A Decomposition Algorithm for Distributionally Robust Chance-Constrained Programs with Polyhedral Ambiguity Set

Soumya Ranjan Pathy*[1] and Hamed Rahimian†[1]

[1]Department of Industrial Engineering, Clemson University, Clemson SC 29634, USA

Abstract

In this paper, we study a distributionally robust optimization approach to chance-constrained stochastic programs to hedge against uncertainty in the distributions of the random parameters. We consider a general polyhedral ambiguity set under finite support and study Wasserstein ambiguity set, total variation distance ambiguity set, and moment-based ambiguity set as examples for our computations. We develop a decomposition-based solution approach to solve the model and take advantage of mixing inequalities to develop custom feasibility cuts. A probability cut framework is also developed to handle the distributionally robust chance constraint. Finally, we present a numerical study to illustrate the effectiveness of the proposed formulations and showcase our results for the chosen ambiguity set examples.

Keywords: Distributionally robust optimization, Chance-constrained programming, Decomposition algorithm, Cutting planes

1 Introduction

The success or failure of a supply chain is determined by customer satisfaction. To enhance customer satisfaction, businesses are endeavoring to optimize the performance of their supply chains by designing a robust and resilient [11, 14]. Such a supply chain allows for the synchronization of supply to meet fluctuations in demand, both peaks and troughs [44]. According to [43], retailers globally suffered a loss of $1.1 trillion because of demand uncertainty.

Demand uncertainty refers to the unpredictability of customer demand for products or services provided by the buyer. Factors such as consumer preferences, competition, and economic fluctuations contribute to this uncertainty [4]. In such volatile market conditions, where businesses must prioritize customer satisfaction, it is pragmatic to take into account the impact of uncertainties on supply chain planning to mitigate their effects [36]. While advanced techniques like time series analysis have been developed to enhance forecasting precision, uncertainties in demand are inevitable due to the constantly evolving market conditions [54]. Overestimating uncertainties can lead to overly cautious decisions, unnecessarily compromising the objective function. Moreover, striving to guarantee feasibility for all possible uncertainty scenarios may also result in excessively conservative decisions. As a result, deterministic models are incapable of offering optimal strategies when faced with demand uncertainty.

Recognizing these facts, there is a necessity to develop decision-making models that account for uncertainty. These models operate under the assumption that the decision-maker possesses some

*spathy@clemson.edu
†hrahimi@clemson.edu

level of information regarding the uncertain demand. This information could entail knowledge of the entire distribution governing these uncertain inputs or, at the very least, the mean and standard deviation of the random demand. Historical data can be leveraged to generate estimates of future demand, which can then inform decision-making processes. In recent decades, numerous such approaches have been suggested to tackle uncertainty, with chance-constrained programming emerging as notably prominent.

Chance-constrained programming (CCP) is a powerful tool for decision-making under uncertainty. Instead of satisfying all system constraints due to the inherent uncertainty in the model, it aims to satisfy the system constraints with high probability. Thus, by specifying a minimum probability for satisfying a constraint, decision-makers can hedge against risk while minimizing costs. The study of CCP has a rich history dating back to the 1950s, with seminal works by [8, 35, 40]. In terms of supply chain management, CCP has been implemented in many past studies. A CCP model is studied in [3] for a production planning problem under demand and price uncertainties and in [45] for a network design problem under demand uncertainty. Similar model is studied for an inventory management problem in [19], where the authors consider a chance constraint to enforce the probability that an inventory shortage occurs during the planning horizon is limited to a maximum acceptable risk level. A risk-averse objective by using a risk measure, such as Conditional Value-at-Risk (CVaR), is considered in some studies to handle the risk of high cost. For example, the authors in [13] develop a CCP model with risk-averse objective, using CVaR risk measure, for a network design problem under demand uncertainty. There are some studies that consider a two stage stochastic programming (2SP) model which consists of decision-making in two stages and the observation of some random event taking place in between. The first stage decision must be made before the realization of the random parameter. While, the second stage decisions are made as recourse actions based on the realized value random parameter. For example, a chance-constrained 2SP model is studied for a network design problem in [41] considering supply uncertainty. Similar chance-constrained 2SP models has been studied in [12, 30, 34, 58].

However, the numerical processing of chance constraints is computationally very hard. The first difficulty in handling CCPs is that calculating the joint probability may require a multidimensional integration. A commonly used approach to handle this issue is sample average approximation (SAA). SAA offers a promising solution to mitigate these challenges by studying an approximation of the optimization problem with probabilistic constraints in which the original distribution of the underlying random vector is replaced with an empirical distribution obtained from a random sample [5, 7, 32, 38]. Consistency results show that the optimal value and optimal solutions of the approximate problem converge to their optimal deterministic counterparts as the sample size goes to infinity. SAA allows us to reformulate CCPs as deterministic optimization problems that can be solved using standard optimization techniques. The deterministic reformulation of CCPs that result from the SAA approach can be reformulated as a deterministic equivalent model by introducing big-M coefficients and adding additional binary variables. The resulting MIP formulation is weak in general as it introduces big-M constraints. However, it can be strengthened by adding valid inequalities obtained from the so-called mixing set substructure [28, 33]. Combining decomposition-based methods and cutting plane techniques, [33] propose a branch-and-cut decomposition algorithm for solving a two-stage chance-constrained program, where the recourse decisions incur no additional costs.

The second difficulty of solving CCPs is due to the nonconvexity of the feasible region of CCP. There has been a great deal of work in developing tractable methods to construct a good feasible solution. One such method is inner approximation approach which deals with developing a deterministic safe convex approximation, i.e., a tractable convex set that is an inner approximation to the chance constraint set. A chance constraint is equivalent to a VaR constraint. Since CVaR is an approximation of VaR, we may consider an approximation scheme for chance constraint by replacing it with a CVaR constraint [39, 42]. An obvious advantage of using CVaR approximation is that the constraint function is convex. Another inner approximation approach is Bernstein approximation [39]. On the other hand, another family of approximation approach involves developing a tractable convex set that is an outer approximation of the feasible region of CCP. One such approach is relaxations obtained by a specific dualization of CCP [1]. CCP can be reformulated into an equivalent formulation by making copies of the decision variables and making them equal. These constraints are known as nonanticipativity constraints. A dual problem is then developed after relaxing the nonanticipativity constraints with Lagrangian multipliers. The resulting dual problem can be decomposed and solved efficiently to obtain lower bound on the optimal value of the original CCP model. Recently, a hybrid of the two approximation approaches, called inner-outer approximation approach, studied in [15, 16, 17] under which two smooth parametric optimization problems are defined whose feasible sets converge to the feasible set of CCP from inside and outside, respectively.

A critical assumption when modelling chance constraints is that the decision makers have perfect knowledge of the probability distribution $\mathbb{P}$. In practice, however, obtaining the true distribution and generating scenarios for approximations is difficult, making this assumption unrealistic and possibly leading to suboptimal policies. To safeguard the chance constraint against uncertainties stemming from variations in the probability distribution, an effective strategy is to embrace a distributionally robust approach. Distributionally robust chance-constrained programming (DRCCP) incorporates a family of potential distributions of the uncertainty, known as the ambiguity set, into the problem and solves it in a computationally tractable way.

The DRCCP framework provides decision makers with the flexibility to customize the size of the ambiguity set. This allows them to express their risk tolerance and incorporate their confidence level regarding the underlying uncertainty. In cases where the ambiguity set is limited to a single distribution or includes all possible distributions, DRCCP produces results that are similar to those obtained through scenario-based stochastic programming and robust optimization, respectively. However, by selecting an ambiguity set size between these two extremes, decision makers can adopt a risk attitude that falls in between. This approach balances the conservatism of robust optimization and the suboptimal out-of-sample performance of stochastic optimization, while leveraging their respective advantages.

The vast literature on DRCCP can be segregated into two primary categories based on the type of ambiguity sets considered. Moment-based ambiguity sets rely on specific moment-based properties of probability distributions. The family of distributions in this category include: distributions conforming to prescribed bounds on first- and second-order moments [6, 50, 52, 59], distributions that meet mean absolute deviation moment constraints [20], distributions that are based on marginal moment constraints [18]. Statistical distance-based ambiguity sets are defined by the statistical proximity of distributions to an empirical distribution. Various metrics and measures are employed to quantify the distance between distributions within ambiguity sets, including: ϕ-divergence [24, 26, 53], Wasserstein distance [10, 25, 26, 48, 51]. Recently, DRCCP is emerging as a popular approach to handle uncertainty in supply chain related problems. For example, a DRCCP model for a network design problem is studied in [55] considering moment based ambiguity set and in [27] considering moment based ambiguity set and Wasserstein distance ambiguity set. A DRCCP model with Wasserstein ambiguity set is considered for a facility location and capacity planning problem in [47].

Within the domain of DRCCP, a plethora of solution approaches has emerged to address the challenges of modeling distributional uncertainty. These approaches can be broadly categorized into two groups of methods: *approximate* and *exact*. A survey on reformulations and approximation approaches to handle CCPs and DRCCPs is reviewed in [29]. Approximate methods are solution approaches that offer computational tractability but potentially sub-optimal solutions. These include reformulation of DRCCP using CVaR [23, 59], or utilizing Bonferroni inequality to develop conservative approximate conic reformulations [9, 39, 49].

Exact methods rigorously resolve DRCCP models, resulting in optimal solutions. These methods include either using the dual form of the worst-case expectation constraint with indicator function or utilizing the structural information of the worst-case probability distribution to reformulate the distributionally robust chance constraint with Wasserstein metric [10, 48]. The resulting reformulation reduces to a MILP formulation when the Wasserstein ambiguity set is constructed via l_1-norm or l_∞-norm and can be solved using any branch and cut algorithm. The exact reformulation of DRCCP over Wasserstein ambiguity sets in [10] has been further studied to develop tighter inequalities [21, 22, 56]. In the case of employing a ϕ-divergence metric, authors in [26] propose a method that transforms distributionally robust chance constraints, characterized by ϕ-divergence from the empirical distribution, into standard chance constraints based on empirical distributions. This conversion enables practitioners to adjust the tolerance level using a perturbed bound on the probability of violation and to utilize regular chance-constrained techniques for solving the problem. A DRCCP with moment-based ambiguity set can be reformulated, using worst-case CVaR, into an mixed-integer conic programming model with second-order cone and semi-definite constraints [27]. The past studies on DRCCPs focus on developing a solution approach for specific ambiguity sets. However, to the best of our knowledge, [46] consider a general ambiguity set under finite support. They study a binary integer program for a chance-constrained assignment problem and the distributionally robust counterpart. They formulate the chance constraints as binary linear constraints using big-M coefficients. In an alternative approach, they reformulate the chance constraints as binary bilinear formulation and develop lifted cover inequalities as feasibility cuts. They also develop a branch-and-cut solution approach with probability cuts, derived from distribution separation problem, to efficiently solve the distributionally robust chance-constrained model. In our study, we consider a general polyhedral ambiguity set under finite support, thus also allowing the framework to be applied to many possible definitions of ambiguity sets. We used the Wasserstein ambiguity set, total variation distance ambiguity set, and moment-based ambiguity set as specific examples in our computations. However, we consider a two-stage stochastic problem with recourse. We develop a big-M free mixed integer formulation and solve the model using a branch and cut algorithm with custom feasibility cuts and probability cuts.

In particular, our main contributions are summarized as below:

- We show that the distributionally robust chance constraint is mixed integer representable with big-M coefficients and additional binary variables. We then derive a deterministic equivalent reformulation of the problem.

- We show that the DRCCP can be represented in a new big-M free mixed integer reformulation, which can be solved with a standard branch and cut algorithm. We develop an algorithm based on the L-shaped method to solve the reformulation.

- We develop custom *feasibility* cuts based on the works in [31]. Taking advantage of the knapsack constraint, we further develop a cut separation scheme for a generic polyhedral ambiguity set. Moreover, in order to take care of the distributionally robust chance constraint, we develop custom *probability* cuts.

- We consider an integrated production and distribution planning problem for a generic supply chain over a single period and conduct a comparative computational study to assess the performance of our proposed decomposition-based algorithm with the mixed integer reformulation solved directly on an off-the-shelf solver.

2 Problem Definition

We investigate a generic DRCCP represented by the optimization problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}}\Big\{\boldsymbol{c}^\top\boldsymbol{x} \mid \min_{\mathbb{P}\in\mathcal{D}}\mathbb{P}\big\{\boldsymbol{x}\in\mathcal{S}_\omega\big\}\geq 1-\epsilon\Big\}\,, \tag{1}$$

where $\boldsymbol{x}\in\mathbb{R}^n$ is the decision variables vector and $\boldsymbol{c}\in\mathbb{R}^n$ denotes the objective function coefficients. We define $\mathcal{S}_\omega$ as the set $\{\boldsymbol{x}\mid \exists\,\boldsymbol{y}: T\boldsymbol{x}+B\boldsymbol{y}\geq \boldsymbol{r}_\omega,\ \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}\geq 0\}$, where Ω is the set of scenarios for the random parameter $\boldsymbol{r}$ with any generic element of the set denoted by ω, and $\mathbb{P}$ represents the probability distribution on Ω. The likelihood of the outcome $\boldsymbol{x}\notin\mathcal{S}_\omega$, over all probability distributions in the set $\mathcal{D}$, is restricted to be less than the given risk tolerance $\epsilon\in(0,1)$. We make several key assumptions throughout the paper:

(A1) Only the right-hand side vector $\boldsymbol{r}$ is random; the technology matrix T is deterministic.

(A2) The random vector $\boldsymbol{r}$ follows a distribution with finite support.

(A3) $\mathcal{S}_\omega$, $\omega\in\Omega$, is a nonempty polyhedral set and have the same recession cone for all $\omega\in\Omega$.

(A4) The ambiguity set $\mathcal{D}$ is a polyhedral set with a generic form $\mathcal{D}=\{\boldsymbol{p}|\boldsymbol{A}\boldsymbol{p}\leq\boldsymbol{b},\ \boldsymbol{p}\geq\boldsymbol{0}\}$.

Since Ω is finite, we use a vector $\boldsymbol{p}=\{p_1,p_2,\ldots,p_{|\Omega|}\}\in\mathbb{R}^{|\Omega|}$ to denote a probability distribution vector.

2.1 Polyhedral Ambiguity Set

In this section, we provide examples of the construction of few polyhedral ambiguity sets considered in our paper. Note that the ambiguity sets are constructed while considering a nominal probability distribution $\boldsymbol{q}$. A Wasserstein distance ambiguity set is represented as

$$\mathcal{D}=\Big\{\boldsymbol{p}\geq\boldsymbol{0}:\sum_{\omega\in\Omega}p_\omega=1,\mathcal{W}(\boldsymbol{p},\boldsymbol{q})\leq\lambda,\Big\}, \tag{2}$$

where the Wasserstein radius is denoted as $\lambda\geq 0$ and Wasserstein distance is defined as

$$\min_{\boldsymbol{p},\boldsymbol{\pi}}\Big\{\boldsymbol{c}^\top\boldsymbol{\pi}\mid\sum_{\omega\in\Omega}\pi_{\omega\omega'}=p_\omega,\sum_{\omega'\in\Omega}\pi_{\omega\omega'}=q_\omega,\boldsymbol{\pi}\geq\boldsymbol{0},\boldsymbol{p}\geq\boldsymbol{0}\Big\}. \tag{3}$$

Next, an example of the total variation ambiguity set is represented as

$$\mathcal{D}=\Big\{\boldsymbol{p}\geq\boldsymbol{0}:\sum_{\omega\in\Omega}p_\omega=1,\frac{1}{2}\sum_{\omega\in\Omega}|p_\omega-q_\omega|\leq\lambda\Big\}, \tag{4}$$

where the total variation radius is denoted as $\lambda\geq 0$. Finally, we provide an example of a moment-based ambiguity set as

$$\begin{aligned}\mathcal{P}=\Big\{\boldsymbol{p}\geq\boldsymbol{0}:(1-\gamma_1)\boldsymbol{\mu}\leq\sum_{\omega=1}^{N}p_\omega\boldsymbol{r}_\omega\leq(1+\gamma_1)\boldsymbol{\mu},\\ \boldsymbol{\mu}^2-\gamma_2\boldsymbol{\sigma}^2\leq\sum_{\omega=1}^{N}p_\omega\boldsymbol{r}_\omega^2\leq\boldsymbol{\mu}^2+\gamma_2\boldsymbol{\sigma}^2\Big\},\end{aligned} \tag{5}$$

where γ_1 and γ_2 denote the metrics on first-order and second-order moments, and $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ denote the mean and variance of the random parameter $\boldsymbol{r}$. Here, $\boldsymbol{\mu}^2$ and $\boldsymbol{\sigma}^2$ are pointwise squared.

2.2 Reformulation

Problem (1) can be reformulated as

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z}} \quad & \boldsymbol{c}^\top \boldsymbol{x} && && (6a)\\
\text{s.t.} \quad & \beta_\omega = 0 \Rightarrow \boldsymbol{x} \in \mathcal{S}_\omega, && \forall\, \omega \in \Omega, && (6b)\\
& \max_{\boldsymbol{p}\in\mathcal{D}} \sum_{\omega\in\Omega} p_\omega \beta_\omega \le \epsilon, && && (6c)\\
& \boldsymbol{x} \in \mathcal{X}, && && (6d)\\
& \boldsymbol{\beta} \in \{0,1\}^{|\Omega|}, && && (6e)
\end{aligned}$$

where β_ω is a binary variable and $\beta_\omega = 0$ implies $x \in \mathcal{S}_\omega$.

A trivial approach to solve the problem (6) is to reformulate the constraint (6b) using "big-M" constraints. Using additional scenario-dependent variables $\boldsymbol{y}_\omega, \omega \in \Omega$, we formulate problem (6) as

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \quad & \boldsymbol{c}^\top \boldsymbol{x} && && (7a)\\
\text{s.t.} \quad & \boldsymbol{T}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_\omega + \beta_\omega M_\omega \ge \boldsymbol{r}_\omega, && \forall\, \omega \in \Omega, && (7b)\\
& \max_{\boldsymbol{p}\in\mathcal{D}} \sum_{\omega\in\Omega} p_\omega \beta_\omega \le \epsilon, && && (7c)\\
& \boldsymbol{x} \in \mathcal{X}, && && (7d)\\
& \boldsymbol{\beta} \in \{0,1\}^{|\Omega|}, && && (7e)\\
& \boldsymbol{y}_\omega \ge 0, && \forall\, \omega \in \Omega, && (7f)
\end{aligned}$$

where $M_\omega \ge 0, \omega \in \Omega$ are sufficiently large to ensure that when $\beta_\omega = 1$, constraints (7b) are not active. On the other hand, when $\beta_\omega = 0$, constraints (7b) enforce $\boldsymbol{x} \in \mathcal{S}_\omega$.

Given that $\mathcal{D}$ is a polyhedral set, using linear programming (LP) duality, $\max_{\boldsymbol{p}\in\mathcal{D}} \sum_{\omega\in\Omega} p_\omega \beta_\omega$ in constraint (7c) is equal to $\min_{\boldsymbol{\pi}\ge 0}\{\boldsymbol{\pi}^\top \boldsymbol{b} : \boldsymbol{\pi}^\top \boldsymbol{A} \ge \boldsymbol{\beta}\}$. Hence, problem (7) can be reformulated as a mixed-integer LP formulation (refer Proposition 2), referred to as Deterministic Equivalent Formulation (DEF), as

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{\pi}} \quad & \boldsymbol{c}^\top \boldsymbol{x} && (8a)\\
\text{s.t.} \quad & (7b), (7d) - (7f) && (8b)\\
& \boldsymbol{\pi}^\top \boldsymbol{b} \le \epsilon, && (8c)\\
& \boldsymbol{\pi}^\top \boldsymbol{A} \ge \boldsymbol{z}, && (8d)\\
& \boldsymbol{\pi} \ge \boldsymbol{0}, && (8e)
\end{aligned}$$

3 Solution Methodology

A major issue in using "big-M" constraints approach is the possibility of weak lower bounds. Moreover, when $|\Omega|$ is large, we get a very large mixed-integer program due to the introduction of the scenario-dependent $\boldsymbol{y}^\omega$ variables. In order to avoid these issues, we develop decomposition-based algorithms to solve the problem (6). We intend to develop a decomposition algorithm based on a first-stage problem that includes the original variables $\boldsymbol{x}$, and the binary variables $\boldsymbol{\beta}$. Our goal is to enforce the constraint (6b) and (6c) by adding strong cutting planes. We formulate a relaxed mixed-binary linear program (LP) for first-stage problem containing only decision variables $(\boldsymbol{x}, \boldsymbol{\beta})$ as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\beta}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & (\boldsymbol{x}, \boldsymbol{\beta}) \in \widetilde{\mathcal{F}}, \\ & (\boldsymbol{p}, \boldsymbol{\beta}) \in \widetilde{\mathcal{P}}. \end{aligned} \tag{9}$$

In order to avoid situations that do not respect the distributionally robust chance constraint (DRCC), we iteratively form a convex polyhedral outer approximation $\widetilde{\mathcal{F}}$, referred to as the set of *feasibility cuts*, for the following set

$$\mathcal{F} = \big\{(\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{X} \times \{0,1\}^{|\Omega|} : (6\text{b}), (6\text{c})\big\}, \tag{10}$$

and $\widetilde{\mathcal{P}}$, referred to as the set of *probability cuts*, for the following set

$$\mathcal{P} = \big\{(\boldsymbol{p}, \boldsymbol{\beta}) \in \mathcal{D} \times \{0,1\}^{|\Omega|} : (6\text{c})\big\}. \tag{11}$$

Given the sets $\mathcal{F}$ and $\mathcal{P}$, problem (6) can be written as

$$\min_{(\boldsymbol{x}.\boldsymbol{\beta}) \in \mathcal{X} \times \{0,1\}^{|\Omega|}} \{\boldsymbol{c}^\top \boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{F},\ (\boldsymbol{p}, \boldsymbol{\beta}) \in \mathcal{P}\}. \tag{12}$$

We now describe how to iteratively generate feasibility and probability cuts, given an optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\beta}})$ to problem (9).

3.1 Feasibility Cut

Proposition 1. *Consider an optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\beta}})$ to problem (9). Suppose that $\hat{\beta}_\omega = 0$ for some $\omega \in \Omega$, where $\hat{\boldsymbol{x}} \notin \mathcal{S}_\omega$ and $\boldsymbol{\sigma}$ is an extreme ray of $\{\boldsymbol{\sigma} : \boldsymbol{B}\boldsymbol{\sigma} \leq \boldsymbol{0}, \boldsymbol{\sigma} \geq \boldsymbol{0}\}$. Let*

$$v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T}) = \min\{\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{r}_\omega,\ \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \geq \boldsymbol{0}\}, \tag{13}$$

and ζ be a permutation of Ω describing a nonincreasing order of $v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T})$, $\omega \in \Omega$. Then, the following inequality is valid for $\mathcal{F}$, defined in (10),

$$\begin{aligned} & \boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} + \sum_{i=1}^{l} \Big(v_{\kappa_i}(\boldsymbol{\sigma}^\top \boldsymbol{T}) - v_{\kappa_{i+1}}(\boldsymbol{\sigma}^\top \boldsymbol{T})\Big)\beta_{\kappa_i} \geq v_{\kappa_1}(\boldsymbol{\sigma}^\top \boldsymbol{T}), \\ & \forall \mathcal{N} = \{\kappa_1, \ldots, \kappa_l\} \subseteq \{\zeta_1, \ldots, \zeta_{j^*}\}, \end{aligned} \tag{14}$$

where $j^* = \max\{j : \max_{\boldsymbol{p} \in \mathcal{D}} \sum_{\omega=1}^{j} p^\omega \leq \epsilon\}$, $v_{\kappa_i}(\boldsymbol{\sigma}^\top \boldsymbol{T}) \geq v_{\kappa_{i+1}}(\boldsymbol{\sigma}^\top \boldsymbol{T})$, $i \in \{1, \ldots, l\}$, and $v_{\kappa_{l+1}}(\boldsymbol{\sigma}^\top \boldsymbol{T}) = v_{\zeta_{j^*+1}}(\boldsymbol{\sigma}^\top \boldsymbol{T})$. Moreover, the most violated cut among (14) is provided by $\mathcal{N} = \{\zeta_1\}$ if $\hat{\beta}_{\zeta_1} = 0$, and $\mathcal{N} = \{\zeta_1, \zeta_m\}$ if $\hat{\beta}_{\zeta_1} \neq 0$, with $m \in \min\{2, \ldots, j^*\}$ such that $\hat{\beta}_{\zeta_m} = 0$.

Proof. Consider scenario ω. From LP duality, $\boldsymbol{x} \in \mathcal{S}_\omega$ when $\hat{\beta}^\omega = 0$ if

$$\boldsymbol{\sigma}^\top (\boldsymbol{r}^\omega - \boldsymbol{T}\boldsymbol{x}) \leq 0. \tag{15}$$

Inequality (15) is of the form $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} \geq \boldsymbol{\mu}_\omega$, with $\boldsymbol{\mu}_\omega := \boldsymbol{\sigma}^\top \boldsymbol{r}_\omega$, and is valid for $\mathcal{S}_\omega$. Moreover, $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{d} \geq 0$ for all $\boldsymbol{d} \in \mathcal{C}$ where $\mathcal{C}$ is the recession cone of $\mathcal{S}_\omega$, $\omega \in \Omega$, defined as $\mathcal{C} := \{\boldsymbol{d} :\ \boldsymbol{x} + \varsigma \boldsymbol{d} \in \mathcal{S}_\omega,\ \forall \boldsymbol{x} \in \mathcal{S}_\omega,\ \varsigma \geq 0\}$. This implies that $\boldsymbol{\sigma}^\top \boldsymbol{T} \in \mathcal{C}^*$, where $\mathcal{C}^*$ is the dual cone of $\mathcal{C}$, defined as

$\mathcal{C}^* := \{\boldsymbol{\rho} : \boldsymbol{\rho}^\top \boldsymbol{d} \geq 0,\ \forall \boldsymbol{d} \in \mathcal{C}\}$. By contradiction, suppose that $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{d} < 0$ for some $\boldsymbol{d} \in \mathcal{C}$. Given that $\boldsymbol{x} + \varsigma \boldsymbol{d} \in \mathcal{S}_\omega$ for all $\varsigma \geq 0$, we cannot find $\boldsymbol{\mu}_\omega$ such that $\boldsymbol{\sigma}^\top \boldsymbol{T}(\boldsymbol{x} + \varsigma \boldsymbol{d}) \geq \boldsymbol{\mu}_\omega$ for all $\varsigma \geq 0$. Hence, $\boldsymbol{\sigma}^\top \boldsymbol{T} \in \mathcal{C}^*$ when $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq \boldsymbol{\mu}_\omega$ is valid for $\mathcal{S}_\omega$.

Next, we show that the inequality of the form

$$\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + \boldsymbol{\sigma}^\top \boldsymbol{r}_\omega \beta_\omega \geq \boldsymbol{\sigma}^\top \boldsymbol{r}_\omega, \tag{16}$$

which is valid for $\{(\boldsymbol{x}, \beta_\omega) \in \mathcal{X} \times \{0,1\} : \beta_\omega = 0 \Rightarrow \boldsymbol{x} \in \mathcal{S}_\omega\}$ for all $\omega \in \Omega$ can be derived from the valid inequality of the form $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq \boldsymbol{\sigma}^\top \boldsymbol{r}_\omega$ for $\mathcal{S}_\omega$. Observe that from (16), we have $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq \boldsymbol{\mu}_\omega$ when $\beta_\omega = 0$ and $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq 0$ when $\beta_\omega = 1$. To show that the inequality (16) holds for all $\boldsymbol{x} \in \mathcal{X}$, we recall that for all $\omega \in \Omega$, we have the same recession cone $\mathcal{C}$, given as $\mathcal{C} = \{\boldsymbol{d} : \boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{d} \geq 0, \boldsymbol{\sigma} \in \Sigma\}$, where Σ is the pointed cone $\Sigma := \{\boldsymbol{\sigma} : \boldsymbol{\sigma}^\top \boldsymbol{B} \leq 0, \boldsymbol{\sigma} \geq \boldsymbol{0}\}$ and is same for all $\omega \in \Omega$. Therefore, for all $\boldsymbol{d} \in \mathcal{C}$, $\boldsymbol{\sigma}^\top \boldsymbol{T} \in \mathcal{C}^*$. Now, the fact that $\boldsymbol{\sigma}^\top \boldsymbol{T} \in \mathcal{C}^*$ and $\boldsymbol{x} \geq \boldsymbol{0}$, implies that for all $\omega \in \Omega$, (13) has a finite and positive optimal value, suggesting that $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq 0,\ \forall\ \boldsymbol{x} \in \mathcal{X}$. Putting these together and noting that $v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T})$, defined in (13), gives a lower bound on $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x},\ \boldsymbol{x} \in \mathcal{X}$, we conclude the validity of $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T})\beta_\omega \geq v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T})$ for $\{(\boldsymbol{x}, \beta_\omega) \in \mathcal{X} \times \{0,1\} : \beta_\omega = 0 \Rightarrow \boldsymbol{x} \in \mathcal{S}_\omega\}$.

Then, we form a mixing set to prove that (14) is valid for $\mathcal{F}$. The above discussion leads us to consider the generic set

$$\mathcal{G}' = \Big\{(\boldsymbol{y}, \boldsymbol{\beta}) \in \mathbb{R} \times \{0,1\}^{|\Omega|} : \boldsymbol{y} + h_i \beta_i \geq h_i,\ i = 1, \ldots, |\Omega|\Big\},$$

by setting $\boldsymbol{y} = \boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x}$ and $h_i = v_i(\boldsymbol{\sigma}^\top \boldsymbol{T})$ for each i. We assume that $h_1 \geq h_2 \geq \ldots \geq h_{|\Omega|}$ and form the star inequalities of the form

$$y + \sum_{i=1}^{l} (h_{t_i} - h_{t_{i+1}})\beta_i \geq h_{t_1},\ \{t_1, t_2, \ldots, t_l\} \subseteq \Omega, \tag{17}$$

where $t_1 \leq t_2 \leq \ldots \leq t_l$, and (17) is facet-defining for $\mathcal{G}'$ when $t_1 = 1$, and are sufficient to define the convex hull of $\mathcal{G}'$ [33, Theorem 2].

We tighten these inequalities in (17) for $\mathcal{G}'$ by using the knapsack constraint $\max_{\boldsymbol{p} \in \mathcal{D}} \sum_{\omega \in \Omega} p_\omega \beta_\omega \leq \epsilon$. Let $j^* = \max\{l : \max_{\boldsymbol{p} \in \mathcal{D}} \sum_{i=1}^{l} p_i \leq \epsilon\}$. Observe that that there has to be at least one $i \in \{1, \ldots, j^* + 1\}$ such that $\beta_i = 0$ and thus we have $y \geq h_i \geq h_{j^*+1}$. Thus, the constraints in $\mathcal{G}'$ redundant for $i = j^* + 1, \ldots, |\Omega|$ is redundant. We can now replace the inequalities in $\mathcal{G}'$ with

$$y + (h_i - h_{j^*+1})\beta_i \geq h_i,\ i = 1, \ldots, j^*.$$

We now prove the last part of the proposition. For simplicity in the notation, let $v_\omega := v_\omega(\boldsymbol{\sigma}^\top \boldsymbol{T})$, $\omega \in \Omega$. The inequalities in (14) are facet-defining for the convex hull of

$$\mathcal{G}' = \Big\{(\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x}, \boldsymbol{\beta}) \in \mathbb{R} \times \{0,1\}^{j^*} : \boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + (v_{\zeta_i} - v_{\zeta_{j^*+1}})\beta_{\zeta_i} \geq v_{\zeta_i},\ i = 1, \ldots, j^*\Big\},$$

if and only if $v_{\kappa_1} = v_{\zeta_1}$, suggesting that $\mathcal{N}$ should always include ζ_1. In particular, we claim that if $\hat{\beta}_{\zeta_1} = 0$, then $\mathcal{N} = \{\kappa_1\}$, where $\kappa_1 = \zeta_1$, provides the most violated cut as $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + (v_{\zeta_1} - v_{\zeta_{j^*+1}})\beta_{\zeta_1} \geq v_{\zeta_1}$. Consider an arbitrary point $(\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{G}'$. Observe that when $\beta_{\zeta_1} = 0$, $\mathcal{N}$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq v_{\kappa_1}$. By contradiction, suppose that $\mathcal{N}' = \{\zeta_1, \zeta_k\}$ where $k \leq j^*$, with $\kappa_1 = \zeta_1$ and $\kappa_2 = \zeta_k$, provides a stronger cut; that is, $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + (v_{\zeta_1} - v_{\zeta_k})\beta_{\zeta_1} + (v_{\zeta_k} - v_{\zeta_{j^*+1}})\beta_{\zeta_k} \geq v_{\zeta_1}$. When $\beta_{\zeta_1} = 0$, $\mathcal{N}'$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} + (v_{\zeta_k} - v_{\zeta_{j^*+1}})\beta_{\zeta_k} \geq v_{\zeta_1}$. Now, if $\beta_{\zeta_k} = 0$, $\mathcal{N}'$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq v_{\zeta_1}$, as in the case of $\mathcal{N} = \{\zeta_1\}$. However, when $\beta_{\zeta_k} = 1$, $\mathcal{N}'$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T} \boldsymbol{x} \geq v_{\zeta_1} - (v_{\zeta_k} - v_{\zeta_{j^*+1}})$. Because

$v_{\zeta_1} \geq v_{\zeta_1} - (v_{\zeta_K} - v_{\zeta_{j^*+1}})$, the distance of the point $(\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x}, \boldsymbol{\beta})$ and hyperplane $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} - v_{\zeta_1} = 0$ is less than or equal to that of the point and hyperplane $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} - v_{\zeta_1} + (v_{\zeta_1} - v_{\zeta_{j^*+1}}) = 0$. As point $(\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{G}'$ is arbitrary, this contradicts that $\mathcal{N}' = \{\zeta_1, \zeta_k\}$ provides a stronger cut for $\mathcal{G}'$. By induction, we can similarly argue that $\mathcal{N}' = \{\kappa_1, \ldots, \kappa_l\} \subseteq \{\zeta_1, \ldots, \zeta_{j^*}\}$ with $\kappa_1 = \zeta_1$ among the class of valid inequalities (14) leads to a weaker cut than when $\mathcal{N} = \{\zeta_1\}$. This proves that $\mathcal{N} = \{\zeta_1\}$ provides the most violated cut for $\mathcal{G}'$.

Let us now consider the case that $\hat{\beta}_{\zeta_1} = 1$. We claim that in this case $\mathcal{N} = \{\kappa_1, \kappa_2\}$ with $\kappa_1 = \zeta_1$ and $\kappa_2 = \zeta_m$ where $m \in \min\{2, \ldots, j^*\}$ such that $\hat{\beta}_{\zeta_m} = 0$, provides the most violated cut as $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} + (v_{\zeta_1} - v_{\zeta_m})\beta_{\zeta_1} + (v_{\zeta_m} - v_{j^*+1})\beta_{\zeta_m} \geq v_{\zeta_1}$. Consider an arbitrary point $(\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{G}'$. Observe that when $\beta_{\zeta_1} = 1$ and $\beta_{\zeta_m} = 0$, $\mathcal{N}$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} \geq v_{\zeta_m}$. By contradiction, suppose that $\mathcal{N}' = \{\kappa_1\}$, where $\kappa_1 = \zeta_1$, provides a stronger cut. When $\beta_{\zeta_1} = 1$, $\mathcal{N}'$ leads to $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} \geq v_{\zeta_{j^*+1}}$. Because $v_{\zeta_m} \geq v_{\zeta_{j^*+1}}$, the distance of the point $(\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x}, \boldsymbol{\beta})$ and hyperplane $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} - v_{\zeta_m} = 0$ is less than or equal to that of the point and hyperplane $\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x} - v_{\zeta_{j^*+1}} = 0$. As point $(\boldsymbol{\sigma}^\top \boldsymbol{T}\boldsymbol{x}, \boldsymbol{\beta}) \in \mathcal{G}'$ is arbitrary, this contradicts that $\mathcal{N}' = \{\zeta_1\}$ provides a stronger cut for $\mathcal{G}'$. With a similar argument as in the case $\hat{\beta}_{\zeta_1} = 0$, it can be shown that any other set $\mathcal{N}'$, with two or more elements, provides a weaker cut than when $\mathcal{N}$ is chosen. This completes the proof. □

Proposition 2. *Index j^* in Proposition 1 to generate the most violated cut can be calculated as the optimal value of the following problem*

$$\max_{\boldsymbol{\pi}, \boldsymbol{z}} \{\sum_{\omega \in \Omega} z^\omega : \boldsymbol{\pi}^\top \boldsymbol{z} \leq \epsilon, \boldsymbol{\pi} \in \mathcal{D}'(\boldsymbol{z}), \boldsymbol{z} \in \{0,1\}^{|\Omega|}\}, \tag{18}$$

where $\boldsymbol{\pi}$ is an extreme points of the dual of the feasible region $\mathcal{D} = \{\boldsymbol{A}\boldsymbol{p} \leq \boldsymbol{b},\ \boldsymbol{p} \geq \boldsymbol{0}\}$ and $\mathcal{D}'(\boldsymbol{z}) = \{\boldsymbol{\pi} : \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{z}, \boldsymbol{\pi} \geq \boldsymbol{0}\}$.

Proof. Recall j^* from Proposition 1 which can be written as $j^* = \max_{\boldsymbol{z}}\{\sum_{\omega \in \Omega} z^\omega : \max_{\boldsymbol{p} \in \mathcal{D}} \sum_{\omega \in \Omega} \boldsymbol{p}^\top \boldsymbol{z} \leq \epsilon, \boldsymbol{z} \in \{0,1\}^{|\Omega|}\}$. From LP duality, we have $j^* = \max_{\boldsymbol{z} \in \mathcal{D}^*} h(\boldsymbol{z})$ where $\mathcal{D}^* = \{\boldsymbol{z} : \min_{\boldsymbol{\pi} \in \mathcal{G}'(\boldsymbol{z})} \boldsymbol{\pi}^\top \boldsymbol{z} \leq \epsilon, \boldsymbol{z} \in \{0,1\}^{|\Omega|}\}$ and $h(\boldsymbol{z}) = \sum_{\omega \in \Omega} z_\omega$. Let $\widetilde{j}$ be an optimal value of the problem (18). Then, we can represent $\widetilde{j} = \max_{(\boldsymbol{\pi}, \boldsymbol{z}) \in \widetilde{\mathcal{D}}} h(\boldsymbol{z})$ where $\widetilde{\mathcal{D}} = \{(\boldsymbol{\pi}, \boldsymbol{z}) : \boldsymbol{\pi}^\top \boldsymbol{z} \leq \epsilon, \boldsymbol{\pi} \in \mathcal{D}'(\boldsymbol{z}), \boldsymbol{z} \in \{0,1\}^{|\Omega|}\}$. We will show that $j^* = \widetilde{j}$.

First, we will look into feasibility context. Suppose that $\hat{\boldsymbol{z}} \in \mathcal{D}^*$. This implies that $\min_{\boldsymbol{\pi} \in \mathcal{D}'(\hat{\boldsymbol{z}})} \boldsymbol{\pi}^\top \hat{\boldsymbol{z}} \leq \epsilon$. In other words, $\exists$ *a* $\boldsymbol{\pi}$ s.t. $\boldsymbol{\pi}^\top \hat{\boldsymbol{z}} \leq \epsilon, \boldsymbol{A}\boldsymbol{\pi} \leq \hat{\boldsymbol{z}}, \boldsymbol{\pi} \geq 0$. This implies that if $\hat{\boldsymbol{z}} \in \mathcal{D}^*$, $\exists$ *a* $\boldsymbol{\pi} \in \mathcal{D}'(\hat{\boldsymbol{z}})$ s.t. $(\boldsymbol{\pi}, \hat{\boldsymbol{z}}) \in \widetilde{\mathcal{D}}$. Next, let $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{z}}) \in \widetilde{\mathcal{D}}$. In other words, $\hat{\boldsymbol{\pi}}^\top \hat{\boldsymbol{z}} \leq \epsilon$ and $\hat{\boldsymbol{\pi}} \in \widetilde{\mathcal{D}}(\hat{\boldsymbol{z}})$. This means $\min_{\boldsymbol{\pi} \in \widetilde{\mathcal{D}}(\hat{\boldsymbol{z}})} \boldsymbol{\pi}^\top \hat{\boldsymbol{z}} \leq \hat{\boldsymbol{\pi}}^\top \hat{\boldsymbol{z}} \leq \epsilon$. Therefore, if $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{z}}) \in \widetilde{\mathcal{D}}$, $\hat{\boldsymbol{z}} \in \mathcal{D}^*$.

Next, we will look into optimality context. Suppose $j^* = h(\boldsymbol{z}^*)$, where $\boldsymbol{z}^*$ is an optimal solution to $\max_{\boldsymbol{z} \in \mathcal{D}^*} h(\boldsymbol{z})$. Also, let $\widetilde{j} = h(\widetilde{z})$, where $(\widetilde{\boldsymbol{\pi}}, \widetilde{z})$ is an optimal solution to $\max_{(\boldsymbol{\pi}, \boldsymbol{z}) \in \widetilde{\mathcal{D}}} h(\boldsymbol{z})$. As per our previous argument, if $(\widetilde{\boldsymbol{\pi}}, \widetilde{z}) \in \widetilde{\mathcal{D}}$, $\widetilde{z} \in \mathcal{D}^*$. However, since $\boldsymbol{z}^*$ is an optimal solution, we have $h(\boldsymbol{z}^*) \geq h(\widetilde{\boldsymbol{z}})$. Moreover, if $\boldsymbol{z}^* \in \mathcal{D}^*$, $\exists$ *a* $\boldsymbol{\pi} \in \mathcal{D}'(\boldsymbol{z}^*)$ s.t. $(\boldsymbol{\pi}, \boldsymbol{z}^*) \in \widetilde{\mathcal{D}}$. However, since $\widetilde{\boldsymbol{z}}$ is an optimal solution, we have $h(\widetilde{z}) \geq h(\boldsymbol{z}^*)$. Therefore, $h(\widetilde{z}) = h(\boldsymbol{z}^*)$, i.e., $j^* = \widetilde{j}$. □

3.2 Probability Cuts

Proposition 3. *Consider an optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}})$ to* (9)*. Suppose that for any $\omega \in \Omega$, if $\hat{z}_\omega = 0$ then $\hat{\boldsymbol{x}} \in \mathcal{S}_\omega$. Then the following inequality is valid for $\mathcal{P}$, defined in* (11)*,*

$$\sum_{\omega \in \Omega} {p^*}_\omega \hat{z}_\omega \leq \epsilon, \tag{19}$$

where $\boldsymbol{p}^*$ is an optimal solution to $\max_{\boldsymbol{p} \in \mathcal{D}} \sum_{\omega \in \Omega} p_\omega \hat{z}_\omega$.

Proof. Consider an arbitrary point $(\boldsymbol{p}^*, \hat{\boldsymbol{z}}) \in \mathcal{P}$; that is, $\sum_{\omega\in\Omega} {p^*}_\omega \hat{z}_\omega = \max_{\boldsymbol{p}\in\mathcal{D}} \sum_{\omega\in\Omega} p_\omega \hat{z}_\omega \leq \epsilon$. Because $\hat{\boldsymbol{p}} \in \mathcal{D}$, then $(\boldsymbol{p}^*, \hat{\boldsymbol{z}}) \in \mathcal{P}$ satisfies $\sum_{\omega\in\Omega} \hat{p}_\omega \hat{z}_\omega \leq \epsilon$, i.e., $\sum_{\omega\in\Omega} \hat{p}_\omega \hat{z}_\omega \leq \sum_{\omega\in\Omega} {p^*}_\omega \hat{z}_\omega$. This completes the proof. □

We propose a decomposition-based cutting plane algorithm to solve (1) and present the pseudocode of the algorithm in Algorithm 1.

3.3 Finiteness

Theorem 1. *Algorithm 1 solves problem* (9) *in finitely many iterations.*

Proof. In order to show that the Algorithm 1 is finitely convergent, we need to show that "while" loop in Algorithm 1 terminates after finite iterations.

Suppose $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}})$ is not a feasible solution to (17), then it means that none of the previously derived feasibility cuts impose $\hat{\boldsymbol{x}} \in \{\boldsymbol{x} | \ \exists\ \boldsymbol{y} \text{ s.t. } \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{r}_\omega - \boldsymbol{T}\boldsymbol{x}\}$ when $\hat{z}_\omega = 0$ for any scenario $\omega \in \Omega$. Therefore, a new set of extreme rays $\boldsymbol{\sigma}$ from the set $\{\boldsymbol{\sigma} : \boldsymbol{B}\boldsymbol{\sigma} \leq \boldsymbol{0}, \boldsymbol{\sigma} \geq \boldsymbol{0}\}$ are obtained to derive an appropriate feasibility cut of the form (14) to cut-off the point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}})$. Since there are finite number of such extreme rays, a finite number of feasibility cuts are possible.

Given Assumption-(A-4), a finite number of probability cuts of the form (19) are possible since each cut corresponds to one of the finitely many extreme points of the ambiguity set $\mathcal{D}$. □

Algorithm 1 Decomposition-based algorithm for distributionally robust chance-constrained problems with polyhedral ambiguity set.

Initialization:
Set $\widetilde{\mathcal{F}} \leftarrow \emptyset$, $\widetilde{\mathcal{P}} \leftarrow \emptyset$, `viol-Feas` = TRUE, `viol-Prob` = TRUE.
Input: ϵ
Output: Optimal solution $\hat{\boldsymbol{x}}$ and optimal value $c^\top \hat{\boldsymbol{x}}$ of problem (9)
while (`viol-Feas` = TRUE or `viol-Prob` = TRUE) **do**
  Solve the problem (9) and obtain the optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\beta}})$.
  for $\omega \in \Omega$ **do**
    if $\hat{\beta^\omega} = 0$ **then**
      if $\hat{\boldsymbol{x}} \notin \{\boldsymbol{x} | \exists \boldsymbol{y} \in \mathbb{R}^d : T\boldsymbol{x} + B\boldsymbol{y} \geq \boldsymbol{r}_\omega, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \geq 0\}$ **then**
        Obtain an extreme ray $\boldsymbol{\sigma}$ from the cone $\{\boldsymbol{\sigma} : B\boldsymbol{\sigma} \leq 0, \boldsymbol{\sigma} \geq 0\}$.
        Calculate $v_\omega(\boldsymbol{\sigma}^\top T)$ for all $\omega \in \Omega$ defined in (13).
        Obtain a feasibility cut of the form (14) and add it to $\widetilde{\mathcal{F}}$
        `viol-Feas` = TRUE
      else
        `viol-Feas` = FALSE
      end if
    end if
  end for
  if `viol-Feas` = FALSE **then**
    Solve $z = \max_{\boldsymbol{p}\in\mathcal{D}} \boldsymbol{p}^\top \hat{\boldsymbol{\beta}}$
    if $z > \epsilon$ **then**
      Obtain a probability cut of the form (19) and add it to $\widetilde{\mathcal{P}}$.
      `viol-Prob` = TRUE
    else
      `viol-Prob` = FALSE
    end if
  end if
end while

4 Numerical Study

In this section, we do an extensive computational study (Section 4.1) and assess performance analysis (Section 4.2). The algorithms were implemented in Python in GUROBI 9.0. For our decomposition-based algorithm, we used the lazy constraint callback function of GUROBI to add the feasibility and probability cuts whenever the algorithm finds a mixed-integer incumbent candidate solution. The algorithms were executed on a single thread of high-performance computing nodes of the Palmetto Cluster with 32 cores and 125 GB of memory, with a time limit of 3600 seconds per instance and algorithm.

For our experiments, we consider an integrated production and distribution planning problem in a supply chain consisting of a set of manufacturers denoted by $i \in \mathcal{I} = \{1, ..., n\}$, and a set of retailers denoted by $j \in \mathcal{J} = \{1, ..., m\}$. We assume that each retailer represents a set of customers and therefore, the retailer demand is the sum of the demand of customers it represents. We develop a chance-constrained stochastic programming problem as follows:

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \Big\{ \boldsymbol{c}^\top \boldsymbol{x} \mid \mathbb{P}\big\{ \boldsymbol{x} \in \mathcal{S}_\omega \big\} \geq 1 - \epsilon \Big\} \ ,$$

where,

$$\mathcal{S}_\omega = \left\{ \boldsymbol{x} \geq \boldsymbol{0} | \ \exists \ \boldsymbol{y} \geq \boldsymbol{0} \text{ s.t. } \sum_{j=1}^{|\mathcal{J}|} y_{ij} \leq x_i, \ \forall i \in \mathcal{I}, \right.$$
$$\left. \sum_{i=1}^{|\mathcal{I}|} \mu_{ij} y_{ij} \geq \boldsymbol{\xi}_{j\omega}, \forall j \in \mathcal{J} \right\} .$$

Here, $\boldsymbol{\xi}_j$ represents the demand at retailer j, and $(1 - \boldsymbol{\mu}_{ij})$ represents the damage rate during transportation of products from manufacturer i to retailer j. The variables $\boldsymbol{x}$ represent the quantity of product manufactured by the manufacturer $i \in \mathcal{I}$. The variables $\boldsymbol{y}$ represent the quantity of product shipped from manufacturer $i \in \mathcal{I}$ to retailer $j \in \mathcal{J}$. A similar CCP model is studied in [31] for a resource allocation problem.

Given a probability distributions ambiguity set $\mathcal{D}$, we represent the DRCCP variant of the problem as:

$$\min_{\boldsymbol{x} \geq 0} \Big\{ \boldsymbol{c}^\top \boldsymbol{x} \mid \min_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\big\{ \boldsymbol{x} \in \mathcal{S}_\omega \big\} \geq 1 - \epsilon \Big\}$$

4.1 Test Instances

For a given problem size (number of manufacturers and retailers), we first generated a single "base instance" consisting of the unit costs of the manufacturing product and a set of base damage rates. The manufacturing cost per unit quantity c_i at each manufacturer i is chosen independently from a normal distribution with mean 1 and standard deviation 0.2. In order to generate damage rate, we first generate $\widetilde{\mu}_j$ for each retailer j independently from a normal distribution with mean 1 and standard deviation 0.2. Then, $\mu_{ij} = \min\{c_i + \widetilde{\mu}_j, 1\}$. The reasoning behind this construction is that manufacturers that charge higher cost will have better packaging and shipping methods that result in higher μ_{ij} and thus low damage rate $(1 - \mu_{ij})$.

For each base instance, we generated five independent samples of retailer demand with $|\Omega|$ realization. The random demand is generated independently from a multivariate normal distribution with mean for each retailer j generated independently from a normal distribution with mean 110 and standard deviation 25. The covariance matrix v is generated as $\frac{1.25}{|J|}\widetilde{v}^\top\widetilde{v}$, where $\widetilde{v}_{jj}$ is generated indepedently from an uniform distribution in [-6.25,25].

4.2 Computational Study

In this section, we present a computational performance to analyze the efficiency of our proposed decomposition-based algorithm, shown in Algorithm 1, and compare it with an off-the-shelf mixed-integer LP solver for the DEF.

For our experiments, we consider various instances with different levels of risk parameter $\epsilon \in \{0.05, 0.1\}$, problem size $(|I|, |J|) \in \{(10, 20), (20, 30), (30, 40)\}$, and sample size $|\Omega| \in \{200, 300, 400, 1000\}$. We consider different radii λ for total variation distance (`TV`) and Wasserstein distance (`WAS`) ambiguity sets (defined in (2) and (4), respectively) and metric γ_2 for moment-based (`MO`) ambiguity set (defined in (5)). For each combination of these parameters, we generated 5 instances.

We present the computational results for `TV` (Table 1), `WAS` (Table 2), and `MO` (Table 3) ambiguity sets. In Tables 1, 2 and 3, "`OV`" column denotes the average optimal value of the instances that were solved to optimality. The columns"`B&C`" and "`DEF`" denote the computational performance of our proposed decomposition-based algorithm and DEF solved directly in GUROBI 9.0. We use the notation "a (b,c)" to report the computational performance, where "a" indicates the average computational time of the instances that were solved to optimality, "b" shows the average relative optimality gap, and "c" shows the number of instances that could not be solved to optimality within the time limit of 3600 seconds.

The observations from Table 1 are the following:

1. The average optimal value increases with the increase in radius λ. As λ increases, the ambiguity set $\mathcal{D}$ becomes larger and includes more probability distributions making the problem more risk averse.

2. We observe that after λ is equal to ϵ, the average optimal value does not change. This implies that the ambiguity set becomes large enough to contain all the probability distributions including the worst-case distribution which is the same as the model without any chance constraints. According to Proposition 4 in [26], DRCCP with total variation distance ambiguity set can be equivalently represented as a classical CCP with a perturbed risk level $\epsilon^{'} = \epsilon - \lambda$. Therefore, when $\lambda = \epsilon$, $\epsilon^{'} = 0$, i.e., a model without chance constraints. Our observation is consistent with their study.

3. At small radius λ, the average computational time is higher. A reason for this is the more computational time spent to generate feasibility cuts. In our experiments, we observed that the average number of feasibility cuts and total time spent in generating feasibility cuts is higher when radius is small. Similar results were observed in [21, 57], and [2].

4. With increase in risk parameter ϵ, the average computational time increases. As ϵ increases, the feasible region increases and more time is spent to find an optimal solution.

5. Increase in sample size $|\Omega|$ results in increase in the average computational time. As sample size increases, more number of subproblems are needed to be solved which results in higher computational time. Observe from Table 1, when $|\Omega| = 1000$, not all instances could be solved to optimality by solving DEF with GUROBI. Whereas, our proposed decomposition-based algorithm was able to solve all the instances to optimality within the time limit.

6. Increase in problem size $(|I|, |J|)$ leads to increase in number of variables and results in higher average computational time.

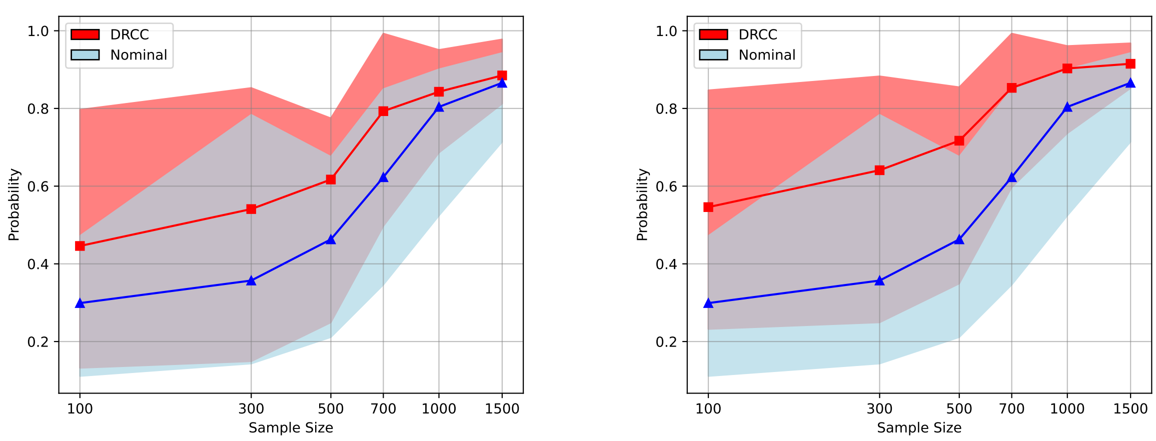
We skip the interpretation of the results for `WAS`, in Table 2, as they remain similar to those mentioned for `TV`. In Table 3, we present computational study for `MO` and vary γ_2 and observe similar observations as varying radius λ for `TV`. The observations related to varying other parameters remain similar to those mentioned for `TV`.

4.3 Out-of-sample Performance

In this section, we compare the out-of-sample performance of our DRCCP with total variation distance and Wasserstein distance ambiguity sets with the classic CCP considering risk tolerance $\epsilon = 0.1$.

In the case of `WAS`, we follow the *holdout method* procedure as described in [37] for the Wasserstein radius λ selection from different candidate radius values via cross-validation. We note that `TV` is a special case of `WAS`, where Wasserstein distance, defined in (3), is constructed with cost $c_{\omega\omega'} = 0$ if $\omega = \omega'$, and $c_{\omega\omega'} = 1$ otherwise. Therefore, we perform out-of-sample experiment for `TV` by using `WAS` with the new Wasserstein distance explained earlier. We generate random problem instances with problem size $(|I|, |J|) = (10, 20)$, and training samples of size $|\Omega| \in \{100, 300, 500, 700, 1000, 1500\}$. In the case of `TV`, we chose candidate radius $\lambda \in \{0, 0.0001, 0.001, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$. Whereas, for `WAS`, the candidate radius were chosen as $\lambda \in \{i * 0.6 : \ i \in \{0, \dots, 12\}\}$.

Figure 1a shows the probability of scenarios that are satisfied by DRCCP with total variance radius chosen selected via *holdout method* (referred to as "DRCC" in the figure) and CCP with $\epsilon = 0.1$ (referred to as "Nominal" in the figure). Our results indicate that as the number of samples increases, both DRCCP and CCP generate solutions that achieve the desired risk tolerance of 0.1, i.e., mean of the upper bounds on probability of scenarios satisfied with the generated solution converges to 90%. Moreover, in comparison to CCP with a nominal risk tolerance, DRCCP appears to result in better solutions, in terms of probability of satisfying desired safety conditions, especially when data are scarce. We skip the interpretation of the results for `WAS`, in Figure 1b, as they remain similar to those mentioned for `TV`.



(a) `TV` ambiguity set

(b) `WAS` ambiguity set

Figure 1: Out-of-sample performance for probability of satisfying safety threshold at $\epsilon = 0.1$.

Table 1: Computational performance of different solution methods in the case of TV.

$(\|I\|,\|J\|)$	ϵ	γ	200 OV	200 B&C	200 DEF	300 OV	300 B&C	300 DEF	400 OV	400 B&C	400 DEF	1000 OV	1000 B&C	1000 DEF
(10,20)	0.05	0.00	898.45	1.65	16.74	938.63	3.00	64.84	918.83	6.43	72.95	1075.59	22.79	2260.84
		0.01	907.73	1.70	55.41	954.27	2.78	86.06	923.65	5.13	110.31	1086.58	20.99	943.27
		0.02	911.18	1.34	13.21	958.37	2.96	71.32	930.32	5.61	100.03	1098.25	29.31	939.86
		0.03	917.06	1.26	27.39	990.88	2.39	34.01	933.69	4.27	134.45	1117.71	26.55	678.87
		0.04	926.94	1.25	10.62	1007.36	1.52	32.32	983.60	3.26	122.41	1142.25	15.56	342.54
		0.05	963.31	1.21	11.48	1073.74	1.78	9.21	1002.93	2.83	31.27	1259.81	9.62	239.76
		0.06	963.31	1.22	16.24	1073.74	1.81	7.12	1002.93	2.83	29.77	1259.81	8.32	236.58
	0.1	0.00	876.68	3.52	22.84	916.38	3.74	62.10	892.72	10.72	98.31	1037.01	47.07	(5,1.83)
		0.02	883.38	2.01	83.71	921.78	3.79	110.12	903.00	9.89	155.02	1050.42	40.19	2393.24(2,0.12)
		0.04	894.70	3.57	31.54	934.18	4.71	57.12	911.82	7.76	184.43	1063.42	28.78	1151.67
		0.06	904.52	1.96	55.14	951.69	2.45	40.23	923.65	4.78	197.29	1086.58	26.71	969.50
		0.08	917.05	1.49	14.89	981.36	1.62	12.22	933.69	4.52	142.92	1117.71	25.63	778.23
		0.10	963.31	1.42	16.20	1073.74	2.08	10.12	1002.93	3.64	33.18	1259.81	10.07	351.19
		0.12	963.31	1.43	21.82	1073.74	1.98	11.34	1002.93	3.57	32.23	1259.81	11.62	249.11
(20,30)	0.05	0.00	1740.76	12.02	80.30	1869.70	26.45	130.12	1699.32	47.82	237.62	1768.56	243.94	(5,6.37)
		0.01	1762.84	7.29	95.05	1879.36	20.11	117.42	1707.44	38.26	179.22	1782.80	129.03	(5,5.18)
		0.02	1765.09	6.25	79.33	1882.40	18.40	112.61	1725.28	26.16	195.19	1811.43	148.56	(5, 0.46)
		0.03	1766.83	7.97	78.73	1898.89	14.86	111.17	1741.24	16.81	140.14	1858.89	77.09	2651.82
		0.04	1810.85	4.56	56.89	1935.50	6.71	97.21	1786.15	14.85	106.47	1891.05	72.79	2447.08
		0.05	1827.72	4.36	13.30	1964.65	4.74	31.57	1852.94	6.15	69.31	2073.15	28.08	2297.58
		0.06	1827.72	4.34	14.40	1964.65	4.71	30.14	1852.94	5.98	74.31	2073.15	28.12	2327.11
	0.1	0.00	1661.53	20.98	48.50	1828.16	27.43	124.17	1646.91	60.44	245.12	1703.14	362.76	(5,9.81)
		0.02	1702.08	27.39	48.62	1841.94	23.16	117.11	1669.28	54.52	112.67	1716.58	408.51	(5,8.87)
		0.04	1728.77	16.80	49.01	1863.60	26.43	114.14	1687.92	30.59	293.81	1741.51	322.86	(5,1.64)
		0.06	1761.28	6.51	99.92	1871.60	25.87	104.32	1707.44	22.19	257.12	1782.80	102.56	3245.67 (4,1.32)
		0.08	1766.83	7.76	110.63	1897.56	11.62	94.12	1741.24	19.25	121.32	1858.89	112.19	3167.78
		0.10	1827.72	4.96	24.32	1964.65	5.09	57.61	1852.94	8.16	98.14	2073.15	30.78	2968.91
		0.12	1827.72	4.65	14.57	1964.65	5.18	41.23	1852.94	8.31	89.14	2073.15	29.22	3018.72
(30,40)	0.05	0.00	2169.78	16.12	125.95	2183.76	25.50	194.25	2135.36	54.22	214.71	2048.84	330.57	(5,12.14)
		0.01	2195.52	11.78	114.33	2216.50	28.86	201.14	2155.97	44.17	204.56	2064.81	153.24	(5,10.66)
		0.02	2202.33	12.44	130.04	2224.27	25.76	257.54	2181.28	31.52	188.42	2097.16	145.79	(5,6.24)
		0.03	2219.98	9.01	112.45	2290.70	13.86	269.14	2241.99	30.78	134.78	2138.81	78.18	(5,2.82)
		0.04	2247.24	8.57	72.42	2324.10	9.89	197.14	2320.32	16.39	112.32	2215.84	69.74	(5,0.64)
		0.05	2371.71	7.49	51.28	2555.65	7.04	110.01	2519.78	16.25	87.64	2358.38	44.94	2913.44
		0.06	2371.71	7.56	49.77	2555.65	7.93	64.17	2519.78	14.38	84.57	2358.38	44.28	2889.17
	0.1	0.00	2084.95	16.77	116.00	2113.49	37.91	205.41	2068.58	72.21	486.29	1965.42	485.75	(5,18.02)
		0.02	2119.38	16.12	109.89	2169.51	33.76	224.31	2073.33	60.46	414.25	1997.08	431.81	(5,16.48)
		0.04	2144.68	13.41	144.22	2196.26	30.51	275.14	2081.18	38.67	349.21	2036.96	349.26	(5,11.76)
		0.06	2195.52	9.86	115.91	2256.06	17.14	241.12	2100.29	28.39	193.28	2064.81	149.26	(5,7.74)
		0.08	2219.58	8.95	83.81	2294.23	10.62	191.01	2132.06	21.25	101.01	2138.52	56.34	(5,4.32)
		0.10	2371.71	7.98	68.76	2555.65	9.84	145.21	2519.78	12.09	97.47	2358.38	45.35	(5,2.12)
		0.12	2371.71	7.91	65.74	2555.65	9.25	77.24	2519.78	12.89	91.14	2358.38	44.82	(5,1.86)

Table 2: Computational performance of different solution methods in the case of `WAS`.

$(\lvert I\rvert, \lvert J\rvert)$	ϵ	γ	N=200 OV	N=200 B&C	N=200 DEF	N=300 OV	N=300 B&C	N=300 DEF	N=400 OV	N=400 B&C	N=400 DEF	N=1000 OV	N=1000 B&C	N=1000 DEF
(10,20)	0.05	0.00	898.45	1.25	1.81	938.63	3.10	5.11	918.83	6.93	25.32	1075.59	176.41	2576.52
		1.20	919.28	1.02	3.01	964.47	2.78	5.47	923.65	6.18	16.87	1114.58	147.09	1921.48
		2.40	950.32	1.04	2.78	978.61	2.96	8.31	958.82	6.11	16.21	1155.27	98.41	1553.62
		3.60	963.31	1.06	2.51	1001.21	2.39	7.59	993.19	5.29	14.59	1204.88	29.02	812.36
		4.80	963.31	0.98	1.21	1073.74	1.52	5.64	1002.93	4.16	12.98	1204.88	26.13	683.12
		6.00	963.31	0.96	1.35	1073.74	1.78	5.13	1002.93	3.87	11.14	1259.81	23.24	291.65
		7.20	963.31	0.97	1.42	1073.74	1.81	5.19	1002.93	3.87	11.07	1259.81	22.11	272.28
	0.1	0.00	876.68	3.62	3.74	916.38	7.89	25.84	901.52	9.16	28.12	1037.01	828.61	3263.71(4,1.68)
		2.40	923.18	2.74	3.79	942.13	6.81	18.39	913.23	10.09	32.47	1082.46	479.69	3178.52(2,1.46)
		4.80	935.43	3.21	4.71	991.06	10.16	22.32	931.62	8.36	21.43	1125.13	301.51	2484.71
		7.20	984.12	2.04	2.45	1007.43	6.65	21.12	963.37	5.71	14.27	1173.77	213.26	1875.67
		9.60	963.31	1.56	1.62	1050.94	3.26	8.12	978.71	5.22	15.01	1204.88	85.57	1139.22
		12.0	963.31	1.52	2.08	1073.74	3.57	7.65	1002.93	4.11	14.63	1259.81	62.48	513.26
		14.40	963.31	1.57	1.98	1073.74	2.19	7.86	1002.93	4.17	14.55	1259.81	61.83	489.75
(20,30)	0.05	0.00	1740.76	17.42	31.12	1869.70	21.25	61.24	1699.32	27.82	87.14	1768.56	562.62	(5,9.22)
		1.20	1765.09	11.14	28.17	1879.36	19.21	55.41	1726.59	19.67	61.24	1807.19	524.63	(5,7.45)
		2.48	1810.85	13.42	33.47	1883.76	14.25	42.98	1777.16	15.16	57.25	1864.64	462.66	(5, 2.75)
		3.60	1827.72	6.23	14.01	1931.76	18.26	33.58	1815.34	19.81	44.21	1897.00	319.63	3364.81(1,0.32)
		4.80	1827.72	2.75	10.18	1964.65	10.13	29.14	1852.94	10.85	33.26	1941.53	198.72	2978.69
		6.00	1827.72	2.61	9.87	1964.65	4.75	11.09	1852.94	6.95	18.15	2027.72	62.52	2629.42
		7.20	1827.72	2.67	9.92	1964.65	4.71	9.86	1852.94	6.98	17.98	2073.15	46.74	2517.82
	0.1	0.00	1661.53	21.98	64.24	1828.16	34.43	201.47	1646.91	33.62	377.36	1703.14	1363.62	(5,14.22)
		2.48	1728.77	18.08	50.19	1863.93	27.86	287.14	1692.49	24.52	326.12	1756.12	1048.72	(5,11.68)
		4.80	1765.09	12.19	41.32	1897.54	25.83	142.13	1732.94	25.59	365.34	1815.31	893.16	(5,10.32)
		7.20	1814.43	7.19	157.04	1912.04	13.88	112.64	1786.26	15.19	284.56	1866.53	625.86	(5,6.28)
		9.60	1827.72	6.03	83.54	1964.65	9.39	70.14	1815.43	11.25	206.34	1902.34	312.68	3327.63(4,3.84)
		12.0	1827.72	3.15	45.12	1964.65	7.35	60.14	1852.94	8.96	112.15	1931.73	198.63	3116.83
		14.40	1827.72	3.27	49.57	1964.65	7.88	60.16	1852.94	8.33	71.24	1964.92	195.47	3112.53
(30,40)	0.05	0.00	2169.78	23.41	66.31	2183.76	28.25	186.32	2135.36	31.82	716.47	2064.81	1368.53	(5,13.66)
		1.20	2199.87	14.11	51.18	2210.37	19.78	201.46	2149.78	29.67	646.22	2082.19	1014.67	(5,10.22)
		2.48	2214.06	17.82	57.84	2245.79	16.55	159.04	2189.16	20.16	531.78	2113.56	916.54	(5,7.66)
		3.60	2229.62	8.13	36.25	2296.16	18.21	164.27	2245.14	21.61	347.85	2196.54	724.86	(5,8.28)
		4.80	2247.24	7.15	33.28	2323.32	10.74	88.74	2310.19	10.24	223.78	2258.93	338.56	(5,6.72)
		6.00	2371.71	7.41	34.10	2555.65	6.67	71.74	2519.78	8.25	159.67	2427.83	204.52	(5,4.42)
		7.20	2371.71	7.57	34.16	2555.65	6.74	59.46	2519.78	8.93	113.96	2463.61	197.65	(5,1.16)
	0.1	0.00	2084.95	22.18	97.49	2113.49	39.11	214.17	2068.58	54.44	1372.64	1965.42	2248.76	(5,27.56)
		2.48	2115.11	19.21	119.78	2164.26	27.86	264.78	2078.89	44.52	1187.41	2011.76	1854.92	(5,22.84)
		4.80	2136.79	16.69	165.47	2183.78	25.83	314.78	2083.94	40.59	780.14	2087.46	1537.98	(5,17.16)
		7.20	2165.43	8.10	178.98	2224.24	18.88	305.19	2096.26	35.19	664.23	2154.98	1466.84	(5,13.42)
		9.60	2197.72	7.54	91.47	2291.55	12.39	156.94	2165.43	25.25	401.47	2236.17	974.47	(5,11.28)
		12.0	2371.71	5.61	86.47	2555.65	10.35	124.78	2519.78	14.96	206.14	2396.58	516.82	(5,8.48)
		14.40	2371.71	5.17	66.71	2555.65	10.88	89.87	2519.78	13.33	214.18	2463.61	422.16	(5,6.56)

Table 3: Computational performance of different solution methods in the case of MO at $\gamma_1 = 0$.

$(\lvert I\rvert, \lvert J\rvert)$	ϵ	γ_2	Sample Size (N)											
			200			300			400			1000		
			OV	B&C	DEF	OV	B&C	DEF	OV	B&C	DEF	OV	B&C	DEF
(10,20)	0.05	0.6	1228.15	2.51	6.35	1254.83	4.78	12.11	1253.93	5.93	18.72	1204.88	26.82	117.15
		0.7	1257.26	2.32	5.91	1309.17	3.71	16.32	1256.25	6.18	17.32	1204.88	26.56	97.97
		0.8	1257.26	2.39	4.57	1309.17	4.76	11.19	1256.25	6.11	15.48	1204.88	25.61	87.33
		0.9	1257.26	3.06	6.41	1309.17	4.39	14.48	1256.25	5.29	16.28	1259.81	23.53	50.25
	0.1	0.7	1190.07	6.26	13.25	1213.58	15.52	48.17	1201.46	39.16	117.58	1155.29	154.52	380.20
		0.8	1192.37	4.29	14.88	1250.53	9.01	32.11	1215.07	18.29	37.12	1168.15	101.71	221.51
		0.9	1192.37	4.71	5.67	1254.82	4.78	15.21	1240.76	8.36	28.71	1185.38	67.78	131.69
		1.0	1228.15	4.54	7.71	1254.83	4.60	12.71	1240.76	8.71	21.14	1194.65	39.44	57.04
(20,30)	0.05	0.6	1766.83	7.05	37.84	1964.35	8.20	46.22	1978.83	9.03	121.01	1941.53	113.84	1174.18
		0.7	1814.43	6.92	29.14	1964.48	6.78	76.04	1985.65	10.11	81.04	2027.72	45.88	612.64
		0.8	1814.43	7.18	32.71	1964.65	7.96	32.17	1985.82	8.11	35.46	2027.72	41.75	398.24
		0.9	1827.72	5.89	24.78	1964.65	6.39	18.19	1985.82	7.19	22.14	2073.15	38.28	180.51
	0.1	0.7	1728.77	39.72	187.45	1912.02	98.74	641.12	1935.18	145.16	723.14	1899.26	243.68	2016.45(1,1.4)
		0.8	1766.83	13.54	89.14	1935.50	34.59	221.34	1941.23	60.19	240.78	1949.68	120.16	1321.24
		0.9	1810.85	10.32	75.34	1960.15	20.71	147.34	1951.62	58.36	189.14	1962.72	88.63	302.19
		1.0	1814.43	9.14	65.66	1964.35	17.45	97.11	1951.62	55.71	201.11	1985.89	65.67	298.85
(30,40)	0.05	0.6	2213.38	8.45	51.10	2336.42	41.01	197.81	2331.24	155.93	583.17	2128.95	712.35	3126.25(3, 2.97)
		0.7	2247.24	9.62	39.89	2457.14	22.78	143.37	2445.65	86.18	445.17	2178.71	501.54	2078.61
		0.8	2263.20	7.74	30.16	2555.78	12.96	120.91	2564.82	76.11	189.11	2178.71	448.96	1979.61
		0.9	2371.71	7.96	28.19	2555.78	12.39	101.32	2564.82	55.29	147.14	2315.82	245.56	1256.43
	0.1	0.7	2188.55	78.63	321.36	2316.12	113.74	671.91	2330.46	209.16	1146.05	2085.12	1218.56	(5, 2.18)
		0.8	2219.98	53.77	257.84	2336.27	253.61	1123.14	2417.18	170.09	858.05	2134.16	864.34	2198.65
		0.9	2247.13	28.47	201.14	2336.27	58.41	127.42	2545.64	78.36	442.14	2134.16	671.34	1986.81
		1.0	2247.24	14.82	151.14	2555.65	34.12	96.12	2545.64	65.63	341.05	2218.16	301.12	1456.62

5 Conclusion and Future Research

We investigate distributionally robust chance-constrained programming with polyhedral ambiguity sets in which randomness appeared only in the right-hand side of the constraints. We consider three types of ambiguity sets, i.e., total variation distance, Wasserstein distance and moment-based. To solve the distributionally robust chance-constrained two-stage stochastic program, we propose a decomposition-based algorithm based on L-shaped method. We conducted extensive numerical experiments to evaluate the computational efficiency of our proposed solution approach with the equivalent mixed-integer linear program formulation solved with an off-the-shelf solver. We also discuss how to choose a proper Wasserstein radius via holdout method and draw insights from out-of-sample analysis considering total variation distance and Wasserstein distance ambiguity sets.

For future research, our proposed model and solution approach can find application in many real-world network design problems. Moreover, in our study, we employ a synthetically generated data. For future study, real historical data from the financial market can be considered.

References

[1] Ahmed, S. and W. Xie (2018). Relaxations and approximations of chance constraints under finite distributions. *Mathematical Programming 170*, 43–65.

[2] Arrigo, A., C. Ordoudis, J. Kazempour, Z. De Grève, J.-F. Toubeau, and F. Vallée (2022). Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation. *European Journal of Operational Research 296*(1), 304–322.

[3] Awudu, I. and J. Zhang (2013). Stochastic production planning for a biofuel supply chain under demand and price uncertainties. *Applied Energy 103*, 189–196.

[4] Bilsel, R. U. and A. Ravindran (2011). A multiobjective chance constrained programming model for supplier selection under uncertainty. *Transportation Research Part B: Methodological 45*(8), 1284–1300.

[5] Calafiore, G. and M. C. Campi (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming 102*, 25–46.

[6] Calafiore, G. C. and L. E. Ghaoui (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications 130*, 1–22.

[7] Campi, M. C. and S. Garatti (2008). The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization 19*(3), 1211–1230.

[8] Charnes, A., W. W. Cooper, and G. H. Symonds (1958). Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science 4*(3), 235–263.

[9] Chen, W., M. Sim, J. Sun, and C.-P. Teo (2010). From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research 58*(2), 470–485.

[10] Chen, Z., D. Kuhn, and W. Wiesemann (2024). Data-driven chance constrained programs over wasserstein balls. *Operations Research 72*(1), 410–424.

[11] Christopher, M. (2000). The agile supply chain: competing in volatile markets. *Industrial marketing management 29*(1), 37–44.

[12] Elçi, Ö. and N. Noyan (2018). A chance-constrained two-stage stochastic programming model for humanitarian relief network design. *Transportation research part B: methodological 108*, 55–83.

[13] Elçi, Ö., N. Noyan, and K. Bülbül (2018). Chance-constrained stochastic programming under variable reliability levels with an application to humanitarian relief network design. *Computers & Operations Research 96*, 91–107.

[14] Fisher, M. L. et al. (1997). What is the right supply chain for your product? *Harvard business review 75*, 105–117.

[15] Geletu, A., A. Hoffmann, M. Kloppel, and P. Li (2017). An inner-outer approximation approach to chance constrained optimization. *SIAM Journal on Optimization 27*(3), 1834–1857.

[16] Geletu, A., A. Hoffmann, and P. Li (2019). Analytic approximation and differentiability of joint chance constraints. *Optimization 68*(10), 1985–2023.

[17] Geletu, A., M. Klöppel, A. Hoffmann, and P. Li (2015). A tractable approximation of non-convex chance constrained optimization with non-gaussian uncertainties. *Engineering Optimization 47*(4), 495–520.

[18] Ghosal, S. and W. Wiesemann (2020). The distributionally robust chance-constrained vehicle routing problem. *Operations Research 68*(3), 716–732.

[19] Gicquel, C. and J. Cheng (2018). A joint chance-constrained programming approach for the single-item capacitated lot-sizing problem with stochastic demand. *Annals of Operations Research 264*, 123–155.

[20] Hanasusanto, G. A., V. Roitch, D. Kuhn, and W. Wiesemann (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research 65*(3), 751–767.

[21] Ho-Nguyen, N., F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee (2021). Distributionally robust chance-constrained programs with right-hand side uncertainty under wasserstein ambiguity. *Mathematical Programming*, 1–32.

[22] Ho-Nguyen, N., F. Kilinç-Karzan, S. Küçükyavuz, and D. Lee (2023). Strong formulations for distributionally robust chance-constrained programs with left-hand side uncertainty under wasserstein ambiguity. *INFORMS Journal on Optimization 5*(2), 211–232.

[23] Hota, A. R., A. Cherukuri, and J. Lygeros (2019). Data-driven chance constrained optimization under wasserstein ambiguity sets. In *2019 American Control Conference (ACC)*, pp. 1501–1506. IEEE.

[24] Hu, Z., L. J. Hong, and A. M.-C. So (2013). Ambiguous probabilistic programs. *Available at Optimization Online*.

[25] Ji, R. and M. A. Lejeune (2021). Data-driven distributionally robust chance-constrained optimization with wasserstein metric. *Journal of Global Optimization 79*(4), 779–811.

[26] Jiang, R. and Y. Guan (2016). Data-driven chance constrained stochastic program. *Mathematical Programming 158*(1-2), 291–327.

[27] Jiang, Z., R. Ji, and Z. S. Dong (2023). A distributionally robust chance-constrained model for humanitarian relief network design. *OR Spectrum 45*(4), 1153–1195.

[28] Küçükyavuz, S. (2012). On mixing sets arising in chance-constrained programming. *Mathematical programming 132*(1-2), 31–56.

[29] Küçükyavuz, S. and R. Jiang (2021). Chance-constrained optimization: A review of mixed-integer conic formulations and applications. *arXiv preprint arXiv:2101.08746*, 2.

[30] Li, Y., J. Shu, M. Song, J. Zhang, and H. Zheng (2017). Multisourcing supply network design: two-stage chance-constrained model, tractable approximations, and computational results. *INFORMS Journal on Computing 29*(2), 287–300.

[31] Luedtke, J. (2014). A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming 146*(1-2), 219–244.

[32] Luedtke, J. and S. Ahmed (2008). A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization 19*(2), 674–699.

[33] Luedtke, J., S. Ahmed, and G. L. Nemhauser (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming 122*(2), 247–272.

[34] Meng, L., X. Wang, J. He, C. Han, and S. Hu (2023). A two-stage chance constrained stochastic programming model for emergency supply distribution considering dynamic uncertainty. *Transportation Research Part E: Logistics and Transportation Review 179*, 103296. https://doi.org/10.1016/j.tre.2023.103296.

[35] Miller, B. L. and H. M. Wagner (1965). Chance constrained programming with joint constraints. *Operations Research 13*(6), 930–945.

[36] Mitra, K., R. D. Gudi, S. C. Patwardhan, and G. Sardar (2008). Midterm supply chain planning under uncertainty: A multiobjective chance constrained programming framework. *Industrial & Engineering Chemistry Research 47*(15), 5501–5511.

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

[38] Nemirovski, A. and A. Shapiro (2006). Scenario approximations of chance constraints. *Probabilistic and randomized methods for design under uncertainty*, 3–47.

[39] Nemirovski, A. and A. Shapiro (2007). Convex approximations of chance constrained programs. *SIAM Journal on Optimization 17*(4), 969–996.

[40] Prekopa, A. (1973). Contributions to the theory of stochastic programming. *Mathematical Programming 4*, 202–221.

[41] Quddus, M. A., S. Chowdhury, M. Marufuzzaman, F. Yu, and L. Bian (2018). A two-stage chance-constrained stochastic programming model for a bio-fuel supply chain network. *International Journal of Production Economics 195*, 27–44.

[42] Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of risk 2*, 21–42.

[43] Ryan, T. (2015). Retailers suffer the high cost of overstocks and out-ofstocks. *Retail Wire*. https://retailwire.com/discussion/retailers-suffer-the-high-cost-of-overstocks-and-out-of-stocks/.

[44] Sabath, R. (1995). Volatile demand calls for quick response: the integrated supplychain. *Logistics Information Management 8*(2), 49–52.

[45] Sun, H., Z. Gao, W. Y. Szeto, J. Long, and F. Zhao (2014). A distributionally robust joint chance constrained optimization model for the dynamic network design problem under demand uncertainty. *Networks and Spatial Economics 14*(3), 409–433.

[46] Wang, S., J. Li, and S. Mehrotra (2022). A solution approach to distributionally robust joint-chance-constrained assignment problems. *INFORMS Journal on Optimization 4*(2), 125–147.

[47] Wang, Z., K. You, Z. Wang, and K. Liu (2021). Multi-period facility location and capacity planning under ∞-wasserstein joint chance constraints in humanitarian logistics. *arXiv preprint arXiv:2111.15057*.

[48] Xie, W. (2021). On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming 186*(1-2), 115–155.

[49] Xie, W., S. Ahmed, and R. Jiang (2022). Optimized bonferroni approximations of distributionally robust joint chance constraints. *Mathematical Programming 191*(1), 79–112.

[50] Xu, H., C. Caramanis, and S. Mannor (2012). Optimization under probabilistic envelope constraints. *Operations Research 60*(3), 682–699.

[51] Yang, I. (2020). Wasserstein distributionally robust stochastic control: A data-driven approach. *IEEE Transactions on Automatic Control 66*(8), 3863–3870.

[52] Yang, W. and H. Xu (2016). Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming 155*, 231–265.

[53] Yanıkoğlu, İ. and D. den Hertog (2013). Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing 25*(4), 666–681.

[54] You, F. and I. E. Grossmann (2008). Design of responsive supply chains under demand uncertainty. *Computers & Chemical Engineering 32*(12), 3090–3111.

[55] Zang, Y., M. Wang, H. Liu, and M. Qi (2023). Moment-based distributionally robust joint chance constrained optimization for service network design under demand uncertainty. *Optimization and Engineering*, 1–53. https://doi.org/10.1007/s11081-023-09858-0.

[56] Zhang, Y. and J. Dong (2022). Building load control using distributionally robust chance-constrained programs with right-hand side uncertainty and the risk-adjustable variants. *INFORMS Journal on Computing 34*(3), 1531–1547.

[57] Zhao, S. and K. Zhang (2020). A distributionally robust stochastic optimization-based model predictive control with distributionally robust chance constraints for cooperative adaptive cruise control under uncertain traffic conditions. *Transportation Research Part B: Methodological 138*, 144–178.

[58] Zhao, Y., Q. Xue, Z. Cao, and X. Zhang (2018). A two-stage chance constrained approach with application to stochastic intermodal service network design problems. *Journal of Advanced Transportation 2018*, 1–18.

[59] Zymler, S., D. Kuhn, and B. Rustem (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming 137*, 167–198.