Stochastic programming

Lorenzo Castelli, Università degli Studi di Trieste.

UNIVERSITÀ
DEGLI STUDI DI TRIESTE

What is Stochastic Programming?

- Since the real world is uncertain, it becomes imperative to consider uncertainty in decision-making.
- "*...but in the world nothing can be said to be certain except death and taxes.*" - Ben Franklin.
- Stochastic programming (SP) is a field of optimisation that deals with decision-making problems involving data uncertainties.

SP has several applications:

- Manufacturing (e.g. Supply chain planning).
- Transportation (e.g. Airline scheduling).
- Telecommunications (e.g. Network design).
- Electricity power generation (e.g. Power adequacy planning).
- Health care (e.g. Patient & resource scheduling).
- Agriculture (e.g. Farm planning under weather uncertainty)
- Forestry (e.g. Wildfire emergency response planning).
- Finance (e.g. Portfolio optimisation).
- etc.

Sources of Uncertainty

Example sources of uncertainty:

- Market (product, stocks) related.
- Financial related.
- Technology related.
- Competition related.
- Weather related (e.g. airline rescheduling).
- Catastrophic events (accidents, war, 9/11, COVID-19, etc.).

Types of Uncertainty

- Exogenous uncertainty – ‘external’ uncertainty, is not influenced by the SP decisions.
  - Weather related uncertainty: airline scheduling decisions.
- Endogenous uncertainty – ‘internal’ uncertainty, is influenced by the SP decisions.
  - e.g. Wildfire behaviour: fire suppression planning decisions.

Does Uncertainty Matter?

- As a general rule, stochastic programs are more difficult to formulate and solve than deterministic mathematical programs.
- Given the availability of post-optimality analysis, it can be tempting to ease the process by relying on sensitivity analysis to investigate the impact of uncertainty.
- In many cases, sensitivity analysis is not an appropriate tool for this investigation.

Dakota furniture

The Dakota Furniture Company manufactures desks, tables, and chairs. The manufacture of each type of furniture requires lumber and two types of skilled labour: finishing and carpentry. The cost of each resource and the amount needed to make each type of furniture is given in the table below.

Resource	Cost (€)	Desk	Table	Chair
Lumber ($\boldsymbol{m}^3$)	2	8	6	1
Finishing (hrs.)	4	4	2	1.5
Carpentry (hrs.)	5.2	2	1.5	0.5
Demand		150	125	300
Selling price (€)		60	40	10

In order to maximise total profit, how much of each item should be produced, and what are the corresponding resource requirements?

Deterministic linear programming

Decision variables

x_l: Number of m^3 of lumber acquired.

x_f: Number of labour hours acquired for finishing.

x_c: Number of labours hours acquired for carpentry.

y_d: Number of desks produced.

y_t: Number of tables produced.

y_c: Number of chairs produced.

Formulation

$$\max 60y_d + 40y_t + 10y_c - 2x_l - 4x_f - 5.2x_c$$

$$y_d \leq 150$$

$$y_t \leq 125$$

$$y_c \leq 300$$

$$8y_d + 6y_t + y_c - x_l \leq 0$$

$$4y_d + 2y_t + 1.5y_c - x_f \leq 0$$

$$2y_d + 1.5y_t + 0.5y_c - x_c \leq 0$$

$$y_d, y_t, y_c, x_b, x_f, x_c \geq 0.$$

Solution

Objective value = €**4165**.

x_l	Number of m^3 of lumber acquired	1950
x_f	Number of labour hours acquired for finishing	850
x_c	Number of labours hours acquired for carpentry	487.5
y_d	Number of desks produced	150
y_t	Number of tables produced	125
y_c	Number of chairs produced	0

Introducing uncertainty

- Even though the basic data might change, the structure of the problem remains the same.
- For example, if the selling price for a chair is increased from €10 to €11, what happens?
- This is the observation that gives rise to the post-optimality investigation of the solution known as 'sensitivity analysis.'
- In this case, if the selling price of a chair increased to €11, the production of chairs would become profitable, and the entire demand (300) is produced.
- On the other hand, if the price of the chair were to remain at €10 but the demand for chairs changed from 300 to 200, there would be no impact on the nature of the solution.
- How should you proceed if the solution varies when the data is changed?

Introducing uncertainty

- Sensitivity analysis is used to study robustness of the solution to an LP model
- Sensitivity analysis offers a false sense of security when there is uncertainty in the problem data.
- Stochastic programming models are intended for problems in which data elements are difficult to predict or estimate.
- In the Dakota problem, what elements can be random?
  - Selling prices
  - Resource costs
  - Product demands
- We can reasonably assume that there is no uncertainty in the resource requirements for each of the three products.

Introducing uncertainty

- In order to illustrate the power of stochastic programming as a modelling paradigm, we will focus on the demand, and explore the impact of uncertainty in these values.
- That is, suppose that for each product we are given three potential demand scenarios, which we identify as being the 'low', 'most likely', and 'high' values.

	Low value	Most likely value	High value
Desks	50	150	250
Tables	20	110	250
Chairs	200	225	500
Probability	$p_l = 0.3$	$p_m = 0.4$	$p_h = 0.3$

Introducing uncertainty

- Now that we have a basic understanding of the deterministic model, and we've been given a collection of demand scenarios and their corresponding probabilities, what should we do next?
- When faced with this question, most people seem to come up with one or both of the following suggestions:
  - Use the expected demand and proceed.
  - Solve the problem for each of the demand scenarios, and somehow combine the solutions.

Expected demand

- The expected demand is actually the demand we initially used!
  - desks $= \mathbf{0.3} \times \mathbf{50} + \mathbf{0.4} \times \mathbf{150} + \mathbf{0.3} \times \mathbf{250} = \mathbf{150}$.
  - tables $= \mathbf{0.3} \times \mathbf{20} + \mathbf{0.4} \times \mathbf{110} + \mathbf{0.3} \times \mathbf{250} = \mathbf{125}$.
  - chairs $= \mathbf{0.3} \times \mathbf{200} + \mathbf{0.4} \times \mathbf{225} + \mathbf{0.3} \times \mathbf{500} = \mathbf{300}$.
- So the solution for the deterministic case is the same as the expected solution.
- Many deterministic models may actually represent the "average" case.

Scenario Analysis

In this case we will create copies of resource and production decisions for each scenario $s \in \{l, m, h\}$ where l, m and h correspond to the scenarios 'Low', 'Most likely' and 'High', respectively. The decision variables are:

x_l^s: Number of m^3 of lumber acquired under scenario s.

x_f^s: Number of labour hours acquired for finishing under scenario s.

x_c^s: Number of labour hours acquired for carpentry under scenario s.

y_d^s: Number of desks produced under scenario s.

y_t^s: Number of tables produced under scenario s.

y_c^s: Number of chairs produced under scenario s.

Demand Data: $(d_d^s; d_t^s; d_c^s)$ for each scenario s is given in slide (10)

Scenario s LP formulation

$$
\begin{aligned}
\max\ & 60y_d^s + 40y_t^s + 10y_c^s - 2x_l^s - 4x_f^s - 5.2x_c^s \\
& y_d^s \leq d_d^s \\
& y_t^s \leq d_t^s \\
& y_c^s \leq d_c^s \\
& 8y_d^s + 6y_t^s + y_c^s - x_l^s \leq 0 \\
& 4y_d^s + 2y_t^s + 1.5y_c^s - x_f^s \leq 0 \\
& 2y_d^s + 1.5y_t^s + 0.5y_c^s - x_c^s \leq 0 \\
& y_d^s, y_t^s, y_c^s, x_l^s, x_f^s, x_c^s \geq 0.
\end{aligned}
$$

Scenario Analysis - Solution

Resources	Demand			
	Expected	Low	Most likely	High
Lumber (m^3)	1950	520	1860	3500
Finishing labour (hrs.)	850	240	820	1500
Carpentry labour (hrs.)	487.5	130	465	875
Desks	150	50	150	250
Tables	125	20	110	250
Chairs	0	0	0	0
Profit (€)	4165	1124	3982	7450
Probability		0.3	0.4	0.3

Scenario Analysis - comments

- The basic structure of the LP that we've used hasn't changed.
- All cases indicate that chairs should not be produced as it still costs more to produce a chair than the revenue that we receive from the sale of a chair.
- The resource and production quantities in the first column are simply the expected values of the resource and production quantities in each of the individual scenario columns.
- The expected value of the optimal solutions obtained from the individual scenarios is an optimal solution to the expected demand problem

Scenario Analysis - comments

These following observations may offer some consolation

- The basic structure of all of the solutions is the same – no chairs are produced.
- There is a direct relationship between the collection of scenario solutions and the solution to the problem in which expected values are used.

This would seem to suggest that if they are interested in maximising the expected profit, Dakota should simply use the solution that appears in the first column.

Scenario Analysis - comments

- This solution would have Dakota produce 150 desks and 125 tables, and purchase exactly enough resource to accomplish this task.
- **BUT** if 150 desks are produced, there is a 30% chance that Dakota will produce more than they are able to sell.
- Similarly, if 125 tables are produced, there is a 70% chance Dakota will produce more than they are able to sell.
- If Dakota plans for the expected demand, and the demand turns out to be 'low', there will be 100 desks and 105 tables produced that will not be sold.
- Instead of the €4165 profit that the model suggests they should expect, they will realise a net loss of €6035 (= 3800 - 9835) .... and there is a 30% chance that this painful situation will occur.

Scenario Analysis - comments

Why does this happen?

- The solutions that we have looked at so far are, on an individual basis, best for one particular demand scenario
- As Dakota embarks on this planning adventure, they do not know what the demand will be, and it is important to obtain a solution that balances the impact of the various scenarios that might be encountered
- Because the model that we have used looks at only one demand scenario at a time, it is not able to provide this type of balance.

We need a model that explicitly considers the various scenarios . . . not individually, but collectively.

This is precisely the need that stochastic programming models fulfil.

Timing of decisions

In the Dakota problem, there are two types of decisions that are made

- Production ($\boldsymbol{y}$): How many desks, tables, and chairs should be produced?
- Resource ($\boldsymbol{x}$): How much lumber, finishing labour, and carpentry labour should be procured?

However there is still a critical piece of information that has not been specified

When do the various decisions have to be made?

Timing of decisions

- In complicated problems, ‘decisions’ and ‘uncertainty’ are often interwoven over time.
- That is, we make a decision, we observe what happens, more decisions are made, we observe again, etc.
- Neither the mean value nor the scenario problems are able to capture this interplay between ‘decisions’ and ‘uncertainty.’

Before a stochastic programming model can be developed, the timing of the decisions, relative to the resolution of uncertainty, must be specified.

Recourse

Recourse

Decisions that can be delayed until after information about the uncertain data is available offer an opportunity to adjust or adapt to the information that is received. We often refer to this adaption as recourse

- Models in which some decisions are delayed until after some information about the uncertain data has been obtained are known as **recourse models**
- Decision variables that are permitted to vary with the scenario are sometimes referred to as **recourse variables**.

A Recourse Model

- Recourse models result when some of the decisions must be fixed before information relevant to the uncertainties is available, while some of them can be delayed until after-ward.
- In the Dakota problem, suppose that the resource quantities must be determined fairly early, but production quantities can be delayed until after the demand is known.
- In some sense, we may think of the production quantities as being 'flexible' or 'adaptive' while the resource quantities are not.

Recourse variables

- To model the production quantities as being dependent on the demand scenario, the production variables that have been used thus far, $(\boldsymbol{y_d}, \boldsymbol{y_t}, \boldsymbol{y_c})$ will be replaced by $\{(\boldsymbol{y}_{\boldsymbol{d},\omega}, \boldsymbol{y}_{\boldsymbol{t},\omega}, \boldsymbol{y}_{\boldsymbol{c},\omega})\}_{\omega \in \{\boldsymbol{l},\boldsymbol{m},\boldsymbol{h}\}}$.
- For example, the variable $\boldsymbol{y_{t,h}}$ will denote the number of tables that will be produced if the demand is 'high' and the other variables are similarly defined.
- On the other hand, the resource variables do not vary with the scenario, and thus will remain as $\boldsymbol{x_l}$, $\boldsymbol{x_f}$, and $\boldsymbol{x_c}$.
- Since production quantities are being determined after the demand scenario is known, Dakota will not run the risk of producing items that it is not able to sell.

Recourse formulation

We may state the recourse version of Dakota's problem as follows

$$\max_{x_l, x_f, x_c \geq 0} -2x_l - 4x_f - 5.2x_c + E[h(x, \tilde{d})]$$

where $\tilde{d} = \{\tilde{d}_d, \tilde{d}_t, \tilde{d}_c\}$ represents the uncertain demand for desks, tables and chairs, and for each scenario $\omega \in \{l, m, h\}$,

$$\begin{aligned}
h(x, d_\omega) = \max & 60y_{d,\omega} + 40y_{t,\omega} + 10y_{c,\omega} \\
& 8y_{d,\omega} + 6y_{t,\omega} + y_{c,\omega} \leq x_l \\
& 4y_{d,\omega} + 2y_{t,\omega} + 1.5y_{c,\omega} \leq x_f \\
& 2y_{d,\omega} + 1.5y_{t,\omega} + 0.5y_{c,\omega} \leq x_c \\
& y_{d,\omega} \leq d_{d,\omega} \\
& y_{t,\omega} \leq d_{t,\omega} \\
& y_{c,\omega} \leq d_{c,\omega} \\
& y_{d,\omega}, y_{t,\omega}, y_{c,\omega} \geq 0
\end{aligned}$$

Recourse formulation - comments

- The subproblem determines the optimal allocation of resource to products, after the demand for products is known.
- The resource quantities $(\boldsymbol{x_l}, \boldsymbol{x_f}, \boldsymbol{x_c})$ are 'inputs' to the subproblem and appear on the right hand side of the constraints.

Non decomposed (or extended) formulation

$$
\begin{aligned}
\max -2x_l - 4x_f - 5.2x_c + \sum_{\omega \in \{l,m,h\}} p_\omega \times \{60y_{d,\omega} + 40y_{t,\omega} + 10y_{c,\omega}\} \\
-x_l + 8y_{d,\omega} + 6y_{t,\omega} + y_{c,\omega} \leq 0 \quad \forall \omega \\
-x_f + 4y_{d,\omega} + 2y_{t,\omega} + 1.5y_{c,\omega} \leq 0 \quad \forall \omega \\
-x_c + 2y_{d,\omega} + 1.5y_{t,\omega} + 0.5y_{c,\omega} \leq 0 \quad \forall \omega \\
y_{d,l} \leq 50 \qquad y_{d,m} \leq 150 \qquad y_{d,h} \leq 250 \\
y_{t,l} \leq 20 \qquad y_{t,m} \leq 110 \qquad y_{t,h} \leq 250 \\
y_{c,l} \leq 200 \qquad y_{c,m} \leq 225 \qquad y_{c,h} \leq 500 \\
x_l, x_f, x_c, y_{i,\omega} \geq 0, \quad i \in \{d,t,c\}, \quad \omega \in \{l,m,h\}
\end{aligned}
$$

Comparison of Dakota models

We've looked at three different models of the Dakota problem:

- one in which the random variables are replaced by their expected values,
- one in which each scenario is considered separately
- the recourse model

Comparison of Dakota models

Resource quantities	Mean value	Scenario solutions			Recourse		
		low	*m.likely*	*high*			
Lumber	1950	520	1860	3500	1300		
Fin. labour	850	240	820	1500	540		
Car. labour	487.5	130	465	875	325		
Production Quantities					*Demand*		
					low	*m.likely*	*high*
Desks	150	50	150	250	50	80	80
Tables	125	20	110	250	20	110	110
Chairs	0	0	0	0	200	0	0
Obj. Value	4165	1124	3982	7450	1730		

Recourse solution - comments

In terms of the resource quantities,

- it is difficult to identify a meaningful connection between the solutions to the scenario problems and the solution to the recourse problem;
- this contrasts sharply with the solution to the mean value problem, which is just the expected value of the individual scenario solutions.

Recourse solution - comments

In terms of the objective function values, where values range from 1124 to 7450

- Scenario solutions are based on the assumption of 'perfect information.' If we knew for sure that demand would be low, Dakota could obtain a maximum profit of 1124.
- The scenario problems consider only one scenario at a time. They do not consider the impact of planning for one scenario and having another scenario materialise.
- Similarly, the mean value solution plans only for the 'average' demand, and does not consider its impact under the various scenarios.

Recourse solution - comments

- Only the recourse model is designed to encourage a solution that balances the potential impact of the various scenarios.

  That is, the objective values for the mean value problem and the scenario problems are unrealistically optimistic – they do not include considerations of the impact of variability in the demand.
- Consequently, they fail to provide a meaningful representation of the expected profit.
- Unlike the mean value and scenario problems, the objective values of the recourse problems correspond to expected profit.

Recourse solution - comments

In terms of the production quantities

- It is the only model that suggests that chairs should be made, and then only when the demand for Dakota's products is low.
- The solution to the recourse model purchases sufficient resources to satisfy demand at some of the higher levels. That is, enough lumber and labour are procured to enable Dakota to address, at least partially, the higher demand scenarios.

Recourse solution - comments

- If it turns out that the demand is low, Dakota is able to recover much of their resource expense by producing chairs.
- In the low demand scenario, some of the lumber and carpentry labour will go unused, but Dakota will make maximum use of the resources that they have purchased.
- **In all cases, Dakota will be able to sell all items produced**.
- Thus, we see that in reality, chairs are useful as a fallback item. When demand is low, Dakota can cut its losses by producing and selling chairs, which is preferable to producing and not selling desks and tables!

Recourse solution - comments

We compare the objective value associated with the individual scenario problems and the corresponding expected profits in the recourse setting.

Demand Scenario	Profit suggested by Scenario Problem	Expected Profit
Recourse		1730
Expected value	4165	1545
Low	1124	1124
Most likely	3982	1702
High	7450	- 2050