Optimization under uncertainty: modeling and solution methods

Paolo Brandimarte
Dipartimento di Scienze Matematiche
Politecnico di Torino

e-mail: paolo.brandimarte@polito.it
URL: http://staff.polito.it/paolo.brandimarte

Lecture 1: Optimization modeling under uncertainty
Dealing with uncertainty
Stochastic Linear Programming
Robust optimization
Multistage SP models with recourse

REFERENCES

Dealing with uncertainty
Stochastic Linear Programming
Robust optimization
Multistage SP models with recourse

Outline

1. Dealing with Uncertainty
2. Stochastic Linear Programming
3. Robust Optimization
4. Multistage SP models with recourse
We consider finite-dimensional mathematical programming problems like \( \min_{x \in S} f(x) \), where \( S \subseteq \mathbb{R}^n \).

There is a huge literature on the topic, and when \( f \) is a convex function and \( S \) is a convex set, the problem is (at least reasonably) solvable.

A quite lucky case is linear programming:

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

for which extremely efficient and robust solvers are commercially available.

In many practically relevant cases, however, problem data may be subject to uncertainty. This may be represented by considering data as random variables \( c(\omega), A(\omega), b(\omega) \).
Dealing with uncertainty makes life much harder:

- We have to find how to state the problem in a meaningful way.
- We should worry about solution feasibility as well as optimality.
- A safe bet is that the resulting model, whatever it is, will be much harder to solve.

So, why should we make our life a misery?
Can’t we just take the expected values of the uncertain data and solve a deterministic model?
Dealing with uncertainty
Stochastic Linear Programming
Robust optimization
Multistage SP models with recourse

Should we bother about uncertainty?
The newsvendor problem
Is sensitivity analysis of any use?
A glimpse of the real world: The Obermeyer case

THE NEWSVENDOR PROBLEM
The Newsvendor/Newsboy Problem

Prototype of problems with fashion items and perishability/obsolescence risk.

- We must decide the order quantity $q$ for an item before knowing the realization of random demand $D$ during the sales time window.
- Buy items at unit price $c$ and sell at full retail price $p > c$ within the sales time window.
- Loss of profit opportunity if $q < D$, but if $q > D$ mark down and sell leftover items at $p_u < c$.
- After a strike in 1899, newsboys obtained buyback contracts (a form of risk sharing).
The newsvendor problem

Does “maximizing profit” make sense?

- Not really, since profit is a random variable $\pi(q, D)$.
- When we choose $q$, we just select the probability distribution of profit.
- How can we rank distributions? The (seemingly) obvious way is to maximize expected profit (see Lecture 6 for alternatives):

$$\max_q \mathbb{E}_D [\pi(q, D)]$$

If demand distribution is continuous, the objective function is an integral depending on $q$.

- So, why don’t we just resort to solving the trivial problem

$$\max_q \pi(q, \mathbb{E}[D])$$
THE NEWSVENDOR PROBLEM: A TOY EXAMPLE

- Demand is discrete and uniformly distributed between 5 and 15: expected value is 10 and each outcome has probability 1/11
- \( c = 20, \ p = 25, \ p_u = 0 \) (no salvage value)

Profit is

\[
\begin{cases}
(p - c)q & \text{if } q \leq D \\
pD - cq & \text{if } q > D
\end{cases}
\]

Hence, expected profit is:

\[
EP = \frac{1}{11} \left[ \sum_{d=5}^{q} (pd - cq) + \sum_{d=q+1}^{15} (p - c)q \right]
\]
The newsvendor problem: a toy example

For instance, if we choose $q = 5$, profit is 25 in any demand scenario. If we choose $q = 6$, profit is 5 when demand is 5, 30 otherwise, resulting in the expected profit

$$\frac{5 + 30 \times 10}{11} = 27.73$$

To choose the best order quantity, let us tabulate the expected profit:

<table>
<thead>
<tr>
<th>$q$</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP</td>
<td>25.00</td>
<td>27.73</td>
<td>28.18</td>
<td>26.36</td>
<td>22.27</td>
<td>15.91</td>
</tr>
<tr>
<td>$q$</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>EP</td>
<td>7.27</td>
<td>-3.64</td>
<td>-16.82</td>
<td>-32.27</td>
<td>-50.00</td>
<td></td>
</tr>
</tbody>
</table>

We see that the optimal solution is not the expected value of demand (10), but a more conservative value (7). Why?
DEALING WITH UNCERTAINTY

Stochastic Linear Programming
Robust optimization
Multistage SP models with recourse

Should we bother about uncertainty?
The newsvendor problem
Is sensitivity analysis of any use?
A glimpse of the real world: The Obermeyer case

THE NEWSVENDOR PROBLEM

In the case of continuous distributions, it is easy to show that the optimal solution solves the equation:

\[ F_D(q^*) = \frac{m}{m + c_u}, \]

where \( F_D(x) \equiv P\{D \leq x\} \) is the cumulative distribution function (CDF) of demand, \( m \equiv p - c \) is profit margin, and \( c_u \equiv c - p_u \) is the cost of unsold items.

We see that the economics of the problem define a service level, i.e., the probability of satisfying the whole demand, and the optimal solution is the corresponding quantile.

If demand is normal, \( D \sim \mathcal{N}(\mu, \sigma^2) \), then \( q^* = \mu + z\sigma \), where \( z \) is the standard normal quantile corresponding to the above ratio. We see the different role of uncertainty, depending on the sign of the quantile \( z \).

Does this correspond to our personal experience?
The newsvendor problem: A variation

In basic Operations Research courses, the role of sensitivity analysis in Linear Programming is emphasized. Is this really useful to cope with uncertainty?

Let us consider an artificial newsvendor-like problem, borrowed from (King and Wallace, 2012).

The newsvendor sells three kinds of journal, featuring the data below:

<table>
<thead>
<tr>
<th></th>
<th>political</th>
<th>business</th>
<th>regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>quantity</td>
<td>$x_p$</td>
<td>$x_b$</td>
<td>$x_r$</td>
</tr>
<tr>
<td>demand</td>
<td>$D_p$</td>
<td>$1000 - D_p$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>profit margin</td>
<td>€1.30</td>
<td>€1.20</td>
<td>€1.00</td>
</tr>
</tbody>
</table>

Note: The newsvendor can order up to 1,000 journals, and within this capacity there is no limit to sales of the regional newspaper.
THE NEWSVENDOR PROBLEM: A VARIATION

Assuming deterministic demand $\hat{D}_p$, possibly a forecast, an LP model for the problem is

\[
\begin{align*}
\max & \quad 1.30x_p + 1.20x_b + 1.00x_r \\
\text{s.t.} & \quad x_p + x_b + x_r \leq 1000 \\
& \quad x_p \leq \hat{D}_p \\
& \quad x_b \leq 1000 - \hat{D}_p \\
& \quad x_p, x_b, x_r \geq 0
\end{align*}
\]

Clearly, the optimal solution is $x_p^* = \hat{D}_p$, $x_b^* = 1000 - \hat{D}_p$, $x_r^* = 0$.

The structure of this solution does not change if we change the forecast.

The above solution will almost never satisfy the constraints, when we plug the actually realized demand $D_p$!
The NewsVendor Problem: A Variation

- The above model formulation does not make any sense.
- We are confusing what we order and what we sell.
- These two decisions pertain to different decisional stages.
- Stages are separated by the acquisition of information: in this example, sales decisions are made after observing demand.
- What if customers may accept a substitute product? Let us assume that there are two types of customers, who prefer either the political or the business journal, respectively, but are willing to take the regional one in case of a stockout.
- In such a case, there is some value in ordering the regional journal as well, but sensitivity analysis cannot show that. Only a stochastic model can help.
- We need a modeling framework allowing for different decision stages, taking into account the flow of information.
A Glimpse of the Real World: The Obermeyer Case

- This is a real-life business case in fashion industry.¹
- Capacitated problem with multiple items and minimum lot sizes.
- Production in Far East and long time between start of design process (Feb year \( t \)) and retail period (Sep year \( t + 1 \)).
- High demand volatility and full scale production starts in Feb year \( t + 1 \).
- Flexibility allows two production runs, with early sales in between.
- It is possible to take advantage of correlation between early orders and overall demand.
- Use of newsvendor model and risk-based sequencing strategy.

¹ Problem described in HBS case 9-695-022, Sport Obermeyer Ltd.
Consider the “stochastic model”

\[
\begin{align*}
\text{min} & \quad c(\omega)^T x \\
\text{s.t.} & \quad A(\omega)x = b(\omega) \\
& \quad x \geq 0
\end{align*}
\]

Stated as such, the model is not even posed in a sensible way, and we cannot just take expected values. There are serious issues with feasibility. Can we require

\[
A(\omega)x = b(\omega) \quad \forall \omega \in \Omega?
\]

The case with inequality constraints is a bit easier. Requiring

\[
A(\omega)x \geq b(\omega) \quad \forall \omega \in \Omega
\]

may lead to a feasible solution; however, this fat solution may be overly expensive.
One possibility is to settle for a solution with a probabilistic guarantee:

\[ P \{ \mathbf{A}(\omega) \mathbf{x} \geq \mathbf{b}(\omega) \} \geq 1 - \alpha \]

for a suitably small \( \alpha \).

This idea leads to chance-constrained models, which have a sensible interpretation and may be useful. However, chance-constrained models have definite limitations:

- Per se, they do not model the flow of information and the sequence of decision stages.
- They can lead to quite risky solutions.
- Even if we take a very small \( \alpha \), can we trust our ability to estimate very small probabilities?
- In general, chance-constrained models are not convex (reason: the union of convex sets need not be convex).
We have seen the opportunity of distinguishing decision stages.

- The first-stage decisions $\mathbf{x} \geq 0$ must satisfy immediate constraints $A\mathbf{x} = \mathbf{b}$ and have immediate (first-stage) cost $\mathbf{c}^T\mathbf{x}$.

- At the second stage, a random event $\omega$ occurs, associated with random data. Given this information, a set of second-stage (recourse) actions $\mathbf{y}(\omega) \geq 0$ are taken.

- The second stage decisions are related to first-stage decisions by constraints $W\mathbf{y}(\omega) + T(\omega)\mathbf{x} = \mathbf{h}(\omega)$.

- The second stage decisions result in a cost $\mathbf{q}(\omega)^T\mathbf{y}(\omega)$.

- We want to minimize the sum of the first-stage cost and the expected value of second-stage cost.
We obtain the following optimization model:

$$\begin{align*}
\min & \quad c^T x + E_\omega \left[ q(\omega)^T y(\omega) \right] \\
\text{s.t.} & \quad Ax = b \\
& \quad Wy(\omega) + T(\omega)x = h(\omega) \\
& \quad x, y(\omega) \geq 0
\end{align*}$$

When the \textit{recourse matrix} \( W \) is deterministic, as above, we have a \textit{fixed recourse} problem. The more general case \( W(\omega) \) may present additional difficulties (see later...).
We may also introduce the recourse function $Q(x)$ and rewrite the model as the deterministic equivalent

$$
\begin{align*}
\min & \quad c^T x + Q(x) \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
$$

where

$$
Q(x) \equiv \mathbb{E}[Q(x, \xi(\omega))]
$$

and

$$
Q(x, \xi(\omega)) \equiv \min_y \left\{ q(\xi(\omega))^T y \mid Wy = h(\xi(\omega)) - T(\xi(\omega))x, y \geq 0 \right\}
$$

Here, we make the presence of random variables $\xi(\omega)$ explicit. This formulation shows that Stochastic Linear Programming is, in general, a nonlinear programming problem.
The recourse function $Q(x)$ looks like a “hopeless function”:

- It is an expectation, with respect to the joint distribution of $\xi(\omega)$; hence, it is a multidimensional integral, if random variables are continuous.
- It is a multidimensional integral of a function that we do not really know, as it is implicitly defined by an optimization problem.

Luckily, in many cases of practical interest, we can prove interesting properties of the recourse function, most notably convexity, which implies continuity (on an open domain).

In some cases, $Q(x)$ is differentiable; in other cases it is polyhedral.

This does not imply that it is easy to evaluate the recourse function, but we may resort to statistical sampling (scenario generation) and take advantage of both convexity and problem structure.
We may represent uncertainty by a discrete probability distribution, resulting the following scenario tree (fan), where $\omega_s$ is the event corresponding to scenario $s$.

First-stage decisions $\mathbf{x}$ must be made here and now, at the root of the tree; second-stage decisions $\mathbf{y}(\omega)$ are scenario-dependent.
A discrete probability distribution, where $S$ is the set of scenarios and $\pi_s$ is the probability of scenario $s \in S$, yields the following LP:

$$\min \ c^T x + \sum_{s \in S} \pi_s q_s^T y_s$$

s.t.  
$$Ax = b$$

$$W y_s + T_s x = h_s \quad \forall s \in S$$

$$x, y_s \geq 0$$

This is a plain LP, even though a possibly large-scale one. Decomposition methods may take advantage of the problem structure (see Lecture 3).
Assemble-to-Order (ATO) envirnments

Let us consider a (simplified) manufacturing management example.

When end items are available in a large number of configurations, it is not possible to keep them in inventory (make-to-stock), as such a strategy is too expensive and risky.

However, a pure make-to-order strategy may not be acceptable either, as it may result in large delivery lead times.

Therefore, when a huge number of configurations stems from assembly of a limited number of basic modules (features/options), an Assemble-to-Order (ATO) strategy may be pursued.

Modules are ordered under demand uncertainty, but assembly is carried out on order.

In practice, common components offer flexibility and help in hedging demand uncertainty.
ATO example: Technological data

The following stylized example is taken from (Brandimarte, 2011).

Consider three end items \((A_1, A_2, A_3)\), whose assembly requires five components/modules \((c_1, c_2, c_3, c_4, c_5)\).

We have a flat, two-level bill of materials:

<table>
<thead>
<tr>
<th></th>
<th>(c_1)</th>
<th>(c_2)</th>
<th>(c_3)</th>
<th>(c_4)</th>
<th>(c_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(A_2)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(A_3)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
ATO example: Technological data

Components require processing on three machining centers. Processing times and the associated component costs are as follows:

<table>
<thead>
<tr>
<th></th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>$c_2$</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>$c_3$</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>$c_4$</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>$c_5$</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>availability</td>
<td>800</td>
<td>700</td>
<td>600</td>
<td></td>
</tr>
</tbody>
</table>

Processing times and resource availability are given, e.g., in hours. We assume that assembly is never a bottleneck, and we also disregard assembly cost.
Say that there are three demand scenarios, associated with the same probability:

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>average</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>100</td>
<td>50</td>
<td>120</td>
<td>90</td>
<td>80</td>
</tr>
<tr>
<td>$A_2$</td>
<td>50</td>
<td>25</td>
<td>60</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>$A_3$</td>
<td>100</td>
<td>110</td>
<td>60</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

The rightmost column gives the selling price for each item.
ATO example: The expected value model

Let us write newsvendor-like models, where we maximize profit, assuming continuous decision variables for the sake of simplicity.

If we disregard uncertainty and consider the expected value of demand, we obtain the “expected value” (EV) model:

\[
\text{max } - \sum_{i=1}^{5} C_i x_i + \sum_{j=1}^{3} P_j y_j \\
\text{s.t. } \sum_{i=1}^{5} T_{im} x_i \leq L_m \quad m = 1, 2, 3 \\
y_j \leq \bar{d}_j \quad j = 1, 2, 3 \\
\sum_{j=1}^{3} G_{ij} y_j \leq x_i \quad i = 1, 2, 3, 4, 5 \\
x_i, y_j \geq 0
\]
ATO example: The expected value model

Solving the EV model yields the following EV solution:

Optimal Solution with Objective Value: 3233.3333
Make[c1] = 116.6667
Make[c2] = 116.6667
Make[c3] = 26.6667
Make[c4] = 0.0000
Make[c5] = 90.0000
Sell[a1] = 26.6667
Sell[a2] = 0.0000
Sell[a3] = 90.0000

Can you interpret the solution?

Is this a robust or a speculative solution?
ATO example: The stochastic solution

If we include uncertainty, assembly and sales decisions depend on scenarios, and we obtain a two-stage SLP with recourse:

\[
\begin{align*}
\text{max} & \quad -\sum_{i=1}^{5} C_i x_i + \sum_{s=1}^{3} \pi^s \left( \sum_{j=1}^{3} P_j y_j^s \right) \\
\text{s.t.} & \quad \sum_{i=1}^{5} T_{im} x_i \leq L_m \quad m = 1, 2, 3 \\
& \quad y_j^s \leq d_j^s \quad j = 1, 2, 3; \quad s = 1, 2, 3 \\
& \quad \sum_{j=1}^{3} G_{ij} y_j^s \leq x_i \quad i = 1, 2, 3, 4, 5; \quad s = 1, 2, 3 \\
& \quad x_i, y_j^s \geq 0
\end{align*}
\]
The SLP model yields the following solution:

Optimal Solution with Objective Value: 2885.7142
Make\[c1\] = 115.7143
Make\[c2\] = 115.7143
Make\[c3\] = 52.8571
Make\[c4\] = 2.8571
Make\[c5\] = 62.8571
Sell\[a1,1\] = 52.8571
Sell\[a1,2\] = 50.0000
Sell\[a1,3\] = 52.8571
Sell\[a2,1\] = 0.0000
Sell\[a2,2\] = 2.8571
Sell\[a2,3\] = 2.8571
Sell\[a3,1\] = 62.8571
Sell\[a3,2\] = 62.8571
Sell\[a3,3\] = 60.0000
ATO example: In-sample check

- The solution of the recourse problem (RP) is less extreme and speculative.
- The production of common components does not really change (risk pooling).
- Apparently, the RP solution has lower profit than the EV solution, 2885.7142 instead of 3233.3333, but this kind of comparison does not make sense.
- We should plug the EV solution into each scenario, and solve the corresponding second stage problem, and then compute the “true” value of the EV solution, which we may call Expected value of the EV solution (EEV).
ATO example: In-sample check

The check points out the fragility of the EV solution:

Optimal Solution with Objective Value: 2333.3339
Sell[a1,1] = 26.6667
Sell[a1,2] = 26.6667
Sell[a1,3] = 26.6667
Sell[a2,1] = 0.0000
Sell[a2,2] = 0.0000
Sell[a2,3] = 0.0000
Sell[a3,1] = 90.0000
Sell[a3,2] = 90.0000
Sell[a3,3] = 60.0000

The difference $2885.7142 - 2333.3339 = 552.3803$ is the Value of the Stochastic Solution (VSS).
ATO example: A few remarks

The example is quite easy to solve, but we should pay attention to a few issues:

- How can we represent demand uncertainty by a limited number of scenarios?
- We should run out-of-sample tests and check solution stability (see Lecture 5).
- What if we require integrality of decision variables?
- What if we introduce multiple time periods?
- What if we introduce product/component substitutability?
A GLIMPSE OF THE REAL WORLD: HEWLETT-PACKARD

- Despite its simplicity, the ATO example points out the use of common component as a tool to hedge against uncertainty.
- The risk pooling effect stems from aggregating sources of uncertainty and is used in many other settings.
- Hewlett-Packard is one example of firm that has adopted product design based on risk pooling effects (Design for Supply Chain Management).
- Does HP standardize toner cartridges for laser printers? Why?
- More generally, component standardization is a mantra of Japanese automotive industry.
- What is the downside of using a common platform for many car models?
VSS vs. EVPI

Let us define the individual scenario problem:

$$\min \ z(x, \xi(\omega)) \equiv c^T x + \min \{ q^T \xi(\omega) y \mid W y = h - T \xi(\omega) x, \ y \geq 0 \}$$

subject to:

$$Ax = b$$

$$x \geq 0$$

where $\xi(\omega)$ is a vector of random variables with expected value $\bar{\xi}$. Note that this scenario problem assumes knowledge of the future event $\omega$.

The recourse problem we have just considered amounts to solving

$$RP = \min_x E_{\xi}[z(x, \xi(\omega))].$$

Solving a deterministic problem, based on the expected value $\bar{\xi}$ of the data, corresponds to the expected value problem:

$$EV = \min_x z(x, \bar{\xi}),$$

which yields a solution $\bar{x}(\bar{\xi})$. 
As we have seen, the EV solution should be checked in the real context; this means that we should evaluate the expected cost of using the EV solution, which calls for some adjustments anyway:

\[ EEV = E_{\xi}[z(\bar{x}(\bar{\xi}), \xi(\omega))] \].

The VSS is defined as

\[ VSS = EEV - RP. \]

It can be shown that \( VSS \geq 0 \). A large VSS value suggests that solving the stochastic problem is well worth the effort; a small value suggests the opportunity to take the much simpler deterministic approach.

In decision theory, a related but different concept is often mentioned, the expected value of perfect information (EVPI)
VSS vs. EVPI

The EVPI is based on the solution of the wait-and-see (WS) problem:

$$WS = \mathbb{E}_\xi [\min_x z(x, \xi(\omega))].$$

Then we define

$$EVPI = WS - RP$$

Clearly, the VSS is more operational than the EVPI, but the two concepts measure different things:

- EVPI measures the maximum amount a decision maker would be willing to pay to obtain perfect information.
- VSS measures how much we lose by disregarding uncertainty.
A simple stochastic representation of uncertainty is not always feasible or viable, as the following complications may arise:

- We do not really know the probabilities of events that may occur (uncertainty about uncertainty).
- We do not really know all of the events that may occur (unk-unks and black swans).
- Uncertainty is not purely exogenous:
  - impact of stocking decisions on demand in supply chain management;
  - market impact of large trades on thin markets;
  - information asymmetries and herding behavior.
Robustness and Black Swans

Some authors draw the line between decisions under risk (e.g., dice throwing) and more radical (Knightian) uncertainty.

An alternative decision framework is robust optimization.
Actually, there are different robustness concepts (see Lecture 7):

- We may consider a stochastic programming problem under an uncertain probability measure $\mathbb{Q}$ within a set $\Pi$:

$$\min_{x \in X} \left\{ \max_{Q \in \Pi} \mathbb{E}_Q[f(x, \omega)] \right\} \quad (10)$$

- Alternatively, we may assume that uncertain parameters $\xi$ are within an uncertainty set $\Xi$ and optimize in the worst-case sense:

$$\min_{x \in X} \left\{ \max_{\xi \in \Xi} f(x, \xi) \right\} \quad (11)$$

- In other cases, robustness may be related to feasibility, rather than optimality of the solution.
Consider a pharmaceutical industry, using two raw materials of uncertain quality to produce two drugs.\(^2\)

<table>
<thead>
<tr>
<th>Parameter (for 1,000 boxes)</th>
<th>DrugI</th>
<th>DrugII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales price [$]</td>
<td>6,200</td>
<td>6,900</td>
</tr>
<tr>
<td>Active principle needed [g]</td>
<td>0.50</td>
<td>0.60</td>
</tr>
<tr>
<td>Labor [h]</td>
<td>90.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Machine capacity [h]</td>
<td>40.0</td>
<td>50.0</td>
</tr>
<tr>
<td>Production cost [$]</td>
<td>700</td>
<td>800</td>
</tr>
</tbody>
</table>

\(^2\)This numerical example is borrowed from Ben-Tal et al., 2009.
Uncertainty in raw material quality: worst-case robust approach

<table>
<thead>
<tr>
<th>Raw material</th>
<th>Cost [$/kg]</th>
<th>Active principle [g/kg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RawI</td>
<td>100.00</td>
<td>0.01</td>
</tr>
<tr>
<td>RawII</td>
<td>199.90</td>
<td>0.02</td>
</tr>
</tbody>
</table>

- Budget $100,000
- Labor available 2,000h
- Machine available 800h
- Inventory capacity 1,000kg
Uncertainty in raw material quality: worst-case robust approach

If we disregard uncertainty in quality, i.e., the amount of active principle obtained from each type of raw material, we solve a nominal model and obtain:

- RawI = 0, RawII = 438.789 [kg]
- DrugI = 17,552, DrugII = 0 [box]
- Profit = $8819.658

A more realistic model should account for uncertainty in the active principle content.

Let us assume that the uncertain parameter has a range: ±0.5% for RawI ([0.00995, 0.01005]) and ±2% for RawII ([0.0196, 0.0204]).
If we adopt the solution of the nominal model, and the parameter for RawII takes the worst-case value, actual production is reduced by 2% (17,201 boxes), with a corresponding reduction in profit: $6,888.987 (-21.89%).

We may pursue a conservative approach and solve the problem assuming the worst-case values, which yields:

- RawI = 877.732, RawII = 0 [kg]
- DrugI = 17,467, DrugII = 0 [box]
- Profit = $8294.567

Profit, according to the model, is reduced by 5.954%, which is much less than the potential loss from the adoption of solution suggested by the nominal model.
A FEW REMARKS

In this example:

- there is no account for corrective actions;
- a handful of values determine the solution.

This is typical of worst-case robust modeling approaches.

We may build hybrid models, merging stochastic and robust optimization.

Robust optimization often relies on recent developments in convex optimization, such as conic programming/duality and semidefinite programming (see Lecture 2).
The following example, borrowed from (Williams, 1999), illustrates very well one possible approach to guarantee robustness in terms of feasibility.

- We have a set of power generators to meet demand for electricity within a day.
- Each day has been partitioned in five time periods (12pm–6am, 6am–9am, 9am–3pm, 3pm–6pm, 6pm–12pm).
- For each time period, we have energy demand (MW) to satisfy.
Robust feasible solutions: Power generation

Three types of generators, indexed by $i$, are available, with the following features:

- number available, $a_i$;
- minimum and maximum power level (MW), $m_i$ and $M_i$;
- cost per hour at minimum, translated into $E_{it}$;
- cost per hour per MW above minimum, translated into $C_{it}$;
- setup cost, $F_i$.

We want to meet demand at minimum cost, accounting for the periodic character of the problem.

There is an additional requirement: We must be able to satisfy a 15% increase in demand without switching additional generators on.
The decision variables are:

- \( n_{it} \leq a_i \): the (integer) number of generators of type \( i \) active in period \( t \);
- \( s_{it} \): the (integer) number of generators of type \( i \) switched on at the beginning of period \( t \);
- \( x_{it} \): the (non-negative) output of generators of type \( i \) in period \( t \).

Question: Should we model each individual generator?

The objective function is:

\[
\min \sum_{i,t} C_{it} (x_{it} - m_i n_{it}) + \sum_{i,t} E_{it} n_{it} + \sum_{i,t} F_i s_{i,t}
\]
As to the constraints:

- Demand must be met:
  \[ \sum_i x_{it} \geq D_t, \quad \forall t \]

- Output is limited:
  \[ m_in_{it} \leq x_{it} \leq M_i n_{it}, \quad \forall i, t \]

- Link setup variables:
  \[ s_{it} \geq n_{it} - n_{i,t-1}, \quad \forall i, t = 2, 3, 4, 5 \]
  \[ s_{i1} \geq n_{i1} - n_{i5}, \quad \forall i \]

- Robustness requirement:
  \[ \sum_i M_i n_{it} \geq 1.15D_t, \quad \forall t \]
This is a simplified model (nonlinearities are disregarded) which may be easily solved.

It is important to notice that:

- The decision variables are *not* at the same level:
  - $n_{it}$ and $s_{it}$ are *design variables*;
  - $x_{it}$ are *control variables*.

- We are adopting a simple modeling framework:
  - no explicit second stage scenarios;
  - no cost for adjustment: emphasis is on robust feasibility.

- Clear tradeoff: We do not need to model uncertainty in a complicated way, and we get a computationally efficient model; however, the approach may not be always appropriate.
Multistage stochastic programming formulations arise naturally as a generalization of two-stage models.

- At the beginning of the first time period (at time $t = 0$) we select the decision vector $x_0$.
- This decision has a deterministic immediate cost $c_0^T x_0$ and must satisfy constraints
  $$A_{00} x_0 = b_0$$
- At the beginning of the second time period we observe random data $(A_{10}, A_{11}, c_1, b_1)$ depending on event $\omega_1$.
- Then, on the basis of this information, we make decision $x_1$.
- This second decision has an immediate cost $c_1^T x_1$ and must satisfy the constraint
  $$A_{10} x_0 + A_{11} x_1 = b_1$$
We repeat the same scheme for time periods up to $H - 1$, where $H$ is our planning horizon.

At the beginning of the last time period $H$, we observe random data $(A_{H,H-1}, A_{HH}, c_H, b_H)$ depending on event $\omega_H$.

Then, on the basis this information we make decision $x_H$, which has an immediate cost $c_H^T x_H$ and must satisfy the constraint

$$A_{H,H-1} x_{H-1} + A_{HH} x_H = b_H$$

From the point of view of time period $t = 0$, the decisions $x_1, \ldots, x_H$ are random variables, as they will be adapted to the realization of the underlying stochastic data process.

However, the only information we may use in making each decision consists on the history so far: no clairvoyance is allowed.
We end up with the following recursive formulation of the multistage problem:

\[
\begin{align*}
\min & \quad c_0^T x_0 + E \\
A_{00} x_0 &= b_0 \\
x_0 &\geq 0
\end{align*}
\]

\[
\begin{align*}
\min & \quad c_1^T x_1 \\
A_{10} x_0 + A_{11} x_1 &= b_1 \\
x_1 &\geq 0
\end{align*}
\]

\[
\begin{align*}
+ E \begin{array}{c}
\vdots \\
\end{array} + E \\
\begin{align*}
\min & \quad c_H^T x_H \\
A_{H,H-1} x_{H-1} + A_{HH} x_H &= b_H \\
x_H &\geq 0
\end{align*}
\end{align*}
\]

Here, decision \(x_t\) depends directly only on the previous decisions \(x_{t-1}\). In general, decisions may depend on all of the past history, leading to a slightly more complicated model. However, we may often introduce additional state variables, such that the above formulation applies.
The previous formulation is not quite precise, actually:

- We do not clearly see the role of conditioning with respect to random variables: the decision sequence is actually a stochastic process.

- Sometimes, a recursive formulation, quite similar to Dynamic Programming equations (see Lecture 8), is possible.

- The delicate point is the inter-stage dependence in terms of the stochastic data process $\xi_t$:
  - The easy case is when there is interstage independence.
  - The fairly easy case is when the process is Markovian, so only the link between $\xi_{t-1}$ and $\xi_t$ matters.
  - In the most general case the whole history up to time $t$, often denoted by $\xi_{[t]}$, must be taken into account.

For the moment, let us just try to understand multistage SP with a simple toy example.
We consider an example borrowed from (Birge and Louveaux, 2011)

- We are given an initial wealth $W_0 = 55,000$ that we may invest in stocks and bonds.
- Our aim is to generate enough wealth to pay for a liability $L = 80,000$ in $T = 3$ years.
- The portfolio is rebalanced at the beginning of each year, and there is no transaction cost.
- Asset returns are uncertain: in good years, the (total) return is 1.25 for stocks and 1.14 for bonds; in bad years, the (total) return is 1.06 for stocks and 1.12 for bonds.
- Our aim is to meet the liability and to keep some surplus if possible; however, we are risk-averse and do not want to end up with any shortfall.
One way to represent risk aversion is by a concave utility function (see Lecture 6). To keep it simple, we consider a piecewise linear utility function.

\[
U(W) = \begin{cases} 
0 & \text{when } W = L \\
q & \text{for any surplus} \\
q + r(W - L) & \text{for any shortfall}
\end{cases}
\]

The utility is zero when the terminal wealth \( W \) matches the liability \( L \) exactly.

The slope \( r = 4 \) penalizes the shortfall, and is larger than the reward \( q = 1 \) for any surplus.
Uncertainty is represented by the following scenario tree:

- **t=0**
  - **n₀**
- **t=1**
  - **n₁**
    - **n₂**
  - **n₃**
    - **n₄**
  - **n₅**
  - **n₆**
- **t=2**
  - **n₇**
  - **n₈**
  - **n₉**
  - **n₁₀**
  - **n₁₁**
  - **n₁₂**
  - **n₁₃**
- **t=3**
  - **n₁₄**
Each node $n_k$ corresponds to an event, where we should make some decision.

The initial node $n_0$ corresponds to the initial (here-and-now) asset allocation decision at time $t = 0$.

Then, for each event node, we have two branches, each labeled by a conditional probability of occurrence, $P\{n_k \mid n_i\}$, where $n_i = a(n_k)$ is the immediate predecessor of node $n_k$.

Each node of the tree, apart from $n_0$, is associated with the set of asset returns during the corresponding time period.

We have two nodes at time $t = 1$ and four at time $t = 2$, where we may rebalance our portfolio on the basis of the previous asset returns.

Finally, in the eight nodes corresponding to $t = 3$, we just compare our final wealth to the liability and we evaluate our utility function.
A scenario consists of an event sequence, i.e., a sequence of asset returns.

We have eight scenarios in the tree. For instance, scenario 2 consists of the node sequence \((n_0, n_1, n_3, n_8)\).

The probability of each scenario depends on the conditional probability of each node on its path.

In the example, each branch at each node is equiprobable, i.e., the conditional probability is 1/2. Therefore, each scenario in the figure has probability 1/8.

In practice, multistage scenario generation is a delicate and complicated task (especially for financial applications, as scenarios must be arbitrage-free).
The branching factor may be arbitrary in principle; the more branches we use, the better our ability to model uncertainty; unfortunately, the number of nodes grows exponentially with the number of stages, as well as the computational effort.

In practice, we are interested in the decisions that must be implemented here and now, i.e., those corresponding to the first node of the tree; the other (recourse) decision variables are instrumental to the aim of devising a robust plan, but they are not implemented in practice, as the multistage model is solved on a rolling horizon basis.

This suggests that, in order to model the uncertainty as accurately as possible with a limited computational effort, a possible idea is to branch many paths from the initial node, and less from the subsequent nodes.

There are two basic ways to build a multistage stochastic programming model: the split-variable and the compact model formulation.
In the split-variable approach we define the decision variables $x_{it}^s$, the amount invested in asset $i$ at the beginning of time period $t$ in scenario $s$.

$R_{it}^s$ is the (total) return of asset $i$ in scenario $s = 1, \ldots, S$ during time period $t$.

It is important to understand that, if we define the decision variables in this way, we must enforce the non-anticipativity constraint explicitly.

We have a set of decision variables for each node; however, the decision variables corresponding to different scenarios at the same time $t$ must be equal if the two scenarios are indistinguishable at time $t$ (non-anticipativity).
The issue may be understood by looking at the following figure, where vertical dotted lines correspond to non-anticipativity requirements:
The initial portfolio must be the same for all scenarios:

\[ x_{i0}^s = x_{i0}^{s'}, \quad i = 1, \ldots, l; \ s, s' = 1, \ldots, S. \]

Consider node \( n_1 \). Scenarios \( s = 1, 2, 3, 4 \) pass through this node and are indistinguishable at time \( t = 1 \). Hence, we must have:

\[ x_{i1}^1 = x_{i1}^2 = x_{i1}^3 = x_{i1}^4, \quad i = 1, \ldots, l. \]

Node \( n_1 \) corresponds to four nodes in the split tree.

By the same token, at time \( t = 2 \) we have constraints like

\[ x_{i2}^5 = x_{i2}^6, \quad i = 1, \ldots, l. \]
More generally, it is customary to denote by $\{s\}_t$ the set of scenarios which are not distinguishable from $s$ up to time $t$.

For instance:

\[ \{1\}_0 = \{1, 2, 3, 4, 5, 6, 7, 8\} \]
\[ \{2\}_1 = \{1, 2, 3, 4\} \]
\[ \{5\}_2 = \{5, 6\} \]

Then, the non-anticipativity constraints may be written as

\[ x_{it}^s = x_{it}^{s'} \quad \forall i, t, s, s' \in \{s\}_t. \]

This is not the only way of expressing the non-anticipativity requirement, and the best approach depends on the chosen solution algorithm.
ALM: Split-variable model formulation

\[
\begin{align*}
\text{max} & \quad \sum_{s} p^{s}(qw^{s}_{+} - rw^{s}_{-}) \\
\text{s.t.} & \quad \sum_{i} x_{i0}^{s} = W_{0} \quad \forall s \in S \\
& \quad \sum_{i} R_{i}^{s}x_{i,t-1}^{s} = \sum_{i} x_{it}^{s} \quad \forall s \in S; \ t = 1, \ldots, T \\
& \quad \sum_{i} R_{i,T+1}^{s}x_{iT}^{s} = L + w_{+}^{s} - w_{-}^{s} \quad \forall s \in S \\
& \quad x_{it}^{s} = x_{it}^{s'} \quad \forall i, t, s, s' \in \{s\}_{t} \\
& \quad x_{it}^{s}, w_{+}^{s}, w_{-}^{s} \geq 0.
\end{align*}
\]
\[ w^s_+ \text{ is the surplus at the end of the planning horizon, with reward } q, \text{ and } w^s_- \text{ is the shortfall, with penalty } r. \]

The objective function (12) is the expected value of the utility function; \( p^s \) is the probability of each scenario.

Equation (13) states that our initial wealth \( W_0 \) is allocated among the different assets.

The portfolio rebalancing constraints (14) say that the wealth at time \( t \) is reallocated.

In equation (15) we evaluate how we did, by comparing the final wealth with the liability \( L \), and setting the proper surplus and shortfall values.

Then we add non-anticipativity and non-negativity constraints.
ALM: Compact Model Formulation

Let us introduce the following notation:

- $N$ is the set of event nodes, in our case
  \[ N = \{ n_0, n_1, n_2, \ldots, n_{14} \}. \]

- Each node $n \in N$, apart from the root node $n_0$, has a unique direct predecessor node, denoted by $a(n)$: for instance, $a(n_3) = n_1$.

- There is a set $S \subset N$ of leaf (terminal) nodes, in our case
  \[ S = \{ n_7, \ldots, n_{14} \}; \]
  for each node $s \in S$ we have surplus and shortfall variables $w^s_+$ and $w^s_-$. 

- There is a set $T \subset N$ of intermediate nodes, where portfolio rebalancing may occur after the initial allocation in node $n_0$; in our case
  \[ T = \{ n_1, \ldots, n_6 \}; \]
  for each node $n \in \{ n_0 \} \cup T$ there is an investment variable $x_{in}$, corresponding to the amount invested in asset $i$ at node $n$. 

ALM: Compact Model Formulation

\[
\begin{align*}
\text{max} & \quad \sum_{s \in S} \rho^s (qw^s_+ - rw^s_-) \\
\text{s.t.} & \quad \sum_{i=1}^l x_{i,n_0} = W_0 \\
& \quad \sum_{i=1}^l R_{i,n} x_{i,a(n)} = \sum_{i=1}^l x_{in} \quad \forall n \in T \\
& \quad \sum_{i=1}^l R_{is} x_{i,a(s)} = L + w^s_+ - w^s_- \quad \forall s \in S \\
& \quad x_{in}, w^s_+, w^s_- \geq 0,
\end{align*}
\]
In the compact model formulation

- $R_{i,n}$ is the total return for asset $i$ during the period that leads to node $n$;
- $p^s$ is the probability of reaching the terminal node $s \in S$;
- this probability is the product of all the conditional probabilities on the path that leads from node $n_0$ to $s$.

The compact formulation uses less decision variables than the split-variable formulation. The choice between the two frameworks depends on the solution algorithm we want to apply.
In this toy example, there is really no difference between the two formulations, and we get the following plan:

<table>
<thead>
<tr>
<th>Node</th>
<th>Stocks</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0$</td>
<td>41.4793</td>
<td>13.5207</td>
</tr>
<tr>
<td>$n_1$</td>
<td>65.0946</td>
<td>2.16814</td>
</tr>
<tr>
<td>$n_2$</td>
<td>36.7432</td>
<td>22.368</td>
</tr>
<tr>
<td>$n_3$</td>
<td>83.8399</td>
<td>0</td>
</tr>
<tr>
<td>$n_4$</td>
<td>0</td>
<td>71.4286</td>
</tr>
<tr>
<td>$n_5$</td>
<td>0</td>
<td>71.4286</td>
</tr>
<tr>
<td>$n_6$</td>
<td>64</td>
<td>0</td>
</tr>
</tbody>
</table>
Note that in the last period the portfolio is not diversified, since the whole wealth is allocated to one asset, and we should wonder if this makes sense. Actually, it is a consequence of two features of this toy model:

- We are approximating a nonlinear utility function by a piecewise linear function, and this may imply “local” risk neutrality, so that we only care about expected return; we should use either a nonlinear programming model or a more accurate representation of utility with more linear pieces.

- The scenario tree has a very low branching factor, and this does not represent uncertainty accurately.

However, the portfolio allocation in the last time period is not necessarily a critical output of the model: the real stuff is the initial portfolio allocation.

As we pointed out, the decision variables for future stages have the purpose of avoiding a myopic policy, but they are not meant to be implemented.
A STOCHASTIC PRODUCTION PLANNING MODEL

The following model is reported exactly as stated in the original paper: M. Akif Bakir, M.D. Byrne. Stochastic linear optimization of an MPMP production planning model. *Int. J. of Production Economics*, 55(1998)87-96.

\[
\begin{align*}
\text{max} & \quad \sum_{t=1}^{T} \left[ \sum_{i=1}^{N} c_i x_{it} \right] - E_\xi \left[ \sum_{t=1}^{T} \sum_{i=1}^{N} Q(x, \xi^L) \right] \\
\text{s.t.} & \quad \sum_{i=1}^{N} a_{ik} x_{it} \leq MC_{kt} \\
& \quad x_{it} \geq 0 \\
\text{min} & \quad (q_{it})^T l_{it} \\
\text{s.t.} & \quad WI_{it} = Ax_{it} + WI_{it-1} - d_{it} \\
& \quad l_{it}^- = l_{it}^+ - d_{it} \\
& \quad l_{it} \geq 0 
\end{align*}
\]
A stochastic production planning model

Questions:
- Is the model two- or multi-stage?
- What is the shape of the scenario tree? Is it the same as in the ALM model?
- Any other comment?

Key message
Stages are related to the flow of information, and should *not* be confused with time periods.
A GLIMPSE OF THE REAL WORLD: NOKIA VS. ERICSSON

Philips Semiconductor had a manufacturing plant in Albuquerque, producing chips for Nokia and Ericsson cell phones.

In March 2000, a lightning stroke and in the ensuing fire, all of the available inventory was destroyed or contaminated, blocking production for months.

Nokia immediately noticed a slowdown in supply, investigated the causes, and reacted promptly: 4 out of the 5 chips supplied by Philips were replaced by adapting the design, and the remaining one was produced by Philips at plants in China and The Netherlands.

The Ericsson management noticed the problem after 4 weeks: this resulted in $400 million of lost sales.

According to some reports, because of this and other issues, the Ericsson cell phones division lost $1.68 billion, practically going out of market.
A rigid, one-to-one association between products and plant is not quite robust, but total flexibility is expensive and difficult to manage.

When designing a supply chain (or a telecommunication network), we should look for a suitable compromise accounting for uncertainty.
PLANT LOCATION MODEL

In the classical plant location model we have a bipartite network, consisting of a set $S$ of potential source nodes and a set $D$ of demand nodes:

- for each $i \in S$ we have a fixed opening cost $f_i$ and a capacity level $R_i$;
- for each $j \in D$ we have a demand $d_j$ (we assume a single product type for simplicity);
- for each arc $(i, j)$ in the network we have a unit transportation cost $c_{ij}$.

The aim is satisfying demand at minimum cost.

Note that we are not including time, so the above cost should refer to a selected time bucket.
Plant Location Model

Let us define the decision variables:

\[ y_i = \begin{cases} 
1 & \text{if source node } i \text{ is opened} \\
0 & \text{otherwise}
\end{cases} \]

and \( x_{ij} \geq 0 \), the amount of flow from source node \( i \) to destination node \( j \).

\[
\min \sum_{i \in S} f_i y_i + \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij}
\]

s.t. \( \sum_{i \in S} x_{ij} \geq d_j, \quad \forall j \in D \)

\( \sum_{j \in D} x_{ij} \leq R_i y_i, \quad \forall i \in S \)

\( x_{ij} \geq 0, \quad y_i \in \{0, 1\} \)
The classical plant location model disregards:
- nonlinearities in transportation costs;
- uncertainty in demand and capacity.

Let us introduce demand uncertainty and a two-stage decision structure:
- $d_j^s$ is the demand at retail store $j$ under scenario $s$;
- $y_i$ is as before, a design variable (first-stage);
- $x_{ij}^s$ is now a scenario-dependent control variable (second-stage).
Let us consider the following two-stage MILP model with recourse:

\[
\begin{align*}
\min & \quad \sum_{i \in S} f_i y_i + \sum_s \pi^s \left( \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij}^s \right), \\
s.t. & \quad \sum_{i \in S} x_{ij}^s \geq d_j^s \quad \forall s, \forall j \in D, \\
& \quad \sum_{j \in D} x_{ij}^s \leq R_i y_i \quad \forall s, \forall i \in S, \\
& \quad x_{ij}^s \geq 0, \quad y_i \in \{0, 1\}.
\end{align*}
\]

Is anything wrong with this model?

What if an extreme scenario with large demand is included?
Sometimes, we should resort to *elastic* model formulations allowing for suitably penalized constraint violations.

Let $z_j^s \geq 0$ be the amount of unmet demand at node $j$ under scenario $s$; these decision variables are included in the objective function multiplied by a penalty coefficient $\beta_j$:

$$\min \sum_{i \in S} f_i y_i + \sum_s \pi^s \left( \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij}^s \right) + \sum_s \pi^s \left( \sum_{j \in D} \beta_j z_j^s \right),$$

s.t. $$\sum_{i \in S} x_{ij}^s + z_j^s = d_j^s \quad \forall s, \forall j \in D,$$

$$\sum_{j \in D} x_{ij}^s \leq R_i y_i \quad \forall s, \forall i \in S,$$

$$x_{ij}^s, z_j^s \geq 0, \ y_i \in \{0, 1\}.$$
PLANT LOCATION MODEL

- Is the plant location model a single or a multiple period model?
- What if we want to account for intertemporal dependence in demand?
- What if we want to include possible demand trends?
- What if we want to include inventory variables?
- How can we account for nonlinear costs? [See Lecture 2 on convexity]
HOMEWORK

For Ph.D. students who formally need credits:

1. Reformulate problem (1) as a two-stage stochastic LP with recourse. What is the form of the recourse matrix?

2. Using your favorite LP solver (possibly Excel), evaluate the VSS for the ATO example when we force integrality of the decision variables. How much do we gain by postponing assembly decisions?