Towards Quasi-Monte Carlo scenario generation in stochastic programming

W. Römisch

Humboldt-University Berlin

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

(H. Heitsch)

CMS 2012, Imperial College London, April 18–20, 2012

Introduction

- Standard approach for solving stochastic programs are variants of Monte Carlo (MC) for generating scenarios (i.e., samples).
- Recent alternative approaches to scenario generation in stochastic programming besides MC:
  (a) Optimal quantization of probability distributions (Pflug-Pichler 2010).
  (b) Quasi-Monte Carlo (QMC) methods (Koivu-Pennanen 05, Homem-de-Mello 06).
  (c) Sparse grid quadrature rules (Chen-Mehrotra 08).
- While the justification of MC and (a) may be based on available stability results for stochastic programs, there is almost no reasonable justification of applying (b) and (c).
- Known convergence rates: MC $O(n^{-\frac{1}{2}})$, (a) $O(n^{-\frac{1}{d}})$
  (b) $O(n^{-1}(\log n)^d)$, recently: $O(n^{-1+\delta})$ (δ small)
  (d dimension of random vector, n number of scenarios).

Two-stage linear stochastic programs

Two-stage stochastic programs arise as deterministic equivalents of improperly posed random linear programs

$$\min\{\langle c, x\rangle : x \in X,\, Tx = \xi\},$$

where X is a convex polyhedral subset of $\mathbb{R}^m$, T a matrix, ξ is a d-dimensional random vector.
A possible deviation $\xi - Tx$ is compensated by additional costs $\Phi(x, \xi)$ whose mean with respect to the probability distribution P of ξ is added to the objective. We assume that the additional costs represent the optimal value of a *second-stage program*, namely,

$$\Phi(x, \xi) = \inf\{\langle q, y\rangle : y \in \mathbb{R}^{\bar{m}},\, Wy = \xi - Tx, y \geq 0\},$$

where $q \in \mathbb{R}^{\bar{m}}$, W a $(d, \bar{m})$-matrix (having rank d).
The *deterministic equivalent* then is of the form

$$\min\left\{\langle c, x\rangle + \int_{\mathbb{R}^d} \Phi(x, \xi) P(d\xi) : x \in X\right\}.$$

We assume that the additional costs are of the form

$$\Phi(x,\xi) = \varphi(\xi - Tx)$$

with the second-stage optimal value function

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

There exist vertices v^j of the dual feasible set $\mathcal{D}$ and polyhedral cones $\mathcal{K}_j$, $j = 1,\ldots,\ell$, decomposing $\operatorname{dom}\varphi$ such that

$$\varphi(t) = \langle v^j, t\rangle, \ \forall t \in \mathcal{K}_j, \quad \text{and} \quad \varphi(t) = \max_{j=1,\ldots,\ell} \langle v^j, t\rangle.$$

Hence, the integrands are of the form

$$f(\xi) = \max_{j=1,\ldots,\ell} \langle v^j, \xi - Tx\rangle \quad \text{if} \quad \xi - Tx \in W(\mathbb{R}^{\bar m}_+).$$

Quasi-Monte Carlo methods

We consider the approximate computation of

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

by a QMC algorithm

$$Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\xi^i) \quad \text{or} \quad Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\xi^i)\rho_d(\xi^i)$$

with (non-random) points ξ^i, $i = 1, \ldots, n$, from $[0,1]^d$ or $\mathbb{R}^d$.

We assume that f belongs to a linear normed space $\mathbb{F}_d$ with norm $\|\cdot\|_d$ and unit ball $\mathbb{B}_d$. Worst-case error of $Q_{n,d}$ over $\mathbb{B}_d$:

$$e(Q_{n,d}) = \sup_{f \in \mathbb{B}_d} |I_d(f) - Q_{n,d}(f)|$$

Example: F_d is a weighted tensor product Sobolev space, a particular kernel reproducing Hilbert space.

Problem: Integrands in stochastic programming are not in F_d.

ANOVA decomposition of multivariate functions

Idea: Decompositions of f may be used, where most of the terms are smooth, but hopefully only some of them relevant.

Let $D = \{1, \ldots, d\}$ and $f \in L_{1,\rho_d}(\mathbb{R}^d)$ with $\rho_d(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j)$, where

$$f \in L_{p,\rho_d}(\mathbb{R}^d) \quad \text{iff} \quad \int_{\mathbb{R}^d} |f(\xi)|^p \rho_d(\xi) d\xi < \infty \quad (p \geq 1).$$

Let the projection P_k, $k \in D$, be defined by

$$(P_k f)(\xi) := \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d).$$

Clearly, $P_k f$ is constant with respect to ξ_k. For $u \subseteq D$ we write

$$P_u f = \Big(\prod_{k \in u} P_k \Big)(f),$$

where the product means composition, and note that the ordering within the product is not important because of Fubini's theorem. The function $P_u f$ is constant with respect to all x_k, $k \in u$.

ANOVA-decomposition of f:

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

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

$$f_u = P_{-u}(f) - \sum_{v \subseteq u} f_v$$

or (due to Kuo-Sloan-Wasilkowski-Woźniakowski 10)

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

where P_{-u} and P_{u-v} mean integration with respect to ξ_j, $j \in D \setminus u$ and $j \in u \setminus v$, respectively. The second representation motivates that f_u is essentially as smooth as $P_{-u}(f)$.

If f belongs to $L_{2,\rho_d}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq D}$ are orthogonal in $L_{2,\rho_d}(\mathbb{R}^d)$.

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

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

Sobol's global sensitivity indices of f:

$$\bar{S}_u = \frac{1}{\sigma^2(f)} \sum_{v \cap u \neq \emptyset} \sigma_v^2(f).$$

Owen's dimension distribution (superposition or truncation) of f: Probability measure ν_S (ν_T) defined on the power set of D

$$\nu_S(s) := \sum_{|u|=s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \qquad \left(\nu_T(s) = \sum_{\max\{j : j \in u\} = s} \frac{\sigma_u^2(f)}{\sigma^2(f)}\right) \;\; (s \in D).$$

Mean superposition dimension of f:

$$\bar{d}_S = \sum_{\emptyset \neq u \subseteq D} |u| \frac{\sigma_u^2(f)}{\sigma^2(f)} = \sum_{i=1}^{d} S_{\{i\}}.$$

Efficient truncation dimension $d_T(\varepsilon)$ of f is the $(1-\varepsilon)$-quantile of ν_T.

ANOVA decomposition of two-stage integrands

Assumption:
(A1) $W(\mathbb{R}^{\bar{m}}_+) = \mathbb{R}^d$ (*complete recourse*).
(A2) $\mathcal{D} \neq \emptyset$ (*dual feasibility*).
(A3) $\int_{\mathbb{R}^d} \|\xi\| P(d\xi) < \infty$.
(A4) P has a density of the form $\rho_d(\xi) = \prod_{j=1}^d \rho_j(\xi_j)$ ($\xi \in \mathbb{R}^d$) with $\rho_j \in C(\mathbb{R})$, $j = 1, \ldots, d$.

(A1) and (A2) imply that $\operatorname{dom} \varphi = \mathbb{R}^d$ and $\mathcal{D}$ is bounded and, hence, it is the convex hull of its vertices. Furthermore, the cones $\mathcal{K}_j$ are the normal cones to $\mathcal{D}$ at the vertices v^j, i.e.,

$$\begin{aligned} \mathcal{K}_j &= \{t \in \mathbb{R}^d : \langle t, z - v^j \rangle \leq 0, \forall z \in \mathcal{D}\} \quad (j = 1, \ldots, \ell) \\ &= \{t \in \mathbb{R}^d : \langle t, v^i - v^j \rangle \leq 0, \forall i = 1, \ldots, \ell, \, i \neq j\}. \end{aligned}$$

It holds that $\cup_{j=1,\ldots,\ell} \mathcal{K}_j = \mathbb{R}^d$ and for $j \neq j'$ the intersection $\mathcal{K}_j \cap \mathcal{K}_{j'}$ is a common closed face of dimension $d-1$ iff the two cones are adjacent. The intersection is contained in

$$\{t \in \mathbb{R}^d : \langle t, v^{j'} - v^j \rangle = 0\}.$$

To compute projections $P_k(f)$ for $k \in D$. Let $\xi_i \in \mathbb{R}$, $i = 1, \dots, d$, $i \neq k$, be given. We set $\xi^k = (\xi_1, \dots, \xi_{k-1}, \xi_{k+1}, \dots, \xi_d)$ and

$$\xi_s = (\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d) \in \mathbb{R}^d = \cup_{j=1,\dots,\ell} \mathcal{K}_j.$$

Assuming (A1)–(A4) it is possible to derive an explicit representation of $P_k(f)$ depending on ξ^k and on the finitely many points at which the one-dimensional affine subspace $\{\xi_s : s \in \mathbb{R}\}$ meets the common face of two adjacent cones. This leads to

Proposition:

Let $k \in D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different kth components.

The kth projection $P_k f$ is infinitely differentiable if the density ρ_k is in $C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$, in particular, if ρ_k is the normal density.

Theorem:

Let $u \subset D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different kth components for some $k \in D \setminus u$.

The ANOVA term f_u belongs to $C^\infty(\mathbb{R}^{d-|u|})$ if $\rho_k \in C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$.

Example:

Let $\bar{m} = 3$, $d = 2$, P denote the two-dimensional standard normal distribution and let the following vector q and matrix W

$$W = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \qquad q = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

be given. Then (A1) and (A2) are satisfied and the dual feasible set $\mathcal{D}$ is the triangle (in $\mathbb{R}^2$)

$$\mathcal{D} = \{z \in \mathbb{R}^2 : -z_1 + z_2 \leq 1, z_1 + z_2 \leq 1, -z_2 \leq 0\},$$

with the vertices

$$v^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v^2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad v^3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

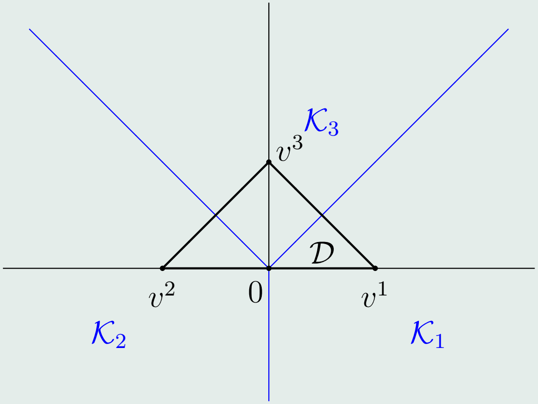

Figure 1: Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

Hence, the second component of the two adjacent vertices v^1 and v^2 coincides. The function φ is of the form

$$\varphi(t) = \max_{i=1,2,3} \langle v^i, t\rangle = \max\{t_1, -t_1, t_2\} = \max\{|t_1|, t_2\}$$

and the integrand is

$$f(\xi) = \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}$$

The ANOVA projection $P_1 f$ is in C^∞, but $P_2 f$ is not differentiable.

Remark: Under the assumptions of the theorem the function

$$f_{d-1}(\xi) = \sum_{|u| \le d-1} f_u = f - f_D$$

is in $C^\infty(\mathbb{R}^d)$ if $\rho_k \in C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$ for every $k \in D$. For which two-stage stochastic programs is $\|f_D\|_{L_2}$ small, i.e., the efficient truncation dimension is less than $d-1$?

Remark: If ξ is normal with covariance matrix Σ, there exists an orthogonal matrix Q such that $\Sigma = QDQ^\top$ with a diagonal matrix D containing the eigenvalues. Hence, we may assume that $h(\xi)$ is of the form

$$h(\xi) = Q\xi \quad \text{with } \xi \text{ satisfying (A4).}$$

Then the geometric condition on the vertices of $\mathcal{D}$ is generically satisfied in the following sense: The set of all orthogonal matrices Q such that $Q\mathcal{D}$ satisfies the geometric condition is representable as the countable intersection of open dense subsets.

Sensitivity and the reduction of efficient dimension

Proposition:
Assume (A1)–(A4) and let σ_i^2 denote the variance of ξ_i, $i = 1, \ldots, d$. Then it holds

$$\bar{S}_{\{i\}} \leq \frac{\sigma_i^2}{\sigma^2(f)} \max_{j=1,\ldots,\ell} |v_i^j|^2 \quad (i = 1, \ldots, d),$$

where v^j, $j = 1, \ldots, \ell$, are the vertices of the dual polyhedron.

Hence, the transformation of a $\mathcal{N}(\mu, \Sigma)$ random vector in the form $\Sigma = B\,B^\top$ should be organized such that the σ_i are decreasing and the first few variances σ_i are (strongly) dominating if possible.

Standard Cholesky decomposition $B = L$ is **not useful**.
Principal component analysis (PCA), i.e., $B = (\sqrt{\lambda_1} v_1, \ldots, \sqrt{\lambda_d} v_d)$, where $\lambda_1 \geq \cdots \geq \lambda_d$ are the eigenvalues of Σ in decreasing order and v_i, $i = 1, \ldots, d$, the corresponding orthonormal eigenvectors, is **very useful** in financial applications (Wang-Fang 03, Wang-Sloan 07).

Conclusions

- The results provide a theoretical basis for applying QMC accompanied by efficient dimension reduction techniques to stochastic programs with low efficient dimension.
- The results are extendable and will be extended to more general two-stage and to multi-stage situations.
- Numerical experiments using randomly shifted lattice rules (Kuo, Sloan) and digitally shifted polynomial lattice rules (Dick, Pillichshammer) are in preparation.

Thank you !

References

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

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

M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* (to appear).

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.

J. Imai and K. S. Tan: Minimizing effective dimension using linear transformation, in *Monte Carlo and Quasi-Monte Carlo Methods 2002* (H. Niederreiter Ed.), Springer, Berlin, 2004, 275–292.

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.

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

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, B. J. Waterhouse: Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands, *Journal of Complexity* 26 (2010), 135–160.

A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.

D. Nuyens and R. Cools: Fast algorithms for component-by-component constructions of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, *Mathematics of Computation* 75 (2006), 903-922.

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

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.), Springer, 2011.

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

I. H. Sloan, F. Y. Kuo and 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.

I. M. Sobol' and S. Kucherenko: Derivative based global sensitivity measures and their link with global sensitivity indices, *Mathematics and Computers in Simulation* 79 (2009), 3009-3017.

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

X. Wang and I. H. Sloan: Brownian bridge and principal component analysis: towards removing the curse of dimensionality, *IMA Journal of Numerical Analysis* 27 (2007), 631–654.

X. Wang and I. H. Sloan: Low discrepancy sequences in high dimensions: How well are their projections distributed ? *Journal of Computational and Applied Mathematics* 213 (2008), 366–386.