in the worst-case CVaR approximation are selected judiciously. Zymler et al.

(2013) show that the worst-case CVaR approximation is indeed exact for distributionally robust chance constrained programs over Chebyshev ambiguity sets if the scaling factors are selected optimally. This result has been extended to non-linear safety conditions by Yang and Xu (2016). Selecting the scaling factors optimally, however, amounts to solving a non-convex optimization problem. For further information, we refer the reader to the surveys by Ben-Tal and Nemirovski (2001), Nemirovski (2012) and Hanasusanto et al. (2015).

This paper complements the literature with the following two contributions.

1. We show that the CVaR offers a tight convex approximation to certain disjunctive constraints underlying the reformulations of Propositions 1 and 2. This provides a theoretical justification for the popularity of this approximation scheme.
2. We also show that the CVaR approximation admits an intuitive interpretation: it accounts for transportation *savings* alongside the transportation costs in the computation of the Wasserstein distance. This complements existing interpretations of chance constraints over Wasserstein balls based on the CVaR (see, *e.g.*, Ho-Nguyen and Wright 2022, Xie 2021).
3. We show that both the CVaR and Bonferroni approximations may deliver solutions that are severely inferior to the optimal solution of problem (1) in data-driven settings. In addition, these two approximation schemes are generally incomparable with each other. This contrasts earlier results for moment ambiguity sets, where the CVaR approximation is known to outperform the Bonferroni approximation under optimally chosen scaling factors.

Notation. Boldface uppercase (resp., lowercase) letters denote matrices (resp., vectors). Special vectors of an appropriate dimension include $\mathbf{0}$ and $\mathbf{e}$, which represent the zero vector and the vector of all ones, respectively. We let $[N] = \{1, 2, \ldots, N\}$ denote the set of positive integers up to N, and $\|\cdot\|_*$ represents the dual norm of a general norm $\|\cdot\|$. Given a possibly fractional real number $\ell \in [0, N]$, the partial sum of the first ℓ values in $\{k_i\}_{i\in[N]}$ is defined as $\sum_{i=1}^{\ell} k_i = \sum_{i=1}^{\lfloor \ell \rfloor} k_i + (\ell - \lfloor \ell \rfloor) k_{\lfloor \ell \rfloor + 1}$. For $x \in \mathbb{R}$, we define $(x)^+ = \max\{x, 0\}$. Finally, we denote random vectors by tilde signs (*e.g.*, $\tilde{\boldsymbol{\xi}}$) and their realizations by the same symbols without tildes (*e.g.*, $\boldsymbol{\xi}$).

2. CVaR Approximation

We propose a systematic approach to constructing safe convex approximations for problem (1) under individual and joint chance constraints in Sections 2.1 and 2.2, respectively. We show that the CVaR approximation to the respective chance constraint represents the best among a low-parametric class of approximations, and we elucidate how the CVaR approximation can be interpreted as assigning transportation *savings* to certain atoms of the empirical distribution.

2.1. Individual Chance Constraints

Consider an instance of problem (1) with an individual chance constraint corresponding to the safety set $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid (\boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{a})^\top \boldsymbol{x} < \boldsymbol{b}^\top \boldsymbol{\xi} + b\}$. As discussed in Section 1, we can assume that $\boldsymbol{A}^\top \boldsymbol{x} \neq \boldsymbol{b}$ for all $\boldsymbol{x} \in \mathcal{X}$. By Proposition 1, the distributionally robust chance constrained program (1) is thus equivalent to the deterministic optimization problem

$$Z^\star_{\text{ICC}} = \min_{(\boldsymbol{x},\boldsymbol{s},t) \in \mathcal{C}_{\text{ICC}}} \boldsymbol{c}^\top \boldsymbol{x}, \tag{5}$$

whose feasible region is given by

$$\mathcal{C}_{\text{ICC}} = \left\{ (\boldsymbol{x},\boldsymbol{s},t) \in \mathcal{X} \times \mathbb{R}^N_+ \times \mathbb{R} \;\middle|\; \begin{array}{l} \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x})^+ \geq t - s_i \quad \forall i \in [N] \end{array} \right\}.$$

Since $\mathcal{C}_{\text{ICC}}$ is non-convex, it is natural to replace it with tractable conservative (inner) approximations. We next show that any convex inner approximation of $\mathcal{C}_{\text{ICC}}$ is dominated, in the sense of set inclusion, by a convex set of the form

$$\mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa}) = \left\{ (\boldsymbol{x},\boldsymbol{s},t) \in \mathcal{X} \times \mathbb{R}^N_+ \times \mathbb{R} \;\middle|\; \begin{array}{l} \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ \kappa_i((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x}) \geq t - s_i \quad \forall i \in [N] \end{array} \right\}$$

parameterized by a vector of slope parameters $\boldsymbol{\kappa} \in [0,1]^N$.

Proposition 3. *For any convex set $\mathcal{W} \subseteq \mathcal{C}_{\text{ICC}}$, there exists $\boldsymbol{\kappa} \in [0,1]^N$ with $\mathcal{W} \subseteq \mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa}) \subseteq \mathcal{C}_{\text{ICC}}$.*

Proposition 3 implies that amongst all convex conservative approximations to problem (5) it is sufficient to focus on those that are induced by a feasible set of the form $\mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa})$ for some $\boldsymbol{\kappa} \in [0,1]^N$. Thus, it is sufficient to focus on the family of approximate problems of the form

$$Z^\star_{\text{ICC}}(\boldsymbol{\kappa}) = \min_{(\boldsymbol{x},\boldsymbol{s},t) \in \mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa})} \boldsymbol{c}^\top \boldsymbol{x} \tag{6}$$

parameterized by $\boldsymbol{\kappa} \in [0,1]^N$. The following proposition asserts that the best approximation within this family is exact.

Proposition 4. *We have $Z^\star_{\text{ICC}} = \min_{\boldsymbol{\kappa} \in [0,1]^N} Z^\star_{\text{ICC}}(\boldsymbol{\kappa})$.*

Proposition 4 implies that the family (6) of tractable upper bounding problems contains an instance $\boldsymbol{\kappa}^\star \in \arg\min_{\boldsymbol{\kappa} \in [0,1]^N} Z^\star_{\text{ICC}}(\boldsymbol{\kappa})$ that recovers an optimal solution of the ambiguous chance constrained program (5), which is known to be NP-hard (Xie and Ahmed 2020, Theorem 12). We may thus conclude that computing $\boldsymbol{\kappa}^\star$ is also NP-hard. The complexity of computing the best upper bound within the family (6) can be reduced by restricting attention to uniform slope parameters of the form $\boldsymbol{\kappa} = \kappa \mathbf{e}$ for some $\kappa \in [0,1]$. Within this subset, the choice $\boldsymbol{\kappa} = \mathbf{e}$ turns out to be optimal.

PROPOSITION 5. *We have* $\min_{\kappa \in [0,1]} Z^\star_{\mathrm{ICC}}(\kappa \mathbf{e}) = Z^\star_{\mathrm{ICC}}(\mathbf{e})$.

Next, we demonstrate that the approximate problem (6) corresponding to $\boldsymbol{\kappa} = \mathbf{e}$ can also be obtained by approximating the worst-case chance constraint in (1) with a worst-case CVaR constraint. To see this, note first that

$$\begin{aligned} \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \mathcal{S}(\boldsymbol{x})] \geq 1 - \varepsilon \quad &\Longleftrightarrow \quad \mathbb{P}[(\boldsymbol{A}\tilde{\boldsymbol{\xi}} + \boldsymbol{a})^\top \boldsymbol{x} \geq \boldsymbol{b}^\top \tilde{\boldsymbol{\xi}} + b] \leq \varepsilon \\ &\Longleftrightarrow \quad \mathbb{P}\text{-VaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) \leq 0 \\ &\Longleftarrow \quad \mathbb{P}\text{-CVaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) \leq 0 \end{aligned}$$

for any $\mathbb{P} \in \mathcal{F}(\theta)$, where the ε-value-at-risk (VaR) and the ε-CVaR of a measurable loss function $\ell(\boldsymbol{\xi})$ are defined as $\mathbb{P}\text{-VaR}_\varepsilon(\ell(\boldsymbol{\xi})) = \inf\{\gamma \in \mathbb{R} \mid \mathbb{P}[\gamma \leq \ell(\tilde{\boldsymbol{\xi}})] \leq \varepsilon\}$ and $\mathbb{P}\text{-CVaR}_\varepsilon(\ell(\tilde{\boldsymbol{\xi}})) = \inf\{\tau + \mathbb{E}_\mathbb{P}[(\ell(\tilde{\boldsymbol{\xi}}) - \tau)^+]/\varepsilon \mid \tau \in \mathbb{R}\}$, respectively. The first equivalence above follows from the definition of the safety set, the second equivalence holds due to the definition of the VaR, and the last implication exploits the fact that the CVaR upper bounds the VaR. Thus, the worst-case CVaR constrained program

$$Z^\star_{\mathrm{CVaR}} = \begin{cases} \min\limits_{\boldsymbol{x} \in \mathcal{X}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \ \sup\limits_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) \leq 0 \end{cases} \tag{7}$$

constitutes a conservative approximation for the worst-case chance constrained program (1), that is, $Z^\star_{\mathrm{ICC}} \leq Z^\star_{\mathrm{CVaR}}$. We are now ready to prove that $Z^\star_{\mathrm{CVaR}} = Z^\star_{\mathrm{ICC}}(\mathbf{e})$.

PROPOSITION 6. *We have* $Z^\star_{\mathrm{CVaR}} = Z^\star_{\mathrm{ICC}}(\mathbf{e})$.

REMARK 1. Using similar arguments as in Proposition 6, one can show that problem (6) with $\boldsymbol{\kappa} = \kappa \mathbf{e}$ for any $\kappa \in (0,1]$ is equivalent to a worst-case CVaR constrained program of the form (7), where the Wasserstein radius θ is inflated to θ/κ. This observation reconfirms that $\kappa = 1$ is the least conservative choice amongst all uniform slope parameters in (6); see Proposition 5.

The intimate links between the worst-case CVaR approximation (7) and the worst-case chance constrained program (1) can also be studied through the lens of Theorem 1. To this end, recall that the ambiguous chance constraint in problem (1) is equivalent to the deterministic constraint

$$\frac{1}{N} \sum_{i=1}^{\varepsilon N} \mathbf{dist}(\hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})}, \bar{\mathcal{S}}(\boldsymbol{x})) \geq \theta.$$

We define the *signed distance* between a point $\boldsymbol{\xi}$ and a closed set $\mathcal{C}$ as $\mathbf{sgn\text{-}dist}(\boldsymbol{\xi}, \mathcal{C}) = \mathbf{dist}(\boldsymbol{\xi}, \mathcal{C})$ if $\boldsymbol{\xi} \notin \mathcal{C}$ and $\mathbf{sgn\text{-}dist}(\boldsymbol{\xi}, \mathcal{C}) = -\mathbf{dist}(\boldsymbol{\xi}, \mathrm{cl}(\bar{\mathcal{C}}))$ otherwise. Here, $\mathrm{cl}(\bar{\mathcal{C}})$ denotes the closure of the open set $\bar{\mathcal{C}} = \mathbb{R}^K \setminus \mathcal{C}$. We then obtain the following result.

PROPOSITION 7. *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$, we have*

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) \leq 0 \iff \frac{1}{N} \sum_{i=1}^{\varepsilon N} \textbf{sgn-dist}(\hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})}, \bar{\mathcal{S}}(\boldsymbol{x})) \geq \theta,$$

where $\boldsymbol{\pi}(\boldsymbol{x})$ permutes the data points $\hat{\boldsymbol{\xi}}_i$ into ascending order of their signed distances to $\bar{\mathcal{S}}(\boldsymbol{x})$.

Theorem 1 and Proposition 7 show that both the ambiguous chance constraint in problem (1) and its worst-case CVaR approximation (7) impose lower bounds on the costs of moving a fraction ε of the training samples to the unsafe set. Moreover, since $\textbf{sgn-dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x})) \leq \textbf{dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x}))$ by construction, the worst-case CVaR constraint conservatively approximates the ambiguous chance constraint. In fact, we have $\textbf{sgn-dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x})) = \textbf{dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x}))$ for safe scenarios $\hat{\boldsymbol{\xi}}_i \in \mathcal{S}(\boldsymbol{x})$, whereas $\textbf{sgn-dist}(\hat{\boldsymbol{\xi}}_j, \bar{\mathcal{S}}(\boldsymbol{x})) < 0$ even though $\textbf{dist}(\hat{\boldsymbol{\xi}}_j, \bar{\mathcal{S}}(\boldsymbol{x})) = 0$ for (strictly) unsafe scenarios $\hat{\boldsymbol{\xi}}_j \in \text{int}(\bar{\mathcal{S}}(\boldsymbol{x}))$. In other words, the worst-case CVaR approximation (7) assigns *fictitious transportation savings* to training samples that are contained in the unsafe set. This leads to the following insight.

COROLLARY 1. *The worst-case CVaR approximation is exact, that is, $Z^\star_{\text{CVaR}} = Z^\star_{\text{ICC}}$, under either of the following conditions.*

(i) We have $\hat{\boldsymbol{\xi}}_i \in \mathcal{S}(\boldsymbol{x}^\star)$ for all $i \in [N]$, where $\boldsymbol{x}^\star$ is optimal in (1)*.*

(ii) We have $\varepsilon \leq 1/N$.

REMARK 2. Propositions 4, 5 and 6 show that the worst-case CVaR approximation, despite being a best-in-class approximation among the inner linearizations $\mathcal{C}_{\text{ICC}}$ with uniform slopes, is in general not exact, that is,

$$Z^\star_{\text{ICC}} \;=\; \min_{\boldsymbol{\kappa} \in [0,1]^N} Z^\star_{\text{ICC}}(\boldsymbol{\kappa}) \;\leq\; \min_{\kappa \in [0,1]} Z^\star_{\text{ICC}}(\kappa \mathbf{e}) \;=\; Z^\star_{\text{ICC}}(\mathbf{e}) \;=\; Z^\star_{\text{CVaR}}.$$

Corollary 1 identifies sufficient conditions for the worst-case CVaR approximation to be exact.

2.2. Joint Chance Constraints with Right-Hand Side Uncertainty

Consider now an instance of problem (1) with a joint chance constraint corresponding to the safety set $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{a}_m^\top \boldsymbol{x} < \boldsymbol{b}_m^\top \boldsymbol{\xi} + b_m \;\; \forall m \in [M]\}$. As discussed in Section 1, we can assume that $\boldsymbol{b}_m \neq \mathbf{0}$ for all $m \in [M]$. By Proposition 2, the distributionally robust chance constrained program (1) is thus equivalent to the deterministic optimization problem

$$Z^\star_{\text{JCC}} = \min_{(\boldsymbol{x}, \boldsymbol{s}, t) \in \mathcal{C}_{\text{JCC}}} \boldsymbol{c}^\top \boldsymbol{x}, \tag{8}$$

whose feasible region is given by

$$\mathcal{C}_{\text{JCC}} = \left\{ (\boldsymbol{x}, \boldsymbol{s}, t) \in \mathcal{X} \times \mathbb{R}^N_+ \times \mathbb{R} \;\middle|\; \begin{array}{l} \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \\ \left(\min\limits_{m \in [M]} \left\{ \dfrac{\boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i + b_m - \boldsymbol{a}_m^\top \boldsymbol{x}}{\|\boldsymbol{b}_m\|_*} \right\} \right)^+ \geq t - s_i \quad \forall i \in [N] \end{array} \right\}.$$

In analogy to Section 2.1, one can again show that any convex inner approximation of $\mathcal{C}_{\rm JCC}$ is weakly dominated by a polyhedron of the form

$$\mathcal{C}_{\rm JCC}(\boldsymbol{\kappa}) = \left\{ (\boldsymbol{x}, \boldsymbol{s}, t) \in \mathcal{X} \times \mathbb{R}^N_+ \times \mathbb{R} \;\middle|\; \begin{array}{l} \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \\ \kappa_i \left(\dfrac{\boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i + b_m - \boldsymbol{a}_m^\top \boldsymbol{x}}{\|\boldsymbol{b}_m\|_*} \right) \geq t - s_i \quad \forall m \in [M],\; i \in [N] \end{array} \right\}$$

for some vector of slope parameters $\boldsymbol{\kappa} \in [0,1]^N$. The following assertion is akin to Proposition 3 and formalizes this statement. Its proof is omitted for the sake of brevity.

Proposition 8. *For any convex set $\mathcal{W} \subseteq \mathcal{C}_{\rm JCC}$, there exists $\boldsymbol{\kappa} \in [0,1]^N$ with $\mathcal{W} \subseteq \mathcal{C}_{\rm JCC}(\boldsymbol{\kappa}) \subseteq \mathcal{C}_{\rm JCC}$.*

Proposition 8 implies that amongst all convex conservative approximations to problem (8) it is sufficient to consider the family of linear programs

$$Z^\star_{\rm JCC}(\boldsymbol{\kappa}) = \min_{(\boldsymbol{x}, \boldsymbol{s}, t) \in \mathcal{C}_{\rm JCC}(\boldsymbol{\kappa})} \boldsymbol{c}^\top \boldsymbol{x} \tag{9}$$

parameterized by $\boldsymbol{\kappa} \in [0,1]^N$. One can show that the best approximation within this family is exact. The proof of this result is similar to that of Proposition 4 and thus omitted.

Proposition 9. *We have $Z^\star_{\rm JCC} = \min_{\boldsymbol{\kappa} \in [0,1]^N} Z^\star_{\rm JCC}(\boldsymbol{\kappa})$.*

Unfortunately, finding the best slope parameters $\boldsymbol{\kappa}^\star \in [0,1]^N$ is again NP-hard, but optimizing over the subclass of uniform slope parameters $\boldsymbol{\kappa} = \kappa \mathbf{e}$ for $\kappa \in [0,1]$ is easy, and $\boldsymbol{\kappa} = \mathbf{e}$ is optimal. This result is reminiscent of Proposition 5, and thus its proof is omitted for the sake of brevity.

Proposition 10. *We have $\min_{\kappa \in [0,1]} Z^\star_{\rm JCC}(\kappa \mathbf{e}) = Z^\star_{\rm JCC}(\mathbf{e})$.*

We now demonstrate that $\mathcal{C}_{\rm JCC}(\mathbf{e})$ can again be interpreted as the feasible set of a worst-case CVaR constraint. To see this, denote by $\Delta^M_{++} = \{\boldsymbol{w} \in (0,1)^M \mid \mathbf{e}^\top \boldsymbol{w} = 1\}$ the relative interior of the probability simplex and observe that for any vector of scaling factors $\boldsymbol{w} \in \Delta^M_{++}$ we have

$$\begin{aligned} \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \mathcal{S}(\boldsymbol{x})] \geq 1 - \varepsilon \quad &\iff \quad \mathbb{P}\Big[\max_{m \in [M]} \{ w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \geq 0 \Big] \leq \varepsilon \\ &\iff \quad \mathbb{P}\text{-VaR}_\varepsilon \Big(\max_{m \in [M]} \{ w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \Big) \leq 0 \\ &\Longleftarrow \quad \mathbb{P}\text{-CVaR}_\varepsilon \Big(\max_{m \in [M]} \{ w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \Big) \leq 0, \end{aligned}$$

where the first equivalence follows from the definition of the safety set $\mathcal{S}(\boldsymbol{x})$. We emphasize that the exact reformulations of the joint chance constraint in the first two lines of the above expression are unaffected by the particular choice of $\boldsymbol{w}$ (that is, for any $\boldsymbol{w}, \boldsymbol{w}' \in \Delta^M_{++}$, a decision $\boldsymbol{x}$ is feasible for $\boldsymbol{w}$ if and only if it is feasible for $\boldsymbol{w}'$), while the CVaR approximation changes with $\boldsymbol{w}$. Thus, the quality of the CVaR approximation can be tuned by varying $\boldsymbol{w} \in \Delta^M_{++}$ (see Ordoudis et al.

2021 for an application of this tuning in energy and reserve dispatch). Note also that the overall normalization $\mathbf{e}^\top \boldsymbol{w} = 1$ is non-restrictive because the CVaR is positive homogeneous.

We now introduce a family of worst-case CVaR constrained programs

$$Z^\star_{\text{CVaR}}(\boldsymbol{w}) = \begin{cases} \min\limits_{\boldsymbol{x} \in \mathcal{X}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \ \sup\limits_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon \Big(\max\limits_{m \in [M]} \{ w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \Big) \leq 0 \end{cases} \tag{10}$$

parameterized by $\boldsymbol{w} \in \Delta^M_{++}$, all of which conservatively approximate the ambiguous chance constrained program (1), that is, $Z^\star_{\text{JCC}} \leq Z^\star_{\text{CVaR}}(\boldsymbol{w})$. In fact, the family (10) contains an instance that is equivalent to the best bounding problem of the form (9) with uniform slope parameters.

Proposition 11. *We have $Z^\star_{\text{CVaR}}(\boldsymbol{w}^\star) = Z^\star_{\text{JCC}}(\mathbf{e})$ for $\boldsymbol{w}^\star \in \Delta^M_{++}$ defined through*

$$w^\star_m = \frac{\|\boldsymbol{b}_m\|_*^{-1}}{\sum_{\ell \in [M]} \|\boldsymbol{b}_\ell\|_*^{-1}} \quad \forall m \in [M].$$

As the quality of the CVaR approximation in (10) depends on the choice of $\boldsymbol{w}$, it would be desirable to identify the best (least conservative) approximation by solving $\min_{\boldsymbol{w} \in \Delta^M_{++}} Z^\star_{\text{CVaR}}(\boldsymbol{w})$. This could be achieved, for instance, by treating $\boldsymbol{w} \in \Delta^M_{++}$ as an additional decision variable in (19). Unfortunately, the resulting optimization problem involves bilinear terms in $\boldsymbol{x}$ and $\boldsymbol{w}$ and is thus non-convex. Finding the best CVaR approximation therefore appears to be computationally challenging. Even if the optimal scaling parameters were known, we will see in Section 3 that the corresponding instance of problem (10) would generically provide a *strict* upper bound on $Z^\star_{\text{JCC}}$.

The CVaR approximation (10) can again be interpreted as imposing a lower bound on the costs of moving training samples to the unsafe set. To see this, we define the *minimum signed distance* between a point $\boldsymbol{\xi}$ and a family of closed sets $\mathcal{C}_m$, $m \in [M]$, as $\textbf{min-dist}(\boldsymbol{\xi}, \{\mathcal{C}_m\}_{m \in [M]}) = \min_{m \in [M]} \textbf{sgn-dist}(\boldsymbol{\xi}, \mathcal{C}_m)$. We then obtain the following result (*cf.* Proposition 7).

Proposition 12. *If $\boldsymbol{w}$ is set to $\boldsymbol{w}^\star$ as defined in Proposition 11, then we have*

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon \Big(\max_{m \in [M]} \{ w^\star_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \Big) \leq 0 \iff \frac{1}{N} \sum_{i=1}^{\varepsilon N} \textbf{min-dist}(\hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})}, \{\mathcal{H}_m(\boldsymbol{x})\}_{m \in [M]}) \geq \theta,$$

where $\boldsymbol{\pi}(\boldsymbol{x})$ orders the data points $\hat{\boldsymbol{\xi}}_i$ by their minimum signed distances to $\mathcal{H}_m(\boldsymbol{x})$, $m \in [M]$.

The proof of Proposition 12 closely resembles that of Proposition 7 and is therefore omitted.

Corollary 2. *If $\boldsymbol{w}$ is set to $\boldsymbol{w}^\star$ as defined in Proposition 11, then the worst-case CVaR approximation is exact, that is, $Z^\star_{\text{CVaR}}(\boldsymbol{w}^\star) = Z^\star_{\text{JCC}}$, under either of the following conditions.*

(i) We have $\hat{\boldsymbol{\xi}}_i \in \mathcal{S}(\boldsymbol{x}^\star)$ for all $i \in [N]$, where $\boldsymbol{x}^\star$ is optimal in (1).

(ii) We have $\varepsilon \leq 1/N$.

The proof is similar to that of Corollary 1 and is thus omitted.

REMARK 3. In analogy to the individual chance constrained programs, Propositions 9, 10 and 11 establish the best-in-class performance of the worst-case CVaR approximation among the inner linearizations $\mathcal{C}_{\text{JCC}}$ with uniform slopes, and Corollary 2 identifies sufficient conditions under which the worst-case CVaR approximation is exact.

3. Bonferroni Approximation

We next investigate the Bonferroni approximation, which applies to joint chance constrained programs with right-hand side uncertainty, that is, a subclass of problem (1) where $\mathcal{S}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{A}\boldsymbol{x} < \boldsymbol{b}(\boldsymbol{\xi})\}$ for $\boldsymbol{A} \in \mathbb{R}^{M\times L}$ and an affine function $\boldsymbol{b} : \mathbb{R}^K \to \mathbb{R}^M$. In this context, note that Bonferroni's inequality (which is also known as the union bound) implies that

$$\mathbb{P}[\tilde{\boldsymbol{\xi}} \notin \mathcal{S}(\boldsymbol{x})] = \mathbb{P}[\boldsymbol{a}_1^\top \boldsymbol{x} \geq \boldsymbol{b}_1^\top \tilde{\boldsymbol{\xi}} + b_1 \quad \text{or} \quad \cdots \quad \text{or} \quad \boldsymbol{a}_M^\top \boldsymbol{x} \geq \boldsymbol{b}_M^\top \tilde{\boldsymbol{\xi}} + b_M] \leq \sum_{m\in[M]} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m].$$

Taking the supremum over all distributions in the Wasserstein ball then yields the bound

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\tilde{\boldsymbol{\xi}} \notin \mathcal{S}(\boldsymbol{x})] \leq \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \sum_{m\in[M]} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m] \leq \sum_{m\in[M]} \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m]. \tag{11}$$

For any collection of risk thresholds $\varepsilon_m \geq 0$, $m \in [M]$, such that $\sum_{m\in[M]} \varepsilon_m = \varepsilon$, the family of *individual* chance constraints

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m] \leq \varepsilon_m \quad \forall m \in [M] \tag{12}$$

thus provides a conservative approximation for the original *joint* chance constraint in (1) because

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\tilde{\boldsymbol{\xi}} \notin \mathcal{S}(\boldsymbol{x})] \leq \sum_{m\in[M]} \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m] \leq \sum_{m\in[M]} \varepsilon_m = \varepsilon,$$

where the two inequalities follow from (11) and (12), respectively. We thus refer to (12) as the *Bonferroni approximation* of the original chance constraint in problem (1). The Bonferroni approximation is attractive because the individual chance constraints in (12) are equivalent to simple linear inequalities. To see this, note that each individual chance constraint in (12) can be written as

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}[\boldsymbol{a}_m^\top \boldsymbol{x} \geq \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}} + b_m] \leq \varepsilon_m \quad &\Longleftrightarrow \quad \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\varepsilon_m}(\boldsymbol{a}_m^\top \boldsymbol{x} - b_m - \boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}}) \leq 0 \\
&\Longleftrightarrow \quad \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\varepsilon_m}(-\boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}}) + \boldsymbol{a}_m^\top \boldsymbol{x} - b_m \leq 0 \\
&\Longleftrightarrow \quad \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\varepsilon_m}(-\boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}}) \leq b_m - \boldsymbol{a}_m^\top \boldsymbol{x},
\end{aligned}$$

where the second equivalence holds because the value-at-risk is translation invariant. The m^{th} individual chance constraint in (12) thus simplifies to the linear inequality $\boldsymbol{a}_m^\top \boldsymbol{x} \leq b_m - \eta_m$, where the constant $\eta_m = \sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_{\varepsilon_m}(-\boldsymbol{b}_m^\top \tilde{\boldsymbol{\xi}})$ is independent of $\boldsymbol{x}$ and can thus be computed offline. Specifically, by using Corollary 5.3 of Mohajerin Esfahani and Kuhn (2018), we can express η_m as the optimal value of a deterministic optimization problem, that is,

$$\eta_m = \left\{ \begin{array}{lll} \min\limits_{\boldsymbol{\alpha},\beta,\boldsymbol{w},\eta} & \eta & \\ \text{s.t.} & \theta\beta + \dfrac{1}{N}\sum\limits_{i\in[N]} \alpha_i \leq \varepsilon_m & \\ & \alpha_i \geq 1 - w_i(\eta + \boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i) & \forall i\in[N] \\ & \beta \geq w_i \|\boldsymbol{b}_m\|_* & \forall i \in [N] \\ & \boldsymbol{\alpha}\geq \boldsymbol{0},\ \boldsymbol{w}\geq\boldsymbol{0}. & \end{array} \right.$$

The product of η and w_i in the second constraint group renders this problem non-convex. As the problem reduces to a linear program for any fixed value of the scalar decision variable η, however, η_m can be computed efficiently to any accuracy by a line search along η. In summary, under the Bonferroni approximation the chance constrained program (1) thus reduces to a highly tractable linear program. However, the quality of the approximation relies on the choice of the individual risk thresholds $\{\varepsilon_m\}_{m\in[M]}$. It is often recommended to set $\varepsilon_m = \varepsilon/M$ for all $m\in[M]$, but Chen et al. (2007) have shown that this choice can be conservative when the safety conditions are positively correlated. Optimizing over all admissible choices of $\{\varepsilon_m\}_{m\in[M]}$ is impractical because η_m generically displays a non-convex dependence on ε_m. Moreover, we will see that the Bonferroni approximation can be very conservative even if the risk thresholds $\{\varepsilon_m\}_{m\in[M]}$ are chosen optimally.

4. Incomparability of the Two Approximation Schemes

We close this paper by demonstrating that the CVaR approximation is generally incomparable to the Bonferroni approximation for ambiguous joint chance constraints over Wasserstein balls. In particular, we compare the two approximation schemes in the context of joint chance constrained programs with right-hand side uncertainty, and we provide two examples where either of the two approximations is strictly less conservative then the other one. This is in stark contrast to Chebyshev ambiguity sets, where the worst-case CVaR approximation is known to dominate the Bonferroni approximation (see Chen et al. 2010 and Zymler et al. 2013).

Example 1. Consider the following instance of the distributionally robust problem (1):

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & x_1 \\ \text{s.t.} & \mathbb{P}[x_1 > \tilde{\xi}_1,\ x_2 > \tilde{\xi}_2] \geq 1-\varepsilon \quad \forall \mathbb{P}\in\mathcal{F}(\theta) \\ & \underline{x}_1 \leq x_1 \leq \overline{x}_1,\ x_2 \geq 0. \end{array} \tag{13}$$

Here, we assume that $0 < \underline{x}_1 \leq \overline{x}_1 < 1$ and that the true data-generating distribution $\mathbb{P}_0$ is a two-point distribution which satisfies $\mathbb{P}_0[(\tilde{\xi}_1,\tilde{\xi}_2) = (1,0)] = p$ and $\mathbb{P}_0(\tilde{\xi}_1,\tilde{\xi}_2) = (0,0)] = 1-p$ for $p\in(0,1)$.

Proposition 13. *Let $p \in (\overline{x}_1\varepsilon, \varepsilon)$. As $N \to \infty$, with probability going to 1, we have for any vanishing sequence of Wasserstein radii $\theta(N)$ that*

(i) the Bonferroni approximation to problem (13) *that replaces the joint chance constraint with*

$$\mathbb{P}[x_1 > \tilde{\xi}_1] \geq 1 - \varepsilon_1 \ \forall \mathbb{P} \in \mathcal{F}(\theta), \quad \mathbb{P}[x_2 > \tilde{\xi}_2] \geq 1 - \varepsilon_2 \ \forall \mathbb{P} \in \mathcal{F}(\theta)$$

becomes exact if the risk thresholds $(\varepsilon_1, \varepsilon_2)$ are sufficiently close to $(\varepsilon, 0)$;

(ii) the worst-case CVaR approximation to (13) *that replaces the joint chance constraint with*

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon\big(\max\big\{w_1(\tilde{\xi}_1 - x_1), w_2(\tilde{\xi}_2 - x_2)\big\}\big) \leq 0$$

becomes infeasible for any choice of scaling factors (w_1, w_2).

Example 2. Consider the following instance of the distributionally robust problem (1):

$$\begin{aligned} \min_{\boldsymbol{x}} \ & x_3 \\ \text{s.t.} \ & \mathbb{P}[x_1 > \tilde{\xi}, \ x_2 > \tilde{\xi}] \geq 1 - \varepsilon \qquad \forall \mathbb{P} \in \mathcal{F}(\theta) \\ & \underline{x} \leq x_1, x_2, x_3 \leq 1, \ x_3 \geq x_1, \ x_3 \geq x_2. \end{aligned} \tag{14}$$

Here, we assume $\frac{1}{2} < \underline{x} \leq 1$ and that the true data-generating distribution $\mathbb{P}_0$ is a two-point distribution which satisfies $\mathbb{P}_0[\tilde{\xi} = 1] = p$ and $\mathbb{P}_0[\tilde{\xi} = 0] = 1 - p$ for $p \in (0, 1)$.

Proposition 14. *Let $p \in (\varepsilon/2, \underline{x}\varepsilon]$. As $N \to \infty$, with probability going to 1, we have for any vanishing sequence of Wasserstein radii $\theta(N)$ that*

(i) the worst-case CVaR approximation to (14) *that replaces the joint chance constraint with*

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon\big(\max\big\{w_1(\tilde{\xi} - x_1), w_2(\tilde{\xi} - x_2)\big\}\big) \leq 0$$

becomes exact if the scaling factors (w_1, w_2) are $(\frac{1}{2}, \frac{1}{2})$;

(ii) the Bonferroni approximation to problem (14) *that replaces the joint chance constraint with*

$$\mathbb{P}[x_1 > \tilde{\xi}] \geq 1 - \varepsilon_1 \ \forall \mathbb{P} \in \mathcal{F}(\theta), \quad \mathbb{P}[x_2 > \tilde{\xi}] \geq 1 - \varepsilon_2 \ \forall \mathbb{P} \in \mathcal{F}(\theta)$$

becomes infeasible for any choice of the risk thresholds $(\varepsilon_1, \varepsilon_2)$.

Acknowledgments

The authors gratefully acknowledge financial support from the ECS grant CityU_21502820, the SNSF grant BSCGI0_157733 and the EPSRC grant EP/N020030/1.

References

Ben-Tal, Aharon, Arkadi Nemirovski. 2001. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM.

Chen, Wenqing, Melvyn Sim, Jie Sun, Chung-Piaw Teo. 2010. From CVaR to uncertainty set: implications in joint chance-constrained optimization. *Operations Research* **58**(2) 470–485.

Chen, Xin, Melvyn Sim, Peng Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Chen, Zhi, Daniel Kuhn, Wolfram Wiesemann. 2022. Data-driven chance constrained programs over Wasserstein balls. Forthcoming in *Operations Research*.

Hanasusanto, Grani A, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151**(1) 35–62.

Ho-Nguyen, Nam, Fatma Kılınç-Karzan, Simge Küçükyavuz, Dabeen Lee. 2020. Strong formulations for distributionally robust chance-constrained programs with left-hand side uncertainty under Wasserstein ambiguity. *arXiv preprint arXiv:2007.06750*.

Ho-Nguyen, Nam, Fatma Kılınç-Karzan, Simge Küçükyavuz, Dabeen Lee. 2021. Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity. *Mathematical Programming* 1–32.

Ho-Nguyen, Nam, Stephen Wright. 2022. Adversarial classification via distributional robustness with Wasserstein ambiguity. *Mathematical Programming* 1–37.

Ji, Ran, Miguel Lejeune. 2021. Data-driven distributionally robust chance-constrained optimization with Wasserstein metric. *Journal of Global Optimization* **79**(4) 779–811.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1-2) 1–52.

Nemirovski, Arkadi. 2012. On safe tractable approximations of chance constraints. *European Journal of Operational Research* **219**(3) 707–718.

Ordoudis, Christos, Viet Anh Nguyen, Daniel Kuhn, Pierre Pinson. 2021. Energy and reserve dispatch with distributionally robust joint chance constraints. *Operations Research Letters* **49**(3) 291–299.

Shen, Haoming, Ruiwei Jiang. 2022. Chance-constrained set covering with Wasserstein ambiguity. *Mathematical Programming* 1–54.

Xie, Weijun. 2021. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming* **186**(1) 115–155.

Xie, Weijun, Shabbir Ahmed. 2020. Bicriteria approximation of chance-constrained covering problems. *Operations Research* **68**(2) 516–533.

Xie, Weijun, Shabbir Ahmed, Ruiwei Jiang. 2019. Optimized Bonferroni approximations of distributionally robust joint chance constraints. *Mathematical Programming* **191**(1) 79–112.

Yang, Wenzhuo, Huan Xu. 2016. Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming* **155**(1-2) 231–265.

Zymler, Steve, Daniel Kuhn, Berç Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* **137**(1-2) 167–198.

Proofs

Proof of Proposition 3. It is clear that $\mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa}) \subseteq \mathcal{C}_{\text{ICC}}$ for every $\boldsymbol{\kappa} \in [0,1]^N$. Next, we show that for every $i \in [N]$ there exists $\kappa_i \in [0,1]$ such that the constraint $\kappa_i((\boldsymbol{b}-\boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x}) \geq t - s_i$ is valid for $\mathcal{W}$. The resulting set $\mathcal{C}_{\text{ICC}}(\boldsymbol{\kappa})$ is thus a convex outer approximation of $\mathcal{W}$.

To determine κ_i, consider the sets $\mathcal{W}_i = \{(\boldsymbol{x}, s_i, t) \mid (\boldsymbol{x}, \boldsymbol{s}, t) \in \mathcal{W}\}$ and $\mathcal{V}_i = \{(\boldsymbol{x}, s_i, t) \mid ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x})^+ < t - s_i\}$. By construction, $\mathcal{W}_i$ and $\mathcal{V}_i$ are intersection-free and convex. Thus, they admit a separating hyperplane. The same holds true if we replace $\mathcal{W}_i$ with

$$\overline{\mathcal{W}}_i = \text{conv}\big(\mathcal{W}_i \cup \{(\boldsymbol{x}, s_i, t) \mid ((\boldsymbol{b}-\boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x}, t - s_i) = (0,0)\}\big).$$

The separating hyperplane between $\overline{\mathcal{W}}_i$ and $\mathcal{V}_i$ must satisfy $t - s_i = 0$ whenever $(\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x} = 0$. In other words, the separating hyperplane must be of the form $\kappa_i((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x}) = t - s_i$ for some $\kappa_i \in [0,1]$. Thus, the claim follows. □

Proof of Proposition 4. It follows from Proposition 1 that

$$\min_{\boldsymbol{\kappa}\in[0,1]^N} Z^\star_{\text{ICC}}(\boldsymbol{\kappa}) = \begin{cases} \min\limits_{\boldsymbol{s},t,\boldsymbol{x},\boldsymbol{\kappa}} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{ s.t.} & \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ & \kappa_i((\boldsymbol{b}-\boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x}) \geq t - s_i \quad \forall i \in [N] \\ & \boldsymbol{s} \geq \mathbf{0},\ \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\kappa} \in [0,1]^N. \end{cases} \tag{15}$$

For any fixed $(\boldsymbol{x}, \boldsymbol{s}, t)$, the optimal (that is, least restrictive) choice of $\boldsymbol{\kappa}$ satisfies

$$\kappa_i = \begin{cases} 1 & \text{if } (\boldsymbol{b}-\boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x} \geq 0, \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in [N]. \tag{16}$$

Eliminating $\boldsymbol{\kappa}$ from (15) by substituting (16) into (15) converts the second constraint group to

$$((\boldsymbol{b}-\boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x})^+ \geq t - s_i \quad \forall i \in [N],$$

which shows that (15) is equivalent to (5). Thus, the claim follows. □

Proof of Proposition 5. We first show that problem (6) is infeasible for $\boldsymbol{\kappa} = \mathbf{0}$, that is, $Z^\star_{\text{ICC}}(\mathbf{0}) = \infty$. Indeed, by the definition of $\mathcal{C}_{\text{ICC}}(\mathbf{0})$ we have

$$Z^\star_{\text{ICC}}(\mathbf{0}) = \begin{cases} \min\limits_{\boldsymbol{s},t,\boldsymbol{x}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{ s.t. } \ \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ \qquad \boldsymbol{s} \geq t\mathbf{e},\ \boldsymbol{s} \geq \mathbf{0},\ \boldsymbol{x} \in \mathcal{X}. \end{cases}$$

Any feasible solution of the above problem satisfies $\varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \leq \varepsilon N t - N \max\{t, 0\} \leq 0$, where the first inequality follows from the constraints $\boldsymbol{s} \geq t\mathbf{e}$ and $\boldsymbol{s} \geq \mathbf{0}$. As $\theta > 0$, the constraint $\varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq$

$\theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_*$ is thus satisfied only if $\boldsymbol{s} = \boldsymbol{0}$, $t = 0$ and $\boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{b}$. However, the last equality contradicts our standing assumption that $\boldsymbol{A}^\top \boldsymbol{x} \neq \boldsymbol{b}$ for all $\boldsymbol{x} \in \mathcal{X}$, confirming that the above problem is infeasible and $Z^\star_{\rm ICC}(\boldsymbol{0}) = \infty$. Thus, $Z^\star_{\rm ICC}(\kappa \mathbf{e})$ is minimized by some $\kappa \in (0,1]$.

If $\kappa \in (0,1]$, we can use the variable substitution $t \leftarrow \kappa t$ and $\boldsymbol{s} \leftarrow \kappa \boldsymbol{s}$ to re-express problem (6) as

$$Z^\star_{\rm ICC}(\kappa \mathbf{e}) = \begin{cases} \min\limits_{\boldsymbol{s},t,\boldsymbol{x}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \ \ \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \dfrac{\theta N}{\kappa} \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ \qquad (\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_i + b - \boldsymbol{a}^\top \boldsymbol{x} \geq t - s_i \quad \forall i \in [N] \\ \qquad \boldsymbol{s} \geq \boldsymbol{0}, \ \boldsymbol{x} \in \mathcal{X}. \end{cases}$$

From this formulation it is evident that $\kappa^\star = 1$ is the best (least restrictive) choice of $\kappa \in (0,1]$. □

Proof of Proposition 6. Using now standard techniques, the worst-case CVaR in (7) can be re-expressed as the optimal value of a finite conic program,

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) = \begin{cases} \min\limits_{\boldsymbol{\alpha},\beta,\tau} \ \tau + \dfrac{1}{\varepsilon}\Big(\theta\beta + \dfrac{1}{N}\sum\limits_{i \in [N]} \alpha_i\Big) \\ \text{s.t.} \ \ \alpha_i \geq \boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \hat{\boldsymbol{\xi}}_i - \tau \quad \forall i \in [N] \\ \qquad \boldsymbol{\alpha} \geq \boldsymbol{0}, \ \beta \geq \|\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b}\|_*, \end{cases}$$

see Mohajerin Esfahani and Kuhn (2018, § 5.1 and § 7.1) for a detailed derivation. Substituting this reformulation into the worst-case CVaR constrained program (7) yields

$$Z^\star_{\rm CVaR} = \begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{\alpha},\beta,\tau} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \ \ \tau + \dfrac{1}{\varepsilon}\Big(\theta\beta + \dfrac{1}{N}\sum\limits_{i \in [N]} \alpha_i\Big) \leq 0 \\ \qquad \alpha_i \geq \boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \hat{\boldsymbol{\xi}}_i - \tau \quad \forall i \in [N] \\ \qquad \boldsymbol{\alpha} \geq \boldsymbol{0}, \ \beta \geq \|\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b}\|_*, \ \boldsymbol{x} \in \mathcal{X}. \end{cases}$$

As $\theta > 0$ and $\varepsilon > 0$, it is clear that $\beta = \|\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b}\|_*$ at optimality, and this insight allows us to eliminate β from the above optimization problem. Multiplying the first constraint with the positive constant εN while renaming $\boldsymbol{\alpha}$ as $\boldsymbol{s}$ and τ as $-t$ then shows that $Z^\star_{\rm CVaR} = Z^\star_{\rm ICC}(\mathbf{e})$. □

Proof of Proposition 7. It follows from the proof of Proposition 6 that the worst-case CVaR constraint $\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon(\boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \tilde{\boldsymbol{\xi}}) \leq 0$ holds if and only if

$$0 \geq \begin{cases} \min\limits_{\boldsymbol{\alpha},\tau} \ \tau + \dfrac{1}{\varepsilon}\Big(\theta\|\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b}\|_* + \dfrac{1}{N}\sum\limits_{i \in [N]} \alpha_i\Big) \\ \text{s.t.} \ \ \alpha_i \geq \boldsymbol{a}^\top \boldsymbol{x} - b + (\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b})^\top \hat{\boldsymbol{\xi}}_i - \tau \quad \forall i \in [N] \\ \qquad \boldsymbol{\alpha} \geq \boldsymbol{0}, \end{cases}$$

which, by multiplying the objective function by the positive constant εN while renaming $\boldsymbol{\alpha}$ as $\boldsymbol{s}$ and τ as $-t$, is equivalent to

$$\exists \boldsymbol{s} \geq \mathbf{0},\, t \in \mathbb{R} : \begin{cases} \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ (\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x} \geq t - s_i \quad \forall i \in [N]. \end{cases} \tag{17}$$

This constraint system is satisfiable by $t \in \mathbb{R}$ and *some* $\boldsymbol{s} \geq \mathbf{0}$ if and only if it is satisfiable by t and $\boldsymbol{s}^\star(t)$ defined by $s_i^\star(t) = (t - ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x}))^+$, $i \in [N]$. Since the second constraint in (17) is automatically satisfied by $\boldsymbol{s}^\star(t)$, we thus conclude that (17) holds if and only if

$$\begin{aligned} &\exists t \in \mathbb{R} : \varepsilon N t - \sum_{i \in [N]} (t - ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x}))^+ \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_* \\ \Longleftrightarrow\ &\max_{t \in \mathbb{R}} \left\{ \varepsilon N t - \sum_{i \in [N]} (t - ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x}))^+ \right\} \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_*. \end{aligned} \tag{18}$$

The objective function of the embedded maximization problem on the left-hand side of (18) is piecewise affine and concave in t. Moreover, by construction of $\boldsymbol{\pi}(\boldsymbol{x})$, we have

$$(\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x} \ \leq\ (\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_j(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x} \qquad \forall 1 \leq i \leq j \leq N.$$

The first-order optimality condition for non-smooth optimization then implies that the maximum on the left-hand side of (18) is attained by $t^\star = (\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_{\lfloor \varepsilon N \rfloor + 1}(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x}$, which results in the equivalent constraint

$$\sum_{i=1}^{\varepsilon N} ((\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x})^\top \hat{\boldsymbol{\xi}}_{\pi_i(\boldsymbol{x})} + b - \boldsymbol{a}^\top \boldsymbol{x}) \geq \theta N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_*.$$

The result now follows if we divide both sides of the constraint by $N \|\boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{x}\|_*$. □

Proof of Corollary 1. The first condition immediately follows from Theorem 1 and Proposition 7 since $\textbf{sgn-dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x})) = \textbf{dist}(\hat{\boldsymbol{\xi}}_i, \bar{\mathcal{S}}(\boldsymbol{x}))$ whenever $\hat{\boldsymbol{\xi}}_i \in \mathcal{S}(\boldsymbol{x})$. The second condition guarantees that $\hat{\boldsymbol{\xi}}_i \in \mathcal{S}(\boldsymbol{x})$, $i \in [N]$, for any solution $\boldsymbol{x} \in \mathcal{X}$ that satisfies the ambiguous chance constraint in problem (1). This, in turn, implies that the first condition of the corollary is satisfied as well. □

Proof of Proposition 11. Using techniques introduced by Mohajerin Esfahani and Kuhn (2018), the worst-case CVaR in (10) can be re-expressed as

$$\begin{aligned} &\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\text{-CVaR}_\varepsilon \left(\max_{m \in [M]} \{ w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \boldsymbol{\xi} - b_m) \} \right) \\ &= \begin{cases} \min\limits_{\boldsymbol{\alpha}, \beta, \tau} \ \tau + \dfrac{1}{\varepsilon} \left(\theta \beta + \dfrac{1}{N} \sum\limits_{i \in [N]} \alpha_i \right) \\ \text{s.t.} \ \ \alpha_i \geq w_m (\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i - b_m) - \tau \quad \forall m \in [M],\ i \in [N] \\ \qquad \beta \geq w_m \|\boldsymbol{b}_m\|_* \quad \forall m \in [M] \\ \qquad \boldsymbol{\alpha} \geq \mathbf{0}. \end{cases} \end{aligned}$$

Substituting this reformulation into (10) yields

$$Z^\star_{\mathrm{CVaR}}(\boldsymbol{w}) = \begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{\alpha},\beta,\tau} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \tau + \frac{1}{\varepsilon}\Big(\theta\beta + \frac{1}{N}\sum\limits_{i\in[N]} \alpha_i\Big) \le 0 \\ & \alpha_i \ge w_m(\boldsymbol{a}_m^\top \boldsymbol{x} - \boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i - b_m) - \tau \quad \forall m \in [M],\ i \in [N] \\ & \beta \ge w_m \|\boldsymbol{b}_m\|_* \quad \forall m \in [M] \\ & \boldsymbol{\alpha} \ge \mathbf{0},\ \boldsymbol{x} \in \mathcal{X}. \end{cases} \tag{19}$$

As $\theta > 0$ and $\varepsilon > 0$, it is clear that $\beta = \max_{m\in[M]}\{w_m\|\boldsymbol{b}_m\|_*\}$ at optimality, and this insight allows us to eliminate β from the above optimization problem. Multiplying the first constraint by the positive constant $\varepsilon N/\beta$ and the second constraint group by the positive constant $1/\beta$ while applying the variable substitutions $\boldsymbol{s} \leftarrow \boldsymbol{\alpha}/\beta$ and $-t \leftarrow \tau/\beta$, we obtain

$$Z^\star_{\mathrm{CVaR}}(\boldsymbol{w}) = \begin{cases} \min\limits_{\boldsymbol{s},t,\boldsymbol{x}} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \varepsilon N t - \mathbf{e}^\top \boldsymbol{s} \ge \theta N \\ & \dfrac{w_m\|\boldsymbol{b}_m\|_*}{\max\limits_{m\in[M]}\{w_m\|\boldsymbol{b}_m\|_*\}} \dfrac{(\boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i + b_m - \boldsymbol{a}_m^\top \boldsymbol{x})}{\|\boldsymbol{b}_m\|_*} \ge t - s_i \quad \forall m \in [M],\ i \in [N] \\ & \boldsymbol{s} \ge \mathbf{0},\ \boldsymbol{x} \in \mathcal{X}. \end{cases} \tag{20}$$

Replacing $\boldsymbol{w}$ with $\boldsymbol{w}^\star$, the second constraint group in problem (20) simplifies to

$$\min_{m\in[M]} \left\{ \frac{(\boldsymbol{b}_m^\top \hat{\boldsymbol{\xi}}_i + b_m - \boldsymbol{a}_m^\top \boldsymbol{x})}{\|\boldsymbol{b}_m\|_*} \right\} \ge t - s_i \ \forall i \in [N],$$

which reveals that the feasible set of problem (20) coincides with $C_{\mathrm{JCC}}(\mathbf{e})$. This observation implies the postulated assertion that $Z^\star_{\mathrm{CVaR}}(\boldsymbol{w}^\star) = Z^\star_{\mathrm{JCC}}(\mathbf{e})$. □

Proof of Proposition 13. We proceed in three steps. We first derive the optimal value of the classical chance constrained program associated with (13) under the true data-generating distribution $\mathbb{P}_0$ (Step 1). This value serves as a lower bound on the optimal value of problem (13). We then show that with probability going to 1 as $N \to \infty$, the Bonferroni approximation achieves this bound (Step 2), whereas the worst-case CVaR approximation becomes infeasible (Step 3).

Step 1. Since $p < \varepsilon$, the feasible region of the classical chance constrained program

$$\begin{aligned} \min_{\boldsymbol{x}} \ & x_1 \\ \text{s.t.} \ & \mathbb{P}_0[x_1 > \tilde{\xi}_1,\ x_2 > \tilde{\xi}_2] \ge 1 - \varepsilon \\ & \underline{x}_1 \le x_1 \le \overline{x}_1,\ x_2 \ge 0 \end{aligned}$$

under the true data-generating distribution $\mathbb{P}_0$ is $\{(x_1,x_2) \in \mathbb{R}^2 \mid x_1 \in [\underline{x}_1, \overline{x}_1],\ x_2 > 0\}$. Hence, the optimal value of this problem is $\underline{x}_1 > 0$, which is attained by any $(x_1,x_2) \in \{\underline{x}_1\} \times (\mathbb{R}_+ \setminus \{0\})$.

Step 2. Fix any $(x_1,x_2)\in[\underline{x}_1,\overline{x}_1]\times(\mathbb{R}_+\backslash\{0\})$, and denote by $\mathcal{S}_1(\boldsymbol{x})=\{\boldsymbol{\xi}\mid x_1>\xi_1\}$ and $\mathcal{S}_2(\boldsymbol{x})=\{\boldsymbol{\xi}\mid x_2>\xi_2\}$ the two safety sets of the Bonferroni approximation. If $\hat{\boldsymbol{\xi}}_i=(1,0)^\top$, then $\hat{\boldsymbol{\xi}}_i\in\bar{\mathcal{S}}_1(\boldsymbol{x})\cap\mathcal{S}_2(\boldsymbol{x})$ with $\mathbf{dist}(\hat{\boldsymbol{\xi}}_i,\bar{\mathcal{S}}_1(\boldsymbol{x}))=0$ and $\mathbf{dist}(\hat{\boldsymbol{\xi}}_i,\bar{\mathcal{S}}_2(\boldsymbol{x}))=x_2$. Likewise, if $\hat{\boldsymbol{\xi}}_i=(0,0)^\top$, then $\hat{\boldsymbol{\xi}}_i\in\mathcal{S}_1(\boldsymbol{x})\cap\mathcal{S}_2(\boldsymbol{x})$ with $\mathbf{dist}(\hat{\boldsymbol{\xi}}_i,\bar{\mathcal{S}}_1(\boldsymbol{x}))=x_1$ and $\mathbf{dist}(\hat{\boldsymbol{\xi}}_i,\bar{\mathcal{S}}_2(\boldsymbol{x}))=x_2$. Under the appropriate permutations $\boldsymbol{\pi}^1(\boldsymbol{x})$ and $\boldsymbol{\pi}^2(\boldsymbol{x})$, Theorem 1 then implies that $\boldsymbol{x}$ satisfies both chance constraints of the Bonferroni approximation if and only if

$$\frac{1}{N}\sum_{i=1}^{\varepsilon_1 N}\mathbf{dist}(\hat{\boldsymbol{\xi}}_{\pi_i^1(\boldsymbol{x})},\bar{\mathcal{S}}_1(\boldsymbol{x}))=\frac{1}{N}(\varepsilon_1 N-I)^+x_1\geq\theta(N)\quad\text{and}\quad\frac{1}{N}\sum_{i=1}^{\varepsilon_2 N}\mathbf{dist}(\hat{\boldsymbol{\xi}}_{\pi_i^2(\boldsymbol{x})},\bar{\mathcal{S}}_2(\boldsymbol{x}))=\varepsilon_2 x_2\geq\theta(N),\tag{21}$$

where I denotes the number of samples $\hat{\boldsymbol{\xi}}_i$, $i\in[N]$, that satisfy $\hat{\boldsymbol{\xi}}_i=(1,0)^\top$.

Choose $\varepsilon_1\in(p,\varepsilon)$ and $\varepsilon_2=\varepsilon-\varepsilon_1$, as well as $x_1=\underline{x}_1$ and any $x_2\geq\theta(N)/\varepsilon_2$. This choice of $(\varepsilon_1,\varepsilon_2)$ and $\boldsymbol{x}$ satisfies the second constraint in (21) by construction. To see that the first constraint in (21) is also satisfied with high probability as $N\to\infty$, we note that $\frac{1}{N}(\varepsilon_1 N-I)^+x_1=(\varepsilon_1-I/N)^+x_1\to(\varepsilon_1-p)^+x_1$ almost surely as $N\to\infty$ due to the strong law of large numbers. We thus conclude that $\frac{1}{N}(\varepsilon_1 N-I)^+x_1>0$ with high probability as $N\to\infty$, and thus this quantity will exceed $\theta(N)$, which goes to zero as N approaches infinity.

Step 3. Using similar arguments as in the proofs of Propositions 7 and 11, one can show that the worst-case CVaR approximation is satisfied for a fixed decision (x_1,x_2) if and only if

$$\max_{t\in\mathbb{R}}\left\{\varepsilon Nt-\sum_{i\in[N]}\left(t-\min\left\{\frac{w_1(x_1-\hat{\xi}_{i,1})}{\max\{w_1,w_2\}},\frac{w_2(x_2-\hat{\xi}_{i,2})}{\max\{w_1,w_2\}}\right\}\right)^+\right\}\geq\theta(N)N.\tag{22}$$

Here, $\hat{\xi}_{i,1}$ (resp., $\hat{\xi}_{i,2}$) refers to the first (resp., second) component of the vector $\hat{\boldsymbol{\xi}}_i$. The first-order optimality condition for non-smooth optimization then implies that the maximum on the left-hand side of this constraint is attained by

$$t^\star=\min\left\{\frac{w_1(x_1-\hat{\xi}_{\pi_{\lfloor\varepsilon N\rfloor+1}(\boldsymbol{x}),1})}{\max\{w_1,w_2\}},\frac{w_2(x_2-\hat{\xi}_{\pi_{\lfloor\varepsilon N\rfloor+1}(\boldsymbol{x}),2})}{\max\{w_1,w_2\}}\right\},$$

where we make use of the permutation $\boldsymbol{\pi}(\boldsymbol{x})$ that orders the data points $\hat{\boldsymbol{\xi}}_i$, $i\in[N]$ in ascending order of the expressions

$$\min\left\{\frac{w_1(x_1-\hat{\xi}_{i,1})}{\max\{w_1,w_2\}},\frac{w_2(x_2-\hat{\xi}_{i,2})}{\max\{w_1,w_2\}}\right\},\quad i\in[N].$$

This implies that the worst-case CVaR constraint (22) holds if and only if

$$\sum_{i=1}^{\varepsilon N}\min\left\{\frac{w_1(x_1-\hat{\xi}_{\pi_i(\boldsymbol{x}),1})}{\max\{w_1,w_2\}},\frac{w_2(x_2-\hat{\xi}_{\pi_i(\boldsymbol{x}),2})}{\max\{w_1,w_2\}}\right\}\geq\theta(N)N.\tag{23}$$

Note that $w_1/\max\{w_1, w_2\} \leq 1$ in the first term inside the minimum. Hence, a necessary condition for the inequality (23) to hold for any scaling factors (w_1, w_2) is that $\sum_{i=1}^{\varepsilon N}(x_1 - \hat{\xi}_{\pi_i(\boldsymbol{x}),1}) \geq \theta(N)N$; otherwise, the sum of the first terms inside the minima is smaller than $\theta(N)N$. Note that for any permutation $\boldsymbol{\pi}(\boldsymbol{x})$, the strong law of large numbers implies that $\frac{1}{N}\sum_{i=1}^{\varepsilon N}\hat{\xi}_{\pi_i(\boldsymbol{x}),1}$ converges to a number smaller than or equal to p almost surely as N approaches infinity. Since $\frac{1}{N}\sum_{i=1}^{\varepsilon N} x_1 = \varepsilon x_1$, we thus conclude that $\frac{1}{N}\sum_{i=1}^{\varepsilon N}(x_1 - \hat{\xi}_{\pi_i(\boldsymbol{x}),1})$ converges to a number not exceeding $\varepsilon x_1 - p$ almost surely as N approaches infinity. Since $\overline{x}_1\varepsilon < p$ by assumption, this implies that the inequality (23) is violated for all $x_1 \in [\underline{x}_1, \overline{x}_1]$ with high probability as N approaches infinity. □

Proof of Proposition 14. We proceed in three steps. We first derive the optimal value of the classical chance constrained program associated with (14) under the true data-generating distribution $\mathbb{P}_0$ (Step 1). This value serves as a lower bound on the optimal value of problem (14). We then show that with probability going to 1 as $N \to \infty$, the worst-case CVaR approximation achieves this bound (Step 2), whereas the Bonferroni approximation becomes infeasible (Step 3).

Step 1. Since $p < \varepsilon$, a similar argument as in the proof of Proposition 13 allows us to conclude that the optimal value of the classical chance constrained program under the true data-generating distribution $\mathbb{P}_0$ is $\underline{x}$, which is attained by the solution $(x_1, x_2, x_3) = (\underline{x}, \underline{x}, \underline{x})$.

Step 2. By Proposition 12, the solution $\boldsymbol{x} = (x_1, x_2, x_3) = (\underline{x}, \underline{x}, \underline{x})$ is feasible in the worst-case CVaR approximation with scaling factors $(w_1, w_2) = (\frac{1}{2}, \frac{1}{2})$ if and only if

$$\frac{1}{N}\sum_{i=1}^{\varepsilon N}\textbf{min-dist}(\hat{\xi}_{\pi_i(\boldsymbol{x})}, \mathcal{H}_1(\boldsymbol{x}), \mathcal{H}_2(\boldsymbol{x})) \geq \theta(N), \tag{24}$$

where $\mathcal{H}_1(\boldsymbol{x}) = \mathcal{H}_2(\boldsymbol{x}) = \{\xi \mid \xi \geq \underline{x}\}$, and the permutation $\boldsymbol{\pi}(\boldsymbol{x})$ orders the data points $\hat{\xi}_i$ so that $\hat{\xi}_1, \ldots, \hat{\xi}_I = 1$, $I \in [N] \cup \{0\}$, and $\hat{\xi}_{I+1}, \ldots, \hat{\xi}_N = 0$. Since $\textbf{min-dist}(\hat{\xi}_i, \mathcal{H}_1(\boldsymbol{x}), \mathcal{H}_2(\boldsymbol{x})) = \underline{x} - 1$ for $i = 1, \ldots, I$ and $\textbf{min-dist}(\hat{\xi}_i, \mathcal{H}_1(\boldsymbol{x}), \mathcal{H}_2(\boldsymbol{x})) = \underline{x}$ for $i = I+1, \ldots, N$, (24) holds if and only if

$$\frac{1}{N}\left(\min\{\varepsilon N, I\}(\underline{x} - 1) + (\varepsilon N - I)^+\underline{x}\right) \geq \theta(N).$$

Note that $I/N \to p$ as $N \to \infty$ by the strong law of large numbers. Since $p < \varepsilon$ and $\theta(N) \to 0$ as $N \to \infty$, the above inequality is thus satisfied with probability approaching 1 as $N \to \infty$ as long as $p(\underline{x} - 1) + (\varepsilon - p)\underline{x} = \varepsilon\underline{x} - p$ is strictly positive. This is the case since $p < \underline{x}\varepsilon$ by assumption.

Step 3. Observe that the Bonferroni approximation is infeasible if $\varepsilon_1 \leq I/N$ because the first individual chance constraint $\mathbb{P}[x_1 > \tilde{\xi}] \geq 1 - \varepsilon_1 \ \forall \mathbb{P} \in \mathcal{F}(\theta)$ is already violated under the empirical distribution. For the same reason, the Bonferroni approximation is infeasible if $\varepsilon_2 \leq I/N$. We next show that when $N \to \infty$, with probability approaching to 1, any pair of Bonferroni weights $(\varepsilon_1, \varepsilon_2)$ satisfying $\varepsilon_1 + \varepsilon_2 = \varepsilon$ also satisfies $\min\{\varepsilon_1, \varepsilon_2\} \leq I/N$, that is, at least one of the two individual chance constraints is violated. Indeed, we have $\min\{\varepsilon_1, \varepsilon_2\} \leq \varepsilon/2$ and $p > \varepsilon/2$ by assumption, and $I/N \to p$ as $N \to \infty$ by the strong law of large numbers. □