A bilevel optimization approach to optimal scenario generation in two-stage stochastic programming

W. Römisch

Humboldt-University Berlin
Institute of Mathematics

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

R. Henrion (WIAS Berlin)

ICBO 16, Dempe 60, Dresden, 4.5.-6.5. 2016

Happy Birthday, Stephan!

Introduction

Many stochastic programming models may be traced back to minimizing an expectation functional on some closed subset of a Euclidean space or, eventually in addition, relative to some expectation constraint. Their general form is

$$\text{(SP)} \qquad \min\left\{\int_{\Xi} f_0(x,\xi)P(d\xi) : x \in X,\ \int_{\Xi} f_1(x,\xi)P(d\xi) \le 0\right\}$$

where X is a closed subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^s$, P is a Borel probability measure on Ξ abbreviated by $P \in \mathcal{P}(\Xi)$. The functions f_0 and f_1 from $\mathbb{R}^m \times \Xi$ to the extended reals $\overline{\mathbb{R}} = [-\infty, \infty]$ are normal integrands.

For example, typical integrands in linear two-stage stochastic programming models are

$$f_0(x,\xi) = \begin{cases} g(x) + \Phi(q(\xi), h(x,\xi)) & , q(\xi) \in D \\ +\infty & , \text{else} \end{cases} \quad \text{and} \quad f_1(x,\xi) \equiv 0,$$

where X and Ξ are convex polyhedral, $g(\cdot)$ is a linear function, $q(\cdot)$ is affine, $D = \{q \in \mathbb{R}^{\bar{m}} : \{z \in \mathbb{R}^r : W^\top z - q \in Y^\star\} \neq \emptyset\}$ denotes the convex polyhedral dual feasibility set, $h(\cdot,\xi)$ is affine for fixed ξ and $h(x,\cdot)$ is affine for fixed x, and Φ denotes the infimal function of the linear (second-stage) optimization problem

$$\Phi(q,t) := \inf\{\langle q, y\rangle : Wy = t, y \in Y\}$$

with $(r, \bar{m})$ matrix W and convex polyhedral cone $Y \subset \mathbb{R}^{\bar{m}}$.

Typical integrands f_1 appearing in chance constrained programming are of the form

$$f_1(x,\xi) = p - \mathbf{1}_{\mathcal{P}(x)}(\xi),$$

where $\mathbf{1}_{\mathcal{P}(x)}$ is the characteristic function of the polyhedron $\mathcal{P}(x) = \{\xi \in \Xi : h(x,\xi) \leq 0\}$ depending on x.

For general continuous multivariate probability distributions P such stochastic optimization models are not solvable in general.

Many approaches for solving such optimization models computationally are based on discrete approximations of the probability measure P, i.e., on finding a discrete probability measure P_n in

$$\mathcal{P}_n(\Xi) := \Big\{ \sum_{i=1}^{n} p_i \delta_{\xi^i} : \xi^i \in \Xi,\, p_i \geq 0,\, i = 1,\ldots,n,\, \sum_{i=1}^{n} p_i = 1 \Big\}$$

for some $n \in \mathbb{N}$, which approximates P in a *suitable* way.

The atoms ξ^i, $i = 1,\ldots,n$, of P_n are often called scenarios in this context. Of course, the notion *suitable* should at least include that the distance of infima

$$|v(P) - v(P_n)|$$

becomes resonably small.

Stability-based scenario generation

Let $v(P)$ and $S(P)$ denote the infimum and solution set of (SP). We are interested in their dependence on the underlying probability distribution P.

To state a stability result we introduce the following sets of functions and of probability distributions (both defined on Ξ)

$$\mathcal{F} = \{f_j(x,\cdot) : j = 0, 1, \; x \in X\},$$

$$\mathcal{P}_{\mathcal{F}} = \left\{Q \in \mathcal{P}(\Xi) : -\infty < \int_{\Xi} \inf_{x\in X} f_j(x,\xi) Q(d\xi), \sup_{x \in X} \int_{\Xi} f_j(x,\xi) Q(d\xi) < +\infty, \forall j\right\}$$

and the (pseudo-) distance on $\mathcal{P}_{\mathcal{F}}$

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

At first sight the set $\mathcal{P}_{\mathcal{F}}$ seems to have a complicated structure. For typical applications, however, like for linear two-stage and chance constrained models, the sets $\mathcal{P}_{\mathcal{F}}$ or appropriate subsets allow a simple characterization, for example, as subsets of $\mathcal{P}(\Xi)$ satisfying certain moment conditions.

Proposition: We consider (SP) for $P \in \mathcal{P}_{\mathcal{F}}$, assume that X is compact and

(i) the function $x \to \int_{\Xi} f_0(x,\xi)P(d\xi)$ is Lipschitz continuous on X,

(ii) the set-valued mapping $y \rightrightarrows \left\{x \in X : \int_{\Xi} f_1(x,\xi)P(d\xi) \leq y\right\}$ satisfies the Aubin property at $(0,\bar{x})$ for each $\bar{x} \in S(P)$.

Then there exist constants $L > 0$ and $\delta > 0$ such that the estimates

$$\begin{aligned} |v(P) - v(Q)| &\leq L\, d_{\mathcal{F}}(P,Q) \\ \sup_{x \in S(Q)} d(x, S(P)) &\leq \Psi_P(L\, d_{\mathcal{F}}(P,Q)) \end{aligned}$$

hold whenever $Q \in \mathcal{P}_{\mathcal{F}}$ and $d_{\mathcal{F}}(P,Q) < \delta$. The real-valued function Ψ_P is given by $\Psi_P(r) = r + \psi_P^{-1}(2r)$ for all $r \in \mathbb{R}_+$, where ψ_P is the growth function

$$\psi_P(\tau) = \inf_{x \in X} \left\{ \int_{\Xi} f_0(x,\xi)P(d\xi) - v(P) \,:\, d(x, S(P)) \geq \tau,\, x \in X, \right. \\ \left. \int_{\Xi} f_1(x,\xi)P(d\xi) \leq 0 \right\}.$$

Note that in case $f_1 \equiv 0$ the estimates hold for $L = 1$ and any $\delta > 0$ and that Ψ_P is lower semicontinuous and increasing on $\mathbb{R}_+$ with $\Psi_P(0) = 0$.

The stability result suggests to choose discrete approximations from $\mathcal{P}_n(\Xi)$ for solving (SP) such that they solve the best approximation problem

$$\text{(OSG)} \qquad \min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n)\,.$$

Determining the scenarios of some solution to (OSG) may be called optimal scenario generation. This choice of discrete approximations was already suggested in (Römisch 03), but also characterized there as challenging task which is not solvable in most cases in reasonable time.

It was suggested in (Rachev-Römisch 02) to eventually enlarge the function class $\mathcal{F}$ such that $d_{\mathcal{F}}$ becomes a metric distance and has further nice properties. Following this suggestion, however, may lead to nonconvex nondifferentiable minimization problems (OSG) for determining the optimal scenarios and to unfavorable convergence rates of the sequence

$$\left(\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \right)_{n \in \mathbb{N}}.$$

A typical example is the choice of $\mathcal{F}$ as the unit ball in the Banach space of Lipschitz functions on Ξ equipped with the Lipschitz norm $\|\cdot\|_L$ which refers to the smallest Lipschitz modulus.

Monte Carlo, Quasi-Monte Carlo and optimal quantization

Monte Carlo: Let $\xi^i(\cdot)$, $i \in \mathbb{N}$, denote independent and identically distributed random vectors with common distribution P and P_n be the empirical measure

$$P_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi^i(\cdot)} \quad (n \in \mathbb{N})$$

defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The law of large numbers implies that the sequence $(P_n(\cdot))_{n\in\mathbb{N}}$ converges $\mathbb{P}$-almost surely weakly to P.
To study the convergence rate one considers the empirical process

$$\{\beta_n(P_n(\cdot) - P)f\}_{f\in\mathcal{F}} \;\; (n \in \mathbb{N})$$

indexed by a function class $\mathcal{F}$ with sequence (β_n), where $Qf = \int_\Xi f(\xi)Q(d\xi)$ for any Borel probability measure Q on Ξ. The latter is called bounded in probability with tail function $\tau_\mathcal{F}$ if for all $\varepsilon > 0$ and $n \in \mathbb{N}$ the estimate

$$\mathbb{P}(\{\beta_n d_\mathcal{F}(P_n(\cdot), P) \geq \varepsilon\}) \leq \tau_\mathcal{F}(\varepsilon)$$

holds. Whether the empirical process is bounded in probability, depends on the size of the class $\mathcal{F}$ measured in terms of covering numbers in $L_2(\Xi, P)$. Typically, on has an exponential tail $\tau_\mathcal{F}(\varepsilon) = C(\varepsilon)\exp(-\varepsilon^2)$ and $\beta_n = \sqrt{n}$.

Quasi-Monte Carlo: The basic idea of Quasi-Monte Carlo (QMC) methods is to use deterministic points that are (in some way) uniformly distributed in $[0,1]^d$ and to consider first the approximate computation of

$$I_d(f) = \int_{[0,1]^s} f(\xi) d\xi$$

by a QMC algorithm with (non-random) points ξ^i, $i = 1, \ldots, n$, from $[0,1]^s$:

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

The uniform distribution property of point sets may be defined in terms of the so-called L_p-discrepancy of $\xi^1, \ldots, \xi^n$ for $1 \leq p \leq \infty$

$$d_{p,n}(\xi^1, \ldots, \xi^n) = \left(\int_{[0,1]^s} |\mathrm{disc}(\xi)|^p d\xi \right)^{\frac{1}{p}}, \quad \mathrm{disc}(\xi) := \prod_{j=1}^{d} \xi_j - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,\xi)}(\xi^i) \, .$$

A sequence $(\xi^i)_{i \in \mathbb{N}}$ is called uniformly distributed in $[0,1]^s$ if

$$d_{p,n}(\xi^1, \ldots, \xi^n) \to 0 \quad \text{for} \quad n \to \infty$$

There exist sequences (ξ^i) in $[0,1]^s$ such that for all $\delta \in (0, \frac{1}{2}]$

$$d_{\infty,n}(\xi^1, \ldots, \xi^n) = O(n^{-1} (\log n)^s) \quad \text{or} \quad d_{\infty,n}(\xi^1, \ldots, \xi^n) \leq C(d, \delta) n^{-1+\delta} \, .$$

Optimal quantization: Determine the best approximation to P from $\mathcal{P}_n(\Xi)$ with respect to the L_p-Wasserstein or L_p-minimal metric ℓ_p, $1 \le p < \infty$,

$$\ell_p(P,Q) = \inf\left\{\left(\int_{\Xi\times\Xi} \|\xi-\tilde{\xi}\|^p \eta(d\xi, d\tilde{\xi})\right)^{\frac{1}{p}} : \eta\pi_1^{-1} = P,\ \eta\pi_2^{-1} = Q\right\}.$$

Due to the Kantorovich-Rubinstein duality theorem it holds

$$\min_{P_n\in\mathcal{P}_n(\Xi)} \ell_p(P,P_n) \quad \leftrightarrow \quad \min_{\xi\in\Xi^n} \varphi_{p,n}(\xi^1,\ldots,\xi^n) = \int_\Xi \min_{i=1,\ldots,n} \|\xi-\xi^i\|^p P(d\xi),$$

where ξ^i, $i = 1,\ldots,n$, are the scenarios and $\|\cdot\|$ is a norm in $\mathbb{R}^s$.

It is known (Graf-Luschgy 2000) that $\varphi_{p,n}$ is continuous on Ξ^n and has one-sided directional derivatives into all directions for all $n \in \mathbb{N}$ and any norm. Moreover, it is nonconvex in general for $n \ge 2$, but minima exist in Ξ^n for all $n \in \mathbb{N}$.

Furthermore, due to a classical result by (Dudley 69), the estimate

$$c\,n^{-\frac{1}{s}} \le \ell_1(P,P_n) \le \ell_p(P,P_n)$$

holds for each $P_n \in \mathcal{P}_n(\Xi)$, sufficiently large n and some constant $c > 0$ if P has a density on Ξ. The convergence rate $O(n^{-\frac{1}{s}})$ is clearly worse than the Monte Carlo rate $O(n^{-\frac{1}{2}})$ if $s > 2$.

Optimal scenario generation for linear two-stage models

We consider linear two-stage stochastic programs as introduced earlier and impose the following conditions:

(A0) X is a bounded polyhedron and Ξ is convex polyhedral.
(A1) $h(x,\xi) \in W(Y)$ and $q(\xi) \in D$ are satisfied for every pair $(x,\xi) \in X \times \Xi$,
(A2) P has a second order absolute moment.
Then the infima $v(P)$ and $v(P_n)$ are attained and the estimate

$$\begin{aligned} |v(P) - v(P_n)| &\leq \sup_{x \in X} \left| \int_\Xi f_0(x,\xi) P(d\xi) - \int_\Xi f_0(x,\xi) P_n(d\xi) \right| \\ &= \sup_{x \in X} \left| \int_\Xi \Phi(q(\xi), h(x,\xi)) P(d\xi) - \int_\Xi \Phi(q(\xi), h(x,\xi)) P_n(d\xi) \right| \end{aligned}$$

holds due to the stability result for every $P_n \in \mathcal{P}_n(\Xi)$.
Hence, an appropriate formulation of the optimal scenario generation problem (OSG) in this case is: Determine $P_n^* \in \mathcal{P}_n(\Xi)$ such that it solves the best uniform approximation problem

$$\min_{(\xi^1,\ldots,\xi^n) \in \Xi^n} \sup_{x \in X} \left| \int_\Xi \Phi(q(\xi), h(x,\xi)) P(d\xi) - \frac{1}{n} \sum_{i=1}^n \Phi(q(\xi^i), h(x,\xi^i)) \right|.$$

The class of functions $\{\Phi(q(\cdot), h(x,\cdot)) : x \in X\}$ from Ξ to $\overline{\mathbb{R}}$ enjoys specific properties. All functions are finite, continuous and piecewise linear-quadratic on Ξ. They are linear-quadratic on each convex polyhedral set

$$\Xi_j(x) = \{\xi \in \Xi : (q(\xi), h(x,\xi)) \in \mathcal{K}_j\} \quad (j = 1, \ldots, \ell),$$

where the convex polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, represent a decomposition of the domain of Φ, which is itself a convex polyhedral cone in $\mathbb{R}^{\bar{m}+r}$.

Theorem: Assume (A0)–(A2). Then (OSG) is equivalent to the generalized semi-infinite program

$$\min_{t \geq 0, (\xi^1,\ldots,\xi^n) \in \Xi^n} \left\{ t \;\middle|\; \begin{array}{c} \frac{1}{n}\sum_{i=1}^n \langle h(x,\xi^i), z_i\rangle \leq t + F_P(x) \\ F_P(x) \leq t + \frac{1}{n}\sum_{i=1}^n \langle q(\xi^i), y_i\rangle \\ \forall (x,y,z) \in \mathcal{M}(\xi^1,\ldots,\xi^n) \end{array} \right\},$$

where the set $\mathcal{M} = \mathcal{M}(\xi^1,\ldots,\xi^n)$ and the function $F_P : X \to \mathbb{R}$ are given by

$$\mathcal{M} = \{(x,y,z) \in X \times Y^n \times \mathbb{R}^{rn} : Wy_i = h(x,\xi^i), W^\top z_i - q(\xi^i) \in Y^*, \forall i\},$$

$$F_P(x) := \int_\Xi \Phi(q(\xi), h(x,\xi)) P(d\xi).$$

The latter is the convex expected recourse function of the two-stage model.

Generalized semi-infinite programming

Generalized semi-infinite optimization problems are of the form

$$\min\{f(x) : x \in M\} \quad \text{with} \quad M = \{x \in \mathbb{R}^n : g_i(x, y) \leq 0,\ y \in Y(x),\ i \in I\},$$

where

$$Y(x) = \{y \in \mathbb{R}^m : h_j(x, y) \leq 0,\ j \in J\}$$

and all functions f, g_i, $i \in I$, h_j, $j \in J$, are real-valued and continuous and I and J are finite index sets.
Moreover, the set-valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is assumed to be locally bounded. The latter implies that $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is upper semicontinuous.

Proposition: (Stein 03)
Let $\mathcal{G}_i = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x, y) \leq 0\}$ and
$\mathcal{Y} = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : h_j(x, y) \leq 0,\ j \in J\}$.
Then $M = \bigcap_{i \in I} [\mathrm{pr}_x(\mathcal{Y} \cap \mathcal{G}_i^c)]^c$, where A^c denotes the set complement of a set A.

Remark: If g_i, $i \in I$, and h_j, $j \in J$, are affine in (x, y), $\mathcal{Y}$ is a polyhedron and $\mathcal{G}_i^c$ are open halfspaces. Hence, M may not be closed even in this case.

Proposition: (Stein 03)
M is closed if, in addition, the set-valued mapping Y is lower semicontinuous.

Proposition: (Still 01)
Assume that g_i, $i \in I$, are convex in (x, y) on $\mathbb{R}^{n+m}$ and that for all x, $\tilde{x}$ in $\mathbb{R}^n$ and $0 < \alpha < 1$ holds that

$$Y(\alpha x + (1-\alpha)\tilde{x}) \subseteq \alpha Y(x) + (1-\alpha) Y(\tilde{x}).$$

Then the feasible set M is convex.

Reformulations:
(i) If $Y(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$, the generalized semi-infinite program is equivalent to the bilevel optimization problem

$$\min\{f(x) : x \in \mathbb{R}^n,\ g_i(x, y) \leq 0,\ i \in I,\ y \in \arg\min\{F(x, y) : y \in Y(x)\}\}$$

by setting

$$F(x, y) = \max_{i \in I} g_i(x, y)\,.$$

Observe that $F(x, y) \leq 0$ is equivalent with $g(x, y) \leq 0$.

(ii) MPEC reformulation: (Stein 03)

$$\min_{x\in X}\{f(x) : g_i(x,y^i) \le 0, \nabla_y L_i(x,y^i,\lambda^i) = 0, 0 \le -h(x,y^i) \perp \lambda^i \ge 0, i \in I\},$$

where L_i is the Lagrangian of the ith lower level problem

$$(Q^i(x)) \qquad \max\{g_i(x,y) : y \in Y(x)\},$$

i.e., $L_i(x,y^i,\lambda^i) = g_i(x,y^i) + \langle \lambda^i, h(x,y^i)\rangle$, $i \in I$, and the lower level problems are convex for all $x \in \mathbb{R}^n$ and $i \in I$. However, the MPEC is degenerate since the Mangasarian-Fromovitz constraint qualification is violated everywhere in the feasible set.

(iii) Lifted lower level Wolfe duality reformulation:

$$\min_{x\in X}\{f(x) : L_i(x,y^i,\lambda^i) \le 0, \nabla_y L_i(x,y^i,\lambda^i) = 0, \lambda^i \ge 0, i \in I\},$$

which is a non-degenerate reformulation under the same assumptions as above.

(Diehl-Houska-Stein-Steuermann 13)

Convexity of optimal scenario generation for two-stage models

Theorem:
Let the function h be affine and assume (A0)–(A2).
Then the set-valued mapping $\mathcal{M} : \Xi^n \rightrightarrows \mathbb{R}^m \times \mathbb{R}^{\bar{m}n} \times \mathbb{R}^{rn}$ has convex polyhedral graph and is Hausdorff Lipschitz continuous on Ξ^n.
The feasible set M is closed and convex.
$\mathcal{M}$ is locally bounded if, in addition, $\ker W = \{0\}$ and the dual feasible set $\{z \in \mathbb{R}^r : W^\top z - q(\xi) \in Y^\star\}$ is bounded for each $\xi \in \Xi$.

We note that $F_P(x)$ can only be calculated approximately even if the probability measure P is completely known. For example, this could be done by Monte Carlo or Quasi-Monte Carlo methods with a large sample size $N > n$. Let

$$F_P(x) \approx \frac{1}{N} \sum_{j=1}^{N} \Phi(q(\hat{\xi}^j), h(x, \hat{\xi}^j))$$

be such an approximate representation of $F_P(x)$ based on a sample $\hat{\xi}^j$, $j = 1, \ldots, N$. Hence, in a sense (OSG) may be characterized as scenario clustering problem.

Solution approach to optimal scenario generation

Polyhedrality and Hausdorff Lipschitz contimuity of Y offer the applicability of a discretization method, i.e., of determining a set

$$\mathcal{M}_k(\xi^1,\ldots,\xi^n) = \{(x^j, y^j(\xi^1,\ldots,\xi^n), z^j(\xi^1,\ldots,\xi^n)) : j \in J_k\}$$

of vertices of $\mathcal{M}(\xi^1,\ldots,\xi^n)$, by exchanging and augmenting vertices for increasing k and by determining solutions $(\xi^{k,1},\ldots,\xi^{k,n})$ of

$$\min_{t\geq 0,(\xi^1,\ldots,\xi^n)\in\Xi^n} \left\{ t \,\middle|\, \begin{array}{c} \frac{1}{n}\sum_{i=1}^n \langle h(x,\xi^i), z_i\rangle \leq t + F_P(x) \\ F_P(x) \leq t + \frac{1}{n}\sum_{i=1}^n \langle q(\xi^i), y_i\rangle \\ \forall (x,y,z) \in \mathcal{M}_k(\xi^1,\ldots,\xi^n) \end{array} \right\},$$

which represents a linear program.

Theorem: (Proof based on (Still 01))

Assume (A0)–(A2), let h be affine, Ξ be compact and $\mathcal{M}$ be locally bounded. Assume that

$$\lim_{k\to\infty} d(\mathcal{M}_k(\xi^1,\ldots,\xi^n), \mathcal{M}(\xi^1,\ldots,\xi^n)) = 0 \quad \text{uniformly on } \Xi^n .$$

Then the sequence $((\xi^{k,1},\ldots,\xi^{k,n}))_{k\in\mathbb{N}}$ has an accumulation point in Ξ^n and each such point solves (OSG).

Example: The newsboy problem

A newsboy must place a daily order for a number x of copies of a newspaper. He has to pay r dollars for each copy and sells a copy at c dollars, where $0 < r < c$. The daily demand ξ is a real random variable with (discrete) probability distribution $P \in \mathcal{P}(\mathbb{N})$, $\Xi = \mathbb{R}$, and the remaining copies $y(\xi) = \max\{0, x - \xi\}$ have to be removed. The newsboy might wish that decision x maximizes his expected profit or, equivalently, minimizes his expected costs, i.e.,

$$f_0(x, \xi) = (r - c)x + c\max\{0, x - \xi\} \quad ((x, \xi) \in \mathbb{R} \times \mathbb{R}).$$

The model may be reformulated as a linear two-stage stochastic program with the optimal value function $\Phi(t) = \max\{0, -t\}$. Starting from

$$\Phi(t) = \inf\{\langle q, y\rangle : Wy = t,\ y \geq 0\} = \sup\{\langle t, z\rangle : W^\top z \leq q\}$$

with $W = (w_{11}, w_{12})$ and $q = (q_1, q_2)^\top$, we choose $W = (-1, 1)$, $q = (0, c)$, $h(x, \xi) = \xi - x$, obtain $\{z \in \mathbb{R} : -z \leq 0, z \leq c\} = [0, c]$, and

$$\int_{\mathbb{R}} f_0(x, \xi) dP(\xi) = rx - cx \sum_{\substack{k \in \mathbb{N} \\ k \geq x}} \pi_k - \sum_{\substack{k \in \mathbb{N} \\ k < x}} \pi_k k\,,$$

where π_k is the probability of demand $k \in \mathbb{N}$. The unique (integer) solution is the minimal $k \in \mathbb{N}$ such that $\sum_{i=k}^{\infty} \pi_i \geq \frac{r}{c}$.

The corresponding optimal scenario generation problem (OSG) is of the form

$$\min_{t\geq 0,(\xi^1,\dots,\xi^n)\in\mathbb{R}^n}\left\{ t \;\middle|\; \begin{array}{c} \frac{1}{n}\sum_{i=1}^n(\xi^i - x)z_i \leq t + F_P(x) \\ F_P(x) \leq t + \frac{c}{n}\sum_{i=1}^n y_{2i} \\ \forall (x,y,z)\in\mathbb{R}_+\times\mathbb{R}_+^{2n}\times\mathbb{R}^n: \\ y_{2i}-y_{1i}=\xi^i - x,\; 0\leq z_i\leq c,\; i=1,\dots,n \end{array}\right\},$$

where

$$F_P(x)=\sum_{k=1}^{\infty}\pi_k\, c\,\max\{0,x-k\}\,.$$

If $\xi^i - x \geq 0$ one has $y_{2i}=\xi^i - x$, $y_{1i}=0$, else in case $\xi^i - x\leq 0$, one has $y_{2i}=0$, $y_{1i}=-(\xi^i - x)$. Hence, (OSG) is equivalent with

$$\min_{t\geq 0,(\xi^1,\dots,\xi^n)\in\mathbb{R}^n}\left\{ t \;\middle|\; \begin{array}{c} \frac{c}{n}\sum_{i=1}^n \max\{0,x-\xi^i\}\leq t+F_P(x) \\ F_P(x)\leq t+\frac{c}{n}\sum_{i=1}^n\max\{0,x-\xi^i\} \\ \forall x\in\mathbb{R}_+ \end{array}\right\}.$$

and

$$\min_{(\xi^1,\dots,\xi^n)\in\mathbb{R}^n}\sup_{x\in\mathbb{R}_+}\left| F_P(x)-\frac{c}{n}\sum_{i=1}^n\max\{0,x-\xi^i\}\right|.$$

Conclusions

- Quantitative stability results motivate the best uniform approximation of the underlying probability distribution with respect to discrete measures from $\mathcal{P}_n(\Xi)$ and the minimal function class $\mathcal{F}$.

- Optimal scenario generation for two-stage models are reformulated as a convex generalized semi-infinite optimization model.

- Discretization and exchange methods seem to be favorable for such optimal scenario generation problems. They require the solution of a number of linear programs.

- The elaboration of an exchange method, numerical tests and comparisons with randomized QMC are planned as next step.

References

M. Diehl, R. Houska, O. Stein and P. Steuermann: A lifting method for generalized semi-infinite programs based on lower level Wolfe duality, *Computational Optimization and Applications* 54 (2013), 189–210.

R. M. Dudley: The speed of mean Glivenko-Cantelli convergence, *The Annals of Mathematical Statistics* 40 (1969), 40–50.

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

F. Guerra Vázquez, J.-J. Rückmann, O. Stein and G. Still: Generalized semi-infinite programming: A tutorial, *Journal of Computational and Applied Mathematics* 217 (2008), 394–419.

H. Leövey and W. Römisch: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, *Mathematical Programming* 151 (2015), 315–345.

S. T. Rachev and W. Römisch: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.

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.

O. Stein: *Bi-level Strategies in Semi-infinite Programming*, Kluwer, Boston, 2003.

O. Stein and G. Still: On generalized semi-infinite optimization and bilevel optimization, *European Journal of Operational Research* 142 (2002), 444–462.

G. Still: Generalized semi-infinite programming: Numerical aspects, *Optimization* 49 (2001), 223–242.