

Optimal scenario generation and reduction in stochastic programming

R. Henrion · W. Römisch

the date of receipt and acceptance should be inserted later

Abstract Scenarios are indispensable ingredients for the numerical solution of stochastic optimization problems. Earlier approaches for optimal scenario generation and reduction are based on stability arguments involving distances of probability measures. In this paper we review those ideas and suggest to make use of stability estimates based on distances containing minimal information, i.e., on data appearing in the optimization model only. For linear two-stage stochastic programs we show that the optimal scenario generation problem can be reformulated as best approximation problem for the expected recourse function and as generalized semi-infinite program, respectively. The latter model turns out to be convex if either right-hand sides or costs are random. We also review the problems of optimal scenario reduction for two-stage models and of optimal scenario generation for chance constrained programs. Finally, we consider scenario generation and reduction for the classical newsvendor problem.

1 Introduction

Most numerical solution approaches in stochastic programming require to replace the underlying multivariate probability distribution by a discrete probability measure with a finite number of realizations or *scenarios*. The mostly used approach so far is Monte Carlo sampling (see, for example, [44, Chapter 6]). Another more classical approach for two-stage models uses discrete probability measures leading to lower and upper bounds for the expected recourse function. They are obtained by means of moment problems (see [25, Section 4.7.2]). More recently optimal quantization techniques (see [15, 33]) and (randomized) Quasi-Monte Carlo methods (see [3, 28, 32]) are employed for solving two-stage stochastic programs. For a survey on scenario generation in stochastic programming we refer to [41].

Jitka Dupačová was one of the pioneers for scenario generation and reduction. We recall her earlier paper [8] and the influential work [9, 10].

Here, we study a problem-based approach to scenario generation and reduction for stochastic programming models without information constraints. A general form of such models is [25, 44, 46]

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

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 (in the sense of [39, Chapter 14]). For example, typical integrands f_0 in linear two-stage stochastic programming models are of the form [52], [44, Chapt. 2]

$$f_0(x, \xi) = \begin{cases} g(x) + \Phi(q(\xi), h(x, \xi)), & q(\xi) \in D \\ +\infty, & \text{else} \end{cases} \quad \text{and} \quad f_1(x, \xi) \equiv 0, \quad (2)$$

where X and Ξ are convex polyhedral, $g(\cdot)$ is a linear function, $q(\cdot)$ is affine, $D = \{q \in \mathbb{R}^m : \{z \in \mathbb{R}^r : W^\top z - q \in Y^*\} \neq \emptyset\}$ 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

$$\Phi(q, t) := \inf \{ \langle q, y \rangle : Wy = t, y \in Y \} \quad (3)$$

with (\bar{m}, r) matrix W and convex polyhedral cone $Y \subset \mathbb{R}^{\bar{m}}$. Another example of practical interest is that Φ is the infimal function of a linear-quadratic optimization problem. Typical integrands f_1 appearing in chance constrained programming are of the form $f_1(x, \xi) = p - \mathbb{1}_{\mathcal{P}(x)}(\xi)$, where $\mathbb{1}_{\mathcal{P}(x)}$ is the characteristic function of the polyhedron $\mathcal{P}(x) = \{\xi \in \Xi : h(x, \xi) \leq 0\}$ depending on x .

Let $v(P)$ and $S(P)$ denote the infima and solution set of (1). The notation indicates that their dependence on the underlying probability distribution is of particular interest. For general continuous multivariate probability distributions P such stochastic optimization models are not solvable in general. Even the computation of the involved integrals requires multivariate numerical integration methods. Many approaches for solving optimization models (1) numerically are based on discrete approximations of the probability measure P , i.e., on finding a discrete probability measure P_n in

$$\mathcal{P}_n(\Xi) := \left\{ \sum_{i=1}^n w_i \delta_{\xi^i} : \xi^i \in \Xi, i = 1, \dots, n, (w_1, \dots, w_n) \in \mathcal{S}_n \right\}$$

for some $n \in \mathbb{N}$, which approximates P in a suitable way. Here, \mathcal{S}_n denotes the standard simplex $\mathcal{S}_n = \{w \in \mathbb{R}_+^n : \sum_{i=1}^n w_i = 1\}$ and $\xi^i, i = 1, \dots, n$, the scenarios. Of course, the notion 'suitable' should at least imply that the distance of infima

$$|v(P) - v(P_n)| \quad (4)$$

becomes reasonably small. The latter is a consequence of stability results for stochastic programming problems which explore the behavior of infima and solution sets if the probability distribution is perturbed. To state a version of such results we

introduce the following sets of functions and of probability distributions (both defined on Ξ)

$$\mathcal{F} = \{f_j(x, \cdot) : j = 0, 1, x \in X\}, \quad (5)$$

$$\mathcal{P}_{\mathcal{F}} = \left\{ Q \in \mathcal{P}(\Xi) : -\infty < \int_{\Xi} \inf_{x \in X} f_j(x, \xi) Q(d\xi), \sup_{x \in X} \int_{\Xi} f_j(x, \xi) Q(d\xi) < +\infty, j = 0, 1 \right\}$$

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

$$d_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} \left| \int_{\Xi} f(\xi)(P - Q)(d\xi) \right| \quad (P, Q \in \mathcal{P}_{\mathcal{F}}). \quad (6)$$

The distance $d_{\mathcal{F}}$ is based on minimal information of the underlying optimization model (1). It is nonnegative, symmetric and satisfies the triangle inequality. 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 set $\mathcal{P}_{\mathcal{F}}$ or appropriate subsets allow a simple characterization. For example as subsets of $\mathcal{P}(\Xi)$ satisfying certain moment conditions.

The following result is a consequence of [40, Theorems 5 and 9].

Proposition 1 *We consider (1) for $P \in \mathcal{P}_{\mathcal{F}}$, assume that X is compact and*

- (i) *the function $x \rightarrow \int_{\Xi} f_0(x, \xi)P(d\xi)$ is Lipschitz continuous on X ,*
- (ii) *the set-valued mapping $y \mapsto \{x \in X : \int_{\Xi} f_1(x, \xi)P(d\xi) \leq y\}$ has the Aubin property at $\bar{y} = 0$ for each $\bar{x} \in S(P)$ (see [39, Definition 9.36]).*

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

$$|v(P) - v(Q)| \leq L d_{\mathcal{F}}(P, Q) \quad (7)$$

$$\sup_{x \in S(Q)} d(x, S(P)) \leq \Psi_P(L d_{\mathcal{F}}(P, Q)) \quad (8)$$

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 near the solution set $S(P)$ and $\psi_P(\tau)$ is defined for $\tau \geq 0$ as

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

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 estimates (7) and (8) in Proposition 1 suggest to choose discrete approximations from $\mathcal{P}_n(\Xi)$ for solving (1) such that they solve the *best approximation problem*

$$\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \quad (9)$$

in order to bound (4) as tight as possible. Determining the scenarios of some solution to (9) may be called *optimal scenario generation*. This choice of discrete approximations was already suggested in [40, Section 4.2], but characterized there as a challenging task which is not solvable in most cases in reasonable time.

It is recommended in [35, 40] to eventually enlarge the function class \mathcal{F} such that $d_{\mathcal{F}}$ becomes a metric distance and has further nice properties. Following

this idea, however, leads to coarse estimates of the original minimal information distance and, hence, may lead to unfavorable convergence rates of the sequence

$$\left(\min_{P_n \in \mathcal{P}_n(\Xi)} d_{\mathcal{F}}(P, P_n) \right)_{n \in \mathbb{N}} \quad (10)$$

and to nonconvex nondifferentiable minimization problems (9) for determining the optimal scenarios.

In linear two-stage stochastic programming the class \mathcal{F} contains piecewise linear-quadratic functions defined on Ξ if condition (A1) (see Section 2) is satisfied. If the linear two-stage model has even random recourse, \mathcal{F} may contain more general piecewise polynomial functions (see [42]). Hence, a suitably enlarged class of functions may be chosen as the set

$$\mathcal{F}_r = \left\{ f : \Xi \mapsto \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq \max\{1, \|\xi\|, \|\tilde{\xi}\|\}^{r-1} \|\xi - \tilde{\xi}\|, \forall \xi, \tilde{\xi} \in \Xi \right\} \quad (11)$$

of all locally Lipschitzian functions on Ξ with polynomially growing normalized local Lipschitz constants. Here, $\|\cdot\|$ denotes any norm on \mathbb{R}^s and $r \geq 1$ characterizes the growth of the Lipschitz moduli. The corresponding distance

$$\zeta_r(P, Q) = d_{\mathcal{F}_r}(P, Q) \quad (12)$$

is defined on the set $\mathcal{P}_r(\Xi)$ of all probability measures on Ξ having r th order central moments and is called Fortet-Mourier metric of order r (see [34, Section 5.1]). The Fortet-Mourier metric has a dual representation as transshipment problem (see [34, Section 5.3]). If Ξ is compact, ζ_r admits even a dual representation as transportation problem (see [36, Section 4.3]), namely, it holds

$$\zeta_r(P, Q) = \inf \left\{ \int_{\Xi \times \Xi} c_r(\xi, \tilde{\xi}) \eta(d\xi, d\tilde{\xi}) : \eta \circ \pi_1^{-1} = P, \eta \circ \pi_2^{-1} = Q \right\}, \quad (13)$$

where η is a probability measure on $\Xi \times \Xi$, π_1 and π_2 are the projections from $\Xi \times \Xi$ to the first and second component, respectively, c_r is a metric on \mathbb{R}^s and $c_r(\xi, \tilde{\xi})$ is defined as

$$\inf \left\{ \sum_{i=1}^{n-1} \max\{1, \|\xi_i\|, \|\xi_{i+1}\|\}^{r-1} \|\xi_i - \xi_{i+1}\| : n \in \mathbb{N}, \xi_i \in \Xi, \xi_1 = \xi, \xi_n = \tilde{\xi} \right\}$$

for all $\xi, \tilde{\xi} \in \Xi$. The representation (13) implies, in particular, that the best approximation problem (9) for $\mathcal{F} = \mathcal{F}_r$ is equivalent to

$$\min_{(\xi^1, \dots, \xi^n) \in \Xi^n} \int_{\Xi} \min_{i=1, \dots, n} c_r(\xi, \xi^i) P(d\xi), \quad (14)$$

where ξ^i , $i = 1, \dots, n$, are the scenarios of $P_n \in \mathcal{P}_n(\Xi)$. This follows similarly as in [15, Lemma 4.2]. The probabilities w_i of ξ^i can be computed by $w_i = P(A_i)$, $i = 1, \dots, n$, where the collection $\{A_i : i = 1, \dots, n\}$ is a Voronoi partition of Ξ , i.e., A_i is Borel measurable and a subset of

$$\left\{ \xi \in \Xi : \|\xi - \xi^i\| = \min_{j=1, \dots, n} \|\xi - \xi^j\| \right\} \quad (i = 1, \dots, n).$$

Note that the objective function in (14) is continuous and inf-compact on Ξ^n . Hence, the minimization problem (14) is solvable, but nonconvex for $n \geq 2$ even for $r = 1$. Furthermore, due to a classical result (see [6, Proposition 2.1]), the estimate

$$cn^{-\frac{1}{s}} \leq \zeta_1(P, P_n) \leq \zeta_r(P, P_n)$$

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 . Hence, the convergence rate (10) for $\mathcal{F} = \mathcal{F}_r$ is worse than the Monte Carlo rate $O(n^{-\frac{1}{2}})$ if the dimension s of Ξ is greater than two.

The approach to *optimal scenario reduction* for linear two-stage stochastic programs developed in [10] is based on Monge-Kantorovich functionals and applies to the Fourier-Mourier metric ζ_r (see (12)) due to the representation (13). Starting with a discrete probability measure P based on a large number N of scenarios, it selects a smaller number n of scenarios out of the original set of scenarios together with new probabilities such that the new discrete probability measure represents the best approximation to P with respect to ζ_r . More precisely, let P have the scenarios ξ^i with probabilities p_i , $i = 1, \dots, N$. Using the dual representation (13) of ζ_r the best approximation problem

$$\min_{Q \in \mathcal{P}_n(\text{supp } P)} \zeta_r(P, Q)$$

can be rewritten as the mixed-integer linear program

$$\min \left\{ \sum_{i,j=1}^N x_{ij} p_i c_r(\xi^i, \xi^j) \left| \begin{array}{l} \sum_{j=1}^N x_{ij} = 1, 0 \leq x_{ij} \leq u_j, i = 1, \dots, N \\ \sum_{j=1}^N u_j = n, u_j \in \{0, 1\}, j = 1, \dots, N \end{array} \right. \right\}, \quad (15)$$

which is known as n -median problem (see [4]). In [26] it is shown that problem (15) is NP-hard. If J denotes a subset of $\{1, \dots, N\}$ with cardinality $|J| = n$, the best approximation problem can be decomposed into finding the optimal index set J of remaining scenarios and into determining the optimal discrete probability measure given J . With P_J denoting any probability measure with support consisting of the scenarios ξ^j , $j \in J$, the best approximation problem has a solution P_J^* such that

$$D_J := \zeta_r(P, P_J^*) = \min_{P_J} \zeta_r(P, P_J) = \sum_{i \notin J} p_i \min_{j \in J} c_r(\xi^i, \xi^j) \quad (16)$$

$$\text{with } P_J^* \text{ given by } P_J^* = \sum_{j \in J} \pi_j \xi^j, \text{ where } \pi_j = p_j + \sum_{i \in I_j} p_i \quad (\forall j \in J) \quad (17)$$

and the index sets I_j , $j \in J$, are defined by $I_j := \{i \in \{1, \dots, N\} \setminus J : j = j(i)\}$ with $j(i) \in \arg \min_{j \in J} c_r(\xi^i, \xi^j)$, $\forall i \notin J$. The formula (17) for the optimal weights is called redistribution rule in [10, 18] where the results (16) and (17) are proved, too. The combinatorial optimization problem is of the form

$$\min \{D_J : J \subset \{1, \dots, N\}, |J| = n\}, \quad (18)$$

where D_j is given by (16). Problem (18) may be reformulated as the following binary program

$$\min \left\{ \sum_{j=1}^N u_j \sum_{i \in I_j} p_i c_r(\xi^i, \xi^{j(i)}) : \sum_{j=1}^N u_j = n, u_j \in \{0, 1\}, j = 1, \dots, N \right\}. \quad (19)$$

For a survey of theory and algorithms for n -median problems we refer the interested reader to [4]. Presently local search heuristics [1] and a novel approximation algorithm [29] are the most favorable algorithms with best approximation guarantees. Simple alternatives are the forward and backward greedy heuristics developed and tested in [17, Algorithms 2.2 and 2.4], [18]. The scenario reduction approach described above has been extended to discrepancy distances in [20, 19]. The latter distances are of the form

$$\alpha(P, Q) = \sup_{B \in \mathcal{B}} |P(B) - Q(B)| \quad (P, Q \in \mathcal{P}(\Xi)), \quad (20)$$

where \mathcal{B} is a suitable class of Borel subsets of Ξ . Such distances are relevant for chance constrained stochastic programs if \mathcal{B} contains the relevant sets (for example, the polyhedra $\mathcal{P}(x)$). We recall, however, that employing probability metrics like (12) and (20) means that decisions on reducing scenarios are based on *coarse estimates* of the minimal information distances (6) and, thus, do essentially not depend on the specific stochastic program.

We will show in this paper that the optimal scenario generation problem (9) may have favorable solution properties if the set \mathcal{F} remains as small as possible, i.e., as chosen in (5). In Section 2 we demonstrate this for linear two-stage stochastic programs. First we show that (9) can be formulated as generalized semi-infinite program (Theorem 1) which is convex in some cases (Theorem 2) and enjoys stability (Theorem 3). In Section 3 we revisit the problem of optimal scenario reduction for two-stage models and provide a new formulation based on the minimal information distance (6) as mixed-integer linear semi-infinite program. The latter decomposes into solving binary and linear semi-infinite programs recursively. Section 4 presents a mixed-integer linear semi-infinite program for optimal scenario generation in chance constrained programming. Finally we illustrate the approach to scenario generation for the classical newsvendor problem and finish with conclusions.

2 Optimal scenario generation for two-stage models by generalized semi-infinite programming

We consider a linear two-stage stochastic program (1) with the integrand (2), a probability distribution P on \mathbb{R}^s and with Φ denoting the infimal value (3) of the second-stage program. Furthermore, we impose the following conditions in addition to the general assumptions made in Section 1:

- (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 Proposition 1 for every $P_n \in \mathcal{P}_n(\Xi)$. Hence, an appropriate formulation of the optimal scenario generation problem for (1), (2) consists in solving the best uniform approximation problem

$$\min_{\substack{(\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \sup_{x \in X} \left| \int_{\Xi} \Phi(q(\xi), h(x, \xi)) P(d\xi) - \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x, \xi^i)) \right|. \quad (21)$$

It means that the convex expected recourse function $F_P : X \rightarrow \mathbb{R}$

$$F_P(x) := \int_{\Xi} \Phi(q(\xi), h(x, \xi)) P(d\xi) \quad (22)$$

has to be approximated uniformly on X by the best convex combination of n convex polyhedral functions appearing as integrand in F_P .

Note that the minimal class $\mathcal{F} = \{\Phi(q(\cdot), h(x, \cdot)) : x \in X\}$ of functions from Ξ to \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

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

where the convex polyhedral cones \mathcal{K}_j , $j = 1, \dots, \ell$, represent a decomposition of the domain of Φ , which is itself a convex polyhedral cone in \mathbb{R}^{m+r} . The latter decomposition depends only on the matrix W [51]. In particular, the functions $\Phi(q(\cdot), h(x, \cdot))$ are locally Lipschitz continuous where the Lipschitz constants on the balls $\{\xi \in \Xi : \|\xi\| \leq \rho\}$ grow linearly with ρ and can be chosen uniform with respect to $x \in X$ (see [40, Proposition 22]).

Next we reformulate (21) as generalized semi-infinite program which allows for specific solution approaches and even turns out to be convex in some cases.

Theorem 1 *Assume (A0)–(A2) and that the function h is affine. Then (21) is equivalent to the generalized semi-infinite program*

$$\min_{\substack{t \geq 0 \\ (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} \sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i \rangle \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n) \end{array} \right. \right\}, \quad (23)$$

where the set $\mathcal{M} = \mathcal{M}(\xi^1, \dots, \xi^n)$ is given by

$$\{(x, y, z) \in X \times Y^n \times \mathbb{R}^{rn} : W y_i = h(x, \xi^i), W^\top z_i - q(\xi^i) \in Y^*, i = 1, \dots, n\}. \quad (24)$$

and $F_P : X \rightarrow \mathbb{R}$ is the convex expected recourse function of the two-stage model. The feasible set of (23) is closed.

Proof By the standard way of rewriting best uniform approximation problems one obtains first by introducing the auxiliary variable t that the semi-infinite program

$$\min_{\substack{t \geq 0 \\ (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x, \xi^i)) \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \Phi(q(\xi^i), h(x, \xi^i)) \\ \forall x \in X \end{array} \right. \right\} \quad (25)$$

is equivalent to (21). Next we exploit the duality relation

$$\Phi(q, t) = \inf\{\langle q, y \rangle : Wy = t, y \in Y\} = \sup\{\langle t, z \rangle : W^\top z - q \in Y^*\}$$

of the second-stage program for all pairs $(q, t) \in D \times W(Y)$, where Y^* denotes the polar cone of Y . Then the primal and dual program are both solvable, too. Due to (A1) the semi-infinite program (25) may be reformulated as

$$\min_{\substack{t \geq 0 \\ (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} \sum_{i=1}^n w_i \sup\{\langle h(x, \xi^i), z \rangle : W^\top z - q(\xi^i) \in Y^*\} \leq t + F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \inf\{\langle q(\xi^i), y \rangle : Wy = h(x, \xi^i), y \in Y\} \\ \forall x \in X \end{array} \right. \right\}. \quad (26)$$

Next we introduce $2n$ new variables $y_i \in Y$ with $Wy_i = h(x, \xi^i)$ and $z_i \in \mathbb{R}^r$ with $W^\top z_i - q(\xi^i) \in Y^*$, $i = 1, \dots, n$, and consider the generalized linear semi-infinite program (23). Then any $t \geq 0$ and $(\xi^1, \dots, \xi^n) \in \Xi^n$ solving problem (26) satisfies the constraints of (23). On the other hand, if $t \geq 0$ and $(\xi^1, \dots, \xi^n) \in \Xi^n$ attain the minimum in (23), the two inequalities

$$\sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle \leq t + F_P(x) \quad \text{and} \quad F_P(x) \leq t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i \rangle$$

are satisfied for all $(x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n)$. Hence, the inequalities

$$\begin{aligned} \sum_{i=1}^n w_i \sup\{\langle h(x, \xi^i), z \rangle : W^\top z - q(\xi^i) \in Y^*\} &\leq t + F_P(x) \\ F_P(x) &\leq t + \sum_{i=1}^n w_i \inf\{\langle q(\xi^i), y \rangle : Wy = h(x, \xi^i), y \in Y\} \end{aligned}$$

are satisfied for all $x \in X$. Hence, programs (26) and (23) are equivalent.

To show that the feasible set of (23) is closed, we know from [47, Corollary 3.1.21] (see also [16, Proposition 3.4]) that the lower semicontinuity of \mathcal{M} on Ξ^n is a sufficient condition. Since the graph $\text{gph } \mathcal{M}$ of \mathcal{M} is of the form

$$\text{gph } \mathcal{M} = \{(\xi^1, \dots, \xi^n, x, y, z) \in \Xi^n \times X \times Y^n \times \mathbb{R}^{rn} : Wy_i = h(x, \xi^i), \\ W^\top z_i - q(\xi^i) \in Y^*, i = 1, \dots, n\},$$

it is convex polyhedral. Such set-valued mappings are even Hausdorff Lipschitz continuous on its domain (see, for example, [39, Example 9.35]) and, hence, on Ξ^n due to (A1). The proof is complete. \square

When the weights w_i , $i = 1, \dots, n$, are fixed, the generalized semi-infinite program (23) is linear. But, even then the optimization model is not convex in general. Nevertheless generalized linear semi-infinite programs allow for a number of solution approaches that exploit the particular structure [47–49]. However, the optimization model (23) is even convex if either only right-hand sides or only costs are random.

Theorem 2 *Assume (A0)–(A2), let the function h be affine, the weights w_i , $i = 1, \dots, n$, be fixed and let either h or q be random. Then the feasible set of (23) is closed and convex.*

Proof Let q be nonrandom. Then the feasible set M of (23) is of the form

$$M = \left\{ (t, \xi^1, \dots, \xi^n) \in \mathbb{R}_+ \times \Xi^n \left| \begin{array}{l} \sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle - t \leq F_P(x) \\ F_P(x) \leq t + \sum_{i=1}^n w_i \langle q, y_i \rangle \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n) \end{array} \right. \right\}. \quad (27)$$

Let $\alpha \in [0, 1]$ and $\xi_j = (\xi_j^1, \dots, \xi_j^n) \in \Xi^n$, $t_j \in \mathbb{R}_+$, be such that $(t_j, \xi_j) \in M$, $j = 1, 2$. We have to show that $\alpha(t_1, \xi_1) + (1 - \alpha)(t_2, \xi_2)$ belongs to M , too.

Let $x \in X$ and $z_i \in \{z \in \mathbb{R}^r : W^\top z - q \in Y^*\}$ for $i = 1, \dots, n$ be chosen arbitrarily. Then we have

$$\begin{aligned} & \sum_{i=1}^n w_i \langle h(x, \alpha \xi_1^i + (1 - \alpha) \xi_2^i), z_i \rangle - \alpha t_1 - (1 - \alpha) t_2 \\ &= \alpha \left(\sum_{i=1}^n w_i \langle h(x, \xi_1^i), z_i \rangle - t_1 \right) + (1 - \alpha) \left(\sum_{i=1}^n w_i \langle h(x, \xi_2^i), z_i \rangle - t_2 \right) \\ &\leq \alpha F_P(x) + (1 - \alpha) F_P(x) = F_P(x). \end{aligned}$$

Now, let $y_{ij} \in \{y \in Y : Wy = h(x, \xi_j^i)\}$ for $j = 1, 2$, $i = 1, \dots, n$, be chosen arbitrarily in addition. We obtain $\alpha y_{i1} + (1 - \alpha) y_{i2} \in \{y \in Y : Wy = h(x, \alpha \xi_1^i + (1 - \alpha) \xi_2^i)\}$ and, hence,

$$\begin{aligned} & \alpha t_1 + (1 - \alpha) t_2 + \sum_{i=1}^n w_i \langle q, \alpha y_{i1} + (1 - \alpha) y_{i2} \rangle \\ &= \alpha \left(t_1 + \sum_{i=1}^n w_i \langle q, y_{i1} \rangle \right) + (1 - \alpha) \left(t_2 + \sum_{i=1}^n w_i \langle q, y_{i2} \rangle \right) \\ &\geq \alpha F_P(x) + (1 - \alpha) F_P(x) = F_P(x). \end{aligned}$$

This means $\alpha(t_1, \xi_1) + (1 - \alpha)(t_2, \xi_2) \in M$ and M is convex. If q is random, but h nonrandom, the proof is similar. The closedness of M follows from Theorem 1. \square

The generalized semi-infinite program (23) for determining the optimal scenarios ξ^i , $i = 1, \dots, n$, for given $n \in \mathbb{N}$ is of dimension $ns + 1$ and, thus, large scale in most cases. A difficulty of (23) is that the set $\mathcal{M}(\xi^1, \dots, \xi^n)$ is always unbounded even if (A0) is satisfied. The reason is that the constraint set of each dual pair (y_i, z_i) is unbounded.

We note that $F_P(x)$ can only be calculated approximately even if the probability

measure P is completely known. Hence, it becomes important that the optimization model (23) behaves stable when the function F_P is perturbed. The following result shows that even Lipschitz stability of the optimal values can be expected.

Theorem 3 *Assume (A0)–(A2) and that the optimal value $v(F_P)$ of (23) is positive. Let the function h be affine, let either h or q be random and the weights w_i , $i = 1, \dots, n$, be fixed. Then there exist $\kappa > 0$ and $\delta > 0$ such that*

$$|v(F_P) - v(F)| \leq \kappa \sup_{x \in X} |F_P(x) - F(x)|, \quad (28)$$

for each continuous function F on X such that $\sup_{x \in X} |F_P(x) - F(x)| < \delta$. Here, $v(F)$ denotes the optimal value of (23) with F_P replaced by F .

Proof As in the proof of Theorem 2 we assume without loss of generality that q is nonrandom. We consider the set-valued mapping $(t, \xi^1, \dots, \xi^n) \mapsto \Lambda(t, \xi^1, \dots, \xi^n)$ from $\mathbb{R}_+ \times \Xi^n$ to the Banach space $C(X)$ of real-valued continuous functions on X with the standard norm $\|\cdot\|_\infty$, where

$$\Lambda(t, \xi^1, \dots, \xi^n) = \left\{ f \in C(X) \left| \begin{array}{l} \sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle - t - F_P(x) \leq f(x) \\ f(x) \leq t + \sum_{i=1}^n w_i \langle q, y_i \rangle - F_P(x) \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n) \end{array} \right. \right\}.$$

First, we show that the graph $\text{gph } \Lambda$ of the set-valued mapping Λ is convex. Let $\alpha \in [0, 1]$ and $f_j \in C(X)$, $\xi_j = (\xi_j^1, \dots, \xi_j^n) \in \Xi^n$, $t_j \in \mathbb{R}_+$, be such that $(t_j, \xi_j, f_j) \in \text{gph } \Lambda$, $j = 1, 2$. Then we obtain as in the proof of Theorem 2

$$\begin{aligned} & \sum_{i=1}^n w_i \langle h(x, \alpha \xi_1^i + (1-\alpha)\xi_2^i), z_i \rangle - \alpha t_1 - (1-\alpha)t_2 - F_P(x) \\ &= \alpha \left(\sum_{i=1}^n w_i \langle h(x, \xi_1^i), z_i \rangle - t_1 - F_P(x) \right) + (1-\alpha) \left(\sum_{i=1}^n w_i \langle h(x, \xi_2^i), z_i \rangle - t_2 - F_P(x) \right) \\ &\leq \alpha f_1(x) + (1-\alpha) f_2(x) \quad \text{and} \\ & \alpha t_1 + (1-\alpha)t_2 + \sum_{i=1}^n w_i \langle q, \alpha y_{i1} + (1-\alpha)y_{i2} \rangle - F_P(x) \\ &\geq \alpha f_1(x) + (1-\alpha) f_2(x), \end{aligned}$$

where $x \in X$, $z_i \in \{z \in \mathbb{R}^r : W^\top z - q \in Y^*\}$ and $y_{ij} \in \{y \in Y : Wy = h(x, \xi_j^i)\}$ for $j = 1, 2$, $i = 1, \dots, n$, are chosen arbitrary. This proves that $\text{gph } \Lambda$ is convex. It is also closed as subset of $\mathbb{R}^{ns+1} \times C(X)$. Furthermore, we know that the null function $0 \in C(X)$ belongs to the range of Λ and that $\Lambda^{-1}(0)$ is just the feasible set of (23). Thus, there exists $(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n) \in \mathbb{R}_+ \times \Xi^n$ such that $0 \in \Lambda(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n)$. Due to the assumption $v(F_P) > 0$ we know that $\bar{t} > 0$. Then we choose δ such that $0 < \delta < \bar{t}$ and conclude that the closed ball $\mathbb{B}(0, \delta)$ in $C(X)$ is contained in the range of Λ . The Robinson-Ursescu theorem (see [38, Theorem 2]) then implies that the inverse multifunction Λ^{-1} has the Aubin property at $f = 0$ for any point $(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n) \in \Lambda^{-1}(0)$ with $\bar{t} > 0$. This means that there exist neighborhoods V of 0 and W of $(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n)$, and a constant $\kappa \in \mathbb{R}_+$ such that

$$\Lambda^{-1}(f) \cap W \subseteq \Lambda^{-1}(\tilde{f}) + \kappa \|f - \tilde{f}\|_\infty \mathbb{B} \quad (29)$$

holds for all $f, \tilde{f} \in V$, where \mathbb{B} is the unit ball in \mathbb{R}^{ns+1} . Next we choose $f = F - F_P$ with $F \in C(X)$ and $\tilde{f} = 0$.

Let $\varepsilon > 0$ and $(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n) \in \Lambda^{-1}(f) \cap W$ such that $v(f) \leq \bar{t} \leq v(f) + \varepsilon$. Then there exists an element $(\tilde{t}, \tilde{\xi}^1, \dots, \tilde{\xi}^n) \in \Lambda^{-1}(0)$ such that

$$\|(\bar{t}, \bar{\xi}^1, \dots, \bar{\xi}^n) - (\tilde{t}, \tilde{\xi}^1, \dots, \tilde{\xi}^n)\| \leq \kappa \|f\|_\infty$$

holds for all $f \in V$ due to the Aubin property (29) of Λ^{-1} at 0. We note that $\Lambda^{-1}(F - F_P)$ is the constraint set of (23) with F_P replaced by F , respectively, and obtain that

$$v(F_P) - v(F) \leq \bar{t} - v(F) \leq |\bar{t} - \tilde{t}| - \varepsilon \leq \kappa \|F - F_P\|_\infty - \varepsilon.$$

holds for all $F \in C(X)$ with $F - F_P \in V$. Since the latter estimate is valid for any $\varepsilon > 0$, we obtain $v(F_P) - v(F) \leq \kappa \|F - F_P\|_\infty$ if $F - F_P \in V$. In the same way we can derive the estimate $v(F) - v(F_P) \leq L \|F - F_P\|_\infty$ if $F - F_P \in V$. It remains to select $\delta > 0$ such that the open ball around 0 with radius δ in $C(X)$ is contained in V and to require $\|F - F_P\|_\infty < \delta$. Finally, we note that starting with the Aubin property of Λ^{-1} at $0 \in C(X)$ the proof followed classical arguments of quantitative stability in optimization (see [27, Theorem 1]). \square

For example, F_P can be approximated 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 $\hat{\xi}^j, j = 1, \dots, N$. Inserting this approximation into (23) and exploiting again the duality relation then leads to the following approximate version of (23)

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} \sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle \leq t + \frac{1}{N} \sum_{j=1}^N \langle q(\hat{\xi}^j), \hat{y}_j \rangle \\ \frac{1}{N} \sum_{j=1}^N \langle h(x, \hat{\xi}^j), \hat{z}_j \rangle \leq t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i \rangle \\ \forall (x, \hat{y}, \hat{z}) \in \mathcal{M}(\hat{\xi}^1, \dots, \hat{\xi}^N) \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n) \end{array} \right. \right\}, \quad (30)$$

where the sample $\hat{\xi}^j, j = 1, \dots, N$ is given. The latter problem may also be characterized as *scenario clustering problem*: Given a large scenario set $\hat{\xi}^j, j = 1, \dots, N$, we are looking for a smaller scenario set $\xi^i, i = 1, \dots, n$, where each scenario ξ^j corresponds to a cluster $\hat{\xi}^i, i \in I_j$, of the original scenarios.

The specific structure of (23) and (30) as generalized semi-infinite programs is promising and allows specific solution algorithms (see [14] for a comprehensive monograph on the linear semi-infinite case and [49, 48, 16] for work on generalized semi-infinite programs and solution methods).

Finally, we discuss the possible use of lower and upper bounds of $F_P(x)$ for scenario generation. There is a well-developed theory for deriving lower and upper bounds of expectation functionals of convex-concave integrands. While lower bounds are due to Jensen's classical result (e.g., see [7, Theorem 10.2.6]), upper

bounds are known as Edmundson-Madansky bounds. They were further developed in the context of stochastic programming, for example, in [2, 5, 11–13, 24]. Many upper bounds are derived via generalized moment problems appearing as duals of semi-infinite programs [12, 24] (see also [25, Section 3.2.1]).

Let $l_P(x)$ and $u_P(x)$ denote lower and upper bounds of $F_P(x)$, respectively. Then the following optimization problem (derived from (23)) computes upper bounds of the infima to (21) or (23), respectively:

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} t \left\{ \begin{array}{l} \sum_{i=1}^n w_i \langle h(x, \xi^i), z_i \rangle \leq t + l_P(x), \\ u_P(x) \leq t + \sum_{i=1}^n w_i \langle q(\xi^i), y_i \rangle, \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^n) \end{array} \right\}. \quad (31)$$

If $l_P(x)$ and $u_P(x)$ are exchanged, the optimization problem (31) provides lower bounds of the infima to (23). These observations are of interest for the numerical solution of (23) if it is nonconvex.

3 Optimal scenario reduction for two-stage models

Next we discuss the scenario reduction approach for two-stage models based on the minimal information distance (5) and the best approximation problem (9).

As in Section 1 let ξ^i , $i = 1, \dots, N$, be a large set of scenarios with probabilities p_i , $i = 1, \dots, N$, that define a discrete probability measure P . For prescribed $n \in \mathbb{N}$, $n < N$, we intend to determine an index set $J \subset \{1, \dots, N\}$ of cardinality $|J| = n$ and new weights $\bar{\pi}_j$, $j \in J$, such that the probability measure

$$P_J^* = \sum_{j \in J} \bar{\pi}_j \delta_{\xi^j}$$

solves the optimal scenario reduction problem

$$\min \left\{ \sup_{x \in X} \left| \sum_{j \in J} \pi_j \varphi_j(x) - \sum_{i=1}^N p_i \varphi_i(x) \right| : J \subset \{1, \dots, N\}, |J| = n, \pi \in \mathcal{S}_n \right\}, \quad (32)$$

where the functions $\varphi_i(x) = \Phi(q(\xi^i), h(x, \xi^i))$, $i = 1, \dots, N$, are convex polyhedral on X . Problem (32) represents a mixed-integer semi-infinite program. Compared with (15), (32) is based on Proposition 1 and, hence, on a (much) smaller upper bound for the difference of the optimal values. In addition, the solution of problem (32) depends on the data of the two-stage stochastic program.

Problem (32) decomposes into finding the optimal index set J of remaining scenarios and into determining the optimal probability measure given J . The outer combinatorial optimization problem

$$\min \{D(J, P) : J \subset \{1, \dots, N\}, |J| = n\}, \quad (33)$$

determines the index set J and can be reformulated as binary optimization problem similar to (19). Here $D(J, P)$ denotes the infimum of the inner program

$$\min_{\pi \in \mathcal{S}_n} \sup_{x \in X} \left| \sum_{j \in J} \pi_j \varphi_j(x) - \sum_{i=1}^N p_i \varphi_i(x) \right|. \quad (34)$$

For linear two-stage stochastic programs satisfying (A0)–(A2) the optimization model (34) contains finite functions and is equivalent to the reduced linear semi-infinite program

$$\min_{t \geq 0, \pi \in \mathcal{S}_n} \left\{ t \left| \begin{array}{l} \sum_{j \in J} \pi_j \varphi_j(x) \leq t + \sum_{i=1}^N p_i \varphi_i(x) \\ \sum_{i=1}^N p_i \varphi_i(x) \leq t + \sum_{j \in J} \pi_j \varphi_j(x) \\ \forall x \in X \end{array} \right. \right\} \quad (35)$$

or to

$$\min_{t \geq 0, \pi \in \mathcal{S}_n} \left\{ t \left| \begin{array}{l} \sum_{j \in J} \pi_j \langle h(x, \xi^j), z_j \rangle \leq t + \sum_{i=1}^N p_i \langle q(\xi^i), y_i \rangle \\ \sum_{i=1}^N p_i \langle h(x, \xi^i), z_i \rangle \leq t + \sum_{j \in J} \pi_j \langle q(\xi^j), y_j \rangle \\ \forall (x, y, z) \in \mathcal{M}(\xi^1, \dots, \xi^N) \end{array} \right. \right\}. \quad (36)$$

Problem (35) means: For a given convex combination of many convex polyhedral functions $\varphi_i(\cdot)$ on X one is looking for the best convex combination of a given subset of convex polyhedral functions that approximates the former uniformly.

To rewrite (35) in the standard form of a linear semi-infinite program, we introduce the index set $I = \{1, 2\}$, the real numbers

$$b(1; x) = - \sum_{i=1}^N p_i \varphi_i(x), \quad b(2; x) = -b(1; x), \quad x \in X,$$

and the elements $a(i; x) \in \mathbb{R}^n$, $i \in I$, $x \in X$, with the components

$$a_j(1; x) = \varphi_j(x), \quad a_j(2; x) = -\varphi_j(x), \quad j \in J, x \in X.$$

Then (35) is equivalent to

$$\min_{(\pi, t) \in \mathcal{S}_n \times \mathbb{R}_+} \left\{ t : a(i; x)^\top \pi + b(i; x) - t \leq 0, \forall (i; x) \in I \times X \right\}. \quad (37)$$

Next we make use of an approach to duality for (37) developed in [45]. It works without structural assumptions on $I \times X$. We consider the linear space Λ of all real-valued functions λ defined on $I \times X$ and the convex cone

$$K = \{ \lambda \in \Lambda : \lambda(i; x) \leq 0, \forall (i; x) \in I \times X \}.$$

We associate with Λ a dual space Λ^* of linear functionals, namely, the linear space of real-valued functions λ^* defined on $I \times X$ such that its support $\text{supp } \lambda^*$ is finite, and define the dual pairing by

$$\langle \lambda^*, \lambda \rangle = \sum_{(i; x) \in \text{supp } \lambda^*} \lambda^*(i; x) \lambda(i; x).$$

The polar cone K^* to K is

$$K^* = \{ \lambda^* \in \Lambda^* : \langle \lambda^*, \lambda \rangle \leq 0, \forall \lambda \in \Lambda \} = \{ \lambda^* \in \Lambda^* : \lambda^*(i; x) \geq 0, \forall (i; x) \in I \times X \}.$$

We reformulate the program (37) by setting $f(t, \pi) = t$ and

$$G(t, \pi; i; x) = a(i; x)^\top \pi + b(i; x) - t = -t + G(0, \pi; i; x)$$

and arrive at the (primal) problem

$$\min_{(\pi, t) \in \mathcal{S}_n \times \mathbb{R}_+} \{f(t, \pi) : G(t, \pi; \cdot; \cdot) \in K\}. \quad (38)$$

The corresponding Lagrangian is

$$\begin{aligned} L(t, \pi; \lambda^*) &= f(t, \pi) + \langle \lambda^*, G(t, \pi; \cdot, \cdot) \rangle \quad (\lambda^* \in \Lambda^*) \\ &= t(1 - \langle \lambda^*, 1 \rangle) + \langle \lambda^*, G(0, \pi; \cdot, \cdot) \rangle \end{aligned}$$

and the dual problem is of the form

$$\max_{\lambda^* \in K^*} \inf_{(\pi, t) \in \mathcal{S}_n \times \mathbb{R}_+} L(t, \pi; \lambda^*) \quad (39)$$

or, equivalently,

$$\max_{\substack{\lambda^* \in K^* \\ \langle \lambda^*, 1 \rangle = 1}} \inf_{\pi \in \mathcal{S}_n} \sum_{(i; x) \in \text{supp } \lambda^*} \lambda^*(i; x) (a(i; x)^\top \pi + b(i; x)) \quad (40)$$

We denote by $v(P)$ and $v(D)$ the optimal values of (38) and (39), respectively, and by $S(P)$ the solution set of (38). Since $S(P)$ is nonempty and bounded, we have strong duality, i.e., $v(P) = v(D)$ (see [45, Proposition 3.4]).

Excellent sources of theory and numerical methods for linear semi-infinite programs we refer to the survey papers [21, 37] and the monographs [22, 14].

4 Scenario generation for chance constrained programs

We consider a chance constrained program

$$\min \{g(x) : x \in X, P(\mathcal{P}(x)) \geq p\},$$

where $\mathcal{P}(x) = \{\xi \in \Xi : h(x, \xi) \leq 0\}$ is a polyhedron depending on x , g is a linear objective g , X and Ξ are polyhedral, h a function as described in Section 1 and $p \in (0, 1)$ a given probability level. Then we have $f_0(x, \xi) = g(x)$ and $f_1(x, \xi) = p - \mathbb{1}_{\mathcal{P}(x)}(\xi)$, and the best approximation problem (9) is of the form

$$\min_{t \geq 0, P_n \in \mathcal{P}_n(\Xi)} \left\{ t \left| \begin{array}{l} P(\mathcal{P}(x)) \leq t + P_n(\mathcal{P}(x)) \\ P_n(\mathcal{P}(x)) \leq t + P(\mathcal{P}(x)) \\ \forall x \in X \end{array} \right. \right\} \quad (41)$$

and, thus,

$$P_n(\mathcal{P}(x)) = \sum_{i=1}^n w_i \mathbb{1}_{\mathcal{P}(x)}(\xi^i) = \sum_{i=1}^n w_i \mathbb{1}_{\mathbb{R}_-^r}(h(x, \xi^i)) \quad (x \in X).$$

It is well-known that chance constrained optimization models with discrete probability distributions are nonconvex in general (see, for example, [25, Section 2.2.2]),

but can be reformulated as mixed-integer programs. We follow here the presentation in [25, Section 2.2.2] and choose a constant $M > 0$ such that

$$h(x, \xi^i) - Me \leq 0 \quad \forall x \in X, \quad (42)$$

holds for each $i = 1, \dots, n$, where $e = (1, \dots, 1)^\top \in \mathbb{R}^r$. Such constant M always exists as X is compact. This allows to introduce binary variables $z_i \in \{0, 1\}$ such that $z_i = 0$ if $h(x, \xi^i) \leq 0$ for all $x \in X$ and $z_i = 1$ otherwise, $i = 1, \dots, n$.

Then it is possible to reformulate (41) as mixed-integer semi-infinite program

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \Xi^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n \\ (z_1, \dots, z_n) \in \{0, 1\}^n}} t \left\{ \begin{array}{l} P(\mathcal{P}(x)) \leq t + \sum_{i=1}^n w_i(1 - z_i) \\ \sum_{i=1}^n w_i(1 - z_i) \leq t + P(\mathcal{P}(x)) \\ h(x, \xi^i) - z_i Me \leq 0, \quad i = 1, \dots, n \\ \forall x \in X \end{array} \right\}. \quad (43)$$

If the weights w_i , $i = 1, \dots, n$, are fixed, problem (43) is a mixed-integer linear semi-infinite optimization model.

Since mixed-integer linear programs containing 'big-M' type constraints are often difficult to solve, one is interested in strengthening the formulation of (43) by incorporating valid inequalities. A possible way consists in introducing precedence constraints based on partial orders \preceq on the index set $\{1, \dots, n\}$. Such a partial order \preceq is called strongly consistent for (43) in [43] if for all $x \in X$

$$i \preceq j \wedge h(x, \xi^j) \leq 0 \Rightarrow h(x, \xi^i) \leq 0.$$

It follows as in [43] that the constraints

$$z_i \leq z_j \quad \text{for all } i, j \in \{1, \dots, n\} \text{ such that } i \preceq j$$

are valid inequalities if \preceq is a strongly consistent order for (43).

If the function h is of the special form $h(x, \xi) = \xi - T(\xi)x$ with a linear (s, m) -matrix function $T(\cdot)$, a strongly consistent order is $i \preceq j \Leftrightarrow \xi^i - T(\xi^i)x \leq \xi^j - T(\xi^j)x$, for all $x \in X$, where \leq is the component-wise inequality between elements of \mathbb{R}^s . For the special function $h(x, \xi) = \xi - Tx$ and fixed weights w_i , $i = 1, \dots, n$, problem (43) is a mixed-integer linear semi-infinite program of the form

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \Xi^n \\ (z_1, \dots, z_n) \in \{0, 1\}^n}} t \left\{ \begin{array}{l} P(\mathcal{P}(x)) \leq t + \sum_{i=1}^n w_i(1 - z_i) \\ \sum_{i=1}^n w_i(1 - z_i) \leq t + P(\mathcal{P}(x)) \\ \xi^i - Tx - z_i Me \leq 0, \quad i = 1, \dots, n \\ z_i \leq z_j \text{ if } \xi^i \leq \xi^j, \quad i, j = 1, \dots, n \\ \forall x \in X \end{array} \right\}. \quad (44)$$

The papers [30, 50, 53] are sources for deriving further valid inequalities.

5 Newsvendor with random demand: An illustration

We consider the classical newsvendor problem to illustrate the approach to scenario generation and reduction developed Sections 2 and 3. We recall that a newsvendor must place a daily order for a number of copies x of a newspaper. He has to pay r monetary units 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 newsvendor wishes that the order x maximizes his expected profit or, equivalently, minimizes his expected costs, i.e.,

$$\mathbb{E}[f_0(x, \xi)] = \int_{\mathbb{R}} f_0(x, \xi) dP(\xi) = (r - c)x + c \int_{\mathbb{R}} \max\{0, x - \xi\} P(d\xi) \quad (x \in \mathbb{R}).$$

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

$$\mathbb{E}[f_0(x, \xi)] = rx - cx \sum_{\substack{k \in \mathbb{N} \\ k \geq x}} p_k - \sum_{\substack{k \in \mathbb{N} \\ k < x}} k p_k, \quad (45)$$

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

The corresponding optimal scenario generation problem is of the form

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \mathbb{R}^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} \sum_{i=1}^n w_i (\xi^i - x) z_i \leq t + F_P(x) \\ F_P(x) \leq t + c \sum_{i=1}^n w_i y_{2i} \\ \forall (x, y, z) \in \mathbb{R}_+ \times \mathbb{R}_+^{2n} \times \mathbb{R}^n : \\ y_{2i} - y_{1i} = \xi^i - x, 0 \leq z_i \leq c, i = 1, \dots, n \end{array} \right. \right\}, \quad (46)$$

where F_P is the expected recourse function

$$F_P(x) = c \sum_{k=1}^{\infty} p_k \max\{0, x - k\}. \quad (47)$$

We note that Theorems 2 and 3 apply to (46) if the weights w_i are fixed.

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, (46) is equivalent with

$$\min_{\substack{t \geq 0, (\xi^1, \dots, \xi^n) \in \mathbb{R}^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \left\{ t \left| \begin{array}{l} c \sum_{i=1}^n w_i \max\{0, x - \xi^i\} \leq t + F_P(x) \\ F_P(x) \leq t + c \sum_{i=1}^n w_i \max\{0, x - \xi^i\} \\ \forall x \in \mathbb{R}_+ \end{array} \right. \right\}.$$

By incorporating F_P from (47), (46) is equivalent with the best approximation problem

$$\min_{\substack{(\xi^1, \dots, \xi^n) \in \mathbb{R}^n \\ (w_1, \dots, w_n) \in \mathcal{S}_n}} \sup_{x \in \mathbb{R}_+} \left| \sum_{k=1}^{\infty} p_k \max\{0, x - k\} - \sum_{i=1}^n w_i \max\{0, x - \xi^i\} \right|. \quad (48)$$

We assume that the series representation in (47) is not infinite. Let $N \in \mathbb{N}$ be such that $p_k = 0$ for all $k > N$. Then F_P is piecewise linear convex on \mathbb{R}_+ with possible kinks at any $k \in \mathbb{N}$, $k \leq N$. The slope of F_P at k is $c \sum_{i=1}^k p_i$ and it holds $F_P(x) = c(x - \mathbb{E}[\xi])$ for $x \geq N$ where $\mathbb{E}[\xi]$ is the mean value of ξ , i.e., $\mathbb{E}[\xi] = \sum_{k=1}^N p_k k$. Then the best possible choice of the scenarios is $\xi^i = k_i \in \{1, 2, \dots, N\}$, $i = 1, \dots, n$ and (48) represents a scenario reduction problem.

6 Conclusions

The generation of scenarios is an important issue for solving applied stochastic programming models. Presently Monte Carlo sampling methods are the preferred approach (see [23]), but besides Quasi-Monte Carlo and sparse grid methods also best approximation methods are in use. The latter utilize metric distances of probability measures and suggest to determine discrete measures as best approximations to the underlying probability distribution (see [31, 33]). Existing scenario reduction methods [10, 17] are based on the same theoretical background. However, we pointed out in Section 1 that stability results indicate that such probability metrics only lead to coarse estimates of distances of optimal values and solutions. Decisions on scenario generation and reduction based on such estimates appear somewhat questionable and should at least be further examined. This is supported by slow convergence rates in terms of such probability metrics. But a stability result like Proposition 1 also suggests to make use of the minimal information distance $d_{\mathcal{F}}$ (see (6), (5)) as a basis for best approximation methods. This observation served as the guideline for the present paper. It turned out that at least for linear two-stage models the best approximation problem for scenario generation has favorable properties. It represents a best uniform approximation problem of the expected recourse function by a convex combination of certain polyhedral functions and can be rewritten as generalized semi-infinite optimization model. If either only right-hand sides or only costs are random the optimization model is convex. In any case there exists a well-developed theory and a number of solution algorithms for such models (see [16, 47–49]). Scenario reduction problems for linear two-stage models can be decomposed into solving a combinatorial optimization problem and a linear semi-infinite program, where the first determines the remaining scenarios and the second their new probabilities.

The characterization of scenario generation with respect to the distance $d_{\mathcal{F}}$ as best approximation problem for the expected recourse function provides a link to bounding schemes for the expected recourse (see [25, Section 3.2.1]). It reveals the close relationship of scenario generation, scenario reduction and bounding.

The aim of the present paper consisted in showing that employing minimal information distances for scenario generation and reduction leads to interesting optimization models. Their solution should result in improved decisions for scenario

generation and reduction. In a next step we are planning to confirm this by numerical experiments.

References

1. Arya, V.; Garg, N.; Khandekar, R.; Meyerson, A.; Munagala, K.; Pandit, V.: Local search heuristics for k -median and facility location problems, *SIAM Journal on Computing* 33 (2004), 544–562.
2. Birge, J.R., Wets, R.J-B: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, *Mathematical Programming Study* 27 (1986), 54–102.
3. Dick, J.; Kuo, F. Y.; Sloan, I. H.: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.
4. Dohan, D.; Karp, S.; Matejek, B.: K -median algorithms: Theory in practice, 2015.
5. Dokov, S.P.; Morton, D.P.: Second-order lower bounds on the expectation of a convex function, *Mathematics of Operations Research* 30 (2005), 662–677.
6. Dudley, R. M.: The speed of mean Glivenko-Cantelli convergence, *The Annals of Mathematical Statistics* 40 (1969), 40–50.
7. Dudley, R.M.: *Real Analysis and Probability*, Chapman & Hall, New York 1989.
8. Dupačová, J.: Scenario based stochastic programs: Strategies for deleting scenarios, IIASA Working Paper WP-95-014, 1995.
9. Dupačová, J.; Consigli, G.; Wallace, S. W.: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.
10. Dupačová, J.; Gröwe-Kuska, N.; Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* 95 (2003), 493–511.
11. Edirisinghe, N.C.P.: New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming, *Operations Research* 44 (1996), 909–922.
12. Edirisinghe, N.C.P.; Ziemba, W.T.: Bounding the expectation of a saddle function with application to stochastic programming, *Mathematics of Operations Research* 19 (1994), 314–340.
13. Frauendorfer, K.: Solving SLP recourse problems with arbitrary multivariate distributions – The dependent case, *Mathematics of Operations Research* 13 (1988), 377–394.
14. Goberna, M.A.; López, M.A.: *Linear Semi-infinite Optimization*, Wiley, Chichester 1998.
15. Graf, S.; Luschgy, H.: *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.
16. Guerra Vázquez, F.; Rückmann, J.-J.; Stein, O.; Still, G.: Generalized semi-infinite programming: A tutorial, *Journal of Computational and Applied Mathematics* 217 (2008), 394–419.
17. Heitsch, H.; Römisch, W.: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.
18. Heitsch, H.; Römisch, W.: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.
19. Henrion, R.; Küchler, C.; Römisch, W.: Scenario reduction in stochastic programming with respect to discrepancy distances, *Comp. Optim. Appl.* 43 (2009), 67–93.
20. Henrion, R.; Küchler, C.; Römisch, W.: Discrepancy distances and scenario reduction in two-stage stochastic integer programming, *Journal of Industrial and Management Optimization* 4 (2008), 363–384.
21. Hettich, R.; Kortanek, K. O.: Semi-infinite programming: theory, methods, and applications, *SIAM Review* 35 (1993), 380–429.
22. Hettich, R.; Zencke, P.: *Numerische Methoden der Approximation und Semi-Infiniten Optimierung*, Teubner, Stuttgart, 1982.
23. Homem-de-Mello, T.; Bayraksan, G.: Monte Carlo sampling-based methods for stochastic optimization, *Surveys of Operations Research and Management Science* 19 (2014), 56–85.
24. Kall, P.: Stochastic programming with recourse: Upper bounds and moment problems, in *Advances in Mathematical Optimization* (J. Guddat et al. eds.), Akademie-Verlag, Berlin 1988.
25. Kall, P.; Mayer, J.: *Stochastic Linear Programming*, Springer 2010 (Second Edition).

26. Kariv, O.; Hakimi, S. L.: An algorithmic approach to network location problems, II: The p -medians, *SIAM Journal Applied Mathematics* 37 (1979), 539–560.
27. Klatte, D.: A note on quantitative stability results in nonlinear optimization, in Proceedings of the 19. Jahrestagung Mathematical Optimization (Lommatzsch, K. ed.), Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 90, 1987, 77–86.
28. Leövey, H.; Römisches, W.: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, *Mathematical Programming* 151 (2015), 315–345.
29. Li, S.; Svensson, O.: Approximating k -median via pseudo-approximation, *SIAM Journal on Computing* 45 (2016), 530–547.
30. Luedtke, J.; Ahmed, S.; Nemhauser, G. L.: An integer programming approach for linear programs with probabilistic constraints, *Mathematical Programming* 122 (2010), 247–272.
31. Pagès, G.: A space vector quantization method for numerical integration, *Journal of Computational and Applied Mathematics* 89 (1997), 1–38.
32. Pennanen, T.; Koivu, M.: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.
33. Pflug, G. Ch.; Pichler, A.: 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.
34. Rachev, S. T.: *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.
35. Rachev, S. T.; Römisches, W.: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.
36. Rachev, S. T.; Rüschendorf, L.: *Mass Transportation Problems*, Vol. I, Springer, Berlin 1998.
37. Reemtsen, R.; Görner, S.: Numerical methods for semi-infinite programming: A survey, in: *Semi-Infinite Programming*, R. Reemtsen and J.-J. Rückmann (eds.), Kluwer, Boston, 1998, pp. 195–275.
38. Robinson, S. M.: Regularity and stability for convex multivalued functions, *Mathematics of Operations Research* 1 (1976), 130–143.
39. Rockafellar, R. T.; Wets, R. J-B: *Variational Analysis*, Springer-Verlag, Berlin, 1998.
40. Römisches, W.: Stability of stochastic programming problems, in: *Stochastic Programming* (Ruszczynski, A., Shapiro, A. eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.
41. Römisches, W.: Scenario generation, in: *Wiley Encyclopedia of Operations Research and Management Science* (J.J. Cochran ed.), Wiley, 2010.
42. Römisches, W.; Wets, R. J-B: Stability of ε -approximate solutions to convex stochastic programs, *SIAM J. Optim.* 18 (2007), 961–979.
43. Ruszczyński, A.: Probabilistic programming with discrete distributions and precedence constraints knapsack polyhedra, *Mathematical Programming* 93 (2002), 195–215.
44. Ruszczyński, A.; Shapiro, A. (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.
45. Shapiro, A.: On duality theory of convex semi-infinite programming, *Optimization* 54 (2005), 535–543.
46. A. Shapiro, D. Dentcheva and A. Ruszczyński: *Lectures on Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2014 (Second Edition).
47. Stein, O.: *Bi-level Strategies in Semi-infinite Programming*, Kluwer, Boston, 2003.
48. Stein, O.; Still, G.: On generalized semi-infinite optimization and bilevel optimization, *European Journal of Operational Research* 142 (2002), 444–462.
49. Still, G.: Generalized semi-infinite programming: Numerical aspects, *Optimization* 49 (2001), 223–242.
50. Vielma, J. P.; Ahmed, S.; Nemhauser, G. L.: Mixed-integer linear programming formulations for probabilistic constraints, *Operations Research Letters* 40 (2012), 153–158.
51. Walkup, D.; Wets, R. J-B: Lifting projections of convex polyhedra, *Pacific Journal of Mathematics* 28 (1969), 465–475.
52. Wets, R.J-B: Stochastic programs with fixed recourse: The equivalent deterministic program, *SIAM Rev.* 16 (1974), 309–339.
53. Zhao, M.; Huang, K.; Zeng, B.: A polyhedral study on chance constrained programs with random right-hand side, *Mathematical Programming* (to appear).