ORIGINAL PAPER

Distributionally robust optimization with decision dependent ambiguity sets

Fengqiao Luo[1] · Sanjay Mehrotra[1]

Abstract

We study decision dependent distributionally robust optimization models, where the ambiguity sets of probability distributions can depend on the decision variables. These models arise in situations with endogenous uncertainty. The developed framework includes two-stage decision dependent distributionally robust stochastic programming as a special case. Decision dependent generalizations of five types of ambiguity sets are considered. These sets are based on bounds on moments, covariance matrix, Wasserstein metric, Phi-divergence and Kolmogorov–Smirnov test. For the finite support case, we use linear, conic or Lagrangian duality to give reformulations of these models with a finite number of constraints. Reformulations are also given for the continuous support case for moment, covariance, Wasserstein and Kolmogorov–Smirnov based models. These reformulations allow solutions of such problems using global optimization techniques within the framework of a cutting surface algorithm. The importance of decision dependence in the ambiguity set is demonstrated with the help of a numerical example modeling simultaneous determination of order quantity and selling price for a newsvendor problem.

Keywords Distributionally robust optimization · Decision dependent ambiguity set · Conic duality

1 Introduction

The uncertain characteristics of a system's performance often depend on its design decisions. This type of uncertainty is called endogenous uncertainty. For example in

✉ Fengqiao Luo
   fengqiaoluo2014@u.northwestern.edu

   Sanjay Mehrotra
   mehrotra@northwestern.edu

[1] Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL, USA

a newsvendor model product demand function may depend on its selling price [1]. Additional examples of decision problems with endogeneous uncertainty from finance, resource management, process design, and network design are given in Sect. 2. The goal of this paper is to present decision dependent ambiguity frameworks to model problems involving endogenous uncertainty. The main contribution is in showing that the dualization of a certain inner problem continues to be applicable in this more general setting. This dualization has a unique advantage for the problems under consideration. It allows application of algorithms from nonlinear global optimization to solve the resulting reformulations.

Specifically, we study the optimization problems in which the ambiguity set of distributions may depend on the decisions in the following modeling framework:

$$\min_{\boldsymbol{x} \in X} \left\{ f(\boldsymbol{x}) + \max_{P \in \mathcal{P}(\boldsymbol{x})} \mathbb{E}_P[h(\boldsymbol{x}, \boldsymbol{\xi})] \right\}. \tag{D^{3}RO}$$

Here $\boldsymbol{x}$ is the vector of decision variables with the feasible set $X \subseteq \mathbb{R}^n$, and $\boldsymbol{\xi}$ is the vector of uncertain model parameters, which is defined on a measurable space $(\varXi, \mathcal{F})$; $\varXi$ is the support in $\mathbb{R}^d$, and $\mathcal{F}$ is a σ-algebra. We may allow ξ to also depend on $\boldsymbol{x}$, i.e., replace $h(\boldsymbol{x}, \boldsymbol{\xi})$ in (D^{3}RO) with $h(\boldsymbol{x}, \boldsymbol{\xi}(\boldsymbol{x}))$ but we have not done so here for notation simplicity. For a given $\boldsymbol{x}$, the ambiguity set $\mathcal{P}(\boldsymbol{x})$ of the unknown probability distribution depends on the decision variables $\boldsymbol{x}$, and $\mathcal{P}(\boldsymbol{x}) \subseteq \mathcal{P}(\varXi, \mathcal{F})$, where $\mathcal{P}(\varXi, \mathcal{F})$ is the set of probability distributions defined on $(\varXi, \mathcal{F})$. The function $f(\boldsymbol{x})$ is the deterministic part of the objective with no uncertain parameters. Keeping this function in (D^{3}RO) allows us to consider decision models involving two-stage decision making.

We denote the inner problem $\max_{P \in \mathcal{P}(\boldsymbol{x})} \mathbb{E}_P[h(\boldsymbol{x}, \boldsymbol{\xi})]$ as (D^{3}RO)-inner. Note that if $h(\boldsymbol{x}, \xi)$ is a recourse function in a two-stage stochastic program, i.e.,

$$h(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y} \in \mathbb{R}^q} \ g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}), \quad \text{s.t. } \psi_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}) \geq 0 \quad \forall i \in [m], \tag{1}$$

where $g(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi})$ and $\psi_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi})$ are bounded and continuous functions of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\xi}$, then (D^{3}RO) becomes a two-stage decision dependent distributionally robust stochastic program (TSD^{3}SP), which is an important application of (D^{3}RO). We assume that the minimization problem in (1) is feasible for any $\boldsymbol{x} \in X$ and $\boldsymbol{\xi} \in \varXi$, and $h(\boldsymbol{x}, \boldsymbol{\xi})$ is finite. In other words, we assume that (D^{3}SP) has complete recourse [2]. Moreover, we assume a certain Slater-type condition for the set $\mathcal{P}(\boldsymbol{x})$, as needed.

The ambiguity set $\mathcal{P}(\boldsymbol{x})$ can be constructed in many different ways. The reformulations given in this chapter are for the decision dependent generalizations of the most common types of ambiguity sets proposed in the distributionally robust optimization literature [3]. In Sect. 3 we investigate the reformulation of (D^{3}RO) for five different possible specifications of $\mathcal{P}(\boldsymbol{x})$: (i) the ambiguity sets defined by using componentwise moment inequalities and bounds on the scenario probabilities [4,5]; (ii) ambiguity sets defined by using the mean vector and covariance matrix inequalities [6]; (iii) ambiguity sets defined by using the Wasserstein metric [7–10]; (iv) ambiguity sets defined by using ϕ-divergence [11–16]; and (v) the ambiguity sets defined by using the multi-

variate Kolmogorov–Smirnov test [17]. The reformulations are given in Sects. 3.1 to 3.5, respectively. The choice of the specific ambiguity set when modeling a problem is context dependent. A choice depends on the data being represented by the set, and the needs of the modeler. Moments based sets might be natural when their estimates are available from a prior statistical analysis. However, Wasserstein-distance based sets have an underlying polyhedral structure that makes them more amenable to algorithm development within the linear programming or linear conic duality frameworks. Total variance based sets specify bounds on the probabilities of each scenario more naturally. Kolmogorov–Smirnov test is commonly used in hypothesis testing that compares probability distributions.

The basic concept used in arriving at the reformulations of (D^3RO)-inner is to use linear programming duality or conic duality as needed in the specific settings. Lagrangian duality is used for the situations where considering the saddle point problem appears more suitable.

We note that the computational complexity of the reformulated problem is not the main motivation of this paper. Our goal is towards studying the modeling frameworks that more realistically represent the underlying phenomenon. Here we also do not focus on developing any new (possibly more efficient) algorithms, as that is left for future studies. In general, we refer to the global optimization techniques for solving the non-convex optimization problems resulting from our reformulations [18]. Moreover, to simplify the presentation, we consider the finite support case in the main text. Reformulations (without proof) are also given for the ambiguity sets (i)–(iii) allowing for continuous support. In these cases semi-infinite programming reformulation of the corresponding models are given. These semi-infinite reformulations allow the use of a cutting surface algorithm, given in Sect. 4.1, to solve the semi-infinite problems. The models having first-stage binary variables allow for the use of a reformulation technique that reformulates the product of a binary variable with a continuous variable (Lagrange multiplier). The central idea behind this technique is given in Sect. 4.2. A newsvendor example illustrating the reformulations and the relevance of decision dependent ambiguity is given in Sect. 5. The construction of the decision dependent parameters appearing in the specification of the ambiguity set is discussed briefly when making the concluding remarks.

2 Literature review on optimization with decision dependent uncertainty

The endogenous uncertainty has been considered in dynamic programming [19], stochastic programming [20], robust optimization [21,22], with applications in financial market modeling [23–25], resource management [26], stochastic traffic assignment [27], oil (natural gas) exploration [28–30], and robust network design [31,32].

In the framework of stochastic optimization, the endogenous uncertainty affects the underlying probability distribution and the scenario tree. Jonsbråten et al. [33] studied stochastic programming problems with decision dependent scenario distributions, where the distribution is indexed by a Boolean vector. They provided an

implicit enumeration algorithm for solving these problems based on a branch-and-bound scheme. This model and the proposed branch-and-bound method was applied to an optimal selection and sequencing of oil well exploration problem under reservoir capacity uncertainty [28]. Ahmed [31] investigated a class of single stage stochastic programs with discrete candidate probability distributions that are based on Luce's choice axiom [34]. The decision affects utility functions of the choices, and hence the probability distribution [35]. These types of problems arise from network design and server selection applications [31]. It is shown that stochastic programs of this class can be reformulated as 0-1 hyperbolic programs. Viswanath et al. [32] investigated a two-stage shortest path problem in a stochastic network which arises from disaster relief services. Here the first stage investment decisions can reduce the failure probability of links in the network, and a shortest path is identified based on the post-event network. Held et al. [36] developed a heuristic algorithm to solve a two-stage stochastic network interdiction problem, where the interdictor is the first-stage decision maker whose objective is to maximize the probability that the minimum path length exceeds a certain value after the interdiction. The interdictor's decision changes the network topology and the uncertainty description. The structure of single-stage stochastic programming with a decision dependent probability distribution was also studied by [37]. Lee et al. [38] investigated a newsvendor model under decision dependent uncertainty, where sequential decisions are made after a re-estimation of the demand distribution. They provide conditions under which the estimation and decision process converges.

Grossmann et al. [20,30] developed a disjunctive programming reformulation for multistage decision dependent stochastic programs. They investigated this problem with finitely many scenarios, where exogenous and endogenous uncertain parameters are involved. In their models, endogenous parameters are resolved after the operational decisions are made (e.g., a facility is installed or an investment is made). A branch-and-bound algorithm is developed to solve the disjunctive program by branching on the logic variables involved in the disjunctive clauses [20,39], and a lower bound is obtained at each node by solving a Lagrangian dual sub-problem [40]. More solution strategies for the disjunctive program are given in [41]. This framework is applied to model and solve the offshore oil or gas field infrastructure multi-stage planning problem with uncertainty in estimating parameters that are not immediately realized [29]. The framework is also applied to optimize process network synthesis problems with yield uncertainty that can be reduced by investing in pilot plants [42]. Tarhan et al. [43] developed a computational strategy that combines global optimization and outer-approximation to solve multistage nonlinear mixed-integer programs with decision dependent uncertainty.

Decision dependent uncertainty is also considered in the framework of robust optimization by letting the uncertainty set depend on the decision variables. Spacey et al. [44] studied a problem of minimizing the run time of a computer program by assigning code segments to execution locations where the scheduling of code segment execution depends on the assignment. In robust combinatorial optimization, decision dependent uncertainty set is used to ensure the same relative protection level of all binary decision vectors [21]. To model a robust task scheduling problem with uncertainty in the processing time, Vujanic et al. [45] proposed a decision dependent uncertainty set as a Minkowski sum of some static sets such that the uncertain completion time interval

of a task can naturally depend on the starting time of the task. Hu et al. [1] studied a newsvendor model where the product demand may depend on the selling price. Since the analytical relationship between the demand and the selling price is unknown, they construct a family of decreasing and convex functions from historical data as the functional ambiguity set of the true demand function, and solve the functionally robust optimization problem of this model using a univariate reformulation. Nohadani et al. [22] investigated robust linear programs with decision dependent budget-type uncertainty (RLP-DDU). They showed that this problem is NP-hard even in the case where the uncertainty set is a polyhedron and the decision dependence is affine. RLP-DDU can be reformulated as a mixed integer linear program (MILP), if the decision variables affect uncertain variables by controlling the upper bounds of the uncertain variables. This concept is demonstrated in a robust shortest-path problem, where the uncertainty is resolved progressively when approaching the destination.

3 Reformulation of (D^{3}RO)

We investigate the dual of the inner problem of (D^{3}RO) under the assumption that the probability distributions are specified over a finite support on $\varXi$. Reformulations for the case where $\varXi$ is continuous are also given for the first three cases. We make the following assumption to simplify the presentation of our key observations:

Assumption 1 Every $P \in \mathcal{P}(x)$ has a decision independent finite support $\varXi := \{\boldsymbol{\xi}^k\}_{k=1}^N, \forall x \in X$, for a fixed N.

The finite support assumption is commonly used when developing computational algorithms for solving general two-stage stochastic optimization. The finite support framework arises either due to an application of sample average approximation approach when solving problem having parameters with continuous support, or in a data-driven setting when the data provides the nominal samples for the model. For a more general model, the set $\varXi$ can be a compact set in the Euclidean space. The reformulations given in Sects. 3.1–3.3 that have finitely many constraints due to Assumption 1 (specified by $\boldsymbol{\xi}^k, k = 1, \ldots, N$) become semi-infinite programs, where the constraints are parameterized over the continuous support $\varXi$. Dual formulations for finite support case provide a bridge to understanding dual formulations for the case where the distributions have continuous support. In the finite support case, because of LP duality, often we do not need Slater-type conditions. However, these conditions are required for the more general case. It is important to be able to see such distinctions when considering applications. Moreover, all reformulations adapt to the case where $\boldsymbol{\xi}^k$ is replaced with $\boldsymbol{\xi}^k(x)$.

An interpretation of Assumption 1 is as follows. In this case, the candidate probability distributions in $\mathcal{P}(x)$ are represented as a vector $\boldsymbol{w} \in \mathbb{R}^N$ such that $w_i(x)$ is the probability assigned to scenario $\boldsymbol{\xi}^i$ ($i \in [N]$) and $\|\boldsymbol{w}(x)\|_1 = 1$ for all $x \in X$. By judicious specification of x, the support of the distribution (turning a scenario on or off) may also be allowed to change with x by forcing certain scenarios to have zero probability in the specification of the set $\mathcal{P}(x)$. Therefore, this finite support framework allows for a great deal of generality.

3.1 Ambiguity sets defined by simple measure and moment inequalities

We consider the moment robust set defined as follows:

$$
\mathcal{P}^{SM}(\boldsymbol{x}) := \left\{ P \in \mathcal{M}(\varXi, \mathcal{F}) \,\middle|\, \nu_1(\boldsymbol{x}) \preceq P \preceq \nu_2(\boldsymbol{x}), \ \int_\varXi f_i(\boldsymbol{\xi}) P(d\xi) \in [l_i(\boldsymbol{x}), u_i(\boldsymbol{x})] \ \ i \in [m] \right\},
\tag{2}
$$

where $\mathcal{M}(\varXi, \mathcal{F})$ is the set of positive measures defined on $(\varXi, \mathcal{F})$, $\nu_1(\boldsymbol{x}), \nu_2(\boldsymbol{x}) \in \mathcal{M}(\varXi, \mathcal{F})$ are two given measures for a fixed $\boldsymbol{x}$ that are lower and upper bounds of candidate probability measures, and $\boldsymbol{f} := [f_1(\boldsymbol{\xi}), \ldots, f_m(\boldsymbol{\xi})]$ is a vector of moment functions. The ordering "$\preceq$" is defined as follows: For any two measures (not necessarily probability measures) $\mathcal{M}_1$ and $\mathcal{M}_2$ defined on $(\varXi, \mathcal{F})$, $\mathcal{M}_1 \preceq \mathcal{M}_2$ is equivalent to $\mathcal{M}_1(S) \leq \mathcal{M}_2(S) \ \forall S \in \mathcal{F}$. If Assumption 1 is satisfied, then $\mathcal{M}_1 \preceq \mathcal{M}_2$ is equivalent to $\mathcal{M}_1(\boldsymbol{\xi}^k) \leq \mathcal{M}_2(\boldsymbol{\xi}^k) \ \forall k \in [N]$. To ensure that P is a probability distribution, we set $l_1(\boldsymbol{x}) = u_1(\boldsymbol{x}) = 1$ and $f_1(\boldsymbol{\xi}) = 1$ in the above definition of $\mathcal{P}^{SM}(\boldsymbol{x})$. For any $\boldsymbol{\xi} \in \varXi$, let $\boldsymbol{\xi} := [\xi_1, \ldots, \xi_d]$. When standard moments are used, the ith ($i \in [m]$) entry of $\boldsymbol{f}$ has the form: $f_i(\boldsymbol{\xi}) := (\xi_1)^{k_{i1}} \cdot (\xi_2)^{k_{i2}} \cdots (\xi_d)^{k_{id}}$, where k_{ij} is a nonnegative integer indicating the power of ξ_j for the ith moment function. The framework also allows the use of generalized moments by choosing alternative base functions. Note that the constraint $\int_\varXi f_1(\xi) P(d\xi) \in [l_1(\boldsymbol{x}), u_1(\boldsymbol{x})]$ in (2) is used to ensure that P is a probability distribution. The ambiguity set (2) is a generalization of the set in [46] for the decision dependent case. The following theorem gives a reformulation of (D^{3}RO) with moment robust ambiguity set $\mathcal{P}^{SM}(\boldsymbol{x})$.

Theorem 3.1 *Let Assumption 1 hold, and therefore the ambiguity set (2) be given by* $\mathcal{P}_0^{SM}(\boldsymbol{x}) = \{\boldsymbol{p} \in \mathbb{R}^N : \boldsymbol{l}(\boldsymbol{x}) \leq \sum_{k=1}^N p_k f(\boldsymbol{\xi}^k) \leq \boldsymbol{u}(\boldsymbol{x}), \ \underline{p}_k(\boldsymbol{x}) \leq p_k \leq \overline{p}_k(\boldsymbol{x}) \ \forall k \in [N]\}$, *where* $\underline{p}_k(\boldsymbol{x})$ *and* $\overline{p}_k(\boldsymbol{x})$ *are given lower and upper bounds of each* p_k. *If for any* $\boldsymbol{x} \in X$, $h(\boldsymbol{x}, \boldsymbol{\xi}^k)$ *is finite for any* $k \in [N]$ *and the ambiguity set* $\mathcal{P}_0^{SM}(\boldsymbol{x})$ *is nonempty, then the* (D^{3}RO) *problem with the ambiguity set* $\mathcal{P}_0^{SM}(\boldsymbol{x})$ *can be reformulated as the following nonlinear program:*

$$
\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}} \quad & f(\boldsymbol{x}) + \boldsymbol{\alpha}^T \boldsymbol{l}(\boldsymbol{x}) + \boldsymbol{\beta}^T \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{\gamma}^T \underline{\boldsymbol{p}}(\boldsymbol{x}) + \boldsymbol{\mu}^T \overline{\boldsymbol{p}}(\boldsymbol{x}) \\
\text{s.t.} \quad & (\boldsymbol{\alpha} + \boldsymbol{\beta})^T \boldsymbol{f}(\boldsymbol{\xi}^k) + \gamma_k + \mu_k \geq h(\boldsymbol{x}, \boldsymbol{\xi}^k) \quad \forall k \in [N], \\
& \boldsymbol{x} \in X, \ \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^m, \ \boldsymbol{\gamma}, \boldsymbol{\mu} \in \mathbb{R}^N, \ \boldsymbol{\alpha} \leq \boldsymbol{0}, \ \boldsymbol{\beta} \geq \boldsymbol{0}, \ \boldsymbol{\gamma} \leq \boldsymbol{0}, \ \boldsymbol{\mu} \geq \boldsymbol{0}.
\end{aligned}
\tag{3}
$$

Proof Under Assumption 1, the inner problem of (D^{3}RO) becomes the following linear program:

$$\max_{\boldsymbol{p} \in \mathbb{R}^N} \sum_{k=1}^{N} p_k h(\boldsymbol{x}, \boldsymbol{\xi}^k)$$

$$\text{s.t.} \quad \boldsymbol{l}(\boldsymbol{x}) \le \sum_{k=1}^{N} p_k \boldsymbol{f}(\boldsymbol{\xi}^k) \le \boldsymbol{u}(\boldsymbol{x}),$$

$$\underline{p}_k(\boldsymbol{x}) \le p_k \le \overline{p}_k(\boldsymbol{x}) \quad \forall k \in [N]. \tag{4}$$

Based on the hypothesis of the theorem, the above linear program is feasible for any $\boldsymbol{x} \in X$, and its objective value is bounded. We take the dual of (4) using strong duality and combining the dual problem with the outer problem to get the desired reformulation. □

A reformulation for the two-stage case is given in the following corollary.

Corollary 3.1 *If $h(\cdot, \cdot)$ is a recourse function defined in* (1), *and the ambiguity set $\mathcal{P}_0^{SM}(\boldsymbol{x})$ is non-empty for any $\boldsymbol{x} \in X$, then the* (D³RO) *problem with the ambiguity set $\mathcal{P}_0^{SM}(\boldsymbol{x})$ can be formulated as follows:*

$$\min_{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}} \quad f(\boldsymbol{x}) + \boldsymbol{\alpha}^T \boldsymbol{l}(\boldsymbol{x}) + \boldsymbol{\beta}^T \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{\gamma}^T \underline{\boldsymbol{p}}(\boldsymbol{x}) + \boldsymbol{\mu}^T \overline{\boldsymbol{p}}(\boldsymbol{x})$$

$$\text{s.t.} \quad (\boldsymbol{\alpha} + \boldsymbol{\beta})^T \boldsymbol{f}(\boldsymbol{\xi}^k) + \gamma_k + \mu_k \ge g(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k) \quad \forall k \in [N]$$

$$\psi_i(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k) \ge 0 \quad \forall i \in [m], \ \forall k \in [N], \tag{5}$$

$$\boldsymbol{x} \in X, \ \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^m, \ \boldsymbol{\gamma}, \boldsymbol{\mu} \in \mathbb{R}^N, \ \boldsymbol{\alpha} \le \boldsymbol{0}, \ \boldsymbol{\beta} \ge \boldsymbol{0}, \ \boldsymbol{\gamma} \le \boldsymbol{0}, \ \boldsymbol{\mu} \ge \boldsymbol{0}.$$

Proof Recall that the complete recourse assumption ensures that $h(\boldsymbol{x}, \boldsymbol{\xi}^k)$ is finite of any given $\boldsymbol{x}$. One can take dual of the inner problem and use strong duality as in the proof of Theorem 3.1. □

Theorem 3.3, which is an analogue of Theorem 3.1 for the continuous case, can be proved using conic duality theory in functional spaces.

Theorem 3.2 (conic duality in functional spaces [47]) *Let X and Y be linear spaces, $C \subset X$ and $K \subset Y$ be convex cones, $b \in Y$ and $A : X \to Y$ be a linear mapping. The spaces X and Y are paired with linear spaces (dual spaces) X' and Y', respectively, in the sense that bilinear forms $\langle \cdot, \cdot \rangle : X' \times X \to \mathbb{R}$ and $\langle \cdot, \cdot \rangle : Y' \times Y \to \mathbb{R}$ are defined. Consider the following pair of conic linear optimization problem and its dual:*

$$\inf_{x \in C} \langle c, x \rangle \quad \text{s.t. } Ax + b \in K, \tag{P}$$

$$\sup_{y \in -K^*} \langle y, b \rangle \quad \text{s.t. } A^* y + c \in C^*, \tag{D}$$

where C^* is the polar (positive dual) cone of C, K^* is the polar cone of K, and A^* is the adjoint mapping of A. Suppose that X and Y are Banach spaces, the cones C and K are closed and $\langle c, \cdot \rangle$ and $A : X \to Y$ are continuous and $b \in \text{ri}[A(C) - K]$. Then $\text{val}(P) = \text{val}(D)$, and an optimal solution of (D) exists.

Theorem 3.3 *Let $\varXi$ be a closed and bounded set in the Euclidean space. Let the probability measure P and the positive measures $v_1(x)$, $v_2(x)$ be defined on the measurable space $(\varXi, \mathcal{F})$, where the σ-algebra $\mathcal{F}$ contains all singleton subsets, i.e., $\{\xi\} \in \mathcal{F}$ for all $\xi \in \varXi$. If for any $x \in \varXi$, the inner problem of (D^3RO) with the ambiguity set $\mathcal{P}^{SM}(x)$ has a non-empty relative interior, then the inner problem has the following dual reformulation:*

$$\min_{\alpha, \beta, \gamma, \mu} \alpha^\top l(x) + \beta^\top u(x) + \int_\varXi \gamma^\top(\xi) v_1(x, \xi) d\xi + \int_\varXi \mu^\top(\xi) v_2(x, \xi) d\xi$$
$$\text{s.t. } (\alpha + \beta)^\top f(\xi) + \gamma(\xi) + \mu(\xi) = h(x, \xi) \quad \forall \xi \in \varXi,$$
$$x \in X, \ \alpha, \beta \in \mathbb{R}^m, \ \gamma(\xi), \mu(\xi) \in \mathbb{R} \ \forall \xi \in \varXi, \ \alpha \le 0, \ \beta \ge 0,$$
$$\gamma(\xi) \le 0, \ \mu(\xi) \ge 0.$$

Furthermore, the strong duality holds and a dual optimal solution exists.

Proof For notational convenience, we omit the x variable in the proof. Let $V^+(\varXi, \mathcal{F})$ be the cone of of positive measures on $(\varXi, \mathcal{F})$. The inner problem of the original model is a conic linear program in a functional space, which we rewrite as follows:

$$\sup_P \ \mathbb{E}_P[h(\xi)]$$
$$\text{s.t. } v_1 \preceq P \preceq v_2, \tag{6}$$
$$l_i \le \int_\varXi f_i(\xi) P(d\xi) \le u_i, \quad i \in [m],$$

where $P \in V^+(\varXi, \mathcal{F})$. We write a dual of (6) as follows:

$$\inf_{\alpha, \beta, \gamma, \mu} \ \alpha^\top l + \beta^\top u + \int_\varXi \gamma(\xi) dv_1(\xi) + \int_\varXi \mu(\xi) dv_2(\xi)$$
$$\text{s.t. } (\alpha + \beta)^\top f(\xi) + \gamma(\xi) + \mu(\xi) = h(\xi) \quad \forall \xi \in \varXi,$$
$$\alpha, \beta \in \mathbb{R}^m, \ \alpha \le 0, \ \beta \ge 0, \ \gamma(\xi) \le 0, \ \mu(\xi) \ge 0 \ \forall \xi \in \varXi. \tag{7}$$

We now verify that the conditions in Theorem 3.2 hold for (6) and (7). The primal decision variable P is an element in the the following Banach space of finite signed measures over $(\varXi, \mathcal{F})$:

$$\mathcal{M}(\varXi, \mathcal{F}) = \{\mu \mid \mu \text{ is a finite signed measure on } (\varXi, \mathcal{F})\}.$$

The Banach space $\mathcal{M}(\varXi, \mathcal{F})$ is equipped with the norm $\|\mu\| = |\mu|(X)$, where $\mu = \mu^+ - \mu^-$, $|\mu| = \mu^+ + \mu^-$. The μ^+, μ^- are the positive and negative components of μ. The dual decision variable $[\alpha, \beta, \gamma, \mu]$ is an element in the Banach space $\mathbb{R}^m \times \mathbb{R}^m \times L \times L$, where L is a Banach space defined as

$$L = \{\psi \mid \psi \text{ is a measurable function in } (\varXi, \mathcal{F}), \text{ such that } \max_{\xi \in \varXi} |\psi(\xi)| < \infty\}.$$

The primal problem can be rewritten as the following linear conic optimization problem:

$$
\begin{aligned}
&\sup_{P \in \mathcal{M}(\varXi, \mathcal{F})} \quad \langle h, P \rangle \\
&\text{s.t. } \langle f_i, P \rangle - l_i \in \mathbb{R}_+ \quad i \in [m], \qquad -\langle f_i, P \rangle + u_i \in \mathbb{R}_+ \quad i \in [m], \\
&\quad\;\; P - v_1 \in \mathcal{M}^+(\varXi, \mathcal{F}), \qquad -P + v_2 \in \mathcal{M}^+(\varXi, \mathcal{F}), \qquad P \in \mathcal{M}^+(\varXi, \mathcal{F}),
\end{aligned}
\tag{8}
$$

where $\langle \psi, P \rangle = \int_{\varXi} \psi\, dP$. The linear mapping $A : \mathcal{M}(\varXi, \mathcal{F}) \to \mathbb{R}^m \times \mathbb{R}^m \times \mathcal{M}(\varXi, \mathcal{F}) \times \mathcal{M}(\varXi, \mathcal{F})$ is given as:

$$
A \circ P := [\langle f, P \rangle,\ -\langle f, P \rangle,\ P,\ -P].
$$

The dual problem can be rewritten as the following linear conic optimization problem:

$$
\begin{aligned}
&\inf_{\substack{[\alpha,\beta,\gamma,\mu] \in \\ \mathbb{R}^m \times \mathbb{R}^m \times L \times L}} \quad \langle [l, u, v_1, v_2], [\alpha, \beta, \gamma, \mu] \rangle \\
&\text{s.t. } (\alpha + \beta)^\top f(\xi) + \gamma(\xi) + \mu(\xi) - h(\xi) \in \{0\} \quad \forall \xi \in \varXi, \\
&\quad\;\; [\alpha, \beta, \gamma, \mu] \in \mathbb{R}^m_- \times \mathbb{R}^m_+ \times L^- \times L^+,
\end{aligned}
\tag{9}
$$

where $L^+ = \{\psi \in L \mid \psi(\xi) \geq 0 \ \forall \xi \in \varXi\}$ and $L^- = \{\psi \in L \mid \psi(\xi) \leq 0 \ \forall \xi \in \varXi\}$.

By definition, it is clear that the cones $\mathcal{M}^+(\varXi, \mathcal{F})$ and $\mathbb{R}^m_+ \times \mathbb{R}^m_+ \times \mathcal{M}^+(\varXi, \mathcal{F}) \times \mathcal{M}^+(\varXi, \mathcal{F})$ are closed. To show that the linear functional $\langle h, \cdot \rangle$ and the linear mapping A is continuous, it suffices to show that if $P_n \to 0$, we have $\langle h, P_n \rangle \to 0$ and $A \circ P_n \to 0$ as $n \to \infty$. Indeed, if $P_n \to 0$, we have

$$
|\langle h, P_n \rangle| = \left| \int_{\varXi} h(\xi)\, dP_n(\xi) \right| \leq \max_{\xi \in \varXi} h(\xi) \cdot P_n(\varXi) \to 0, \text{ and similarly, } |\langle f_i, P_n \rangle| \to 0.
$$

Therefore, $\langle h, P_n \rangle \to 0$ and $A \circ P_n \to 0$ as $n \to \infty$. Since the primal problem has a non-empty relative interior, by Theorem 3.2, the strong duality holds and the dual problem has an optimal solution. □

3.2 Covariance matrix based ambiguity set

We now consider a distributional ambiguity set with multi-variate bounds defined as follows:

$$
\mathcal{P}^{DY}(x) := \left\{ P \in \mathcal{P}(\varXi, \mathcal{F}) \;\middle|\; \begin{array}{c} (\mathbb{E}_P[\xi] - \mu(x))^T Q(x)^{-1} (\mathbb{E}_P[\xi] - \mu(x)) \leq \alpha(x) \\ \mathbb{E}_P[(\xi - \mu(x))(\xi - \mu(x))^T] \preceq \beta(x) Q(x) \end{array} \right\}.
\tag{10}
$$

This set is a generalization of the set used in [6] for the decision dependent case. In the finite support case this set is written as follows:

$$
\mathcal{P}_0^{DY}(x)
$$

$$
:= \left\{ p \in [N] \left|
\begin{array}{l}
\displaystyle\sum_{i=1}^{N} p_i = 1, \quad \left(\displaystyle\sum_{i=1}^{N} p_i' \xi^i - \mu(x)\right)^T Q(x)^{-1} \left(\displaystyle\sum_{i=1}^{N} p_i \xi^i - \mu(x)\right) \le \alpha(x), \\[2mm]
\displaystyle\sum_{i=1}^{N} p_i (\xi^i - \mu(x))(\xi^i - \mu(x))^T \preceq \beta(x) Q(x)
\end{array}
\right. \right\}
$$

(11)

Note that in a special case of (10), $\mu(x)$ and $Q(x)$ may not depend on x and only $\alpha(x)$, $\beta(x)$ depend on the decision variables. In this case, the decision only affects the size of the ambiguity set. Note that in $\mathcal{P}_0^{DY}(x)$ we can also allow the probabilities p_i to depend on x as in $\mathcal{P}^{DY}(x)$. This possible generalization is ignored here for simplicity. The following theorem gives a reformulation of (D³RO) with the ambiguity set $\mathcal{P}_0^{DY}(x)$. The following theorem is based on Lemma 1 in [6] for the finite support and decision dependent case.

Theorem 3.4 *Let Assumption 1 hold. Suppose that Slater's constraint qualification conditions are satisfied, i.e., for any $x \in X$, there exist a vector p' in the interior of $\mathcal{P}^{DY}(x)$. Then the (D³RO) problem with the ambiguity set $\mathcal{P}^{MB}(x)$ can be reformulated as:*

$$
\min_{x,s,u,z,Y} \quad f(x) + s + [\sqrt{\alpha(x)}, -(Q(x)^{-1/2}\mu(x))^T]z + \beta(x)Q(x) \bullet Y
$$
$$
\text{s.t.} \quad s - (\xi^i)^T Q(x)^{-1/2} z_1 + (\xi^i - \mu(x))(\xi^i - \mu(x))^T \bullet Y \ge h(x, \xi^i) \quad \forall i \in [N],
$$
$$
x \in X, \ z := [z_0, z_1] \in \mathcal{N}_{SOC}, \ Y \in \mathbb{R}^{N \times N}, \ Y \succeq 0,
$$

(12)

where $\mathcal{N}_{SOC}$ is a second order cone defined as $\mathcal{N}_{SOC} := \left\{ y := [y_0, y_1, \ldots, y_d] \left\| y_0 \ge \sqrt{\sum_{i=1}^{d} y_i^2} \right. \right\}$, $z := [z_0, \ z_1]$ with $z_0 \in \mathbb{R}$ and $z_1 \in \mathbb{R}^d$, and $A \bullet B = Tr(A^T B)$ for matrices A and B.

Proof With the ambiguity set $\mathcal{P}_0^{DY}$ the inner problem of (D³RO) is given as follows:

$$
\max_{p,\tau} \quad \sum_{i=1}^{N} p_i h(x, \xi^i)
$$
$$
\text{s.t.} \quad \sum_{i=1}^{N} p_i = 1, \qquad\qquad\qquad\qquad : s \in \mathbb{R},
$$

$$\boldsymbol{\tau} = \sum_{i=1}^{N} p_i \boldsymbol{\xi}^i, \qquad\qquad\qquad : \boldsymbol{u} \in \mathbb{R}^d,$$

$$(\boldsymbol{\tau} - \boldsymbol{\mu}(\boldsymbol{x}))^T \boldsymbol{Q}(\boldsymbol{x})^{-1} (\boldsymbol{\tau} - \boldsymbol{\mu}(\boldsymbol{x})) \le \alpha(\boldsymbol{x}), \qquad : \boldsymbol{z} \in \mathcal{N}_{\text{SOC}},$$

$$\sum_{i=1}^{N} p_i (\boldsymbol{\xi}^i - \boldsymbol{\mu}(\boldsymbol{x})) (\boldsymbol{\xi}^i - \boldsymbol{\mu}(\boldsymbol{x}))^T \preceq \beta(\boldsymbol{x}) \boldsymbol{Q}(\boldsymbol{x}), \qquad : \boldsymbol{Y} \succeq 0,$$

$$\boldsymbol{p} := [p_1, \ldots, p_N] \in \mathbb{R}_+^N, \qquad\qquad\qquad \forall i \in [N]. \qquad (13)$$

Under the Slater's condition using the strong duality for semi-definite programming, we reformulate the dual problem as:

$$\min_{s, \boldsymbol{u}, \boldsymbol{z}, \boldsymbol{Y}} \quad s + \beta(\boldsymbol{x}) \boldsymbol{Q}(\boldsymbol{x}) \bullet \boldsymbol{Y} + \left[\sqrt{\alpha(\boldsymbol{x})}, - \left(\boldsymbol{Q}(\boldsymbol{x})^{-1/2} \boldsymbol{\mu}(\boldsymbol{x}) \right)^T \right] \boldsymbol{z}$$

$$\text{s.t.} \quad s + \boldsymbol{u}^T \boldsymbol{\xi}^i + (\boldsymbol{\xi}^i - \boldsymbol{\mu}(\boldsymbol{x})) (\boldsymbol{\xi}^i - \boldsymbol{\mu}(\boldsymbol{x}))^T \bullet \boldsymbol{Y} \ge h(\boldsymbol{x}, \boldsymbol{\xi}^i) \quad \forall i \in [N], \quad (14)$$

$$\boldsymbol{u} + \boldsymbol{Q}(\boldsymbol{x})^{-1/2} \boldsymbol{z}_1 = 0,$$

$$\boldsymbol{z} \in \mathcal{N}_{\text{SOC}}, \ \boldsymbol{Y} \in \mathbb{R}^{d \times d}, \ \boldsymbol{Y} \succeq 0.$$

Substituting (14) in (D^{3}RO), we obtain (12). $\qquad\qquad\qquad\qquad\qquad \square$

Corollary 3.2 *If $h(\cdot, \cdot)$ is a recourse function defined in (1) and the ambiguity set $\mathcal{P}_0^{DY}(\boldsymbol{x})$ is non-empty for any $\boldsymbol{x} \in X$, then the (D^{3}RO) problem with the ambiguity set $\mathcal{P}_0^{DY}(\boldsymbol{x})$ can be reformulated as follows:*

$$\min_{\boldsymbol{x}, s, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{z}, \boldsymbol{Y}} \quad f(\boldsymbol{x}) + s + [\sqrt{\alpha(\boldsymbol{x})}, -(\boldsymbol{Q}(\boldsymbol{x})^{-1/2} \boldsymbol{\mu}(\boldsymbol{x}))^T] \boldsymbol{z} + \beta(\boldsymbol{x}) \boldsymbol{Q}(\boldsymbol{x}) \bullet \boldsymbol{Y}$$

$$\text{s.t.} \quad s - (\boldsymbol{\xi}^k)^T \boldsymbol{Q}(\boldsymbol{x})^{-1/2} \boldsymbol{z}_1 + (\boldsymbol{\xi}^k - \boldsymbol{\mu}(\boldsymbol{x})) (\boldsymbol{\xi}^k - \boldsymbol{\mu}(\boldsymbol{x}))^T \bullet \boldsymbol{Y} \ge g(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k)$$

$$\forall k \in [N],$$

$$\psi_i(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k) \ge 0 \quad \forall i \in [m], \ \forall k \in [N],$$

$$\boldsymbol{x} \in X, \ \boldsymbol{z} := [z_0, \boldsymbol{z}_1] \in \mathcal{N}_{\text{SOC}}, \ \boldsymbol{Y} \succeq 0.$$

$$(15)$$

The following theorem, which is an analogue of Theorem 3.4, can be proved using conic duality.

Theorem 3.5 *Let Ξ be a closed and bounded set in the Euclidean space. Suppose the σ-algbra $\mathcal{F}$ contains all singleton sets, i.e., $\{\xi\} \in \mathcal{F}$ for all $\xi \in \Xi$. Suppose the following Slater-type conditions are satisfied: For all $\boldsymbol{x} \in X$, there exists a probability measure $P \in \mathcal{P}(\Xi, \mathcal{F})$ satisfying $(\mathbb{E}_P[\boldsymbol{\xi}] - \boldsymbol{\mu}(\boldsymbol{x}))^T \boldsymbol{Q}(\boldsymbol{x})^{-1} (\mathbb{E}_P[\boldsymbol{\xi}] - \boldsymbol{\mu}(\boldsymbol{x})) < \alpha(\boldsymbol{x})$ and $\mathbb{E}_P[(\boldsymbol{\xi} - \boldsymbol{\mu}(\boldsymbol{x}))(\xi - \boldsymbol{\mu}(\boldsymbol{x}))^T] \prec \beta(\boldsymbol{x}) \boldsymbol{Q}(\boldsymbol{x})$. Then problem (D^{3}RO) with the ambiguity*

set $\mathcal{P}^{DY}(\pmb{x})$ can be reformulated as

$$
\min_{\pmb{x},s,\pmb{u},z,\pmb{Y}} \quad f(\pmb{x}) + s + [\sqrt{\alpha(\pmb{x})}, -(\pmb{Q}(\pmb{x})^{-1/2}\pmb{\mu}(\pmb{x}))^T]z + \beta(\pmb{x})\pmb{Q}(\pmb{x}) \bullet \pmb{Y}
$$
$$
s.t. \quad s - \pmb{\xi}^T \pmb{Q}(\pmb{x})^{-1/2} z_1 + (\pmb{\xi} - \pmb{\mu}(\pmb{x}))(\pmb{\xi} - \pmb{\mu}(\pmb{x}))^T \bullet \pmb{Y} \geq h(\pmb{x}, \pmb{\xi}) \quad \forall \pmb{\xi} \in \Xi,
$$
$$
\pmb{x} \in X, \ \ z := [z_0, z_1] \in \mathcal{N}_{\text{SOC}}, \ \pmb{Y} \in \mathbb{R}^{d \times d}, \ \pmb{Y} \succeq 0.
$$

Proof For the notational convenience, we omit the $\pmb{x}$ argument in all functions and parameters in the proof. Let $\mathcal{M}^+(\Xi, \mathcal{F})$ be the cone of positive measures defined on the measurable space $(\Xi, \mathcal{F})$. The inner problem of (D³RO) under the uncertainty set $\mathcal{P}^{DY}$ is written as:

$$
\begin{aligned}
&\max \ \mathbb{E}_P[h(\xi)] \\
&\text{s.t. } (\mathbb{E}_P[\xi] - \mu)^\top Q^{-1}(\mathbb{E}_P[\xi] - \mu) \leq \alpha, \\
&\quad\quad \mathbb{E}_P[(\xi - \mu)(\xi - \mu)^\top] \preceq \beta Q, \\
&\quad\quad \mathbb{E}_P[1] = 1, \ P \in \mathcal{M}^+(\Xi, \mathcal{F}).
\end{aligned} \tag{16}
$$

Define the linear mapping $\mathcal{A}_1 : \mathcal{M}^+(\Xi, \mathcal{F}) \longmapsto \mathbb{R}$ as $\mathcal{A}_1 P := \int_\Xi 1 dP(\xi)$, define the linear mapping $\mathcal{A}_2 : \mathcal{M}^+(\Xi, \mathcal{F}) \longmapsto \mathbb{R}^{m+1}$ as $\mathcal{A}_2 P := [0, Q^{-1/2} \int_\Xi \xi dP(\xi)]$, and define the linear mapping $\mathcal{A}_3 : \mathcal{M}^+(\Xi, \mathcal{F}) \longmapsto \mathbb{R}^{m \times m}$ as $\mathcal{A}_3 P := -\int_\Xi (\xi - \mu)(\xi - \mu)^\top dP(\xi)$. Define the linear functional $\langle h, \cdot \rangle : \mathcal{M}^+(\Xi, \mathcal{F}) \longmapsto \mathbb{R}$ as $\langle h, P \rangle := \int_\Xi h(\xi) dP(\xi)$. Let $b = [\sqrt{\alpha}, \ 0]$ be a constant vector in $\mathbb{R}^{m+1}$. Then (16) can be reformulated as the following conic linear program:

$$
\begin{aligned}
&\max \ \langle h, P \rangle \\
&\text{s.t. } \mathcal{A}_1 P - 1 = 0, \\
&\quad\quad \mathcal{A}_2 P + b \in \mathcal{K}_{SOC}, \\
&\quad\quad \mathcal{A}_3 P + \beta Q \in \mathcal{K}_{SD}, \\
&\quad\quad P \in \mathcal{M}^+(\Xi, \mathcal{F}),
\end{aligned} \tag{17}
$$

where $\mathcal{K}_{SOC}$ and $\mathcal{K}_{SD}$ are the second-order cone and the positive semi-definite cone, respectively. Applying the duality theory [47] of the conic linear program in the functional space, we can obtain the dual of (16) as given in the statement of the theorem. The slater-type condition is equivalent to that there exists a probability measure P' such that $\mathcal{A}_2 P' + b \in \text{int}(\mathcal{K}_{SOC})$, and $\mathcal{A}_3 P' + \beta Q \in \text{int}(\mathcal{K}_{SD})$. Based on (23) in [47], the statement after (23), and Proposition 2.9 of [47], the strong duality holds. $\qquad\square$

3.3 Ambiguity sets defined by Wasserstein metric

Instead of using moment based definitions of the ambiguity set, we may define this set using a statistical distance, such as the Wasserstein metric. We now study the (D³RO) problem with a decision dependent ambiguity set defined using the L_1-Wasserstein metric as follows:

$$\mathcal{P}^W(\boldsymbol{x}) := \{P \in \mathcal{P}(\varXi, \mathcal{F}) \mid \mathcal{W}(P, P_0) \le r(\boldsymbol{x})\}, \tag{18}$$

where P_0 is a nominal probability distribution, and $\mathcal{W}(\cdot, \cdot) : \mathcal{P}(\varXi, \mathcal{F}) \times \mathcal{P}(\varXi, \mathcal{F}) \to \mathbb{R}$ is the L_1-Wasserstein metric defined in [48]:

$$\mathcal{W}(P_1, P_2) := \inf_{K \in \mathcal{S}(P_1, P_2)} \int_{\varXi \times \varXi} \|\boldsymbol{s}_1 - \boldsymbol{s}_2\| K(d\boldsymbol{s}_1 \times d\boldsymbol{s}_2), \tag{19}$$

where $\mathcal{S}(P_1, P_2) := \{K \in \mathcal{P}(\varXi \times \varXi, \mathcal{F} \times \mathcal{F}) : K(A \times \varXi) = P_1(A), \ K(\varXi \times A) = P_2(A), \ \forall A \in \mathcal{F}\}$ is the set of all joint probability distributions whose marginals are P_1 and P_2, and $\| \cdot \|$ is an arbitrary norm defined on $\mathbb{R}^d$. The ambiguity set (18) is a generalization of the one considered in [7–10] for the decision dependent case. As a special case of (18), under Assumption 1, $\mathcal{P}^W(\boldsymbol{x})$ is written as:

$$\mathcal{P}^W(\boldsymbol{x}) = \left\{ \boldsymbol{p} \in \mathbb{R}^N \mid \mathcal{W}(\boldsymbol{p}, \hat{\boldsymbol{p}}) \le r(\boldsymbol{x}), \ \sum_{i=1}^N p_i = 1, \ p_i \ge 0, \ \forall i \in [N] \right\}, \tag{20}$$

where $\hat{\boldsymbol{p}}$ is a given empirical probability distribution on $\varXi$, and the Wasserstein metric can be simplified as

$$\mathcal{W}(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \left\{ \min_{\boldsymbol{w}} \sum_{i=1}^N \sum_{j=1}^N \|\boldsymbol{\xi}^i - \boldsymbol{\xi}^j\| w_{ij} \ \middle| \ \sum_{j=1}^N w_{ij} = p_i \ \forall i \in [N], \right.$$
$$\left. \sum_{i=1}^N w_{ij} = \hat{p}_j \ \forall j \in [N], \ w_{ij} \ge 0 \ \forall i, j \in [N] \right\},$$

where $\hat{p}_j$ is the probability of scenario $\boldsymbol{\xi}^j$. The following theorem gives a reformulation of (D³RO) for the ambiguity set (20).

Theorem 3.6 *Let Assumption 1 hold. If for any $\boldsymbol{x} \in X$, $h(\boldsymbol{x}, \boldsymbol{\xi}^i)$ is finite for any $i \in [N]$ and the ambiguity set* (20) *is nonempty, then* (D³RO) *problem with the ambiguity set* (20) *can be reformulated as:*

$$\min_{\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \gamma, \eta} f(\boldsymbol{x}) - \sum_{i=1}^N \hat{p}_i \beta_i - r(\boldsymbol{x})\gamma + \eta$$
$$\text{s.t. } -\alpha_i - \mu_i + \eta \ge h(\boldsymbol{x}, \boldsymbol{\xi}^i) \quad \forall i \in [N],$$
$$-\alpha_i + \beta_j + \|\boldsymbol{\xi}^i - \boldsymbol{\xi}^j\|\gamma + \lambda_{ij} \le 0 \quad \forall i \in [N], \ \forall j \in [N],$$
$$\boldsymbol{x} \in X, \ \alpha_i \in \mathbb{R}, \ \beta_i \in \mathbb{R}, \ \mu_i \ge 0, \ \lambda_{ij} \ge 0, \ \gamma \le 0, \ \eta \in \mathbb{R}$$
$$\forall i \in [N], \ \forall j \in [N]. \tag{21}$$

Proof Since $\varXi$ is finite, the (D³RO)-inner problem with ambiguity set $\mathcal{P}^W(\boldsymbol{x})$ can be formulated as the following linear program:

$$\max_{\boldsymbol{p},\boldsymbol{w}} \ \sum_{i=1}^{N} h(\boldsymbol{x},\boldsymbol{\xi}^i)p_i$$

$$\text{s.t.} \ \sum_{j=1}^{N} w_{ij} = p_i \quad \forall i \in [N],$$

$$\sum_{i=1}^{N} w_{ij} = \hat{p}_j \quad \forall j \in [N], \tag{22}$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \|\boldsymbol{\xi}^i - \boldsymbol{\xi}^j\| w_{ij} \le r(\boldsymbol{x}),$$

$$\sum_{i=1}^{N} p_i = 1, \ p_i \ge 0, \ w_{ij} \ge 0 \ \forall i \in [N], \ \forall j \in [N],$$

where $\boldsymbol{w}$ is a joint probability distribution with two marginal distributions given by $\boldsymbol{p}$ and $\hat{\boldsymbol{p}}$, respectively. The dual of the above linear program is:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda},\gamma,\eta} \ -\sum_{i=1}^{N} \hat{\boldsymbol{p}}_i \beta_i - r(\boldsymbol{x})\gamma + \eta$$

$$\text{s.t.} \ -\alpha_i - \mu_i + \eta \ge h(\boldsymbol{x},\boldsymbol{\xi}^i) \quad \forall i \in [N], \tag{23}$$

$$-\alpha_i + \beta_j + \|\boldsymbol{\xi}^i - \boldsymbol{\xi}^j\|\gamma + \lambda_{ij} \le 0 \quad \forall i \in [N], \ \forall j \in [N],$$

$$\alpha_i \in \mathbb{R}, \ \beta_i \in \mathbb{R}, \ \mu_i \ge 0, \ \lambda_{ij} \ge 0, \ \gamma \le 0, \ \eta \in \mathbb{R} \ \forall i \in [N], \ \forall j \in [N].$$

After substituting (23) into (D^{3}RO), we obtain the desired reformulation (21).   □

A reformulation for the two-stage case is given in the following corollary.

Corollary 3.3 *If $h(\cdot,\cdot)$ is a recourse function defined in* (1), *and the ambiguity set $\mathcal{P}^W(\boldsymbol{x})$ is non-empty for any $\boldsymbol{x} \in X$, then the* (D^{3}RO) *problem with the ambiguity set $\mathcal{P}^W(\boldsymbol{x})$ can be reformulated as follows:*

$$\min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\lambda},\gamma} \ f(\boldsymbol{x}) - \sum_{k=1}^{N} \hat{\boldsymbol{p}}_k \beta_k - r(\boldsymbol{x})\gamma$$

$$\text{s.t.} \ -\alpha_k - \mu_k \ge g(\boldsymbol{x},\boldsymbol{y}^k,\boldsymbol{\xi}^k) \quad \forall k \in [N],$$

$$-\alpha_i + \beta_j + \|\boldsymbol{\xi}^i - \boldsymbol{\xi}^j\|\gamma + \lambda_{ij} \le 0 \quad \forall i \in [N], \ \forall j \in [N], \tag{24}$$

$$\psi_i(\boldsymbol{x},\boldsymbol{y}^k,\boldsymbol{\xi}^k) \ge 0 \ \forall i \in [m], \ \forall k \in [N],$$

$$\boldsymbol{x} \in X, \ \alpha_i \in \mathbb{R}, \ \beta_i \in \mathbb{R}, \ \mu_i \ge 0, \ \lambda_{ij} \ge 0, \ \gamma \le 0 \ \forall i \in [N], \ \forall j \in [N].$$

A generalization of (21) of (D^{3}RO) for the continuous support case is given in Appendix A.

3.4 Ambiguity sets defined using ϕ-divergence

We now study the (D³RO) problem using a decision dependent ambiguity set defined using the notion of ϕ-divergence:

$$\mathcal{P}^\phi(\mathbf{x}) := \left\{ P \in \mathcal{P}(\Xi, \mathcal{F}) : \ \mathcal{D}_\phi(P \| P_0) \leq \eta(\mathbf{x}) \right\}, \tag{25}$$

where $\mathcal{D}_\phi(P \| P_0) = \int_\Omega \phi\left(\frac{dP}{dP_0}\right) dP_0$, and ϕ is a non-negative and convex function. This type of ambiguity set is a generalization of the one considered in [11,12,14–16,49] for the decision dependent case. Under Assumption 1, and letting $\hat{p}_i$ be the probability of scenario $\boldsymbol{\xi}_i$, the ambiguity set (25) is written as:

$$\mathcal{P}^\phi(\mathbf{x}) = \left\{ P = \sum_{i=1}^N p_i \delta_{\boldsymbol{\xi}^i} : \ \sum_{i=1}^N \hat{p}_i \phi(p_i/\hat{p}_i) \leq \eta(\mathbf{x}), \ \sum_{i=1}^N p_i = 1, \ p_i \geq 0 \ \forall i \leq N \right\}. \tag{26}$$

Two reformulations of (D³RO) with ambiguity set $\mathcal{P}^\phi(\mathbf{x})$ are given in the following theorem.

Theorem 3.7 *Let Assumption 1 hold, and ϕ be a non-negative convex function. Assume that the following Slater's condition is satisfied for every $\mathbf{x} \in X$: there exist a $\mathbf{p} \in \mathbb{R}^N$ such that $p_i > 0$, $\sum_{i=1}^N p_i = 1$ and $\sum_{i=1}^N \hat{p}_i \phi(p_i/\hat{p}_i) < \eta(\mathbf{x})$. Then (D³RO) with the ambiguity set $\mathcal{P}^\phi(\mathbf{x})$ can be reformulated as the following semi-infinite program:*

$$\min_{\mathbf{x}, \mathbf{p}, \alpha, \beta, \lambda, z} z$$

$$\text{s.t. } z \geq \sum_{i=1}^N h(\mathbf{x}, \boldsymbol{\xi}^i) p_i + \alpha\left(\frac{1}{N} \sum_{i=1}^N \phi(N p_i) - \eta(\mathbf{x})\right)$$

$$+ \beta\left(\sum_{i=1}^N p_i - 1\right) + \sum_{i=1}^N p_i \lambda_i \quad \forall \mathbf{p} \in S, \tag{27}$$

$$\mathbf{x} \in X, \ \alpha \leq 0, \ \beta \in \mathbb{R}, \ \lambda_i \geq 0, \ p_i \geq 0 \ \forall i \in [N],$$

where $S = \{\mathbf{p} \in \mathbb{R}^N : \ \sum_{i=1}^N p_i = 1, \ p_i \geq 0 \ \forall i \in [N]\}$. Alternatively, (27) also has the reformulation:

$$\min_{\mathbf{x}, \mathbf{p}, \alpha, \beta, \lambda} f(\mathbf{x}) + \sum_{i=1}^N p_i h(\mathbf{x}, \boldsymbol{\xi}^i) + \alpha\left(\frac{1}{N} \sum_{i=1}^N \phi(N p_i) - \eta(\mathbf{x})\right)$$

$$+ \beta\left(\sum_{i=1}^N p_i - 1\right) + \sum_{i=1}^N \lambda_i p_i \tag{28}$$

$$\text{s.t. } \alpha \phi'(N p_i) + \beta + h(\mathbf{x}, \boldsymbol{\xi}^i) + \lambda_i = 0 \quad \forall i \in [N],$$

$$\mathbf{x} \in X, \ \alpha \leq 0, \ \beta \in \mathbb{R}, \ \lambda_i \geq 0, \ p_i \geq 0 \ \forall i \in [N].$$

Proof The (D^3RO) problem can be written as $\min_{x \in X} f(x) + \Phi(x)$, where the function $\Phi(x)$ is the optimal objective of the following optimization problem:

$$\Phi(x) = \max_{p} \sum_{i=1}^{N} p_i h(x, \xi^i)$$

$$\text{s.t.} \frac{1}{N} \sum_{i=1}^{N} \phi(Np_i) \leq \eta(x), \quad \sum_{i=1}^{N} p_i = 1, \tag{29}$$

$$p_i \geq 0 \ \forall i \leq N, \ x \in X.$$

Since ϕ is convex, (29) is a convex program with respect to the decision variable p. For a fixed $x \in X$, the Lagrangian dual of (29) is written as follows:

$$\min_{\alpha, \beta, \lambda} \max_{p} \mathcal{L}(p; \alpha, \beta, \lambda)$$

$$\text{s.t.} \alpha \leq 0, \ \beta \in \mathbb{R}, \ \lambda_i \geq 0, \quad \forall i \in [N], \tag{30}$$

where $\mathcal{L}(p; \alpha, \beta, \lambda) = \sum_{i=1}^{N} h(x, \xi^i) p_i + \alpha \left(\frac{1}{N} \sum_{i=1}^{N} \phi(Np_i) - \eta(x) \right) + \beta \left(\sum_{i=1}^{N} p_i - 1 \right) + \sum_{i=1}^{N} p_i \lambda_i$, and α, β, λ are the Lagrange multipliers. Since Slater's condition is satisfied for any $x \in X$, strong duality holds. The inner maximization problem of (30) is equivalent to {max z, s.t. $z \geq \mathcal{L}(p; \alpha, \beta, \lambda) \ \forall p \in S$}, which gives the reformulation (27).

Note that the inner problem of (30) is an unconstrained convex optimization problem. Using the KKT optimality conditions we have:

$$\frac{\partial \mathcal{L}}{\partial p_i} = \alpha \phi'(Np_i) + h(x, \xi^i) + \beta + \lambda_i = 0, \quad \forall i \in [N]. \tag{31}$$

After substituting the expression of the Lagrangian in (30), adding the optimality condition and using strong duality, we obtain the reformulation given in (28). □

We remark that in the cutting-surface algorithm (Algorithm 1) given in Sect. 4, the separation problem is specified over the set S, and it maximize $\sum_{i=1}^{N} h(x, \xi^i) p_i + \alpha \left(\frac{1}{N} \sum_{i=1}^{N} \phi(Np_i) - \eta(x) \right) + \beta \left(\sum_{i=1}^{N} p_i - 1 \right) + \sum_{i=1}^{N} p_i \lambda_i$ over $p \in S$, where $x, \alpha, \beta, \lambda$ are the current solution obtained from solving the master problem at an iteration. Since $\alpha \leq 0$, and ϕ is a convex function, the separation problem is maximizing a concave function, which is a convex optimization problem.

A reformulation for the two-stage stochastic optimization case is given in the following corollary.

Corollary 3.4 *If $h(\cdot, \cdot)$ is a recourse function defined in (1) and the ambiguity set $\mathcal{P}^\phi(x)$ is non-empty for any $x \in X$, then the (D^3RO) problem with the ambiguity set $\mathcal{P}^\phi(x)$ can be reformulated as follows:*

$$\min_{\substack{\boldsymbol{x},\boldsymbol{y},z,\boldsymbol{p},\\ \alpha,\beta,\boldsymbol{\lambda}}} \quad z$$

$$\begin{aligned}
\text{s.t. } z \geq f(\boldsymbol{x}) &+ \sum_{k=1}^{N} p_k g(\boldsymbol{x}, \boldsymbol{y}^k, \xi^k) \\
&+ \alpha\Big(\frac{1}{N}\sum_{k=1}^{N}\phi(Np_k) - \eta(\boldsymbol{x})\Big) + \beta\Big(\sum_{k=1}^{N} p_k - 1\Big) + \sum_{k=1}^{N}\lambda_k p_k \ \forall \boldsymbol{p} \in S,
\end{aligned} \tag{32}$$

$$\psi_i(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k) \geq 0 \quad \forall i \in [m], \ \forall k \in [N],$$

$$z \in \mathbb{R}, \ \boldsymbol{x} \in X, \ \alpha \leq 0, \ \beta \in \mathbb{R}, \ \lambda_k \geq 0, \ p_k \geq 0 \ \forall k \in [N],$$

where $S = \{\boldsymbol{p} \in \mathbb{R}^N : \sum_{k=1}^{N} p_k = 1, \ p_k \geq 0 \ \forall k \in [N]\}$.

3.5 Ambiguity sets defined based on the Kolmogorov–Smirnov test

The K–S distance was used by [17] in defining an ambiguity set for a data-driven distributionally robust optimization model. For two univariate probability distributions P_1 and P_2, let F_1 and F_2 be their cumulative distribution functions. The Kolmogorov–Smirnov (KS) distance is defined as:

$$D(P_1, P_2) = \sup_s |F_1(s) - F_2(s)|. \tag{33}$$

We now study the (D³RO) problem with the ambiguity set defined based on the KS-distance. Note that although (33) is defined for a univariate random variable, this definition can be directly generalized for the probability distribution of a random vector with a finite support. Specifically, under Assumption 1, let $P_0 = \sum_{i=1}^{N} \hat{p}_i \delta_{\boldsymbol{\xi}^i}$ be an empirical probability distribution, where $\delta_{\boldsymbol{\xi}^i}$ is the indicator function. The KS-distance between a discrete probability distribution $P = \sum_{i=1}^{N} p_i \delta_{\boldsymbol{\xi}^i}$ and P_0 can be written as:

$$D(P, P_0) = \sup_{k \in [N]} \Big| \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} \hat{p}_i \Big|. \tag{34}$$

The decision dependent ambiguity set of probability distributions is constructed using the KS-distance as follows:

$$\mathcal{P}^{KS}(\boldsymbol{x}) = \left\{ \boldsymbol{p} \in \mathbb{R}^N : \ \sup_{k \in [N]} \Big| \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} \hat{p}_i \Big| \leq \eta(\boldsymbol{x}), \ \sum_{i=1}^{N} p_i = 1, \ p_i \geq 0 \ \forall i \in [N] \right\}. \tag{35}$$

A reformulation of the (D³RO) problem is given in the following theorem.

Theorem 3.8 *Let Assumption [1] hold. If for any $\boldsymbol{x} \in X$, $h(\boldsymbol{x}, \boldsymbol{\xi}^k)$ is finite for any $k \in [N]$ and the ambiguity set (35) is nonempty, then the (D³RO) problem with the ambiguity set (35) can be reformulated as:*

$$
\min_{\boldsymbol{x},\lambda,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} f(\boldsymbol{x}) + \lambda + \sum_{i=1}^{N}\sum_{k=1}^{N}(\alpha_k + \beta_k)\hat{p}_i + \sum_{k=1}^{N}(\alpha_k - \beta_k)\eta(\boldsymbol{x})
$$

$$
\text{s.t. } \lambda + \sum_{k=i}^{N}(\alpha_k + \beta_k) + \gamma_i \geq h(\boldsymbol{x}, \xi^i) \quad \forall i \in [N],
$$

$$
\lambda \in \mathbb{R}, \ \alpha_i \geq 0, \ \beta_i \leq 0, \ \gamma_i \leq 0 \ \forall i \in [N].
$$

(36)

Proof The (D³RO) problem with the ambiguity set (35) can be written as:

$$
\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \max_{\boldsymbol{p}} \sum_{i=1}^{N} h(\boldsymbol{x}, \xi^i) p_i
$$

$$
\text{s.t. } \sup_{k \in [N]} \left| \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} \hat{p}_i \right| \leq \eta(\boldsymbol{x}),
$$

$$
\sum_{i=1}^{N} p_i = 1, \quad p_i \geq 0 \ \forall i \in [N].
$$

(37)

Note that the inner problem of (37) can be reformulated as the following linear program:

$$
\max_{\boldsymbol{p}} \sum_{i=1}^{N} h(\boldsymbol{x}, \xi^i) p_i
$$

$$
\text{s.t. } \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} \hat{p}_i \leq \eta(\boldsymbol{x}) \quad \forall k \in [N],
$$

$$
\sum_{i=1}^{k} p_i - \sum_{i=1}^{k} \hat{p}_i \geq -\eta(\boldsymbol{x}) \quad \forall k \in [N],
$$

$$
\sum_{i=1}^{N} p_i = 1, \quad p_i \geq 0 \ \forall i \in [N].
$$

(38)

After taking the dual of the above linear program and combining it with the outer problem, we obtain (36). □

A reformulation for the two-stage stochastic optimization case is given in the following corollary.

Corollary 3.5 *If $h(\cdot, \cdot)$ is a recourse function defined in* (1) *and the ambiguity set is given by* (35). *If for any $\boldsymbol{x} \in X$, $h(\boldsymbol{x}, \boldsymbol{\xi}^k)$ is finite for any $k \in [N]$ and the ambiguity set* (35) *is nonempty, then the* (D^3RO) *problem with the ambiguity set* (35) *can be reformulated as follows:*

$$\min_{\boldsymbol{x},\boldsymbol{y},\lambda,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} f(\boldsymbol{x}) + \lambda + \sum_{i=1}^{N}\sum_{k=1}^{N}(\alpha_k + \beta_k)\hat{p}_i + \sum_{k=1}^{N}(\alpha_k - \beta_k)\eta(\boldsymbol{x})$$

$$\text{s.t.} \lambda + \sum_{k=i}^{N}(\alpha_k + \beta_k) + \gamma_i \geq g(\boldsymbol{x}, \boldsymbol{y}^i, \xi^i) \quad \forall i \in [N], \tag{39}$$

$$\psi_i(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\xi}^k) \geq 0 \quad \forall i \in [m], \ \forall k \in [N],$$

$$\boldsymbol{x} \in X, \ \lambda \in \mathbb{R}, \ \alpha_i \geq 0, \ \beta_i \leq 0, \ \gamma_i \leq 0 \ \forall i \in [N].$$

The reformulation (36) of (D^3RO) with the ambiguity set defined using the K–S distance can be generalized for the case where the support Ξ is continuous. The details of this generalization are given in Appendix B.

4 Solution approaches

4.1 Models with continuous support

For models with continuous and compact support Ξ, reformulations given in Sect. 3.1 (without the measure bound constraint) Sect. 3.2, Appendix A, B, and that in Sect. 3.4 are semi-infinite programs (gen-SIP):

$$\begin{aligned} \min_{x} \ & f(x) \\ \text{s.t.} \ & g(x, t) \leq 0, \ \forall t \in T, \\ & x \in X, \end{aligned} \tag{gen-SIP}$$

by appropriately specifying $g(x, t)$. Here $X \subseteq \mathbb{R}^{k_1}$ and $T \subseteq \mathbb{R}^{k_2} \times \mathbb{Z}^{k_3}$. A cutting-surface algorithm from [10] (Algorithm 1) is now presented to solve such problems. We assume to have an oracle that can solve a deterministic non-convex optimization problem with finitely many constraints. The basic cutting-surface algorithm solves a finite constraint relaxation problem at each iteration:

$$\min_{x \in X} \{f(x) : \text{s.t.} \ g(x, t) \leq 0, \ t \in T'\}, \tag{40}$$

where T' are finite samples from T. This relaxation is called the master problem. The algorithm generates desired samples as the algorithm progresses. These samples are identified from the support by solving a separation problem. See for example, the algorithm given in [10]. Specifically, a new constraint is added at each iteration of the master problem. This constrained is identified by finding a $t \in T$ for which the

incumbent solution from the master problem is violated. Given the incumbent solution $\hat{x} \in X$, the following separation problem

$$\max_{t \in T} g(\hat{x}, t), \tag{41}$$

is solved for the purposes of generating the next constraint. The algorithm outputs an ε-optimal solution (Definition 4.1) to (gen-SIP) in a finite number of iterations (Theorem 4.1).

Definition 4.1 For a general semi-infinite program in the form of (gen-SIP), a point $x_0 \in X$ is an ε-feasible solution of (gen-SIP) if $\max_{t \in T} g(x_0, t) \leq \varepsilon$. A point $x_0 \in X$ is an ε-optimal solution of (gen-SIP) if x_0 is an ε-feasible solution of (gen-SIP) and $f(x_0) \leq \text{Val}(\text{gen-SIP})$.

Algorithm 1 A cutting-surface algorithm (modified exchange algorithm) to solve (gen-SIP).

Prerequisites: An oracle that generates the optimal solution to the master problem (40) and an oracle that generates an ε-optimal solution to the separation problem (41).
Output: An ε-optimal solution of (gen-SIP).
Step 1 Set $T_0 \leftarrow \emptyset, k \leftarrow 0$.
Step 2 Determine an optimal solution x_k of the problem $\min_{x \in X} \{f(x) : \text{ s.t. } g(x, t) \leq 0, \ t \in T_k\}$.
Step 3 Determine a $\frac{\varepsilon}{2}$-optimal solution t_{k+1} of the problem $\max_{t \in T} g(x_k, t)$. If $g(x_k, t_{k+1}) \leq \frac{\varepsilon}{2}$, stop and return x_k; otherwise let $T_{k+1} \leftarrow T_k \cup \{t_{k+1}\}, k \leftarrow k + 1$ and go to Step 2

Theorem 4.1 (Theorem 3.2 in [10]) *If $X \times T$ is compact, and $g(x, t)$ is continuous on $X \times T$, then Algorithm 1 terminates in finitely many iterations and returns an ε-optimal solution of* (gen-SIP).

We note that the oracle problem (41) in the cutting-surface algorithm is simply a function evaluation problem for the finite support case. In this case the algorithm can be adapted by sequentially adding constraints as violated inequalities are identified. This may be useful when N is large.

4.2 Models with binary decision variables and finite support

The reformulations developed in Sect. 3 have nonlinear terms that are products of dual variables and primal variables. These reformulations are in general non-convex programming problems. A solution approach is needed to handle this nonlinearity. However, when applying the reformulations to problems where first-stage decision variables $\boldsymbol{x}$ are binary and the dual variables/Lagrange multipliers are bounded, the nonlinear terms may become a product of a binary variable and a bounded continuous variable. For example, this is the case when in Sect. 3.1 the expressions $\boldsymbol{l}(\boldsymbol{x}), \boldsymbol{u}(\boldsymbol{x})$, $\underline{p}_k(\boldsymbol{x}), \underline{p}_k(\boldsymbol{x})$ are linear functions of $\boldsymbol{x}$.

Such terms can be linearized as follows. Consider a bilinear term xy where $x \in \{0, 1\}$ is a binary variable and y is a bounded continuous variable with the bounds $a \leq y \leq b$. We introduce a continuous variable z and the following constraints to represent xy:

$$ax \leq z \leq bx, \qquad y - b(1 - x) \leq z \leq y - a(1 - x). \tag{42}$$

When $x = 1$, the second constraint of (42) is equivalent to $z = y$, and the first constraint $a \leq z \leq b$ becomes redundant. When $x = 0$, the first constraint of (42) is equivalent to $z = 0$, and the second constraint $y - b \leq z \leq y - a$ becomes redundant. With the above linearization technique, problems where the nonlinear terms in the dual formulations in Sects. 3.1–3.4 involve a bilinear product of continuous and binary variables, and $h(x\xi)$ is convex in x can be reformulated as a mixed 0-1 convex program. As an example, in [50], we investigate a distributionally-robust service center location problem with decision dependent utilities. We show that this problem can be reformulated as a two-stage stochastic mixed 0-1 second-order cone program using such as technique.

5 Illustrative numerical example

We now use a numerical example to illustrate the relevance of incorporating endogenous ambiguity in a distributionally-robust optimization model. Specifically, we consider a distributionally-robust newsvendor model:

$$\max_{x \in \mathcal{X}, q \in \mathcal{Q}} \min_{P \in \mathcal{P}(x)} \mathbb{E}_P[x \cdot \min(q, D)] - cq, \tag{DNV}$$

where c is the unit cost, D is the stochastic demand, x is the selling price, and q is the number of units ordered/purchased by the vendor. The variables x and q are decision variables. The sets $\mathcal{X}$ and $\mathcal{Q}$ are the feasible sets of selling price and order quantity. The probability distribution of the demand D depends on the selling price. The set $\mathcal{P}(x)$ is a decision dependent ambiguity set of the probability distribution of D. The objective is to maximize the risk-averse profit. Product demand is a decreasing function of the selling price. For every candidate selling price x_i, there is a nominal deterministic demand value $\widehat{D}_i$ corresponding to x_i. In the numerical illustration, we consider a finite support Ξ for D which is given by $\Xi = \{d_i\}_{i=1}^8$ where $d_1 = 5$, $d_2 = 10$, $d_3 = 15$, $d_4 = 20$, $d_5 = 25$, $d_6 = 30$, $d_7 = 35$, $d_8 = 40$. The set $\mathcal{X}$ of candidate selling price is given by $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ where $x_1 = 1.2$, $x_2 = 1.4$, $x_3 = 1.6$, $x_4 = 1.8$. The set $\mathcal{Q}$ of order quantities is given by $\mathcal{Q} = \{q_i\}_{i=1}^8$ where $q_i = d_i$ for all $i \in \{1, \ldots, 8\}$. The cost is $c = 1.0$. The nominal demand values are given by $\widehat{D}_1 = d_8$, $\widehat{D}_2 = d_6$, $\widehat{D}_3 = d_4$, $\widehat{D}_4 = d_2$. Let $\mathcal{P}_i = \mathcal{P}(x_i)$ for $i \in \{1, \ldots, 4\}$. The nominal optimal solution (when no ambiguity exits) is $x^* = x_2 = 1.4, q^* = q_6 = 30$, and the nominal optimal value is $V^* = 11.30$. We consider different ways to define $\mathcal{P}_i$ and study the optimal solution of (DNV) under these definitions. In the current illustration, we enumerate all the possible combinations of $(x, q) \in \mathcal{X} \times \mathcal{Q}$ and

solve (DNV) to determine the optimal solution. A future work addresses this problem through a systematic algorithm development [51].

5.1 Definition of $\mathcal{P}_i$ using mean and variance

Consider the case of defining $\mathcal{P}_i$ using mean and variance as described in Sect. 3.2. Specifically, the $\mathcal{P}_i$ is defined as

$$\mathcal{P}_i = \left\{ \boldsymbol{p} \in \mathbb{R}_+^K \;\middle|\; \begin{array}{c} \widehat{D}_i - \delta_i \leq \sum_{k=1}^{K} p_k d_k \leq \widehat{D}_i + \delta_i \\ \sum_{k=1}^{K} p_k (d_k - \widehat{D}_i)^2 \leq \gamma_i \\ \sum_{k=1}^{K} p_k = 1 \end{array} \right\}, \tag{43}$$

where $K = 8$. Similarly, we can enumerate all the possible combinations of $(x, q) \in \mathcal{X} \times \mathcal{Q}$ and solve (DNV) in every case to determine the optimal solution. In the case that $(x, q) = (x_i, q_j)$ the objective value of (DNV) is given by

$$
\begin{aligned}
V(x_i, q_j) = \min \; & \sum_{k=1}^{K} p_k x_i \cdot \min(q_j, d_k) - c q_j \\
\text{s.t.} \; & \sum_{k=1}^{K} p_k = 1, \\
& \widehat{D}_i - \delta_i \leq \sum_{k=1}^{K} p_k d_k \leq \widehat{D}_i + \delta_i, \\
& \sum_{k=1}^{K} p_k (d_k - \widehat{D}_i)^2 \leq \gamma_i, \\
& p_k \geq 0 \quad \forall k \in [K].
\end{aligned}
\tag{44}
$$

The optimization problem in (44) is a linear program of the decision variables $\boldsymbol{p}$. In the first setting of the ambiguity, we set $\delta_i = 2$ and $\gamma_i = 5$ for $i \in \{1, 2, 3, 4\}$. The optimal solution is $x^* = x_2 = 1.4$, $q^* = q_6 = 30$, and the optimal value is $V^* = 10.60$. In the second setting of the ambiguity, we increase the ambiguity by setting $\delta_i = 3$ and $\gamma_i = 10$ for $i \in \{1, 2, 3, 4\}$. The optimal solution becomes $x^* = x_2 = 1.4$, $q^* = q_5 = 25$, and the optimal value is $V^* = 9.30$. Note that although the optimal selling price has not changed, the optimal order quantity changes with decision dependent ambiguity.

5.2 Definition of $\mathcal{P}_i$ using the Wasserstein metric

The $\mathcal{P}_i$ can be defined using the Wasserstein metic (20). Specifically, let $\mathcal{P}_i$ be defined as

$$\mathcal{P}_i = \{\boldsymbol{p} \in \mathbb{R}^K \mid \mathcal{W}(\boldsymbol{p}, \hat{\boldsymbol{p}}^i) \le r_i, \sum_{k=1}^{K} p_k = 1, \ p_k \ge 0, \ \forall k \in [K]\} \qquad \forall i \in [I], \quad (45)$$

where $I = 4$, $K = 8$, and $\hat{\boldsymbol{p}}^i$ is a nominal probability distribution on Ξ satisfying

$$\hat{p}_k^i = \begin{cases} 1 \text{ if } d_k = \widehat{D}_i, \\ 0 \text{ otherwise.} \end{cases} \quad (46)$$

In the case that $(x, q) = (x_i, q_j)$ the objective value of (DNV) is given by

$$\begin{aligned}
V(x_i, q_j) = \min \quad & \sum_{k=1}^{K} p_k x_i \cdot \min(q_j, d_k) - cq_j \\
\text{s.t.} \quad & \sum_{k=1}^{K} p_k = 1, \\
& \sum_{k=1}^{K} t_{kl} = \hat{p}_l^i \qquad \forall l \in [K], \\
& \sum_{l=1}^{K} t_{kl} = p_k \qquad \forall k \in [K], \\
& \sum_{k=1}^{K} \sum_{l=1}^{K} |d_k - d_l| t_{kl} \le r_i, \\
& p_k \ge 0, \ t_{kl} \ge 0 \qquad \forall k, l \in [K],
\end{aligned} \quad (47)$$

Notice that (47) is a linear program of the decision variables p_k and t_{kl} ($k, l \in [K]$). In the first setting of the ambiguity, we set $r_i = 0.5$ for $i \in \{1, 2, 3, 4\}$. The optimal solution is $x^* = x_2 = 1.4$, $q^* = q_6 = 30$, and the optimal value is $V^* = 11.30$. In the second setting of the ambiguity, we increase the ambiguity of the demand at the price x_2 by setting $r_2 = 0.6$ and $r_i = 0.5$ for $i \in \{1, 3, 4\}$. The optimal solution becomes $x^* = x_3 = 1.6$, $q^* = q_4 = 20$, and the optimal value is $V^* = 11.20$. Note that in this case both the optimal selling price and the order quantity changes with the ambiguity.

6 Concluding remarks

We have established a framework for reformulating the distributionally robust optimization problems with important types of decision dependent ambiguity sets. These ambiguity sets contain decision dependent parameters. For example, the moment robust ambiguity set (10) contains parameters $\alpha(\boldsymbol{x})$, $\beta(\boldsymbol{x})$, $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{Q}(\boldsymbol{x})$, which are functions of the decision $\boldsymbol{x}$. We now briefly discuss the estimation of these functions using a data-driven approach. Ambiguity sets for $\boldsymbol{\xi}$ under an arbitrary decision $\boldsymbol{x}$ can be constructed if such information is available from past decisions, or if it is possible for

us to experiment with trial decisions $\{x^i\}_{i=1}^k$ and collect samples of the random vector ξ under each decision x^i. From these samples we can establish the analytical relation between the parameters in defining the ambiguity set and the decision using statistical learning models. We can subsequently extrapolate this analytical relation to a general decision x to obtain an empirical decision dependent ambiguity set description.

The goal of this chapter was to show that it is possible to extend the dual formulations in DRO even when the ambiguity sets are decision dependent. The analysis suggests that the situations for which DRO models admit a dual reformulation also allow for dual reformulations for the decision dependent case. The reformulated models are generally non-convex optimization problems requiring further investigation towards developing efficient algorithms for the specific situations. The non-convex optimization problems may have further structure when additional assumptions on decision dependent parameters and the feasible set X are imposed. This structure may be exploited for further refined reformulations and the development of efficient algorithms.

A Reformulation of (D^3RO) with Wasserstein metric and continuous support of random parameters

We study the reformulation of (D^3RO) with Wasserstein metric and continuous support of random parameters. In contrast to the case studied in Sect. 3.3, we do not assume that Assumption 1 holds. As a consequence, the support Ξ of the decision dependent random parameters ξ can be continuous. Suppose at a decision x_0, we have observed N samples of the random variable ξ, written as $\{\xi^i\}_{i=1}^N$. We construct an empirical distribution as: $P_0 = \sum_{i=1}^N \frac{1}{N} \delta_{\xi^i}$. Setting the empirical distribution as the center of the Wasserstein ball, we can define the decision dependent ambiguity set as:

$$\mathcal{P}_C^W(x) := \{P \in \mathcal{P}(\Xi, \mathcal{F}) \mid \mathcal{W}(P, P_0) \leq r(x)\}, \tag{48}$$

where $\mathcal{W}(P, P_0)$ is defined in (19). The reformulation of (D^3RO) with the ambiguity set $\mathcal{P}_C^W(x)$ is given by the following theorem.

Theorem A.1 *The* (D³RO) *model with the ambiguity set* $\mathcal{P}_C^W(\boldsymbol{x})$ *can be reformulated as the following semi-infinite program if* $\mathcal{P}_C^W(\boldsymbol{x})$ *is non-empty for any* $\boldsymbol{x} \in X$:

$$
\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{v}} \quad & f(\boldsymbol{x}) + \frac{1}{N}\sum_{i=1}^{N} v_i + r(\boldsymbol{x}) \cdot v_{N+1} \\
\text{s.t.} \quad & h(\boldsymbol{x}, \boldsymbol{s}) - v_i - v_{N+1} \cdot \|\boldsymbol{s} - \boldsymbol{\xi}^i\| \leq 0, \qquad \forall \boldsymbol{s} \in \Xi, \forall i \in [N] \\
& \boldsymbol{x} \in \mathcal{X}, \quad v_1, \ldots, v_N \in \mathbb{R}, \quad v_{N+1} \geq 0.
\end{aligned}
\tag{49}
$$

Proof From Theorem 3.6 of [10], the inner problem of D³RO with the ambiguity set defined by the Wasserstein metric (48) is equivalent to the following conic linear program:

$$
\begin{aligned}
\max_{\mu} \quad & \int_{\boldsymbol{s} \in \Xi} h(\boldsymbol{x}, \boldsymbol{s})\mu(d\boldsymbol{s} \times \Xi) \\
\text{s.t.} \quad & \mu(\Xi \times \{\boldsymbol{\xi}^i\}) = 1/N, \qquad\qquad \forall i \in [N] \\
& \mu(\Xi \times \Xi') = 0, \\
& \sum_{i \in [N]} \int_{\boldsymbol{s} \in \Xi} \|\boldsymbol{s} - \boldsymbol{\xi}^i\|\mu(d\boldsymbol{s} \times \{\boldsymbol{\xi}^i\}) \leq r(\boldsymbol{x}), \\
& \mu \succeq 0,
\end{aligned}
\tag{50}
$$

where $\Xi' := \Xi \setminus \{\boldsymbol{\xi}^i\}^N$, and $\mu \succeq 0$ denotes that μ is a positive measure. Based on Theorem 3.7 of [10], we can apply the conic duality theory from [47] to (50), and obtain the following dual formulation of (50):

$$
\begin{aligned}
\min_{\boldsymbol{v}} \quad & \frac{1}{N}\sum_{i=1}^{N} v_i + r(\boldsymbol{x}) \cdot v_{N+1} \\
\text{s.t.} \quad & h(\boldsymbol{x}, \boldsymbol{s}) - v_i - v_{N+1} \cdot \|\boldsymbol{s} - \boldsymbol{\xi}^i\| \leq 0 \quad \forall \boldsymbol{s} \in \Xi, \forall i \in [N], \\
& v_1, \ldots, v_N \in \mathbb{R}, \quad v_{N+1} \geq 0.
\end{aligned}
\tag{51}
$$

After combining (51) with the outer minimization problem over x, we obtain the desired reformulation (49). □

B Reformulation of (D³RO) with K–S distance and continuous support of the random parameters

We now investigate the reformulation of (D³RO) with the ambiguity set defined by K–S distance where the support Ξ is not finite. We assume that Ξ is contained in a hyper-rectangle $[\boldsymbol{a}, \boldsymbol{b}] := [a_1, b_1] \times \ldots \times [a_d, b_d]$. The definition of K–S distance

(33) can be generalized for two multivariate cumulative distribution functions as follows:

$$D(P_1, P_2) = \sup_{s \in \mathbb{R}^d} |F_1(s) - F_2(s)|, \tag{52}$$

where $s = [s_1, \ldots, s_d]$ and the cumulative function F_i ($i = 1, 2$) is defined as: $F_i(s) = \int_{-\infty}^{s_1} \ldots \int_{-\infty}^{s_d} P_1(x_1, \ldots, x_d) dx_1 \ldots dx_d$. Suppose for a $x_0 \in X$, we have observed N samples of the random vector ξ, written as $\{\xi^i\}_{i=1}^N$. We define the empirical distribution as $P_0 := \sum_{i=1}^N \frac{1}{N} \delta_{\xi_i}$ and denote the cumulative distribution function of P_0 as F_0. Let $\mathcal{P}([a, b], \mathcal{B})$ denote the set of probability distributions on $[a, b]$ with the Borel sigma algebra $\mathcal{B}$. For any $P \in \mathcal{P}([a, b], \mathcal{B})$, let F^P denote the cumulative distribution function of P. The decision dependent ambiguity set based on K–S distance can be constructed as:

$$\mathcal{P}_C^{KS}(x) := \left\{ P \in \mathcal{P}([a, b], \mathcal{B}) : \sup_{s \in [a, b]} |F^P(s) - F_0(s)| \le \alpha(x) \right\}. \tag{53}$$

We now reformulate the ambiguity set (53) into finitely many expectation constraints of indicator functions by partitioning the hyper-rectangle $[a, b]$ into hyper-rectangular cells. Specifically, let ξ_k^i be the k-th component of the i-th observed sample. For each component k ($k \in [d]$), we sort the observed samples $\{\xi_k^i\}_{i=1}^N$ based on the k-th component in the ascending order, and suppose that the sorted sample components are labeled as $\{\xi_k^{[i]}\}_{i=1}^N$ such that $a_k < \xi_k^{[1]} < \xi_k^{[2]} < \cdots < \xi_k^{[N]} < b_k$. Let us divide each interval $[a_k, b_k]$ into $N + 1$ sub-intervals as: $I_0^k = [a_k, \xi_k^{[1]})$, $I_i^k = [\xi_k^{[i]}, \xi_k^{[i+1]})$ for $i \in [N - 1]$ and $I_N^k = [\xi_k^{[N]}, b_k]$, and create an N-dimensional grid based on the sub-intervals for each dimension k to partition $[a, b]$ into $(N + 1)^d$ sub-rectangular cells. Based on this partition and using the convention that $\xi_k^{[0]} = a_k$ for $k \in [d]$, the reference CDF can be written as:

$$F_0(s) = \frac{N_{j_1 j_2 \ldots j_d}}{N} \quad \text{for } s \in I_{j_1}^1 \times I_{j_2}^2 \times \ldots \times I_{j_d}^d, \quad j_r \in \{0, 1, \ldots, N\}, \quad r \in [d], \tag{54}$$

where $N_{j_1 j_2 \ldots j_d}$ is the number of observed samples within the hyper-rectangle $[a_1, \xi_1^{[j_1]}] \times [a_2, \xi_2^{[j_2]}] \times \cdots \times [a_d, \xi_d^{[j_d]}]$. For simplicity of notations, we let $I_{j_1 j_2 \ldots j_d} := I_{j_1}^1 \times I_{j_2}^2 \times \cdots \times I_{j_d}^d$, then the ambiguity set (53) can be reformulated as

$$\mathcal{P}_C^{KS}(x) = \left\{ P \in \mathcal{P}([a, b], \mathcal{B}) : \left| P(s \in I_{j_1 j_2 \ldots j_d}) - \frac{N_{j_1 \ldots j_d}}{N} \right| \right.$$
$$\left. \le \alpha(x) \text{ for } j_r \in \{0, 1, \ldots, N\}, \ r \in [d] \right\}. \tag{55}$$

Reformulation of the (D^{3}RO) with the ambiguity set (53) is given in the following theorem:

Theorem B.1 *If $h(x, s)$ is continuous in $s \in [a, b]$ for any $x \in X$ and (53) is non-empty for any $x \in X$, then the (D^{3}RO) with the ambiguity set (53) can be reformulated as the following semi-infinite program:*

$$
\min_{x, \overline{\lambda}, \underline{\lambda}} \ f(x) + \sum_{j_1=0}^{N} \cdots \sum_{j_d=0}^{N} \left(\frac{N_{j_1 \ldots j_d}}{N} + \alpha(x) \right) \overline{\lambda}_{j_1 \ldots j_d}
$$

$$
+ \sum_{j_1=0}^{N} \cdots \sum_{j_d=0}^{N} \left(\frac{N_{j_1 \ldots j_d}}{N} - \alpha(x) \right) \underline{\lambda}_{j_1 \ldots j_d} + \gamma \tag{56}
$$

s.t. $\overline{\lambda}_{j_1 \ldots j_d} + \underline{\lambda}_{j_1 \ldots j_d} + \gamma \geq h(x, s) \quad \forall s \in \mathrm{cl}(I_{j_1 \ldots j_d}), \ \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d],$

$\gamma \in \mathbb{R}, \ \overline{\lambda}_{j_1 \ldots j_d} \leq 0, \ \underline{\lambda}_{j_1 \ldots j_d}(s) \geq 0, \ \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d].$

Proof Note that the probability $P(s \in I_{j_1 j_2 \ldots j_d})$ in (55) can be written as the expectation of the indicator function $\mathbf{1}_{I_{j_1 j_2 \ldots j_d}}(s)$ with respect to P. The inner problem of (D^{3}RO) can be reformulated as the following conic linear program:

$$
\max_{P \in \mathcal{P}([a,b], \mathcal{B})} \ \mathbb{E}_P[h(x, s)]
$$

s.t. $\mathbb{E}_P[\mathbf{1}_{I_{j_1 \ldots j_d}}(s)] \leq \dfrac{N_{j_1 \ldots j_d}}{N} + \alpha(x) \quad \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d], \quad \overline{\lambda}_{j_1 \ldots j_d} \geq 0$

$\mathbb{E}_P[\mathbf{1}_{I_{j_1 \ldots j_d}}(s)] \geq \dfrac{N_{j_1 \ldots j_d}}{N} - \alpha(x) \quad \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d], \quad \underline{\lambda}_{j_1 \ldots j_d} \leq 0$

$\mathbb{E}_P[\mathbf{1}_{[a,b]}(s)] = 1, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \gamma \in \mathbb{R}.$

(57)

Applying conic duality [47] to (57) we obtain the following dual problem of (57):

$$
\min_{\overline{\lambda}, \underline{\lambda}, \gamma} \ \sum_{j_1=0}^{N} \cdots \sum_{j_d=0}^{N} \left(\frac{N_{j_1 \ldots j_d}}{N} + \alpha(x) \right) \overline{\lambda}_{j_1 \ldots j_d}
$$

$$
+ \sum_{j_1=0}^{N} \cdots \sum_{j_d=0}^{N} \left(\frac{N_{j_1 \ldots j_d}}{N} - \alpha(x) \right) \underline{\lambda}_{j_1 \ldots j_d} + \gamma \tag{58}
$$

s.t. $(\overline{\lambda}_{j_1 \ldots j_d} + \underline{\lambda}_{j_1 \ldots j_d}) \mathbf{1}_{I_{j_1 \ldots j_d}}(s) + \gamma \mathbf{1}_{[a,b]} \geq h(x, s) \quad \forall s \in [a, b],$

$\gamma \in \mathbb{R}, \ \overline{\lambda}_{j_1 \ldots j_d} \geq 0, \ \underline{\lambda}_{j_1 \ldots j_d}(s) \leq 0, \ \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d].$

Note that by partitioning the range of the vector s, the first constraint of (58) can be reformulated as the following semi-infinite constraints:

$$
\overline{\lambda}_{j_1 \ldots j_d} + \underline{\lambda}_{j_1 \ldots j_d} + \gamma \geq h(x, s) \quad \forall s \in I_{j_1 \ldots j_d}, \ \forall j_r \in \{0, 1, \ldots, N\}, \ r \in [d].
$$

(59)

Since $h(x, s)$ is continuous in $s \in [a, b]$ for any $x \in X$, we can replace $I_{j_1 \ldots j_d}$ with the closure $\mathrm{cl}(I_{j_1 \ldots j_d})$ in the above semi-infinite constraints. Then by Proposition 2.8(iii) of [47], the optimal objective of (57) equals the optimal objective of (58). After combining (58) with the outer minimization problem over $x \in X$, we can reformulate (D^3RO) into (56). □

References

1. Hu, J., Li, J., Mehrotra, S.: A data driven functionally robust approach for coordinating pricing and order quantity decisions with unknown demand function. Oper. Res. (2015). http://www.optimization-online.org/DB_HTML/2015/07/5016.html **(to appear)**
2. Birge, J., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York (1997)
3. Rahimian, H., Mehrotra, S.: Distributionally-robust optimization: a review (2019). Available on optimization online. http://www.optimization-online.org/DB_FILE/2019/08/7332.pdf
4. Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.: Models for minimax stochastic linear optimization problems with risk aversion. Math. Oper. Res. **35**(3), 580–602 (2010)
5. Mehrotra, S., Zhang, H.: Models and algorithms for distributionally robust least squares problems. Math. Program. **146**, 123–141 (2014)
6. Delage, E., Ye, Y.: Distributional robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 595–612 (2010)
7. Gao, R., Kleywegt, A.J.: Distributionally robust stochastic optimization with Wasserstein distance (2016). https://arxiv.org/pdf/1604.02199.pdf
8. Esfahan, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. Math. Program. **171**, 115–166 (2018)
9. Shafieezadeh Abadeh, S., Mohajerin Esfahani, P.M., Kuhn, D.: Distributionally robust logistic regression. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 28, pp. 1576–1584. Curran Associates, Inc. (2015). http://papers.nips.cc/paper/5745-distributionally-robust-logistic-regression.pdf
10. Luo, F.Q., Mehrotra, S.: Decomposition algorithms for distributionally robust optimization using Wasserstein metric (2017). http://www.optimization-online.org/DB_HTML/2017/04/5946.html
11. Ben-Tal, A., Hertog, D., Waegenaere, A.D., Melenberg, B., Rennen, G.: Robust solutions of optimization problems affected by uncertain probabilities. Manag. Sci. **59**, 341–357 (2013)
12. Calafiore, G.C.: Ambiguous risk measures and optimal robust portfolios. SIAM J. Optim. **18**(3), 853–877 (2007)
13. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. Math. Program. **158**, 291–327 (2016)
14. Love, D.K., Bayraksan, G.: Phi-divergence constrained ambiguous stochastic programs for data-driven optimization (2016). http://www.optimization-online.org/DB_HTML/2016/03/5350.html
15. Wang, Z., Glynn, P.W., Ye, Y.: Likelihood robust optimization for data-driven problems. Comput. Manag. Sci. **13**, 241–261 (2016)
16. Yanıkoğlu, İ., Hertog, D.: Safe approximations of ambiguous chance constraints using historical data. INFORMS J. Comput. **25**(4), 666–681 (2013)
17. Bertsimas, D., Gupta, V., Kallus, N.: Data-driven robust optimization. Math. Program. **167**, 235–292 (2018)
18. Floods, C.A., Pardalos, P.M.: Recent Advances in Global Optimization. Princeton University Press, Princeton (2014)
19. Webster, M., Santen, N., Parpas, P.: An approximate dynamic programming framework for modeling global climate policy under decision-dependent uncertainty. Comput. Manag. Sci. **9**, 339–362 (2012)
20. Goel, V., Grossmann, I.E.: A class of stochastic programs with decision dependent uncertainty. Math. Program. **108**, 355–394 (2006)
21. Poss, M.: Robust combinatorial optimization with variable budgeted uncertainty. J. Oper. Res. **11**, 75–92 (2013)
22. Nohadani, O., Sharma, K.: Optimization under decision-dependent uncertainty. SIAM J. Optim. **28**, 1773–1795 (2018)

23. Kurz, M.: The Kesten–Stigum model and the treatment of uncertainty in equilibrium theory. In: Balch, M.S., McFadden, D.L., Wu, S.Y. (eds.) Essays on Economic Behavior Under Uncertainty. North Holland, Amsterdam (1974)

24. Kurz, M.: Symposium: rational beliefs and endogenous uncertainty. Econ. Theory **8**, 383–397 (1996)

25. Kurz, M., Motolese, M.: Endogenous uncertainty and market volatility. Econ. Theory **17**, 497–544 (2001)

26. Tsur, Y., Zemel, A.: Endangered aquifers: groundwater management under threats of catastrophic events. Water Resour. Res. **40**, W06S20 (2004)

27. Shao, H., Lam, W.H.K., Tam, M.L.: A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Netw. Spat. Econ. **6**, 173–204 (2006)

28. Jonsbråten, T.W.: Optimization models for petroleum field exploitation. Ph.D. thesis, Norwegian School of Economics and Business Administration, Bergen, Norway (1998)

29. Tarhan, B., Grossmann, I.E., Goel, V.: Stochastic programming approach for planning of offshore oil or gas field infrastructure under decision-dependent uncertainty. Ind. Eng. Chem. Res. **48**, 3078–3097 (2009)

30. Goel, V., Grossmann, I.E.: A stochastic programming approach to planning of offshore gas filed developments under uncertainty in reserves. Comput. Chem. Eng. **108**, 1409–1429 (2004)

31. Ahmed, S.: Strategic planning under uncertainty: stochastic integer programming approaches. Ph.D. thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA (2000)

32. Viswanath, K., Peeta, S., Salman, F.S.: Investing in the links of a stochastic network to minimize expected shortest path length. Technical report, Purdue University (2004)

33. Jonsbråten, T.W., Wets, R.J.B., Woodruff, D.L.: A class of stochastic programs with decision dependent random elements. Ann. Oper. Res. **82**, 83–106 (1998)

34. Luce, R.D.: The choice axiom after twenty years. J. Math. Psychol. **15**(3), 215–233 (1977)

35. Hu, J., Mehrotra, S.: Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization. IIE Trans. **47**, 358–372 (2015)

36. Held, H., Woodruff, D.L.: Heuristics for multi-stage interdiction of stochastic networks. J. Heuristics **11**, 483–500 (2005)

37. Dupačová, J.: Optimization under exogenous and endogenous uncertainty. In: Mathematical Methods in Economics (2006)

38. Lee, S., de Mello, T.H., Kleywegt, A.J.: Newsvendor-type models with decision-dependent uncertainty. Math. Methods Oper. Res. **76**, 189–221 (2012)

39. Goel, V., Grossmann, I.E., El-Bakry, A.S., Mulkay, E.L.: A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. Comput. Chem. Eng. **30**, 1076–1092 (2006)

40. Goel, V., Grossmann, I.E.: A Lagrangian duality based branch and bound for solving linear stochastic programs with decision dependent uncertainty. In: Puigjaner, L., Espuña , A. (eds.) European Symposium on Computer Aided Process Engineering, vol. 15. Elsevier Science B. V. (2005)

41. Gupta, V., Grossmann, I.E.: Solution strategies for multistage stochastic programming with endogenous uncertainty. Comput. Chem. Eng. **35**, 2235–2247 (2011)

42. Tarhan, B., Grossmann, I.E.: A multistage stochastic programming approach with strategies for uncertainty reduction in the synthesis of process networks with uncertain yields. Comput. Chem. Eng. **32**, 766–788 (2008)

43. Tarhan, B., Grossmann, I.E., Goel, V.: Computational strategies for non-convex multistage minlp models with decision-dependent uncertainty and gradual uncertainty resolution. Ann. Oper. Res. **203**, 141–166 (2013)

44. Spacey, S.A., Wiesemann, W., Kuhn, D., Luk, W.: Robust software partitioning with multiple instantiation. INFORMS J. Comput. **24**(3), 500–515 (2012)

45. Vujanic, R., Goulart, P., Morari, M.: Robust optimization of schedules affected by uncertain events. J. Optim. Theory Appl. **171**, 1033–1054 (2016)

46. Mehrotra, S., Papp, D.: A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. SIAM J. Optim. **24**(4), 1670–1697 (2015)

47. Shapiro, A.: On duality theory of conic linear problems. In: Goberna, M.Á., López, M.A. (eds.) Semi-InfiniteProgramming. Nonconvex Optimization and Its Applications, vol. 57, pp. 135–165. Springer, Boston (2001)

48. Givens, C.R., Shortt, R.M.: A class of wasserstein metrics for probability distributions. Mich. Math. J. **31**(2), 231–240 (1984)

49. Jiang, R., Guan, Y.: Risk-averse two-stage stochastic program with distributional ambiguity. Oper. Res. **66**, 1390–1405 (2018)
50. Luo, F.Q., Mehrotra, S.: Distributionally robust service center location problem with decision dependent utilities. Technical report, Northwestern University (2019)
51. Rahimian, H., Mehrotra, S.: Multi-product newsvendor problem with decision dependent demand. Working paper, Northwestern University (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.