A Distributionally-Robust Service Center Location Problem with Decision Dependent Demand Induced from a Maximum Attraction Principle

Fengqiao Luo

Department of Industrial Engineering and Management Science, Northwestern University, Evanston, Illinois 60208,
fengqiaoluo2014@u.northwestern.edu

This paper establishes and analyzes a service center location model with a simple but novel decision-dependent demand induced from a maximum attraction principle. The model formulations are investigated in the distributionally-robust optimization framework for the capacitated and uncapacitated cases. A statistical model that is based on the maximum attraction principle for estimating customer demand and utility gain from service is established and analyzed. The numerical experiments show that the model admits high computational efficiency in solving mid- and large-size instances.

Key words: service center location, distributionally-robust optimization, decision-dependent demand, maximum attraction principle

1. Introduction

The facility location problem is one of the most fundamental problems investigated in operations research. In this problem, a decision maker needs to decide locations of a limited number of facilities (factories, retail centers, power plants, service centers, etc.), and determine coverage of demand from different sites by the located facilities. The objective is to minimize the facility setup cost and the cost of production and delivery. This problem provides a basic framework to formulate related problems in resource allocation, supply chain management and logistics, etc. The facility location models in the stochastic and robust (distributionally-robust) optimization framework have also received sufficient investigation. In the stochastic optimization framework, the customer demand can be random parameters following a known or partially known probability distributions, while in the robust (distributionally-robust) optimization framework customer demand (the probability distribution of customer demand) can have a certain level of uncertainty (ambiguity). Incorporating randomness or uncertainty in customer demand estimation is a sensible and realistic model improvement, as the decision maker has no perfect prediction on the demand in practice. Recently, there is a trend of research on distributionally-robust optimization problems with decision-dependent ambiguity, in which the ambiguity set is specified by decision dependent parameters. Introducing decision dependency has certain merit in situations when model parameters are naturally functions

of decision variables. For example, in a basic pricing and revenue management problem the demand of a product can be a decreasing function of the price to be determined. The notion of modeling decision-dependency in the robust (distributionally-robust) optimization framework can be applied to investigate a variety of specific problems.

In this paper, we investigate a distributionally-robust service center location problem with decision-dependent customer demand. In the model, customers need to physically access a service center in order to receive service. This problem is motivated from clinic or medical test center (such as for COVID-19 screen test) allocation. The objective is to maximize the total utility gain of all customers who have decided to receive service from the opened service centers. The decision dependence is decoupled from the ambiguity set in our model, which is different from (Basciftci et al. 2020). Specifically, we impose the decision dependency on the demand itself but not on the uncertainty of demand, and the decision-dependent demand is induced from a maximum attraction principle based on a ranking of opened service centers in the neighborhood. Our investigation shows that the decoupling approach leads to high computational efficiency of solving the problem and it can be more data-driven in practice. Furthermore, the novel approach of modeling decision dependency introduce additional combinatorial properties to the model which is of independent interest.

The contribution of this paper is summarized as follows:

1. A novel approach of modeling decision dependency of the customer demand has been established for the service center location problem, with a possible extension on the modeling of decision dependency. In this approach, the decision dependency is decoupled from the ambiguity set which admits highly computational tractable reformulations.

2. A learning model has been established to estimate utilities and demand based on data from survey.

3. The numerical experience with this model shows that mid- and large-size instances can be solved very efficiently due to decoupling of decision dependency from the ambiguity set.

Literature review

Facility location problems with non-deterministic demand have received plentiful investigation in the framework of stochastic optimization and robust (distributionally-robust) optimization. In the stochastic optimization framework, customer demand are independent random parameters with known probability distributions and the problem can be formulated as a two-stage stochastic program in which the location vector is in the first-stage decision that needs to be made before realization of customer demand (Louveaux and Peeters 1992, Albareda-Sambola et al. 2011). In a special case when the customer demand rates and service rates of each facility are assumed to

follow exponential probability distributions, the problem of minimizing long term average cost can be reformulated as a deterministic optimization problem with corresponding Poisson rates to characterize the demand and service. This setting has been applied to an allocation problem of ATMs (Wang et al. 2002). The probability distribution of demand can also be used to define chance constraints to ensure a required service level under possible stockout and supply disruption (Murali et al. 2012, Gülpınar et al. 2013, Lim et al. 2013, Li et al. 2017).

In the robust (distributionally-robust) optimization framework for facility location problems, the information of demand is partially known to the decision maker, and the goal is to find a robust optimal location vector that optimizes the objective after a worst-case realization of demand information (Snyder 2006). Following this direction of research, Baron and Milner 2010 investigated a robust multi-period facility location problem with box and ellipsoid sets of uncertainty. Gourtani et al. 2020 investigated a distributionally-robust two-stage facility location problem with an ambiguity set defined corresponding to the mean and covariance matrix of a random parameter for expressing the demand. Facility location problems with demand uncertainty have been studied with a variety of novel application background, which includes but not limited to medication coverage and delivery under a large-scale bio-terror attack (Murali et al. 2012), medical equipment (defibrillators) location problem to reduce cardiopulmonary resuscitation (CPR) risk (Chan et al. 2017), humanitarian relief logistics (Döyen et al. 2012), and hazardous waste transportation (Berglund and Kwon 2014), etc.

In a lot of real world problems, the uncertainty of parameters can interplay with the decision to be made. This behavior is well observed especially in a sequential (multi-stage) decision-making process, in which information about system parameters are gradually revealed and the decisions made up until the current stage can reshape the uncertainty in future (Goel and Grossmann 2006). It motivates the research on multi-stage stochastic optimization with decision-dependent uncertainty. Solution strategies based on Lagrangian duality and novel branch-and-bound methods are developed to solve this family of problems (Goel and Grossmann 2005, Gupta and Grossmann 2011, Tarhan et al. 2013), and it has a rich application in oil-chemical industrial (Goel and Grossmann 2004, Tarhan and Grossmann 2008, Goel et al. 2006, Tarhan et al. 2009). An approximation scheme (Vayanos et al. 2011) is proposed to tackle the high complexity in solving the multi-stage problems with decision-dependent information discovery. Decision-dependent uncertainty has also been considered in stochastic optimization problems such as resource management (Tsur and Zemel 2004), stochastic traffic assignment modeling (Shao et al. 2006), and robust network design (Ahmed 2000, Viswanath et al. 2004).

Imposing decision-dependent uncertainty for robust (distributionally-robust) optimization has received great attention in recently years (Nohadani and Sharma 2018, Luo and Mehrotra 2020,

Noyan et al. 2017, Basciftci et al. 2020, Luo and Mehrotra 2019a). Nohadani and Sharma 2018 studied robust linear programs with decision dependent budget-type uncertainty and its generalization with a polyhedral uncertainty set. This concept is demonstrated in a robust shortest-path problem, where the uncertainty is resolved progressively when approaching the destination. Noyan et al. 2017 investigated a family of distributionally-robust optimization problems with an ambiguity set defined using earth mover's distances (including total variation distance and the Wasserstein metric) with decision-dependent parameters such as the nominal probability distribution and the radius, and focused on understanding which settings can lead to tractable formulation. Royset and Wets 2017 provided a variational principle analysis for optimization under stochastic ambiguity, which gives fruitful tools for analyzing solution quality and price of robustness. Some novel applications are studied in (Spacey et al. 2012) for robust software partition, and in (Nohadani and Roy 2017) for radiation therapy design.

The work in (Basciftci et al. 2020) and (Luo and Mehrotra 2019b) are mostly related to the work of this paper. Basciftci et al. 2020 investigated a distributionally-robust facility location problem with decision-dependent ambiguity, where the ambiguity set is defined using disjointed lower and upper bounds on the mean and variance of the candidate probability distributions for customer demand, and these bounds are defined as linear functions of the location vector to admit a tractable MILP reformulation. The model considered in (Luo and Mehrotra 2019b) assumed deterministic customer demand and imposed a linear decision dependency on the moment bounds for candidate probability distributions of utility, which leads to a MISOCP reformulation.

2.   Problem Formulation

We consider a problem of allocating a set of service centers (facilities) in a region to meet the customer demand from the region. Similarly to the traditional facility location problem, the region consists of customer sites and candidate locations for the service centers. Each customer site has certain demand that needs to be fulfilled by a service center from its neighborhood. Especially, we consider the case that the demand of each customer site is a random parameter which depends on the locations of service centers. In most literature on decision-dependent robust (distributionally-robust) optimization, the decision-dependency is formulated as a linear function. In this problem, we consider a special form of decision dependency that is based on an assumption of maximum attraction principle, which will be clear in Section 2.1.

2.1.   A deterministic model

We first formulate a deterministic model of this problem, in which we treat the demands as deterministic parameters that depend on the locations of service centers and introduce the notion of maximum attraction principle. The notations used in this model is given by the following list:

$\mathcal{S}$	the set of customer sites;
$\mathcal{F}$	the set of candidate locations for service centers;
b_j	the cost of opening a service center at location $j \in \mathcal{F}$;
B	the budget of allocating service centers;
C_j	the service capacity of a service center located at $j \in \mathcal{F}$;
u_{ij}	the utility gain obtained by a customer from site $i \in \mathcal{S}$ who gets service from the service center at $j \in \mathcal{F}$;
y_j	the binary decision variable of opening a service center at location $j \in \mathcal{F}$;
$D_i(y)$	the demand from customer site i that depends on the decision vector y;
x_{ij}	the demand flow from customer site $i \in \mathcal{S}$ to the service center at $j \in \mathcal{F}$.

The deterministic model is formulated as follows:

$$
\begin{aligned}
\max \ & \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}} u_{ij} x_{ij} \\
\text{s.t.} \ & \sum_{j \in \mathcal{F}} b_j y_j \leq B, \\
& \sum_{i \in \mathcal{S}} x_{ij} \leq C_j y_j \qquad \forall j \in \mathcal{F}, \\
& \sum_{j \in \mathcal{F}} x_{ij} \leq D_i(y) \qquad \forall i \in \mathcal{S}, \\
& y_j \in \{0,1\}, \ x_{ij} \geq 0 \ \forall i \in \mathcal{S}, j \in \mathcal{F}.
\end{aligned}
\tag{1}
$$

The first three constraints represent the limit of budget, capacity and the amount of demand, respectively. Note that in general the demand from site $i \in \mathcal{S}$ can depend on the pattern of service center locations which is represented by the decision vector y. In this paper, we investigate a specific type of location dependent demand which is simple but sensible in depicting customers' behavior in practice. This type of location dependency of demand is referred as the *maximum attraction principle* in this paper. We now give a detailed description of this principle. First, each customer site i is associated with a preferable subset $\mathcal{F}_i$ of candidate facility locations. When there is only one service center opened at $j \in \mathcal{F}_i$, it will attract D_{ij} amount of demand from i. If there are multiple service centers opened at locations in $\mathcal{F}_i$, the maximum demand that can be attracted from i is equal to the maximum D_{ij} for $j \in \mathcal{F}_i$ that has a facility. This principle is formally described by the following equation:

$$
D_i(y) = \begin{cases} 0 & \text{if } \sum_{j \in \mathcal{F}} \mathbb{1}\{j \in \mathcal{F}_i\} y_j = 0 \\ D_{ij} & \text{if } y = e_j \text{ for some } j \in \mathcal{F}_i \\ \max_{j \in \mathcal{F}_i(y)} D_{ij} & \text{for other cases,} \end{cases}
\tag{2}
$$

where e_j is the $|\mathcal{F}_i|$-dimensional vector with the j-th entry being 1 and other entries being 0, and the set $\mathcal{F}_i(y)$ is defined as $\mathcal{F}_i(y) := \{j \in \mathcal{F}_i \,|\, y_j = 1\}$. The maximum attraction principle matches with our intuition from practice. This principle first assumes that locations of service centers that are not within the preferable location set $\mathcal{F}_i$ are not attractive to customers from site i at all.

The most natural way of establishing $\mathcal{F}_i$ is based on the distance from site i to the candidate locations. This is a reasonable assumption because customers usually will not consider visiting a service center that is beyond a certain distance from their living place. The maximum attraction principle further assumes that within the preferable opened service centers $\mathcal{F}_i(y)$, there exists one service center j^* (or multiple centers) that is (are) most attractive to the customers from site i in the sense that $j^* \in \mathrm{argmax}_{j \in \mathcal{F}_i(y)} D_{ij}$, and the presence of multiple service centers in $\mathcal{F}_i(y)$ including j^* attracts the same amount of demand from site i as the presence of just a single service center at j^*. This assumption is also consistent with our intuition from practice. If every service center is identical in the sense of scale and service quality, the most attractive one to site i is likely the one that is most close to i. Furthermore, customers who are willing to visit further service centers are also willing to visit the one that is most close to their living place. Note that the maximum attraction principle can be extended to incorporate the impact of the number of opened service centers in $\mathcal{F}_i$ on the demand. This extension is discussed in Section 2.2. We also define the subset $\mathcal{S}_j$ as $\mathcal{S}_j := \{i \in \mathcal{S} \mid j \in \mathcal{F}_i\}$ for all $j \in \mathcal{F}$. Based on the maximum attraction principle, the service center location problem (1) is written as the follows:

$$
\begin{aligned}
\max \quad & \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}} u_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j \in \mathcal{F}} b_j y_j \le B, \\
& \sum_{i \in \mathcal{S}} x_{ij} \le C_j y_j \qquad \forall j \in \mathcal{F}, \\
& \sum_{j \in \mathcal{F}_i} x_{ij} \le \max_{j \in \mathcal{F}_i(y)} D_{ij} \qquad \forall i \in \mathcal{S}, \\
& y_j \in \{0,1\}, \ x_{ij} \ge 0 \ \forall i \in \mathcal{S}, j \in \mathcal{F}.
\end{aligned} \tag{3}
$$

To go one step further, we can linearize the term $\max_{j \in \mathcal{F}_i(y)} D_{ij}$ by introducing some continuous auxiliary variables q_{ij} to form it as a convex combination of D_{ij} for $j \in \mathcal{F}_i$. After this transformation, we obtain the following equivalent formulation of (3):

$$
\max \quad \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}} u_{ij} x_{ij} \tag{DDSL.1}
$$

$$
\text{s.t.} \quad \sum_{j \in \mathcal{F}} b_j y_j \le B, \tag{DDSL.2}
$$

$$
\sum_{i \in \mathcal{S}} x_{ij} \le C_j y_j \qquad \forall j \in \mathcal{F}, \tag{DDSL.3}
$$

$$
\sum_{j \in \mathcal{F}_i} x_{ij} \le \sum_{j \in \mathcal{F}_i} D_{ij} q_{ij} \qquad \forall i \in \mathcal{S}, \tag{DDSL.4}
$$

$$
\sum_{j \in \mathcal{F}_i} q_{ij} \le 1 \qquad \forall i \in \mathcal{S}, \tag{DDSL.5}
$$

$$
q_{ij} \le y_j \qquad \forall i \in \mathcal{S}, \forall j \in \mathcal{F}_i, \tag{DDSL.6}
$$

$$
y_j \in \{0,1\}, \ x_{ij} \ge 0, \ q_{ij} \ge 0 \ \forall i \in \mathcal{S}, j \in \mathcal{F}_i. \tag{DDSL.7}
$$

Notice that in the above reformulation, the auxiliary variables $\{q_{ij} \mid j \in \mathcal{F}_i\}$ are used select the most attractive candidate location driven by the sense of maximizing the objective. The constraint $q_{ij} \leq y_j$ ensures that only opened service centers in $\mathcal{F}_i$ are involved in the maximum attraction principle. Since $\sum_{j \in \mathcal{F}_i} q_{ij} \leq 1$, the model will set $q_{ij^*} = 1$ and $q_{ij'} = 0$ for all $j' \in \mathcal{F}_i \setminus \{j^*\}$ to relax the constraint $\sum_{j \in \mathcal{F}_i} x_{ij} \leq \sum_{j \in \mathcal{F}_i} D_{ij} q_{ij}$ as possible.

2.2. Some remarks on the model

In the model setting, we implicitly assume that the customers are willing to corporate with the decision maker to maximize the total utility gain. Although this assumption is highly impractical, the rationality of this model depends on what metric we use to measure the system performance. If the goal is to estimate what is expectation of total utility in practice, then the following questions should be addressed and the corresponding aspects should be properly modeled: What is the service policy used by each service center (FIFO or some other policies)? How to characterize customers' behavior and the mechanism of competing for the limited service capacity? If their most favored service center does not have any capacity, are they willing to accept the service from the less preferred locations with less utility gain? Incorporating all these factors into the model can easily make it very complicated, and meanwhile a large amount of customers' information are needed to drive this approach of modeling. On the other side, if the goal is to access what is the maximum potential utility that can be achieved by the system in the most ideal situation, then the (DDSL) model can be used to give an estimation. A possible modification of modeling the decision-dependent demand is to add a perturbation term to the demand based on the number of opened service centers. Specifically, we can modify the constraint $\sum_{i \in \mathcal{F}_i} x_{ij} \leq \max_{j \in \mathcal{F}_i} D_{ij}$ to be $\sum_{i \in \mathcal{F}_i} x_{ij} \leq \max_{j \in \mathcal{F}_i} D_{ij} + a_i(\sum_{j \in \mathcal{F}_i} y_j)$, where a_i is a parameter that measures the influence of opening one more service center on the demand. In this way of modeling, it is assumed that the marginal increase of demand may depend on the number of opened service centers. In reality, when potential customers see more chain stores are opened in the neighborhood, they may have higher intention to try one of them. In this case, the parameter a_i can be a random parameter with certain level of ambiguity in the distributionally-robust extension of the model.

2.3. A distributionally-robust two-stage stochastic extension

The deterministic service center location model with decision dependent demand can be further extended to a distribtuionally-robust two-stage stochastic program after imposing an ambiguity set on the pairwise demand D_{ij}. This extension is motivated by the fact that estimation of the demand parameters D_{ij} could be inaccurate. Therefore, we can treat D_{ij} as random parameters with an unknown joint probability distribution, and apply the distributionally-robust framework on this service center location problem with uncertainty. We assume a finite support of the joint

probability distribution of the pairwise random demand vector $D := \{D_{ij} \mid i \in \mathcal{S}, j \in \mathcal{F}_i\}$. Let the finite support be based on $|\Omega|$ samples written as $D^\omega = \{D_{ij}^\omega \mid i \in \mathcal{S}, j \in \mathcal{F}_i\}$ for all $\omega \in \Omega$. In this case, any probability distribution of D can be represented as a $|\Omega|$-dimensional vector. We define a nominal probability distribution μ_0 of D as

$$\mu_0(D = D^\omega) = \mu_0^\omega \qquad \forall \omega \in \Omega. \tag{5}$$

In the vector representation, we write it as $\mu_0 = [\mu_0^\omega : \omega \in \Omega]$. The ambiguity set of candidate joint probability distribution of D is defined based on the total variation distance between two probability distributions. Specifically we consider an ambiguity set of the following form:

$$\mathcal{P} := \{\mu \in \mathbb{R}^{|\Omega|} \mid \|\mu - \mu_0\|_1 \le d\}. \tag{6}$$

Based on the above definition of ambiguity set, the distributionally-robust two-stage stochastic extension of the service center location model can be formulated as follows:

$$\max_y \min_{\mu \in \mathcal{P}} \mathbb{E}_{D \sim \mu}[Q(y, D)]$$
$$\text{s.t.} \sum_{j \in \mathcal{F}} b_j y_j \le B, \tag{DRO-FL}$$
$$y_j \in \{0, 1\} \, \forall j \in \mathcal{F},$$

where the recourse function $Q(y, D^\omega)$ for scenario ω is given by

$$Q(y, D^\omega) = \max \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}_i} u_{ij} x_{ij}^\omega \tag{7a}$$

$$\text{s.t.} \sum_{i \in \mathcal{S}} x_{ij}^\omega \le C_j y_j \qquad \forall j \in \mathcal{F}, \tag{7b}$$

$$\sum_{j \in \mathcal{F}_i} x_{ij}^\omega \le \sum_{j \in \mathcal{F}_i} D_{ij}^\omega q_{ij}^\omega \qquad \forall i \in \mathcal{S}, \tag{7c}$$

$$\sum_{j \in \mathcal{F}_i} q_{ij}^\omega \le 1 \qquad \forall i \in \mathcal{S}, \tag{7d}$$

$$q_{ij}^\omega \le y_j \qquad \forall i \in \mathcal{S}, \forall j \in \mathcal{F}_i, \tag{7e}$$

$$x_{ij}^\omega \ge 0, \, q_{ij}^\omega \ge 0 \, \forall i \in \mathcal{S}, \, \forall j \in \mathcal{F}_i. \tag{7f}$$

Using a standard technique for two-stage stochastic programming, we can decompose (DRO-FL) into a master problem and scenario sub-problems. The original problem (DRO-FL) can be solved iteratively. In each iteration, we solve the current master problem in the space of y and pass the current master solution to each scenario sub-problem. After solving each scenario sub-problem for the fixed first-stage solution, we can generate a valid inequality for each scenario using the optimal dual values associated with constraints of the scenario sub-problem. Then we aggregate the valid inequalities from all scenario sub-problems using the worst-case measure of scenarios to get a single

cut for the master problem which is an optimality cut. The optimality cut is added to the master problem in the next iteration. Specifically, the master problem at iteration n can be represented as

$$\max \eta$$

$$\text{s.t. } \sum_{j \in \mathcal{F}} b_j y_j \leq B,$$

$$\eta \leq \sum_{\omega \in \Omega} \mu_\omega^{(k)} \Big(r_\omega^{(k)} + \sum_{j \in \mathcal{F}} t_{\omega,j}^{(k)} y_j \Big) \qquad \forall k \in [n-1], \qquad \text{(Master)}$$

$$y \in \{0,1\}^{|\mathcal{F}|}, \ \eta \geq 0,$$

where $\mu^{(k)}$ is the iteration based worst-case probability measure on scenarios at iteration $k \in [n-1]$. The way of determining this worst-case probability measure is given in (10). At iteration n we solve the master problem (Master) and obtain the current optimal first-stage solution $y^{(n)}$. This $y^{(n)}$ is input into every second-stage scenario sub-problem. The second-stage linear program is solved and let $\alpha_{\omega,j}^{(n)} \geq 0$, $\beta_{\omega,i}^{(n)} \geq 0$, $\gamma_{\omega,i}^{(n)} \geq 0$, and $\tau_{\omega,ij}^{(n)} \geq 0$ for all $i \in \mathcal{S}$, $j \in \mathcal{F}_i$ are optimal dual values corresponding to the constraints (7b), (7c), (7d) and (7e), respectively. Using the standard technique of Bender's decomposition [cite ref], we can obtain the following valid inequality on the value function $Q(y, D^\omega)$:

$$Q(y, D^\omega) \leq \sum_{i \in \mathcal{S}} \gamma_{\omega,i}^{(n)} + \sum_{j \in \mathcal{F}} \Big(C_j \alpha_{\omega,j}^{(n)} + \sum_{i \in \mathcal{S}_j} \tau_{\omega,ij}^{(n)} \Big) y_j. \qquad (8)$$

The strong duality implies that when evaluating $Q(y, D^\omega)$ at $y^{(n)}$, we get

$$Q(y^{(n)}, D^\omega) = \sum_{i \in \mathcal{S}} \gamma_{\omega,i}^{(n)} + \sum_{j \in \mathcal{F}} \Big(C_j \alpha_{\omega,j}^{(n)} + \sum_{i \in \mathcal{S}_j} \tau_{\omega,ij}^{(n)} \Big) y_j^{(n)}. \qquad (9)$$

The worst-case probability measure $\mu^{(n)}$ is an optimal solution of the following linear program:

$$\min_\mu \ \sum_{\omega \in \Omega} \mu_\omega Q(y^{(n)}, D^\omega)$$

$$\text{s.t } \ \|\mu - \mu_0\|_1 \leq d, \qquad (10)$$

$$\sum_{\omega \in \Omega} \mu_\omega = 1, \ \mu_\omega \geq 0 \ \forall \omega \in \Omega,$$

where $Q(y^{(n)}, D^\omega)$ is the value function evaluated at the first stage solution $y^{(n)}$ and scenario ω. Once the current worst-case probability measure is obtained, the following inequality will be added to the first-stage master problem:

$$\eta \leq \sum_{\omega \in \Omega} \mu_\omega^{(n)} \Big[\sum_{i \in \mathcal{S}} \gamma_{\omega,i}^{(n)} + \sum_{j \in \mathcal{F}} \Big(C_j \alpha_{\omega,j}^{(n)} + \sum_{i \in \mathcal{S}_j} \tau_{\omega,ij}^{(n)} \Big) y_j \Big]. \qquad (11)$$

The algorithm and convergence property for solving (DRO-FL) are given in Appendix 5. It can be shown that for a given first-stage solution y, the scenario sub-problem can be solved using a greedy algorithm.

2.4. A single-stage reformulation of (DRO-FL) for the uncapacitated case

We consider a special case of (DRO-FL) in which each candidate service center has sufficient capacity for service. In this case, the capacity constraints (7b) can be removed from the second-stage scenario problems. We will show that this leads to a simplified scenario problem that admits a closed form optimal solution and optimal objective for the second-stage problem with a mild regularity condition on the parameters. Then after dualizing the inner minimization problem over the probability measure on scenarios, (DRO-FL) can be reformulated as a mixed 0-1 linear program. The single-stage reformulation result is given by the following theorem.

DEFINITION 1. The utility gain and demand are *consistent* if for every $\omega \in \Omega$, $i \in \mathcal{S}$ and $\mathcal{F}' \subseteq \mathcal{F}_i$ there exists a $j \in \mathcal{F}'$ such that $u_{ij} = \max_{j' \in \mathcal{F}'} u_{ij'}$ and $D_{ij}^{\omega} = \max_{j' \in \mathcal{F}'} D_{ij'}^{\omega}$.

The consistency condition for utility gain and demand says that the location that attracts the most demand over other locations should also correspond to the highest utility.

THEOREM 1. *Suppose the utility gain and demand are consistent. In the uncapacitated case, the distributionally-robust service center location problem* (DRO-FL) *with decision dependent demand based on the maximum attraction principle and the ambiguity set* (6) *can be reformulated as the following mixed 0-1 linear program:*

$$
\begin{aligned}
\max \quad & \sum_{\omega \in \Omega} \mu_0^{\omega}(\alpha^{\omega} - \beta^{\omega}) - \lambda d - \gamma \\
\text{s.t.} \quad & \sum_{j \in \mathcal{F}} b_j y_j \le B, \\
& \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} s_{ij} + \alpha^{\omega} - \beta^{\omega} + \gamma \ge 0 \qquad \forall \omega \in \Omega, \\
& s_{ij} \le y_j \qquad \forall i \in \mathcal{S}, \forall j \in \mathcal{F}_i, \\
& \sum_{j \in \mathcal{F}_i} s_{ij} \le 1 \qquad \forall i \in \mathcal{S}, \\
& \lambda \ge 0, \ \alpha^{\omega} \ge 0, \ \beta^{\omega} \ge 0, \ \gamma \in \mathbb{R}, \ \forall \omega \in \Omega \\
& y_j \in \{0,1\} \ \forall j \in \mathcal{F}, \ 0 \le s_{ij} \le 1 \ \forall i \in \mathcal{S}, \forall j \in \mathcal{F}_i.
\end{aligned}
\tag{12}
$$

Proof. Without capacity constraints, it is easy to see that the optimal objective value of the scenario problem is given by

$$
\sum_{i \in \mathcal{S}} \left(\max_{j \in \mathcal{F}_i} u_{ij} y_j \right) \left(\max_{k \in \mathcal{F}_i} D_{ik}^{\omega} y_k \right) = \sum_{i \in \mathcal{S}} \max_{j,k \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j y_k = \sum_{i \in \mathcal{S}} \max_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j,
\tag{13}
$$

where we use the assumption that the utility gain and demand are consistent. The problem (DRO-FL) becomes the following:

$$
\begin{aligned}
\max_{y} \min_{\mu} \quad & \sum_{\omega \in \Omega} \mu^{\omega} \sum_{i \in \mathcal{S}} \max_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j \\
\text{s.t.} \quad & \sum_{j \in \mathcal{F}} b_j y_j \leq B, \\
& \sum_{\omega \in \Omega} |\mu^{\omega} - \mu_0^{\omega}| \leq d, \\
& y_j \in \{0,1\} \, \forall j \in \mathcal{F}.
\end{aligned}
\tag{14}
$$

The terms involved in the inner minimization problem of (14) can be written as

$$
\begin{aligned}
\min_{\mu} \quad & \sum_{\omega \in \Omega} \mu^{\omega} \left(\sum_{i \in \mathcal{S}} \max_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j \right) \\
\text{s.t.} \quad & \sum_{\omega \in \Omega} \rho^{\omega} \leq d, \\
& \mu^{\omega} - \mu_0^{\omega} \leq \rho^{\omega} \qquad \forall \omega \in \Omega, \\
& \mu_0^{\omega} - \mu^{\omega} \leq \rho^{\omega} \qquad \forall \omega \in \Omega, \\
& \sum_{\omega \in \Omega} \mu^{\omega} = 1, \\
& \mu^{\omega} \geq 0, \, \rho^{\omega} \geq 0 \, \forall \omega \in \Omega.
\end{aligned}
\tag{15}
$$

Taking the dual of the above linear program with respect to the probability measure μ, we obtain the following inner problem:

$$
\begin{aligned}
\max \quad & \sum_{\omega \in \Omega} \mu_0^{\omega}(\alpha^{\omega} - \beta^{\omega}) - \lambda d - \gamma \\
\text{s.t.} \quad & \sum_{i \in \mathcal{S}} \max_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j + \alpha^{\omega} - \beta^{\omega} + \gamma \geq 0 \qquad \forall \omega \in \Omega, \\
& \lambda - \alpha^{\omega} - \beta^{\omega} \geq 0 \qquad \forall \omega \in \Omega, \\
& \lambda \geq 0, \, \alpha^{\omega} \geq 0, \, \beta^{\omega} \geq 0, \, \gamma \in \mathbb{R}.
\end{aligned}
$$

To linearize the term $\max_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j$, we can introduce binary indicator variables s_{ij}, and reformulate the first constraint as follows

$$
\begin{aligned}
& \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} s_{ij} + \alpha^{\omega} - \beta^{\omega} + \gamma \geq 0, \\
& s_{ij} \leq y_j, \, \sum_{j \in \mathcal{F}_i} s_{ij} = 1, \, 0 \leq s_{ij} \leq 1,
\end{aligned}
$$

where we implicitly use the consistency condition of utility and demand which implies $\operatorname{argmax}_{j \in \mathcal{F}_i} u_{ij} D_{ij}^{\omega} y_j$ is scenario independent and hence the variable s_{ij}. Incorporate with the outer maximization yields the reformulation (12). □

3. A Statistical Model for Utilities and Demand Estimation

We establish a regression model based on the maximum attraction principle for estimating utility value u_{ij} and demand D_{ij} that are input parameters to the service center location model (DRO-FL). As a by-product, this regression model can also be used to estimate the size parameter d of the ambiguity set (6). The regression model requires samples of survey among potential customers from a site $i \in \mathcal{S}$ on their ratings of multiple candidate service center locations. Before formulating the regression model, we first introduce how the samples of survey are collected for fitting the regression model. Suppose N residents have been randomly selected from the site i, and they are viewed as potential customers of service centers under planning. Each of them is asked to give a score in the range $\{0, 1, \ldots, q\}$ to each candidate service center location in $\mathcal{F}_i$. The score that a customer is assigned to a location $j \in \mathcal{F}_i$ is taken as the potential utility gained by the customer if going to the service center at j. Score value 0 means the customer is unwilling to get service from the corresponding location. Since the regression model structure is identical for each $i \in \mathcal{S}$, we omit the customer site index i and re-write $\mathcal{F}_i$ as $\mathcal{G}$ in the following modeling and analysis. Suppose the indices of location in $\mathcal{G}$ are labeled as $\mathcal{G} := \{1, 2, \ldots, g\}$, the N residents are labeled as $\{1, 2, \ldots, N\}$, and the score assigned to a candidate location $j \in \mathcal{G}$ by the resident k is denoted as a_{kj}.

The maximum attraction principle described in (2) may not be satisfied in reality. But it is possible to establish a utility-demand estimation model that approximately meets the maximum attraction principle. For example, we can verify whether the samples from survey satisfy the maximum attraction principle by grouping the customers who have taken the survey as follows:

$$V_j = \{k \in [N] \,|\, a_{kj} \geq 1\} \qquad \forall j \in \mathcal{G}, \tag{16}$$

where $[N] := \{1, \ldots, N\}$. In words, V_j is the set of residents who are willing to go to a service center located at j. According to (2), the samples satisfy the maximum attraction principle exactly if there exists a permutation σ on the indices in $\mathcal{G}$ such that the following inclusive condition holds:

$$V_{\sigma(1)} \subseteq V_{\sigma(2)} \subseteq \ldots \subseteq V_{\sigma(g)}. \tag{17}$$

A logic behind the maximum attraction principle is that if two candidate locations have similar features, then if a customer is attracted by one location, the customer should also be attracted by the other location with high chance. In this case, the difference in the attractability of the two candidate locations viewed by the customer is reflected in the score assigned to the two locations by the customer. The scores in this case are both non-zero indicating that the customer is willing to visit any of them. Only in the case that two candidate locations have some substantial differences (i.e., one is too far away from the customer site or one is located at a bad community), a customer will be willing to visit one location (assigning a non-zero score to it) while unwilling to visit the other one (assigning a zero score to it).

3.1. An inclusive chain representation of the maximum attraction principle

We establish an inclusive chain model to represent the set-level realization of the maximum attraction principle. First we build the subsets (16) using collected samples, and the we sort the indices in $\mathcal{G}$ such that $|V_{\sigma(1)}| \leq |V_{\sigma(2)}| \leq \ldots \leq |V_{\sigma(g)}|$, where σ is a permutation on $\mathcal{G}$ that makes this condition hold. For clarity we assume that $|V_1| \leq |V_2| \leq \ldots \leq |V_g|$ without loss of generality. The inclusive chain model assumes that the score assigned to the candidate locations in $\mathcal{G}$ by a customer from a fixed location should match with one of the following patterns:

$$[0,\ldots,0,a_i^{(i)},\ldots,a_g^{(i)}] \qquad \text{for } i \in \{1,\ldots,g+1\}, \tag{18}$$

where in the i-th pattern, first $i-1$ scores are all zero and the remaining $g-i+1$ scores are all non-zero. By convention, the $(g+1)$-th pattern is just $[0,\ldots,0]$. In a probabilistic flavor, the inclusive chain model (ICM) for the score vector can be formally established as a statistical model presented as

$$
\begin{aligned}
&\xi \stackrel{\mathcal{D}}{=\!=} \sum_{i=1}^{g+1} \mathbf{1}\{B=i\}[0,\ldots,0,Q_i^{(i)},\ldots,Q_g^{(i)}], \\
&P(Q_j^{(i)}=r) = \pi_{jr}^{(i)} \qquad \forall i \in \{1,\ldots,g\}, \ \forall j \in \{i,\ldots,g\}, \ \forall r \in \{1,\ldots,q\}, \\
&P(B=i) = p_i \ \forall i \in \{1,\ldots,g+1\}, \ \sum_{i=1}^{g+1} p_i = 1, \\
&\sum_{r=1}^{q} \pi_{jr}^{(i)} = 1 \quad \forall i \in \{1,\ldots,g\}, \ \forall j \in \{i,\ldots,g\},
\end{aligned}
\tag{ICM}
$$

where ξ is the random score vector, B is an indicator random variable that selects the pattern matching with the score vector, p_j is the probability of score vector matching with the j-th pattern, and $\left\{\pi_{jr}^{(i)} \,\middle|\, r=1,\ldots,q\right\}$ gives the probability distribution of the random score $Q_j^{(i)}$. Note that p_i and $\pi_{jr}^{(i)}$ are model parameters that can be determined by fitting the inclusive chain model with collected samples. The following proposition connects the inclusive chain model with the set-level maximum attraction principle.

PROPOSITION 1. *If the score vector of every customer taking the survey follows the inclusive chain model, then the set-level maximum attraction principle $V_1 \subseteq V_2 \subseteq \ldots \subseteq V_g$ is satisfied by the collection of samples almost surely.*

Proof. We prove it contradiction. Suppose there exist subsets V_i and V_j (with $i < j$) such that $V_i \setminus V_j \neq \emptyset$ with some positive probability. Suppose k is the customer who is willing to visit V_i but not V_j, and let ξ^k be the score vector of this customer. The definition of V_i implies that $\xi_i \geq 1$ but $\xi^k \notin V_j$ implies $\xi_j^k = 0$. On the other side, for every pattern vector ζ we should have $\mathbf{1}\{\zeta_i > 0\} \leq \mathbf{1}\{\zeta_j > 0\}$ almost surely, which implies that $\mathbf{1}\{\xi_i^k > 0\} \leq \mathbf{1}\{\xi_j^k > 0\}$ almost surely. But this contradicts to $\xi_i^k \geq 1$ and $\xi_j^k = 0$. $\square$

3.2. An adjusted inclusive chain model for incorporating defective score vectors

We consider the problem of fitting the inclusive chain model with the collected samples (score vectors). The first step is to determine the rank (level of attracability) of candidate locations $\mathcal{G}$ involved in the inclusive chain. An empirical method by simply sorting the cardinality of V_i can be used to achieve this. The probability guarantee of this method which will be discussed in a moment. For now, assume that the order has been identified, and suppose $|V_1| \leq |V_2| \leq \ldots |V_g|$ without loss of generality. So empirically, this implies that the rank is $1, 2, \ldots, g$ sorted by level of attractability from low to high. Based on this information, we can establish an inclusive chain model as (ICM), but this model is not capable to handle score vectors in which there is at least one zero-value entry between two non-zero value entries, i.e., $[0, \ldots, 0, 5, 7, 0, 0, 7, 6, 9]$. We call a score vector *defective* if it can not match with any pattern vector in (18) valid for the inclusive chain model. Note that to convert the defective score vector $[0, \ldots, 0, 5, 7, 0, 0, 7, 6, 9]$ into a qualified pattern one can either add two 1 score entries to the middle ($[0, \ldots, 0, 5, 7, 0, 0, 7, 6, 9] \rightarrow [0, \ldots, 0, 5, 7, 1, 1, 7, 6, 9]$) or remove the first two nonzero entries ($[0, \ldots, 0, 5, 7, 0, 0, 7, 6, 9] \rightarrow [0, \ldots, 0, 7, 6, 9]$). The first approach has added a total 2 units of score while the second one has removed a total 12 units of score, and it can be verified that the first approach is the one that makes the minimum total score change among all possible changes that can convert this score vector into a qualified pattern. Based on this rule, we can define the defective score units for an arbitrary score vector as follows.

DEFINITION 2. For a given arbitrary score vector v, the *defective score quantity* is given as

$$\min \|v - w\|_1 \quad \text{s.t. } w \text{ is a qualified pattern of the inclusive chain model.} \tag{19}$$

OBSERVATION 1. The defective score quantity has an upper bound $g - 1$.

Let H_s be the event that the defective score quantity of score vector is s, and m_s is the probability of this event. After introducing these additional parameters, the inclusive chain model can then be adjusted to incorporate all defective score vectors as follows

$$
\begin{aligned}
&\mathbb{E}[\xi|H_0] \overset{\mathcal{D}}{=} \sum_{i=1}^{g+1} \mathbf{1}\{B = i\}[0, \ldots, 0, Q_i^{(i)}, \ldots, Q_g^{(i)}], \\
&P(Q_j^{(i)} = r) = \pi_{jr}^{(i)} \qquad \forall i \in \{1, \ldots, g\}, \ \forall j \in \{i, \ldots, g\}, \ \forall r \in \{1, \ldots, q\}, \\
&P(B = i) = p_i \ \forall i \in \{1, \ldots, g+1\}, \ \sum_{i=1}^{g+1} p_i = 1, \\
&\sum_{r=1}^{q} \pi_{jr}^{(i)} = 1 \quad \forall i \in \{1, \ldots, g\}, \ \forall j \in \{i, \ldots, g\}, \\
&P(\xi \in H_s) = m_s \quad \forall s \in \{0, 1, \ldots, g-1\}, \ \sum_{s=0}^{g-1} m_s = 1.
\end{aligned}
\tag{AICM}
$$

The above model is referred as the adjusted inclusive chain model (AICM).

3.2.1.   A theoretical probability guarantee of identifying the order of the inclusive chain model As mentioned at the beginning of Section 3.2, a simple sorting method can be used to identify the rank of the candidate locations in $\mathcal{G}$ for the AICM. We sort the cardinality of subsets V_j' $(j \in \mathcal{G})$ defined as

$$V_j' = \{k \in [N] \,|\, a_{kj} \geq 1\},$$

where a_k is the score vector of the k-th customer participated in the survey, and a_{kj} is its j-th element. The following proposition provides a probability guarantee on identifying the order of inclusive chain model given that samples obey the AICM.

PROPOSITION 2. *Suppose the set of N samples collected from survey follow the AICM, Let the set of candidate locations be $\mathcal{G} = \{1, \ldots, g\}$. Let E be the event that the rank of inclusive chain model can be successfully identified by sorting the cardinality of V_j' $(j \in \mathcal{G})$. Let $p^* = \min_{j \in \mathcal{G}} p_j$. If $p^* > (1 - m_0)/m_0$, then*

$$P(E) \geq 1 - (|\mathcal{G}| - 1)\exp\Big(-\frac{1}{2}N[m_0(1 + p^*) - 1]^2\Big). \tag{20}$$

Proof.   We first show that with the given assumptions, the rank of the inclusive chain model can be identified almost surely as the total number of samples N goes to infinity. In the analysis, we let σ be the permutation that gives the rank of locations, i.e., $\sigma(1) < \ldots < \sigma(g)$ is the rank of locations from low to high. Let N_j be the number of samples that match with the j-th pattern $[0, \ldots, 0, Q_{\sigma(j)}^{(j)}, \ldots, Q_{\sigma(g)}^{(j)}]$. Based on the probability setting of AICM, and the law of large numbers we should have the following asymptotic relation hold:

$$\frac{N_j}{N} \to m_0 p_j \qquad a.s. \text{ as } N \to \infty \qquad \forall j \in \mathcal{G}. \tag{21}$$

This implies that $N_j > 0$ almost surely as $N \to \infty$. By the definition of V_j', we should have the following bounds for every $|V_j'|$

$$\sum_{k=1}^{\sigma(j)} N_k \leq |V_j'| \leq \sum_{k=1}^{\sigma(j)} N_k + |\cup_{s=1}^{g-1} H_s| \qquad \forall j \in \mathcal{G}. \tag{22}$$

For any i, j satisfying $\sigma(i) < \sigma(j)$, it suffices to show that $|V_{\sigma(i)}'| < |V_{\sigma(j)}'|$ as $N \to \infty$. Indeed we have

$$\frac{|V_{\sigma(j)}'| - |V_{\sigma(i)}'|}{N} \geq \frac{1}{N}\Big(\sum_{k=1}^{\sigma(j)} N_k - \sum_{k=1}^{\sigma(i)} N_k - |\cup_{s=1}^{g-1} H_s| \Big)$$

$$= \frac{1}{N}\Big(\sum_{k=\sigma(i)+1}^{\sigma(j)} N_k - |\cup_{s=1}^{g-1} H_s| \Big) \to \sum_{k=\sigma(i)+1}^{\sigma(j)} mp_k - (1 - m_0) \qquad \text{as } N \to \infty,$$

and $\sum_{k=\sigma(i)+1}^{\sigma(j)} m p_k - (1 - m_0) \geq m p^* - (1 - m_0) > 0$ by assumption. Therefore the rank of the inclusive chain can be identified almost surely as N goes to infinity. Now we prove the non-asymptotic result. Note the probability $P(E)$ can be lower bounded as follows

$$P(E) = P\Big(\bigcap_{i=1}^{g-1} \big\{|V'_{\sigma(i+1)}| - |V'_{\sigma(i)}| > 0\big\}\Big) = 1 - P\Big(\bigcup_{i=1}^{g-1} \big\{|V'_{\sigma(i+1)}| - |V'_{\sigma(i)}| \leq 0\big\}\Big)$$

$$\geq 1 - \sum_{i=1}^{g-1} P\Big(|V'_{\sigma(i+1)}| - |V'_{\sigma(i)}| \leq 0\Big) \geq 1 - \sum_{i=1}^{g-1} P\Big(N_{\sigma(i+1)} - |\cup_{s=1}^{g-1} H_s| \leq 0\Big).$$

Since we have $N_{\sigma(i+1)} - |\cup_{s=1}^{g-1} H_s| = \sum_{k=1}^{N} I_k$, where $\{I_k\}_{k=1}^{N}$ are i.i.d. random variables satisfying $P(I_k = 1) = m_0 p_{i+1}$, $P(I_k = 0) = m_0(1 - p_{i+1})$ and $P(I_k = -1) = 1 - m_0$, which gives $\mathbb{E}[I_k] = m_0(1 + p_{i+1}) - 1 > 0$ by assumption. Let $\bar{I} = \frac{1}{N} \sum_{k=1}^{N} I_k$ Applying Hoeffding's inequality gives the following bound

$$P\Big(N_{\sigma(i+1)} - |\cup_{s=1}^{g-1} H_s| \leq 0\Big) = P\Big(\sum_{k=1}^{N} I_k \leq 0\Big)$$

$$= P\Big(\bar{I} - \mathbb{E}[\bar{I}] \leq 1 - m_0(1 + p_{i+1})\Big) \leq \exp\Big(-\frac{1}{2}N[m_0(1 + p_{i+1}) - 1]^2\Big)$$

$$\leq \exp\Big(-\frac{1}{2}N[m_0(1 + p^*) - 1]^2\Big).$$

Then it follows that $P(E) \geq 1 - (|\mathcal{G}| - 1)\exp\Big(-\frac{1}{2}N[m_0(1 + p^*) - 1]^2\Big)$. □

3.2.2. Model fitting and parameter estimation To avoid redundant notations in the following analysis, we assume that candidate locations are ranked as $1 < \ldots < g$ based on attractability. To fit the AICM with samples, a likelihood function and some regularity terms need to be established. Note that without consideration of any regularity, all model parameters can be estimated naturally using a frequency count approach based on the given samples. For example, an empirical estimation of p_j can be $\hat{p}_j = N_j/N'$, and an empirical estimation of $\pi_{jr}^{(i)}$ can be $\hat{\pi}_{jr}^{(i)} = \frac{N_{jr}^{(i)}}{\sum_{r=1}^{q} N_{jr}^{(i)}}$, where $N_{jr}^{(i)}$ is the number of samples that have $Q_j^{(i)} = r$. However, this estimation can lead to overfitting of new samples. We now focus on establishing a likelihood function and introduce some regularity terms to prevent overfitting. First, the parameter m_s can be estimated as $\widehat{m}_s = \frac{\#\{k \in [N]: \xi_k \in H_s\}}{N}$ for all $s \in \{1, \ldots, g-1\}$. All defective samples can be further used to estimate other parameters with more information. If a sample ξ_k is defective, it can be converted into a qualified score vector ξ'_k by imposing a defective score quantity (Definition 2) to ξ_k, and hence the original samples $\{\xi_k\}_{k=1}^{N}$ can then be converted into a set of qualified score vectors $\{\xi'_k\}_{k=1}^{N}$. We can evaluate the probability of generating ξ'_k by the model using the model parameters, and denote this probability as $P(\xi'_k|p, \pi)$. Then the (logarithmic) likelihood function of generating $\{\xi'_k\}_{k=1}^{N}$ can then be formulated as

$$\mathcal{L}(p, \pi) = \frac{1}{N} \sum_{k=1}^{N} \log P(\xi'_k|p, \pi). \tag{23}$$

Note that maximizing the above likelihood function with only probability-normalization constraints leads to a frequency estimation of all parameters. Some regularity terms can be added to balance this trend. First, for score vectors in the same pattern, parameters that lead to smaller variance of score on a given candidate location are preferred, and hence the following regularity terms can be added (with some regularity coefficients):

$$R_{ij}^1(\pi) := \mathbb{Var}[Q_j^{(i)}] = \sum_{r=1}^q \left(r^2 \pi_{jr}^{(i)}\right) - \left(\sum_{r=1}^q r \pi_{jr}^{(i)}\right)^2$$
$$\forall i \in \{1, \ldots, g\}, \ \forall j \in \{i, \ldots, g\}.$$

Second, consider any two consecutive locations j and $j+1$ in the inclusive chain. The two locations are both included in patterns $H_1, \ldots, H_j$. In the pattern H_i $(1 \le i \le j)$, the mean score difference between the two locations is given by

$$w_j^i = \mathbb{E}[Q_{j+1}^{(i)} - Q_j^{(i)}] = \sum_{r=1}^q r(\pi_{j+1,r}^{(i)} - \pi_{j,r}^{(i)}).$$

For every pattern that contains locations j and $j+1$, we expect the above mean score difference to be similar. To enforce this similarity, we can add the following regularity terms (with some regularity coefficients):

$$R_j^2(\pi) := \frac{1}{j} \sum_{i=1}^j (w_j^i)^2 - \frac{1}{j^2}\left(\sum_{i=1}^j w_j^i\right)^2 \qquad \forall j \in \{1, \ldots, g-1\}.$$

After incorporating the above regularity terms, the parameters can be estimated by solving the following constrained minimization problem of the loss function.

$$\min \ L(p, \pi) := \frac{2\lambda_1}{g(g+1)} \sum_{i=1}^g \sum_{j=i}^g R_{ij}^1(\pi) + \frac{\lambda_2}{g-1} \sum_{j=1}^{g-1} R_j^2(\pi) - \frac{1}{N} \sum_{k=1}^N \log P(\xi_k'|p, \pi)$$
$$\text{s.t.} \ \sum_{r=1}^q \pi_{jr}^{(i)} = 1 \qquad \forall i \in \{1, \ldots, g\}, \ \forall j \in \{i, \ldots, g\}, \tag{24}$$
$$\sum_{j=1}^{g+1} p_j = 1,$$

where λ_1 and λ_2 are regularity coefficients that can be tuned using a cross-validation approach. Note that we add a negative sign to the logarithmic likelihood function when it is incorporated into the loss function $L(p, \pi)$. Notice that (24) is a nonconvex optimization problem with continuous parameters and simple constraints, and it can be solved to local optimality very efficiently using numerical optimization solvers.

3.3. Generation of utility and demand samples guided by AICM

Once the regression problem (24) is solved to give an estimation of the AICM parameters, the AICM can be used to estimate the (mean) utility value, guide the generation of demand samples. First, the utility value $u_{\cdot,j}$ (the customer site index is ignored since the AICM is established for any given customer site) can be estimated as follows:

$$\hat{u}_{\cdot,j} = \sum_{i=1}^{j} \hat{p}_i \Big(\sum_{r=1}^{q} r \hat{\pi}_{jr}^{(i)} \Big), \tag{25}$$

where $\hat{p}$ and $\hat{\pi}$ are the estimation of AICM parameters. Generating a number of joint demand samples and use them to construct the ambiguity set can be a challenging problem in practice. There are two major obstacles in achieving this. (a) Scaling factor: since only a limited amount of survey samples are collected, how to scale the demand estimation based on the samples to roughly estimate the total demand from a customer site? (b) Demand catenation: how to catenate the estimated demand samples for each customer site to get joint demand vectors (for all customer sites) as samples that can be directly input to the ambiguity set? We try to give a few guidelines for dealing with these problems in this paper.

For (a), the total potential demand from a customer site could be estimated via a different channel which is not the focus of this paper. Some investigation on the demand estimation and especially the healthcare service demand estimation is conducted in (Bajari and Benkard 2005, Cote and Stephen 2001, Doi et al. 2017, Griffin et al. 2008). The survey of scoring on candidate locations are only used to quantify the attractability of each candidate location that can lead to different estimated demand for different candidate locations when the total potential demand is given. In the next paragraph for addressing (b), we provide a estimation of demand associated with each candidate location based on assuming the total potential demand is N (the number of collected samples). In the case that we have a separate estimation A (can be a random variable) of total potential demand the scaling factor can be defined as $\frac{A}{N}$, and it can be used to scale up estimated demand associated with different candidate locations. In the following discussion for (b), we assume that the demand samples generated (using Algorithm 1-2) will be multiplied by the scaling factor (could be customer site dependent), and omit this factor in these algorithms.

For (b), one approach is to assume the distributions of demand $D_i := \{D_{ij} \,|\, j \in \mathcal{F}_i\}$ for different customer sites i are independent. In this case we can generate demand samples for each i and combine them (in an outer-product manner) to construct samples for the random vector $D := \{D_{ij} \,|\, i \in \mathcal{S}, j \in \mathcal{F}_i\}$. For example, if we have constructed $|\Omega_i|$ samples for D_i which are denoted as $\{D_i^{\omega} \,|\, \omega \in \Omega_i\}$, then we can get samples for D as $\{[D_i^{\omega_i} : i \in \mathcal{S}] \,|\, \omega_i \in \Omega_i \,\forall i \in \mathcal{S}\}$. In this manner, the number of samples for D can be as large as $\prod_{i \in \mathcal{S}} |\Omega_i|$. To reduce the sample size for D, one can

select a desired number N of combinations of samples for each D_i at random. Another way is to assume the demand from different costumer sites are strongly correlated. As a consequence, we can divide the samples for each D_i into several bands according to the magnitude of the demand. Then we select samples from the same band for different D_i's to construct a sample for D. A possible way of defining the bands can be based on different intervals of quantiles. For example, one can define bands corresponding to 5 intervals of percentage $[0, 0.2], [0.2, 0.4], \ldots, [0.8, 1.0]$ for each D_i. This way of constructing samples (scenarios) has been used in (Mehrotra et al. 2020). Now let us focus on how to generate samples for each D_i. The pseudo code for generating a sample for D_i is given in Algorithm 1. Note that by repeatedly call this algorithm, one can generate any given number of samples for D_i. A pseudo-code for generating a sample for D is given as Algorithm 2.

Algorithm 1 An algorithm for generating a mixed-defective sample for D_i ($i \in \mathcal{S}$) following AICM.

1: Input data: score vector samples $Y_i = \{\widehat{\xi}^k\}_{k=1}^N$ that have been collected from the survey at the customer site i.

2: Determine the rank of ICM based on the score vectors in Y_i. Sort the indices in $\mathcal{F}_i$ (from low to high) based on this rank.

3: Based on this rank, partition Y_i into qualified subset Y_i^q of samples and defective subset Y_i^d of samples. Let $N^q = |Y_i^q|$, $N^d = |Y_i^d|$ and $\rho = N^q/(N^q + N^d)$.

4: Solve (24) to get an optimal estimation of probability parameters denoted as p^*, π^* and substitute p^*, π^* into (ICM).

5: Generate N i.i.d. score vectors denoted as $\{\xi^k\}_{k=1}^N$ such that each score vector is generated using the way given as follows: with probability ρ sample ξ^k following the fitted (ICM), and with probability $1 - \rho$ sample ξ^k uniformly from Y_i^d.

6: Let $\widehat{D}_{ij} = \mathrm{card}\{k \mid \xi_j^k \geq 1\}$ for all $j \in \mathcal{F}_i$.

7: Return the sample $\widehat{D}_i = [\widehat{D}_{ij} : j \in \mathcal{F}_i]$.

Algorithm 2 An algorithm for generating a mixed-defective sample for D following AICM.

1: Input data: (1) same input as Algorithm 1 for all $i \in \mathcal{S}$. (2) sample size N_i for every $i \in \mathcal{S}$ and N.

2: Generate N_i i.i.d. mixed-defective samples for every $i \in \mathcal{S}$ using Algorithm 1.

3: Catenate the mixed-defective demand samples $\{\widehat{D}_i^k\}_{k=1}^{N_i}$ over $i \in \mathcal{S}$ (either in an independent approach or a strong correlated approach) and draw N mixed-defective samples $\{\widehat{D}^k\}_{k=1}^N$ for D.

4: Return the mixed-defective samples $\{\widehat{D}^k\}_{k=1}^N$.

4. Numerical Investigation

The numerical investigation consists of two major parts: the numerical analysis of the AICM and the computational performance of the (DRO-FL) model. In Section 4.1, we provide numerical analysis of the AICM, where we simulate certain amount of score vector samples and use them to fit the AICM. In Section 4.2, we show the computational performance of solving generated instances of the (DRO-FL) model, and provide sensitivity analysis of the robust optimal locations with respect to the sample size and the choice of radius in the ambiguity set.

4.1. Numerical analysis of the AICM

For the numerical analysis of AICM, we focus on a specific customer site with 4 candidate service center locations labeled as $L_1 \sim L_4$ (the rank of attractability is unknown). We consider 6 possible scores $\{0, 1, \ldots, 5\}$. We simulate N i.i.d. score vectors with some defected ones and use them to fit the AICM. The generation of each score vector is based on the following procedures. First, we simulate a 4-dimensional real vector v following the Gaussian distribution $\mathcal{N}(w, \Sigma)$ with $w = [1, 3, 4.5, 2]$ corresponding to the mean score of $L_1 \sim L_4$ respectively, and $\Sigma = \text{diag}[0.8^2, 1.5^2, 1, 1.5^2]$ as the covariance matrix. Next, each entry of v is rounded to the closest integer from $\{0, 1, \ldots, 5\}$. Finally, with probability 0.8 we accept this score vector, while with probability 0.2 we randomly select an entry of v, set it to be zero and accept the resulting score vector. We follow the steps instructed in Section 3.2 to fit the AICM. An essential step is to solve the nonlinear optimization problem (24). This problem is implemented in Julia 1.5.2 (Bezanson et al. 2017) with optimization package JuMP dev (Dunning et al. 2017) as the modeling interface and Ipopt 3.13.2 (Wächter and Biegler 2006) as the solver. For simplicity, we let $\lambda_1 = \lambda_2 = \lambda$ in (24). We use a 4-fold cross validation approach to select an appropriate regularization parameter λ. The parameter selection consists of two rounds. The first round is a rough selection, where the candidate $\lambda^{(1)}$ is selected from $\{0, 0.1, 0.2, \ldots, 5\}$ based on the out-of-sample performance. Suppose $\lambda_*^{(1)}$ is the selected value after the first round. In the second round, we select an optimal λ^* from the set $\left\{\lambda_*^{(1)} \pm 0.01k \,\middle|\, k = 0, 1, \ldots, 10\right\}$ based on the out-of-sample performance. The out-of-sample performance is measured by the averaged logarithm of likelihood. The first-round selection results are given in Figure 1 for three different sample sizes $N = 50, 100, 200$. The second-round selection results are given in Table 2, from which we see that the optimal λ is 2.98, 2.98 and 0.37 for $N = 50, 100, 200$, respectively. The curves in Figure 1 indicate that for a small sample size ($N = 50$) a larger regularization parameter lead to a better out-of-sample performance, while for a larger sample size ($N = 200$) the best out-of-sample performance is achieved at a smaller regularization parameter. This behavior is intuitive since training the model with a small sample size can lead to over fitting if the model parameters are not sufficiently regularized. When more samples are available to train the model, the regularization can be less restrictive.

Table 1 Mean scores for each location at different sample sizes.

Samples	Loc. 1	Loc. 2	Loc. 3	Loc. 4
50	0.83	2.79	3.86	1.67
100	0.88	2.91	3.89	1.87
200	0.87	2.88	3.94	1.78
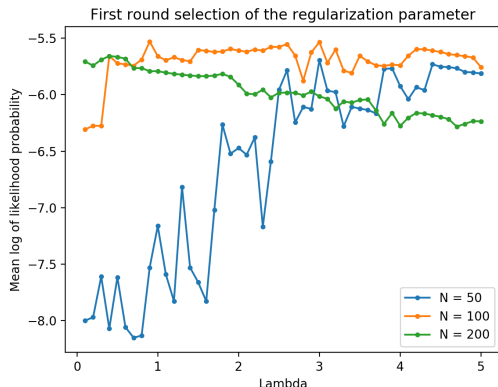
Figure 1 First round of λ selection.

Table 2 Second round of λ selection.

| N	50	100	200
range	$[2.9, 3.0]$	$[2.9, 3.0]$	$[0.3, 0.4]$
λ^*	2.98	2.98	0.37

Finally, we fit the AICM using N samples and the optimal λ parameter. The probability parameters p and π from AICM can be used to estimate the probability distribution on scores for each candidate location. The results are given in Figure 2 for the 4 candidate locations. The probability distribution on score values reveal that the rank of attractability from high to low is: Location 3 > Location 2 > Location 4 > Location 1, which matches with the underline stochastic model that generates the scores. The mean scores corresponding to the four locations are given in Table 1.

4.2. Computational performance of the (DRO-FL) model

We investigate the computational performance of solving numerical instances of the (DRO-FL) model. To generate numerical instances, we create a 100×100 two-dimensional square as the map. The customer sties are points inside the square, and the coordinates of each customer site are uniformly drawn from the range $[0, 100]^2$. We let $\mathcal{F} = \mathcal{S}$, which means each customer site can be a candidate location of a service center. We sort the set of distances $W = \{d(i,j) \mid \forall i \in \mathcal{S}, \forall j \in \mathcal{S}, i < j\}$ from small to large, and let d_c be the 5% quantile of the sorted list. For every customer site i, we select all customer sites (including i) that are within d_c distance from i to form the set $\mathcal{F}_i$. We let the utility completely depend on the distance. Specifically, we set $u_{ii} = 5$ for all $i \in \mathcal{S}$. We also let the first element in W correspond to utility 5, the last element in W correspond to utility 0.5, and any other elements correspond to utility values obtained using a linear interpolation. For a sample ω, the demand D_{ij}^ω is drawn from the normal distribution $\mathcal{N}(120u_{ij}, (12u_{ij})^2)$. The capacity of each candidate service center is set to be 1000. We assume the cost of opening each service center
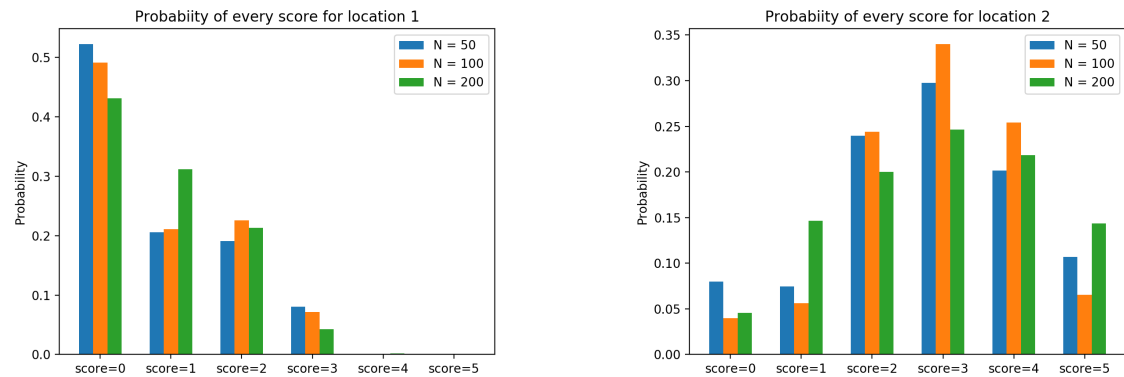
(a) Probability distribution on scores of location 1.

(b) Probability distribution on scores of location 2.
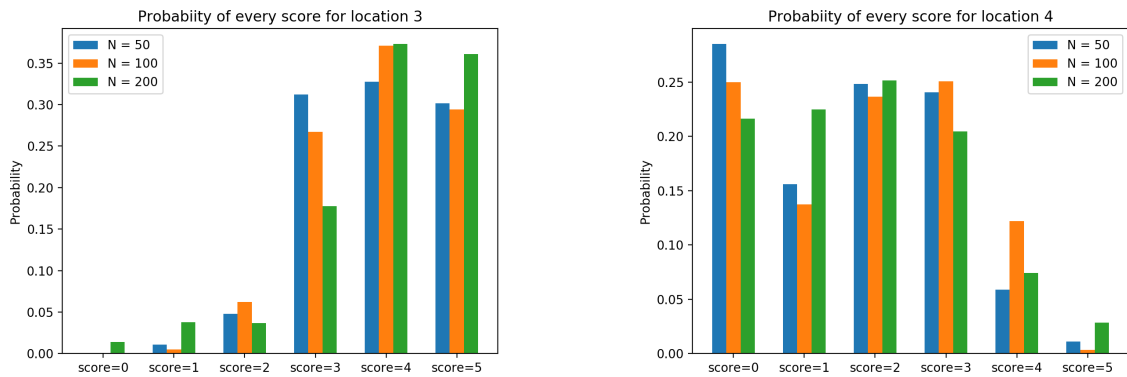
(c) Probability distribution on scores of location 3.

(d) Probability distribution on scores of location 4.

Figure 2   **Probability distribution on scores for each location at different sample sizes.**

is identical (an unit cost), and the budget B in this case is equivalent to the maximum number of service centers that can be opened. We have generated 18 instances of (DRO-FL), and each instance is characterized by the number $|\mathcal{S}|$ of customer sites (also candidate locations) and the budget B (number of maximum service centers that can be allocated). For each instance, we have tried three different sample-size options ($N = 100, 200, 500$), and the total variation distance is set to be 0.2 in these calculation. The (DRO-FL) model is implemented and solved using a Gurobi Python interface (Python 3.7 and Gurobi 9.0). For every instance, the master problem and all scenario sub-problems are solved using a single 2.50GHz CPU. The computational results for the 12 instances are shown in Table 3. It is shown that all instances have been solved to optimality within 15 master iterations (see Algorithm 3). The first 9 instances are solved within 120 seconds, and large instances are solved within 1.5 hours where the majority of computational time is spent on the scenario sub-problems. The range of the robust optimal objective is less than 0.7% across

Table 3 Computational performance of solving 12 numerical instances of (DRO-FL). **In all instances, the total variation distance for the ambiguity set is set to be 0.2. All instances are solved to optimality.**

		$N = 100$			$N = 200$			$N = 500$		
$|\mathcal{S}|$	B	Obj.$(\times 10^3)$	Iters.	T(s)	Obj$(\times 10^3)$	Iters.	T(s)	Obj$(\times 10^3)$	Iters.	T(s)
20	3	9.60	7	5.2	9.66	6	5.8	9.61	6	10.2
20	5	15.57	9	2.3	15.57	7	4.8	15.6	8	11.2
20	7	22.68	6	1.6	22.55	8	5.1	22.7	12	17.0
40	5	17.52	10	8.2	17.55	11	17.8	17.53	10	43.4
40	10	36.30	12	9.8	36.38	9	15.9	36.34	11	49.7
40	15	54.44	11	9.2	54.46	11	19.1	54.38	12	55.7
60	10	35.74	7	14.6	35.74	8	36.5	35.74	6	67.5
60	15	54.64	8	17.4	54.70	10	46.8	54.70	10	120.2
60	20	67.62	8	20.2	67.70	8	37.8	67.61	7	84.6
80	15	60.62	13	61.3	60.47	7	74.3	60.56	6	158.1
80	20	80.13	9	41.6	80.02	8	79	80.11	6	136
80	25	98.3	14	70.6	98.29	14	127.5	98.38	10	220.5
100	20	78.93	10	92.5	78.90	12	204.3	78.89	7	293.1
100	25	100.69	13	120	100.54	11	185.5	100.53	15	637.6
100	30	117.3	15	143.9	117.16	11	186.3	117.15	15	694.3
200	40	173.24	12	1092.9	173.28	7	1184	173.41	11	4678.1
200	50	215.74	13	1102.2	215.74	15	2549.9	215.78	9	3829.4
200	60	256.8	14	1214.2	256.9	8	1360.1	256.78	13	5515.9

three different sample-size options for each instance, and we have also observed that the robust optimal solution remains the same for different sample-size options.

We further investigate the impact of the total variation distance d on the robust optimal decision and the objective value. In this study, we focus on the two instances $(|\mathcal{S}|, B) = (60, 20)$ and $(|\mathcal{S}|, B) = (60, 10)$. For a given sample size N (with $N = 100, 200, 500$), we tune the total variation distance d from 0 to 0.4 with an increment 0.05 and investigate the change of optimal objective of (DRO-FL). The results are shown in Figure 3. Obviously, the objective should decrease with d as a larger d value can allow the worse scenarios get more weight and hence makes the objective decrease. With d increasing from 0 to 0.4, the decrement of the objective is 1.5% (1000 units) for the case $B = 20$, and 1.9% (700 units) for the case $B = 10$. The relative decrement is about one magnitude smaller compared to the 10% relative deviation in generating samples for the demand. It is also observed that the optimal decision of (DRO-FL) does not change with the sample size or the total variation distance.

5. Concluding Remarks

The distributionally-robust service center location problem investigated in this research possesses decision-dependent demand induced naturally from a maximum attraction principle and the number of opened service centers in a neighborhood. The ambiguity set considered in this work is a decision-independent one defined using the total-variational distance while many other metics such as the Wasserstein distance, Phi-divergences, etc. can also be applied to define the ambiguity set with minor modification in the formulation. The modeling approach in this work decouples the endogenous impact (decision dependency) and distributional ambiguity. As a result, the model is

(a) $|\mathcal{S}| = 60$ and $B = 20$.   (b) $|\mathcal{S}| = 60$ and $B = 10$
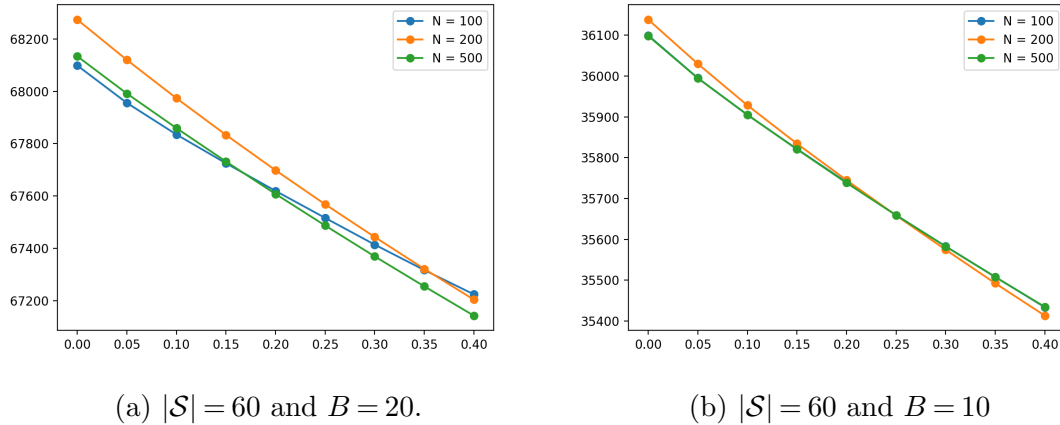
Figure 3   **The impact of total variation distance on the robust optimal objective of** (DRO-FL).

highly computationally efficient for mid- and large scale instances compared to models with an endogenous uncertainty (ambiguity) set. In practice, such decoupling is likely to be more data driven, as there is a higher chance to identify data-based evidence that supports decision-dependent demand than decision-dependent uncertainty of demand.

Appendix. A cutting-plane algorithm for solving (DRO-FL)

The model (DRO-FL) is a special case of the distributionally-robust two-stage stochastic program with a polyhedral ambiguity set.

Algorithm 3 A cutting-plane algorithm for solving (DRO-FL).

1: Initialization: $\eta^* \leftarrow \infty$, $y^* \leftarrow 0^{|\mathcal{F}|}$, $n \leftarrow 0$, $y^{(n)} \leftarrow 0^{|\mathcal{F}|}$, $\eta^{(n)} \leftarrow \infty$ and $SolSet \leftarrow \emptyset$.

2: **while** $y^{(n)}$ is not in $SolSet$. **do**

3:     $SolSet \leftarrow SolSet \cup \{y^{(n)}\}$.

4:     Set $n \leftarrow n+1$.

5:     Solve the master problem (Master) for iteration n, and let $y^{(n)}$ be the optimal solution.

6:     Evaluate the value function $Q(y^{(n)}, D^\omega)$ for each $\omega \in \Omega$.

7:     Solve the linear program (10) to get the worst-case probability measure $\mu^{(n)}$.

8:     Add the inequality (11) to (Master).

9: **end while**

10: Set $y^* \leftarrow y^{(n)}$, and $\eta^* \leftarrow \eta^{(n)}$.

11: Return y^* as the optimal solution and η^* as the optimal objective.

THEOREM 2 (**Bansal et al. 2018**). *Algorithm 3 terminates in a finite number of iterations and return an optimal solution and optimal objective of* (DRO-FL).

Acknowledgments

References

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

Albareda-Sambola M, Fernández E, da Gama FS (2011) The facility location problem with Bernoulli demands. *Omega* 39:335–345.

Bajari P, Benkard CL (2005) Demand estimation with heterogeneous consumers and unobserved product characteristics: a hedonic approach. *Journal of Political Economy* 113(6):1239–1276.

Bansal M, Huang KL, Mehrotra S (2018) Decomposition algorithm for two-stage distributionally robust mixed binary programs. *SIAM Journal on Optimizaion* 28(3):2360–2383.

Baron O, Milner J (2010) Facility location: a robust optimization approach. *Production and Operations Management* 20(5):772–785.

Basciftci B, Ahmed S, Shen S (2020) *Distributionally robust facility location problem under decision-dependent stochastic demand.* https://doi.org/10.1016/j.ejor.2020.11.002.

Berglund PG, Kwon C (2014) Robust facility location problem for hazardous waste transportation. *Netw. Spat. Econ.* 14:91–116.

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM review* 59(1):65–98, URL https://doi.org/10.1137/141000671.

Chan T, Shen Z, Siddiq A (2017) Robust defibrillator deployment under cardiac arrest location uncertainty via row-and-column generation. *Operations Research* 66:358–379.

Cote M, Stephen T (2001) Four methodologies to improve healthcare demand forecasting. *Healthcare Financial Management* 55(5):54.

Doi S, Ide H, Takeuchi K, Fujita S, Takabayashi K (2017) Estimation and evaluation of future demand and supply of healthcare services based on a patient access area model. *International Journal of Environmental Research and Public Health* 14(11):1367.

Döyen A, Aras N, Barbarosoğlu G (2012) A two-echelon stochastic facility location model for humanitarian relief logistics. *Optim. Lett.* 6:1123–1145.

Dunning I, Huchette J, Lubin M (2017) Jump: A modeling language for mathematical optimization. *SIAM Review* 59(2):295–320, URL http://dx.doi.org/10.1137/15M1020575.

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

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

Goel V, Grossmann IE (2006) A class of stochastic programs with decision dependent uncertainty. *Math. Program.* 108:355–394.

Goel V, Grossmann IE, El-Bakry AS, Mulkay EL (2006) A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. *Comp. Chem. Eng.* 30:1076–1092.

Gourtani A, Nguyen TD, Xu H (2020) A distributionally robust optimization approach for two-stage facility location problems. *EURO Journal on Computational Optimization* 8:141–172.

Griffin PM, Scherrer CR, Swann JL (2008) Optimization of community health center locations and service offerings with statistical need estimation. *IIE Transactions* 40(9):880–892.

Gülpınar N, Pachamanova D, Çanakoğlu E (2013) Robust strategies for facility location under uncertainty. *European Journal of Operational Research* 225:21–35.

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

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

Lim MK, Bassamboo A, Chopra S, Daskin MS (2013) Facility location decisions with random disruptions and imperfect estimation. *Manufacturing & Service Operations Management* 15(2):239–249.

Louveaux FV, Peeters D (1992) A dual-based procedure for stochastic facility location. *Operations Research* 40(3):564–573.

Luo FQ, Mehrotra S (2019a) Distributionally robust service center location problem with decision dependent utilities. Tech. report, Northwestern University.

Luo FQ, Mehrotra S (2019b) *Service center location problem with decision dependent utilities.* `https://arxiv.org/pdf/1910.09765.pdf`.

Luo FQ, Mehrotra S (2020) Distributionally robust optimization with decision-dependent ambiguity set. *Optimization Letters* 14:2565–2594.

Mehrotra S, Rahimian H, Barah M, Luo FQ, Schantz K (2020) A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. *Naval Research Logistics* 67(5):303–320.

Murali P, Ordóñez F, Dessouky MM (2012) Facility location under demand uncertainty: response to a large-scale bio-terror attack. *Socio-Economic Planning Sciences* 46:78–87.

Nohadani O, Roy A (2017) Robust optimization with time-dependent uncertainty in radiation therapy. *IISE Transactions on Healthcare Systems Engineering* 7(2):81–92.

Nohadani O, Sharma K (2018) Optimization under decision-dependent uncertainty. *SIAM Journal on Optimization* 28(2):1773–1795.

Noyan N, Rudolf G, Lejeune M (2017) *Distributionally robust optimization with decision-dependent ambiguity set.* `http://www.optimization-online.org/DB_FILE/2018/09/6821.pdf`.

Royset JO, Wets RJB (2017) Variational theory for optimization under stochastic ambiguity. *SIAM Journal on Optimization* 27(2):1118–1149.

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

Snyder LV (2006) Facility location under uncertainty: a review. *IIE Transactions* 38:537–554.

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

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

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

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

Tsur Y, Zemel A (2004) Endangered aquifers: Ggroundwater management under threats of catastrophic events. *Water Resources Research* 40(W06S20).

Vayanos P, Kuhn D, Rustem B (2011) Decision rules for information discovery in multi-stage stochastic programming. *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC).*

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

Wächter A, Biegler LT (2006) On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming* 106(1):25–57.

Wang Q, Batta R, Rump CM (2002) Algorithms for a facility location problem with stochastic customer demand and immobile servers. *Annals of Operations Research* 111:17–34.