Stochastic Optimization
Introduction and Examples

Alireza Ghaffari-Hadigheh

Azarbaijan Shahid Madani University (ASMU)

hadigheha@azaruniv.edu

Fall 2017

Overview

The farmer's problem

- Consider a farmer who specializes in raising wheat, corn, and sugar beets on his 500 acres of land.
- During the winter, he wants to decide how much land to devote to each crop.
- At least 200 tons (T) of wheat and 240 T of corn are needed for cattle feed
- Can be raised on the farm or bought from a wholesaler.
- Any production in excess of the feeding requirement would be sold.
- Over the last decade, mean selling prices have been $170 and $150 per ton of wheat and corn, respectively. The purchase prices are 40% more than this due to the wholesalers margin and transportation costs.

The farmer's problem: Assumptions

- The sugar beet is expected to be sold at \$36/T.
- The European Commission imposes a quota on sugar beet production. Any amount in excess of the quota can be sold only at \$10/T.
- The farmer's quota for next year is 6000 T.
- Based on past experience, the farmer knows that the mean yield on his land is roughly 2.5 T, 3 T, and 20 T per acre for wheat, corn, and sugar beets, respectively.

	Wheat	Corn	Sugar Beets
Yield (T/acre)	2.5	3	20
Planting cost (\$/acre)	150	230	260
Selling price (\$/T)	170	150	36 under 6000 T 10 above 6000 T
Purchase price (\$/T)	238	210	–
Minimum requirement (T)	200	240	–
Total available land: 500 acres			

The farmer's problem: Mathematical Model

- Decision variables:

$$
\begin{aligned}
x_1 &= \textit{acres of land devoted to wheat},\\
x_2 &= \textit{acres of land devoted to corn},\\
x_3 &= \textit{acres of land devoted to sugar beets},\\
w_1 &= \textit{tons of wheat sold},\\
y_1 &= \textit{tons of wheat purchased},\\
w_2 &= \textit{tons of corn sold},\\
y_2 &= \textit{tons of corn purchased},\\
w_3 &= \textit{tons of sugar beets sold at the favorable price},\\
w_4 &= \textit{tons of sugar beets sold at the lower price}.
\end{aligned}
$$

- Mathematical Model

$$
\begin{array}{ll}
\min & 150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4\\
s.t. & x_1 + x_2 + x_3 \leq 500\\
& 2.5x_1 + y_1 - w_1 \geq 200\\
& 3x_2 + y_2 - w_2 \geq 240\\
& w_3 + w_4 \leq 20x_3\\
& w_3 \leq 6000, x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0.
\end{array}
$$

The farmer's problem:Optimal Solution

- Optimal Solution

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	120	80	300
Yield (T)	300	240	6000
Sales (T)	100	–	6000
Purchase (T)	–	–	–
Overall profit: $118,600			

- Worries.
  - Over different years, the same crop, quite different yields because of changing weather conditions.
  - Most crops need rain during the few weeks after seeding or planting, then sunshine is welcome for the rest of the growing period.
  - Sunshine should not turn into drought, which causes severe yield reductions.
  - Dry weather is again beneficial during harvest.
  - From all these factors, yields varying 20 to 25% above or below the mean yield are not unusual.

A scenario representation

- Assume some correlation among the yields of the different crops.
- Assume that years are good, fair, or bad for all crops, resulting in above average, average, or below average yields for all crops.
- To fix these ideas, above and below average indicate a yield 20% above or below the mean yield.
- Weather conditions and yields for the farmer do not have a significant impact on prices.
- The farmer wishes to know whether the optimal solution is sensitive to variations in yields.
- He decides to run two more optimizations based on above average and below average yields.

A scenario representation: Optimal Solutions

Optimal solution based on above average yields (+ 20%).

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	183.33	66.67	250
Yield (T)	550	240	6000
Sales (T)	350	–	6000
Purchase (T)	–	–	–
Overall profit: $167,667			

Optimal solution based on below average yields (-20%).

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	100	25	375
Yield (T)	200	60	6000
Sales (T)	–	–	6000
Purchase (T)	–	180	–
Overall profit: $59,950			

A scenario representation: Formulation

- The farmer now realizes that he is unable to make a perfect decision that would be best in all circumstances.
- He would, therefore, want to assess the benefits and losses of each decision in each situation.
- Decisions on land assignment (x_1, x_2, x_3) have to be taken now, but sales and purchases $(w_i, i = 1, \ldots, 4, y_j, j = 1, 2)$ depend on the yields.
- Index those decisions by a scenario index $s = 1, 2, 3$ corresponding to above average, average, or below average yields, respectively.
- This creates a new set of variables of the form $w_{is}, i = 1, 2, 3, 4, s = 1, 2, 3$ and $y_{js}, j = 1, 2, s = 1, 2, 3$.
- As an example, w_{32} represents the amount of sugar beets sold at the favorable price if yields are average.
- The farmer wants to maximize long-run profit, it is reasonable for him to seek a solution that maximizes his expected profit.
- Three scenarios have an equal probability of $1/3$.

A scenario representation: Mathematical Model (The extensive form)

$$
\begin{aligned}
\min\ & 150x_1 + 230x_2 + 260x_3 \\
& \quad - \tfrac{1}{3}(170w_{11} - 238y_{11} + 150w_{21} - 210y_{21} + 36w_{31} + 10w_{41}) \\
& \quad - \tfrac{1}{3}(170w_{12} - 238y_{12} + 150w_{22} - 210y_{22} + 36w_{32} + 10w_{42}) \\
& \quad - \tfrac{1}{3}(170w_{13} - 238y_{13} + 150w_{23} - 210y_{23} + 36w_{33} + 10w_{43}) \\
\text{s.t. } & x_1 + x_2 + x_3 \le 500\ ,\ 3x_1 + y_{11} - w_{11} \ge 200\ , \\
& 3.6x_2 + y_{21} - w_{21} \ge 240\ ,\ w_{31} + w_{41} \le 24x_3\ ,\ w_{31} \le 6000\ , \\
& 2.5x_1 + y_{12} - w_{12} \ge 200\ ,\ 3x_2 + y_{22} - w_{22} \ge 240\ , \\
& w_{32} + w_{42} \le 20x_3\ ,\ w_{32} \le 6000\ ,\ 2x_1 + y_{13} - w_{13} \ge 200, \\
& 2.4x_2 + y_{23} - w_{23} \ge 240,\ w_{33} + w_{43} \le 16x_3\ , \\
& w_{33} \le 6000,\ x, y, w \ge 0\ .
\end{aligned}
$$

- It explicitly describes the second-stage decision variables for all scenarios.

A scenario representation: Mathematical Model (The extensive form)

		Wheat	Corn	Sugar Beets
First Stage	Area (acres)	170	80	250
$s = 1$ Above	Yield (T)	510	288	6000
	Sales (T)	310	48	6000 (favor. price)
	Purchase (T)	–	–	–
$s = 2$ Average	Yield (T)	425	240	5000
	Sales (T)	225	–	5000 (favor. price)
	Purchase (T)	–	–	–
$s = 3$ Below	Yield (T)	340	192	4000
	Sales (T)	140	–	4000 (favor. price)
	Purchase (T)	–	48	–
Overall profit: \$108,390				

- The top line gives the planting areas, which must be determined before realizing the weather and crop yields. This decision is called the first stage.
- The other lines describe the yields, sales, and purchases in the three scenarios. They are called the second stage.
- The bottom line shows the overall expected profit.

A scenario representation: Illustration of the solution

- The most profitable decision for sugar beet land allocation is the one that always avoids sales at the unfavorable price even if this implies that some portion of the quota is unused when yields are average or below average.
- The area devoted to corn is such that it meets the feeding requirement when yields are average. This implies sales are possible when yields are above average and purchases are needed when yields are below average.
- Finally, the rest of the land is devoted to wheat. This area is large enough to cover the minimum requirement. Sales then always occur.
- This solution illustrates that it is impossible, under uncertainty, to find a solution that is ideal under all circumstances.
- Selling some sugar beets at the unfavorable price or having some unused quota is a decision that would never take place with a perfect forecast.
- Such decisions can appear in a stochastic model because decisions have to be balanced or hedged against the various scenarios.

A scenario representation: The hedging effect

- Suppose yields vary over years but are cyclical. A year with above average yields is always followed by a year with average yields and then a year with below average yields.
  - The farmer would then take optimal solutions as given in Table 3, then Table 2, then Table 4, respectively.
  - This would leave him with a profit of \$167,667 the first year, \$118,600 the second year, and \$59,950 the third year.
  - The mean profit over the three years (and in the long run) would be the mean of the three figures, namely \$115,406 per year.
- Assume again that yields vary over years, but on a random basis.
  - If the farmer gets the information on the yields before planting, he will again choose the areas on the basis of the solution in Table 2, 3, or 4, depending on the information received.
  - In the long run, if each yield is realized one third of the years, the farmer will get again an expected profit of \$115,406 per year.
  - This is the situation under perfect information.

A scenario representation: The hedging effect

- The farmer unfortunately does not get prior information on the yields.
- The best he can do in the long run is to take the solution as given by Table 5.
- This leaves the farmer with an expected profit of \$108,390.
- The difference between this figure and the value, \$115,406, in the case of perfect information, namely \$7016, represents what is called the expected value of perfect information (EVPI).
- Another approach the farmer may have is to assume expected yields and always to allocate the optimal planting surface according to these yields, as in Table 2.
- This approach represents the expected value solution.
- It is common in optimization but can have unfavorable consequences.
- The loss by not considering the random variations is the difference between this and the stochastic model profit from Table 5.
- This value, \$108,390- 107,240=\$1,150, is the value of the stochastic solution (VSS), the possible gain from solving the stochastic model.
- Note that it is not equal to the expected value of perfect information, and, as we shall see in later models, may in fact be larger than the EVPI .

General model formulation

- A set of decisions to be taken without full information on some random events. These decisions are called first-stage decisions and are usually represented by a vector x.
- In the farmer example, they are the decisions on how many acres to devote to each crop.
- Later, full information is received on the realization of some random vector ξ .
- Then, second-stage or corrective actions y are taken.
- The functional form, such as $\xi(\omega)$ or $y(s)$, to show explicit dependence on an underlying element, ω or s.

$$\begin{array}{ll} \min & c^T x + E_\xi Q(x,\xi) \\ s.t. & Ax = b, \\ & x \geq 0 \end{array} \tag{1}$$

where $Q(x,\xi) = \min\{q^T y | Wy = h - Tx, y \geq 0\}$, ξ is the vector formed by the components of q^T, h^T, and T , and E_ξ denote mathematical expectation with respect to ξ. We assume here that W is fixed (fixed recourse).

Farmer's Example: Revisited

- The random vector is a discrete variable with only three different values.
- Only the T matrix is random.
- A second-stage problem for one particular scenario s is

$$
\begin{aligned}
Q(x,s) = \min \ & \{238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4\} \\
\text{s. t. } & t_1(s)x_1 + y_1 - w_1 \geq 200 \ , \\
& t_2(s)x_2 + y_2 - w_2 \geq 240 \ , \\
& w_3 + w_4 \leq t_3(s)x_3 \ , \\
& w_3 \leq 6000 \ , \\
& y, w \geq 0 \ ,
\end{aligned}
$$

- $t_i(s)$ represents the yield of crop i under scenario s (or state of nature s).
- The random vector $\xi = (t_1, t_2, t_3)$ is formed by the three yields and that ξ can take on three different values, say ξ_1, ξ_2, and ξ_3, which represent $(t_1(1), t_2(1), t_3(1)), ({}_t1(2), t_2(2), t_3(2))$, and $(t_1(3), t_2(3), t_3(3))$, respectively
- The random vector $\xi(s)$ depends on the scenario s , which takes on three different values.

Continuous random variables

- Assumption: yields for the different crops are independent.
- Consider a continuous random vector for the yields.
- Assume that the yield for each crop i can be appropriately described by a uniform random variable, inside some range $[l_i, u_i]$.
- For the sake of comparison, we may take l_i to be 80% of the mean yield and u_i to be 120% of the mean yield so that the expectations for the yields will be the same.
- the decisions on land allocation are first-stage decisions because they are taken before knowledge of the yields.
- Second-stage decisions are purchases and sales after the growing period.
- The second-stage formulation can again be described as $Q(x) = E_\xi Q(x, \xi)$, where $Q(x, \xi)$ is the value of the second stage for a given realization of the random vector.

computation of $Q(x,\xi)$

- Can be separated among the three crops due to independence of the random vector.

$$E_\xi Q(x,\xi) = \sum_{i=1}^{3} E_\xi Q_i(x_i,\xi) = \sum_{i=1}^{3} \mathcal{Q}_i(x_i),$$

  where $Q_i(x_i,\xi)$ is the optimal second-stage value of purchases and sales of crop i.
- Sugar beet sales: for a given value $t_3(\xi)$ of the sugar beet yield, one obtains the following second-stage problem:

$$\begin{aligned} Q_3(x_3,\xi) = \min\ & -36w_3(\xi) - 10w_4(\xi) \\ \text{s. t. }\ & w_3(\xi) + w_4(\xi) \le t_3(\xi)x_3\ , \\ & w_3(\xi) \le 6000\ , \\ & w_3(\xi), w_4(\xi) \ge 0\ . \end{aligned}$$

The optimal decisions of this problem

- The optimal decisions for this problem are clearly to sell as many sugar beets as possible at the favorable price, and to sell the possible remaining production at the unfavorable price:

$$\begin{aligned} w_3(\xi) &= \min[6000, t_3(\xi)x_3], \\ w_4(\xi) &= \max[t_3(\xi)x_3 - 6000, 0]. \end{aligned}$$

- Second-stage value:

$$Q_3(x_3, \xi) = -36 \min[6000, t_3(\xi)x_3] - 10 \max[t_3(\xi)x_3 - 6000, 0].$$

- First assume that the surface x_3 devoted to sugar beets will not be so large that the quota would be exceeded for any possible yield or so small that production would always be less than the quota for any possible yield. In other words

$$l_3 x_3 \leq 6000 \leq u_3 x_3$$

The expected value of the second stage

- The expected value of the second stage for sugar beet sales is

$$\begin{aligned}\mathcal{Q}_3(x_3) &= E_\xi Q_3(x_3,\xi_3)\\ &= -\int_{l_3}^{6000/x_3} 36tx_3 f(t)dt\\ &\quad -\int_{6000/x_3}^{u_3}(216000+10tx_3-6000)f(t)dt\end{aligned}$$

  $f(t)$ denotes the density of the random yield $t_3(\xi)$.

- After some computation,

$$\begin{aligned}\mathcal{Q}_3(x_3) &= -18\frac{(u_3^2-l_3^2)x_3}{u_3-l_3}+\frac{13(u_3x_3-6000)^2}{x_3(u_3-l_3)}\\ &= -36\bar{t}_3x_3+\frac{13(u_3x_3-6000)^2}{x_3(u_3-l_3)}\end{aligned}$$

- $\bar{t}_3$ denotes the expected yield for sugar beet production, which is $\frac{u_3+l_3}{2}$ for a uniform density.

Alternative Assumption

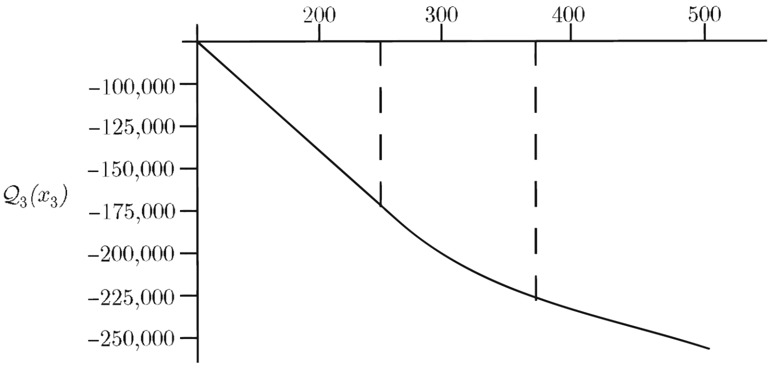
- If the surface x_3 is such that the production exceeds the quota for any possible yield $(l_3 x_3 > 6000)$, then the optimal second-stage decisions are

$$\begin{aligned} w_3(\xi) &= 6000, \\ w_4(\xi) &= t_3(\xi)x_3 - 6000, \forall \xi \end{aligned}$$

- The second-stage value for a given ξ is

$$\mathcal{Q}_3(x_3, \xi) = -216000 - 10(t_3(\xi)x_3 - 6000) = -156000 - 10t_3(\xi)x_3$$

- The expected value is $\mathcal{Q}_3(x_3) = -156000 - 10\bar{t}_3 x_3$
- If the surface devoted to sugar beets is so small that for any yield the production is lower than the quota, the second-stage value function is $\mathcal{Q}_3(x_3) = -36\bar{t}_3 x_3$.



- The graph of the function $\mathcal{Q}_3(x_3)$ for all possible values of x_3. Note that with our assumption of $\bar{t}_3 = 20$, we would then have the limits on x_3 as $250 \leq x_3 \leq 375$.
- The function has three different pieces. Two of these pieces are linear and one is nonlinear, but the function $\mathcal{Q}_3(x_3)$ is continuous and convex.

- Other two values.

$$\mathcal{Q}_1(x_1) = \begin{cases} 47600 - 595x_1 & \text{for } x_1 \leq 200/3\,, \\ 119\frac{(200-2x_1)^2}{x_1} - 85\frac{(200-3x_1)^2}{x_1} & \text{for } \frac{200}{3} \leq x_1 \leq 100\,, \\ 34000 - 425x_1 & \text{for } x_1 \geq 100\,, \end{cases}$$

$$\mathcal{Q}_2(x_2) = \begin{cases} 50400 - 630x_2 & \text{for } x_2 \leq 200/3\,, \\ 87.5\frac{(240-2.4x_2)^2}{x_2} - 62.5\frac{(240-3.6x_2)^2}{x_2} & \text{for } 200/3 \leq x_2 \leq 100\,, \\ 36000 - 450x_2 & \text{for } x_2 \geq 100\,. \end{cases}$$

- The global problem

$$\begin{array}{ll} \min & 150x_1 + 230x_2 + 260x_3 + \mathcal{Q}_1(x_1) + \mathcal{Q}_2(x_2) + \mathcal{Q}_3(x_3) \\ s.t. & x_1 + x_2 + x_3 \leq 500, \\ & x_1, x_2, x_3 \geq 0 \end{array}$$

Solution Approach

- The three functions $\mathcal{Q}_i(x_i)$ are convex, continuous, and differentiable functions and the first-stage objective is linear, this problem is a convex program for which Karush-Kuhn-Tucker (K-K-T) conditions are necessary and sufficient for a global optimum.
- Denoting by λ the multiplier of the surface constraint and as before by c_i the first-stage objective coefficient of crop i , the K-K-T conditions require

$$x_i\left[c_i+\frac{\partial\mathcal{Q}_i(x_i)}{\partial x_i}+\lambda\right]=0,$$

$$c_i+\frac{\partial\mathcal{Q}_i(x_i)}{\partial x_i}+\lambda\geq 0, x_i\geq 0, i=1,2,3$$

$$\lambda[x_1+x_2+x_3-500]=0, x_1+x_2+x_3\leq 500, \lambda\geq 0$$

- Assume the optimal solution is such that $100\leq x_1$, $\frac{200}{3}\leq x_2\leq 100$, and $250\leq x_3\leq 375$ with $\lambda\neq 0$

Optimal solution

$$\begin{cases} -275 + \lambda = 0 \,, \\ -76 - \frac{1.44\ 10^6}{x_2^2} + \lambda = 0 \,, \\ 476 - \frac{5.85\ 10^7}{x_3^2} + \lambda = 0 \,, \\ x_1 + x_2 + x_3 = 500 \,. \end{cases}$$

- $\lambda = 275.00, x_1 = 135.83, x_2 = 85.07, x_3 = 279.10.$
- Satisfies all the required conditions and is therefore optimal.

The news vendor problem

- A news vendor goes to the publisher every morning and buys x newspapers at a price of c per paper.
- This number is usually bounded above by some limit u, representing either the news vendors purchase power or a limit set by the publisher to each vendor.
- The vendor then walks along the streets to sell as many newspapers as possible at the selling price q .
- Any unsold newspaper can be returned to the publisher at a return price r , with $r < c$.
- Help the news vendor decide how many newspapers to buy every morning.
- Demand for newspapers varies over days and is described by a random variable ξ.
- It is assumed here that the news vendor cannot return to the publisher during the day to buy more newspapers. Other news vendors would have taken the remaining newspapers.
- Readers also only want the last edition.

Formulation of the problem

- Define y as the effective sales and w as the number of newspapers returned to the publisher at the end of the day.
- Mathematical formulation

$$\min_{0\leq x\leq u} cx + \mathcal{Q}(x),$$

where

$$\mathcal{Q}(x) = E_\xi Q(x,\xi),$$

$$\begin{aligned} Q(x,\xi) \quad = \quad \min \quad & -qy(\xi) - rw(\xi) \\ s.t. \quad & y(\xi) \leq \xi \\ & y(\xi) + w(\xi) \leq x \\ & y(\xi), w(\xi) \geq 0 \end{aligned}$$

- $-\mathcal{Q}(x)$ is the expected profit on sales and returns, while $-Q(x,\xi)$is the profit on sales and returns if the demand is at level ξ.

Simple Rules

- The model illustrates the two-stage aspect of the news vendor problem.
  - The buying decision has to be taken before any information is given on the demand.
  - When demand is known in the so-called second stage, which represents the end of the sales period of a given edition, the profit can be computed.
- This is done using the following simple rule:

$$
\begin{aligned}
y(\xi) &= \min(\xi, x), \\
w(\xi) &= \max(x - \xi, 0)
\end{aligned}
$$

- Sales can never exceed the number of available newspapers or the demand. Returns occur only when demand is less than the number of newspapers available.
- The second-stage expected value function is

$$\mathcal{Q}(x) = E_{\xi}[-q \min(\xi, x) - r \max(x - \xi, 0)].$$

Solution Approach

- This function is convex and continuous.
- It is also differentiable when ξ is a continuous random vector.
- The optimal solution of the news vendors problem is

$$\begin{cases} x = 0 & \textit{if } \ c + \mathcal{Q}'(0) > 0, \\ x = u & \textit{if } \ c + \mathcal{Q}'(u) < 0, \\ \textit{a solution of } \ c + \mathcal{Q}'(x) = 0 & \textit{otherwise} \end{cases}$$

where $\mathcal{Q}'(x)$ denotes the first order derivative of $\mathcal{Q}(x)$ evaluated at x.
- $\mathcal{Q}(x)$ can be computed as

$$\begin{aligned} \mathcal{Q}(x) &= \int_{-\infty}^{x} (-q\xi - r(x-\xi))dF(\xi) + \int_{x}^{\infty} -qxdF(\xi) \\ &= -(q-r)\int_{-\infty}^{x} \xi dF(\xi) - rxF(x) - qx(1-F(x)) \end{aligned}$$

- $F(\xi)$ represents the cumulative probability distribution of ξ.

Optimal Solution

- Integrating by parts,

$$\int_{-\infty}^{x} \xi dF(\xi) = xF(x) - \int_{-\infty}^{x} F(\xi)d\xi$$

$$\mathcal{Q}(x) = -qx + (q-r)\int_{-\infty}^{x} F(\xi)d\xi$$

$$\mathcal{Q}'(x) = -q + (q-r)F(x)$$

- Optimal solution:

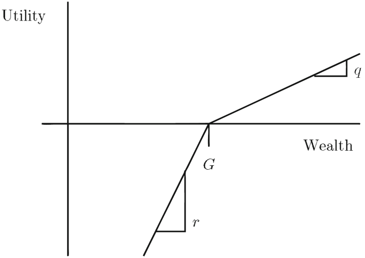
$$\begin{cases} x = 0 & \mathit{if}\ \frac{q-c}{q-r} < F(0), \\ x = u & \mathit{if}\ \frac{q-c}{q-r} > F(u), \\ x = F^{-1}(\frac{q-c}{q-r}) & \mathit{otherwise}, \end{cases}$$

- where $F^{-1}(\alpha)$is the α -quantile of F.

Financial Planning and Control: an example

- The essence of financial planning is the incorporation of risk into investment decisions.
- This example involves randomness in the constraint matrix instead of the right-hand side elements.
- We wish to provide for a child's college education Y years from now.
- We currently have \$ b to invest in any of I investments. After Y years, we will have a wealth that we would like to have exceed a tuition goal of \$ G . We suppose that we can change investments every u years, so we have $H = Y/u$ investment periods.
- We ignore transaction costs and taxes on income although these considerations would be important in reality.
- We suppose that exceeding \$ G after Y years would be equivalent to our having an income of $q\%$ of the excess while not meeting the goal would lead to borrowing for a cost r % of the amount short.

Utility function of wealth



- The major uncertainty in this model is the return on each investment i within each period t .
- We describe this random variable as $\xi(i,t) = \xi(i,t,\omega)$ where ω is some underlying random element.
- The decisions on investments will also be random. We describe these decisions as $x(i,t) = x(i,t,\omega)$.
- From the randomness of the returns and investment decisions, our final wealth will also be a random variable.

A key point

- We cannot completely observe the random element ω when we make all our decisions $x(i, t, \omega)$.
- We can only observe the returns that have already taken place.
- In stochastic programming, we say that we cannot anticipate every possible outcome so our decisions are nonanticipative of future outcomes.
- Before the first period, this restriction corresponds to saying that we must make fixed investments, $x(i, 1)$, for all $\omega \in \Omega$, the space of all random elements or, more specifically, returns that could possibly occur.

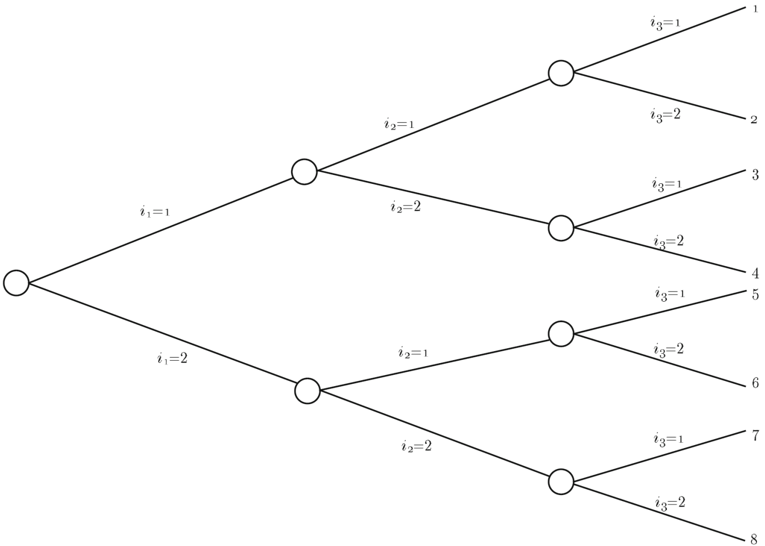
The effects of including stochastic outcomes

- The effects of including stochastic outcomes as well as modeling effects from choosing the time horizon Y and the coarseness of the period approximations H
- Consider a simple example with two possible investment types,
  - Stocks ($i = 1$)
  - Government securities (bonds) ($i = 2$).
- Set Y at 15 years and allow investment changes every five years so that $H = 3$.
- Assume that, over the three decision periods, eight possible scenarios may occur. indicate the scenarios by an index $s = 1, \cdots, 8$, which represents a collection of the outcomes ω that have common characteristics (such as returns) in a specific model.
- The scenarios correspond to independent and equal likelihoods of having (inflation-adjusted) returns over the five-year period
  - 1.25 for stocks and 1.14 for bonds
  - 1.06 for stocks and 1.12 for bonds.

Introduction

- Assign probabilities for each s , $p(s) = 0.125$.
- The returns are $\xi(1, t, s) = 1.25$, $\xi(2, t, s) = 1.14$ for $t = 1$ and $s = 1, \ldots, 4$ for $t = 2$, $s = 1, 2, 5, 6$, and for $t = 3$, $s = 1, 3, 5, 7$. In the other cases, $\xi(1, t, s) = 1.06$, $\xi(2, t, s) = 1.12$
- The scenario tree divides into branches corresponding to different realizations of the random returns.
- Because Scenarios 1 to 4, for example, have the same return for $t = 1$, they all follow the same first branch.
- Scenarios 1 and 2 then have the same second branch and finally divide completely in the last period.
- To show this more explicitly, we may refer to each scenario by the history of returns indexed by st for periods $t = 1, 2, 3$ as indicated on the tree in Figure.
- Scenario 1 may also be represented as $(s_1, s_2, s_3) = (1, 1, 1)$.

Tree Representation



Mathematical program

- We need only have a decision vector for each node of the tree.
- The decisions at $t = 1$ are just $x(1,1)$ and $x(2,1)$ for the amounts invested in stocks (1) and bonds (2) at the outset.
- For $t = 2$, we would have $x(i, 2, s_1)$ where $i = 1, 2$ for the type of investment and $s_1 = 1, 2$ for the first-period return outcome.
- The decisions at $t = 3$ are $x(i, 3, s_1, s_2)$.
- A mathematical program to maximize expected utility.
- Because the concave utility function 1 is piecewise linear, we just need to define deficit or shortage and excess or surplus variables, $w(i_1, i_2, i_3)$ and $y(i_1, i_2, i_3)$, and we can maintain a linear model.

Objective function and Constraints

- The objective is a probability- and penalty-weighted sum of these terms

$$\sum_{s_H} \cdots \sum_{s_1} p(s_1, \dots, s_H)(-rw(s_1, \dots, s_H) + qy(s_1, \dots, s_H))$$

- The first-period constraint is to invest the initial wealth:

$$\sum_i x(i, 1) = b.$$

- The constraints for periods $t = 2, \dots, H$ are, for each $s_1, \dots, s_{t-1}$

$$\sum_i -\xi(i, t-1, s_1, \dots, s_{t-1})x(i, t-1, s_1, \dots, s_{t-2}) + \sum_i x(i, t, s_1, \dots, s_{t-1}) = 0,$$

- The constraints for period H

$$\sum_i \xi(i, H, s_1, \dots, s_H)x(i, H, s_1, \dots, s_{H-1}) - y(s_1, \dots, s_H) + w(s_1, \dots, s_H) = G.$$

- Other constraints restrict the variables to be non-negative.

Specifying the model

- Initial wealth, $b = 55,000$; target value, $G = 80,000$; surplus reward, $q = 1$; and shortage penalty, $r = 4$

$$\max z = \sum_{s_1=1}^{2} \sum_{s_2=1}^{2} \sum_{s_3=1}^{2} 0.125(y(s_1,s_2,s_3) - 4w(s_1,s_2,s_3))$$

$$\begin{aligned}
\text{s. t.} \quad & x(1,1) + x(2,1) &= 55\,,\\
& -1.25x(1,1) - 1.14x(2,1) + x(1,2,1) + x(2,2,1) &= 0\,,\\
& -1.06x(1,1) - 1.12x(2,1) + x(1,2,2) + x(2,2,2) &= 0\,,\\
& -1.25x(1,2,1) - 1.14x(2,2,1) + x(1,3,1,1) + x(2,3,1,1) &= 0\,,\\
& -1.06x(1,2,1) - 1.12x(2,2,1) + x(1,3,1,2) + x(2,3,1,2) &= 0\,,\\
& -1.25x(1,2,2) - 1.14x(2,2,2) + x(1,3,2,1) + x(2,3,2,1) &= 0\,,\\
& -1.06x(1,2,2) - 1.12x(2,2,2) + x(1,3,2,2) + x(2,3,2,2) &= 0\,,\\
& 1.25x(1,3,1,1) + 1.14x(2,3,1,1) - y(1,1,1) + w(1,1,1) &= 80\,,\\
& 1.06x(1,3,1,1) + 1.12x(2,3,1,1) - y(1,1,2) + w(1,1,2) &= 80\,,\\
& 1.25x(1,3,1,2) + 1.14x(2,3,1,2) - y(1,2,1) + w(1,2,1) &= 80\,,\\
& 1.06x(1,3,1,2) + 1.12x(2,3,1,2) - y(1,2,2) + w(1,2,2) &= 80\,,\\
& 1.25x(1,3,2,1) + 1.14x(2,3,2,1) - y(2,1,1) + w(2,1,1) &= 80\,,\\
& 1.06x(1,3,2,1) + 1.12x(2,3,2,1) - y(2,1,2) + w(2,1,2) &= 80\,,\\
& 1.25x(1,3,2,2) + 1.14x(2,3,2,2) - y(2,2,1) + w(2,2,1) &= 80\,,\\
& 1.06x(1,3,2,2) + 1.12x(2,3,2,2) - y(2,2,2) + w(2,2,2) &= 80\,,\\
& x(i,t,s_1,\ldots,s_{t-1}) \geq 0\,,\ y(s_1,s_2,s_3) \geq 0\,,\ w(s_1,s_2,s_3) &\geq 0\,,
\end{aligned}$$

Introduction

- Solving the problem yields an optimal expected utility value of -1.514 .

Period, Scenario	Stock	Bonds
1,1-8	41.5	13.5
2,1-4	65.1	2.17
2,5-8	36.7	22.4
3,1-2	83.8	0.00
3,3-4	0.00	71.4
3,5-6	0.00	71.4
3,7-8	64.0	0.00
Scenario	Above G	Below G
1	24.8	0.00
2	8.87	0.00
3	1.43	0.00
4	0.00	0.00
5	1.43	0.00
6	0.00	0.00
7	0.00	0.00
8	0.00	12.2

Interpretation of the results

- The initial investment is heavily in stock ($41,500) with only $13,500 in bonds.
- In the case of Scenarios 1 to 4, stocks are even more prominent, while Scenarios 5 to 8 reflect a more conservative government security portfolio.
- In the last period, notice how the investments are either completely in stocks or completely in bonds.
- This is a general trait of one-period decisions. It occurs here because in Scenarios 1 and 2, there is no risk of missing the target.
- In Scenarios 3 to 6, stock investments may cause one to miss the target, so they are avoided.
- In Scenarios 7 and 8, the only hope of reaching the target is through stocks.

Comparison of the results to a deterministic model

- All random returns are replaced by their expectation.
- because the expected return on stock is 1.155 in each period, while the expected return on bonds is only 1.13 in each period, the optimal investment plan places all funds in stocks in each period.
- If we implement this policy each period, but instead observed the random returns, we would have an expected utility called the expected value solution, or EV .
- In this case, we would realize an expected utility of EV = -3.788 , while the stochastic program value is again RP = -1.514.
- The difference between these quantities is the value of the stochastic solution:

$$VSS = RPEV = -1.514 - (-3.788) = 2.274.$$

- This comparison gives us a measure of the utility value in using a decision from a stochastic program compared to a decision from a deterministic program.

New formulation

- The formulation can become quite cumbersome as the time horizon H , increases and the decision tree grows quite bushy.
- Another modeling approach to this type of multistage problem is to consider the full horizon scenarios, s , directly, without specifying the history of the process.
- We substitute a scenario set S for the random elements Ω .
- Probabilities, $p(s)$, returns, $\xi(i, t, s)$, and investments, $x(i, t, s)$, become functions of the H -period scenarios and not just the history until period t.
- The difficulty is that, when we have split up the scenarios, we may have lost nonanticipativity of the decisions because they would now include knowledge of the outcomes up to the end of the horizon.
- To enforce nonanticipativity, we add constraints explicitly in the formulation.

The new general formulation

- First, the scenarios that correspond to the same set of past outcomes at each period form groups, $S^t_{s_1,\ldots,s_{t-1}}$, for scenarios at time t .
- Now, all actions up to time t must be the same within a group. We do this through an explicit constraint.

$$
\begin{aligned}
\max z &= \sum_s p(s)(qy(s) - rw(s)) \\
\text{s. t. } & \sum_{i=1}^{I} x(i,1,s) = b\ , \ \forall s \in S\ , \\
& \sum_{i=1}^{I} \xi(i,t,s)x(i,t-1,s) - \sum_{i=1}^{I} x(i,t,s) = 0\ , \ \forall s \in S\ , \\
& \qquad\qquad t = 2,\ldots,H\ , \\
& \sum_{i=1}^{I} \xi(i,H,s)x(i,H,s) - y(s) + w(s) = G\ , \\
& \left(\sum_{s' \in S^t_{J(s,t)}} p(s')x(i,t,s')\right) - \left(\sum_{s' \in S^t_{J(s,t)}} p(s')\right) x(i,t,s) = 0\ , \\
& \qquad\qquad \forall 1 \le i \le I\ , \forall 1 \le t \le H\ , \forall s \in S\ , \\
& x(i,t,s) \ge 0\ , \ y(s) \ge 0\ , \ w(s) \ge 0\ , \\
& \qquad\qquad \forall\ 1 \le i \le I\ , \forall\ 1 \le t \le H\ , \forall\ s \in S\ ,
\end{aligned}
$$

- $J(s,t) = \{s_1, \ldots, s_{t-1}\}$ such that $s \in S^t_{s_1,\ldots,s_{t-1}}$.
- the last equality constraint indeed forces all decisions within the same group at time t to be the same.
- These nonanticipativity constraints are the only constraints linking the separate scenarios.
- Without them, the problem would decompose into a separate problem for each s , maintaining the structure of that problem.

Capacity Expansion

- Optimal choices of the timing and levels of investments to meet future demands of a given product.
- Power plant expansion for electricity generation: Find optimal levels of investment in various types of power plants to meet future electricity demand.
- Static deterministic analysis of the electricity generation problem.
  - Static means that decisions are taken only once.
  - Deterministic means that the future is supposed to be fully and perfectly known.
- Properties of a given power plant i
  - The investment cost r_i,
  - The operating cost q_i,
  - The availability factor a_i (the percent of time the power plant can effectively be operated).

Demand

- Demand for electricity can be considered a single product, but the level of demand varies over time.
- The demand in terms of load duration curve that describes the demand over time in decreasing order of demand level.

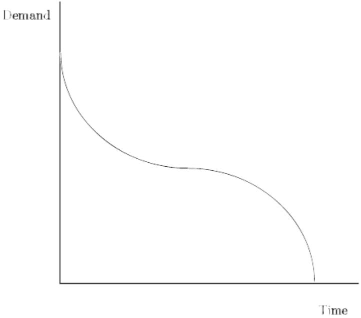


Fig. 5 The load duration curve.

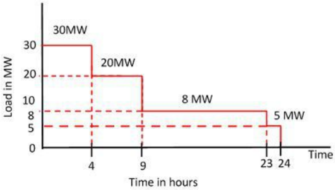
- The curve gives the time τ that each demand level D, is reached.
- Because we are concerned with investments over the long run, the load duration curve we consider is taken over the life cycle of the plants.

The load duration curve

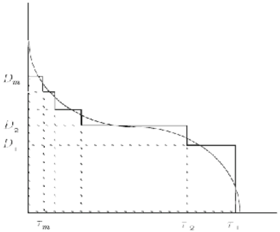
Time	Load in MW
6.00 am to 8.00am	8
8.00 am to 1.00 noon	20
1.00 noon to 2.00 noon	5
12.30 noon to 6.00 pm	30
6.00 pm to 6.00 am	8

Solution: The data available from the load curve are tabulated as follows. Here the total time is 24 hours or 100%.

Load in MW	Hours in a day	Time in percentage
30	4	4/5×100=16.67%
20	4+5	9/24×100=37.5%
8	2+4+5+12 =23	23/24×100=95.83%
5	4+5+2+12+1 = 24	24/24×100=100%



Load Duration Curve



The problem in the static situation

- Finding the optimal investment for each mode j , i.e., to find the particular type of power plant $i, i = 1, \dots, n$, that minimizes the total cost of effectively producing 1 MW (megawatt) of electricity during the time τ_j.

$$i(j) = \mathbf{argmin}_{i=1,\dots,n} \left\{ \frac{r_i + q_i \tau_j}{a_i} \right\}, \tag{3.1}$$

- n is the number of available technologies.
- Four elements justify considering a *dynamic or multistage model*:
  - the long-term evolution of equipment costs;
  - the long-term evolution of the load curve;
  - the appearance of new technologies;
  - the obsolescence of currently available equipment.

Multistage model

- $t = 1, \ldots, H$ index the periods or stages;
- $i = 1, \ldots, n$ index the available technologies;
- $j = 1, \ldots, m$ index the operating modes in the load duration curve.
- $a_i =$ availability factor of i;
- $L_i =$ lifetime of i ;
- g_i^t = existing capacity of i at time t , decided before $t = 1$;
- r_i^t = unit investment cost for i at time t (assuming a fixed plant life cycle for each type i of plant);
- q_i^t = unit production cost for i at time t;
- d_j^t = maximal power demanded in mode j at time t;
- τ_j^t = duration of mode j at time t.

The set of decisions

- $x_i^t =$ new capacity made available for technology i at time t;
- $w_i^t =$ total capacity of i available at time t;
- y_{ij}^t = capacity of i effectively used at time t in mode j.

The electricity generation H-stage problem

$$\min_{x,y,w} \sum_{t=1}^{H} \left(\sum_{i=1}^{n} r_i^t \cdot w_i^t + \sum_{i=1}^{n} \sum_{j=1}^{m} q_i^t \cdot \tau_j^t \cdot y_{ij}^t \right) \qquad (3.2)$$

$$\text{s. t. } w_i^t = w_i^{t-1} + x_i^t - x_i^{t-L_i} , \qquad i = 1,\ldots,n , \ t = 1,\ldots,H , \qquad (3.3)$$

$$\sum_{i=1}^{n} y_{ij}^t = d_j^t , \qquad j = 1,\ldots,m , \ t = 1,\ldots,H , \qquad (3.4)$$

$$\sum_{j=1}^{m} y_{ij}^t \le a_i (g_i^t + w_i^t) , \qquad i = 1,\ldots,n , \ t = 1,\ldots,H , \qquad (3.5)$$

$$x, y, w \ge 0 .$$

- Decisions in each period t involve new capacities x_i^t made available in each technology
- Capacities y_{ij}^t operated in each mode for each technology.
- Newly decided capacities increase the total capacity w_i^t made available, as given by (3.3), where the equipments becoming obsolete after its lifetime is also considered.

- Assumption $x_i^t = 0$ if $t \le 0$, so equation (3.3) only involves newly decided capacities.
- By (3.4), the optimal operation of equipment must be chosen to meet demand in all modes using available capacities, which by (3.5) depend on capacities g_i^t decided before $t = 1$, newly decided capacities x_i^t, and the availability factor.
- The objective function (3.2) is the sum of the investment plus maintenance costs and operating costs.

Stochastic model

- The main difference is in the definition of the variables x_i^t and w_i^t.
- x_i^t represents the new capacity of i decided at time t, which becomes available at time $x_i^{t+\triangle i}$,($\triangle i$ is the construction delay for equipment i).
- To have extra capacity available at time t, it is necessary to decide at $t - \triangle i$, when less information is available on the evolution of demand and equipment costs.
- Assumption: each decision is now a random variable. Instead of writing an explicit dependence on the random element, ω, we again use boldface notation to denote random variables.
  - x_i^t = new capacity decided at time t for equipment $i, i = 1, \ldots, n$;
  - w_i^t = total capacity of i available and in order at time t;
  - ξ = the vector of random parameters at time t.

The stochastic model

$$\min \mathrm{E}_{\boldsymbol{\xi}} \sum_{t=1}^{H} \left(\sum_{i=1}^{n} \mathbf{r}_i^t \mathbf{w}_i^t + \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{q}_i^t \, \tau_j^t \, \mathbf{y}_{ij}^t \right) \tag{3.6}$$

$$\text{s. t. } \mathbf{w}_i^t = \mathbf{w}_i^{t-1} + \mathbf{x}_i^t - \mathbf{x}_i^{t-L_i} , \qquad i = 1, \ldots, n \, , \; t = 1, \ldots, H \, , \tag{3.7}$$

$$\sum_{i=1}^{n} \mathbf{y}_{ij}^t = \mathbf{d}_j^t \, , \qquad j = 1, \ldots, m \, , \; t = 1, \ldots, H \, , \tag{3.8}$$

$$\sum_{j=1}^{m} \mathbf{y}_{ij}^t \le a_i (g_i^t + \mathbf{w}_i^{t-\Delta_i}) \, , \qquad i = 1, \ldots, n \, , \; t = 1, \ldots, H \, , \tag{3.9}$$

$$\mathbf{w}, \mathbf{x}, \mathbf{y} \ge 0 \, ,$$

- Random vector $\xi = (\xi^2, \ldots, \xi^H)$.
- the elements forming ξ^t are the demands, $(d_1^t, \ldots, d_k^t)$, and the cost vectors, (r^t, q^t) .
- In some cases, ξ^t can also contain the lifetimes L_i , the delay factors $\triangle_i$, and the availability factors a_i, depending on the elements deemed uncertain in the future.

An Example (Louveaux and Smeers [1988])

$$
\begin{aligned}
\min\ & 10x_1^1 + 7x_2^1 + 16x_3^1 + 6x_4^1 + \mathrm{E}_{\boldsymbol{\xi}}\Big[\sum_{j=1}^{3} \tau_j^2 (4\mathbf{y}_{1j}^2 + 4.5\mathbf{y}_{2j}^2 \\
& \qquad\qquad + 3.2\mathbf{y}_{3j}^2 + 5.5\mathbf{y}_{4j}^2)\Big] \\
\text{s. t. }\ & 10x_1^1 + 7x_2^1 + 16x_3^1 + 6x_4^1 \le 120\,, \\
& -x_i^1 + \sum_{j=1}^{3} \mathbf{y}_{ij}^2 \le 0\,, \qquad i = 1,\dots,4\,, \\
& \sum_{i=1}^{y} \mathbf{y}_{i1}^2 = \xi\,, \\
& \sum_{i=1}^{y} \mathbf{y}_{ij}^2 = d_j^2\,, \qquad j = 2,3\,, \\
& x_1^1 \ge 0\,, \quad x_2^1 \ge 0\,, \quad x_3^1 \ge 0\,, \quad x_4^1 \ge 0\,, \\
& \mathbf{y}_{ij}^2 \ge 0\,, \quad i = 1,\dots,4\,, \quad j = 1,2,3\,.
\end{aligned} \tag{3.11}
$$

- $n = 4$ technologies.
- $\triangle_i = 1$ period of construction delay.
- Full availabilities, $a \equiv 1$.
- No existing equipment, $g \equiv 0$.
- The only random variable is $d_1 = \xi$.
- $d_2 = 3$ and $d_3 = 2$.
- Investment costs are $r^1 = (10, 7, 16, 6)^T$.
- Production costs $q^2 = (4, 4.5, 3.2, 5.5)^T$.
- Load durations $\tau^2 = (10, 6, 1)^T$.
- Budget constraint: keep all investment below 120.

Design for Manufacturing Quality

- Assuming that ξ takes on the values 3 , 5 , and 7 with probabilities 0.3 , 0.4 , and 0.3 , respectively, an optimal stochastic programming solution: $x^{1*} = (2.67, 4.00, 3.33, 2.00)^T$ with an optimal objective value of 381.85 .
- Consider the expected value solution, which would substitute $\xi \equiv 5$in (3.11). An optimal solution here (again not unique) is $\bar{x}^1 = (0.00, 3.00, 5.00, 2.00)^T$. The objective value, if this single event occurs, is 365 .
- If we use this solution in the stochastic problem, with probability 0.3, demand cannot be met. This would yield an infinite value of the stochastic solution.

System's reliability to meet demand

$$\mathrm{P}[\sum_{i=1}^{n-1} a_i(g_i^t + w_i^t) \geq \sum_{j=1}^{m} \mathbf{d}_j^t] \geq \alpha , \qquad \forall t , \tag{3.12}$$

- $0 < \alpha \leq 1$.
- Chance or probabilistic constraint
- Deterministic equivalent:

$$\sum_{i=1}^{n-1} a_i(g_i^t + w_i^t) \geq (F^t)^{-1}(\alpha) , \qquad \forall t , \tag{3.13}$$

- F^t is the (assumed continuous) distribution function of $\sum_{j=1}^{m} d_j^t$ and $F^{-1}(\alpha)$ is the α -quantile of F.