

RESEARCH ARTICLE

Robust reformulations of ambiguous chance constraints with discrete probability distributions

İhsan Yanıkoğlu 

Department of Industrial Engineering, Özyegin University, İstanbul, Turkey
ihsan.yanikoglu@ozyegin.edu.tr

ARTICLE INFO

Article History:

Received 04 May 2018

Accepted 06 May 2019

Available 31 July 2019

Keywords:

Robust optimization

Chance constraint

Ambiguous chance constraint

AMS Classification 2010:

90C11; 90C15; 90C35

ABSTRACT

This paper proposes robust reformulations of ambiguous chance constraints when the underlying family of distributions is discrete and supported in a so-called “p-box” or “p-ellipsoidal” uncertainty set. Using the robust optimization paradigm, the deterministic counterparts of the ambiguous chance constraints are reformulated as mixed-integer programming problems which can be tackled by commercial solvers for moderate sized instances. For larger sized instances, we propose a safe approximation algorithm that is computationally efficient and yields high quality solutions. The associated approach and the algorithm can be easily extended to joint chance constraints, nonlinear inequalities, and dependent data without introducing additional mathematical optimization complexity to that of the original robust reformulation. In numerical experiments, we first present our approach over a toy-sized chance constrained knapsack problem. Then, we compare optimality and computational performances of the safe approximation algorithm with those of the exact and the randomized approaches for larger sized instances via Monte Carlo simulation.



1. Introduction

Chance constrained optimization is introduced by [1] (also see, [2–4]) and it ensures that the probability of satisfying an uncertain constraint is greater than or equal to a certain threshold while minimizing or maximizing a given objective function. A *chance constraint* is given by

$$\Pr_{\zeta \sim \mathbb{P}} \{ \zeta : f(\mathbf{x}, \zeta) \leq 0 \} \geq 1 - \epsilon, \quad (1)$$

where $f(\mathbf{x}, \zeta)$ denotes a function of a decision vector $\mathbf{x} \in \mathbf{X} \subseteq \mathbb{R}^n$ (\mathbf{X} is the set of feasible decisions) and a vector of uncertainty parameters $\zeta \in \mathbb{R}^L$, $\epsilon \in [0, 1]$ is the predetermined probability threshold, and \mathbb{P} is the known probability distribution of ζ . It can be shown that the feasible set of (1) is convex when ζ follows a Gaussian distribution and $f(\mathbf{x}, \zeta)$ is linear in \mathbf{x} and ζ ; see [5–7]. Additional convexity results for (1) can be shown when $f(\mathbf{x}, \zeta)$ is additively separable and ζ follows a log-concave distribution ([8, 9]). Even though the chance constraint is tractable for the above

mentioned problem classes, it is generally computationally intractable because the feasible set of the chance constraint is non-convex or it is computationally intractable to compute the left-hand side (LHS) of the constraint even when the feasible set is convex. In the latter case, one may use a Monte Carlo simulation to check the feasibility of the chance constraint, nevertheless, the simulation approach can also be too costly at high accuracies.

The chance constraint approach can be extended to multiple constraints, i.e., referred to as the joint chance constraint:

$$\Pr_{\zeta \sim \mathbb{P}} \{ \zeta : f_k(\mathbf{x}, \zeta) \leq 0 \quad \forall k \in K \} \geq 1 - \epsilon. \quad (2)$$

Notice that (2) is at least as computationally challenging as (1).

Computationally tractable safe approximations of the chance constraint have been proposed to overcome the difficulties that are mentioned above. A safe approximation method replaces the chance

constraint with a set of constraints that yields a solution set that is a subset of the feasible set of the chance constraint. Nemirovski and Shapiro [10] propose a computationally tractable approximation of a chance constrained problem where constraints are affine in the uncertainty parameters that are independent with known support. Ben-Tal and Nemirovski [11] translate the existing stochastic uncertainties to ‘uncertain-but-bounded’ sets under mild assumptions. Namely, a feasible solution $\mathbf{x} \in \mathbf{X}$ for the tractable reformulation of the safe approximation

$$f(\mathbf{x}, \boldsymbol{\zeta}) \leq 0 \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \quad (3)$$

satisfies the chance constraint (1) with at least $1 - \epsilon$ probability where \mathcal{Z} denotes a ‘bounded’ uncertainty set.

The accuracies of the obtained approximations in [10, 11] are good when the number of uncertainty parameters is high. Associated approximations are famous because they yield tractable robust reformulations for (1) using modern *robust optimization* (RO) techniques. We shall delve into details of these RO methods in the next section. Chen et al. [12] propose a conservative approximation of a joint chance constraint in terms of a worst-case conditional value-at-risk (CVaR) constraint. The resulting approximation outperforms the Bonferroni approximation that is known to be pessimistic ([10, 12–14]). Zymler et al. [14] propose a method for approximating joint chance constraints when the first- and second-order moments together with the support of the uncertainty parameter are known. Similar to the safe approximations methods that are mentioned above, extension of our approach to the joint chance constraint is also straightforward; later in §3, the associated extension shall be mentioned. Calafiore and Campi [15] and Campi and Garatti [16] substitute the chance constraint with a finite number of constraints which are randomly sampled from the original constraint according to the known distribution of the uncertainty parameter. The authors show that the resulting solution that comes out of the randomized approach fails to satisfy the chance constraint with the given confidence level provided that a sufficient number of samples is drawn.

In practice, we usually have partial or no information on the probability distribution \mathbb{P} , since it needs to be estimated from historical data. This is why it makes sense to pass to *ambiguous* chance constraint. The term ‘ambiguous’ stands for the uncertainty in the probability distribution. In other words, the distribution of the uncertainty

parameters is itself uncertain ([25]). The ambiguous chance constraint can be formulated as follows

$$\Pr_{\boldsymbol{\zeta} \sim \mathbb{P}} \{ \boldsymbol{\zeta} : f(\mathbf{x}, \boldsymbol{\zeta}) \leq 0 \} \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P}, \quad (4)$$

where \mathbb{P} belongs to a family \mathcal{P} of distributions and the chance constraint is satisfied for all (\forall) probability distributions in \mathcal{P} . This introduces an additional computational complexity in solving the problem aside from the existing difficulties that are mentioned above. To the best of our knowledge, there is no systematic and exact way of solving ambiguous chance constraint problems for general family classes with continuous distributions. Formulating ambiguity in the probability distribution has taken attention of scholars from different fields. In the absence of full information on the probability distribution or when only a set of possible distributions \mathcal{P} is known, it is natural to optimize the expectation function corresponding to the worst-case probability distribution in \mathcal{P} . This lead to the following formulation: $[\min_{\mathbf{x} \in \mathbf{X}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \boldsymbol{\zeta})]]$. For more details, we refer to [17–20]. Moreover, ambiguity in the probability distribution is also addressed by [21–23] in economics.

Ambiguity in the context of the chance constrained optimization has been studied by [10, 24, 25]. It is important to point out that the ambiguous chance constraint is ‘severely’ intractable compared to the regular chance constraint. Good news is that, as it is pointed out by [26, 27], robust reformulation methods for chance constraints can be straightforwardly extended to the ambiguous chance constraints and this is why the adjective ‘ambiguous’ is generally skipped. Erdoğan and Iyengar [24] define the distribution family \mathcal{P} in (4) using the Prohorov metric. The authors propose a robust sampled problem that is a good approximation for the associated ambiguous chance constrained problem with high probability. Yanıkoğlu et al. [25] propose an iterative algorithm that constructs the uncertainty set yielding a tractable robust counterpart that safely approximates the ambiguous chance constraint; the authors use the ϕ -divergence metric and historical data on the uncertainty parameters to define the distribution family \mathcal{P} . For further details on such approximations, we refer reader to [25–27]. In this paper, we focus on robust reformulations of ambiguous chances constraints with discrete probability distributions. More precisely, \mathbb{P} corresponds to a discrete probability distribution, i.e., we have a finite set of scenarios for uncertainty realizations $\{\boldsymbol{\zeta}^1, \boldsymbol{\zeta}^2, \dots, \boldsymbol{\zeta}^{|S|}\}$ where each scenario

$s \in S$ or realization ζ^s has an individual probability $p_s \geq 0$ of being realized; needless to say all probabilities are summed to 1, i.e., $\sum_{s \in S} p_s = 1$. Finite supports are often faced in practice when data at hand is discrete, some examples are, demand, the number of customers in a queue, the amount of inventory in a production facility, the time and quantity of returning products in a reverse logistics network, the number of quality grades in remanufacturing and so on. On the other hand, empirical distributions are often used when the data at hand is scarce so that no continuous distribution can be fitted; or when the information is based on an expert opinion. Last but not least, continuous supports may be reduced to finite ones in order to yield computational tractability for the problem at hand (e.g., see [28]).

When ζ follows a discrete probability distribution, the chance constraint (1) can be *equivalently* reformulated as

$$\begin{aligned} f(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) & \forall s \in S \\ \sum_{s \in S} p_s y_s &\geq 1 - \epsilon, \end{aligned}$$

where binary variable $y_s \in \{0, 1\}$ is 1 if solution $\mathbf{x} \in \mathbf{X}$ satisfies the constraint $[f(\mathbf{x}, \zeta^s) \leq 0]$ for realization ζ^s or 0 otherwise, and M denotes a large constant; also see [29,30] that adopt the associated big-M reformulation technique and propose branch-and-bound/cut solution approach to chance constrained problems under finite support. As it is pointed out above, the probability distribution \mathbb{P} that has to be known is often not (exactly) known in practice; the probability vector $\mathbf{p} \in [0, 1]^{|S|}$ of the discrete uncertainty parameter $\zeta \in \mathbb{R}^L$ often comes from an expert opinion or a forecast, i.e., the probability vector that determines the structure of the uncertainty is in itself uncertain. This is why, working with so-called a nominal probability vector may cause a lot of problems if the associated ambiguity in the discrete probability distribution is not taken into account. The ambiguous chance constraint (4) can be equivalently reformulated as the following semi-infinite mixed-integer problem when \mathbb{P} follows a discrete probability distribution.

$$\begin{aligned} f(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) & \forall s \in S \\ \sum_{s \in S} p_s y_s &\geq 1 - \epsilon & \forall \mathbf{p} \in \mathcal{U}, \end{aligned} \quad (5)$$

where \mathcal{U} denotes the ambiguity (or uncertainty) set that supports the family of distributions \mathcal{P} in the ambiguous chance constraint. As pointed out above, (5) is a semi-infinite optimization problem,

i.e., it has finitely many decision variables and infinitely many constraints (see, $\forall \mathbf{p} \in \mathcal{U}$) that is intractable in its current form. Using the robust optimization paradigm, the tractable reformulations of (5) shall be proposed in this paper. The associated approach is exact, i.e., it equivalently reformulates the ambiguous chance constraint by exploiting the deterministic structure of the distribution, when the random perturbation in \mathbf{p} is independent; and we propose safe approximations when \mathbf{p} is dependent.

Hanasusanto et al. [31] derive explicit conic representations of ambiguous chance constraints for tractable classes and efficiently solvable conservative approximations for the intractable ones using tools from *distributionally robust optimization* (DRO). The authors, derive tractable reformulations of the ambiguous individual chance constraint when the ambiguity set is Markov. For the joint case, the tractability is obtained only for conic moment ambiguity sets. The authors propose a conservative approximation algorithm for the intractable cases which is based on improving the fixed decision at each stage of the algorithm. Hanasusanto et al. [32] study conic representable reformulations of ambiguous joint chance constraints when the mean and the support, and the upper bound on the dispersion of the uncertainty parameters are known. The authors also provide the conic representable reformulation of the optimistic chance constraint for specific classes. Jiang and Guan [33] study distributionally robust chance constraints when the family of distributions is based on a phi-divergence measure. The problems are efficiently solvable by using strong cutting planes and hence a branch-and-cut algorithm. Chen et al. [34] propose distributionally robust reformulations of data-driven chance constraints using the Wasserstein ball uncertainty set that is often used in DRO framework. The resulting RC is mixed-integer linear program that can be solved for moderate sized instances. The method can also be extended to joint chance constraints when the uncertainty is in the right-hand side (RHS). Similarly, Ji and Lejeune [35] study distributionally robust chance-constrained programming with data-driven Wasserstein ambiguity sets. They strengthen the formulations by adopting valid inequalities. The resulting RC is a mixed-integer second-order cone programming (MISOCP) reformulation for the exact model with RHS uncertainty. The authors also propose a MISOCP relaxation for models with random technology vector. Zhang et al. [36] study

distributionally robust reformulation of chance-constrained bin packing problem. Using two moments of the uncertainty, they construct the ambiguity set and the resulting robust reformulation of the problem. To strengthen second-order RC formulation, the authors adopt valid polymatroid inequalities that improves the computational performance the off-the-self solvers such as GUROBI and CPLEX for the given test instances. Cheng et al. [37] consider distributionally robust version of the quadratic knapsack problem where the knapsack constraint coincides with an ambiguous chance constraint such that the two moments of the distributions in the ambiguity sets are partly known. The resulting RC is a semi-definite programming (SDP) relaxation reformulation. The joint case is more challenging and the authors propose two tractable methods to find upper and lower bounds for the SDP relaxation. Hu and Hong [38] propose a DRO framework that adopts the Kullback-Leibler phi-diverge measure to model the ambiguity set of the unknown probability distribution. The authors show that the associated approach can also be used to reformulate ambiguous chances constraints when the confidence level of the phi-divergence measure is set to the probability threshold of the chance constraint. Xie and Ahmed [39] propose robust reformulations of ambiguous chance constrained problems when the ambiguity set is specified by convex moment constraints. They show that distributionally robust chance constrained problem can be modeled as a convex optimization problem under certain conditions. Xie et al. [40] propose a Bonferroni approximation approach to solve distributionally robust joint chance constrained programs. The author shows that the associated approximation is convex and tractable when the family of distributions is specified by moments or by marginal distributions. Xie [41] studies exact and approximate reformulations of distributionally robust chance constrained problems where the ambiguity set is determined by Wasserstein metric. The author adopts a branch and cut algorithm to solve the resulting reformulations.

Finite dimensional uncertainty parameters are often faced in regression analysis and worst-case expectation problems with Wasserstein ambiguity sets often yield efficient second-order conic formulation; we refer reader to [42] for distributionally robust regression. Similarly, Özmen et al. [43] robustify an extension of multivariate adaptive regression splines (MARS) under polyhedral uncertainty (i.e., so-called RCMARS) by adopting RO; we refer reader to [44] for MARS and to [45] for its deterministic extension, i.e., so-called CMARS.

For robust reformulations of various risk measures that are often used in financial optimization, we refer to [12, 46–48], and references therein.

The remainder of the paper is organized as follows. In §2, we first give a brief introduction to the RO paradigm (§2.1), then we propose the robust reformulations of the ambiguous chance constraints with discrete probability distributions (§2.2). In §2.3, we propose a safe approximation algorithm for the robust reformulations. In §3, we present extensions of our approach. In §4, we present numerical experiments. Finally, we give our concluding remarks and future research directions in §5.

Notation. Bold-face, lower-case letters and numbers represent vectors, e.g., $\mathbf{0}$ denotes a vector of zeros and \mathbf{e} is the all-one vector. Bold-face, upper-case letters represent matrices. The dimensions of the vectors and matrices will usually be clear from the context. Lower-case letters refer to vector elements, e.g., p_s denotes the s^{th} element of vector \mathbf{p} . A vector or matrix superscript indicates either the transpose (\top) or the element-wise power of a given vector or matrix, e.g., \mathbf{q}^{-2} and \mathbf{Q}^2 ; only exceptions are indexes of probability vectors (\mathbf{p}^0 , $\mathbf{p}^{(j)}$) and uncertainty parameter realizations (ζ^s). \mathbb{P} denotes a probability distribution and \mathcal{P} denotes a family of distributions. The uncertainty set is given by \mathcal{Z} where $\zeta \in \mathcal{Z} \subseteq \mathbb{R}^L$ denotes the uncertainty parameter and L is the number of uncertain parameters. The ambiguity set is given by \mathcal{U} where $\xi \in \mathcal{U} \subseteq \mathbb{R}^{|S|}$ denotes the perturbation vector of the probability vector $\mathbf{p} \in [0, 1]^{|S|}$ and S denotes the set of indexes for the uncertainty parameter realizations, i.e., the components of the probability vector. The subscripts B and E for the ambiguity set denote the specific properties of the set, namely, p-box (\mathcal{U}_B) and p-ellipsoidal (\mathcal{U}_E). Finally, “...” denotes that the remainder of an expression shall continue in the next line.

2. Methodology

In this section, we first present the three core steps of the RO paradigm. Later we adopt RO to ambiguous chance constraints with discrete probability distributions and we to derive the tractable robust reformulations of the ambiguous chance constraints.

2.1. Introduction to robust optimization

For the sake of simplicity, we assume that LHS of the constraint $[f(\mathbf{x}, \zeta) \leq 0]$ is a simple linear function $[f(\mathbf{x}, \zeta) = (\mathbf{a} + \mathbf{B}\zeta)^\top \mathbf{x}]$ and the uncertainty parameter ζ resides in a polyhedral uncertainty set $[\mathcal{U} = \mathbf{D}\zeta + \mathbf{d} \geq 0]$, i.e., the constraint can be

formulated as follows:

$$(\mathbf{a} + \mathbf{B}\zeta)^\top \mathbf{x} \leq 0 \quad \forall \zeta : \mathbf{D}\zeta + \mathbf{d} \geq \mathbf{0}, \quad (6)$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^{n \times L}$, $\mathbf{D} \in \mathbb{R}^{m \times L}$, $\mathbf{d} \in \mathbb{R}^m$. Next, we present a three step procedure to remove the universal quantifier (\forall) in (6); the resulting deterministic equivalent reformulation is often referred to as the ‘tractable’ *robust counterpart* (RC).

Step 1: The worst-case formulation (6) can be equivalently reformulated as

$$\mathbf{a}^\top \mathbf{x} + \max_{\zeta: \mathbf{D}\zeta + \mathbf{d} \geq \mathbf{0}} (\mathbf{B}^\top \mathbf{x})^\top \zeta \leq 0. \quad (7)$$

We now have a nested optimization problem because we have a maximization inside the constraint. Next, we take the dual the inner maximization problem.

Step 2: Note that by strong duality

$$\begin{aligned} & \max \left\{ (\mathbf{B}^\top \mathbf{x})^\top \zeta : \mathbf{D}\zeta + \mathbf{d} \geq \mathbf{0} \right\} \\ & = \min \left\{ \mathbf{d}^\top \mathbf{y} : \mathbf{D}^\top \mathbf{y} = -\mathbf{B}^\top \mathbf{x}, \mathbf{y} \geq \mathbf{0} \right\}. \end{aligned}$$

Hence (7) is equivalent to

$$\mathbf{a}^\top \mathbf{x} + \min_{\mathbf{y} \geq \mathbf{0}} \{ \mathbf{d}^\top \mathbf{y} : \mathbf{D}^\top \mathbf{y} = -\mathbf{B}^\top \mathbf{x} \} \leq 0. \quad (8)$$

Step 3: We can remove the *minimization* in (8) because a feasible solution for (8) yields an upper bound for the maximization problem by weak duality (i.e., equivalent at optimality by strong duality because the problem is convex).

$$\exists \mathbf{y} : \mathbf{a}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \leq \beta, \quad \mathbf{D}^\top \mathbf{y} = -\mathbf{B}^\top \mathbf{x}, \quad \mathbf{y} \geq \mathbf{0}.$$

Finally, we obtain the tractable RC of the initial semi-infinite problem (6).

Table 1 presents the RCs of (6) with respect to a class of uncertainty sets. The only difference is that the type of duality in Step 3 changes according to the uncertainty set at hand, e.g., when the uncertainty set is a cone, we use conic duality to obtain the RC; see, the third row of the Table 1.

Table 1. The tractable robust counterparts of uncertain linear constraints

Set	\mathcal{Z}	Robust Counterpart	Tractability
Box	$\ \zeta\ _\infty \leq \rho$	$\mathbf{a}^\top \mathbf{x} + \rho \ \mathbf{B}^\top \mathbf{x}\ _1 \leq 0$	LP
Ball	$\ \zeta\ _2 \leq \rho$	$\mathbf{a}^\top \mathbf{x} + \rho \ \mathbf{B}^\top \mathbf{x}\ _2 \leq 0$	CQP
Cone	$\mathbf{D}\zeta + \mathbf{d} \in K$	$\begin{aligned} \mathbf{a}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} &\leq 0 \\ \mathbf{D}^\top \mathbf{y} &= -\mathbf{B}^\top \mathbf{x} \\ \mathbf{y} &\in K^* \end{aligned}$	Conic

Notice that the three step procedure shall be similarly applied in our setting to derive the tractable RC reformulations of the semi-infinite probability

threshold constraint:

$$\sum_{s \in S} p_s y_s \geq \epsilon \quad \forall \mathbf{p} \in \mathcal{U} \quad (\text{SI})$$

for the two different classes of the ambiguity set \mathcal{U} . We refer reader to [26, 49] and the references therein for further details on the RO paradigm.

2.2. Robust reformulation of ambiguous chance constraint

In this section, we focus on the robust reformulations of the ambiguous chance constraints for two classes of uncertainty sets, namely, ‘p-box’ and ‘p-ellipsoidal’ uncertainty sets that are specifically created for the uncertain probability vector $\mathbf{p} \in \{0, 1\}^{|S|}$ that resides in the plane sections of box and ellipsoidal regions. To this end, we shall develop the tractable RC reformulations of the semi-infinite probability constraint (SI) in

$$\begin{aligned} f(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) \quad \forall s \in S \\ \sum_{s \in S} p_s y_s &\geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}. \end{aligned} \quad (\text{re-5})$$

The formal representations of p-box (\mathcal{U}_B) and p-ellipsoidal (\mathcal{U}_E) ambiguity sets are given as

$$\begin{aligned} \mathcal{U}_B &= \left\{ \mathbf{p} = \mathbf{p}^0 + \boldsymbol{\xi} \in \mathbb{R}^{|S|} : \boldsymbol{\ell} \leq \boldsymbol{\xi} \leq \mathbf{u}, \mathbf{e}^\top \boldsymbol{\xi} = 0 \right\} \\ \text{and} \\ \mathcal{U}_E &= \left\{ \mathbf{p} = \mathbf{p}^0 + \boldsymbol{\xi} \in \mathbb{R}^{|S|} : \|\mathbf{Q}\boldsymbol{\xi}\|_2 \leq \sigma, \mathbf{e}^\top \boldsymbol{\xi} = 0 \right\}, \end{aligned}$$

where \mathbf{p}^0 denotes the nominal probability vector given that $\mathbf{p}^0 \geq \mathbf{0}$ and $\sum_{s \in S} p_s^0 = 1$; $\boldsymbol{\xi}$ is the uncertainty vector that determines the dispersion from the nominal probability vector; $\boldsymbol{\ell}$ and \mathbf{u} are the upper and lower bound vectors for the p-box; σ denotes the radius (or the worst-case dispersion from the nominal vector \mathbf{p}^0) in the p-ellipsoidal, $\mathbf{Q} \in \mathbb{R}^{|S| \times |S|} (:= \text{diag}(\mathbf{q}))$ is a diagonal matrix where vector $\mathbf{q} \in \mathbb{R}^{|S|}$ contains the scaling parameters that determine the shape of the ellipsoid, e.g., when $\mathbf{q} = \mathbf{e}$, the resulting uncertainty set is a ball. Notice that the sum of the elements of the uncertainty vector $\boldsymbol{\xi}$ is zero (i.e., $[\sum_{s \in S} \xi_s = 0]$) in order to guarantee that the sum of the elements of the unknown (or ambiguous) probability vector to be 1 (i.e., $[\sum_{s \in S} p_s = 1]$).

Independent versus dependent data. In some cases, it may be known that the uncertainty parameters ζ_j^s are independent for $j \in \{1, \dots, L\}$. If this is the case, then the joint probability p_s of the uncertainty vector $[\zeta_1^s, \zeta_2^s, \dots, \zeta_L^s]$ is being realized is equivalent to the product of the marginal probabilities of the uncertainty parameters for the given scenario, i.e., $[p_s = p_1^{(s)} \dots p_L^{(s)}]$. This results in complex ambiguity sets. We shall delve into details of such reformulations later in §3.3.

In the remainder, for the sake of simplicity, we assume that the elements of the ambiguous probability vector \mathbf{p} are independent and the uncertainty parameter (ζ^s) is realized as a vector. Other assumptions are listed below.

Assumptions. (A1) We assume that $f(\mathbf{x}, \zeta)$ is linear in the decision variable \mathbf{x} (extended in §3.2). (A2) Without loss of generality, we assume that $\mathcal{U}_{B,E} \subseteq [0, 1]^{|S|}$, i.e., the parameters $\ell, \mathbf{u}, \mathbf{Q}$, and σ are selected such that $[\mathbf{p} \geq 0 \quad \forall \mathbf{p} \in \mathcal{U}_{B,E}]$ is satisfied.

Selecting the parameters of p-ellipsoidal uncertainty set. The standard form of the ellipsoid inequality is

$$\frac{\xi_1^2}{\alpha_1^2} + \dots + \frac{\xi_{|S|}^2}{\alpha_{|S|}^2} \leq 1,$$

where α_i denotes the half length of the i^{th} principal semiaxis of the ellipsoid, i.e., $\alpha_i > 0$ for all $i \in \{1, \dots, |S|\}$. Let parameter " $\gamma_i \in (0, 1]^{|S|}$ " be the percentage of dispersion $[(1-\gamma_i)p_i^0, (1+\gamma_i)p_i^0]$ from the nominal probability p_i^0 for $i = \{1, \dots, |S|\}$, then we may set the half length α_i to the associated dispersion, i.e., $\alpha_i = \gamma_i p_i^0$ for all $i \in \{1, \dots, |S|\}$. Eventually, the p-ellipsoidal uncertainty set \mathcal{U}_E that satisfies (A2) can be generated by setting $\sigma = 1$ and

$$q_i = \frac{1}{\gamma_i p_i^0} \quad \forall i \in \{1, \dots, |S|\},$$

where $\mathbf{Q} = \text{diag}(\mathbf{q})$. Notice that one can also set σ and \mathbf{Q} using other methods as long as (A2) is satisfied; here we present a systematic method that can be easily implemented in practice.

Theorem 1 and 2 provide the tractable RCs of (SI) for the p-box and p-ellipsoidal uncertainty sets; respectively.

Theorem 1. *The vector $\mathbf{y} \in \{0, 1\}^{|S|}$ satisfies*

$$\sum_{s \in S} p_s y_s \geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}_B \quad (\text{SI}_B)$$

if and only if $\mathbf{y}, \boldsymbol{\theta}, \mathbf{w}$ and λ satisfy the following RC problem:

$$\begin{aligned} \mathbf{y}^\top \mathbf{p}^0 + \mathbf{u}^\top \mathbf{w} + \boldsymbol{\ell}^\top \boldsymbol{\theta} &\geq 1 - \epsilon \\ \lambda \mathbf{e} + \mathbf{w} + \boldsymbol{\theta} - \mathbf{y} &= \mathbf{0} \\ \boldsymbol{\theta} \geq \mathbf{0}, \mathbf{w} \leq \mathbf{0}, \lambda &\text{ urs,} \end{aligned}$$

where \mathbf{e} is the all-one vector, and $\mathbf{w}, \boldsymbol{\theta}$, and λ denote the dual variables.

Proof. As in Step 1 of the three step procedure in §2.1, (SI_B) can be equivalently reformulated as

$$\min_{\mathbf{p} \in \mathcal{U}_B} \sum_{s \in S} p_s y_s \geq 1 - \epsilon.$$

The above expression is equivalent to the following

$$\min_{\boldsymbol{\xi}} \left\{ \sum_{s \in S} \xi_s y_s : \boldsymbol{\ell} \leq \boldsymbol{\xi} \leq \mathbf{u} \right\} \geq 1 - \epsilon - \mathbf{y}^\top \mathbf{p}^0.$$

Notice that $\mathbf{y}^\top \mathbf{p}^0$ is taken out of the minimization because it is a fixed term for given \mathbf{y} . Let us now focus on the minimization problem in the constraint LHS:

$$\begin{aligned} \min_{\boldsymbol{\xi}} \quad & \sum_{s \in S} \xi_s y_s \\ \text{s.t.} \quad & \sum_{s \in S} \xi_s = 0 \quad (\lambda) \\ & \boldsymbol{\xi} \leq \mathbf{u} \quad (\mathbf{w}) \\ & \boldsymbol{\xi} \geq \boldsymbol{\ell} \quad (\boldsymbol{\theta}), \end{aligned}$$

where each term in the parentheses denotes the dual variable of the associated constraint.

Next we take the dual of the problem as in Step 2:

$$\begin{aligned} \max_{\mathbf{w} \leq \mathbf{0}, \boldsymbol{\theta} \geq \mathbf{0}, \lambda} \quad & \mathbf{u}^\top \mathbf{w} + \boldsymbol{\ell}^\top \boldsymbol{\theta} \quad (\geq 1 - \epsilon - \mathbf{y}^\top \mathbf{p}^0) \\ \text{s.t.} \quad & \lambda + w_s + \theta_s = y_s \quad \forall s \in S. \end{aligned}$$

Finally, we remove the maximization and obtain the tractable reformulation of (SI_B) \square

Consequently, the tractable RC of the robust reformulation of the ambiguous chance constraint (5) with respect to the p-box uncertainty set becomes

$$\begin{aligned} f(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) \quad \forall s \in S \\ \mathbf{y}^\top \mathbf{p}^0 + \mathbf{u}^\top \mathbf{w} + \boldsymbol{\ell}^\top \boldsymbol{\theta} &\geq 1 - \epsilon \\ \lambda \mathbf{e} + \mathbf{w} + \boldsymbol{\theta} - \mathbf{y} &= \mathbf{0} \quad (\text{RC}_B) \\ \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \{0, 1\}^{|S|} \\ \mathbf{w} \leq \mathbf{0}, \boldsymbol{\theta} \geq \mathbf{0}, \lambda &\text{ urs.} \end{aligned}$$

The resulting problem (RC_B) is a *mixed-integer linear problem* (MILP) that can be solved by commercial solvers (e.g., CPLEX) for moderate sized instances; given that there exists a linear objective function $[\mathbf{c}^\top \mathbf{x}]$ that we either minimize or maximize.

Similarly, Theorem 2 presents the tractable RC of (SI) with respect to the p-ellipsoidal uncertainty set.

Theorem 2. *The vector $\mathbf{y} \in \{0, 1\}^{|S|}$ satisfies*

$$\sum_{s \in S} p_s y_s \geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}_E \quad (\text{SI}_E)$$

if and only if \mathbf{y} , β , and λ satisfy the following RC problem:

$$\begin{aligned}
 & -\mathbf{y}_\beta^\top \mathbf{p}^0 + 0.25 (\lambda^2 \mathbf{e} + \mathbf{y} + 2\mathbf{y}_\lambda)^\top \mathbf{q}^{-2} + \dots \\
 & \quad \beta^2 \sigma^2 + \beta(1 - \epsilon) \leq 0 \\
 & \lambda \mathbf{e} - M(\mathbf{e} - \mathbf{y}) \leq \mathbf{y}_\lambda \leq \lambda \mathbf{e} + M(\mathbf{e} - \mathbf{y}) \\
 & \beta \mathbf{e} - M(\mathbf{e} - \mathbf{y}) \leq \mathbf{y}_\beta \leq M\mathbf{y} \\
 & -M\mathbf{y} \leq \mathbf{y}_\lambda \leq M\mathbf{y} \\
 & \mathbf{0} \leq \mathbf{y}_\beta \leq \beta \mathbf{e} \\
 & \beta \geq \epsilon, \lambda \text{ urs},
 \end{aligned}$$

where \mathbf{e} is the all-one vector, \mathbf{y}_λ and \mathbf{y}_β are the auxiliary variables to linearize $\lambda \mathbf{y}$ and $\beta \mathbf{y}$; respectively, β and λ are the dual variables, and ϵ is a small positive constant.

Proof. Similar to the proof of Theorem 1, (\mathbf{SI}_E) can be reformulated as

$$\min_{\mathbf{p} \in \mathcal{U}_E} \sum_{s \in S} p_s y_s \geq 1 - \epsilon.$$

The Lagrangian dual problem of the minimization problem in the constraint LHS is given by

$$\max_{\beta \geq 0, \lambda} \min_{\boldsymbol{\xi}} \boldsymbol{\xi}^\top \mathbf{y} + \beta (\|\mathbf{Q}\boldsymbol{\xi}\|_2^2 - \sigma^2) + \lambda \sum_{s \in S} \xi_s$$

where $\beta \geq 0$ and λ denote the dual variables for the p-ellipsoidal constraints $[\|\mathbf{Q}\boldsymbol{\xi}\|_2^2 \leq \sigma^2]$ and $[\sum_{s \in S} \xi_s = 0]$; respectively. From the first-order condition, we have $[\mathbf{y} + 2\beta \mathbf{Q}^2 \boldsymbol{\xi} + \lambda \mathbf{e} = 0]$ ($\mathbf{Q}^\top \mathbf{Q} = \mathbf{Q}^2$ because \mathbf{Q} is a diagonal matrix), i.e., the optimal $\boldsymbol{\xi}^*$ is

$$\boldsymbol{\xi}^* = -\frac{\mathbf{Q}^{-2}(\lambda \mathbf{e} + \mathbf{y})}{2\beta}.$$

Consequently, we substitute $\boldsymbol{\xi}^*$ for $\boldsymbol{\xi}$ in the dual problem. Next, we can remove the maximization as in Step 3 and the LHS becomes

$$\begin{aligned}
 & -\frac{1}{2\beta} \left(\lambda \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{y} + \mathbf{y}^\top \mathbf{Q}^{-2} \mathbf{y} + \dots \right. \\
 & \quad \left. \lambda^2 \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{e} + \lambda \mathbf{y}^\top \mathbf{Q}^{-2} \mathbf{e} \right) + \dots \\
 & \frac{1}{4\beta} \left(\lambda^2 \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{e} + 2\lambda \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{y} + \dots \right. \\
 & \quad \left. \mathbf{y}^\top \mathbf{Q}^{-2} \mathbf{y} \right) - \beta \sigma^2
 \end{aligned}$$

which is $[\geq 1 - \epsilon - \mathbf{y}^\top \mathbf{p}^0]$. It is important to point out that assuming $\beta > 0$ does not affect the optimum solution because the first constraint is binding at optimality $[\|\mathbf{Q}\boldsymbol{\xi}^*\|_2 - \sigma^2 = 0]$ (also see Assumption A2), i.e., the complementary condition $[\beta^*(\|\mathbf{Q}\boldsymbol{\xi}^*\|_2 - \sigma^2) = 0]$ is satisfied for $\beta^* > 0$. Next we multiply the both sides of the inequality

by $\beta (> 0)$:

$$\begin{aligned}
 & -\left(\frac{\mathbf{y}^\top \mathbf{Q}^{-2} \mathbf{y} + \lambda^2 \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{e}}{4\beta} + \frac{\lambda \mathbf{e}^\top \mathbf{Q}^{-2} \mathbf{y}}{2\beta} + \sigma^2 \beta \right) \\
 & \geq 1 - \epsilon - \mathbf{y}^\top \mathbf{p}^0 \quad (\times \beta \text{ both sides})
 \end{aligned}$$

which results in the following inequality

$$\begin{aligned}
 & 0.25 (\mathbf{y}^2 + \lambda^2 \mathbf{e} + 2\lambda \mathbf{y})^\top \mathbf{q}^{-2} + \dots \\
 & \beta^2 \sigma^2 + \beta(1 - \epsilon) - \beta \mathbf{y}^\top \mathbf{p}^0 \leq 0.
 \end{aligned}$$

Notice that since \mathbf{y} is a vector of binary variables, $\mathbf{y}^\top \mathbf{y}$ is equivalent to $\mathbf{y}^\top \mathbf{e}$, and $\mathbf{Q} = \text{diag}(\mathbf{q})$. Moreover, we introduce a vector of auxiliary variables $\mathbf{y}_\beta \in \mathbb{R}_+^{|\mathcal{S}|}$ ($:= \beta \mathbf{y}$) and the following group of constraints to linearize $\beta \mathbf{y}$:

$$\begin{aligned}
 & \mathbf{y}_\beta \geq \beta \mathbf{e} - M(\mathbf{e} - \mathbf{y}) \\
 & \mathbf{y}_\beta \leq M\mathbf{y} \\
 & \mathbf{y}_\beta \leq \beta \mathbf{e} \\
 & \mathbf{y}_\beta \geq \mathbf{0},
 \end{aligned}$$

where M is a large constant. Similarly, we define $\mathbf{y}_\lambda \in \mathbb{R}^{|\mathcal{S}|}$ ($:= \lambda \mathbf{y}$) and the following constraints to linearize $\lambda \mathbf{y}$:

$$\begin{aligned}
 & \mathbf{y}_\lambda \leq M\mathbf{y} \\
 & \mathbf{y}_\lambda \geq -M\mathbf{y} \\
 & \mathbf{y}_\lambda \leq \lambda \mathbf{e} + M(\mathbf{e} - \mathbf{y}) \\
 & \mathbf{y}_\lambda \geq \lambda \mathbf{e} - M(\mathbf{e} - \mathbf{y}).
 \end{aligned}$$

The final formulation of the RC is obtained after the above mentioned substitutions are made and the additional constraints are included. \square

Eventually, the tractable RC of (5) with respect to the p-ellipsoidal uncertainty set becomes

$$\begin{aligned}
 & f(\mathbf{x}, \boldsymbol{\zeta}^s) \leq M(1 - y_s) \quad \forall s \in S \\
 & \sigma^2 \beta^2 + \beta(1 - \epsilon) - \mathbf{y}_\beta^\top \mathbf{p}^0 + \dots \\
 & 0.25 (\mathbf{y} + \lambda^2 \mathbf{e} + 2\mathbf{y}_\lambda)^\top \mathbf{q}^{-2} \leq 0 \\
 & \lambda \mathbf{e} - M(\mathbf{e} - \mathbf{y}) \leq \mathbf{y}_\lambda \\
 & \lambda \mathbf{e} + M(\mathbf{e} - \mathbf{y}) \geq \mathbf{y}_\lambda \tag{RC_E} \\
 & \beta \mathbf{e} - M(\mathbf{e} - \mathbf{y}) \leq \mathbf{y}_\beta \leq M\mathbf{y} \\
 & -M\mathbf{y} \leq \mathbf{y}_\lambda \leq M\mathbf{y} \\
 & \mathbf{0} \leq \mathbf{y}_\beta \leq \beta \mathbf{e} \\
 & \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \{0, 1\}^{|\mathcal{S}|}, \beta \geq \epsilon, \lambda \text{ urs}.
 \end{aligned}$$

Notice that (\mathbf{RC}_E) is a *mixed-integer second-order cone problem* (MISOCP) that yields a convex relaxation problem; SOCP is a tractable mathematical optimization class and supported by commercial solvers such as CPLEX.

Remark 1. We assume that M is large enough to guarantee that big- M constraints are equivalent to the constraints in the original formulation. On the other hand, they are also small enough

to avoid excessive branching, i.e., they must be systematically generated in the numerical experiments.

Remark 2. The second constraint in (\mathbf{RC}_E) can be equivalently reformulated as the following second-order cone constraint:

$$\left\| 1 - \left(\mathbf{y}_\beta^\top \mathbf{p}^0 - 0.25 (\mathbf{y} + 2\mathbf{y}_\lambda)^\top \mathbf{q}^{-2} - \beta(1 - \epsilon) \right) / 2; \right. \\ \left. 0.5\lambda \mathbf{q}^{-1}; \sigma\beta \right\|_2 \leq 1 + \left(\mathbf{y}_\beta^\top \mathbf{p}^0 - 0.25 (\mathbf{y} + 2\mathbf{y}_\lambda)^\top \mathbf{q}^{-2} - \beta(1 - \epsilon) \right) / 2.$$

In summary, using Theorem 1 and 2, one may obtain *exact* robust reformulations of *unique* ambiguous chance constraints with discrete probability distributions when the unknown scenario probabilities $[p_1, p_2, \dots, p_{|S|}]$ are *independent*. The mathematical optimization complexity of the resulting RC is MILP for the p-box and is MISOCP for the p-ellipsoidal uncertainty set given that f is linear in the decision variable $\mathbf{x} \in \mathbf{X}$. In §§3.1-3.3, we shall relax some of these properties in order to tackle with more general classes of ambiguous chance constraints. Good news is that our approach straightforwardly extends for these new classes even though it is not necessarily exact but a *safe approximation* method when the elements of the probability vector are *dependent*.

2.3. Safe Approximation Algorithm

Notice that when the number of scenarios ($|S|$) increases, (\mathbf{RC}_B) and (\mathbf{RC}_E) may be challenging mathematical optimization problems to be solved in a given time limit. This is why, in this section, we present an efficient iterative algorithm that shall yield a safe approximation for ambiguous chance constraints with discrete probability distributions. We illustrate our approach using the following (implicit) notation of the unique ambiguous chance constrained problem:

$$\min_{\mathbf{x} \in \mathbf{X}} \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } \Pr_{\zeta \sim \mathbb{P}} \{ \zeta : f(\mathbf{x}, \zeta) \leq 0 \} \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P},$$

where $1 - \epsilon$ is the prescribed probability, and ζ is the *primitive* uncertainty vector. The four main steps of the algorithm are explained in detail below.

Step 0. We scale the uncertainty parameters to $[-1, 1]$. W.l.o.g., an uncertainty parameter $\hat{\zeta}_i = [\ell_i, u_i]$ for $i \in S$ can be scaled to $[-1, 1]$ by adopting the following linear transformation:

$$\hat{\zeta}_i = \text{NV}_i + \text{HL}_i \times \zeta_i, \quad (9)$$

where $\text{NV}_i = 0.5(\ell_i + u_i)$, $\text{HL}_i = 0.5(u_i - \ell_i)$, and $\zeta_i \in \{-1, 1\}$ is the so-called primitive uncertainty parameter.

Step 1. Derive the RC of the problem at hand

$$\max_{\mathbf{x} \in \mathbf{X}} \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } f(\mathbf{x}, \zeta) \leq 0 \quad \forall \zeta \in \mathcal{U} \quad (10)$$

where $\mathcal{U} \equiv \{ \zeta \in [-1, 1]^{|L|} : \|\zeta\|_2 \leq \Omega \}$. It is important to point out that the associated RC yields SOCP when the f is linear in \mathbf{x} and ζ ; and it is also tractable for the nonlinear case when f is concave in ζ for any given \mathbf{x} (see, [50]).

Step 2. Next, we find the uncertainty parameter realizations that are satisfied for the optimal solution \mathbf{x}^* of (10). Namely,

$$y_s^* = 1 \text{ if } f(\mathbf{x}^*, \zeta^s) \leq 0; = 0 \text{ o.w. } \forall s \in S.$$

Step 3. Fix $\mathbf{y} = \mathbf{y}^*$ and solve (\mathbf{RC}_B) if the ambiguity set of the uncertain probability vector \mathbf{p} is the p-box; or solve (\mathbf{RC}_E) if the associated set is the p-ellipsoidal. If there exists a feasible solution for the RC at hand (for given $\mathbf{y} = \mathbf{y}^*$) then we terminate the algorithm and record \mathbf{x}^* as the final solution of the algorithm. Else, we go to Step 4.

Step 4. We increase Ω by the step size ω and go to Step 1.

The table representation of the so-called *stepwise ellipsoidal algorithm* (SEA) that safely approximates the ambiguous chance constrained problem is given below.

Algorithm 1: Stepwise Ellipsoidal Algorithm

Inputs: Probability threshold ϵ , nominal probability distribution \mathbf{p}^0 , step size ω , radius of the ambiguity set σ , realizations $\zeta^1, \dots, \zeta^{|S|}$, optimization problem.

Outputs: Robust optimal solution \mathbf{x}^* .

Step 0: Scale uncertainty parameters (ζ_i^s) to $[-1, 1]$ and set $\Omega = 0$.

Step 1: Solve the robust counterpart of the given problem with respect to the uncertainty set $\|\zeta\|_2 \leq \Omega$ and find the optimal solution \mathbf{x}^* .

Step 2: Check the satisfied scenarios by \mathbf{x}^* : $f(\mathbf{x}^*, \zeta^s) \leq M(1 - y_s) \quad \forall s \in S$. Record these scenarios to vector \mathbf{y}^* .

Step 3: Solve (\mathbf{RC}_E) for given $\epsilon, \sigma, \mathbf{p}^0$ when $\mathbf{y} = \mathbf{y}^*$.

If (\mathbf{RC}_E) yields a feasible solution **then** terminate the algorithm.

Else go to Step 4.

Step 4: Set $\Omega = \Omega + \omega$ and **go to** Step 1.

Extensions of Algorithm 1. 1) The associated algorithm can be easily extended to joint chance constraints and non-linear uncertain inequalities

as long as the original problem is tractable, i.e., the algorithm does not introduce additional mathematical optimization complexity to that of the original problem. 2) One may also adopt a box uncertainty set in Step 1. We prefer using the ellipsoidal uncertainty set while we are generating our ambiguity set because it yields more diversified solution to the RC compared with the box uncertainty set; also see, [51]. 3) If the ambiguity set is p-box, in Step 3 we solve (\mathbf{RC}_B) instead of (\mathbf{RC}_E) .

Complexity. 1) It is important to point that the mathematical optimization complexities of (\mathbf{RC}_B) and (\mathbf{RC}_E) in Step 3 are reduced to LP and SOCP; respectively, because we fix the binary variable \mathbf{y} to \mathbf{y}^* that comes out of Step 2. 2) In an L -dimensional uncertainty space, the radius of the ellipsoid (or ball) Ω can be at most \sqrt{L} since the ball shall be larger than the support of the uncertainty parameter ζ , i.e., given by $\zeta \in [-1, 1]^L$, when it gets larger than \sqrt{L} . Therefore, Ω is updated in at most $O(\omega^{-1}\sqrt{L})$ iterations of the algorithm. In addition, we need the check $|S|$ scenarios at each iteration of the algorithm in Step 2. Eventually, the total number of iterations of the algorithm is bounded above by $O(\omega^{-1}\sqrt{L}|S|)$.

3. Extensions

In this section, we present extensions of our approach to joint chance constraints, nonlinear inequalities, and dependent probability vectors. It is important to stress that these extensions are not straightforward in general, nevertheless, our approach can be easily extended to the joint chance constraint with slight modification and it is unaffected from nonlinear inequalities, and we propose safe approximations for the dependent case.

3.1. Joint chance constraints

If we have more than one uncertain constraint and aim to find a solution that jointly satisfies these constraints with at least probability $1 - \epsilon$, then we have the following joint (ambiguous) chance constrained formulation:

$$\Pr_{\zeta \sim \mathbb{P}} \{ \zeta : f_k(\mathbf{x}, \zeta) \leq 0 \ \forall k \in K \} \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P}, \quad (11)$$

where K denotes the set of constraint indexes. Unlike classic approaches, our approach straightforwardly extends to a joint chance joint chance constraint. More precisely, one can equivalently

reformulate (11) as

$$\begin{aligned} f_k(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) \quad \forall k \in K, \ \forall s \in S \\ \sum_{s \in S} p_s y_s &\geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}. \end{aligned} \quad (12)$$

The tractable RC of (12) when \mathcal{U} is the p-box uncertainty set is given in Corollary 1.

Corollary 1. *The vectors $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \{0, 1\}^{|S|}$ satisfy*

$$\begin{aligned} f_k(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) \quad \forall k \in K, \ \forall s \in S \\ \sum_{s \in S} p_s y_s &\geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}_B \end{aligned}$$

if and only if \mathbf{x} , \mathbf{y} , $\boldsymbol{\theta}$, \mathbf{w} , and λ satisfy the following RC problem:

$$\begin{aligned} f_k(\mathbf{x}, \zeta^s) &\leq M(1 - y_s) \quad \forall k \in K, \ \forall s \in S \\ \mathbf{y}^\top \mathbf{p}_0 + \mathbf{u}^\top \mathbf{w} + \boldsymbol{\ell}^\top \boldsymbol{\theta} &\geq 1 - \epsilon \\ \lambda \mathbf{e} + \mathbf{w} + \boldsymbol{\theta} - \mathbf{y} &= \mathbf{0} \\ \boldsymbol{\theta} \geq \mathbf{0}, \mathbf{w} \leq \mathbf{0}, \lambda & \text{ urs,} \end{aligned} \quad (\mathbf{jRC}_B)$$

where \mathbf{w} , $\boldsymbol{\theta}$, and λ denote the dual variables.

Proof. The first group of constraints $[f_k(\mathbf{x}, \zeta^s) \leq M(1 - y_s) \ \forall k \in K, \ \forall s \in S]$ will not be affected while we are deriving the RC. The tractable RC of the semi-infinite constraint $[\sum_{s \in S} p_s y_s \geq 1 - \epsilon \ \forall \mathbf{p} \in \mathcal{U}_B]$ is given in Theorem 1. \square

The mathematical optimization complexity of the robust reformulation of the joint chance constraint (\mathbf{jRC}_B) is the same with that of (\mathbf{RC}_B) ; RC of (11) for the p-ellipsoidal ambiguity set can be similarly derived by adopting Theorem 2.

3.2. Nonlinear inequalities

The mathematical optimization complexity of our approach is unaffected when we have nonlinearity in the uncertainty parameter ζ , e.g., when f is linear in \mathbf{x} . In this case, optimization complexity of the resulting RCs for the p-box and the p-ellipsoidal ambiguity sets are MILP and MISOCP; respectively, as in (\mathbf{RC}_B) and (\mathbf{RC}_E) . On the other hand, when f is nonlinear in \mathbf{x} , the structure of nonlinearity in \mathbf{x} determines the problem complexity, i.e., (\mathbf{SI}) does not introduce additional mathematical optimization complexity to that of the original problem.

3.3. Dependent probability vector

In the previous reformulations, we have assumed that the elements of the ambiguous joint probability vector \mathbf{p} that reside in the ambiguity set \mathcal{U} to be independent, e.g., the off-diagonal elements of \mathbf{Q} are 0 in p-ellipsoidal ambiguity set. Nevertheless, the probabilities in the ambiguity set \mathcal{U} may be dependent in two ways:

- (D1) probabilities $[\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{|S|}]$ of some realizations $\{\zeta^{(1)}, \zeta^{(2)}, \dots, \zeta^{(|S|)}\}$ are correlated (e.g., when the probability of high demand increases in period t , the probability of high demand may increase or decrease in period $t + 1$), or
- (D2) uncertainty parameters ζ_i are independent for $i = \{1, \dots, L\}$ and the marginal probabilities follow ambiguous discrete probability distribution.

The dependency in (D1) can be easily formulated by setting the value of \mathbf{Q} such that $(\mathbf{Q}^\top \mathbf{Q})^{-1}$ corresponds to the covariance matrix in the p-ellipsoidal ambiguity set and the RC may be derived similar to Theorem 2; for the sake of brevity, it is skipped. However, the latter dependency (D2) is somewhat more complex, and we propose safe approximations for the associated case in the remainder of this section.

Notice that, when the uncertainty parameters are independent as in (D2), the joint probability is equivalent to the product of the individual probabilities, i.e., probabilities of different uncertainty realizations become dependent because joint probabilities $[\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{|S|}]$ share common or complementary terms.

Extension of the semi-infinite representation (5) of the ambiguous chance constraint for (D2) is given as follows

$$f(\mathbf{x}, \zeta^{i_1, i_2, \dots, i_L}) \leq M(1 - Y_{i_1, i_2, \dots, i_L}) \\ \forall (i_1, i_2, \dots, i_L) \in S$$

$$\sum_{(i_1, i_2, \dots, i_L) \in S} \prod_{j=1}^L \left(p_{i_j}^{(j)} + \xi_{i_j}^{(j)} \right) Y_{i_1, i_2, \dots, i_L} \geq 1 - \epsilon \\ \forall \xi^{(j)} \in \mathcal{U}_{B,E}^{(j)}, j \in \{1, 2, \dots, L\},$$

where an aggregate scenario is indexed by (i_1, i_2, \dots, i_L) for $i_j \in S_j$ for all $j \in \{1, \dots, L\}$; S_j denotes the set of indexes (i_j) for possible realizations of the j^{th} uncertainty parameter $\zeta_{i_j}^{(j)}$, i.e., an aggregate scenario set S is equivalent to the Cartesian product of the individual sets, i.e., $S := S_1 \times S_2 \dots \times S_L$; an aggregate uncertainty vector $\zeta^{i_1, i_2, \dots, i_L} = [\zeta_{i_1}^{(1)}, \zeta_{i_2}^{(2)}, \dots, \zeta_{i_L}^{(L)}]$ is a collection of uncertainty parameter realizations at scenario (i_1, i_2, \dots, i_L) ; $p_{i_j}^{(j)}$ denotes the probability of the i_j^{th} realization of the j^{th} uncertainty parameter $\zeta_{i_j}^{(j)}$ is being realized; $\xi_{i_j}^{(j)}$ denotes the random perturbation in probability $p_{i_j}^{(j)}$ where $\mathcal{U}_B^{(j)}$ and $\mathcal{U}_E^{(j)}$ are the associated uncertainty sets; Y_{i_1, i_2, \dots, i_L} is a 0-1 variable that takes value 1 if the solution \mathbf{x} satisfies the constraint for the associated scenario or 0 otherwise. The formal representations

of the p-box and p-ellipsoidal ambiguity sets for probability distributions of the independent uncertainty parameters are as follows.

$$\mathcal{U}_B^{(j)} = \left\{ \mathbf{p}^{(j)} = \mathbf{p}^{0(j)} + \boldsymbol{\xi}^{(j)} \in \mathbb{R}^{|S_j|} : \dots \\ \ell^{(j)} \leq \boldsymbol{\xi}^{(j)} \leq \mathbf{u}^{(j)}, \dots \\ \sum_{i_j \in S_j} \xi_{i_j}^{(j)} = 0 \quad \forall j \in \{1, \dots, L\} \right\}$$

and

$$\mathcal{U}_E^{(j)} = \left\{ \mathbf{p}^{(j)} = \mathbf{p}^{0(j)} + \boldsymbol{\xi}^{(j)} \in \mathbb{R}^{|S_j|} : \dots \\ \left\| \mathbf{Q}^{(j)} \boldsymbol{\xi}^{(j)} \right\|_2 \leq \sigma_j, \dots \\ \sum_{i_j \in S_j} \xi_{i_j}^{(j)} = 0 \quad \forall j \in \{1, \dots, L\} \right\}.$$

The probability that the scenario (i_1, i_2, \dots, i_L) , i.e., $[\zeta_{i_1}^{(1)}, \zeta_{i_2}^{(2)}, \dots, \zeta_{i_L}^{(L)}]$, is being realized is equivalent to the product of the individual probabilities:

$$P_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L \left(p_{i_j}^{(j)} + \xi_{i_j}^{(j)} \right),$$

Eventually, the semi-infinite probability threshold constraint for the joint distribution can be reformulated as the following mathematical optimization problem for the p-box uncertainty set.

$$(\mathbf{IB}) \min_{\mathbf{P}, \mathbf{p}^{(j)}} \sum_{(i_1, i_2, \dots, i_L) \in S} P_{i_1, i_2, \dots, i_L} Y_{i_1, i_2, \dots, i_L} \\ \text{s.t. } P_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L \left(p_{i_j}^{(j)} + \xi_{i_j}^{(j)} \right) \\ \forall (i_1, i_2, \dots, i_L) \in S \quad (13)$$

$$\sum_{i \in S_j} \xi_{i_j}^{(j)} = 0 \quad \forall j \in \{1, \dots, L\} \quad (14)$$

$$\ell_{i_j}^{(j)} \leq \xi_{i_j}^{(j)} \leq u_{i_j}^{(j)} \\ \forall i \in S_j, \forall j \in \{1, \dots, L\}, \quad (15)$$

where (13) formulates the joint probability as an equality constraint for the ease of exposition, (14) ensures that probability of each independent uncertainty parameter is summed to 1, and (15) determines the bounds of the box uncertainty sets, and $[\mathbf{val}(\mathbf{IB}) \geq 1 - \epsilon]$. It is easy to see that (IB) is nonconvex and highly nonlinear because of (13), and this is why it cannot be dualized in its current form. The following theorem relaxes the nonlinear structure of (IB) and provides a lower bound for $\mathbf{val}(\mathbf{IB})$.

Theorem 3. Let $S = S_1 \times S_2 \dots \times S_L$, $\ell_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L \left(p_{i_j}^{(j)} + \ell_{i_j}^{(j)} \right)$ and $u_{i_1, i_2, \dots, i_L} =$

$\prod_{j=1}^L (p_{i_j}^{(j)} + u_{i_j}^{(j)}) \quad \forall i \in S_j \quad \forall j \in \{1, \dots, L\}$
 the following mathematical optimization problem:

$$\left\{ \begin{array}{l} \min_{\mathbf{P}} \sum_{s \in S} P_s Y_s : \sum_{s \in S} P_s = 1 \\ \ell_s \leq P_s \leq u_s \quad \forall s \in S \end{array} \right\} \quad (\text{rIB})$$

is a relaxation of (IB).

Proof. Let $(\mathbf{P}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(L)})$ be a feasible solution of (IB). If we prove that $\mathbf{P} \in \mathbb{R}^{|S_1| \times |S_2| \times \dots \times |S_L|}$ is feasible for (rIB), then we can conclude that (rIB) is a reduced relaxation of (IB).

To begin with, let $\hat{p}_{i_j}^{(j)} = p_{i_j}^{(j)} + \xi_{i_j}^{(j)}$, (14) implies that $\sum_{i_j \in S_{i_j}} \hat{p}_{i_j}^{(j)} = 1$ for all $j \in \{1, \dots, L\}$ given that $0 \leq \hat{p}_{i_j}^{(j)} \leq 1$ for all $i_j \in S_{i_j}$ for all $j \in \{1, \dots, L\}$ (by Assumption 2). Notice that, probability P_{i_1, i_2, \dots, i_L} is the product of independent probabilities $\hat{\mathbf{p}}^{(j)}$, i.e., the summation of the elements of \mathbf{P} over all combinations of individual probabilities is equivalent to 1. Constraints (13) and (14) imply that $P_{i_1, i_2, \dots, i_L} \geq \prod_{j=1}^L (p_{i_j}^{(j)} + \ell_{i_j}^{(j)})$ and $P_{i_1, i_2, \dots, i_L} \leq \prod_{j=1}^L (p_{i_j}^{(j)} + u_{i_j}^{(j)})$ for all $i_j \in S_{i_j}$ for all $j \in \{1, \dots, L\}$. As a result, \mathbf{P} of (IB) satisfies all the constraints in (rIB). \square

Next, the semi-infinite probability threshold constraint for the joint distribution can be reformulated as the following mathematical optimization problem for the p-ellipsoidal uncertainty set.

$$(\text{IE}) \min_{\mathbf{P}, \mathbf{p}^{(c)}} \sum_{(i_1, i_2, \dots, i_L) \in S} P_{i_1, i_2, \dots, i_L} Y_{i_1, i_2, \dots, i_L}$$

$$\text{s.t. } P_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L (p_{i_j}^{(j)} + \xi_{i_j}^{(j)})$$

$$\forall (i_1, i_2, \dots, i_L) \in S \quad (16)$$

$$\sum_{i \in S_j} \xi_{i_j}^{(j)} = 0 \quad \forall j \in \{1, \dots, L\} \quad (17)$$

$$\| \mathbf{Q}^{(j)} \boldsymbol{\xi}^{(j)} \|_2 \leq \sigma_j \quad \forall j \in \{1, \dots, L\}, \quad (18)$$

Similarly, (IE) is a nonlinear and non-convex optimization problem that is intractable in its current form. The following theorem yields a tractable reduced relaxation of (IE).

Theorem 4. Let $S = S_1 \times S_2 \dots \times S_L$, $\ell_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L (p_{i_j}^{(j)} - \sigma_j / q_{i_j}^{(j)})$ and $u_{i_1, i_2, \dots, i_L} = \prod_{j=1}^L (p_{i_j}^{(j)} + \sigma_j / q_{i_j}^{(j)}) \quad \forall i \in S_j \quad \forall j \in \{1, \dots, L\}$ then (rIB) is a relaxation of (IE).

Proof. Constraints (17) and (18) imply that the largest value that $\xi_{i_j}^{(j)}$ can get is $\sigma_j / q_{i_j}^{(j)}$ for all $i_j \in S_{i_j}$ for all $j \in \{1, \dots, L\}$; similarly, the smallest value is $-\sigma_j / q_{i_j}^{(j)}$. The remainder of the proof follows similar to Theorem 3. \square

Remark 3. The ambiguity set \mathcal{U} of the unknown true probability vector \mathbf{p} for the independent uncertainty parameters may also be defined by first calculating the nominal joint probabilities of the uncertainty parameter realizations and then adopting the p-box or p-ellipsoidal ambiguity set over the nominal joint probability vector; notice that this shall result in the RCs that are proposed in §2. The extension that is presented in §3.2 defines the ambiguity sets with respect to the marginal probability vectors of the individual uncertainty parameters and propose safe approximations of (SI) for the resulting complex ambiguity set of the joint probability vector.

4. Numerical example: knapsack problem

In this section, we present validity or our approach using toy sized instances of an ambiguous chance constrained knapsack problem. We compare optimality and feasibility performances of the classic and the ambiguous chance constrained versions of the problem.

We focus on a chance constrained knapsack problem:

$$\max_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^n v_i x_i$$

$$\text{s.t. } \Pr_{\zeta \sim \mathbb{P}} \left\{ \zeta : \sum_{i=1}^n \zeta_i x_i \leq W \right\} \geq 1 - \epsilon,$$

where v_i denotes the value and ζ_i denotes the uncertain weight of item $i = \{1, \dots, m\}$, x_i is the decision variable that is 1 if item i is included in the knapsack or 0 otherwise, and W is the capacity of the knapsack. The uncertainty parameter ζ follows a discrete probability distribution \mathbb{P} :

$$\frac{\zeta \in \{\zeta^1 \dots \zeta^{|S|}\}}{\mathbb{P}(\zeta)} = \frac{p_1 \dots p_{|S|}}{|S|}$$

where the true probability distribution \mathbf{p} is unknown or ambiguous but is assumed to reside in a p-box (\mathcal{U}_B) or a p-ellipsoidal (\mathcal{U}_E) ambiguity set:

$$\mathcal{U}_B = \left\{ \boldsymbol{\xi} \in \mathbb{R}^L : \mathbf{p} = \mathbf{p}^0 + \boldsymbol{\xi}, \sum_{s \in S} \xi_s = 0, \dots \right.$$

$$\left. (1 - \gamma)\mathbf{p}^0 \leq \mathbf{p} \leq (1 + \gamma)\mathbf{p}^0 \right\},$$

and

$$\mathcal{U}_E = \left\{ \boldsymbol{\xi} \in \mathbb{R}^L : \mathbf{p} = \mathbf{p}^0 + \boldsymbol{\xi}, \sum_{s \in S} \xi_s = 0, \dots, \|\mathbf{p} - \mathbf{p}^0\|_2 \leq \sigma \right\}.$$

The p-box uncertainty set is symmetric around the nominal probability distribution \mathbf{p}^0 and its size is determined by the scaling parameter $\gamma \in [0, 1]$; σ denotes the radius of the p-ellipsoidal uncertainty set that is again centered at the nominal data \mathbf{p}^0 .

Notice that the ambiguous chance constrained knapsack problem with a discrete probability distribution can be equivalently reformulated as

$$\begin{aligned} \max_{\mathbf{x} \in \{0,1\}} \quad & \sum_{i=1}^n v_i x_i \\ & \sum_{i=1}^n \zeta_i^s x_i - W \leq M(1 - y_s) \quad \forall s \in S \\ & \sum_{s \in S} p_s y_s \geq 1 - \epsilon \quad \forall \mathbf{p} \in \mathcal{U}_B \text{ or } E, \end{aligned}$$

where $M = (\max\{\mathbf{e}^\top \boldsymbol{\zeta}^1, \dots, \mathbf{e}^\top \boldsymbol{\zeta}^{|S|}\} - W)$. The tractable RCs of the knapsack problem can be derived as in (RC_B) and (RC_E).

Illustrative example. We solve an instance of the knapsack problem with ten items (i.e., $n = 10$). Scenarios for the item weights (i.e., $S = \{1, \dots, 10\}$) are given in Table 2; the realization probabilities are given in the last row of the table (see, \mathbf{p}^0). The objective coefficients are $v = [37; 43; 53; 67; 44; 57; 69; 45; 54; 66]$, the knapsack capacity is $W = 60$, and the probability threshold is $1 - \epsilon = 0.75$.

Table 2. Uncertainty parameter realizations with probabilities

	ζ^1	ζ^2	ζ^3	ζ^4	ζ^5	ζ^6	ζ^7	ζ^8	ζ^9	ζ^{10}
ζ_1	10	4	5	8	9	5	6	6	7	5
ζ_2	9	4	10	4	9	9	7	8	6	9
ζ_3	8	8	9	7	10	7	8	7	9	7
ζ_4	9	7	10	4	10	9	7	4	9	9
ζ_5	6	5	6	8	8	10	5	7	6	5
ζ_6	9	9	6	6	7	5	5	6	5	5
ζ_7	8	10	5	10	7	5	8	8	6	7
ζ_8	4	4	9	4	8	9	9	8	10	6
ζ_9	10	7	8	10	5	10	5	10	10	5
ζ_{10}	4	5	7	9	8	6	7	9	8	5
\mathbf{p}^0	0.05	0.025	0.1	0.025	0.15	0.15	0.1	0.2	0.1	0.1

Notice that when $\mathbf{p} = \mathbf{p}^0$ in the knapsack problem, we have the classic chance constrained version of the problem. The optimal solution and the objective function value for the classic problem are given Table 3. The optimal solution for

the chance constrained knapsack problem satisfies the constraint with $1 - \epsilon^* = 0.775$ probability.

Table 3. Optimal solution for the chance constrained problem

x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*	x_7^*	x_8^*	x_9^*	x_{10}^*	Obj.
1	0	1	1	1	1	1	0	1	1	<u>447</u>
y_1^*	y_2^*	y_3^*	y_4^*	y_5^*	y_6^*	y_7^*	y_8^*	y_9^*	y_{10}^*	$1 - \epsilon$
0	1	1	0	0	1	1	1	1	1	0.775

Next we solve the ambiguous chance constrained knapsack problem with respect to the p-box (\mathcal{U}_B) and p-ellipsoidal (\mathcal{U}_E) ambiguity sets when $\gamma = 0.4$; respectively. Notice that σ is determined according to the tightest ball inside the box region, i.e., it is less conservative than the box. The optimal solutions are given in Table 4.

It is easy to see that optimal objective function value for the chance constrained problem that uses the nominal probability vector is greater than these of the ambiguous chance constrained version of the problem with respect to p-box and p-ellipsoidal ambiguity sets. To point out, the classic chance constraint approach is not immunized against the ambiguity in the probability distribution and this is why it yields a progressive objective function value. On the other hand, when the worst-case probability distribution \mathbf{p}^* is realized, the nominal solution cannot satisfy the given probability threshold $1 - \epsilon = 0.75$. More precisely, the LHS value for the probability threshold constraint $[\sum_{s \in S} p_s^* y_s \geq 1 - \epsilon]$ is 0.65 for the p-box and 0.746 for the p-ellipsoidal uncertainty sets when \mathbf{y} is the nominal optimal solution; compare these values with “ $\sum_{s \in S} p_s^* y_s$ ” column in Table 4.

Table 4. Optimal solutions for the ambiguous chance constrained problem

Set	\mathbf{x}^*	Obj.
p-box	(1, 0, 1, 1, 1, 1, 1, 1, 0, 1)	<u>438</u>
p-ellip.	(1, 1, 1, 1, 0, 1, 1, 0, 1, 1)	<u>446</u>
	\mathbf{y}^*	$\sum_{s \in S} p_s^* y_s$
p-box	(1, 1, 1, 1, 0, 1, 1, 1, 1, 1)	0.79
p-ellip.	(0, 1, 1, 1, 0, 1, 1, 1, 1, 1)	0.76

As it is anticipated, the classic approach outperforms the robust approach when the nominal probability distribution is used and the robust approach outperforms the classic approach when the worst-case distribution is used. Nevertheless, it may be known that both approaches are based on unique probability vector realizations, namely, the nominal and the worst-case, out of infinitely many possible options in the ambiguity

set, i.e., technically both data points might never be realized. This is why, we also test the average feasibility performances of the nominal and robust solutions via Monte Carlo sampling. Even though, the robust approach does not require any distributional information, we assume in the experiment that probabilities are uniform between $[(1 - \gamma_i)p_i^0, (1 + \gamma_i)p_i^0]$ for the sampling. Algorithm 1 and Algorithm 2 are used to sample from p-box and p-ellipsoidal ambiguity sets.

Algorithm 2: p-box sampling algorithm

Inputs: the nominal probability vector $\mathbf{p}^0 \in [0, 1]^S$, the scaling vector $\gamma \in [0, 1]^S$, *sample size*

Output: *sample*

While # of sampled vectors \leq *sample size*

for $i = 1 : 1 : |S| - 1$

sample $p_i \sim \text{Uniform}[(1 - \gamma_i)p_i^0, (1 + \gamma_i)p_i^0]$

end

calculate $p_{|S|} = 1 - \sum_{i=1}^{|S|-1} p_i$

If $(1 - \gamma_i)p_{|S|}^0 \leq p_{|S|} \leq (1 + \gamma_i)p_{|S|}^0$

add \mathbf{p} to *sample*; # of sampled vectors ++

end

end

We sample 1000 probability vectors from p-box and p-ellipsoidal uncertainty sets and the simulation outcomes are presented in Table 5 & 6. Table 5 reports the average LHS value of the probability threshold constraint $[\sum_{i=1}^{1000} \sum_{s \in S} p_s^i y_s]$ when \mathbf{y} is fixed to the solution at hand. It is easy to see that the optimal solution for the p-box and the p-ellipsoidal uncertainty sets outperform the feasibility performance of the nominal solution.

Algorithm 3: p-ellipsoidal sampling algorithm

Inputs: the nominal probability vector $\mathbf{p}^0 \in [0, 1]^S$, the radius of the ellipsoid $\sigma \in \mathbb{R}_+$, *sample size*

Output: *sample*

Initialization: set $\gamma_i = \sigma/p_i^0 \quad \forall i \in S$

While # of sampled vectors \leq *sample size*

for $i = 1 : 1 : |S| - 1$

sample $p_i \sim \text{Uniform}[(1 - \gamma_i)p_i^0, (1 + \gamma_i)p_i^0]$

end

calculate $p_{|S|} = 1 - \sum_{i=1}^{|S|-1} p_i$

If $\|\mathbf{p} - \mathbf{p}^0\|_2 \leq \sigma$

add \mathbf{p} to *sample*; # of sampled vectors ++

end

end

As we have pointed out before the p-box uncertainty set is larger than the p-ellipsoidal and this is why its solution is immunized against uncertainty more than that for the p-ellipsoidal uncertainty set. If we compare the average feasibility performance of the p-box approach, we also see that its average performance is better than that of the p-ellipsoidal approach. It is important to point out that all approaches, namely,

p-box, p-ellipsoidal, and nominal, yield (on average) feasible results that are (on average) far from being binding to the given probability threshold $1 - \epsilon = 0.75$.

Table 5. Feasibility performance of solutions

sample sets	solution		
	box	ellip	nom
p-box	0.85	0.79	0.76
p-ellip	0.85	0.80	0.76

On the other hand, solutions are not feasible for all of the sampled instances. In Table 6, we report the average performances of the solutions for the violating instances. The first three columns in Table 6 denote the number of violating instances (out of 1000). E.g., the numerical result at “row = p-box” and “column = #Vio(ellip)” denotes the number of violated constraints when the solution is fixed to the optimal solution of the RC with respect to the p-box uncertainty set, i.e., the optimal solution of the ambiguous chance constraint for the p-box uncertainty is tested when the probability vectors are sampled from the p-ellipsoidal uncertainty set. The last three columns report the average LHS value of the probability threshold constraint for the violating instances.

Table 6. Feasibility performance of solutions

\mathcal{U}	#Vio(box)	#Vio(ellip)	#Vio(nom)
p-box	0	83	283
p-ellip	0	0	25
\mathcal{U}	$\overline{\text{Vio}}(\text{box})$	$\overline{\text{Vio}}(\text{ellip})$	$\overline{\text{Vio}}(\text{nom})$
p-box	-	0.73	0.72
p-ellip	-	-	0.74

The numerical results show that the nominal solution cannot satisfy nearly one third of the instances when data comes from p-box uncertainty set. Since the ellipsoidal uncertainty set is less conservative with 83 violations, the feasibility performance of the nominal solution for the p-ellipsoidal uncertainty is less conservative; compare the number of violations in the third column. As it is anticipated, the optimal solution with respect p-box uncertainty set does not violate the probability threshold constraint when data comes from the p-ellipsoidal because p-box is a superset; and both approaches are robust with respect to their own ambiguity sets, e.g., the p-box solution is robust against the p-box ambiguity set by definition. The average violation of the LHS value of the probability constraint for

the violating instances ranges between %1 to %3; see the last two columns where p-box solution is fully robust against the uncertainty in the probability distribution; p-ellipsoidal approach violates the constraints for the p-box uncertainty (on average) with %2; and the nominal solution violates the constraint for p-box uncertainty with %3 and for the p-ellipsoidal uncertainty with %1.

Monte Carlo simulation. In the remainder of this section, we shall adopt a Monte Carlo simulation study to compare the optimality and computational performances of the *stepwise ellipsoidal algorithm* (SEA) that is proposed in §2.3 with those of the exact reformulation (\mathbf{RC}_E) and the *randomized algorithm* (RA) that is proposed by [15] for combinations of the instance parameters of the chance constrained knapsack problem. RA is based on random extraction of N realizations ($\{\hat{\zeta}^1, \dots, \hat{\zeta}^N\}$) of the uncertainty parameter ζ using the known probability distribution of the parameter, i.e., a discrete probability distribution in our case. Each extraction coincides with a discrete constraint and instead of solving the exact problem, the authors solve the mathematical optimization problem with N constraints, i.e., $[f(x, \hat{\zeta}^i) \leq 0 \ \forall i \in \{1, \dots, N\}]$. They show that the optimal solution of the randomized approach satisfies the following probabilistic guarantee:

$$\Pr\{V(x^*) > \epsilon\} \leq \sum_{i=0}^{L-1} \text{Cb}(N, i)\epsilon^i(1 - \epsilon)^{N-i},$$

where the LHS of the inequality is associated with the probability that the chance constraint is violated for the solution at hand (\mathbf{x}_N^*) when N samples are drawn for an L -dimensional decision space, and the RHS coincides with the worst-case bound for the associated probability, e.g., when $\epsilon = 0.2$, $L = 5$ and $\text{RHS} = 0.01$, then $N = 54$ samples have to be drawn. We refer reader to [15] for further details on RA; and to [16] for the formal proof on tightness of the given probability bound.

Data used in the experiment is generated as follows: 1) we randomly sample the lower (ℓ_i) and upper (u_i) bounds of the intervals of the item weights from $U[1-10]$ and $U[11-20]$; respectively, i.e., $\zeta_i \in [\ell_i, u_i] \ \forall i \in L$. 2) Next, we randomly sample $|S|$ scenarios as vectors $\{\zeta^1, \zeta^2, \dots, \zeta^{|S|}\}$ using the associated intervals. 3) For the sake of simplicity, we assume that all scenarios are equally likely, i.e., $p_s^0 = 1/|S|$ for all $s \in S$. 4) Item values (\mathbf{v}) are randomly sampled from $U[10-20]$ and the bag capacity (W) is set to %80 of the total nominal weights of the items. The confidence level $[\sum_{i=0}^{L-1} \text{Cb}(N, i)\epsilon^i(1 - \epsilon)^{N-i}]$ of RA is

set to 0.01; the number of samples to be drawn for $\epsilon \in \{0.1, 0.2\}$. 5) L denotes the number of items and $|S|$ denotes the number of scenario realizations used in discrete probability distribution. For each $L/|S|$ pair 20 different data sets have been generated by following the data generation structure mentioned above. Using the associated data, the optimality and CPU performances of RA, SEA (Algorithm 1) and the exact reformulation (\mathbf{RC}_E) have been compared in Tables 7 & 8.

As we have pointed out above we have generated 20 instances for each $L/|S|$ combination. The asterisk symbol (*) denotes that the associated algorithm solves all instances to optimality for the given parameter combination or it denotes the best performing algorithm when the exact reformulation cannot be solved. The tilde symbol (\sim) denotes that optimality cannot be attained in 7200 seconds. “Gap” denotes the (average) optimality gap percentage of an algorithm for the instances those cannot be solved to optimality or the (average) optimality gap percentage of an algorithm with respect to the best performing algorithm when the exact reformulation cannot be solved in 7200 seconds; the hash tag symbol (#) denotes the number of instances (out of 20) those cannot be solved to optimality (or worse than the best performing algorithm) by the given algorithm. Finally, CPU is the total computation time of a given algorithm in seconds. All experiments are run on a 64-bit Windows machine equipped with an Intel Quad-Core i7-6700HQ processor with 16 GB of RAM.

Table 7. Comparison of solution approaches when $\epsilon = 0.1$

$L/ S $	Output	Exact	RA	SEA
5/10	Gap (#)	*	*	*
	CPU	0.10	0.02	0.01
10/20	Gap (#)	*	%7.6 (18)	%1 (1)
	CPU	76	0.1	0.02
25/50	Gap (#)	-	%6 (19)	%1 (1)
	CPU	\sim	0.4	0.02

Table 8. Comparison of solution approaches when $\epsilon = 0.2$

$L/ S $	Output	Exact	RA	SEA
5/10	Gap (#)	*	%15 (15)	*
	CPU	0.10	0.2	0.02
10/20	Gap (#)	*	%12.2 (19)	*
	CPU	22	0.3	0.02
25/50	Gap (#)	~	%9.8 (20)	*
	CPU	~	0.4	0.02

The numerical results in Table 7 & 8 show that SEA outperforms RA at all instances both in CPU time and the optimality gap. Improved optimality performance of SEA may be an anticipated result since the algorithm systematically includes the uncertain parameters in the uncertainty set in order to yield the tightest ambiguity set and hence a better objective function value compared with RA that is fully randomized without a systematic objective performance consideration. Improved computational and optimality performances of SEA become more significant when probability threshold ϵ , and hence the size of the feasible region, increases; it is easy to see that when we compare the numerical results (Gap, CPU) in Tables 7 & 8. Notice that $L/|S|=5/10$ instances in Table 7 are kind of redundant since we need to satisfy all uncertainty realizations (i.e., robust) in this case to satisfy $1 - \epsilon = 0.9$ probability threshold. Nevertheless, these are the only instance where we see no significant difference among all alternative approaches. According to these numerical results, we can conclude that SEA is an efficient approximation algorithm to solve the ambiguous chance constrained problem for medium to large sized instances. SEA is almost as good as the exact reformulation (except for 1 instance in Table 7 at $L/|S| = 10/20$ with %1 gap) in optimality performance while yielding high quality solutions in less than a second. Moreover, SEA significantly outperforms RA in increased dimensions where we cannot adopt the exact approach (see rows $L/|S| = 25/50$ in Tables 8 & 9).

5. Conclusion

In this paper, we have proposed robust reformulations of ambiguous chance constraints with discrete probability distributions. We have derived the tractable robust counterparts of the associated class of ambiguous chance constraints using p-box and p-ellipsoidal uncertainty sets that support the ambiguous family of distributions. Our approach can be easily applied in practice where one works with a finite number of scenarios that

follow a discrete probability distribution. The associated probabilities are ambiguous by nature because they are forecast or decided by an expert opinion. Proposed methodology aims to find a solution that satisfies the given probability requirement of being feasible while maximizing the utility (or minimizing the cost) and at the same time taking into account the distributional ambiguity. The proposed approach can be easily extended to joint chances constraints, nonlinear inequalities as well as different data structures without introducing additional mathematical optimization complexity to that of tackling a unique ambiguous chance constraint with discrete probability distribution using our method.

Acknowledgment

The author thanks to two anonymous referees for their insightful comments and suggestions which significantly improved the overall quality of the paper. The author also thanks to Farzad Avishan who has generated the data used in the numerical experiments and made the final proofreading.

References

- [1] Charnes, A. & Cooper, W.W. (1959). Chance-constrained programming. *Management Science*, 6(1), 73–79.
- [2] Charnes, A., Cooper, W.W. & Symonds, G.H. (1958). Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science*, 4(3), 235–263.
- [3] Miller, B.L. & Wagner, H.M. (1965). Chance constrained programming with joint constraints. *Operations Research*, 13(6), 930–945.
- [4] Prékopa, A. (1970). Efficient robust optimization of metal forming processes using a sequential metamodel based strategy. In *Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton, NJ, 113–138.
- [5] Burkauskas, A. (1986). On the convexity problem of probabilistic constraint stochastic programming problems. *Alkalmazott Matematikai Lapok*, 12, 77–90.
- [6] Prékopa, A. (1974). Programming under probabilistic constraints with a random technology matrix. *Mathematische Operationsforschung und Statistik*, 5, 109–116.
- [7] Van de Panne, C. & Popp, W., (1963). Minimum-cost cattle feed under probabilistic protein constraints. *Management Science*, 9(3), 405–430.

- [8] Prékopa, A. (1973). On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34, 335–343.
- [9] Prékopa, A. (1995). *Stochastic Programming* Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [10] Nemirovski, A. & Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4), 969–996.
- [11] Ben-Tal, A. & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3), 411–424.
- [12] Chen, W., Sim, M., Sun, J. & Teo, C.P. (2010). From CVaR to uncertainty set: implications in joint chance-constrained optimization. *Operations Research*, 58(2), 470–485.
- [13] Chen, X., Sim, M. & Sun, P. (2007). A robust optimization perspective on stochastic programming. *Operations Research* 55(6), 1058–1107.
- [14] Zymler S., Kuhn, D. & Rüstem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2), 167–198.
- [15] Calafiore, G.C. & Campi, M.C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102, 25–46.
- [16] Campi, M.C. & Garatti, S. (2008). The exact feasibility of randomized solutions of robust convex programs. *SIAM Journal on Optimization*, 19(3), 1211–1230.
- [17] Birge, J.R. & Wets, R.J.-B. (1986). Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Mathematical Programming Studies*, 27, 54–102.
- [18] Dupačová, J. (2001). Stochastic Programming: minimax approach. In: *Encyclopedia of Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [19] Shapiro, A. & Ahmed, S. (2004). On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14(4), 1237–1252.
- [20] Shapiro, A. & Kleywegt, A.J. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17, 523–542.
- [21] Epstein, L.G. & Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1), 1–31.
- [22] Epstein, L.G. & Schneider, M. (2007). Learning under ambiguity. *Review of Economic Studies*, 74(4), 1275–1303.
- [23] Hansen, L.P. & Sargent, T.J. (2001). Robust control and model uncertainty. *American Economic Review*, 91, 60–66.
- [24] Erdoğan, E. & Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1–2), 37–61.
- [25] Yanikoğlu, İ, den Hertog, D. (2013). Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4), 666–681.
- [26] Ben-Tal, A., El Ghaoui, L., Nemirovski, A. (2009). *Robust Optimization*. Princeton Press, Princeton, NJ.
- [27] Nemirovski, A. (2012). On safe tractable approximations of chance constraints. *European Journal of Operations Research*, 219(3), 707–718
- [28] Yanikoğlu, İ & den Hertog, D. & Kleijnen, J.P.C. (2016). Robust dual-response optimization. *IIE Transactions*, 48(3), 298–312.
- [29] 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.
- [30] Song, Y., Luedtke, J. R., & Küçükyavuz, S. (2014). Chance-constrained binary packing problems. *INFORMS Journal on Computing*, 26(4), 735–747.
- [31] Hanasusanto, G.A., Roitch, V., Kuhn, D. & Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1), 35–62.
- [32] Hanasusanto, G.A., Roitch, V., Kuhn, D. & Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3), 751–767.
- [33] Jiang, R. & Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1-2), 291–327.
- [34] Chen, Z., Kuhn, D., & Wiesemann, W. (2018). Data-driven chance constrained programs over Wasserstein Balls. *arXiv preprint*, arXiv:1809.00210.
- [35] Ji, R. & Lejeune, M. (2018). Data-driven distributionally robust chance-constrained programming with Wasserstein metric. Available at *Optimization Online*.
- [36] Zhang, Y., Jiang, R. & Shen, S. (2016). Distributionally robust chance-constrained bin packing. Available at *Optimization Online*.
- [37] Cheng, J., Delage, E. & Lisser, A. (2014). Distributionally robust stochastic knapsack

- problem. *SIAM Journal on Optimization*, 24(3), 1485–1506.
- [38] Hu, Z. & Hong, L.J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. Available at *Optimization Online*.
- [39] Xie, W. & Ahmed, S. (2018). On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2), 1151–1182.
- [40] Xie, W., Ahmed, S. & Jiang, R. (2017). Optimized Bonferroni approximations of distributionally robust joint chance constraints. Available at *Optimization Online*.
- [41] Xie, W. (2018). On distributionally robust chance constrained program with wasserstein distance. *arXiv preprint*, arXiv:1806.07418.
- [42] Mehrotra, S. & Zhang, H. (2014). Models and algorithms for distributionally robust least squares problems. *Mathematical Programming*, 146(1-2), 123–141.
- [43] Özmen, A., Weber, G.W., Batmaz, İ. & Kropat, E. (2011). RCMARS: Robustification of CMARS with different scenarios under polyhedral uncertainty set. *Communications in Nonlinear Science and Numerical Simulation*, 16(12), 4780–4787.
- [44] Friedman, J.H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19(1), 1–141.
- [45] Weber, G.W., Batmaz, İ., Köksal, G., Taylan, P. & F. Yerlikaya-Özkurt, (2012). CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, 20(3), 371–400.
- [46] Bemis, C., Hu, X., Lin, W., Moazeni, S., Wang, L., Wang, T. & Zhang, J. (2009). Robust portfolio optimization using a simple factor model. *IMA Preprint Series*, No. 2284, University of Minnesota, Minneapolis, MN.
- [47] Kara, G., Özmen, A., & Weber, G. W. (2019). Stability advances in robust portfolio optimization under parallelepiped uncertainty. *Central European Journal of Operations Research*, 27(1), 241–261.
- [48] Postek, K., den Hertog, D. & Melenberg, B. (2016). Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review*, 58(4), 603–650.
- [49] Gorissen, B.L., Yanıkoğlu, İ & den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53, 124–137.
- [50] Ben-Tal, A., den Hertog, D. & and Vial, J.-P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2), 265–299.
- [51] Yanıkoğlu, İ. (2018). Selected Topics in Robust Optimization. In: *Optimization Techniques for Problem Solving in Uncertainty*. IGI Global, PA, USA.

İhsan Yanıkoğlu is an Assistant Professor of Industrial Engineering at Özyegin University, İstanbul, Turkey. He obtained his Ph.D. in Econometrics and Operations Research at Tilburg University in 2014. His research focuses on developing robust optimization methodologies and applying these methods in real-life problems. In 2017, he has received the Operations Engineering and Analytics Best Paper Award from IIE Transactions Journal with his publication titled “Robust Dual-response Optimization” together with Dick den Hertog and Jack Kleijnen. His research has been published in journals such as *SIAM Journal on Optimization*, *INFORMS Journal on Computing*, *OMEGA*, *IIE Transactions*, and *EJOR*.

