02435 - Decision-making under Uncertainty

Assignment 1

Reda LAHLOU s192431

02435 - Decision-making under Uncertainty
March 6, 2020

1 Part 1 - Modelling and duality

1.1 Task 1 - Modelling a mixed-integer linear program

1) Below is an explicit model formulation of the optimization problem considered in this exercise:

Variables

x_1	Space allocated to roses (m^2)
x_2	Space allocated to dahlias (m^2)
x_3	Space allocated to garden pinks (m^2)

$Max Z = 25x_1 + 30x_2 + 20x_3$	Objective function (maximization of the income from selling the flowers)
$s.t.\ \ x_1 + x_2 + x_3 \leq 5000$	Respect maximum space available in the nursery
$2.5x_1 + 5x_2 + 2x_3 \leq 20000$	Respect the maximum number of working hours in the year
$90x_1 + 100x_2 + 120x_3 \leq 700000$	Respect the maximum amount of water available in the year
$0.3x_1 + 0.5x_2 + 0.2x_3 \leq 2500$	Respect the maximum amount of fertilizer available in the year
$x_1, x_2, x_3 \geq 0$	Positive variables

2) The extended model can be formulated as follows (changes from previous model in red):

Variables

x_1	Space allocated to roses (m^2)
x_2	Space allocated to dahlias (m^2)
x_3	Space allocated to garden pinks (m^2)
y_1	$\begin{cases} 1 & \text{if } x_1 > 0 \\ 0 & \text{otherwise} \end{cases}$

Max Z $= 25x_1 + 30x_2 + 20x_3 - 20000y_1$
$s.t.\ \ x_1 + x_2 + x_3 \leq 5000$
$x_1 \leq 700y_1$
$2.5x_1 + 5x_2 + 2x_3 \leq 20000$
$90x_1 + 100x_2 + 120x_3 \leq 700000$
$0.3x_1 + 0.5x_2 + 0.2x_3 \leq 2500$
$x_1, x_2, x_3 \geq 0$
$y_1 \in \{0, 1\}$

The new binary variable y_1 indicates whether or not the nursery has decided to plant roses. A new cost of 20000 is added in the objective function if this decision is taken. The new constraint indicates that roses can only be planted in this greenhouse (if roses are planted, they only have access to the 700 m^2 inside the greenhouse). Once roses are planted, the rest of the space (including the hypothetical space left in the greenhouse) can be filled with the other flowers.

3) The extended model can be formulated as follows (changes from previous model in red):

Variables	x_1	Space allocated to roses (m^2)
	x_2	Space allocated to dahlias (m^2)
	x_3	Space allocated to garden pinks (m^2)
	y_1	$\begin{cases} 1 & \text{if } x_1 > 0 \\ 0 & \text{otherwise} \end{cases}$

Max Z $= 25x_1 + 30x_2 + 20x_3 - 20000y_1$
$s.t.\ \ x_1 + x_2 + x_3 \leq 5000$
$200y_1 \leq x_1 \leq 700y_1$
$2.5x_1 + 5x_2 + 2x_3 \leq 20000$
$90x_1 + 100x_2 + 120x_3 \leq 700000$
$0.3x_1 + 0.5x_2 + 0.2x_3 \leq 2500$
$x_1, x_2, x_3 \geq 0$
$y_1 \in \{0, 1\}$

The new constraint indicates that, if flowers are planted (i.e. if y_1 is equal to 1), then at least 200 m^2 of them has to be planted.

4) The extended model can be formulated as follows (changes from previous model in red):

Variables	x_1	Space allocated to roses (m^2)
	x_2	Space allocated to dahlias (m^2)
	x_3	Space allocated to garden pinks (m^2)
	y_1	$\begin{cases} 1 & \text{if } x_1 > 0 \\ 0 & \text{otherwise} \end{cases}$

Max Z $= 25x_1 + 30x_2 + 20x_3 - 20000y_1$
$s.t.\ \ x_1 + x_2 + x_3 \leq 5000$
$200y_1 \leq x_1 \leq 700y_1$
$x_2 \leq 5000 - 700y_1$
$2.5x_1 + 5x_2 + 2x_3 \leq 20000$
$90x_1 + 100x_2 + 120x_3 \leq 700000$
$0.3x_1 + 0.5x_2 + 0.2x_3 \leq 2500$
$x_1, x_2, x_3 \geq 0$
$y_1 \in \{0, 1\}$

The new constraint indicates that, if roses are planted (i.e. if y_1 is equal to 1), dahlias cannot be planted in the greenhouse. This means that they only have access to a space of 5000 m^2 (total space available) minus 700 m^2 (space inside the greenhouse).

1.2 Task 2 - Primal-dual-transformation

The dual LP of this LP can be formulated as follows:

Max $5y_1 + 11y_2 + 8y_3 + y_4$
$s.t.\ 2y_1 + 4y_2 - 3y_3 + 6y_4 \leq -5$
$-3y_1 - y_2 + 4y_3 - 5y_4 \leq 4$
$-y_1 + 2y_2 + 2y_3 + y_4 = -3$
$y_1, y_3 \leq 0$
$y_2 \geq 0$
$y_4 \in \mathbb{R}$

2 Part 2 - Stochastic programming

2.1 Task 3 - General questions

1) In a two-stage program, there are two types of variables: *first-stage variables* and *second-stage variables*.

- *First-stage variables* correspond to decisions which have to be taken before the uncertainty of the parameters is revealed. Their value is optimized by taking into account the probabilities of the different outcomes.
- *Second-stage variables* correspond to variables which values can be chosen after the uncertainty is revealed. Their optimal value is found by solving the linear program model which results from the outcome of the uncertainty (in this model, first-stage variables are set to the values which have been determined during the first stage).

2) In a linear program, it can happen that some of the data is uncertain, and that some variables have to be optimized before the uncertainty is revealed (no *wait-and see* approach possible). Unlike stochastic programming, deterministic programming does not allow to take into account the eventual outcome of the uncertainty. In a deterministic model, all variables would be optimized before the uncertainty is revealed (for example by taking the expected values of the uncertain parameters). This leads to a loss of information, since a stochastic model would have allowed to optimize late stage decisions depending on the revealed uncertainty.
In this situation, a stochastic model would give better results that a deterministic one (i.e. lower objective value in case of minimization, or higher objective value in case of maximization).

3) In a two-stage stochastic program, the ”most important” decisions are the ones made *here-and-now*. Indeed, the other decisions (*wait-and-see decisions*) can not be taken before the uncertainty is revealed: they will simply be optimized once the outcome is known. The results a decision-maker is the most interested in having when building a two-stage stochastic program model are the values of *here-and-now* variables, as they are the only ones he can have an influence on before the uncertainty is revealed - this is also where the probabilities of the different outcomes come into play.

2.2 Task 4 - Modelling a two-stage stochastic program

Below is a general formulation of the two-stage stochastic program for the furniture producer:

Sets

$\mathscr{T}$	$[\![1,10]\!]$	Planning horizon (10 years)
$\mathscr{F}$	$[\![1,6]\!]$	Type of furniture piece
$\mathscr{M}$	$[\![1,15]\!]$	Market
$\mathscr{I}$	$[\![1,5]\!]$	Location
$\mathscr{L}$	$[\![0,4]\!]$	Capacity level
$\mathscr{S}$	$[\![1,5]\!]$	Demand scenario

Parameters

$b_{m,f,t,s}$	Forecasted demand per market m, product f, time t and scenario s
$d^M_{i,m}$	Distance between location i and market m (in km)
c_i	Yearly operation costs of location i
$d^L_{i,j}$	Distance between locations i and j (in km)
D	Minimal distance between two opened locations (in km)
c^T	Cost of transporting one product unit from the facility to the market or between facilities (per km)
c^B_l	Building cost for level l
k^P_l	Production capacity for level l (per period of time)
k^S_l	Storage capacity for level l (per period of time)
c^T	Cost of transporting the furniture from the facility to the market (per km)
π_s	Probability of scenario s
Up	Big value used for conditional constraints (value set to 100

Variables

$y_{l_{i,l}}$	$[\![0,1]\!]$	1 if facility i is at level l, 0 otherwise	First-stage
$y_{fac_{i,m}}$	$[\![0,1]\!]$	1 if market m is supplied by facility i, 0 otherwise	First-stage
y_{open_i}	$[\![0,1]\!]$	1 if facility i is open, 0 otherwise	First-stage
$p_{i,f,t,s}$	$\mathbb{R}^+$	Quantity of product f produced by facility i at time t in scenario s	Second-stage
$mar_{i,m,f,t,s}$	$\mathbb{R}^+$	Quantity of product f sent by facility i to market m at time t in scenario s	Second-stage
$ex_{i,j,f,t,s}$	$\mathbb{R}^+$	Quantity of product f sent by facility i to facility j at time t in scenario s	Second-stage
$st_{i,f,t,s}$	$\mathbb{R}^+$	Quantity of product f found in the stock of facility i at the end of time t	Second-stage

$$\text{Min C} = \sum_{i\in\mathscr{I},l\in\mathscr{L}} c^B_l * y_{l_{i,l}} + \sum_{i\in\mathscr{I},t\in\mathscr{T}} c_i * y_{open_i} +$$

$$c^T * \sum_{i\in\mathscr{I},m\in\mathscr{M},f\in\mathscr{F},t\in\mathscr{T},s\in\mathscr{S}} \pi_s * mar_{i,m,f,t,s} * d^M_{i,m} + c^T * \sum_{i\in\mathscr{I},j\in\mathscr{I},f\in\mathscr{F},t\in\mathscr{T},s\in\mathscr{S}} \pi_s * ex_{i,j,f,t,s} * d^L_{i,j}$$

s.t.

$$b_{m,f,t,s} = \sum_{i \in \mathscr{I}} mar_{i,m,f,t,s} \qquad (1) \qquad \forall (\boldsymbol{m, f, t, s}) \in \mathscr{M} \times \mathscr{F} \times \mathscr{T} \times \mathscr{S}$$

$$\sum_{f \in \mathscr{F}} p_{i,f,t,s} \leq \sum_{l \in \mathscr{L}} k_l^P * y_{l_{i,l}} \qquad (2) \qquad \forall (\boldsymbol{i, f, t, s}) \in \mathscr{I} \times \mathscr{T} \times \mathscr{S}$$

$$\sum_{f \in \mathscr{F}} st_{i,f,t,s} \leq \sum_{l \in \mathscr{L}} k_l^S * y_{l_{i,l}} \qquad (3) \qquad \forall (\boldsymbol{i, f, t, s}) \in \mathscr{I} \times \mathscr{T} \times \mathscr{S}$$

$$y_{open_i} = 1 - y_{open_j} \text{ if } d_{i,j}^M < D \qquad (4) \qquad \forall (\boldsymbol{i, j}) \in \mathscr{I} \times \mathscr{I}, \boldsymbol{i \neq j}$$

$$\sum_{l \in \mathscr{L}} y_{l_{i,l}} = 1 \qquad (5) \qquad \forall (\boldsymbol{i}) \in \mathscr{I}$$

$$y_{l_{i,0}} = 1 - y_{openi} \qquad (6) \qquad \forall (\boldsymbol{i}) \in \mathscr{I}$$

$$\sum_{i \in \mathscr{I}} y_{fac_{i,m}} = 1 \qquad (7) \qquad \forall (\boldsymbol{m}) \in \mathscr{M}$$

$$mar_{i,m,f,t,s} \leq Up * y_{fac_{i,m}} \qquad (8) \qquad \forall (\boldsymbol{i, m, f, t, s}) \in \mathscr{I} \times \mathscr{M} \times \mathscr{F} \times \mathscr{T} \times \mathscr{S}$$

$$st_{i,f,t,s} = st_{i,f,t-1,s} + p_{i,f,t,s} - \Big(\sum_{j \in \mathscr{I}} (ex_{i,j,f,t,s} - ex_{j,i,f,t,s}) + \sum_{m \in \mathscr{M}} mar_{i,m,f,t,s} \Big) \qquad (9) \qquad \forall (\boldsymbol{i, f, t, s}) \in \mathscr{I} \times \mathscr{F} \times [\![\boldsymbol{2, 10}]\!] \times \mathscr{S}$$

$$st_{i,f,1,s} = p_{i,f,1,s} - \big(\textstyle\sum_{j \in \mathscr{I}} (ex_{i,j,f,1,s} - ex_{j,i,f,t,s}) + \sum_{m \in \mathscr{M}} mar_{i,m,f,1,s}\big) \qquad (10) \qquad \forall (\boldsymbol{i, f, s}) \in \mathscr{I} \times \mathscr{F} \times \mathscr{S}$$

- The objective function represents the total for the facilities. The cost c_i is applied every year only if the facility i is open. The cost c_l^B is applied if the facility has a level l. The rest of the equation relates to second-stage variables. Quantities supplied from facility i to market m represent a cost of c^T multiplied by the distance between i and m. The same reasoning is made for quantities supplied from one facility i to another facility j.
- **Constraint (1)** ensures that the demand from the markets are fulfilled.
- **Constraint (2)** ensures that a facility cannot produce more that its maximum capacity - which is determined by its level l.
- **Constraint (3)** ensures that the stock of a facility at the end of a time frame is no more than its storage capacity - which is determined by its level l.
- **Constraint (4)** ensures that 2 facilities i and j cannot be opened at the same time if the distance between them is lower than D.
- **Constraint (5)** ensures that each facility has exactly one level.

- **Constraint (6)** ensures that a facility with a level l of 0 is closed.
- **Constraint (7)** ensures that each market m is assigned to exactly 1 facility.
- **Constraint (8)** ensures that a facility i cannot supply a market m it is not assigned to.
- **Constraint (9)** is the storage balance constraint of facilities after the second time frame.
- **Constraint (10)** is the initial storage balance of facilities for the first time frame.

Before showing the results of the model, it should be noted that the original data provided in the Julia code for the following parameters have been modified: c_l^B, k_l^P and k_l^S. For each of these vectors, a value of 0 has been appended in the beginning to account for the level 0 (no cost, no production capacity and no storage capacity). Since the Julia language indexes its vector beginning from 1, the l index has been replaced by l+1 in the code in c_l^B, k_l^P and k_S^P - respectively found in the objective function, constraints (2) and constraint (3).
Furthermore, the model allows - in theory - to assign a closed facility to a market. However, this facility could not produce any furniture, nor store them. It could possibly import furniture from other facilities and export them back - either to a facility or a market it has been assigned to. In practice, this would never happen, as, within a time frame, a straight line between 2 facilities - or a facility and a market - is a shorter way than going through a closed facility in between.

Below are 2 results tables showing the values found by the model.

As shown on figure 1, 3 facilities are opened, with facility 5 supplying nearly half of the markets. Figure 2 confirms that the closed facilities (2 and 3) don't participate in any way in the model. It is interesting to notice the different roles played by the facilities. Facility 1 is rather isolated from the others: it hardly exports anything, and nearly all of its production goes to supply the markets it is assigned to. Facility 4 is only assigned to 2 markets, and a quite large part of its production is exported to facility 5, which needs to stock some furniture in order to supply all the markets it is assigned to.

```
Optimal solution found

Location: 1     open            4
Location: 2     not open        0
Location: 3     not open        0
Location: 4     open            3
Location: 5     open            4
Market : 1      5.0
Market : 2      5.0
Market : 3      5.0
Market : 4      1.0
Market : 5      5.0
Market : 6      1.0
Market : 7      4.0
Market : 8      5.0
Market : 9      1.0
Market : 10     4.0
Market : 11     5.0
Market : 12     1.0
Market : 13     5.0
Market : 14     1.0
Market : 15     1.0

Objective value: 173211.125
```

Figure 1: Main results, showing the state and the level of each location, as well as the facilities assigned to each market

```
Location: 1
                s1      s2      s3      s4      s5
Production:     72081.0 72041.0 74169.0 72160.0 70439.0
Max production: 7883.0  8000.0  8000.0  8000.0  7710.0
Supp market:    72081.0 72041.0 74169.0 72412.0 70439.0
Export:         0.0     0.0     0.0     0.0     0.0
Max stock:      0.0     574.0   658.0   232.0   0.0


Location: 2
                s1      s2      s3      s4      s5
Production:     0.0     0.0     0.0     0.0     0.0
Max production: 0.0     0.0     0.0     0.0     0.0
Supp market:    0.0     0.0     0.0     0.0     0.0
Export:         0.0     0.0     0.0     0.0     0.0
Max stock:      0.0     0.0     0.0     0.0     0.0


Location: 3
                s1      s2      s3      s4      s5
Production:     0.0     0.0     0.0     0.0     0.0
Max production: 0.0     0.0     0.0     0.0     0.0
Supp market:    0.0     0.0     0.0     0.0     0.0
Export:         0.0     0.0     0.0     0.0     0.0
Max stock:      0.0     0.0     0.0     0.0     0.0


Location: 4
                s1      s2      s3      s4      s5
Production:     32331.0 28193.0 26395.0 25281.0 30440.0
Max production: 4886.0  5000.0  3996.0  4843.0  5000.0
Supp market:    23407.0 21462.0 23390.0 21833.0 26198.0
Export:         8924.0  6731.0  3005.0  3448.0  4242.0
Max stock:      0.0     0.0     0.0     0.0     0.0


Location: 5
                s1      s2      s3      s4      s5
Production:     80000.0 80000.0 80000.0 80000.0 80000.0
Max production: 8000.0  8000.0  8000.0  8000.0  8000.0
Supp market:    88924.0 86731.0 83005.0 83196.0 84242.0
Export:         0.0     0.0     0.0     0.0     0.0
Max stock:      1683.0  2046.0  2208.0  2450.0  2610.0
```

Figure 2: Detailed results showing for each facility in each scenario its total production, exports to other facilities and supply to markets, as well as its maximum production and storage level over the time frames

2.3 Task 5 - Evaluation of stochastic programs

3 different approaches were used to solve the proposed optimization problem:

- A **wait-and-see** approach. The 3 scenarios are solved separately in a deterministic model. Taking the expected value of the objective values gives a value called WS.
- An **expected value solution**. The expected values of the uncertain parameters are calculated (in this case, the heat demand). These values are used in a deterministic model, giving a value for the objective value called EEV.
- A **stochastic solution**. The first-stage variables (the ones which have to be taken *here and now*) are optimized by taking into the probabilities of each of the scenarios. Second-stage variables are then optimized to give the best result if the scenario they correspond to happens. The value of the objective value is called RP.

For a minimization problem, these values are ordered as follows: $WS \leq RP \leq EEV$.

Table 1 shows the results of the different approaches for the heat optimization problem proposed.

	WS	EEV	RP
Values	52718	157451	66794

Table 1: Solutions of the different approaches

Measures of the information gap between the various solutions can be calculated (here for a minimization problem):

- $EVPI = RP - WS$ is the **exepected value of perfect information**.
- $VSS = EEV - RP$ is the **value of stochastic solution**.

	EVPI	VSS
Values	14076	90656

Table 2: Evaluation of information gap between the different approaches

The value for EVPI means that knowing by advance the outcome of the uncertainty would help saving 14076 euros compared to having to deal with uncertainty (and using a stochastic model). In other words, a decision-maker would be willing to pay up to 14076 euros to get by advance the eventual values of the uncertain parameters.
The value for VSS means that a decision-maker would save 90656 euros if he chose to use a stochastic model rather than using expected value for the uncertain parameters.