The multi-handler knapsack problem under uncertainty

The Multi-Handler Knapsack Problem under Uncertainty

Guido Perboli[a,b], Roberto Tadei[a,*], Luca Gobbato[a]

[a] *Politecnico di Torino, Turin, Italy*
[b] *CIRRELT, Montreal, Canada*

Abstract

The Multi-Handler Knapsack Problem under Uncertainty is a new stochastic knapsack problem where, given a set of items, characterized by volume and random profit, and a set of potential handlers, we want to find a subset of items which maximizes the expected total profit. The item profit is given by the sum of a deterministic profit plus a stochastic profit due to the random handling costs of the handlers. On the contrary of other stochastic problems in the literature, the probability distribution of the stochastic profit is unknown. By using the asymptotic theory of extreme values, a deterministic approximation for the stochastic problem is derived. The accuracy of such a deterministic approximation is tested against the two-stage with fixed recourse formulation of the problem. Very promising results are obtained on a large set of instances in negligible computing time.

Keywords: knapsack problem, stochastic profit, multiple handlers, deterministic approximation.

1. Introduction

The increasing competition due to both globalization of production processes and the movement of large quantities of freight between continents and countries creates the need for new tools for strategic and tactical decisions, that are able to deal with the stochastic nature of the processes involved. While this leads to new location and transportation problems [1, 2], only a few studies deal with the stochastic study of packing problems [3]. This is mainly due to the peculiarities of the literature related to packing. In fact, even if packing problems play a central role in transportation and logistics, the problems presented in the literature are mainly related to operational issues [4, 5]. Moreover, the parameter uncertainty affecting the final solutions such as the profit associated to item delivery or the cost of container renting is usually more evident in the planning phases rather than in the operational ones.

In this paper we introduce a new stochastic variant of the knapsack problem, the Multi-Handler Knapsack Problem under Uncertainty (MHKP$_u$). Given a set of items, characterized by volume and random profit, and a set of potential logistics handlers, the problem consists in finding a subset of items which maximizes the expected total profit. The profit is given by the sum of a deterministic profit and a stochastic profit oscillation, with unknown probability distribution, due to the random handling costs of the handlers.

A large number of real-life situations can be satisfactorily modeled as a MHKP$_u$, e.g. in financial and resource allocation. The general idea is to think of the capacity of the knapsack as the available amount of a resource (i.e. budget) and the items as activities to which this resource can be allocated (i.e. shares). Moreover, these items present profits which are random variables. The MHKP$_u$ may also appear as a subproblem of larger optimization problems.

A specific application of the MHKP$_u$ can be found in the automotive sector [6]. There the delivery of cars from manufacturers to dealers is not managed by the manufacturers themselves, but is delegated to specialized companies. These companies manage both the finishing operations on the cars (e.g. removal of the protective wax, installation of specific accessories, etc.) and the logistics operations linked to delivery to the dealers. In order to have a more flexible structure, the fleet of auto-carriers used to deliver the cars is only partially owned by each company, while a substantial part of the deliveries is sub-contracted to micro-companies with highly

*Corresponding author:
Department of Control and Computer Engineering, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino (Italy)
Tel.: +39 0110907032, Fax: +39 0110907099 E-mail: roberto.tadei@polito.it

variable random costs. Moreover, the auto-carriers have different capacities due to the presence of specific technical features. From the point of view of the cars that must be delivered, the net profit for the company is affected by different factors, including delays in the finishing operations, additional costs due to violations of the negotiated deadlines or additional transportation costs.

Another example of real-world applications of the MHKP$_u$ comes from trans-continental naval shipping operations, where freight transportation from eastern ports to Europe and North America is managed by specialized companies. The competition between the transportation companies, as well as the possibility of managing the port cranes by different operators, force the companies to consider both the profit given by the shipped items and the additional costs due to the logistics operations.

The problem can be also seen as a relaxation of container loading problems where the capacity of the given containers are collapsed into one single container. This leads to an approximation of the real problem suitable in strategic and tactical planning, where the stochastic nature of the profits in more relevant than the actual loading of the items. Moreover, it is required to obtain accurate solutions within a limited computational effort in order to explore multiple scenarios of the underlying business model.

Other applications come from the domain of Smart City and City Logistics, and particularly in the last mile delivery. One of the trends is to substitute traditional single-echelon routing structures with two and three-echelon ones. The reason is the willingness to use ''green'' vehicles inside the city, whilst consolidating the freight in medium and small sized transshipment depots, called satellites [7, 8, 9]. In the satellites different sequences of consolidation operations are done by different workers. The different skill levels of the workers can cause delay in the operations reducing the overall profit. A similar problem is present in yard management, where the profit oscillations are given by the operations done by workers working for yard management companies, different in skills and reliability.

In general, the MHKP$_u$ arises in logistics and production scheduling applications, where a single item can be managed by several handlers (third-party logistics providers or sub-contractors), whose costs affect the net profit of the item itself. The large number of possible handler cost scenarios and the difficulty to measure the associated handler costs suggest the representation of these net profits as stochastic variables with unknown probability distribution.

This paper introduces the formulation of the stochastic problem. From this formulation, a deterministic approximation is derived. In particular, under a mild hypothesis on the unknown probability distribution, the deterministic approximation becomes a knapsack problem where the total expected profit of the loaded items is proportional to the logarithm of the total accessibility of those items to the set of handlers. Moreover, at optimality, the percentage of an item handled by any handler is given by a multinomial Logit model.

The paper is organized as follows. The literature review is introduced in Section 2. In Section 3 the model of the MHKP$_u$ is given. Section 4 is devoted to presenting the deterministic approximation of the MHKP$_u$, while in Section 5 its two-stage program with fixed recourse is given. Finally, in Section 6 the deterministic approximation and the two-stage program with fixed recourse are tested and compared on a set of newly introduced instances. The conclusion of our work is reported in Section 7.

2. Literature review

While different variants of the stochastic knapsack problem are present in the literature, the MHKP$_u$ is absent. For this reason, we will consider some relevant literature on similar problems, highlighting the main differences with the problem faced in this paper.

A first group of studies consider deterministic profits and random volumes, with the goal of maximizing the total expected value of selected items, while ensuring that the probability to satisfy the knapsack capacity is limited by some upper bounds. Usually, heavy assumptions on the distribution of the random volumes are considered (e.g. [10], [11] where item volumes have a Bernoulli distribution, and [12], [13] where the distribution is a Normal one). These assumptions on the distributions heavily limit the possibility to extend the results to other variants of the problem.

A second group of studies deals with random profits and the goal to assign a set of items to the knapsack in order to maximize the probability of achieving some target total value. They are usually more related to financial and economic issues than to the impact of the operations on the final revenue [14, 15, 16]. Unfortunately, these problems differ from MHKP$_u$ because they consider the random profit associated only to the item, while in MHKP$_u$

the randomness is given by the interaction between the item and the handler managing the loading/unloading operations.

Finally, from a methodological point of view, the study most similar to the present paper is [3], where the authors consider the stochastic version of the Generalized Bin Packing Problem, a recently introduced packing problem where, given a set of bins characterized by volume and cost and a set of items characterized by volume and profit (which also depends on bins), a subset of items is selected for loading into a subset of bins which maximizes the total net profit, while satisfying the volume and bin availability constraints [17]. Similarly to MHKP$_u$, the item profits are random variables and the probability distribution of these random variables is assumed to be unknown.

3. The MHKP$_u$

In the MHKP$_u$ the item profits are random variables. In fact, they are composed by a deterministic profit plus a random term, which represents the profit oscillation due to the handling costs occurred by the different handlers for preparing items for loading. In practice, such profit oscillations randomly depend on the handling scenarios adopted by the handlers for preparing items for loading and are actually very difficult to be measured. This implies that the probability distribution of these random terms must be assumed as unknown.

Let it be

- I: set of items
- J: set of handlers
- L: set of handling scenarios for loading items into the knapsack
- p_i: non-negative deterministic profit of item i
- p_{ij}: non-negative deterministic profit of item i when loaded by handler j
- $\tilde{\theta}^{jl}$: random profit oscillation of any item when it is loaded by handler j under scenario $l \in L$
- $\tilde{p}_{ij}(\tilde{\theta}^{jl}) = p_{ij} + \tilde{\theta}^{jl}$: random profit of item i when loaded by handler j under scenario l
- y_i: boolean variable equal to 1 if item i is loaded, 0 otherwise
- x_{ij}: percentage of item i handled by handler j
- w_i: volume of item i
- W: knapsack capacity.

The MHKP$_u$ is formulated as follows

$$\max_{\{y,x\}} \quad \sum_{i\in I} p_i y_i + \mathbb{E}_{\{\tilde{\theta}^{jl}\}}\left[\sum_{i\in I}\sum_{j\in J}\sum_{l\in L} \tilde{p}_{ij}(\tilde{\theta}^{jl}) x_{ij}\right] \tag{1}$$

subject to

$$\sum_{i\in I} w_i y_i \leq W \tag{2}$$

$$\sum_{j\in J} x_{ij} = y_i \quad i \in I \tag{3}$$

$$y_i \in \{0,1\} \quad i \in I \tag{4}$$

$$x_{ij} \geq 0 \quad i \in I, \quad j \in J. \tag{5}$$

The objective function (1) expresses the maximization of the profit of the items loaded into the knapsack plus the expected value of the handling profit; constraint (2) ensures that the capacity of the knapsack is not exceeded;

constraints (3) guarantee that any item is completely processed by some handlers only if it is loaded. Finally, (4)-(5) are the integrality and non-negativity constraints, respectively.

Let us assume that $\tilde{\theta}^{jl}$ are independent and identically distributed (i.i.d.) random variables with a common and unknown probability distribution

$$F(x) = Pr\{\tilde{\theta}^{jl} \leq x\}. \tag{6}$$

Let us define with $\tilde{\theta}^{j}$ the maximum of the random profit oscillations $\tilde{\theta}^{jl}$ for handler j among the alternative scenarios $l \in L$

$$\tilde{\theta}^{j} = \max_{l \in L} \tilde{\theta}^{jl} \qquad j \in J. \tag{7}$$

Because F(x) is unknown, $\tilde{\theta}^{j}$ is still of course a random variable with unknown probability distribution given by

$$B_j(x) = \Pr\left\{\tilde{\theta}^{j} \leq x\right\} \qquad j \in J. \tag{8}$$

As, for any handler j, $\tilde{\theta}^{j} \leq x \Longleftrightarrow \tilde{\theta}^{jl} \leq x,\ l \in L$ and $\tilde{\theta}^{jl}$ are independent, using (6) one gets

$$B_j(x) = \prod_{l \in L} \Pr\left\{\tilde{\theta}^{jl} \leq x\right\} = \prod_{l \in L} F(x) = [F(x)]^{|L|} \qquad j \in J. \tag{9}$$

We assume that the knapsack loading is efficiency-based so that, for any item i and any handler j, among the alternative scenarios $l \in L$ the one which maximizes the random profit $\tilde{p}_{ij}(\tilde{\theta}^{jl})$ will be selected. This does not mean that, in the stochastic problem, we select the model scenario which maximizes the profit, but that, when the actual profits become known (e.g. in day-by-day operations, the profits of a given day), the choice among the different alternatives is done by taking the most profitable one.

Then, the random profit of item i when it is loaded (i.e. $y_i = 1$) by handler j becomes

$$\tilde{p}_{ij}(\tilde{\theta}^{j}) = \max_{l \in L} \tilde{p}_{ij}(\tilde{\theta}^{jl}) = p_{ij} + \max_{l \in L} \tilde{\theta}^{jl} = p_{ij} + \tilde{\theta}^{j} \qquad i \in I : y_i = 1, j \in J. \tag{10}$$

The maximum profit oscillation $\tilde{\theta}^{j}$ can be either positive or negative, but, in practice, its absolute value does not overcome the profit p_{ij}, so that $\tilde{p}_{ij}(\tilde{\theta}^{j})$ is always non negative.

The expected maximum total profit of the loaded items is obtained by solving the following problem

$$\mathbb{E}_{\{\tilde{\theta}\}}\left[\max_{\{x\}} \sum_{i \in I : y_i = 1} \sum_{j \in J} \tilde{p}_{ij}(\tilde{\theta}^{j}) x_{ij}\right] \tag{11}$$

$$\sum_{j \in J} x_{ij} = 1 \quad i \in I : y_i = 1 \tag{12}$$

$$x_{ij} \geq 0 \quad i \in I : y_i = 1, \quad j \in J. \tag{13}$$

The objective function (11) maximizes the expected total profit for the loaded items. Constraints (12) guarantee that each loaded item is completely processed by some handlers, while (13) are the non-negativity constraints.

For each item i, let us consider the handler $j = i^*$ (for the sake of simplicity, we assume it is unique), which gives the maximum random profit for loading the item.

The maximum random profit for loading item i then becomes

$$\tilde{p}_i(\tilde{\theta}^{i^*}) = \max_{j \in J} \tilde{p}_{ij}(\tilde{\theta}^{j}) \qquad i \in I : y_i = 1 \tag{14}$$

and the optimal variables $\left\{x_{ij}\right\}$ are

$$x_{ij} = \begin{cases} 1, & \text{if} \quad j = i^* \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

which satisfy (12) and (13).

Using (14), (15), and the linearity of the expected value operator $\mathbb{E}$, the objective function (11) becomes

$$\mathbb{E}_{\{\tilde{\theta}^*\}}\left[\sum_{i\in I:y_i=1}\tilde{p}_i(\tilde{\theta}^{i^*})\right]=\sum_{i\in I:y_i=1}\mathbb{E}_{\{\tilde{\theta}^*\}}\left[\tilde{p}_i(\tilde{\theta}^{i^*})\right]=\sum_{i\in I:y_i=1}\hat{p}_i \tag{16}$$

where

$$\hat{p}_i=\mathbb{E}_{\{\tilde{\theta}^*\}}\left[\tilde{p}_i(\tilde{\theta}^{i^*})\right]\qquad i\in I:y_i=1. \tag{17}$$

The MHKP$_u$ (1)-(5) then becomes

$$\max_{\{y\}}\quad\sum_{i\in I}(p_i+\hat{p}_i)y_i \tag{18}$$

$$\sum_{i\in I}w_iy_i\le W \tag{19}$$

$$y_i\in\{0,1\}\qquad i\in I. \tag{20}$$

However, the calculation of $\hat{p}_i$ in (18) requires to know the probability distribution of the maximum random profit for loading item i, i.e. $\tilde{p}_i(\tilde{\theta}^{i^*})$ in (17), which will be derived in the next section.

4. The deterministic approximation of the MHKP$_u$

By (10) and (14), let

$$G_i(x)=Pr\left\{\tilde{p}_i(\tilde{\theta}^{i^*})\le x\right\}=Pr\left\{\max_{j\in J}\left[p_{ij}+\tilde{\theta}^j\right]\le x\right\}\qquad i\in I \tag{21}$$

be the probability distribution of the maximum random profit for loading item i.

As, for any item i, $\max_{j\in J}\left[p_{ij}+\tilde{\theta}^j\right]\le x\iff\left[p_{ij}+\tilde{\theta}^j\right]\le x,\quad j\in J$, and the random variables $\tilde{\theta}^j$ are independent (because $\tilde{\theta}^{jl}$ are independent), due to (8) and (9), $G_i\{x\}$ in (21) becomes a function of the total number $|L|$ of handling scenarios for loading as follows

$$\begin{aligned}G_i(x,|L|)\quad&=Pr\left\{\max_{j\in J}\left[p_{ij}+\tilde{\theta}^j\right]\le x\right\}=\prod_{j\in J}Pr\left\{\left[p_{ij}+\tilde{\theta}^j\right]\le x\right\}\\&=\prod_{j\in J}Pr\left\{\tilde{\theta}^j\le x-p_{ij}\right\}=\prod_{j\in J}B_j\left(x-p_{ij}\right)\\&=\prod_{j\in J}\left[F\left(x-p_{ij}\right)\right]^{|L|}\qquad i\in I.\end{aligned} \tag{22}$$

First, let us consider the following aspect: the optimal solution of problem (11)-(13) does not change if any arbitrary constant is added or subtracted to the random variables $\tilde{\theta}^j$.

Let us choose this constant as the root a of the equation

$$1-F(a)=1/|L|. \tag{23}$$

Let us assume that $|L|$ is large enough to use the asymptotic approximation $\lim_{|L|\to+\infty}G_i(x,|L|)$ as a good approximation of $G_i(x)$, i.e.

$$G_i(x)=\lim_{|L|\to+\infty}G_i(x,|L|))\qquad i\in I. \tag{24}$$

The calculation of the limit in (24) would require to know the probability distribution $F(.)$ in (6), which is still unknown. From [3], we know that under a mild assumption on the shape of the unknown probability distribution $F(.)$ (i.e. it is asymptotically exponential in its right tail), the limit in (24) tends towards the following Gumbel [18] probability distribution

$$G_i(x) = \lim_{|L|\to+\infty} G_i(x, |L|)) = \exp\left(-A_i e^{-\beta x}\right) \qquad i \in I \tag{25}$$

where $\beta > 0$ is a parameter to be calibrated and

$$A_i = \sum_{j\in J} e^{\beta p_{ij}} \qquad i \in I \tag{26}$$

is the *accessibility*, in the sense of Hansen [19], of item i to the set of handlers.

The accessibility in the sense of Hansen is defined as the potential of opportunities for interaction and is a measure of the intensity of the possibility of interaction. Here the interaction is between item and handlers. (26) shows that the accessibility of an item to the set of handlers is proportional to a function of profits associated to the different handlers.

Using the probability distribution $G_i(x)$ given by (25), after some manipulations, $\hat{p}_i$ in (17) becomes

$$\hat{p}_i = \int_{-\infty}^{+\infty} x dG_i(x) = \int_{-\infty}^{+\infty} x \exp\left(-A_i e^{-\beta x}\right) A_i e^{-\beta x} \beta dx = 1/\beta(\ln A_i + \gamma) \qquad i \in I \tag{27}$$

where $\gamma \simeq 0.5772$ is the Euler constant.

By (27), the MHKP$_u$ (18)-(20) becomes

$$\begin{aligned} & \max_{\{y\}} \sum_{i\in I} p_i y_i + \frac{1}{\beta} \sum_{i\in I} y_i \ln A_i + \frac{\gamma}{\beta} \sum_{i\in I} y_i = \\ = \; & \max_{\{y\}} \sum_{i\in I} p_i y_i + \frac{1}{\beta} \ln \prod_{i\in I} A_i^{y_i} + \frac{\gamma}{\beta} \sum_{i\in I} y_i = \\ = \; & \max_{\{y\}} \sum_{i\in I} p_i y_i + \frac{1}{\beta} \ln \Phi + \frac{\gamma}{\beta} \sum_{i\in I} y_i \end{aligned} \tag{28}$$

subject to (19)-(20), where $\Phi = \prod_{i\in I} A_i^{y_i}$ is the *total accessibility* of the loaded items to the set of handlers.

It is interesting to observe that the total expected profit of the loaded items is proportional to the logarithm of the total accessibility of those items to the set of handlers.

In the following, we will refer to the deterministic approximation of the MHKP$_u$ as DA-MHKP$_u$.

The following theorem holds

Theorem 1. *At optimality, the percentage of each item i handled by handler j, x_{ij}, is given by*

$$x_{ij} = \frac{e^{\beta p_{ij}}}{\sum_{j'\in J} e^{\beta p_{ij'}}}, \quad i \in I, j \in J. \tag{29}$$

Proof. At optimality, the probability that item i is handled by handler j is equal to the probability that handler j is that one of maximum profit. Then, from the Total Probability Theorem [20], one obtains

$$\begin{aligned} x_{ij} &= \int_{-\infty}^{+\infty} \prod_{v\neq j} \exp\left[-e^{-\beta(x-p_{iv})}\right] d\left[\exp\left(-e^{-\beta(x-p_{ij})}\right)\right] = \\ &= e^{\beta p_{ij}} \int_{-\infty}^{+\infty} \beta e^{-\beta x} exp(-A_i e^{-\beta x}) dx = \\ &= e^{\beta p_{ij}} \int_0^{+\infty} e^{-A_i t} dt = \frac{e^{\beta p_{ij}}}{A_i} = \\ &= \frac{e^{\beta p_{ij}}}{\sum_{j'\in J} e^{\beta p_{ij'}}} \quad i \in I, j \in J \end{aligned} \tag{30}$$

where $t = e^{-\beta x}$. □

□

It is trivial to check for x_{ij} the satisfaction of constraints (12) and (13).

Expression (29) represents a multinomial Logit model, which is widely used in choice theory [21]. In our case, it describes how the optimal handling of item i is split among different handlers j, due to the stochastic handling profit of item i.

5. The MHKP$_u$ as a two-stage program with fixed recourse

Approximating the profit stochasticity by discretizing the probability distributions and generating a set of scenarios $S \subseteq L$, the MHKP$_u$ (1)-(5) may be interpreted as a two-stage program with fixed recourse.

Let be the variables

- y_i : first stage decision variable equals to 1 if item i is loaded, 0 otherwise
- x_{ij} : first stage decision variable which represents the percentage of item i handled by handler j
- y_i^{+s} : second stage decision variable equals to 1 if item i is loaded under scenario s, 0 otherwise
- y_i^{-s} : second stage decision variable equals to 1 if item i is unloaded under scenario s, 0 otherwise
- x_{ij}^{+s} : second stage decision variable which represents the percentage of loaded item i handled by handler j under scenario s
- x_{ij}^{-s} : second stage decision variable which represents the percentage of unloaded item i handled by handler j under scenario s.

Moreover, we define by

$$\pi_{ij}^{+s} = p_i + p_{ij} + \tilde{\theta}^{js} \tag{31}$$

and

$$\pi_{ij}^{-s} = -\pi_{ij}^{+s} - \pi'_{ij} \tag{32}$$

the stochastic profits related to loading and unloading operations in the second stage, respectively, where $-\pi'_{ij}$ represents an extra cost to be paid for unloading item i by handler j in the second stage.

Finally, given the probability ρ_s of each second-stage scenario s, the two-stage program with fixed recourse, named 2S-MHKP$_u$, is formulated as follows

$$\max_{\{y,x\}} \quad \sum_{i\in I} p_i y_i + \sum_{i\in I}\sum_{j\in J} p_{ij}x_{ij} + \sum_{s\in S}\rho_s\left[\sum_{i\in I}\sum_{j\in J}\pi_{ij}^{+s}x_{ij}^{+s} + \sum_{i\in I}\sum_{j\in J}\pi_{ij}^{-s}x_{ij}^{-s}\right] \tag{33}$$

$$\sum_{i\in I} w_i y_i \le W \tag{34}$$

$$\sum_{j\in J} x_{ij} = y_i \quad i \in I \tag{35}$$

$$\sum_{i\in I} w_i y_i + \sum_{i\in I}\sum_{s\in S} w_i y_i^{+s} - \sum_{i\in I}\sum_{s\in S} w_i y_i^{-s} \le W \tag{36}$$

$$\sum_{j\in J} x_{ij}^{+s} = y_i^{+s} \quad i \in I, \quad s \in S \tag{37}$$

$$\sum_{j\in J} x_{ij}^{-s} = y_i^{-s} \quad i \in I, \quad s \in S \tag{38}$$

$$y_i^{+s} \le 1 - y_i \quad i \in I, \quad s \in S \tag{39}$$

$$y_i^{-s} \le y_i \quad i \in I, \quad s \in S \tag{40}$$

$$x_{ij}^{+s} = x_{ij}^{+s'} \quad i \in I, \quad j \in J, \quad s, s' \in S \tag{41}$$

$$x_{ij}^{-s} = x_{ij}^{-s'} \quad i \in I, \quad j \in J, \quad s, s' \in S \tag{42}$$

$$y_i \in \{0,1\} \quad i \in I \tag{43}$$

$$y_i^{+s} \in \{0,1\} \quad i \in I, \quad s \in S \tag{44}$$

$$y_i^{-s} \in \{0,1\} \quad i \in I, \quad s \in S \tag{45}$$

$$x_{ij} \ge 0 \quad i \in I, \quad j \in J \tag{46}$$

$$x_{ij}^{+s} \ge 0 \quad i \in I, \quad j \in J, \quad s \in S \tag{47}$$

$$x_{ij}^{-s} \ge 0 \quad i \in I, \quad j \in J, \quad s \in S. \tag{48}$$

The objective function (33) expresses the maximization of the total profit, given by the sum of the first stage profit plus the expected profit of the items handled in the second stage. Note that constraints (34) and (35) are the first stage constraints, while constraints (36)-(42) are the second stage ones. In particular, constraints (34) and (36) ensure that the capacity of the knapsack is not exceeded in first and second stages, respectively. Constraints (35) guarantee that any item is completely processed by some handlers only if it is loaded. Constraints (37) and (38) guarantee that if an item is loaded or unloaded in the second stage it is completely processed by some handlers. Constraints (39) establish that no item can be handled for loading in the second stage if it has already been loaded in the first stage. Similarly, constraints (40) establish that no item can be handled for unloading in the second stage if it has not been loaded in the first stage. Constraints (41) and (42) are the non-anticipativity constraints. Finally, constraints (43)-(45) and (46)-(48) are the integrality and the non-negativity constraints, respectively.

The optimal solutions of the two-stage model 2S-MHKP$_u$ and the deterministic approximation DA-MHKP$_u$ are strictly related. Let us suppose to have an optimal solution of model DA-MHKP$_u$. This model gives us a feasible approximation of the first-stage variables y_i, while it gives, by (29), a continuous relaxation of the assignment variables x_{ij}. Notice that, given the values of the variables x_{ij} in (29), one can derive a feasible first-stage solution of 2S-MHKP$_u$ by fixing to one, for each item with $y_i = 1$, any x_{ij} variable associated to it (for example, the one with the greatest value). This means that the information given by model DA-MHKP$_u$ is related to the first level only, while to obtain the possible recursion we need to force the first-level solution in the two-stage model.

6. Computational results

In this section, we present and analyze the results of the computational experiments. The first goal is to assess the behavior of the 2S-MHKP$_u$, the two-stage program with fixed recourse for the MHKP$_u$ proposed. The second is to evaluate the effectiveness of the deterministic approximation of the MHKP$_u$ we derived. Moreover, we want to calculate and evaluate the handling costs obtained by using our approximated results as first-stage decisions of the 2S-MHKP$_u$.

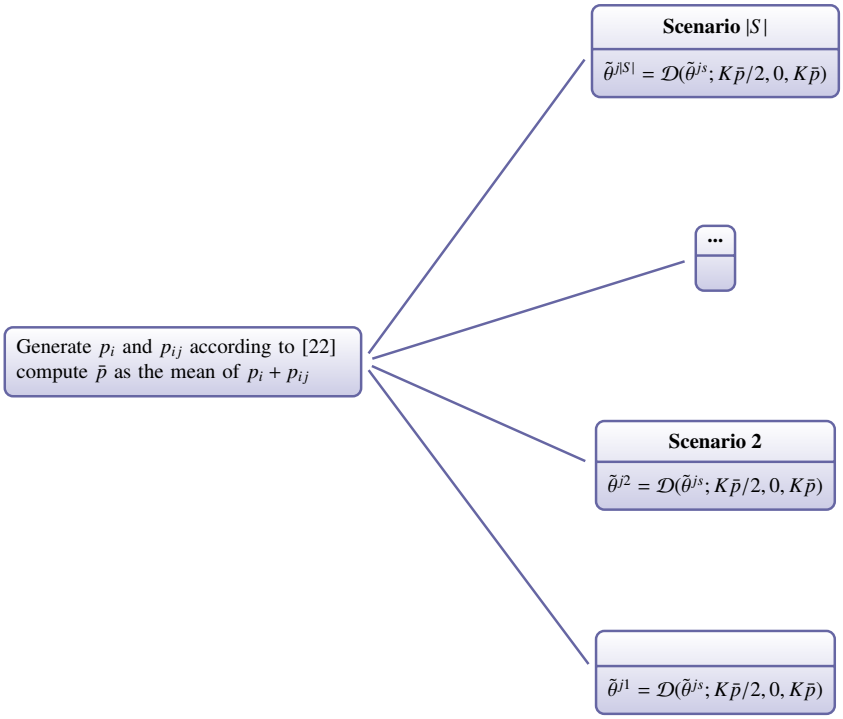

Figure 1: Scenario tree generation

The two-stage program with fixed recourse was implemented with `CPLEX 12.3`. Experiments were performed on a Intel i7 2 Ghz workstation with 6 GB of RAM.

Section 6.1 introduces the instance set. The calibration of the deterministic approximation of the MHKP$_u$ is described in Section 6.2, whilst the comparison between the two-stage and approximated solutions is given in Section 6.3. Finally, we study the impact of the approximated results on the two-stage program with fixed recourse in Section 6.4.

6.1. Instance set

No instances are present in the literature for this stochastic version of the knapsack problem. We then generated instances, partially based on those available for the deterministic knapsack problem [22].

Instances were created with the goal of providing the means to explore the impact of both the correlation between volume and profit of the items and the different probability distributions of the profit oscillations. Thus, the instances are characterized by various correlation strengths, as well as different probability distributions. Ten instances were randomly generated for each combination of the parameters.

- Number of items in the interval [100, 1000].
- Number of handlers in the interval [3, 5].
- Item volume uniformly distributed in the interval [1, R], where $R = 1000$.
- Deterministic item profits generated according to the following three rules

  UC: the deterministic item profits are uncorrelated to item volumes. They are uniformly generated in the interval [1, R] [5, 22].

  SC: the deterministic item profits are strongly correlated with the item volumes. The profit is defined as w_i+R/10, where w_i is the item volume [5, 22].

  PC: the deterministic item profits are proportionally correlated with the item volumes.The profit is defined as αw_i, where α is uniformly drawn from the interval [1, 5].

- Capacity of the knapsack was computed according to $\frac{h}{H+1}\sum_{i\in I} w_i$, where H is the number of instances for a set of parameters and $h \in [1, H]$ is the identification of an instance in that subset. This approach covers a large number of cases, diversifying the correlation between the parameters and the maximum capacity of the knapsack.

- Scenario generation. For each combination of the parameters described above, we first generate, for all the scenarios, the deterministic the item profits according to the above three rules. Given the average value of the deterministic profits, let $\bar{p}$, let $P_K = K\bar{p}$ be the maximum profit oscillation, where K belongs to the set {0.1, 0.3, 0.5}. The item profit oscillations were generated as $\tilde{\theta}^{j|S|} = \mathcal{D}(\tilde{\theta}^{js}; K\bar{p}/2, 0, K\bar{p})$, where $\mathcal{D}(\tilde{\theta}^{js}; \mu, min, max)$ is the distribution $\mathcal{D}$ with mean μ and truncated between the values *min* and *max* (see Figure 1). In our tests we used the Uniform and the Gumbel distributions.

Having solved the instances 10 times each and computed the standard deviation and the mean of the optima over the runs, we derived that the appropriate number of scenarios is 50. For each instance, this value ensures a maximum ratio between the standard deviation and the mean for the optima which is less than 1%.

The parameters were also chosen to reflect realistic cases of supply chain applications. In details, the different levels of correlation between item volumes and profits have the double effect to explore more challenging instances from the computation point of view [22] and explore price policies quite common in transportation [23]. The interval of item volumes and their link to the knapsack capacity is derived from Pisinger [22]. Finally, the bound of the stochastic oscillations has been set as in Tadei et al. [2] and reflects typical boundaries for profit oscillations in logistics.

6.2. Calibration of the β parameter

The deterministic approximation of the MHKP$_u$ given by (28) requires an appropriate value of the positive parameter β. This parameter describes the propensity of the model to choose among the set of the handlers characterized by different handling profits.

β is obtained by calibration as follows. Let us consider the standard Gumbel distribution $G(x) = \exp(e^{-x})$. If an approximation error of 2‰ is accepted, then $G(x) = 1 \Leftrightarrow x = 6.08$ and $G(x) = 0 \Leftrightarrow x = -1.76$. Let us consider the range $[m, M]$ ($[0, P_K]$ in our case) where the stochastic profit oscillations are drawn from. The following equations hold

$$\beta(m - \zeta) = -1.76 \tag{49}$$

$$\beta(M - \zeta) = 6.08 \tag{50}$$

where ζ is the mode of the Gumbel distribution $G(x) = \exp\left(e^{-\beta(x-\zeta)}\right)$. From (49) and (50) one gets the corresponding value of the parameter β

$$\beta = \frac{7.84}{M - m} = \frac{7.84}{P_K} = \frac{7.84}{K\bar{p}}. \tag{51}$$

More sophisticated methods to calibrate β can be found in [24].

6.3. Comparison of two-stage program and deterministic approximation results

The two-stage program solutions showed a common trend: the 2S-MHKP$_u$ reserves half of the total knapsack capacity in the second stage. For this reason, no items are unloaded in 99% of the instances. This means that the 2S-MHKP$_u$ uses the knowledge given by the scenarios in order to forecast what items can be immediately loaded, while preserving a proper loading space for items to be arranged in the recourse. This leads to a drastic reduction of the unloading and rearranging operations. Thus, the 2S-MHKP$_u$ policy is quite far from the usual supply chain approach of almost fully loading the knapsack in advance, while the unloading/rearranging operations are made at a later time and the percentage gap with the optimal solution can easily overcome 10%.

Here we summarize the results for all instances and different combinations of the parameters. The performance, in terms of optimality gap, is defined by the relative percentage error of the approximated solution when compared to the optimum. Moreover, we estimate the solution likelihood as the percentage of items loaded by the approximated solution which are also present in the optimal solution.

Note that the comparison results do not consider the number of handlers, which does not seem to affect the average performance of the deterministic approximation.

Table 1 reports the percentage optimality gap and the solution likelihood of the deterministic approximation for all combinations of the parameters, while varying the probability distribution (either Gumbel or Uniform). The first column displays the number of items, while Columns 2-3 and 5-6 report the mean and variance of the optimality gap, and Columns 4 and 7 show the mean percentage solution likelihood. The best mean values are obtained for the Gumbel distribution, that is very close to the optimum (0.12% for larger instances) and characterized by a negligible variance. Moreover, increasing the number of items gives better results for the deterministic approximation. Similarly, in terms of likelihood, this method guarantees results close to the optimum for both distributions.

Table 1: Optimality gap and solution likelihood of the deterministic approximation

IT	**GUMBEL**			**UNIFORM**		
	mean	**var**	**likelihood**	**mean**	**var**	**likelihood**
100	0.25	0.05	97.12	1.33	0.74	96.26
1000	0.12	0.01	98.54	1.33	0.52	97.31

Table 2 reports the percentage optimality gap and the solution likelihood performance of the deterministic approximation while varying the correlation between profits and volumes of items (Column 2) and the probability distribution (Columns 3-8). The results indicate that the strongly correlated instances (SC) yield the worst gaps, as well as the worst solution likelihood, with an average optimality gap of about 0.27% and 1.54% for the Gumbel and the Uniform distributions, respectively. For uncorrelated instances (UC) and the Gumbel distribution, some solutions of the deterministic approximation exactly match the two-stage program solutions.

Table 2: Optimality gap and solution likelihood for the profit correlation rules

IT	**PROFIT**	**GUMBEL**			**UNIFORM**		
		mean	**var**	**likelihood**	**mean**	**var**	**likelihood**
100	**UC**	0.23	0.05	98.78	1.20	0.57	98.21
	SC	0.27	0.06	95.28	1.54	0.93	95.19
	PC	0.25	0.05	97.31	1.26	0.68	95.19
1000	**UC**	0.10	0.01	99.31	1.14	0.38	98.92
	SC	0.15	0.02	98.06	1.56	0.55	96.40
	PC	0.11	0.01	98.23	1.29	0.55	96.62

The analysis of the impact of the maximum profit oscillations on the results accuracy is proposed in Table 3, considering different probability distributions (Columns 3-8). Recalling the definition of the maximum profit oscillation $P_K = K\bar{p}$, column 2 represents the percentage K of the mean profit $\bar{p}$ of the instances. The gap and the solution likelihood are clearly inversely proportional to the range of the oscillations. Indeed, the best mean values are obtained for $K = 0.1$.

Table 3: Optimality gap and solution likelihood for the maximum profit oscillations

IT	**K**	**GUMBEL**			**UNIFORM**		
		mean	**var**	**likelihood**	**mean**	**var**	**likelihood**
100	**0.1**	0.09	0.01	97.56	0.53	0.04	96.67
	0.3	0.30	0.06	97.02	1.33	0.27	96.54
	0.5	0.36	0.06	96.80	2.13	0.64	95.58
1000	**0.1**	0.04	0.00	98.81	0.55	0.04	97.88
	0.3	0.12	0.01	98.56	1.39	0.13	97.29
	0.5	0.30	0.01	98.23	2.09	0.26	96.76

In conclusion, the results are very promising. The procedure performs very well for all types of instances and distributions and guarantees a high accuracy. The best performance is obtained if the random profits have a Gumbel distribution, that is usually the case for real market oscillations. Moreover, the variance of the results is tight and in some cases close to zero. With respect to the solution likelihood, the mean values are all greater than 95% and increase according to the number of items.

As we expected, the mean optimality gap slightly increases for instances with Uniform distributed profit oscillations, but results are stable for each combination of the parameters and improve with respect to the number of items. Furthermore, it is interesting to note that the hardest subset of instances are the stronglly correlated ones. In fact, this kind of instance is characterized by a peculiar profit-volume correlation, as shown in [22].

From a computational point of view, the average CPU-times to solve to optimality the 2S-MHKP$_u$ and to compute the deterministic approximation are about 120 seconds and less than one second, respectively.

6.4. Usage of the approximated model as a decision tool

In the last part of this computational analysis, we analyze the losses, in terms of optimality gap, obtained by plugging the solution of the deterministic approximation of the MHKP$_u$ into the first-stage decision of the 2S-MHKP$_u$ [25]. In this way we can measure the accuracy of the approximated model when used not only as a method to calculate the optimum, but also as a decision tool to actually choose the items to be loaded. This means that the only degrees of freedom to maximize the objective function are the item-to-handler assignments and the handling operations in the second stage. Indeed, the effect of this strategy is the increase of the unloading operations, which does not exceed 6% of total operations, however.

Next, we present the comparison results organized as in Section 6.3. Tables 4, 5, and 6 summarize the average gap for all combinations of the parameters, for the profit correlations and for the maximum profit oscillations, respectively.

Table 4: Optimality gap with fixed first stage decision

IT	**GUMBEL**		**UNIFORM**	
	mean	**var**	**mean**	**var**
100	1.18	0.44	1.99	1.40
1000	1.22	0.48	1.99	1.15

Table 5: Optimality gap with fixed first stage decision for the profit correlation rules

IT	**PROFIT**	**GUMBEL**		**UNIFORM**	
		mean	**var**	**mean**	**var**
100	**UC**	1.05	0.31	1.77	1.06
	SC	1.36	0.59	2.31	1.74
	PC	1.12	0.38	1.88	1.28
1000	**UC**	1.01	0.28	1.75	0.88
	SC	1.49	0.64	2.32	1.25
	PC	1.15	0.42	1.91	1.18

Table 6: Optimality gap with fixed first stage decision for the maximum profit oscillations

IT	**K**	**GUMBEL**		**UNIFORM**	
		mean	**var**	**mean**	**var**
100	**10**	0.47	0.03	0.81	0.09
	30	1.24	0.10	2.01	0.48
	50	1.82	0.26	3.13	0.93
1000	**10**	0.47	0.02	0.82	0.08
	30	1.26	0.10	2.08	0.29
	50	1.92	0.26	3.07	0.53

As expected, the best performance is obtained by instances with the Gumbel distribution (Table 4). With respect to the profit correlations, the observed gap of SC instances is worse than other rules (Table 5). Finally, the gap increases according to the maximum range of the random profits (Table 6).

7. Conclusions

In this paper we have addressed the Multi-Handler Knapsack Problem under Uncertainty, which consists in finding a subset of items which maximizes the expected total profit, given by the sum of deterministic profits plus stochastic profit oscillations. One of the main features of this problem is that the profit oscillations are random variables with unknown probability distribution.

From a theoretical perspective, the paper shows that, under a mild assumption, the probability distribution of the maximum random profit for loading any item becomes a Gumbel distribution. Moreover, the total expected profit of the loaded items is proportional to the logarithm of the total accessibility of those items to the set of handlers and, at optimality, any item is handled by the set of handlers according to a multinomial Logit model.

The deterministic approximation of the stochastic model obtained provides very promising results on a large set of instances in negligible computing time.

In conclusion, the performance of the methodology proposed is particularly good when the probability distribution of the random profits of the stochastic model is a Gumbel distribution, even if good results are also provided by the Uniform distribution. This feature makes our deterministic approximation a good predictive tool for considering stochastic handling costs in supply chain problems.

[1] R. Tadei, N. Ricciardi, G. Perboli, The stochastic p-median problem with unknown cost probability distribution, Operations Research Letters 37 (2009) 135–141.

[2] R. Tadei, G. Perboli, M. M. Baldi, The capacitated transshipment location problem with stochastic handling costs at the facilities, International Transactions in Operational Research.

[3] G. Perboli, R. Tadei, M. Baldi, The stochastic generalized bin packing problem, Discrete Applied Mathematics 160 (2012) 1291–1297.

[4] T. Crainic, G. Perboli, R. Tadei, Recent advances in multi-dimensional packing problems, in: C. Volosencu (Ed.), New Technologies - Trends, Innovations and Research. ISBN: 978-953-51-0480-3, InTech, 2012, pp. 91–110.

[5] S. Martello, P. Toth, The 0-1 knapsack problem, in: N. Christofides, A. Mingozzi, P. Toth, C. Sandi (Eds.), Combinatorial Optimization, John Wiley & Sons, 1979, pp. 237–279.

[6] R. Tadei, G. Perboli, F. Della Croce, A heuristic algorithm for the auto-carrier transportation problem, Transportation Science 36 (1) (2002) 55–62.

[7] T. Crainic, N. Ricciardi, G. Storchi, Models for evaluating and planning city logistics systems, Transportation Science 43 (2009) 432–454.

[8] T. G. Crainic, S. Mancini, G. Perboli, R. Tadei, Two-echelon vehicle routing problem: A satellite location analysis, PROCEDIA - Social and Behavioral Sciences 2 (2010) 5944–5955.

[9] G. Perboli, R. Tadei, D. Vigo, The two-echelon capacitated vehicle routing problem: Models and math-based heuristics, Transportation Science forthcoming.

[10] J. Kleinberg, Y. Rabani, E. Tardos, Allocating bandwidth for bursty connections, SIAM Journal on Computing 30 (2000) 191–217.

[11] A. Goel, P. Indyk, Stochastic load balancing and related problems, in: 40th Annual Symposium on Foundations of Computer Science, 1999, pp. 579–586.

[12] Y. Merzifonluoglu, J. Geunes, H. E. Romeijn, The static stochastic knapsack problem with normally distributed item sizes, Mathematical Programming forthcoming.

[13] C. Barnhart, A. M. Cohn, The stochastic knapsack problem with random weights: A heuristic approach to robust transportation planning, in: Proceedings of the Triennial Symposium on Transportation Analysis, 1998, pp. 17–23.

[14] A. Lisser, R. Lopez, Stochastic quadratic knapsack with recourse, Electronic Notes in Discrete Mathematics 36 (2010) 97–104.

[15] M. Henig, Risk criteria in a stochastic knapsack problem, Operations Research 38 (1990) 820–825.

[16] E. Steinberg, M. Parks, A preference order dynamic program for a knapsack problem with stochastic rewards, Journal of the Operational Research Society 30 (1979) 141–147.

[17] M. M. Baldi, T. G. Crainic, G. Perboli, R. Tadei, The generalized bin packing problem, Transportation Research Part E 48 (2012) 1205–1220.

[18] E. J. Gumbel, Statistics of Extremes, Columbia University Press, 1958.

[19] W. Hansen, How accessibility shapes land use, Journal of the American Institute of Planners 25 (1959) 73–76.

[20] M. DeGroot, M. Schervish, Probability and Statistics (Third Edition), Addison Wesley, Reading, 2002.

[21] T. Domencich, D. McFadden, Urban travel dynamics: a behavioral analysis, North Holland, Amsterdam, 1975.

[22] D. Pisinger, Where are the hard knapsack problems?, Computers & Operations Research 32 (9) (2005) 2271?–2284.

[23] T. G. Crainic, G. Perboli, W. Rei, R. Tadei, Efficient lower bounds and heuristics for the variable cost and size bin packing problem, Computers & Operations Research 38 (2011) 1474–1482.

[24] J. Galambos, J. Lechner, E. Simiu, Extreme Value Theory and Applications, Kluwer, 1994.

[25] F. Maggioni, S. Wallace, Analyzing the quality of the expected value solution in stochastic programming, Annals of Operations Research 200 (2012) 37–54.