

---

# Distributionally Robust Bottleneck Combinatorial Problems: Uncertainty Quantification and Robust Decision Making

Weijun Xie · Jie Zhang · Shabbir Ahmed

February 21, 2021

**Abstract** In a bottleneck combinatorial problem, the objective is to minimize the highest cost of elements of a subset selected from the combinatorial solution space. This paper studies data-driven distributionally robust bottleneck combinatorial problems (DRBCP) with stochastic costs, where the probability distribution of the cost vector is contained in a ball of distributions centered at the empirical distribution specified by the Wasserstein distance. We study two distinct versions of DRBCP from different applications: (i) Motivated by the multi-hop wireless network application, we first study the uncertainty quantification of DRBCP (denoted by DRBCP-U), where decision-makers would like to have an accurate estimation of the worst-case value of DRBCP. The difficulty of DRBCP-U is to handle its max-min-max form. Fortunately, similar to the strong duality of linear programming, the alternative forms of the bottleneck combinatorial problems using clutters and blocking systems allow us to derive equivalent deterministic reformulations, which can be computed via mixed-integer programs. In addition, by drawing the connection between DRBCP-U and its sampling average approximation counterpart under empirical distribution, we show that the Wasserstein radius can be chosen in the order of negative square root of sample size, improving the existing known results; and (ii) Next, motivated by the ride-sharing application, decision-makers choose the best service-and-passenger matching that minimizes the unfairness. That is, we study the decision-making DRBCP, denoted by DRBCP-D. For DRBCP-D, we show that its optimal solution is also optimal to its sampling average approximation counterpart, and the Wasserstein radius can be chosen in a similar order as DRBCP-U. When the sample size is small, we propose to use the optimal value of DRBCP-D to construct an indifferent solution space and propose an alternative decision-robust model, which finds the best indifferent solution to minimize the empirical variance. We further show that the decision robust model can be recast as a mixed-integer conic program. Finally, we extend the proposed models and solution approaches to the distributionally robust  $L$ -sum bottleneck combinatorial problem (DR $L$ BCP), where decision-makers are interested in minimizing the worst-case sum of  $L$  highest costs of elements.

---

Weijun Xie  
Virginia Tech, Blacksburg, VA  
E-mail: wxie@vt.edu

Jie Zhang  
Virginia Tech, Blacksburg, VA  
E-mail: jiezhang@vt.edu

Shabbir Ahmed  
Georgia Institute of Technology, Atlanta, GA  
E-mail: sahmed@isye.gatech.edu

## 1 Introduction

### 1.1 Setting

In a combinatorial problem, there are  $n$  elements, indexed by  $[n] := \{1, \dots, n\}$ , and a collection of subsets  $X \subseteq 2^{[n]}$ , where each element  $j \in [n]$  bears a cost  $c_j$ . The objective of a bottleneck combinatorial problem (BCP) [28] is to seek the best subset  $x \in X$  that minimizes its highest cost of elements (i.e., bottleneck cost), defined as  $\min_{x \in X} \max_{j \in x} c_j$ , i.e., a BCP can be formulated as

$$\text{(BCP)} \quad Z(\mathbf{c}) = \min_{x \in X} \max_{j \in x} c_j. \quad (1)$$

The BCP (1) covers many interesting bottleneck problems, e.g., bottleneck shortest path [26], bottleneck spanning tree [9], bottleneck assignment [41], etc.

Oftentimes, the cost vector is random, denoted by  $\tilde{\mathbf{c}}$ , and its probability distribution  $\mathbb{P}$  is also difficult to characterize. Therefore, this paper considers the Distributionally Robust Combinatorial Bottleneck Problem (DRBCP). Let ambiguity set  $\mathcal{P}$  represent a distributional family. In this paper, we focus on two interesting variants of DRBCP: uncertainty quantification (DRBCP-U) and decision-making (DRBCP-D), i.e.,

$$\text{(DRBCP-U)} \quad v_U = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [Z(\tilde{\mathbf{c}})], \quad \text{and} \quad (2a)$$

$$\text{(DRBCP-D)} \quad v_D = \min_{x \in X} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[ \max_{j \in x} \tilde{c}_j \right]. \quad (2b)$$

For DRBCP-U (2a), the objective is to determine the value of the worst-case bottleneck cost, while in DRBCP-D (2b), it aims to find the best subset to minimize the worst-case bottleneck cost. Both DRBCP-U (2a) and DRBCP-D (2b) are motivated from real-world applications. For example, in a wireless sensor network (e.g., multi-hop network) [18, 31, 47, 48], there is usually a pair of transmitter (Tx) and receiver (Rx) and throughout the communication network, there is a minimum data rate requirement to be strictly enforced. Thus, it is important for the wireless network coordinator to determine the data rate between Tx and Rx, i.e., solve DRBCP-U (2a). Another example is that in the bottleneck matching problem [41] (e.g., online matching between ride-sharing service providers and passengers), since the travel time is often stochastic, it is important to find a fair matching between ride-sharing service providers and passengers which minimizes the longest waiting time of passengers, i.e., solve DRBCP-D (2b). Besides, DRBCP-D (2b) can be also applied to organ transplanting [34] to fairly distribute available organs to the patients, or to nurses' scheduling problems [32] to minimize the longest overtime workloads.

### 1.2 Wasserstein Ambiguity Set

Given an empirical distribution  $\mathbb{P}_{\tilde{\zeta}}$  constructed using i.i.d. historical data  $\mathcal{Z} = \{\tilde{\mathbf{c}}^k\}_{k \in [N]}$  such that  $\mathbb{P}_{\tilde{\zeta}}\{\tilde{\zeta} = \tilde{\mathbf{c}}^k\} = 1/N$ , this paper considers data-driven Wasserstein ambiguity set (see, e.g., [1, 6–8, 10, 11, 15, 17, 21, 29, 44]) as below:

$$\mathcal{P}_q = \left\{ \mathbb{P} : W_q \left( \mathbb{P}, \mathbb{P}_{\tilde{\zeta}} \right) \leq \theta \right\}, \quad (3)$$

where  $\theta \geq 0$  denotes the Wasserstein radius and for any  $q \in [1, \infty]$ , the Wasserstein distance  $W_q(\cdot, \cdot)$  is defined as

$$W_q(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \sqrt[q]{\int_{\Xi \times \Xi} \|\xi_1 - \xi_2\|^q \mathbb{Q}(d\xi_1, d\xi_2)} : \mathbb{Q} \text{ is a joint distribution of } \tilde{\xi}_1 \text{ and } \tilde{\xi}_2 \right. \\ \left. \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \right\}.$$

In particular, when  $q = \infty$ , it reduces to the  $\infty$ -Wasserstein distance

$$W_\infty(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ Q - \text{ess sup } \|\tilde{\xi}_1 - \tilde{\xi}_2\| : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \tilde{\xi}_1 \text{ and } \tilde{\xi}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \end{array} \right\}.$$

Above,  $Q$ -ess sup  $\|\cdot\|$  is the essential supreme of  $\|\cdot\|$  with respect to the joint distribution  $Q$ , which is formally defined as

$$Q - \text{ess sup } \|\tilde{\xi}_1 - \tilde{\xi}_2\| \triangleq \inf \left\{ \Delta : Q[\|\tilde{\xi}_1 - \tilde{\xi}_2\| > \Delta] = 0 \right\}.$$

It has been shown in [5] that its corresponding  $\infty$ -Wasserstein ambiguity set  $\mathcal{P}_\infty$  has the following equivalent form

$$\mathcal{P}_\infty = \left\{ \frac{1}{N} \sum_{k \in [N]} \delta(\tilde{c} - c^k) : \exists c^k, \|c^k - \tilde{c}^k\| \leq \theta \right\}, \quad (4)$$

where  $\delta(\cdot)$  is the Dirac delta function. This neat representation has a straightforward interpretation, i.e., the worst-case distribution is also supported by  $N$  points and each support point can only deviate at most  $\theta$  amount from one of the empirical data  $\mathcal{Z} = \{c^k\}_{k \in [N]}$ .

Since  $W_\infty(\cdot, \cdot)$  has the similar statistical performance as  $W_q(\cdot, \cdot)$  with  $q \in [1, \infty)$  [16, 43], we will focus on  $\infty$ -Wasserstein ambiguity set, i.e., throughout this paper, we make the following assumption:

**Assumption 1** Suppose that  $\mathcal{P} = \mathcal{P}_\infty$ .

The main results derived in this paper can be also extended to the other  $q$ -Wasserstein ambiguity set with  $q \in [1, \infty)$ , which is shown in the appendix. The advantages of using  $\infty$ -Wasserstein ambiguity set can be found in recent works (see, e.g., [5, 17, 45]). Particularly, this paper will show that for DRBCP-U (2a), the resulting formulation using  $\infty$ -Wasserstein ambiguity set can be decomposed into  $N$  separate mixed-integer programs, and thus can be computed very effectively, while for the other  $q$ -Wasserstein ambiguity set with  $q \in [1, \infty)$ , the decomposition is not straightforward and requires additional computational efforts.

### 1.3 Relevant Literature

Bottleneck combinatorial problems (BCPs) have been widely studied in literature [9, 26, 28, 40, 41]. Existing approaches mainly focused on BCPs with deterministic costs and developing polynomial-time algorithms. Recently, some development has been made to study BCPs with random costs, which are mainly focused on algorithms and approximations of some special variants of stochastic BCPs. We separate these works according to DRBCP-U (2a) and DRCBP-D (2b), respectively.

Uncertainty Quantification, i.e., DRBCP-U (2a). Existing works attempted to derive the asymptotic behavior of stochastic BCP (see, e.g., [36]), where the probability distribution of the random costs is given, and the number of elements  $n$  tends to be infinite. In [41], the author derived all of the asymptotic moments of the stochastic bottleneck assignment problem in which edge costs are i.i.d. from a continuous distribution. Similar results can be found in [36], where the author assumed the distribution is uniform. The work [3] investigated the asymptotic gap between the minimum pessimistic bottleneck cost (i.e.,  $\min_{x \in X} \max_{j \in x} \tilde{c}_j$ ) and the minimum optimistic bottleneck cost (i.e.,  $\min_{x \in X} \min_{j \in x} \tilde{c}_j$ ) under the probabilistic setting. All of these works assumed that the probability distribution is known, and the random cost of each element is i.i.d., which might not be realistic. In this paper, we do not assume known probability distribution, and instead, decision-makers can obtain empirical data to construct an ambiguity set.

Decision Making, i.e., DRBCP-D (2b). Many existing works focused on stochastic counterpart of DRBCP-D (2b), where a probability distribution is assumed to be known. For example, in [28], the authors proposed new criteria to generalize stochastic BCPs with discrete random costs. The complexity analysis and approximation results for several special cases were also provided. A similar result can be found in [27]. In [35], the authors developed algorithms to solve minimax regret bottleneck path problems with interval uncertainty. The work [42] studied risk-averse BCPs and reduced them to deterministic BCPs. In [39], the authors formulated the stochastic bottleneck spanning trees problem with independently distributed edge weights as a chance constrained program and provided efficient solution algorithms. All of the works mentioned above assumed the known probability distribution, while this paper studies DRBCP-D (2b) with unknown probability distributions. As far as we are concerned, only in [33], the authors studied a generalization of DRBCP-D to minimize the worst-case sum of  $\Gamma$  highest costs of elements with known marginal distributions, while different from [33], this paper studies both DRBCP-U and DRBCP-D under the less conservative Wasserstein ambiguity set.

As an effective way for the decision-making under uncertainty without fully knowing the probability distribution, distributionally robust optimization has attracted much attention [8, 10, 12, 17, 22, 23, 44, 46, 49]. Interested readers are referred to [37] for a comprehensive review. However, existing works mainly studied deterministic reformulations of distributionally robust optimization, where the objective function is in the form of either piece-wise convex, piece-wise concave or two-stage recourse function. This paper presents two unique aspects for distributionally robust optimization: (i) for DRBCP-U (2a), existing deterministic reformulation techniques are not directly applicable due to the complicated minimax operator and thus we develop new techniques, allowing us to represent DRBCP-U (2a) as mixed-integer program (MIP); and (ii) for DRBCP-D (2b), we observe that the sampling average approximation of a stochastic BCP is optimal, which, however, might not generate robust decisions. Thus, we present a better alternative for DRBCP-D (2b), which simultaneously guarantees the robustness of decision as well as the conservatism of objective value.

#### 1.4 Summary of Contributions

Since in general solving a BCP is NP-hard [24], this paper will focus on the deterministic reformulations of DRBCP-U (2a) and DRBCP-D (2b) under  $\infty$ -Wasserstein ambiguity set  $\mathcal{P}_\infty$  and derive their mixed integer-programming representations. Our main contributions are summarized as below:

- (i) By refining set  $X$  as an equivalent collection of mutually noncomparable subsets, we are able to derive an equivalent form of BCP (1). Using the equivalent representations of BCP (1) and  $\mathcal{P}_\infty$ , we show that DRBCP-U (2a) can be formulated as  $N$  separate MIPs.
- (ii) By drawing the connection between DRBCP-U (2a) and its sampling average approximation counterpart, we show that the Wasserstein radius can be chosen in the order of  $O(N^{-1/2})$ , i.e.,  $\theta = O(N^{-1/2})$ . This observation demonstrates that one does not need too many samples to ensure that with high probability, the optimal value of DRBCP-U (2a) provides an upper bound to the true stochastic BCP.
- (iii) We observe that the optimal decision of DRBCP-D (2b) is always equal to its sampling average approximation counterpart. Thus, we propose a new decision-robust model, which simultaneously guarantees the robustness of the objective value and the robustness of the decision.
- (iv) We show that the proposed results can be extended to the distributionally robust  $\Gamma$ -sum bottleneck combinatorial problem (DR $\Gamma$ BCP), whose objective is to minimize the worst-case sum of  $\Gamma$ -highest costs of elements.

**Notation.** The following notation is used throughout the paper. We use bold-letters (e.g.,  $\mathbf{x}$ ,  $\mathbf{A}$ ) to denote vectors and matrices and use corresponding non-bold letters to denote their components. Given a vector or matrix  $\mathbf{x}$ , its zero norm  $\|\mathbf{x}\|_0$  denotes the number of its nonzero elements. We let  $\mathbf{e}$  be the vector or matrix of all ones, and let  $\mathbf{e}_i$  be the  $i$ th standard basis vector. Given an integer  $n$ , we let  $[n] := \{1, 2, \dots, n\}$ , and use  $\mathbb{R}_+^n := \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$ . Given a real number  $t$ , we let  $(t)_+ := \max\{t, 0\}$ . Given a finite

set  $I$ , we let  $|I|$  denote its cardinality. We let  $\tilde{\xi}$  denote a random vector and denote its realizations by  $\xi$ . Given a probability distribution  $\mathbb{P}$  on  $\Xi$ , we use  $\mathbb{P}\{A\}$  to denote  $\mathbb{P}\{\tilde{\xi} : \text{condition } A(\tilde{\xi}) \text{ holds}\}$  when  $A(\tilde{\xi})$  is a condition on  $\tilde{\xi}$ , and to denote  $\mathbb{P}\{\tilde{\xi} : \tilde{\xi} \in A\}$  when  $A \subseteq \Xi$ . Given a set  $S$  and a positive integer  $\Gamma$ , we let  $\binom{S}{\Gamma}$  denote the collection of all the size- $\Gamma$  subsets of  $S$ . We use superscript  $k \in [N]$  denote the index of scenario  $k$ . We let  $\delta(\cdot)$  denote a Dirac delta function, roughly speaking,  $\delta(t) = \begin{cases} \infty, & \text{if } t = 0, \\ 0, & \text{otherwise.} \end{cases}$  Additional notation will be introduced as needed.

**Organization.** The remainder of the paper is organized as follows. In Section 2, we introduce the definition of clutters and blocking systems and an equivalent formulation of BCP. Sections 3 and 4 show the equivalent reformulations and properties of RBCP-U (2a) and DRBCP-D (2b) under  $\infty$ -Wasserstein ambiguity set  $\mathcal{P}_\infty$ . Section 5 shows the numerical illustration and Section 6 extends the results to distributionally robust  $\Gamma$ -sum bottleneck problems. Finally, Section 7 concludes this paper.

## 2 Clutter and Equivalent Dual Representation of BCP (1)

In this section, we will introduce important concepts and preliminary results useful for the main derivations in the following sections.

### 2.1 Clutter and Blocking System

Following from [14, 24, 38], we first define clutters and blocking systems as follows.

**Definition 1** (Clutter) A clutter  $\bar{X} \subseteq 2^{[n]}$  on the set  $[n]$  is a family of mutually noncomparable subsets of  $[n]$ , i.e.,

$$\bar{X} = \{h \subseteq [n] : \text{there is no subset } q \in \bar{X} \text{ such that } q \subsetneq h\}.$$

**Definition 2** (Blocking System) Given a clutter  $\bar{X}$  on  $[n]$ , suppose that the unique set  $F$  consists of the minimal subsets of  $[n]$  that have non-empty intersection with every member of  $\bar{X}$ , i.e.,

$$F = \{y \subseteq [n] : y \cap h \neq \emptyset, \forall h \in \bar{X} \text{ and } y \text{ is minimal with this property}\}.$$

Then the pair  $(\bar{X}, F)$  is called a *blocking system* on  $[n]$  and  $F = b(\bar{X})$  is called the *blocker* of  $\bar{X}$ .

According to [14], for any clutter  $\bar{X}$  on  $[n]$ , its blocker always exists and is unique.

The following examples from [14, 24] list some commonly-used blocking systems.

*Example 1* (Blocking System of the Cardinality Set) Consider a collection of size- $m$  subsets  $\bar{X} = \{x \subseteq [n] : |x| = m\}$ , then its blocker  $F = b(\bar{X}) = \{y \subseteq [n] : |y| = n - m + 1\}$ .

*Example 2* (Blocking System of  $s - t$  Paths) Suppose the clutter  $\bar{X}$  is a collection of all the simple paths from vertex  $s$  to vertex  $t$  in a graph  $G(V, E)$ , then its blocker  $F = b(\bar{X})$  is the collection of all the  $s - t$  cuts separating vertices  $s$  and  $t$  in the graph. Here, a  $s - t$  cut is defined as a collection of edges between a partition of nodes  $V$ , i.e., two subsets of nodes  $V_1, V_2$  such that  $V_1 \cap V_2 = \emptyset, V_1 \cup V_2 = V$ , and  $s \in V_1, t \in V_2$ .

*Example 3* (Blocking System of Spanning Trees) Suppose the clutter  $\bar{X}$  is a collection of all the spanning trees in an undirected graph  $G(V, E)$ , then its blocker  $F = b(\bar{X})$  is the collection of all the cuts in the graph. Here, a cut is defined as a collection of edges between a partition of nodes  $V$ , i.e., two subsets of nodes  $V_1, V_2$  such that  $V_1 \cap V_2 = \emptyset, V_1 \cup V_2 = V$ .

*Example 4* (Blocking System of the Assignments) Suppose the clutter  $\bar{X}$  is a collection of all the possible  $[m] \times [m]$  assignments, then its blocker  $F = b(\bar{X}) = \{\text{All the } h \times k \text{ submatrices} : h, k \subseteq [m], |h| + |k| = m + 1\}$ .

## 2.2 Equivalent Representation of BCP (1)

To derive an equivalent representation of BCP (1), let us first define a set  $\bar{X} = X \setminus \{x \in X : \exists y \in X, y \subsetneq x\}$ , i.e.,  $\bar{X}$  is the minimal subset of  $X$  that each member of  $X$  is either a subset or equal to a member of  $\bar{X}$  but no member of  $\bar{X}$  is strictly contained in another member of  $\bar{X}$ . Clearly, we have

**Lemma 1** *Set  $\bar{X}$  is a clutter.*

*Proof:* For any  $x \in \bar{X}$ , there is no  $y \in \bar{X}$  such that  $y \subsetneq x$ . Thus,  $\bar{X}$  is a clutter.  $\square$

Further, in BCP (1), we can achieve the same optimal value if we replace  $X$  by  $\bar{X}$ .

**Lemma 2** *BCP (1) is equivalent to*

$$Z(\mathbf{c}) = \min_{h \in \bar{X}} \max_{i \in h} c_i. \quad (5)$$

*Proof:* Let us denote  $\hat{Z}(\mathbf{c}) = \min_{h \in \bar{X}} \max_{i \in h} c_i$ . Since  $\bar{X} \subseteq X$ , we must have  $\hat{Z}(\mathbf{c}) \geq Z(\mathbf{c})$ .

On the other hand, according to the definition of  $\bar{X}$ , for each  $x \in X$ , there must exist an  $h \in \bar{X}$  such that  $h \subseteq x$ , which implies that  $\max_{i \in h} c_i \leq \max_{i \in x} c_i$ . Thus, we must have  $\min_{h \in \bar{X}} \max_{i \in h} c_i \leq \min_{x \in X} \max_{i \in x} c_i$ , i.e.,  $\hat{Z}(\mathbf{c}) \leq Z(\mathbf{c})$ .  $\square$

Finally, we present the equivalent formulation of (5).

**Theorem 1** *Suppose  $F$  is the blocker of  $\bar{X}$ . Then BCP (1) is equivalent to*

$$Z(\mathbf{c}) = \max_{y \in F} \min_{j \in y} c_j. \quad (6)$$

*Proof:* According to Lemma 1,  $\bar{X}$  is a clutter and its blocker  $F$  exists and is unique. Then the conclusion follows from Lemma 2 and the theorem on page 300 in [14].  $\square$

The result in Theorem 1 is key to the reformulation of DRBCP-U (2a), i.e., this result allows us to reformulate max-min-max into max-max-min and to switch the outer maximum operators. It is worthy of noting that for a general set  $X$ , it can take  $O(2^n)$  to find set  $\bar{X}$  by enumeration, and it can take up to  $O(2^n)$  to find a member of its blocker  $F$  according to the definition; however, for special examples in the previous subsection, we have closed-form representations of blocking systems. Thus, we recommend readers to derive closed-form representations if possible.

## 3 DRBCP-U (2a): Equivalent Reformulation, MIP Representation, and Confidence Bounds

In this section, we will derive the equivalent reformulation of DRBCP-U (2a), which motivates us to develop a MIP representation and to study its connection to the sampling average approximation counterpart. Similar conclusions follow for DRBCP-U (2a) under  $q$ -Wasserstein ambiguity set, which can be found in Appendix A except the more computationally expensive reformulations due to the additional dual multiplier.

### 3.1 Equivalent Reformulation

Using the representation (4) of  $\infty$ -Wasserstein ambiguity set  $\mathcal{P}_\infty$ , we first derive a deterministic representation of DRBCP-U (2a).

**Theorem 2** *Suppose  $\|\cdot\| = \|\cdot\|_r$  with  $r \geq 1$ , then DRBCP-U (2a) is equivalent to*

$$v_U = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} t_*^k(y), \quad (7)$$

where  $F$  is defined in Theorem 1 and  $t_*^k(y)$  is defined as

$$t_*^k(y) := \max_{t^k} \left\{ t^k : \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \leq \theta^r \right\} \quad (8)$$

with set  $I^k(t, y) = \{j \in y : t - \bar{c}_j^k \geq 0\}$ .

*Proof:* We split the proof into the following two steps.

**Step 1. Equivalent Representation of DRBCP-U (2a).** According to the representation (4) and Theorem 1, we have

$$v_U = \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathcal{P}} [Z(\tilde{c})] = \frac{1}{N} \sum_{k \in [N]} \max_{\|\mathbf{c}^k - \tilde{\mathbf{c}}^k\|_r \leq \theta} \max_{y \in F} \min_{j \in y} c_j^k.$$

By swapping the two maximization operators and linearizing the inner minimization operator, we further have

$$v_U = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{\|\mathbf{c}^k - \tilde{\mathbf{c}}^k\|_r \leq \theta} \min_{j \in y} c_j^k = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{\|\mathbf{c}^k - \tilde{\mathbf{c}}^k\|_r \leq \theta, t^k \leq c_j^k, \forall j \in y} t^k.$$

To simplify the inner maximization, we let  $\beta_j^k = c_j^k - t^k$  for all  $j \in y$  and define set  $I^k(t, y) = \{j \in y : t - \bar{c}_j^k \geq 0\}$ , then we have

$$\begin{aligned} v_U &= \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{\beta^k, t^k} \left\{ t^k : \sum_{j \in y} |t^k + \beta_j^k - \bar{c}_j^k|^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in y \right\} \\ &= \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k} \left\{ t^k : \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \leq \theta^r \right\} \end{aligned}$$

where the second equality is due to the optimality condition that at optimality, if  $j \in y \setminus I^k(t^k, y)$ , we must have  $\beta_j^k = -t^k + \bar{c}_j^k > 0$ ; otherwise, for  $j \in I^k(t^k, y)$ , given that  $t - \bar{c}_j^k \geq 0$  and  $\beta_j^k \geq 0$ , the optimal  $\beta_j^k = 0$ .

**Step 2. Obtaining Optimal  $t^k$ .** Next, suppose that  $y = \{y_1, \dots, y_{|y|}\}$  such that  $-\infty < \bar{c}_{y_1}^k \leq \dots \leq \bar{c}_{y_{|y|}}^k < \bar{c}_{y_{|y|+1}}^k := +\infty$ . Also, suppose that

$$t_*^k(y) \in \arg \max_{t^k} \left\{ t^k : \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \leq \theta^r \right\}.$$

Clearly,  $t_*^k(y)$  must satisfy

$$\bar{c}_{y_{|I^k(t_*^k(y), y)|}}^k \leq t_*^k(y) < \bar{c}_{y_{|I^k(t_*^k(y), y)|+1}}^k, \quad (9a)$$

$$\sum_{j \in I^k(t_*^k(y), y)} (\bar{c}_{y_{|I^k(t_*^k(y), y)|}}^k - \bar{c}_j^k)^r \leq \theta^r, \quad (9b)$$

$$\sum_{j \in I^k(t_*^k(y), y)} (\bar{c}_{y_{|I^k(t_*^k(y), y)|+1}}^k - \bar{c}_j^k)^r > \theta^r. \quad (9c)$$

Now, we are going to show that

*Claim* the optimal  $t_*^k(y)$  is unique.

*Proof:* To prove the uniqueness of  $t_*^k(y)$ , it is sufficient to show that the set  $I^k(t_*^k(y), y)$  is unique. We prove it by contradiction. Suppose that  $I^k(t_*^k(y), y)$  is not unique, then there exists two optimal solutions  $t_{1*}^k(y), t_{2*}^k(y)$  such that  $I^k(t_{1*}^k(y), y) \subsetneq I^k(t_{2*}^k(y), y)$ . Since  $t_{1*}^k(y)$  satisfies (9c), we have

$$\sum_{j \in I^k(t_{1*}^k(y), y)} (\bar{c}_{y|I^k(t_{1*}^k(y), y)|+1}^k - \bar{c}_j^k)^r > \theta^r.$$

Since  $t_{2*}^k(y)$  satisfies (9b) and  $I^k(t_{1*}^k(y), y) \subsetneq I^k(t_{2*}^k(y), y)$ , we have the following two inequalities

$$\sum_{j \in I^k(t_{2*}^k(y), y)} (\bar{c}_{y|I^k(t_{2*}^k(y), y)|}^k - \bar{c}_j^k)^r \leq \theta^r, \quad \bar{c}_{y|I^k(t_{1*}^k(y), y)|+1}^k \leq \bar{c}_{y|I^k(t_{2*}^k(y), y)|}^k,$$

which imply that

$$\theta^r < \sum_{j \in I^k(t_{1*}^k(y), y)} (\bar{c}_{y|I^k(t_{1*}^k(y), y)|+1}^k - \bar{c}_j^k)^r \leq \sum_{j \in I^k(t_{2*}^k(y), y)} (\bar{c}_{y|I^k(t_{2*}^k(y), y)|}^k - \bar{c}_j^k)^r \leq \theta^r,$$

a contradiction.  $\diamond$

$\square$

The result in Theorem 2 shows that DRBCP-U (2a) is equivalent to sum of  $N$  robust BCPs. Thus, to obtain the value  $v_U$  of DRBCP-U (2a), we can solve  $N$  separate robust BCPs, which can be done in parallel. Another observation is that if we choose  $L_1$  norm in Theorem 2, we can further simplify its result.

**Corollary 1** Suppose  $\|\cdot\| = \|\cdot\|_1$  and for each  $y \in F$ ,  $y = \{y_1, \dots, y_{|y|}\}$  such that  $-\infty < \bar{c}_{y_1}^k \leq \dots \leq \bar{c}_{y_{|y|}}^k < \bar{c}_{y_{|y|+1}}^k := +\infty$ . Then DRBCP-U (2a) is equivalent to

$$v_U = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \frac{(\sum_{j \in I^k(y)} \bar{c}_j^k) + \theta}{|I^k(y)|}, \quad (10)$$

where  $I^k(y) = \{y_1, \dots, y_{|I^k(y)|}\}$  and  $\bar{c}_{y_{|I^k(y)|}}^k \leq \frac{\sum_{j \in I^k(y)} \bar{c}_j^k + \theta}{|I^k(y)|} < \bar{c}_{y_{|I^k(y)|+1}}^k$ .

According to (10), the value of DRBCP-U is equal to the average of robust values from  $N$  scenarios, and within each scenario, we find the subset that maximizes the average of its several smallest elements instead of its bottleneck value. This result will indeed ensure the robustness of the DRBCP-U value.

Finally, we remark the characterization of the worst-case distribution of DRBCP-U (2a), i.e., the worst-case  $\{c^k\}_{k \in [N]}$  defined in (4).

**Corollary 2** For each  $k \in [N]$ , let  $y_*^k \in \arg \max_{y \in F} t_*^k(y)$  and

$$c_{j*}^k = \begin{cases} t_*^k(y_*^k), & \text{if } j \in I^k(t_*^k(y_*^k), y_*^k), \\ \bar{c}_j^k, & \text{otherwise.} \end{cases}$$

Then the worst-case probability distribution of DRBCP-U (2a) is  $\mathbb{P}^* = \frac{1}{N} \sum_{k \in [N]} \delta(\bar{c} - c_*^k)$ .

The result in Corollary 2 shows that the worst-case distribution has  $N$  support points, where each support point slightly deviates from one of the empirical data  $\mathcal{Z} = \{\bar{c}^k\}_{k \in [N]}$  to enforce the robustness of the DRBCP-U value.



### 3.2 Mixed-Integer Programming Representation

Another side product of Theorem 2 is that to compute the optimal value  $v_U$  of DRBCP-U (2a) is equivalent to the average of  $N$  separate MIPs.

**Proposition 1** *Suppose that the blocker  $F$  admits a binary programming representation  $\widehat{F} \subseteq \{0, 1\}^n$ . Then  $v_U = 1/N \sum_{k \in [N]} v_U^k$  and  $v_U^k$  is equivalent to*

$$v_U^k = \max_{z^k \in \widehat{F}, t^k, \beta^k} \left\{ t^k : \sum_{j \in [n]} |t^k - \bar{c}_j^k + \beta_j^k|^r \leq \theta^r, \beta_j^k \geq -M_j^k(1 - z_j^k), \forall j \in [n] \right\}, \quad (11)$$

where  $M_j^k = \theta + \min_{\tau \in [n]} \bar{c}_\tau^k - \bar{c}_j^k$ .

*Proof:* According to the proof of Theorem 2,  $v_U = 1/N \sum_{k \in [N]} v_U^k$ , where

$$v_U^k := \max_{y \in F, t^k, \beta^k} \left\{ t^k : \sum_{j \in y} |t^k - \bar{c}_j^k + \beta_j^k|^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in y \right\}.$$

Above, we can augment the vector  $\beta^k$  by defining free variable  $\beta_j^k$  for each  $j \in [n] \setminus y$  and rewrite it as

$$v_U^k := \max_{y \in F, t^k, \beta^k} \left\{ t^k : \sum_{j \in [n]} |t^k - \bar{c}_j^k + \beta_j^k|^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in y, \beta_j^k \in \mathbb{R}, \forall j \in [n] \setminus y \right\},$$

since at optimality, we must have  $\beta_j^k = -t^k + \bar{c}_j^k$  for each  $j \in [n] \setminus y$ .

As the blocker  $F$  admits a binary programming representation  $\widehat{F}$ , the value  $v_U^k$  is further equal to

$$v_U^k := \max_{z^k \in \widehat{F}, t^k, \beta^k} \left\{ t^k : \sum_{j \in [n]} |t^k - \bar{c}_j^k + \beta_j^k|^r \leq \theta^r, z_j^k \beta_j^k \geq 0, \forall j \in [n] \right\}. \quad (12)$$

To prove that (12) and (11) are equivalent, we only need to show that  $\beta_j^k \geq -M_j^k$  for each  $j \in [n]$ . Since at optimality,  $\beta_j^k$  is equal to 0 or  $-t^k + \bar{c}_j^k$ , it is sufficient to find an upper bound of  $t^k$ . According to (8), we have

$$\left( t^k - \min_{\tau \in [n]} \bar{c}_\tau^k \right)^r \leq \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \leq \theta^r,$$

which implies that  $t^k \leq \theta + \min_{\tau \in [n]} \bar{c}_\tau^k$ . Thus, we can choose  $M_j^k = \theta + \min_{\tau \in [n]} \bar{c}_\tau^k - \bar{c}_j^k$ .  $\square$

We remark that: (i) to obtain  $v_U$ , one needs to solve  $N$  MIPs (11), which can be done in parallel; and (ii) as long as  $r$  is a rational number, using the conic quadratic representation results in [4], the MIP (11) essentially can be formulated as a mixed-integer second order conic program (MISOCP).

The following example illustrates the MIP formulation proposed in Proposition 1.

*Example 5* Given a graph  $G(V, E)$ , suppose that  $X$  denotes all  $s - t$  paths, then according to Example 2, its blocker  $F$  denotes all the  $s - t$  cuts. To compute  $v_U$  of distributionally robust bottleneck shortest path, it is equal to  $1/N \sum_{k \in [N]} v_U^k$ , where

$$v_U^k = \max_{z^k, w^k, \beta^k, t^k} t^k \quad (13a)$$

$$\text{s.t. } \|t^k \mathbf{e} + \beta^k - \bar{\mathbf{c}}^k\|_r \leq \theta, \quad (13b)$$

$$\beta_{ij}^k \geq -M_{ij}^k(1 - z_{ij}^k), \forall (i, j) \in E, \quad (13c)$$

$$z_{ij}^k \leq 1 - w_i^k, w_j^k - z_{ij}^k \leq w_i^k, w_j^k - z_{ij}^k \geq 0, \forall (i, j) \in E, \quad (13d)$$

$$w_s^k = 0, w_t^k = 1, w_i^k \in \{0, 1\}, \forall i \in V, \quad (13e)$$

where  $M_{ij}^k = \theta + \min_{\tau \in E} \bar{c}_\tau^k - \bar{c}_{ij}^k$ .

### 3.3 Confidence Bounds

In this subsection, we compare the optimal value  $v_U$  of DRBCP-U (2a) with its sampling average approximation counterpart, which further motivates us on how to choose the Wasserstein radius  $\theta$ .

Let us define the sampling average approximation of stochastic BCP with empirical distribution  $\mathbb{P}_{\tilde{\zeta}}$  as below:

$$v_U^{SAA} = \frac{1}{N} \sum_{k \in [N]} \min_{x \in X} \max_{j \in x} \tilde{c}_j^k. \quad (14a)$$

Suppose that the true distribution is  $\mathbb{P}^T$ , then the stochastic BCP under the true distribution is defined as

$$v_U^T = \mathbb{E}_{\mathbb{P}^T} \left[ \min_{j \in X} \max_{j \in h} \tilde{c}_j \right]. \quad (14b)$$

Our goal is to determine a proper  $\theta$  such that  $v_U^T + 2\theta \geq v_U \geq v_U^T$  with a high probability. To begin with, we would like to show the relationship between  $v_U$  and  $v_U^{SAA}$ .

**Proposition 2** *Let  $v_U^{SAA}$  be defined in (14a). Then*

$$\frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}} \leq v_U - v_U^{SAA} \leq \theta.$$

*Proof:*

- (i) To prove  $v_U - v_U^{SAA} \leq \theta$ , according to Theorem 1 and Theorem 2, it is sufficient to show that for each  $y \in F$ , we must have

$$t_*^k(y) \leq \min_{j \in y} \tilde{c}_j^k + \theta.$$

In (8), we have

$$\left( t_*^k(y) - \min_{j \in y} \tilde{c}_j^k \right)^r \leq \sum_{j \in I^k(t_*^k(y), y)} (t_*^k(y) - \tilde{c}_j^k)^r \leq \theta^r,$$

which implies that  $t_*^k(y) \leq \min_{j \in y} \tilde{c}_j^k + \theta$ .

- (ii) To prove  $v_U - v_U^{SAA} \geq \frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}}$ , we first show that for each  $y \in F$ , we must have

$$t_*^k(y) \geq \min_{j \in y} \tilde{c}_j^k + \frac{\theta}{|y|^{\frac{1}{r}}}.$$

In (8), we first observe that at optimality, the inequality constraint must be equality:

$$\sum_{j \in I^k(t_*^k(y), y)} (t_*^k(y) - \tilde{c}_j^k)^r = \theta^r;$$

otherwise, we can further increase  $t_*^k(y)$ . Thus,

$$\theta^r = \sum_{j \in I^k(t_*^k(y), y)} (t_*^k(y) - \tilde{c}_j^k)^r \leq |I^k(t_*^k(y), y)| \left( t_*^k(y) - \min_{j \in y} \tilde{c}_j^k \right)^r \leq |y| \left( t_*^k(y) - \min_{j \in y} \tilde{c}_j^k \right)^r,$$

where the first inequality is due to  $\tilde{c}_j^k \geq \min_{j \in y} \tilde{c}_j^k$  for each  $j \in I^k(t_*^k(y), y)$  and the second inequality is because  $|I^k(t_*^k(y), y)| \leq |y|$ .

Next, according to Theorem 1 and Theorem 2, we have

$$v_U = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} t_*^k(y) \geq \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \left( \min_{j \in y} \tilde{c}_j^k + \frac{\theta}{|y|^{\frac{1}{r}}} \right)$$

$$\geq \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \min_{j \in y} c_j^k + \frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}} := v_U^{SAA} + \frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}},$$

where the last inequality is due to  $|y|^{\frac{1}{r}} \leq \max_{h \in F} |h|^{\frac{1}{r}}$  for all  $y \in F$ .

This completes the proof.  $\square$

Now we are ready to show the following main result of the relationship between the value of DRBCP-U  $v_U$  and true value  $v_U^T$ .

**Theorem 3** *Suppose that there exists a positive  $\sigma$  such that  $\mathbb{E}_{\mathbb{P}^T}[\exp((Z(\tilde{c}) - v_U^T)^2/\sigma^2)] \leq e$  a.s. Given  $\epsilon \in (0, 1)$  and  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon)\max_{y \in F}|y|^{\frac{1}{r}}} = O(N^{-\frac{1}{2}})$ , then we have*

(i)

$$\mathbb{P}^T \{v_U \geq v_U^T\} \geq 1 - \epsilon; \text{ and}$$

(ii)

$$\mathbb{P}^T \{v_U \leq v_U^T + 2\theta\} \geq 1 - \epsilon.$$

*Proof:* Since the proof of Part (ii) is very similar to that of Part (i), we will only prove Part (i). According to Proposition 2, it is sufficient to prove

$$\mathbb{P}^T \left\{ v_U^{SAA} \geq v_U^T - \frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}} \right\} \geq 1 - \epsilon.$$

Using Lemma 2 in [20] with  $d_k = v_U^T - Z(\tilde{c}^k)$  for each  $k \in [N]$ , we have

$$\mathbb{P}^T \left\{ v_U^{SAA} < v_U^T - \frac{\lambda\sigma}{\sqrt{N}} \right\} \leq e^{-\frac{\lambda^2}{3}},$$

i.e.,

$$\mathbb{P}^T \left\{ v_U^{SAA} \geq v_U^T - \frac{\lambda\sigma}{\sqrt{N}} \right\} \geq 1 - e^{-\frac{\lambda^2}{3}}.$$

Letting  $e^{-\frac{\lambda^2}{3}} := \epsilon$ , i.e.,  $\lambda = \sqrt{-3\log(\epsilon)}$  and  $\theta := \lambda\sigma N^{-\frac{1}{2}}\max_{y \in F}|y|^{\frac{1}{r}}$ , we arrive at the conclusion.  $\square$

We make the following remarks for Theorem 3.

- (a) If we focus on convergence of the optimal objective value instead of probability distribution, then the Wasserstein radius should be chosen in the order of  $O(N^{-\frac{1}{2}})$  rather than  $O(N^{-\frac{1}{n}})$  shown in [15, 16].
- (b) The asymptotic rate  $O(N^{-\frac{1}{2}})$  has been observed in [8] from a different angle. However, our rate is non-asymptotic.
- (c) The presumption of Theorem 3 follows from that of proposition 1 in [20]. One can also directly assume that the random vector  $\tilde{c}$  is sub-Gaussian. Together with the facts that  $|z(\tilde{c})| \leq \max_{j \in [n]} |\tilde{c}_j|$  and maximum of finite sub-Gaussian variables is still sub-Gaussian, the same conclusion follows.
- (d) Parts (i) and (ii) together show that with high probability, the value of DRBCP-U  $v_U$  is no less than the true value  $v_U^T$  and is at most  $2\theta$  larger than  $v_U^T$ . This implies that an appropriate  $\theta$  will ensure less conservatism of DRBCP-U value  $v_U$ .
- (e) In the numerical study, we can further improve the radius in Theorem 3 by selecting a list of  $\theta$  values and choosing the smallest  $\theta$  from this list such that  $v_U \geq v_U^T$  with high probability.

#### 4 DRBCP-D (2b): Equivalent Reformulation, and Alternative Decision-Robust Form

In this section, we derive an equivalent reformulation of DRBCP-D (2b) and propose an alternative decision-robust model to ensure the robustness of decision.

#### 4.1 Equivalent Reformulation and Confidence Bounds

In this subsection, we show that for DRBCP-D (2b), its sampling average approximation counterpart always provides an optimal solution, which motivates us a choice of Wasserstein radius  $\theta$ .

Let us first define the sampling average approximation of stochastic decision-making BCP as below:

$$v_D^{SAA} = \min_{x \in X} \frac{1}{N} \sum_{k \in [N]} \max_{j \in x} \tilde{c}_j^k. \quad (15a)$$

Suppose that the true distribution is  $\mathbb{P}^T$ , then the stochastic decision-making BCP under the true distribution is defined as

$$v_D^T = \min_{x \in X} \mathbb{E}_{\mathbb{P}^T} \left[ \max_{j \in x} \tilde{c}_j \right]. \quad (15b)$$

We first observe that DRBCP-D (2b) and model (15a) have the same optimal solution.

**Proposition 3** *Let  $v_D^{SAA}$  be the optimal value in (15a). Then*

$$v_D = v_D^{SAA} + \theta.$$

*Proof:* According to the representation (4), we have

$$v_D = \min_{x \in X} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}} \left[ \max_{j \in x} \tilde{c}_j \right] = \min_{x \in X} \frac{1}{N} \sum_{k \in [N]} \max_{\|c^k - \bar{c}^k\|_r \leq \theta} \max_{j \in x} c_j^k = \min_{x \in X} \frac{1}{N} \sum_{k \in [N]} \max_{j \in x} (\bar{c}_j^k + \theta) = v_D^{SAA} + \theta,$$

where the third equality is due to swapping two maximization operators and the definition of dual norm.  $\square$

This result is quite interesting and motivates us to seek an alternative model for DRBCP-D (2b). Besides, we make the following remarks.

- (i) The result in Proposition 3 holds for any  $q$ -Wasserstein ambiguity set with  $q \in [1, \infty)$ .
- (ii) This result shows that applying distributionally robustness into the objective function may indeed only ensure the robustness of objective function. In next subsection, we will propose an alternative decision-robust model for DRBCP-D (2b).
- (iii) One might want to strengthen the result in Proposition 3 by adding support of the random vector  $\tilde{c}$ . However, if the support is large enough, we can still obtain the same result.

Similar to Proposition 1, we show that  $v_D^{SAA}$  can be obtained by solving an MIP, and according to Proposition 3, so is  $v_D$ . The proof is based upon straightforward linearization and is thus omitted.

**Proposition 4** *Suppose that set  $X$  admits a binary programming representation  $\widehat{X} \subseteq \{0, 1\}^n$ . Then model (15a) is equivalent to the following MIP:*

$$\min_{z^k \in \widehat{X}, v_D^k, \forall k \in [N]} \left\{ \frac{1}{N} \sum_{k \in [N]} v_D^k : v_D^k \geq (\bar{c}_j^k - L^k) z_j^k + L^k, \forall j \in [n], k \in [N] \right\}, \quad (16)$$

where  $L^k := \min_{\tau \in [n]} \bar{c}_\tau^k$ .

Next, we will show the relationship between the value of DRBCP-D  $v_D$  and the true value  $v_D^T$  defined in (15b).

**Theorem 4** *Suppose that there exists a positive  $\sigma$  such that  $\mathbb{E}_{\mathbb{P}^T} [\exp((\max_{j \in x} \tilde{c}_j - \mathbb{E}_{\mathbb{P}^T} [\max_{j \in x} \tilde{c}_j])^2 / \sigma^2)] \leq e$  for each  $x \in X$ . Given  $\epsilon \in (0, 1)$ , then*

(i) let  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon) + 3n\log(2)} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_D \geq v_D^T\} \geq 1 - \epsilon;$$

(ii) let  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon) + 3n\log(2)} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_D \leq v_D^T + 2\theta\} \geq 1 - \epsilon.$$

*Proof:* Since the proof of Part (ii) is very similar to that of Part (i), we only prove Part (i). According to Proposition 2, it is sufficient to prove

$$\mathbb{P}^T \{v_D^{SAA} \geq v_D^T - \theta\} \geq 1 - \epsilon.$$

Using Lemma 2 in [20] with  $d_k = \mathbb{E}_{\mathbb{P}^T} [\max_{j \in x} \tilde{c}_j] - \max_{j \in x} \tilde{c}_j^k$  for each  $k \in [N]$ , we have

$$\mathbb{P}^T \left\{ \frac{1}{N} \sum_{k \in [N]} \max_{j \in x} \tilde{c}_j^k < \mathbb{E}_{\mathbb{P}^T} \left[ \max_{j \in x} \tilde{c}_j \right] - \frac{\lambda\sigma}{\sqrt{N}} \right\} \leq e^{-\frac{\lambda^2}{3}},$$

for each  $x \in X$ . Thus, applying the union bound, we have

$$\mathbb{P}^T \left\{ v_D^{SAA} \geq v_D^T - \frac{\lambda\sigma}{\sqrt{N}} \right\} \geq 1 - |X|e^{-\frac{\lambda^2}{3}} \geq 1 - 2^n e^{-\frac{\lambda^2}{3}},$$

where the last inequality is due to  $|X| \leq 2^n$ . Letting  $2^n e^{-\frac{\lambda^2}{3}} := \epsilon$ , i.e.,  $\lambda = \sqrt{-3\log(\epsilon) + 3n\log(2)}$  and  $\theta := \lambda\sigma N^{-\frac{1}{2}}$ , we arrive at the conclusion.  $\square$

In Theorem 4, we note that (i) by choosing  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon)}$ , we can use the results to estimate the theoretical  $(1 - 2\epsilon)$ -confidence interval of a given solution  $x$  as  $[1/N \sum_{k \in [N]} \max_{j \in x} \tilde{c}_j^k - \theta, 1/N \sum_{k \in [N]} \max_{j \in x} \tilde{c}_j^k + \theta]$ ; and (ii) if the sample size is large enough, then the decision of DRBCP-D (2b) will be close to that of the true problem (15b). This might not hold if the sample size is small (i.e.,  $N$  is small). Thus, to address this issue, in the next subsection, we propose an alternative decision-robust model.

Finally, we remark that the worst-case distribution of DRBCP-D (2b), i.e., the worst-case scenarios  $\{\mathbf{c}^k\}_{k \in [N]}$  are defined in the representation (4).

**Corollary 3** Given  $x \in X$ , for each  $k \in [N]$ , let  $\ell_*^k \in \arg \max_{j \in x} \tilde{c}_j^k$  and

$$\mathbf{c}_{j_*}^k = \begin{cases} \tilde{c}_j^k + \theta, & \text{if } j = \ell_*^k, \\ \tilde{c}_j^k, & \text{otherwise.} \end{cases}$$

Then the worst-case probability distribution of DRBCP-D (2b) is  $\mathbb{P}^* = \frac{1}{N} \sum_{k \in [N]} \delta(\tilde{\mathbf{c}} - \mathbf{c}_{j_*}^k)$ .

The result in Corollary 3 shows that the worst-case distribution has  $N$  support points, where each support point coincides with one of the empirical data  $\mathcal{Z} = \{\tilde{\mathbf{c}}^k\}_{k \in [N]}$  except one coordinate.

## 4.2 An Alternative Decision-Robust Model

Proposition 3 shows that the optimal solution of the model (15a) is always optimal to DRBCP-D (2b), which might not be ideal, especially when the sample size is small. Alternatively, we consider the following lower level set

$$\mathcal{L}_\theta(X) = \left\{ x \in X : \frac{1}{N} \sum_{k \in [N]} \max_{j \in x} \tilde{c}_j^k \leq v_D := v_D^{SAA} + \theta \right\}, \quad (17a)$$

which corresponds to the set of indifferent solutions to model (15a). Here, Wasserstein radius  $\theta$  provides an indifferent level and needs to be properly chosen (we will discuss this issue at the end of this section). Next, we choose the most robust subset  $x$  that minimizes its sampling variance within the lower level set  $\mathcal{L}_\theta(X)$ , i.e., we consider the following “decision-robust” model:

$$v_{DR} = \min_{x \in \mathcal{L}_\theta(X)} \mathbb{V}_{\mathbb{P}_\xi} \left[ \max_{j \in x} \tilde{\zeta}_j \right], \quad (17b)$$

where  $\mathbb{V}_{\mathbb{P}}[\cdot]$  denotes the variance function with probability distribution  $\mathbb{P}$ .

Our next result shows that the decision-robust model (17b) can be formulated as an MISOCP.

**Proposition 5** *Suppose that set  $X$  admits a binary programming representation  $\hat{X}$ . Then the decision-robust model (17b) is equivalent to the following MISOCP*

$$\begin{aligned} v_{DR} = \min_{z \in \hat{X}, v, \lambda} \quad & \frac{1}{N} \sum_{k \in [N]} \left( v^k - \frac{1}{N} \sum_{\tau \in [N]} v^\tau \right)^2 \\ \text{s.t.} \quad & \frac{1}{N} \sum_{k \in [N]} v^k \leq v_D^{SAA} + \theta, \\ & v^k \geq (\bar{c}_j^k - L^k) z_j + L^k, \forall j \in [n], k \in [N], \\ & v^k \leq \lambda_j^k \bar{c}_j^k + (1 - \lambda_j^k) M^k, \forall k \in [N], \\ & \sum_{j \in [n]} \lambda_j^k = 1, \forall k \in [N], \\ & \lambda_j^k \leq z_j, \forall j \in [n], \forall k \in [N], \\ & \lambda_j^k \in [0, 1], \forall j \in [n], k \in [N], \end{aligned} \quad (18)$$

where  $L^k := \min_{\tau \in [n]} \bar{c}_\tau^k$  and  $M^k := \max_{\tau \in [n]} \bar{c}_\tau^k$ .

*Proof:* Since set  $X$  admits a binary programming representation  $\hat{X}$ , according to the definition, the decision-robust model (17b) is equivalent to

$$v_{DR} = \min_{z \in \hat{X}} \left\{ \frac{1}{N} \sum_{k \in [N]} \left( \max_{j \in [n]} \bar{c}_j^k z_j - \frac{1}{N} \sum_{\tau \in [N]} \left[ \max_{j \in [n]} \bar{c}_j^\tau z_j \right] \right)^2 : \frac{1}{N} \sum_{k \in [N]} \max_{j \in [n]} \bar{c}_j^k z_j \leq v_D^{SAA} + \theta \right\}. \quad (19)$$

Next, for each  $k \in [N]$ , we let  $v^k = \max_{j \in [n]} \bar{c}_j^k z_j$  and linearize it by introducing additional variables  $\lambda_j^k \in [0, 1]$  for all  $j \in [n]$ . Thus, we arrive at (18).  $\square$

We make the following remarks about decision-robust model (17b).

- (i) To solve the decision-robust model (17b), one needs to solve model (15a) first and obtain its optimal value  $v_D^{SAA}$ .
- (ii) The decision-robust model (17b) is different from mean-variance optimization [2], where the objective of the latter is to jointly minimize the weighted mean and variance.
- (iii) The Wasserstein radius  $\theta$  here means the indifferent level. One can choose  $\theta$  using the results in Theorem 4. Alternatively, suppose that the optimal solution of model (15a) is  $x^{SAA}$ , and choose  $\theta := 1.645 \sqrt{N^{-1} \mathbb{V}_{\mathbb{P}_\xi} [\max_{j \in x^{SAA}} \tilde{\zeta}_j]}$ , which is approximately equal to the upper bound of 95% confidence level of  $v_D^{SAA}$ . In numerical study, we will use cross-validation method to choose a proper  $\theta$ .
- (iv) As long as  $\theta \rightarrow 0$  as  $N \rightarrow \infty$ , the optimal solution to decision-robust model (17b) will eventually converge to the optimal solution of true decision model (15b).

The following example shows how to use Proposition 5 to reformulate decision-robust model (17b) as an MISOCP.

*Example 6* Suppose that there are  $m$  ride-sharing service providers and  $m$  passengers waiting for a matching, and the travel time  $\tilde{c}_{ij}$  from driver  $i$  to passenger  $j$  is stochastic. Our objective is to find a decision-robust fair matching, which minimizes the largest travel time of the matching. Following Proposition 5, this decision-robust fair matching can be formulated as

$$\begin{aligned}
v_{DR} = \min_{\mathbf{v}, \mathbf{z}, \boldsymbol{\lambda}} & \frac{1}{N} \sum_{k \in [N]} \left( v^k - \frac{1}{N} \sum_{\tau \in [N]} v^\tau \right)^2 \\
\text{s.t.} & \frac{1}{N} \sum_{k \in [N]} v^k \leq v_D^{SAA} + \theta, \\
& v^k \geq (\bar{c}_{ij}^k - L^k) z_{ij} + L^k, \forall i, j \in [m], \forall k \in [N], \\
& v^k \leq \lambda_{ij}^k \bar{c}_{ij}^k + (1 - \lambda_{ij}^k) M^k, \forall i, j \in [m], \forall k \in [N], \\
& \sum_{i \in [m]} \sum_{j \in [m]} \lambda_{ij}^k = 1, \forall k \in [N], \\
& \sum_{i \in [m]} z_{ij} = 1, \forall j \in [m], \\
& \sum_{j \in [m]} z_{ij} = 1, \forall i \in [m], \\
& \lambda_{ij}^k \leq z_{ij}, \forall i, j \in [m], \forall k \in [N], \\
& \lambda_{ij}^k \in [0, 1], z_{ij} \in \{0, 1\} \forall i, j \in [m], \forall k \in [N],
\end{aligned} \tag{20a}$$

where  $L^k = \min_{i, j \in [m]} \bar{c}_{ij}^k$ ,  $M^k = \max_{i, j \in [m]} \bar{c}_{ij}^k$  and  $v_D^{SAA}$  is the optimal value of

$$v_D^{SAA} = \min_{\mathbf{v}, \mathbf{z}} \left\{ \frac{1}{N} \sum_{k \in [N]} v^k : v^k \geq \bar{c}_{ij}^k z_{ij}, \sum_{i \in [m]} z_{ij} = 1, \sum_{j \in [m]} z_{ij} = 1, z_{ij} \in \{0, 1\}, \forall i, j \in [m], \forall k \in [N] \right\}. \tag{20b}$$

## 5 Numerical Illustrations

In this section, we provide numerical illustrations for both DRBCP-U (2a) and DRBCP-D (2b). All the instances are coded in Python 3.0 with calls to Gurobi 7.5 on a personal computer with 2.3 GHz Intel Core i5 processor and 8G of memory.

### 5.1 Numerical Illustration of DRBCP-U (2a)

In this numerical study, we consider a wireless sensor network (i.e., multi-hop network), where the bandwidth has been normalized to 1 unit. Suppose that there is a channel (Tx-Rx) with transmitter (Tx) and receiver (Rx), which needs to meet a data rate (i.e., minimum capacity) requirement. Since data rate governs the speed of data transmission of the channel, in order to know how fast we can send data (in bits per second) over the channel, it is important to decide whether this requirement can be guaranteed. We also suppose each data packet is not separable, i.e., there is no flow split, and the link capacity is measured periodically and is broadcast across the network. The uncertainty of link capacity is mainly because of (i) receiver's signal to noise ratio (SNR), and (ii) the network delay. For each channel  $i \in [n]$ , we use Shannon capacity to calculate the theoretical highest data rate  $\tilde{c}_i$  following the work [19], i.e.,

$$\tilde{c}_i = B_i \log_2 (1 + \text{SNR}_i \times \xi_i),$$

where  $B_i$  is the bandwidth,  $\xi_i \sim \exp(1)$ ,  $\text{SNR}_i$  is the signal-to-noise ratio. Specifically,  $\text{SNR}_i$  is computed as

$$\text{SNR}_i = p_i \sigma_i^{-2} 10^{-12.81 - 3.76 \times \log_{10} d_{\text{macro}}},$$

where  $p_i$  is the transmission power,  $\sigma^2$  is the noise level,  $d_i$  is the distance from the transmitter to the receiver. In this numerical study, we suppose that  $B_i = 1\text{HZ}$ ,  $p_i \sim \text{uniform}(0.1, 0.2)W$ ,  $\sigma^2 = 1 \times 10^{-10}W$ ,  $d_{\text{macro}} \sim \text{uniform}(0.03, 0.07)$  km. In addition, we suppose that the network constitutes of 20 nodes and 190 edges, i.e., it is a 20-node complete undirected graph. As the objective of the multi-hop network is to find the smallest data rate, DRBCP-U (2a) becomes  $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [\max_{x \in X} \min_{j \in x} \tilde{c}_j]$  with  $X$  denoting all the  $s - t$  paths, where the results in this paper straightforwardly follow.

To demonstrate the effectiveness of Model (13), we generate three different instances with sample size  $N \in \{100, 250, 500\}$ . For each instance, we consider the Wasserstein radius  $\theta \in \{0.00, 0.02, \dots, 0.20\}$ , norm  $\|\cdot\|_1$ , and norm  $\|\cdot\|_2$  (i.e.,  $r \in \{1, 2\}$ ). The computational results can be found in Table 1, where ‘‘Time’’ represents the total running time for solving an instance with a particular parametric setting, ‘‘95% of  $v_U^{SAA}$ ’’ represents the 95% confidence interval (CI) of the optimal value of Model (14a), and ‘‘ $\theta^*$ ’’ is the smallest Wasserstein radius such that its corresponding  $v_U$  is below the 95% asymptotic CI of  $v_U^{SAA}$ . Here, CI is computed using samples and asymptotic formula, i.e.,

$$95\% \text{CI} = \text{sample mean} \pm 1.96 \times \text{sample standard deviation} / \sqrt{N}.$$

We also display the theoretical confidence interval, denoted by ‘‘Theoretical 95% CI’’, which is computed using the results in Theorem 3 with  $\epsilon = 0.025$ . Note that since we are not able to obtain a closed-form expression of  $\sigma$ , a key parameter in Theorem 3, we estimate it by drawing 1000 i.i.d. samples.

In Table 1, we see that by exploring the decomposition structure, each instance can be solved within 3 minutes. On the other hand, there is no much difference between norm  $\|\cdot\|_1$  and norm  $\|\cdot\|_2$  except the computational time. Thus, in practice, we recommend using  $\|\cdot\|_1$  (i.e.,  $r = 1$ ) in our reformulations. The optimal Wasserstein radius is decreasing as  $N$  increases, which is consistent with the intuition that the Wasserstein radius shrinks as the sample size grows. We also see that the theoretical CI varies as the underlying norm changes, as expected in Theorem 3. Particularly, both asymptotic CI and theoretical CI are quite similar and do not dominate each other.

Finally, we also test DRBCP-U (2a) under  $q$ -Wasserstein ambiguity set with  $q = 2$  using the same instances, where the formulation can be found in the appendix. We display the numerical results in Table 3 in Appendix A.4 due to the page limit, where we let  $\lambda^*$  denote the best dual multiplier found. For norm  $\|\cdot\|_2$  (i.e.,  $r = 2$ ), each instance takes more than an hour to solve, and is thus not included in Table 3. Compared to the results in Table 1, it is seen that the running time of DRBCP-U (2a) under  $\infty$ -Wasserstein ambiguity set is much shorter than that of DRBCP-U (2a) under 2-Wasserstein ambiguity set, which is mainly because the reformulation of the former does not have a dual multiplier. In contrast, the reformulation of the latter has. On the other hand, we can also see that by choosing a proper Wasserstein radius, DRBCP-U under each ambiguity set delivers a similar objective value, respectively. Thus, this implies that in practice, we recommend using the  $\infty$ -Wasserstein ambiguity set due to its shorter running time and similar statistical performances.

## 5.2 Numerical Illustrations of Decision Robust Model (17b)

We generate two sets of numerical study to illustrate the decision robust model (17b).

**First Set of Numerical Study: Using Real-world Dataset-** To demonstrate the effectiveness of decision robust model (17b), in our first set of numerical study, we test the bottleneck matching Example 6. We suppose that there are 9 drivers and 9 passengers, where the travel time between each pair of driver and passenger is stochastic. We use the 2018 New York taxi record data from <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>, where we treat average monthly travel time as a sample and thus, there are 12 empirical samples in total. We use the following cross-validation method to choose



Table 1: Numerical Results of DRBCP-U (2a) with Application to the Multi-hop Network.

$N$	$\theta$	$r = 1$			$r = 2$			95% CI of $v_U^{SAA}$	Theoretical 95% CI	
		Time	$v_U$	$\theta^*$	Time	$v_U$	$\theta^*$		$r = 1$	$r = 2$
100	0.00	5.210	5.411		5.535	5.411				
	0.02	6.045	5.391		11.679	5.391				
	0.04	6.137	5.372		12.019	5.371				
	0.06	6.241	5.353		12.999	5.352				
	0.08	6.222	5.335		12.970	5.332				
	0.10	6.179	5.316	0.16	12.318	5.313	0.16	[5.279, 5.543]	[5.286, 5.554]	[5.284, 5.369]
	0.12	6.260	5.299		12.714	5.294				
	0.14	6.290	5.281		12.889	5.275				
	0.16	6.290	5.264		12.317	5.257				
	0.18	6.285	5.247		12.658	5.238				
0.20	6.320	5.230		12.355	5.220					
250	0.00	14.057	5.417		14.292	5.417				
	0.02	16.462	5.398		31.270	5.397				
	0.04	16.473	5.378		31.542	5.378				
	0.06	16.762	5.359		33.610	5.358				
	0.08	16.823	5.341		33.901	5.339				
	0.10	16.734	5.322	0.10	32.767	5.319	0.10	[5.335, 5.500]	[5.335, 5.504]	[5.337, 5.391]
	0.12	16.978	5.304		33.183	5.300				
	0.14	16.939	5.286		33.352	5.281				
	0.16	16.999	5.268		33.135	5.262				
	0.18	16.982	5.251		33.149	5.244				
0.20	17.099	5.234		32.443	5.225					
500	0.00	29.429	5.452		32.524	5.452				
	0.02	34.224	5.433		65.434	5.432				
	0.04	34.619	5.413		66.014	5.413				
	0.06	35.205	5.394		68.920	5.393				
	0.08	35.266	5.375		69.517	5.374				
	0.10	34.839	5.357	0.06	67.238	5.354	0.06	[5.397, 5.508]	[5.393, 5.513]	[5.395, 5.433]
	0.12	35.302	5.339		68.342	5.335				
	0.14	35.585	5.321		69.470	5.316				
	0.16	35.526	5.303		70.626	5.297				
	0.18	35.470	5.286		69.109	5.279				
0.20	35.668	5.269		68.809	5.260					

the optimal  $\theta$ : (i) We randomly choose 8 samples as the training dataset and the other 4 samples as the testing dataset; (ii) Given each  $\theta \in \{0, 0.05, 0.1, \dots, 1.0\}$ , we solve the decision robust model (20a) using the training dataset and evaluate the mean and variance of its unique optimal solution using the testing dataset; and (iii) We repeat steps (i) and (ii) 50 times and compute the 95% confidence intervals of the estimated mean and variance. The computational results are illustrated in Figure 1.

From Figure 1, it is seen that the expected objective value slightly increases as  $\theta$  grows. On the other hand, the variance reduces significantly as  $\theta$  increases, in particular, when  $\theta$  is small. Thus, in general, we recommend users choosing a proper  $\theta$  such that the objective value of its corresponding optimal solution is relatively the same as SAA, but its variability is much smaller. In this numerical instance, we recommend using  $\theta = 0.65$ , where the expected objective value increases only by 3.3%, but its variance reduces by 64.2%.

We also study distributionally robust fair matching problem in Example 6 under total variation based ambiguity set (see theorem 1 in [25]), the detailed formulation can be found in Appendix A.5 and the numerical results are shown in Figure 2. As observed in [13, 30], the formulation using total variation (e.g., a special case of  $\phi$ -divergence) based ambiguity set (e.g., a special case of  $\phi$ -divergence) can help reduce the variance of the decision. However, comparing Figure 1 with Figure 2, it is seen that our decision robust model tends to find solutions with much lower variances while the objective value is almost the same as that of SAA.

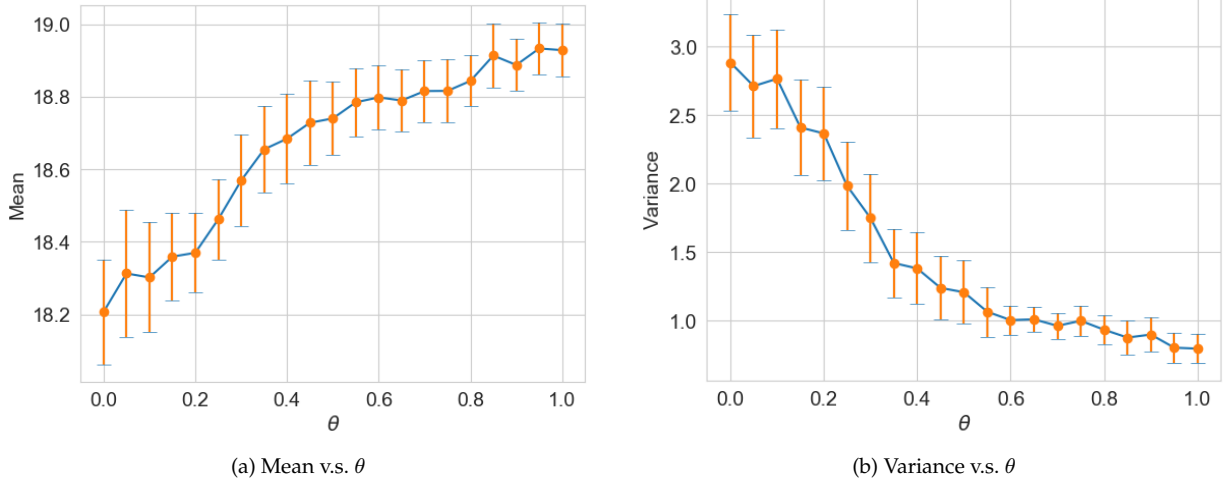


Fig. 1: Numerical Illustration of Decision Robust Model (17b): Using Real-world Dataset

**Second Set of Numerical Study: Using Hypothetical Dataset** In our second set of numerical study, similar to the first set of numerical study, we consider 9 drivers and 9 passengers. Their travel times are independent Gaussian variables truncated at the interval  $[0, \infty)$ , where their means are equal to the average travel times of the 12 empirical samples of 2018 New York record data, denoted by  $\mu_{NY}$ , and their standard deviations are proportional to the empirical standard deviations of the 2018 New York data,  $\alpha\sigma_{NY}$  with  $\alpha > 1$ . In particular, we consider two cases with  $\alpha = 10$  and  $\alpha = 100$ . For each case, we generate three different testing instances with sample size  $N \in \{10, 20, 40\}$ , respectively. For the case  $\alpha = 10$ , we consider the candidate  $\theta$  ranging from  $\{0, 1, \dots, 10\}$ , while for the case  $\alpha = 100$ , we consider the candidate  $\theta$  ranging from  $\{0, 5, \dots, 50\}$ . In addition, for each instance, we evaluate the unique optimal solution of the decision robust model (20a) and its 95% confidence interval by generating  $10^5$  i.i.d. samples. We also compute the theoretical confidence interval, denoted by “Theoretical 95% CI”, according to Theorem 4 with  $\epsilon = 0.025$ . Note that as we are not able to obtain a closed-form expression of  $\sigma$ , a key parameter in Theorem 4, we estimate it using the same  $10^5$  i.i.d. samples. The computational results are displayed in Table 2.

In Table 2, “Time” represents the total running time for solving an instance, “95% CI of Solutions” represents the 95% confidence interval (CI) of evaluating the optimal solution of the decision robust model (20a). Note that  $\theta = 0$  represents SAA model. It is seen that the SAA solution can be inferior to the solution of the decision robust model (20a), even for the instances with the larger sample size  $N = 40$ . The phenomenon is even more striking when the variance of the true distribution increases. Since the deviations after positive-negative sign in the column “95% CI of Solutions” are proportional to the sample standard deviation of  $10^5$  samples, it is seen that for most of the instances with  $\theta > 0$ , their confidence intervals are smaller than SAA model, which is consistent with the objective of the decision robust model (20a). On the other hand, since the decision robust model (20a) involves big-M coefficients and more auxiliary variables, its running time is often longer than that of the SAA. It is seen that the theoretical CI is often wider than the asymptotic one, implying that the proposed theoretical results might not be tight. In practice, if the sample size is small and the distributionally robust optimization model does not offer better solutions, we recommend using decision robust model as a better alternative.

We also test distributionally robust fair matching problem in Example 6 under total variation based ambiguity set in Appendix A.5 and the numerical results are shown in Table 4. Comparing Table 2 with Table 4, we can see that for most of the instances, our decision robust model can find better decisions with lower cross-validation objective values than those of distributionally robust counterpart under total varia-

Table 2: Numerical Results of Decision Robust Model (17b): Using Hypothetical Dataset

$N$	Case 1: $\mathcal{N}(\mu_{NY}, 10^2 \sigma_{NY})$					Case 2: $\mathcal{N}(\mu_{NY}, 10^4 \sigma_{NY})$				
	$\theta$	Time	$v_{DR}$	95% CI of Solutions	Theoretical 95% CI	$\theta$	Time	$v_{DR}$	95% CI of Solutions	Theoretical 95% CI
10	0	2.181	345.082	66.646 ± 0.130	66.646 ± 0.220	0	2.027	25974.710	497.251 ± 1.382	497.251 ± 2.345
	1	2.335	266.640	66.198 ± 0.139	66.198 ± 0.235	5	1.923	19726.124	478.010 ± 1.253	478.010 ± 2.127
	2	1.768	128.644	68.351 ± 0.127	68.351 ± 0.216	10	1.750			
	3	2.076				15	1.419			
	4	1.805	86.938	63.478 ± 0.130	63.478 ± 0.220	20	1.475			
	5	1.518	84.734	66.635 ± 0.127	66.635 ± 0.215	25	1.635	15999.190	489.546 ± 1.236	489.546 ± 2.097
	6	1.310				30	2.121	12968.835	473.883 ± 1.242	473.883 ± 2.108
	7	1.416				35	1.851			
	8	1.698	76.076	63.968 ± 0.128	63.968 ± 0.217	40	1.656	12879.386	<b>453.492 ± 1.251</b>	453.492 ± 2.123
	9	1.458	66.968	<b>60.016 ± 0.119</b>	60.016 ± 0.203	45	1.267	12303.694	463.123 ± 1.246	463.123 ± 2.115
10	1.598	52.736	67.945 ± 0.131	67.945 ± 0.222	50	1.447	12088.447	474.230 ± 1.226	474.230 ± 2.081	
20	0	5.975	129.040	64.258 ± 0.165	64.258 ± 0.280	0	6.483	17812.189	427.099 ± 1.149	427.099 ± 1.950
	1	5.767				5	8.392			
	2	4.982	10	6.694						
	3	43.247	15	54.704	14899.214	425.33 ± 1.159	425.33 ± 1.968			
	4	5.733	20	7.281						
	5	84.842	108.433	61.118 ± 0.159	61.118 ± 0.271	25	7.193	12579.868	406.824 ± 1.030	406.824 ± 1.749
	6	7.103	106.171	63.520 ± 0.169	63.520 ± 0.287	30	45.405	12433.038	473.365 ± 1.644	473.365 ± 2.790
	7	7.136				35	34.292			
	8	63.746				102.174	63.133 ± 0.120	63.133 ± 0.203	40	6.049
	9	9.318	89.397	<b>60.904 ± 0.122</b>	60.904 ± 0.207	45	59.523			
10	9.459	106.171	63.520 ± 0.169	63.520 ± 0.287	50	50.737				
40	0	83.605	166.063	61.118 ± 0.159	61.118 ± 0.271	0	146.536	29951.644	412.923 ± 1.086	412.923 ± 1.843
	1	66.774				5	123.786	21742.854	412.998 ± 1.063	412.998 ± 1.805
	2	94.462	148.079	<b>56.372 ± 0.103</b>	56.372 ± 0.174	10	139.989	16679.337	397.311 ± 1.007	397.311 ± 1.710
	3	110.405				15	118.074			
	4	33.100				20	151.350			
	5	160.091				25	130.666			
	6	282.649	30	141.449						
	7	293.787	35	135.817						
	8	511.723	133.785	63.109 ± 0.120	63.109 ± 0.204	40	136.939	13742.891	<b>396.782 ± 1.010</b>	396.782 ± 1.715
	9	1263.980	45	193.149						
10	49.186	148.760	67.394 ± 0.125	67.394 ± 0.244	50	160.060				

tion based ambiguity set. This further demonstrates that our proposed decision-robust model proposed in Section 4 can be indeed more reliable than the alternatives.

## 6 Extension: Distributionally Robust $\Gamma$ -sum Bottleneck Combinatorial Problems (DR $\Gamma$ BCP)

In this section, we extend the developed results into distributionally robust  $\Gamma$ -sum bottleneck combinatorial problems (DR $\Gamma$ BCP), where the decision-makers are interested in the worst-case sum of  $\Gamma$ - largest elements instead of the bottleneck element. Thus, following the notation of BCP (1), DRBCP-U (2a), and DRBCP-D (2a), their  $\Gamma$ -sum counterparts, i.e.,  $\Gamma$ BCP, DR $\Gamma$ BCP-U, DR $\Gamma$ BCP-D, can be formulated as below:

$$(\Gamma\text{BCP}) \quad Z_{\Gamma}(\mathbf{c}) = \min_{x \in X} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} c_j, \quad (21a)$$

$$(\text{DR}\Gamma\text{BCP-U}) \quad v_{U\Gamma} = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [Z_{\Gamma}(\tilde{\mathbf{c}})], \quad (21b)$$

$$(\text{DR}\Gamma\text{BCP-D}) \quad v_{D\Gamma} = \min_{x \in X} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[ \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j \right], \quad (21c)$$

where we assume that all the elements in  $X$  have the size at least  $\Gamma$ . Please note that in [33], the authors studied DR $\Gamma$ BCP-D with known marginal distributions, while this paper studies both DR $\Gamma$ BCP-U and DR $\Gamma$ BCP-D under the less conservative Wasserstein ambiguity set.

### 6.1 DR $\Gamma$ BCP-U (21b): Equivalent Reformulations and Confidence Bounds

Following the same definition of  $\bar{X}$  in Section 2.2, we can show that the following result. The proof is similar to Lemma 2 and is thus omitted.

**Lemma 3**  $\Gamma$ BCP (21a) is equivalent to

$$Z_\Gamma(\mathbf{c}) = \min_{h \in \bar{X}} \max_{s \subseteq h, |s|=\Gamma} \sum_{j \in s} c_j. \quad (22)$$

Next, let us define  $\Gamma$ -blocker  $F_\Gamma$  as

$$F_\Gamma = \left\{ y \subseteq 2^{[N]} : \forall h \in \bar{X} \text{ such that } y \cap \binom{h}{\Gamma} \neq \emptyset \text{ and } y \text{ is minimal with this property} \right\},$$

where for any given set  $S$ , we let  $\binom{S}{\Gamma}$  denote the collection of all its size- $\Gamma$  subsets. According to its definition, there exists a unique  $\Gamma$ -blocker  $F_\Gamma$ , where the existence is because  $\binom{[N]}{\Gamma}$  satisfies the definition, and the uniqueness is due to the minimality property. Another observation is that we must have  $F_\Gamma \subseteq \binom{[N]}{\Gamma}$  according to the nonempty intersection requirement and minimality property. Then we present the equivalent dual formulation of  $\Gamma$ BCP (22).

**Theorem 5** Suppose  $F$  is the  $\Gamma$ -blocker of  $\bar{X}$ . Then  $\Gamma$ BCP (21a) is equivalent to

$$Z_\Gamma(\mathbf{c}) = \max_{y \in F_\Gamma} \min_{s \in y} \sum_{j \in s} c_j. \quad (23)$$

*Proof:* Let  $\hat{Z}_\Gamma(\mathbf{c})$  denote the value of the right-hand side of (23). We separate the proof into two steps.

(i)  $Z_\Gamma(\mathbf{c}) \geq \hat{Z}_\Gamma(\mathbf{c})$ . For any  $y \in F_\Gamma$  and  $x \in \bar{X}$ , we have  $\binom{x}{\Gamma} \cap y \neq \emptyset$ . Thus,

$$\min_{s \in y} \sum_{j \in s} c_j \leq \min_{s \in \binom{x}{\Gamma} \cap y} \sum_{j \in s} c_j \leq \max_{s \in \binom{x}{\Gamma} \cap y} \sum_{j \in s} c_j \leq \max_{s \in \binom{x}{\Gamma}} \sum_{j \in s} c_j,$$

where the first inequality is due to  $\binom{x}{\Gamma} \cap y \subseteq y$ , the second inequality is due to the fact that minimization is no larger than maximization, and the third one is due to  $\binom{x}{\Gamma} \cap y \subseteq \binom{x}{\Gamma}$ . Since the above inequalities hold for any  $y \in F_\Gamma$  and  $x \in \bar{X}$ , we must have  $Z_\Gamma(\mathbf{c}) \geq \hat{Z}_\Gamma(\mathbf{c})$ .

(ii)  $Z_\Gamma(\mathbf{c}) \leq \hat{Z}_\Gamma(\mathbf{c})$ . Let set  $\{s_i\}_{i \in \binom{[n]}{\Gamma}} := \binom{[n]}{\Gamma}$  be such that their corresponding objective values are sorted in ascending order, i.e.,  $\sum_{j \in s_{i_1}} c_j \leq \dots \leq \sum_{j \in s_{i_2}} c_j \leq \dots \leq \sum_{j \in s_{i_3}} c_j$ . Let  $i_1 \in \binom{[n]}{\Gamma}$  be the smallest index such that  $\binom{x}{\Gamma} \subseteq \{s_i\}_{i \in [i_1]}$  for some  $x \in \bar{X}$ . Clearly, for any such  $x$ , according to our construction, we must have  $s_{i_1} \subseteq x$ , otherwise, we can reduce  $i_1$ , a contradiction. Thus, we have

$$Z_\Gamma(\mathbf{c}) = \sum_{j \in s_{i_1}} c_j.$$

On the other hand, since  $\binom{x}{\Gamma} \not\subseteq \{s_i\}_{i \in [i_1-1]}$  for any  $x \in \bar{X}$ , for any  $x \in \bar{X}$ , there must exist  $s_{i_x} \subseteq x$  for some  $i_x \in \{i_1, \dots, \binom{[n]}{\Gamma}\}$ . Therefore, according to the minimality property of the elements of  $\Gamma$ -blocker  $F_\Gamma$ , there exists at least one  $\hat{y} \in F_\Gamma$  such that  $\hat{y} \subseteq \{s_i\}_{i \in [i_1, \dots, \binom{[n]}{\Gamma}]}$ . Hence,

$$\hat{Z}_\Gamma(\mathbf{c}) = \max_{y \in F_\Gamma} \min_{s \in y} \sum_{j \in s} c_j \geq \min_{s \in \hat{y}} \sum_{j \in s} c_j \geq \sum_{j \in s_{i_1}} c_j = Z_\Gamma(\mathbf{c}).$$

□

We are ready to present the main results of reformulations and confidence bounds.

**Theorem 6 (Reformulations)** Suppose  $\|\cdot\| = \|\cdot\|_r$  with  $r \geq 1$ , then

(i) DR $\Gamma$ BCP-U (21b) is equivalent to

$$v_{UG} = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \max_{\beta^k} \left\{ \min_{s \in y} \sum_{j \in s} (\bar{c}_j^k + \beta_j^k) : \sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s \right\}. \quad (24a)$$

(ii) Let  $v_{UG}^{SAA} = \frac{1}{N} \sum_{k \in [N]} \min_{x \in X} \max_{s \subseteq x, |s| = \Gamma} \sum_{j \in s} \bar{c}_j^k$  denote the sampling average approximation counterpart of DR $\Gamma$ BCP-U (21b). Then

$$v_{UG}^{SAA} + \frac{\Gamma \theta}{\max_{y \in F_\Gamma} |\bigcup_{s \in y} s|^{\frac{1}{r}}} \leq v_{UG} \leq v_{UG}^{SAA} + \Gamma^{\frac{r-1}{r}} \theta. \quad (24b)$$

*Proof:*

(i) According to the representation (4) and Theorem 5, we have

$$v_{UG} = \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}} [Z_\Gamma(\tilde{\mathbf{c}})] = \frac{1}{N} \sum_{k \in [N]} \max_{\|\mathbf{c}^k - \tilde{\mathbf{c}}^k\|_r \leq \theta} \max_{y \in F_\Gamma} \min_{s \in y} \sum_{j \in s} c_j^k.$$

By swapping the two maximization operators and letting  $\beta_j^k = c_j^k - \bar{c}_j^k$ , we further have

$$v_{UG} = \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \max_{\|\beta^k\|_r \leq \theta} \min_{s \in y} \sum_{j \in s} (\bar{c}_j^k + \beta_j^k).$$

According to the optimality condition, we must have  $\beta_j^k \geq 0$  for all  $j \in \bigcup_{s \in y} s$ , and we arrive at (24a).

(ii) According to (24a), we note that

$$\begin{aligned} v_{UG} &\leq \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \min_{s \in y} \sum_{j \in s} \bar{c}_j^k + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \max_{\sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s} \max_{s \in y} \sum_{j \in s} \beta_j^k \\ &= v_{UG}^{SAA} + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \max_{s \in y} \max_{\sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s} \sum_{j \in s} \beta_j^k \\ &= v_{UG}^{SAA} + \Gamma^{\frac{r-1}{r}} \theta \end{aligned}$$

where the first inequality is due to

$$\min_{s \in y} \sum_{j \in s} (\bar{c}_j^k + \beta_j^k) \leq \min_{s \in y} \sum_{j \in s} \bar{c}_j^k + \max_{s \in y} \sum_{j \in s} \beta_j^k,$$

the first equality is due to the definition of  $v_{UG}^{SAA}$  and the interchange of maximization operators, and the second one is due to the definition of the dual norm.

On the other hand, replacing the inner minimization in (24a) by its lower bound, we have

$$\begin{aligned} v_{UG} &\geq \frac{1}{N} \sum_{k \in [N]} \max_{y \in F_\Gamma} \min_{s \in y} \sum_{j \in s} \bar{c}_j^k + \frac{1}{N} \sum_{k \in [N]} \min_{y \in F_\Gamma} \max_{\sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s} \min_{s \in y} \sum_{j \in s} \beta_j^k \\ &\geq v_{UG}^{SAA} + \frac{1}{N} \sum_{k \in [N]} \min_{y \in F_\Gamma} \max_{\sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s} H(\beta) \end{aligned}$$

where

$$H(\beta) := \min_{\lambda^k} \left\{ \sum_{j \in \bigcup_{s \in y} s} \lambda_j^k \beta_j^k : \sum_{j \in \bigcup_{s \in y} s} \lambda_j^k = \Gamma, \lambda_j^k \in [0, 1], \forall j \in \bigcup_{s \in y} s \right\}.$$

Note that  $H(\beta)$  is a symmetric concave function and the constraint system  $\{\beta : \sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s\}$  is also symmetric. Thus, the optimal value of

$$\max_{\beta} \left\{ H(\beta) : \sum_{j \in \bigcup_{s \in y} s} (\beta_j^k)^r \leq \theta^r, \beta_j^k \geq 0, \forall j \in \bigcup_{s \in y} s \right\}$$

is achieved by  $\beta_j^k = \frac{\theta}{|\bigcup_{s \in y} s|^{\frac{1}{r}}}$  for each  $j \in \bigcup_{s \in y} s$ , i.e.,

$$v_{U\Gamma} \geq v_{U\Gamma}^{SAA} + \frac{\Gamma\theta}{\max_{y \in F_T} |\bigcup_{s \in y} s|^{\frac{1}{r}}}.$$

□

We remark that: (a) If there are a finite number of elements in  $F_T$ , then for each  $y \in F_T$ , we can convert the inner maximin problem in formulation (24a) as a convex program; and (b) In general, formulation (24a) can be difficult to solve. Thus, we recommend obtaining  $v_{U\Gamma}^{SAA}$  instead and using inequalities (24b) to estimate  $v_{U\Gamma}$ .

The following results on confidence bounds follow the similar proof techniques in Theorem 7, and thus their proofs are omitted for brevity.

**Theorem 7** Let  $v_{U\Gamma}^T = \mathbb{E}_{\mathbb{P}^T}[\min_{x \in X} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j]$  denote stochastic  $\Gamma$ BCP under the true distribution  $\mathbb{P}^T$ . Suppose that there exists a positive  $\sigma$  such that  $\mathbb{E}_{\mathbb{P}^T}[\exp(-(Z_{\Gamma}(\tilde{c}) - v_{U\Gamma}^T)^2/\sigma^2)] \leq e$ . Given  $\epsilon \in (0, 1)$ , then

(i) let  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon)}\Gamma^{-1}\max_{y \in F_T} |\bigcup_{s \in y} s|^{\frac{1}{r}} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_{U\Gamma} \geq v_{U\Gamma}^T\} \geq 1 - \epsilon;$$

(ii) let  $\theta = N^{-\frac{1}{2}}\sigma\sqrt{-3\log(\epsilon)}\Gamma^{-\frac{r-1}{r}} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_{U\Gamma} \leq v_{U\Gamma}^T + 2\theta\} \geq 1 - \epsilon.$$

## 6.2 DR $\Gamma$ BCP-D (21c): Equivalent Reformulations, Confidence Bounds, and Alternative Decision-Robust Model

In this subsection, we will provide equivalent reformulations, and confidence bounds of DR $\Gamma$ BCP-D (21c). Similarly, DR $\Gamma$ BCP-D (21c) may not yield robust decisions when the sample size is small, and therefore, we propose an alternative decision-robust model. The proofs in this subsection are almost identical to those in Section 4 and are thus omitted.

We first summarize the reformulation results.

**Theorem 8 (Reformulations)** Let  $v_{D\Gamma}^{SAA} = \min_{x \in X} \frac{1}{N} \sum_{k \in [N]} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j^k$  denote sampling average approximation of stochastic decision-making  $\Gamma$ BCP with empirical distribution  $\mathbb{P}_{\tilde{c}}$ . Suppose  $\|\cdot\| = \|\cdot\|_r$  with  $r \geq 1$ , then

(i) DR $\Gamma$ BCP-D (21c) is equivalent to

$$v_{D\Gamma} = v_{D\Gamma}^{SAA} + \Gamma^{\frac{r-1}{r}}\theta \quad (25a)$$

(ii) Suppose that set  $X$  admits a binary programming representation  $\hat{X} \subseteq \{0, 1\}^n$ . Then  $v_{D\Gamma}^{SAA}$  is equal to

$$v_{D\Gamma}^{SAA} = \min_{z^k \in \hat{X}, \alpha^k, \pi^k, \forall k \in [N]} \frac{1}{N} \sum_{k \in [N]} \left( \alpha^k \Gamma + \sum_{j \in [n]} \pi_j^k \right) \quad (25b)$$

s.t.  $\alpha^k + \pi_j^k \geq (\tilde{c}_j^k - L^k)z_j^k + L^k, \pi_j^k \geq 0, \forall j \in [n], k \in [N],$

where  $L^k := \min_{\tau \in [n]} \tilde{c}_{\tau}^k$ .

Next, we summarize the confidence bounds.

**Theorem 9** Let  $v_{D\Gamma}^T = \min_{x \in X} \mathbb{E}_{\mathbb{P}^T} [\sum_{k \in [N]} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j]$  denote stochastic decision-making  $\Gamma$ BCP with true distribution  $\mathbb{P}^T$ . Suppose that there exists a positive  $\sigma$  such that

$$\mathbb{E}_{\mathbb{P}^T} \left[ \exp \left( \sigma^{-2} \left( \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j - \mathbb{E}_{\mathbb{P}^T} \left[ \sum_{k \in [N]} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j \right] \right)^2 \right) \right] \leq e$$

for each  $x \in X$ . Given  $\epsilon \in (0, 1)$  and  $r \in [1, \infty]$ , then

(i) let  $\theta = \Gamma^{-\frac{r-1}{r}} N^{-\frac{1}{2}} \sigma \sqrt{-3 \log(\epsilon) + 3n \log(2)} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_{D\gamma} \geq v_{D\gamma}^T\} \geq 1 - \epsilon;$$

(ii) let  $\theta = \Gamma^{-\frac{r-1}{r}} N^{-\frac{1}{2}} \sigma \sqrt{-3 \log(\epsilon) + 3n \log(2)} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_{D\gamma} \leq v_{D\gamma}^T + 2\theta\} \geq 1 - \epsilon.$$

We propose the following decision-robust model as an alternative of DR $\Gamma$ BCP-D (21c).

$$v_{DR}^\gamma = \min_{x \in \mathcal{L}_{\theta\gamma}(X)} \mathbb{V}_{\mathbb{P}_\xi} \left[ \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j \right], \quad (26)$$

where

$$\mathcal{L}_{\theta\Gamma}(X) = \left\{ x \in X : \frac{1}{N} \sum_{k \in [N]} \max_{s \subseteq x, |s|=\Gamma} \sum_{j \in s} \tilde{c}_j^k \leq v_{D\Gamma} := v_{D\Gamma}^{SAA} + \Gamma^{\frac{r-1}{r}} \theta \right\}. \quad (27)$$

Finally, we show that the decision-robust model (26) can be formulated as an MISOCP.

**Proposition 6** Suppose that set  $X$  admits a binary programming representation  $\widehat{X}$ . Then the decision-robust model (17b) is equivalent to the following MISOCP

$$\begin{aligned} v_{DR}^\Gamma &= \min_{z \in \widehat{X}, v, \alpha, \pi, \lambda} \frac{1}{N} \sum_{k \in [N]} \left( v^k - \frac{1}{N} \sum_{\tau \in [N]} v^\tau \right)^2 \\ \text{s.t.} \quad & \frac{1}{N} \sum_{k \in [N]} v^k \leq v_{D\Gamma}^{SAA} + \Gamma^{\frac{r-1}{r}} \theta, \\ & v^k = \alpha^k \Gamma + \sum_{j \in [n]} \pi_j^k \geq (\tilde{c}_j^k - L^k) z_j + L^k, \forall j \in [n], k \in [N], \\ & v^k \leq \sum_{j \in [n]} \lambda_j^k \tilde{c}_j^k, \forall k \in [N], \\ & \sum_{j \in [n]} \lambda_j^k = \Gamma, \forall k \in [N], \\ & \lambda_j^k \leq z_j, \forall k \in [N], j \in [n], \\ & \lambda_j^k \in [0, 1], \forall j \in [n], k \in [N], \end{aligned} \quad (28)$$

where  $L^k := \min_{\tau \in [n]} \tilde{c}_\tau^k$ .

## 7 Conclusion

This paper studies distributionally robust bottleneck combinatorial problems (DRCBP) under the Wasserstein ambiguity set. Motivated by different applications, this paper focuses on two variants of DRCBP: DRCBP-U and DRCBP-D: (i) Motivated by the multi-hop wireless network, for DRCBP-U, the decision-makers would like to have a reliable estimation of the values of the stochastic bottleneck combinatorial problem; and (ii) Motivated by the ride-sharing transportation problem, for DRCBP-D, the decision-makers seek the best decision to minimize the worst-case bottleneck cost. In both models, the underlying true probability distributions are assumed to be not fully specified and are contained in an ambiguity set defined by a ball centered at empirical distribution within a certain Wasserstein distance. For both DRCBP-U and DRCBP-D, we derive their equivalent reformulations, demonstrate that they can be computed effectively as mixed-integer programs, and show that the Wasserstein radius can be chosen less conservatively by establishing their connections to the sampling average approximation counterparts. We show that the results can be extended to distributionally robust  $\Gamma$ -sum bottleneck combinatorial problems (DR $\Gamma$ CBP), where the decision-makers would like to know the worst-case sum of  $\Gamma$  largest elements instead of the bottleneck element.

## Acknowledgments

The authors would like to thank Editor-in-Chief Dr. Sven Leyffer and two anonymous reviewers for their valuable comments on improving this paper.

## References

1. S. S. Abadeh, V. A. Nguyen, D. Kuhn, and P. M. M. Esfahani. Wasserstein distributionally robust kalman filtering. In *Advances in Neural Information Processing Systems*, pages 8474–8483, 2018.
2. S. Ahmed. Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming*, 106(3):433–446, 2006.
3. H. Albrecher. A note on the asymptotic behaviour of bottleneck problems. *Operations research letters*, 33(2):183–186, 2005.
4. A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. SIAM, 2001.
5. D. Bertsimas, S. Shtern, and B. Sturt. A data-driven approach for multi-stage linear optimization. *Optimization Online*.
6. J. Blanchet, Y. Kang, and K. Murthy. Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019.
7. J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.
8. J. Blanchet, K. Murthy, and N. Si. Confidence regions in wasserstein distributionally robust estimation. *arXiv preprint arXiv:1906.01614*, 2019.
9. P. M. Camerini. The min-max spanning tree problem and some extensions. *Information Processing Letters*, 7(1):10–14, 1978.
10. Z. Chen, D. Kuhn, and W. Wiesemann. Data-driven chance constrained programs over wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.
11. Z. Chen and W. Xie. Sharing the value-at-risk under distributional ambiguity. *Available at SSRN 3400033*, 2019.
12. E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.
13. J. C. Duchi and H. Namkoong. Variance-based regularization with convex objectives. *J. Mach. Learn. Res.*, 20:68–1, 2019.



14. J. Edmonds and D. R. Fulkerson. Bottleneck extrema. *Journal of Combinatorial Theory*, 8(3):299–306, 1970.
15. P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.
16. N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.
17. R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.
18. Y. Gao, D.-M. Chiu, and J. Lui. Determining the end-to-end throughput capacity in multi-hop networks: methodology and applications. In *ACM SIGMETRICS Performance Evaluation Review*, volume 34, pages 39–50. ACM, 2006.
19. A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.
20. V. Guigues, A. Juditsky, and A. Nemirovski. Non-asymptotic confidence bounds for the optimal value of a stochastic program. *Optimization Methods and Software*, 32(5):1033–1058, 2017.
21. G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018.
22. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151:35–62, 2015.
23. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.
24. W.-L. Hsu and G. L. Nemhauser. Easy and hard bottleneck location problems. *Discrete Applied Mathematics*, 1(3):209–215, 1979.
25. R. Jiang and Y. Guan. Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research*, 66(5):1390–1405, 2018.
26. V. Kaibel and M. Peinhardt. On the bottleneck shortest path problem. 2006.
27. A. Kasperski and P. Zieliński. Possibilistic bottleneck combinatorial optimization problems with ill-known weights. *International Journal of Approximate Reasoning*, 52(9):1298–1311, 2011.
28. A. Kasperski and P. Zieliński. Bottleneck combinatorial optimization problems with uncertain costs and the owa criterion. *Operations Research Letters*, 41(6):639–643, 2013.
29. D. Kuhn, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019.
30. H. Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4):1090–1105, 2019.
31. X. Liu, K. Ravindran, and D. Loguinov. Multi-hop probing asymptotics in available bandwidth estimation: stochastic analysis. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 15–15. USENIX Association, 2005.
32. S. Martin, D. Ouelhadj, P. Smet, G. V. Berghe, and E. Özcan. Cooperative search for fair nurse rosters. *Expert Systems with Applications*, 40(16):6674–6683, 2013.
33. K. Natarajan, D. Shi, and K.-C. Toh. A probabilistic model for minmax regret in combinatorial optimization. *Operations Research*, 62(1):160–181, 2013.
34. B. Parent and A. L. Caplan. Fair is fair: We must re-allocate livers for transplant. *BMC medical ethics*, 18(1):26, 2017.
35. F. Pérez-Galarce, A. Candia-Véjar, C. Astudillo, and M. Bardeen. Algorithms for the minmax regret path problem with interval data. *Information Sciences*, 462:218–241, 2018.
36. U. Pferschy. The random linear bottleneck assignment problem. *RAIRO-Operations Research*, 30(2):127–142, 1996.
37. H. Rahimian and S. Mehrotra. Distributionally robust optimization: A review. Available at [Optimization Online](https://optimization-online.org/), 2019.

38. A. Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.
39. S. Shen, M. Kurt, and J. Wang. Chance-constrained programming models and approximations for general stochastic bottleneck spanning tree problems. *INFORMS Journal on Computing*, 27(2):301–316, 2015.
40. T. Shinn and T. Takaoka. Efficient graph algorithms for network analysis. In *First International Conference on Resource Efficiency in Interorganizational Networks-ResEff 2013*, page 236, 2013.
41. M. Z. Spivey. Asymptotic moments of the bottleneck assignment problem. *Mathematics of Operations Research*, 36(2):205–226, 2011.
42. I. Stancu-Minasian, R. Caballero, E. Cerdá, and M. Muñoz. The stochastic bottleneck linear programming problem. *Top*, 7(1):123–143, 1999.
43. N. G. Trillos and D. Slepčev. On the rate of convergence of empirical measures in  $\infty$  transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.
44. W. Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, pages 1–41, 2019.
45. W. Xie. Tractable reformulations of distributionally robust two-stage stochastic programs with  $\infty$ -wasserstein distance. *arXiv preprint arXiv:1908.08454*, 2019.
46. W. Xie and S. Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182, 2018.
47. J.-Y. Yoo and J. Kim. Maximum end-to-end throughput of chain-topology wireless multi-hop networks. In *2007 IEEE Wireless Communications and Networking Conference*, pages 4279–4283. IEEE, 2007.
48. L. Zhang, S. Chen, and Y. Jian. Achieving global end-to-end maxmin in multihop wireless networks. In *2008 The 28th International Conference on Distributed Computing Systems*, pages 225–232. IEEE, 2008.
49. S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137:167–198, 2013.

## Appendix A. DRBCP-U (2a) under $q$ -Wasserstein Ambiguity Set: Equivalent Reformulation, MIP Representation, and Confidence Bounds

For the completeness of this paper, in this appendix, we provide parallel results for DRBCP-U (2a) under  $q$ -Wasserstein ambiguity set (3) with  $q \in [1, \infty)$ .

### A.1 Equivalent Reformulation

We first derive a deterministic representation of DRBCP-U (2a) under  $q$ -Wasserstein ambiguity set.

**Theorem 10** *Suppose  $\|\cdot\| = \|\cdot\|_r$  with  $r \geq 1$ , then DRBCP-U (2a)  $q$ -Wasserstein ambiguity set is equivalent to*

$$v_U = \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k} \left\{ t^k - \lambda \left( \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \right)^{\frac{q}{r}} \right\} \right\}, \quad (29)$$

where  $F$  is defined in Theorem 1 and set  $I^k(t, y) = \{j \in y : t - \bar{c}_j^k \geq 0\}$ .

*Proof:* According to Theorem 1 and theorem 1 in [17], we have

$$v_U = \sup_{\mathbb{P} \in \mathcal{P}_q} \mathbb{E}_{\mathcal{P}} [Z(\bar{c})] = \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{c^k} \max_{y \in F} \left( \min_{j \in y} c_j^k - \lambda \|c^k - \bar{c}^k\|_r^q \right) \right\}.$$

By swapping the two maximization operators and linearizing the inner minimization operator, we further have

$$v_U = \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k \leq c_j^k, \forall j \in y} \left( t^k - \lambda \|c^k - \bar{c}^k\|_r^q \right) \right\}.$$

To simplify the inner maximization, we let  $\beta_j^k = c_j^k - t^k$  for all  $j \in y$  and define set  $I^k(t, y) = \{j \in y : t - \bar{c}_j^k \geq 0\}$ , then we have

$$\begin{aligned} v_U &= \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{\beta^k, t^k} \left\{ (t^k - \lambda \|t^k \mathbf{e} + \beta^k - \bar{c}^k\|_r^q : \beta_j^k \geq 0, \forall j \in y) \right\} \right\} \\ &= \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k} \left\{ t^k - \lambda \left( \sum_{j \in I^k(t^k, y)} (t^k - \bar{c}_j^k)^r \right)^{\frac{q}{r}} \right\} \right\}, \end{aligned} \quad (30)$$

where the second equality is due to the optimality condition that if  $j \in y \setminus I^k(t^k, y)$ , we must have  $\beta_j^k = -t^k + \bar{c}_j^k > 0$ ; otherwise,  $\beta_j^k = 0$ . □

### A.2 Mixed Integer Programming Representation

Note that in Theorem 10, the outer minimization problem is a univariate convex program and thus can be solved efficiently via the bisection method. We then observe that for any given  $\lambda$ , the inner maximizations in DRBCP-U (29) can be decomposed into  $N$  separate MIPs.

**Proposition 7** Suppose that the blocker  $F$  admits a binary programming representation  $\widehat{F} \subseteq \{0, 1\}^n$ . Then  $v_U = \min_{\lambda \geq 0} \{\lambda \theta^q + 1/N \sum_{k \in [N]} v_U^k(\lambda)\}$  and  $v_U^k(\lambda)$  is equivalent to

$$v_U^k(\lambda) = \max_{z^k \in \widehat{F}, t^k, \beta^k, \theta^k} \{t^k - \lambda(\theta^k)^q : \|t^k \mathbf{e} - \bar{c}_j^k + \beta^k\|_r \leq \theta^k, \beta_j^k \geq -M_j^k(1 - z_j^k), \forall j \in [n]\}, \quad (31)$$

where  $M_j^k = \widehat{M}^k - \bar{c}_j^k$  and  $\widehat{M}^k = \max_{\tau \in [n]} \bar{c}_\tau^k + (\lambda q)^{-1/(q-1)}$ .

*Proof:* According to the proof of Theorem 10,  $v_U = \min_{\lambda \in [0, 1]} \{\lambda \theta^q + 1/N \sum_{k \in [N]} v_U^k(\lambda)\}$ , where

$$v_U^k(\lambda) := \max_{y \in F, t^k, \beta^k, \theta^k} \{t^k - \lambda(\theta^k)^q : \|t^k \mathbf{e} - \bar{c}_j^k + \beta^k\|_r \leq \theta^k, \beta_j^k \geq 0, \forall j \in y\}.$$

Above, we can augment the vector  $\beta^k$  by defining free variable  $\beta_j^k$  for each  $j \in [n] \setminus y$  and rewrite  $v_U^k(\lambda)$  as

$$v_U^k(\lambda) := \max_{y \in F, t^k, \beta^k, \theta^k} \{t^k - \lambda(\theta^k)^q : \|t^k \mathbf{e} - \bar{c}_j^k + \beta^k\|_r \leq \theta^k, \beta_j^k \geq 0, \forall j \in y, \beta_j^k \in \mathbb{R}, \forall j \in [n] \setminus y\},$$

since at optimality, we must have  $\beta_j^k = -t^k + \bar{c}_j^k$  for each  $j \in [n] \setminus y$ .

As the blocker  $F$  admits a binary programming representation  $\widehat{F}$ , the value  $v_U^k(\lambda)$  is further equal to

$$v_U^k(\lambda) := \max_{z^k \in \widehat{F}, t^k, \beta^k, \theta^k} \{t^k - \lambda(\theta^k)^q : \|t^k \mathbf{e} - \bar{c}_j^k + \beta^k\|_r \leq \theta^k, z_j^k \beta_j^k \geq 0, \forall j \in [n]\}. \quad (32)$$

To prove that (32) and (31) are equivalent, we only need to show that  $\beta_j^k \geq -M_j^k$  for each  $j \in [n]$ . Since at optimality,  $\beta_j^k$  is equal to 0 or  $-t^k + \bar{c}_j^k$ . Thus, it is sufficient to find an upper bound of  $t^k$ . In (29), suppose the optimal  $t^k \geq \max_{\tau \in [n]} \bar{c}_\tau^k$ , and then  $I^k(t, y) = y$ . The first order optimality condition of inner maximization of (29) yields

$$1 = \lambda q \left( \sum_{j \in y} (t^k - \bar{c}_j^k)^r \right)^{\frac{q-r}{r}} \left( \sum_{j \in y} (t^k - \bar{c}_j^k)^{r-1} \right) \geq \lambda q \left( t^k - \max_{\tau \in [n]} \bar{c}_\tau^k \right)^{q-1},$$

where the first inequality is due to  $t^k - \bar{c}_j^k \geq t^k - \max_{\tau \in [n]} \bar{c}_\tau^k$  and  $|y| \geq 1$ . Thus, we can define  $\widehat{M}^k$  as

$$\widehat{M}^k = \max_{\tau \in [n]} \bar{c}_\tau^k + (\lambda q)^{-1/(q-1)}.$$

□

We remark that: (i) to obtain  $v_U$ , one needs to first fix  $\lambda$  and then solve  $N$  separate MIPs (31), and then optimize  $\lambda$ , which requires more computational time compared to that of DRBCP-U (2a) under  $\infty$ -Wasserstein ambiguity set; and (ii) as long as  $r$  is a rational number, using the conic quadratic representation results in [4], then the MIP (31) can be formulated as a mixed integer second order conic program (MISOCP).

### A.3 Confidence Bounds

In this subsection, we plan to compare  $v_U$  with its sampling average approximation counterpart, which further motivates us a choice of Wasserstein radius  $\theta$ . Recall that  $v_U^{SAA}$  and  $v_U^T$  are defined in (14a) and (14b), respectively. Our goal is to determine a proper  $\theta$  such that  $v_U \geq v_U^T$  with a high probability. To begin with, we would like to show the relationship between  $v_U$  and  $v_U^{SAA}$ .

**Proposition 8** Let  $v_U^{SAA}$  be defined in (14a). Then

$$\frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}} \leq v_U - v_U^{SAA} \leq \theta.$$

*Proof:*

(i) To prove  $v^U - v^{SAA} \leq \theta$ , using the fact that  $t^k \geq \min_{j \in y} \bar{c}_j^k$ , then we can upper bound  $v^U$  as

$$v_U \leq \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k \geq \min_{j \in y} \bar{c}_j^k} \left\{ t^k - \lambda \left( t^k - \min_{j \in y} \bar{c}_j^k \right)^q \right\} \right\}. \quad (33)$$

To optimize the inner maximization, there are two cases:

Case 1. If  $q = 1$ , then the inequality (33) is equivalent to

$$v_U \leq \min_{\lambda \geq 0} \left\{ \lambda \theta + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \min_{j \in y} \bar{c}_j^k : \lambda \geq 1 \right\} = v^{SAA} + \theta.$$

Case 2. If  $q > 1$ , then the inequality (33) is equivalent to

$$\begin{aligned} v_U &\leq \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{q-1}{q^{\frac{q}{q-1}}} \lambda^{-\frac{1}{q-1}} + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \min_{j \in y} \bar{c}_j^k \right\} \\ &= v^{SAA} + \theta. \end{aligned}$$

(ii) To prove  $v^U - v^{SAA} \geq \frac{\theta}{\max_{y \in F} |y|^{\frac{q}{r}}}$ , using the fact that  $t^k \geq \min_{j \in y} \bar{c}_j^k$  again, then we can bound  $v^U$  from the below as

$$v_U \geq \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \max_{t^k \geq \min_{j \in y} \bar{c}_j^k} \left\{ t^k - \lambda |y|^{\frac{q}{r}} \left( t^k - \min_{j \in y} \bar{c}_j^k \right)^q \right\} \right\}. \quad (34)$$

To optimize the inner maximization, there are two cases:

Case 1. If  $q = 1$ , then the inequality (34) is equivalent to

$$\begin{aligned} v_U &\geq \min_{\lambda \geq 0} \left\{ \lambda \theta + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \min_{j \in y} \bar{c}_j^k : \lambda \geq \frac{1}{\min_{h \in F} |h|^{\frac{q}{r}}} \right\} = v^{SAA} + \frac{\theta}{\min_{y \in F} |y|^{\frac{q}{r}}} \\ &\geq v^{SAA} + \frac{\theta}{\max_{y \in F} |y|^{\frac{q}{r}}}. \end{aligned}$$

Case 2. If  $q > 1$ , using the fact that  $|y|^{\frac{q}{r}} \leq \max_{h \in F} |h|^{\frac{q}{r}}$  for all  $y \in F$ , then the right-hand side of inequality (34) is further lower bounded by

$$\begin{aligned} v_U &\geq \min_{\lambda \geq 0} \left\{ \lambda \theta^q + \frac{q-1}{q^{\frac{q}{q-1}}} \left( \frac{1}{\lambda \max_{h \in F} |h|^{\frac{q}{r}}} \right)^{\frac{1}{q-1}} + \frac{1}{N} \sum_{k \in [N]} \max_{y \in F} \min_{j \in y} \bar{c}_j^k \right\} \\ &= v^{SAA} + \frac{\theta}{\max_{y \in F} |y|^{\frac{1}{r}}}. \end{aligned}$$

□

Now we are ready to show the following main result on the relationship between the value of DRBCP-U  $v_U$  and true value  $v_U^T$ .

**Theorem 11** *Suppose that there exists a positive  $\sigma$  such that  $\mathbb{E}_{\mathbb{P}^T}[\exp((Z(\tilde{c}) - v_U^T)^2/\sigma^2)] \leq e$ . Given  $\epsilon \in (0, 1)$ , then*

(i) let  $\theta = N^{-\frac{1}{2}}\sigma q^{-1/q}\sqrt{-3\log(\epsilon)}\max_{y\in F}|y|^{\frac{1}{r}} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_U \geq v_U^T\} \geq 1 - \epsilon;$$

(ii) let  $\theta = N^{-\frac{1}{2}}\sigma q^{-1/q}\sqrt{-3\log(\epsilon)} = O(N^{-\frac{1}{2}})$ . Then we have

$$\mathbb{P}^T \{v_U \leq v_U^T + 2\theta\} \geq 1 - \epsilon.$$

*Proof:* The proof is almost identical to Theorem 3 and is thus omitted.  $\square$

#### A.4 Numerical Illustration of DRBCP-U (2a) under 2–Wasserstein Ambiguity Set

Table 3: Numerical Results of DRBCP-U (2a) under 2–Wasserstein Ambiguity Set with Application to the Multi-hop Network.

$N$	$\theta$	$r = 1$				95% CI of $v_U^{SAA}$	Theoretical 95% CI
		Time	$v_U$	$\lambda^*$	$\theta^*$		
100	0.00	565.993	5.411	9761.176			
	0.02	477.811	5.391	24.607			
	0.04	610.274	5.371	12.077			
	0.06	506.139	5.352	7.993			
	0.08	607.456	5.333	5.892			
	0.10	478.448	5.314	4.660	0.14	[5.279, 5.543]	
	0.12	446.110	5.296	3.843			
	0.14	430.717	5.278	3.262			
	0.16	431.357	5.259	2.823			
	0.18	615.299	5.241	2.475			
	0.20	450.941	5.224	2.196			
250	0.00	1482.088	5.417	6180.340			
	0.02	1617.128	5.398	24.728			
	0.04	1549.721	5.378	12.235			
	0.06	1612.423	5.359	8.012			
	0.08	1725.868	5.339	6.002			
	0.10	1700.600	5.320	4.668	0.10	[5.335, 5.499]	
	0.12	1715.304	5.302	3.857			
	0.14	1571.976	5.283	3.272			
	0.16	1571.375	5.264	2.823			
	0.18	1778.676	5.246	2.491			
	0.20	1734.860	5.228	2.233			
500	0.00	2924.362	5.452	6180.340			
	0.02	3012.862	5.432	24.879			
	0.04	3492.183	5.413	12.407			
	0.06	3170.741	5.393	8.026			
	0.08	3496.983	5.374	6.002			
	0.10	3423.280	5.355	4.677	0.06	[5.397, 5.508]	
	0.12	3416.526	5.336	3.897			
	0.14	3187.321	5.318	3.287			
	0.16	3284.556	5.300	2.823			
	0.18	3538.916	5.281	2.509			
	0.20	2845.976	5.263	2.255			

A.5 Numerical Illustration of DRBCP-D (2b) under  $\phi$ -divergence

Alternatively, we can consider DRBCP-D (2b) under a  $\phi$ -divergence ambiguity set  $\mathcal{P}$ . In particular, we construct the ambiguity set based on the total-variational distance with radius  $d \in [0, 2]$ . According to [25], given  $d \in [0, 2]$ , the distributionally robust fair matching problem in Example 6 under total variation based ambiguity set admits the following form

$$\begin{aligned}
v_{D\phi} = \min_{\mathbf{v}, \bar{\mathbf{v}}, \mathbf{z}, \beta, r} & (1 - d/2)\beta + \frac{1}{N} \sum_{k \in [N]} \bar{v}^k + dr/2 \\
\text{s.t. } & \bar{v}^k \geq v^k - \beta, \bar{v}^k \geq 0, \forall k \in [N] \\
& r \geq v^k, \forall k \in [N], \\
& v^k \geq \bar{c}_{ij}^k z_{ij}, \sum_{i \in [m]} z_{ij} = 1, \sum_{j \in [m]} z_{ij} = 1, z_{ij} \in \{0, 1\}, \forall i, j \in [m], \forall k \in [N], \\
& z_{ij} \in \{0, 1\} \forall i, j \in [m], \forall k \in [N].
\end{aligned} \tag{35}$$

The numerical results using the same datasets as those in Section 5.2 are displayed in Figure 2 and Table 4.

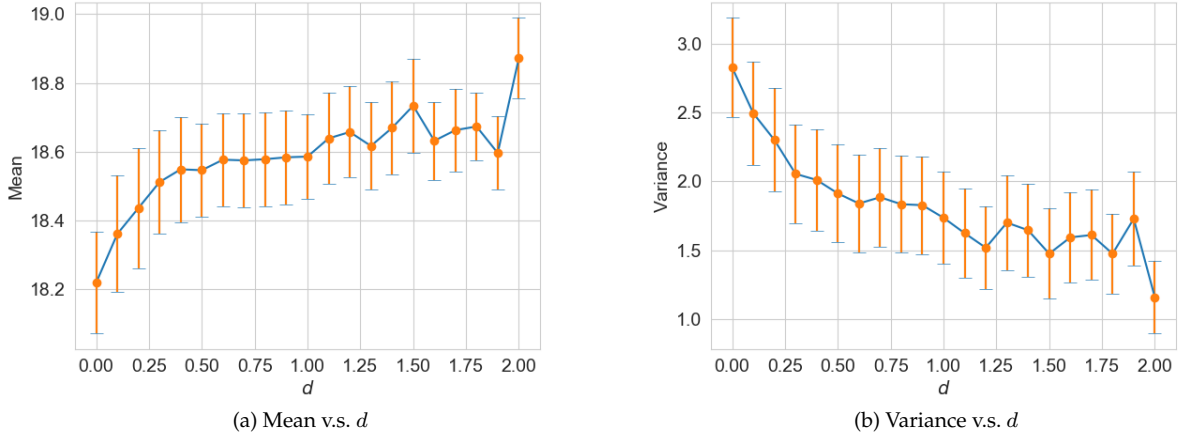


Fig. 2: Numerical Illustration of Fair Matching Problem under Total Variation-based Ambiguity Set (35): Using Real-world Dataset in Section 5.2

Table 4: Numerical Results of Fair Matching Problem under Total Variation-based Ambiguity Set (35): Using Hypothetical Dataset in Section 5.2

$N$	Case 1: $\mathcal{N}(\mu_{NY}, 10^2\sigma_{NY})$					Case 2: $\mathcal{N}(\mu_{NY}, 10^4\sigma_{NY})$						
	$d$	Time	$v_{D\phi}$	95% CI of Solutions	Theoretical 95% CI	$d$	Time	$v_{D\phi}$	95% CI of Solutions	Theoretical 95% CI		
10	0.0	0.678	40.800	$66.646 \pm 0.130$	$66.646 \pm 0.220$	0.0	0.571	242.350	$497.251 \pm 1.382$	$497.251 \pm 2.345$		
	0.2	0.579	45.790	$68.351 \pm 0.127$	$68.351 \pm 0.216$	0.2	0.363	297.520	$478.010 \pm 1.253$	$478.010 \pm 2.127$		
	0.4	0.394	49.070			0.4	0.375	337.720	$468.365 \pm 1.240$	$468.365 \pm 2.104$		
	0.6	0.501	52.330			0.6	0.551	369.560				
	0.8	0.735	55.180			0.8	0.308	397.080				
	1.0	0.350	57.340	$63.478 \pm 0.130$	$63.478 \pm 0.220$	1.0	0.310	417.970	$467.371 \pm 1.276$	$467.371 \pm 2.166$		
	1.2	0.327	59.350	<b><math>63.362 \pm 0.131</math></b>	$63.362 \pm 0.222$	1.2	0.332	434.730				
	1.4	0.339	61.200	$63.478 \pm 0.130$	$63.478 \pm 0.220$	1.4	0.334	447.180				
	1.6	0.281	62.590			1.6	0.264	453.730				
	1.8	0.050	63.400	$66.635 \pm 0.127$	$66.635 \pm 0.215$	1.8	0.079	458.400			$505.127 \pm 1.355$	$505.127 \pm 2.300$
2.0	0.027	63.400	$63.362 \pm 0.131$	$63.362 \pm 0.222$	2.0	0.113	458.400	$467.371 \pm 1.276$			$467.371 \pm 2.166$	
20	0.0	1.606	45.930	$64.258 \pm 0.165$	$64.258 \pm 0.280$	0.0	1.685	304.910			<b><math>427.099 \pm 1.149</math></b>	$427.099 \pm 1.950$
	0.2	1.1906	49.520			0.2	2.357	352.965				
	0.4	0.920	52.950			0.4	0.952	393.750				
	0.6	1.082	55.600			0.6	1.324	428.775				
	0.8	0.800	57.875			0.8	0.913	461.800				
	1.0	1.018	60.035			1.0	0.784	487.835				
	1.2	0.750	61.890	$61.118 \pm 0.159$	$61.118 \pm 0.271$	1.2	0.548	502.480	$501.812 \pm 1.662$	$501.812 \pm 2.821$		
	1.4	0.865	63.420	<b><math>59.228 \pm 0.118</math></b>	$59.228 \pm 0.201$	1.4	0.491	513.220	$431.159 \pm 1.127$	$431.159 \pm 1.912$		
	1.6	0.574	64.470			1.6	0.521	519.565				
	1.8	0.319	64.825			1.8	0.348	522.805				
2.0	0.034	64.900	2.0			0.036	523.000					
40	0.0	4.829	49.4825	$61.118 \pm 0.159$	$61.118 \pm 0.271$	0.0	4.217	342.862	$412.923 \pm 1.086$	$412.923 \pm 1.843$		
	0.2	4.427	55.518			0.2	2.897	405.390	$420.407 \pm 1.133$	$420.407 \pm 1.922$		
	0.4	3.860	61.025			0.4	1.636	455.753	<b><math>397.311 \pm 1.007</math></b>	$397.311 \pm 2.493$		
	0.6	2.379	65.990			0.6	1.661	493.1175	$415.991 \pm 1.084$	$415.991 \pm 1.840$		
	0.8	2.576	70.675			0.8	1.193	526.960				
	1.0	1.604	75.030			1.0	1.239	556.585				
	1.2	2.050	78.975	1.2	1.256	582.030						
	1.4	3.237	81.685	<b><math>56.372 \pm 0.103</math></b>	$56.372 \pm 0.174$	1.4	1.189	600.633	$466.052 \pm 1.262$	$466.052 \pm 2.143$		
	1.6	1.553	83.970	1.6	0.954	612.990						
	1.8	1.189	85.838	$63.167 \pm 0.130$	$63.167 \pm 0.221$	1.8	0.622	621.578				
2.0	0.062	86.300	$60.244 \pm 0.118$	$60.244 \pm 0.201$	2.0	0.044	626.500	$429.502 \pm 1.092$			$429.502 \pm 1.854$	