INFORMS Optimization Society
2018 Khachiyan Prize

Expository Article:
Stochastic Programming: Approximations and Scenarios

Werner Römisch
Institute of Mathematics
Humboldt-University Berlin
romisch@math.hu-berlin.de

Published online by the INFORMS Optimization Society on May 24, 2021.
http://connect.informs.org/optimizationsociety/home

It is truly a great honor to have been awarded the 2018 Khachiyan prize of the INFORMS Optimization Society. I wish to thank the members of the prize committee and my colleagues and friends for nominating me. I would like to take this opportunity to briefly sketch essential parts of my interest in the area of stochastic programming. In 1976 I finished my dissertation on theory and approximations of random equations at Humboldt-University Berlin. Shortly after that I listened there to a lecture by Peter Kall (University of Zürich) on some recent advances in stochastic programming in the late seventies. Having in mind the progress of parametric programming achieved at that time and later in my department [3], I got interested in connecting both areas. During the International Conference on Stochastic Programming at Köszeg (Hungary) in 1981 I obtained further inspiration on approximation concepts in optimization from Roger Wets (UC Davis) and met my future cooperators Jitka Dupačová (Charles University Prague) and Georg Pflug (University of Vienna) for the first time. Additional motivation appeared later when practical electricity optimization models under uncertainty [4, 9, 27] required both the statistical estimation of the underlying probability distribution from available data and their numerical solution based on scenario approximations. The following is a short version of my presentation in the prize session.

1 Introduction

A number of stochastic programming models can be cast into the general form

$$\min \left\{ \int_{\Xi} f_0(x,\xi)P(d\xi) : x \in X,\ \int_{\Xi} f_1(x,\xi)P(d\xi) \leq 0 \right\} \tag{1}$$

where X is a closed subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^d$, 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, where we adopt here and in the following the notation from the monograph [33].

Many approaches to the computational solution of (1) are based on finding a discrete probability measure P_n belonging to

$$\mathcal{P}_n(\Xi) := \left\{ \sum_{i=1}^{n} p_i \delta_{\xi^i} : \xi^i \in \Xi,\ p_i \geq 0,\ i = 1, \ldots, n,\ \sum_{i=1}^{n} p_i = 1 \right\} \tag{2}$$

for some $n \in \mathbb{N}$, which approximates P at least such that the corresponding infima are close. The Dirac measures δ_{ξ^i} in (2) place unit mass at the atoms ξ^i, $i = 1, \ldots, n$, of P_n. The latter are often called scenarios in this context.

Typical integrands f_0 and f_1 appear in two-stage stochastic programming and in chance constrained models.

Example 1. Linear two-stage stochastic programs:

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

where X and Ξ are convex polyhedral, $g(\cdot)$ is a linear function, $q(\cdot)$ is affine, $h(\cdot,\xi)$ 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) program

$$\Phi(q,t) := \inf\{\langle q, y\rangle : Wy = t, y \geq 0\}$$

with W denoting the $(s, \bar{m})$ recourse matrix W and $D = \{q \in \mathbb{R}^{\bar{m}} : \{z \in \mathbb{R}^s : W^\top z \leq q\} \neq \emptyset\}$ the convex polyhedral dual feasibility set. We note that already the evaluation of the objective function of two-stage models is known to be #P-hard in general if P is a continuous multivariate probability distribution [12].

Example 2. Chance constrained programs:

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

where $p \in (0,1)$ is a probability level and $\mathbb{1}_{\mathfrak{P}(x)}$ is the characteristic function of the polyhedron $\mathfrak{P}(x) = \{\xi \in \Xi : h(x, \xi) \leq 0\}$ depending on x, where Ξ and h have the same properties as in the preceding example.

The general model (1) covers a great variety of stochastic optimization problems. In addition to both examples, (1) contains mixed-integer two-stage stochastic programs [39], optimization problems with stochastic dominance constraints [5], problems containing risk functionals [29] in objective and constraints etc. For more information on stochastic programming we refer to the monographs [19, 40].

2 Stability-based scenario generation

Let $v(P)$ and $S(P)$ denote the infimum and solution set of (1). We are interested in conditions implying their continuous dependence on the underlying probability distribution P in terms of an appropriate distance for probability measures. To this end we introduce the following sets of functions and of probability distributions (both defined on Ξ)

$$\begin{aligned} \mathcal{F} &= \left\{ f_j(x, \cdot\,) : j = 0, 1,\, x \in X \right\}, \\ \mathcal{P}_{\mathcal{F}} &= \left\{ Q \in \mathcal{P}(\Xi) : \sup_{x \in X} \int_{\Xi} |f_j(x, \xi)| Q(d\xi) < +\infty,\, j = 0, 1 \right\}, \end{aligned}$$

and the (semi-) distance on $\mathcal{P}_{\mathcal{F}}$ defined by

$$d_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} \left| \int_{\Xi} f(\xi) P(d\xi) - \int_{\Xi} f(\xi) Q(d\xi) \right| \quad (P, Q \in \mathcal{P}_{\mathcal{F}}). \tag{3}$$

The mapping $d_{\mathcal{F}} : \mathcal{P}_{\mathcal{F}} \times \mathcal{P}_{\mathcal{F}} \to \overline{\mathbb{R}}$ is finite, non-negative, symmetric and satisfies the triangle inequality, but $d_{\mathcal{F}}(P, Q) = 0$ does not imply $P = Q$ in general unless the class $\mathcal{F}$ is rich enough. For typical applications like for linear two-stage and chance constrained models, the sets $\mathcal{P}_{\mathcal{F}}$ or appropriate subsets allow a simpler characterization, for example, as subsets of $\mathcal{P}(\Xi)$ satisfying certain moment conditions. Next we state a result on continuity properties of infima and solutions sets with respect to the distance $d_{\mathcal{F}}$. It is a simplified version of a general result in [32].

Proposition 1. *Consider* (1) *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

$$
\begin{aligned}
|v(P) - v(Q)| &\leq L\, d_{\mathcal{F}}(P, Q) && (4)\\
\sup_{x \in S(Q)} d(x, S(P)) &\leq \Psi_P(L\, d_{\mathcal{F}}(P, Q)) && (5)
\end{aligned}
$$

hold whenever $Q \in \mathcal{P}_{\mathcal{F}}$ and $d_{\mathcal{F}}(P, Q) < \delta$. Here $d(x, S(P))$ denotes the distance of x to the solution set $S(P)$ of (1) *and the function Ψ_P is given by $\Psi_P(t) = t + \psi_P^{-1}(2t)$ for all $t \in \mathbb{R}_+$, where ψ_P is the growth function*

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

In case $f_1 \equiv 0$ only lower semicontinuity is needed in (i) and the estimates (4) *and* (5) *hold with $L = 1$ and for any $\delta > 0$. Furthermore, Ψ_P is lower semicontinuous and increasing on $\mathbb{R}_+$ with $\Psi_P(0) = 0$.*

We note that all assumptions refer to the original probability distribution, where we tried to avoid restrictive differentiability assumptions on objective and constraint functions. Condition (ii), for example, represents a constraint qualification for the original problem (1) (see also [33, Section 9F]).

The stability result suggests to choose discrete approximations from $\mathcal{P}_n(\Xi)$ for solving (1) such that they solve the best approximation problem

$$\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \tag{6}$$

at least approximately. Determining the scenarios of some solution to (6) may be called optimal scenario generation. This optimal choice of discrete approximations is computationally challenging and hard to compute in general. However, for linear two-stage models we argue later that problem (6) allows a reformulation as linear semi-infinite program in some cases. Classical discrete approximations in stochastic programming are based on Monte Carlo approximations for which P_n are random discrete measures. Its convergence properties are well understood due to the well known properties of empirical processes [34, 40, 41]. More recently, (randomized) Quasi-Monte Carlo methods [6, 22], optimal quantization techniques [28], sparse grid quadrature rules [10] and moment-based methods [25] have been studied.

In [32] it is suggested to eventually enlarge the function class $\mathcal{F}$ such that $d_{\mathcal{F}}$ becomes a metric distance and has further nice properties. This may lead, however, to coarse upper bounds in (4) and (5), to challenging minimization problems (6) for determining the optimal scenarios and to unfavorable convergence rates of

$$\left(\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \right)_{n \in \mathbb{N}}. \tag{7}$$

Typical examples are bounded subsets $\mathcal{F}$ of Banach spaces $C^{k,\alpha}(\Xi)$ of functions which are k-times differentiable and all kth order derivatives are Hölder continuous with exponent α for $k \in \mathbb{N}_0$, $\alpha \in (0, 1]$. For such sets $\mathcal{F}$ the convergence rate of (7) is at best $O(n^{-\frac{k+\alpha}{d}})$ where $d \in \mathbb{N}$ is the dimension of Ξ [26].

3 The road of probability metrics

For linear two-stage models the function $f_0(x,\cdot)$ is locally Lipschitz continuous in general. Hence, one might resort to a bounded subset of a space of locally Lipschitz continuous functions as enlarged class $\mathcal{F}$. This idea leads to the so-called Fortet-Mourier metrics (of order $r \geq 1$)

$$\zeta_r(P,Q) = d_{\mathcal{F}_r(\Xi)}(P,Q) = \sup_{f\in\mathcal{F}_r(\Xi)} \left| \int_\Xi f(\xi)P(d\xi) - \int_\Xi f(\xi)Q(d\xi) \right|, \tag{8}$$

where the function class $\mathcal{F}_r(\Xi)$ is given by

$$\begin{aligned} \mathcal{F}_r(\Xi) &= \left\{ f : \Xi \mapsto \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq c_r(\xi,\tilde{\xi}),\ \forall \xi,\tilde{\xi} \in \Xi \right\}, \\ c_r(\xi,\tilde{\xi}) &= \max\{1, \|\xi\|, \|\tilde{\xi}\|\}^{r-1} \|\xi - \tilde{\xi}\| \quad (\xi,\tilde{\xi} \in \Xi). \end{aligned}$$

Such Fortet-Mourier metrics admit a dual representation as Kantorovich-Rubinstein functional (see [31]). If the so-called reduced cost function $\hat{c}_r$ is defined by

$$\hat{c}_r(\xi,\tilde{\xi}) = \inf \left\{ \sum_{i=0}^{k-1} c_r(\xi^{l_i}, \xi^{l_{i+1}}) : k \in \mathbb{N}, \xi^{l_i} \in \Xi, i = 0,\ldots,k, \xi^{l_0} = \xi, \xi^{l_k} = \tilde{\xi} \right\},$$

all functions $f \in \mathcal{F}_r$ satisfy $f(\xi) - f(\tilde{\xi}) \leq \hat{c}_r(\xi,\tilde{\xi})$, $\forall \xi,\tilde{\xi} \in \Xi$, and $\hat{c}_r$ is a metric on Ξ satisfying $\hat{c}_r \leq c_r$. Hence, the next result is a consequence of [31, Section 6.1].

Proposition 2. *The Fortet-Mourier metric* (8) *admits the dual representation*

$$\zeta_r(P,Q) = \inf \left\{ \int_{\Xi\times\Xi} \hat{c}_r(\xi,\tilde{\xi}) \eta(d\xi, d\tilde{\xi}) : \eta \circ \pi_1^{-1} = P, \eta \circ \pi_2^{-1} = Q \right\} \tag{9}$$

as Monge-Kantorovich transportation problem based on reduced costs if the metric space $(\Xi, \hat{c}_r)$ is separable. The latter is true for $r = 1$ and for $r > 1$ if Ξ is bounded.

Using Fortet-Mourier metrics the problem (6) of optimal scenario generation reads

$$\min_{P_n \in \mathcal{P}_n(\Xi)} \zeta_r(P, P_n) \tag{10}$$

or, equivalently,

$$\min_{(\xi^1,\ldots,\xi^n)\in\Xi^n} \int_\Xi \min_{j=1,\ldots,n} \hat{c}_r(\xi,\xi^j) P(d\xi).$$

The function $(\xi^1,\ldots,\xi^n) \mapsto \int_\Xi \min_{j=1,\ldots,n} \hat{c}_r(\xi,\xi^j)P(d\xi)$ is continuous on Ξ^n and has compact level sets, but is nonconvex and nondifferentiable in general. Hence, optimal scenarios exist, but their computation seems to be challenging.

Let P itself be discrete with a large number $N \gg n$ of scenarios. In case of $\Xi = \mathbb{R}^d$ the optimal scenario generation problems (6) and (10) are called continuous scenario reduction in [38] while in case $\Xi = \mathrm{supp}(P)$ problems (6) and (10) reduce to the classical (discrete) scenario reduction problem [8].

We consider the latter case and let first P and Q be two general discrete distributions, where ξ^i are the scenarios with probabilities p_i, $i = 1,\ldots,N$, of P and $\tilde{\xi}^j$ the scenarios and q_j, $j = 1,\ldots,n$, the probabilities of Q. If Ξ denotes the union of both scenario sets, we have due to Proposition 2

$$
\begin{aligned}
\zeta_r(P,Q) &= \inf\Big\{\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{ij}\hat{c}_r(\xi^i,\tilde{\xi}^j) : \sum_{j=1}^{n}\eta_{ij}=p_i, \sum_{i=1}^{N}\eta_{ij}=q_j, \eta_{ij}\ge 0, \\
&\qquad\qquad i=1,\dots,N,\ j=1,\dots,n\Big\} \\
&= \sup\Big\{\sum_{i=1}^{N}p_iu_i-\sum_{j=1}^{n}q_jv_j : u_i-v_j\le \hat{c}_r(\xi^i,\tilde{\xi}^j), i=1,\dots,N, j=1,\dots,n\Big\}
\end{aligned}
$$

primal and dual representations of $\zeta_r(P,Q)$ as primal and dual forms of the linear tranportation problem.

If the discrete distribution Q is supported by a subset ξ^j, $j\in J$, of the support of P with probabilities q_j, $j\in J$, $\emptyset\neq J\subset\{1,\dots,N\}$ and $\Xi=\mathrm{supp}(P)$, the best approximation of P with respect to ζ_r by such a distribution Q exists. If it is denoted by Q^* its ζ_r-distance to P is

$$D_J := \zeta_r(P,Q^*) = \sum_{i\notin J} p_i \min_{j\in J}\hat{c}_r(\xi^i,\xi^j) \tag{11}$$

and its probabilities are $q_j^* = p_j + \sum_{i\in I_j} p_i$, $\forall j\in J$, where $I_j := \{i\notin J : j=j(i)\}$ and $j(i)\in \arg\min_{j\in J}\hat{c}_r(\xi^i,\xi^j)$, $\forall i\notin J$ (optimal redistribution [8]).

Determining the optimal scenario index set J with prescribed cardinality n is a so-called metric n-median problem and of the form

$$\min\{D_J : J\subset\{1,...,N\}, |J|=n\}\,. \tag{12}$$

Metric n-median problems and, hence, the problem of finding the optimal index set J of remaining scenarios are known to be NP-hard [20] and polynomial time algorithms for solving (12) are not available.

Problem (12) may be reformulated as combinatorial program

$$
\begin{array}{rcl}
\min & \displaystyle\sum_{i,j=1}^{N} p_i x_{ij}\hat{c}_r(\xi^i,\xi^j) & \text{subject to} \\
\displaystyle\sum_{i=1}^{N} x_{ij} & = & 1 \quad (j=1,\dots,N), \quad \displaystyle\sum_{i=1}^{N} y_i \le n\,, \\
x_{ij} & \le & y_i, \quad y_i\in\{0,1\}, \quad x_{ij}\in\{0,1\} \quad (i,j=1,\dots,N).
\end{array}
$$

The variable y_i decides whether scenario ξ^i remains and x_{ij} indicates whether scenario ξ^j, $j\neq i$, minimizes the $\hat{c}_r$-distance to ξ^i. The combinatorial program could, of course, be tackled by standard software, e.g., by LP-based branch and bound. There is also a well developed theory of polynomial-time approximation algorithms for such programs. The currently available best approximation algorithms for metric n-median problems are local search heuristics [1] and pseudo-approximation [23]. The latter provides an approximation ratio of $1+\sqrt{3}+\varepsilon$. Approximation ratio $\alpha>1$ means that such an algorithm produces for all instances of the program a solution whose value is within a factor of α of the infimum.

The simplest algorithms are so-called greedy heuristics, namely, backward (or reverse) and forward heuristics. In the present context such backward and forward heuristics are described below (see [13, 14]).

Backward and foward starting points for $n = N-1$ and $n = 1$, respectively, are

$$\min_{l\in\{1,\ldots,N\}} p_l \min_{j\neq l} \hat{c}_r(\xi_l,\xi_j) \qquad \text{and} \qquad \min_{u\in\{1,\ldots,N\}} \sum_{k=1}^{N} p_k \hat{c}_r(\xi_k,\xi_u).$$

Backward reduction algorithm:

$$\begin{aligned}
&\textbf{Step [0]:} && J^{[0]} := \emptyset\,.\\
&\textbf{Step [i]:} && l_i \in \arg\min_{l\notin J^{[i-1]}} \sum_{k\in J^{[i-1]}\cup\{l\}} p_k \min_{j\notin J^{[i-1]}\cup\{l\}} \hat{c}_r(\xi_k,\xi_j).\\
& && J^{[i]} := J^{[i-1]} \cup \{l_i\}\,.\\
&\textbf{Step [N-n+1]:} && \text{Optimal redistribution.}
\end{aligned}$$

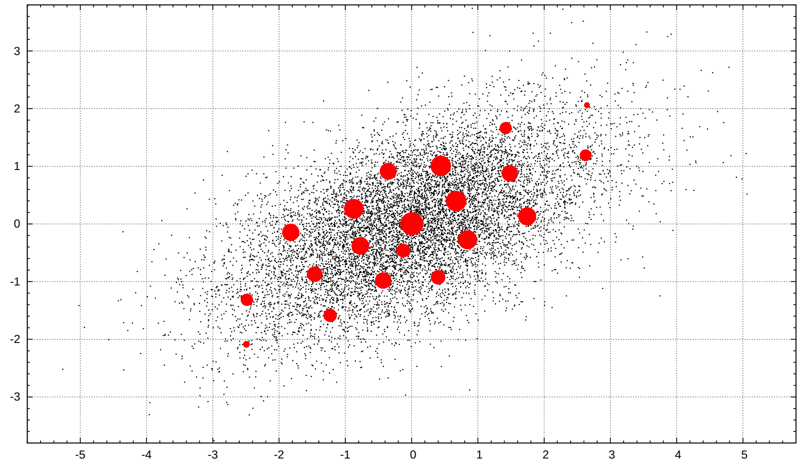
Forward selection algorithm:

$$\begin{aligned}
&\textbf{Step [0]:} && J^{[0]} := \{1,\ldots,N\}.\\
&\textbf{Step [i]:} && u_i \in \arg\min_{u\in J^{[i-1]}} \sum_{k\in J^{[i-1]}\setminus\{u\}} p_k \min_{j\notin J^{[i-1]}\setminus\{u\}} \hat{c}_r(\xi_k,\xi_j),\\
& && J^{[i]} := J^{[i-1]} \setminus \{u_i\}\,.\\
&\textbf{Step [n+1]:} && \text{Optimal redistribution.}
\end{aligned}$$

In case $r = 1$ and if the support of P is contained in the Euclidean unit ball in $\mathbb{R}^d$ it is shown in [38] that the infimum in (10) is less than or equal to $2\sqrt{\frac{N-n}{N-1}}$ where the latter bound is sharp in general. Moreover, it is proved in [38] that the approximation ratio of forward selection for solving (10) is not bounded for unbounded N. Instead, a $(5+\varepsilon)$-approximation algorithm based on [1] is developed and tested there. The authors report that warmstarting the algorithm using forward selection can significantly reduce its runtime. In addition, forward selection worked well in test cases [13] and many practical instances reported in the literature (see, for example, [2, 30] and Example 4). Together with its extension to construct scenario trees for solving multi-stage stochastic programs, forward selection has been implemented in GAMS/Scenred2 by my colleague Holger Heitsch (WIAS Berlin, formerly Humboldt-University Berlin).

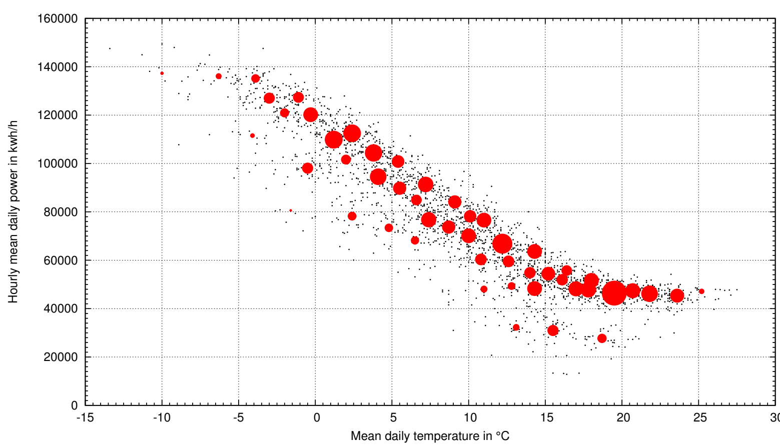
Example 3. Sample reduction for a two-dimensional normal distribution

We consider a two-dimensional normal distribution with correlated components and generate $N = 10^4$ Monte Carlo points with identical probability $\frac{1}{N}$ using the Mersenne Twister [24]. Then the optimal scenario reduction problem (10) for $r = 1$ and $n = 20$ is solved approximately by the forward selection heuristic. The result is shown below where the new probabilities after redistribution are proportional to the diameters of the red balls that represent the remaining scenarios.



Example 4. Generating exit gas flow scenarios

We consider the stationary state of the network of Germany's largest gas transport company. It contains a large number of exit and a low number of entry nodes. In its isothermal case gas pressure and flow satisfy a large system of nonlinear equations and inequalities. A huge number of hourly gas flow data is available at all exit nodes. Temperature classes are introduced and a corresponding filtering of the daily mean gas flows at all exit points is performed according to a daily reference temperature. Based on this a multivariate probability distribution for the gas flow at all exits is estimated. Then $N = 2340$ temperature depending gas flow samples with identical probability $\frac{1}{N}$ are generated as randomized Sobol' points for several hundred exits and later reduced by forward selection for $r = 1$ to $n = 50$ scenarios. The result is shown below for a typical exit where the new probabilities after redistribution are proportional to the diameters of the red balls representing the remaining scenarios. For details we refer the reader to [21, Chapters 13, 14].



While Fortet-Mourier metrics serve as upper bounds for the distance $d_{\mathcal{F}}$ in case of two-stage models, this is not the case for mixed-integer two-stage and chance constrained models. As shown in [17, 35, 36] discrepancy distances of the form

$$\alpha_{\mathcal{B}}(P,Q) = \sup_{B\in\mathcal{B}} |P(B) - Q(B)| \quad (P,Q \in \mathcal{P}(\Xi)), \tag{13}$$

with appropriate classes $\mathcal{B}$ of Borel sets are relevant for both types of models. To study scenario reduction in case of such distances let P be again a discrete distribution with scenarios ξ^i and probabilities p_i, $i = 1, \ldots, N$, and Q be supported by a subset ξ^j of scenarios with probabilities q_j, $j \in J$, where $J \subset \{1, \ldots, N\}$ is an index set with cardinality $|J| = n$. Then (13) admits the representation

$$\alpha_{\mathcal{B}}(P,Q) = \min \left\{ t \;\middle|\; \begin{array}{l} -\sum\limits_{j \in I \cap J} q_j \leq t - \sum\limits_{i \in I} p_i, \; I \in \mathcal{I}_{\mathcal{B}} \\ \sum\limits_{j \in I \cap J} q_j \leq t + \sum\limits_{i \in I} p_i, \; I \in \mathcal{I}_{\mathcal{B}} \end{array} \right\} \tag{14}$$

as linear program, where the set $\mathcal{I}_{\mathcal{B}}$ contains all index sets I_B of scenarios which belong to B for each $B \in \mathcal{B}$, i.e., $\mathcal{I}_{\mathcal{B}} = \{I_B = \{i \in \{1, \ldots, N\} : \xi^i \in B\} : B \in \mathcal{B}\}$. The set $\mathcal{I}_{\mathcal{B}}$ may have up to 2^N elements and is, hence, too large in general for solving the linear program (14) and for determining (t, q) if the set J of remaining scenarios is known. Instead the system $\mathcal{I}^\star_{\mathcal{B}}(J) = \{I \cap J : I \in \mathcal{I}_{\mathcal{B}}\}$ of reduced index sets with at most 2^n elements and the quantities

$$\overline{\gamma}(I^\star) = \max_{I \in \mathcal{I}_{\mathcal{B}}} \left\{ \sum_{i \in I} p_i : I \cap J = I^\star \right\}, \quad \underline{\gamma}(I^\star) = \min_{I \in \mathcal{I}_{\mathcal{B}}} \left\{ \sum_{i \in I} p_i : I \cap J = I^\star \right\}.$$

are introduced for all $I^\star \in \mathcal{I}^\star_{\mathcal{B}}(J)$. Then $\alpha_{\mathcal{B}}(P,Q)$ may be represented as

$$\alpha_{\mathcal{B}}(P,Q) = \min \left\{ t \;\middle|\; \begin{array}{l} -\sum\limits_{j \in I^\star} q_j \leq t - \overline{\gamma}(I^\star), \; I^\star \in \mathcal{I}^\star_{\mathcal{B}}(J) \\ \sum\limits_{j \in I^\star} q_j \leq t + \underline{\gamma}(I^\star), \; I^\star \in \mathcal{I}^\star_{\mathcal{B}}(J) \end{array} \right\}. \tag{15}$$

For classes $\mathcal{B}$ of closed cells (i.e., of all sets $\xi + \mathbb{R}^d_-$), of rectangular (i.e., of all axis-parallel boxes) and of polyhedral subsets (having at most k vertices) of $\mathbb{R}^d$ the concept of supporting sets is developed and its intimate relationship to determining $\mathcal{I}^\star_{\mathcal{B}}(J)$ and $\overline{\gamma}(I^\star)$ and $\underline{\gamma}(I^\star)$, respectively, is unveiled in [15, 16]. Numerical experience for scenario reduction with respect to discrepancy distances is available in [15, 16] only for low dimensions d. It is known that even the computation of the cell discrepancy is NP-hard [7, Section 10.3.3]. We refer to [7, Section 10.3.4] for promising alternative approaches to calculate the cell discrepancy.

4 Problem-based scenario generation for linear two-stage models

Next we show that for linear two-stage stochastic programs the use of enlarged function classes $\mathcal{F}$ for scenario generation and reduction can be avoided. Instead the minimal information distance (3) can be utilized to develop a problem-based approach. To this end, we consider two-stage models with the notation introduced in Example 1 and impose the following conditions:
(A0) X is a bounded polyhedron.
(A1) $h(x, \xi) \in W(\mathbb{R}^{\bar{m}}_+)$ 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 Proposition 1 for every $P_n \in \mathcal{P}_n(\Xi)$. Hence, the optimal scenario generation problem (6) with uniform weights may be reformulated as:

Determine $P_n^* \in \mathcal{P}_n(\Xi)$ such that its scenarios ξ^i, $i = 1, \ldots, n$, with uniform weight $\frac{1}{n}$ solve the best uniform approximation problem

$$\min_{(\xi^1,\ldots,\xi^n)\in\Xi^n} \sup_{x\in X} \left| \int_\Xi \Phi(q(\xi), h(x,\xi))P(d\xi) - \frac{1}{n}\sum_{i=1}^n \Phi(q(\xi^i), h(x,\xi^i)) \right| \tag{16}$$

for the expected recourse function $F_P(x) := \int_\Xi \Phi(q(\xi), h(x,\xi))P(d\xi)$. All functions belonging to the class $\{\Phi(q(\cdot), h(x,\cdot)) : x \in X\}$ from Ξ to $\mathbb{R}$ are finite, continuous and piecewise linear-quadratic on Ξ. They are linear-quadratic on each convex polyhedral set

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

where the convex polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, represent a decomposition of the domain $D \times W(\mathbb{R}_+^{\bar{m}})$ of Φ, which is itself a convex polyhedral cone in $\mathbb{R}^{\bar{m}+s}$. Problem (16) is equivalent to a generalized semi-infinite problem in which the infinite index set of the constraints depends on the decisions.

Proposition 3. *Assume (A0)–(A2). Then* (16) *is equivalent to the generalized semi-infinite program*

$$\min_{t\geq 0, \hat{\xi}=(\xi^1,\ldots,\xi^n)\in\Xi^n} \left\{ t \;\middle|\; \begin{array}{c} \frac{1}{n}\sum\limits_{i=1}^n \langle h(x,\xi^i), z_i\rangle \leq t + F_P(x) \\ F_P(x) \leq t + \frac{1}{n}\sum\limits_{i=1}^n \langle q(\xi^i), y_i\rangle \\ \forall (x,y,z) \in \mathcal{M}(\xi^1,\ldots,\xi^n) \end{array} \right\}, \tag{17}$$

where the constraint-index set mapping $\mathcal{M} : \Xi^n \to \mathbb{R}^{m+(\bar{m}+s)n}$ is given by

$$\mathcal{M}(\hat{\xi}) = \left\{(x,y,z) \in X \times \mathbb{R}_+^{\bar{m}n} \times \mathbb{R}^{sn} : Wy_i = h(x,\xi^i), W^\top z_i \leq q(\xi^i), i = 1, \ldots, n\right\}.$$

We note that theory and numerical methods for generalized semi-infinite programs are well-developed (see the survey [11]). If either right-hand sides or costs of the two-stage model are random, problem (17) is convex and can be further simplified.

Proposition 4. *Assume (A0)–(A2). Let the function h be affine and let either h or q be random. Then the generalized semi-infinite program* (17) *is convex and can be transformed into an equivalent (standard) linear semi-infinite program.*

Of course, the expected recourse function F_P can only be calculated approximately even if the probability measure P is completely known. For numerical computations F_P has to be replaced by its Monte Carlo or Quasi-Monte Carlo approximation with a large sample size $N > n$. For proofs and details it is referred to [18].

Acknowledgement: At different stages of my scientific career I owe much to several persons. I wish to thank my collaborators, many of whom became friends, my Post-docs and PhD students who obtained excellent results and provided a pleasant working environment.

References

[1] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit: Local search heuristics for k-median and facility location problems, *SIAM Journal Computing* 33 (2004), 544–562.

[2] M. Armstrong, A. Ndiaye, R. Razanatsimba, and A. Galli: Scenario reduction applied to geostatistical simulations, *Mathematical Geosciences* 45 (2013), 165–182.

[3] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer: *Nonlinear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.

[4] D. Dentcheva and W. Römisch: Optimal power generation under uncertainty via stochastic programming, in: Stochastic Programming Methods and Technical Applications (K. Marti and P. Kall eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer-Verlag, Berlin, 1998, 22–56.

[5] D. Dentcheva and A. Ruszczyński: Optimization with stochastic dominance constraints, *SIAM Journal Optimization* 14 (2003), 548–566.

[6] J. Dick, F. Y. Kuo, and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.

[7] C. Doerr, M. Gnewuch and M. Wahlström: Calculation of discrepancy and applications, in: *A Panorama of Discrepancy Theory* (W. W. L. Chen, A. Srivastav, and G. Travaglini eds.), Lecture Notes in Mathematics vol. 2107, Springer, 2014, 621–678.

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

[9] A. Eichhorn, W. Römisch, and I. Wegner: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, IEEE St. Petersburg Power Tech 2005.

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

[11] 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.

[12] G. A. Hanasusanto, D. Kuhn, and W. Wiesemann: A comment on "computational complexity of stochastic programming problems", *Mathematical Programming* 159 (2016), 557–569.

[13] H. Heitsch and W. Römisch: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.

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

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

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

[17] R. Henrion and W. Römisch: Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, *Mathematical Programming* 100 (2004), 589–611.

[18] R. Henrion and W. Römisch: Problem-based optimal scenario generation and reduction in stochastic programming, *Mathematical Programming* (appeared online).

[19] P. Kall and J. Mayer: *Stochastic Linear Programming* (Second Edition), Springer, New York, 2010.

[20] O. Kariv and S. L. Hakimi: An algorithmic approach to network location problems, II: The p-medians, *SIAM Journal Applied Mathematics* 37 (1979), 539–560.

[21] T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe (eds.): Evaluating Gas Network Capacities, SIAM-MOS Series on Optimization, Philadelphia, 2015.

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

[23] S. Li and O. Svensson: Approximating k-median via pseudo-approximation, *SIAM Journal on Computing* 45 (2016), 530–547.

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

[25] S. Mehrotra and D. Papp: Generating moment matching scenarios using optimization techniques, *SIAM Journal Optimization* 23 (2013), 963–999.

[26] E. Novak: Quadrature and widths, *Journal Approximation Theory* 47 (1986), 195–202.

[27] M.P. Nowak and W. Römisch: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty, *Annals of Operations Research* 100 (2000), 251–272.

[28] G. Ch. Pflug and A. Pichler: Approximations for probability distributions and stochastic optimization problems, in: *Stochastic Optimization Methods in Finance and Energy* (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.), Springer, 2011, 343–387.

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

[30] S. Pineda and A. J. Conejo: Scenario reduction in risk-averse electricity trading, *IET Generation, Transmission & Distribution* 4 (2010), 694–705.

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

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

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

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

[35] W. Römisch and R. Schultz: Stability analysis for stochastic programs, *Annals of Operations Research* 30 (1991), 241–266.

[36] W. Römisch and S. Vigerske: Quantitative stability of fully random mixed-integer two-stage stochastic programs, *Optimization Letters* 2 (2008), 377–388.

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

[38] N. Rujeerapaiboon, K. Schindler, D. Kuhn, and W. Wiesemann: Scenario reduction revisited: Fundamental limits and guarantees, *Mathematical Programming* (appeared online).

[39] R. Schultz: Stochastic programming with integer variables, *Mathematical Programming* 97 (2003), 285–309.

[40] A. Shapiro, D. Dentcheva, and A. Ruszczyński: *Lectures on Stochastic Programming* (Second Edition). MPS-SIAM Series on Optimization, Philadelphia, 2014.

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