Distributionally Robust Discrete Optimization with Entropic Value-at-Risk

Daniel Zhuoyu Long

Department of SEEM, The Chinese University of Hong Kong, zylong@se.cuhk.edu.hk

Jin Qi

NUS Business School, National University of Singapore, qijin@nus.edu.sg

Abstract

We study the discrete optimization problem under the distributionally robust framework. We optimize the Entropic Value-at-Risk, which is a coherent risk measure and is also known as Bernstein approximation for the chance constraint. We propose an efficient approximation algorithm to resolve the problem via solving a sequence of nominal problems. The computational results show that the number of nominal problems required to be solved is small under various distributional information sets.

Keywords: robust optimization; discrete optimization; coherent risk measure.

1. Introduction

Robust optimization has been widely applied in decision making as it could address data uncertainty while preserving the computational tractability. The majority of studies has been focusing on developing modeling and solving techniques for the convex optimization problems. Its application on discrete optimization problems, however, is rather limited. The difficulty mainly arises from the computational perspective. As it is well known that the discrete deterministic optimization in general is hard to solve, how to tackle the stochastic version of the problem without tremendously increasing the computational complexity is challenging. In this work, our primary goal is to provide an efficient approximation scheme to optimize a coherent risk measure in discrete optimization problems, where the distributional information is not exactly known.

For discrete optimization under robust framework, Kouvelis and Yu [1] describe the uncertainties by a finite number of scenarios and optimize the worst-case performance. They show that even the shortest path problem with only two scenarios on the cost vector, however, is already NP-hard. Averbakh [2] investigates the matroid optimization problem where each uncertainty belongs to an interval, and shows that it can be polynomially solvable. However, the conclusion cannot be generalized to other discrete optimization problems. Bertsimas and Sim [3, 4] introduce the budget of uncertainty to control the conservatism level. Under their framework, the shortest path problem under uncertainty could be solved by $O(n)$ deterministic shortest path problem, where n is the number of arcs. Nevertheless, the above approaches use little information of the uncertainties. When more distributional information is available, e.g., mean or variance, the solutions from those approaches might be unnecessarily and overly conservative. There have also been literatures using stochastic programming approach on discrete optimization (see, for instance, [5]), but the computational issue is no less severe.

In this paper, we study the discrete optimization problem with more general distributional information by optimizing the Entropic Value-at-Risk (EVaR). EVaR is first proposed by Nemirovski and Shapiro [6] as a convex approximation of the chance constraint problem. It is also a coherent risk measure that could properly evaluate the risk of the uncertain outcomes; see [7]. We then provide an efficient approximation algorithm to solve it via a sequence of nominal problems. Our computational results clearly demonstrate its benefit and suggest that the number of nominal problems we need to solve is relatively small.

The paper is structured as follows. In Section 2, we present our general framework. In Section 3, we discuss the solution procedure where we approximate EVaR using piecewise linear functions. In Section 4, we show how to find the linear segments efficiently. In Section 5, we provide several examples on the distributional information to show that the algorithms can work quite efficiently.

Notations: Vectors are represented as boldface characters, for example, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ where x_i denotes the ith element of the vector $\boldsymbol{x}$. We use a tilde to represent uncertain quantities, such as random variable $\tilde{z}$ or random vector $\tilde{\boldsymbol{c}}$.

2. Model

We consider the following nominal discrete optimization problem,

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X}, \end{array}$$

where $\mathcal{X} \subseteq \{0,1\}^n$. In practice, we may not have deterministic cost parameters $\boldsymbol{c}$. For example, in the shortest path problem, the travel time on each arc might be uncertain. To incorporate these uncertainties, we let the cost associated to each x_i be $\tilde{c}_i$, $i = 1, \ldots, n$.

Given a probability level $\epsilon \in (0,1)$, it is natural to minimize the corresponding quantile by solving the following problem,

$$\begin{array}{ll} \min & \tau \\ \text{s.t.} & \text{Prob}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x} > \tau\right) \leq \epsilon, \\ & \boldsymbol{x} \in \mathcal{X}, \end{array}$$

where Prob(A) represents probability of an event A. Essentially, this is the same problem as minimizing the Value-at-Risk (VaR) of $\tilde{\boldsymbol{c}}'\boldsymbol{x}$,

$$\min_{\boldsymbol{x} \in \mathcal{X}} \text{VaR}_{1-\epsilon}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x}\right), \tag{1}$$

where

$$\text{VaR}_{1-\epsilon}\left(\tilde{z}\right) = \inf\left\{t : \text{Prob}\left(\tilde{z} \leq t\right) \geq 1-\epsilon\right\}$$

for any random variable $\tilde{z}$. While VaR is a widely used risk measure, it is not convex and hence Problem (1) suffers from computational intractability. Indeed, Nemirovski [8] has pointed out that the feasible set of Problem (1) may not be convex even under the relaxation of discrete decision variables. Besides, VaR only accounts for the frequency but pays no attention to the magnitude, which is actually non-negligible in many practical problems. To tackle these issues, we next introduce an invariant of VaR.

Definition 1. *The Entropic Value-at-Risk (EVaR) of a random variable $\tilde{z}$, whose distribution $\mathbb{P}$ is only known to lie in a distributional ambiguity set $\mathcal{P}$, is defined as*

$$EVaR_{1-\epsilon}\left(\tilde{z}\right) = \inf_{\alpha>0}\left\{\alpha \ln \sup_{\mathbb{P}\in\mathcal{P}} \text{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{z}}{\alpha}\right)\right] - \alpha \ln \epsilon\right\}.$$

Note that the definition of EVaR incorporates the situation where we may not have full knowledge of the probability distribution. Hence, the situation of knowing the exact distribution is only a special case with $\mathcal{P}$ being a singleton. This function is first proposed by Nemirovski and Shapiro [6] as a convex and safe approximation for the chance constraint and named as Bernstein approximation. They show that $\text{EVaR}_{1-\epsilon}(\tilde{z}) \leq 0$ implies $\text{Prob}_{\mathbb{P}}(\tilde{z} > 0) \leq \epsilon$ for any $\mathbb{P} \in \mathcal{P}$, where $\text{Prob}_{\mathbb{P}}$ is the probability of the event under distribution $\mathbb{P}$. With known distribution, Ahmadi-Javid [7] studies EVaR from the perspective of risk measure, and proves that EVaR is a coherent risk measure, i.e., it has the properties of translation invariance, subadditivity, monotonicity, and positive homogeneity. With general distributional information, we can also easily check that EVaR in Definition 1 remains to be a coherent risk measure.

Instead of VaR, we minimize EVaR as follows:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \text{EVaR}_{1-\epsilon}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x}\right). \tag{2}$$

From now on, we assume that for any $i = 1, \ldots, n$, the distribution $\mathbb{P}_i$ of uncertain cost $\tilde{c}_i$ belongs to a distributional ambiguity set $\mathcal{P}_i$. Besides, $\tilde{c}_i, i = 1, \ldots, n$, are independent bounded random variables, and the distributional ambiguity set $\mathcal{P}$ of all possible distributions of $\tilde{\boldsymbol{c}}$ can be written as $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_n$. We also impose the condition on the expectation of uncertain costs such that $\sup_{\mathbb{P} \in \mathcal{P}} \text{E}_{\mathbb{P}}[\tilde{\boldsymbol{c}}] \geq \boldsymbol{0}$. In particular, when $\mathcal{P}_i = \{\mathbb{P} \,|\, \text{E}_{\mathbb{P}}(\tilde{c}_i) = 0, \mathbb{P}(\tilde{c}_i \in [-1, 1]) = 1\}$ for all $i = 1, \ldots, n$, Ben-Tal et al. [9] have shown that Bernstein approximation is a less conservative approximation to the ambiguous chance constraint compared with the budgeted approximation proposed by Bertsimas and Sim [4].

Remark 1 When the uncertain costs $\tilde{\boldsymbol{c}}$ independently follow normal distribution with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$, we could observe that

$$\begin{aligned}
\text{EVaR}_{1-\epsilon}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x}\right) &= \inf_{\alpha > 0} \left\{ \alpha \ln \text{E}_{\mathbb{P}} \left[\exp\left(\frac{\tilde{\boldsymbol{c}}'\boldsymbol{x}}{\alpha} \right) \right] - \alpha \ln \epsilon \right\} \\
&= \inf_{\alpha > 0} \left\{ \sum_{i=1}^{n} \left(\mu_i + \frac{\sigma_i^2}{2\alpha} \right) x_i - \alpha \ln \epsilon \right\} \\
&= \boldsymbol{\mu}'\boldsymbol{x} + \sqrt{2 \ln(1/\epsilon)} \sqrt{\sum_{i=1}^{n} \sigma_i^2 x_i},
\end{aligned}$$

where the optimal α is achieved at $\sqrt{\frac{\sum_{i=1}^n \sigma_i^2 x_i}{2\ln(1/\epsilon)}}$. In this case, Problem (2) is equivalent to the classical mean-standard deviation model,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \quad \boldsymbol{\mu}'\boldsymbol{x} + \sqrt{2\ln(1/\epsilon)}\sqrt{\sum_{i=1}^n \sigma_i^2 x_i}.$$

3. Solution procedure

We now focus on the solution procedure for Problem (2), and start with decomposing the uncertain costs.

Proposition 1. *Given any $\epsilon \in (0,1)$, we have*

$$Z_\epsilon^* := \min_{\boldsymbol{x}\in\mathcal{X}} \; EVaR_{1-\epsilon}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x}\right) = \inf_{\alpha>0,\boldsymbol{x}\in\mathcal{X}} \left\{ \sum_{i=1}^n C_i(\alpha)x_i - \alpha\ln\epsilon \right\}, \qquad (3)$$

where for any $i = 1,\ldots,n$,

$$C_i(\alpha) = \sup_{\mathbb{P}\in\mathcal{P}_i} \alpha \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha}\right)\right] = \alpha \ln \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha}\right)\right]$$

is a convex nonincreasing function in $\alpha > 0$, and $\lim_{\alpha\to+\infty} C_i(\alpha) = \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left(\tilde{c}_i\right)$.

With Proposition 1, we first consider a special case of Problem (3) where the nominal problem is an optimization problem over a greedoid. In this situation, the optimal solution could be exactly found by solving a polynomial number of nominal problems.

Corollary 1. *When the nominal problem is an optimization problem over a greedoid, and the uncertain cost $\tilde{c}_i, i = 1,\ldots,n$, have the structure as $\tilde{c}_i = a_i + b_i\tilde{z}_i$, where $\tilde{z}_i$ are independent and identically distributed and $b_i > 0$. Then the optimal solution of Problem (3) could be derived by solving at most $\binom{n}{2} + 1 = O(n^2)$ nominal problems.*

The above corollary, however, cannot be applied to Problem (3) in general. We next provide an approximation scheme for Problem (3) by approximating the function $C_i(\alpha)$ with a piecewise linear function.

Suppose we know the optimal α in Problem (3) is bounded by $[\alpha_{\min}, \alpha_{\max}]$, then we choose $\alpha_1 < \alpha_2 < \cdots < \alpha_{m+1}$ such that $\alpha_1 = \alpha_{\min}$, $\alpha_{m+1} = \alpha_{\max}$. For any $i = 1, \ldots, n$, we denote $C_{ij} = C_i(\alpha_j)$ and define the function

$$\hat{C}_{ij}(\alpha) = C_{ij} + \frac{C_{i(j+1)} - C_{ij}}{\alpha_{j+1} - \alpha_j} \left(\alpha - \alpha_j\right), j = 1, \ldots, m.$$

In other words, $\hat{C}_{ij}(\alpha)$ is a linear function connecting the points $(\alpha_j, C_i(\alpha_j))$ and $(\alpha_{j+1}, C_i(\alpha_{j+1}))$. In addition, we define the piecewise function

$$\hat{C}_i(\alpha) = \max_{j=1,\ldots,m} \hat{C}_{ij}(\alpha).$$

We plan to use the piecewise linear function, $\hat{C}_i(\alpha)$, to approximate the concave function $C_i(\alpha)$ and solve the following problem:

$$\hat{Z}^*_\epsilon = \min_{\alpha \in [\alpha_{\min}, \alpha_{\max}], \boldsymbol{x} \in \mathcal{X}} \left\{ \sum_{i=1}^{n} \hat{C}_i(\alpha) x_i - \alpha \ln \epsilon \right\}. \tag{4}$$

We next discuss the solution procedure of Problem (4).

Proposition 2. *$\hat{Z}^*_\epsilon$ can be equivalently calculated as*

$$\hat{Z}^*_\epsilon = \min_{j=1,\ldots,m+1} U(j),$$

where

$$U(j) = -\alpha_j \ln \epsilon + \min_{\boldsymbol{x} \in \mathcal{X}} \sum_{i=1}^{n} C_{ij} x_i.$$

Notice that $U(j)$ can be derived via solving a deterministic version of the nominal problem. Therefore, $\hat{Z}^*_\epsilon$ can be calculated with $(m+1)$ underlying deterministic problems. To see how close $\hat{Z}^*_\epsilon$ can approximate Z^*_ϵ, we need the following proposition.

Proposition 3. *Suppose there exists $\beta > 0$ such that $\hat{C}_i(\alpha) \leq (1+\beta) C_i(\alpha)$ for any $i = 1, \ldots, n$ and $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, we then have*

$$Z^*_\epsilon \leq \hat{Z}^*_\epsilon \leq (1+\beta) Z^*_\epsilon.$$

Therefore, we can use $\hat{Z}^*_\epsilon$ to arbitrarily approximate Z^*_ϵ. Clearly, the approximation level, $(1+\beta)$, would affect the number of linear segments, m, which in turn is critical to the computational complexity of the approximation problem.

4. Guidelines for Selecting Break Points

In this section, we show potential approaches to find the bound $[\alpha_{\min}, \alpha_{\max}]$ and discuss how to search for the linear segments to approximate function $C_i(\alpha)$ within $(1+\beta)$ level.

Setting $\alpha_{\min} = 0$ is obviously one potential approach to set the lower bound of optimal α. However, this will incur many linear segments at the beginning of the search. So we next provide another approach to find better $\alpha_{\min}$ and $\alpha_{\max}$. We denote $\partial C_i(\alpha)$ as the subdifferential of $C_i(\alpha)$ at α for $i = 1, \ldots, n$.

Proposition 4. *Consider any $\alpha^*_{\min} \in \mathcal{A}_{\min}$, $\alpha^*_{\max} \in \mathcal{A}_{\max}$, where*

$$\begin{aligned}
\mathcal{A}_{\min} &= \left\{ \alpha > 0 \,\middle|\, \max_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{n} s_i x_i \le \ln\epsilon,\ s_i \in \partial C_i(\alpha),\ i = 1, \ldots, n \right\}, \\
\mathcal{A}_{\max} &= \left\{ \alpha > 0 \,\middle|\, \min_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{n} s_i x_i \ge \ln\epsilon,\ s_i \in \partial C_i(\alpha),\ i = 1, \ldots, n \right\}
\end{aligned}$$

are both convex sets. Then the optimal α in Problem (3) must be in the range $[\alpha^*_{\min}, \alpha^*_{\max}]$.

Remark 2 There are a few points we need to highlight with respect to Proposition 4. First, as both $\mathcal{A}_{\min}$ and $\mathcal{A}_{\max}$ are convex, we can decrease (increase) the value of any initial α until it becomes an element of $\mathcal{A}_{\min}$ ($\mathcal{A}_{\max}$) and then we find a lower (upper) bound of the optimal α. Second, the set $\mathcal{A}_{\min}$ might possibly be empty, in which case we cannot use Proposition 4 to find the lower bound and hence we need other alternatives, such as simply setting $\alpha_{\min} = 0$. Third, the conditions in sets $\mathcal{A}_{\min}$ and $\mathcal{A}_{\max}$ involve a discrete optimization problem which might be computationally intractable. In that case, we can relax $\boldsymbol{x} \in \mathcal{X}$ to $\boldsymbol{x} \in Conv(\mathcal{X})$, where $Conv(\cdot)$ represents the convex hull. We then would have a stricter but easier condition to check, and still guarantee the bound of optimal α.

We now study how to choose α_j, $j = 1, \ldots, m$ for any approximation level $(1+\beta)$ in Proposition 3. Denote by $\lceil\cdot\rceil$ as the ceiling, i.e., $\lceil z \rceil = \min\{z^o \in \mathbb{Z} : z^o \ge z\}$.

Proposition 5. *Given any $\epsilon > 0$ and $i = 1, \ldots, n$, we let $\bar{\mu}_i = \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}[\tilde{c}_i]$, and select*

$$m = \left\lceil \max_{i=1,\ldots,n} \frac{s_i(\alpha_{\min} - \alpha_{\max})}{\bar{\mu}_i \beta} \right\rceil,$$

and $\alpha_j = \alpha_{\min} + \frac{j-1}{m}(\alpha_{\max} - \alpha_{\min})$ *for all* $j = 2, \ldots, m$*, where* $s_i \in \partial C_i(\alpha_{\min})$*. We then have* $\hat{C}_i(\alpha) \le (1+\beta) C_i(\alpha)$ *for all* $\alpha \in [\alpha_{\min}, \alpha_{\max}]$*,* $i = 1, \ldots, n$*.*

Proposition 5 provides an upper bound on the number of linear segments needed to achieve the $(1+\beta)$-approximation for an arbitrary level of β. However, the scheme just simply divides the set $[\alpha_{\min}, \alpha_{\max}]$ evenly, and may lead to an unnecessarily high value of m. Indeed, we can use the following algorithm to find the value of m and the corresponding break points $\alpha_1, \ldots, \alpha_m$ efficiently.

Algorithm for Break Points

Initialization: $k = 1$, $\alpha_1 = \alpha_{\min}$.

repeat

1. Solve

$$\begin{array}{rlll} \alpha_{k+1} := & \max & \alpha^o & \\ & \text{s.t.} & f_i(\alpha_k, \alpha^o) \le 0, & i = 1, \ldots, n \\ & & \alpha_k < \alpha^o \le \alpha_{\max}, & \end{array} \tag{5}$$

where

$$f_i(\alpha_k, \alpha^o) = \max_{\alpha \in [\alpha_k, \alpha^o]} g_i(\alpha_k, \alpha^o, \alpha),$$

and

$$g_i(\alpha_k, \alpha^o, \alpha) = C_i(\alpha_k) + \frac{C_i(\alpha^o) - C_i(\alpha_k)}{\alpha^o - \alpha_k}(\alpha - \alpha_k) - (1+\beta) C_i(\alpha).$$

2. Update $k = k + 1$.

until $\alpha_k = \alpha_{\max}$.

Output: $m = k$, $\alpha_1, \ldots, \alpha_m$.

Proposition 6. *Consider any* $\alpha_k \ge \alpha_{\min}$*,* $i = 1, \ldots, n$*.* $f_i(\alpha_k, \alpha^o)$ *is nondecreasing in* $\alpha^o \in [\alpha_k, \alpha_{\max}]$*, and* $g_i(\alpha_k, \alpha^o, \alpha)$ *is concave in* α*.*

According to Proposition 6, to calculate α_{k+1} in Problem (5), we can use a binary search on α^o. At any given value of α^o, we check the sign of $\max_{i=1,\ldots,n} f_i(\alpha_k, \alpha^o)$, where $f_i(\alpha_k, \alpha^o)$ can be obtained from a univariate concave maximization problem that can be efficiently solved. Interestingly, when $\tilde{c}_i$, $i = 1, \ldots, n$, are all normally distributed, we can analytically calculate the breakpoints based on the Algorithm for Break Points.

Corollary 2. *When the random vector* $\tilde{\mathbf{c}}$ *follows normal distribution with mean* $\boldsymbol{\mu}$ *and standard deviation* $\boldsymbol{\sigma}$*, we have for any* $\alpha > 0$*,*

$$C_i(\alpha) = \mu_i + \frac{\sigma_i^2}{2\alpha}, \qquad i = 1, \ldots, n.$$

Based on the Algorithm for Break Points, we could calculate the break points as

$$\alpha_{k+1} = \frac{1}{\left(\sqrt{\frac{1+\beta}{\alpha_k}} - \sqrt{\frac{2\beta\mu_j}{\sigma_j^2} + \frac{\beta}{\alpha_k}}\right)^2}, \qquad k = 1, \ldots, m-1.$$

where

$$j = \arg\min_{i=1,\ldots,n} \left\{\frac{\mu_i}{\sigma_i^2}\right\}.$$

The number of linear segments m *can be bounded above by*

$$\left\lceil \frac{\ln(\alpha_{\max}/\alpha_{\min})}{2\ln\left(\sqrt{1+\beta}+\sqrt{\beta}\right)} \right\rceil + 1.$$

Corollary 2 suggests that the calculation of break points under normal distribution assumption only depends on the uncertain cost $\tilde{c}_j$ which has the smallest mean-variance ratio. Hence, to guarantee the $(1+\beta)-$approximation level, the number of linear segments is not related to the number of decision variables n. It is proportionally increasing with function $1/(2\ln(\sqrt{1+\beta}+\sqrt{\beta}))$. Table 1 shows this value when β varies.

β	0.1	0.01	0.001	0.0001
$1/\left(2\ln\left(\sqrt{1+\beta}-\sqrt{\beta}\right)\right)$	1.607	5.008	15.81	50.00

Table 1: Bound for the number of linear segments under normal distribution.

However, for other distributional information sets, it is not easy to find the closed form formulation of the break points and the number of linear segments. Next, we will conduct some computational studies to show the calculation of break points.

5. Computational studies

In this section, we consider a stochastic shortest path problem to show the calculation of break points, and let the approximation level, β be 0.001 in all computational studies. We consider three different types of distributional information on the uncertainties $\tilde{\boldsymbol{c}}$, including the distributional ambiguity set, normal distribution, and uniform distribution. The evaluation of $C_i(\alpha)$ for the case of normal distribution has been discussed in Corollary 2.

Distributional ambiguity set. Nemirovski and Shapiro [6] have shown the calculation of function $C_i(\alpha)$ under various distributional ambiguity sets. Since the calculation procedure is quite similar, we would only take a specific one as an example. We assume that the random variable $\tilde{c}_i$ takes values in $[\underline{c}_i, \overline{c}_i]$, and its mean equals μ_i, i.e.,

$$\mathcal{P}_i = \left\{ \mathbb{P} \,|\, \mathrm{E}_{\mathbb{P}}\left(\tilde{c}_i\right) = \mu_i, \mathbb{P}\left(\tilde{c}_i \in [\underline{c}_i, \overline{c}_i]\right) = 1 \right\}.$$

Then we have for $i = 1, \ldots, n$,

$$C_i(\alpha) = \alpha \ln \sup_{\mathbb{P} \in \mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha}\right)\right] = \alpha \ln\left(\frac{\overline{c}_i - \mu_i}{\overline{c}_i - \underline{c}_i}\exp\left(\frac{\underline{c}_i}{\alpha}\right) + \frac{\mu_i - \underline{c}_i}{\overline{c}_i - \underline{c}_i}\exp\left(\frac{\overline{c}_i}{\alpha}\right)\right).$$

Uniform distribution. When the random variable $\tilde{c}_i$ follows uniform distribution $U(\underline{c}_i, \overline{c}_i)$, then

$$C_i(\alpha) = \alpha \ln\left(\alpha \frac{\exp\left(\overline{c}_i/\alpha\right) - \exp\left(\underline{c}_i/\alpha\right)}{\overline{c}_i - \underline{c}_i}\right), \; i = 1, \ldots, n.$$

We first investigate the same graph studied by Bertsimas and Sim [3] which has 300 nodes and 1475 arcs. They also specify the bound of $\tilde{c}_i$, which we denote by $[\underline{c}_i, \overline{c}_i]$. For the case of distributional ambiguity set, $\underline{c}_i$ and $\overline{c}_i$ represent the lower and upper bound, and we consider the mean μ_i at different values. For the case of normal distribution, we define $\mu_i = (\underline{c}_i + \overline{c}_i)/2$ and vary the standard deviation σ_i, while for uniform distribution, we just let $\tilde{c}_i$ follow $U(\underline{c}_i, \overline{c}_i)$. The numbers of break points under these three situations are summarized in Table 2. We can observe that these numbers are relatively small. It implies that we only need to solve a small collection of nominal problems to achieve $(1+\beta)-$approximation of the stochastic problems even β is as low as 0.001.

We next study how the size of graph would influence the number of break points. To this end, we randomly generate directed graphs following the procedure in [3]. The origin node is at $(0, 0)$, and the destination node is at $(1, 1)$. We randomly generate the two coordinates of the nodes, and calculate their Euclidean distance as $\underline{c}_i$. For each arc i, we generate a parameter γ

Distributional information	Distributional ambiguity set			Normal			Uniform
	μ_i			σ_i			
	$\frac{3\underline{c}_i+\bar{c}_i}{4}$	$\frac{\underline{c}_i+\bar{c}_i}{2}$	$\frac{\underline{c}_i+3\bar{c}_i}{4}$	$\frac{\bar{c}_i-\underline{c}_i}{2}$	$\frac{\bar{c}_i-\underline{c}_i}{4}$	$\frac{\bar{c}_i-\underline{c}_i}{6}$	
$\epsilon = 0.01$	13	9	6	28	21	18	19
$\epsilon = 0.05$	11	8	5	26	19	16	19

Table 2: Number of break points under different distributional information.

from the uniform distribution $U(1, 9)$, and let $\bar{c}_i = \gamma\underline{c}_i$. We then construct the distributional information similar to our first study, and let $\mu_i = \frac{\underline{c}_i+\bar{c}_i}{2}$ in the case of distributional ambiguity set and $\sigma_i = \frac{\bar{c}_i-\underline{c}_i}{4}$ in the case of normal distribution. Table 3 shows that the number of break points, and hence the computational complexity of the approximated problem, would not increase with the size of graphs. Compared with [3], which requires solving $O(n)$ nominal problems, our method is more appealing for the large networks.

ϵ	Distributional Information	(number of nodes, number of arcs)					
		$(30, 150)$	$(100, 500)$	$(200, 1000)$	$(300, 1500)$	$(500, 2500)$	$(800, 5000)$
0.01	Distributional ambiguity set	19	12	11	9	8	7
	Normal	25	25	21	26	24	28
	Uniform	27	23	20	22	22	23
0.05	Distributional ambiguity set	17	11	9	8	6	6
	Normal	23	23	19	24	22	26
	Uniform	23	21	19	21	20	22

Table 3: Number of break points under different graphs.

Reference

[1] P. Kouvelis, G. Yu, Robust Discrete Optimization and Its Applications, vol. 14, Springer, 1997.

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

[3] D. Bertsimas, M. Sim, Robust discrete optimization and network flows, Mathematical programming 98 (1-3) (2003) 49–71.

[4] D. Bertsimas, M. Sim, The price of robustness, Operations Research 52 (1) (2004) 35–53.

[5] R. Schultz, L. Stougie, M. H. Van Der Vlerk, Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis, Mathematical Programming 83 (1-3) (1998) 229–252.

[6] A. Nemirovski, A. Shapiro, Convex approximations of chance constrained programs, SIAM Journal on Optimization 17 (4) (2006) 969–996.

[7] A. Ahmadi-Javid, Entropic Value-at-Risk: A New Coherent Risk Measure, Journal of Optimization Theory and Applications 155 (3) (2012) 1105–1123.

[8] A. Nemirovski, On safe tractable approximations of chance constraints, European Journal of Operational Research 219 (3) (2012) 707–718.

[9] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust optimization, Princeton University Press, 2009.

[10] R. Kaas, M. Goovaerts, J. Dhaene, M. Denuit, Modern actuarial risk theory, vol. 328, Springer, 2001.

Appendix

Proof of Proposition 1

Observe that for any given $\boldsymbol{x} \in \mathcal{X}$, $\alpha > 0$,

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \alpha \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{\boldsymbol{c}}'\boldsymbol{x}}{\alpha}\right)\right] &= \alpha \ln\left(\prod_{i=1}^{n} \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i x_i}{\alpha}\right)\right]\right) \\
&= \sum_{i=1}^{n} \sup_{\mathbb{P}\in\mathcal{P}_i} \alpha \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i x_i}{\alpha}\right)\right] \\
&= \sum_{i=1}^{n} \sup_{\mathbb{P}\in\mathcal{P}_i} \left(\alpha \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha}\right)\right] \cdot x_i\right) \\
&= \sum_{i=1}^{n} C_i(\alpha)\, x_i,
\end{aligned}
$$

where the first two equalities follow from the independence of $\tilde{c}_i$ and $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_n$, and the third equality holds since $x_i \in \{0,1\}$. Therefore,

$$Z_\epsilon^* = \min_{\boldsymbol{x}\in\mathcal{X}} \text{ EVaR}_{1-\epsilon}\left(\tilde{\boldsymbol{c}}'\boldsymbol{x}\right) = \min_{\boldsymbol{x}\in\mathcal{X}} \inf_{\alpha>0} \left\{ \sum_{i=1}^{n} C_i\left(\alpha\right) x_i - \alpha \ln \epsilon \right\}.$$

To prove the convexity of $C_i(\alpha)$, $i = 1, \ldots, n$, we consider any $\alpha_1, \alpha_2 > 0$. Define $\alpha_\lambda = \lambda\alpha_1 + (1-\lambda)\alpha_2$ for any $\lambda \in [0,1]$. We have

$$\begin{aligned}
& \lambda\alpha_1 \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_1}\right)\right] + (1-\lambda)\alpha_2 \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_2}\right)\right] \\
= \ & \alpha_\lambda \ln\left(\mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_1}\right)\right]\right)^{\frac{\lambda\alpha_1}{\alpha_\lambda}} + \alpha_\lambda \ln\left(\mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_2}\right)\right]\right)^{\frac{(1-\lambda)\alpha_2}{\alpha_\lambda}} \\
= \ & \alpha_\lambda \ln\left(\left(\mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_1}\right)\right]\right)^{\frac{\lambda\alpha_1}{\alpha_\lambda}} \times \left(\mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_2}\right)\right]\right)^{\frac{(1-\lambda)\alpha_2}{\alpha_\lambda}}\right) \\
\geq \ & \alpha_\lambda \ln \mathrm{E}_{\mathbb{P}}\left[\left(\exp\left(\frac{\tilde{c}_i}{\alpha_1}\right)\right)^{\frac{\lambda\alpha_1}{\alpha_\lambda}} \times \left(\exp\left(\frac{\tilde{c}_i}{\alpha_2}\right)\right)^{\frac{(1-\lambda)\alpha_2}{\alpha_\lambda}}\right] \\
= \ & \alpha_\lambda \ln \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{c}_i}{\alpha_\lambda}\right)\right],
\end{aligned}$$

where the inequality follows from Hölder's inequality. As pointwise supremum preserves convexity, we know that $C_i(\alpha)$ is convex.

For the special case that $\mathcal{P}$ is a singleton, Kaas et al. [10] have shown the nonincreasing property and provided the asymptotic analysis of $C_i(\alpha)$. In the general case, we can easily extend the result to complete the proof. □

Proof of Corollary 1

For any $i = 1, \ldots, n$, we let

$$h_i(\alpha) = \alpha \ln \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{z}_i}{\alpha}\right)\right].$$

Since $\tilde{z}_i$ are independent and identically distributed, we could easily get that $h_1(\alpha) = \ldots = h_n(\alpha)$. For simplicity, we define $h(\alpha) = h_1(\alpha)$. Hence, we could write function $C_i(\alpha)$ in terms of function $h(\alpha)$ as

$$\begin{aligned} C_i(\alpha) &= \alpha \ln \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{a_i + b_i\tilde{z}_i}{\alpha}\right)\right] \\ &= a_i + \alpha \ln \sup_{\mathbb{P}\in\mathcal{P}_i} \mathrm{E}_{\mathbb{P}}\left[\exp\left(\frac{\tilde{z}_i}{\alpha/b_i}\right)\right] \\ &= a_i + b_i h\left(\alpha/b_i\right). \end{aligned}$$

We next show that when α varies from 0 to $+\infty$, the two functions $C_i(\alpha)$ and $C_j(\alpha)$ either are equal to each other or have at most one intersection point. Consider any $i \neq j, i, j \in \{1, \ldots, n\}$. If $b_i = b_j$, then it is obvious to justify our argument. If $b_i \neq b_j$, without loss of generality, we assume $b_j > b_i$, and let $l(\alpha) = C_j(\alpha) - C_i(\alpha)$ for any $\alpha > 0$, then we have for any $\Delta > 0$,

$$\begin{aligned} & l(\alpha+\Delta) - l(\alpha) \\ = \; & (C_j(\alpha+\Delta) - C_j(\alpha)) - (C_i(\alpha+\Delta) - C_i(\alpha)) \\ = \; & b_j\left(h\left(\frac{\alpha+\Delta}{b_j}\right) - h\left(\frac{\alpha}{b_j}\right)\right) - b_i\left(h\left(\frac{\alpha+\Delta}{b_i}\right) - h\left(\frac{\alpha}{b_i}\right)\right). \end{aligned}$$

When Δ decreases to 0, we could get

$$\begin{aligned} & \lim_{\Delta\downarrow 0} \frac{l(\alpha+\Delta) - l(\alpha)}{\Delta} \\ = \; & \lim_{\Delta\downarrow 0} \frac{b_j\left(h\left(\frac{\alpha+\Delta}{b_j}\right) - h\left(\frac{\alpha}{b_j}\right)\right) - b_i\left(h\left(\frac{\alpha+\Delta}{b_i}\right) - h\left(\frac{\alpha}{b_i}\right)\right)}{\Delta} \\ = \; & \lim_{\Delta\downarrow 0} \frac{h\left(\frac{\alpha}{b_j} + \frac{\Delta}{b_j}\right) - h\left(\frac{\alpha}{b_j}\right)}{(\Delta/b_j)} - \lim_{\Delta\downarrow 0} \frac{h\left(\frac{\alpha}{b_i} + \frac{\Delta}{b_i}\right) - h\left(\frac{\alpha}{b_i}\right)}{(\Delta/b_i)} \\ = \; & \lim_{\Delta\downarrow 0} \frac{h\left(\frac{\alpha}{b_j} + \Delta\right) - h\left(\frac{\alpha}{b_j}\right)}{\Delta} - \lim_{\Delta\downarrow 0} \frac{h\left(\frac{\alpha}{b_i} + \Delta\right) - h\left(\frac{\alpha}{b_i}\right)}{\Delta} \\ = \; & d_{\alpha/b_j} - d_{\alpha/b_i}, \end{aligned}$$

where d_{α/b_j} and d_{α/b_i} are the upper bound of subdifferential of h at α/b_j and α/b_i, respectively. Based on the definition of subdifferential and the convexity of function h, we have

$$\begin{aligned} h(\alpha/b_j) - h(\alpha/b_i) \geq d_{\alpha/b_i}(\alpha/b_j - \alpha/b_i), \\ h(\alpha/b_i) - h(\alpha/b_j) \geq d_{\alpha/b_j}(\alpha/b_i - \alpha/b_j). \end{aligned}$$

Combining the two inequalities together, and $\alpha/b_j < \alpha/b_i$, we get

$$d_{\alpha/b_j} \leq d_{\alpha/b_i}.$$

Hence, we have $\lim_{\Delta\downarrow 0} \frac{l(\alpha+\Delta)-l(\alpha)}{\Delta} \leq 0$, which implies that $l(\alpha)$ is a nonincreasing function in $\alpha > 0$. $C_j(\alpha)$ and $C_i(\alpha)$ either are overlap or have at most one intersection point.

Therefore, the cost functions $C_i(\alpha), i = 1, \ldots, n$ can only have at most $\begin{pmatrix} n \\ 2 \end{pmatrix} + 1 = O(n^2)$ possible ordering. Observe that the nominal problem is an optimization problem over a greedoid, the optimal solution depends only on the ordering of the cost. Hence, the optimal $\boldsymbol{x}$ in Problem (3) can be obtained from enumerating all $O(n^2)$ possible orderings of $C_i(\alpha)$. $\square$

Proof of Proposition 2

According to the convexity of $C_i(\alpha)$ and the definition of $\hat{C}_i(\alpha)$, we know $\hat{C}_i(\alpha) = \hat{C}_{ij}(\alpha)$ if $\alpha \in [\alpha_j, \alpha_{j+1}]$. Therefore,

$$\hat{Z}^*_\epsilon = \min_{\boldsymbol{x}\in\mathcal{X}} \min_{j=1,\ldots,m} \min_{\alpha\in[\alpha_j,\alpha_{j+1}]} \left\{ \sum_{i=1}^n \hat{C}_{ij}(\alpha)x_i - \alpha \ln \epsilon \right\}. \tag{6}$$

It is noted that $\hat{C}_{ij}(\alpha)$ is linear in α. Therefore, the optimal solution for $\min_{\alpha\in[\alpha_j,\alpha_{j+1}]} \left\{ \sum_{i=1}^n \hat{C}_{ij}(\alpha)x_i - \alpha \ln \epsilon \right\}$ can be achieved at the extreme points, i.e., α_j or α_{j+1}. Hence,

$$\begin{aligned}
& \min_{j=1,\ldots,m} \min_{\alpha\in[\alpha_j,\alpha_{j+1}]} \left\{ \sum_{i=1}^n \hat{C}_{ij}(\alpha)x_i - \alpha \ln \epsilon \right\} \\
= \quad & \min_{j=1,\ldots,m} \min \left\{ \sum_{i=1}^n \hat{C}_{ij}(\alpha_j)x_i - \alpha_j \ln \epsilon \ , \ \sum_{i=1}^n \hat{C}_{ij}(\alpha_{j+1})x_i - \alpha_{j+1} \ln \epsilon \right\} \\
= \quad & \min_{j=1,\ldots,m} \min \left\{ \sum_{i=1}^n C_{ij}x_i - \alpha_j \ln \epsilon \ , \ \sum_{i=1}^n C_{i(j+1)}x_i - \alpha_{j+1} \ln \epsilon \right\} \\
= \quad & \min_{j=1,\ldots,m+1} \left\{ \sum_{i=1}^n C_{ij}x_i - \alpha_j \ln \epsilon \right\}.
\end{aligned}$$

Therefore, Problem (6) can be reformulated as

$$\begin{aligned}
\hat{Z}^*_\epsilon &= \min_{\boldsymbol{x}\in\mathcal{X}} \min_{j=1,\ldots,m+1} \sum_{i=1}^n (C_{ij}x_i - \alpha_j \ln\epsilon) \\
&= \min_{j=1,\ldots,m+1} \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \sum_{i=1}^n C_{ij}x_i - \alpha_j \ln\epsilon \right\} \\
&= \min_{j=1,\ldots,m+1} U(j).
\end{aligned}$$

□

Proof of Proposition 3

Consider any $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, there exists $j \in \{1,\ldots,m\}$ such that $\alpha \in [\alpha_j, \alpha_{j+1}]$. Since $C_i(\alpha)$ is convex, we then have $C_i(\alpha) \le \hat{C}_{ij}(\alpha) = \hat{C}_i(\alpha) \le (1+\beta)C_i(\alpha)$. Therefore, for any $\boldsymbol{x} \in \mathcal{X}$, $\epsilon \in (0,1)$,

$$\sum_{i=1}^n C_i(\alpha)x_i - \alpha\ln\epsilon \le \sum_{i=1}^n \hat{C}_i(\alpha)x_i - \alpha\ln\epsilon \le \sum_{i=1}^n (1+\beta)C_i(\alpha)x_i - (1+\beta)\alpha\ln\epsilon,$$

where the inequalities hold since $\boldsymbol{x}$ is nonnegative and $\ln\epsilon < 0$. Taking the minimum on each side we get $Z^*_\epsilon \le \hat{Z}^*_\epsilon \le (1+\beta)Z^*_\epsilon$. □

Proof of Proposition 4

The convexity of the sets $\mathcal{A}_{\min}$ and $\mathcal{A}_{\max}$ follows from the convexity of the functions $C_i(\alpha)$, which is supported by Proposition 1.

For any $\boldsymbol{x} \in \mathcal{X}$ and $0 < \alpha \le \alpha^*_{\min} \in \mathcal{A}_{\min}$, we can find $s_i \in \partial C_i(\alpha^*_{\min})$, $i = 1,\ldots,n$ such that $\max_{\boldsymbol{y}\in\mathcal{X}} \sum_{i=1}^n s_i y_i \le \ln\epsilon$. Hence we have

$$\begin{aligned}
& \sum_{i=1}^n C_i(\alpha)x_i - \alpha\ln\epsilon \\
\ge\ & \sum_{i=1}^n \left(C_i(\alpha^*_{\min}) + s_i(\alpha - \alpha^*_{\min})\right)x_i - \alpha^*_{\min}\ln\epsilon + (\alpha^*_{\min} - \alpha)\ln\epsilon \\
\ge\ & \sum_{i=1}^n C_i(\alpha^*_{\min})x_i - \alpha^*_{\min}\ln\epsilon - (\alpha^*_{\min} - \alpha)\cdot\max_{\boldsymbol{y}\in\mathcal{X}}\sum_{i=1}^n s_i y_i + (\alpha^*_{\min} - \alpha)\ln\epsilon \\
\ge\ & \sum_{i=1}^n C_i(\alpha^*_{\min})x_i - \alpha^*_{\min}\ln\epsilon,
\end{aligned}$$

where the first inequality is due to the convexity of $C_i(\alpha)$, the second and third inequalities hold since $\sum_{i=1}^n s_i x_i \le \max_{\boldsymbol{y}\in\mathcal{X}} \sum_{i=1}^n s_i y_i \le \ln\epsilon$. Therefore, we can find optimal α to be no less than $\alpha^*_{\min}$.

To show the other way, consider any $\alpha \ge \alpha^*_{\max} \in \mathcal{A}_{\max}$, we then can find $s_i \in \partial C_i\left(\alpha^*_{\max}\right),\ i = 1,\ldots,n$ such that $\min_{\boldsymbol{y}\in\mathcal{X}} \sum_{i=1}^n s_i y_i \ge \ln\epsilon$. Similar to the previous analysis, we have

$$
\begin{aligned}
& \sum_{i=1}^{n} C_i(\alpha)x_i - \alpha\ln\epsilon \\
\ge\ & \sum_{i=1}^{n} \left(C_i(\alpha^*_{\max}) + s_i\left(\alpha-\alpha^*_{\max}\right)\right)x_i - \alpha^*_{\max}\ln\epsilon - \left(\alpha-\alpha^*_{\max}\right)\ln\epsilon \\
\ge\ & \sum_{i=1}^{n} C_i(\alpha^*_{\max})x_i - \alpha^*_{\max}\ln\epsilon + \left(\alpha-\alpha^*_{\max}\right)\cdot\min_{\boldsymbol{y}\in\mathcal{Y}}\sum_{i=1}^{n} s_i y_i - \left(\alpha-\alpha^*_{\max}\right)\ln\epsilon \\
\ge\ & \sum_{i=1}^{n} C_i(\alpha^*_{\max})x_i - \alpha^*_{\max}\ln\epsilon.
\end{aligned}
$$

Therefore, we can find optimal α to be no greater than $\alpha^*_{\max}$. □

Proof of Proposition 5

Consider any $\alpha \in [\alpha_j, \alpha_{j+1}]$, $j = 1,\ldots,m$. We can observe that

$$
\begin{aligned}
\hat{C}_i(\alpha) &\le C_{ij} \\
&\le C_{i(j+1)} - s_i\left(\alpha_{j+1}-\alpha_j\right) \\
&\le C_i(\alpha) + \frac{C_i(\alpha)}{\bar{\mu}_i}\left(-s_i\right)\left(\alpha_{j+1}-\alpha_j\right) \\
&= C_i(\alpha)\left(1 + \frac{s_i}{\bar{\mu}_i}\frac{\alpha_{\min}-\alpha_{\max}}{m}\right) \\
&\le C_i(\alpha)\left(1+\beta\right),
\end{aligned}
$$

where the first inequality follows from the monotonicity of $\hat{C}_i(\alpha)$, the second inequality follows from the convexity of $C_i(\alpha)$, and the third inequality follows from the monotonicity of $C_i(\alpha)$ and $C_i(\alpha) \ge \bar{\mu}_i > 0$ for all α. □

Proof of Proposition 6

To prove that $f_i(\alpha_k, \alpha^o)$ is nondecreasing in $\alpha^o \in [\alpha_k, \alpha_{\max}]$, we consider any $\alpha^o, \bar{\alpha}^o$ such that $\alpha_k < \alpha^o \le \bar{\alpha}^o \le \alpha_{\max}$. In that case, $\alpha^o = \lambda\alpha_k + (1-\lambda)\bar{\alpha}^o$

for some $\lambda \in [0,1)$. We then have

$$\frac{C_i(\alpha_k) - C_i(\alpha^o)}{\alpha^o - \alpha_k} \geq \frac{C_i(\alpha_k) - \lambda C_i(\alpha_k) - (1-\lambda) C_i(\bar{\alpha}^o)}{\lambda \alpha_k + (1-\lambda)\bar{\alpha}^o - \alpha_k} = \frac{C_i(\alpha_k) - C_i(\bar{\alpha}^o)}{\bar{\alpha}^o - \alpha_k}. \tag{7}$$

Therefore,

$$\begin{aligned} f_i(\alpha_k, \bar{\alpha}^o) &\geq \max_{\alpha \in [\alpha_k, \alpha^o]} g_i(\alpha_k, \bar{\alpha}^o, \alpha) \\ &= \max_{\alpha \in [\alpha_k, \alpha^o]} \left\{ C_i(\alpha_k) - \frac{\alpha - \alpha_k}{\bar{\alpha}^o - \alpha_k} \left(C_i(\alpha_k) - C_i(\bar{\alpha}^o) \right) - (1+\beta) C_i(\alpha) \right\} \\ &\geq \max_{\alpha \in [\alpha_k, \alpha^o]} \left\{ C_i(\alpha_k) - \frac{\alpha - \alpha_k}{\alpha^o - \alpha_k} \left(C_i(\alpha_k) - C_i(\alpha^o) \right) - (1+\beta) C_i(\alpha) \right\} \\ &= f_i(\alpha_k, \alpha^o), \end{aligned}$$

where the first inequality holds since $\alpha^o \leq \bar{\alpha}^o$, and the second inequality is due to the result in the inequality (7). The concavity of $g_i(\alpha_k, \alpha^o, \alpha)$ in α follows immediately from the convexity of $C_i(\alpha)$. □

Proof of Corollary 2

Since $\tilde{c}_i$ follows normal distribution, from the moment generating function, we have for any $\alpha > 0$,

$$C_i(\alpha) = \mu_i + \frac{\sigma_i^2}{2\alpha}, \qquad i = 1, \ldots, n.$$

Hence, in Algorithm for Break Points, we could get

$$\begin{aligned}
& f_i(\alpha_k, \alpha^o) \\
= & \max_{\alpha\in[\alpha_k,\alpha^o]} g_i(\alpha_k, \alpha^o, \alpha) \\
= & \max_{\alpha\in[\alpha_k,\alpha^o]} \left\{ C_i(\alpha_k) + \frac{C_i(\alpha^o) - C_i(\alpha_k)}{\alpha^o - \alpha_k}(\alpha - \alpha_k) - (1+\beta)C_i(\alpha) \right\} \\
= & C_i(\alpha_k) - \frac{C_i(\alpha^o) - C_i(\alpha_k)}{\alpha^o - \alpha_k}\alpha_k + \max_{\alpha\in[\alpha_k,\alpha^o]} \left\{ \frac{C_i(\alpha^o) - C_i(\alpha_k)}{\alpha^o - \alpha_k}\alpha - (1+\beta)C_i(\alpha) \right\} \\
= & \mu_i + \frac{\sigma_i^2}{2\alpha_k} + \frac{\sigma_i^2}{2\alpha^o} + \max_{\alpha\in[\alpha_k,\alpha^o]} \left\{ -\frac{\sigma_i^2}{2\alpha_k\alpha^o}\alpha - (1+\beta)\left(\mu_i + \frac{\sigma_i^2}{2\alpha}\right) \right\} \\
= & \mu_i + \frac{\sigma_i^2}{2\alpha_k} + \frac{\sigma_i^2}{2\alpha^o} - \mu_i(1+\beta) + \begin{cases} -\sqrt{\frac{1+\beta}{\alpha_k\alpha^o}}\sigma_i^2, & \text{if } \alpha^o \geq (1+\beta)\alpha_k, \\ -\frac{\sigma_i^2}{2\alpha_k} - \frac{(1+\beta)\sigma_i^2}{2\alpha^o}, & \text{if } \alpha_k \leq \alpha^o \leq (1+\beta)\alpha_k \end{cases} \\
= & \begin{cases} \frac{\sigma_i^2}{2}\left(\frac{1}{\sqrt{\alpha^o}} - \sqrt{\frac{1+\beta}{\alpha_k}}\right)^2 - \beta\mu_i - \frac{\beta\sigma_i^2}{2\alpha_k}, & \text{if } \alpha^o \geq (1+\beta)\alpha_k, \\ -\frac{\beta\sigma_i^2}{2\alpha^o} - \beta\mu_i, & \text{if } \alpha_k \leq \alpha^o \leq (1+\beta)\alpha_k \end{cases}
\end{aligned}$$

Given the closed form of function $f_i(\alpha_k, \alpha^o)$, we could solve Problem (5), and get

$$\alpha_{k+1} = \frac{1}{\left(\sqrt{\frac{1+\beta}{\alpha_k}} - \sqrt{\frac{2\beta\mu_j}{\sigma_j^2} + \frac{\beta}{\alpha_k}}\right)^2},$$

where

$$j = \arg\min_{i=1,\ldots,n} \left\{ \frac{\mu_i}{\sigma_i^2} \right\}.$$

It is easy to show that

$$\alpha_{k+1} = \frac{1}{\left(\sqrt{\frac{1+\beta}{\alpha_k}} - \sqrt{\frac{2\beta\mu_j}{\sigma_j^2} + \frac{\beta}{\alpha_k}}\right)^2} \geq \frac{1}{(\sqrt{1+\beta} - \sqrt{\beta})^2}\alpha_k = (\sqrt{1+\beta}+\sqrt{\beta})^2\alpha_k,$$

where the inequality holds because we have $\mu_j + \frac{\sigma_j^2}{2\alpha_k} \geq (1+\beta)\mu_j$ and $\frac{\mu_j}{\sigma_j^2} > 0$.

□