Approximation of stochastic optimization problems and scenario generation

W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

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

Numerical Methods for PDE Constrained Optimization with Uncertain Data,
Oberwolfach, January 27–February 2, 2013

DFG Research Center MATHEON
Mathematics for key technologies

Introduction

What is Stochastic Programming ?

- Mathematics for Decision Making under Uncertainty
- subfield of Mathematical Programming (MSC 90C15)

Stochastic programs are **optimization models**

- having special properties and structures,
- depending on the underlying probability distribution,
- requiring specific approximation and numerical approaches,
- having close relations to practical applications.

Selected recent monographs:

A. Ruszczynski, A. Shapiro (eds.): Stochastic Programming, Handbook, Elsevier, 2003
S.W. Wallace, W.T. Ziemba (eds.): Applications of Stochastic Programming, MPS-SIAM, 2005,
P. Kall, J. Mayer: Stochastic Linear Programming, Kluwer, 2005,
A. Shapiro, D. Dentcheva, A. Ruszczyński: Lectures on Stochastic Programming, MPS-SIAM, 2009.
G. Infanger (ed.): Stochastic Programming - The State-of-the-Art, Springer, 2010.

Approaches to optimization models under stochastic uncertainty

Let us consider the optimization model

$$\min\{f(x,\xi) \,:\, x \in X,\, g(x,\xi) \leq 0\}\,,$$

where $\xi : \Omega \to \Xi$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, Ξ and X are closed subsets of $\mathbb{R}^s$ and $\mathbb{R}^m$, respectively, $f : X \times \Xi \to \mathbb{R}$ and $g : X \times \Xi \to \mathbb{R}^d$ are lower semicontinuous.

Aim: Finding optimal decisions before knowing the random outcome of ξ (here-and-now decision).

Main approaches:

- Replace the objective by $\mathbb{E}[f(x,\xi)]$ or by $\mathbb{F}[f(x,\xi)]$, where $\mathbb{E}$ denotes expectation (w.r.t. $\mathbb{P}$) and $\mathbb{F}$ some functional on the space of real random variables (e.g., playing the role of a *risk functional*).

- (i) Replace the random constraints by the constraint

$$\mathbb{P}(\{\omega \in \Omega : g(x, \xi(\omega)) \leq 0\}) = \mathbb{P}(g(x, \xi) \leq 0) \geq p$$

  where $p \in [0, 1]$ denotes a probability level, **or** (ii) go back to the *modeling stage* and introduce a recourse action to compensate constraint violations and add the optimal recourse cost to the objective.

The first variant leads to stochastic programs with probabilistic or chance constraints:

$$\min\{\mathbb{E}[f(x, \xi)] : x \in X, \, \mathbb{P}(g(x, \xi) \leq 0) \geq p\}$$

The second variant leads to two-stage stochastic programs with recourse:

$$\min\{\mathbb{E}[f(x, \xi)] + \mathbb{E}[q(y, \xi)] : x \in X, \, y \in Y, \, g(x, \xi) + h(y, \xi) \leq 0\}.$$

or $\mathbb{E}$ replaced by a risk functional $\mathbb{F}$.

Stability of stochastic programs

Consider the stochastic programming model

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

$$M(P) := \left\{ x \in X : \int_\Xi f_j(x,\xi)P(d\xi) \le 0, j = 1, ..., r \right\}$$

where f_j from $\mathbb{R}^m \times \Xi$ to the extended reals $\overline{\mathbb{R}}$ are normal integrands, X is a nonempty closed subset of $\mathbb{R}^m$, Ξ is a closed subset of $\mathbb{R}^d$ and P is a Borel probability measure on Ξ.

(f is a normal integrand if it is Borel measurable and $f(\xi, .)$ is lower semicontinuous $\forall \xi \in \Xi$.)

Let $\mathcal{P}(\Xi)$ the set of all Borel probability measures on Ξ and by

$$v(P) = \inf_{x \in M(P)} \int_\Xi f_0(x,\xi)P(d\xi) \quad \text{(optimal value)}$$

$$S_\varepsilon(P) = \left\{ x \in M(P) : \int_\Xi f_0(x,\xi)P(d\xi) \le v(P) + \varepsilon \right\}$$

$$S(P) = S_0(P) = \arg\min_{x \in M(P)} \int_\Xi f_0(x,\xi)P(d\xi) \quad \text{(solution set).}$$

The underlying probability distribution P is often incompletely known in applied models and/or has to be approximated (estimated, discretized).
Hence, the stability behavior of stochastic programs becomes important when changing (perturbing, estimating, approximating) the probability distribution P on Ξ.

Stability refers to (quantitative) continuity properties of the optimal value function $v(.)$ and of the set-valued mapping $S_\varepsilon(.)$ at P, where both are regarded as mappings given on certain subset of $\mathcal{P}(\Xi)$ equipped with some probability metric.

(The corresponding subset of probability measures is determined by imposing certain moment conditions that are related to growth properties of the integrands f_j with respect to ξ.)

Examples: Two-stage and chance constrained stochastic programs.

Survey:
W. Römisch: Stability of stochastic programming problems, in: Stochastic Programming (A. Ruszczynski, A. Shapiro eds.), Handbook, Elsevier, 2003.

Weak convergence in $\mathcal{P}(\Xi)$

$$P_n \to_w P \;\text{ iff }\; \int_\Xi f(\xi)P_n(d\xi) \to \int_\Xi f(\xi)P(d\xi) \;\; (\forall f \in C_b(\Xi)),$$
$$\text{iff }\; P_n(\{\xi \le z\}) \to P(\{\xi \le z\}) \text{ at continuity points } z \text{ of } P(\{\xi \le \cdot\}).$$

Probability metrics on $\mathcal{P}(\Xi)$ (Monographs: Rachev 91, Rachev/Rüschendorf 98)

Metrics with ζ-structure:

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

where $\mathcal{F}$ is a suitable set of measurable functions from Ξ to $\overline{\mathbb{R}}$ and P, Q are probability measures in some set $\mathcal{P}_{\mathcal{F}}$ on which $d_{\mathcal{F}}$ is finite.
If $\mathcal{F}$ is a P-uniformity class, $P_n \to_w P$ implies $d_{\mathcal{F}}(P_n, P) \to 0$.

Examples (of $\mathcal{F}$): Sets of locally Lipschitzian functions on Ξ or of piecewise (locally) Lipschitzian functions.

There exist canonical sets $\mathcal{F}$ and metrics $d_{\mathcal{F}}$ for each specific class of stochastic programs!

Quantitative stability results

To simplify matters, let X be compact (otherwise, consider localizations).

$$\begin{aligned} \mathcal{F} &:= \{f_j(x,\cdot) \,:\, x \in X, j = 0,\ldots,r\}, \\ \mathcal{P}_{\mathcal{F}} &:= \Big\{Q \in \mathcal{P}(\Xi) : \int_\Xi \inf_{x\in X} f_j(x,\xi)Q(d\xi) > -\infty, \\ &\qquad\qquad \sup_{x\in X} \int_\Xi f_j(x,\xi)Q(d\xi) < \infty,\ j = 0,\ldots,r\Big\}, \end{aligned}$$

and the probability (semi-) metric on $\mathcal{P}_{\mathcal{F}}$:

$$d_{\mathcal{F}}(P,Q) = \sup_{x\in X} \max_{j=0,\ldots,r} \left| \int_\Xi f_j(x,\xi)(P-Q)(d\xi) \right|.$$

Lemma:
The functions $(x,Q) \mapsto \int\limits_\Xi f_j(x,\xi)Q(d\xi)$ are lower semicontinuous on $X \times \mathcal{P}_{\mathcal{F}}$.

Theorem: (Rachev-Römisch 02)

If $d \geq 1$, let the function $x \mapsto \int_\Xi f_0(x,\xi)P(d\xi)$ be Lipschitz continuous on X, and, let the function

$$(x,y) \mapsto d\Big(x, \Big\{\tilde{x} \in X : \int_\Xi f_j(\tilde{x},\xi)P(d\xi) \leq y_j, j = 1, ..., r\Big\}\Big)$$

be locally Lipschitz continuous around $(\bar{x}, 0)$ for every $\bar{x} \in S(P)$ (metric regularity condition).

Then there exist constants L, $\delta > 0$ such that

$$\begin{aligned} |v(P) - v(Q)| &\leq Ld_{\mathcal{F}}(P,Q) \\ S(Q) &\subseteq S(P) + \Psi_P(Ld_{\mathcal{F}}(P,Q))\mathbb{B} \end{aligned}$$

holds for all $Q \in \mathcal{P}_{\mathcal{F}}$ with $d_{\mathcal{F}}(P,Q) < \delta$.

Here, $\Psi_P(\eta) := \eta + \psi^{-1}(\eta)$ and $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is given by

$$\psi(\tau) := \min\Big\{\int_\Xi f_0(x,\xi)P(d\xi) - v(P) : d(x, S(P)) \geq \tau, x \in M(P)\Big\}.$$

(Proof by appealing to general perturbation results see Klatte 94 and Rockafellar/Wets 98.)

Home Page
Title Page
Contents
Page 9 of 43
Go Back
Full Screen
Close
Quit

Convex case and $r := 0$:

Assume that $f_0(\cdot, \xi)$ is convex on $\mathbb{R}^m$ for each $\xi \in \Xi$.

Theorem: (Römisch-Wets 07)

Then there exist constants L, $\bar{\varepsilon} > 0$ such that

$$d\!l_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{L}{\varepsilon} d_{\mathcal{F}}(P, Q)$$

for every $\varepsilon \in (0, \bar{\varepsilon})$ and $Q \in \mathcal{P}_{\mathcal{F}}$ such that $d_{\mathcal{F}}(P, Q) < \varepsilon$.

Here, $d\!l_\infty$ is the Pompeiu-Hausdorff distance of nonempty closed subsets of $\mathbb{R}^m$, i.e.,

$$d\!l_\infty(C, D) = \inf\{\eta \geq 0 : C \subseteq D + \eta\mathbb{B}, D \subseteq C + \eta\mathbb{B}\}.$$

(Proof using a perturbation result see Rockafellar/Wets 98)

The (semi-) distance $d_{\mathcal{F}}$ plays the role of a minimal probability metric implying quantitative stability.

Furthermore, the result remains valid when bounding $d_{\mathcal{F}}$ from above by another distance and when reducing the set $\mathcal{P}_{\mathcal{F}}$ to a subset on which this distance is defined and finite.

Idea: Enlarge $\mathcal{F}$, but maintain the analytical (e.g., (dis)continuity) properties of $f_j(x,\cdot)$, $j = 0, \ldots, r$!

This idea may lead to well-known probability metrics, for which a well developed theory is available !

Example: (Fortet-Mourier-type metrics)

We consider the following classes of locally Lipschitz continuous functions (on Ξ)

$$\mathcal{F}_H := \{f : \Xi \to \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq \max\{1, H(\|\xi\|), H(\|\tilde{\xi}\|)\} \cdot \|\xi - \tilde{\xi}\|, \forall \xi, \tilde{\xi} \in \Xi\},$$

where $H : \mathbb{R}_+ \to \mathbb{R}_+$ is nondecreasing, $H(0) = 0$. The corresponding distances are

$$d_{\mathcal{F}_H}(P,Q) = \sup_{f \in \mathcal{F}_H} \left| \int_\Xi f(\xi)P(d\xi) - \int_\Xi f(\xi)Q(d\xi) \right| =: \zeta_H(P,Q)$$

so-called Fortet-Mourier-type metrics defined on

$$\mathcal{P}_H(\Xi) := \{Q \in \mathcal{P}(\Xi) : \int_\Xi \max\{1, H(\|\xi\|)\} \|\xi\| Q(d\xi) < \infty\}$$

Important special case: $H(t) := t^{p-1}$ for $p \geq 1$ leading to the notation $\mathcal{F}_p$, $\mathcal{P}_p(\Xi)$ and ζ_p, respectively.

(Convergence with respect to ζ_p means weak convergence of the probability measures and convergence of the p-th order moments (Rachev 91))

Two-stage stochastic programming models with recourse

Consider a linear program with stochastic parameters of the form

$$\min\{\langle c, x\rangle : x \in X,\, T(\xi)x = h(\xi)\},$$

where $\xi : \Omega \to \Xi$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $c \in \mathbb{R}^m$, Ξ and X are polyhedral subsets of $\mathbb{R}^s$ and $\mathbb{R}^m$, respectively, and the $d \times m$-matrix $T(\cdot)$ and vector $h(\cdot) \in \mathbb{R}^d$ are affine functions of ξ.

Idea: Introduce a recourse variable $y \in \mathbb{R}^{\overline{m}}$, recourse costs $q(\xi) \in \mathbb{R}^{\overline{m}}$, a fixed recourse $d \times \overline{m}$-matrix W, a polyhedral cone $Y \subseteq \mathbb{R}^{\overline{m}}$, and solve the second-stage or recourse program

$$\min\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}.$$

Add the expected minimal recourse costs $\mathbb{E}[\Phi(x,\xi)]$ (depending on the first-stage decision x) to the original objective and consider

$$\min\left\{\langle c, x\rangle + \mathbb{E}[\Phi(x,\xi)] : x \in X\right\},$$

where $\Phi(x,\xi) := \inf\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}$.

Assumptions:

(A1) *relatively complete recourse*: for any $(\xi, x) \in \Xi \times X$, $h(\xi) - T(\xi)x \in W(Y)$;

(A2) *dual feasibility*: $q(\xi) \in D(\xi) = \{z : W^\top z - q(\xi) \in Y^*\}$ holds for all $\xi \in \Xi$ (with Y^* denoting the polar cone to Y).

(A3) *finite second order moment*: $\int_\Xi \|\xi\|^2 P(d\xi) < \infty$.

Note that (A1) is satisfied if $W(Y) = \mathbb{R}^d$ (complete recourse). In general, (A1) and (A2) impose a condition on the support of P.

Proposition:

Assume (A1)–(A3). Then the deterministic equivalent of the two-stage model represents a nondifferentiable convex program (with polyhedral constraints). An element $x \in X$ minimizes the convex program if and only if

$$0 \in \int_\Xi \partial\Phi(x, \xi) P(d\xi) + N_X(x)\,,$$

where the subdifferential in the integrand is given by

$$\partial\Phi(x, \xi) = c - T(\xi)^\top \arg\max_{z \in D(\xi)} z^\top (h(\xi) - T(\xi)x).$$

Stability of two-stage models

We set

$$f_0(x,\xi) = \langle c, x\rangle + \Phi(x,\xi)$$

for all pairs $(x,\xi) \in X \times \Xi$ such that $h(\xi) - T(\xi)x \in W(Y)$ and $q(\xi) \in \mathcal{D}$ and $f_0(x,\xi) = +\infty$ otherwise.

Proposition:

Assume (A1) and (A2). Then there exist $\hat{L} > 0$ such that

$$\begin{aligned} |f_0(x,\xi) - f_0(x,\tilde{\xi})| &\leq \hat{L}\max\{1, \|\xi\|, \|\tilde{\xi}\|\}\|\xi - \tilde{\xi}\| \\ |f_0(x,\xi) - f_0(\tilde{x},\xi)| &\leq \hat{L}\max\{1, \|\xi\|^2\}\|x - \tilde{x}\| \end{aligned}$$

for all $\xi, \tilde{\xi} \in \Xi$, $x, \tilde{x} \in X$.

Theorem:

Assume (A1)–(A3) and let X be compact. Then there exist $L > 0$, $\bar{\varepsilon}$, $\delta > 0$ such that

$$\begin{aligned} |v(P) - v(Q)| &\leq L\zeta_2(P,Q), \\ S(Q) &\subseteq S(P) + \Psi_P(L\zeta_2(P,Q))\mathbb{B}, \\ d\!l_\infty(S_\varepsilon(P), S_\varepsilon(Q)) &\leq \frac{L}{\varepsilon}\zeta_2(P,Q), \end{aligned}$$

whenever Q satisfies $\zeta_2(P,Q) < \delta$, $\varepsilon \in (0, \bar{\varepsilon}]$, $\Psi_P(\eta) := \eta + \psi^{-1}(\eta)$ and

$$\psi(\tau) := \min\left\{ \int_\Xi f_0(x,\xi)P(d\xi) - v(P) : d(x, S(P)) \geq \tau, x \in X \right\}.$$

Note ψ has quadratic growth (near 0) in a number of cases (Schultz 94) and linear growth if P is discrete.

Discrete approximations of two-stage stochastic programs

Replace the (original) probability measure P by measures P_n having (finite) discrete support $\{\xi_1, \ldots, \xi_n\}$ $(n \in \mathbb{N})$, i.e.,

$$P_n = \sum_{i=1}^{n} p_i \delta_{\xi_i},$$

and insert it into the infinite-dimensional stochastic program:

$$\min\{\langle c, x\rangle + \sum_{i=1}^{n} p_i \langle q(\xi_i), y_i\rangle : x \in X, y_i \in Y,\ i = 1, \ldots, n,$$

$$\begin{array}{cccccc} W y_1 & & & +T(\xi_1)x & = & h(\xi_1) \\ & W y_2 & & +T(\xi_2)x & = & h(\xi_2) \\ & & \ddots & \vdots & = & \vdots \\ & & W y_n & +T(\xi_n)x & = & h(\xi_n)\} \end{array}$$

Hence, we arrive at a (finite-dimensional) large scale block-structured linear program which allows for specific decomposition methods.

(Ruszczyński/Shapiro, Handbook, 2003)

Empirical or Monte Carlo approximations of stochastic programs

Given a probability distribution $P \in \mathcal{P}(\Xi)$, we consider a sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ of independent, identically distributed Ξ-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having the common distribution P.

We consider the empirical measures

$$P_n(\omega) := \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(\omega)}$$

for every $n \in \mathbb{N}$.

Empirical or sample average approximation of stochastic programs (replacing P by $P_n(\cdot)$):

$$\min \left\{ \frac{1}{n} \sum_{i=1}^{n} f_0(\xi_i, x) : x \in X, \frac{1}{n} \sum_{i=1}^{n} f_j(\xi_i, x) \leq 0, j = 1, \ldots, r \right\}$$

To study convergence of empirical approximations, one may use the quantitative stability results by deriving estimates of the (uniform) distances

$$d_{\mathcal{F}}(P, P_n(\cdot))$$

Tool: Empirical process theory, in particular, the size of $\mathcal{F}$ as subset of $L_p(\Xi, P)$ measured by covering numbers, where

$$\mathcal{F} = \{f_j(x, \cdot) : x \in X, j = 0, \dots, r\}.$$

Empirical process (indexed by some class of functions):

$$\left\{n^{\frac{1}{2}}(P_n(\cdot) - P)f = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(f(\xi_i(\cdot)) - \int_{\Xi} f(\xi)P(d\xi)\right)\right\}_{f \in \mathcal{F}}$$

Desirable estimate:

$$\mathbb{P}\left(\left\{\omega : n^{\frac{1}{2}} d_{\mathcal{F}}(P, P_n(\omega)) \geq \varepsilon\right\}\right) \leq C_{\mathcal{F}}(\varepsilon) \quad (\forall \varepsilon > 0, n \in \mathbb{N})$$

for some tail function $C_{\mathcal{F}}(\cdot)$ defined on $(0, +\infty)$ and decreasing to 0, in particular, exponential tails $C_{\mathcal{F}}(\varepsilon) = K\varepsilon^r \exp(-2\varepsilon^2)$.

If $N(\varepsilon, L_p(Q))$ denotes the minimal number of open balls $\{g : \|g - f\|_{Q,p} < \varepsilon\}$ needed to cover $\mathcal{F}$, then an estimate of the form

$$\sup_Q N(\varepsilon, L_2(Q)) \leq \left(\frac{R}{\varepsilon}\right)^r$$

for some $r, R \geq 1$ and all $\varepsilon > 0$, is needed to obtain exponential tails.

(Literature: Talagrand 94, van der Vaart/Wellner 96, van der Vaart 98)

Typical result for optimal values:

$$\mathbb{P}\left(|v(P) - v(P_n)| \geq \varepsilon\, n^{-\frac{1}{2}}\right) \leq C_{\mathcal{F}}\left(\min\{\delta, \varepsilon\, L^{-1}\}\right)$$

Such results are available for two-stage (mixed-integer) and chance constrained stochastic programs (Römisch 03).

Desirable results for optimal values: Limit theorems

$$n^{\frac{1}{2}}(v(P_n(\cdot)) - v(P)) \longrightarrow z,$$

where z is a real random variable and the convergence is *convergence in distribution*.

Such results can be derived if $\mathcal{F}$ is a Donsker class of functions. Donsker classes can also be characterized via covering numbers.

Examples for available limit theorems:

- Limit theorem for optimal values of mixed-integer two-stage stochastic programs (Eichhorn/Römisch 07).
- Limit theorem for optimal values of kth order stochastic dominance constrained stochastic programs for $k \geq 2$

  (Dentcheva/Römisch 12).

(Chapters by Shapiro and Pflug in the Handbook 2003; recent work of Shapiro, Xu and coworkers)

Scenario generation methods

Assume that we have to solve a stochastic program with a class $\mathcal{F} = \{f_j(x,\cdot) : x \in X, j = 1,\dots,r\}$ of functions on $\Xi \subseteq \mathbb{R}^d$ and probability (semi-) metric

$$d_{\mathcal{F}}(P,Q) = \sup_{f\in\mathcal{F}} \Big| \int_{\Xi} f(\xi)(P-Q)(d\xi) \Big|.$$

Optimal scenario generation:

For given $n \in \mathbb{N}$ and probabilities $p_i = \frac{1}{n}$, $i = 1,\dots,n$, the best possible choice of scenarios $\xi_i \in \Xi$, $i = 1,\dots,n$, is obtained by solving the best approximation problem

$$\min \Big\{ d_{\mathcal{F}}\big(P, \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \delta_{\xi_i}\big); \xi_i \in \Xi, i = 1,\dots,n \Big\}.$$

However, this is a large-scale, nonsmooth and nonconvex minimization problem (of dimension $n \cdot d$) and often extremely difficult to solve. Note that, in addition, function calls for $f_j(x,\cdot)$ are often expensive and the appropriate choice of $n \in \mathbb{N}$ is difficult.

Next we discuss 4 specific scenario generation methods for stochastic programs (*without information constraints*) based on (high-dimensional) numerical integration methods:

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

(b) Optimal quantization of probability distributions (Pflug-Pichler 11).

(c) Quasi-Monte Carlo methods (Koivu-Pennanen 05, Homem-de-Mello 08).

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

Given an integral

$$I_d(f) = \int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi \quad \text{or} \quad I_d(f) = \int_{[0,1]^d} f(\xi)d\xi$$

a numerical integration method means

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

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 $1 \le a < M$ and $gcd(a, M) = 1$, and $c \in Z_M = \{0, 1, \dots, M-1\}$. Starting with $y_0 \in Z_M$ a sequence is generated by

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

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 that 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

Assume that the underlying stochastic program behaves stable with respect to a distance d of probability measures on $\mathbb{R}^d$.

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).

Note, however, $\ell_r(P, Q_n) \geq cn^{-\frac{1}{d}}$ for some $c > 0$ and all $n \in \mathbb{N}$.

In general, the function

$$\Psi_d(\xi^1, \ldots, \xi^n) := d\Big(P, \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi^i}\Big)$$

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

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).

Asymptotically optimal quantizers can be determined explicitly in a number of cases (Pflug-Pichler 11).

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 $\xi^1,\ldots,\xi^n$

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

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

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

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

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

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

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

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 $V_{\mathrm{HK}}(f)$ in the sense of Hardy and Krause, it holds

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

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

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

however, the constant depends on the dimension d.

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.

Specific sequences: Faure, Sobol', Niederreiter and Niederreiter-Xing sequences (Lemieux 09, Dick-Pillichshammer 10).

Recent development: Scrambled (t, m, d)-nets, where the digits are randomly permuted (Owen 95).

Second general construction: (Korobov 59, Sloan-Joe 94)

Lattice rules: 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 is the *fractional part* of $z \in \mathbb{R}_+$, i.e., $\{z\} = z - \lfloor z \rfloor \in [0, 1)$.
The generator g is chosen such that the lattice rule has good convergence properties.

Such lattice rules may achieve better convergence rates $O(n^{-k+\delta})$, $k \in \mathbb{N}$, for smooth integrands.

Recent development: Randomized lattice rules.

Randomly shifted lattice points:

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

where $\triangle$ is uniformly distributed in $[0,1]^d$.

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

$$e(Q_{n,d}) \leq C(\delta) n^{-1+\delta} \quad (n \in \mathbb{N})$$

is achieved if the integrand f belongs to the tensor product Sobolev space

$$\mathbb{F}_d = W_2^{(1,\ldots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1])$$

equipped with a weighted norm. Since the space $\mathbb{F}_d$ is a kernel reproducing Hilbert space, a well developed technique for estimating the quadrature error can be used.

(Hickernell 96, Sloan/Wožniakowski 98, Sloan/Kuo/Joe 02, Kuo 03)

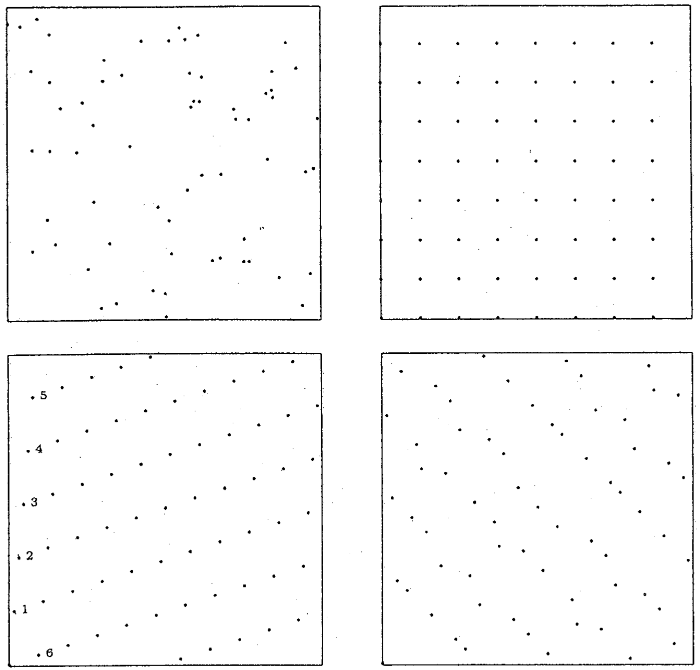


Fig. 5.3 Four different point sets with $n = 64$: random (top left), rectangular grid (top right), Korobov lattice (bottom left), and Sobol' (bottom right).

Is QMC efficient in stochastic programming ?

Problem: Typical integrands in linear stochastic programming are not of bounded variation in the HK sense and nonsmooth and, hence, do not belong to the relevant function space $\mathbb{F}_d$ in general.

Idea: Study the ANOVA decomposition and efficient dimension of two-stage integrands.

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 in $[0,1]$, $j \in D \setminus u$ and $D = \{1, \ldots, d\}$. Hence, f_u is essentially as smooth as $I_{-u}(f)$ and does not depend on ξ^{-u}.

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 superposition dimension d_s of f is the smallest $d_s \in \mathbb{N}$ with

$$\sum_{|u| \leq d_s} \|f_u\|_{L_2}^2 \geq (1-\varepsilon)\sigma^2(f) \quad \text{(where } \varepsilon \in (0,1) \text{ is small).}$$

Then

$$\|f - \sum_{|u| \leq d_s} f_u\|_{L_2}^2 \leq \varepsilon \sigma^2(f).$$

Result:

All ANOVA terms f_u, $u \subset D$, $u \neq D$, of integrands in two-stage stochastic programming belong to C^∞ if the underlying marginal densities belong to $C_b^\infty(\mathbb{R})$ and certain geometric condition is satisfied (Heitsch/Leovey/Römisch 12).

Hence, after reducing the efficient superposition dimension of f such that (at least) $d_s \leq d-1$ holds, QMC methods should have optimal rates.

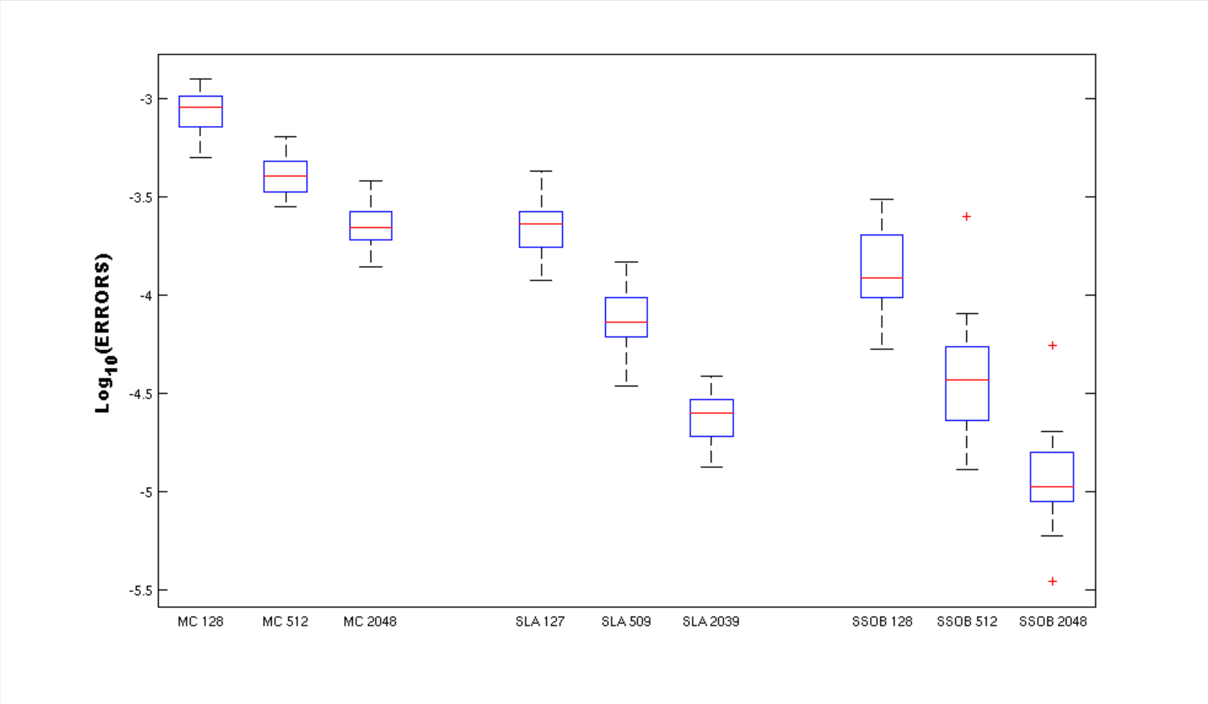
Some computational experience

We considered a two-stage production planning problem for maximizing the expected revenue while satisfying a fixed demand in a time horizon with $d = T = 100$ time periods and stochastic prices for the second-stage decisions. It is assumed that the probability distribution of the prices ξ is log-normal. The model is of the form

$$\max\left\{\sum_{t=1}^{T}\Big(c_t^\top x_t + \int_{\mathbb{R}^T} q_t(\xi)^\top y_t P(d\xi)\Big) : Wy + Vx = h, y \geq 0, x \in X\right\}$$

The use of PCA for decomposing the covariance matrix has led to efficient truncation dimension $d_T(0.01) = 2$. As QMC methods we used a randomly scrambled Sobol sequence (SSobol)(Owen, Hickernell) with $n = 2^7, 2^9, 2^{11}$ and a randomly shifted lattice rule (Sloan-Kuo-Joe) with $n = 127, 509, 2039$, weights $\gamma_j = \frac{1}{j^2}$ and used for MC the Mersenne-Twister. 10 runs were performed for the error estimates and 30 runs for plotting relative errors.

Average rate of convergence for QMC: $O(n^{-0.9})$ and $O(n^{-0.8})$.
Instead of $n = 2^7$ SSobol samples one would need $n = 10^4$ MC samples to achieve a similar accuracy as SSobol.



Shown are the Log_{10} of relative RMSE with PCA factorization of covariance matrix. Results for Mersenne Twister MC and randomly scrambled Sobol' QMC $128, 512$ and 2048 points (MC 128,... or SSOB 128,...), and randomly shifted lattice rules QMC $127, 509$ and 2039 lattice points (SLA 127,...)

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.,

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

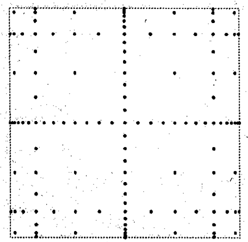
for example, the dyadic grid

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

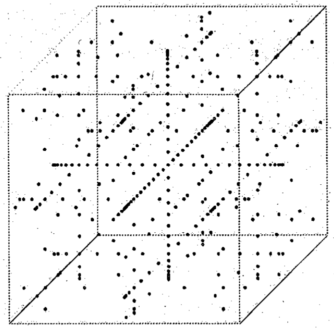
Then the point set suggested by Smolyak (Smolyak 63)

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

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 $\mathbb{F}_d = W_2^{(r,\ldots,r)}([0,1]^d)$, it holds

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

References:

S. T. Rachev: *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.

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

A. Prekopa: *Stochastic Programming*, Kluwer, Dordrecht, 1995.

A. W. van der Vaart, J. Wellner: *Weak Convergence and Empirical Processes*, Springer, New York, 1996.

R. T. Rockafellar and R. J-B. Wets: *Variational Analysis*, Springer, Berlin, 1998.

A. W. van der Vaart: *Asymptotic Statistics*, Cambridge University Press, Cambridge 1998.

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

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

A. Müller and D. Stoyan: *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester, 2002.

A. Ruszczyński and A. Shapiro (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

H. Föllmer, A. Schied: Stochastic Finance: An Introduction in Discrete Time, de Gruyter, Berlin 2004 (Third Edition 2011).

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

S. W. Wallace and W. T. Ziemba (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2005.

G. Ch. Pflug, W. Römisch: Modeling, Measuring and Managing Risk, World Scientific, Singapore, 2007.

A. Genz, F. Bretz: *Computation of multivariate normal and t probabilities*, Lecture Notes in Statistics Vol. 195, Springer, Berlin, 2009.

A. Shapiro, D. Dentcheva and A. Ruszczyński: *Lectures on Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2009.

J. Dick and F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, 2010.