Assignment - Alternative A

Deadline: Tuesday, 26th July 2022, 23.59 CET

General information - READ CAREFULLY

- **Please choose between Alternative A or Alternative B**. You must submit only one report to either answering Alternative A or answering Alternative B!
- This assignment is part of the overall assessment of this course and, therefore, your answer counts for the final grade.
- This assignment must be solved and submitted as a group work before the above mentioned deadline closes.
- The assignment has to be submitted via DTU Learn using the according Assignment system. Use the entry `Assignments` in the course content and upload your files to the corresponding assignment. In case of technical problems with DTU Learn, please send your files to `dngk@dtu.dk` before the deadline.
- **The submission must consist of one pdf-document containing the answers to the tasks below. Furthermore, program code and scripts have to be uploaded as well**.
- You may use Julia or Python as the programming language to solve tasks. However, all input code is given in Julia only.
- **Name your report `Group<#>-Report-Alternative<A/B>.pdf`**
- **Name your codefiles `Group<#>-Task<#>.jl` (analog for python)**

Part 1 - Stochastic programming

In Task 1-4 you are optimizing the logistics operations of the biomass supplier *BioFuel*.

Task 1 - Scenario generation (10 points)

The file `demands_scengen.csv` contains the biomass demand (in thousand tons) for one customer for the last 20 years with 4 data points per year. Use one of the bootstrapping variants from the course for scenario generation (and reduction via k-medoid clustering, if necessary) and generate 5 representative scenarios for the coming 8 periods. Initialize the random number generator with `Random.seed!(1234)`.

1. Describe the steps of your scenario generation with short descriptions and plots. Report the probabilities of the 5 scenarios.
2. Briefly argue why the chosen method is an appropriate method for scenario generation in this case.

Task 2 - Two-stage Stochastic programming (35 points)

Currently, *BioFuel* has one central location (the depot) where they gather all biomass from the set of forest regions and distribute it to their customers which are power producers and district heating companies. However, *BioFuel* is thinking about taking more forest regions and customers on board. The setup would contain the customers $\mathcal{K}$ and forest regions $\mathcal{F}$. Therefore, they want to open additional regional locations that can take over the same tasks as the central location. The biomass supplier selected already a set of potential regional locations $\mathcal{L}$. Now they want use a stochastic program to decide the size of the locations (and indirectly also which locations to open at all), i.e., the handling capacity and the storage capacity that should be installed to minimize the overall costs.
For the optimization you should take the following requirements into account:

- The data for the next two years is divided in 8 seasons $\mathcal{T}$ to represent the demand and supply fluctuations over the year.
- Each forest region $f \in \mathcal{F}$ has a limited supply $e_{f,t,s}$ per season (given in thousand tons) and the supply is uncertain but defined on a set of possible scenarios $\mathcal{S}$ with given probabilities π_s.
- There is a loss when processing the incoming biomass at the locations, i.e., the efficiency of the locations η is less than 1.0 and given by parameter. The efficiency is the same for all locations. E.g. if the efficiency $\eta = 0.9$, only 90% of the incoming biomass can be used for customers.
- Each customer $k \in \mathcal{K}$ has an uncertain demand $d_{k,t,s}$ per season (given in thousand tons), but it is defined on a set of possible scenarios $\mathcal{S}$ with given probabilities π_s.
- One scenario s contains the information about supply and demand combined, i.e, there is no need to distinguish the scenarios.
- The biomass can be transported from forest regions to locations and from locations to customers (locations include the already opened central depot). This means transport directly from forest regions to customers and in between locations, customers or forest regions is not allowed.
- Each opened location has a limitation on the amount of biomass it can handle per season that is measured by the incoming biomass flow (in thousand tons). The company has to decide on this capacity when opening the location at the beginning of the planning horizon. There is an investment cost C_l^H (EUR/thousand ton) associated with each additional ton of handling capacity. If a location is opened, the minimum capacity $\underline{U}$ (in thousand tons) has to be installed. There is also a maximum capacity $\overline{U}$ (in thousand tons) that can be installed.
- To account for some flexibility, the company may plan for some flexible handling capacity per opened location and period (short-term hires) with a certain cost C^E per period (EUR per thousand tons) of capacity. The extra capacity can differ per period but has to be planned now. There is an upper limit of extra capacity per location and period (given in thousand tons) $\overline{U}^E$ (i.e. each location can have up to $\overline{U}^E$ extra capacity per period).
- The biomass can be stored from one season to the next at the opened locations. However, each location only has a limited capacity per season (given in thousand tons). The company has to decide on this capacity (with a given maximum $\overline{U}^S$) when opening the location at the beginning of the planning horizon. There is investment cost (EUR per thousand tons) C_l^S associated with each additional ton of storage capacity.
- The following costs should be minimized: costs of installing the capacity at a location (storage + handling), costs for extra handling capacity and operational handling costs of biomass (EUR per thousand tons, measured by inflow from forest regions to locations). The operational cost C_l^{OP} [EUR per thousand ton] have to be paid for normal handling and handling using the extra capacity.
- Be aware that it could be the case that not all customer demands can be fulfilled for all scenarios. Handle this appropriately in your model to still get a solution. The missing biomass is accounted with a penalty cost factor ϕ (EUR / thousand tons).
- The data contains initial storage I_c^S and handling capacity I_c^H for the central location (use index "c") which is fixed. Only extra handling capacity can be added for the central location. The storage at the central location also has some initial biomass left from the previous seasons.

In all following tasks, you have to define and describe all necessary sets, variables, parameters and constraints. The data is given in Julia file `TaskA2-data.jl` using the files `demands.csv` and `supplies.csv` as input. Do not use your own scenarios from Task 1.

1. Formulate a general two-stage stochastic program to help the biomass supplier. Define and describe all necessary sets, variables, parameters and constraints.
2. Solve your model with the provided data `TaskA2-data.jl`. Report the solution and present your observations regarding the results within the context of the planning problem.
3. Briefly answer: What are the most important decisions for the company in this planning problem (in terms of first-stage or second-stage) and why?

Task 3 - EVPI and VSS (12 points)

Evaluate your model from task 2 with respect to the Expected Value of Perfect Information (EVPI) and the Value of Stochastic Solution (VSS). Use the same data file as in Task 2.

1. Calculate EVPI and VSS and interpret the two values in your own words also with respect to the problem setting (what do the numbers mean). State also the expected values of the wait-and-see solution and expected value solution that you used for calculation.
2. Answer briefly: Why are wait-and-see solutions theoretical solutions and rather irrelevant for application in practice?

Task 4 - Out-of-sample test (10 points)

1. Evaluate the stochastic solution and expected value solution in an out-of-sample test using the samples in `demands_outofsample.csv` and `supplies_outofsample.csv`. There are some functions included in `TaskA2-data.jl` that read these csv files and give the you the specific demand/supply per period, customer/forest and sample.
2. Report your results in reasonable and concise manner including plots. Which method would you use in practice and why?
3. Why is important to run an out-of-sample test and not only rely on EVPI and VSS?

Part 2 - Robust optimization

In Task 5 you are optimizing the energy production of the distict heating company *GreenDH* using the biomass from *BioFuel*.

Task 5 - Robust optimization (25 points)

GreenDH is a customer of *BioFuel* and a district heating company that supplies heat for a small municipality. They need your help to optimize their heat production of their units $i \in I$ for one entire day (24 hours) $t \in T = \{1, \ldots, 24\}$ in order to maximize their profits from the electricity market given the electricity prices for each hour of the day (λ_t). The heat production plant consists of two combined heat and power (CHP) units. One of the units runs using natural gas ($i =$ G) and the other runs using biomass ($i =$ B). For each MWh of heat, the natural gas CHP unit produces 0.9 MWh of power ($\phi_G = 0.9$) and the biomass CHP unit produces 0.85 MWh ($\phi_B = 0.85$). The maximum heat production for each unit is 5 MWh ($Q_G = Q_B = 5$MWh) and the cost for producing one MWh of heat is 400 DKK for the natural gas CHP and 250 DKK for the biomass CHP ($C_G = 400$ and $C_B = 250$). Due to market requirements, the units must be scheduled one day in advance and the production of power (and therefore heat) must remain as scheduled. The main goal of *GreenDH* is to satisfy the heat demand (D_t, hard constraint) of the costumers at minimal cost. Excess production of heat in each hour can be cooled away free of charge. The data for the heat demand (D_t) and the electricity prices (λ_t) for 24 hours period are given. All data is given in `TaskA5-data.jl`.
In the following tasks provide a general formulation using the suggested notation in the task description. State the model formulations by describing all variables, sets, parameters and constraints. Besides the model formulations and solution values attach also the code files to your final submission. Although the model should be general, you can refer to the specific types of units by using the indices (B, G).

1. Each of the fuels has a different heating value (η_i) when producing heat by the CHP units. This means: Although we produce and pay for 10MWh heat, only $\eta_i 10$MWh heat are going to the district heating network to fulfill the demand. We know that the heating value of natural gas is almost 100% ($\eta_G = 1$). However, due to the characteristics of the fuel delivered by *BioFuel*, the heating value of the biomass is uncertain and varies in every new delivery. Based on observations, we know that the value varies between 65% and 90% ($\widetilde{\eta}_B \sim U[0.65, 0.9]$).

   Formulate a linear robust optimization (RO) problem that provides a feasible solution for every possible realization of the uncertainty. Start with the robust counterpart and describe the steps that you used to reformulate the robust counterpart to its linear formulation including definition of the uncertainty set. State also the final linear formulation.
2. Solve the model for the given data set in `TaskA5-data.jl` and state the objective value and heat production for each CHP unit and hour of the day (table or plot).

3. Extend your model from the previous task by the following requirements: Rasmus is the new manager of *GreenDH*. He doesn't like the idea of planning the heat production one day in advance without knowing the exact heating value of the biomass. He proposes to buy an electric peak boiler ($i = E$) that only produces heat. The boiler has the flat cost of 150 DKK per MWh having a heating value of 100% and a maximum capacity of 2 MWh ($C_E = 150$, $\eta_E = 1$ and $Q_E = 2$MWh). The production of the electric boiler doesn't need to be planned one day in advance, and it can be adapted to the realization of the uncertainty. You can use the index E of the boiler explicitly in the model.

   Formulate an adjustable robust optimization problem using linear decision rules for the problem setting described in this task. Define the linear decision rule and describe it. Start with the robust counterpart and describe the steps that you used to reformulate the robust counterpart to its linear formulation. State also the final linear formulation.

4. Solve the model for the given data set and state the objective value and heat production for each CHP unit and hour of the day (table or plot).

5. Compare the RO and ARO solutions and describe your observations.

Task 6 - Reformulation to robust linear models (8 points)

Consider the following basis model:

$$\begin{array}{lrcrcrcrl}
\text{Max} & 10x_1 & + & 20x_2 & + & 15x_3 & - & 17x_4 & \\
\text{s.t.} & 4x_1 & + & 3x_2 & + & 5x_3 & - & x_4 \leq & 13 \\
& 2x_1 & + & 5x_2 & - & 4x_3 & - & 2x_4 \geq & 5 \\
& & & & & & & \multicolumn{2}{r}{-3 \leq x_1 \leq 15} \\
& & & & & & & \multicolumn{2}{r}{-2 \leq x_2 \leq 10} \\
& & & & & & & \multicolumn{2}{r}{-10 \leq x_3 \leq 2} \\
& & & & & & & \multicolumn{2}{r}{2 \leq x_4 \leq 5}
\end{array}$$

In the following tasks some of the parameters are defined as uncertain parameters within given uncertainty sets. In each task you need to reformulate the model to a robust linear model that can be solved using a general purpose linear programming solver.
For each task you have to document all the steps of reformulation (writing down just the final robust linear model is not enough).
You do not need solve the models.

1. Parameters $\tilde{a}_2$, $\tilde{a}_3$ and $\tilde{a}_4$ are uncertain:

$$\begin{array}{lrcrcrcrl}
\text{Max} & 10x_1 & + & 20x_2 & + & 15x_3 & - & 17x_4 & \\
\text{s.t.} & 4x_1 & + & \tilde{a}_2x_2 & + & \tilde{a}_3x_3 & + & \tilde{a}_4x_4 \leq & 13 \\
& 2x_1 & + & 5x_2 & - & 4x_3 & - & 2x_4 \geq & 5 \\
& & & & & & & \multicolumn{2}{r}{-3 \leq x_1 \leq 15} \\
& & & & & & & \multicolumn{2}{r}{-2 \leq x_2 \leq 10} \\
& & & & & & & \multicolumn{2}{r}{-10 \leq x_3 \leq 2} \\
& & & & & & & \multicolumn{2}{r}{2 \leq x_4 \leq 5}
\end{array}$$

   $\tilde{a}_2$, $\tilde{a}_3$ and $\tilde{a}_4$ can take values in the uncertainty set $U = \Big\{ -4 \leq \tilde{a}_2 + \tilde{a}_3 + \tilde{a}_4 \leq 4, \tilde{a}_2 \leq 0, \tilde{a}_3 \leq 0, \tilde{a}_4 \geq 0 \Big\}$.

2. Parameters $\tilde{a}_2$ $\tilde{a}_3$ and $\tilde{a}_4$ are uncertain:

$$
\begin{array}{lrcrcrcrl}
\text{Max} & 10x_1 & + & 20x_2 & + & 15x_3 & - & 17x_4 & \\
\text{s.t.} & 4x_1 & + & \tilde{a}_2 x_2 & + & \tilde{a}_3 x_3 & + & \tilde{a}_4 x_4 \leq & 13 \\
 & 2x_1 & + & 5x_2 & - & 4x_3 & - & 2x_4 \geq & 5 \\
 & & & & & & & -3 \leq x_1 \leq 15 & \\
 & & & & & & & -2 \leq x_2 \leq 10 & \\
 & & & & & & & -10 \leq x_3 \leq 2 & \\
 & & & & & & & 2 \leq x_4 \leq 5 &
\end{array}
$$

$\tilde{a}_2, \tilde{a}_3$ and $\tilde{a}_4$ are uniform distributed with $\tilde{a}_2 \sim \mathcal{U}(1,5)$, $\tilde{a}_3 \sim \mathcal{U}(4,6)$, $\tilde{a}_4 \sim \mathcal{U}(-3,-1)$. At most two of the three parameters will deviate from their mean value.