Approximations of Stochastic Programs. Scenario Tree Reduction and Construction

W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

www.mathematik.hu-berlin.de/~romisch

(J. Dupačová, N. Gröwe-Kuska, H. Heitsch)

GAMS Workshop, Heidelberg, Sept. 1-3, 2003

DFG research center Berlin
mathematics for key technologies

1 Introduction

Let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process defined on some probability space $(\Omega, \mathcal{F}, P)$ and with ξ_1 deterministic. The stochastic decision x_t at period t is assumed to depend only on $(\xi_1, \ldots, \xi_t)$ (nonanticipativity).

Typical financial and production planning model:

$$\min\{\mathbb{E}[\sum_{t=1}^{T} c_t(\xi_t, x_t)] \ : \ x_t \in X_t\,,\ x_t \text{ nonanticipative,}$$
$$A_{tt}(\xi_t)x_t + A_{t,t-1}(\xi_t)x_{t-1} \geq g_t(\xi_t)\}$$

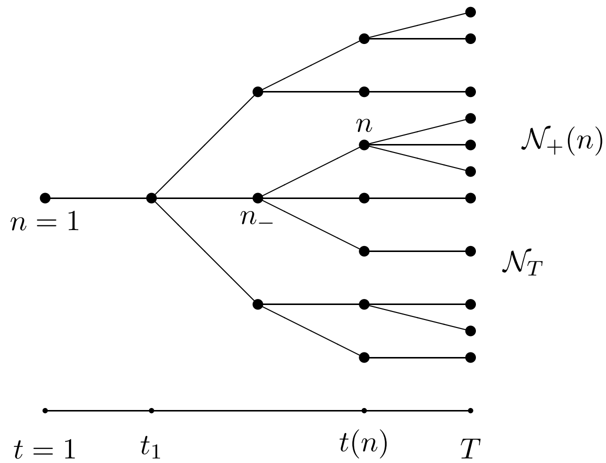
Alternative for the minimization of expected costs: Minimizing some risk measure $\mathbb{F}$ of the stochastic cost process $\{c_t(\xi_t, x_t)\}_{t=1}^T$ (risk management).

First step of its numerical solution:

Approximation of $\{\xi_t\}_{t=1}^T$ by finitely many scenarios with certain probabilities. Nonanticipativity leads to a scenario tree structure of the approximation.

2 Data process approximation by scenario trees

The data process $\xi = \{\xi_t\}_{t=1}^T$ is approximated by a process forming a scenario tree which is based on a finite set $\mathcal{N}$ of nodes.

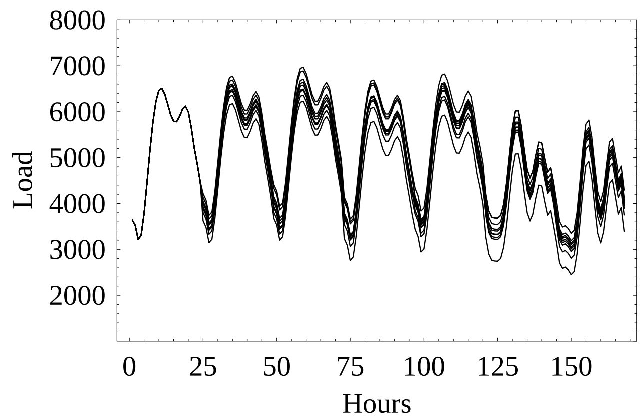


Scenario tree with $t_1 = 2$, $T = 5$, $|\mathcal{N}| = 23$ and 11 leaves

The root node $n = 1$ stands for period $t = 1$. Every other node n has a unique predecessor n_- and a set $\mathcal{N}_+(n)$ of successors. Let $\mathrm{path}(n)$ be the set $\{1, \ldots, n_-, n\}$ of nodes from the root to node n, $t(n) := |\mathrm{path}(n)|$ and $\mathcal{N}_T := \{n \in \mathcal{N} : \mathcal{N}_+(n) = \emptyset\}$ the set of leaves. A scenario corresponds to $\mathrm{path}(n)$ for some $n \in \mathcal{N}_T$. With the given scenario probabilities $\{\pi_n\}_{n \in \mathcal{N}_T}$, we define recursively node probabilities $\pi_n := \sum_{n_+ \in \mathcal{N}_+(n)} \pi_{n_+}$, $n \in \mathcal{N}$.

3 Generation of scenario trees

(i) Development of a stochastic model for the data process ξ
(parametric [e.g. time series model], nonparametric [e.g. resampling])



Scenarios for the weekly electrical load

and generation of simulation scenarios;

(ii) Construction of a scenario tree out of the stochastic model
or of the simulation scenarios;

(iii) optional scenario tree reduction.

Approaches for (ii):

(1) Barycentric scenario trees (conditional expectations w.r.t. a decomposition of the support into simplices) (Frauendorfer 96,...);

(2) Fitting of trees with prescribed structure to given moments (Hoyland/Wallace 01, Hoyland/Kaut/Wallace 03);

(3) Conditional sampling by (Quasi) Monte Carlo methods (QMC means low discrepancy sequences) (Morton 03, Koivu/Pennanen 02, 03);

(4) Clustering methods for bundling scenarios (Philpott/Craddock/Waterer 00);

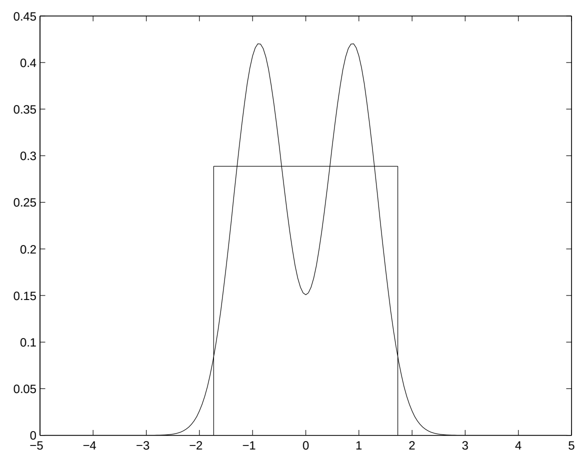
(5) Scenario tree construction based on optimal approximations w.r.t. certain probability metrics (Pflug 01, Hochreiter/Pflug 02, Gröwe-Kuska/Heitsch/Römisch 03).

Recent reference: Kaut/Wallace 03

Example: (Hochreiter/Pflug 02)
Let P denote the uniform distribution on $[-\sqrt{3}, \sqrt{3}]$ and $\tilde{P}$ be the distribution of $Z := c_1 Z_1 + c_2 Z_2$, where Z_1 is discrete with two equally probable scenarios -1 and 1, Z_2 is standard normal, i.e., $Z_2 \in N(0,1)$, and c_1 and c_2 are normalizing constants ($c_1 := \sqrt[4]{\frac{3}{5}}$, $c_2 := \sqrt{1 - \sqrt{\frac{3}{5}}}$). Then the first four (central) moments coincide

$$\int_{I\!R} \xi^i P(d\xi) = \int_{I\!R} \xi^i \tilde{P}(d\xi) = 0,\ 1,\ 0,\ \frac{9}{5}, \quad i = 1,\ 2,\ 3,\ 4.$$

However, the densities of P and $\tilde{P}$ have the following form



and, thus, are quite different.

4 Distances of probability distributions

Let P denote the probability distribution of the stochastic process $\{\xi_t\}_{t=1}^T$ with ξ_t in $\mathbb{R}^r$, i.e., P has support $\Xi \subseteq \mathbb{R}^{rT} = \mathbb{R}^s$.
The Kantorovich functional or transportation metric takes the form

$$\mu_c(P, Q) := \inf\{\int_{\Xi\times\Xi} c(\xi, \tilde{\xi})\eta(d\xi, d\tilde{\xi}) : \pi_1\eta = P,\ \pi_2\eta = Q\},$$

where $c : \Xi \times \Xi \to \mathbb{R}$ is a certain cost function and the minimum ist taken w.r.t. all probability measures η on $\Xi \times \Xi$ having (fixed) marginals P and Q.
Example: $c_p(\xi, \tilde{\xi}) := \max\{1, \|\xi - \xi_0\|^{p-1}, \|\tilde{\xi} - \xi_0\|^{p-1}\}\|\xi - \tilde{\xi}\|$
($p \geq 1$, $\xi_0 \in \Xi$ fixed)
We consider the following convex stochastic program

$$\min\{\int_{\Xi} f_0(x, \xi)P(d\xi) : x \in X\}$$

with a normal convex integrand f_0 and denote by
$v(P) := \inf_{x\in X} \int_{\Xi} f_0(x, \xi)P(d\xi)$ and $S(P) := \arg\min_{x\in X} \int_{\Xi} f_0(x, \xi)P(d\xi)$
its optimal value and solution set, respectively.

We choose c such that the property

$|f_0(x,\xi) - f_0(x,\tilde{\xi})| \leq L(\|x\|)c(\xi,\tilde{\xi}), \ \forall \xi, \tilde{\xi} \in \Xi, x \in X,$

holds with some function $L(\cdot)$ depending on $\|x\|$.
This means that c plays the role of a continuity modulus of the function $f_0(x,\cdot)$ from Ξ to $I\!R$ (for each $x \in X$).
Typically, f_0 is continuous and piecewise polynomial.

Theorem: (Stability)
Under weak conditions on the stochastic program the optimal values are Lipschitz continuous w.r.t. μ_c, i.e.,

$$|v(P) - v(Q)| \leq \hat{L}\mu_c(P,Q),$$

and the solution sets are upper semicontinuous. In particular, if $S(P) = \{\bar{x}\}$ any element of the approximate solution set $S(Q)$ is close to $\bar{x}$ if $\mu_c(P,Q)$ is small.
(Rachev/Römisch 02, Römisch 03)

Choice of $p \geq 1$ in $c = c_p$:

- two-stage with random right-hand side: $p = 1$.
- general two-stage with fixed recourse: $p = 2$.
- multi-stage with random right-hand side: $p = 1.^*$
- general multi-stage with T stages: $p = T.^*$

(* present conjecture valid under appropriate assumptions on the dependence structure;

not valid for mixed-integer models; in that case f_0 is piecewise continuous !)

Approach:
Select a probability metric a function $c : \Xi \times \Xi \to I\!R$ such that the underlying stochastic optimization model is stable w.r.t. μ_c.
Given P and a tolerance $\varepsilon > 0$, determine a scenario tree such that its probability distribution P_{tr} has the property

$$\mu_c(P, P_{tr}) \leq \varepsilon \, .$$

Distances of discrete distributions

P: scenarios ξ_i with probabilities p_i, $i = 1, \ldots, N$,
Q: scenarios $\tilde{\xi}_j$ with probabilities q_j, $j = 1, \ldots, M$.

Then

$$\mu_c(P,Q) = \sup\{\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{M} q_j v_j : u_i + v_j \leq c(\xi_i, \tilde{\xi}_j)\ \forall i,j\}$$

$$= \inf\{\sum_{i,j} \eta_{ij} c(\xi_i, \tilde{\xi}_j) : \eta_{ij} \geq 0, \sum_j \eta_{ij} = p_i, \sum_i \eta_{ij} = q_j\}$$

(*optimal value of linear transportation problems*)

(a) Distances of distributions can be computed by solving specific linear programs.

(b) The principle of optimal scenario generation can be formulated as a best approximation problem with respect to μ_c. However, it is nonconvex and difficult to solve.

(c) The best approximation problem simplifies considerably if the scenarios are taken from a specified finite set.

5 Scenario Reduction

We consider discrete distributions P with scenarios ξ_i and probabilities p_i, $i = 1, \ldots, N$, and Q having a subset of scenarios ξ_j, $j \in J \subset \{1, \ldots, N\}$, of P, but different probabilities q_j, $j \in J$.

Optimal reduction of a given scenario set J:
The best approximation of P with respect to μ_c by such a distribution Q exists and is denoted by $\bar{Q}$. It has the distance

$$D_J = \mu_c(P, Q) = \sum_{i \in J} p_i \min_{j \notin J} c(\xi_i, \xi_j)$$

and the probabilities $\bar{q}_j = p_j + \sum_{i \in J_j} p_i$, $\forall j \notin J$, where $J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} c(\xi_i, \xi_j)$, $\forall i \in J$, i.e., the optimal redistribution consists in adding the deleted scenario weight to that of some of the closest scenarios.

However, finding the optimal scenario set with a fixed number n of scenarios is a combinatorial optimization problem.

6 Fast reduction heuristics

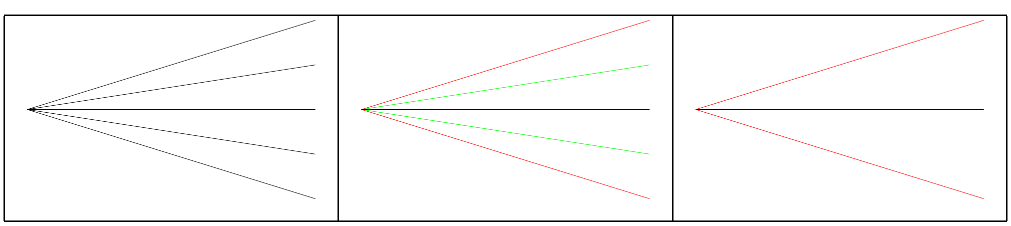
Starting point ($n = N - 1$): $\min_{l \in \{1,\dots,N\}} p_l \min_{j \neq l} c(\xi_l, \xi_j)$

Algorithm 1: (Simultaneous backward reduction)

Step [0]: Sorting of $\{c(\xi_j, \xi_k) : \forall j\}, \forall k$,
$J^{[0]} := \emptyset$.

Step [i]: $l_i \in \arg \min_{l \notin J^{[i-1]}} \sum_{k \in J^{[i-1]} \cup \{l\}} p_k \min_{j \notin J^{[i-1]} \cup \{l\}} c(\xi_k, \xi_j).$
$J^{[i]} := J^{[i-1]} \cup \{l_i\}$.

Step [N-n+1]: Optimal redistribution.



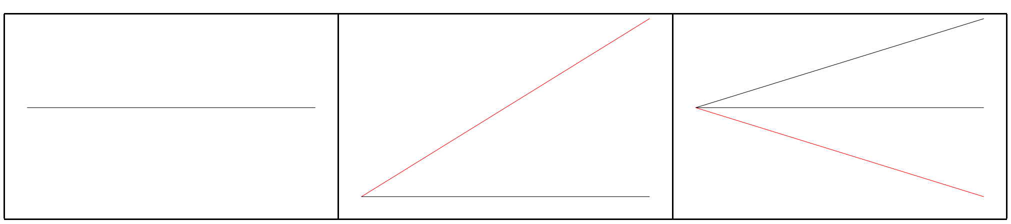
Starting point ($n = 1$): $\min_{u \in \{1,\dots,N\}} \sum_{k=1}^{N} p_k c(\xi_k, \xi_u)$

Algorithm 2: (Fast forward selection)

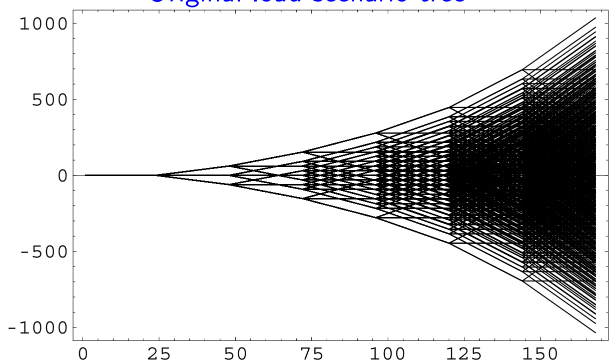
Step [0]: Compute $c(\xi_k, \xi_u)$, $k, u = 1, \dots, N$,
$J^{[0]} := \{1, \dots, N\}$.

Step [i]: $u_i \in \arg \min_{u \in J^{[i-1]}} \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k \min_{j \notin J^{[i-1]} \setminus \{u\}} c(\xi_k, \xi_j)$,
$J^{[i]} := J^{[i-1]} \setminus \{u_i\}$.

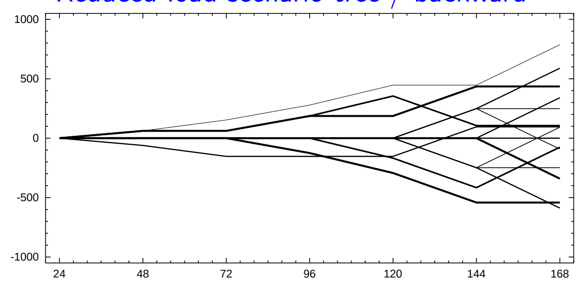
Step [n+1]: Optimal redistribution.



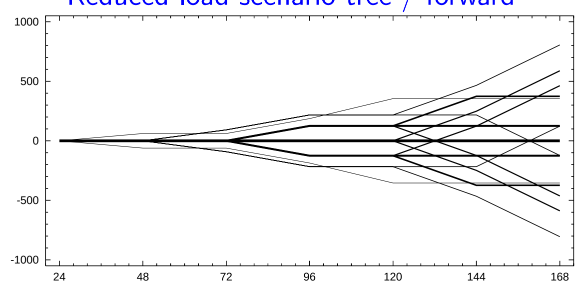
Original load scenario tree



Reduced load scenario tree / backward



Reduced load scenario tree / forward



Binary test scenario tree

Let a binary scenario tree have $N := 2^{T-1}$ scenarios $\xi_i = (\xi_i^1, \ldots, \xi_i^T)$, $i = 1, \ldots, N$, with equal probabilities $p_i = \frac{1}{N}$, $i = 1, \ldots, N$, and $\xi_1^1 = \ldots = \xi_N^1$ as its root node.
Such a scenario tree is called regular if, for each $t \in \{1, \ldots, T\}$, $\delta_1^t := -\delta^t$ and $\delta_2^t := \delta^t$ with $\delta^t \in \mathbb{R}_+$ and

$$\xi_i^t = \sum_{\tau=1}^{t} \delta_{i_\tau}^{\tau} \quad (t \in \{1, \ldots, T\})$$

where to each index $i = 1, \ldots, N$ there corresponds a T-tupel of indices $(i_1, \ldots, i_T) \in \{1, 2\}^T$.

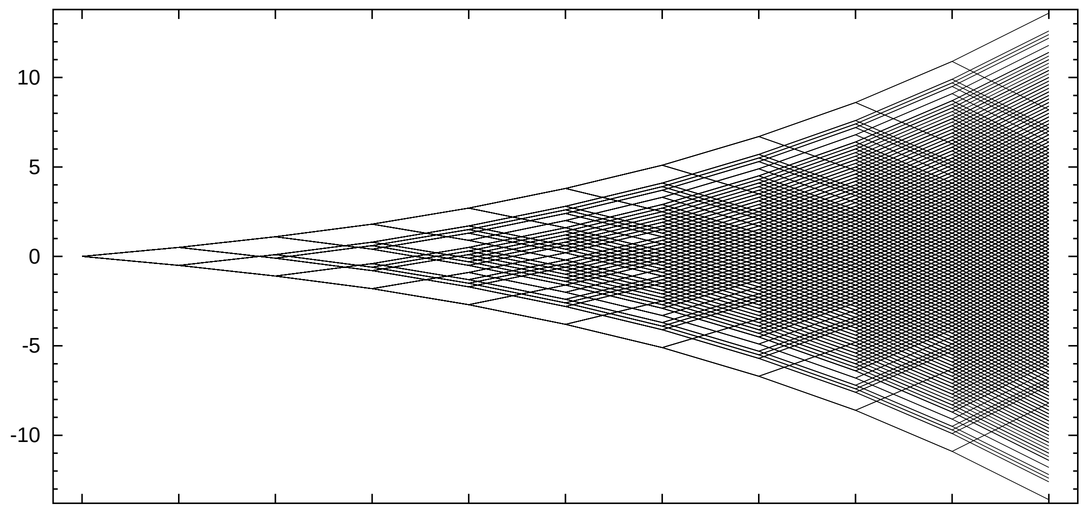
Proposition:

Let a regular binary scenario tree with $N = 2^{T-1}$ scenarios and $T \geq 4$ be given. Let $t_0 \in \arg\min_{2 \leq t \leq T} \delta^t$, $t_0 \leq T-2$ and $\max\{\delta^{t_0+1}, \delta^{t_0+2}\} \leq 2\delta^{t_0}$.
Then it holds for each $n \in \mathbb{N}$ with $\frac{N}{4} \leq n < N$:

$$D_n^{opt} := \min\{D_J : \#J = N - n\} = \frac{N-n}{N} 2\delta^{t_0}.$$

Here, c is defined by $c(\xi, \tilde{\xi}) := \|\xi - \tilde{\xi}\|_\infty$ ($\xi, \tilde{\xi} \in \Xi$).

Example: (regular binary scenario tree)
$T = 11$, $N = 2^{10} = 1024$,
$(\delta^1, \ldots, \delta^{11}) = (0, 0.5, 0.6, 0.7, 0.9, 1.1, 1.3, 1.6, 1.9, 2.3, 2.7)$,
$D_n^{opt} = \frac{N-n}{N}$ for each $\frac{N}{4} \leq n < N$.



Relative accuracy:

$$\mu_c^{rel}(P, Q) := \frac{\mu_c(P, Q)}{\mu_c(P, \delta_{\xi_{l_*}})}$$

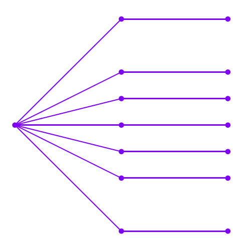
$\mu_c(P, \delta_{\xi_{l_*}}) = \min\{D_J : \#J = N - 1\}$ and $c(\cdot, \cdot) := \| \cdot - \cdot \|_\infty$.

Number n of Scenarios	**Backward of Scenario Sets**		**Simultaneous Backward**		**Fast Forward**		Minimal Distance
	ζ_c^{rel}	Time	ζ_c^{rel}	Time	ζ_c^{rel}	Time	
1	116.01 %	2 s	111.93 %	96 s	100.00 %	2 s	100.00 %
2	102.86 %	2 s	75.45 %	96 s	79.16 %	2 s	*
3	78.54 %	2 s	66.54 %	96 s	63.96 %	2 s	*
4	66.35 %	2 s	61.69 %	96 s	59.04 %	3 s	*
5	64.81 %	2 s	57.95 %	96 s	54.51 %	3 s	*
10	53.68 %	2 s	48.21 %	95 s	44.39 %	4 s	*
20	39.16 %	2 s	40.15 %	95 s	35.84 %	7 s	*
30	35.61 %	2 s	34.70 %	94 s	31.56 %	10 s	*
50	31.55 %	2 s	29.11 %	93 s	26.75 %	15 s	*
100	22.68 %	2 s	21.73 %	89 s	20.97 %	27 s	*
150	18.48 %	2 s	18.16 %	85 s	18.02 %	38 s	*
200	16.70 %	2 s	16.50 %	81 s	16.11 %	48 s	*
250	15.23 %	2 s	15.21 %	76 s	14.55 %	56 s	*
260	14.97 %	2 s	14.97 %	75 s	14.26 %	58 s	14.04 %
270	14.75 %	2 s	14.75 %	74 s	14.00 %	60 s	13.86 %
280	14.53 %	2 s	14.53 %	72 s	13.76 %	61 s	13.67 %
290	14.30 %	2 s	14.30 %	71 s	13.54 %	63 s	13.49 %
300	14.08 %	2 s	14.08 %	70 s	13.32 %	64 s	13.30 %
350	12.98 %	2 s	12.98 %	64 s	12.39 %	71 s	12.39 %
400	11.88 %	2 s	11.88 %	57 s	11.47 %	76 s	11.47 %
450	10.78 %	2 s	10.78 %	51 s	10.55 %	81 s	10.55 %
500	9.67 %	2 s	9.67 %	45 s	9.63 %	85 s	9.63 %
600	7.79 %	2 s	7.79 %	33 s	7.79 %	91 s	7.79 %
700	5.95 %	2 s	5.95 %	22 s	5.95 %	95 s	5.95 %
800	4.12 %	2 s	4.12 %	12 s	4.12 %	97 s	4.12 %

Computational results for the binary scenario tree

7 Constructing scenario trees from data scenarios

Let a fan of data scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i)$ with probabilities π^i, $i = 1, \ldots, N$, be given, i.e., all scenarios coincide at the starting point $t = 1$, i.e., $\xi_1^1 = \ldots = \xi_1^N =: \xi_1^*$. Hence, it has the form



$t = 1$ may be regarded as the root node of the scenario tree consisting of N scenarios (leaves).
Now, P is the (discrete) probability distribution of ξ. Let c be adapted to the underlying stochastic program containing P.

We describe an **algorithm** that may produce, for each $\varepsilon > 0$, a scenario tree with distribution P_ε, root node ξ_1^*, less nodes than P and

$$\mu_c(P, P_\varepsilon) < \varepsilon.$$

Recursive reduction algorithm:

Let $\varepsilon_t > 0$, $t = 1, \ldots, T$, be given such that $\sum_{t=1}^{T} \varepsilon_t \leq \varepsilon$, set $t := T$, $I_{T+1} := \{1, \ldots, N\}$, $\pi^i_{T+1} := \pi^i$ and $P_{T+1} := P$.

For $\mathbf{t = T, \ldots, 2}$:

Step t: Determine an index set $I_t \subseteq I_{t+1}$ such that

$$\mu_{c_t}(P_t, P_{t+1}) < \varepsilon_t \, ,$$

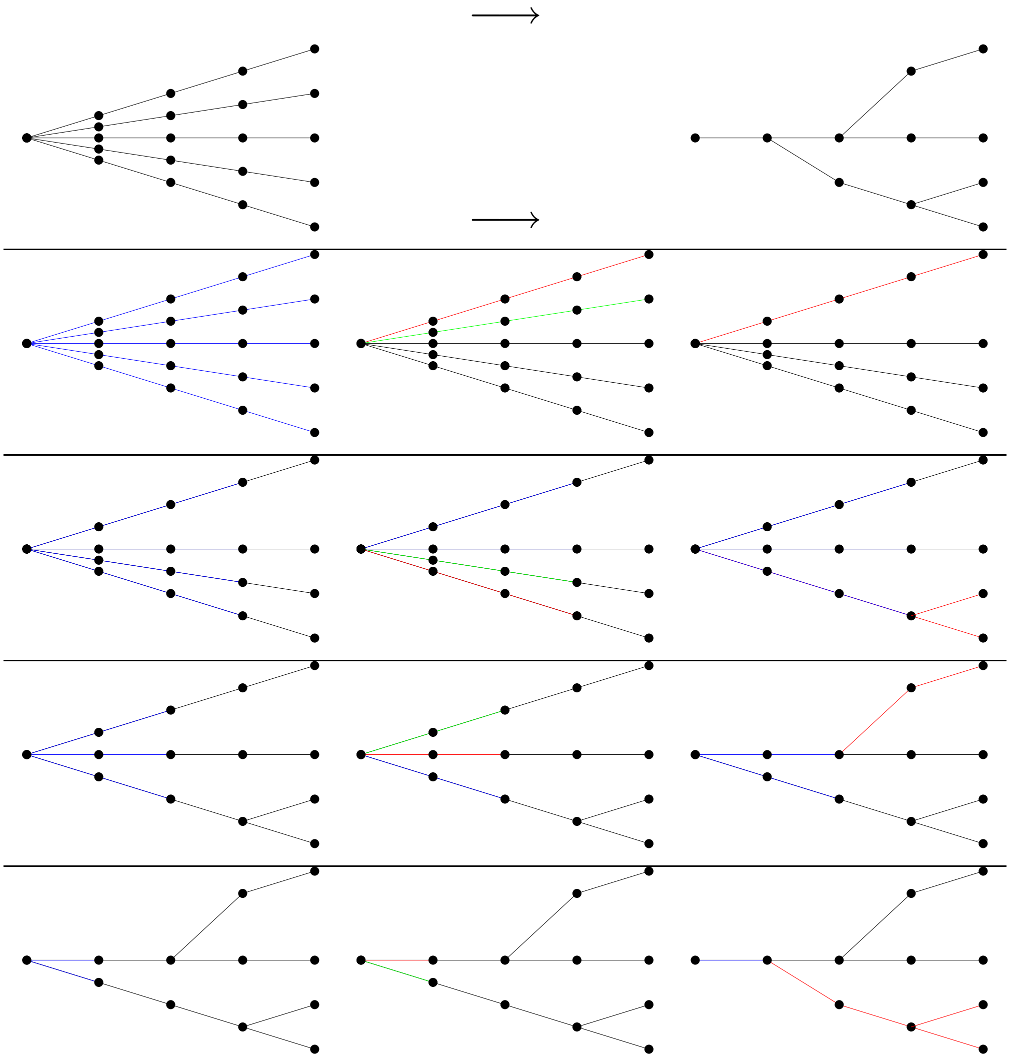
where $\{\xi^i\}_{i \in I_t}$ is the support of P_t and c_t is defined by $c_t(\xi, \tilde{\xi}) := c((\xi_1, \ldots, \xi_t, 0, \ldots, 0), (\tilde{\xi}_1, \ldots, \tilde{\xi}_t, 0, \ldots, 0))$;

(scenario reduction w.r.t. the time horizon $[1, t]$)

Step 1: Determine a probability measure P_ε such that its marginal distributions $P_\varepsilon \Pi_t^{-1}$ are $\delta_{\xi_1^*}$ for $t = 1$ and

$$P_\varepsilon \Pi_t^{-1} = \sum_{i \in I_t} \pi_t^i \delta_{\xi_t^i} \quad \text{and} \quad \pi_t^i := \pi_{t+1}^i + \sum_{j \in J_{t,i}} \pi_{t+1}^j \, ,$$

where $J_{t,i} := \{j \in I_{t+1} \setminus I_t : i_t(j) = i\}$, $i_t(j) \in \arg\min_{i \in I_t} c_t(\xi^j, \xi^i)\}$ are the index sets according to the redistribution rule.



Blue: compute c-distances of scenarios; delete the green scenario & add its weight to the red one

Application:

ξ is the bivariate weekly data process having the components

a) electrical load,

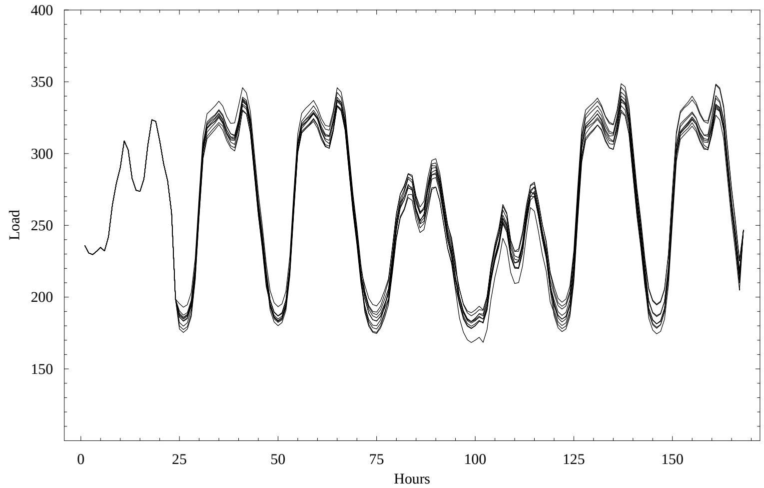
b) hourly electricity spot prices (at EEX).

Data scenarios are obtained from a stochastic model calibrated to the historical load data of a (small) German power utility and historical price data of the European Energy Exchange (EEX) at Leipzig.

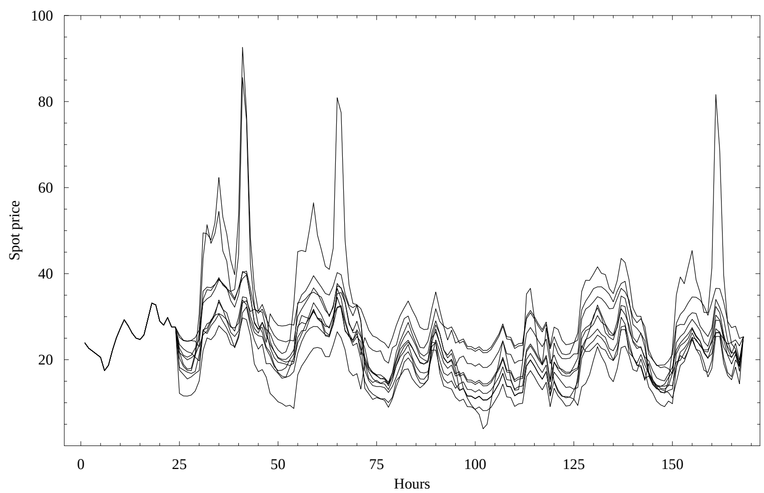
We choose $N = 50$, $T = 7$, $\varepsilon = 0.05$, $\varepsilon_t = \frac{\varepsilon}{T}$, and arrive at a tree with 4608 nodes (instead of 8400 nodes of the original fan).

t	hours	$\lvert I_t \rvert$
1	$1 \cdots 24$	1
2	$25 \cdots 48$	12
3	$49 \cdots 72$	23
4	$73 \cdots 96$	31
5	$97 \cdots 120$	37
6	$121 \cdots 144$	42
7	$145 \cdots 168$	46

Scenario tree for the electrical load



Scenario tree for hourly spot prices



8 GAMS/SCENRED

- GAMS/SCENRED introduced to GAMS Distribution 20.6 (May 2002)

- SCENRED is a collection of C++ routines for the optimal reduction of scenarios or scenario trees

- GAMS/SCENRED provides the link from GAMS programs to the scenario reduction algorithms. The reduced problems can then be solved by a deterministic optimization algorithm provided by GAMS.

- SCENRED contains three reduction algorithms:
  - FAST BACKWARD method
  - Mix of FAST BACKWARD/FORWARD methods
  - Mix of FAST BACKWARD/BACKWARD methods
  Automatic selection (best expected performance w.r.t. running time)

Details: www.scenred.de, www.scenred.com