Stochastic optimization

Fabian Bastin
fabian.bastin@cirrelt.ca
Université de Montréal – CIRRELT

Introduction

Consider the general deterministic program

$$\begin{aligned} &\min g_0(x) \\ \text{s.t. } &g_i(x) \le 0, i = 1, \ldots, m \\ &x \in X \subset \mathcal{F}^n. \end{aligned}$$

All the parameters are assumed to be perfectly known.
Realistic?

- measurement errors;
- uncertainties on the future;
- data unavailable;
- ...

Mathematical programming and stochastic programming

- Mathematical programming (optimization): typically: decision problem (where the meaning of the term "decision" is broad).
- Stochastic programming concerns decision under uncertainty, the uncertainty being represented by means of random parameters.

$$
\begin{aligned}
&\text{“}\min_{x \in X}\text{”}\, g_0(x, \xi) \\
&\text{s.t. } g_i(x, \xi) \le 0,\ i = 1, \dots, m,
\end{aligned}
$$

where ξ is a random vector. Meaning of "min"?

Assumption: we can represent the uncertainty by means of the (joint) probability distribution.

The farmer problem

Source: Birge et Louveaux, Section 1.1.

Scenarios approach

- One assumes that the random vector is finite. Each of the realization is a scenario.
- Even if the random vector is continuous, a discrete approximation is often useful.

A European farmer has 500 acres of land and cultivates wheat, corn and sugar beets.

At least 200T of wheat and 240T of corn are needed to feed his livestock. Any additional production can be sold, but in case of underproduction, he has to buy the complement, with a purchase cost 40% greater than the sale cost. The farmer can sold the sugar beets at $36T for the first 6000 tons, and $10T after, due to European quotas.

The farmer problem II

Culture	Wheat	Corn	Sugar beets
Average return (T)	2.5	3	20
Plantation cost ($/acre)	150	230	260
Selling price ($/T)	170	150	36 ($\leq$ 6000T), 10
Buying price ($/T)	238	210	-
Minimum required (T)	200	240	-

Notations:

- x_1, x_2, x_3: acres for wheat, corn, sugar beets;
- y_1, y_2: tons of bought wheat and corn;
- w_1, w_2: tons of sold wheat and corn;
- w_3, w_4: tons of sold sugar beets, at high price and at low price.

How to decide the surface to allocate to each plant?

The farmer problem: deterministic version

Linear program:

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

The farmer problem: deterministic solution

Total (expected) profit: \$118600. Details:

Culture	Wheat	Corn	Sugar beets
Surface (acres)	120	80	300
Production (T)	300	240	6000
Sales (T)	100	-	6000
Purchase (T)	-	-	-

The production is however dependant on the weather, and can increase or decrease by 20% to 25%.

In a very simplified setting, assume three possible cases: good year (for every plant, the production is 20% higher), average year, and bad year (for every plant, the production is 20% lower). The prices do not change.

The farmer problem: scenario solutions

New optimal solutions?

Good year. Total profit: \$167667.

Culture	Wheat	Corn	Sugar beets
Surface (acres)	183.33	66.67	250
Production (T)	550	240	6000
Sales (T)	350	-	6000
Purchases (T)	-	-	-

Bad year. Total profit: \$59950.

Culture	Wheat	Corn	Sugar beets
Surface (acres)	100	25	375
Production (T)	200	60	6000
Sales (T)	-	-	6000
Purchases (T)	-	180	-

The farmer problem: scenarios

The decisions considerably change with the weather conditions, but how to know them when deciding what to plant?
The decisions (x_1, x_2, x_3) have to be made now, but sales and purchases $(w_i, i = 1, \ldots, 4, y_j, j = 1, 2)$ depend on yields.

Scenarios.
Index $s = 1, 2, 3$, designing yields higher than the average, equal to the average, and lower than the average, respectively.
New variables w_{is} and y_{is}.

The farmer problem: extensive form

We now want to maximize the expected profit. Assuming that the 3 scenarios are equiprobable, we can form the new program

$$\min 150x_1 + 230x_2 + 260x_3 +$$

$$+ \sum_{s=1}^{3} \frac{1}{3}(238y_{1s} - 170w_{1s} + 210y_{2s} - 150w_{2s} - 36w_{3s} - 10w_{4s})$$

$$\text{t.q. } x_1 + x_2 + x_3 \leq 500;$$

$$3x_1 + y_{11} - w_{11} \geq 200; 2.5x_1 + y_{12} - w_{12} \geq 200; 2x_1 + y_{13} - w_{13} \geq 200;$$

$$3.6x_2 + y_{21} - w_{21} \geq 240; 3x_2 + y_{22} - w_{22} \geq 240;$$

$$2.4x_2 + y_{23} - w_{23} \geq 240;$$

$$w_{31} + w_{41} \leq 24x_3; w_{32} + w_{42} \leq 20x_3; w_{33} + w_{43} \leq 16x_3;$$

$$w_{31} \leq 6000; w_{32} \leq 6000; w_{33} \leq 6000;$$

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

$\rightarrow$ extensive form.

The farmer problem: stages

The seeding decisions are called first-stage decisions, while the sale and purchase decisions are called second-stage decisions.

Total profit: \$108390.

	Culture	Wheat	Corn	Sugar beets
First stage	Surface (acres)	170	80	250
$s = 1$	Productions (T)	510	288	6000
	Sales (T)	310	48	6000
	Purchases (T)	-	-	-
$s = 2$	Productions (T)	425	240	5000
	Sales (T)	225	-	5000
	Purchases (T)	-	-	-
$s = 3$	Productions (T)	340	192	4000
	Sales (T)	140	-	4000
	Purchases (T)	-	48	-

Observations

The optimal decision has changed!!!

Decision under perfect information: if the farmer could know the scenario in advance, or wait to observe the realization of the random variables (wait-and-see approach), the average annual profit would be \$115406. The difference with the optimal decision under uncertainty is called expected value of perfect information (EVPI): profit loss due to uncertainty.

Observations: value of the stochastic solution

- If the farmer only uses the average information, i.e. he replaces the random variables (r.v.) by their expectations, the average profit would be \$107240.
- Replacing the r.v. by their expectation leads to the expected value (EV) problem, delivering the expected value solution.
- Here, the expectation of scenarios is the average year, but in general, the expectation will not necessarily correspond to a pre-existent scenario.
- The expectation of the expected value (EEV) problem is obtained by computing the expected profit over the scenarios when the expected value solution is always used.
- Here, it leads to a loss of \$1150 with respect to the solution of the stochastic problem. This difference is known as value of the stochastic solution (VSS).

Example: the newsvendor problem

Source: Birge and Louveaux, Section 1.1.

- A newsvendor has to decide how many newspapers to buy in order to maximize his profit. However he does not know in advance how many newspapers he will be able to sell during a day (the demand).
- Each newspaper costs c, and can be sold at a price q.
- The newsvendor can turn back the unsold newspapers at the end of the day, and obtain a price r for each of them
- Knowing the probability distribution $F(t) = P(\omega \leq t)$, how many newspapers should the newsvendor buy in order to maximize his revenue?

The newsvendor problem (cont'd)

- With the previous definitions, the newsvendor would like to solve the following optimization problem:

$$\max_{x \geq 0} -cx + \mathcal{Q}(x),$$

- $\mathcal{Q}(x)$ is the expected sale amount if the newsvendor buy x newspapers:

$$\mathcal{Q}(x) = E_{\boldsymbol{\omega}}[Q(x, \boldsymbol{\omega})].$$

- Here $Q(x, \omega)$ is the amount of money obtained by the newsvendor if he buys x newspaper and the demand is ω.

The newsvendor problem (cont'd)

- As previously, we could construct an equivalent linear problem (presented in Birge and Louveaux).
- Can we simplify? It is easy to see that

$$Q(x,\omega) = \begin{cases} qx & \text{if } x \leq \omega, \\ q\omega + r(x-\omega) & \text{if } x \geq \omega. \end{cases}$$

Therefore,

$$\begin{aligned} \mathcal{Q}(x) = E_{\boldsymbol{\omega}}[Q(x,\boldsymbol{\omega})] &= \int_{-\infty}^{\infty} Q(x,\omega) dF(\omega) \\ &= \int_{-\infty}^{x} (q\omega + r(x-\omega)) dF(\omega) + \int_{x}^{\infty} qx dF(\omega). \end{aligned}$$

The newsvendor problem (cont'd)

Therefore, we have

$$\begin{aligned}\mathcal{Q}(x) &= (q-r)\int_{-\infty}^{x} \omega dF(\omega) + rx\int_{-\infty}^{x} dF(\omega) + qx\int_{x}^{\infty} dF(\omega)\\ &= (q-r)\int_{-\infty}^{x} \omega dF(\omega) + rxF(x) + qx(1-F(x))\\ &= (q-r)\left[\int_{-\infty}^{x} \omega dF(\omega) - xF(x)\right] + qx.\end{aligned}$$

Integration by parts

Assume that F satisfies $\lim_{t\to-\infty} tF(t) = 0$.

We can then integrate by parts to obtain:

$$\int_{-\infty}^{x} \omega dF(\omega) = \omega F(\omega)|_{-\infty}^{x} - \int_{-\infty}^{x} F(\omega)d\omega$$
$$= xF(x) - \int_{-\infty}^{x} F(\omega)d\omega.$$

Thus,

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

Solution of the second stage

Recall the initial problem...

$$\max_{x \geq 0} -cx + \mathcal{Q}(x),$$

We have to solve this problem. We will consider the associated optimality conditions.

Assuming $x \neq 0$, the solution of the second-stage is obtained by computing the zero of the objective gradient. As

$$\frac{d}{dx}\mathcal{Q}(x) = q - (q - r)F(x),$$

we have

$$-c + q - (q - r)F(x) = 0$$

The newsvendor problem (cont'd)

The solution x^* is therefore

$$x^* = F^{-1}\left(\frac{q-c}{q-r}\right).$$

Example: $c = 0.15$, $q = 0.25$, $r = 0.02$, $\omega \sim N(650, 80^2)$. Alors

$$x^* = N^{-1}_{(650,80^2)}(0.1/0.23).$$

Since $N(650, 80^2) \sim 80\Phi + 650$, where Φ is the distribution function of a $N(0, 1)$, it easy to show that

$$x^* = 80\Phi^{-1}(0.1/0.23) + 650 \approx 636.86.$$

In Julia, we can compute this value as

```
using Distributions
d = Normal(650,80)
quantile(d, 0.1/0.23)
```

Marginal revenue

Other interpretation, more intuitive: assume that the vendor has bought t journaux. What is the expected marginal revenue if he buys an additional newspaper? On an economical point of view, we would like this marginal revenue to be equal to 0.

The expected marginal revenue (MR) is

$$\begin{aligned} MR(t) &= -c + qP[\omega \geq t] + rP[\omega \leq t] \\ &= -c + q(1 - F(t)) + rF(t). \end{aligned}$$

If we set the marginal revenue to 0, we get

$$MR(t) = 0 \text{ iff } F(t) = \frac{q-c}{q-r},$$

and we recover the previous solution.

Formulation with recourse

More generally, we consider the (linear) program

$$\begin{aligned}\min\ & c^T x + E_\omega[q(\omega)^T y(\omega)]\\ \text{t.q. } & Ax = b,\\ & T(\omega)x + Wy(\omega) = h(\omega),\ \forall \omega \in \Omega\\ & x \in X,\\ & y(\omega) \in Y,\ \forall \omega.\end{aligned}$$

Fixed recourse: W does not change with the scenario.

How to decide over y?

Some definitions

$$\min_{x \in X | Ax = b} \left\{ c^T x + E_\omega \left[\min_{y \in Y} q(\omega)^T y | Wy = h(\omega) - T(\omega) x \right] \right\}.$$

- Second stage function, or recourse function $v : \Omega \times \mathbb{R}^m \to \mathbb{R}$:

$$v(\omega, z) \stackrel{def}{=} \min_{y \in Y} \{ q(\omega)^T y | Wy = z \};$$

- Expected value function, or recourse of minimum expectation $\mathcal{Q} : \mathbb{R}^n \to \mathbb{R}$:

$$\mathcal{Q}(x) = E_\omega [v(\omega, h(\omega) - T(\omega) x)].$$

  It describes the expected recourse cost, for any first-stage decision $x \in \mathbb{R}^n$.

The two-stage (linear) stochastic program

One can reformulate our optimization problem as

$$\min_{x \in X} \left\{ c^T x + \mathcal{Q}(x) \mid Ax = b \right\}.$$

It is a (nonlinear) optimization problem in $\mathbb{R}^n$.

In terms of y's:

$$\begin{aligned} \min_{x,\ y(\omega)} \ & E_\omega[c^T x + q(\omega)^T y(\omega)] & \\ \text{s.t. } & Ax = b & \text{first-stage constraints} \\ & T(\omega)x + Wy(\omega) = h(\omega),\ \forall \omega \in \Omega & \text{second-stage constraints} \\ & x \in X,\ y(\omega) \in Y. & \end{aligned}$$

Consider the (discrete) case where $\Omega = \{\omega_1,\ \omega_2, \ldots, \omega_S\} \subset \mathbb{R}^r$.

$$P(\omega = \omega_s) = p_s,\ s = 1, 2, \ldots \ldots, S$$
$$T_s = T(\omega),\ h_s = h(\omega)$$

Deterministic equivalent

$$\min_{x,y_1,\ldots,y_S} \; c^T x + p_1 q_1^T y_1 + p_2 q_2^T y_2 + \ldots p_S q_S^T y_S$$

t.q.

$$\begin{array}{llllll}
Ax & & & & & = b \\
T_1 x & +Wy_1 & & & & = h_1 \\
T_2 x & & +Wy_2 & & & = h_2 \\
\vdots & & & \ddots & & \\
T_S x & & & & +Wy_s & = h_s
\end{array}$$

$$x \in X, \; y_1 \in Y, \; y_2 \in Y, \ldots, y_s \in Y.$$

Deterministic equivalent (II)

- $y_s = y(\omega_s)$ is the recourse action to take if the scenario ω_s occurs.
- *Advantage*: it is a linear program.
- *Disadvantage*: it is a linear program of (very) high dimension:
  - $n + pS$ variables;
  - $m_1 + mS$ constraints.

  But the constraints matrix has a staircase structure.
  It is possible to exploit it (L-Shaped algorithm – Benders decomposition).

Large scale,… and?

Assume that we have r random variables ($\Omega \subset \mathbb{R}^r$).

- Consider the following problem (source: Linderoth). A Telecom company want to expand its network in order to meet an unknown (random) demand.
- There are 86 unknown demands. Each demand is independant and take a value in a set of 7 values. Consequently

$$S = |\Omega| = 7^{86} \approx 4.77 \times 10^{72}.$$

  …number of subatomic particles in the universe!
- It can be even worse…
  If Ω is not finite, but holds an infinite number of elements? It is especially true with continuous random variables. Our "deterministic equivalent" would have an infinite number of variables and constraints!
- We can solve an approximate problem, obtained by sampling over the random vector.

Decomposition methods

General principle: the nonlinear term in the objective, that is the recourse function $\mathcal{Q}(x)$, requires to solve all the linear second-stage programs.

Is it possible to avoid the repeated second-stage functions evaluations?

Idea: build a master problem in x, but compute the complete objective function (involving first- and second-stage decision) only as a subproblem.