# A Probabilistic Model for Minmax Regret in Combinatorial Optimization

Karthik Natarajan* Dongjian Shi† Kim-Chuan Toh‡

May 11, 2012

**Abstract**

In this paper, we propose a 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 standard 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. 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 $R(\boldsymbol{x}, \boldsymbol{c}) \geq 0$. Let $\Omega$ represent a deterministic uncertainty set of all the possible realizations of the vector $\boldsymbol{c}$. The maximum value of regret for a decision $\boldsymbol{x}$ corresponding to the uncertainty set $\Omega$ is given as:

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

Savage [33] 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 [23]. The computational complexity of solving minmax regret problems have been extensively studied therein under the following two representations of $\Omega$ :

(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 [23]). 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 this paper, we restrict our attention to the interval uncertainty representation. 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. [37] and Kasperski [20]). 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 [5] 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 [21]. Their algorithm is based on a mid-point scenario approach where the deterministic combinatorial optimization problem is solved with an objective coefficient vector $(\underline{\boldsymbol{c}} + \overline{\boldsymbol{c}})/2$. It is natural to contrast this with the absolute robust counterpart where the deterministic combinatorial optimization problem is solved with the objective coefficient vector $\underline{\boldsymbol{c}}$. Kasperski and Zieliński [22] 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 [4] 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 [12] designed a faster algorithm to solve this problem with running time $O(N \min(K, N-K))$.

These regret models for combinatorial optimization are based on the original approach of Savage [33] where the uncertainty is represented in a deterministic manner. To quantify the impact of probabilistic information on regret, consider the graph in Figure 1. In this graph, there are three paths connecting node 1 to node 4: $1-4$, $2-5$ and $1-3-5$. Consider a decision-maker who wants to go from node 1 to node 4 in the shortest possible time by choosing among the three paths. The mean $\mu_i$ and range $[\underline{c}_i, \overline{c}_i]$

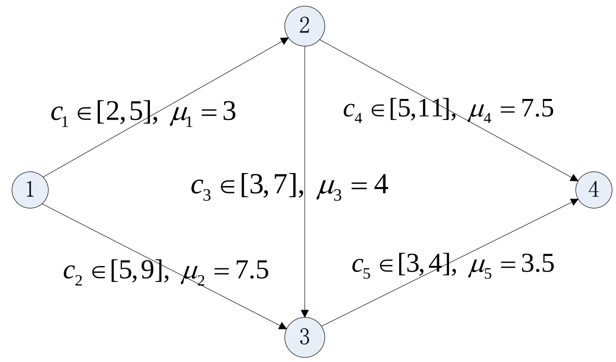


Figure 1: Find a Shortest Path from Node 1 to Node 4

for each edge $i$ in Figure 1 denotes the average time and the range of possible times in hours to traverse the edge. In the minmax regret model, the optimal decision is path $2-5$. The maximum regret from choosing this path is 6 hours and lesser than the maximum regret of 8 and 7 hours from choosing paths $1-4$ and $1-3-5$, respectively. However, on average this path takes 0.5 hours more than the other two paths. In terms of expected regret, the optimal decision is either of the paths $1-4$ or $1-3-5$. 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 probabilistic model for minmax regret in combinatorial optimization is proposed. The concept of conditional value-at-risk is reviewed and worst-case conditional value-at-risk of regret is used as a new probabilistic decision criterion. The interval uncertainty representation of data is supplemented with information on the marginal distributions. We consider two models of distributional uncertainty in the objective coefficients with: (a) known marginal distributions and (b) known marginal moments.

2. In Section 3, we develop a 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 formulation is extended to more general convex piecewise linear regret functions.

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.

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 polynomial complexity result of Averbakh [4] and Conde [12] which is based only on range information.

5. In Section 6, numerical examples for the shortest path and subset selection problems are provided.

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

Our model is inspired from recent works by Hayashi [17] and Stoye [35] on minmax regret with multiple priors in the area of economics. Hayashi [17] and Stoye [35] proposed an axiomatic representation of a decision maker who is driven by anticipated ex-post regrets. In these models, both aversion to regret and probabilities over the future states of the world are present. 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 $\Omega$ 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 $\Omega$. 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) is equivalent to solving the deterministic optimization problem where the random objective is replaced with the mean vector $\boldsymbol{\mu}$:

$$\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\mu}^T\boldsymbol{x}.$$

Formulation (2.1) however does not capture the degree of regret aversion. Furthermore, if the mean of the objective coefficients is specified, then the optimal decision in (2.1) is independent of 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 $\mathbb{E}_Q[|\tilde{r}|] < \infty$. For a given $\alpha \in (0,1)$, the value-at-risk is defined as the lower $\alpha$ quantile of the random variable $\tilde{r}$:

$$\text{VaR}_\alpha(\tilde{r}) = \inf \{v \mid Q(\tilde{r} \leq v) \geq \alpha\}. \tag{2.2}$$

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

**Definition 1** (Rockafellar and Uryasev [31, 32], Acerbi and Tasche [1])**.** *For* $\alpha \in (0,1)$*, the conditional value-at-risk* (CVaR) *at level* $\alpha$ *of a random variable* $\tilde{r}(\omega) : \Pi \to \Re$ *is the average of the highest* $1-\alpha$ *of the outcomes:*

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

*An equivalent representation for* CVaR *is:*

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

From an axiomatic perspective, CVaR is an example of a coherent risk measure (see Artzner et. al. [3], Föllmer and Schied [13] and Fritelli and Gianin [14]) and satisfies the following four properties:

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

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

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

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

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. [6] 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 [31] and inventory control [2] among other stochastic optimization problems. Combinatorial optimization problems under the CVaR measure has been studied by So et. al. [34]:

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

The negative sign in Formulation (2.5) denotes 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. [34] 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 [39], Natarajan et. al. [27])**.** *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 $\alpha$ of a random variable $\tilde{r}$ with respect to $\mathbb{Q}$ is defined as:*

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

From an axiomatic perspective, WCVaR can also be shown to be a coherent risk measure under mild assumptions on the set of distributions (see the discussions in Zhu and Fukushima [39] and Natarajan et. al. [27]). WCVaR has been used as a risk measure in distributional robust portfolio optimization [39, 27] and joint chance constrained optimization problems [11, 41]. In this paper, 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:

$$\min_{\boldsymbol{x}\in\mathcal{X}}\text{WCVaR}_\alpha(R(\boldsymbol{x},\tilde{\boldsymbol{c}})) \quad = \quad \min_{\boldsymbol{x}\in\mathcal{X}}\inf_{v\in\Re}\left(v + \frac{1}{1-\alpha}\sup_{P\in\mathbb{P}}\mathbb{E}_P\left[R(\boldsymbol{x},\tilde{\boldsymbol{c}})-v\right]^+\right). \tag{2.7}$$

The specification of the set of distributions $\mathbb{P}(\Omega)$ is provided next.

## 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 - the moment vector is in the interior of the set of feasible moments (see Isii [18]). 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. [8] and Lasserre [24] for a detailed description on this topic.

The moment representation of uncertainty in distributions has been used in the minimax regret newsvendor problem [38, 29]. A newsvendor needs to choose an order quantity of a product before the exact value of demand is known by balancing the costs of under-ordering and over-ordering. Yue et. al. [38] solved the minmax regret model analytically where only the mean and variance of demand are known. Roels and Perakis [29] generalized this model to incorporate additional moments and information on the shape of the demand. There are two major differences between these model and the model proposed in this paper. While the newsvendor optimization deals with a single demand variable, the optimization settings in this paper deals with multiple random variables. The marginal moment model forms the natural extension from single to multiple random variables. Also, the newsvendor minimizes the maximum ex-ante regret of not knowing the right distribution while in this paper, the decision-maker minimizes the maximum ex-post regret of not knowing the right objective coefficients.

The new probabilistic regret model can be related to standard minmax regret. In the marginal moment model, if only the range 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 $\delta_{\hat{\boldsymbol{c}}(\boldsymbol{x})}$ 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:

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

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 (2.7) 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 $\alpha$ 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 (2.7) 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 at least NP-hard since the minmax regret problem is NP-hard [5], . The parameter $\alpha$ allows for the flexibility to vary the degree of regret aversion.

We provide a definition of dominance in the regret aversion sense next.

**Definition 3.** *Consider two decisions $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$. Then $\boldsymbol{x}_1$ dominates $\boldsymbol{x}_2$ in regret aversion sense at level $\alpha$ if* $\mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})) \leq \mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}}))$*, and this is denoted by*

$$\boldsymbol{x}_1 \geq_{ra} \boldsymbol{x}_2.$$

**Definition 4.** *Consider two decisions $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$. Then $\boldsymbol{x}_1$ dominates $\boldsymbol{x}_2$ if $\boldsymbol{c}^T\boldsymbol{x}_1 \geq \boldsymbol{c}^T\boldsymbol{x}_2$ for all $\boldsymbol{c}\in\Omega$, and this is denoted by*

$$\boldsymbol{x}_1 \geq_s \boldsymbol{x}_2.$$

Definition 4 implies that the decision $\boldsymbol{x}_1$ is better than $\boldsymbol{x}_2$ for all realizations of the uncertainty. If one decision is better than another in all the scenarios, it is natural to expect that this decision is preferred in the regret aversion sense. The following lemma proves the validity of this.

**Lemma 1.** *For two decisions $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$, if $\boldsymbol{x}_1 \geq_s \boldsymbol{x}_2$, then $\boldsymbol{x}_1 \geq_{ra} \boldsymbol{x}_2$.*

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

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

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 $\mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_1,\tilde{\boldsymbol{c}})) \leq \mathrm{WCVaR}_\alpha(R(\boldsymbol{x}_2,\tilde{\boldsymbol{c}}))$. Hence by the definition, $\boldsymbol{x}_1 \geq_{ra} \boldsymbol{x}_2$. □

# 3 Computation of the Worst-case Conditional Value-at-risk of Regret

In this section, we compute the WCVaR of regret 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 [26] and later extended in the works of Klein Haneveld [16], Weiss [36], Birge and Maddox [10] and Bertsimas et. al. [8]. 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 [26] 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:

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

Meilijson and Nadas [26] 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:

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

With partial marginal distribution information, Klein Haneveld [16], Birge and Maddox [10] and Bertsimas et al. [8] 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 [16] 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 this section.

**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 $\Omega_i=[\underline{c}_i,\overline{c}_i]$ is given. Let $\boldsymbol{\mu}=\mathbb{E}(\tilde{\boldsymbol{c}})$ and $v\geq 0$. For $\boldsymbol{x}\in\mathcal{X}\subseteq\{0,1\}^N$, 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*

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

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

**Proof.** Define

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

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

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

Since for all $\boldsymbol{d} \in \Re^N \setminus \Omega$, we can find a $\boldsymbol{d}^* \in \Omega$, with

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

This modification will not increase the objective value of (3.4). Hence $\bar{\phi}_0(\boldsymbol{x}, v)$ can be expressed as:

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

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:

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

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

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

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

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

There exists an optimal $\boldsymbol{d}$ in the compact set $\Omega$ and optimal $t = \max\left(Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v\right)$ 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. 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 $I_1 = \{i \in [N] : \underline{c}_i < d_i < \overline{c}_i\}$ 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

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

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:

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

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

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

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

$$
\begin{aligned}
\phi_0(\boldsymbol{x}, v) &\geq \sum_{i=1}^{N} \int c_i \mathbb{I}_{[d_i,\bar{c}_i]}(c_i) f_i(c_i) dc_i + sv \\
&= \sum_{i=1}^{N} \int (c_i - d_i) \mathbb{I}_{[d_i,\bar{c}_i]}(c_i) f_i(c_i) dc_i + \sum_{i=1}^{N} d_i \int \mathbb{I}_{[d_i,\bar{c}_i]}(c_i) f_i(c_i) dc_i \\
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + \sum_{i=1}^{N} d_i \Big( \sum_{\boldsymbol{y} \in \mathcal{X}: y_i = 1} \lambda(\boldsymbol{y}) + s x_i \Big) + sv \qquad \text{(by (3.10))} \\
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + \sum_{\boldsymbol{y} \in \mathcal{X}} \lambda(\boldsymbol{y}) \boldsymbol{d}^T \boldsymbol{y} + s(\boldsymbol{d}^T \boldsymbol{x} + v) \\
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + \sum_{\boldsymbol{y} \in \mathcal{X}} \lambda(\boldsymbol{y}) \big( \max \big( Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v \big) \big) + s(\boldsymbol{d}^T \boldsymbol{x} + v) \qquad \text{(by (3.8))} \\
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + (1 - s) \big( \max \big( Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v \big) \big) + s(\boldsymbol{d}^T \boldsymbol{x} + v) \qquad \text{(by (3.7))} \\
&= \sum_{i=1}^{N} \mathbb{E}_{P_i}[\tilde{c}_i - d_i]^+ + \max \big( Z(\boldsymbol{d}), \boldsymbol{d}^T \boldsymbol{x} + v \big) \qquad \text{(by (3.9))} \\
&= \bar{\phi}_0(\boldsymbol{x}, v).
\end{aligned}
$$

□

It is useful to contrast the regret bound in Theorem 1 with the earlier bound of Meilijson and Nadas [26] 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.

**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 $\underline{c}_i \leq c_i \leq \bar{c}_i$ 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*

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

*with $0 \leq a_1 < a_2 < \ldots < a_J$. Let*

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

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

**Proof.**

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

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

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

The first inequality is due to the subadditivity of $Z(\cdot)$, and the second 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 $\phi(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) \geq \bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b})$.**

To prove the converse inequality, consider the dual problem of (3.11) and (3.12). Since the moments lie interior to the set of feasible moment vectors, strong duality holds (see Isii [18]). Hence

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

$$\begin{aligned}\bar{\phi}(\boldsymbol{x},\boldsymbol{a},\boldsymbol{b}) = \quad \min \quad & \left(\max_{j\in[J]}[a_j(Z(\boldsymbol{d}_j) - \boldsymbol{d}_j^T\boldsymbol{x}) + b_j] + \sum_{i=1}^{N}\bar{y}_{i0} + \sum_{i=1}^{N}\sum_{k=1}^{K_i}\bar{y}_{ik}m_{ik}\right) \\ \text{s.t.} \quad & \bar{p}_{i1}(c_i) := \bar{y}_{i0} + \sum_{k=1}^{K_i}\bar{y}_{ik}f_{ik}(c_i) - a_j(c_i - d_{ji})(1 - x_i) \geq 0, \quad \forall c_i \in \Omega_i, i \in [N], j \in [J], \\ & \bar{p}_{i2}(c_i) := \bar{y}_{i0} + \sum_{k=1}^{K_i}\bar{y}_{ik}f_{ik}(c_i) + a_j(c_i - d_{ji})x_i \geq 0, \quad \forall c_i \in \Omega_i, i \in [N], j \in [J], \\ & \bar{p}_{i3}(c_i) := \bar{y}_{i0} + \sum_{k=1}^{K_i}\bar{y}_{ik}f_{ik}(c_i) \geq 0, \quad \forall c_i \in \Omega_i, i \in [N], j \in [J].\end{aligned} \tag{3.14}$$

Note the third constraint $\bar{p}_{i3}(c_i) \geq 0$ of problem (3.14) is redundant. 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). Set $\bar{y}_{ik} = y_{ik}^*, k \in [K_i], i \in [N]$. Having fixed these $\bar{y}_{ik}$, solve problem (3.14), and let $\bar{y}_{i0}^*$ and $d_{ji}^*$, $i \in [N], j \in [J]$ be the optimal values for the remaining variables. For any $i \in [N]$, choose $\bar{y}_{i0}^*$ to be the minimal value such that $\bar{p}_{i3}(c_i)$ is nonnegative over $\Omega_i$. To see why this can be done, suppose there exists $\epsilon > 0$ such that we decrease $\bar{y}_{i0}^*$ by $\epsilon$ and $\bar{p}_{i3}(c_i) \geq 0, \forall c_i \in \Omega_i$. We can increase $d_{ji}^*$ by $\frac{\epsilon}{a_j}$ if $x_i = 0$ and $a_j > 0$ such that $\bar{p}_{i1}(c_i)$ remain unchanged. Similarly, decrease $d_{ji}^*$ by $\frac{\epsilon}{a_j}$ if $x_i = 1$ and $a_j > 0$ such that $\bar{p}_{i2}(c_i)$ remain unchanged. Then all the constraints of (3.14) are still satisfied. With this modification, $\max_{j\in[J]}(a_j(\max_{\boldsymbol{y}\in\mathcal{X}} \boldsymbol{d}_j^T\boldsymbol{y} - \boldsymbol{d}_j^T\boldsymbol{x}) + b_j)$ will increase by at most $\epsilon$. Hence the above modification will not increase the objective value of (3.14), thus we can decrease this $\bar{y}_{i0}^*$ till it is minimal.

Let $\bar{y}_{i0}^*$ be the minimum value such that $\bar{p}_{i3}(c_i)$ is nonnegative over $\Omega_i$. Similarly, for all $i \in [N], j \in [J]$, having fixed $\bar{y}_{i0}^*$, choose $d_{ji}^*$ to be the minimum value such that $\bar{p}_{i1}(c_i)$ is nonnegative over $\Omega_i$ if $x_i = 0$, and choose $d_{ji}^*$ to be the maximal value such that $\bar{p}_{i2}(c_i)$ is nonnegative over $\Omega_i$ if $x_i = 1$. 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 $\Omega_i$ 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 $\Omega_i$, 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 $\Omega$ is $\sum_{i=1}^{N} a_{i0}$. By 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

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

and 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 [26] and Bertsimas et al. [8] 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 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*

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

**Proof.** From the definition of WCVaR in (2.7):

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

Applying Theorems 1 and 2, we have:

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

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 conditional value-at-risk 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 $i \in [N]$, $\sup_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}([\tilde{c}_i - d_i]^+ - \tilde{c}_i x_i)$ is computable in polynomial time for a fixed $d_i$ and $x_i$.*

*Then for a given solution $\boldsymbol{x} \in \mathcal{X}$, the worst-case conditional value-at-risk of regret under the marginal model is computable in polynomial time.*

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

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

Under assumption (a), the feasibility of the constraint $t \geq Z(\boldsymbol{d})$ can be verified in polynomial time for a fixed $\boldsymbol{d}, t$. Under assumption (b), the objective function is efficiently computable for a given $\boldsymbol{d}, t, \boldsymbol{x}$. Using standard arguments on the equivalence of separation and optimization, it follows that the worst-case conditional value-at-risk of regret under the marginal model is computable in polynomial time. □

Examples include the longest path problem on a directed acyclic graph, spanning tree problems and assignment problems. In the next section, we provide conic mixed integer programs to minimize the worst-case conditional value-at-risk of regret in such problems.

# 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 term $\boldsymbol{d}^T\boldsymbol{x}$. Using standard linearization techniques from Glover [15] it is possible to reformulate the problem as a stochastic convex mixed integer program. For all $i \in [N]$, and $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^N$,

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

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

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

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. Under this assumption, the deterministic combinatorial optimization problem is solvable in polynomial time as a compact linear program:

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

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

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

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

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{d},\boldsymbol{z},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2} \quad & \left(\boldsymbol{b}^T\boldsymbol{\lambda}_1 + \boldsymbol{e}^T\boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha}\sum_{i=1}^N z_i + \frac{1}{1-\alpha}H(\boldsymbol{x},\boldsymbol{d})\right) \\
\text{s.t.} \quad & \overline{c}_i x_i \ge z_i \ge \underline{c}_i x_i, \quad i \in [N], \qquad ① \\
& d_i - \underline{c}_i(1-x_i) \ge z_i \ge d_i - \overline{c}_i(1-x_i), \quad i \in [N], \quad ② \\
& \boldsymbol{d} \in \Omega,\ \boldsymbol{A}^T\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 \ge \boldsymbol{d},\ \boldsymbol{\lambda}_2 \ge 0,\ \boldsymbol{x} \in \mathcal{X}. \qquad ③
\end{aligned} \tag{4.7}$$

The constraints in formulation (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 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 $i \in [N]$. In this case:

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

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

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{d},\boldsymbol{z},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\boldsymbol{t}} \quad & \left(\boldsymbol{b}^T\boldsymbol{\lambda}_1 + \boldsymbol{e}^T\boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha}\sum_{i=1}^N z_i + \frac{1}{1-\alpha}\left(\sum_{i=1}^N\sum_{j=1}^{J_i} t_{ij}p_{ij} - \boldsymbol{\mu}^T\boldsymbol{x}\right)\right) \\
\text{s.t.} \quad & t_{ij} \ge c_{ij} - d_i, t_{ij} \ge 0, j \in [J_i], i \in [N], \\
& ①, ② \text{ and } ③.
\end{aligned} \tag{4.8}$$

## 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 [28] and Bertsimas and Popescu [9] 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:

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

In this case, the optimal distribution to the problem $\sup\limits_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}\left[(\tilde{c}_i - d_i)^+ - \tilde{c}_i x_i\right]$ is known explicitly (see Madansky [25] and Ben-Tal and Hochman [7]):

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

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:

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

For feasibility the mean absolute deviation satisfies $\delta_i \leq \frac{2(\overline{c}_i - \mu_i)(\mu_i - \underline{c}_i)}{\overline{c}_i - \underline{c}_i}$. The optimal distribution for $\sup\limits_{P_i \in \mathbb{P}_i} \mathbb{E}_{P_i}\left[(\tilde{c}_i - d_i)^+ - \tilde{c}_i x_i\right]$ has been identified by Ben-Tal and Hochman [7]:

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

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 the worst-case CVaR of regret can be formulated in this case as the mixed integer SOCP:

$$\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{z}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \boldsymbol{y}, \boldsymbol{\tau}} \quad & \left( \boldsymbol{b}^T \boldsymbol{\lambda}_1 + \boldsymbol{e}^T \boldsymbol{\lambda}_2 + \frac{\alpha}{1-\alpha} \sum_{i=1}^N z_i + \frac{1}{1-\alpha} \sum_{i=1}^N \left( y_{i0} + \mu_i y_{i1} + (\mu_i^2 + \sigma_i^2) y_{i2} \right) - \frac{1}{1-\alpha} \boldsymbol{\mu}^T \boldsymbol{x} \right) \\
\text{s.t.} \quad & \tau_{i1} \geq 0, \quad y_{i0} + d_i + \underline{c}_i \overline{c}_i \tau_{i1} \geq 0, \quad y_{i2} + \tau_{i1} \geq 0, \quad \forall\, i \in [N], \\
& \tau_{i2} \geq 0, \quad y_{i0} + \underline{c}_i \overline{c}_i \tau_{i2} \geq 0, \quad y_{i2} + \tau_{i2} \geq 0, \quad \forall\, i \in [N], \\
& \left\| \begin{matrix} y_{i1} - 1 - (\underline{c}_i + \overline{c}_i)\tau_{i1} \\ y_{i0} + d_i + (\underline{c}_i \overline{c}_i - 1)\tau_{i1} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + d_i + (\underline{c}_i \overline{c}_i + 1)\tau_{i1} + y_{i2}, \quad \forall\, i \in [N], \\
& \left\| \begin{matrix} y_{i1} - (\underline{c}_i + \overline{c}_i)\tau_{i2} \\ y_{i0} + (\underline{c}_i \overline{c}_i - 1)\tau_{i2} - y_{i2} \end{matrix} \right\|_2 \leq y_{i0} + (\underline{c}_i \overline{c}_i + 1)\tau_{i2} + y_{i2}, \quad \forall\, i \in [N], \\
& ①, ② \text{ and } ③.
\end{aligned} \tag{4.11}$$

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

$$a_i(\lambda_0) = \operatorname*{argmin}_{d_i \in \Omega_i} F_i(d_i, 1, \lambda_0), \qquad b_i(\lambda_0) = \operatorname*{argmin}_{d_i \in \Omega_i} F(d_i, 0, \lambda_0).$$

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

$$d_i^*(x_i, \lambda_0) = a_i(\lambda_0)x_i + b_i(\lambda_0)(1 - x_i).$$

For simplicity, we will denote $a_i(\lambda_0), b_i(\lambda_0)$ and $d_i^*(x_i, \lambda_0)$ by $a_i, b_i$ 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

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

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

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

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 $\lambda_0$, 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 $\lambda_0$ can be restricted to a finite set.

**Proposition 2.** *Assume that the marginal distribution of $\tilde{c}_i$ is discrete and*

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

*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:

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

For a fixed $\boldsymbol{d}$, the minimizer $\lambda_0$ 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:

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

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 $\lambda_0$ is the $K$-th largest component of the optimal $\boldsymbol{d}$, the result holds. □

By combining Proposition 2 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 for subset selection.

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

**Output**: Optimal decision $\boldsymbol{x}$, the minimum WCVaR objective

Sort $\{c_{ij}\}_{j\in[J_i], i\in[N]}$ as a increasing sequence in the set $\Lambda$.
Delete the repeated numbers in $\Lambda$ to get a new set $\Lambda_0$.
$\boldsymbol{x} = 0$, $obj = \infty$
**for** $\lambda_0 \in \Lambda_0$ **do**
    **for** $i = 1, \ldots, N$ **do**
        $a_i = \mathrm{argmin}_{d_i\in\Omega_i} F(d_i, 1, \lambda_0)$, $b_i = \mathrm{argmin}_{d_i\in\Omega_i} F(d_i, 0, \lambda_0)$,
        $h_i = (a_i - \lambda_0)^+ - (b_i - \lambda_0)^+ + \frac{\alpha}{1-\alpha} a_i + \frac{1}{1-\alpha} \sum_{j=1}^{J_i} [(c_{ij} - a_i)^+ - (c_{ij} - b_i)^+] p_{ij} - \frac{1}{1-\alpha} \mu_i$,
    **end**
    $h_0 = \sum_{i=1}^{N} (b_i - \lambda_0)^+ + \frac{1}{1-\alpha} \sum_{i=1}^{N} \sum_{j=1}^{J_i} (c_{ij} - b_i)^+ p_{ij} + K\lambda_0$.
    $\boldsymbol{y} = \mathrm{argmin}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{h}^T \boldsymbol{x}$, $val = \boldsymbol{h}^T \boldsymbol{y} + h_0$.
    **if** $val < obj$ **then**
        $\boldsymbol{x} = \boldsymbol{y}$, $obj = val$.
    **end**
**end**

**Proposition 3.** *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 $O(N^2 J_{max}^2)$. 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 $\lambda_0$. The complexity of evaluating $a_i$ is thus $O(J_i)$. Likewise for $b_i$. 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 conditional value-at-risk 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 [4] and Conde [12]). Algorithm 1 is related to the earlier algorithms of Averbakh [4] and Conde [12] 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 $\delta_{\hat{\boldsymbol{c}}(\boldsymbol{x})}$, 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 $a_i = \underline{c}_i$, $b_i = \overline{c}_i$, $h_i = [\underline{c}_i - \lambda_0]^+ - [\overline{c}_i - \lambda_0]^+ - \underline{c}_i, i \in [N]$, 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 $\lambda_0$ is the $K$th largest value of the optimal $d_i^*(x_i) = \underline{c}_i(x_i) + \overline{c}_i(x_i)$, the feasible set $\Lambda_0$ can be further reduced to a smaller set with cardinality $2K$ (see the discussion in Conde [12]). 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 [4]). Algorithm 1 is thus a generalization of these algorithms for the minmax regret subset selection problem.

# 6 Computational Experiments

## 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 [19, 40, 5, 21, 30]. Some of the common methods 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 [19] 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. **Robust Shortest Path:** A robust shortest 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 [40] 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. [37].

To compare the minimum regret WCVaR path with these "shortest paths," consider the following example from Reich and Lopes [30].

**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 $t$.*

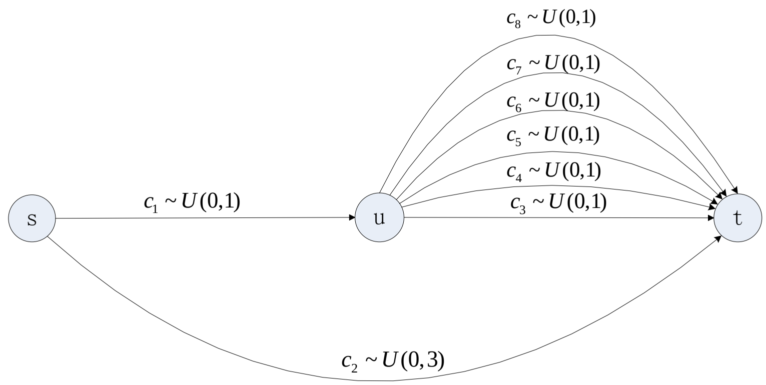


Figure 2: Network for Example 1

The choice 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. Hence in the sense of (1) Expected shortest path, (3) Robust shortest path, and (4) Minmax regret 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[30]). 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 is always one of the paths passing through the intermediate node u, which is the same as the decision of (1), (3) and (4). 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 of the lengths given.*

The network in Example 2 is the Wheatstone bridge network with the objective of finding the shortest path from node 1 to node 4. The solutions identified from the expected shortest path, robust shortest path, minmax regret path and minmax regret WCVaR methods are provided in Table 1.

While the expected approach path uses only the mean and the robust and minmax regret approaches uses only range information, the probabilistic regret model uses both the mean and range information. As the probability level $\alpha$ is varied, the minimum WCVaR regret decision changes. This is consistent

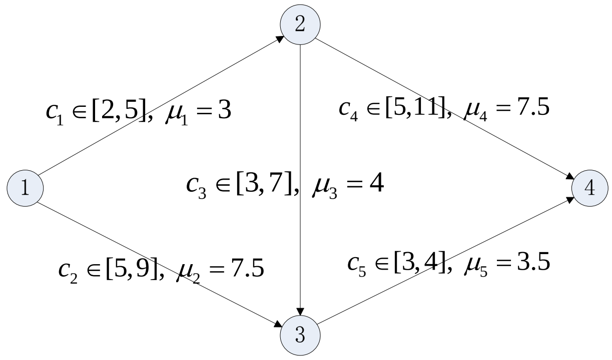


Figure 3: Network for Example 2

Table 1: The stochastic "shortest path".

| Methods | "Shortest path" | Information |
|---|---|---|
| Expected shortest path | $1-4$ or $1-3-5$ | Mean |
| Robust shortest path | $2-5$ | Range |
| Minmax regret path | $2-5$ | Range |
| Minimum WCVaR regret path | $2-5$ if $0.6667 \leq \alpha < 1$<br>$1-3-5$ if $0 < \alpha \leq 0.6667$ | Mean and range |

with the observation that $\alpha$ captures the decision-maker's aversion to regret where a larger $\alpha$ 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.

### 6.2 Subset Selection

**Example 3.** *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 $\mu_i$ are given and (b) range $[\underline{c}_i, \overline{c}_i]$, mean $\mu_i$ and mean absolute deviation $\delta_i$ are given.*

The computational studies were implemented in Matlab R2010b on an Intel Core 2 Duo CPU 2.53GHz laptop with 4 GB of RAM. To compare the efficiency of Algorithm 1 with CPLEX's MIP solver (version 12.2), randomly generated instances were tested for different $\alpha$'s and $K$'s. We compare

Table 2: 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 3: 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 |

the CPU times of the two methods in the following tables. First, we fix the value of $\alpha$ 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 $\alpha$ 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 Tables 2, 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 3). The CPU time for the algorithm is relatively insensitive to the parameters $K$ and the probability level $\alpha$ (see Figure 4), indicating that Algorithm 1 is very robust and efficient.

# 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 [26] to the

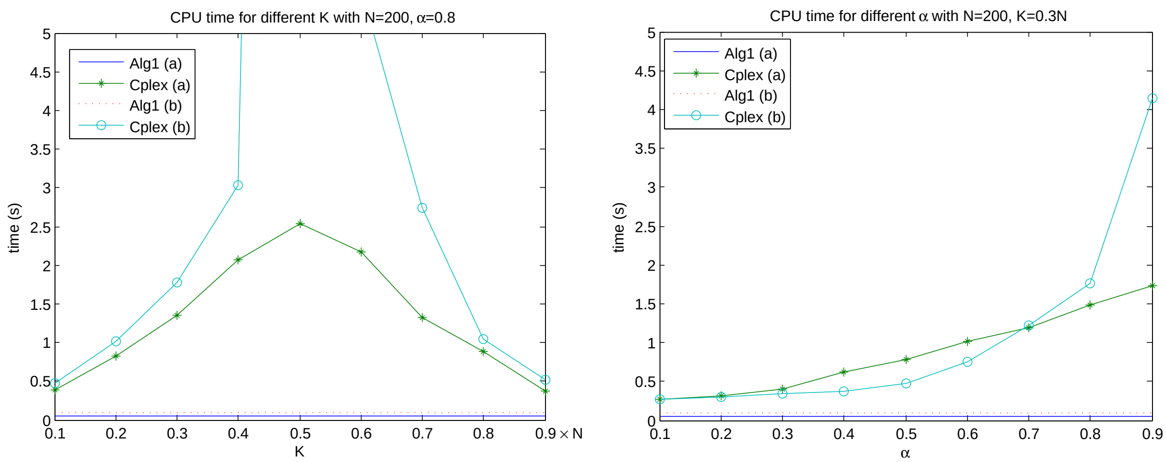


Figure 4: Sensitivity to the parameters $K$ and $\alpha$

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.

# References


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

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

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

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

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

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

[7] 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.

[8] 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, 2005.

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

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

[11] 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.

[12] 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.

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

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

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

[16] 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.

[17] T. Hayashi. Regret aversion and opportunity dependence. *Journal of Economic Theory*, 139(1):242–268, 2008.

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

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

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

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

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

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

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

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

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

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

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

[34] 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.

[35] J. G. Stoye. Axioms for minimax regret choice correspondences. *Journal of Economic Theory*, 146(6):2226–2251, 2011.

[36] 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.

[37] 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.

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

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

[40] 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.

[41] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *To appear in Mathematical Programming*, 2012.