Scenario reduction in stochastic programming with respect to discrepancy distances

**René Henrion · Christian Küchler ·
Werner Römisch**

Abstract Discrete approximations to chance constrained and mixed-integer two-stage stochastic programs require moderately sized scenario sets. The relevant distances of (multivariate) probability distributions for deriving quantitative stability results for such stochastic programs are $\mathcal{B}$-discrepancies, where the class $\mathcal{B}$ of Borel sets depends on their structural properties. Hence, the optimal scenario reduction problem for such models is stated with respect to $\mathcal{B}$-discrepancies. In this paper, upper and lower bounds, and some explicit solutions for optimal scenario reduction problems are derived. In addition, we develop heuristic algorithms for determining nearly optimally reduced probability measures, discuss the case of the cell discrepancy (or Kolmogorov metric) in some detail and provide some numerical experience.

Keywords Stochastic programming · Chance constraints · Two-stage ·
Mixed-integer · Scenario reduction · Discrepancy · Kolmogorov metric

1 Introduction

Quantitative stability studies in stochastic programming (see [20] for a recent survey) indicate which probability metrics (i.e., distances of probability distributions) are relevant and, in some sense, even canonical for the stability of specific classes of stochastic programs. In particular, Fortet-Mourier and Wasserstein metrics are relevant for two-stage stochastic programs [19, 21], $\mathcal{B}$-discrepancies are canonical for (static) chance constrained models [9, 10, 22] and also relevant for two-stage mixed-integer models [19, 24]. The class $\mathcal{B}$ of Borel sets is chosen as small as possible

R. Henrion (✉)
Weierstrass Institute for Applied Analysis and Stochastics, 10117 Berlin, Germany
e-mail: henrion@wias-berlin.de

C. Küchler · W. Römisch
Institute of Mathematics, Humboldt-University Berlin, 10099 Berlin, Germany

but large enough to contain all sets appearing in chance constrained models or all closures of continuity regions of integrands in two-stage mixed-integer models.

Fortet-Mourier metrics have recently been used for the (nearly) optimal reduction of discrete probability distributions in two-stage stochastic programs without integrality requirements [4, 6, 8]. Furthermore, a multiperiod recursive extension of the reduction technique is developed in [7] for generating scenario trees as inputs of multistage stochastic programs. Inserting discrete probability distributions into chance constrained and two-stage mixed-integer stochastic programs represents an often used approximation technique for solving these models (cf. [23]). Such discrete approximations lead to mixed-integer programs in both cases (see [17, Chap. 11.9] for chance constrained models), whose dimensions grow rapidly with the number of scenarios involved. Hence, moderately sized scenario sets that represent good approximations of the underlying probability distribution are even of greater importance than for two-stage models without integrality requirements.

The present paper aims at paving some roads for optimal scenario reduction in chance constrained and mixed-integer two-stage stochastic programming models. Let P be a discrete probability measure on $\mathbb{R}^s$ with support $\{\xi^1, \ldots, \xi^N\}$ and $P(\xi^i) = p_i > 0$ for $i = 1, \ldots, N$. We consider the problem of finding another discrete probability measure Q on $\mathbb{R}^s$ which is supported on a subset of $\{\xi^1, \ldots, \xi^N\}$ and which deviates from P as little as possible with respect to some discrepancy. We recall from [11, 12] that, for given Borel probability measures P, Q on $\mathbb{R}^s$ and for a given system $\mathcal{B}$ of Borel subsets of $\mathbb{R}^s$, the $\mathcal{B}$-discrepancy between P and Q is defined as

$$\alpha_{\mathcal{B}}(P, Q) := \sup_{B \in \mathcal{B}} |P(B) - Q(B)|.$$

Important examples are the systems $\mathcal{B}_{cl}$ of all closed subsets, $\mathcal{B}_{conv}$ of all closed, convex subsets, $\mathcal{B}_{polyk}$ of all polyhedra having at most k vertices, $\mathcal{B}_{rect}$ of all closed, s-dimensional rectangles $\times_{i=1}^{s} I_i$ with I_i, $i = 1, \ldots, s$, denoting a closed interval in $\mathbb{R}$, and $\mathcal{B}_{cell}$ of all closed cells (i.e., sets of the form $\xi + \mathbb{R}_-^s$ with $\xi \in \mathbb{R}^s$) of $\mathbb{R}^s$. Evidently, one has that

$$\alpha_{\mathcal{B}_{cell}} \leq \alpha_{\mathcal{B}_{rect}} \leq \alpha_{\mathcal{B}_{polyk}} \leq \alpha_{\mathcal{B}_{conv}} \leq \alpha_{\mathcal{B}_{cl}}, \tag{1}$$

where for the inequality $\alpha_{\mathcal{B}_{rect}} \leq \alpha_{\mathcal{B}_{polyk}}$ one has to require that $k \geq 2^s$ (in order to ensure that $\mathcal{B}_{rect} \subseteq \mathcal{B}_{polyk}$). Any $\mathcal{B}$-discrepancy is a semimetric on the space of all probability measures on $\mathbb{R}^s$, i.e., it is non-negative, symmetric and satisfies the triangle equality. The discrepancy $\alpha_{\mathcal{B}_{cell}}$ (and, thus, all discrepancies in (1)) are metrics as, in addition, $\alpha_{\mathcal{B}_{cell}}(P, Q) = 0$ implies $P = Q$. A sequence (P_n) of probability measures converges to P with respect to $\alpha_{\mathcal{B}}$ with $\mathcal{B} \subseteq \mathcal{B}_{cl}$ iff it converges weakly to P and $P(\partial B) = 0$ holds for each $B \in \mathcal{B}$ (with ∂B denoting the boundary of B) [2]. We refer to the monograph [1] for further background on weak convergence of probability measures.

In the literature (cf. [18]), $\alpha_{\mathcal{B}_{cell}}$ is also called *uniform or Kolmogorov metric* as $\alpha_{\mathcal{B}_{cell}}(P, Q)$ is just the uniform distance of the probability distribution functions of P and Q on $\mathbb{R}^s$. The distance $\alpha_{\mathcal{B}_{conv}}$ is known as *isotrope discrepancy* [14] and $\alpha_{\mathcal{B}_{cl}}$ as *total variation* [18]. Consistently, the distance $\alpha_{\mathcal{B}_{cell}}$ ($\alpha_{\mathcal{B}_{rect}}$, $\alpha_{\mathcal{B}_{polyk}}$) will be called

cell (*rectangular*, *polyhedral*) *discrepancies*. Some of these discrepancies have been extensively used for studying properties of uniformly distributed sequences in the s-dimensional unit cube $U^s = [0, 1]^s$ [13] and, more recently, for developing Quasi-Monte Carlo methods [15]. Converse inequalities to (1), e.g., for the isotrope and rectangular discrepancies of probability measures P and Q on $\mathbb{R}^s$, were also derived [11, 13, 14, 16]. For instance, the estimate

$$\alpha_{\mathcal{B}_{conv}}(P, Q) \leq s \left(\frac{4Ms}{s-1} \right)^{\frac{s-1}{s}} \alpha_{\mathcal{B}_{rect}}(P, Q)^{\frac{1}{s}}$$

holds if P has a density (with respect to the Lebesgue measure on $\mathbb{R}^s$) which is bounded by M [16]. In the context of quantitative stability of stochastic programs it is worth noting that the polyhedral, rectangular and cell discrepancies are of special importance for linear chance constrained and mixed-integer two-stage models [10, 19–21, 24].

Denoting by δ_ξ the Dirac-measure placing mass one at the point ξ, one may write the discrete measure P introduced above as

$$P = \sum_{i=1}^{N} p_i \delta_{\xi^i}, \tag{2}$$

where $\sum_{i=1}^{N} p_i = 1$. Now, the reduction problem formulated above can be restated as the following optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \alpha_{\mathcal{B}}(P, Q) = \alpha_{\mathcal{B}} \left(\sum_{i=1}^{N} p_i \delta_{\xi^i}, \sum_{j=1}^{n} q_j \delta_{\eta^j} \right), \\
\text{subject to} \quad & \{\eta^1, \ldots, \eta^n\} \subset \{\xi^1, \ldots, \xi^N\}, \\
& q_j \geq 0 \quad (j = 1, \ldots, n), \qquad \sum_{j=1}^{n} q_j = 1.
\end{aligned}
\tag{3}
$$

The variables to be optimally adjusted here are $\eta^1, \ldots, \eta^n$ and $q_1, \ldots, q_n$ and altogether they define the desired reduced discrete measure Q via

$$Q = \sum_{j=1}^{n} q_j \delta_{\eta^j}. \tag{4}$$

The optimization problem (3) may be decomposed into two subproblems: a combinatorial optimization problem for determining the scenario set $\eta = \{\eta^1, \ldots, \eta^n\}$ and a (linear) program for fixing $q = (q_1, \ldots, q_n)$. To describe this in more detail, we denote by $\alpha_{\mathcal{B}}(P, (\eta, q))$ the $\mathcal{B}$-discrepancy between P and Q, and by S_n the standard simplex in $\mathbb{R}^n$, i.e.,

$$\alpha_{\mathcal{B}}(P, (\eta, q)) := \alpha_{\mathcal{B}} \left(\sum_{i=1}^{N} p_i \delta_{\xi^i}, \sum_{j=1}^{n} q_j \delta_{\eta^j} \right),$$

$$S_n := \left\{ q \in \mathbb{R}^n | q_j \geq 0, \ j = 1, \ldots, n, \ \sum_{j=1}^{n} q_j = 1 \right\}.$$

Now, the optimization problem (3) is of the form

$$\min_{\eta} \left\{ \min_{q \in S_n} \alpha_{\mathcal{B}}(P, (\eta, q)) | \eta \subset \{\xi^1, \ldots, \xi^N\}, \#\eta = n \right\}, \tag{5}$$

with the inner optimization model

$$\min\{\alpha_{\mathcal{B}}(P, (\eta, q)) | q \in S_n\} \tag{6}$$

for fixed scenario set η. While (5) represents a specific clustering problem, a so-called k-median problem of combinatorial optimization, the problem (6) will turn out as a linear program. Both problems will be further discussed in Sect. 3. In Sect. 2, we derive upper and lower bounds of the optimal value of (3) and discuss some particular cases, which allow for an explicit solution. In Sect. 4, we provide some preliminary numerical experience for optimal scenario reduction with respect to the cell discrepancy (or Kolmogorov distance).

2 Bounds and specific solutions

In this section, we shall derive a specific solution for problem (3) in the case of the closed set discrepancy $\alpha_{\mathcal{B}} = \alpha_{\mathcal{B}_{cl}}$ as well as universal bounds for the optimal value of (3) in case of general discrepancies. By 'universal' we mean a bound that just depends on the probabilities p_i of the original discrete measure P but not on its support. In particular, these bounds do not depend on the geometry of the support or the space dimension s. Hence, in contrast to the exact solution of (3), these bounds are very easy to compute for a quite general class of discrepancies.

2.1 Ordered solution and upper bound

Intuitively, approximating the original discrete measure P by some other measure Q which is supported by a subset of the support of P, requires well to approximate those supporting points of P having large probability. In this section, we assume, without loss of generality, that $p_1 \geq \cdots \geq p_N$. Then, a naive idea for solving (3) would be to put in the definition (4) of Q:

$$\eta^j := \xi^j \quad (j = 1, \ldots, n), \qquad q_j := p_j \quad (j = 1, \ldots, n-1), \qquad q_n := \sum_{i=n}^{N} p_i. \tag{7}$$

This means that Q selects its support as the atoms of P having largest probability and, that the assignment of probabilities is adopted from the original measure except at the last atom, where the new probability is modified to make all q_j sum up to one. Evidently, this simple approximating probability measure Q is feasible in (3). From now on, it shall be called the *ordered solution*. Although it may be a poor

approximation in general, Q has the interesting feature that it realizes a universal (with respect to any discrepancy), easy to calculate upper bound in (3), which is actually sharp in case of the closed set discrepancy. This is shown by the following proposition and the subsequent corollary.

Proposition 2.1 *As before, we assume, without loss of generality, that $p_1 \geq \cdots \geq p_N$. Denote by $\Delta_{\mathcal{B}}$ the optimal value of* (3), *where $\mathcal{B}$ is any system of Borel measurable subsets of $\mathbb{R}^s$. Then, one has that*

1. $\Delta_{\mathcal{B}} \leq \sum_{i=n+1}^{N} p_i$.
2. $\Delta_{\mathcal{B}_{cl}} = \sum_{i=n+1}^{N} p_i$.

Proof Define Q in (4) as the ordered solution according to (7). Let $B \in \mathcal{B}$ be arbitrary and put

$$C := \{\xi^i : i = n, n+1, \ldots, N\}.$$

Since P and Q coincide on $\{\xi^i | i = 1, \ldots, N\} \setminus C$, we have

$$|P(B) - Q(B)| = |P(B \cap C) - Q(B \cap C) + P(B \setminus C) - Q(B \setminus C)|$$
$$= |P(B \cap C) - Q(B \cap C)|$$

and, therefore, we may assume that $B \subset C$. Thus, we can write

$$|P(B) - Q(B)| = \left| \sum_{i=n+1}^{N} p_i \delta_{\xi^i}(B) + (p_n - q_n)\delta_{\xi^n}(B) \right|$$

$$= \begin{cases} \left| \sum_{i=n+1}^{N} p_i \delta_{\xi^i}(B) - \sum_{i=n+1}^{N} p_i \right| & \text{if } \xi^n \in B, \\ \sum_{i=n+1}^{N} p_i \delta_{\xi^i}(B) & \text{if } \xi^n \notin B. \end{cases}$$

Due to

$$0 \leq \sum_{i=n+1}^{N} p_i \delta_{\xi^i}(B) \leq \sum_{i=n+1}^{N} p_i,$$

one arrives at

$$|P(B) - Q(B)| \leq \sum_{i=n+1}^{N} p_i.$$

Since $B \in \mathcal{B}$ was arbitrary, there follows assertion 1:

$$\Delta_{\mathcal{B}} \leq \alpha_{\mathcal{B}}(P, Q) = \sup_{B \in \mathcal{B}} \left| P(B) - Q(B) \right| \leq \sum_{i=n+1}^{N} p_i.$$

Concerning assertion 2, let Q in (4) be any discrete measure which is feasible in problem (3) with respect to the special discrepancy distance $\alpha_{\mathcal{B}} = \alpha_{\mathcal{B}_{cl}}$. Feasibility of Q in (3) implies that $\{\eta^1, \ldots, \eta^n\} \subseteq \{\xi^1, \ldots, \xi^N\}$. Therefore, $\eta^j = \xi^{i_j}$ for certain

selections $i_j \in \{1, \ldots, N\}$ and $j = 1, \ldots, n$. Since $B := \{\xi^1, \ldots, \xi^N\} \setminus \{\eta^1, \ldots, \eta^n\}$ is a closed set, one derives that $B \in \mathcal{B}_{cl}$ and thus:

$$\alpha_{\mathcal{B}_{cl}}(P, Q) \geq |P(B) - Q(B)|$$

$$= \left| \sum_{i \in \{i_1, \ldots, i_n\}} p_i \delta_{\xi^i}(B) + \sum_{i \notin \{i_1, \ldots, i_n\}} p_i \delta_{\xi^i}(B) - \sum_{j=1}^{n} q_j \delta_{\eta^j}(B) \right|$$

$$= \sum_{i \notin \{i_1, \ldots, i_n\}} p_i \geq \sum_{i=n+1}^{N} p_i,$$

where in the last inequality the assumed decreasing order of the p_i was exploited. As Q was supposed to be arbitrary feasible in (3), one gets that

$$\Delta_{\mathcal{B}_{cl}} = \inf\{\alpha_{\mathcal{B}_{cl}}(P, Q) | Q \text{ feasible in (3)}\} \geq \sum_{i=n+1}^{N} p_i.$$

Taking into account the reverse inequality, already proved in assertion 1 for an arbitrary discrepancy, there follows assertion 2.   □

Corollary 2.1 *Every probability measure Q satisfying*

$$Q = \sum_{j=1}^{n} q_j \delta_{\xi^j} \quad and \quad q_j \geq p_j, \ j = 1, \ldots, n \tag{8}$$

is an optimal solution of problem (3) for the closed set discrepancy $\alpha_{\mathcal{B}_{cl}}$. In particular, this is true for the ordered solution defined in (7).

Proof On the one hand, $\sum_{i=n+1}^{N} p_i$ is the optimal value in (3) for the closed set discrepancy $\alpha_{\mathcal{B}_{cl}}$ (see assertion 2 in Proposition 2.1). On the other hand, we proceed as in the first part of the proof of Proposition 2.1, to show that for any discrepancy $\alpha_{\mathcal{B}}$, each probability measure Q with (8) realizes an objective value in (3) which is not larger than $\sum_{i=n+1}^{N} p_i$. To this end, we consider such a Q fulfilling (8) and $B \in \mathcal{B}_{cl}$ to obtain

$$|P(B) - Q(B)| = \left| \sum_{i \in \{1, \ldots, N\} \cap B} p_i - \sum_{j \in \{1, \ldots, n\} \cap B} q_j \right|$$

$$= \left| \sum_{i \in \{1, \ldots, n\} \cap B} (p_i - q_i) + \sum_{i \in \{n+1, \ldots, N\} \cap B} p_i \right|$$

$$\leq \sum_{i=n+1}^{N} p_i.$$

The inequality holds since from (8) it follows that the sums under the norms have different signs and, furthermore, both $|\sum_{i\in\{1,...,n\}\cap B}(p_i - q_i)|$ and $\sum_{i\in\{n+1,...,N\}\cap B} p_i$ are less than $\sum_{i=n+1}^{N} p_i$. $\qquad\qquad\square$

The last corollary shows, that in case of the closed set discrepancy, an explicit solution of problem (3) can be found without any computational effort. The same does not hold true for the weaker discrepancies mentioned in the introduction. Nevertheless, for those other discrepancies too, one may benefit from the upper bound information for the optimal value in (3) provided by the first statement in Proposition 2.1. For instance, from the (ordered) values of the original probabilities p_i, one can directly read off the number of atoms $n < N$ required for the approximating measure Q, in order to make the approximation error $\alpha_{\mathcal{B}}(P, Q)$ not exceed a prescribed tolerance $\varepsilon > 0$. In the special case of $p_i = N^{-1}$ $(i = 1, \ldots, N)$, one derives the condition

$$\frac{n}{N} \geq 1 - \varepsilon.$$

For instance, a tolerance of 10% ($\varepsilon = 0.1$) can be satisfied, if n is at least 90% of N. Of course, such linear relation between tolerance and size of distribution is not very satisfactory. Indeed, the second assertion of Proposition 2.1 tells us, that, in the assumed equi-distributed case, one actually observes this undesirable linear relation for the closed set discrepancy. However, there is some hope, that a better behaviour can be observed for the weaker discrepancies, which are more appropriate for the stability of chance constrained and mixed-integer stochastic programs (cf. Sect. 1). This, however, comes at the price that a simple solution of (3) is no longer available and, actually, cannot even be obtained computationally for relevant problem sizes in an exact sense.

The following example complements Corollary 2.1 by showing that the ordered solution need not be optimal for a discrepancy different from $\alpha_{\mathcal{B}_{cl}}$:

Example 2.1 Define the original measure P on $\mathbb{R}$ by

$$\xi^1 := 1, \qquad \xi^2 := 3, \qquad \xi^3 := 2, \qquad \xi^4 := 4,$$
$$p_1 := p_2 := 0.4, \qquad p_3 := p_4 := 0.1.$$

We are looking for the optimally reduced measure Q in problem (3) which selects $n := 2$ atoms from the original measure. As far as this is done with respect to the closed set discrepancy $\alpha_{\mathcal{B}_{cl}}$, Corollary 2.1 guarantees that the ordered solution Q defined by

$$\eta^1 := 1, \qquad \eta^2 := 3, \qquad q_1 := 0.4, \qquad q_2 := 0.6$$

is optimal and, by Proposition 2.1, realizes the minimal discrepancy $\Delta_{\mathcal{B}_{cl}} = p_3 + p_4 = 0.2$. For the convex set discrepancy $\alpha_{\mathcal{B}_{conv}}$ (see introduction), this ordered solution realizes the same value $\alpha_{\mathcal{B}_{conv}}(P, Q) = 0.2$. However, considering the reduced measure Q^* defined by

$$\eta^1 := 1, \qquad \eta^2 := 3, \qquad q_1 := 0.5, \qquad q_2 := 0.5,$$

it follows that $\alpha_{\mathcal{B}_{conv}}(P, Q^*) = 0.1$. Consequently, the ordered solution is not optimal in (3) with respect to $\alpha_{\mathcal{B}_{conv}}$. At the same time, this is an example for a strict inequality in statement 1 of Proposition 2.1.

2.2 Lower bound

In this section, we want to find a universal lower bound for the optimal value of problem (3). For this purpose, we will access on the following property.

Definition 2.1 We call a system $\mathcal{B}$ of Borel subsets of $\mathbb{R}^s$ *isolating* if for any finite subset $\{x^1, \ldots, x^p\} \subseteq \mathbb{R}^s$ there exist sets $B^i \in \mathcal{B}$ for $i = 1, \ldots, p$ with

$$B^i \cap \{x^1, \ldots, x^p\} = \{x^i\} \quad (i = 1, \ldots, p).$$

Clearly, the systems $\mathcal{B}_{rect}$, $\mathcal{B}_{conv}$, $\mathcal{B}_{cl}$ and $\mathcal{B}_{polyk}$ (for $k \geq 2^s$) mentioned in the introduction are isolating, whereas $\mathcal{B}_{cell}$, for instance, is not.

Theorem 2.1 *Let $\mathcal{B}$ be an isolating system of Borel subsets of $\mathbb{R}^s$. In (3), let $n < N$. Then, assuming as before that the p_i are decreasingly ordered, the optimal value $\Delta_{\mathcal{B}}$ in problem (3) has the lower bound*

$$\Delta_{\mathcal{B}} \geq \max\left\{ p_{n+1}, n^{-1} \sum_{i=n+1}^{N} p_i \right\}.$$

Proof Each measure Q defined by (4) which is feasible in problem (3) induces an injective selection mapping $\sigma : \{1, \ldots, n\} \to \{1, \ldots, N\}$ with

$$\eta^i = \xi^{\sigma(i)} \quad (i = 1, \ldots, n).$$

Applying Definition 2.1 to the support $\{\xi^1, \ldots, \xi^N\}$ of the original measure P, we derive the existence of sets $B^i \in \mathcal{B}$ for $i = 1, \ldots, N$ such that

$$B^i \cap \{\xi^1, \ldots, \xi^N\} = \{\xi^i\} \quad (i = 1, \ldots, N).$$

Then,

$$|P(B^{\sigma(i)}) - Q(B^{\sigma(i)})| = |P(\{\xi^{\sigma(i)}\}) - Q(\{\eta^i\})| = |p_{\sigma(i)} - q_i| \quad (i = 1, \ldots, n),$$

$$|P(B^i) - Q(B^i)| = p_i \quad (i \in C_\sigma),$$

where $C_\sigma := \{1, \ldots, N\} \setminus \{\sigma(1), \ldots, \sigma(n)\}$. It follows for the discrepancy that

$$\alpha_{\mathcal{B}}(P, Q) \geq \max_{i=1,\ldots,N} |P(B^i) - Q(B^i)| = \max\left\{ \max_{i \in C_\sigma} p_i, \max_{i=1,\ldots,n} |p_{\sigma(i)} - q_i| \right\}.$$

Note, that the variation of Q among the feasible measures in (3) amounts to variation of the selection mapping σ and to variation of coefficients $q_i \geq 0$ subject to the constraints $\sum_{i=1}^{n} q_i = 1$. This allows to write

$$\Delta_{\mathcal{B}} = \inf\{\alpha_{\mathcal{B}}(P, Q) | Q \text{ feasible in (3)}\}$$

$$\geq \inf\{\varphi(\sigma) | \sigma : \{1, \dots, n\} \to \{1, \dots, N\} \text{ injective}\}, \tag{9}$$

where

$$\varphi(\sigma) := \max\left\{\max_{i \in C_\sigma} p_i, \psi(\sigma)\right\}$$

$$\psi(\sigma) := \inf\left\{\max_{i=1,\dots,n} |p_{\sigma(i)} - q_i| \,\middle|\, q_i \geq 0 \ (i = 1, \dots, n), \ \sum_{i=1}^{n} q_i = 1\right\}.$$

Next, we want to develop the expression for $\psi(\sigma)$. Since $p_i > 0$ for $i = 1, \dots, N$ and $n < N$, by assumption, it follows that

$$\gamma := \sum_{i=1}^{n} p_{\sigma(i)} < 1.$$

Note that the infimum in the definition of $\psi(\sigma)$ is always realized as a minimum. We claim that $\hat{q} \in \mathbb{R}^n$ defined by

$$\hat{q}_i = p_{\sigma(i)} + n^{-1}(1 - \gamma) \quad (i = 1, \dots, n) \tag{10}$$

provides this minimum. We have $\hat{q}_i \geq 0$ for $i = 1, \dots, n$ due to $\gamma < 1$ and $\sum_{i=1}^{n} \hat{q}_i = 1$, hence $\hat{q}$ is feasible in the definition of $\psi(\sigma)$. Now, let $q' \in \mathbb{R}^n$ be any other feasible n-tuple. Then, by

$$\sum_{i=1}^{n} q'_i = \sum_{i=1}^{n} \hat{q}_i = 1,$$

it is excluded that $q'_i < \hat{q}_i$ holds true for all $i = 1, \dots, n$. Consequently, there exists some $k \in \{1, \dots, n\}$ with $q'_k \geq \hat{q}_k$. From the relation $\hat{q}_k \geq p_{\sigma(k)}$ (see (10)), one derives that $|p_{\sigma(k)} - q'_k| \geq |p_{\sigma(k)} - \hat{q}_k|$. Thus,

$$\max_{i=1,\dots,n} |p_{\sigma(i)} - q'_i| \geq |p_{\sigma(k)} - \hat{q}_k| = n^{-1}(1 - \gamma) = \max_{i=1,\dots,n} |p_{\sigma(i)} - \hat{q}_i|.$$

This shows that indeed $\hat{q}$ realizes the infimum in the definition of $\psi(\sigma)$ and so, by (10) and by definition of C_σ, one gets that

$$\psi(\sigma) = n^{-1}(1 - \gamma) = n^{-1}\left(1 - \sum_{i=1}^{n} p_{\sigma(i)}\right) = n^{-1} \sum_{i \in C_\sigma} p_i.$$

Now, we continue (9) as

$$\Delta_{\mathcal{B}} \geq \inf\left\{\max\left\{\max_{i \in C_\sigma} p_i, n^{-1} \sum_{i \in C_\sigma} p_i\right\} \,\middle|\, \sigma : \{1, \dots, n\} \to \{1, \dots, N\} \text{ injective}\right\}.$$

Identifying the set of all selections as given in this relation with the system of all subsets of $\{1, \ldots, N\}$ having cardinality n, one obtains the reformulation

$$\Delta_{\mathcal{B}} \geq \inf\left\{\max\left\{\max_{i \in A} p_i, n^{-1}\sum_{i \in A} p_i\right\} \,\middle|\, A \subseteq \{1, \ldots, N\}, \#A = N - n\right\}.$$

As the p_i are decreasingly ordered, both expressions

$$\max_{i \in A} p_i \quad \text{and} \quad n^{-1}\sum_{i \in A} p_i$$

are simultaneously minimized by the set $A^* := \{n + 1, \ldots, N\}$. Therefore,

$$\Delta_{\mathcal{B}} \geq \max\left\{\max_{i \in \{n+1,\ldots,N\}} p_i, n^{-1}\sum_{i=n+1}^{N} p_i\right\}.$$

Owing to $\max\{p_i | i \in \{n + 1, \ldots, N\}\} = p_{n+1}$, the assertion of the theorem is proved.   □

Remark 2.1 The lower bound from Theorem 2.1 can be interpreted as follows. Consider an arbitrary reduced measure Q. Since $\mathcal{B}$ is isolating, the $\mathcal{B}$-discrepancy between P and Q is not smaller than the maximal difference of P and Q on a singleton. Over all common mass points of P and Q, this maximum is at least $n^{-1}\sum_{i=n+1}^{N} p_i$, over all points without Q-mass it is not less than p_{n+1}.

Corollary 2.2 *Under the assumptions of Theorem 2.1, the following holds true:*

1. *If $n \geq \frac{N}{2}$, then the lower bound in Theorem 2.1 reduces to p_{n+1}.*
2. *If $n = 1$, then $\Delta_{\mathcal{B}} = 1 - p_1$ and an optimal solution of (3) is given by the measure Q placing unit mass on the atom realizing maximum probability with respect to the original measure P.*

   If $n = N - 1$, then $\Delta_{\mathcal{B}} = p_N$, and any measure Q_j of the form

   $$Q_j = \sum_{i=1, i \neq j}^{N-1} p_i \delta_{\xi^i} + (p_j + p_N)\delta_{\xi^j}, \quad j \in \{1, \ldots, N - 1\},$$

   is an optimal solution of (3).

Proof The decreasing order of the p_i implies assertion 1 (by Theorem 2.1) as well as the estimate

$$np_{n+1} \geq (N - n)p_{n+1} \geq \sum_{i=n+1}^{N} p_i,$$

which proves the first statement.

In both cases $n = 1$ and $n = N - 1$, Theorem 2.1 provides that $\Delta_{\mathcal{B}} \geq \sum_{i=n+1}^{N} p_i$. Now, the upper bound in Proposition 2.1 turns this inequality into an equality:

$$\Delta_{\mathcal{B}} = \sum_{i=n+1}^{N} p_i = \begin{cases} 1 - p_1 & \text{if } n = 1, \\ p_N & \text{if } n = N - 1. \end{cases}$$

From the proof of statement 1 in Proposition 2.1 we know that the ordered solution always realizes a discrepancy not larger than $\sum_{i=n+1}^{N} p_i$, where this last value was just recognized to be optimal for $n = 1$ and $n = N - 1$. Consequently, the ordered solution must be optimal in these cases. For $n = 1$, the ordered solution places unit mass on the atom with highest probability in the original measure P. For $n = N - 1$, the ordered solution corresponds to the measure Q_{N-1}. Moreover, by construction any measure Q_j, $j \in \{1, \ldots, N - 1\}$ realizes a discrepancy of p_N and is therefore optimal, too. $\qquad\square$

Unfortunately, the results in Corollary 2.2 are lost for the cell discrepancy $\alpha_{\mathcal{B}_{cell}}$ as the next example shows.

Example 2.2 Consider the probability measure $P = \sum_{i=1}^{3} p_i \delta_{\xi^i}$ on $\mathbb{R}^2$ with ordered probabilities $p_1 \geq p_2 \geq p_3$ and $\xi^1 = 1, \xi^2 = 0, \xi^3 = 2$. For $n = 1$ we obtain $\Delta_{\mathcal{B}_{cell}} = \alpha_{\mathcal{B}_{cell}}(P, \delta_{\xi^1}) = p_2$, which shows that the assertion of Corollary 2.2 does not hold in this case.

3 Solution techniques

As mentioned in Sect. 1, problem (3) can be tackled by a bilevel approach: in an outer iteration, the support selection is carried out by solving the combinatorial optimization problem (5), whereas in an inner iteration optimal probabilities q_j are determined conditional to the fixed support by solving (6). Since k-median problems are known to be *NP*-hard [5], we resort to applying heuristic approaches for solving problem (5). For two-stage models certain forward selection and backward reduction techniques are developed in [6]. In the context of the present paper, their analogues represent recursive extensions of the cases $n = 1$ and $n = N - 1$ in Corollary 2.2. The forward and backward algorithms determine index subsets $J^{[n]}$ and $J^{[N-n]}$, respectively, of $\{1, \ldots, N\}$. Both index sets are of cardinality n and the corresponding scenario sets form the support of the reduced probability measure Q.

Algorithm 3.1 (Forward selection)

> **Step [0]:** $J^{[0]} := \emptyset$.
> **Step [i]:** $l_i \in \operatorname{argmin}_{l \notin J^{[i-1]}} \inf_{q \in S_i} \alpha_{\mathcal{B}}(P, (\{\xi^{l_1}, \ldots, \xi^{l_{i-1}}, \xi^l\}, q))$,
> $\qquad\quad J^{[i]} := J^{[i-1]} \cup \{l_i\}$.
> **Step [n + 1]:** Minimize $\alpha_{\mathcal{B}}(P, (\{\xi^{l_1}, \ldots, \xi^{l_n}\}, q))$ subject to $q \in S_n$.

Algorithm 3.2 (Backward reduction)

> **Step [0]:** $J^{[0]} := \{1, \ldots, N\}$.
> **Step [i]:** $u_i \in \operatorname{argmin}_{u \in J^{[i-1]}} \inf_{q \in S_{N-i}} \alpha_{\mathcal{B}}(P, (\{\xi^j | j \in J^{[i-1]} \setminus \{u\}\}, q))$,
> $\quad\quad\quad\quad J^{[i]} := J^{[i-1]} \setminus \{u_i\}$.
> **Step [N − n + 1]:** Minimize $\alpha_{\mathcal{B}}(P, (\{\xi^j | j \in J^{[N-n]}\}, q))$ subject to $q \in S_n$.

Note that, at each Step [i], $i = 1, \ldots, n$, one has to solve $N - i$ linear programs, which are of dimension i and $N - i$ in Algorithms 3.1 and 3.2, respectively. Hence, forward selection seems to be preferable in most relevant cases. Although both algorithms do not lead to optimality in (3) in general, the performance evaluation of their implemented analogues for transportation distances in [6, 8] is encouraging.

3.1 Formulation as a linear optimization problem

In this section, we consider the inner iteration problem of optimizing the probability distribution conditional to a fixed support. Without loss of generality, we may assume that $\{\eta^1, \ldots, \eta^n\} = \{\xi^1, \ldots, \xi^n\}$. Of course, we may no longer maintain then the assumption of ordered probabilities p_i from the previous section without restricting the generality. Anyway, ordered probabilities p_i are no longer relevant in the sequel. Then, problem (6) is of the form:

$$
\text{minimize} \quad \alpha_{\mathcal{B}}(P, (\{\xi^1, \ldots, \xi^n\}, q)) = \alpha_{\mathcal{B}}\left(\sum_{i=1}^{N} p_i \delta_{\xi^i}, \sum_{j=1}^{n} q_j \delta_{\xi^j}\right), \quad (11)
$$

$$
\text{subject to} \quad q \in S_n.
$$

In the following, we are going to reformulate (11) as a linear optimization problem. To this end, we define for $B \in \mathcal{B}$ a 'critical index set' $I(B) \subseteq \{1, \ldots, N\}$ by

$$
I(B) := \{i \in \{1, \ldots, N\} \mid \xi^i \in B\}.
$$

Then,

$$
|P(B) - Q(B)| = \left| \sum_{i \in I(B)} p_i - \sum_{j \in I(B) \cap \{1, \ldots, n\}} q_j \right|. \quad (12)
$$

Obviously, this value does not depend on the concrete structure of the set B but is uniquely determined by the index set $I(B)$. That is why, for calculating the discrepancy $\alpha_{\mathcal{B}}(P, Q)$, it suffices to know all (finitely many) critical index sets which may occur when B varies in $\mathcal{B}$. We define the system of critical index sets as

$$
\mathcal{I}_{\mathcal{B}} := \{I \subseteq \{1, \ldots, N\} \mid \exists B \in \mathcal{B} : I = I(B)\}.
$$

For the closed set discrepancy, for instance, one has $\mathcal{I}_{\mathcal{B}_{cl}} = 2^{\{1, \ldots, N\}}$, because for each arbitrary subset $I \subseteq \{1, \ldots, N\}$ the set $B := \bigcup_{i \in I} \{\xi^i\}$ is closed with and $I(B) = I$. For the other systems $\mathcal{B}$ considered in the introduction, all one usually gets, is the strict inclusion $\mathcal{I}_{\mathcal{B}} \subset 2^{\{1, \ldots, N\}}$.

As soon as for some concrete $\mathcal{B}$ the system $\mathcal{I}_\mathcal{B}$ of critical index sets is known, the discrepancy between P and Q may be calculated according to (12) by using the formula

$$\alpha_\mathcal{B}(P, Q) = \max_{I \in \mathcal{I}_\mathcal{B}} \left| \sum_{i \in I} p_i - \sum_{j \in I \cap \{1,...,n\}} q_j \right|.$$

We recall the well-kown fact that minimizing a function $|f(x)|$ in terms of the variable x is equivalent to minimizing the function t subject to the constraints $f(x) \leq t$ and $-f(x) \leq t$ in terms of the variables (x, t). This allows to solve (11) by means of the following linear optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & t, \\
\text{subject to} \quad & q \in S_n, \\
& \left. \begin{array}{l} -\sum_{j \in I \cap \{1,...,n\}} q_j \leq t - \sum_{i \in I} p_i \\ \sum_{j \in I \cap \{1,...,n\}} q_j \leq t + \sum_{i \in I} p_i \end{array} \right\} \ I \in \mathcal{I}_\mathcal{B}.
\end{aligned}
\tag{13}
$$

The variables to be optimized here, are t and the q_j. If (q^*, t^*) is an optimal solution of (13), then q^* is an optimal solution of the original problem (11), whereas t^* indicates the optimal value attained by q^* in (13), i.e., t^* provides the minimal discrepancy $\alpha_\mathcal{B}(P, Q)$ between the original measure P and any measure Q whose support coincides with the first n points of the support of P.

Unfortunately, the size of (13) is too large to be useful, in general. Indeed, since $\mathcal{I}_{\mathcal{B}_{cl}} = 2^{\{1,...,N\}}$, as observed above, the number of constraints in (13) amounts to $2^{N+1} + n + 1$. On the other hand, one recognizes from (13), that many inequalities are just copies of themselves as far as the involved coefficients q_j are concerned, because many different index sets $I \in \mathcal{I}_\mathcal{B}$ may lead to the same intersection $I \cap \{1, \ldots, n\}$. The only term which varies then for those sets I, is the right-hand side of the inequalities in (13). Consequently, one may pass to the minimum of these right-hand sides corresponding to one and the same intersection $I \cap \{1, \ldots, n\}$ which will drastically reduce the number of inequalities. In order to do so formally correct, we introduce a reduced system of critical index sets as

$$\mathcal{I}_\mathcal{B}^* := \{I \cap \{1, \ldots, n\} | I \in \mathcal{I}_\mathcal{B}\}.$$

Each member $J \in \mathcal{I}_\mathcal{B}^*$ of the reduced system generates a set $\varphi(J)$ of members in the original system $\mathcal{I}_\mathcal{B}$ all of which share the same intersection with $\{1, \ldots, n\}$:

$$\varphi(J) := \{I \in \mathcal{I}_\mathcal{B} | J = I \cap \{1, \ldots, n\}\} \quad (J \in \mathcal{I}_\mathcal{B}^*). \tag{14}$$

Now, introducing the quantities

$$\gamma^J := \max_{I \in \varphi(J)} \sum_{i \in I} p_i \quad \text{and} \quad \gamma_J := \min_{I \in \varphi(J)} \sum_{i \in I} p_i \quad (J \in \mathcal{I}_\mathcal{B}^*), \tag{15}$$

(13) may be rewritten as

$$\begin{aligned}
\text{minimize} \quad & t, \\
\text{subject to} \quad & q \in S_n, \\
& \left.\begin{array}{r}
-\sum_{j \in J} q_j \leq t - \gamma^J \\
\sum_{j \in J} q_j \leq t + \gamma_J
\end{array}\right\} J \in \mathcal{I}_{\mathcal{B}}^*.
\end{aligned} \tag{16}$$

This corresponds indeed to passing to the minimum on the right-hand sides of the inequalities in (13). Since $\mathcal{I}_{\mathcal{B}}^*$ is a subset of $\{1, \ldots, n\}$, the number of inequalities in (16) is not larger than $2^{n+1} + n + 1$. Having in mind that often $n \ll N$, this results in a drastic reduction of size in the linear optimization problem (13). Moreover, at least for small dimensions s the size of $\mathcal{I}_{\mathcal{B}}^*$ will be significantly smaller than 2^n, see the Tables 1 and 2 in Sect. 4.

A condition on $q \in S_n$ ensuring optimality in (16) can be obtained by the following considerations. The linear constraints for each $J \in \mathcal{I}_{\mathcal{B}}^*$ imply that every feasible t of problem (16) satisfies $t \geq \frac{1}{2}(\gamma^J - \gamma_J)$ and, thus, one obtains the lower bound

$$\frac{1}{2} \max_{J \in \mathcal{I}_{\mathcal{B}}^*} (\gamma^J - \gamma_J) \leq \inf_{q \in S_n} \alpha_{\mathcal{B}}(P, (\{\xi^1, \ldots, \xi^n\}, q)).$$

Equality holds if and only if there exists a $q^* \in S_n$ satisfying the inequalities in (16) with t being equal to this threshold, i.e.

$$\gamma^J - \frac{1}{2} \max_{J \in \mathcal{I}_{\mathcal{B}}^*} (\gamma^J - \gamma_J) \leq \sum_{j \in J} q_j^* \leq \frac{1}{2} \max_{J \in \mathcal{I}_{\mathcal{B}}^*} (\gamma^J - \gamma_J) + \gamma_J \quad (J \in \mathcal{I}_{\mathcal{B}}^*). \tag{17}$$

Consequently, such a q^* would be optimal for (16). In particular, evaluation of (17) in $J_* \in \arg\max_{J \in \mathcal{I}_{\mathcal{B}}^*} (\gamma^J - \gamma_J)$ implies

$$\sum_{j \in J_*} q_j^* = \frac{1}{2}(\gamma^{J_*} + \gamma_{J_*}).$$

3.2 The special case of the cell discrepancy $\alpha_{\mathcal{B}_{cell}}$

The main challenge in the solution of (16) is not the solution of the linear program itself but the computational determination of the reduced critical index set $\mathcal{I}_{\mathcal{B}}^*$ and of the coefficients γ^J and γ_J introduced in (15). As these strongly depend on the geometric structure of the chosen system $\mathcal{B}$ of Borel subsets, there is no general procedure available for their determination. In this section, an algorithmic approach for dealing with the special case of the cell discrepancy $\alpha_{\mathcal{B}_{cell}}$ shall be presented. The same methodology can be carried over to the rectangle discrepancy $\alpha_{\mathcal{B}_{rect}}$ though with higher computational effort. The more general discrepancies $\alpha_{\mathcal{B}_{polyk}}$ and $\alpha_{\mathcal{B}_{conv}}$ of polyhedra and closed convex sets, would require more sophisticated approaches which are outside the scope of this paper.

Recall that the system of cells is defined by

$$\mathcal{B}_{cell} = \{z + \mathbb{R}^s_- | z \in \mathbb{R}^s\},$$

where $\mathbb{R}^s_- = \{x \in \mathbb{R}^s | x_j \leq 0 \ (j = 1, \ldots, s)\}$ is the negative orthant of $\mathbb{R}^s$. For the purpose of abbreviation, we put $[z] := z + \mathbb{R}^s_-$ for $z \in \mathbb{R}^s$.

Since the support $\{\xi^1, \ldots, \xi^N\}$ of the measure P is finite, it is contained in an open rectangle (c, d) for some $c, d \in \mathbb{R}^s$. We introduce an artificial point set $\{r^1, \ldots, r^s\}$ in order to control the boundary of the support, where r^j is defined by

$$r^j_j := d_j \quad \text{and} \quad r^j_k := c_k \quad \text{if } k \neq j. \tag{18}$$

Recall that, by assumption, the support of any feasible reduced measure Q in (11) is given by the set $\{\xi^1, \ldots, \xi^n\}$ of first n atoms of the support of P. We will show that it is sufficient to consider those cells, which are bounded in every direction by an element of

$$R := \{\xi^1, \ldots, \xi^n\} \cup \{r^1, \ldots, r^s\}$$

in the following sense.

Definition 3.1 A cell $[z]$ is called *supporting*, if there exists a subset $\{x^1, \ldots, x^s\} \subseteq R$ such that $\{x^1, \ldots, x^s\} \subseteq [z]$ and $x^j_j = z_j$ for $j = 1, \ldots, s$.

Indeed, we can restrict ourselves on supporting cells as shown by the following proposition, which relies on the fact that every cell can be enlarged until it is a supporting one. As the proof of this proposition is rather technical, it is given in the Appendix.

Proposition 3.1 *For any* $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$, *there exists a supporting cell* $[z]$ *such that* $\gamma^J = P(\text{int}[z])$ *and*

$$\bigcup_{j \in J} \{\xi^j\} = \{\xi^1, \ldots, \xi^n\} \cap \text{int}[z]. \tag{19}$$

Before drawing some essential conclusions from Proposition 3.1, we mention the obvious fact that for any $z \in \mathbb{R}^s$ and any finite subset $F \subseteq \mathbb{R}^s$, there exists some $\varepsilon > 0$, such that

$$[z^\varepsilon] \cap F = \text{int}[z] \cap F, \tag{20}$$

where z^ε is defined by $z^\varepsilon_j := z_j - \varepsilon$ for $j = 1, \ldots, s$.

Corollary 3.1 *Define*

$$\mathcal{Z} := \{z \in \mathbb{R}^s | [z] \text{ is a supporting cell}\}.$$

Then,

$$\mathcal{I}^*_{\mathcal{B}_{cell}} = \{J \subseteq \{1, \ldots, n\} | \exists z \in \mathcal{Z} : (19) \text{ holds true}\},$$

$$\gamma^J = \max\{P(\text{int}[z]) | z \in \mathcal{Z}, (19) \text{ holds true}\} \quad \forall J \in \mathcal{I}^*_{\mathcal{B}_{cell}}.$$

Proof The inclusion '$\subseteq$' in the first identity follows directly from the second statement of Proposition 3.1. Similarly, the inequality '$\leq$' in the second identity follows directly from the first statement of Proposition 3.1. For the reverse direction of the first identity, let $z \in \mathcal{Z}$ be given such that (19) holds true for some $J \subseteq \{1, \ldots, n\}$. Now, we apply (20) to $F := \{\xi^1, \ldots, \xi^n\}$ to derive the existence of some z^ε such that

$$\{\xi^1, \ldots, \xi^n\} \cap [z^\varepsilon] = \{\xi^1, \ldots, \xi^n\} \cap \text{int}[z] = \bigcup_{j \in J} \{\xi^j\}. \tag{21}$$

Since $[z^\varepsilon] \in \mathcal{B}_{cell}$, we observe upon recalling the definition of $I(B)$ for $B \in \mathcal{B}_{cell}$ in the beginning of Sect. 3.1, that

$$I([z^\varepsilon]) = \{i \in \{1, \ldots, N\} | \xi^i \in [z^\varepsilon]\} = J \cup \{i \in \{n+1, \ldots, N\} | \xi^i \in [z^\varepsilon]\}.$$

Therefore,

$$I([z^\varepsilon]) \cap \{1, \ldots, n\} = J \cap \{1, \ldots, n\} = J, \tag{22}$$

which provides $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$ via the definition of $\mathcal{I}^*_{\mathcal{B}_{cell}}$. This shows the inclusion '$\supseteq$' in the first identity. Concerning the reverse direction of the second identity, let $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$ and $z \in \mathcal{Z}$ be arbitrary, such that (19) holds true. Applying again (20), this time consecutively to $\{\xi^1, \ldots, \xi^n\}$ and to $\{\xi^1, \ldots, \xi^N\}$, one deduces the existence of some z^ε such that (21) and

$$\{\xi^1, \ldots, \xi^N\} \cap [z^\varepsilon] = \{\xi^1, \ldots, \xi^N\} \cap \text{int}[z] \tag{23}$$

hold true simultaneously. From (21) it follows (22) as in the lines above. Therefore, $I([z^\varepsilon]) \in \varphi(J)$ (see (14)) and

$$\gamma^J \geq \sum_{i \in I([z^\varepsilon])} p_i = \sum_{\xi^i \in [z^\varepsilon]} P(\xi^i) = P([z^\varepsilon]) = P(\text{int}[z]),$$

where the last equality relies on (23). Since $z \in \mathcal{Z}$ was chosen arbitrarily such that (19) holds true, this shows the inequality '$\geq$' in the second identity. $\square$

Corollary 3.1 suggests that one can calculate the index family $\mathcal{I}^*_{\mathcal{B}_{cell}}$ as well as all upper coefficients γ^J for $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$, as soon as one knows the set $\mathcal{Z}$ of supporting cells which is finite. Indeed, it follows from Definition 3.1, that each supporting cell is defined by an s-tuple $\{x^1, \ldots, x^s\}$. Consequently, one may extract the finite system of all s-tuples $\{x^1, \ldots, x^s\}$ out of the set $R = \{\xi^1, \ldots, \xi^n\} \cup \{r^1, \ldots, r^s\}$ and check for each of its $\binom{n+s}{s}$ members, whether this s-tuple defines a supporting cell. If so, this cell contributes to the calculation of $\mathcal{I}^*_{\mathcal{B}_{cell}}$ and of γ^J for $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$. Referring again to the Tables 1 and 2 in Sect. 4 we indicate that, depending on s, the number of critical index sets may be significantly smaller than $\binom{n+s}{s}$. It remains to determine the lower coefficients γ_J for $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$.

Proposition 3.2 *For all $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$, one has $\gamma_J = \sum_{i \in I} p_i$, where*

$$I := \{i \in \{1, \ldots, N\} | \xi^i_k \leq \max_{j \in J} \xi^j_k \ (k = 1, \ldots, s)\}.$$

Proof Completely analogous to the derivation of (24) in the Appendix, one obtains that

$$\gamma_J = \min\left\{P([y]) | [y] \cap \{\xi^1, \ldots, \xi^n\} = \bigcup_{j \in J} \{\xi^j\}\right\}.$$

Define $z \in \mathbb{R}^s$ by $z_k := \max_{j \in J} \xi^j_k$ for $k = 1, \ldots, s$. Then, $\xi^j \in [z]$ for all $j \in J$ and, thus,

$$\bigcup_{j \in J} \{\xi^j\} \subseteq [z] \cap \{\xi^1, \ldots, \xi^n\}.$$

Assume that this inclusion is strict. Then, there is some $i \in \{1, \ldots, n\} \setminus J$ such that $\xi^i \in [z]$. $J \in \mathcal{I}^*_{\mathcal{B}_{cell}}$ means that there exists some $B \in \mathcal{B}_{cell}$ with $J = I(B) \cap \{1, \ldots, n\}$. Consequently, $\xi^j \in B$ for all $j \in J$, which entails that $[z] \subseteq B$, by construction of z. We derive that $\xi^i \in B$ and, hence, $i \in I(B)$. On the other hand, $i \in \{1, \ldots, n\} \setminus J$, which is a contradiction. It follows that

$$\bigcup_{j \in J} \{\xi^j\} = [z] \cap \{\xi^1, \ldots, \xi^n\}$$

and, thus, $\gamma_J \leq P([z])$. On the other hand, if $y \in \mathbb{R}^s$ is arbitrary feasible in the definition of γ_J, then $\xi^j \in [y]$ for all $j \in J$, and so, $[z] \subseteq [y]$ again by construction of z. Now, $P([z]) \leq P([y])$ which, upon passing to the minimum over all feasible y, provides $P([z]) \leq \gamma_J$. We may conclude that

$$\gamma_J = P([z]) = \sum_{\xi^i \in [z]} p_i,$$

which proves the assertion of the proposition. □

4 Algorithm and numerical results

The results of the previous sections suggest the following algorithmic approach for the solution of problem (11) in the special case of the cell discrepancy $\alpha_{\mathcal{B}_{cell}}$, given the problem data $N, n, s, \{\xi^1, \ldots, \xi^N\}, \{p_1, \ldots, p_N\}$, and the fixed support $\{\xi^1, \ldots, \xi^n\}$.

Algorithm 4.1

Step [1]: To find an open rectangle (c, d) such that $\{\xi^1, \ldots, \xi^N\} \subseteq (c, d)$, define

$$c_j := \min_{i=1,\ldots,N} \xi^i_j - 1, \ d_j := \max_{i=1,\ldots,N} \xi^i_j + 1 \ (j = 1, \ldots, s).$$

Calculate the set R from Definition 3.1 with points r^j defined in (18).

Put $\mathcal{I}_{\mathcal{B}}^* := \{\emptyset\}$, $\mathcal{A}' := \{\emptyset\}$ and $\gamma^J := 0$ for all $J \subseteq \{1, \ldots, n\}$.

Let $\mathcal{A} := \{A \subseteq \{1, \ldots, n+s\} \mid \#A = s\}$.

Step [2]: If $\mathcal{A}' = \mathcal{A}$, then go to Step [7], else select $A = \{i_1, \ldots, i_s\} \in \mathcal{A} \setminus \mathcal{A}'$.

Put $\mathcal{A}' := \mathcal{A}' \cup \{A\}$ and

$$x^j := \begin{cases} \xi^{i_j} & \text{if } i_j \leq n, \\ r^{i_j - n} & \text{if } i_j > n \end{cases} \quad (j = 1, \ldots, s).$$

Then, $x^j \in R$ for $j = 1, \ldots, s$.

Step [3]: Define a 'potential' supporting cell $[z]$ for some $z \in \mathbb{R}^s$ by

$$z_k := \max_{j=1,\ldots,s} x_k^j \ (k = 1, \ldots, s).$$

Check whether it is indeed a supporting cell according to Definition 3.1.

If not so, then go to Step [2].

Step [4]: Define J by the relation (19), i.e.,

$$J := \{i \in \{1, \ldots, n\} \mid \xi_k^i < z_k \ (k = 1, \ldots, s)\}.$$

If $\gamma^J = 1$, then go to Step [2]. Else if $J \in \mathcal{I}_{\mathcal{B}}^*$ then go to Step [6].

Step [5]: $\mathcal{I}_{\mathcal{B}}^* := \mathcal{I}_{\mathcal{B}}^* \cup \{J\}$. Calculate γ_J according to Proposition 3.2.

Step [6]: Calculate $\alpha := \sum_{i \in I} p_i$, where $I := \{i \in \{1, \ldots, N\} \mid \xi_k^i < z_k \ (k = 1, \ldots, s)\}$.

If $\alpha > \gamma^J$, then $\gamma^J := \alpha$. Go to Step [2].

Step [7]: With the additional data $\mathcal{I}_{\mathcal{B}}^*$ and γ_J, γ^J for all $J \in \mathcal{I}_{\mathcal{B}}^*$ solve the linear

optimization problem (16).

Remark 4.1 When using Algorithm 4.1 repeatedly with varying support, e.g. within a forward selection, it is desirable to reuse some of the data $\underline{\mathcal{I}}_{\mathcal{B}}^*$ and $\underline{\gamma}_J, \underline{\gamma}^J$ for $J \in \underline{\mathcal{I}}_{\mathcal{B}}^*$ that has been computed for preceding support. To this end, we assume without loss of generality that the previous support was given by

$$(\xi^1, \ldots, \xi^{n_1}, \xi^{n+1}, \xi^{n+2}, \ldots, \xi^{n_2}) \quad \text{with } 0 < n_1 \leq n \leq n_2$$

and the new support has the form $(\xi^1, \ldots, \xi^n)$, i.e. the set $A := \{\xi^{n_1+1}, \ldots, \xi^n\}$ has been added to and $R := \{\xi^{n+1}, \ldots, \xi^{n_2}\}$ has been removed from the previous support. For $J \subset \{1, \ldots, n\}$, we define $[J]$ as the smallest closed cell containing $\{\xi^j : j \in J\}$. It is easy to see that we can get the values γ_J, γ^J from older ones, whenever one of the following two conditions holds for every point $\xi^m \in R \cup A$:

(i) $[J]$ is separated from ξ^m, i.e., there exists $\xi^k, k \in \{1, \ldots, n_1\}$ such that $\xi^k \notin [J]$ and $\xi^k \in [J \cup \{m\}]$.
(ii) $\xi^m \in \text{int}[J]$.

If this is the case, it follows that

$$\underline{J} \triangleq (J \setminus \{i | \xi^i \in A \text{ and (ii) holds for } \xi^i\}) \cup \{i | \xi^i \in R \text{ and (ii) holds for } \xi^i\} \in \underline{\mathcal{I}}_{\mathcal{B}}^*,$$

so we can set $\gamma_J = \underline{\gamma}_J$ and $\gamma^J = \underline{\gamma}^{\underline{J}}$.

Algorithm 4.1 remains unchanged until Step [4], where the parts b and c are inserted:

Step [4a]: Define J by the relation $J := \{i \in \{1, \ldots, n\} | \xi_k^i < z_k \ (k = 1, \ldots, s)\}$.

Step [4b]: If $J \in \mathcal{I}_{\mathcal{B}}^*$ and J is completed, go to Step [2].

Step [4c]: If either (i) or (ii) is fulfilled get γ_J, γ^J from the old data as described above, mark J as completed, add J to $\mathcal{I}_{\mathcal{B}}^*$ (if not already contained) and go to Step [2].

Step [4d]: If $J \in \mathcal{I}_{\mathcal{B}}^*$ go to Step [6].

This avoids the computation of some γ_J and many α in Steps [5] and [6], an effect that becomes noticeable with increasing number of scenarios N. In numerical examples reductions of the forward selection's running time of up to $70\% \, (N = 10\,000)$ have been observed.

Numerical experiments show, that the main effort in Algorithm 4.1 is spent for the determination of supporting cells (compared to this, the time consumed by the solution of the linear program in the last step is negligible). Supporting cells are identified in this algorithm by checking all subsets of cardinality s in a set of cardinality $n + s$. Therefore, the complexity of the algorithm is mainly determined by the binomial coefficient $\binom{n+s}{s}$. This suggests that the practical value of the algorithm is limited to small dimension s of the random distributions and moderate cardinality n of the reduced support or to small n and moderate s (due to the identity $\binom{n+s}{s} = \binom{n+s}{n}$). One might wonder if there is a more efficient way of determining supporting cells than just by crude enumeration. It seems, however, that basically all mentioned subsets are potential candidates for realizing the upper bound in the cell discrepancy between two measures. Indeed, recall that the cell discrepancy coincides with the supremum distance of distribution functions (the Kolmogorov distance). Figure 1 plots the difference of two discrete distribution functions having 30 and 18 atoms, respectively. It can be recognized at how many different regions, the maximum (positive or negative) deviation may occur.

On the other hand, the computational effort is basically linear in the cardinality N of the original support, so larger values of N seem to be no problem. This is easy to see, since the impact of N on Algorithm 4.1 is limited on the number of summands that constitute γ_J in Step [5] and α in Step [6].

In the following, we present some numerical results which are based on a $C++$ implementation of Algorithm 4.1, where the linear program (6) has been solved with *CPLEX 10.0* [3]. Calculations have been carried out on a 3 GHz PC with 1 GB RAM. Table 1 compiles some computing times of Algorithm 4.1 under different problem sizes. As indicated above, the dependence of running time on the size N of the original support is moderate (basically linear), whereas it quickly grows with the size n of the reduced support and with the dimension s. Table 2 shows the corresponding numbers of supporting cells, determining the number of constraints of the linear program

Fig. 1 Difference between the
distribution functions of two
discrete probability measures
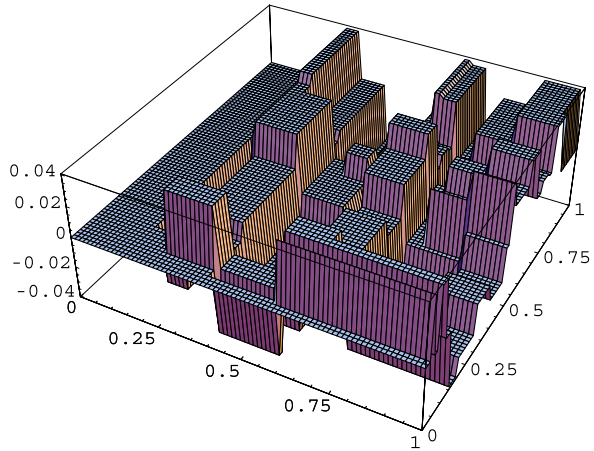(having 30 and 18 atoms,
respectively)

Table 1 Running time (in seconds) of Algorithm 4.1 for different problem parameters

	$N = 100$		$N = 200$		$N = 300$	
	$s = 4$	$s = 8$	$s = 4$	$s = 8$	$s = 4$	$s = 8$
$n = 10$	0.01	0.17	0.01	0.23	0.01	0.22
$n = 20$	0.30	29.58	0.29	27.52	0.82	23.45
$n = 30$	3.88	3175.53	3.95	3322.24	9.20	1792.74

Table 2 Number of reduced critical index sets $\#\mathcal{I}_{\mathcal{B}}^{*}$ for different problem parameters

	$N = 100$		$N = 200$		$N = 300$		$\binom{n+s}{s}$	
	$s = 4$	$s = 8$	$s = 4$	$s = 8$	$s = 4$	$s = 8$	$s = 4$	$s = 8$
$n = 10$	106	419	88	462	91	360	1,001	43,758
$n = 20$	747	14,525	740	13,844	1,489	12,500	10,626	3,108,105
$n = 30$	3,213	157,461	3,469	293,386	5,625	150,882	46,376	48,903,492

(16). Obviously, the vast majority of the $\binom{n+s}{s}$ potential supporting cells turn out not
to be supporting.

Figure 2 shows the decrease of the minimal discrepancy and the increase of the
running time in the course of a forward selection procedure (Algorithm 3.1), where
the initial 2-dimensional measure supported by 10 000 points sampled from a normal
distribution is approximated by measures having up to 25 atoms.

Figure 3 shows possible reductions of a randomly generated (w.r.t. support and
probabilities) 2-dimensional measure with $N = 1000$ atoms. A reduction on $n = 50$
atoms is considered. The left diagram of the figure illustrates the ordered solution de-
fined in (7). By definition, this solution selects the 50 atoms of the original measure
realizing the largest probabilities. The first 49 atoms even keep the original probabil-
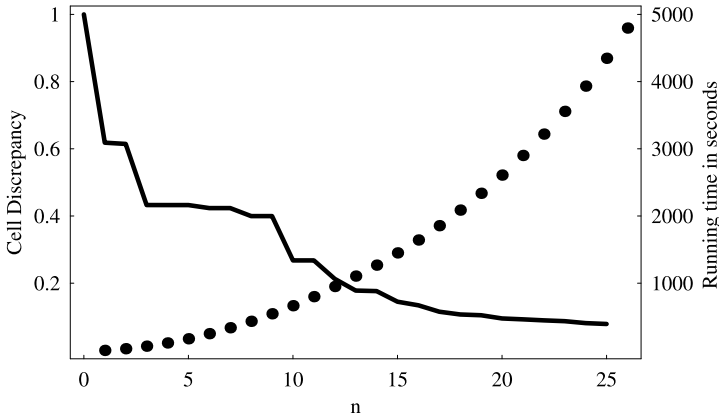
Fig. 2 Cell discrepancy and running time in the course of a forward selection procedure

Fig. 3 Reduction of a 2-dimensional measure with 1000 atoms (*thin lines*) to 50 atoms (*thick lines*). Heights correspond to probabilities. The *left diagram* illustrates the ordered solution, whereas the *right diagram* shows the measure which, on the fixed support of the ordered solution, is best approximating in the sense of the cell discrepancy. (The height of the leading atom in the left diagram has been cut off to about one tenth)
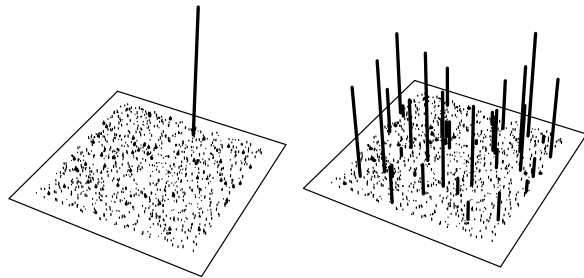
ities, whereas the last one is arranged to complement the sum of theses probabilities to one in order to make the selected measure a probability measure as well. Since each of the selected 50 atoms has rather low probability whereas the set of all non-selected 950 atoms has high probability, the ordered solution places almost all mass on atom no. 50. Although such choice would be optimal in case of the closed-set discrepancy $\alpha_{\mathcal{B}_{cl}}$ according to Proposition 2.1, it is intuitively clear from the picture that such solution may not be meaningful for other discrepancy distances. Indeed, the ordered solution realizes a closed-set discrepancy of $\alpha_{\mathcal{B}_{cl}} = 0.90$. Recalling, that discrepancies always take values between 0 and 1, this value is certainly not satisfactory. It illustrates the effect of 'linear decrease' mentioned in Sect. 2.1: a reduction of the support to 5% of its atoms leads to a reduction of the discrepancy which is no more than 10%. On the other hand, Proposition 2.1 tells us, that the same value of 0.90 is an upper bound for any solution of any other discrepancy. Now, measuring the ordered solution in the cell discrepancy instead, yields the slightly smaller value of $\alpha_{\mathcal{B}_{cell}} = 0.81$, which is still far from satisfactory. However, there is no reason, why the ordered solution should be optimal with respect to the cell discrepancy, neither from the choice of the support nor from the assignment of probabilities. We illus-

trate this fact by keeping the same support as the ordered solution but optimizing the probabilities according to Algorithm 4.1. The resulting solution is illustrated in the right diagram of Fig. 3. It realizes the cell discrepancy $\alpha_{\mathcal{B}_{cell}} = 0.08$, which is optimal for the chosen fixed support (of course, even better solutions might exist for a different support). This value of 8% discrepancy obtained by 5% of the original atoms highlights the potential of Algorithm 4.1 for scenario reduction.

Finally, we consider the following simple chance constrained optimization problem:

$$\min_{x,y \in \mathbb{R}} x + y \quad \text{s.t.} \quad P(\xi \le (x, y)) \ge p,$$

for some $p \in [0, 1]$. Thereby, $\xi = (\xi_1, \xi_2)$ is a 2-dimensional random vector, given by 1 000 equally weighted points sampled from a standard normal distribution.

We compare different scenario reduction techniques for several values of p and denote the optimal values of the initial and the approximated problem by v and $\tilde{v}$, respectively. In Fig. 4, the relative deviations $\frac{|v - \tilde{v}|}{|v|}$ of the optimal values are plotted for ξ being approximated by measures supported by up to $n = 50$ atoms.

The bold line stands for the forward selection together with Algorithm 4.1 and the thin line is the average error of 10 000 samples of dimension n. The dashed line represents an Quasi-Monte Carlo (QMC) approach, based on the first n points of the Halton sequence with bases 2 and 3, cf., e.g., [15]. The dotted line is obtained by readjusting the probabilities of the QMC points with Algorithm 4.1. It can be seen, that in most cases the measures obtained by forward selection produce (sometimes drastically) smaller errors than those constructed by sampling or QMC. Moreover, in many cases the quality of the QMC approximation could be significantly improved by simply readjusting the probabilities by a single run of Algorithm 4.1. This last approach is of course much faster than forward selection. On the other hand, there is no clear indication of which method provides better solutions.

Appendix

Proof of Proposition 3.1 Let $J \in \mathcal{I}_{\mathcal{B}_{cell}}^*$ be arbitrary. By definition of $\varphi(J)$ in (14), for any $I \in \varphi(J)$ there exists some $B \in \mathcal{B}_{cell}$ such that $I = I(B)$ and $J = I(B) \cap \{1, \ldots, n\}$. Then, by definition of $I(B)$,

$$\sum_{i \in I} p_i = \sum_{i \in I(B)} p_i = P(B),$$

whence

$$\gamma^J = \max_{I \in \varphi(J)} \sum_{i \in I} p_i = \max\{P(B) | B \in \mathcal{B}_{cell}, \ J = I(B) \cap \{1, \ldots, n\}\}$$

$$= \max\left\{ P(B) | B \in \mathcal{B}_{cell}, \ B \cap \{\xi^1, \ldots, \xi^n\} = \bigcup_{j \in J} \{\xi^j\} \right\}$$

$$= \max\left\{ P([y]) | y \in (c,d), \ [y] \cap \{\xi^1, \ldots, \xi^n\} = \bigcup_{j \in J} \{\xi^j\} \right\}, \qquad (24)$$

where in the last equation it was used that the support of P is contained in (c,d). Let $y^{(0)}$ be a maximizing cell, i.e.,

$$\gamma^J = P([y^{(0)}]), \quad y^{(0)} \in (c,d), \quad [y^{(0)}] \cap \{\xi^1, \ldots, \xi^n\} = \bigcup_{j \in J} \{\xi^j\}.$$

In the following, the cell $y^{(0)}$ will be successive enlarged in every dimension, until it becomes a supporting one. To this end, for $t \geq 0$ we put $y(t) := (y_1^{(0)} + t, y_2^{(0)}, \ldots, y_s^{(0)})$ and consider the enlargement $[y(t)]$ of the cell $[y^{(0)}]$ along the first coordinate direction. Put

$$\tau := \sup\{t | [y(t)] \cap R = [y^{(0)}] \cap R\}.$$
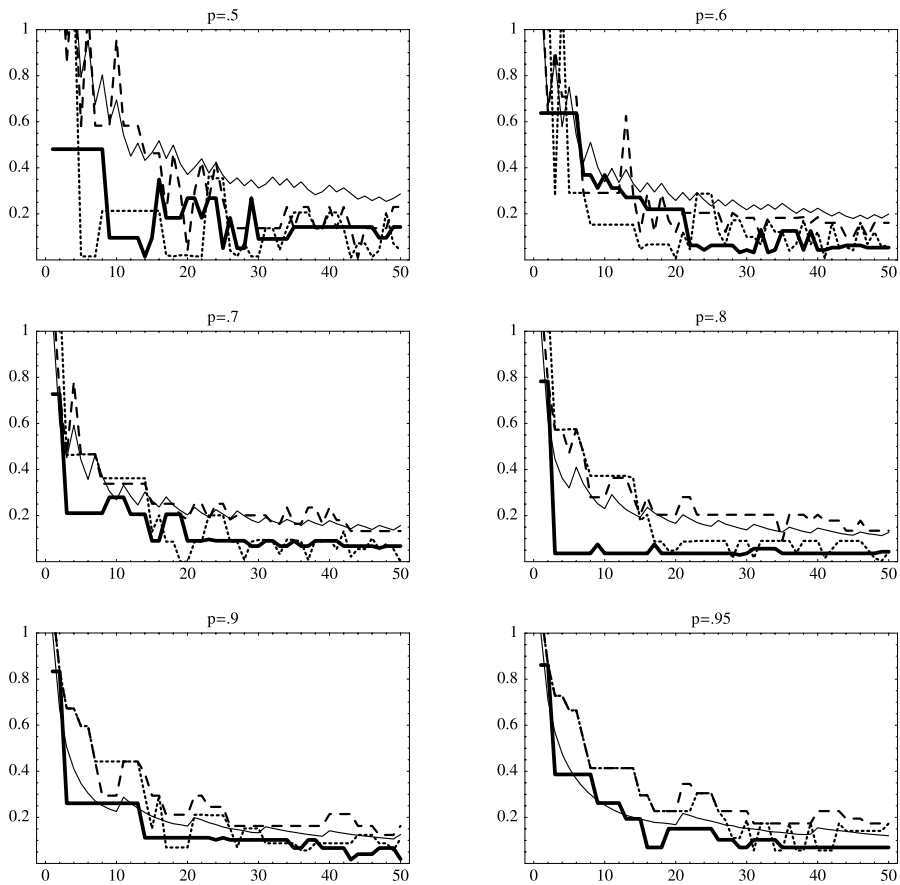
Fig. 4 Relative perturbation of the optimal value $\frac{|v - \tilde{v}|}{|v|}$, depending on n for forward selection (*bold*), sampling (*thin*), Quasi-Monte Carlo (*dashed*) and Algorithm 4.1 applied on Quasi-Monte Carlo (*dotted*)

Since the intersection of $[y(t)]$ with the finite set R remains constant for small $t > 0$, it follows that $\tau > 0$. From $y^{(0)} \in (c, d)$, we derive that $y_1^{(0)} < d_1 = r_1^1$, whence $r^1 \notin [y^{(0)}] \cap R$. On the other hand, with $t^* := d_1 - y_1^{(0)}$, one gets that $r^1 \in [y(t^*)] \cap R$. Consequently, $\tau \leq t^* < \infty$. We put $y^{(1)} := y(\tau)$ and arrive at

$$[y^{(1)}] \cap R = \{[y^{(1)}] \cap \{z | z_1 < y_1^{(1)}\} \cap R\} \cup \underbrace{\{[y^{(1)}] \cap \{z | z_1 = y_1^{(1)}\} \cap R\}}_{=:\Delta_1}$$

$$= \left\{ \bigcup_{t < \tau} [y(t)] \cap R \right\} \cup \Delta_1 = \{[y^{(0)}] \cap R\} \cup \Delta_1.$$

Once more, due to $[y(\tau + t)] \cap R$ remaining constant for small $t > 0$ and to the definition of τ, it follows that $[y^{(1)}] \cap R \neq [y^{(0)}] \cap R$. Consequently, $\Delta_1 \neq \varnothing$, which means that there exists some $x^1 \in [y^{(1)}] \cap R$ such that $x_1^1 = y_1^{(1)}$.

Now, in the second step, we extend the construction above for the second coordinate by defining

$$y(t) := (y_1^{(1)}, y_2^{(1)} + t, y_3^{(1)}, \ldots, y_s^{(1)})$$

$$\tau := \sup\{t | [y(t)] \cap R = [y^{(1)}] \cap R\}.$$

Upon observing that $y_2^{(1)} = y_2^{(0)} < d_2$ and replacing r^1 by r^2, we may repeat the same argumentation as before, in order to verify that $0 < \tau < \infty$, which allows to put $y^{(2)} := y(\tau)$. As before, the definition of τ allows to find that

$$[y^{(2)}] \cap R = \{[y^{(1)}] \cap R\} \cup \Delta_2 = \{[y^{(0)}] \cap R\} \cup \Delta_1 \cup \Delta_2,$$

where

$$\Delta_2 := \{[y^{(2)}] \cap \{y | y_2 = y_2^{(2)}\} \cap R\}.$$

With the same argument as in the first step, one infers that $\Delta_2 \neq \varnothing$, which means that there exists some $x^2 \in [y^{(2)}] \cap R$ such that $x_2^2 = y_2^{(2)}$.

Continuing the construction in this way for all coordinates, we finally arrive at points $y^{(j)}$ and x^j for $j = 1, \ldots, s$, such that

$$[y^{(j)}] \cap R = \{[y^{(0)}] \cap R\} \cup \Delta_1 \cdots \cup \Delta_j$$

$$\Delta_j = \{[y^{(j)}] \cap \{y | y_j = y_j^{(j)}\} \cap R\} \quad (j = 1, \ldots, s)$$

and

$$x^j \in [y^{(j)}] \cap R, \quad x_j^j = y_j^{(j)} \quad (j = 1, \ldots, s).$$

We put $z := y^{(s)}$. Then,

$$[z] \cap R = \{[y^{(0)}] \cap R\} \cup \Delta_1 \cdots \cup \Delta_s. \tag{25}$$

By the very construction, $[y^{(j)}] \subseteq [y^{(s)}]$ and $z_j = y_j^{(s)} = y_j^{(j)}$ for $j = 1, \ldots, s$. Consequently,

$$x^j \in [z] \cap R, \quad x_j^j = z_j \quad (j = 1, \ldots, s),$$

which means that $[z]$ is a supporting cell.

To prove the remaining statements of the proposition, note that the equalities $x_j^j = z_j$ for $x^j \in R$ and $j = 1, \ldots, s$ imply that $c_j \leq z_j \leq d_j$ for $j = 1, \ldots, s$ because the same bounds apply for all points of R. Again by construction, $y_j^{(j)} > y_j^{(j-1)}$ for $j = 1, \ldots, s$, which entails that $z_j > y_j^{(0)}$ for $j = 1, \ldots, s$. Now, we define a sequence of cells $[z^m]$ by

$$z_j^m := z_j - 1/m \quad (j = 1, \ldots, s). \tag{26}$$

It follows that, for m large enough, $z^m \in (c, d)$ and

$$z_j^m > y_j^{(0)} \quad (j = 1, \ldots, s). \tag{27}$$

We conclude from (26) and (25) that, for all $m \in \mathbb{N}$,

$$[z^m] \cap R \subseteq [z] \cap R = \{[y^{(0)}] \cap R\} \cup \Delta_1 \cdots \cup \Delta_s.$$

The definition of Δ_j shows that all elements of this subset have jth coordinate equal to $y_j^{(j)} = z_j$. Combining this with (26) yields

$$[z^m] \cap \Delta_j = \varnothing \quad (j = 1, \ldots, s).$$

Therefore, we may continue by

$$[z^m] \cap R \subseteq [y^{(0)}] \cap R.$$

On the other hand, $[z^m] \supseteq [y^{(0)}]$ by (27), hence $[z^m] \cap R = [y^{(0)}] \cap R$. We recall the fact that $y^{(0)} \in (c, d)$, whence—by definition of the set $\{r^1, \ldots, r^s\}$ in (18)— $[y^{(0)}] \cap \{r^1, \ldots, r^s\} = \varnothing$. With the same reasoning, the inclusion $z^m \in (c, d)$ stated above, yields that $[z^m] \cap \{r^1, \ldots, r^s\} = \varnothing$. Owing to the definition of $y^{(0)}$, we may continue as

$$[z^m] \cap \{\xi^1, \ldots, \xi^n\} = [z^m] \cap R = [y^{(0)}] \cap R = [y^{(0)}] \cap \{\xi^1, \ldots, \xi^n\}$$
$$= \bigcup_{j \in J} \{\xi^j\} \quad (m \in \mathbb{N}). \tag{28}$$

Clearly,

$$\text{int}[z] = \bigcup_{m \in \mathbb{N}} [z^m], \tag{29}$$

so (28) yields that

$$\text{int}[z] \cap \{\xi^1, \ldots, \xi^n\} = \bigcup_{j \in J} \{\xi^j\},$$

which is (19). Finally, (24) and the inclusion $[z^m] \supseteq [y^{(0)}]$ lead to

$$\gamma^J \geq P([z^m]) \geq P([y^{(0)}]) = \gamma^J \quad (m \in \mathbb{N}),$$

hence, actually equality holds true here. Since $[z^m]$ is an increasing sequence of cells in the union (29), one gets that

$$P(\text{int}[z]) = P\left(\bigcup_{m \in \mathbb{N}} [z^m]\right) = \lim_{m \to \infty} P([z^m]) = \gamma^J. \qquad \square$$

References

1. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
2. Billingsley, P., Topsøe, F.: Uniformity in weak convergence. Z. Wahrscheinlichkeitstheor. Verw. Geb. **7**, 1–16 (1967)
3. ILOG CPLEX 10.0, http://www.ilog.com/products/cplex
4. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics. Math. Program. **95**, 493–511 (2003)
5. Garey, M.R., Johnson, D.S.: Computers and Intractability—A Guide to the Theory of NP-Completeness. Freeman, New York (1979)
6. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comput. Optim. Appl. **24**, 187–206 (2003)
7. Heitsch, H., Römisch, W.: Scenario tree modelling for multistage stochastic programs. Preprint No. 296, DFG Research Center Matheon "Mathematics for key technologies" (2005, to appear in Math. Program.)
8. Heitsch, H., Römisch, W.: A note on scenario reduction for two-stage stochastic programs. Oper. Res. Lett. **35**, 731–738 (2007)
9. Henrion, R., Römisch, W.: Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. Math. Program. **84**, 55–88 (1999)
10. Henrion, R., Römisch, W.: Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints. Math. Program. **100**, 589–611 (2004)
11. Hlawka, E.: Zur Definition der Diskrepanz. Acta Arith. **18**, 233–241 (1971)
12. Hlawka, E., Niederreiter, H.: Diskrepanz in kompakten abelschen Gruppen I. Manuscr. Math. **1**, 259–288 (1969)
13. Kuipers, L., Niederreiter, H.: Uniform Distribution of Sequences. Wiley, New York (1974)
14. Mück, P., Philipp, W.: Distances of probability measures and uniform distribution mod 1. Math. Z. **142**, 195–202 (1975)
15. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia (1992)
16. Niederreiter, H., Wills, J.M.: Diskrepanz und Distanz von Maßen bezüglich konvexer und Jordanscher Mengen. Math. Z. **144**, 125–134 (1975)
17. Prékopa, A.: Stochastic Programming. Kluwer, Dordrecht (1995)
18. Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, Chichester (1991)
19. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics. Math. Oper. Res. **27**, 792–818 (2002)
20. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10, pp. 483–554. Elsevier, Amsterdam (2003)

21. Römisch, W., Schultz, R.: Stability analysis for stochastic programs. Ann. Oper. Res. **30**, 241–266 (1991)
22. Römisch, W., Wakolbinger, A.: Obtaining convergence rates for approximations in stochastic programming. In: Guddat, J., Jongen, H.T., Kummer, B., Nožička, F. (eds.) Parametric Optimization and Related Topics, pp. 327–343. Akademie Verlag, Berlin (1987)
23. Ruszczyński, A., Shapiro, A. (eds.): In: Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
24. Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse. SIAM J. Optim. **6**, 1138–1152 (1996)