A Probabilistic Model for Minmax Regret in Combinatorial Optimization

Karthik Natarajan* Dongjian Shi† Kim-Chuan Toh‡

May 21, 2013

Abstract

In this paper, we propose a new probabilistic model for minimizing the anticipated regret in combinatorial optimization problems with distributional uncertainty in the objective coefficients. The interval uncertainty representation of data is supplemented with information on the marginal distributions. As a decision criterion, we minimize the worst-case conditional value-at-risk of regret. The proposed model includes the interval data minmax regret as a special case. For the class of combinatorial optimization problems with a compact convex hull representation, a polynomial sized mixed integer linear program (MILP) is formulated when (a) the range and mean are known, and (b) the range, mean and mean absolute deviation are known while a mixed integer second order cone program (MISOCP) is formulated when (c) the range, mean and standard deviation are known. For the subset selection problem of choosing K elements of maximum total weight out of a set of N elements, the probabilistic regret model is shown to be solvable in polynomial time in the instances (a) and (b) above. This extends the current known polynomial complexity result for minmax regret subset selection with range information only.

*Engineering Systems and Design, Singapore University of Technology and Design, 20 Dover Drive, Singapore 138682. Email: natarajan_karthik@sutd.edu.sg

†Department of Mathematics, National University of Singapore, 10, Lower Kent Ridge Road, Singapore 119076. Email: shidongjian@nus.edu.sg

‡Department of Mathematics, National University of Singapore, 10, Lower Kent Ridge Road, Singapore 119076. Email: mattohkc@nus.edu.sg

1 Minmax Regret Combinatorial Optimization

Let $Z(\boldsymbol{c})$ denote the optimal value to a linear combinatorial optimization problem over a feasible region $\mathcal{X} \subseteq \{0,1\}^N$ for the objective coefficient vector $\boldsymbol{c}$:

$$Z(\boldsymbol{c}) = \max \left\{ \boldsymbol{c}^T \boldsymbol{y} \mid \boldsymbol{y} \in \mathcal{X} \subseteq \{0,1\}^N \right\}. \tag{1.1}$$

Consider a decision-maker who needs to decide on a solution $\boldsymbol{x} \in \mathcal{X}$ before knowing the actual value of the objective coefficients. Let Ω represent a deterministic uncertainty set that captures all the possible realizations of the vector $\boldsymbol{c}$. Under the regret criterion, the decision-maker experiences an ex-post regret of possibly not choosing the optimal solution. The value of regret in absolute terms is given by:

$$R(\boldsymbol{x}, \boldsymbol{c}) = Z(\boldsymbol{c}) - \boldsymbol{c}^T \boldsymbol{x}, \tag{1.2}$$

where $R(\boldsymbol{x}, \boldsymbol{c}) \geq 0$. The maximum value of regret for a decision $\boldsymbol{x}$ corresponding to the uncertainty set Ω is given as:

$$\max_{\boldsymbol{c} \in \Omega} R(\boldsymbol{x}, \boldsymbol{c}). \tag{1.3}$$

Savage [39] proposed the use of the following minimax regret model, where the decision $\boldsymbol{x}$ is chosen to minimize the maximum regret over all possible realizations of the uncertainty:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{c} \in \Omega} R(\boldsymbol{x}, \boldsymbol{c}). \tag{1.4}$$

One of the early references on minmax regret models in combinatorial optimization is the work of Kouvelis and Yu [27]. The computational complexity of solving the minmax regret problem have been extensively studied therein under the following two representations of Ω :

(a) Scenario uncertainty: The vector $\boldsymbol{c}$ lies in a finite set of M possible discrete scenarios:

$$\Omega = \{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_M}\}.$$

(b) Interval uncertainty: Each component c_i of the vector $\boldsymbol{c}$ takes a value between a lower bound $\underline{c}_i$ and upper bound $\overline{c}_i$. Let $\Omega_i = [\underline{c}_i, \overline{c}_i]$ for $i = 1, \ldots, N$. The uncertainty set is the Cartesian product of the sets of intervals:

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_N.$$

For the discrete scenario uncertainty, the minmax regret counterpart of problems such as the shortest path, minimum assignment and minimum spanning tree problems are NP-hard even when the scenario set contains only two scenarios (see Kouvelis and Yu [27]). This indicates the difficulty of solving regret problems to optimality since the original deterministic optimization problems are solvable in polynomial time in these instances. These problems are weakly NP-hard for a constant number of scenarios while they become strongly NP-hard when the number of scenarios is non-constant.

In the interval uncertainty case, for any $\boldsymbol{x} \in \mathcal{X}$, let $S_{\boldsymbol{x}}^{+}$ denote the scenario in which $c_i = \overline{c}_i$ if $x_i = 0$, and $c_i = \underline{c}_i$ if $x_i = 1$. It is straightforward to see that the scenario $S_{\boldsymbol{x}}^{+}$ is the worst-case scenario that maximizes the regret in (1.3) for a fixed $\boldsymbol{x} \in \mathcal{X}$. For deterministic combinatorial optimization problems with a compact convex hull representation, this worst-case scenario can be used to develop compact MILP formulations for the minmax regret problem (1.4) (refer to Yaman et. al. [44] and Kasperski [24]). As in the scenario uncertainty case, the minmax regret counterpart is NP-hard under interval uncertainty for most classical polynomial time solvable combinatorial optimization problems. Averbakh and Lebedev [6] proved that the minmax regret shortest path and minmax regret minimum spanning tree problems are strongly NP-hard with interval uncertainty. Under the assumption that the deterministic problem is polynomial time solvable, a 2-approximation algorithm for minmax regret was designed by Kasperski and Zieliński [25]. Their algorithm is based on a mid-point scenario approach where the deterministic combinatorial optimization problem is solved with an objective coefficient vector $(\underline{\boldsymbol{c}} + \overline{\boldsymbol{c}})/2$. Kasperski and Zieliński [26] developed a fully polynomial time approximation scheme under the assumption that a pseudopolynomial algorithm is available for the deterministic problem. A special case where the minmax regret problem is solvable in polynomial time is the subset selection problem. The deterministic subset selection problem is: Given a set of elements $[N] := \{1, \ldots, N\}$ with weights $\{c_1, \ldots, c_N\}$, select a subset of K elements of maximum total weight. The deterministic problem can be solved by a simple sorting algorithm. With an interval uncertainty representation of the weights, Averbakh [5] designed a polynomial time algorithm to solve the minmax regret problem to optimality with a running time of $O(N \min(K, N-K)^2)$. Subsequently, Conde [14] designed a faster algorithm to solve this problem with running time $O(N \min(K, N-K))$.

A related model that has been analyzed in discrete optimization is the absolute robust approach (see Kouvelis and Yu [27] and Bertsimas and Sim [11]) where the decision-maker chooses a decision $\boldsymbol{x}$ that maximizes the minimum objective over all possible realizations of the uncertainty:

$$\max_{\boldsymbol{x} \in \mathcal{X}} \min_{\boldsymbol{c} \in \Omega} \boldsymbol{c}^T \boldsymbol{x}. \tag{1.5}$$

Problem (1.5) is referred to as the absolute robust counterpart of the deterministic optimization problem. The formulation for the absolute robust counterpart should be contrasted with the minmax regret formulation which can be viewed as the relative robust counterpart of the deterministic optimization problem. For the discrete scenario uncertainty, the absolute robust counterpart of the shortest path problem is NP-hard as in the regret setting (see Kouvelis and Yu [27]). However for the interval uncertainty case, the absolute robust counterpart retains the complexity of the deterministic problem unlike the minmax regret counterpart. This follows from the observation that the worst case realization of the uncertainty in absolute terms is to set the objective coefficient vector to the lower bound $\underline{\boldsymbol{c}}$ irrespective of the solution $\boldsymbol{x}$. The minmax regret version in contrast is more difficult to solve since the worst case realization depends on the solution $\boldsymbol{x}$. However this also implies that the minmax regret solution is less conservative as it considers both the best and worst case. For illustration, consider the binary decision problem of deciding whether to invest or not in a single project with payoff c:

$$Z(c) = \max \{cy \mid y \in \{0, 1\}\}.$$

The payoff is uncertain and takes a value in the range $c \in [\underline{c}, \overline{c}]$ where $\underline{c} < 0$ and $\overline{c} > 0$. The absolute robust solution is to not invest in the project since in the worst case the payoff is negative. On the other hand, the minmax regret solution is to invest in the project if $\overline{c} > -\underline{c}$ (the best payoff is more than the magnitude of the worst loss) and not invest in the project otherwise. Since the regret criterion evaluates the performance with respect to the best decision, it is not as conservative as the absolute robust solution. However the computation of the minmax regret solution is more difficult than the absolute robust solution. The minmax regret models handle support information and assumes that the decision-maker uses the worst scenario (in terms of regret) to make the decision. However if additional probabilistic information is known or can be estimated from data, it is natural to incorporate this information into the regret model. To quantify the impact of probabilistic information on regret, consider the graph in Figure 1. In this graph, there are three paths connecting node A to node D: $1-4$, $2-5$ and $1-3-5$. Consider a decision-maker who wants to go from node A to node D in the shortest possible time by choosing among the three paths. The mean μ_i and range $[\underline{c}_i, \overline{c}_i]$ for each edge i in Figure 1 denotes the average time and the range of possible times in hours to traverse the edge. The comparison of the different paths are shown in the following table:

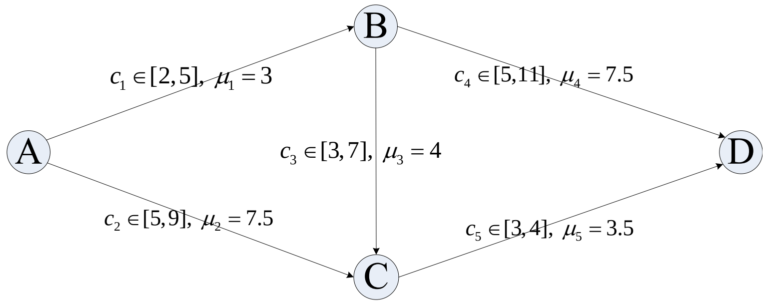


Figure 1: Find a Shortest Path from Node A to Node D

Table 1: Comparison of paths

Criterion	Regret			Absolute Robust		Average
Path	$(c_1, c_2, c_3, c_4, c_5)$	Best Path	Max Regret	$(c_1, c_2, c_3, c_4, c_5)$	Max Time	Expected Time
$1-4$	$(5, 5, 3, 11, 3)$	$2-5$	8	$(5, 9, 7, 11, 4)$	16	10.5
$1-3-5$	$(5, 5, 7, 5, 4)$	$2-5$	7	$(5, 9, 7, 11, 4)$	16	10.5
$2-5$	$(2, 9, 3, 5, 4)$	$1-4$	6	$(5, 9, 7, 11, 4)$	13	11

In the minmax regret model, the optimal decision is the path $2-5$ with regret of 6 hours. However, on average this path takes 0.5 hours more than the other two paths. In terms of expected cost, the optimal decision is either of the paths $1-4$ or $1-3-5$. Note that only the range information is used in the minmax regret model, and only mean information is used to minimize the expected cost. Clearly, the choice of an "optimal" path is based on the decision criterion and the available data that guides the decision process. In this paper, we propose an integer programming approach for probabilistic regret in combinatorial optimization that incorporates partial distributional information such as the mean and variability of the random coefficients and provides flexibility in modeling the decision-maker's aversion to regret.

The structure and the contributions of the paper are summarized next:

1. In Section 2, a new probabilistic model for minmax regret in combinatorial optimization is proposed. As a decision criterion, the worst-case conditional value-at-risk of cost has been previously studied. We extend this to the concept of regret which can be viewed as a relative cost with respect to the best solution. The proposed model incorporates limited probabilistic information on the uncertainty such as the knowledge of the mean, mean absolute deviation or standard deviation while also providing flexibility to model the decision-maker's attitude to regret. In the special

case, the probabilistic regret criterion reduces to the traditional minmax regret criterion and the expected objective criterion respectively.

2. In Section 3, we develop a tractable formulation to compute the worst-case conditional value-at-risk of regret for a fixed solution $\boldsymbol{x} \in \mathcal{X}$. The worst-case conditional value-at-risk of regret is shown to be computable in polynomial time if the deterministic optimization problem is solvable in polynomial time. This generalizes the current result for the interval uncertainty model, where the worst-case regret for a fixed solution $\boldsymbol{x} \in \mathcal{X}$ is known to be computable in polynomial time when the deterministic optimization problem is solvable in polynomial time.

3. In Section 4, we formulate conic mixed integer programs to solve the probabilistic regret model. For the class of combinatorial optimization problems with a compact convex hull representation, a polynomial sized MILP is developed when (a) range and mean are given, and (b) range, mean and mean absolute deviation are given. If (c) range, mean and standard deviation are given, we develop a polynomial sized MISOCP. The formulations in this section generalizes the polynomial sized MILP formulation to solve the minmax regret model when only the the range is given.

4. In Section 5, we provide a polynomial time algorithm to solve the probabilistic regret counterpart for subset selection when (a) range and mean, and (b) range, mean and mean absolute deviation are given. This extends the current results of Averbakh [5] and Conde [14] who proved the polynomial complexity of the minmax regret counterpart of subsect selection which uses range information only. Together, Sections 3, 4 and 5 generalize several of the results for the interval uncertainty regret model to the probabilistic regret model.

5. In Section 6, numerical examples for the shortest path and subset selection problems are provided. The numerical results provide illustration on the structure of the solutions generated by the probabilistic regret model while identifying the computational improvement obtained by the polynomial time algorithm proposed in Section 5 relative to a generic MILP formulation.

2 Worst-case Conditional Value-at-risk of Regret

Let $\tilde{\boldsymbol{c}}$ denote the random objective coefficient vector with a probability distribution P that is itself unknown. P is assumed to lie in the set of distributions $\mathbb{P}(\Omega)$ where Ω is the support of the random vector. In the simplest model, the decision-maker minimizes the anticipated regret in an expected sense:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{P\in\mathbb{P}(\Omega)} \mathbb{E}_P[R(\boldsymbol{x},\tilde{\boldsymbol{c}})]. \tag{2.1}$$

Model (2.1) includes two important subcases: (a) $\mathbb{P}(\Omega)$ is the set of all probability distributions with support Ω. In this case (2.1) reduces to the standard minmax regret model (1.4) and (b) The complete distribution is given with $\mathbb{P} = \{P\}$. In this case (2.1) reduces to solving the deterministic optimization problem where the random objective is replaced with the mean vector $\boldsymbol{\mu}$, since

$$\operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_P[Z(\tilde{\boldsymbol{c}}) - \tilde{\boldsymbol{c}}^T\boldsymbol{x}] = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left(\mathbb{E}_P[Z(\tilde{\boldsymbol{c}})] - \boldsymbol{\mu}^T\boldsymbol{x}\right) = \operatorname*{argmax}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\mu}^T\boldsymbol{x}.$$

Formulation (2.1) however does not capture the degree of regret aversion. Furthermore, as long as the mean vector is fixed, the optimal decision in (2.1) is the optimal solution to the deterministic problem with the mean objective. Thus the solution is insensitive to other distributional information such as variability. To address this, we propose use of the conditional value-at-risk measure that has been gaining popularity in the risk management literature.

2.1 Conditional Value-at-risk Measure

Conditional value-at-risk is also referred to as average value-at-risk or expected shortfall in the risk management literature. We briefly review this concept here. Consider a random variable $\tilde{r}$ defined on a probability space $(\Pi, \mathcal{F}, Q)$, i.e. a real valued function $\tilde{r}(\omega) : \Pi \to \Re$ with finite second moment $\mathbb{E}[\tilde{r}^2] < \infty$. This ensures that the conditional-value-at-risk is finite. For the random variables that we consider in this paper, the finiteness of the second moment is guaranteed as the random variables are assumed to lie within a finite range. For a given $\alpha \in (0,1)$, the value-at-risk is defined as the lower α quantile of the random variable $\tilde{r}$:

$$\mathrm{VaR}_\alpha(\tilde{r}) = \inf\{v \mid Q(\tilde{r} \le v) \ge \alpha\}. \tag{2.2}$$

The definition of conditional value-at-risk is provided next.

Definition 1 (Rockafellar and Uryasev [36, 37], Acerbi and Tasche [1])**.** *For $\alpha \in (0,1)$, the conditional value-at-risk* (CVaR) *at level α of a random variable $\tilde{r}(\omega) : \Pi \to \Re$ is the average of the highest $1-\alpha$ of the outcomes:*

$$\mathrm{CVaR}_\alpha(\tilde{r}) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta(\tilde{r})d\beta. \tag{2.3}$$

An equivalent representation for CVaR *is:*

$$\mathrm{CVaR}_\alpha(\tilde{r}) = \inf_{v \in \Re} \left(v + \frac{1}{1-\alpha} \mathbb{E}_Q[\tilde{r} - v]^+ \right). \tag{2.4}$$

From an axiomatic perspective, CVaR is an example of a coherent risk measure (see Artzner et. al. [4], Föllmer and Schied [16] and Fritelli and Gianin [17]) and satisfies the following four properties:

1. Monotonicity: If $\tilde{r}_1(\omega) \geq \tilde{r}_2(\omega)$ for each outcome, then $\mathrm{CVaR}_\alpha(\tilde{r}_1) \geq \mathrm{CVaR}_\alpha(\tilde{r}_2)$.

2. Translation invariance: If $c \in \Re$, then $\mathrm{CVaR}_\alpha(\tilde{r}_1 + c) = \mathrm{CVaR}_\alpha(\tilde{r}_1) + c$.

3. Convexity: If $\lambda \in [0, 1]$, then $\mathrm{CVaR}_\alpha(\lambda\tilde{r}_1 + (1-\lambda)\tilde{r}_2) \leq \lambda \mathrm{CVaR}_\alpha(\tilde{r}_1) + (1-\lambda)\mathrm{CVaR}_\alpha(\tilde{r}_2)$.

4. Positive homogeneity: If $\lambda \geq 0$, then $\mathrm{CVaR}_\alpha(\lambda\tilde{r}_1) = \lambda \mathrm{CVaR}_\alpha(\tilde{r}_1)$.

Furthermore, CVaR is an attractive risk measure for stochastic optimization since it is convexity preserving unlike the VaR measure. However the computation of CVaR might still be intractable (see Ben-Tal et. al. [7] for a detailed discussion on this). An instance when the computation of CVaR is tractable is for discrete distributions with a polynomial number of scenarios. Optimization with the CVaR measure has been used in portfolio optimization [36] and inventory control [3] among other stochastic optimization problems. Combinatorial optimization problems under the CVaR measure has been studied by So et. al. [42]:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \mathrm{CVaR}_\alpha \left(-\tilde{\boldsymbol{c}}^T \boldsymbol{x} \right). \tag{2.5}$$

The negative sign in Formulation (2.5) capture the feature that higher values of $\boldsymbol{c}^T\boldsymbol{x}$ are preferred to lower values. Formulation (2.5) can be viewed as regret minimization problem where the regret is defined with respect to an absolute benchmark of zero. Using a sample average approximation method, So et. al. [42] propose approximation algorithms to solve (2.5) for covering, facility location and Steiner tree problems. In the distributional uncertainty representation, the concept of conditional value-at-risk is extended to the concept of worst-case conditional value-at-risk through the following definition.

Definition 2 (Zhu and Fukushima [46], Natarajan et. al. [31])**.** *Suppose the distribution of the random variable $\tilde{r}$ lies in a set $\mathbb{Q}$. For $\alpha \in (0, 1)$, the worst-case conditional value-at-risk* (WCVaR) *at level α of a random variable $\tilde{r}$ with respect to $\mathbb{Q}$ is defined as:*

$$\mathrm{WCVaR}_\alpha(\tilde{r}) = \sup_{Q \in \mathbb{Q}} \inf_{v \in \Re} \left(v + \frac{1}{1-\alpha} \mathbb{E}_Q[\tilde{r} - v]^+ \right). \tag{2.6}$$

From an axiomatic perspective, WCVaR has also be shown to be a coherent risk measure under mild assumptions on the set of distributions (see the discussions in Zhu and Fukushima [46] and Natarajan et. al. [31]). WCVaR has been used as a risk measure in distributional robust portfolio optimization [46, 31] and joint chance constrained optimization problems [13, 48]. Zhu and Fukushima [46] and Natarajan et. al. [31] also provide examples of sets of distributions $\mathbb{Q}$ where the position of sup and inf can be exchanged in formula (2.6). Since the objective is linear in the probability measure (possibly infinite-dimensional) over which it is maximized and convex in the variable v over which it is minimized, the saddle point theorem from Rockafellar [38] is applicable. Applying Theorem 6 in [38] implies the following lemma:

Lemma 1. *Let $\alpha \in (0,1)$, and the distribution of the random variable $\tilde{r}$ lies in a set $\mathbb{Q}$. If $\mathbb{Q}$ is a convex set of the probability distributions defined on a closed convex support set $\Omega \subseteq \Re^n$, then*

$$\mathrm{WCVaR}_\alpha(\tilde{r}) = \inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}\sup_{Q\in\mathbb{Q}} \mathbb{E}_Q[\tilde{r}-v]^+\right). \tag{2.7}$$

For all sets of distributions $\mathbb{P}(\Omega)$ that are studied in this paper, the condition in Lemma 1 is satisfied. We propose the use of worst-case conditional value-at-risk of regret as a decision criterion in combinatorial optimization problems. The central problem of interest to solve is:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}})) \quad = \quad \min_{\boldsymbol{x}\in\mathcal{X}}\inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}\sup_{P\in\mathbb{P}(\Omega)} \mathbb{E}_P\left[R(\boldsymbol{x},\tilde{\boldsymbol{c}})-v\right]^+\right). \tag{2.8}$$

2.2 Marginal Distribution and Marginal Moment Models

To generalize the interval uncertainty model supplemental marginal distributional information of the random vector $\tilde{\boldsymbol{c}}$ is assumed to be given. The random variables are however not assumed to be independent. The following two models are considered:

(a) Marginal distribution model: For each $i \in [N]$, the marginal probability distribution P_i of $\tilde{c}_i$ with support $\Omega_i = [\underline{c}_i, \overline{c}_i]$ is assumed to be given. Let $\mathbb{P}(P_1,\ldots,P_N)$ denote the set of joint distributions with the fixed marginals. This is commonly referred to as the Fréchet class of distributions.

(b) Marginal moment model: For each $i \in [N]$, the probability distribution P_i of $\tilde{c}_i$ with support $\Omega_i = [\underline{c}_i, \overline{c}_i]$ is assumed to belong to a set of probability measures $\mathbb{P}_i$. The set $\mathbb{P}_i$ is defined through moment equality constraints on real-valued functions of the form $\mathbb{E}_{P_i}[f_{ik}(\tilde{c}_i)] = m_{ik}, k \in [K_i]$. If $f_{ik}(c_i) = c_i^k$, this reduces to knowing the first K_i moments of $\tilde{c}_i$. Let $\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)$ denote the set of multivariate joint distributions compatible with the marginal probability distributions

$P_i \in \mathbb{P}_i$. Throughout the paper, we assume that mild Slater type conditions hold on the moment information to guarantee that strong duality is applicable for moment problems. One such simple sufficient condition is that the moment vector is in the interior of the set of feasible moments (see Isii [22]). With the marginal moment specification, the multivariate moment space is the product of univariate moment spaces. Ensuring that Slater type conditions hold in this case is relatively straightforward since it reduces to Slater conditions for univariate moment spaces. The reader is referred to Bertsimas et. al. [9] and Lasserre [28] for a detailed description on this topic.

The moment representation of uncertainty in distributions has been used in the minmax regret newsvendor problem [45, 34]. A newsvendor needs to choose an order quantity q of a product before the exact value of demand is known by balancing the costs of under-ordering and over-ordering. The random demand is represented by $\tilde{d}$ with a probability distribution P. The unit selling price is p, the unit cost is c and the salvage value for any unsold product is 0. A risk neutral firm chooses its quantity to maximize its expected profit:

$$\max_{q \geq 0} \left(p\mathbb{E}_P[\min(q, \tilde{d})] - cq \right),$$

where $\min(q, \tilde{d})$ is the actual quantity of units sold which depends on the demand realization. In the minmax regret version of this problem studied in [45, 34], the newsvendor chooses the order quantity where the demand distribution is not exactly known. The demand distribution is assumed to belong to a set of probability measure $P \in \mathbb{P}$ typically characterized with moment information. The objective is to minimize the maximum loss in profit from not knowing the full distribution:

$$\min_{q \geq 0} \max_{P \in \mathbb{P}} \left[\max_{s \geq 0} \left(p\mathbb{E}_P[\min(s, \tilde{d})] - cs \right) - \left(p\mathbb{E}_P[\min(q, \tilde{d})] - cq \right) \right].$$

Yue et. al. [45] solved this model analytically where only the mean and variance of demand are known. Roels and Perakis [34] generalized this model to incorporate additional moments and information on the shape of the demand. On the other hand, if the demand is known with certainty, the optimal order quantity is exactly the demand. The maximum profit would be $(p - c)\tilde{d}$ and the regret model as proposed in this paper is:

$$\min_{q \geq 0} \inf_{v \in \Re} \left(v + \frac{1}{1 - \alpha} \sup_{P \in \mathbb{P}(\Omega)} \mathbb{E}_P \left[(p - c)\tilde{d} - \left(p \min(q, \tilde{d}) - cq \right) - v \right]^+ \right),$$

where α is the parameter that captures aversion to regret. There are two major differences between the minmax regret newsvendor model in [45, 34] and the regret model proposed in this paper. The first difference is that in [45, 34] the newsvendor minimizes the maximum ex-ante regret (with respect to distributions) of not knowing the right distribution, while in this paper, the decision-maker minimizes the ex-post regret (with respect to cost coefficient realizations) of not knowing the right objective coefficients. The second difference is that the newsvendor problem deals with a single demand variable. However in the multi-dimensional case, the marginal model forms the natural extension and is a more tractable formulation.

The new probabilistic regret model can be related to standard minmax regret. In the marginal moment model, if only the range information of each random variable $\tilde{c}_i$ is given, then the WCVaR of regret reduces to the maximum regret. Consider the random vector whose distribution is a Dirac measure $\delta_{\hat{\mathbf{c}}(\boldsymbol{x})}$ with $\hat{c}_i(\boldsymbol{x}) = \overline{c}_i(1 - x_i) + \underline{c}_i x_i$ for $i \in [N]$. Then WCVaR of the regret satisfies:

$$\begin{aligned}
\inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}\sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)}\mathbb{E}_P[R(\boldsymbol{x},\tilde{\boldsymbol{c}}) - v]^+\right) &\geq \inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}\mathbb{E}_{\delta_{\hat{\mathbf{c}}}}[R(\boldsymbol{x},\tilde{\boldsymbol{c}}) - v]^+\right) \\
&= \inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}[R(\boldsymbol{x},\hat{\boldsymbol{c}}) - v]^+\right) \\
&= R(\boldsymbol{x},\hat{\boldsymbol{c}}) \\
&= \max_{\boldsymbol{c}\in\Omega} R(\boldsymbol{x},\boldsymbol{c}).
\end{aligned}$$

The last equality is valid since $\hat{\boldsymbol{c}}(\boldsymbol{x})$ is the worst-case scenario for a given $\boldsymbol{x} \in \mathcal{X}$. Moreover, the WCVaR of the regret cannot be larger than the maximum value of regret. Hence, they are equal in this case. When $\alpha = 0$, problem (2.8) reduces to minimizing the worst-case expected regret,

$$\min_{\boldsymbol{x}\in\mathcal{X}}\ \sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)}\mathbb{E}_P[R(\boldsymbol{x},\tilde{\boldsymbol{c}})].$$

On the other hand, as α converges to 1, $\mathrm{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}}))$ converges to $\max_{\boldsymbol{c}\in\Omega} R(\boldsymbol{x},\boldsymbol{c})$, and problem (2.8) reduces to the traditional interval uncertainty minmax regret model. This implies that the problem of minimizing the WCVaR of the regret in this probabilistic model is NP-hard since the minmax regret problem is NP-hard [6]. The parameter α allows for the flexibility to vary the degree of regret aversion.

If a decision $\boldsymbol{x}_1$ is preferred to decision $\boldsymbol{x}_2$ for each realization of the uncertainty, it is natural to conjecture that $\boldsymbol{x}_1$ is preferred to $\boldsymbol{x}_2$ in the regret model. The following lemma validates this monotonicity property for the chosen criterion.

Lemma 2. *For two decisions $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$, if $\boldsymbol{x}_1$ dominates $\boldsymbol{x}_2$ in each realization of the uncertainty, i.e. $\boldsymbol{c}^T\boldsymbol{x}_1 \geq \boldsymbol{c}^T\boldsymbol{x}_2$ for all $\boldsymbol{c} \in \Omega$, then the decision $\boldsymbol{x}_1$ is preferred to $\boldsymbol{x}_2$, i.e.* $\mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})) \leq \mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}}))$.

Proof. Since $\boldsymbol{c}^T\boldsymbol{x}_1 \geq \boldsymbol{c}^T\boldsymbol{x}_2$ for all $\boldsymbol{c} \in \Omega$,

$$R(\boldsymbol{x}_1,\boldsymbol{c}) = \max_{\boldsymbol{y}\in\mathcal{X}}\boldsymbol{c}^T\boldsymbol{y} - \boldsymbol{c}^T\boldsymbol{x}_1 \leq \max_{\boldsymbol{y}\in\mathcal{X}}\boldsymbol{c}^T\boldsymbol{y} - \boldsymbol{c}^T\boldsymbol{x}_2 = R(\boldsymbol{x}_2,\boldsymbol{c}), \quad \forall\boldsymbol{c}\in\Omega.$$

Thus $[R(\boldsymbol{x}_1,\boldsymbol{c})-v]^+ \leq [R(\boldsymbol{x}_2,\boldsymbol{c})-v]^+, \ \ \forall \boldsymbol{c}\in\Omega, v\in\Re$. Hence for any distribution $P\in\mathbb{P}$, $\mathbb{E}_P[R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})-v]^+ \leq \mathbb{E}_P[R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}})-v]^+$. This implies that $\sup_{P\in\mathbb{P}}\mathbb{E}_P[R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})-v]^+ \leq \sup_{P\in\mathbb{P}}\mathbb{E}_P[R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}})-v]^+$. Therefore,

$$\inf_{v\in\Re}\left(v+\frac{1}{1-\alpha}\sup_{P\in\mathbb{P}}\mathbb{E}_P[R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})-v]^+\right) \leq \inf_{v\in\Re}\left(v+\frac{1}{1-\alpha}\sup_{P\in\mathbb{P}}\mathbb{E}_P[R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}})-v]^+\right),$$

that is $\text{WCVaR}_\alpha(R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})) \leq \text{WCVaR}_\alpha(R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}}))$. □

3 Computation of the WCVaR of Regret and Cost

In this section, we compute the WCVaR of regret and cost for a fixed $\boldsymbol{x}\in\mathcal{X}$ in the marginal distribution and marginal moment model. This is motivated by bounds in PERT networks that were proposed by Meilijson and Nadas [30] and later extended in the works of Klein Haneveld [21], Weiss [43], Birge and Maddox [12] and Bertsimas et. al. [9]. In a PERT network, let $[N]$ represent the set of activities. Each activity $i\in[N]$ is associated with a random activity time $\tilde{c}_i$ and marginal distribution P_i. Meilijson and Nadas [30] computed the worst-case expected project tardiness $\sup_{P\in\mathbb{P}(P_1,\ldots,P_N)}\mathbb{E}_P[Z(\tilde{\boldsymbol{c}})-v]^+$ where $Z(\boldsymbol{c})$ denotes the time to complete the project and v denotes a deadline for the project. Their approach can be summarized as follows. For all $\boldsymbol{d}\in\Re^N$ and $\boldsymbol{c}\in\Omega$:

$$\begin{aligned}
[Z(\boldsymbol{c})-v]^+ &= \left[\max_{\boldsymbol{y}\in\mathcal{X}}(\boldsymbol{d}+\boldsymbol{c}-\boldsymbol{d})^T\boldsymbol{y}-v\right]^+ \\
&\leq \left[\max_{\boldsymbol{y}\in\mathcal{X}}\boldsymbol{d}^T\boldsymbol{y}-v\right]^+ + \left[\max_{\boldsymbol{y}\in\mathcal{X}}(\boldsymbol{c}-\boldsymbol{d})^T\boldsymbol{y}\right]^+ \\
&\leq [Z(\boldsymbol{d})-v]^+ + \sum_{i=1}^N [c_i-d_i]^+.
\end{aligned}$$

Taking expectation with respect to a distribution $P\in\mathbb{P}(P_1,\ldots,P_N)$ and minimizing over $\boldsymbol{d}\in\Re^N$ gives the bound:

$$\mathbb{E}_P[Z(\tilde{\boldsymbol{c}})-v]^+ \leq \inf_{\boldsymbol{d}\in\Re^N}\left([Z(\boldsymbol{d})-v]^+ + \sum_{i=1}^N\mathbb{E}_{P_i}[\tilde{c}_i-d_i]^+\right), \quad \forall P\in\mathbb{P}(P_1,\ldots,P_N).$$

Meilijson and Nadas [30] constructed a multivariate probability distribution that is consistent with the marginal distributions such that the upper bound is attained. This leads to their main observation that the worst-case expected project tardiness is obtained by solving the following convex minimization problem:

$$\sup_{P\in\mathbb{P}(P_1,\ldots,P_N)}\mathbb{E}_P[Z(\tilde{\boldsymbol{c}})-v]^+ = \inf_{\boldsymbol{d}\in\Re^N}\left([Z(\boldsymbol{d})-v]^+ + \sum_{i=1}^N\mathbb{E}_{P_i}[\tilde{c}_i-d_i]^+\right). \tag{3.1}$$

With partial marginal distribution information, Klein Haneveld [21], Birge and Maddox [12] and Bertsimas et al. [9] extended the convex formulation of the worst-case expected project tardiness to:

$$\sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_P[Z(\tilde{\boldsymbol{c}})-v]^+ = \inf_{\boldsymbol{d}\in\Re^N}\Big([Z(\boldsymbol{d})-v]^+ + \sum_{i=1}^N \sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}[\tilde{c}_i-d_i]^+\Big). \tag{3.2}$$

Klein Haneveld [21] estimated a project deadline v that balances the expected project tardiness with respect to the most unfavorable distribution and the cost of choosing the deadline for the project. This can be formulated as a two stage recourse problem:

$$\inf_{v\in\Re}\left(v+\frac{1}{1-\alpha}\sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)}\mathbb{E}_P\left[Z(\tilde{\boldsymbol{c}})-v\right]^+\right). \tag{3.3}$$

where $\alpha\in(0,1)$ is the tradeoff parameter between the two costs. Formulation (3.3) is clearly equivalent to estimating the worst-case conditional value-at-risk of the project completion time. We extend these results to the regret framework in the following subsection.

3.1 Worst-case Conditional Value-at-risk of Regret

To compute the WCVaR of regret, we first consider the subproblem $\sup_{P\in\mathbb{P}(\Omega)}\mathbb{E}_P\left[Z(\tilde{\boldsymbol{c}})-\tilde{\boldsymbol{c}}^T\boldsymbol{x}-v\right]^+$ for the central problem (2.8). The proof of Theorem 1 is inspired from proof techniques in Doan and Natarajan [15] and Natarajan et. al. [32].

Theorem 1. *For each $i\in[N]$, assume that the marginal distribution P_i of the continuously distributed random variable $\tilde{c}_i$ with support $\Omega_i=[\underline{c}_i,\bar{c}_i]$ is given. For $\boldsymbol{x}\in\mathcal{X}\subseteq\{0,1\}^N$ and $v\geq 0$, define*

$$\phi(\boldsymbol{x},v):=\sup_{P\in\mathbb{P}(P_1,\ldots,P_N)}\mathbb{E}_P\left[Z(\tilde{\boldsymbol{c}})-\tilde{\boldsymbol{c}}^T\boldsymbol{x}-v\right]^+,$$

and

$$\bar{\phi}(\boldsymbol{x},v):=\min_{\boldsymbol{d}\in\Omega}\Big(\left[Z(\boldsymbol{d})-\boldsymbol{d}^T\boldsymbol{x}-v\right]^+ + (\boldsymbol{d}-\boldsymbol{\mu})^T\boldsymbol{x}+\sum_{i=1}^N\mathbb{E}_{P_i}[\tilde{c}_i-d_i]^+\Big).$$

Then $\phi(\boldsymbol{x},v)=\bar{\phi}(\boldsymbol{x},v)$.

Proof. Define

$$\begin{aligned}\phi_0(\boldsymbol{x},v) &:= \sup_{P\in\mathbb{P}(P_1,\ldots,P_N)}\mathbb{E}_P\left[\max\left(Z(\tilde{\boldsymbol{c}}),\tilde{\boldsymbol{c}}^T\boldsymbol{x}+v\right)\right]\\ \bar{\phi}_0(\boldsymbol{x},v) &:= \min_{\boldsymbol{d}\in\Omega}\Big(\max\left(Z(\boldsymbol{d}),\boldsymbol{d}^T\boldsymbol{x}+v\right)+\sum_{i=1}^N\mathbb{E}_{P_i}[\tilde{c}_i-d_i]^+\Big).\end{aligned}$$

Since $\max\left(Z(\boldsymbol{c}), \boldsymbol{c}^T\boldsymbol{x} + v\right) = [Z(\boldsymbol{c}) - \boldsymbol{c}^T\boldsymbol{x} - v]^+ + \boldsymbol{c}^T\boldsymbol{x} + v$, to prove $\phi(\boldsymbol{x}, v) = \bar{\phi}(\boldsymbol{x}, v)$ is equivalent to proving that $\phi_0(\boldsymbol{x}, v) = \bar{\phi}_0(\boldsymbol{x}, v)$.

Step 1: Prove that $\phi_0(\boldsymbol{x}, v) \leq \bar{\phi}_0(\boldsymbol{x}, v)$.

For any $\boldsymbol{c} \in \Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_N$, the following holds:

$$\begin{aligned}
\max\left(Z(\boldsymbol{c}), \boldsymbol{c}^T\boldsymbol{x} + v\right) &= \max\left(\max_{\boldsymbol{y}\in\mathcal{X}} (\boldsymbol{c} - \boldsymbol{d} + \boldsymbol{d})^T\boldsymbol{y}, (\boldsymbol{c} - \boldsymbol{d} + \boldsymbol{d})^T\boldsymbol{x} + v\right) \\
&\leq \max\left(\max_{\boldsymbol{y}\in\mathcal{X}} \boldsymbol{d}^T\boldsymbol{y} + \max_{\boldsymbol{y}\in\mathcal{X}} (\boldsymbol{c} - \boldsymbol{d})^T\boldsymbol{y}, \boldsymbol{d}^T\boldsymbol{x} + v + (\boldsymbol{c} - \boldsymbol{d})^T\boldsymbol{x}\right) \\
&\leq \max\left(Z(\boldsymbol{d}) + \sum_{i=1}^{n} [c_i - d_i]^+, \boldsymbol{d}^T\boldsymbol{x} + v + \sum_{i=1}^{n} [c_i - d_i]^+\right) \\
&= \max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T\boldsymbol{x} + v\right) + \sum_{i=1}^{n} [c_i - d_i]^+.
\end{aligned}$$

Taking expectation with respect to the probability measure $P \in \mathbb{P}(P_1, \ldots, P_N)$ and minimum with respect to $\boldsymbol{d} \in \Omega$, we get

$$\mathbb{E}_P\left[\max\left(Z(\tilde{\boldsymbol{c}}), \tilde{\boldsymbol{c}}^T\boldsymbol{x} + v\right)\right] \leq \bar{\phi}_0(\boldsymbol{x}, v), \quad \forall\, P \in \mathbb{P}(P_1, \ldots, P_N).$$

Taking supremum with respect to $P \in \mathbb{P}(P_1, \ldots, P_N)$, implies $\phi_0(\boldsymbol{x}, v) \leq \bar{\phi}_0(\boldsymbol{x})$.

Step 2: Prove that $\phi_0(\boldsymbol{x}, v) \geq \bar{\phi}_0(\boldsymbol{x}, v)$.

First, we claim that

$$\bar{\phi}_0(\boldsymbol{x}, v) = \min_{\boldsymbol{d}\in\Re^N} \left(\max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T\boldsymbol{x} + v\right) + \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+\right). \tag{3.4}$$

Since for all $\boldsymbol{d} \in \Re^N \setminus \Omega$, we can choose $\boldsymbol{d}^* \in \Omega$:

$$d_i^* = \begin{cases} d_i, \text{ if } d_i \in [\underline{c}_i, \overline{c}_i], \\ \overline{c}_i, \text{ if } d_i > \overline{c}_i, \\ \underline{c}_i, \text{ if } d_i < \underline{c}_i. \end{cases}$$

such that the objective value will be lesser than or equal to the objective value at $\boldsymbol{d}$. The reason is that if $d_i > \overline{c}_i$, by setting $d_i^* = \overline{c}_i$, the second term of the objective function in (3.4) will not change while the first term will decrease or stay constant. If $d_i < \underline{c}_i$, by setting $d_i^* = \underline{c}_i$, the second term will decrease by $\underline{c}_i - d_i$, and the first term will increase by at most $\underline{c}_i - d_i$. Hence $\bar{\phi}_0(\boldsymbol{x}, v)$ can be expressed as:

$$\begin{aligned}
\bar{\phi}_0(\boldsymbol{x}, v) \quad &= \min_{\boldsymbol{d},t} \quad t + \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ \\
&\quad \text{s.t.} \quad t \geq \boldsymbol{d}^T\boldsymbol{y}, \quad \forall\, \boldsymbol{y} \in \mathcal{X}
\end{aligned} \tag{3.5}$$

$$t \geq \boldsymbol{d}^T \boldsymbol{x} + v.$$

For a fixed $\boldsymbol{x} \in \mathcal{X}$, (3.5) is a convex programming problem in decision variables $\boldsymbol{d}$ and t. The Karush-Kuhn-Tucker (KKT) conditions for (3.5) are given as follows:

$$\lambda(\boldsymbol{y}) \geq 0, t \geq \boldsymbol{d}^T \boldsymbol{y}, \forall \boldsymbol{y} \in \mathcal{X}, \text{ and } s \geq 0, t \geq \boldsymbol{d}^T \boldsymbol{x} + v \tag{3.6}$$

$$\sum_{\boldsymbol{y} \in \mathcal{X}} \lambda(\boldsymbol{y}) + s = 1 \tag{3.7}$$

$$\lambda(\boldsymbol{y})\Big(\max\big(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v\big) - \boldsymbol{d}^T \boldsymbol{y}\Big) = 0, \ \forall \ \boldsymbol{y} \in \mathcal{X} \tag{3.8}$$

$$s\Big(\max\big(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v\big) - \boldsymbol{d}^T \boldsymbol{x} - v\Big) = 0 \tag{3.9}$$

$$P(\tilde{c}_i \geq d_i) = \sum_{\boldsymbol{y} \in \mathcal{X}: y_i = 1} \lambda(\boldsymbol{y}) + s x_i. \tag{3.10}$$

There exists an optimal $\boldsymbol{d}$ in the compact set Ω and optimal $t = \max\big(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v\big)$ to problem (3.5). Under the standard Slater's conditions for strong duality in convex optimization, there exist dual variables $s, \lambda(\boldsymbol{y})$ such that these optimal $\boldsymbol{d}, t, s, \lambda(\boldsymbol{y})$ that satisfy the KKT conditions. For the rest of the proof, we let $\boldsymbol{d}, t, s, \lambda(\boldsymbol{y})$ denote the optimal solution that satisfy the KKT conditions. We construct a distribution $\bar{P}$ as follows:

(a) Generate a random vector $\tilde{\boldsymbol{y}}$ which takes the value $\boldsymbol{y} \in \mathcal{X}$ with probability $\lambda(\boldsymbol{y})$ if $\boldsymbol{y} \neq \boldsymbol{x}$, and takes the value $\boldsymbol{x} \in \mathcal{X}$ with probability s. Note that $\lambda(\boldsymbol{x}) = 0$ from the KKT condition (3.8).

(b) Define the set $I_1 = \{i \in [N] : \underline{c}_i < d_i < \bar{c}_i\}$ and $I_2 = [N] \setminus I_1$. For $i \in I_1$, generate the random variable $\tilde{c}_i$ with the conditional probability density function

$$\bar{f}_i(c_i | \tilde{\boldsymbol{y}} = \boldsymbol{y}) = \begin{cases} \frac{1}{P(\tilde{c}_i \geq d_i)} \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) & \text{if } y_i = 1, \\ \frac{1}{P(\tilde{c}_i < d_i)} \mathbb{I}_{[\underline{c}_i, d_i)}(c_i) f_i(c_i) & \text{if } y_i = 0, \end{cases}$$

and for $i \in I_2$ generate the random variable $\tilde{c}_i$ with the conditional probability density function $\bar{f}_i(c_i | \tilde{\boldsymbol{y}} = \boldsymbol{y}) = f_i(c_i)$.

For $i \in I_2$, the probability density function for each $\tilde{c}_i$ under $\bar{P}$ is $\bar{f}_i(c_i) = f_i(c_i)$. For $i \in I_1$, the probability density function is:

$$\begin{aligned} \bar{f}_i(c_i) &= \sum_{\boldsymbol{y} \in \mathcal{X}} \lambda(\boldsymbol{y}) \bar{f}_i(c_i | \tilde{\boldsymbol{y}} = \boldsymbol{y}) + s \cdot \bar{f}_i(c_i | \tilde{\boldsymbol{y}} = \boldsymbol{x}) \\ &= \sum_{\boldsymbol{y} \in \mathcal{X}: y_i = 1} \lambda(\boldsymbol{y}) \frac{1}{P(\tilde{c}_i \geq d_i)} \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) \ + \ s x_i \frac{1}{P(\tilde{c}_i \geq d_i)} \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) \end{aligned}$$

$$
\begin{aligned}
& + \sum_{\boldsymbol{y}\in\mathcal{X}:y_i=0} \lambda(\boldsymbol{y}) \frac{1}{P(\tilde{c}_i < d_i)} \mathbb{I}_{[\underline{c}_i, d_i)}(c_i) f_i(c_i) \;+\; s(1-x_i) \frac{1}{P(\tilde{c}_i < d_i)} \mathbb{I}_{[\underline{c}_i, d_i)}(c_i) f_i(c_i) \\
=\; & \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) + \mathbb{I}_{[\underline{c}_i, d_i)}(c_i) f_i(c_i) \\
=\; & f_i(c_i).
\end{aligned}
$$

The probability density function constructed hence belongs to $\mathbb{P}(P_1, \ldots, P_n)$. Therefore,

$$
\begin{aligned}
\phi_0(\boldsymbol{x}, v) \;\geq\; & \mathbb{E}_{\bar{P}} \left[\max\left(Z(\tilde{\boldsymbol{c}}), \tilde{\boldsymbol{c}}^T \boldsymbol{x} + v \right) \right] \\
=\; & \sum_{\boldsymbol{y}\in\mathcal{X}} \lambda(\boldsymbol{y}) \mathbb{E}_{\bar{P}} \left[\max\left(Z(\tilde{\boldsymbol{c}}), \tilde{\boldsymbol{c}}^T \boldsymbol{x} + v \right) \mid \tilde{\boldsymbol{y}} = \boldsymbol{y} \right] + s \mathbb{E}_{\bar{P}} \left[\max\left(Z(\tilde{\boldsymbol{c}}), \tilde{\boldsymbol{c}}^T \boldsymbol{x} + v \right) \mid \tilde{\boldsymbol{y}} = \boldsymbol{x} \right] \\
\geq\; & \sum_{\boldsymbol{y}\in\mathcal{X}} \lambda(\boldsymbol{y}) \mathbb{E}_{\bar{P}} \left[Z(\tilde{\boldsymbol{c}}) \mid \tilde{\boldsymbol{y}} = \boldsymbol{y} \right] + s \mathbb{E}_{\bar{P}} \left[\tilde{\boldsymbol{c}}^T \boldsymbol{x} + v \mid \tilde{\boldsymbol{y}} = \boldsymbol{x} \right] \\
\geq\; & \sum_{\boldsymbol{y}\in\mathcal{X}} \lambda(\boldsymbol{y}) \mathbb{E}_{\bar{P}} \left[\tilde{\boldsymbol{c}}^T \boldsymbol{y} \mid \tilde{\boldsymbol{y}} = \boldsymbol{y} \right] + s \mathbb{E}_{\bar{P}} \left[\tilde{\boldsymbol{c}}^T \boldsymbol{x} + v \mid \tilde{\boldsymbol{y}} = \boldsymbol{x} \right] \\
=\; & \sum_{\boldsymbol{y}\in\mathcal{X}:y_i=1} \lambda(\boldsymbol{y}) \Big(\sum_{i\in I_1} \int c_i \frac{1}{P(\tilde{c}_i \geq d_i)} \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + \sum_{i\in I_2} \int c_i f_i(c_i) dc_i \Big) \\
& + s \Big(\sum_{i\in I_1} \int c_i x_i \frac{1}{P(\tilde{c}_i \geq d_i)} \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + \sum_{i\in I_2} \int c_i x_i f_i(c_i) dc_i \Big) + sv \\
=\; & \sum_{i\in I_1} \int c_i \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + \sum_{i\in I_2} \int P(\tilde{c}_i \geq d_i) c_i f_i(c_i) dc_i + sv. \qquad \text{(by (3.10))}
\end{aligned}
$$

Since $P(\tilde{c}_i \geq d_i) = 1$ or 0 for $i \in I_2$, hence $\int P(\tilde{c}_i \geq d_i) c_i f_i(c_i) dc_i = \int c_i \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i, \forall i \in I_2$. Then, we obtain

$$
\begin{aligned}
\phi_0(\boldsymbol{x}, v) \;\geq\; & \sum_{i=1}^{N} \int c_i \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + sv \\
=\; & \sum_{i=1}^{N} \int (c_i - d_i) \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + \sum_{i=1}^{N} d_i \int \mathbb{I}_{[d_i, \bar{c}_i]}(c_i) f_i(c_i) dc_i + sv \\
=\; & \sum_{i=1}^{N} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ + \sum_{i=1}^{N} d_i \Big(\sum_{\boldsymbol{y}\in\mathcal{X}:y_i=1} \lambda(\boldsymbol{y}) + s x_i \Big) + sv \qquad \text{(by (3.10))} \\
=\; & \sum_{i=1}^{N} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ + \sum_{\boldsymbol{y}\in\mathcal{X}} \lambda(\boldsymbol{y}) \boldsymbol{d}^T \boldsymbol{y} + s(\boldsymbol{d}^T \boldsymbol{x} + v) \\
=\; & \sum_{i=1}^{N} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ + \sum_{\boldsymbol{y}\in\mathcal{X}} \lambda(\boldsymbol{y}) \big(\max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v \right) \big) + s(\boldsymbol{d}^T \boldsymbol{x} + v) \qquad \text{(by (3.8))} \\
=\; & \sum_{i=1}^{N} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ + (1-s) \big(\max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v \right) \big) + s(\boldsymbol{d}^T \boldsymbol{x} + v) \qquad \text{(by (3.7))}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + \max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T\boldsymbol{x} + v\right) \qquad \text{(by (3.9))} \\
&= \bar{\phi}_0(\boldsymbol{x}, v).
\end{aligned}
$$

□

It is useful to contrast the regret bound in Theorem 1 with the earlier bound of Meilijson and Nadas [30] in (3.1). In Theorem 1, the worst-case joint distribution depends on the solution $\boldsymbol{x} \in \mathcal{X}$ and the scalar v. The worst-case joint distribution in Formulation (3.1) however depends on the scalar v only. The proof of Theorem 1 can be extended directly to discrete marginal distributions by replacing the integrals with summations and using linear programming duality. This result generalizes to the marginal moment model and piecewise linear convex functions as illustrated in the next theorem. The proof of Theorem 2 is inspired from proof techniques in Bertsimas et. al. [9] and Natarajan et. al. [32].

Theorem 2. *For $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^N$, consider the marginal moment model:*

$$
\mathbb{P}_i = \{P_i \mid \mathbb{E}_{P_i}[f_{ik}(\tilde{c}_i)] = m_{ik}, k \in [K_i], \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i, \bar{c}_i]}(\tilde{c}_i)] = 1\},
$$

where $\mathbb{I}_{[\underline{c}_i,\bar{c}_i]}(c_i) = 1$ if $\underline{c}_i \le c_i \le \bar{c}_i$ and 0 otherwise. Assume that the moments lie interior to the set of feasible moment vectors. Define

$$
\phi(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b}) := \sup_{P \in \mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_P\left[g\left(Z(\tilde{\boldsymbol{c}}) - \tilde{\boldsymbol{c}}^T\boldsymbol{x}\right)\right] \tag{3.11}
$$

where $g(\cdot)$ is a non-decreasing piecewise linear convex function defined by

$$
g(z) = \max_{j \in [J]} \left(a_j z + b_j\right),
$$

with $0 \le a_1 < a_2 < \ldots < a_J$. Let

$$
\bar{\phi}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b}) := \min_{\boldsymbol{d}_1,\ldots,\boldsymbol{d}_J \in \Omega} \left(g\left(Z(\boldsymbol{d}_j) - \boldsymbol{d}_j^T\boldsymbol{x}\right) + \sum_{i=1}^{N} \sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}\left[\max_{j \in [J]} a_j \left([\tilde{c}_i - d_{ji}]^+ - [\tilde{c}_i - d_{ji}]x_i\right)\right]\right). \tag{3.12}
$$

Then $\phi(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b}) = \bar{\phi}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b})$.

Proof.

Step 1: Prove that $\phi(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b}) \le \bar{\phi}(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b})$.

For any $\boldsymbol{c} \in \Omega$, and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_J \in \Omega$, the following holds:

$$
\begin{aligned}
g(Z(\boldsymbol{c}) - \boldsymbol{c}^T\boldsymbol{x}) &= \max_{j \in [J]} \left[a_j \left(\max_{\boldsymbol{y} \in \mathcal{X}} (\boldsymbol{c} - \boldsymbol{d}_j + \boldsymbol{d}_j)^T\boldsymbol{y} - \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}_j^T\boldsymbol{x} - \boldsymbol{d}_j^T\boldsymbol{x}\right) + b_j\right] \\
&\le \max_{j \in [J]} \left[a_j (\max_{\boldsymbol{y} \in \mathcal{X}} \boldsymbol{d}_j^T\boldsymbol{y} - \boldsymbol{d}_j^T\boldsymbol{x}) + b_j\right] + \max_{j \in [J]} a_j \left[\max_{\boldsymbol{y} \in \mathcal{X}} (\boldsymbol{c} - \boldsymbol{d}_j)^T\boldsymbol{y} - (\boldsymbol{c} - \boldsymbol{d}_j)^T\boldsymbol{x}\right] \\
&\le g(Z(\boldsymbol{d}_j) - {\boldsymbol{d}_j}^T\boldsymbol{x}) + \max_{j \in [J]} a_j \sum_{i=1}^{N} \left[(c_i - d_{ji})^+ - (c_i - d_{ji})x_i\right].
\end{aligned}
$$

The first inequality is due to the subadditivity of $Z(\cdot)$, and the second one follows from the fact that $\max_{\boldsymbol{y}\in\mathcal{X}}(\boldsymbol{c}-\boldsymbol{d}_j)^T\boldsymbol{y} \leq \sum_{i=1}^N (c_i - d_{ji})^+$ and $a_j \geq 0$ for all $j \in [J]$. For any distribution P, taking expectation on both sides of the above inequality gives

$$\begin{aligned}\mathbb{E}_P[g(Z(\boldsymbol{c})-\boldsymbol{c}^T\boldsymbol{x})] &\leq g(Z(\boldsymbol{d}_j)-\boldsymbol{d}_j{}^T\boldsymbol{x}) + \mathbb{E}_P\left(\max_{j\in[J]} a_j \sum_{i=1}^N [(\tilde{c}_i - d_{ji})^+ - (\tilde{c}_i - d_{ji})x_i]\right)\\ &\leq g(Z(\boldsymbol{d}_j)-\boldsymbol{d}_j{}^T\boldsymbol{x}) + \sum_{i=1}^N \mathbb{E}_{P_i}\left(\max_{j\in[J]} a_j [(\tilde{c}_i - d_{ji})^+ - (\tilde{c}_i - d_{ji})x_i]\right).\end{aligned}$$

Note that the last inequality follows from the fact that $\max_{j\in[J]}(a_j \sum_{i=1}^N [(\tilde{c}_i - d_{ji})^+ - (\tilde{c}_i - d_{ji})x_i]) \leq \sum_{i=1}^N \max_{j\in[J]}(a_j[(\tilde{c}_i - d_{ji})^+ - (\tilde{c}_i - d_{ji})x_i])$. The above inequality holds for any distribution $P \in \mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)$ and any $\boldsymbol{d}_1,\ldots,\boldsymbol{d}_J \in \Omega$. Taking supremum with respect to $P \in \mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)$, and taking minimum with respect to $\boldsymbol{d}_1,\ldots,\boldsymbol{d}_J \in \Omega$, we get

$$\begin{aligned}\phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) &= \sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)} \mathbb{E}_P[g(Z(\boldsymbol{c})-\boldsymbol{c}^T\boldsymbol{x})]\\ &\leq \min_{\boldsymbol{d}_1,\ldots,\boldsymbol{d}_J\in\Omega}\left(\max_{j\in[J]}[a_j(Z(\boldsymbol{d}_j)-\boldsymbol{d}_j^T\boldsymbol{x})+b_j] + \sum_{i=1}^N \sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}\left[\max_{j\in[J]} a_j\left([\tilde{c}_i - d_{ji}]^+ - [\tilde{c}_i - d_{ji}]x_i\right)\right]\right)\\ &= \bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}).\end{aligned}$$

Step 2: Prove that $\phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) \geq \bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b})$.

Consider the dual problem of (3.11) and the dual of the supremum problem in (3.12). Since the moments lie interior to the set of feasible moment vectors, strong duality holds (see Isii [22]). Hence

$$\begin{aligned}\phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) = \quad &\min \quad y_{00} + \sum_{i=1}^N\sum_{k=1}^{K_i} y_{ik}m_{ik}\\ &\text{s.t.} \quad y_{00} + \sum_{i=1}^N\sum_{k=1}^{K_i} y_{ik}f_{ik}(c_i) - [a_j(\boldsymbol{c}^T\boldsymbol{y}-\boldsymbol{c}^T\boldsymbol{x})+b_j] \geq 0, \quad \forall \boldsymbol{c}\in\Omega, \boldsymbol{y}\in\mathcal{X}, j\in[J].\end{aligned} \tag{3.13}$$

$$\begin{aligned}\bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) = \quad &\min \quad \left(\max_{j\in[J]}[a_j(Z(\boldsymbol{d}_j)-\boldsymbol{d}_j^T\boldsymbol{x})+b_j] + \sum_{i=1}^N \bar{y}_{i0} + \sum_{i=1}^N\sum_{k=1}^{K_i}\bar{y}_{ik}m_{ik}\right)\\ &\text{s.t.} \quad \bar{p}_{i1}(c_i) := \bar{y}_{i0} + \sum_{k=1}^{K_i}\bar{y}_{ik}f_{ik}(c_i) - a_j(c_i - d_{ji})(1-x_i) \geq 0, \quad \forall c_i\in\Omega_i, i\in[N], j\in[J],\\ &\qquad \bar{p}_{i2}(c_i) := \bar{y}_{i0} + \sum_{k=1}^{K_i}\bar{y}_{ik}f_{ik}(c_i) + a_j(c_i - d_{ji})x_i \geq 0, \quad \forall c_i\in\Omega_i, i\in[N], j\in[J].\end{aligned} \tag{3.14}$$

Let $y_{00}^*, y_{ik}^*, k\in[K_i], i\in[N]$ be the optimal solution to (3.13). Now generate a feasible solution to (3.14) as follows. Set $\bar{y}_{ik} = y_{ik}^*, k\in[K_i], i\in[N]$. Having fixed $\bar{y}_{ik}$, choose $\bar{y}_{i0}$ and d_{ji} based on the value $x_i = 1$ or 0 in the following manner:

1. If $x_i = 1$, we choose $\bar{y}_{i0}^*$ to be the minimal value such that $\bar{p}_{i1}(c_i)$ is nonnegative over Ω_i. Namely, there exists some $c_i^* \in \Omega_i$ such that $\bar{p}_{i1}(c_i^*) = 0$. Then for all $j \in [J]$, choose d_{ji}^* to be the maximal value such that $\bar{p}_{i2}(c_i)$ is nonnegative over Ω_i. This value can be chosen such that $d_{ji}^* \in \Omega_i$. To verify this observe that since $x_i = 1$, $\bar{p}_{i2}(c_i) = \bar{p}_{i1}(c_i) + a_j(c_i - d_{ji})$. If $a_j = 0$ the result is obvious; if $a_j > 0$ then $d_{ji} = \underline{c}_i$ is feasible since $\bar{p}_{i1}(c_i) \geq 0, \forall c_i \in \Omega_i$. Hence $d_{ji}^* \geq \underline{c}_i$ since it is chosen as the maximal value such that $\bar{p}_{i2}(c_i)$ is nonnegative over Ω_i. Moreover, $d_{ji}^* \leq \bar{c}_i$ or else $p_{i2}(c_i^*) = a_j(c_i^* - d_{ji}) < 0$. Hence $d_{ji}^* \in \Omega_i$.

2. If $x_i = 0$, we choose $\bar{y}_{i0}^*$ to be the minimal value such that $\bar{p}_{i2}(c_i)$ is nonnegative over Ω_i. Then for all $j \in [J]$, choose d_{ji}^* to be the minimal value such that $\bar{p}_{i1}(c_i)$ is nonnegative over Ω_i. A similar argument to shows that d_{ji}^* can be restricted to the set Ω_i.

Now, for any $j \in [J]$, and any $\boldsymbol{y} \in \mathcal{X}$, from the constraints of (3.14), we obtain that

$$\bar{y}_{i0}^* + \sum_{k=1}^{K_i} y_{ik}^* f_{ik}(c_i) - a_j(c_i - d_{ji}^*)(y_i - x_i) \geq 0, \quad \forall c_i \in \Omega_i, i \in [N]. \tag{3.15}$$

By the choice of $\bar{y}_{i0}^*$ and d_{ji}^*, the value $\bar{y}_{i0}^* + a_j d_{ji}^*(y_i - x_i)$ is the minimal value such that the above inequality holds over Ω_i for all $i \in [N]$. Take the summation of these n inequalities:

$$\sum_{i=1}^{N} \bar{y}_{i0}^* + \sum_{i=1}^{N}\sum_{k=1}^{K_i} y_{ik}^* f_{ik}(c_i) - a_j(\boldsymbol{c} - \boldsymbol{d}_j^*)^T(\boldsymbol{y} - \boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{c} \in \Omega, \boldsymbol{y} \in \mathcal{X}, j \in [J]. \tag{3.16}$$

Note that in general, given N univariate functions $\bar{p}_i(c_i) = \sum_{k=1}^{K_i} a_{ik} f_{ik}(c_i) + a_{i0}$ such that a_{i0} is the minimal value for $\bar{p}_i(c_i)$ to be nonnegative over Ω_i, the minimal value of a_{00} for the multivariate function $\bar{p}(\boldsymbol{c}) = \sum_{i=1}^{N}\sum_{k=1}^{K_i} a_{ik} f_{ik}(c_i) + a_{00}$ to be nonnegative over Ω is $\sum_{i=1}^{N} a_{i0}$. By setting $a_{i0} = \bar{y}_{i0}^* + a_j d_{ji}^*(y_i - x_i)$, $a_{00} = y_{00}^* - b_j$, and comparing (3.16) with the constraint of (3.13):

$$y_{00}^* + \sum_{i=1}^{N}\sum_{k=1}^{K_i} y_{ik}^* f_{ik}(c_i) - [a_j(\boldsymbol{c}^T\boldsymbol{y} - \boldsymbol{c}^T\boldsymbol{x}) + b_j] \geq 0, \quad \forall \boldsymbol{c} \in \Omega, \boldsymbol{y} \in \mathcal{X}, j \in [J].$$

This leads to the following result

$$y_{00}^* - b_j \geq \sum_{i=1}^{N} \bar{y}_{i0}^* + a_j \boldsymbol{d}_j^{*T}(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathcal{X}, j \in [J],$$

which is equivalent to

$$y_{00}^* \geq \sum_{i=1}^{N} \bar{y}_{i0}^* + g(Z(\boldsymbol{d}_j^*) - \boldsymbol{d}_j^{*T}\boldsymbol{x}).$$

Therefore

$$\begin{aligned}\bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) &\leq g(Z(\boldsymbol{d}_j^*)-\boldsymbol{d}_j^{*T}\boldsymbol{x})+\sum_{i=1}^N \bar{y}_{i0}^*+\sum_{i=1}^N\sum_{k=1}^{K_i} y_{ik}^* m_{ik}\\ &\leq y_{00}^*+\sum_{i=1}^N\sum_{k=1}^{K_i} y_{ik}^* m_{ik} \;=\; \phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}).\end{aligned}$$

□

The next proposition provides an extension of the results in Meilijson and Nadas [30] and Bertsimas et al. [9] to the worst-case conditional value-at-risk of regret.

Proposition 1. *Consider the marginal distribution model with* $\mathbb{P}_i = \{P_i\}, i \in [N]$ *or the marginal moment model with* $\mathbb{P}_i = \{P_i : \mathbb{E}_{P_i}[f_{ik}(\tilde{c}_i)] = m_{ik}, k \in [K_i], \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i,\bar{c}_i]}(\tilde{c}_i)] = 1\}, i \in [N]$*. For* $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^N$*, the worst-case* CVaR *of regret can be computed as*

$$\mathrm{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}})) = \min_{\boldsymbol{d}\in\Omega}\left(Z(\boldsymbol{d})+\frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x}+\frac{1}{1-\alpha}\sum_{i=1}^N \sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}([\tilde{c}_i-d_i]^+-\tilde{c}_i x_i)\right). \tag{3.17}$$

Proof. From the definition of WCVaR in (2.8):

$$\mathrm{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}})) = \inf_{v\in\Re}\left(v+\frac{1}{1-\alpha}\sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)}\mathbb{E}_P\left[Z(\tilde{\boldsymbol{c}})-\tilde{\boldsymbol{c}}^T\boldsymbol{x}-v\right]^+\right).$$

Applying Theorems 1 and 2, we have:

$$\sup_{P\in\mathbb{P}(\mathbb{P}_1,\ldots,\mathbb{P}_N)}\mathbb{E}_P\left[Z(\tilde{\boldsymbol{c}})-\tilde{\boldsymbol{c}}^T\boldsymbol{x}-v\right]^+ = \min_{\boldsymbol{d}\in\Omega}\left(\left[Z(\boldsymbol{d})-\boldsymbol{d}^T\boldsymbol{x}-v\right]^+ + \sum_{i=1}^N\sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}\left([\tilde{c}_i-d_i]^+-[\tilde{c}_i-d_i]x_i\right)\right). \tag{3.18}$$

The worst-case CVaR of regret is thus computed as:

$$\inf_{v\in\Re}\min_{\boldsymbol{d}\in\Omega}\left(v+\frac{1}{1-\alpha}[Z(\boldsymbol{d})-\boldsymbol{d}^T\boldsymbol{x}-v]^+ + \frac{1}{1-\alpha}\sum_{i=1}^N\sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}([\tilde{c}_i-d_i]^+-[\tilde{c}_i-d_i]x_i)\right). \tag{3.19}$$

In formulation (3.19), the optimal decision variable is $v^* = Z(\boldsymbol{d})-\boldsymbol{d}^T\boldsymbol{x}$ which results in the desired formulation. □

This formulation is appealing computationally since it exploits the marginal distributional representation of the uncertainty. The next result identifies conditions under which the worst-case CVaR of regret is computable in polynomial time for a fixed solution $\boldsymbol{x} \in \mathcal{X}$.

Theorem 3. *Assume the following two conditions hold:*

(a) The deterministic combinatorial optimization is solvable in polynomial time and

(b) For each $i \in [N]$, $G_i(d_i) := \sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i)$ and its subgradient with respect to d_i are computable in polynomial time for a fixed $d_i \in \Omega_i$ and x_i.

Then for a given solution $\boldsymbol{x} \in \mathcal{X}$, the worst-case CVaR *of regret under the marginal distribution or marginal moment models is computable in polynomial time.*

Proof. From Proposition 1, the worst-case CVaR of regret is computed as:

$$\begin{aligned}
\text{WCVaR}_\alpha(R(\boldsymbol{x}, \tilde{\boldsymbol{c}})) = \quad & \min_{\boldsymbol{d}, t, \boldsymbol{s}} \quad \left(t + \frac{\alpha}{1-\alpha} \boldsymbol{d}^T \boldsymbol{x} + \frac{1}{1-\alpha} \sum_{i=1}^{N} s_i \right) \\
& \text{s.t.} \quad t \geq Z(\boldsymbol{d}), \\
& \qquad\ \ s_i \geq \sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i), \quad i = 1, \ldots, N, \\
& \qquad\ \ \boldsymbol{d} \in \Omega.
\end{aligned} \tag{3.20}$$

Denote the feasible set of (3.20) by $\mathcal{K}$. We consider the separation problem of (3.20): given $(\boldsymbol{d}^*, t^*, \boldsymbol{s}^*)$, decide if $(\boldsymbol{d}^*, t^*, \boldsymbol{s}^*) \in \mathcal{K}$, and if not, find a hyperplane which separates $(\boldsymbol{d}^*, t^*, \boldsymbol{s}^*)$ from $\mathcal{K}$. Under assumption (a) and (b), we can check if $(\boldsymbol{d}^*, t^*, \boldsymbol{s}^*) \in \mathcal{K}$ in polynomial time, and if not we consider the following two situations.

1. If $t^* < Z(\boldsymbol{d}^*)$, we can find $\boldsymbol{y}^* \in \mathcal{X}$ such that $Z(\boldsymbol{d}^*) = {\boldsymbol{d}^*}^T \boldsymbol{y}^*$ in polynomial time. It follows that the hyperplane $\{(\boldsymbol{d}, t, \boldsymbol{s}) : {\boldsymbol{y}^*}^T \boldsymbol{d} = t\}$ separates $(\boldsymbol{d}^*, t^*, \boldsymbol{s}^*)$ from $\mathcal{K}$.

2. If $s_i^* < G_i(d_i^*)$ for some $i \in [N]$, then we can find the separating hyperplane in polynomial time, since the subgradient of $G_i(d_i)$ is computable in polynomial time. The remaining constraints $\boldsymbol{d} \in \Omega$ are $2N$ linear constraints that are easy to enforce. Hence, the separation problem of (3.20) can be solved in polynomial time. It follows that the WCVaR of regret under the marginal model is computable in polynomial time. □

Many combinatorial optimization problems satisfy the assumption (a) in Theorem 3. Examples include the longest path problem on a directed acyclic graph, spanning tree problems and assignment problems. Moreover, in the marginal distribution model and several instances of the marginal moment model, Assumption (b) in Theorem 3 is easy to verify. For both the continuous and discrete marginal distribution model, $G_i(d_i)$ is a convex function of d_i and a subgradient of the function is given by $-P(\tilde{c}_i \geq d_i)$. For the marginal moment model when (a) the range and mean are given, or (b) the range, mean and mean absolute deviation are given, $G_i(d_i)$ is a piecewise linear convex function that is efficiently computable (see Madansky [29] and Ben-Tal and Hochman [8]). If $P_i^* \in \mathbb{P}_i$ denotes the extremal distribution that attains the bound $\sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i)$ in these instances, then a subgradient of the function is given by $-P_i^*(\tilde{c}_i \geq d_i)$.

3.2 Worst-case Conditional Value-at-risk of Cost

In this subsection, we apply the previous results to combinatorial optimization problems with an objective of minimizing the worst-case CVaR of the cost:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha\left(-\tilde{\boldsymbol{c}}^T\boldsymbol{x}\right). \tag{3.21}$$

Applying similar arguments as in Proposition 1, we obtain the following proposition under the marginal model.

Proposition 2. *Consider the marginal distribution model with* $\mathbb{P}_i = \{P_i\}, i \in [N]$ *or the marginal moment model with* $\mathbb{P}_i = \{P_i : \mathbb{E}_{P_i}[f_{ik}(\tilde{c}_i)] = m_{ik}, k \in [K_i], \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i,\overline{c}_i]}(\tilde{c}_i)] = 1\}, i \in [N]$*. For* $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^N$*, the worst-case* CVaR *of cost can be computed as*

$$\mathrm{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x}) = \min_{\boldsymbol{d}\in\Omega}\left(\frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x} + \frac{1}{1-\alpha}\sum_{i=1}^N \sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i)\right). \tag{3.22}$$

Since the objective function in (3.22) is separable of d_i, it can be expressed as:

$$\mathrm{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x}) = \sum_{i=1}^N \hat{h}_i(x_i), \tag{3.23}$$

where the function $\hat{h}_i(x_i)$ is defined as:

$$\hat{h}_i(x_i) = \min_{d_i\in\Omega_i}\left(\frac{\alpha}{1-\alpha}d_i x_i + \frac{1}{1-\alpha}\sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i)\right).$$

Define the parameter h_i as the optimal value to a univariate convex programming problem:

$$h_i = \min_{d_i\in\Omega_i}\left(\frac{\alpha}{1-\alpha}d_i + \frac{1}{1-\alpha}\sup_{P_i\in\mathbb{P}_i}\mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i)\right).$$

Then the worst-case CVaR of the cost can be expressed as

$$\mathrm{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x}) = \sum_{i=1}^N h_i x_i. \tag{3.24}$$

To see why this is true observe that if $x_i = 1$, then $\hat{h}_i(1) = h_i$ and if $x_i = 0$, then $\hat{h}_i(0) = 0$. Hence the problem of minimizing the worst-case CVaR of the cost can be formulated as the deterministic combinatorial optimization problem:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathcal{X}}\sum_{i=1}^N h_i x_i. \tag{3.25}$$

In the next section, we provide conic mixed integer programs to minimize the worst-case CVaR of regret.

4 Mixed Integer Programming Formulations

From Proposition 1, the problem of minimizing the WCVaR of regret is formulated as:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}})) \;=\; \min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{d}\in\Omega}\left(Z(\boldsymbol{d}) + \frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x} + \frac{1}{1-\alpha}H(\boldsymbol{x},\boldsymbol{d})\right), \tag{4.1}$$

where

$$H(\boldsymbol{x},\boldsymbol{d}) \;:=\; \sum_{i=1}^{N} \sup_{P_i\in\mathbb{P}_i} \mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i). \tag{4.2}$$

Formulation (4.1) is a stochastic nonconvex mixed integer programming problem where the nonconvexity appears in the bilinear term $\boldsymbol{d}^T\boldsymbol{x}$. For bilinear terms, several linearization techniques have been proposed in the literature by Glover [18], Glover and Woolsey [19, 20], Sherali and Alameddine [41] and Adams and Sherali [2] among others. These alternative linearization techniques vary significantly in terms of their computational performance. We adopt the simplest linearization technique from Glover [18] to handle the bilinear terms where one set of variables is restricted to be binary. For all $i \in [N]$, and $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^N$,

$$z_i = d_i x_i \Leftrightarrow \begin{cases} \overline{c}_i x_i \geq z_i \geq \underline{c}_i x_i \\ d_i - \underline{c}_i(1-x_i) \geq z_i \geq d_i - \overline{c}_i(1-x_i). \end{cases} \tag{4.3}$$

By applying the linearization technique, (4.1) is reformulated as the following stochastic convex mixed integer program:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{d},\boldsymbol{z}} \quad & \left(Z(\boldsymbol{d}) + \frac{\alpha}{1-\alpha}\sum_{i=1}^{N} z_i + \frac{1}{1-\alpha}H(\boldsymbol{x},\boldsymbol{d})\right) \\ \text{s.t.} \quad & \overline{c}_i x_i \geq z_i \geq \underline{c}_i x_i, \quad i \in [N], \\ & d_i - \underline{c}_i(1-x_i) \geq z_i \geq d_i - \overline{c}_i(1-x_i), \quad i \in [N], \\ & \boldsymbol{d} \in \Omega, \quad \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{4.4}$$

The objective function in (4.4) is convex with respect to $\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{z}$ since convexity is preserved under the expectation and maximization operation.

Assume that the feasible region $\mathcal{X}$ is described in the compact form:

$$\mathcal{X} = \left\{\boldsymbol{y} \in \{0,1\}^N \mid \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}\right\},$$

where $\boldsymbol{A}$ is a given integer matrix and $\boldsymbol{b}$ is a given integer vector. For the rest of this section, we assume that matrix $\boldsymbol{A}$ is totally unimodular, namely each square submatrix of $\boldsymbol{A}$ has determinant equal to 0,

+1, or 1. Under this assumption the deterministic combinatorial optimization problem is solvable in polynomial time as a compact linear program (see Schrijver [40]):

$$Z(\boldsymbol{d}) = \max\left\{\boldsymbol{d}^T\boldsymbol{y} \mid \ \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b},\ 0 \leq y_i \leq 1, i \in [N]\right\}. \tag{4.5}$$

Many polynomially solvable 0-1 optimization problems fall under this category including subset selection, longest path on a directed acyclic graph and linear assignment problems. Let $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ be the vectors of dual variables associated with the constraints of (4.5). The dual linear program of (4.5) is given by

$$Z(\boldsymbol{d}) = \min\left\{\boldsymbol{b}^T\boldsymbol{\lambda}_1 + \boldsymbol{e}^T\boldsymbol{\lambda}_2 \mid \ \boldsymbol{A}^T\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 \geq \boldsymbol{d},\ \boldsymbol{\lambda}_2 \geq 0\right\}, \tag{4.6}$$

where $\boldsymbol{e}$ is the vector of all ones. By solving the dual formulation of $Z(\boldsymbol{d})$ in (4.4), we get:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{d},\boldsymbol{z},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2} \quad & \left(\boldsymbol{b}^T\boldsymbol{\lambda}_1 + \boldsymbol{e}^T\boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha}\sum_{i=1}^N z_i + \frac{1}{1-\alpha}H(\boldsymbol{x},\boldsymbol{d})\right) && (4.7) \\ \text{s.t.} \quad & \bar{c}_i x_i \geq z_i \geq \underline{c}_i x_i, \quad i \in [N], && (4.7a) \\ & d_i - \underline{c}_i(1-x_i) \geq z_i \geq d_i - \bar{c}_i(1-x_i), \quad i \in [N], && (4.7b) \\ & \boldsymbol{d} \in \Omega,\ \boldsymbol{A}^T\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 \geq \boldsymbol{d},\ \boldsymbol{\lambda}_2 \geq 0,\ \boldsymbol{x} \in \mathcal{X}. && (4.7c) \end{aligned}$$

The constraints in problem (4.7) are all linear except for the integrality restrictions in the description of $\mathcal{X}$. To convert this to a conic mixed integer program, we apply standard conic programming methods to evaluate $H(\boldsymbol{x},\boldsymbol{d})$ in the objective function.

4.1 Marginal Discrete Distribution Model

Assume that the marginal distributions of $\tilde{\boldsymbol{c}}$ are discrete:

$$\tilde{c}_i \sim c_{ij} \text{ with probability } p_{ij},\ j \in [J_i], i \in [N]$$

where $\sum_{j\in[J_i]} p_{ij} = 1$ and $\sum_{j\in[J_i]} c_{ij}p_{ij} = \mu_i$ for each $i \in [N]$. The input specification for the marginal discrete distribution model needs $J_1 + J_2 + \ldots + J_N$ probabilities which is typically much smaller than the size of the input needed to specify the joint distribution that needs up to $J_1 \times J_2 \times \ldots \times J_N$ probabilities. In this case:

$$H(\boldsymbol{x},\boldsymbol{d}) \ = \ \sum_{i=1}^N \sum_{j=1}^{J_i} (c_{ij} - d_i)^+ p_{ij} - \boldsymbol{\mu}^T\boldsymbol{x} \ = \min_{t_{ij} \geq c_{ij} - d_i, t_{ij} \geq 0} \sum_{i=1}^N \sum_{j=1}^{J_i} t_{ij}p_{ij} - \boldsymbol{\mu}^T\boldsymbol{x},$$

The problem of minimizing WCVaR is thus formulated as the compact MILP:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{d},\boldsymbol{z},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\boldsymbol{t}} \quad & \left(\boldsymbol{b}^T\boldsymbol{\lambda}_1 + \boldsymbol{e}^T\boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha}\sum_{i=1}^{N} z_i + \frac{1}{1-\alpha}\left(\sum_{i=1}^{N}\sum_{j=1}^{J_i} t_{ij}p_{ij} - \boldsymbol{\mu}^T\boldsymbol{x} \right)\right) \\ \text{s.t.} \quad & t_{ij} \geq c_{ij} - d_i, t_{ij} \geq 0, j \in [J_i], i \in [N], \\ & (4.7a), (4.7b) \text{ and } (4.7c). \end{aligned} \tag{4.8}$$

4.2 Marginal Moment Model

In the standard representation of the marginal moment model, $H(\boldsymbol{x}, \boldsymbol{d})$ is evaluated through conic optimization. This is based on the well-known duality theory of moments and nonnegative polynomials for univariate models. The reader is referred to Nesterov [33] and Bertsimas and Popescu [10] for details. We restrict attention to instances of the marginal moment model where (4.7) can be solved as a MILP or MISOCP. The advantage of these formulations is that the probabilistic regret model can be solved with standard off the shelf solvers such as CPLEX. The details are listed next:

(a) Range and Mean are Known:

Assume the interval range and mean of the random vector $\tilde{\boldsymbol{c}}$ are given:

$$\mathbb{P}_i = \{P_i : \mathbb{E}_{P_i}[\tilde{c}_i] = \mu_i, \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i,\bar{c}_i]}(\tilde{c}_i)] = 1\}.$$

In this case, the optimal distribution to the problem $\sup_{P_i\in\mathbb{P}_i} \mathbb{E}_{P_i}\left[(\tilde{c}_i - d_i)^+ - \tilde{c}_i x_i\right]$ is known explicitly (see Madansky [29] and Ben-Tal and Hochman [8]):

$$\tilde{c}_i = \begin{cases} \bar{c}_i, \text{ with probability } \frac{\mu_i - \underline{c}_i}{\bar{c}_i - \underline{c}_i}, \\ \underline{c}_i, \text{ with probability } \frac{\bar{c}_i - \mu_i}{\bar{c}_i - \underline{c}_i}. \end{cases}$$

The worst-case marginal distribution is a two point distribution and can be treated as a special case of the discrete marginal distribution. The probabilistic regret model is solved with the MILP (4.8).

(b) Range, Mean and Mean Absolute Deviation are Known:

Assume the interval range, mean and the mean absolute deviation of the random vector $\tilde{\boldsymbol{c}}$ are given:

$$\mathbb{P}_i = \{P_i : \mathbb{E}_{P_i}(\tilde{c}_i) = \mu_i, \mathbb{E}_{P_i}(|\tilde{c}_i - \mu_i|) = \delta_i, \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i,\bar{c}_i]}(\tilde{c}_i)] = 1\}.$$

For feasibility the mean absolute deviation satisfies $\delta_i \leq \frac{2(\bar{c}_i-\mu_i)(\mu_i-\underline{c}_i)}{\bar{c}_i-\underline{c}_i}$. The optimal distribution for $\sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}\left[(\tilde{c}_i - d_i)^+ - \tilde{c}_i x_i\right]$ has been identified by Ben-Tal and Hochman [8]:

$$\tilde{c}_i = \begin{cases} \underline{c}_i, \text{ with probability } \frac{\delta_i}{2(\mu_i - \underline{c}_i)} =: p_i, \\ \overline{c}_i, \text{ with probability } \frac{\delta_i}{2(\overline{c}_i - \mu_i)} =: q_i, \\ \mu_i, \text{ with probability } 1 - p_i - q_i. \end{cases}$$

This is a three point distribution and the MILP reformulation (4.8) can be used.

(c) Range, Mean and Standard Deviation are Known:

Assume the range, mean and the standard deviation of the random vector $\tilde{\boldsymbol{c}}$ are given:

$$\mathbb{P}_i = \{P_i : \mathbb{E}_{P_i}(\tilde{c}_i) = \mu_i, \mathbb{E}_{P_i}(\tilde{c}_i^2) = \mu_i^2 + \sigma_i^2, \mathbb{E}_{P_i}[\mathbb{I}_{[\underline{c}_i, \overline{c}_i]}(\tilde{c}_i)] = 1\}.$$

By using duality theory, we have:

$$\begin{aligned} \sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}\left[(\tilde{c}_i - d_i)^+ - \tilde{c}_i x_i\right] = \quad \min \quad & y_{i0} + \mu_i y_{i1} + (\mu_i^2 + \sigma_i^2) y_{i2} - \mu_i x_i \\ \text{s.t.} \quad & y_{i0} + y_{i1} c_i + y_{i2} c_i^2 - (c_i - d_i) \geq 0, \quad \forall c_i \in [\underline{c}_i, \overline{c}_i], \\ & y_{i0} + y_{i1} c_i + y_{i2} c_i^2 \geq 0, \quad \forall c_i \in [\underline{c}_i, \overline{c}_i]. \end{aligned} \tag{4.9}$$

By applying the S-lemma to the constraints of the above problem, problem (4.9) can be formulated as

$$\begin{aligned} \min \quad & y_{i0} + \mu_i y_{i1} + (\mu_i^2 + \sigma_i^2) y_{i2} - \mu_i x_i \\ \text{s.t.} \quad & \tau_{i1} \geq 0, \quad y_{i0} + d_i + \underline{c}_i \overline{c}_i \tau_{i1} \geq 0, \quad y_{i2} + \tau_{i1} \geq 0, \\ & \tau_{i2} \geq 0, \quad y_{i0} + \underline{c}_i \overline{c}_i \tau_{i2} \geq 0, \quad y_{i2} + \tau_{i2} \geq 0, \\ & \left\| \begin{matrix} y_{i1} - 1 - (\underline{c}_i + \overline{c}_i)\tau_{i1} \\ y_{i0} + d_i + (\underline{c}_i \overline{c}_i - 1)\tau_{i1} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + d_i + (\underline{c}_i \overline{c}_i + 1)\tau_{i1} + y_{i2}, \\ & \left\| \begin{matrix} y_{i1} - (\underline{c}_i + \overline{c}_i)\tau_{i2} \\ y_{i0} + (\underline{c}_i \overline{c}_i - 1)\tau_{i2} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + (\underline{c}_i \overline{c}_i + 1)\tau_{i2} + y_{i2}. \end{aligned} \tag{4.10}$$

The problem of minimizing WCVaR of regret can be formulated in this case as the mixed integer SOCP:

$$\begin{aligned} \min_{\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{z}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \boldsymbol{y}, \boldsymbol{\tau}} \quad & \left(\boldsymbol{b}^T \boldsymbol{\lambda}_1 + \boldsymbol{e}^T \boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha} \sum_{i=1}^N z_i + \frac{1}{1-\alpha} \sum_{i=1}^N \left(y_{i0} + \mu_i y_{i1} + (\mu_i^2 + \sigma_i^2) y_{i2} \right) - \frac{1}{1-\alpha} \boldsymbol{\mu}^T \boldsymbol{x} \right) \\ \text{s.t.} \quad & \tau_{i1} \geq 0, \quad y_{i0} + d_i + \underline{c}_i \overline{c}_i \tau_{i1} \geq 0, \quad y_{i2} + \tau_{i1} \geq 0, \quad \forall\, i \in [N], \\ & \tau_{i2} \geq 0, \quad y_{i0} + \underline{c}_i \overline{c}_i \tau_{i2} \geq 0, \quad y_{i2} + \tau_{i2} \geq 0, \quad \forall\, i \in [N], \end{aligned} \tag{4.11}$$

$$\left\| \begin{matrix} y_{i1} - 1 - (\underline{c}_i + \overline{c}_i)\tau_{i1} \\ y_{i0} + d_i + (\underline{c}_i\overline{c}_i - 1)\tau_{i1} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + d_i + (\underline{c}_i\overline{c}_i + 1)\tau_{i1} + y_{i2}, \quad \forall\, i \in [N],$$

$$\left\| \begin{matrix} y_{i1} - (\underline{c}_i + \overline{c}_i)\tau_{i2} \\ y_{i0} + (\underline{c}_i\overline{c}_i - 1)\tau_{i2} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + (\underline{c}_i\overline{c}_i + 1)\tau_{i2} + y_{i2}, \quad \forall\, i \in [N],$$

$(4.7a), (4.7b)$ and $(4.7c)$.

The regret formulations identified in this section are compact size mixed integer conic programs and generalize to higher order moments using mixed integer semidefinite programs.

5 Polynomial Solvability for Regret in Subset Selection

In this section, we identify a polynomial time algorithm to solve the probabilistic regret model for subset selection. Assume that the weight vector $\tilde{\boldsymbol{c}}$ for a set of items $\{1, \ldots, N\}$ is random. The marginal distribution of each $\tilde{c}_i$ is given as P_i. In the deterministic subset selection problem, the objective is to choose a subset of K items of maximum total weight. In the probabilistic regret model, the objective is to minimize the worst-case conditional value-at-risk of regret. This problem is formulated as

$$\min_{x \in \mathcal{X}} \mathrm{WCVaR}_\alpha \left(Z(\tilde{\boldsymbol{c}}) - \tilde{\boldsymbol{c}}^T \boldsymbol{x} \right), \tag{5.1}$$

where the feasible region is:

$$\mathcal{X} = \left\{ x \in \{0,1\}^N : \sum_{i=1}^{N} x_i = K \right\}.$$

For the subset selection problem, $Z(\cdot)$ is computed as the optimal objective value to the linear program:

$$Z(\boldsymbol{c}) = \max \left\{ \boldsymbol{c}^T \boldsymbol{y} \mid \boldsymbol{e}^T \boldsymbol{y} = K, 0 \leq \boldsymbol{y} \leq \boldsymbol{e} \right\}.$$

Strong duality of linear programming implies that it can be reformulated as:

$$Z(\boldsymbol{c}) = \min \left\{ \boldsymbol{e}^T \boldsymbol{\lambda} + K\lambda_0 \mid \boldsymbol{\lambda} \geq \boldsymbol{c} - \lambda_0 \boldsymbol{e}, \boldsymbol{\lambda} \geq 0 \right\} = \min_{\lambda_0} \sum_{i=1}^{N} (c_i - \lambda_0)^+ + K\lambda_0.$$

Using Proposition 1, the probabilistic regret model for subset selection is formulated as:

$$\min_{\lambda_0, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{d} \in \Omega} \left(\sum_{i=1}^{N} [d_i - \lambda_0]^+ + K\lambda_0 + \frac{\alpha}{1-\alpha} \boldsymbol{d}^T \boldsymbol{x} - \frac{1}{1-\alpha} \boldsymbol{\mu}^T \boldsymbol{x} + \frac{1}{1-\alpha} \sum_{i=1}^{N} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ \right). \tag{5.2}$$

Observe that for a fixed λ_0, the objective function of (5.2) is separable in d_i. Define

$$F_i(d_i, x_i, \lambda_0) = [d_i - \lambda_0]^+ + \frac{\alpha}{1-\alpha} d_i x_i + \frac{1}{1-\alpha} \mathbb{E}_{P_i} [\tilde{c}_i - d_i]^+ - \frac{1}{1-\alpha} \mu_i x_i.$$

Then problem (5.2) is expressed as:

$$\min_{\lambda_0, \boldsymbol{x}\in\mathcal{X}, \boldsymbol{d}\in\Omega} \sum_{i=1}^{N} F_i(d_i, x_i, \lambda_0) + K\lambda_0. \tag{5.3}$$

For fixed λ_0 and x_i, $F_i(d_i, x_i, \lambda_0)$ is a convex function of d_i. Denote a minimizer of this function as $d_i^*(x_i, \lambda_0) = \text{argmin}_{d_i\in\Omega_i} F_i(d_i, x_i, \lambda_0)$. Define the minimizers:

$$\beta_i(\lambda_0) = \underset{d_i\in\Omega_i}{\text{argmin}}\, F_i(d_i, 1, \lambda_0), \qquad \gamma_i(\lambda_0) = \underset{d_i\in\Omega_i}{\text{argmin}}\, F_i(d_i, 0, \lambda_0).$$

Since $x_i \in \{0, 1\}$, this implies:

$$d_i^*(x_i, \lambda_0) = \beta_i(\lambda_0)x_i + \gamma_i(\lambda_0)(1 - x_i).$$

For simplicity, we will denote $\beta_i(\lambda_0), \gamma_i(\lambda_0)$ and $d_i^*(x_i, \lambda_0)$ by β_i, γ_i and d_i^* by dropping the explicit dependence on the parameters. By substituting in the expression for d_i^* with the observation that $x_i \in \{0, 1\}$, we have

$$\begin{aligned} F_i(d_i^*, x_i, \lambda_0) \;&=\; (\beta_i - \lambda_0)^+ x_i + (\gamma_i - \lambda_0)^+(1 - x_i) + \frac{\alpha}{1-\alpha}\beta_i x_i \\ &\quad + \frac{1}{1-\alpha}\mathbb{E}_{P_i}\Big[(\tilde{c}_i - \beta_i)^+ x_i + (\tilde{c}_i - \gamma_i)^+(1 - x_i)\Big] - \frac{1}{1-\alpha}\mu_i x_i \\ &=\; \Big((\beta_i - \lambda_0)^+ - (\gamma_i - \lambda_0)^+ + \frac{\alpha}{1-\alpha}\beta_i + \frac{1}{1-\alpha}\mathbb{E}_{P_i}[(\tilde{c}_i - \beta_i)^+ - (\tilde{c}_i - \gamma_i)^+] - \frac{1}{1-\alpha}\mu_i\Big)x_i \\ &\quad + (\gamma_i - \lambda_0)^+ + \frac{1}{1-\alpha}\mathbb{E}_{P_i}[\tilde{c}_i - \gamma_i]^+. \end{aligned}$$

Define an N dimensional vector $\boldsymbol{h}(\lambda_0)$ and a scalar $h_0(\lambda_0)$ with

$$\begin{aligned} h_i(\lambda_0) =& (\beta_i - \lambda_0)^+ - (\gamma_i - \lambda_0)^+ + \frac{\alpha}{1-\alpha}\beta_i + \frac{1}{1-\alpha}\mathbb{E}_{P_i}[(\tilde{c}_i - \beta_i)^+ - (\tilde{c}_i - \gamma_i)^+] - \frac{1}{1-\alpha}\mu_i, \quad i \in [N], \\ h_0(\lambda_0) =& \sum_{i=1}^{N}(\gamma_i - \lambda_0)^+ + \frac{1}{1-\alpha}\sum_{i=1}^{N}\mathbb{E}_{P_i}[\tilde{c}_i - \gamma_i]^+ + K\lambda_0. \end{aligned}$$

Problem (5.2) is thus reformulated as:

$$\min_{\lambda_0} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{h}(\lambda_0)^T\boldsymbol{x} + h_0(\lambda_0). \tag{5.4}$$

For a fixed λ_0, the inner optimization problem over the uniform matroid can be done efficiently in $O(N)$ time. The next proposition shows that for discrete marginal distributions, the search for the optimal value of λ_0 can be restricted to a finite set.

Proposition 3. *Assume that the marginal distribution of $\tilde{c}_i$ is discrete and*

$$\tilde{c}_i \sim c_{ij} \textit{ with probability } p_{ij}, \quad j \in [J_i], i \in [N].$$

The objective function of (5.4) attains its minimum in the finite set:

$$\lambda_0 \in \{c_{ij} \mid j \in [J_i], i \in [N]\}.$$

Proof. For discrete marginal distributions, problem (5.2) is formulated as:

$$\min_{\lambda_0, \boldsymbol{x}\in\mathcal{X}, \boldsymbol{d}\in\Omega} \left(\sum_{i=1}^{N} [d_i - \lambda_0]^+ + K\lambda_0 + \frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x} - \frac{1}{1-\alpha}\boldsymbol{\mu}^T\boldsymbol{x} + \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{J_i}(c_{ij} - d_i)^+ p_{ij} \right). \quad (5.5)$$

For a fixed vector $\boldsymbol{d}$, sort the components of the vector such that $d^{(1)} \geq d^{(2)} \geq \dots d^{(N)}$. Let $\lambda_0^* = d^{(K)}$ be the K-th largest component of $\boldsymbol{d}$. Then

$$\min_{\lambda_0} \left(\sum_{i=1}^{N} [d_i - \lambda_0]^+ + K\lambda_0 \right) = \max \left\{ \boldsymbol{d}^T\boldsymbol{y} : \sum_{i=1}^{N} y_i = K, 0 \leq y_i \leq 1, i \in [N] \right\} = \sum_{i=1}^{K} d^{(i)} = \sum_{i=1}^{N} [d_i - \lambda_0^*]^+ + K\lambda_0^*,$$

where the first equality comes from linear programming duality. Hence the minimizer λ_0 can be chosen as the K-th largest component of $\boldsymbol{d}$. We claim that for each $i \in [N]$, the i-th component of all the optimal $\boldsymbol{d}$ can be chosen in the set $\{c_{ij} \mid j \in [J_i]\}$. To prove this claim, the problem of minimizing the worst-case conditional value-at-risk is formulated as:

$$\begin{aligned}
&\min_{\boldsymbol{x}\in\mathcal{X}} \min_{\boldsymbol{d}\in\Omega} \max_{\boldsymbol{y}\in\mathrm{conv}(\mathcal{X})} \left(\boldsymbol{d}^T\boldsymbol{y} + \frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x} - \frac{1}{1-\alpha}\boldsymbol{\mu}^T\boldsymbol{x} + \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{J_i} p_{ij}[c_{ij} - d_i]^+ \right) \\
= \;& \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathrm{conv}(\mathcal{X})} \min_{\boldsymbol{d}\in\Omega} \left(\boldsymbol{d}^T\boldsymbol{y} + \frac{\alpha}{1-\alpha}\boldsymbol{d}^T\boldsymbol{x} - \frac{1}{1-\alpha}\boldsymbol{\mu}^T\boldsymbol{x} + \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{J_i} p_{ij}[c_{ij} - d_i]^+ \right) \\
= \;& \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathrm{conv}(\mathcal{X})} \left(\sum_{i=1}^{N} \min_{d_i\in\Omega_i} \left(d_i \left(y_i + \frac{\alpha}{1-\alpha}x_i \right) + \frac{1}{1-\alpha}\sum_{j=1}^{J_i} p_{ij}[c_{ij} - d_i]^+ \right) - \frac{1}{1-\alpha}\mu_i x_i \right), \quad (5.6)
\end{aligned}$$

where $\mathrm{conv}(\mathcal{X})$ is the convex hull of the set $\mathcal{X}$. For fixed $\boldsymbol{x}$ and $\boldsymbol{y}$, the function $d_i(y_i + \frac{\alpha}{1-\alpha}x_i) + \frac{1}{1-\alpha}\sum_{j=1}^{J_i} p_{ij}[c_{ij} - d_i]^+$ is a piecewise linear function in d_i, and its minimum value over $d_i \in \Omega_i$ occurs at one of the break points $\{c_{ij} \mid j \in [J_i]\}$. Since the optimal λ_0 is the K-th largest component of the optimal $\boldsymbol{d}$, the result holds. □

By combining Proposition 3 and formulation (5.4), we provide a polynomial time algorithm to minimize the WCVaR of regret for the subset selection problem. The algorithm is described as follows:

Algorithm 1: Minimization of WCVaR of regret for subset selection.

Input: K, probability level α, discrete marginal distribution $c_{ij}, p_{ij}, j \in [J_i], i \in [N]$.

Output: Optimal decision $\boldsymbol{x}$, the minimum WCVaR of regret obj.

1 Sort $\{c_{ij}\}_{j\in[J_i],i\in[N]}$ as a increasing sequence in the set Λ.

2 Delete the repeated numbers in Λ to get a new set Λ_0.

3 $\boldsymbol{x} = 0$, $obj = \infty$

4 **for** $\lambda_0 \in \Lambda_0$ **do**

5   **for** $i = 1, \ldots, N$ **do**

6         $\beta_i = \text{argmin}_{d_i \in \Omega_i} F(d_i, 1, \lambda_0)$, $\gamma_i = \text{argmin}_{d_i \in \Omega_i} F(d_i, 0, \lambda_0)$,

7         $h_i = (\beta_i - \lambda_0)^+ - (\gamma_i - \lambda_0)^+ + \frac{\alpha}{1-\alpha}\beta_i + \frac{1}{1-\alpha}\sum_{j=1}^{J_i}[(c_{ij} - \beta_i)^+ - (c_{ij} - \gamma_i)^+]p_{ij} - \frac{1}{1-\alpha}\mu_i$,

8   **end**

9     $h_0 = \sum_{i=1}^{N}(\gamma_i - \lambda_0)^+ + \frac{1}{1-\alpha}\sum_{i=1}^{N}\sum_{j=1}^{J_i}(c_{ij} - \gamma_i)^+ p_{ij} + K\lambda_0$.

10     $\boldsymbol{y} = \text{argmin}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{h}^T\boldsymbol{x}$, $val = \boldsymbol{h}^T\boldsymbol{y} + h_0$.

11   **if** $val < obj$ **then**

12         $\boldsymbol{x} = \boldsymbol{y}, \quad obj = val$.

13   **end**

14 **end**

Proposition 4. *The running time of Algorithm 1 is* $O(N^2 J_{max}^2)$ *where* $J_{max} = \max_{i\in[N]} J_i$. *This solves formulation (5.1) to optimality.*

Proof. Sorting in step 1 can be done in $O(NJ_{max}\log(NJ_{max}))$. The function $F(d_i, 1, \lambda_0)$ is a piecewise linear function with respect to d_i. To get the optimal d_i, the values of $F(d_i, 1, \lambda_0)$ are evaluated at the break points c_{ij}, $j \in [J_i]$ and λ_0. The complexity of evaluating β_i is thus $O(J_i)$. Likewise for γ_i. The complexity of evaluating the vector $\boldsymbol{h}(\lambda_0)$ and the scalar $h_0(\lambda_0)$ in steps 5 to 9 is thus $O(NJ_{max})$. For subset selection, the complexity of finding $\text{argmin}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{h}^T\boldsymbol{x}$ is $O(N)$. Moreover, $|\Lambda_0| \leq NJ_{max}$, hence the total computational complexity for Algorithm 1 is $O(N^2 J_{max}^2)$. □

In the marginal moment model, if (a) the mean and range are given, the worst-case marginal distribution is a two-point discrete distribution; and (b) the mean, range and mean absolute deviation are given, the worst-case distribution is a three-point discrete distribution. The worst-case marginal distributions as discussed in Section 4.2 are fixed and can hence be treated as a special case of the discrete marginal distribution model. Thus in these two cases, Algorithm 1 solves the problem to optimality which brings us to the following result.

Theorem 4. *The problem of minimizing the worst-case* CVaR *of regret for the subset selection problem is solvable in polynomial time when (a) the range and mean, and (b) the range, mean and mean absolute deviation are given.*

This extends the polynomial complexity result when only the range is given (see Averbakh [5] and Conde [14]). Algorithm 1 is related to the earlier algorithms of Averbakh [5] and Conde [14] for the range case. When only the $[\underline{c}_i, \overline{c}_i], i \in [N]$ of each $\tilde{c}_i$ is known, the problem of minimizing the worst-case CVaR of the regret reduces to the interval uncertainty minmax regret problem. In this case, the worst-case marginal distribution the Dirac measure $\delta_{\hat{\mathbf{c}}(\boldsymbol{x})}$, where $\hat{c}_i(\boldsymbol{x}) = \underline{c}_i x_i + \overline{c}_i(1 - x_i)$. It is easy to check that the variables in Algorithm 1 are then $\beta_i = \underline{c}_i$, $\gamma_i = \overline{c}_i$, $h_i = [\underline{c}_i - \lambda_0]^+ - [\overline{c}_i - \lambda_0]^+ - \underline{c}_i, i \in [N]$, and $h_0 = \sum_{i=1}^N [\overline{c}_i - \lambda_0]^+ + K\lambda_0$. The running time of Algorithm 1 is $O(N^2)$ algorithm for the minmax regret subset selection problem in this case. Since the optimal λ_0 is the Kth largest value of the optimal $d_i^*(x_i) = \underline{c}_i(x_i) + \overline{c}_i(x_i)$, the feasible set Λ_0 can be further reduced to a smaller set with cardinality $2K$ (see the discussion in Conde [14]). Furthermore, if $K > N/2$ the problem can be transformed in $O(N)$ time to an equivalent problem with $K' \leq N/2$ (see Averbakh [5]). Algorithm 1 is thus a generalization of these algorithms for the minmax regret subset selection problem.

6 Numerical Examples

6.1 Shortest Path

Consider a directed, acyclic network $G = (V, A)$ with a finite set of vertices V and a finite set of arcs A. Associated with each arc, is the duration (length) of that arc. The goal is to find the shortest path from a fixed source node to the sink node. When the arc lengths are deterministic, the shortest path problem can be solved efficiently. However, when the arc lengths are random, the definition of a "shortest path" has to be suitably modified.

Shortest paths under a stochastic setting is a well studied problem [23, 47, 6, 25, 35]. Some of the possible approaches to determine the "shortest path" in the stochastic framework are discussed next.

1. **Expected Shortest Path:** The classical approach chooses the path with the shortest length in an expected sense.

2. **Most Likely Path:** Kamburowski [23] defined the optimality index of a path to be the probability that it is the shortest path. The "shortest path" in this case is defined as the path with the greatest optimality index and is termed as the most likely path. Unlike the expected shortest path, computing the most likely path is highly challenging even for moderate size networks.

3. **Absolute Robust Path:** An absolute robust path is defined as the path that is the shortest under the worst-case scenario. In the interval uncertainty model, this path is found by solving the shortest path problem on the graph when the arc lengths are replaced by the largest length for each arc.

4. **Minmax Regret Path:** In recent years, the shortest path with the minmax regret criterion has been proposed as an alternative decision criterion. In the interval uncertainty case, Zieliński [47] showed that the minmax regret shortest path problem is NP-hard even when the graph is restricted to be directed, acyclic and planar with vertex degrees at most three. Mixed integer linear programs to solve the interval uncertainty minmax regret path have been developed in Yaman et. al. [44].

5. **Minimum WCVaR Cost Path:** Choose the path by minimizing the WCVaR of the cost.

6. **Minimum WCVaR Regret Path:** Choose the path by minimizing the WCVaR of regret.

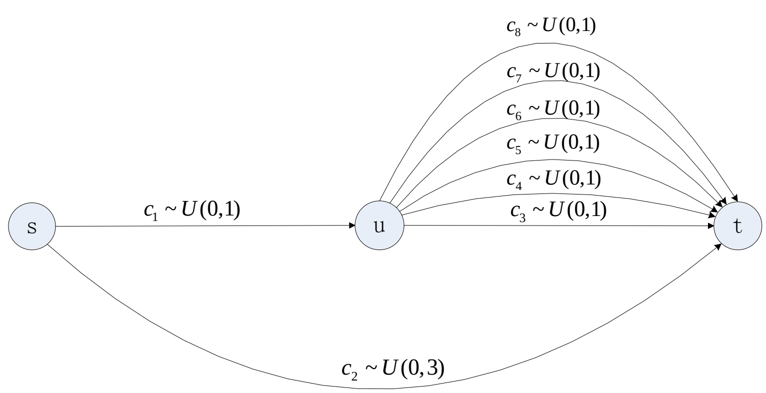


Figure 2: Network for Example 1

Example 1. *In figure 2, arc length* $\tilde{c}_2 \sim$ *uniform*$(0,3)$, *and the other arc length* $\tilde{c}_i \sim$ *uniform*$(0,1)$, $i \neq 2$. *The goal is to find a shortest path from s to t. This example is from Reich and Lopes [35].*

The choices of paths passing through the intermediate node u have expected length 1, worst-case length 2, and maximum regret 2, while the path consisting of $\tilde{c}_2$ has expected length 1.5, worst-case length

3, and maximum regret 3. In the sense of (1) Expected shortest path, (3) Absolute robust path, (4) Minmax regret path, and (5) Min WCVaR Path, the "shortest path" is any path passing through the intermediate node u. In the sense of (2) Most likely path, the "shortest path" consists of $\tilde{c}_2$ (see Reich and Lopes[35]). To solve the probabilistic regret model, we use only the marginal moment information. Consider the following three cases (a) known range and mean, (b) known range, mean and mean absolute deviation and (c) known range, mean and variance. In all the three cases, by choosing the probability level $\alpha \in [0, 0.99)$ the optimal decision (6) Min WCVaR Regret Path is always one of the paths passing through the intermediate node u, which is the same as the decision of (1), (3), (4) and (5). This result is in agreement with the intuition that while the path consisting of arc $\tilde{c}_2$ is the most likely shortest path, in terms of worst-case value and regret it is not the best one.

Example 2. *Reconsider the example shown in Figure 1 in Section 1 with a network that consists of four nodes and five arcs. All the length of the arcs are known to lie in interval ranges with the means and standard deviations of the lengths given.*

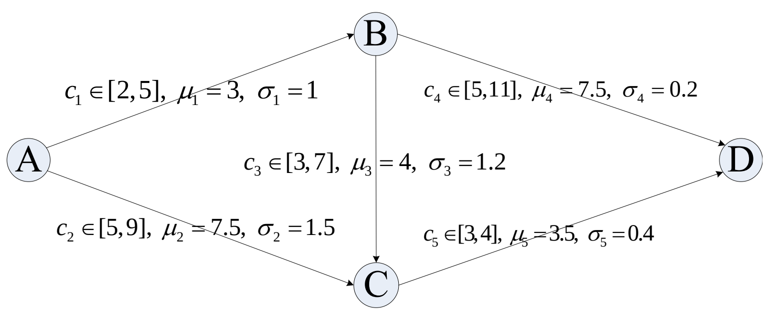


Figure 3: Network for Example 2

The network in Example 2 is the Wheatstone bridge network with the objective of finding the shortest path from node A to node D. The solutions identified from the Expected shortest path, Absolute robust shortest path, Minmax regret path, Min WCVaR cost path and Min WCVaR regret path are provided in Table 2.

While the expected approach path uses only the mean and the absolute robust and minmax regret approaches use only range information, while minimizing the WCVaR of regret can deal with more information. As the probability level α is varied, the minimum WCVaR regret decision changes. This is consistent with the observation that α captures the decision-maker's aversion to regret where a larger α implies higher aversion to regret. If the decision-maker is regret neutral, by setting $\alpha = 0$, the method reduces to the expected shortest path method where the mean is specified for each arc. Moreover, the

Table 2: The stochastic "shortest path"

Methods	"Shortest path"	Information
Expected shortest path	$1-4$ or $1-3-5$	Mean
Absolute robust path	$2-5$	Range
Minmax regret path	$2-5$	Range
Min WCVaR cost path	$2-5$ if $0.5001 \leq \alpha < 1$ $1-3-5$ if $0 < \alpha \leq 0.5000$	Range and mean
Min WCVaR cost path	$2-5$ if $0.8261 \leq \alpha < 1$ $1-4$ if $0 < \alpha \leq 0.8260$	Range, mean and standard deviation
Min WCVaR regret path	$2-5$ if $0.6667 \leq \alpha < 1$ $1-3-5$ if $0 < \alpha \leq 0.6666$	Range and mean
Min WCVaR regret path	$2-5$ if $0.6883 \leq \alpha < 1$ $1-4$ if $0 < \alpha \leq 0.6882$	Range, mean and standard deviation

choice of the solution is sensitive to the probability information available. If we only use the range and mean information, path 1-3-5 always dominates path 2-5 if $\alpha \leq 0.6666$, although they have the same expected traveling time. This should correspond to our intuition since the range of the edge c_4 is significantly larger than the range of edge c_3 and c_5, and there are more edges in path 1-3-5 which can spread more risk than path 1-4. However, when the standard deviation information is also involved, the standard deviation of edge c_4 is much smaller than the standard deviation of c_3 and c_5, which means that the risk of c_4 is smaller. Hence in this case, it is intuitive to expect that path 1-4 dominates path 1-3-5 as indicated in Table 2. The optimal decision of minimizing the WCVaR of cost is similar to the decision of minimizing the WCVaR of regret. In this example, this can be partly explained by the observation that the absolute robust path and minmax regret path are the same. When α is close to 0, Min WCVaR cost path and Min WCVaR regret path reduce to the expected shortest path, and when α goes to 1, the solutions reduce to the absolute robust path and minmax regret path which are the same in this example.

Example 3. *The previous two examples used small-sized networks. We now create a fictitious network in the form of a square grid graph with width and height both equal to H as in Figure 4. There are H^2 nodes and $2H(H-1)$ arcs in the graph. The start node is at the left bottom corner and the end node is at the right upper corner. Each arc on the graph proceeds either towards the right node or the upper node.*

We evaluate the CPU times needed to minimize the WCVaR of regret and cost with randomly

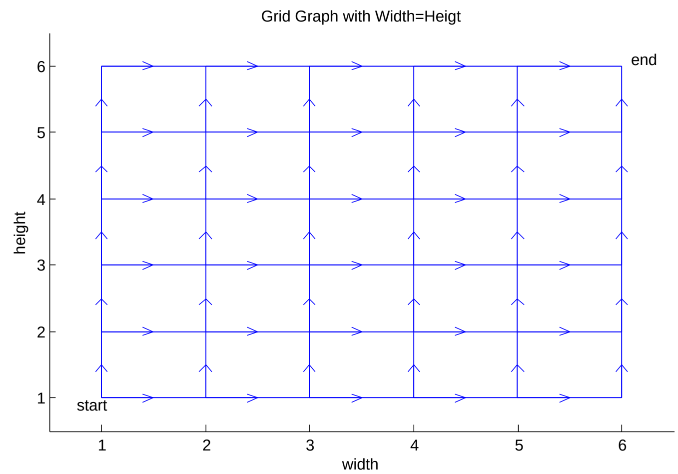
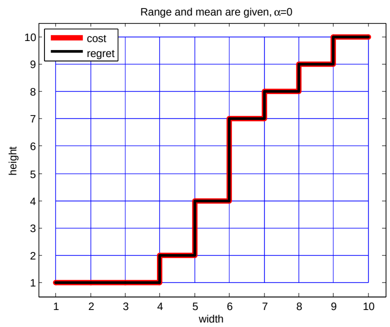


Figure 4: Grid Graph with $H = 6$

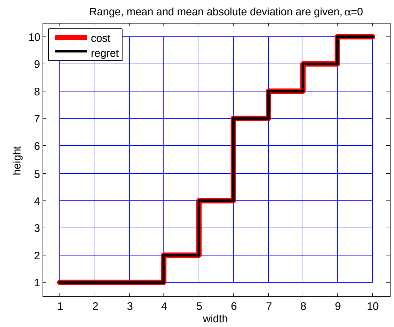
generated data. In this experiment, the interval range for each arc length $[\underline{c}_i, \overline{c}_i]$ is randomly generated with $\underline{c}_i = \min\{a_i, b_i\}$, $\overline{c}_i = \max\{a_i, b_i\}$, where a_i, b_i are chosen from the uniform distribution $U[1, 10]$. The mean is randomly generated as $\mu_i \sim U[\underline{c}_i, \overline{c}_i]$. Define $\bar{\delta}_i = 2\frac{(\overline{c}_i - \mu_i)(\mu_i - \underline{c}_i)}{\overline{c}_i - \underline{c}_i}$ as the largest mean absolute deviation when the mean and range of $\tilde{c}_i$ are given. Let the mean absolute deviation of $\tilde{c}_i$ be randomly generated by $\delta_i \sim U[0, \bar{\delta}_i]$. We report the CPU time taken to minimize the WCVaR of regret and cost for the following two cases of the marginal moment model: (a) range $[\underline{c}_i, \overline{c}_i]$ and mean μ_i are given and (b) range $[\underline{c}_i, \overline{c}_i]$, mean μ_i and mean absolute deviation δ_i are given. The results are shown in Table 3.

The computational studies were implemented in Matlab R2012a on an Intel Core 2 Duo CPU 2.8GHz laptop with 4 GB of RAM. In Table 3, the CPU time (in the format of seconds) is the average execution time for 10 randomly generated instances, and ** indicates that the instances ran out of memory. The CPU time taken to minimize the WCVaR is very small (< 0.03 seconds), since this problem is solvable as a linear programming problem (see (3.22) to (3.25)). To minimize the WCVaR of regret, we use CPLEX to solve the binary integer linear programming problem. When (a) range and mean information are given, we can solve the regret problem to optimality for $H = 25$ (i.e. 625 nodes and 1200 edges) in around 4 seconds; when (b) range, mean and mean absolute deviation are given, we can solve the regret problem to optimality for $H = 23$ (i.e. 529 nodes and 1012 edges) in around 5 seconds.

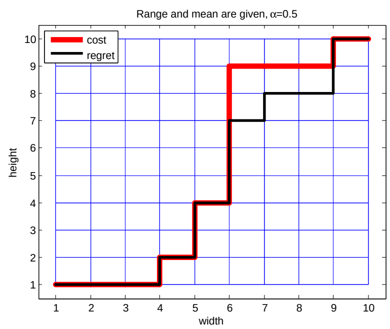
Next we compare the optimal paths obtained from minimizing the WCVaR of cost and regret at different probability levels α. Let $H = 10$, and assume $\underline{c}_i, \overline{c}_i, \mu_i, \delta_i,\ i = 1, 2, \ldots, 2H(H-1)$ are given. The optimal paths that minimize the WCVaR of cost and regret with (a) the range and mean information, and (b) the range, mean and mean absolute deviation information are provided in Figure 5. From



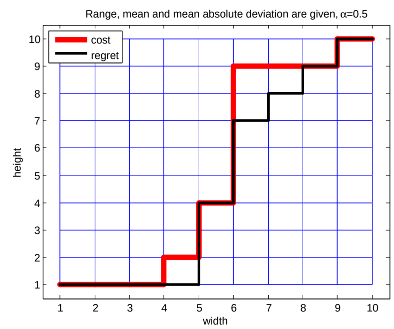
(a) $[\underline{c}_i, \overline{c}_i]$ and μ_i are given, $\alpha = 0$



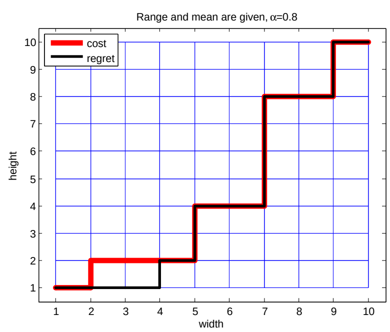
(b) $[\underline{c}_i, \overline{c}_i]$, μ_i and δ_i are given, $\alpha = 0$



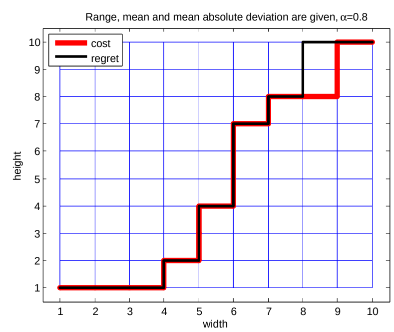
(c) $[\underline{c}_i, \overline{c}_i]$ and μ_i are given, $\alpha = 0.5$



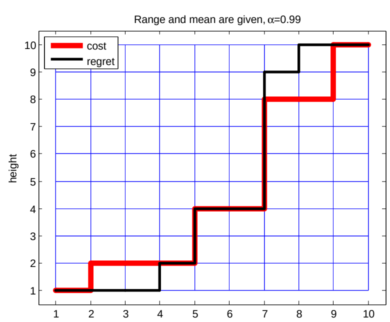
(d) $[\underline{c}_i, \overline{c}_i]$, μ_i and δ_i are given, $\alpha = 0.5$



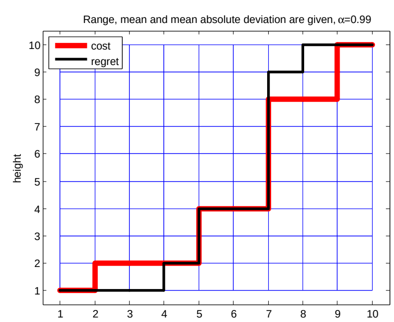
(e) $[\underline{c}_i, \overline{c}_i]$ and μ_i are given, $\alpha = 0.8$



(f) $[\underline{c}_i, \overline{c}_i]$, μ_i and δ_i are given, $\alpha = 0.8$



(g) $[\underline{c}_i, \overline{c}_i]$ and μ_i are given, $\alpha = 0.99$



(h) $[\underline{c}_i, \overline{c}_i]$, μ_i and δ_i are given, $\alpha = 0.99$

Figure 5: Optimal paths that minimize the WCVaR of cost and regret

Table 3: Average CPU time to minimize the WCVaR of cost and regret, $\alpha = 0.8$

H	Nodes	Arcs	$\min_{\boldsymbol{x}\in\mathcal{X}} \text{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x})$		$\min_{\boldsymbol{x}\in\mathcal{X}} \text{WCVaR}_\alpha(Z(\tilde{\boldsymbol{c}}) - \tilde{\boldsymbol{c}}^T\boldsymbol{x})$	
			$[\underline{c}_i, \overline{c}_i], \mu_i$ are given	$[\underline{c}_i, \overline{c}_i], \mu_i, \delta_i$ are given	$[\underline{c}_i, \overline{c}_i], \mu_i$ are given	$[\underline{c}_i, \overline{c}_i], \mu_i, \delta_i$ are given
10	100	180	3.10e-03	0.00e+00	2.95e-01	3.62e-01
12	144	264	1.50e-03	4.70e-03	6.52e-01	8.11e-01
14	196	364	3.00e-03	3.20e-03	6.24e-01	9.63e-01
16	256	480	9.60e-03	3.20e-03	9.95e-01	1.07e+00
18	324	612	1.09e-02	1.60e-03	1.29e+00	1.86e+00
20	400	760	8.00e-03	9.30e-03	1.24e+00	1.96e+00
21	441	840	9.60e-03	1.26e-02	1.86e+00	2.41e+00
22	484	924	1.11e-02	1.40e-02	2.74e+00	3.68e+00
23	529	1012	1.59e-02	1.39e-02	3.31e+00	4.46e+00
24	576	1104	1.57e-02	1.40e-02	3.34e+00	**
25	625	1200	2.05e-02	1.69e-02	3.68e+00	**
26	676	1300	1.99e-02	2.07e-02	**	**

Figure 5, we see that when $\alpha = 0$, the optimal paths that minimize the WCVaR of cost and regret are the same regardless of whether (a) range and mean information are given or (b) range, mean and mean absolute deviation are given. In this case, the two models reduce to the deterministic shortest path problem where every edge length equals to its mean. When α is close to 1, the Min-WCVaR-cost path approaches the absolute robust path, and the Min-WCVaR-regret path approaches the minmax regret path. For intermediate values of α, the Min-WCVaR-cost path and the Min-WCVaR-regret path are different. The optimal paths also differ based on whether information on the mean absolute deviation is available or not.

6.2 Subset Selection

Consider the problem of investing in a subset of K projects with random payoffs from a set of N projects. The potential payoff of each project is random where the statistics of the payoff for each project can be estimated through experts.

Example 4. *Consider the goal of choosing 4 projects from a set of 10 projects so as to maximize the total payoff. The payoff of each project is random with the estimated range, mean and standard deviation provided in Table 4. In this example, we compare the optimal decisions of minimizing the* WCVaR *of regret and cost respectively.*

Table 4: Range, mean, and standard deviation of the project profits

Project	1	2	3	4	5	6	7	8	9	10
Range $[\underline{c}_i, \bar{c}_i]$	[2, 23]	[9, 17]	[8, 16]	[7, 29]	[3, 34]	[3, 34]	[1, 34]	[0, 35]	[3, 30]	[1, 35]
Mean μ_i	18	10	14	14	20	6	8	10	12	8
Standard deviation σ_i	5	2	1	8	10	3	7	10	6	5

The optimal set of projects to invest in for each of the following objectives are:

- Maximize the expected profit: (1, 3, 4, 5).
- Maximize the worst-case profit: (2, 3, 4, 5) or (2, 3, 4, 6) or (2, 3, 4, 9).
- Minimize the maximum regret: (4, 5, 6, 10).

By varying the parameter α, we next compare the decisions obtained from minimizing the WCVaR of cost and WCVaR of regret. The results are provided in Table 6.2.

Table 5: Optimal decisions for different probability level α

α	$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha(-\tilde{\boldsymbol{c}}^T\boldsymbol{x})$		$\min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{WCVaR}_\alpha(Z(\tilde{\boldsymbol{c}})-\tilde{\boldsymbol{c}}^T\boldsymbol{x})$	
	$[\underline{c}_i, \bar{c}_i], \mu_i$ are given	$[\underline{c}_i, \bar{c}_i], \mu_i, \sigma_i$ are given	$[\underline{c}_i, \bar{c}_i], \mu_i$ are given	$[\underline{c}_i, \bar{c}_i], \mu_i, \sigma_i$ are given
0.99	(2,3,4,9)	(2,3,4,9)	(4,5,6,10)	(4,5,6,10)
0.90	(2,3,4,9)	(1,2,3,4)	(4,5,6,10)	(4,5,8,9)
0.80	(2,3,4,9)	(1,2,3,4)	(4,5,8,10)	(1,4,5,8)
0.70	(1,2,3,4)	(1,2,3,4)	(4,5,8,9)	(1,3,4,5)
0.60	(1,2,3,4)	(1,2,3,5)	(1,4,5,9)	(1,3,4,5)
0.50	(1,2,3,4)	(1,2,3,5)	(1,4,5,9)	(1,3,4,5)
0.40	(1,2,3,5)	(1,2,3,5)	(1,3,4,5)	(1,3,4,5)
0.30	(1,2,3,5)	(1,2,3,5)	(1,3,4,5)	(1,3,4,5)
0.20	(1,3,4,5)	(1,3,4,5)	(1,3,4,5)	(1,3,4,5)
0.10	(1,3,4,5)	(1,3,4,5)	(1,3,4,5)	(1,3,4,5)

In the model of minimizing the WCVaR of cost, the parameter α denotes the degree of risk aversion while in the model of minimizing the WCVaR of regret, the parameter α denotes the degree of regret aversion. Both of these models converge to maximizing the expected profit in the case with $\alpha = 0$. However, for larger values of the parameter α, the nature of the optimal decisions from these two models are very different. For example when $\alpha = 0.8$ and only the range and mean information are used, the set of projects that minimize the WCVaR of cost are (2,3,4,9), and the set of projects that minimize the WCVaR of regret are (4,5,8,10). The set of projects (4,5,8,10) are more aggressive than the set of projects (2,3,4,9), since the former set of projects have more profit in the best-case and lesser profit in the worst-case. The regret criterion compares with the optimal set of projects that would be invested in if the payoffs were known and is hence more aggressive as compared to the absolute robust criterion.

Example 5. *In this experiment, the interval range for each item* $[\underline{c}_i, \overline{c}_i]$ *are randomly generated with* $\underline{c}_i = \min\{a_i, b_i\}$, $\overline{c}_i = \max\{a_i, b_i\}$, *with* a_i, b_i *generated from the uniform distribution* $U[0, 100]$. *The mean for each item is randomly generated as* $\mu_i \sim U[\underline{c}_i, \overline{c}_i]$. *Define* $\bar{\delta}_i = 2\frac{(\overline{c}_i - \mu_i)(\mu_i - \underline{c}_i)}{\overline{c}_i - \underline{c}_i}$ *as the largest mean absolute deviation when the mean and range of* $\tilde{c}_i$ *are given. Let the mean absolute deviation of* $\tilde{c}_i$ *be randomly generated by* $\delta_i \sim U[0, \bar{\delta}_i]$. *We test* **Algorithm 1** *for the following two cases of the marginal moment model: (a) range* $[\underline{c}_i, \overline{c}_i]$ *and mean* μ_i *are given and (b) range* $[\underline{c}_i, \overline{c}_i]$, *mean* μ_i *and mean absolute deviation* δ_i *are given.*

To compare the efficiency of Algorithm 1 with CPLEX's MIP solver (version 12.4), randomly generated instances were tested for different α's and K's. We compare the CPU times of the two methods in the following tables. First, we fix the value of α and K, and compare the CPU time for different N. Then, we fix the value of the dimension N, and tested the sensitivity of the running time of Algorithm 1 to the parameters α and K. In the tables, the CPU time (in the format of seconds) taken by Algorithm 1 to solve (5.2) and CPLEX's MIP solver to solve (4.8) are denoted by "time Alg1" and "time Cplex", respectively. The CPU time in the tables was the average execution time of 10 randomly generated instances. The instances with "$**$" indicates that it ran out of memory. From Table 6, it is clear that the CPU time taken by Algorithm 1 is significantly lesser than that taken by CPLEX's MIP solver. Even for extremely large values of N, Algorithm 1 was able to solve the problem to optimality in a reasonable amount of time (see Table 7). The CPU time for the algorithm is relatively insensitive to the parameters K and the probability level α (see Figure 6), indicating that Algorithm 1 is very robust and efficient.

To test the value of knowing probabilistic information, we compare the minmax regret solutions obtained by assuming that range and mean are given and then incorporating dispersion information with the mean absolute deviation. Let $\boldsymbol{x}_1$ denote the optimal solution when only the range and mean information are used, and $\boldsymbol{x}_2$ denote the optimal solution when the range, mean and mean absolute

Table 6: Computational results for $\alpha = 0.3, K = 0.4N$.

	(a) $[\underline{c}_i, \overline{c}_i], \mu_i$ are given		(b) $[\underline{c}_i, \overline{c}_i], \mu_i, \delta_i$ are given	
N	time Alg1	time Cplex	time Alg1	time Cplex
50	7.80e-003	2.06e-001	1.23e-002	1.89e-001
100	1.72e-002	2.76e-001	3.73e-002	2.42e-001
200	5.65e-002	5.76e-001	9.52e-002	3.42e-001
400	1.58e-001	1.53e+000	2.98e-001	7.04e-001
800	5.23e-001	**	9.98e-001	**

Table 7: CPU time of Algorithm 1 for solving large instances ($\alpha = 0.9, K = 0.3N$).

N	(a) $[\underline{c}_i, \overline{c}_i], \mu_i$ are given	(b) $[\underline{c}_i, \overline{c}_i], \mu_i, \delta_i$ are given
5000	1.79e+001	3.55e+001
10000	7.02e+001	1.40e+002
20000	2.56e+002	5.25e+002
40000	1.02e+003	2.28e+003
80000	4.91e+003	1.07e+004

deviation information are used. Define the gap between the two solutions as:

$$\text{gap} = \frac{\text{WCVaR}_\alpha(R(\boldsymbol{x}_1, \tilde{\boldsymbol{c}})) - \text{WCVaR}_\alpha(R(\boldsymbol{x}_2, \tilde{\boldsymbol{c}}))}{\text{WCVaR}_\alpha(R(\boldsymbol{x}_1, \tilde{\boldsymbol{c}}))},$$

where the WCVaR is calculated using range, mean and mean absolute deviation information. Note that while $\boldsymbol{x}_2$ is the optimal solution for $\text{WCVaR}_\alpha(R(\boldsymbol{x}, \tilde{\boldsymbol{c}}))$ in this case, $\boldsymbol{x}_1$ is potentially sub-optimal. Thus gap measures the improvement obtained from knowing the value of mean absolute deviation. We test the instances with randomly generated range $[\underline{c}_i, \overline{c}_i]$ and mean μ_i, $i = 1, \ldots, N$ as before. Let $\bar{\delta}_i = 2\frac{(\overline{c}_i - \mu_i)(\mu_i - \underline{c}_i)}{\overline{c}_i - \underline{c}_i}$ be the largest possible mean absolute deviation. Set $\delta_i = \beta\bar{\delta}_i$, and then vary β to test how the gap changes as β increases from 0 to 1. The gap in Figure 7 is the average gap for 10 randomly generated instances. From Figure 7, it clear that gap decreases to 0 as β increases to 1. This can be explained by observing that as the mean absolute deviation approaches the upper bound $\bar{\boldsymbol{\delta}}$, the decision $\boldsymbol{x}_1$ obtained from using only the range and mean information becomes increasingly tight. The shape of the curve also indicates that the relative improvement is more sensitive to the mean absolute deviation information particularly when it is smaller.

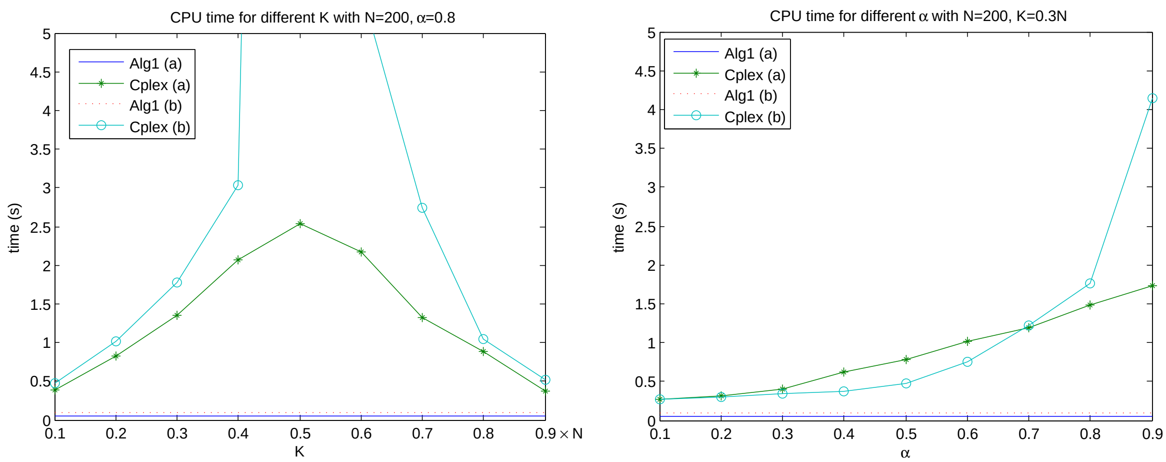


Figure 6: Sensitivity to the parameters K and α

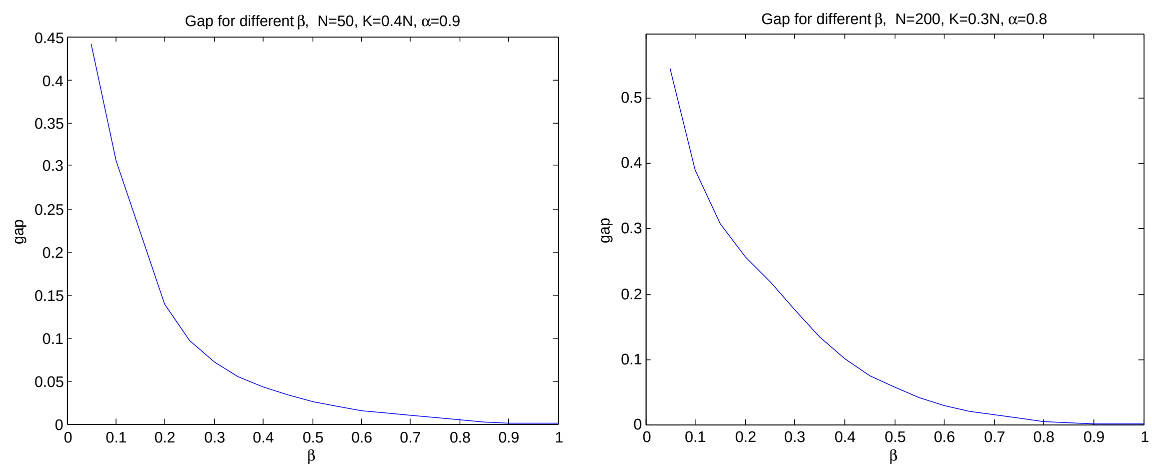


Figure 7: Gap for different β

7 Conclusions

In this paper, we have proposed a new probabilistic model of regret for combinatorial optimization problem. This generalizes the interval uncertainty model, by incorporating additional marginal distribution information on the data. By generalizing the earlier bounds of Meilijson and Nadas [30] to the regret framework, we provide mixed integer LP and mixed integer SOCP formulations for marginal distribution and marginal moment models. For the subset selection problem, a polynomial complexity result for the newly proposed probabilistic model of regret is derived. This polynomial time algorithm works for the case (a) range and mean are given, or (b) range, mean and mean absolute deviation are given. In the case (c) range, mean and standard deviation are given, the complexity of the probabilistic regret problem for subset selection remains an open question.

Acknowledgements

We would like to thank an Associate Editor and two anonymous reviewers for their valuable comments and suggestions to improve the paper.

References

[1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

[2] W. P. Adams and H. D. Sherali. Mixed-integer bilinear programming problems. *Mathematical Programming*, 59(1-3):279–305, 1993.

[3] S. Ahmed, U. Çakmak, and A. Shapiro. Coherent risk measures in inventory problems. *European Journal of Operational Research*, 182(1):226–238, 2007.

[4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[5] I. Averbakh. On the complexity of a class of combinatorial optimization problems with uncertainty. *Mathematical Programming*, 90(2):263–272, 2001.

[6] I. Averbakh and V. Lebedev. Interval data minmax regret network optimization problems. *Discrete Applied Mathematics*, 138(3):289–301, 2004.

[7] A. Ben-Tal, L. El Ghaoui, and A.S. Nemirovskiĭ. *Robust Optimization*. Princeton University Press, 2009.

[8] A. Ben-Tal and E. Hochman. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, 9(4):803–812, 1972.

[9] D. Bertsimas, K. Natarajan, and C.P. Teo. Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM Journal on Optimization*, 15(1):185–209, 2004.

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

[11] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98:49–71, 2003.

[12] J. R. Birge and M. J. Maddox. Bounds on expected project tardiness. *Operations Research*, 43(5):838–850, 1995.

[13] W. Chen, M. Sim, J. Sun, and C.P. Teo. From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58:470–485, 2010.

[14] E. Conde. An improved algorithm for selecting p items with uncertain returns according to the minmax-regret criterion. *Mathematical Programming*, 100(2):345–353, 2004.

[15] X. V. Doan and K. Natarajan. On the complexity of non-overlapping multivariate marginal bounds for probabilistic combinatorial optimization. *Operations Research*, 60(1):138–149, 2012.

[16] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.

[17] M. Frittelli and E. R. Gianin. Putting order in risk measures. *Journal of Banking and Finance*, 26(7):1473–1486, 2002.

[18] F. Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

[19] F. Glover and E. Woolsey. Further reduction of zero-one polynomial programming problems to zero-one linear programming problems. *Operations Research*, 21(1):156–161, 1973.

[20] F. Glover and E. Woolsey. Converting the 0-1 polynomial programming problem to a 0-1 linear program. *Operations Research*, 22(1):180–182, 1974.

[21] W. K. Klein Haneveld. Robustness against dependence in pert: An application of duality and distributions with known marginals. *Mathematical Programming Studies*, 27:153–182, 1986.

[22] K. Isii. On the sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14:185–197, 1963.

[23] J. Kamburowski. A note on the stochastic shortest route problem. *Operations Research*, 33(3):696–698, 1985.

[24] A. Kasperski. *Discrete Optimization with Interval Data: Minmax Regret and Fuzzy Approach*. Springer Verlag, 2008.

[25] A. Kasperski and P. Zieliński. The robust shortest path problem in series-parallel multidigraphs with interval data. *Operations Research Letters*, 34(1):69–76, 2006.

[26] A. Kasperski and P. Zieliński. On the existence of an FPTAS for minmax regret combinatorial optimization problems with interval data. *Operations Research Letters*, 35(4):525–532, 2007.

[27] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications*, volume 14. Kluwer Academic Publishers, 1997.

[28] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[29] A. Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746, 1959.

[30] I. Meilijson and A. Nadas. Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, 16(3):671–677, 1979.

[31] K. Natarajan, M. Sim, and J. Uichanco. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20(4):695–731, 2010.

[32] Song M. Natarajan, K. and C. P. Teo. Persistency model and its applications in choice modeling. *Management Science*, 55(3):453–469, 2009.

[33] Y. Nesterov. *Structure of non-negative polynomials and optimization problems*, volume 200. Kluwer Academic Publishers, 1997.

[34] G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

[35] D. Reich and L. Lopes. The most likely path on series-parallel networks. *Networks*, 2010.

[36] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[37] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[38] R.T. Rockafellar. Saddle-points and convex analysis. *Differential games and related topics,H.W. Kuhn and G.P. Szego, eds.*, pages 109–128, 1971.

[39] L. J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46(253):55–67, 1951.

[40] A. Schrijver. *Theory of linear and integer programming.* Wiley, 1998.

[41] Hanif D Sherali and Amine Alameddine. A new reformulation-linearization technique for bilinear programming problems. *Journal of Global Optimization*, 2(4):379–410, 1992.

[42] A. M-C So, J. Zhang, and Y. Ye. Stochastic combinatorial optimization with controllable risk aversion. *Mathematics of Operations Research*, 34(3):522–537, 2009.

[43] G. Weiss. Stochastic bounds on distributions of optimal value functions with applications to pert, network flows and reliability. *Operations Research*, 34(4):595–605, 1986.

[44] H. Yaman, O. E. Karaşan, and M. C. Pinar. The robust spanning tree problem with interval data. *Operations Research Letters*, 29(1):31–40, 2001.

[45] J. Yue, B. Chen, and M.C. Wang. Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136, 2006.

[46] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

[47] P. Zieliński. The computational complexity of the relative robust shortest path problem with interval data. *European Journal of Operational Research*, 158(3):570–576, 2004.

[48] Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.