

Combinatorial optimization problems with normal random costs

Paolo Serafini

University of Udine, Department of Mathematics and Computer Science

Abstract. We consider combinatorial optimization problems with random jointly normal costs. Target-probability efficient solutions are those that grant a given target at a certain probability and no other solution can grant the same target at a higher probability. Since the costs belong to a location-scale family the efficient solutions have a simple geometric characterization in terms of mean and variance. Computing efficient solutions can be carried out by an iteration scheme calling for the solution of a constrained integer programming problem.

1 Introduction

We consider the following general definition of a combinatorial optimization problem: a finite set E and a family \mathcal{X} of subsets of E are given. Subsets $x \in \mathcal{X}$ are called feasible solutions. For each $e \in E$ a cost c_e is given and the cost of a subset $x \subset E$ is defined as $c(x) = \sum_{e \in x} c_e$. The goal is to find a solution $x^* \in \mathcal{X}$ minimizing $c(x)$ over all feasible solutions. A very large number of combinatorial optimization problems can be framed according to this definition.

In this paper we consider the extension of the previous definition to the case of costs C_e represented by random variables. Then, for each x , the quantity $C(x) = \sum_{e \in x} C_e$ is also a random variable with distribution function $F_x(c) := \Pr \{C(x) \leq c\}$.

A classical and simple approach to solve a stochastic optimization problem turns it into a deterministic one by finding solutions of minimum expected cost. However, relying on minimum expected cost may not be a satisfactory approach especially if there are large fluctuations of the random variables.

The problem is inherently bi-objective because we want to minimize at the same time the cost and the risk that the cost will be worse than expected. A standard approach, applied to financial investment problems, was proposed long ago by Markowitz [4] and earlier by de Finetti [2]. The risk is modeled on the variance of the random variable $C(x)$ and we are faced with the problem of finding mean-variance efficient solutions.

More recently the concept of Value at Risk (VaR) has been defined to approach financial investment problems [5, 3]. Two quantities are taken into consideration, a cost threshold and a probability, and we want to find solutions within the threshold at the given probability. A similar approach has been applied also to combinatorial optimization problems, like in [1, 6].

In this paper we apply this concept to the stochastic combinatorial optimization problems as defined above. Therefore, given a probability value p and a target c , we want to find solutions x such that $F_x(c) = \Pr \{C(x) \leq c\} \geq p$. In particular we are interested in computing efficient solutions according to the following definitions.

Definition 1. A target-probability pair (c^*, p^*) , with $p^* < 1$, is target-probability efficient (tp-efficient or simply efficient) if there exist $x^* \in \mathcal{X}$ such that $F_{x^*}(c^*) = p^*$ and for no $x \in \mathcal{X}$, $c \in R$, $p \in R$

$$F_x(c) = p \geq p^* = F_{x^*}(c^*), \quad c \leq c^*$$

with at least one inequality strict. The solution $x^* \in \mathcal{X}$ is also called target-probability efficient. □

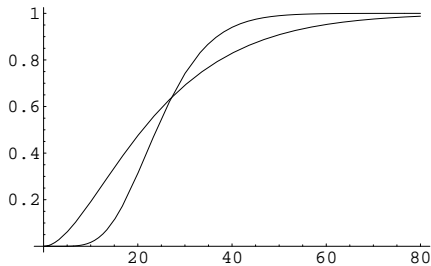


Fig. 1. Distribution functions for the travel times of the two paths

In the definition it is assumed that $p^* < 1$. This is necessary in case there is a common upper bound for the ranges of all C_e . In this case, if we take the target c^* equal to the common upper bound, so that $F_x(c^*) = 1$ for all $x \in \mathcal{X}$, then all $x \in \mathcal{X}$ would be trivially efficient.

Note that, if the distribution functions are strictly monotonic, efficiency can be equivalently defined by saying that x^* is efficient if there exists a target c^* such that x^* maximizes $F_x(c^*)$ w.r.t. $x \in \mathcal{X}$. Hence in this case, we should compute the pointwise maximum of the family of distribution functions $F_x(c)$, $x \in \mathcal{X}$, i.e. the following function of c

$$\Psi(c) := \max_{x \in \mathcal{X}} F_x(c)$$

and the tp-efficient set we want to compute is the set $\{(c; \Psi(c) : c \in R)\}$.

We illustrate the problem with the following example: suppose there are two paths connecting two towns. The travel times of the paths are random variables with Erlang density function $k^k \lambda^k t^{k-1} e^{-k\lambda t} / (k-1)!$. The first path has the values $k = 2$ and $\lambda = 1/24$ and the second path has $k = 8$ and $\lambda = 1/24$. We can think that the first path consists of two arcs with travel times exponentially distributed and mean values 12 for each arc and the second path consists of eight arcs with travel times exponentially distributed and mean values 3 for each arc.

The two distribution functions are shown in Figure 1. They intersect at about the value $t = 26$. The steeper function corresponds to the second path. So, if we are interested in reaching the town within the threshold 30, we see that the second path can guarantee this goal with probability 0.779, whereas the first path can guarantee it with slightly lower probability 0.712. Therefore we prefer to choose the second path. On the contrary if we fix a threshold of 20 (say we have to reach the town within this time otherwise the trip is useless) the first path gives a better probability (0.496) than the second path (0.351). The target-probability efficient set we are looking for is the pointwise maximum of the two distribution functions.

The problem of computing efficient solutions is very difficult in general. It is a tractable problem if the random variables $C(x)$ belong to a location-scale family. However, since the cost of a feasible solution is computed as the sum of the costs of its elements, the randomness of $C(x)$ is induced by the randomness of C_e and, unless the costs C_e are normal, we do not expect the $C(x)$ to be a location-scale family.

Hence we assume that the random variables C_e , $e \in E$, are jointly normal. This in turn implies that also $C(x)$ are normal random variables belonging to a location-scale family. We consider both the case of independent and dependent (jointly distributed) random variables C_e . We note that the assumption of normal variables is quite realistic since, whenever the feasible sets x have many elements then the random variables $C(x)$ are very close to normal.

Under these hypotheses the computation of efficient solutions can be carried out algorithmically in a straightforward fashion. As we shall show, we have to solve iteratively an integer programming problem

closely related to the combinatorial problem we are dealing with. Hence if the underlying combinatorial problem is difficult, the whole procedure is also difficult, but this is an unavoidable feature.

The fact that we model the costs by normal variables may have surprising effects. Even if the problem presents costs which are inherently non negative (like, e.g., travel times of edges in a graph) normal random costs may be negative, although with negligible probability. This fact must be taken into account in interpreting the solutions.

The paper is organized as follows. After having defined target-probability efficiency in this section, in Section 2 we characterize efficient solutions by exploiting the concept of location-scale family. In Section 3 we show an iterative scheme that produces a set of solutions which includes all tp-efficient solutions, and in Section 4 we show how we can extract from this set the efficient solutions. In Section 5 we describe a similar iteration to find only one particular efficient solution for a given target. In Section 6 we hint on how to use mathematical programming techniques to solve the series of problems of Section 3 and Section 5. In Section 7 we address the problem of the number of iterations needed to find the efficient solutions. An example is provided in Section 8. Finally some conclusions follow in Section 9.

2 Efficient solutions for jointly normal costs

Let the random variables C_e be jointly normally distributed with mean μ_e , $e \in E$, and covariance matrix $\sigma_{e,e'}^2$, $e, e' \in E$ (denote $\sigma_e^2 = \sigma_{e,e}^2$). We assume that the means μ_e are integer numbers. This is equivalent to assume rational numbers and multiply them by the l.c.m. of the denominators. Clearly the efficient solutions are not affected by a change of scale.

Then the random variables $C(x) = \sum_{e \in x} C_e$ are normally distributed with mean $\mu(x) = \sum_{e \in x} \mu_e$ and variance $\sigma^2(x) = \sum_{e, e' \in x} \sigma_{e, e'}^2$.

Let

$$\hat{\mu} = \min_{x \in \mathcal{X}} \mu(x)$$

i.e., $\hat{\mu}$ is the minimum expected cost. Given the strict monotonicity of the distribution functions we compute efficient solutions by solving the optimization problem:

$$\max_{x \in \mathcal{X}} F_x(c) \tag{1}$$

maximizing the probability that $C(x) \leq c$ for a fixed target c . We have

$$\Pr \{C(x) \leq c\} = \Pr \{\mu(x) + Y \sigma(x) \leq c\} = \Pr \left\{ Y \leq \frac{c - \mu(x)}{\sigma(x)} \right\}$$

where Y is the standard normal deviate with density function

$$\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

Hence all solutions which share the same value for the expression $y := (c - \mu(x))/\sigma(x)$ have the same probability that their cost is less than c . So if we consider the pairs $(\mu(x), \sigma(x))$ as coordinates of points in the plane, we have that solutions whose $(\mu(x), \sigma(x))$ values lie on the straight line

$$c = \mu + y \sigma$$

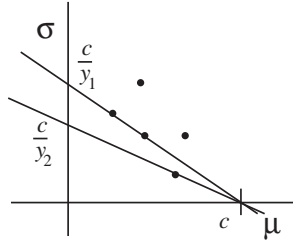


Fig. 2. Two $c = \mu + y\sigma$ straight lines with same c and different y

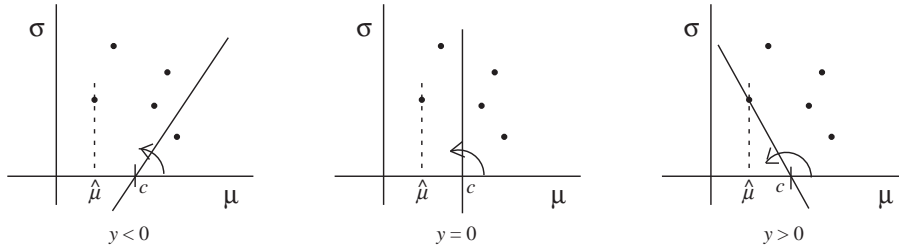


Fig. 3. Rotating the straight line by keeping fixed c and increasing y

have all the same probability $0.5 + 0.5 \operatorname{erf}(y/\sqrt{2})$ that their cost is less than c .

In Figure 2 the five points correspond to the points $(\mu(x), \sigma(x))$ for five different solutions. The two lines correspond to the equations

$$c = \mu + y_1 \sigma, \quad c = \mu + y_2 \sigma, \quad y_2 > y_1$$

The solution whose point lies on the line defined by y_2 has higher probability of having value less than c (and is therefore better than the other two solutions in spite of the fact that its mean is the highest). Hence problem (1) is equivalent to the following problem

$$\max_{x \in \mathcal{X}} \frac{c - \mu(x)}{\sigma(x)} \quad (2)$$

Geometrically, we have to rotate counterclockwise a straight line with rotation center in the point $(c, 0)$, starting from $y = -\infty$, and finding the point $(\mu(x), \sigma(x))$ which is swept last by the line, as shown in Figure 3.

If $c > \hat{\mu}$, then the best y is positive, i.e., we can guarantee the target c with probability larger than one half. If $c < \hat{\mu}$, the best y is negative, i.e., we can guarantee the target c with probability smaller than one half. If $c = \hat{\mu}$, the best y is null and we can guarantee the target c with probability exactly equal to one half. We may think that choosing a target larger than $\hat{\mu}$ corresponds to a risk-averse attitude while a target smaller than $\hat{\mu}$ corresponds to a risk-prone attitude.

Note that a target-probability pair (c^*, p^*) is in a one-to-one correspondence with the pair (c^*, y^*) , that we may also call tp-efficient, where $p^* = 0.5 + 0.5 \operatorname{erf}(y^*/\sqrt{2})$, so that we may also say that $x^* \in \mathcal{X}$ is tp-efficient if there exists y^* such that for no $x \in \mathcal{X}$, $c \in R$,

$$y = \frac{c - \mu(x)}{\sigma(x)} \geq y^* = \frac{c^* - \mu(x^*)}{\sigma(x^*)}, \quad c \leq c^*$$

with at least one inequality strict.

We want to show that solving (2) by varying the parameter c yields all and only tp-efficient solutions.

Theorem 1. *Each tp-efficient solution is optimal in (2) for some c and each optimum in (2) is tp-efficient for any c .*

Proof: Let x^* be an optimal solution in (2) (with mean μ^* and standard deviation σ^*) with chosen target c^* and optimal value $y^* = (c^* - \mu^*)/\sigma^*$. Suppose there exists a solution x and values c and y such that

$$c = \mu(x) + y\sigma(x), \quad c \leq c^*, \quad y \geq y^*, \quad (c^*, y^*) \neq (c, y)$$

Then

$$y^* = \frac{c^* - \mu^*}{\sigma^*} \geq \frac{c^* - \mu(x)}{\sigma(x)} \geq \frac{c - \mu(x)}{\sigma(x)} = y \geq y^*$$

Moreover, since at least one of the two inequalities $c \leq c^*$ and $y \geq y^*$ must be strict, it follows that at least one of the last two inequalities in the chain must be strict, which implies $y^* > y^*$. From the contradiction we have that x^* is tp-efficient.

Conversely, given a tp-efficient (c^*, y^*) associated to a solution x^* with mean μ^* and variance σ^* , consider (2) with $c := c^*$. Assuming there exists a better solution x would imply $y > y^*$ contradicting tp-efficiency of x^* . \square

An obvious property, given the geometrical picture, is provided by the following Lemma which will be useful in the next sections.

Lemma 1. *Given three solutions x_1, x_2 and x_3 , such that $\sigma(x_1) \geq \sigma(x_3) \geq \sigma(x_2)$, if*

$$\mu(x_3)(\sigma(x_1) - \sigma(x_2)) + \sigma(x_3)(\mu(x_2) - \mu(x_1)) > \mu(x_2)\sigma(x_1) - \mu(x_1)\sigma(x_2) \quad (3)$$

then x_3 is not tp-efficient.

Proof: Note that the inequality (3) cannot be verified if $\sigma(x_1) = \sigma(x_3) = \sigma(x_2)$. Hence we may assume $\sigma(x_1) > \sigma(x_2)$. The points $(\mu(x_1), \sigma(x_1))$ and $(\mu(x_2), \sigma(x_2))$ lie on the straight line

$$\mu \cdot (\sigma(x_1) - \sigma(x_2)) + \sigma \cdot (\mu(x_2) - \mu(x_1)) = \mu(x_2)\sigma(x_1) - \mu(x_1)\sigma(x_2)$$

Let us consider the target

$$\bar{c} = \frac{\mu(x_2)\sigma(x_1) - \mu(x_1)\sigma(x_2)}{\sigma(x_1) - \sigma(x_2)}$$

given by the intersection of the straight line with the μ -axis. The inequality (3) can be equivalently written as

$$\frac{\bar{c} - \mu(x_2)}{\sigma(x_2)} = \frac{\bar{c} - \mu(x_1)}{\sigma(x_1)} > \frac{\bar{c} - \mu(x_3)}{\sigma(x_3)}$$

The inequality between the first and the third term is still valid if we replace \bar{c} with any $c \geq \bar{c}$ since $\sigma(x_3) \geq \sigma(x_2)$ and the inequality between the second and the third term is still valid for any $c \leq \bar{c}$ since $\sigma(x_1) \geq \sigma(x_3)$. Hence, for any target c , x_3 is dominated by either x_1 or x_2 and cannot be tp-efficient. \square

3 An iterative scheme to compute all efficient solutions

In principle we could think of solving (2) for all possible values of c . However, this would entail the considerable difficulty of solving a nonlinear integer programming problem. Furthermore, it would not be clear how to sample the target values c in order not to miss any efficient solution. Hence, instead of tackling (2) directly, we solve the following two series of associated constrained problems which yield a set of candidate solutions containing all tp-optima. Then, in the next section, we see how to extract from these solutions the tp-optima. Let

$$\mu_0 := \mu^0 := \max_{x \in \mathcal{X}} \mu(x)$$

The first series, for various values μ_i ($i = 0, 1, \dots$) to be defined below, is given by

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \sigma(x) \\ & \mu(x) \leq \mu_i \end{aligned} \tag{4}$$

and the second series, for various values μ^j ($j = 0, 1, \dots$) to be defined below, is given by

$$\begin{aligned} \max_{x \in \mathcal{X}} \quad & \sigma(x) \\ & \mu(x) \leq \mu^j \end{aligned} \tag{5}$$

Let us first consider the iteration (4). During the iteration we reference to a solution \hat{x} such that $\mu(\hat{x}) = \hat{\mu}$ (i.e., a solution of minimum mean) and the variance $\sigma(\hat{x})$ is minimum. Denote $\hat{\sigma} := \sigma(\hat{x})$. Let x_i denote an optimum in (4). Initially we set $\bar{\mu} := \mu_0$ and $\bar{\sigma} := \sigma(x_0)$. For all i we will have $\bar{\sigma} \leq \sigma(x_i) \leq \hat{\sigma}$. Initially this is true for $i = 0$. Then the values μ_i , $i > 0$, are defined as follows. If

$$\mu(x_i) (\hat{\sigma} - \bar{\sigma}) + \sigma(x_i) (\bar{\mu} - \hat{\mu}) > \hat{\sigma} \bar{\mu} - \bar{\sigma} \hat{\mu}$$

then we disregard x_i as a candidate for tp-optimality due to Lemma 1, leave $\bar{\mu}$ and $\bar{\sigma}$ unchanged and define

$$\mu_{i+1} := \frac{\hat{\sigma} \bar{\mu} - \bar{\sigma} \hat{\mu} - \sigma(x_i) (\bar{\mu} - \hat{\mu})}{\hat{\sigma} - \bar{\sigma}} \tag{6}$$

Note that $\hat{\mu} < \mu_{i+1} < \mu(x_i) \leq \mu_i$. If, on the contrary (this case always occurs with equality for $i = 0$),

$$\mu(x_i) (\hat{\sigma} - \bar{\sigma}) + \sigma(x_i) (\bar{\mu} - \hat{\mu}) \leq \hat{\sigma} \bar{\mu} - \bar{\sigma} \hat{\mu}$$

we keep x_i as a candidate for tp-optimality, and update

$$\bar{\mu} := \mu(x_i), \quad \bar{\sigma} := \sigma(x_i), \quad \mu_i := \mu(x_{i-1}) - \varepsilon$$

where ε is sufficiently small, in the sense that there is no feasible x such that $\mu(x_i) - \varepsilon < \mu(x) < \mu(x_i)$. In both cases x_i is not feasible in the next problem $i + 1$, whereas, if $\mu_{i+1} \geq \hat{\mu}$, \hat{x} is feasible and therefore the condition $\bar{\sigma} \leq \sigma(x_i) \leq \hat{\sigma}$ is valid also in the next iteration.

We briefly comment on the choice of ε . In case the μ_e values are integer, as we have assumed, then $\varepsilon = 1$ suffices. If the means are just rational numbers, we have already observed that we may scale the random variables by multiplying them by the l.c.m. of the denominators of the means, thereby obtaining integer values. However, instead of rescaling the random variables we may keep the rational means and take ε to be equal to the inverse of the l.c.m. The size of ε does not affect the computational complexity of the

problem. The number of iterations is related to the number of different mean-variance efficient solutions of the problem.

For the iteration (5) let \hat{x} such that $\mu(\hat{x}) = \hat{\mu}$ (i.e., again a solution of minimum mean) and the variance $\sigma(\hat{x})$ is maximum. Denote $\hat{\sigma} := \sigma(\hat{x})$. Let x^j denote an optimum in (5). Initially we set $\bar{\mu} = \mu^0$ and $\bar{\sigma} = \sigma(x^0)$. For all j we will have $\bar{\sigma} \geq \sigma(x^j) \geq \hat{\sigma}$. Initially this is true for $j = 0$. Then the values μ^j , $j > 0$ are defined as follows: If

$$\mu(x^j) (\bar{\sigma} - \hat{\sigma}) + \sigma(x^j) (\hat{\mu} - \bar{\mu}) > \bar{\sigma} \hat{\mu} - \hat{\sigma} \bar{\mu}$$

then we disregard x^j as a candidate for tp-optimality due to Lemma 1, leave $\bar{\mu}$ and $\bar{\sigma}$ unchanged and define

$$\mu^{j+1} := \frac{\bar{\sigma} \hat{\mu} - \hat{\sigma} \bar{\mu} - \sigma(x^j) (\hat{\mu} - \bar{\mu})}{\bar{\sigma} - \hat{\sigma}}$$

Note that $\hat{\mu} < \mu^{j+1} < \mu(x^j) \leq \mu^j$. If, on the contrary,

$$\mu(x^j) (\bar{\sigma} - \hat{\sigma}) + \sigma(x^j) (\hat{\mu} - \bar{\mu}) \leq \bar{\sigma} \hat{\mu} - \hat{\sigma} \bar{\mu}$$

we keep x^j as a candidate for tp-optimality, and update

$$\bar{\mu} := \mu(x^j), \quad \bar{\sigma} := \sigma(x^j), \quad \mu^i := \mu(x^{j-1}) - \varepsilon$$

where again ε is sufficiently small, in the sense that there is no feasible x such that $\mu(x^j) - \varepsilon < \mu(x) < \mu(x^j)$. Since x^j is infeasible problem $j + 1$ the condition $\bar{\sigma} \geq \sigma(x^j) \geq \hat{\sigma}$ is valid also in the next iteration.

The two series are finite because the set \mathcal{X} is finite. Let r be the last index such that problem (4) is feasible and necessarily $\mu(x_r) = \hat{\mu}$. Let q be the last index such that problem (5) is feasible and necessarily $\mu(x^q) = \hat{\mu}$. Let us call equivalent two solutions x and x' such that $\mu(x) = \mu(x')$ and $\sigma(x) = \sigma(x')$.

Theorem 2. *Given an efficient solution x^* , such that $\mu(x^*) > \hat{\mu}$, associated to the efficient target pair (c, p) , this solution or an equivalent one is among the accepted optima x_i in (4) if $c > \hat{\mu}$ and among the accepted optima x^j in (5) if $c < \hat{\mu}$.*

Proof: We prove the statement by contradiction. Suppose that x^* is efficient for a target $c > \hat{\mu}$ and assume that $x^* \neq x_i$ for all i and for any equivalent optimum x^* . Since $c > \hat{\mu}$ we have $c > \mu(x^*)$ otherwise x^* could not be optimum in (2).

The solution x^* is feasible in (4) for $i = 0$. By using an inductive argument let us assume that x^* is feasible for a generic i . Then we have by optimality of x_i in (4)

$$\sigma(x^*) \geq \sigma(x_i) \tag{7}$$

Since x^* is efficient it is optimal in (2). We have two cases,

$$\text{either } i) \quad \frac{c - \mu(x^*)}{\sigma(x^*)} = \frac{c - \mu(x_i)}{\sigma(x_i)} \quad \text{or} \quad ii) \quad \frac{c - \mu(x^*)}{\sigma(x^*)} > \frac{c - \mu(x_i)}{\sigma(x_i)}$$

Note that the case $c - \mu(x^*) = c - \mu(x_i)$ and $\sigma(x^*) = \sigma(x_i)$ is excluded by the hypothesis that x_i is not an equivalent optimum of x^* . In case $i)$ we have from (7) and from $c > \mu(x^*)$ that $c - \mu(x^*) > c - \mu(x_i)$, i.e., $\mu(x^*) < \mu(x_i)$. Note that, by Lemma 1, case $i)$ cannot happen if x_i is not accepted because it would imply that also x^* is dominated for any target. Hence, due to the choice of ε , x^* is feasible for problem $(i + 1)$. In case $ii)$, if x_i is accepted, we have

$$\sigma(x_i) (c - \mu(x^*)) > \sigma(x^*) (c - \mu(x_i))$$

By (7) and from $c > \mu(x^*)$ we have

$$\sigma(x^*)(c - \mu(x^*)) \geq \sigma(x_i)(c - \mu(x^*)) > \sigma(x^*)(c - \mu(x_i))$$

so that, since $\sigma(x^*) > 0$ we get again $\mu(x_i) > \mu(x^*)$, i.e., feasibility of x^* for problem $(i + 1)$. If x_i is not accepted then, by Lemma 1

$$\mu(x^*)(\hat{\sigma} - \bar{\sigma}) + \sigma(x^*)(\bar{\mu} - \hat{\mu}) \leq \hat{\sigma}\bar{\mu} - \bar{\sigma}\hat{\mu}$$

otherwise x^* could not be efficient. Hence x^* is feasible in the next problem $(i + 1)$. Feasibility of x^* for problem $(i + 1)$ and hence for all i contradicts the fact that problem $(r + 1)$ is infeasible.

The proof for the iteration (5) is similar and is omitted. \square

Note that the converse of the theorem does not hold, that is, there are accepted optima of (4) and (5) that are not efficient. To detect the efficient solutions among the accepted optima we need the procedure that we present in the next section. The important fact pointed out by the theorem is that we do not miss efficient solutions by solving the sequence of problems (4) and (5) if $\mu(x^*) > \hat{\mu}$.

The case $\mu(x^*) = \hat{\mu}$ is critical. Besides the solutions x_r and x^q which are necessarily efficient there might be other efficient \bar{x} , optima in (2) for a target $c = \hat{\mu}$ such that $\sigma(x_r) < \sigma(\bar{x}) < \sigma(x^q)$. These efficient solutions cannot appear as optima neither in (4) nor in (5). In case $x^q \neq x^r$ and they are not equivalent we may scan all efficient solutions by minimizing $\mu(x)$ and constraining $\sigma(x)$. Note that all these solution have the same mean equal to the minimum expected cost and different variance. They all grant the same target $\hat{\mu}$ with probability 1/2 but having different variance they have different fluctuations. The larger the fluctuations the higher the probability of reaching good values but also bad values.

4 Recovering efficient solutions

It is useful to relabel all accepted optima found in problems (4) and (5) in order of decreasing values of $\sigma(x^j)$ and $\sigma(x_i)$. Hence the new indices h are such that first the accepted solutions from (5) are listed in that order and then follow the accepted solutions from (4) in the reverse order. From now on x_i will denote an accepted optimum in either (4) or (5) with the new index. Let us also denote $\mu_i := \mu(x_i)$, $\sigma_i := \sigma(x_i)$.

In order to detect efficiency of the optima and recover the associated target let us recall that by definition x_i is efficient if there exists a target c such that

$$\frac{c - \mu_i}{\sigma_i} \geq \frac{c - \mu(x)}{\sigma(x)} \quad \text{for all } x \in \mathcal{X}$$

However, we may restrict the check to the subset of efficient solutions, i.e., x_i is efficient if if there exists a target c such that

$$\frac{c - \mu_i}{\sigma_i} \geq \frac{c - \mu(x)}{\sigma(x)} \quad \text{for all efficient } x \in \mathcal{X}$$

Indeed if it were not so, there would exist another solution \bar{x} and in particular an efficient one such that

$$\frac{c - \mu_i}{\sigma_i} < \frac{c - \mu(\bar{x})}{\sigma(\bar{x})}$$

Hence to detect efficiency of x_i we may check among the solutions of (4) and (5) which include all tp-efficient solutions possibly enlarged with all tp-efficient solutions associated to $c = \hat{\mu}$. Therefore we simply check whether there exist target values c such that

$$\frac{c - \mu_i}{\sigma_i} \geq \frac{c - \mu_h}{\sigma_h} \quad \text{for all } h$$

i.e.

$$c(\sigma_i - \sigma_h) \leq \sigma_i \mu_h - \sigma_h \mu_i \quad \text{for all } h$$

Let us define

$$\alpha_{ij} := \frac{\sigma_i \mu_j - \sigma_j \mu_i}{\sigma_i - \sigma_j}$$

Note that $\alpha_{ij} = \alpha_{ji}$. The matrix α_{ij} (undefined on the diagonal) is therefore symmetric. Moreover, due to the index ordering, all values on the upper triangular part have indices i and j ($i < j$) such that $\sigma_i > \sigma_j$, whereas all values on the lower triangular part ($i > j$) have indices i and j such that $\sigma_i < \sigma_j$. Let us define

$$c_i^- := \max \{ \alpha_{ij} : i > j \}, \quad c_i^+ := \min \{ \alpha_{ij} : i < j \}$$

Hence x_i is efficient if and only if $c_i^- \leq c_i^+$ and its associated target is any c such that $c_i^- \leq c \leq c_i^+$ with probability

$$p_i = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{c - \mu_i}{\sigma_i \sqrt{2}} \right)$$

To compute the feasible interval extremes c_i^- and c_i^+ (note that c_i^+ is equal to c_j^- if j corresponds to the next feasible interval) we carry out the following iteration:

```

i := 1;  $c^- := -\infty$ ;
while  $i < n$  do
   $k := \operatorname{argmin}_{j > i} \alpha_{ij}$ ;  $c^+ := \alpha_{ik}$ ;
  output  $(c^-, c^+)$ ;
   $i := k$ ;  $c^- := c^+$ ;
end;
output  $(c^-, +\infty)$ ;

```

In words, the iteration works as follows: starting from the index $i = 1$ the minimum value $\alpha_{1k} = \min_{j > 1} \alpha_{1j}$ is computed. This is the first interval extreme. Then we go to row k . By construction it turns out that $\alpha_{k1} = \max_{j < k} \alpha_{kj}$. Then we compute the minimum $\alpha_{kh} = \min_{j > k} \alpha_{kj}$, record this value as a second interval extreme and move to row h . Again it turns out that α_{hk} is the maximum for α_{hj} , $j < h$. The procedure goes on until the last row is reached.

We note an interesting relation between the tp-efficient solutions, and the mean-variance efficient solutions. The tp-efficient x^* computed in the case $c > \hat{\mu}$ are also mean-variance efficient. In the case $c < \hat{\mu}$ all tp-efficient solutions are strictly dominated in the mean-variance approach. Moreover, there are mean-variance efficient solutions that are not tp-efficient, since all tp-efficient solutions are on the boundary of the convex hull of $\{(\mu(x), \sigma(x)) : x \in \mathcal{X}\}$, whereas there may be in general mean-variance efficient solutions within the convex hull.

5 Computing one efficient solution for a given target

In this section we consider the problem of finding just one tp-efficient solution for a fixed target c , i.e., solving problem (2), which we repeat

$$\max_{x \in \mathcal{X}} \frac{c - \mu(x)}{\sigma(x)} \quad (8)$$

As already observed, it is difficult to tackle this integer nonlinear programming problem directly. We propose again an iterative procedure to find the tp-efficient x^* . The iteration mimics the previous one designed to

X

find all efficient solutions, but it exploits the fact that c is fixed and hence the iteration is faster and the parameters are updated differently.

Again we have to consider two separate cases: either $c > \hat{\mu}$ or $c < \hat{\mu}$ (the case $c = \hat{\mu}$ is easy, see the previous discussion). As before, for the case $c > \hat{\mu}$, we iteratively solve the problem

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \sigma(x) \\ & \mu(x) \leq \mu_i \end{aligned} \tag{9}$$

for various values μ_i , $i = 0, \dots, r$, under the inductive hypothesis that at stage i there is available an incumbent solution \bar{x} and that, if $\bar{x} \neq x^*$, then $\mu(x^*) \leq \mu_i$, i.e., x^* is feasible for problem i .

At the first iteration of (9) we set $\mu_0 := c$. Let \hat{x} be the solution that has minimum mean $\mu(\hat{x}) = \hat{\mu}$ and that, among the solutions with minimum mean, has minimum variance $\sigma(\hat{x})$. We initially set $\hat{\sigma} := \sigma(\hat{x})$. Moreover, we set $\bar{x} := \hat{x}$ as a first incumbent. Hence the inductive hypothesis is verified at the first stage.

Let x' be the optimum in (9) and denote $\mu' = \mu(x')$, $\sigma' = \sigma(x')$. Hence there are no $(\mu(x), \sigma(x))$ pairs such that $\sigma(x) < \sigma'$ and $\mu(x) \leq \mu_i$ and by the inductive hypothesis, if $\bar{x} \neq x^*$, then $(\mu(x^*), \sigma(x^*))$ must belong to the set $C_1 = \{(\mu, \sigma) : \mu \leq \mu_i, \sigma \geq \sigma'\}$. Moreover due to the existence of the incumbent \bar{x} , $(\mu(x^*), \sigma(x^*))$ must belong to the intersection of C_1 with the set

$$C_2 = \left\{ (\mu, \sigma) : \frac{c - \hat{\mu}}{\hat{\sigma}} \leq \frac{c - \mu}{\sigma} \right\}$$

If (case 1, refer to Figure 4 above at left, where the sets to which $(\mu(x^*), \sigma(x^*))$ cannot belong are denoted in grey)

$$\frac{c - \mu'}{\sigma'} < \frac{c - \hat{\mu}}{\hat{\sigma}}$$

then x' is dominated by the incumbent \bar{x} and we set μ_{i+1} for the next problem as (Figure 4 above at right)

$$\mu_{i+1} := c - \frac{\sigma'}{\hat{\sigma}} (c - \hat{\mu})$$

and the inductive hypothesis is satisfied. If (case 2, refer to Figure 4 below at left)

$$\frac{c - \mu'}{\sigma'} \geq \frac{c - \hat{\mu}}{\hat{\sigma}}$$

then the incumbent \bar{x} is dominated by x' if the inequality is strict. In any case we reset $\bar{x} := x'$ and the set C_2 may be strengthened, thanks to the new incumbent \bar{x} , to

$$C_2 = \left\{ (\mu, \sigma) : \frac{c - \mu'}{\sigma'} < \frac{c - \mu}{\sigma} \right\}$$

so that we may reset (Figure 4 below at right)

$$\hat{\sigma} := \sigma' \frac{c - \hat{\mu}}{c - \mu'}$$

and we need to set μ_{i+1} as $\mu_{i+1} := \mu' - \varepsilon$ as we did before. The iteration terminates when $\mu_i < \hat{\mu}$. Then $x^* := \bar{x}$.

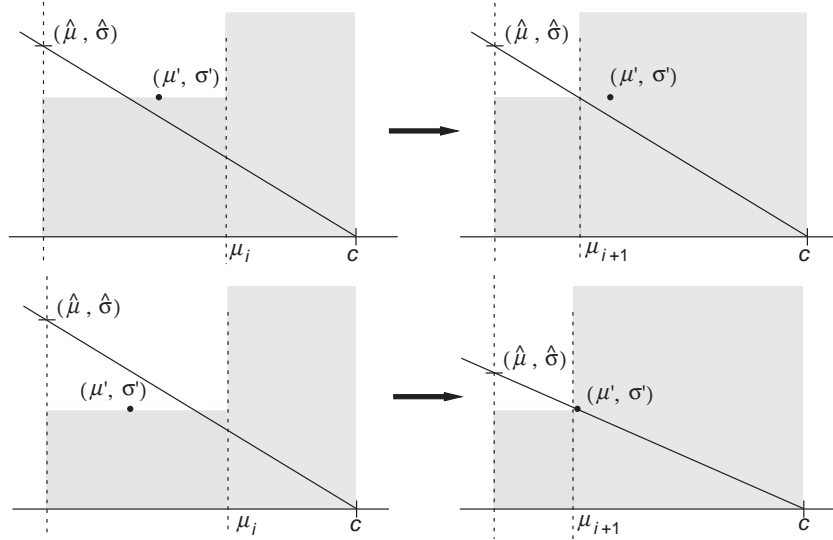


Fig. 4. Iteration update: case 1, above; case 2, below

A similar iteration works if $c < \hat{\mu}$. In this case we have to solve

$$\begin{aligned} \max_{x \in \mathcal{X}} \quad & \sigma(x) \\ & \mu(x) \leq \mu^j \end{aligned} \quad (10)$$

for various values μ^j . The initial value μ^0 is the maximum mean among all feasible solutions and the first incumbent \bar{x} is the solution \hat{x} that has minimum mean $\mu(\hat{x}) = \hat{\mu}$ and that, among the solutions with minimum mean, has maximum variance $\sigma(\hat{x})$. The iteration works exactly as in the previous case. We leave to the reader the easy verification of this fact. It is also not difficult to extend these ideas to the case of computing a single efficient solution granting the best target for a given probability threshold.

6 Solving the associated constrained problems

In both series we can equivalently replace the objective function with $\sigma^2(x)$. If the random variables C_e are independent then

$$\sigma^2(x) = \sum_{e \in x} \sigma_e^2$$

so that (4), (5), (9) and (10) become

$$\begin{aligned} \min (\max) \quad & \sum_{e \in x} \sigma_e^2 \\ & \sum_{e \in x} \mu_e \leq \bar{\mu} \\ & x \in \mathcal{X} \end{aligned} \quad (11)$$

Even if the combinatorial problem $\min_{x \in \mathcal{X}} (\max_{x \in \mathcal{X}} \sum_{e \in x} \sigma_e^2)$ is an easy one, usually adding the constraints $\sum_{e \in x} \mu_e \leq \bar{\mu}$ turns the problem into a NP-hard one. If the combinatorial problem can be modeled as 0-1 linear programming problem, then the constraint can be easily added to the model.

If the random variables are not independent, we have

$$\sigma^2(x) = \sum_{e, e' \in x} \sigma_{e, e'}^2$$

In this case the corresponding 0-1 programming problem has a quadratic objective function, and this is notoriously difficult. If there is a limited number of correlated variables, the problem can be linearized with little added complexity.

7 Computational complexity issues

Besides the inherent difficulty of solving (11) (especially when the variables are correlated) another crucial issue with regard to the overall computation time is related to the number of iterations needed to solve the series of problems (4) and (5) in case we want to compute all tp-efficient solutions or the series of problems (9) and (10) in case we want to compute only one tp-efficient solution for a given target.

Apparently, we need to solve (4) as many times as there are mean-variance different efficient solutions. The only trivial bound on the number of iterations is given by the number of feasible solutions itself. Indeed we can build instances such that this bound is reached. For instance (let $E = \{1, \dots, n\}$ and $\mathcal{X} = 2^E$) assume that $\mu_i = -2^{i-1}$, $i = 1, \dots, n$, and $\sigma_{i, j}^2 = \mu_i \mu_j$, $i \neq j$ and $\sigma_{i, i}^2 = \mu_i^2 (1 + \eta)$, with η like a very small ‘noise’. This is an extreme example with all variables very strongly correlated (the correlation coefficients are close to $1 - \eta$). Then two different feasible solutions have different pairs $(\mu(x), \sigma(x))$ and all of them are mean-variance efficient (all $(\mu(x), \sigma(x))$ pairs lie close to the straight line $\sigma + \mu = 0$ where they belong in the limit as η approaches 0). Hence we have to enumerate all feasible solutions. However, there are only two tp-efficient solutions corresponding to the ones with minimum and maximum mean respectively, i.e., the whole set and the empty set.

If we want to compute just one efficient solution for this example, the number of iterations is smaller and it decreases as the target decreases.

We may argue that, for real data, means and variances are not so strongly correlated and so the number of iterations should be much lower. Indeed the computational experiments show that the number of iterations is usually not very high. The computational bottleneck is more due to the problems (11) themselves rather than to the number of times they must be solved.

8 An example

We illustrate the procedure with an example. Suppose the combinatorial optimization problem is a knapsack problem with $n = 12$ items. This example could refer to a portfolio problem with no possibility of investing fractional amounts of stock options. The data we provide for this example are just random fictitious data. The knapsack capacity, i.e., the available budget, is $K = 13$. The weights, i.e., the amounts that can be invested, are the following deterministic values:

$$w = (3 \ 2 \ 3 \ 4 \ 4 \ 5 \ 5 \ 3 \ 4 \ 3 \ 3 \ 5)$$

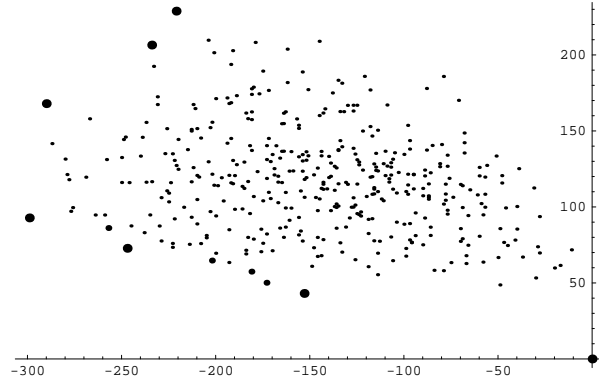


Fig. 5. All knapsack solutions of the example

whereas the costs, which are negative profits in this case, are jointly normal with means

$$\mu = -(17 \ 86 \ 30 \ 28 \ 49 \ 20 \ 67 \ 40 \ 70 \ 94 \ 29 \ 11)$$

and covariance matrix

$$\sigma^2 = \begin{pmatrix} 3786 & 1874 & -375 & -1538 & -281 & -1437 & -2163 & 328 & 1744 & 231 & 319 & -83 \\ 1874 & 7516 & 1173 & -4653 & 116 & -3458 & -4799 & -5276 & -2417 & -3942 & 2666 & -2369 \\ -375 & 1173 & 2840 & -1822 & -920 & -984 & -325 & -1394 & -485 & -202 & 1080 & -938 \\ -1538 & -4653 & -1822 & 4862 & 924 & 3050 & 2943 & 3099 & 444 & 1995 & -172 & 2828 \\ -281 & 116 & -920 & 924 & 2371 & 150 & -135 & -537 & -2337 & -1397 & 1540 & -504 \\ -1437 & -3458 & -984 & 3050 & 150 & 3580 & 1996 & 1811 & 978 & 978 & -840 & 1960 \\ -2163 & -4799 & -325 & 2943 & -135 & 1996 & 3940 & 2711 & 1192 & 2370 & -1743 & 2065 \\ 328 & -5276 & -1394 & 3099 & -537 & 1811 & 2711 & 7299 & 2694 & 6300 & -1413 & 2685 \\ 1744 & -2417 & -485 & 444 & -2337 & 978 & 1192 & 2694 & 5986 & 3053 & -3429 & 2684 \\ 231 & -3942 & -202 & 1995 & -1397 & 978 & 2370 & 6300 & 3053 & 6584 & -1338 & 3331 \\ 319 & 2666 & 1080 & -172 & 1540 & -840 & -1743 & -1413 & -3429 & -1338 & 5447 & -271 \\ -83 & -2369 & -938 & 2828 & -504 & 1960 & 2065 & 2685 & 2684 & 3331 & -271 & 5152 \end{pmatrix}$$

For illustration purposes we show in Figure 5 how the 409 feasible solutions are mapped in the plane (μ, σ) . The optima accepted by the iterative procedure are the largest dots (one of them turns out to be non efficient), whereas the optima discarded by the iterative procedure are the dots of intermediate size.

The minimum and maximum mean values are $\hat{\mu} = -299$ and $\mu^0 = \mu_0 = 0$. Problems (4) and (5) have been solved by using binary quadratic programming. Due to the small size of the instance, the solutions have been found very quickly. Solving the series of problems (4) yields the following 8 optima, with the optima discarded by the procedure shown in italics:

$$\{\}, \{2, 7\}, \{2, 6, 7\}, \{2, 4, 7\}, \{2, 5, 7\}, \{2, 7, 10\}, \{2, 4, 5, 10\}, \{2, 5, 9, 10\}$$

with corresponding (μ_i, σ_i) pairs (σ_i figures rounded to integers)

$$(0, 0), (-153, 43), (-173, 50), (-181, 57), (-202, 65), (-247, 73), (-257, 86), (-299, 93)$$

and solving the series of problems (5) yields the following 4 optima (the last one is the same as the last one of the previous series) all of which are accepted

$$\{1, 8, 9, 10\}, \{3, 8, 9, 10\}, \{2, 8, 9, 10\}, \{2, 5, 9, 10\}$$

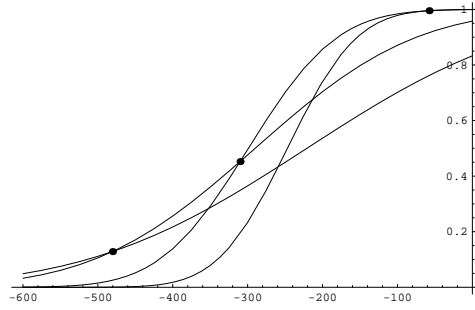


Fig. 6. Distribution functions of four tp-efficient solutions

with corresponding (μ^j, σ^j) pairs:

$$(-221, 229), (-234, 206), (-290, 168), (-299, 93)$$

From these data we compute the α_{ij} matrix as (all figures rounded to integers)

$$\alpha = \begin{pmatrix} & -354 & -480 & -352 & -259 & -137 & 0 \\ -354 & & -534 & -352 & -254 & -132 & 0 \\ -480 & -534 & & -310 & -214 & -106 & 0 \\ -352 & -352 & -310 & & -58 & -26 & 0 \\ -259 & -254 & -214 & -58 & & -17 & 0 \\ -137 & -132 & -106 & -26 & -17 & & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

from which we get the following 5 values for the target intervals

$$\{-480.429, -310.109, -57.7276, -16.496, 0.\}$$

which yield the following 6 efficient solutions together with the target and probability efficient values:

$x^1 = \{1, 8, 9, 10\}$	$-\infty < c \leq -480.429$	$0 < p \leq 0.128437$
$x^2 = \{2, 8, 9, 10\}$	$-480.429 \leq c \leq -310.109$	$0.128437 \leq p \leq 0.452349$
$x^3 = \{2, 5, 9, 10\}$	$-310.109 \leq c \leq -57.7276$	$0.452349 \leq p \leq 0.995344$
$x^4 = \{2, 7, 10\}$	$-57.7276 \leq c \leq -16.496$	$0.995344 \leq p \leq 0.999229$
$x^5 = \{2, 7\}$	$-16.496 \leq c \leq 0$	$0.999229 \leq p \leq 0.999807$
$x^6 = \{\}$	$0 \leq c < \infty$	$0.999807 \leq p \leq 1$

Clearly, for practical purposes, we may disregard x^4 , x^5 and x^6 since they have probability values almost indistinguishable from x^3 , which can grant better targets. The (c, p) efficient set for x^1 , x^2 , x^3 and x^4 is shown in Figure 6. It is the upper envelope of the four distribution functions $F_{x^i}(c)$, $i = 1, \dots, 4$. The breakpoints between the efficient solutions are also shown.

9 Conclusions

We have considered random combinatorial optimization problems with the randomness only in the cost coefficients. In particular the random variables representing the costs are assumed to be jointly normal, so that the cost of a combinatorial object, expressed as the sum of the cost random variables, is also normal. Target-probability efficient solutions are those solutions which grant a given target at a certain probability and no other solution can grant the same target at better probability. Normality of the costs allows for iterative computational schemes to find either all target-probability efficient solutions or just one. The computational scheme is based on a variance minimization/maximization with a constraint on the mean. If the costs are not correlated this is a linear objective function together with a linear constraint which has to be added to the constraints defining the particular combinatorial problem at hand.

References

1. Atamtürk, A and V. Narayanan: Polymatroids and mean-risk minimization in discrete optimization, *Operations Research Letters*, **36**, 618-622 (2008).
2. de Finetti, B.: Il problema dei pieni, *Giornale Istituto Italiano Attuari*, **9**, 1-88 ; English translation by L. Barone available as “The problem of “Full-risk insurances”, Ch. 1 ‘The problem in a single accounting period’, *Journal of Investment Management*, **4**, 19-43, 2006 (1940).
3. Linsmeier, T.J. and N.D. Pearson: Value at Risk, *Financial Analysts Journal*, **56**, 47-67 (2000).
4. Markowitz, H.: Portfolio selection, *Journal of Finance*, **6**, 77-91 (1952).
5. Marshall, C. and M. Siegel: Value at Risk: Implementing a risk measurement standard, *The Journal of Derivatives*, **4**, 91-111 (1997).
6. Nikolova, E.: Approximation Algorithms for Reliable Stochastic Combinatorial Optimization, in: M. Serna, R. Shaltiel, K. Jansen, Klaus and J. Rolim (eds), *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 338-351, LNCS 6302, Springer (2010).