Industrial and
Systems Engineering

Distributionally Robust Facility Location with Bimodal Random Demand

Karmel S. Shehadeh[1] and Ece Sanci [2]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA
[2]School of Management, University of Bath, Bath, UK

ISE Technical Report 20T-21

LEHIGH
UNIVERSITY

Distributionally Robust Facility Location with Bimodal Random Demand

Karmel S. Shehadeh[a,*], Ece Sanci[b]

[a] *Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA*
[b] *School of Management, University of Bath, Bath, UK*

Abstract

In this paper, we consider a decision-maker who wants to determine a subset of locations from a given set of candidate sites to open facilities and accordingly assign customer demand to these open facilities. Unlike classical facility location settings, we focus on a new setting where customer demand is bimodal, i.e., display, or belong to, two spatially distinct probability distributions. We assume that these two distributions are ambiguous, and only their mean values and ranges are known. Therefore, we construct a scenario-wise ambiguity set with two scenarios corresponding to the demand's two distinct distributions. Then, we formulate a distributionally robust facility location (DRFL) model that seeks to find the number and locations of facilities to open that minimize the fixed cost of opening facilities and the worst-case (maximum) expectation of transportation and unmet demand costs. We take the worst-case expectations over all possible demand distributions residing in the scenario-wise ambiguity set. We propose a decomposition-based algorithm to solve our min-max DRFL model and derive lower bound inequalities that accelerate the algorithm's convergence. In a series of numerical experiments, we demonstrate our approach's superior computational and operational performance compared with the stochastic programming approach and a DR approach that does not consider the demand's bimodality. Our results draw attention to the need to consider the multi-modality and ambiguity of the demand distribution in many strategic real-world problems.

Keywords: Facility location; Distributionally robust optimization; Bimodal Demand; Mixed-Integer programming; Decomposition-based algorithm

1. Introduction

In this paper, we consider a decision-maker who wants to determine a subset of locations from a given set of candidate sites to open facilities and accordingly assign customer demand to these open facilities. Different than classical facility location settings, we focus on a new setting in which customer demand is bimodal. We use the term "*bimodal*" in a slightly informal way to refer to the tendency of a random demand to display, or belong to, two spatially distinct probability

*Corresponding author.
Email address: kshehadeh@Lehigh.edu, ksheha@umich.edu (Karmel S. Shehadeh)

distributions. For example, depending on the occurrence of a random event, the demand may follow two distinct distributions, one before the occurrence of the event and one after it takes place. We assume that these two distributions are ambiguous (unknown), and only their mean values and ranges are known. The quality of facility location decisions is a function of the fixed cost of opening facilities and a measure of the costs of transportation and unmet demand.

Determining facility locations is a fundamental managerial problem and has many applications such as transportation, logistics, healthcare, to name a few (Ahmadi-Javid et al., 2017; Melo et al., 2009; Owen and Daskin, 1998; Turkoglu and Genevois, 2019). Customer demand often derives facility location decisions. Unfortunately, the precise volume of customer demand is not known at the time of making facility location decisions. Even in a perfect world where we can forecast, estimate, or obtain an approximation of the expected demand, many random events can change/shift customer demand (from low to high, for example), and the probability of such shift is hard to predict in advance. For different scenarios, the support of the demand could be different. At the same time, conditioning on the event realization, the expectation of the demand and thus its distribution may also differ.

In the carsharing industry such as Zipcar, for example, customers choose vehicles to rent for a short time from the most convenient rental location. A competitive company that offers cheaper rental options, more modern cars, or more convenient locations in a certain service region may reduce customer demand for Zipcar, and the probability of observing such an event is not known at the time when decision-makers locate their Zipcars. Moreover, as pointed out by Hao et al. (2019), customer demand for Zipcar or last-mile transportation services such as Taxi and Uber may be higher on rainy/snowy days than other days (i.e., demand's distribution is bimodal). The future weather information at the demand location is uncertain at the point when companies decide where to locate their vehicles (Hao et al., 2019). A similar problem arises when locating charging stations for electric vehicles (EVs). Mak et al. (2013) and Zhang et al. (2017) draw attention to charging infrastructure planning and location as a pressing issue that needs to be dealt with considering different EV adoption rates. For example, the adoption rate could be low or high, and such bimodal uncertainty must be taken into account when locating EV charging stations.

Demand bimodality is also observed in healthcare settings. For example, Senaldi (2018) shows that the demand for blood supply has bimodal distributions (e.g., one for normal state and another for emergency state) in some of the largest hospitals. Thus, location and inventory management of blood banks, blood centers, or blood storage rooms should consider such bimodality to prevent shortage and wastage. The problem of where to and how to stockpile medical supplies in the presence of multi-modal demand distributions is another prominent issue. As of writing this paper, our world is going through unprecedented times, fighting against the novel coronavirus (COVID-19) pandemic. One of the key challenges has been the shortage of personal protective equipment

(PPE). In April 2020, the UN launched the COVID-19 Supply Chain Task Force, coordinated by the World Health Organization (WHO) and the World Food Program (WFP), for the procurement and delivery of PPE along with several other essential supplies (WHO, 2020). According to the Chief of Operations Support and Logistics at WHO, Paul Molinaro, forecasting demand for PPE is a more challenging task when compared to, for instance, forecasting demand for vaccines (Burki, 2020). Therefore, flexibility is particularly important in managing PPE supply chains and facility locations in the wake of a large second wave of classes (e.g., whether a dramatic surge in cases occurs or not). Finally, demand multi-modality and ambiguity are also observed in other contexts such as newsvendor problems (Hanasusanto et al., 2015), supply chain management in the fashion industry (Vaagen and Wallace, 2008), and appointment scheduling (Shehadeh et al., 2020).

A deterministic approach that relies on estimated demand values, which may be easy to solve from a computational perspective, can produce sub-optimal facility location decisions as it does not capture the bimodality and variability of the demand. By incorporating uncertainty, classical two-stage stochastic programming (SP) models seek to find facility location decisions that minimize the first-stage fixed costs of locating facilities and the expected cost of transportation and unmet demand. Here, the expectation is taken with respect to known probability distributions of random demand. While SP is a powerful approach, it often results in intractable formulations. Furthermore, the assumption of full knowledge about the underlying probability distribution, which is usually estimated using limited sample data, might lead to disappointments when the optimal facility location obtained is implemented with a different sample drawn from the same population (Basciftci et al., 2019; Liu et al., 2019; Saif and Delage, 2020).

In reality, it is unlikely that decision-makers have enough data to infer the actual demand distribution. However, it is possible to construct and hedge against an ambiguity set (i.e., a family) of all possible distributions compatible with the given data or expert knowledge about the demand. In this paper, we address the distributional ambiguity and bimodality of customer demand in a facility location problem via scenario-wise distributionally robust (DR) optimization. Specifically, to model bimodality, we first construct a scenario-wise ambiguity set with two scenarios corresponding to two distinct demand distributions. For example, this ambiguity set could be a family of all possible demand distributions before (distribution 1) and after (distribution 2) the occurrence of a random event. We characterize our ambiguity set by the known mean values and ranges of the demand's unknown distributions. No other distributional information is assumed to be available.

Second, we formulate a *distributionally robust facility location* (DRFL) model that seeks to find the number and locations of facilities to open that minimize the fixed cost of opening facilities and the worst-case (maximum) expectation of transportation and unmet demand costs. Here, we take the worst-case expectation over the scenario-wise ambiguity set. Third, we propose a decomposition-based algorithm to solve the min-max DRFL model and derive lower bound inequalities that accelerate the algorithm's convergence. Additionally, we show that the uncapacitated DRFL admits a mixed-integer linear programming (MILP) reformulation.

Finally, with the intent of justifying the value of a scenario-wise DR approach for DRFL, we conduct an extensive numerical experiment. The results demonstrate the superior computational and operational performance of our DR approach as compared with the SP approach and a DR approach that does not consider the bimodality of the demand. More broadly, our results and insights draw attention to the need to consider the impact of customer demand uncertainty when it does not follow one distinct and known distribution in many strategic real-world problems. Thus, our results motivate the need for new approaches that consider the multi-modality and ambiguity of random parameters' distributions in real-world optimization problems.

To the best of our knowledge, and according to the recent survey of Turkoglu and Genevois (2019), this paper is the first to recognize and address the bimodal ambiguity of the demand distribution in classical facility location using a tractable distributionally robust optimization approach. Although we use the occurrence of a random event to illustrate our ideas, present our approach, and derive useful insights, our approach is valid for other applications of DRFL in which the distribution of the demand is bimodal.

The reminder of this paper is structured as follows. In Section 2, we review relevant literature. In Section 3, we formally define DRFL and its reformulation. In Section 4, we introduce our decomposition algorithm to solve DRFL. In Section 5, we test various instances to demonstrate the computational efficacy and solution performance of our DR model as compared to the SP model. We draw conclusions and discuss future directions in Section 6.

2. Relevant Literature

Service and facility location problem. Service and facility location have been extensively studied in the literature for a wide range of private (e.g., industrial plants, warehouses, distribution centers, etc.) and public sectors (e.g., emergency medical services, fire station, etc.). Various operations research techniques have been developed to handle these problems (Chan, 2001). We refer to ReVelle and Eiselt (2005), Turkoglu and Genevois (2019), and Owen and Daskin (1998) for a comprehensive and comparative survey of service facility location problems. Facility location problems under uncertainty have also received significant attention. We refer to Snyder (2006) for a comprehensive review of facility location problems under random demand, characteristics, and cost parameters. We refer to Ahmadi-Javid et al. (2017) for a thorough review of deterministic and stochastic healthcare facility location problems and future directions. Most of this literature assumes that customer demand follows a fully known probability distribution.

Our paper belongs to the distributionally robust (DR) "*static*" facility location (FL) literature, which consider facilities at fixed locations. Most of this literature assumes that demand distribution is ambiguous and unimodal. Recent DR approaches for the static FL problems considering unimodal and unknown probability distributions include Wu et al. (2015), Luo and Mehrotra (2018), Basciftci et al. (2019), Saif and Delage (2020), Santiváñez and Carlo (2018), Wang et al. (2020) and references therein. Shehadeh (2020) proposed the first distributionally robust optimization approach for a "*mobile*" facility routing and scheduling problem with stochastic and unimodal demand. In contrast to these DR facility location papers, we address both the bimodality and ambiguity of demand in the classical static FL setting. We use the same concepts from duality theory and standard DR reformulation techniques employed in these DR facility location papers to derive a solvable reformulation of our min-max model. Our decomposition-based algorithm for DRFL is based on the same theory and art of cutting plane-based algorithms employed in Shehadeh (2020). Our lower bound inequalities are based on the fact that we know the lower bound of the demand as in Shehadeh (2020). Nevertheless, our lower bound inequalities have a different structure than those derived in Shehadeh (2020) due to the differences in the decision variables and objectives. Finally, inspired by Basciftci et al. (2019) and Wang et al. (2020), we derive a structured MILP reformulation of the uncapacitated DRFL.

Stochastic optimization. There are three frameworks for optimization under uncertainty; stochastic programming (SP), robust optimization (RO), and, more recently, distributionally robust (DR) optimization (Chen et al., 2020). Classical SP extends the linear optimization framework to minimize the total expected cost associated with the optimal *here-and-now* (i.e., first-stage) and *wait-and-see* (i.e., second-stage recourse) decisions under known probability distributions of random parameters. We refer to Birge and Louveaux (2011) and Shapiro et al. (2014) for a thorough discussion on SP. While SP is a powerful approach for modeling uncertainty, its applicability is limited to the cases in which we have sufficiently large data to characterize uncertainty or when the distribution of the underlying uncertainty is fully known. If we calibrate an SP to a data sample from a biased distribution, then the resulting (*optimistically biased*) optimal decisions will have a disappointing out-of-sample performance when implemented with a different sample drawn from the same population. This phenomenon is well-known as the optimizers' curse (Smith and Winkler, 2006). Furthermore, SP approaches suffer from the "curse of dimensionality," and so they are often computationally expensive and intractable.

Classical RO models assume that uncertain parameters reside in a so-called *uncertainty set* of possible outcomes, and optimization is based on the worst-case scenario occurring within the uncertainty set (Bertsimas and Sim, 2004; Ben-Tal et al., 2015; Soyster, 1973). As argued by Chen et al. (2020), Delage and Ye (2010), and Thiele (2010), sometimes classical RO models can yield overly-conservative (pessimistically biased) solutions and poor expected performances because it cannot capture the distributional information of uncertainty.

DR optimization has been developed in recent years and becomes an attractive approach for addressing optimization problems under distributional uncertainty. DR optimization mitigates RO's conservativeness by incorporating more distributional information (e.g., mean, range, correlation, etc.) about uncertainty beyond the worst-case scenario and leverage available distributional information more robustly than SP. In DR optimization, one assumes that the distribution of uncertain parameters resides in a so-called "*ambiguity set*" and optimization is based on the worst-case distribution within the ambiguity set. The ambiguity set is a family of distributions characterized through certain known properties of the unknown distributions (Mohajerin Esfahani and Kuhn, 2018). One can use information that is easy to compute such as the mean and range of random parameters to construct the ambiguity sets and build DR models that better mimic reality and less conservative than RO. Maybe surprisingly, it turns out that DR models, where the distribution of uncertain parameters is a decision variable, are often more tractable than their SP counterparts in many real-world applications (Delage and Ye, 2010).

Recently, Chen et al. (2020) described a new class of DR optimization with "scenario–wise" or "mixture distributions" ambiguity set. This approach is motivated by the fact that for different scenarios (e.g., before the occurrence of an event vs. after the event takes place), the random parameter could be different (e.g., typical vs. high demand), while conditioning on the scenario realization, the expectation, and distribution of random parameter can also be different. To model such multi-modality, Chen et al. (2020) assumes that the probability distribution of a random parameter is a mixture of several distinct distributions, where each mixture component is an ambiguous distribution with known support and moments (see Appendix A). Despite the advantages of using this scenario-wise DR optimization approach in modeling the multi-modality of random parameters in recent real-world applications (see, e.g., Shehadeh et al. (2020)), no paper adapted this approach to address demand's multi-modality in classical static facility location problems.

2.1. Contributions

This paper contributes to both the static facility location and DR optimization literature. We propose the first scenario-wise DR optimization approach that addresses demand's bimodal ambiguity in a static facility location setting. We first construct a special case of Chen et al. (2020)'s scenario-wise ambiguity set with two scenarios corresponding to demand's two distinct probability distributions. Then, we formulate a two-stage min-max DRFL model that seeks to find the number and locations of facilities to open that minimize the fixed cost of opening facilities and the maximum expectation of transportation and unmet demand costs. We take the worst-case expectations over all possible demand distributions residing in the scenario-wise ambiguity set. To solve our min-max DRFL model, we propose a computationally efficient decomposition-based algorithm and derive lower bound inequalities that accelerate the algorithm's convergence. We also derive an equivalent MILP reformulation of the uncapacitated DRFL problem with bimodal demand. To date, no paper addresses the bimodal distributional ambiguity of the demand in both the capacitated and uncapacitated DRFL settings.

We conduct extensive computational experiments that compare our DR approach with an SP-bimodal approach that considers demand's bimodality and SP-plain and DR-plain approaches that ignore demand's bimodality. Our results demonstrate our approach's superior computational and operational performance compared to both the SP-bimodal and the DR-plain and SP- plain approaches. More broadly, our results and insights draw attention to the need to consider the demand distribution's multi-modality and ambiguity in many facility location and other strategic real-world problems.

3. DRFL Formulation and Analysis

We present a distributionally robust facility location (DRFL) problem, in which customer demand is bimodal, and we need to determine a subset of locations from a given set of candidate sites to open facilities and accordingly assign customer demand to these open facilities. The quality of facility location decisions is a function of the fixed cost of opening facilities and a measure of the costs of transportation and unmet demand. In Section 3.1, we formaly define DRFL, introduce our scenario-wise ambiguity set of the demand, and present our min-max DRFL model. Then, in Section 3.2, we derive an equivalent and solvable reformulation of the min-max DRFL model.

Notation: For $a, b \in \mathbb{Z}$, we define $[a] := \{1, 2, \ldots, a\}$ and $[a, b]_{\mathbb{Z}} := \{c \in \mathbb{Z} : a \leq c \leq b\}$. The abbreviations "w.l.o.g." and "w.l.o.o." respectively represent "without loss of generality" and "without loss of optimality." Table 1 summarizes other notation.

3.1. ***DRFL Definitions and formulation***

We consider a set of I candidate locations for building facilities and J customer sites that generate demand. Customer demand at each node j is random and "bimodal," i.e., display, or belong to, two spatially distinct distributions. For example, depending on the occurrence of a random event, the demand may follow two distinct and unknown probability distributions, one before the occurrence of the event and one after it takes place. Note that the bimodality of the demand could be due to other reasons than the occurrence of a random event. For ease of presentation, hereafter, we use the idea that the demand bimodality is a function of a random event to present our models and derive useful insights.

To model the unknown and multi-modal distribution of the demand, we adopt a special case of the so-called "scenario-wise" or "mixture distributions" ambiguity set, recently proposed by Chen et al. (2020). For different scenarios (r), the random variable ξ could be different, while conditioning on the scenario realization, the expectation and distribution of ξ can also be different. Mathematically, Chen et al. (2020) assume that the distribution $\mathbb{P}$ of random parameter ξ is a

Table 1: Notation.

Indices	
i	index of location, $i = 1, \ldots, I$
j	index of costumer, $j = 1, ..., J$
Parameters and sets	
I	number, or set, of locations
J	number, or, set of costumers
f_i	cost of opening a facility
$t_{i,j}$	unit transportation cost from facility i to customer site j
p_j	penalty of not satisfying demand at customer site j
C_i	capacity of facility i
d_j	demand at customer site j
$\underline{d}_j^{\mathrm{B}}/\overline{d}_j^{\mathrm{B}}$	lower/upper bound of the demand before the event at customer site j
$\underline{d}_j^{\mathrm{A}}/\overline{d}_j^{\mathrm{A}}$	lower/upper bound of the demand after the event at customer site j
First-stage decision variables	
y_i	$\begin{cases} 1, & \text{if a facility is open at location } i, \\ 0, & \text{otherwise.} \end{cases}$
Second-stage decision variables	
$x_{i,j}$	amount of demand of customer site j satisfied by facility i
u_j	amount of unsatisfied demand of customer site j

mixture of R distinct distributions, i.e., $\mathbb{P} = \sum_{r=1}^{R} p_r \mathbb{P}_r$ with $\sum_{r=1}^{R} p_r = 1$, where each mixture component $\mathbb{P}_r$ is an ambiguous distribution with known support and moments. For completeness, we provide the high-level details of Chen et al. (2020)'s scenario-wise ambiguity set in Appendix A, and we refer to Chen et al. (2020) for the mathematical details.

Herein, we propose a special case of Chen et al. (2020)'s ambiguity set. In particular, we assume that the distribution $\mathbb{P}_j$ of the demand at customer site j is a mixture of two distinct distributions $\mathbb{P}^{\mathrm{B}}$ and $\mathbb{P}^{\mathrm{A}}$, where $\mathbb{P}^{\mathrm{B}}$ and $\mathbb{P}^{\mathrm{A}}$ are the distributions of the demand before and after the occurrence of a random event, respectively. Accordingly, we let $\mathbb{P}_j = q_j\mathbb{P}^{\mathrm{B}} + (1 - q_j)\mathbb{P}^{\mathrm{A}}$. Furthermore, we model q_j as a 0-1 Bernoulli random variable such that $q_j = 1$ if the event does not occur and $q_j = 0$ otherwise (i.e., $1 - q_j$ is the probability of the event). In other words, $q_j = 1$ if the demand follows $\mathbb{P}^{\mathrm{B}}$ (distribution 1) and $q_j = 0$ if it follows $\mathbb{P}^{\mathrm{A}}$ (distribution 2). Accordingly, demand d_j at each customer site j is $d_j = q_j d_j^{\mathrm{B}} + (1-q_j)d_j^{\mathrm{A}}$, where $d_j^{\mathrm{B}} \sim \mathbb{P}^{\mathrm{B}}$ and $d_j^{\mathrm{A}} \sim \mathbb{P}^{\mathrm{A}}$. Note that given the demand's bimodal marginal distribution at each customer site, the demand's joint probability distribution could be multi-modal.

We further assume that only certain properties are known about the unknown distributions $\mathbb{P}^{\mathrm{B}}$ and $\mathbb{P}^{\mathrm{A}}$ of demand d_j at customer site j. In particular, we assume that we know the support (i.e., upper and lower bound) and the mean values of the random parameters $(q, d^{\mathrm{B}}, d^{\mathrm{A}})$. Mathematically, we consider support $S = S^{\text{q}} \times S^{\text{B}} \times S^{\text{A}}$, where S^{q}, S^{B}, and S^{A} are respectively the supports of random parameters q, d^{B}, and d^{A} defined as follows:

$$
\begin{aligned}
&\mathcal{S}^{\text{q}} := \{0,1\}^J \\
&\mathcal{S}^{\text{B}} := \left\{ d^{\text{B}} \geq 0 : \ \underline{d}^{\text{B}}_j \leq d^{\text{B}}_j \leq \overline{d}^{\text{B}}_j, \forall j \in [J]. \ \right\} \\
&\mathcal{S}^{\text{A}} := \left\{ d^{\text{A}} \geq 0 : \ \underline{d}^{\text{A}}_j \leq d^{\text{A}}_j \leq \overline{d}^{\text{A}}_j, \forall j \in [J]. \ \right\}
\end{aligned}
$$

In addition, we let μ^{q}, μ^{B}, and μ^{A} represent the mean values of q, d^{B}, and d^{A}, respectively. We denote $\xi := [q, d^{\text{B}}, d^{\text{A}}]^\top$ and $\mu := \mathbb{E}_{\mathbb{P}}[\xi] = [\mu^{\text{q}}, \mu^{\text{B}}, \mu^{\ \text{A}}]^\top$ for notational brevity. Then, we consider the following mean-support ambiguity set $\mathcal{F}(S, \mu)$:

$$
\mathcal{F}(S,\mu) := \left\{ \mathbb{P} \in \mathcal{P}(S) : \begin{array}{l} \int_S d\mathbb{P} = 1 \\ \mathbb{E}_{\mathbb{P}}[\xi] = \mu \end{array} \right\} \equiv \left\{ \mathbb{P} \in \mathcal{P}(S) : \begin{array}{ll} \int_S d\mathbb{P} = 1 & \\ \int_S d^{\text{B}}_j \ d\mathbb{P} = \mu^{\text{B}}_j & \forall j = 1, \ldots, [J] \\ \int_S d^{\text{A}}_j \ d\mathbb{P} = \mu^{\text{A}}_j & \forall j = 1, \ldots, [J] \\ \int_S q_j \ d\mathbb{P} = \mu^{\text{q}}_j & \forall j = 1, \ldots, [J] \end{array} \right\} \tag{1}
$$

where $\mathcal{P}(S)$ in $\mathcal{F}(S, \mu)$ represents the set of probability distributions supported on S and each distribution matches the mean values of q, d^{B}, and d^{A}. For all $i \in [I]$, we let binary variable y_i represent the location decision such that $y_i = 1$ if a facility is open at location i and $y_i = 0$ otherwise. For all $i \in [I]$ and $j \in [J]$, we let decision variable $x_{i,j}$ represent the amount of satisfied demand at customer site j by facility i. We let decision variable u_j represent the amount of unsatisfied demand at each customer site $j \in [J]$. Finally, we let parameters C_i, f_i, $t_{i,j}$, and p_j represent the facility capacity at location i, the cost of opening a facility at location i, the unit transportation cost from location i to customer site j, and the penalty of each unit of unsatisfied demand at customer site j, respectively. Using this notation and ambiguity set $\mathcal{F}(S, \mu)$, we formulate DRFL as

$$
\text{(DRFL)} \quad \min_{y \in \mathcal{Y} \subseteq \{0,1\}^I} \left\{ \sum_{i \in I} f_i y_i + \sup_{\mathbb{P} \in \mathcal{F}(S,\mu)} \mathbb{E}_{\mathbb{P}}[Q(y, \xi)] \right\} \tag{2a}
$$

where for a given $y \in \mathcal{Y}$ and a joint realization of uncertain parameters $\xi := [q, d^{\text{B}}, d^{\text{A}}]^\top$

$$
\begin{aligned}
Q(y,\xi) := \min_{x,u} & \Big(\sum_{j \in J} \sum_{i \in I} t_{i,j} x_{i,j} + \sum_{j \in J} p_j u_j \Big) && && \text{(3a)} \\
\text{s.t.} \quad & \sum_{i \in I} x_{i,j} + u_j = q_j d^{\text{B}}_j + (1 - q_j) d^{\text{A}}_j, && \forall j \in [J] && \text{(3b)} \\
& \sum_{j \in J} x_{i,j} \leq C_i y_i, && \forall i \in [I] && \text{(3c)} \\
& u_j, x_{i,j} \geq 0, && \forall i \in [I], \ j \in [J] && \text{(3d)}
\end{aligned}
$$

Formulation DRFL searches for facility location decisions that minimizes the total cost of locating facilities and the maximum worst-case of transportation and unmet demand over a family of distributions characterized by ambiguity set $\mathcal{F}(S, \mu)$. Constraints (3b) ensure that demand at each customer site is either satisfied by other locations or penalized, and constraints (3c) respect the capacity of each open facility. Polyhedron $\mathcal{Y}$ can include any constraints related to facility location decisions y.

3.2. ***Reformulation***

In this section, we use duality theory and follow a standard approach in DR optimization to reformulate the min-max DRFL model in (2) to a one that is solvable. We first consider the inner maximization problem $\sup_{\mathbb{P}\in\mathcal{F}(S,\mu)} \mathbb{E}_{\mathbb{P}}[Q(x,\xi)]$ for a fixed facility location decision $y \in \mathcal{Y}$, where $\mathbb{P}$ is the decision variable, i.e., we are choosing the distribution that maximizes the expected value of $Q(y,\xi)$. For a fixed $y \in \mathcal{Y}$, we formulate $\sup_{\mathbb{P}\in\mathcal{F}(S,\mu)} \mathbb{E}_{\mathbb{P}}[Q(y,\xi)]$ as the following linear functional optimization problem.

$$
\begin{aligned}
\max\ & \mathbb{E}_{\mathbb{P}}[Q(y,\xi)] && \text{(4a)}\\
\text{s.t.}\ & \mathbb{E}_{\mathbb{P}}[\xi] = \mu, && \text{(4b)}\\
& \mathbb{E}_{\mathbb{P}}[\mathbb{1}_S(\xi)] = 1 && \text{(4c)}
\end{aligned}
$$

where $\mathbb{1}_S(\xi) = 1$ if $\xi \in S$ and $\mathbb{1}_S(\xi) = 0$ if $\xi \notin S$. In Proposition 1, we show that problem (4) is equivalent to problem (5) (see Appendix B for detailed proof).

Proposition 1. *For any $y \in \mathcal{Y}$, problem* (4) *is equivalent to*

$$
\min_{\rho,\alpha,\lambda} \left\{ \sum_{j\in J} \mu^B_j \rho_j + \mu^A_j \alpha_j + \mu^q_j \lambda_j + \max_{(q,d^B,d^A)\in S} \left\{ Q(y,q,d^B,d^A) + \sum_{j\in J} -(d^B_j\rho_j + d^A_j\alpha_j + q_j\lambda_j) \right\} \right\} \quad (5)
$$

Note that $Q(y,q,d^{\mathrm{B}},d^{\mathrm{A}})$ is a minimization problem, and thus in problem (5) we have a potentially challenging inner max-min problem. We next analyze the structure of $Q(y,q,d^{\mathrm{B}},d^{\mathrm{A}})$ for a fixed y and a realization of $(q,d^{\mathrm{B}},d^{\mathrm{A}})$. Taking the dual of $Q(y,q,d^{\mathrm{B}},d^{\mathrm{A}})$ lead to the following proposition (see Appendix C for a detailed proof).

Proposition 2. *For fixed $y \in \mathcal{Y}$, $\max_{(q,d^B,d^A)\in S} \left\{ Q(y,q,d^B,d^A) + \sum_{j\in J} -(d^B_j\rho_j + d^A_j\alpha_j + q_j\lambda_j) \right\}$ is equivalent to*

$$
\begin{aligned}
\max_{\beta,v}\ & \Bigg\{ \sum_{j\in J} \max\{\overline{d}^B_j, \overline{d}^A_j\}\beta_j + \sum_{i\in I} C_i y_i v_i + \sum_{j\in J} \Big[-\Big(\overline{d}^B_j\rho_j + (\underline{d}^B_j - \overline{d}^B_j)(\rho_j)^+\Big) - \Big(\overline{d}^A_j\alpha_j + (\underline{d}^A_j - \overline{d}^A_j)(\alpha_j)^+\Big) \\
& + (-\lambda_j)^+ \Big] \Bigg\} && \text{(6a)}\\
s.t.\ & \beta_j + v_i \le t_{i,j}, \qquad \forall i \in [I], \forall j \in [J] && \text{(6b)}\\
& \beta_j \le p_j, \qquad \forall j \in [J] && \text{(6c)}\\
& v_i \le 0, \qquad \forall i \in [I] && \text{(6d)}
\end{aligned}
$$

In view of equation 6, formulation (5) is equivalent to:

$$\min_{\rho,\alpha,\lambda} \Big\{ \sum_{j\in J} \Big[\mu_j^{\mathrm{B}}\rho_j + \mu_j^{\mathrm{A}}\alpha_j + \mu_j^{\mathrm{q}}\lambda_j - \big(\overline{d}_j^{\mathrm{B}}\rho_j + (\underline{d}_j^{\mathrm{B}} - \overline{d}_j^{\mathrm{B}})(\rho_j)^+\big) - \big(\overline{d}_j^{\mathrm{A}}\alpha_j + (\underline{d}_j^{\mathrm{A}} - \overline{d}_j^{\mathrm{A}})(\alpha_j)^+\big) + (-\lambda_j)^+\Big] + F(y)\Big\} \tag{7}$$

where

$$F(y) := \max_{(\beta,v)} \Big\{ \sum_{j\in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\}\beta_j + \sum_{i\in I} C_i y_i v_i \Big\}$$
$$\text{s.t.} \quad (\beta, v) \in \Omega := \{(6\text{b}) - (6\text{d})\}$$

Combining the inner problem in the form of (7) with the outer minimization problem in (2), we derive a reformulation of the DR model in (2) as

$$\min \Big\{ \sum_{i\in I} f_i y_i + \sum_{j\in J} \big(\mu_j^{\mathrm{B}}\rho_j + \mu_j^{\mathrm{A}}\alpha_j + \mu_j^{\mathrm{q}}\lambda_j\big) \sum_{j\in J} -\big(\overline{d}_j^{\mathrm{B}}\rho_j + (\underline{d}_j^{\mathrm{B}} - \overline{d}_j^{\mathrm{B}})w_j\big) - \big(\overline{d}_j^{\mathrm{A}}\alpha_j + (\underline{d}_j^{\mathrm{A}} - \overline{d}_j^{\mathrm{A}})z_j\big) + r_j + \delta \Big\} \tag{8a}$$

$$\text{s.t.} \quad y \in \mathcal{Y}, w_j \geq \rho_j,\ w_j \geq 0,\ z_j \geq \alpha_j,\ z_j \geq 0,\ r_j \geq -\lambda_j,\ r_j \geq 0, \quad \forall j \in [J] \tag{8b}$$

$$\delta \geq F(y) = \max_{(\beta,v)\in\Omega} \{\sum_{j\in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\}\beta_j + \sum_{i\in I} C_i y_i v_i\} \tag{8c}$$

Next, we analyze structural properties of function $F(y)$ as a function of variables $y \in Y$ in Proposition (3) (see Appendix D for a detailed proof). In Appendix E, we show that the special case of uncapacitated DRFL admits an equivalent single MILP reformulation.

Proposition 3. *For any fixed values of variables y, $F(y) < \infty$. Furthermore, $F(y)$ is convex and piecewise linear in y with finite number of pieces.*

4. Solution Approaches

Given the two-stage characteristics of the problem and Proposition (3), it is natural to attempt to solve formulation (8) (or equivalently, the DR model in (2)) with a decomposition algorithm. In Section 4.1, we present our decomposition (cutting–plane) algorithm to solve the DRFL model in (8). Then, in Sections 4.2, we derive valid lower bound inequalities for the master problem.

4.1. ***Decomposition Algorithm***

Proposition 3 suggests that constraint (8c) describes the epigraph of a convex and piecewise linear function of decision variables in formulation (8). This observation facilitates us applying a separation-based decomposition algorithm to solve model (8) as in Jiang et al. (2017), Thiele et al. (2009), and Lei et al. (2016). Algorithm 1 presents DRFL–decomposition algorithm. Algorithm 1 is

Algorithm 1: DRFL–decomposition algorithm.

1. Input. Feasible regions $\mathcal{Y}$ and Ω; Set of cuts $\{L(y,\delta) \geq 0\} = \emptyset$; $LB = -\infty$ and $UB = \infty$.

2. Master Problem. Solve the following master problem

$$Z = \min \Big\{ \sum_{i \in I} f_i y_i + \sum_{j \in J} \Big[\big(\mu_j^{\mathrm{B}} \rho_j + \mu_j^{\mathrm{A}} \alpha_j + \mu_j^{\mathrm{q}} \lambda_j\big) - \big(\overline{d}_j^{\mathrm{B}} \rho_j + (\underline{d}_j^{\mathrm{B}} - \overline{d}_j^{\mathrm{B}}) w_j\big) - \big(\overline{d}_j^{\mathrm{A}} \alpha_j + (\underline{d}_j^{\mathrm{A}} - \overline{d}_j^{\mathrm{A}}) z_j\big) + r_j \Big] + \delta \Big\} \tag{9a}$$

$$\text{s.t.} \quad y \in \mathcal{Y}, w_j \geq \rho_j,\ w_j \geq 0,\ z_j \geq \alpha_j,\ z_j \geq 0,\ r_j \geq -\lambda_j,\ r_j \geq 0, \quad \forall j \in [J] \tag{9b}$$

$$L(y,\delta) \geq 0 \tag{9c}$$

and record the optimal solution (y^*, δ^*) and set $LB = Z^*$.

3. Sub-problem. With y fixed to y^*, solve the following problem

$$W = \max_{(\beta,v) \in \Omega} \Big\{ \sum_{j \in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j + \sum_{i \in I} C_i y_i v_i \Big\} \tag{10a}$$

and record the optimal solution (β^*, v^*) and optimal value W^*. Set $UB = \min\{UB,\ W^* + (LB - \delta^*)\}$

4. if $\delta^* \geq \sum_{j \in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j^* + \sum_{i \in I} C_i y_i^* v_i^*$ **then**

stop and return y^* as the optimal solution to the DR formulation (2)

else

add cut $\delta \geq \sum_{j \in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j^* + \sum_{i \in I} C_i y_i v_i^*$ to the set of cuts $\{L(y,\delta) \geq 0\}$ and go to step 2.

end if

finite because we identify a new piece of the function $F(y) = \max_{(\beta,v) \in \Omega} \{ \sum_{j \in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j + \sum_{i \in I} C_i y_i v_i \}$ each time when we augment the set $\{L(y,\delta) \geq 0\}$ in step 4, and the function $F(y)$ has a finite number of pieces (according to Proposition 3). That is, the set $\{L(y,\delta) \geq 0\}$ consists of a finite number of cuts $\delta \geq \sum_{j \in J} \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j^* + \sum_{i \in I} C_i y_i v_i^*$ generated in step 4.

Note that if a feasible solution y^* to the master problem (9) is not optimal, then it will provide an upper bound on DRFL. That is, if y^* and δ^* obtained from step 1 are not optimal, then $\delta^* \leq W^*$, and so $W^* + LB - \delta^*$ is a valid upper bound.

*4.2. **Lower bound inequalities***

Since the master problem is a relaxation of the DR problem (i.e., provide a lower bound), the tightness of the lower bound is the key to convergence efficiency. In this section, we aim to incorporate more second-stage information into the master problem without adding optimality cuts into the master problem by exploiting the specific characteristics of the second-stage (recourse) problem. In Proposition 4 and Proposition 5, we derive lower bound inequalities for the master problem, which exploits the structure of the recourse problem.

Proposition 4. *Inequality* (11) *is a valid lower bound inequality for DRFL.*

$$\delta \geq \sum_{j \in J} \min\{p_j, \min_{i \in I}\{t_{i,j}\}\} \min\{\underline{d}_j^B, \underline{d}_j^A\} \tag{11}$$

Proof. Recall from the definition of the ambiguity set that the lowest demand of each customer site j equals to $\min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$. Now if we assume that the facilities are uncapacitated (i.e., we relax the capacity restriction), then we will be able to satisfy the demand of each customer site $j \in [J]$ at the lowest transportation cost from the nearest open facility $i \in I' := \{i : y_i = 1\}$. Given that $I' \subseteq I$, then the lowest transportation cost from customer site j to the nearest facility $i \in I'$ must be at least equal to or larger than $\min_{i \in I}\{t_{i,j}\} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$. If $p_j \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\} > \min_{i \in I}\{t_{i,j}\} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$, then the second-stage recourse objective cannot be less than $\sum_{j \in J} \min_{i \in I}\{t_{i,j}\} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$. Otherwise, $\delta \geq \sum_{j \in J} p_j \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$. It follows that (11) is a valid lower bound.

□

Proposition 5. *Inequality* (12) *is a valid lower bound inequality for DRFL.*

$$\delta \geq \min_{j \in J} p_j \Big\{ \sum_{j \in J} \min\{\underline{d}_j^B, \underline{d}_j^A\} - \sum_{i \in I} C_i y_i \Big\} \tag{12}$$

Proof. Note that the lowest demand for all customers is $\sum_{j \in J} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$ whereas the total capacity available over all facilities is $\sum_{i \in I} C_i y_i$. This means that at least $\sum_{j \in J} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\} - \sum_{i \in I} C_i y_i$ demands are not satisfied whenever the lowest demand exceeds the total capacity. Since the minimum unit penalty for unsatisfied demand is $\min_{j \in J} p_j$, (12) is a valid lower bound. □

Remark. Note that the strength of the lower bound provided by inequality (11) and (12) depends on the parameter settings of each DRFL instance. Inequality (11) may be stronger than inequality (12) for some DRFL instances, in which $\sum_{j \in J} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\} \leq \sum_{i \in I} C_i y_i$, and so $\{\sum_{j \in J} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\} - \sum_{i \in I} C_i y_i\} \leq 0$. This could happen in DRFL instances with large facility capacities, small facility opening costs, small lower bounds of the demand (e.g., $\sum_{j \in J} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\} = 0$), or a combination of these settings. In fact, for such instances, inequality (12) may be redundant when $\delta \geq 0$. In contrast, in inequality (11), $\sum_{j \in J} \min\{p_j, \min_{i \in I}\{t_{i,j}\}\} \min\{\underline{d}_j^{\mathrm{B}}, \underline{d}_j^{\mathrm{A}}\}$ is always non-negative, and does not depend on decision variables y_i. It is not straightforward to compare the strength of these bounds for all other instances of DRFL. Finally, it is worth mentioning that (11) does not impact the size of the master problem as it does not depend on DRFL decisions, i.e., it is part of the inputs to the optimization.

5. Computational Experiments

In this section, we generate random instances of the capacitated DRFL and compare our DR approach with the SP approach and draw several insights. Specifically, we compare the optimal

Table 2: DRFL instances. Notation: I is # of locations, J is # of customers.

Inst	J	I	Inst	J	I
1	10	5	7	40	20
2	10	10	8	40	40
3	20	10	9	50	25
4	20	20	10	50	50
5	30	15	11	100	50
6	30	30	12	100	100

solutions of the DR-bimodal model in (8) with those yielded by: (1) SP-bimodal, a SP approach that considers bimodality of demand (see Appendix G for the formulation), (2) DR-plain, a DR model that ignores the bimodality of demand (see Appendix F for the formulation), and (3) SP-plain, a SP approach that ignores the bimodality of demand. In the plain models, we ignore the bimodality of demand and assume that it follows a single probability distribution. However, SP-bimodal and SP-plain only differ in how the demand is sampled; from two known distributions and from a single distribution, respectively. The SP models minimize the fixed cost of opening facilities plus the expected transportation and unmet demand costs via the the sample average approximation (SAA) approach (see, e.g., Kim et al. (2015); Kleywegt et al. (2002) for a detailed discussion on SAA).

We summarize our computational study as follows. We first follow a distributional belief to generate N independent and identically distributed (i.i.d) samples of each random parameter. Second, we compute the upper and lower bounds information from the generated samples and use them to obtain the (in-sample) optimal solutions of the DR model. Third, we solve the SP model using the generated sample and compare (1) solution times of DR and SP, (2) optimal facility locations of DR and SP, and (3) the in-sample and out-of-sample simulation performance of the optimal solutions of DR and SP. Section 5.1 presents the details of data generation and experimental design. In Section 5.2, we compare solution times of the DR and SP models. In Section 5.3, we study the efficiency of lower bound inequalities (11)-(12). In Section 5.4, we compare the optimal solutions of the DR and SP models. In Section 5.5, we compare the out-of-sample performance of these solutions. In Section 5.6, we conduct sensitivity analysis, and derive insights into DRFL.

5.1. ***Experimental Design***

We construct 12 DRFL instances based on the same parameter settings and assumptions made in the literature. We summarize our test instances in Table 2. Each of the 12 DRFL instances is characterized by the number of customers J and number of candidate facilities I. For each DRFL instance, we randomly generate a set of potential facility locations and customer sites as uniformly distributed on a 100 by 100 plane (as in Basciftci et al. (2019), Lei et al. (2014), Lei et al. (2016), Shehadeh (2020), and references therein). We compute the transportation cost, $t_{i,j}$, between each candidate location $i \in [I]$ and customer site $j \in [J]$ based on Euclidean distances

(Basciftci et al., 2019; Lei et al., 2014). We generate fixed opening cost f from uniform distribution as $f_i \in U[2000, 5000]$. We set the capacity $C_i = 150$, for all $i \in [I]$. For most of the experiments, we set the unit penalty cost of unmet demand $p_j > \max_i \{t_{i,j}\}$ at each $j \in [J]$ as in Lei et al. (2014) and Lei et al. (2016). We conduct a sensitivity analysis of these parameters in Section 5.6.

We follow the same procedures in the DR applications literature (see, e.g., Jiang et al. (2017), Shehadeh et al. (2020)) to generate random parameters as follows. We randomly sample the mean value of the demand before the event as $\mu^{\mathrm{B}} \in U[20, 40]$ and after the event as $\mu^{\mathrm{A}} \in U[30, 60]$. We set the standard deviation of the demand before and after the event as $\sigma^{\mathrm{B}} = 0.5\mu^{\mathrm{B}}$ and $\sigma^{\mathrm{A}} = 0.5\mu^{\mathrm{A}}$, respectively. To approximate the lower ($\underline{d}^{\mathrm{B}}$, $\underline{d}^{\mathrm{A}}$) and upper bound ($\overline{d}^{\mathrm{B}}$, $\overline{d}^{\mathrm{A}}$) values of ($d^{\mathrm{B}}$, d^{A}), we respectively use the 20%-quantile and 80%-quantile of the N in-sample data. We generate the in-sample data of d^{B} and d^{A} by following lognormal (LogN) distributions. Specifically, we sample $N = 1000$ realizations of d^{B}_j, for all $j \in [J]$, from LogN with (μ^{B}, σ^{B}), and $N = 1000$ realizations of d^{A}_j, for all $j \in [J]$, from LogN with (μ^{A}, σ^{A}). We generate $(q^n_1, \ldots, q^n_j)$, $n \in [N]$, from Bernoulli distribution with $\mu^{\mathrm{q}} = 0.8$. We generate data for the plain models by following the same steps.

For each instance of DRFL, we optimize the SP model with the generated N scenarios and the DR model with the generated mean and support of random parameters. We implemented the SP, DR, and DRFL–decomposition algorithm using AMPL2016 programming language calling CPLEX V12.6.2 as a solver with default settings and relative MIP gap of 1-2%. We ran all experiments on a computer with an Intel Core i7 processor, 2.5 GHz CPU, and 16 GB (1600MHz DDR3) of memory. We imposed a solver time limit of 1 hour.

5.2. ***CPU Time***

In this section, we compare solution times of SP-bimodal and DR-bimodal models. In addition to the default capacity of $C = 150$, we study solution times of SP and DR with $C = 100$ and a tight capacity of $C \in U[20, 50]$ (i.e., uniformly generated as in Basciftci et al. (2019)). For each of the 12 DRFL instances in Table 2 and choice of C, we randomly generate 5 instances as described in Section 5.1 for a total of 180 SAA instances. As detailed in Section 5.1, we respectively use the 20% and 80% of these 180 SAA instances (in-sample data) to approximate the lower and upper bound on random demand for the DR model. Then, we solve each instance using the SAA formulation of the SP model in Appendix G and our DR model via the DRFL–decomposition algorithm.

In Table 3, we compare the minimum (Min), average (Avg), and maximum (Max) SP and DR solution times (in seconds) of the 180 instances. From Table 3, we first observe that solution times increase as the number of customer and locations increase under all values of C. Second, we observe that the SP model takes a significantly longer time to solve each of the 180 instances than the DR model. The SP model can solve all of the 165 SAAs corresponding to instances 1–11 with solution times ranging from 1 to 2787 seconds. The SP model terminates with no optimal solution ($\sim 10\%$ MIP relative gap) for all of the 15 SAAs corresponding to instance 12.

In contrast, solution times of the DR model ranges from 0.1 to 187 seconds for instances 1–10. The DR model can quickly solve all of the 30 DRFL instances corresponding to instance 11 and 12 with a relaxed tolerance level $\epsilon := \frac{UB-LB}{UB} = 0.01$ in Algorithm 1 (i.e., the algorithm terminates with near-optimal solutions). Note that when solving these instances with $\epsilon = 0.01$, the gap remained at 0.01% for several hours. This could indicate that CPLEX finds good integer solutions early, but examine many additional nodes to prove optimality.

The results in Table 3 also suggest that it is easier and faster to solve DRFL instances with a tighter capacity $C \in [20, 50]$ than with a relaxed capacity of $C = 100, 150$. Consider instance 8 (40, 40), for example. Increasing C from $U[20, 50]$ to 100 increases the average solution times of the DR and SP models respectively from 0.2 and 97 to 4.30 and 499 seconds. One possible explanation for the increase in solution times could be that the relaxed capacity allows for satisfying a larger amount of demand, which may cause both the SP and DR models to search for various alternative combinations of facility locations than under the tighter capacity.

Finally, it is worthy of mentioning that C is the only parameter affecting the solution times significantly. Using other values of the fixed cost, f, and unmet penalty, p, we obtain similar solution times. The DR-plain model, which we solve using a decomposition algorithm similar to Algorithm 1 (see Appendix F), has approximately the same computational performance as the DR-bimodal model. The SP-plain model also has roughly a similar computational performance to the SP-bimodal model.

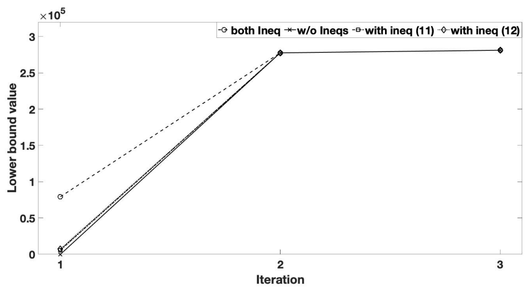
5.3. ***Efficiency of lower bound inequalities***

In this section, we study the efficiency of inequalities (11)–(12). For illustrative purposes and presentation brevity, for this experiment, we focus on instance 8 (40, 40) (we observe similar results for other DRFL instances). We fix $f = 5000$ and $C = \{50, 100, 150\}$ (as in Section 5.2), keep all other parameter settings as described in Section 5.1, and separately solve each instance using our DRFL–decomposition algorithm with and without these inequalities.

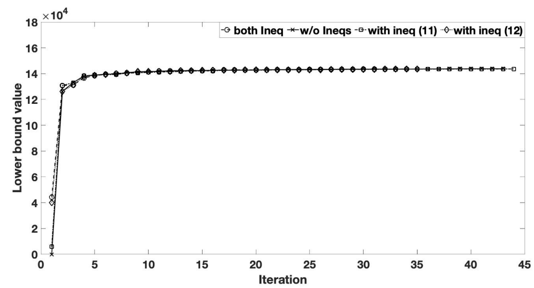
Figure 1 compares the lower bound values obtained by solving instance 8 with both inequalities (both ineqs), without inequalities (w/o ineqs), with inequality (11) alone, and with inequality (12) alone in the master problem of DRFL in (9). From this figure, we first observe that the lower bounds start to converge to roughly similar values after the initial five iterations. Second, we observe that including these inequalities provides a stronger (larger) initial lower bound in the solution process and faster convergence. For example, when $C = 100$, the initial lower bound equals to zero (without the inequalities), 51859 (with both inequalities), 6859 (with inequality (11) only), and 47550 (with inequality (12) only). Note that, in this example, inequality (12) provides a stronger initial lower bound than inequality (11).

Table 3: Solution times (in seconds) using the SP and DR formulations. Instances marked with $*$ are solved with $\epsilon := \frac{UB-LB}{UB} = 0.01$

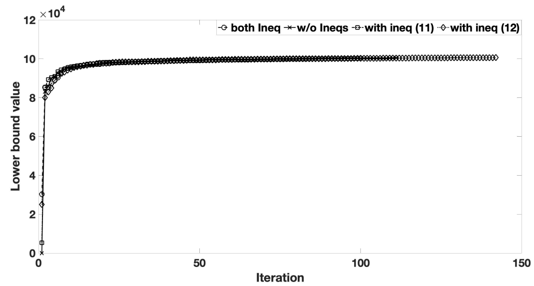
Inst	**Model**	$C = 150$			$C = 100$			$C \in [20, 50]$		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
1 (10, 5)	DR	0.2	0.2	0.3	0.1	0.1	0.1	0.1	0.1	0.1
	SP	3.6	4.5	6.8	4	6	7	1.0	1.0	1
2 (10, 10)	DR	0.3	0.4	0.6	0.4	0.5	6	0.1	0.1	0.1
	SP	8	8	9	17	32	51	2	3	4
3 (20,10)	DR	0.2	0.2	0.3	0.1	0.1	0.1	0.1	0.1	0.1
	SP	89	92	94	5	15	31	4	4	5
4 (20, 20)	DR	2.0	37	96	1.0	1.6	2.1	0.1	0.1	0.2
	SP	657	763	974	208	573	1079	11	14	24
5 (30,15)	DR	0.3	0.7	0.9	0.1	0.1	0.2	0.1	0.1	0.1
	SP	60	75	126	13	41	77	9	10	11
6 (30, 30)	DR	0.3	0.6	0.9	1.8	2.8	5.2	0.1	0.1	0.2
	SP	215	273	427	97	172	285	34.0	45.0	51
7 (40, 20)	DR	0.6	1.6	2.1	0.1	0.1	15	0.1	0.1	0.10
	SP	185.9	291	417	69	171	259	19	22	24
8 (40, 40)	DR	4.2	56	138	1.9	4.3	6.4	0.2	0.2	0.2
	SP	2164	2467	2690	291	499	783	77	97	136
9 (50, 25)	DR	1.0	2.3	5.5	0.1	0.1	0.1	0.1	0.1	0.1
	SP	357.2	481	721	70	331	781	32	37	41
10 (50, 50)	DR	3.7	8.9	16.7	17	104	187	0.1	0.1	0.1
	SP	1183	1183	1889	464	1110	2787	124	145	220
11 (100, 50)*	DR	0.4	0.5	0.6	0.12	0.13	0.14	0.1	0.1	0.07
	SP	211	319	673	242	874	1931	206	230	255
12 (100, 100)*	DR	32	65	114	39.0	63.0	86.0	0.1	0.2	0.2
	SP	-	-	-	-	-	-	-	-	-



(a) Instance 8, $C = 50$



(b) Instance 8, $C = 100$



(c) Instance 8, $C = 150$

Figure 1: Comparison of the lower bound values with and without inequalities (11)–(12).

Table 4: Computational details of solving instance 8 using DRFL–decomposition algorithm with and without inequalities (11)–(12). Notation: Iter is the number of iterations of DRFL–decomposition algorithm, Time is the solution time of each instance, Nodes is the total number of B&B nodes explored, MIPiter is the total number of MIP simplex iterations, and InGap is the initial relative gap between the lower and upper bound values.

	$C = 100$				
	Iter	**Time**	**Nodes**	**MIPiter**	**InGap**
both ineqs	34	32	1583790	3951209	84%
w/o ineqs (11)-(12)	43	210	9508990	9510912	100%
with ineq (11)	11	44	3750610	8960956	97%
with ineq (12)	12	35	3189100	6698171	86%
	$C = 150$				
	Iter	**Time**	**Nodes**	**MIPiter**	**InGap**
both ineqs	84	55	1174810	4466193	88%
w/o ineqs	111	150	2939650	11738402	100%
with ineq (11)	100	113	2459690	9882928	97%
with ineq (12)	142	115	3957100	16632452	91%

Third, we observe that the tightness of the initial lower bound value at the start of the solution process determine the computational performance of the algorithm. Table 4 presents the computational details of solving instance 8 with and without (11)-(12) under $C = 100$ and $C = 150$ (results under $C = 50$ are comparable). From these results we first observe that the number of Algorithm 1 iterations before it converges to the optimum, the solution time, the number of branch-and-bound (B&B) nodes explored, and the number of MIP simplex iterations are significantly larger without inequalities (11)-(12), especially under $C = 150$ and $C = 100$ (which are harder to solve as discussed in Section 5.2). For example, when $C = 100$, the average solution time (in seconds) of instance 8, increase from 32 to 210 (w/o inequalities), 44 (with only inequality (11)), and 35 (with only inequality 12). The number of B&B nodes and MIP iterations under $C = 100$ respectively increases from 1583790 and 3951209 to 9508990 and 9510912 without inequalities (11)-(12), indicating an increase in computational difficulty and justifying the increase in solution time.

These results give an example of how including inequalities (11)–12 improves the solvability of DRFL instances. Finally, we reemphasize that it is not easy to compare the efficiency and strength of inequities (11)–(12) for all possible instances of DRFL (see Remark 1).

5.4. ***Optimal Open Facility Locations***

In this section, we compare the optimal facility location decisions of the DR and SP models. For presentation brevity and illustrative purposes, we consider instances 3 (20, 10), 8 (40, 40), and 12 (100, 100) as examples of relatively small, medium, and large instances, respectively. Table 5 presents the optimal decisions of the SP and DR models under the default parameter settings.

From Table 5 we first observe that to mitigate the ambiguity of the demand, the DR models tend to open more facilities than the SP models. Consider instance 8, for example, the DR-bimodal and DR-plain models respectively open 15 and 10 facilities as compared to 9 and 8 facilities opened by the SP-bimodal and SP-plain models, respectively. DR-bimodal and SP-bimodal open more facilities than their plain counterparts.

Second, we observe that the DR-bimodal approach mitigates the bimodality of the demand by opening more facilities than the DR-plain approach. By opening more facilities, the DR-bimodal satisfies a larger amount of customer demand than the SP models and the DR-plain model (reflected by the zero unsatisfied demand in Tables 6–7 presented later in Section 5.5). Although this comes at a higher (one-time) fixed cost associated with opening facilities, as we show in the next section, it results in a lower total cost and better service quality (in terms of satisfying customers demand) in the long run.

5.5. ***Out-of-Sample Simulation Performance***

In this section, we compare the simulation performance of the optimal solutions of the DR and SP models. Considering that the results obtained for different instances consistently share some common features, for presentation brevity and illustrative purposes, we again present results with 3 (20, 10), 8 (40, 40), and 12 (100, 100) as examples of relatively small, medium, and large instances, respectively.

We evaluate the out-of-sample performance of the optimal DR and SP solutions to these instances (see Table 5) under both perfect (in-sample) and misspecified (out-of-sample) distributional information as follows. First, we fix the optimal first-stage decisions y in the SP model. Then,

Table 5: Optimal Facility Location of SP and DR models.

	3(20, 10)		8 (40, 40)		12 (100, 100)	
Model	# open	location	# open	location	# open	location
DR-bimodal	8	1,3,4,5,7,8,9,10	15	1,9,12,13,15,16, 19,20,22, 24,26, 37,38,39,40	35	1,2,3,4,5,6,8,10, 11,12,13,18,19,20, 21,22,23,26,27,28, 29,31,32,33,34,35, 36,40,41,43,45,46, 47,49,50
SP-bimodal	5	1,3,4,8,9	9	12,13,19,22,24, 26,37,38,40	22	1,2,3,4,6,12,17,18, 22,24,25,27,35,36, 37,38,41,42,43,44, 45,46
SP-plain	5	1,3,5,8,9	8	13,19,22,24,26, 37,38,40	19	1,3,4,6,12,17,18,22, 24,25,27,35,37,41, 42,44,45,46,47
DR-plain	5	1,3,4,8,9	10	9,12,13,15,19,22, 24,37,38,40	25	1,2,3,4,6,9,11,12,17, 18,22,24,25,27,32, 35,36,37,38,41,42, 43,44,45,46

we solve the second-stage recourse problem in (3) using the following two sets of $N' = 10000$ out-of-sample data $(q_1^n, d_1^{\mathrm{B,n}}, d_1^{\mathrm{A,n}}), \ldots, (q_J^n, d_J^{\mathrm{B,n}}, d_J^{\mathrm{A,n}})$, for all $n \in [N']$, to compute the corresponding second-stage and unmet demand costs.

1. *Perfect distributional info:* we use the same parameter settings and distributions (i.e., LogN) that we use for generating the N in-sample data points to generate the N' data points.
2. *Misspecified distributional info:* we use the same mean values (μ^{q}, μ^{B}, μ^{A}) and standard deviations (σ^{B}, σ^{A}) of random parameters (q, d^{B}, d^{A}), but we vary distribution type of (d^{B}, d^{A}) to generate the N' data points. Specifically, we follow Weibull distributions with ranges $[0, \overline{d}^{\mathrm{B}}]$ and $[0, \overline{d}^{\mathrm{A}}]$ to generate $(d_1^{\mathrm{B,n}}, d_1^{\mathrm{A,n}}), \ldots, (d_J^{\mathrm{B,n}}, d_J^{\mathrm{A,n}})$, for all $n \in [N']$. This is to simulate the out-of-sample performance of the DR and SP optimal solutions when the in-sample data is biased.

We use the same sets of N' in-sample and out-of-sample data to simulate the optimal solutions obtained from the DR and SP models. Using the first set of N' data points, we test the performance of the optimal DR and SP solutions under the case that the distributional belief we follow to optimize SP (LogN) well-represent the true distribution or if the true distribution is indeed LogN. Using the second set of N' data points, we test optimal DR and SP solutions' performance under the case that the true distribution is not LogN but is Weibull (another candidate distribution of demand with the generated mean and standard deviation), i.e., the SP model was optimized under misspecified data. Note that these choices are only for illustrative purposes as we do not have actual data on the demand.

Table 6 presents the means and quantiles of second-stage cost (2-stage), cost of unmet demand, and total cost yielded by optimal solutions of the DR and SP models in Table 5 under perfect (in-sample) distributional information. Clearly, by opening more facilities, the DR models satisfy a larger amount of customer demand than their SP counterparts. The DR-bimodal model has the highest fixed cost because it opens the largest number of facilities as compared to other models. However, DR-bimodal has the best performance with zero unmet demand and thus a significantly lower second-stage and total costs on average and at all quantiles than that of the SP-bimodal and the plain models. In fact, DR-bimodal has a significantly better performance at the 0.75- and 0.95- quantiles, especially for instance 8 and 12. By opening more facilities than the SP models, DR-plain satisfy a larger amount of demand and thus has a lower second-stage and total costs. The SP-plain model, which opens the least number of facilities and thus has the lowest fixed cost, yields the highest unmet demand and total costs. These results imply that when the distributional information is accurate, DR- bimodal still yields robust solutions that better hedge against uncertainty and provide a better service quality by satisfying customer demand.

Table 7 presents the out-of-sample performance of the DR and SP models under misspecified distributional information. From these results, we observe that DR-bimodal yields zero unmet demand and the lowest total cost as compared to DR-plain and SP. The cost reductions are reflected in all quantiles of the random second-stage cost and total cost. The DR-plain model has a lower total cost than the two SP models. Moreover, SP-bimodal satisfies a larger amount of demand and yields a lower second-stage and total costs than SP-plain.

These simulation results show how the DR-bimodal approach can produce facility location decisions that are robust (i.e., maintain a good performance under various probability distributions of demand). Satisfying customer demand is a desirable property in many, if not all, real-world applications.

Finally, it is worth mentioning that we chose the LogN distribution to generate the in-sample data, because it is typically used to model demand (and other random parameters). For the out-of-sample data, we use Weibull distribution, because it is another candidate distribution for modeling demand uncertainty. It is well-known that LogN and Weibull models can be used quite effectively to analyze skewed data set. However, as pointed out by Kundu and Manglick (2004) although these two models may provide similar data fit for moderate sample sizes, it is still crucial to select a model that provides an accurate approximation of the true unknown distribution. Choosing the best model among various candidate models to represent random data is, however, not an easy task. Our results confirm these observations and show that the DR approach (which is a distribution-free approach) maintains robust performance under both distributions than the SP approach.

5.6. ***Sensitivity Analysis***

In this section, we study the sensitivity of DR-bimodal and SP-bimodal to different parameter settings. For illustrative purposes, we consider instance 8 (40, 40) for this experiment. For each experiment, we obtain the DR and SP optimal solutions and then simulate their performance under a sample of 10000 scenarios of the demand generated from LogN distribution.

Table 6: Out-of-sample performance of optimal locations given by DR and SP models under perfect distributional information (LogN).

Inst	Model	Metric	Fixed	2-stage	unmet	TC
3	DR- bimodal	mean	26815	7416	0	34231
	SP-bimodal		14080	23912	8699	37992
	SP-plain		14080	23917	8846	37997
	DR-plain		14080	23912	8699	37992
	DR- bimodal	0.75–quantile	26815	9348	0	36163
	SP-bimodal		14080	29117	12600	43197
	SP-plain		14080	28717	12300	42797
	DR-plain		14080	29117	12600	43197
	DR- bimodal	0.95–quantile	26815	10353	0	37168
	SP-bimodal		14080	59773	41700	73853
	SP-plain		14080	61105	43035	75185
	DR-plain		14080	59773	41700	73853
8	DR- bimodal	mean	37975	12643	0	50618
	SP-bimodal		22095	44827	26168	66922
	SP-plain		19432	66287	50100	85719
	DR-plain		28282	25372	4539	53654
	DR- bimodal	0.75–quantile	37975	13353	0	51328
	SP-bimodal		22095	62558	42900	84653
	SP-plain		19432	92138	75000	111570
	DR-plain		28282	25291	4539	53573
	DR- bimodal	0.95–quantile	37975	14292	0	52267
	SP-bimodal		22095	100849	79770	122944
	SP-plain		19432	131562	113400	150994
	DR-plain		28282	57130	34500	85412
12	DR- bimodal	mean	98817	20437	0	119254
	SP-bimodal		58089	149360	120957	207449
	SP-plain		49253	262349	239700	311602
	DR-plain		77368	44519	11861	121887
	DR- bimodal	0.75–quantile	98817	21108	0	119925
	SP-bimodal		58089	185324	155700	243413
	SP-plain		49253	300582	277200	349835
	DR-plain		77368	91453	53700	168821
	DR- bimodal	0.95–quantile	98817	22123	0	120940
	SP-bimodal		58089	245994	213960	304083
	SP-plain		49253	355743	331200	404996
	DR-plain		77368	91453	53700	168821

Impact of event likelihood, μ^{q}

First, we analyze the impact of μ^{q} on the optimal number of open facilities, average total cost, average cost of unmet demand, and average transportation cost. We fix $f = 5000$ and vary $\mu^{\mathrm{q}} \in \{0.3, 0.5, 0.8, 1\}$. Note that $\mu^{\mathrm{q}} = 0.3$ and 0.8 indicate that the event occurs with a probability

Table 7: Out-of-sample performance of optimal locations given by DR and SP models under misspecified distributional information (Weibull).

Inst	Model	Metric	Fixed	2-stage	unmet	TC
3	DR- bimodal	mean	26815	8806	0	35621
	SP-bimodal		14080	24054	8700	38134
	SP-plain		14080	26022	10800	40102
	DR-plain		14080	24054	8700	38134
	DR- bimodal	0.75–quantile	26815	9493	0	36308
	SP-bimodal		14080	29431	12900	43511
	SP-plain		14080	32912	16500	46992
	DR-plain		14080	29431	12900	43511
	DR- bimodal	0.95–quantile	26815	10427	0	37242
	SP-bimodal		14080	58501	40500	72581
	SP-plain		14080	63624	45600	77704
	DR-plain		14080	58501	40500	72581
8	DR- bimodal	mean	37975	11697	0	49672
	SP-bimodal		22095	44759	26100	66854
	SP-plain		19432	74340	58651	93772
	DR-plain		28282	29723	7854	58005
	DR- bimodal	0.75–quantile	37975	12224	0	50199
	SP-bimodal		22095	62558	42900	84653
	SP-plain		19432	95649	79200	115081
	DR-plain		28282	31150	7800	59432
	DR- bimodal	0.95–quantile	37975	12950	0	50925
	SP-bimodal		22095	100879	79800	122974
	SP-plain		19432	132329	114735	151761
	DR-plain		28282	69371	44235	97653
12	DR- bimodal	mean	98817	19084	0	117901
	SP-bimodal		58089	148006	120600	206095
	SP-plain		49253	277469	255600	326722
	DR-plain		77368	51341	16800	128709
	DR- bimodal	0.75–quantile	98817	19650	0	118467
	SP-bimodal		58089	184121	155700	242210
	SP-plain		49253	313893	291300	363146
	DR-plain		77368	61319	25200	138687
	DR- bimodal	0.95–quantile	98817	20406	0	119223
	SP-bimodal		58089	244354	213900	302443
	SP-plain		49253	367534	343800	416787
	DR-plain		77368	122469	84000	199837

of (1-μ^{q})= 70% and 20%, respectively. We keep all other parameter settings as described in Section 5.1. Figures 2a–2d compare the results under different μ^{q}. It is apparent from Figure 2a that DR tends to open more facilities to mitigate the ambiguity and bimodality of the demand as the event is more likely to occur (i.e., smaller values of μ^q). Figure 2b indicates that the average total costs of the optimal DR and SP solutions are approximately equal when μ^q approaches to the lower and upper limits of the specified set. The SP solution results in a slightly better average total cost when $\mu^q = 0.3$ and 1. On the other hand, when $\mu^q = 0.5$, the average total cost reaches to the peak point for the optimal SP solution, which is significantly higher than the average total cost of the optimal DR solution. These results indicate that the average total cost is quite sensitive to μ^q for the SP approach. The SP approach shows poor performance compared to the DR approach, especially when the event's probability is similar to the probability that the event does not occur. Figure 2c and Figure 2d also show that the optimal DR solution provides a better service quality with 0 unmet demand and less or approximately the same transportation cost.

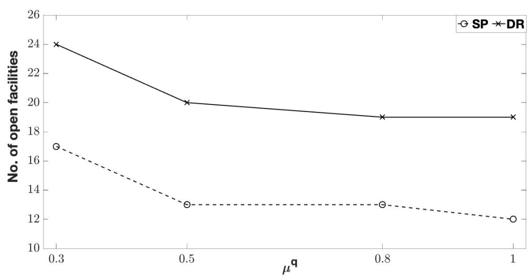
Impact of capacity

Second, we analyze the optimal number of open facilities as a function of the capacity parameter C. Specifically, we fix $f = 5000$ and vary $C \in \{50, 100, 150, 200, 250\}$. Figures 3a–3d compare the results under different C. From Figure 3a, we first observe that, irrespective of C, DR always opens more facilities than SP to better hedge against uncertainty and bimodality. When C increases from 50 to 250, both models open fewer facilities as each facility can satisfy higher demand. The total cost yielded by both models decreases as the capacity increases (see Figure 3b). Given that DR opens a higher number of facilities, it results in a higher fixed cost and a higher total cost than SP when the capacity is tight (because the former opens all 40 facilities) and a slightly higher total cost at a larger capacity. Nevertheless, DR yields a better service quality with substantially lower second-stage cost with 0 unmet demand and significantly less transportation cost than the SP model (see Figures 3c–3d). This could indicate that DR is more conservative in hedging against all possible scenarios of the demand (i.e., hedge better against demand's variability and bimodality) than SP.

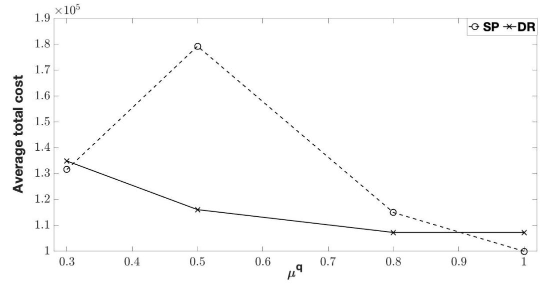
Using these results, practitioners should decide whether to pay the extra one-time fixed cost associated with adopting DR solutions to provide better service quality and thus maintain customer satisfaction or adopt the SP solution, which has a lower fixed cost but performs poorly in terms of satisfying customers demand and transportation cost. These results also show that if one can choose between tight and larger capacity, it may be better and more cost-effective to adopt the latter option.

Impact of variability in demand/demand ranges

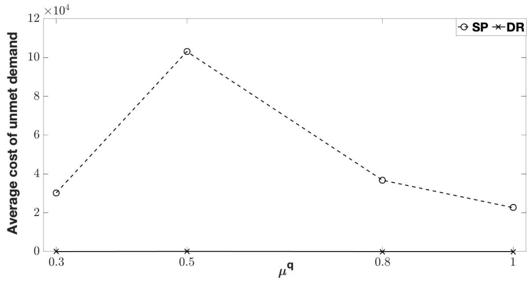
Finally, we analyze the sensitivity of the DR and SP solutions to the variability and volume of the demand range. In addition to the base range (Range 1: $\mu^{\text{B}} \in U[20, 40]$ and $\mu^{\text{A}} \in U[30, 60]$), we consider two additional ranges. In Range 2, we increase the variability (difference between the lower and upper bounds) of μ^{B} and μ^{A} to $\mu^{\text{B}} \in [10, 50]$ and $\mu^{\text{A}} \in [20, 70]$. In Range 3, we keep the difference between the upper and lower bounds of μ^{B} and μ^{A} as in Range 1 (20 and 30, respectively) and increase the demand volume (lower and upper bounds) to $\mu^{\text{B}} \in [30, 50]$ and $\mu^{\text{A}} \in [40, 70]$.



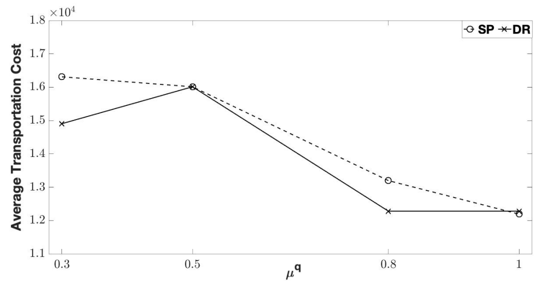
(a) No. of open facilities vs μ^q



(b) Average total cost vs μ^q

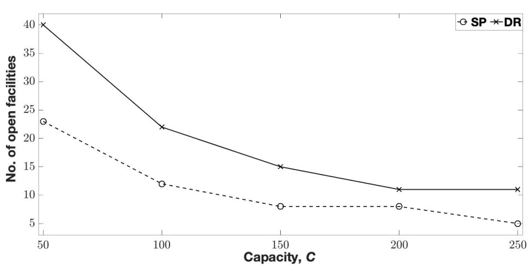


(c) Average cost of unmet demand vs μ^q

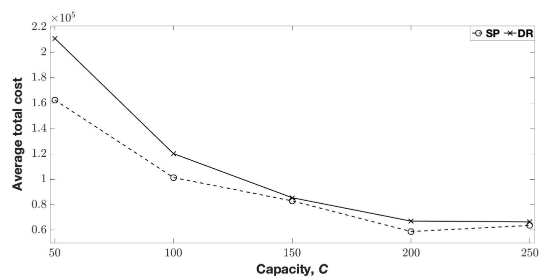


(d) Average transportation cost vs μ^q

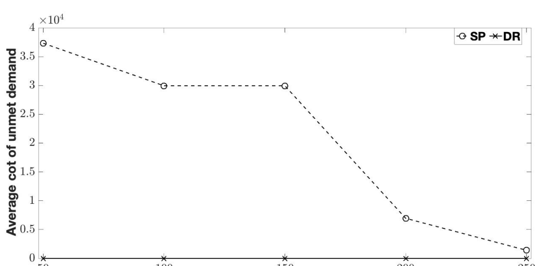
Figure 2: Comparison of the results under different values of μ^q



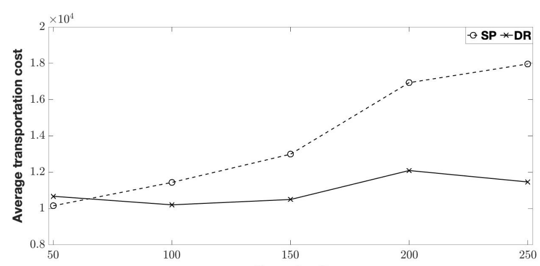
(a) No. of open facilities vs C



(b) Average total cost vs C

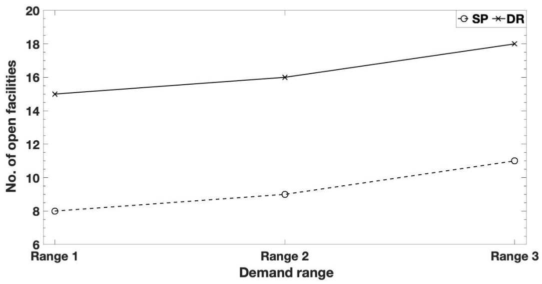


(c) Average cost of unmet demand vs C

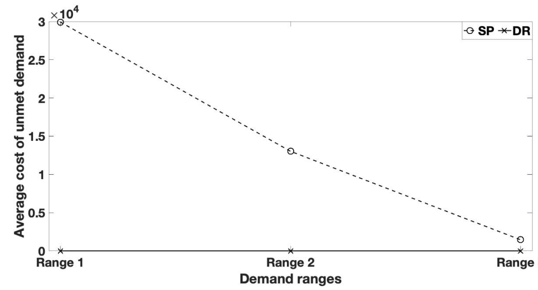


(d) Average transportation cost vs C

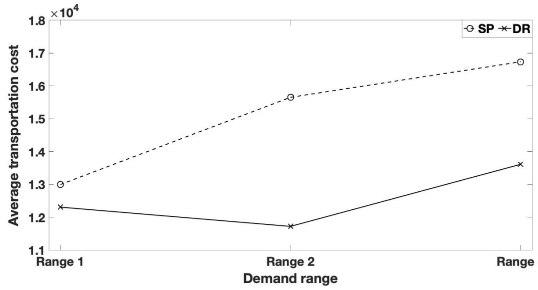
Figure 3: Comparison of the results under different values of capacity.



(a) No. of open facilities



(b) Average cost of unmet demand



(c) Average transportation cost

Figure 4: Effect of demand range

Figure 4 presents the optimal number of open facilities, average cost of unmet demand, and average transportation cost under Range 1–Range 3. It is quite apparent from Figure 4a that both models tend to open more facilities under a higher variability and volume of the demand. By opening more facilities than SP, DR mitigates the increase in the variability and volume of the demand better by maintaining zero unmet demand (see Figure 4b) and significantly lower transportation cost (Figure 4c). The average second-stage cost of the SP model under Range 1, Range 2, and Range 3 are respectively 2.75, 1.5, and 1.5 times higher than that of the DR model.

Our experiments in this section provide an example of how our computationally efficient DR approach can be used to generate robust facility location decisions under different parameter settings.

6. Conclusion

In this paper, we consider a facility location problem, recognizing the bimodality of random demand. That is, customer demand tends to display two spatially distinct probability distributions. We assume that these two distributions are ambiguous, and only their mean values and ranges are known. Therefore, we propose a *distributionally robust facility location* (DRFL) problem that seeks to find a subset of locations from a given set of candidate sites to open facilities to minimize the fixed cost of opening facilities, and worst-case expected costs of transportation and unmet demand over a family of distributions characterized through the known means and support of these distributions. We propose a decomposition-based algorithm to solve DRFL, which include valid lower bound inequalities in the master problem.

Using a set of DRFL instances of various sizes constructed based on prior studies, we conduct a series of numerical experiments to draw insights into DRFL. Specifically, we demonstrate that: (1) our DR-bimodal approach has a superior computational performance as compared to the SP approach, (2) DR-bimodal can produce facility location decisions that satisfy customer demand (providing a better quality of service) and maintain lower unmet demand and transportation costs than the SP-bimodal, SP-plain, and DR-plain models (which fail to satisfy customer demand and have higher transportation costs), under various probability distributions (and extreme scenarios) of the random parameters. Although we use the occurrence of a random event to illustrate our ideas, to present our model and derive useful insights, our approach is valid for other applications of DRFL in which the distribution of the demand is bimodal, i.e., tends to display two spatially distinct and unknown distributions.

We suggest the following areas for future research. First, due to the lack of data, our results are based on assumptions and parameter settings used in prior studies, and we assume that we know the capacity and the number of potential facility locations. We aim to extend our model to optimize the capacity and location of the facilities jointly. Second, we want to extend our approach by incorporating multi-modal probability distributions and higher moments of the demand in a data-driven DR approach. Third, it would be theoretically interesting to extend our approach by considering locating facilities in a country-wide setting (i.e., a larger network) and other non-classical settings such as mobile facility. This may require us to investigate efficient exact methods to solve larger instances of these problems under the general case of multi-modal distribution. Fourth, we aim to incorporate other sources of uncertainty (e.g., random capacity, product usability and lifetime, etc.) and objectives (e.g., holding cost).

Acknowledgments

We want to thank all of our colleagues who have contributed significantly to the related literature. We are grateful to the anonymous reviewers for their insightful suggestions and comments that allowed us to improve the paper. Dr. Karmel S. Shehadeh dedicates her effort in this paper to every little dreamer in the whole world who has a dream so big and so exciting. Believe in your dreams and do whatever it takes to achieve them–the best is yet to come for you.

Appendices

Appendix A. Chen et al. (2020) Scenario-wise Ambiguity Set

For completeness, in what follow we provide the details of the general scenario-wise ambiguity set proposed by Chen et al. (2020). To model multi-modality of random parameter ξ, Chen et al. (2020) assume that the probability distribution $\mathbb{P}$ of ξ is a mixture of R distinct distributions $\mathbb{P} = \sum_{r=1}^{R} p_r \mathbb{P}_r$ with $\sum_{r=1}^{S} p_r = 1$, where each mixture component $\mathbb{P}_r$ is an ambiguous distribution with support $\mathcal{U}_r$ and moments $\mathbb{E}_{\mathbb{P}_r}[\xi] \in \mathcal{L}_r$ and $\phi(\xi)$. As pointed out by Chen et al. (2020), the generalized moments characterized by convex function ϕ can provide useful statistical characterizations of the uncertainty ξ. Accordingly, Chen et al. (2020) consider the following mixture distribution ambiguity set.

$$\mathcal{F} := \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^{I_\xi + I_v} \times [R]) : \begin{array}{ll} ((\widetilde{\xi}, \widetilde{v}), \widetilde{s}) \in \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\widetilde{\xi} | \widetilde{r} = r] \in \mathcal{L}_r & \forall r \in [R] \\ \mathbb{E}_{\mathbb{P}}[\widetilde{v} | \widetilde{r} = r] \leq \boldsymbol{\sigma}_r & \forall s \in [S] \\ \mathbb{P}[(\widetilde{\xi}, \widetilde{v}) \in Z_r | \widetilde{r} = r] = 1 & \forall r \in [R] \\ \mathbb{P}[\widetilde{r} = r] = p_r & \forall r \in [R] \end{array} \right\} \tag{A.1}$$

where $\mathcal{P}(\mathbb{R}^{I_\xi + I_v} \times [R])$ is the set of all distributions on $\mathbb{R}^{I_\xi + I_v} \times [R]$. Primary and auxiliary random variables $\widetilde{\xi}$ and $\widetilde{v}$ jointly reside in the support set $\mathcal{Z}_r := \{(\widetilde{\xi}, \widetilde{v}) | \xi \in \mathcal{U}_r, v \geq \phi(\xi)\}$, for different scenarios $s \in [S]$. Note that any distribution $\mathbb{P}$ in $\mathcal{F}$ can be written as $\sum_{R=1}^{R} p_r \mathbb{P}_r$. In addition, $\mathcal{F}$ incorporates special cases of unimodal (i.e., $R = 1$). For different scenarios (r), the random variable ξ could be different, while conditioning on the scenario realization, the expectation and distributions of ξ can also be different.

In this paper, to motivate the need for considering multimodality of random parameters in real-world optimization problems and to derive a tractable reformulation of DRFL, we propose a special case of ambiguity set (A.1), with two distinct unknown distributions of the demand. We further restrict ourselves to the case in which we only know the mean values and range of the demand. That is, in our case, $R = 2$ and $\mathbb{P}_j = \sum_{r=1}^{2} p_r \mathbb{P}_r$, where $r = 1$ is the demand scenario before the occurrence of the event with distribution $\mathbb{P}_1 = \mathbb{P}^{\text{B}}$, $r = 2$ is the demand scenario after the event with distribution $\mathbb{P}_2 = \mathbb{P}^{\text{A}}$, $p_1 = q$, $p_2 = (1 - q)$, and $\mathcal{L}$ consists of the mean values of the demand under scenarios $r = 1$ and $r = 2$. These assumptions lead to our mean-range ambiguity set $\mathcal{F}(S, \mu)$ in (1).

Appendix B. Proof of Proposition 1

Proof. For a fixed y, we can formulate problem (4) as the following linear functional optimization problem.

$$\max_{\mathbb{P}\geq 0} \int_S Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}})\ d\mathbb{P} \tag{B.1a}$$

$$\text{s.t.} \int_S d_j^{\mathrm{B}}\ d\mathbb{P} = \mu_j^{\mathrm{B}} \qquad \forall j = 1, \ldots, J \tag{B.1b}$$

$$\int_S d_j^{\mathrm{A}}\ d\mathbb{P} = \mu_j^{\mathrm{A}} \qquad \forall j = 1, \ldots, J \tag{B.1c}$$

$$\int_S q_j\ d\mathbb{P} = \mu_j^{\mathrm{q}} \qquad \forall j = 1, \ldots, J \tag{B.1d}$$

$$\int_S d\mathbb{P} = 1 \tag{B.1e}$$

Letting $\rho = [\rho_1, \ldots, \rho_J]^T$, $\alpha = [\alpha_1, \ldots, \alpha_J]^\top$, $\lambda = [\lambda_1, \ldots, \lambda_J]^\top$, and θ be the dual variable associated with constraints (B.1b), (B.1c), (B.1d), and (B.1e), respectively, we present problem (B.1) in its dual form:

$$\min_{(\rho,\alpha,\lambda)\in\mathbb{R}^J, \theta\in\mathbb{R}} \sum_{j\in J} \mu_j^{\mathrm{B}}\rho_j + \mu_j^{\mathrm{A}}\alpha_j + \mu_j^{\mathrm{q}}\lambda_j + \theta \tag{B.2a}$$

$$\text{s.t.} \quad \sum_{j\in J}(d_j^{\mathrm{B}}\rho_j + d_j^{\mathrm{A}}\alpha_j + q_j\lambda_j) + \theta \geq Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) \qquad \forall (q, d^{\mathrm{B}}, d^{\mathrm{A}}) \in S \tag{B.2b}$$

where ρ, α, λ, and θ are unrestricted in sign, and constraint (B.2b) is associated with the primal variable $\mathbb{P}$. Under the standard assumptions that: (1) $\mu_j^{\mathrm{B}}(\mu_j^{\mathrm{A}})$ lies in the interior of the set $\{\int_S d_j^{\mathrm{B}}(d_j^{\mathrm{A}})\ d\mathbb{Q} : \mathbb{Q}$ is a probability distribution over $S\}$, and (2) μ_j^{q} lies in the interior of the set $\{\int_S q_j\ d\mathbb{Q} : \mathbb{Q}$ is a probability distribution over $S\}$ for each customer site j, strong duality hold between (B.1) and (B.2) (Bertsimas and Popescu (2005); Jiang et al. (2017); Mak et al. (2014); Shehadeh (2020)). Note that for fixed $(\rho, \alpha, \lambda, \theta)$, constraint (B.2b) is equivalent to

$$\theta \geq \max_{(q,d^{\mathrm{B}},d^{\mathrm{A}})\in S} \left\{Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) - \sum_{j\in J}(d_j^{\mathrm{B}}\rho_j + d_j^{\mathrm{A}}\alpha_j + q_j\lambda_j)\right\}$$

Since we are minimizing θ in (B.2), the dual formulation of (B.1) is equivalent to:

$$\min_{\rho,\alpha,\lambda} \left\{ \sum_{j\in J} \mu_j^{\mathrm{B}}\rho_j + \mu_j^{\mathrm{A}}\alpha_j + \mu_j^{\mathrm{q}}\lambda_j + \max_{(q,d^{\mathrm{B}},d^{\mathrm{A}})\in S} \left\{Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) + \sum_{j\in J} -(d_j^{\mathrm{B}}\rho_j + d_j^{\mathrm{A}}\alpha_j + q_j\lambda_j)\right\} \right\}$$

□

Appendix C. Proof of Proposition 2

Proof. For fixed $y \in \mathcal{Y}$ and $\xi = (q, d^{\mathrm{B}}, d^{\mathrm{A}})$, the dual of formulation (3) is as follow

$$Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) = \max_{\beta, v} \sum_{j \in J} (q_j d_j^{\mathrm{B}} + (1 - q_j) d_j^{\mathrm{A}}) \beta_j + \sum_{i \in I} C_i y_i v_i \tag{C.1a}$$

$$\text{s.t.} \quad \beta_j + v_i \leq t_{i,j}, \qquad \forall i \in [I], \forall j \in [J] \tag{C.1b}$$

$$\beta_j \leq p_j, \qquad \forall j \in [J] \tag{C.1c}$$

$$v_i \leq 0, \qquad \forall i \in [I] \tag{C.1d}$$

where $\beta = [\beta_1, \dots, \beta_J]^\top$ and $v = [v_1, \dots, v_I]^\top$ are the dual variables associated with constraints (3b) and (3c), respectively. Note that we can rewrite constraints (C.1b) as $v_i \leq \min_{j \in J} \{t_{i,j} - \beta_j\}, \forall i \in [I]$. Given that $v_i \leq 0$ and the objective of maximizing a positive number times β_j, then without loss of optimality, we can assume that $\beta_j \geq 0$ (note that if $\beta_j < 0$ for one j, then $v_i \leq \min_{j' \neq j} \{t_{i,j'} - \beta_{j'}\}$ and $v_i \leq t_{i,j} + |\beta_j| =$ positive number. Given that $v_i \leq 0$ then condition $v_i \leq t_{i,j} + |\beta_j|$ is redundant, i.e., $v_i \leq \min_{j' \neq j} \{t_{i,j'} - \beta_{j'}\}$, and the first term in the objective function will be negative for j). Next, we derive several useful algebraic expressions:

$$\max_{\substack{q_j \in \{0,1\}, \\ d_j^{\mathrm{B}} \in [\underline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{B}}] \\ d_j^{\mathrm{A}} \in [\underline{d}_j^{\mathrm{A}}, \overline{d}_j^{\mathrm{A}}]}} (q_j d_j^{\mathrm{B}} + (1 - q_j) d_j^{\mathrm{A}}) \beta_j = \left\{ \begin{array}{ll} \overline{d}_j^{\mathrm{B}} \beta_j & \textbf{if } q_j = 1 \\ \overline{d}_j^{\mathrm{A}} \beta_j & \textbf{if } q_j = 0 \end{array} \right\} \equiv \max\{\overline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{A}}\} \beta_j. \tag{C.2}$$

$$\max_{d_j^{\mathrm{B}} \in [\underline{d}_j^{\mathrm{B}}, \overline{d}_j^{\mathrm{B}}]} -d_j^{\mathrm{B}} \rho_j = \left\{ \begin{array}{ll} -\overline{d}_j^{\mathrm{B}} \rho_j & \textbf{if } \rho_j \leq 0 \\ -\underline{d}_j^{\mathrm{B}} \rho_j & \textbf{if } \rho_j > 0 \end{array} \right\} \equiv -\Big(\overline{d}_j^{\mathrm{B}} \rho_j + (\underline{d}_j^{\mathrm{B}} - \overline{d}_j^{\mathrm{B}}) (\rho_j)^+ \Big) \tag{C.3}$$

$$\max_{d_j^{\mathrm{A}} \in [\underline{d}_j^{\mathrm{A}}, \overline{d}_j^{\mathrm{A}}]} -d_j^{\mathrm{A}} \alpha_j = \left\{ \begin{array}{ll} -\overline{d}_j^{\mathrm{A}} \alpha_j & \textbf{if } \alpha_j \leq 0 \\ -\underline{d}_j^{\mathrm{A}} \alpha_j & \textbf{if } \alpha_j > 0 \end{array} \right\} \equiv -\Big(\overline{d}_j^{\mathrm{A}} \alpha_j + (\underline{d}_j^{\mathrm{A}} - \overline{d}_j^{\mathrm{A}}) (\alpha_j)^+ \Big) \tag{C.4}$$

$$\max_{q_j \in \{0,1\}} -q_j \lambda_j = \left\{ \begin{array}{ll} -\lambda_j & \textbf{if } \lambda_j \leq 0 \\ 0 & \textbf{if } \lambda_j > 0 \end{array} \right\} \equiv (-\lambda_j)^+ \tag{C.5}$$

Next, we observe that the feasible region $\Omega := \{(\text{C.1b}){-}(\text{C.1d})\}$ of $Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}})$ in (C.1) is bounded polyhedral and thus the optimal solution (β^*, v^*) to (C.1) is an extreme point of Ω. In addition, the support S is bounded. It follows that we can equivalently reformulate $\max_{(q, d^{\mathrm{B}}, d^{\mathrm{A}}) \in S} \Big\{ Q(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) + \sum_{j \in J} -(d_j^{\mathrm{B}} \rho_j + d_j^{\mathrm{A}} \alpha_j + q_j \lambda_j) \Big\}$ as

$$\max_{\substack{(\beta, v) \in \Omega \\ (q, d^{\mathrm{B}}, d^{\mathrm{A}}) \in S}} \Big[\sum_{j \in J} (q_j d_j^{\mathrm{B}} + (1 - q_j) d_j^{\mathrm{A}}) \beta_j + \sum_{i \in I} C_i y_i v_i + \sum_{j \in J} (-d_j^{\mathrm{B}} \rho_j - d_j^{\mathrm{A}} \alpha_j - q_j \lambda_j) \Big]$$

$$\equiv \max_{(\beta,v)\in\Omega} \sum_{j\in J} \max_{\substack{q_j\in\{0,1\},\\ d_j^{\mathrm{B}}\in[\underline{d}_j^{\mathrm{B}},\overline{d}_j^{\mathrm{B}}]\\ d_j^{\mathrm{A}}\in[\underline{d}_j^{\mathrm{A}},\overline{d}_j^{\mathrm{A}}]}} (q_j d_j^{\mathrm{B}} + (1-q_j)d_j^{\mathrm{A}})\beta_j + \sum_{i\in I} C_i y_i v_i$$

$$+ \sum_{j\in J} \max_{d_j^{\mathrm{B}}\in[\underline{d}_j^{\mathrm{B}},\overline{d}_j^{\mathrm{B}}]} -d_j^{\mathrm{B}}\rho_j + \max_{d_j^{\mathrm{A}}\in[\underline{d}_j^{\mathrm{A}},\overline{d}_j^{\mathrm{A}}]} -d_j^{\mathrm{A}}\alpha_j + \max_{q_j\in\{0,1\}} -q_j\lambda_j \tag{C.6}$$

Using (C.2)–(C.5), we can rewrite (C.6) as

$$\max_{(q,d^{\mathrm{B}},d^{\mathrm{A}})\in S} \Big\{Q(y,q,d^{\mathrm{B}},d^{\mathrm{A}}) - \sum_{j\in J}(d_j^{\mathrm{B}}\rho_j + d_j^{\mathrm{A}}\alpha_j + q_j\lambda_j)\Big\}$$

$$\equiv \max_{(\beta,v)\in\Omega} \Big\{\sum_{j\in J} \max\{\overline{d}_j^{\mathrm{B}},\overline{d}_j^{\mathrm{A}}\}\beta_j - \Big(\overline{d}_j^{\mathrm{B}}\rho_j + (\underline{d}_j^{\mathrm{B}} - \overline{d}_j^{\mathrm{B}})(\rho_j)^+\Big) - \Big(\overline{d}_j^{\mathrm{A}}\alpha_j + (\underline{d}_j^{\mathrm{A}} - \overline{d}_j^{\mathrm{A}})(\alpha_j)^+\Big) + (-\lambda_j)^+$$

$$+ \sum_{i\in I} C_i y_i v_i\Big\} \tag{C.7}$$

This complete the proof.

□

Appendix D. Proof of Proposition 3

Proof. First, note that feasible region $\Omega := \{(\text{C.1b}) - (\text{C.1d})\}$ is feasible and bounded because (1) $0 \le \beta_j \le p_j$, for all $j \in [J]$ (see the proof of Proposition 2, (2) given the objective of maximizing a positive number multiply by v_i, then v_i will either equal to 0 or a finite negative number by $v_i \le 0$ for all $i \in [I]$. In addition, Ω is independent of y. Therefore,

$$F(y) = \max_{(\beta,v)\in\Omega} \Big\{\sum_{j\in J} \max\{\overline{d}_j^{\mathrm{B}},\overline{d}_j^{\mathrm{A}}\}\beta_j + \sum_{i\in I} C_i y_i v_i\Big\} < \infty$$

.

Second, for any fixed β and v, $\sum_{j\in J} \max\{\overline{d}_j^{\mathrm{B}},\overline{d}_j^{\mathrm{A}}\}\beta_j + \sum_{i\in I} C_i y_i v_i$ is a linear function of y. It follows that $F(y)$ is a maximum of a set of linear functions of y and hence convex and piecewise linear. Third, it is clear that each linear piece of $F(y)$ is associated with one distinct extreme point of polyhedra Ω. Hence, $F(y)$ is finite because the bounded polyhedra Ω has a finite number of extreme points. This complete the Proof. □

Appendix E. Uncapacitated DRFL

In this section, we consider the uncapacitated version of DRFL, which we call UDRFL. The settings are the same as in Section 3.1, with only the exception that we relax capacity constraints (3c), i.e., we assume that facilities have no capacity limitations. We show that this relaxation facilitates an equivalent reformulation of UDRFL with bimodal demand as a mixed-integer linear program (MILP). Using ambiguity set $\mathcal{F}(\mathcal{S}, \mu)$, we formulate UDRFL as

$$\text{(UDRFL)} \quad \min_{y \in \mathcal{Y} \subseteq \{0,1\}^I} \left\{ \sum_{i \in I} f_i y_i + \sup_{\mathbb{P} \in \mathcal{F}(S,\mu)} \mathbb{E}_{\mathbb{P}}[\mathcal{G}(y, \xi)] \right\} \tag{E.1a}$$

where for a given $y \in \mathcal{Y}$ and a joint realization of uncertain parameters $\xi := [q, d^{\mathrm{B}}, d^{\mathrm{A}}]^\top$

$$\begin{aligned}
\mathcal{G}(y, \xi) := \min_{x,u} & \Big(\sum_{j \in J} \sum_{i \in I} t_{i,j} x_{i,j} + \sum_{j \in J} p_j u_j \Big) && && \text{(E.2a)}\\
\text{s.t.} \quad & \sum_{i \in I} x_{i,j} + u_j = q_j d_j^{\mathrm{B}} + (1 - q_j) d_j^{\mathrm{A}}, && \forall j \in [J] && \text{(E.2b)}\\
& x_{i,j} \leq M_j y_i, && \forall i \in [I],\ j \in [J] && \text{(E.2c)}\\
& u_j, x_{i,j} \geq 0, && \forall i \in [I],\ j \in [J] && \text{(E.2d)}
\end{aligned}$$

Constraints (E.2b) ensure that demand at each customer site is either satisfied by other locations or penalized, and constraints (E.2c) guarantee that customer demand can only be satisfied by open facilities. Recall that the maximum possible demand at customer site j is $\max\{\bar{d}_j^{\mathrm{B}}, \bar{d}_j^{\mathrm{A}}\}$, Therefore, w.l.o.o., we let $M_j = \max\{\bar{d}_j^{\mathrm{B}}, \bar{d}_j^{\mathrm{A}}\}$. Next, we follow the same analysis and reformulation techniques of the capacitated DRFL in Section 3.2 to derive an MILP reformulation of the min-max UDRFL model in (E.1a). First, given that we use the same ambiguity set in DRFL and UDRFL, then it is straightforward to apply the same techniques in the proof of Proposition 1 to verify that $\sup_{\mathbb{P} \in \mathcal{F}(S,\mu)} \mathbb{E}_{\mathbb{P}}[\mathcal{G}(y, \xi)]$ in (E.1a) is equivalent to

$$\min_{\rho, \alpha, \lambda} \left\{ \sum_{j \in J} \mu_j^{\mathrm{B}} \rho_j + \mu_j^{\mathrm{A}} \alpha_j + \mu_j^{\mathrm{q}} \lambda_j + \max_{(q, d^{\mathrm{B}}, d^{\mathrm{A}}) \in S} \Big\{ \mathcal{G}(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) + \sum_{j \in J} -(d_j^{\mathrm{B}} \rho_j + d_j^{\mathrm{A}} \alpha_j + q_j \lambda_j) \Big\} \right\} \tag{E.3}$$

Second, we observe that the recourse problem $\mathcal{G}(y, \xi)$ is decomposable by each customer site j, i.e., $\mathcal{G}(y, \xi) \equiv \sum_{j \in J} \mathcal{G}_j(y, \xi)$, where

$$\begin{aligned}
\mathcal{G}_j(y, \xi) := \min_{x,u} \ & \sum_{i \in I} t_{i,j} x_{i,j} + p_j u_j && \text{(E.4a)}\\
\text{s.t.} \quad & \sum_{i \in I} x_{i,j} + u_j = q_j d_j^{\mathrm{B}} + (1 - q_j) d_j^{\mathrm{A}}, && \text{(E.4b)}
\end{aligned}$$

$$x_{i,j} \leq M_j y_i, \qquad \forall i \in [I] \qquad \text{(E.4c)}$$

$$u_j, x_{i,j} \geq 0, \qquad \forall i \in [I] \qquad \text{(E.4d)}$$

Let Γ and ψ_i represent the dual variables associated with constraints (E.4b) and (E.4c), respectively. The dual of formulation (E.4) is as follows:

$$\mathcal{G}_j(y,\xi) \equiv \max_{\Gamma,\psi} \left\{ q_j d_j^{\mathrm{B}} + (1-q_j) d_j^{\mathrm{A}} \Gamma + \sum_{i \in I} M_j y_i \psi_i \right\} \qquad \text{(E.5a)}$$

$$\text{s.t.} \quad \Gamma + \psi_i \leq t_{i,j}, \qquad \forall i \in [I] \qquad \text{(E.5b)}$$

$$\Gamma \leq p_j \qquad \text{(E.5c)}$$

$$\psi_i \leq 0 \qquad \text{(E.5d)}$$

Note that problem (E.5) is a feasible and bound convex maximization problem. It follows from the fundamental convex analysis (see, e.g., Boyd and Vandenberghe (2004)) that there exists an optimal solution to (E.5) at one of its extreme points. Using the same proof techniques of Basciftci et al. (2019), we obtain the extreme points of (E.5) by counting the number of tight constraints in the following two cases.

- Case 1, $\Gamma = p_j$: in this case, $\psi_i = t_{i,j} - p_j$ or $\psi_i = 0$, for all $i \in [I]$. Given that we assume $p_j > t_{i,j}$, then by constraints (E.5b), we have $\psi_i \leq t_{i,j} - p_j < 0$. Thus, constraints (E.5c) is not binding (i.e., $\psi_i < 0$), and so $\psi_i = t_{i,j} - p_j$ in this extreme point. The associated value of the objective function with this extreme point equals $p_j(q_j d_j^{\mathrm{B}} + (1-q_j) d_j^{\mathrm{A}}) + \sum_{i \in I} M_j y_i (t_{i,j} - p_j)$.
- Case 2, $\Gamma < p_j$: in this case, $\psi_i = t_{i,j} - \Gamma$ or $\psi_i = 0$, for all $i \in [I]$. Note that we have $I + 1$ variables. Therefore, at least $I + 1$ constraints must be satisfied at each extreme point. Given that constraint (E.5c) is not binding in this case, then there exist at least one location i^* such that $\psi_{i^*} = t_{i^*,j} - \Gamma = 0$ and $\Gamma = t_{i^*,j}$ so that at least $I + 1$ constraints are satisfied. For all other locations $i \neq i^*$, $\psi_i = t_{i,j} - t_{i^*,j}$ or $\psi_i = 0$. Since we are maximizing a positive number multiplied by ψ_i, $\psi_i \leq t_{i,j} - t_{i^*,j}$, and $\psi_i \leq 0$, then: (1) $\psi_i = t_{i,j} - t_{i^*,j}$ if $t_{i,j} < t_{i^*,j}$, and (2)$\psi_i = 0$ if $t_{i,j} > t_{i^*,j}$. And so, for location i^*, the objective function is $t_{i^*,j}(q_j d_j^{\mathrm{B}} + (1-q_j) d_j^{\mathrm{A}}) + \sum_{i: t_{i,j} < t_{i^*,j}} M_j y_i (t_{i,j} - t_{i^*,j})$.

Combining the above two cases for each customer site j and summing over all customer sites, we obtain the following closed-from expression for the optimal objective value of the recourse problem.

$$\mathcal{G}(y, q, d^{\mathrm{B}}, d^{\mathrm{A}}) = \sum_{j \in J} \left[\max_{i^*=0,\dots,I} \left\{ t_{i^*,j} \left[q_j d_j^{\mathrm{B}} + (1-q_j) d_j^{\mathrm{A}} \right] + \sum_{i \in I: t_{i,j} < t_{i^*,j}} M_j y_i (t_{i,j} - t_{i^*,j}) \right\} \right] \qquad \text{(E.6)}$$

where $t_{0,j} = p_j, \ \forall k \in [J]$. Accordingly, we derive the following equivalent reformulation of the inner maximization problem in (E.3) as follows:

$$
\begin{aligned}
&\max_{(q,d^{\mathrm{B}},d^{\mathrm{A}})\in S} \Big\{\mathcal{G}(y,q,d^{\mathrm{B}},d^{\mathrm{A}}) + \sum_{j\in J} -(d_j^{\mathrm{B}}\rho_j + d_j^{\mathrm{A}}\alpha_j + q_j\lambda_j)\Big\} && \text{(E.7a)}\\
&\equiv \min \quad \theta && \text{(E.7b)}\\
&\quad \text{s.t.} \quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\overline{d}_j^{\mathrm{B}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \overline{d}_j^{\mathrm{B}}\rho_j - \lambda_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\underline{d}_j^{\mathrm{B}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \underline{d}_j^{\mathrm{B}}\rho_j - \lambda_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\overline{d}_j^{\mathrm{A}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \overline{d}_j^{\mathrm{A}}\alpha_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\underline{d}_j^{\mathrm{A}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \underline{d}_j^{\mathrm{A}}\alpha_j\Big], \forall i^* \in I\cup\{0\}
\end{aligned}
$$

Combing the inner optimization problem in the form of (E.7b) with the outer minimization in (E.3) and the UDRFL model in (E.1), we derive the following MILP reformulation of UDRFL.

$$
\begin{aligned}
&\min_{\substack{y\in\mathcal{Y}\rho,\alpha,\lambda\\ w,z,r,\delta}} \left\{ \sum_{i\in I} f_i y_i + \sum_{j\in J}\big(\mu_j^{\mathrm{B}}\rho_j + \mu_j^{\mathrm{A}}\alpha_j + \mu_j^{\mathrm{q}}\lambda_j\big) + \theta \right\} && \text{(E.8a)}\\
&\quad \text{s.t.} \quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\overline{d}_j^{\mathrm{B}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \overline{d}_j^{\mathrm{B}}\rho_j - \lambda_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\underline{d}_j^{\mathrm{B}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \underline{d}_j^{\mathrm{B}}\rho_j - \lambda_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\overline{d}_j^{\mathrm{A}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \overline{d}_j^{\mathrm{A}}\alpha_j\Big], \forall i^* \in I\cup\{0\}\\
&\qquad\quad \theta \geq \sum_{j\in J}\Big[t_{i^*,j}\underline{d}_j^{\mathrm{A}} + \sum_{i\in I: t_{i,j}<t_{i^*,j}} M_j y_i(t_{i,j}-t_{i^*,j}) - \underline{d}_j^{\mathrm{A}}\alpha_j\Big], \forall i^* \in I\cup\{0\}
\end{aligned}
$$

Appendix F. DR Plain

In this section, we present the DR-plain model in which we ignore the bimodality of the demand and assume that the demand, d, follow a single (unknown) probability distribution with a known mean μ^{d} and range $\mathcal{S}^{\mathrm{d}} := \{d \geq 0 : \underline{d}_j \leq d_j \leq \overline{d}_j\}$. Accordingly, we assume that $\mathbb{P}$ of d belongs to an ambiguity set $\mathcal{F}_2(\mathcal{S}, \mu)$ of possible distributions, which incorporate the known support $\mathcal{S}$ and mean $\mu := \mathbb{E}_{\mathbb{P}} = [\mu_1, \ldots, \mu_J]^\top$:

$$\mathcal{F}_2(S, \mu) := \left\{ \mathbb{P} \in \mathcal{P}(S) : \begin{array}{l} \int_S d\mathbb{P} = 1 \\ \mathbb{E}_{\mathbb{P}}[\xi] = \mu \end{array} \right\} \tag{F.1}$$

Using the ambiguity set $\mathcal{F}_2(S, \mu)$, we formulate the DR-plain model as

$$\min_{y \in \mathcal{Y} \subseteq \{0,1\}^I} \left\{ \sum_{i \in I} f_i y_i + \sup_{\mathbb{P} \in \mathcal{F}(S, \mu)} \mathbb{E}_{\mathbb{P}}[Q(y, \xi)] \right\} \tag{F.2a}$$

where for a given $y \in \mathcal{Y}$ and a realization of uncertain parameters $\xi := [d_1, \ldots, d_J]^\top$

$$\begin{aligned}
Q_2(y, \xi) := \min_{x,u} & \Big(\sum_{j \in J} \sum_{i \in I} t_{i,j} x_{i,j} + \sum_{j \in J} p_j u_j \Big) && && \text{(F.3a)} \\
\text{s.t.} \quad & \sum_{i \in I} x_{i,j} + u_j = d_j, && \forall j \in [J] && \text{(F.3b)} \\
& \sum_{j \in J} x_{i,j} \leq C_i y_i, && \forall i \in [I] && \text{(F.3c)} \\
& u_j, x_{i,j} \geq 0, && \forall i \in [I],\ j \in [J] && \text{(F.3d)}
\end{aligned}$$

We follow the same logic in Section 3.2 to derive the following equivalent reformulation of the min-max DR-plain model in (F.2)

$$\begin{aligned}
\min_{y \in \mathcal{Y}, \eta, g, \delta} & \Big\{ \sum_{i \in I} f_i y_i + \sum_{j \in J} \mu_j^{\mathrm{d}} \eta_j + \sum_{j \in J} -\big(\overline{d}_j \eta_j + (\underline{d}_j - \overline{d}_j) g_j\big) + \Delta \Big\} && \text{(F.4a)} \\
\text{s.t. } & g_j \geq \eta_j,\ g_j \geq 0 \forall j \in [J] && \text{(F.4b)} \\
& \Delta \geq F_2(y) = \max_{\psi, \phi} \{ \sum_{j \in J} \overline{d}_j \psi_j + \sum_{i \in I} C_i y_i \phi_i \} && \text{(F.4c)}
\end{aligned}$$

$$\begin{aligned}
\text{s.t. } & \psi_j + \phi_i \leq t_{i,j}, && \forall i \in [I], \forall j \in [J] \\
& \psi_j \leq p_j, && \forall j \in [J] \\
& \phi_i \leq 0, && \forall i \in [I]
\end{aligned}$$

It is easy to verify that $F_2(y) < \infty$ and further convex and piecewise linear in y with finite number of pieces. As such, we apply a decomposition algorithm similar to the one in Algorithm 1 to solve this model.

Appendix G. SP Formulation

$$\min_{y \in \mathcal{Y} \subseteq \{0,1\}^I x,u} \left\{ \sum_{i \in I} f_i y_i + \frac{1}{N} \sum_{n=1}^{N} \Big[\sum_{j \in J} \sum_{i \in I} t_{i,j} x_{i,j}^n + \sum_{j \in J} p_j u_j^n \Big] \right\} \tag{G.1a}$$

$$\text{s.t.} \quad \sum_{i \in I} x_{i,j}^n + u_j^n = q_j^n d_j^{\mathrm{B,n}} + (1 - q_j^n) d_j^{\mathrm{A,n}}, \qquad \forall j \in [J] \ \forall n \in [N] \tag{G.1b}$$

$$\sum_{j \in J} x_{i,j}^n \leq C_i y_i, \qquad \forall i \in [I], \ \forall n \in [N] \tag{G.1c}$$

$$u_j^n, x_{i,j}^n \geq 0, \qquad \forall i \in [I], \ j \in [J], \ \forall n \in [N] \tag{G.1d}$$

Ahmadi-Javid, A., Seyedi, P., Syam, S. S., 2017. A survey of healthcare facility location. Computers & Operations Research 79, 223–263.

Basciftci, B., Ahmed, S., Shen, S., 2019. Distributionally robust facility location problem under decision-dependent stochastic demand. arXiv preprint arXiv:1912.05577.

Ben-Tal, A., Den Hertog, D., Vial, J.-P., 2015. Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical programming 149 (1-2), 265–299.

Bertsimas, D., Popescu, I., 2005. Optimal inequalities in probability theory: A convex optimization approach. SIAM Journal on Optimization 15 (3), 780–804.

Bertsimas, D., Sim, M., 2004. The price of robustness. Operations research 52 (1), 35–53.

Birge, J. R., Louveaux, F., 2011. Introduction to stochastic programming. Springer Science & Business Media.

Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.

Burki, T., 2020. Global shortage of personal protective equipment. The Lancet Infectious Diseases 20 (7), 785–786.

Chan, Y., 2001. Location theory and decision analysis. Springer.

Chen, Z., Sim, M., Xiong, P., 2020. Robust stochastic optimization made easy with rsome. Management Science.

Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations research 58 (3), 595–612.

Hanasusanto, G. A., Kuhn, D., Wallace, S. W., Zymler, S., 2015. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. Mathematical Programming 152 (1-2), 1–32.

Hao, Z., He, L., Hu, Z., Jiang, J., 2019. Robust vehicle pre-allocation with uncertain covariates. Production and Operations Management.

Jiang, R., Shen, S., Zhang, Y., 2017. Integer programming approaches for appointment scheduling with random no-shows and service durations. Operations Research 65 (6), 1638–1656.

Kim, S., Pasupathy, R., Henderson, S. G., 2015. A guide to sample average approximation. In: Handbook of simulation optimization. Springer, pp. 207–243.

Kleywegt, A. J., Shapiro, A., Homem-de Mello, T., 2002. The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12 (2), 479–502.

Kundu, D., Manglick, A., 2004. Discriminating between the weibull and log-normal distributions. Naval Research Logistics (NRL) 51 (6), 893–905.

Lei, C., Lin, W.-H., Miao, L., 2014. A multicut l-shaped based algorithm to solve a stochastic programming model for the mobile facility routing and scheduling problem. European Journal of operational research 238 (3), 699–710.

Lei, C., Lin, W.-H., Miao, L., 2016. A two-stage robust optimization approach for the mobile facility fleet sizing and routing problem under uncertainty. Computers & Operations Research 67, 75–89.

Liu, K., Li, Q., Zhang, Z.-H., 2019. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. Transportation research part B: methodological 119, 79–101.

Luo, F., Mehrotra, S., 2018. Distributionally robust optimization with decision dependent ambiguity sets. arXiv preprint arXiv:1806.09215.

Mak, H.-Y., Rong, Y., Shen, Z.-J. M., 2013. Infrastructure planning for electric vehicles with battery swapping. Management Science 59 (7), 1557–1575.

Mak, H.-Y., Rong, Y., Zhang, J., 2014. Appointment scheduling with limited distributional information. Management Science 61 (2), 316–334.

Melo, M. T., Nickel, S., Saldanha-Da-Gama, F., 2009. Facility location and supply chain management–a review. European journal of operational research 196 (2), 401–412.

Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming 171 (1-2), 115–166.

Owen, S. H., Daskin, M. S., 1998. Strategic facility location: A review. European journal of operational research 111 (3), 423–447.

ReVelle, C. S., Eiselt, H. A., 2005. Location analysis: A synthesis and survey. European journal of operational research 165 (1), 1–19.

Saif, A., Delage, E., 2020. Data-driven distributionally robust capacitated facility location problem. European Journal of Operational Research.

Santiváñez, J. A., Carlo, H. J., 2018. Reliable capacitated facility location problem with service levels. EURO Journal on Transportation and Logistics 7 (4), 315–341.

Senaldi, E., 2018. Inventory management: connecting one arm to another.

Shapiro, A., Dentcheva, D., Ruszczyński, A., 2014. Lectures on stochastic programming: modeling and theory. SIAM.

Shehadeh, K. S., 2020. A distributionally robust optimization approach for a stochastic mobile facility routing and scheduling problem. arXiv preprint arXiv:2009.10894.

Shehadeh, K. S., Cohn, A. E., Jiang, R., 2020. A distributionally robust optimization approach for outpatient colonoscopy scheduling. European Journal of Operational Research 283 (2), 549–561.

Smith, J. E., Winkler, R. L., 2006. The optimizers curse: Skepticism and postdecision surprise in decision analysis. Management Science 52 (3), 311–322.

Snyder, L. V., 2006. Facility location under uncertainty: a review. IIE transactions 38 (7), 547–564.

Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations research 21 (5), 1154–1157.

Thiele, A., 2010. A note on issues of over-conservatism in robust optimization with cost uncertainty. Optimization 59 (7), 1033–1040.

Thiele, A., Terry, T., Epelman, M., 2009. Robust linear optimization with recourse. Rapport technique, 4–37.

Turkoglu, D. C., Genevois, M. E., 2019. A comparative survey of service facility location problems. Annals of Operations Research, 1–70.

Vaagen, H., Wallace, S. W., 2008. Product variety arising from hedging in the fashion supply chains. International Journal of Production Economics 114 (2), 431–455.

Wang, S., Chen, Z., Liu, T., 2020. Distributionally robust hub location. Transportation Science 54 (5), 1189–1210.

WHO, 2020. How WHO is re-imagining and fixing the links in the supply chains during COVID-19. https://www.who.int/news-room/feature-stories/detail/how-who-is-re-imagining-and-fixing-the-links-in-the-supply-chains-during-covid-19.

Wu, C., Du, D., Xu, D., 2015. An approximation algorithm for the two-stage distributionally robust facility location problem. In: Advances in Global Optimization. Springer, pp. 99–107.

Zhang, A., Kang, J. E., Kwon, C., 2017. Incorporating demand dynamics in multi-period capacitated fast-charging location planning for electric vehicles. Transportation Research Part B: Methodological 103, 5–29.