Scenario generation for multistage models

Jitka Dupačová

Material for GAMS course on Stochastic Programming 2003

To apply the SP models we need to generate scenarios to represent the uncertainty in a *sensible* way, taking into account

- the goal of the model and its structure
- the available information
- computer facilities and available software

For financial applications

$\exists$ many models suitable for exploitation of financial time series and stochastic models with continuous time

$\exists$ various specific problems, e.g.

- various levels of information on various sources of uncertainty
- demographic data, economic factors
- correlations of stock market and bond returns
- problem specific additional random factors (prepayment rates, defaults ...)
- difficulties to model illiquid assets prices, derivatives, ...

How to build *sensible* scenarios?

THE GOAL of scenario generation procedure is **NOT** to get a good approximation of the probability distribution but of the optimal value and of the optimal decisions.

GOOD APPROXIMATION may mean precision (evaluate errors), robustness, etc. Connection with output analysis is evident.

No general recipe exists; it is necessary to exploit past data, experience, models, theory...
$\longrightarrow$ **HARD JOB**

$\exists$ specific requirements for **MULTISTAGE** stochastic programs

Multistage Stochastic Programs

T**-stage stochastic program** ... stochastic data process

$$\omega = \{\omega_1, \dots, \omega_{T-1}\} \quad \text{or} \quad \omega = \{\omega_1, \dots, \omega_T\}$$

whose realizations are (multidimensional) data trajectories and of a vector decision process

$$\boldsymbol{x} = \{\boldsymbol{x}_1, \dots \boldsymbol{x}_T\},$$

a measurable function of ω.

$t \in \{1, \dots, T\}$ ... the stage index

Then the sequence of decisions and observations is

$$\boldsymbol{x}_1, \omega_1, \boldsymbol{x}_2(\boldsymbol{x}_1, \omega_1), \omega_2, \dots, \dots, \boldsymbol{x}_T(\boldsymbol{x}_1, \omega_1, \dots, \omega_{T-1}).$$

The decision process is *nonanticipative* ...
i.e. decisions taken at any stage of the process do not depend on *future realizations* of random parameters or on future decisions.

The basic theoretical assumption reads:

Basic assumption. The probability distribution P of ω is known and independent of the decision $\boldsymbol{x}$.

To generate **SCENARIOS FOR MULTISTAGE SP**

means again to replace the initial probability distribution P of $\omega = (\omega_1, \omega_2, \dots, \omega_T)$ by a discrete distribution carried by a finite number of *atoms*

$$\omega^s = (\omega_1^s, \omega_2^s, \dots, \omega_T^s), s = 1, \dots, S$$

hence, to replace the conditional probability distributions $P(\omega_t|\omega^{t-1,\bullet})$ and the marginal distributions $\mathsf{P}(\omega_t)\,\forall t$ by discrete ones.

Denote $S_t,\ S_t(\omega^{t-1,\bullet})$ the finite supports of $P(\omega_t),\ P(\omega_t|\omega^{t-1,\bullet})$;

Scenarios must fulfil conditions

$\omega_\tau \in S_\tau(\omega^{\tau-1,\bullet})$ and

$$\{\omega : \omega = \omega^s, s = 1, \dots, S\} = \{\omega : \omega_\tau \in S_t(\omega^{t-1,\bullet}), \forall t > 1\}$$

Arc probabilities are $P(\omega_1)$ on S_1 and $P(\omega_t|\omega^{t-1,\bullet})\,\forall t > 1$

Path probabilities

$P(\omega^{t-1,\bullet}) = P(\omega_1) \prod_{\tau=2}^{t-1} P(\omega_\tau|\omega^{\tau-1,\bullet})$

and the probability of scenario $\omega^s = (\omega_1, \dots, \omega_T)$ is

$p_s = P(\omega^s) = P(\omega_1^s) \prod_{t=2}^{T} P(\omega_t^s|\omega_1^s, \dots, \omega_{t-1}^s)$

This information is organized in SCENARIO TREE:

THERE IS EXACTLY ONE ANCESTOR of $\omega_t^s\,\forall s, t$, but multiple descenadants are allowed.

Two special types of scenario tree

• INTERSTAGE INDEPENDENCE – For all stages $P(\omega_\tau|\omega^{\tau-1,\bullet}) = P(\omega_t)$

• For all stages, $S_t(\omega^{t-1,\bullet})$ are SINGLETONS

$$\longrightarrow$$

the tree collapses into FAN of individual scenarios $\omega^s = (\omega_1^s, \ldots, \omega_T^s)$ which occur with probabilities $p_s = P(\omega_1^s)$, $s = 1, \ldots, S$

$\longrightarrow$ MULTIPERIOD TWO-STAGE PROBLEM

In case of COEFFICIENTS defined by scenarios (e.g. prices obtained from interest rates or discount factors), the tree structure is required for coefficients.

The "arborescent" form of the *T-stage scenario-based stochastic linear program with recourse* reads:

Minimize

$$\boldsymbol{c}_1^\top \boldsymbol{x}_1 + \sum_{k_2=2}^{K_2} p_{k_2} \boldsymbol{c}_{k_2}^\top \boldsymbol{x}_{k_2} + \sum_{k_3=K_2+1}^{K_3} p_{k_3} \boldsymbol{c}_{k_3}^\top \boldsymbol{x}_{k_3} + \ldots + \sum_{k_T=K_{T-1}+1}^{K_T} p_{k_T} \boldsymbol{c}_{k_T}^\top \boldsymbol{x}_{k_T}$$

subject to the constraints

$$\begin{array}{lllllll}
\boldsymbol{A}_1 \boldsymbol{x}_1 & & & & & & = \boldsymbol{b}_1 \\
\boldsymbol{B}_{k_2} \boldsymbol{x}_1 & + & \boldsymbol{A}_{k_2} \boldsymbol{x}_{k_2} & & & & = \boldsymbol{b}_{k_2} \\
, & & & & & k_2 = & 2, \ldots, K_2 \\
 & & \boldsymbol{B}_{k_3} \boldsymbol{x}_{a(k_3)} & + \boldsymbol{A}_{k_3} \boldsymbol{x}_{k_3} & & & = \boldsymbol{b}_{k_3} \\
 & & & & & k_3 = & K_2 + 1, \ldots, K_3 \\
 & & \ddots & \ddots & & & \vdots \\
 & & & \boldsymbol{B}_{k_T} \boldsymbol{x}_{a(k_T)} & + \boldsymbol{A}_{k_T} \boldsymbol{x}_{k_T} & & = \boldsymbol{b}_{k_T} \\
 & & & & & k_T = & K_{T-1} + 1, \ldots, K_T
\end{array}$$

$$\boldsymbol{l}_{k_t} \le \boldsymbol{x}_{k_t} \le \boldsymbol{u}_{k_t}, \qquad k_t = K_{t-1} + 1, \ldots, K_t, \quad t = 1, \ldots, T.$$

We denote here by $a(k_t)$ the immediate ancestor of k_t, so that (with $K_1 = 1$) $a(k_2) = 1$, $k_2 = 2, \ldots, K_2$. The problem is based on the used scenarios, i.e., on the $S = K_T - K_{T-1}$ sequences of possible realizations $(\boldsymbol{c}_{k_t}, \boldsymbol{A}_{k_t}, \boldsymbol{B}_{k_t}, \boldsymbol{b}_{k_t}, \boldsymbol{l}_{k_t}, \boldsymbol{u}_{k_t})$ of coefficients in the objective function, in recourse matrices, transition matrices and right-hand sides in the constraints for all stages, and on probabilities

$p_{k_t} > 0 \forall\ k_t,\ \sum_{k_t=K_{t-1}+1}^{K_t} p_{k_t} = 1,\ t = 2, \ldots, T,$ of subsequences of these realizations connecting the root and the node k_t. These *path probabilities* identify the discrete distribution P.

The probabilities p_s of the individual scenarios ω^s, i.e., the path probabilities assigned to the terminal nodes, are obtained by multiplication of the (conditional) *arc* or *transition probabilities* related with the corresponding sequences of realizations. The nonanticipativity constraints are included in an implicit form.

The first-stage decisions consist of all decisions that have to be selected before further information is revealed whereas the second-stage decisions can adapt to this information, etc.

FROM A FAN SCENARIOS TO SCENARIO TREE

ASSUME:

∃ a given structure of the scenario tree, i.e.

- horizon
- time discretization
- stages
- branching scheme

∃ sufficiently many scenarios

Various ways to create scenario tree

- Ad hoc / expert cutting and pasting
- Conditional / importance sampling
- Clustering
- Moments fitting
- Techniques for scenario tree construction by minimization of distances of probability distributions
- Discretization schemes used instead of simulation or sampling

In addition, one should respect **problem specific requirements, cf. NO-ARBITRAGE**

DIFFERENCE BETWEEN TWO-STAGE MULTIPERIOD AND MULTISTAGE STOCHASTIC PROGRAMS

In multiperiod two-stage problems decisions at all time instances $t = 1, \ldots, T$ are made at once, no further information is expected;

Hedging against all considered **unrelated** scenarios of possible developments is assumed;

Except for the first stage no nonanticipativity constraints appear;

The input – no tree structure, just a fan of scenarios.

Why to use multistage formulation?

- Robustness, stability of solutions: similar subscenarios result in similar decisions even for $t > 1$;
- Stochastic specification (interstage dependence) is reflected;
- Number of nodes decreases.

Special case – interstage independence
Importance sampling, etc. reduces just in approximating the marginal probability distributions $P(\omega_t)\,\forall t$

$\exists$ extension for Markov dependence between stages, i.e., for

$$P(\omega_t|\omega_1,\ldots,\omega_{t-1}) = P(\omega_t|\omega_{t-1})$$

Disadvantage – rapid growth of number of scenarios in the scenario tree.

Main references & quotations therein:

Dupačová, Consigli, Wallace (2000), Scenarios for multistage stochastic programs, Annals of OR 100, 25–53.
Dupačová, Hurt, Štěpán (2002), Stochastic Modeling in Economics and Finance (Part II), Kluwer.