Generating and handling scenarios in stochastic programming

W. Römisch

Humboldt-University Berlin
Department of Mathematics

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

IMA-Workshop *Computing with Uncertainty*,

Minneapolis, October 18-22, 2010

DFG Research Center MATHEON
Mathematics for key technologies

Contents

SP and approximation issues

We consider a stochastic program of the form

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

where $X \subseteq \mathbb{R}^m$ is a constraint set, P a probability distribution on $\Xi \subseteq \mathbb{R}^d$, and $f = \Phi(\cdot, x)$ is a decision-dependent integrand.

Any approach to solving such models computationally requires to replace the integral by a quadrature rule

$$Q_{n,d}(f) = \sum_{i=1}^{n} w_i f(\xi^i),$$

with weights $w_i \in \mathbb{R}$ and scenarios $\xi^i \in \Xi$, $i = 1, \ldots, n$.

If the natural condition $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$ is satisfied, $Q_{n,d}(f)$ allows the interpretation as integral with respect to the discrete probability measure Q_n having scenarios ξ^i with probabilities w_i, $i = 1, \ldots, n$.

Example: Linear two-stage stochastic programs

We consider two-stage linear stochastic programs with random right-hand sides:

$$\min\left\{\langle c,x\rangle+\int_{\Xi}\varphi(h(\xi)-Tx)P(d\xi):x\in X\right\}$$

where $c\in\mathbb{R}^m$, X is a polyhedral subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^d$, T a (r,m)-matrix, $h(\cdot)$ an affine mapping from $\mathbb{R}^d$ to $\mathbb{R}^r$, P a Borel probability measure on Ξ and

$$\begin{aligned}\varphi(t)&=\inf\{\langle q,y\rangle:Wy=t,y\geq 0\}\\&=\sup\{\langle t,z\rangle:W^\top z\leq q\}=\sup_{z\in\mathcal{D}}\langle t,z\rangle\,,\end{aligned}$$

where $q\in\mathbb{R}^{\bar{m}}$, W a $(r,\bar{m})$-matrix (having rank r) and t varies in the polyhedral cone $W(\mathbb{R}^{\bar{m}})$. If $\mathcal{D}\neq\emptyset$ there exist vertices v^j of $\mathcal{D}$ and polyhedral cones $\mathcal{K}_j$, $j=1,\ldots,\ell$, decomposing $\operatorname{dom}\varphi$ such that $\varphi(t)=\langle v^j,t\rangle$, $\forall t\in\mathcal{K}_j$, and $\varphi(t)=\max_{j=1,\ldots,\ell}\langle v^j,t\rangle$. Hence

$$\Phi(\xi,x)=\langle c,x\rangle+\max_{j=1,\ldots,\ell}\langle v^j,h(\xi)-Tx\rangle$$

Assumption: P has a density ρ w.r.t. λ^d.

Now, we set $\mathcal{F} = \{\Phi(\cdot, x)\rho(\cdot) : x \in X\}$ and assume that the set $\mathcal{F}$ is a bounded subset of some linear normed space F_d with norm $\|\cdot\|_d$ and unit ball $\mathbb{B}_d = \{f \in F_d : \|f\|_d \leq 1\}$.

The absolute error of the quadrature rule $Q_{n,d}$ is

$$e(Q_{n,d}) = \sup_{f \in \mathbb{B}_d} \left| \int_\Xi f(\xi)d\xi - \sum_{i=1}^n w_i f(\xi^i) \right|$$

and the approximation criterion is based on the relative error and a given tolerance $\varepsilon > 0$, namely, it consists in finding the smallest number $n_{\min}(\varepsilon, Q_{n,d}) \in \mathbb{N}$ such that

$$e(Q_{n,d}) \leq \varepsilon e(Q_{0,d}),$$

holds, where $Q_{0,d}(f) = 0$ and, hence, $e(Q_{0,d}) = \|I_d\|$ with

$$I_d(f) = \int_\Xi f(\xi)d\xi.$$

Alternatively, we look for a suitable set $\mathcal{F}$ of functions such that $\{C\Phi(\cdot, x) : x \in X\} \subseteq \mathcal{F}$ for some constant $C > 0$ and, hence,

$$e(Q_{n,d}) \leq \frac{1}{C} \sup_{f \in \mathcal{F}} \left| \int_\Xi f(\xi) P(d\xi) - \int_\Xi f(\xi) Q_n(d\xi) \right| = D(P, Q_n),$$

and that D is a metric distance between probability distributions.

Example: Fortet-Mourier metric (of order $r \geq 1$)

$$\zeta_r(P, Q) := \sup \left| \int_\Xi f(\xi)(P - Q)(d\xi) : f \in \mathcal{F}_r(\Xi) \right|,$$

where

$$\mathcal{F}_r(\Xi) := \{f : \Xi \mapsto \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq c_r(\xi, \tilde{\xi}), \forall \xi, \tilde{\xi} \in \Xi\},$$

$$c_r(\xi, \tilde{\xi}) := \max\{1, \|\xi\|^{r-1}, \|\tilde{\xi}\|^{r-1}\} \|\xi - \tilde{\xi}\| \quad (\xi, \tilde{\xi} \in \Xi).$$

The behavior of $e(Q_{n,d})$ with respect to $n \in \mathbb{N}$ and of $n_{\min}(\varepsilon, Q_{n,d})$ with respect to ε is of considerable interest. In both cases the dependence on the dimension d of P is often crucial, too.

The behavior of both quantities depends heavily on the normed space F_d and the set $\mathcal{F}$, respectively.
It is desirable that an estimate of the form

$$n_{\min}(\varepsilon, Q_{n,d}) \leq C\, d^q \varepsilon^{-p} \qquad (\text{'tractability'})$$

is valid for some constants $q \geq 0$, $C, p > 0$ and for every $\varepsilon \in (0, 1)$. Of course, $q = 0$ is highly desirable for high-dimensional problems.

Proposition: (Stability)
Let the set X be compact. Then there exists $L > 0$ such that

$$\left| \inf_{x \in X} \int_{\Xi} \Phi(\xi, x)\rho(\xi)d\xi - \inf_{x \in X} \sum_{i=1}^{n} w_i \Phi(\xi^i, x)\rho(\xi^i) \right| \leq L\, e(Q_{n,d}).$$

The solution set mapping is outer semicontinuous at P.

Examples of normed spaces F_d relevant in SP:

(a) The Banach space $F_d = \mathrm{Lip}(\mathbb{R}^d)$ of Lipschitz continuous functions equipped with the norm

$$\|f\|_d = |f(0)| + \sup_{\xi \neq \tilde{\xi}} \frac{|f(\xi) - f(\tilde{\xi})|}{\|\xi - \tilde{\xi}\|}.$$

The best possible convergence rate is $e(Q_{n,d}) = O(n^{-\frac{1}{d}})$.
It is attained for $w_i = \frac{1}{n}$ and certain ξ^i, $i = 1, \ldots, n$, if P has finite moments of order $1 + \delta$ for some $\delta > 0$. (Graf-Luschgy 00)

(b) **Assumption:** $\Xi = [0,1]^d$ (attainable by suitable transformations).
We consider the Banach space $F_d = \mathrm{BV}_{\mathrm{HK}}([0,1]^d)$ of functions having bounded variation in the sense of Hardy and Krause equipped with the norm $\|f\|_d = |f(0)| + V_{\mathrm{HK}}(f)$.
Then for $w_i = \frac{1}{n}$, $i = 1, \ldots, n$, there exist $\xi^i \in [0,1]^d$, $i \in \mathbb{N}$, such that the convergence rate is

$$e(Q_{n,d}) = O\left(\frac{(\log n)^{d-1}}{n}\right).$$

(c) The tensor product Sobolev space

$$F_{d,\gamma} = \mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1])$$

of real functions on $[0,1]^d$ having first order mixed weak derivatives with the (weighted) norm

$$\|f\|_{d,\gamma} = \left(\sum_{u \subset D} \gamma_u^{-1} \int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|}}{\partial \xi^u} f(\xi^u, 1^{-u}) \right|^2 d\xi^u \right)^{\frac{1}{2}},$$

where $D = \{1, \ldots, d\}$, $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_d > 0$, $\gamma_\emptyset = 1$ and

$$\gamma_u = \prod_{j \in u} \gamma_j \quad (u \subseteq D).$$

Note that any $f \in \mathcal{W}_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d)$ is of bounded variation in the sense of Hardy and Krause.

For n prime, $w_i = \frac{1}{n}$, and a suitable choice of (γ_j), points $\xi^i \in [0,1]^d$, $i = 1, \ldots, n$ can be constructed such that

$$e(Q_{n,d}) \leq C(\delta) n^{-1+\delta} \|I_d\|$$

for some $C(\delta) > 0$ (not depending on d) and all $0 < \delta \leq \frac{1}{2}$ (Sloan, Woźniakowski 98, Kuo 03).

Scenario generation methods

We will discuss the following four scenario generation methods for stochastic programs *without nonanticipativity constraints*:

(a) Monte Carlo sampling from the underlying probability distribution P on $\mathbb{R}^d$ (Shapiro 03).

(b) Optimal quantization of probability distributions (Pflug-Pichler 10).

(c) Quasi-Monte Carlo methods (Koivu-Pennanen 05, Homem-de-Mello 06).

(d) Quadrature rules based on sparse grids (Chen-Mehrotra 08).

Monte Carlo sampling

Monte Carlo methods are based on drawing independent identically distributed (iid) Ξ-valued random samples $\xi^1(\cdot), \ldots, \xi^n(\cdot), \ldots$ (defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$) from an underlying probability distribution P (on Ξ) such that

$$Q_{n,d}(\omega)(f) = \frac{1}{n} \sum_{i=1}^{n} f(\xi^i(\omega)),$$

i.e., $Q_{n,d}(\cdot)$ is a random functional, and it holds

$$\lim_{n\to\infty} Q_{n,d}(\omega)(f) = \int_{\Xi} f(\xi) P(d\xi) = \mathbb{E}(f) \quad \mathbb{P}\text{-almost surely}$$

for every real continuous and bounded function f on Ξ.

If P has finite moment of order $r \geq 1$, the error estimate

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(\xi^i(\omega)) - \mathbb{E}(f)\right|^r\right) \leq \frac{\mathbb{E}\left((f - \mathbb{E}(f))^r\right)}{n^{r-1}}$$

is valid. Hence, the mean square convergence rate is

$$\|Q_{n,d}(\omega)(f) - \mathbb{E}(f)\|_{L_2} = \sigma(f)n^{-\frac{1}{2}},$$

where $\sigma^2(f) = \mathbb{E}\left((f - \mathbb{E}(f))^2\right)$.
The latter holds without any assumption on f except $\sigma(f) < \infty$.

Advantages:
(i) MC sampling works *for (almost) all integrands*.
(ii) The machinery of probability theory is available.
(iii) The convergence *rate does not depend on* d.

Deficiencies: (Niederreiter 92)
(i) There exist 'only' *probabilistic error bounds*.
(ii) Possible regularity of the integrand *does not improve* the rate.
(iii) Generating (independent) random samples is *difficult*.

Practically, iid samples are approximately obtained by pseudo random number generators as uniform samples in $[0, 1]^d$ and later transformed to more general sets Ξ and distributions P.

Survey: L'Ecuyer 94.

Classical generators for pseudo random numbers are based on linear congruential methods. As the parameters of this method, we choose a large $M \in \mathbb{N}$ (*modulus*), a *multiplier* $a \in \mathbb{N}$ with $1 \leq a < M$ and $gcd(a, M) = 1$, and $c \in Z_M = \{0, 1, \ldots, M-1\}$. Starting with $y_0 \in Z_M$ a sequence is generated by

$$y_n \equiv a y_{n-1} + c \quad \text{mod } M \qquad (n \in \mathbb{N})$$

and the linear congruential pseudo random numbers are

$$\xi^n = \frac{y_n}{M} \in [0, 1).$$

Excellent pseudo random number generator: Mersenne Twister (Matsumoto-Nishimura 98).

Use only pseudo random number generators having passed a series of statistical tests, e.g., uniformity test, serial correlation test, serial test, coarse lattice structure test etc.

Optimal quantization of probability measures

Let D be a distance of probability measures on $\mathbb{R}^d$ such that the underlying stochastic program behaves stable w.r.t. D (Römisch 03).

Examples:

(a) Fortet-Mourier metric ζ_r of order r,

(b) L_r-minimal metric ℓ_r (or Wasserstein metric), i.e.

$$\ell_r(P,Q) = \inf\{(\mathbb{E}(\|\xi - \eta\|^r))^{\frac{1}{r}} : \mathcal{L}(\xi) = P,\, \mathcal{L}(\eta) = Q\}$$

Let P be a given probability distribution on $\mathbb{R}^d$. We are looking for a discrete probability measure Q_n with support

$$\mathrm{supp}(Q_n) = \{\xi^1, \ldots, \xi^n\} \quad \text{and} \quad Q_n(\{\xi^i\}) = \frac{1}{n},\, i = 1, \ldots, n,$$

such that it is the best approximation to P with respect to D, i.e.,

$$D(P, Q_n) = \min\{D(P,Q) : |\mathrm{supp}(Q)| = n, Q \text{ is uniform}\}.$$

Existence of best approximations, called optimal quantizers, and their convergence rates are well known for ℓ_r (Graf-Luschgy 00).

In general, however, the function

$$\Psi_D(\xi^1, \dots, \xi^n) := D\Big(P, \frac{1}{n}\sum_{i=1}^{n}\delta_{\xi^i}\Big)$$

$$\Psi_{\ell_r}(\xi^1, \dots, \xi^n) = \Big(\int_{\mathbb{R}^d} \min_{i=1,\dots,n} \|\xi - \xi^i\|^r P(d\xi)\Big)^{\frac{1}{r}}$$

is nonconvex and nondifferentiable on $\mathbb{R}^{dn}$.
Hence, the global minimization of Ψ_D is not an easy task.

Algorithmic procedures for minimizing Ψ_{ℓ_r} globally may be based on stochastic gradient algorithms, stochastic approximation methods and stochastic branch-and-bound techniques (e.g. Pflug 01, Hochreiter-Pflug 07, Pagés 97, Pagés et al 04).

However, asymptotically optimal quantizers can be determined explicitly in several cases (Pflug-Pichler 10).

Quasi-Monte Carlo methods

The basic idea of Quasi-Monte Carlo (QMC) methods is to replace random samples in Monte Carlo methods by deterministic points that are uniformly distributed in $[0,1]^d$. The latter property may be defined in terms of the so-called star-discrepancy of $\xi^1, \dots, \xi^n$

$$D_n^*(\xi^1, \dots, \xi^n) := \sup_{\xi \in [0,1]^d} \left| \lambda^d([0,\xi)) - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[0,\xi)}(\xi^i) \right|,$$

by calling a sequence $(\xi^i)_{i \in \mathbb{N}}$ uniformly distributed in $[0,1]^d$

$$D_n^*(\xi^1, \dots, \xi^n) \to 0 \quad \text{for} \quad n \to \infty \,.$$

A **classical result** due to Roth 54 states

$$D_n^*(\xi^1, \dots, \xi^n) \geq B_d \frac{(\log n)^{\frac{d-1}{2}}}{n}$$

for some constant B_d and all sequences (ξ^i) in $[0,1]^d$.

Classical convergence results:

Theorem: (Proinov 88)
If the real function f is continuous on $[0,1]^d$, then there exists $C > 0$ such that

$$|Q_{n,d}(f) - I_d(f)| \leq C\omega_f\left(D_n^*(\xi^1,\ldots,\xi^n)^{\frac{1}{d}}\right),$$

where $\omega_f(\delta) = \sup\{|f(\xi) - f(\tilde{\xi})| : \|\xi - \tilde{\xi}\| \leq \delta,\ \xi, \tilde{\xi} \in [0,1]^d\}$ is the modulus of continuity of f.

Theorem: (Koksma-Hlawka 61)
If f is of bounded variation in the sense of Hardy and Krause, it holds

$$|I_d(f) - Q_{n,d}(f)| \leq V_{\mathrm{HK}}(f) D_n^*(\xi^1,\ldots,\xi^n)\,.$$

for any $n \in \mathbb{N}$ and any $\xi^1,\ldots,\xi^n \in [0,1]^d$.

There exist sequences (ξ^i) in $[0,1]^d$ such that

$$D_n^*(\xi^1,\ldots,\xi^n) = O(n^{-1}(\log n)^{d-1}).$$

First general construction: (Sobol 69, Niederreiter 87)

Elementary subintervals E in base b:

$$E = \prod_{j=1}^{d} \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right),$$

with $a_i, d_i \in \mathbb{Z}_+, 0 \leq a_i < d_i, i = 1, \ldots, d$.

Let $m, t \in \mathbb{Z}_+$, $m > t$.
A set of b^m points in $[0,1]^d$ is a (t, m, d)-net in base b if every elementary subinterval E in base b with $\lambda^d(E) = b^{t-m}$ contains b^t points.

A sequence (ξ^i) in $[0,1]^d$ is a (t, d)-sequence in base b if, for all integers $k \in \mathbb{Z}_+$ and $m > t$, the set

$$\{\xi^i : kb^m \leq i < (k+1)b^m\}$$

is a (t, m, d)-net in base b.

Proposition: $(0, d)$-sequences exist if $d \leq b$.

Theorem:
The star-discrepancy of a $(0, m, d)$-net $\{\xi_i\}$ in base b satisfies

$$D_n^*(\xi^i) \leq A_d(b) \frac{(\log n)^{d-1}}{n} + O\left(\frac{(\log n)^{d-2}}{n}\right).$$

Special cases: Sobol, Faure and Niederreiter sequences.

Second general construction: (Korobov 59, Sloan-Joe 94)
Let $g \in \mathbb{Z}^d$ and consider the lattice points

$$\left\{\xi^i = \left\{\frac{i}{n} g\right\} : i = 1, \ldots, n\right\},$$

where $\{z\}$ is defined componentwise and for $z \in \mathbb{R}_+$ it is the *fractional part* of z, i.e., $\{z\} = z - \lfloor z \rfloor \in [0, 1)$.
Randomly shifted lattice points with a uniform $\triangle$:

$$\left\{\xi^i = \left\{\frac{i}{n} g + \triangle\right\} : i = 1, \ldots, n\right\},$$

There is a component-by-component construction algorithm for g such that for some constant $C(\delta)$ and all $0 < \delta \leq \frac{1}{2}$

$$e(Q_{n,d}) \leq C(\delta) n^{-1+\delta} \|I_d\| \quad \text{(Sloan, Kuo 03).}$$

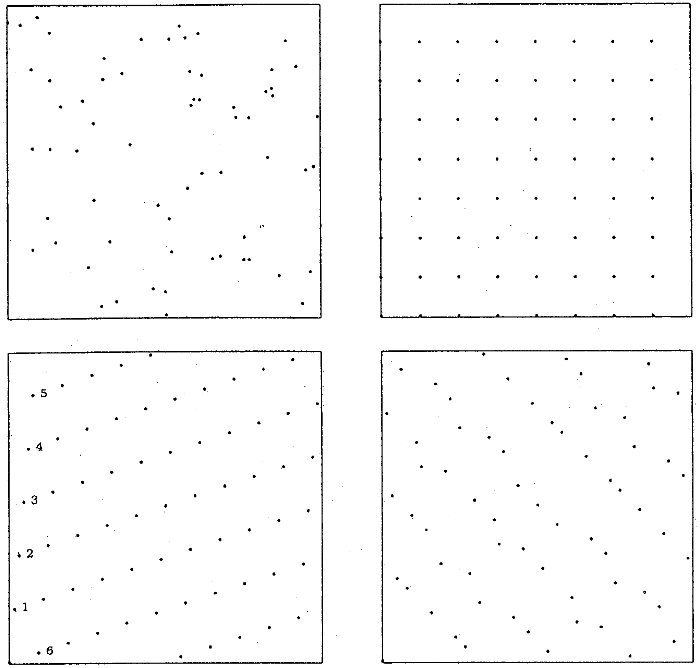


Fig. 5.3 Four different point sets with $n = 64$: random (top left), rectangular grid (top right), Korobov lattice (bottom left), and Sobol' (bottom right).

Quadrature rules with sparse grids

Again we consider the unit cube $[0,1]^d$ in $\mathbb{R}^d$. Let nested sets of grids in $[0,1]$ be given, i.e.,

$$\Xi^i = \{\xi_1^i, \ldots, \xi_{m_i}^i\} \subset \Xi^{i+1} \subset [0,1] \quad (i \in \mathbb{N}),$$

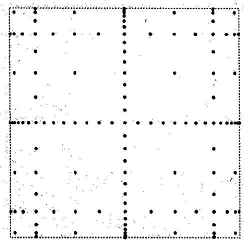
for example, the dyadic grid

$$\Xi^i = \left\{ \frac{j}{2^i} : j = 0, 1, \ldots, 2^i \right\}.$$

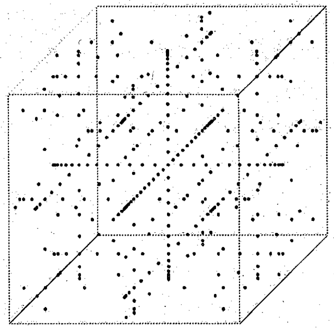
Then the point set suggested by Smolyak

$$H(n,d) := \bigcup_{\sum_{j=1}^d i_j = n} \Xi^{i_1} \times \cdots \times \Xi^{i_d} \qquad (n \in \mathbb{N})$$

is called a sparse grid in $[0,1]^d$. In case of dyadic grids in $[0,1]$ the set $H(n,d)$ consists of all d-dimensional dyadic grids with product of mesh size given by $\frac{1}{2^n}$.



(a) $d = 2$



(b) $d = 3$

The corresponding tensor product quadrature rule for $n \geq d$ on $[0,1]^d$ with respect to the Lebesgue measure λ^d is of the form

$$Q_{n,d}(f) = \sum_{n-d+1 \leq |\mathbf{i}| \leq n} (-1)^{n-|\mathbf{i}|} \binom{d-1}{n-|\mathbf{i}|} \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(\xi_{j_1}^{i_1}, \ldots, \xi_{j_d}^{i_d}) \prod_{l=1}^{d} a_{j_l}^{i_l},$$

where $|\mathbf{i}| = \sum_{j=1}^{d} i_j$ and the coefficients a_j^i ($j = 1, \ldots, m_i$, $i = 1, \ldots, d$) are weights of one-dimensional quadrature rules.

Even if the one-dimensional weights are positive, some of the weights w_i may become negative. Hence, an interpretation as discrete probability measure is no longer possible.

Theorem: (Bungartz-Griebel 04)
If f belongs to $F_d = W_2^{(r,\ldots,r)}([0,1]^d)$, it holds

$$\left| \int_{[0,1]^d} f(\xi) d\xi - \sum_{i=1}^{n} w_i f(\xi^i) \right| \leq C_{r,d} \|f\|_d \frac{(\log n)^{(d-1)(r+1)}}{n^r} .$$

Example (continued)

Proposition: (Owen 05)
Let $d \geq 3$, $b_i \in \mathbb{R}$, $i = 0, 1, \ldots, d$, and we consider for $\xi \in [0,1]^d$

$$f(\xi) = \max\{\langle b, \xi \rangle - b_0, 0\}.$$

If $\{\xi \in [0,1]^d : \langle b, \xi \rangle = b_0\}$ has positive $(d-1)$-dimensional volume and none of $b_1, \ldots, b_d$ is zero, it holds $V_{\mathrm{HK}}(f) = \infty$.

Conclusion: Typical integrands in two-stage linear stochastic programming are not of bounded variation in general.

Alternatives ? (open problem)
(a) Smoothing of stochastic programs ?
(b) Arguing via smoother ANOVA decomposition terms of f and small effective dimension ?

Scenario reduction

Assume that a two-stage stochastic program behaves stable with respect to a Fortet-Mourier metric ζ_r for some $r \geq 1$ (Römisch-Wets 07).

Proposition: (Rachev-Rüschendorf 98)
If Ξ is bounded, ζ_r may be reformulated as transportation problem

$$\zeta_r(P,Q) = \inf\left\{\int_{\Xi\times\Xi} \hat{c}_r(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \pi_1\eta = P, \pi_2\eta = Q\right\},$$

where $\hat{c}_r$ is a metric (reduced cost) with $\hat{c}_r \leq c_r$ and given by

$$\hat{c}_r(\xi,\tilde{\xi}) := \inf\left\{\sum_{i=1}^{n-1} c_r(\xi_{l_i},\xi_{l_{i+1}}) : n \in \mathbb{N}, \xi_{l_i} \in \Xi, \xi_{l_1} = \xi, \xi_{l_n} = \tilde{\xi}\right\}.$$

We consider discrete distributions P with scenarios ξ^i and probabilities p_i, $i = 1,\ldots,N$, and Q being supported by a given subset of scenarios ξ^j, $j \notin J \subset \{1,\ldots,N\}$, of P.

Best approximation given a scenario set J:

The best approximation of P with respect to ζ_r by such a distribution Q exists and is denoted by Q^*. It has the distance

$$D_J := \zeta_r(P, Q^*) = \min_{Q} \zeta_r(P, Q) = \sum_{i \in J} p_i \min_{j \notin J} \hat{c}_r(\xi^i, \xi^j)$$

and the probabilities $q_j^* = p_j + \sum_{i \in J_j} p_i$, $\forall j \notin J$, where

$J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} \hat{c}_r(\xi^i, \xi^j)$, $\forall i \in J$ (optimal redistribution) (Dupačová-Gröwe-Römisch 03).

For mixed-integer two-stage stochastic programs the relevant distance is a polyhedral discrepancy. In that case, the new weights have to be determined by linear programming (Henrion-Küchler-Römisch 08, 09).

Determining the optimal index set J with prescribed cardinality $N-n$ is a clustering problem, thus, a combinatorial optimization problem of n-median type:

$$\min\{D_J : J \subset \{1, ..., N\}, |J| = N-n\}$$

Hence, the problem of finding the optimal set J for deleting scenarios is $\mathcal{NP}$-hard and polynomial time algorithms are not available in general.

$\longrightarrow$ Search for fast heuristics starting from $n=1$ or $n=N-1$.

Fast reduction heuristics

Starting point ($n = N-1$): $\min_{l \in \{1,\dots,N\}} p_l \min_{j \neq l} \hat{c}_r(\xi^l, \xi^j)$

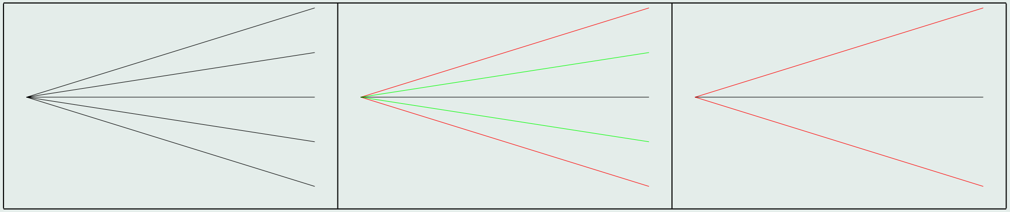
Algorithm 1: (Backward reduction)

Step [0]: $J^{[0]} := \emptyset\,.$

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

$J^{[i]} := J^{[i-1]} \cup \{l_i\}\,.$

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



Home Page
Title Page
Contents
Page 28 of 45
Go Back
Full Screen
Close
Quit

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

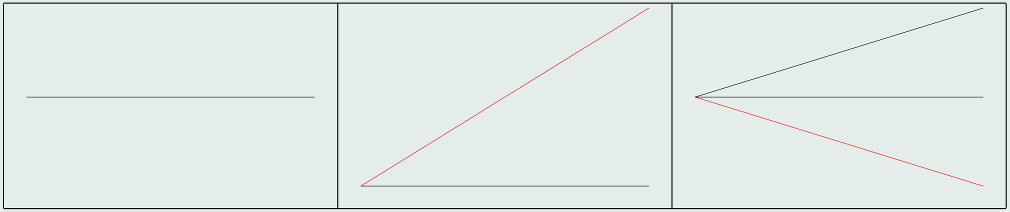
Algorithm 2: (Forward selection)

Step [0]: $J^{[0]} := \{1,\dots,N\}$.

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

$$J^{[i]} := J^{[i-1]} \setminus \{u_i\}\,.$$

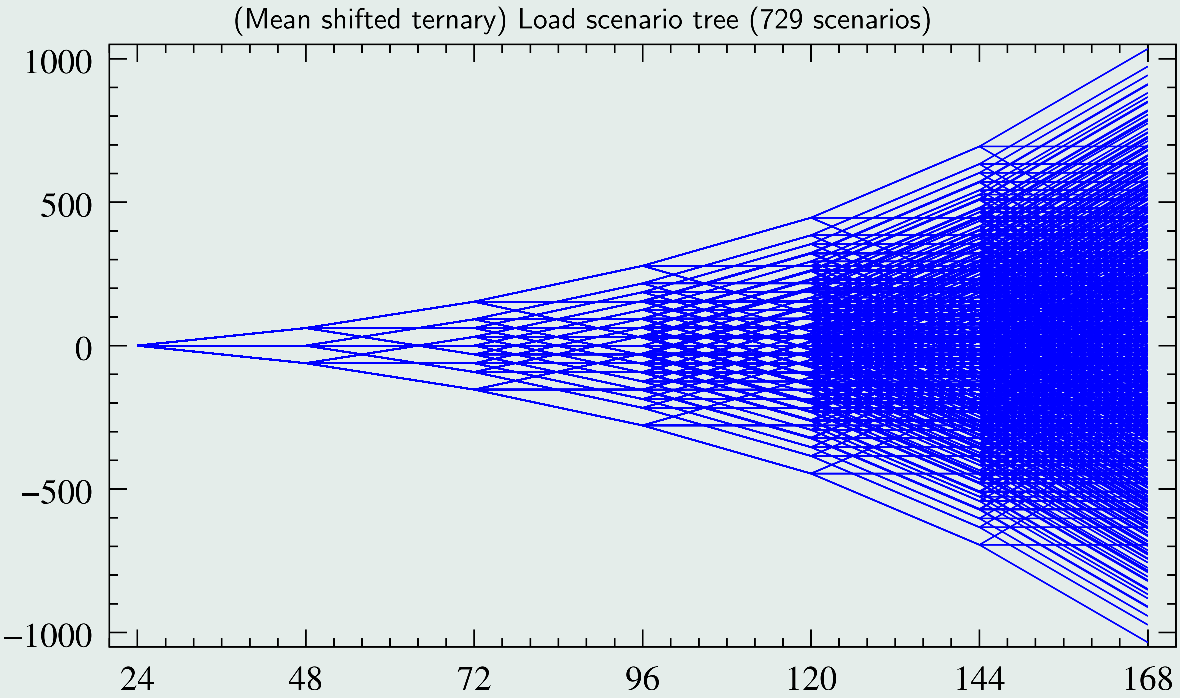
Step [n+1]: Optimal redistribution.



(Heitsch-Römisch 03, 07)

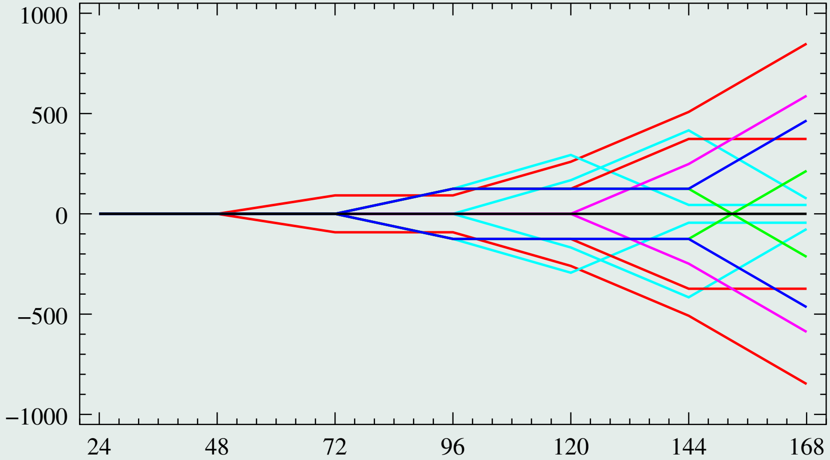
Home Page
Title Page
Contents
Page 29 of 45
Go Back
Full Screen
Close
Quit

Example: (Electrical load scenario tree)

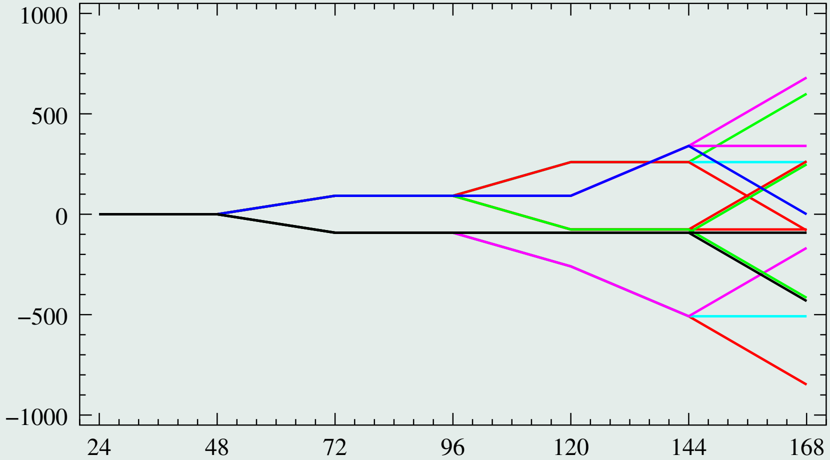


<Start Animation>

Reduced load scenario tree obtained by the forward selection method (15 scenarios)



Reduced load scenario tree obtained by the backward reduction method (12 scenarios)



Generation of scenario trees

In multistage stochastic programs the decisions x have to satisfy the additional information constraint that x_t is measurable with respect to $\mathcal{F}_t = \sigma(\xi_\tau, \tau = 1, \ldots, t)$, $t = 1, \ldots, T$. The increase of the σ-fields $\mathcal{F}_t$ w.r.t. t is reflected by approximating the underlying stochastic process $\xi = (\xi_t)_{t=1}^T$ by scenarios forming a scenario tree.

Some recent approaches:

(1) Bound-based approximation methods: Kuhn 05, Casey-Sen 05.

(2) Monte Carlo-based schemes: Shapiro 03, 06.

(3) Quasi-Monte Carlo methods: Pennanen 06, 09 .

(4) Moment-matching principle: Høyland-Kaut-Wallace 03.

(5) Optimal quantization: Pagés et al. 99.

(6) Stability-based approximations: Hochreiter-Pflug 07, Mirkov-Pflug 07, Pflug-Pichler 10, Heitsch-Römisch 05, 09.

Survey: Dupačová-Consigli-Wallace 00

Theoretical basis of (6):

Quantitative stability results for multi-stage stochastic programs.

(Heitsch-Römisch-Strugarek 06; Mirkov-Pflug 07, Pflug 09)

Scenario tree generation: (Heitsch-Römisch 09)

(i) Generate a number of scenarios by one of the methods discussed earlier.

(ii) Construction of a scenario tree out of these scenarios by recursive scenario reduction and bundling over time such that the optimal value stays within a prescribed tolerance.

Implementation: GAMS-SCENRED 2.0 (developed by H. Heitsch)

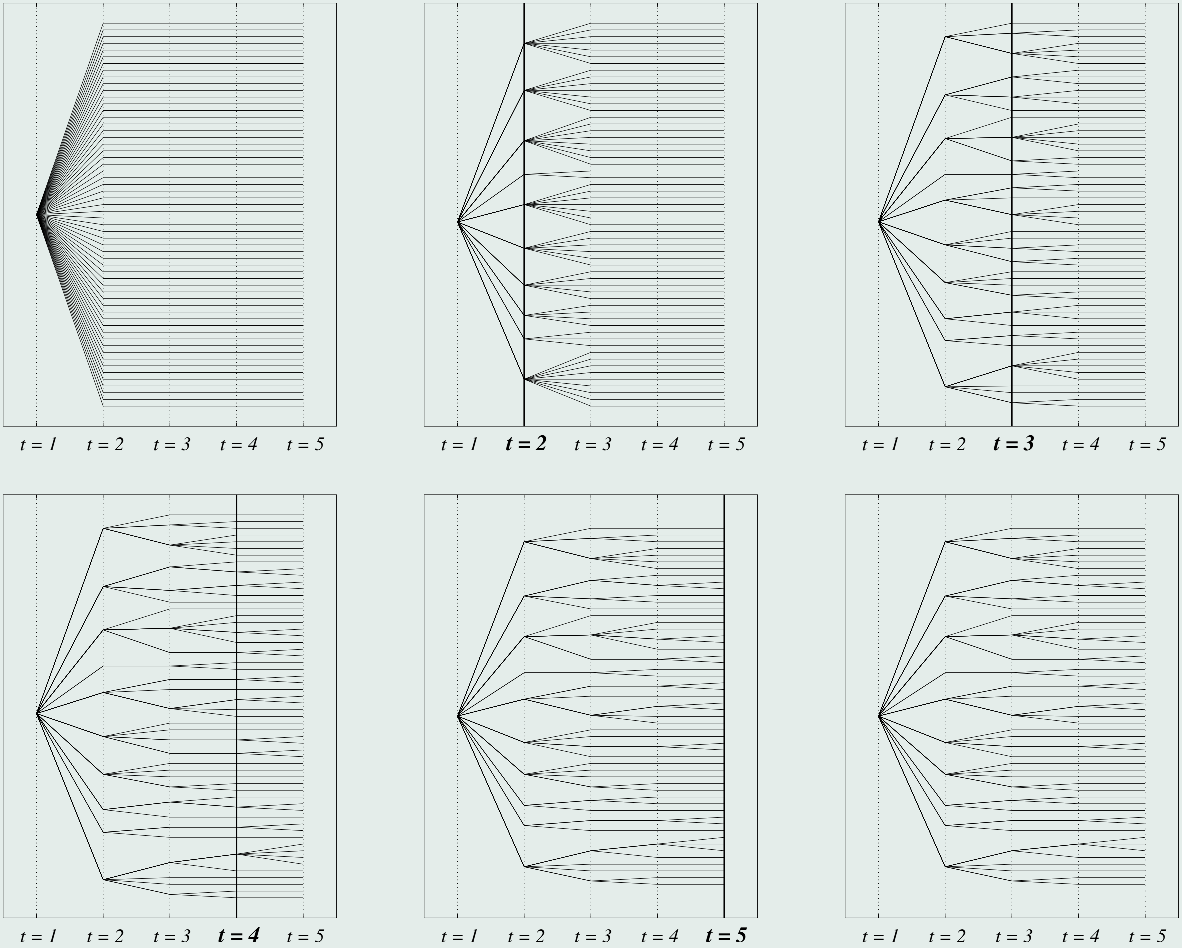


Illustration of the forward tree generation for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>

Appendix A: Functions of bounded variation

Let $D = \{1, \ldots, d\}$ and we consider subsets u of D with cardinality $|u|$. By $-u$ we mean $-u = D \setminus u$.
The expression ξ^u denotes the $|u|$-tuple of the components ξ_j, $j \in u$, of $\xi \in \mathbb{R}^d$. For example, we write

$$f(\xi) = f(\xi^u, \xi^{-u}).$$

We set the d-fold alternating sum of f over the d-dimensional interval $[a, b]$ as

$$\triangle(f; a, b) = \sum_{u \subseteq D} (-1)^{|u|} f(a^u, b^{-u}).$$

Furthermore, we set for any $v \subseteq u$

$$\triangle_u(f; a, b) = \sum_{v \subseteq u} (-1)^{|v|} f(a^v, b^{-v}).$$

Let G_j denote finite grids in $[a_j, b_j)$, $a_j < b_j$, $j = 1, \ldots, d$, and $G = \times_{i=1}^{d} G_j$ a grid in $[a, b) = \times_{i=1}^{d}[a_j, b_j)$. For $g \in G$ let $g^+ = (g_1^+, \ldots, g_d^+)$, where g_j^+ is the successor of g_j in $G_j \cup \{b_j\}$. Then the variation of f over G is

$$V_G(f) = \sum_{g \in G} \left|\triangle(f; g, g^+)\right| .$$

If $\mathcal{G}$ denotes the set of all finite grids in $[a, b)$, the variation of f on $[a, b]$ in the sense of Vitali is

$$V_{[a,b]}(f) = \sup_{G \in \mathcal{G}} V_G(f) .$$

The variation of f on $[a, b]$ in the sense of Hardy and Krause is

$$V_{\mathrm{HK}}(f; a, b) = \sum_{u \subset D} V_{[a^{-u}, b^{-u}]}(f(\xi^{-u}, b^u)) .$$

Bounded variation on $[a, b]$ in the sense of Hardy and Krause then means $V_{\mathrm{HK}}(f; a, b) < \infty$.

Appendix B: Decomposition of multivariate L_2 functions

Idea: If f isn't of bounded variation or smooth, decompositions of f may be used, where only some of the terms are relevant and, hopefully, are of bounded variation or smooth.

ANOVA-decomposition of f:

$$f = \sum_{u \subseteq D} f_u \,,$$

where $f_\emptyset = I_d(f) = I_D(f)$ and recursively

$$f_u = I_{-u}(f) + \sum_{v \subseteq u} (-1)^{|u|-|v|} I_{u-v}(I_{-u}(f)) \,,$$

where I_{-u} means integration with respect to ξ_j, $j \in D \setminus u$. Hence, f_u is essentially as smooth as $I_{-u}(f)$ and does not depend on ξ^{-u}.

Proposition:
The functions $\{f_u\}_{u \subseteq D}$ are orthogonal in $L_2([0,1]^d)$.

We set $\sigma^2(f) = \|f - I_d(f)\|_{L_2}^2$ and have

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_d(f))^2 = \sum_{\emptyset \neq u \subseteq D} \|f_u\|_{L_2}^2 .$$

The truncation dimension d_t of f is the smallest $d_t \in \mathbb{N}$ such that

$$\sum_{u \subseteq \{1,\ldots,d_t\}} \|f_u\|_{L_2}^2 \geq p\sigma^2(f) \quad \text{(where } p \in (0,1) \text{ is close to } 1\text{).}$$

Then it holds $\quad f \approx \sum_{u \subseteq \{1,\ldots,d_t\}} f_u \quad$ (in L_2).

The f_u can be smoother than f under certain conditions

(Griebel-Kuo-Sloan 10).

Problem:

How to determine the truncation dimension in SP ?

(Drew and Homem-de-Mello 06).

Conclusions

- We presented a framework for approximating stochastic programs suitable for a number of scenario generation methods.
- We gave a survey of approaches for scenario generation.
- We outlined that a competitive theoretical basis is still open for applying Quasi-Monte Carlo and sparse grid methods in stochastic programming.
- We discussed strategies for scenario reduction and scenario tree generation.

References

Monographs:

H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

C. Lemieux: *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2009.

J. Dick, F. Pillichshammer: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, 2010 (to appear).

I. H. Sloan, S. Joe: *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford 1994.

S. Graf, H. Luschgy: *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.

H.-J. Bungartz, M. Griebel: Sparse Grids, *Acta Numerica* (2004), 1–123.

Monte Carlo:

P. L'Ecuyer: Uniform random number generation, *Annals of Operations Research* 53 (1994), 77–120.

M. Matsumoto, T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.

A. Genz: Numerical computation of multivariate normal probabilities, J. Comp. Graph. Stat. 1 (1992), 141–149.

A. Shapiro: Monte Carlo sampling methods, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 353–425.

A. Shapiro: Inference of statistical bounds for multistage stochastic programming problems, *Mathematical Methods of Operations Research* 58 (2003), 57–68.

Quantization:

W. Römisch: Stability of stochastic programming problems, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.

G. Pagès, H. Pham, J. Printems: Optimal quantization methods and applications to numerical problems in finance, in *Handbook of Computational and Numerical Methods in Finance* (S.T. Rachev Ed.), Birkhäuser, Boston 2004, 253–297.

G. Ch. Pflug, A. Pichler: Scenario generation for stochastic optimization problems, in: Stochastic Optimization Methods in Finance and Energy (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.) (to appear).

G. Ch. Pflug: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* 89 (2001), 251–271.

Quasi-Monte Carlo:

T. Pennanen, M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.

T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling, *SIAM Journal on Optimization* 19 (2008), 524-551.

S. S. Drew, T. Homem-de-Mello: Quasi-Monte Carlo strategies for stochastic optimization, *Proceedings of the 2006 Winter Simulation Conference*, IEEE, 2006, 774–782.

I. M. Sobol': *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).

M. Götz: Discrepancy and the error in integration, *Monatshefte für Mathematik* 136 (2002), 99-121.

I. H. Sloan, H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.

F. Y. Kuo, I. H. Sloan: Lifting the curse of dimensionality, *Notices of the AMS* 52 (2005), 1320–1328.

A. B. Owen: Multidimensional variation for Quasi-Monte Carlo, in J. Fan, G. Li (Eds.), International Conference on Statistics, World Scientific Publ., 2005, 49–74.

X. Wang, K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

M. Griebel, F. Y. Kuo, I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 2010 (in press).

I. H. Sloan, F. Y. Kuo, S. Joe: On the step-by-step construction of Quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, *Mathematics of Computation* 71 (2002), 1609-1640.

F. Y. Kuo: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces, *Journal of Complexity* 19 (2003), 301-320.

Sparse Grids:

S. A. Smolyak: Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet Math. Dokl.* 4 (1963), 240–243.

E. Novak, K. Ritter: High dimensional integration of smooth functions over cubes, *Numerische Mathematik* 75 (1996), 79–97.

T. Gerstner, M. Griebel: Numerical integration using sparse grids, *Numerical Algorithms* 18 (1998), 209–232.

M. Chen, S. Mehrotra: Epi-convergent scenario generation method for stochastic problems via sparse grid, *SPEPS* 7-2008.

Scenario reduction:

P. Avella, A. Sassano, I. Vasil'ev: Computational study of large-scale p-median problems, *Mathematical Programming* 109 (2007), 89–117.

J. Dupačová, N. Gröwe-Kuska, W. Römisch: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* 95 (2003), 493–511.

W. Römisch and R. J-B Wets: Stability of epsilon-approximate solutions to convex stochastic programs, *SIAM Journal on Optimization* 18 (2007), 961–979.

H. Heitsch, W. Römisch: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.

R. Henrion, C. Küchler, W. Römisch: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming, *Journal of Industrial and Management Optimization* 4 (2008), 363–384.

R. Henrion, C. Küchler, W. Römisch: Scenario reduction in stochastic programming with respect to discrepancy distances, *Computational Optimization and Applications* 43 (2009), 67–93.

S, T. Rachev, L. Rüschendorf: *Mass Transportation Problems*, Vol. I, Springer, Berlin, 1998.

Scenario trees:

J. Dupačová, G. Consigli, S. W. Wallace: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.

M. Casey, S. Sen: The scenario generation algorithm for multistage stochastic linear programming, *Mathematics of Operations Research* 30 (2005), 615–631.

K. Frauendorfer: Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming* Ser. B, 75 (1996), 277–293.

H. Heitsch, W. Römisch: Scenario tree modeling for multistage stochastic programs, *Mathematical Programming* 118 (2009), 371–406.

H. Heitsch, W. Römisch: Scenario tree reduction for multistage stochastic programs, *Computational Management Science* 6 (2009), 117–133.

R. Hochreiter, G. Ch. Pflug: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* 152 (2007), 257–272.

K. Høyland, M. Kaut, S. W. Wallace: A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* 24 (2003), 169–185.

D. Kuhn: *Generalized bounds for convex multistage stochastic programs*, Lecture Notes in Economics and Mathematical Systems, Vol. 548, Springer, Berlin, 2005.

T. Pennanen: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Mathematical Programming* 116 (2009), 461–479.

R. Mirkov, G. Ch. Pflug: Tree approximations of dynamic stochastic programs, *SIAM Journal on Optimization* 18 (2007), 1082–1105.

G. Ch. Pflug: Version independence and nested distributions in multistage stochastic optimization, *SIAM Journal on Optimization* 20 (2009), 1406–1420.