OUU - Assignment

[redacted]
Wessel Boosman (589273)

December 3, 2021

Question 1

(1) *Specify the uncertainty structure you want to use in the formulation (e.g., explain if you want to use the worst-case or probabilistic setup, how and why). You can make assumptions if needed to justify your modeling choices.*

We choose a probabilistic set-up for our uncertainty constraints. This seems reasonable to us because the problem has, in addition to the decision variables which decide on the amount of items produced on forehand, the ability to relocate products between locations after the demand is known. An advantage of the ability to relocate products after the demand is known, is that you could possibly still satisfy the complete demand at all locations, even when production does not meet the demand at individual locations. If there are one or more locations which can compensate for this deficit, we can satisfy the demand at individual locations where the demand was initially higher than the number of items produced.

Therefore, a worst case approach to this problem does not make sense to us because this would result in solutions which are extremely conservative. The worst case solution would imply that the demand at all locations would be maximum, and all locations would produce therefore the maximum amount. The probability of this case is obviously extremely small and the solution is trivial.

A probabilistic approach to the problem makes much more sense, especially because of this relocation ability after the demand is known. Figure 1 shows the relation between the minimum probability for which the single chance constraints should hold and the percentage of feasible solutions. This is done in the following way, first we fix the probability for which the single chance constraints should hold and compute the optimal solution, that means we determine the optimal decision variables (production and relocation). Then, we fix the obtained optimal production plan and simulate 1000 realizations of demand at each location from an appropriate distribution. Later in this report we dive deeper into the specifications of this distribution. Finally, with the fixed production plan and the simulated demand we check the number of cases we obtain a feasible solution. That means, with the production fixed does

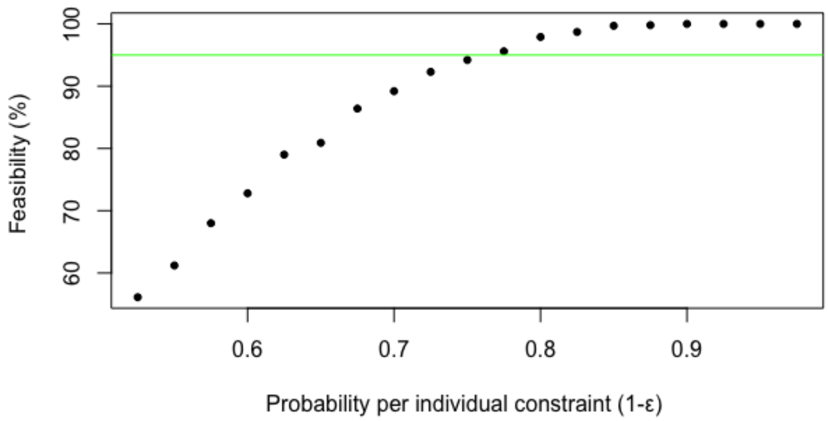

Figure 1: Relation between the minimum probability for which single chance constraints should hold and the feasibility of the problem.

there exist a relocation plan such that all demands can be satisfied. The vertical axis of figure 1 shows the percentage of feasible solutions out of the 1000 realizations, while the horizontal axis shows the minimum probability for which the single chance constraints should hold. We conclude from this empirically obtained figure that if each constraint holds with a probability of at least 77.5%, then the joint probability of all constraints hold approximately with 95.0%. Therefore, we conclude that if we make sure that the individual chance constraints hold with at least 77.5%, the total demand at each of the facilities will be satisfied in 95.0% of the cases which seems a very reasonable level.

(2) *Specify the objective you want to optimize (average cost, worst case cost, something else that sounds reasonable to you).*
We minimize the cost under the condition that the total demand is satisfied in 95.0% of the cases. As explained in the previous question this does not imply that the demand at each individual location should be satisfied by its own production with a probability of at least 95.0%. This approach would result in over conservative constraints, as we would not use the ability to relocate demand. We build our model such that it will result in a feasible solution in 95% of all cases. In case of an infeasible solution, we are not able to do any type of evaluation on the objective. We want to minimize the costs of feasible solutions, given that our model obtains a feasible solution in 95% of the cases.

(3) *Write a mathematical formulation of the uncertain problem you are going to solve. Explain/derive a solution approach for your 'uncertain' mathematical formulation (specify the uncertainty structure in such a way that the resulting problem is feasible).*
The general form of an optimization problem with chance constraints is given in equation 1. The disadvantage of this formulation is that the chance constraints incorporate uncertainty

which can not be easily implemented in common solvers.

$$
\begin{aligned}
&\min_{\mathbf{x}} \mathbf{a}^{0\top}\mathbf{x} \\
&\text{s.t. } \mathbb{P}\left(\left(\mathbf{a}^i + \mathbf{A}^i\mathbf{z}\right)^\top \mathbf{x} \geq b_i, \quad \forall i = 1, \ldots, m\right) \geq 1 - \epsilon
\end{aligned} \tag{1}
$$

Unfortunately, it is hard to evaluate joint probability constraints as in equation 1. Therefore, we rewrite our problem into single chance constraints. The uncertainty (in our case the demand at each location) is assumed to be normally distributed we can simplify it to equation 2. The advantage of reformulating the chance constraints in the form of equation 2 is that it does not contain uncertainty and is therefore easy to implement in common solvers.

$$
\begin{aligned}
&\mathbb{P}\left(\left(\mathbf{a}^i + \mathbf{A}^i\mathbf{z}\right)^\top \mathbf{x} \geq b_i\right) \geq 1 - \hat{\epsilon}, \quad \forall i = 1, \ldots, m, \quad \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\iff \quad &\underbrace{\mathrm{a}^{i\top}\mathbf{x} + \mathbf{x}^\top \mathbf{A}^i \mu}_{\text{mean}} + q_{1-\epsilon} \underbrace{\sqrt{\left(\mathbf{A}^{i\top}\mathbf{x}\right)^\top \boldsymbol{\Sigma}\mathbf{A}^{i\top}\mathbf{x}}}_{\text{standard deviation}} \geq b_i, \quad \forall i = 1, \ldots, m
\end{aligned} \tag{2}
$$

However, this only works of course when we assume that demand is normally distributed which still has to be verified based on historical (provided) data. Table 1 shows the p-values of the Shapiro-Wilk test based on historical demand of each location. In nine out of the ten locations, the p-value is above the 5% significance level and there is no evidence to reject that the demand distribution is normal. At location 5 this is not the case and, based on the Shapiro-Wilk test with 5% significance level, we would reject that demand is normally distributed. However, when we inspect the demand visually as in figure 2 we still assume the demand to be normally distributed due to its similarities. Moreover, if we compare the actual distribution with the normal distribution we can see that the actual demand is a bit skewed towards lower values. Assuming demand to be normally distributed results therefore in more robust solutions. In conclusion, we can safely simplify our problem by this assumption.

Table 1: Table showing the p-values of the Shapiro-Wilk test for each of the locations

Location	1	2	3	4	5	6	7	8	9	10
p-value	0.416	0.594	0.427	0.105	0.023	0.268	0.092	0.063	0.342	0.188

For our capacitated network inventory problem we want to reformulate our constraints into standard form and introduce vector $\mathbf{x}$ and coefficient vector $\mathbf{A}^{i\top}$. The multiplication of this variable and parameter will result in the left-hand side of the constraints.

$$
x^i - \sum_{j \neq i} y_{ij} + \sum_{j \neq i} y_{ji} \geq z_i, \quad \forall i = 1, \ldots, m, \quad \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3}
$$

First we are going to introduce how we constructed variable $\mathbf{x}$ and parameter vector $\mathbf{A}^{i\top}$. The vector $\mathbf{x}$ is defined in equation 4. Using this definition we can construct $\mathbf{A}^{i\top}$ such that, when multiplied with $\mathbf{x}$, we obtain the original left hand side of the chance constraint. The vector

$\mathbf{A}^i$ has the same length and dimensions as $\mathbf{x}$ and is constructed by the following recipe. We create a vector $\mathbf{A}^i$ for every location, where $i = 1, \dots, N$. We set the value on index i equal to 1. On every index corresponding to y_{ik}, $\forall k = 1, \dots, N, k \neq i$, we set the value -1. Lastly, for every index corresponding to y_{ki}, $\forall k = 1, \dots, N, k \neq i$, we set the value 1. All other indices get value 0.

$$\mathbf{x} = [x_1, x_2, \cdots, x_N, y_{11}, y_{12}, \cdots, y_{1N}, y_{21}, y_{22}, \cdots, y_{2N}, \cdots\cdots, y_{N1}, y_{N2}, \cdots, y_{NN}] \tag{4}$$

We have introduced our standard vectors and can now reformulate the problem as in equation 5. We use that the demand is normally distributed to obtain the final expression in which μ_i and σ_i are the mean and the standard deviation of demand at location i respectively. The final result is dependent on the vectors $\mathbf{A}^{i\top}$ and $\mathbf{x}$. Furthermore, there is no uncertainty left in the reformulation. Note that $q_{1-\epsilon}$ corresponds to the quantile of the normal distribution.

$$\begin{aligned}
& x^i - \sum_{j \neq i} y_{ij} + \sum_{j \neq i} y_{ji} \geq z_i, \quad \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\iff \quad & z_i \leq \mathbf{A}^{i\top}\mathbf{x} \\
\iff \quad & \frac{z_i - \mu_i}{\sigma_i} \leq \frac{\left(\mathbf{A}^{i\top}\mathbf{x}\right) - \mu_i}{\sigma_i} \\
\iff \quad & \underbrace{\mathbb{P}\left(\frac{z_i - \mu_i}{\sigma_i} \leq \frac{\left(\mathbf{A}^{i\top}\mathbf{x}\right) - \mu_i}{\sigma_i}\right)}_{\text{CDF of standard normal}} \geq 1 - \epsilon \\
\iff \quad & \frac{\left(\mathbf{A}^{i\top}\mathbf{x}\right) - \mu_i}{\sigma_i} \geq \Phi^{-1}(1 - \epsilon) \\
\iff \quad & \mathbf{A}^{i\top}\mathbf{x} \geq \sigma^i q_{1-\epsilon} + \mu^i
\end{aligned} \tag{5}$$

Last, we want to reformulate the objective and construct $\mathbf{a}^{0\top}$. This vector is easy to interpret and should, when multiplied with $\mathbf{x}$ return the total cost. The first N positions in this vector correspond to the production cost at each location and are in our case equal to 1. Then, each subsequent position should contain the transportation cost corresponding to the arc in $\mathbf{x}$. For example, at the index corresponding to y_{27} in our vector $\mathbf{x}$, we have our transportation cost in $\mathbf{a}^{0\top}$ of moving an item from location 2 to location 7.

In final form our optimisation problem can be easily solved by common solvers and is written as:

$$\begin{aligned}
\min_{\mathbf{x}} \quad & \mathbf{a}^{0\top}\mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}^{i\top}\mathbf{x} \geq \sigma^i q_{1-\epsilon} + \mu^i && \forall i = 1, \dots, n \\
& 0 \leq y_{ij} \leq b_{ij} && \forall i = 1, \dots, n, i \neq j = 1, \dots, n \\
& 0 \leq x_i \leq 20 && \forall i = 1, \dots, n
\end{aligned} \tag{6}$$

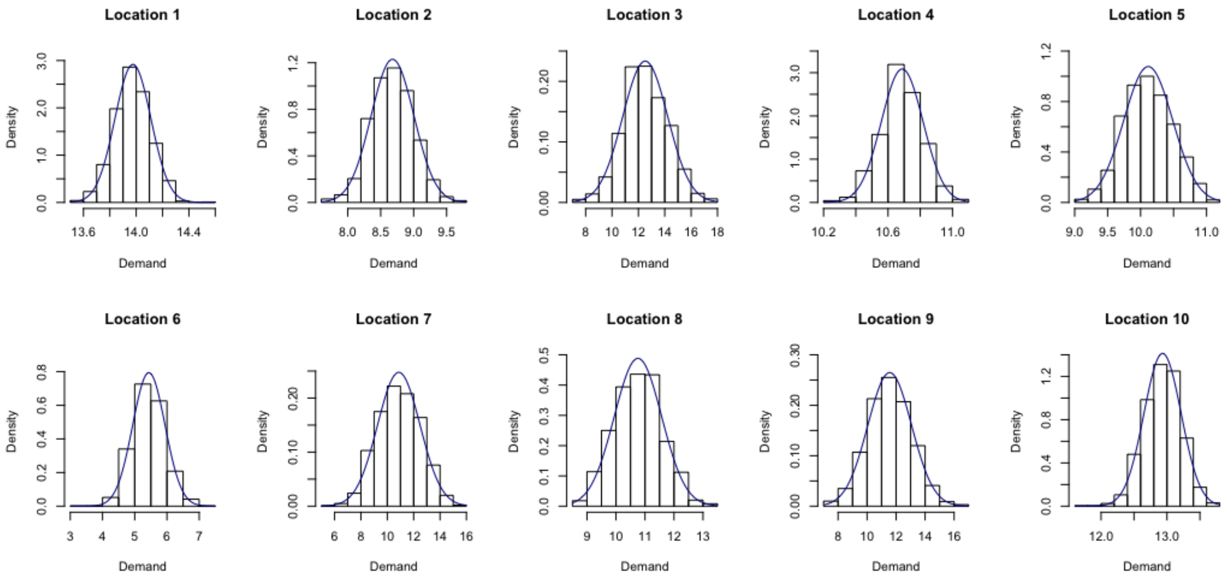

Figure 2: Histogram of demand at each location. Blue line shows the demand approximation with a normal distribution.

(4) *Implement your solution approach using a programming language of your choice.*
See attached code.

(5) *Evaluate the performance of the obtained solution out-of-sample. That is, simulate demand samples from a distribution that would be similar to the data sample but need not be exactly the data sample. Motivate your choice of a distribution for simulations. Describe what you observe out-of-sample about the cost and feasibility of the solutions you obtained.*

Demand at each location is assumed to be normally distributed as explained in question (3). Therefore, we draw samples from this distribution to evaluate the out-of-sample performance of our approach. We evaluated every location individually, that is, the demand at each location has an individual mean and standard deviation. In total we created 1000 samples to test the out-of-sample performance. Each sample consists of ten values for the demand at each corresponding location. We use that every location should be produce enough to satisfy its demand in 77.5% of the cases and we fix the x-variables corresponding to the production. So, the only variables to determine in this question are the y-variables and the corresponding transportation costs. Since the relocation variables are the only variables we optimize over, there are two possible scenarios. Either there exists a feasible solution and we are able to serve all demand (z) by our production (x) and, possibly by transporting between locations (y). Or, the demand can not be satisfied for each of the locations. This can be due to the capacity constraints on the arcs, or simply because the total demand is greater than the total production.

Our results show that in 44 out of 1000 samples, the model is able to find a solution. So, the feasibility level is 95.6%, which satisfies the feasibility level we wanted to achieve. The average cost in the out-of-sample setting is equal to 113.943, where the costs vary between 113.119 (low) and 119.398 (high). The low costs correspond to solutions where we do not need to transport any demand, as each location can serve its own demand. This conclusion can be drawn as the objective of solving our model to obtain x results in the same objective and all y variables are equal to zero in that solution. To conclude, we observe an average cost increase of 0.89% compared to the model in question (4), due to the additional transportation costs.

(6) *In the same out-of-sample setting, evaluate the 'nominal solution', i.e., the solution obtained by solving (1) with sample mean inserted as the value of z.*

The solution of this question shows us what happens if we just use the sample mean of z to construct the model. So, we ignore the possible variance of z when optimizing our x variables. We expect that this results in a lot of infeasibile solutions. The total production ($\sum_{i=1}^{10} x_i$) equals the sum of the average demand of all locations ($\sum_{i=1}^{10} \bar{z}_i$), where $\bar{z}$ is the average demand obtained from the initial instance we were provided with. To determine the out-of-sample performance, we draw random samples from the normal distribution again. The normal distribution is symmetric, and thus we are equally likely to draw values below or above the average. So, as these values are independently drawn, the probability that the total demand of a random sample is greater than the total amount we produce is 0.5. In this case, there can obviously never be a feasible solution. Furthermore, even if total production exceeds total demand, infeasibility can occur due to capacity constraints on the arcs. Therefore, we expect that at least 50% of the solutions are infeasible (for a large sample size).

After implementing this approach, we obtained the following results. In 519 out of 1000 samples, we could not find a feasible solution. The feasibility level is 48.1% confirming our expectation. The average objective of the feasible solutions is equal to 109.378, which is 4.0% lower compared to the average objective found in question (5). However, due to the low feasibility level, this result is distorted and we conclude that this is not a good approach to solve this problem.

(7) *In the same out-of-sample setting, evaluate the solution to the cost of the "perfect hindsight" solution. That is, for each out of sample z you solve the problem (1) separately, which gives you the lowest possible cost you could achieve if you knew z in advance.*

The solution to this question provides us the lowest possible cost, as we solve the model as if demand z is known in advance and we can base the production level x on this demand. As the production cost is equal for all locations and transportation costs are greater than 0, it is always cheaper to produce demand locally. This corresponds to an optimal solution with positive values for x and all y variables equal to zero. In our case specifically, the objective value exactly equals the demand. Again we determine the demand z by sampling from the normal distribution.

The results show us that the average objective is equal to 107.518. As expected, this value is smaller than the objectives we obtained in the previous questions. As we solve the model for every sample individually and can adjust our production amount x accordingly, this average objective is the lowest cost possible. We produce exactly enough to satisfy the (given) demand and we never obtain an infeasible solution. Note here that this conclusion can only be drawn because the (random) demand is truncated and can never exceed the production capacity of our locations.

(8) *Compare all three solutions, describe what you observe and explain your results.*

In conclusion we observe that the lowest costs are achieved when the demand at each location is known beforehand[1] (question (7)). When demand is unknown beforehand and we just produce the nominal demand at each location we observe a slight increase in costs. Using this approach results in a feasibility level slightly below 50%. Using a probabilistic approach to the problem such that the total demand is satisfied for approximately 95% of all cases raises the costs further. However, this comes at the benefit of feasibility. Therefore we conclude that there is a clear trade-off between cost and feasibility. In our case the trade-off favors the probabilistic approach as the the marginal cost of increased feasibility is relatively low. All results are summarized in table 2.

Table 2: Summarizing table of three approaches to the capacitated network inventory problem

	Optimal objective	**Increase w.r.t minimum (%)**	**Feasibility (%)**
Perfect hindsight	107.518	0.00%	100%
Out-of-sample	113.943	5.98%	95.6%
Nominal demand	109.378	1.73%	48.1%

[1]This is a hypothetical situation to determine a benchmark of the absolute minimum costs and we assume this situation does not occur in real-life