Math. Program., Ser. A 100: 589–611 (2004)

Digital Object Identifier (DOI) 10.1007/s10107-004-0507-x

René Henrion · Werner Römisch

Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints

Received: December 17, 2002 / Accepted: January 15, 2004
Published online: March 10, 2004 –

Abstract. We study perturbations of a stochastic program with a probabilistic constraint and r-concave original probability distribution. First we improve our earlier results substantially and provide conditions implying Hölder continuity properties of the solution sets w.r.t. the Kolmogorov distance of probability distributions. Secondly, we derive an upper Lipschitz continuity property for solution sets under more restrictive conditions on the original program and on the perturbed probability measures. The latter analysis applies to linear-quadratic models and is based on work by Bonnans and Shapiro. The stability results are illustrated by numerical tests showing the different asymptotic behaviour of parametric and nonparametric estimates in a program with a normal probabilistic constraint.

Key words. probabilistic constraints – chance constraints – Lipschitz stability – stochastic optimization

1. Introduction

We consider the following optimization problem with chance constraints:

$$(P) \qquad \min\{g(x) \mid x \in X,\ \mathbb{P}(\xi \leq h(x)) \geq p\}.$$

Here ξ is an s-dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $g : \mathbb{R}^m \to \mathbb{R}$ is an objective, $X \subseteq \mathbb{R}^m$ is some abstract constraint set, $h : \mathbb{R}^m \to \mathbb{R}^s$ defines a system of inequalities and $p \in (0, 1)$ is some probability level. The meaning of the probabilistic constraint above is that the system of inequalities $\xi \leq h(x)$ has to be satisfied with probability p at least. The most prominent representative of (P) is given by linear constraints, i.e. $h(x) = Ax$ for some matrix A. By $\mu := \mathbb{P} \circ \xi^{-1} \in \mathcal{P}(\mathbb{R}^s)$ (the space of Borel probability measures on $\mathbb{R}^s$) we denote the probability distribution of ξ. Throughout the paper we shall make the following basic convexity assumptions:

g is convex, X is closed and convex, h has concave components and the probability measure μ is r-concave for some $r < 0$. (BCA)
The latter property means that μ^r is a convex set function, i.e.,

$$\mu^r(\lambda A + (1-\lambda)B) \leq \lambda\mu^r(A) + (1-\lambda)\mu^r(B)$$

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

W. Römisch: Humboldt-University Berlin Institute of Mathematics 10099 Berlin, Germany, e-mail: romisch@mathematik.hu-berlin.de

Mathematics Subject Classification (2000): 90C15, 90C31

holds true for all $\lambda \in [0, 1]$ and for all Borel measurable and convex $A, B \subseteq \mathbb{R}^s$ such that $\lambda A + (1-\lambda)B$ is Borel measurable too. Note that many prominent multivariate distributions (e.g. normal, Pareto, Dirichlet or uniform distribution on convex, compact sets) share the property of being r-concave for some $r < 0$ (see [12]).

Introducing the distribution function of the probability measure μ as $F_\mu(y) = \mu(z \in \mathbb{R}^s | z \leq y)$, the problem (P) can be equivalently rewritten as

$$(P) \qquad \min\{g(x) \mid x \in X,\ F_\mu(h(x)) \geq p\}.$$

Usually, only partial information about μ is available, and (P) is solved on the basis of some estimation $\nu \in \mathcal{P}(\mathbb{R}^s)$ of μ. Typically, ν is chosen as a parametric or nonparametric estimator of μ. Hence, rather than the original program (P), some substitute

$$(P_\nu) \qquad \min\{g(x) \mid x \in X,\ F_\nu(h(x)) \geq p\}$$

is solved. Although, at least in principle, arbitrarily good approximations ν of μ can be obtained, it is by no means obvious that the solutions of (P_ν) will well approximate those of $(P = P_\mu)$ as ν tends to μ. A counterexample illustrating 'wrong convergence' or emptiness of approximating solutions is provided by Example 15 in the appendix.

Although the original data are supposed to be convex, we do not make any assumptions on the data of the perturbed problems (P_ν). This allows to admit the important class of empirical approximations which lack any convexity or smoothness properties. Since, in general, the solutions of (P_ν) are not unique under the assumptions (BCA), we have to deal with solution sets. The dependence of solutions and optimal values on the parameter ν is described by the set-valued mapping $\Psi : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^m$ and the extended-valued function $\varphi : \mathcal{P}(\mathbb{R}^s) \to \bar{\mathbb{R}}$ via

$$\Psi(\nu) = \operatorname{argmin}\{g(x) \mid x \in X,\ F_\nu(h(x)) \geq p\}$$
$$\varphi(\nu) = \inf\{g(x) \mid x \in X,\ F_\nu(h(x)) \geq p\}.$$

We are interested in conditions formulated for the data of the original problem (P) such that Ψ and φ behave stable locally around the fixed measure μ. In order to measure distances among parameters and among solutions, we mostly rely on the Kolmogorov metric between probability measures

$$d_K(\nu_1, \nu_2) = \sup_{z \in \mathbb{R}^s} \left| F_{\nu_1}(z) - F_{\nu_2}(z) \right| \quad (\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^s))$$

and on the Hausdorff distance between closed subsets of $\mathbb{R}^m$

$$d_H(A, B) = \max\left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\} \quad (A, B \subseteq \mathbb{R}^m).$$

Qualitative stability results in the sense of $d_H(\Psi(\mu), \Psi(\nu)) \to 0$ as $d_K(\mu, \nu) \to 0$ have been obtained in [5] based on earlier works like [15] and [6]. These results guarantee that, under the imposed conditions (see Theorem 1 below), cluster points of approximating solutions will be solutions of the original problem and that any solution of the original problem is the limit of a sequence of approximating solutions. For further work in this direction we refer to [4, 9, 11, 17, 18].

Beyond qualitative stability it is of much interest to know how fast solutions or optimal values of approximating problems converge to original solutions, which is a question of quantitative stability. Recall that Ψ is Hausdorff-Hölder continuous with rate $\kappa > 0$ at μ, if there are $L, \delta > 0$ such that

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L\,[d_K(\mu, \nu)]^\kappa \quad \text{for all} \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \delta. \tag{1}$$

There exists an immediate link between Hausdorff-Hölder continuity with rate κ of the solution set mapping and exponential bounds for empirical solution estimates. Consider independent s-dimensional random vectors $\xi_1, \ldots, \xi_N$ on $(\Omega, \mathcal{A}, \mathbb{P})$ having common distribution μ. Then, $\nu_N(\omega) := N^{-1} \sum_{i=1}^N \delta_{\xi_i(\omega)}$ (with δ_z denoting the Dirac measure placing mass one at $z \in \mathbb{R}^s$) is an empirical measure approximating μ as $N \to \infty$. The deviation between the original solution set and its empirical approximation can be estimated as follows (see Proposition 6 in [6]):

$$\exists C > 0\,\forall N \in \mathbb{N}\,\forall \varepsilon > 0$$
$$P(d_H(\Psi(\mu), \Psi(\nu_N)) \geq \varepsilon) \leq C\,[N \cdot \lambda(\varepsilon, \delta, \kappa, L)]^{s-0.5} \exp(-2N \cdot \lambda(\varepsilon, \delta, \kappa, L)), \tag{2}$$

where $\lambda(\varepsilon, \delta, \kappa, L) = \left[\min\{\delta, (\varepsilon/L)^{1/\kappa}\}\right]^2$ and δ, L, κ refer to (1).

Conditions for Hausdorff-Hölder continuity of Ψ at rate $\kappa = 1/2$ were obtained in [6] for the special case of linear chance constraints with convex-quadratic objective and in [7] for the more general setting of the above data assumptions (BCA). The first part of this paper is devoted to a substantial improvement of the previous results in two directions: firstly, the mentioned results relate to so-called localized solution sets rather than to the solution sets themselves. This technical restriction seemed to be a necessary consequence of considering non-convex perturbations of the original convex measure. It turns out, however, that one can exploit additional arguments provided in [5] in order to get rid of localizations. Of course, statements on stability of solution sets themselves as in (1) are easier to interpret than their localized counterparts. Secondly, all previous results on quantitative stability of Ψ essentially relied on the condition

$$\Psi(\mu) \cap \operatorname{argmin}\{g(x) \mid x \in X\} = \emptyset, \tag{3}$$

which means that no solution of (P) is a solution of the relaxed problem with the chance constraint removed and vice versa. In this paper we shall obtain the same results without requiring such kind of strict complementarity condition.

Specializing our setting to linear chance constraints, the best (largest) rate we can obtain is $\kappa = 1/2$ provided that the random variable has at least dimension $s = 2$. Of course, the exponential bound in (2) improves with increasing κ. Thus, it is of much interest to find conditions ensuring even Hausdorff-Lipschitz continuity of Ψ ($\kappa = 1$). It is interesting to note that a Lipschitz rate results for linear chance constraints with $1-$dimensional random variable ξ (but with arbitrary dimension of the decision variable x). Yet, this observation seems to be too restrictive for practical relevance. The second part of the paper investigates more reasonable settings and conditions for Lipschitz rates. Two basic additional requirements turn out to be crucial then: firstly the approximating measures ν can no longer be arbitrary but have to be restricted to class $\mathcal{C}^{1,1}$ in a sense to be precised. Secondly, the strict complementarity condition (3) which was dispensable for the Hölder rate $\kappa = 1/2$, has to be incorporated into the set of conditions now. Doing so, one may derive an upper Lipschitz result for solution sets, but now with respect to a $\mathcal{C}^{1,1}$-type distance between probability measures which is stronger than the previously used Kolmogorov distance. Such a result might be useful for studying the asymptotic behaviour of nonparametric density estimators (cf. [16], Sect. 24) of the original distribution μ.

2. Hölder Stability

The main result of this section is stated with the technical details of proof left to the appendix. We start by recalling a result on qualitative stability of solution sets and quantitative stability of optimal values which is needed for the proof of our main theorem but which is also of independent interest:

Theorem 1. *(see [5], Th. 1) In addition to the basic convexity assumptions (BCA), let the following conditions be satisfied at the fixed probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$:*

1. *$\Psi(\mu)$ is nonempty and bounded.*
2. *There exists some $\hat{x} \in X$ such that $F_\mu(h(\hat{x})) > p$.*

Then, $\Psi : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^m$ is upper semicontinuous in the sense of Berge at μ, and there exist constants $L, \delta > 0$, such that

$$\Psi(\nu) \neq \emptyset \text{ and } |\varphi(\nu) - \varphi(\mu)| \leq L d_K(\nu, \mu) \text{ for all } \nu \in \mathcal{P}(\mathbb{R}^s) \text{ with } d_K(\nu, \mu) < \delta.$$

We note that the Lipschitz estimate for φ in the previous theorem is restricted in the sense that one of the measures (μ) has to be kept fixed. A full Lipschitz result, where both measures are allowed to vary freely around μ does not hold true under the given assumptions (see Example 1 in [8]).

The key for obtaining quantitative stability results for solution set is a two-level decomposition of the parametric program (P_ν). To this aim, we introduce the following objects, where V is an open ball containing $\Psi(\mu)$ under the boundedness assumption of Theorem 1:

$$\begin{aligned} Y_V &= [h(X \cap \operatorname{cl} V) + \mathbb{R}^s_-] \cap F_\mu^{-1}([p/2, 1]) \\ \pi(y) &= \inf\{g(x) \mid x \in X \cap \operatorname{cl} V,\ h(x) \geq y\}, \\ \sigma(y) &= \operatorname{argmin}\{g(x) \mid x \in X \cap \operatorname{cl} V,\ h(x) \geq y\}\ (y \in Y_V) \\ .Y(\nu) &= \operatorname{argmin}\{\pi(y) \mid y \in Y_V,\ F_\nu(y) \geq p\} \quad (\nu \in \mathcal{P}(\mathbb{R}^s)) \end{aligned}$$

Note that σ and π denote the solution set and optimal value, respectively, of a lower level parametric program, the parameter y of which refers to right-hand side perturbations of the inequalities defined by the mapping h. In contrast, the multifunction Y represents the solution set of an upper level parametric program, where the explicit inequality constraint reduces to a description based on distribution functions F_ν. This allows to separate the influence of F_ν and h in the inequality defining (P_ν). The relation between lower and upper level solution sets and optimal values on the one hand and overall solution set and optimal value of (P_ν) is characterized in Proposition 10 in the appendix.

Now, we are in a position to state the main result of this section which is proved in the appendix (following the proof of Prop. 12).

Theorem 2. *In addition to the basic convexity assumptions (BCA), let the following conditions be satisfied at some fixed* $\mu \in \mathcal{P}(\mathbb{R}^s)$*:*

1. $\Psi(\mu)$ *is nonempty and bounded.*
2. *There exists some* $\hat{x} \in X$ *such that* $F_\mu(h(\hat{x})) > p$.
3. F_μ^r *is strongly convex on some convex open neighbourhood* U *of* $Y(\mu)$*, where* $r < 0$ *is chosen from (BCA) such that* μ *is* r*- concave.*
4. σ *is Hausdorff Hölder continuous with rate* κ^{-1} *on* Y_V.

Then, Ψ *is Hausdorff Hölder continuous with rate* $(2\kappa)^{-1}$ *at* μ*, i.e., there are* $L, \delta > 0$ *such that*

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L\, [d_K(\mu, \nu)]^{1/(2\kappa)} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \delta.$$

The first assumption of Theorem 2 is of technical nature. It can be enforced, for instance, by compactness of X (the nonemptiness of the compact constraint set is then a consequence of the second assumption). The second assumption can be interpreted as a Slater condition (see proof of Prop. 12). In special situations, its verification is possible without explicit knowledge of the measure μ. For instance, in the situation of linear chance constraints under nonnegativity restrictions ($h(x) = Ax$, $X = \mathbb{R}^m_+$), it suffices to know that $A \geq 0$ and that A does not contain zero rows. Indeed, for $v := A\mathbf{1}$ with $\mathbf{1} = (1, \ldots, 1)$, one has $v_i > 0$ for all i. Consequently, $\lim_{\lambda\to\infty} F_\mu(\lambda v) = 1$. Since $p < 1$, there is some $\lambda > 0$ such that $F_\mu(\lambda v) > p$. Hence, $F_\mu(A\hat{x}) > p$ for $\hat{x} := \lambda\mathbf{1} \in X$, which is condition 2. in Theorem 2. An alternative situation occurs when $X = \mathbb{R}^m$ and A has linearly independent rows.

The third assumption of Theorem 2 is satisfied for r- concave measures ($r < 0$) for which F_μ^r is strongly convex on bounded, convex sets (because $Y(\mu)$ is compact, see Prop. 10). An example for such measure is the multivariate normal distribution with independent components, as it is shown in Proposition 14 in the appendix. To prove the same result in the correlated case appears to be much more involved. But even measures lacking the mentioned property of 'global' strong convexity may still satisfy the third assumption. For instance, the uniform distribution over multidimensional rectangles is r- concave for any $r < 0$ and F_μ^r is strongly convex on this rectangle. All one has to know then is that $Y(\mu)$ is contained in the rectangle too. Unfortunately, not all uniform distributions over polytopes share the required strong convexity property (e.g., the uniform distribution over the triangle $\mathrm{conv}\{(1,0), (0,1), (1,1)\}$ is r- concave for any $r < 0$ but F_μ^r fails even to be strictly convex on this triangle). If h is linear, i.e., $h(x) = Ax$, then the strong convexity assumption can be simplified in the sense that it is supposed to hold on some convex open neighbourhood U of $A(\Psi(\mu))$.

In the last assumption of Theorem 2, it is assumed that some Hölder continuity of σ with respect to the Hausdorff distance is known. This is the case, for instance, for linear mappings h, polyhedral sets X and convex-quadratic functions g . Then the Hölder rate for σ equals 1 (see Th. 4.2 in [10] or Prop. 2.4 in [7]) and we have the following Corollary to Theorem 2:

Corollary 3. *In addition to the basic convexity assumptions (BCA), let* g *be convex-quadratic,* h *linear and* X *polyhedral. Then, supposing the first three assumptions of*

Theorem 2, Ψ is Hausdorff Hölder continuous with rate $1/2$ at μ, i.e., there are $L, \delta > 0$ such that

$$d_H(\Psi(\mu), \Psi(\nu)) \le L\sqrt{d_K(\mu, \nu)} \quad \text{for all} \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \delta.$$

Apart from this application to linear chance constraints defined by h, there is also a chance of identifying a Hölder rate of σ in the more general situation considered here, when h has concave components (see Prop. 2.4 in [7] for more details).

Example 16 in the appendix demonstrates that the Hölder rate obtained in Theorem 2 and in Corollary 3 is sharp. This observation is refined in Example 17 in the appendix, in order to show that the Hölder rate 1/2 in Corollary 3 is realized, in particular, by discrete approximations of μ. In both of these counter-examples, the objective function was defined by a degenerate convex quadratic form. Using more sophisticated constructions, one could verify the sharpness of the Hölder rate also in case of linear or strongly convex objective functions g (e.g., Example 2.10 in [7]).

On the other hand, all these examples live in $\mathbb{R}^2$. The following Proposition which is proved in the appendix (following the proof of Prop. 13) confirms that the Hölder rates of Theorem 2 and Corollary 3 can be improved as long as the random variable ξ is one-dimensional (the decision variable x is arbitrary). Moreover, in this special case no strong convexity assumption is needed for the measure μ (condition 3. in Theorem 2):

Proposition 4. *In addition to the basic convexity assumptions (BCA), let $s = 1$ and assume conditions 1.,2. and 4. of Theorem 2. Then, Ψ is Hausdorff Hölder continuous with rate κ^{-1} at μ. In the context of Corollary 3, Ψ is even Hausdorff Lipschitz continuous (rate $\kappa = 1$) at μ.* □

3. Lipschitz Stability

The Lipschitz result of Proposition 4 (in the context of Corollary 3) is based on the one-dimensionality of the random variable which is rather restrictive in stochastic programming. In order to derive Lipschitz stability in a multivariate setting, one has to impose further conditions and also to restrict the class of considered measures (for the original as well as the approximating one). The subsequent analysis relies on general stability results obtained in [1, 2]results to the setting which will be of interest here:

Theorem 5. *(see [2], Th. 4.81) Consider the parametric optimization problem*

$$\min\{f(x)|G(x, u) \in K\},$$

where $f : \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^m \times U \to \mathbb{R}^q$, U is a Banach space, $K = \mathbb{R}^{q_1}_- \times \{0\}_{q_2}$, $q_1 + q_2 = q$. Denote by $S(u) := \arg\min\{f(x)|G(x, u) \in K\}$ the parametric solution set and fix some parameter $u_0 \in U$. Let the following conditions hold true:

1. *f and G are $\mathcal{C}^{1,1}$ functions (differentiable with Lipschitz continuous derivative).*
2. *$S(u_0) \neq \emptyset$ and S is uniformly bounded in a neighbourhood of u_0.*

3. *f satisfies a second order growth condition with respect to $S(u_0)$, i.e., there exist a neighbourhood $\mathcal{V}$ of $S(u_0)$ and a constant $c > 0$ such that*

$$f(x) \geq f_0 + c\text{dist}^2(x, S(u_0)) \quad \forall x \in \mathcal{V},\ G(x, u_0) \in K$$
$$(f_0 = \inf\{f(x) | G(x, u_0) \in K\}).$$

4. *For all $x \in S(u_0)$ it holds that*

$$\{\nabla_x G_i(x, u_0)\}_{i=1,\dots,q_2} \cup \{\nabla_x G_j(x, u_0)\}_{j \in I(x)}$$

is a set of linearly independent vectors in $\mathbb{R}^m$, where

$$I(x) = \{j \in \{1, \dots, q_1\} | G_j(x, u_0) = 0\}.$$

Then, S is upper Lipschitz at u_0, i.e., there are constants $L, \delta > 0$ such that

$$\text{dist}(x, S(u_0)) \leq L \|u - u_0\| \quad \forall x \in S(u)\, \forall u \in U,\ \|u - u_0\| < \delta.$$

In order to apply Theorem 5 to our parametric problem (P_ν), we have to interpret the parameter u as distribution functions F_ν where $\nu \in \mathcal{P}(\mathbb{R}^s)$. However, condition 1. requires to restrict the admissible class of measures to those having $\mathcal{C}^{1,1}$ distribution function. More precisely, we introduce the following space:

$$\mathcal{C}_b^{1,1}(\mathbb{R}^n) := \{f \in \mathcal{C}^1(\mathbb{R}^n) | f \text{ is bounded and has a bounded, Lipschitzian derivative}\}$$

With the norm

$$\|f\|_b^{1,1} := \max\left\{ \sup_{x \in \mathbb{R}^n} |f(x)|,\ \sup_{x \in \mathbb{R}^n} \|\nabla f(x)\|,\ \sup_{x, y \in \mathbb{R}^n, x \neq y} \frac{\|\nabla f(x) - \nabla f(y)\|}{\|x - y\|} \right\},$$

$\mathcal{C}_b^{1,1}(\mathbb{R}^n)$ becomes a Banach space.

In the parametric problem (P_ν), let us specify the general convexity assumptions (BCA) in the following sense:

- The objective function g is convex-quadratic, i.e., $g(x) = \langle x, Hx \rangle + \langle c, x \rangle$ for some positive semidefinite (m, m)-matrix H ($H = 0$ possible) and some $c \in \mathbb{R}^m$.
- $h(x) = Ax$, where A is a matrix of order (s, n).
- X is a polyhedron and has an explicit description

$$X = \{x \in \mathbb{R}^m | \langle \alpha_j, x \rangle \leq a_j \ (j = 1, \dots, \tilde{q}_1); \quad \langle \beta_i, x \rangle = b_i \ (i = 1, \dots, \tilde{q}_2)\}.$$

- For some fixed probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ it holds that μ is r-concave for some $r < 0$.

In this setting, the following result was proved in [6] (Th. 8):

Theorem 6. *Let the following conditions be satisfied at μ fixed in the setting above:*

1. *$\Psi(\mu)$ is nonempty and bounded.*
2. *F_μ^r is strongly convex on some convex open neighbourhood U of the compact set $A(\Psi(\mu))$.*

3. *There exists some* $\hat{x} \in X$ *such that* $F_\mu(A\hat{x}) > p$.
4. $\Psi(\mu) \cap \operatorname{argmin}\{g(x) \mid x \in X\} = \emptyset$.

Then, there exist a neighbourhood $\mathcal{V}$ *of* $\Psi(\mu)$ *and a constant* $c > 0$ *such that*

$$g(x) \geq \varphi(\mu) + c\operatorname{dist}^2(x, \Psi(\mu)) \quad \forall x \in \mathcal{V} \cap X,\ F_\mu(Ax) \geq p \quad .$$

Now, we are in a position to formulate the desired stability result:

Theorem 7. *Let the following conditions be satisfied at* μ *fixed in the setting above:*

1. $\Psi(\mu)$ *is nonempty and bounded.*
2. F_μ^r *is strongly convex on some convex open neighbourhood* U *of the compact set* $A(\Psi(\mu))$.
3. $F_\mu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.
4. *For all* $x \in \Psi(\mu)$, *the following set is linearly independent, where* $J(x) = \{j \in \{1, \ldots, \tilde{q}_1\} \mid \langle \alpha_j, x \rangle = a_j\}$:

$$\{\nabla F_\mu(Ax) \cdot A\} \cup \{\alpha_j\}_{j \in J(x)} \cup \{\beta_i\}_{i=1,\ldots,q_2}.$$

5. $\Psi(\mu) \cap \operatorname{argmin}\{g(x) \mid x \in X\} = \emptyset$.

Then, the solution set mapping Ψ *is upper Lipschitz continuous at* μ *in the accordingly restricted class of probability measures, i.e., there are constants* $L, \delta > 0$ *such that*

$$\begin{aligned} \operatorname{dist}(x, \Psi(\mu)) \ &\leq L \left\| F_\mu - F_\nu \right\|_b^{1,1} \quad \forall x \in \Psi(\nu)\, \forall \nu \in \mathcal{P}(\mathbb{R}^s),\, F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s), \\ &\left\| F_\mu - F_\nu \right\|_b^{1,1} < \delta. \end{aligned}$$

Proof. We are going to apply Theorem 5 with $U := \mathcal{C}_b^{1,1}(\mathbb{R}^s)$, $u_0 := F_\mu$, $q_1 := \tilde{q}_1 + 1$, $q_2 := \tilde{q}_2$, $G_1(x, u) := p - u(Ax)$, $G_j(x, u) := \langle \alpha_{j-1}, x \rangle$ $(j = 2, \ldots, \tilde{q}_1 + 1)$, $G_i(x, u) := \langle \beta_i, x \rangle$ $(i = 1, \ldots, \tilde{q}_2)$. Then, obviously, the constraint sets in Theorems 5 and 7 coincide for all $u := F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$, $\nu \in \mathcal{P}(\mathbb{R}^s)$:

$$G(x, u) \in K \Longleftrightarrow x \in X,\ u(Ax) \geq p.$$

In particular, $S(u) = \Psi(\nu)$. The partial derivatives of G are calculated as

$$\nabla_x G(x, u) = \begin{pmatrix} -\nabla u(Ax)A \\ \alpha_j^T \ (j = 1, \ldots, \tilde{q}_1) \\ \beta_i^T \ (i = 1, \ldots, \tilde{q}_2 \end{pmatrix}, \quad \nabla_u G(x, u) = \begin{pmatrix} L \\ 0_{\tilde{q}_1 + \tilde{q}_2} \end{pmatrix},$$

where $Lu = -u(Ax)$. From the definition of $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$ one easily verifies that G belongs to the class $\mathcal{C}^{1,1}$, hence, assumption 1. of Theorem 5 is satisfied.

Next, we show:

$$\text{there is some } \hat{x} \in X \text{ such that } F_\mu(A\hat{x}) > p. \tag{4}$$

To this aim, choose some $x \in \Psi(\mu)$ according to condition 1. in our theorem. Then, $x \in X$ and $F_\mu(Ax) \geq p$. Owing to condition 4., there is a solution v of the linear system

$$\langle \nabla(F_\mu \circ A)(x), v \rangle = 1, \quad \langle \alpha_j, v \rangle = \langle \beta_i, v \rangle = 0 \ (j \in J(x), i = 1, \ldots, \tilde{q}_2).$$

Then, for $\varepsilon > 0$ sufficiently small, $\hat{x} := x + \varepsilon v$ satisfies (4). Now, (4) along with condition 1. entails upper semicontinuity of Ψ at μ via Theorem 1, whence assumption 2. of Theorem 5. The quadratic growth of f required in assumption 3. of Theorem 5 follows from Theorem 6 together with (4) upon taking into account that $f_0 = \varphi(\mu)$. Finally, assumption 4. of Theorem 5 follows immediately from condition 4. in our theorem (recall that $\nabla_x G_1(x, u_0) = -\nabla F_\mu(Ax) \cdot A$). □

When comparing the last Theorem with Corollary 3 which imposes the same data requirements, the stronger Lipschitz result is mainly based on two additional assumptions (leaving apart the condition 4. of linear independence in Theorem 7 which can be understood as a modification of the Slater type condition in the previous results): firstly, condition 5. requires that the chance constraint $F_\mu(Ax) \geq p$ affects the solution of the problem. If this condition is violated, no Lipschitz rate can be expected for solutions even when all remaining assumptions of Theorem 7 hold true. This can be seen from a small modification of Example 16 upon replacing the uniform distribution there by some bivariate normal distribution with independent components in order to meet the data requirements of Theorem 7. In that example, the solution set of the fixed problem with chance constraint is the same as the solution set of the unconstrained problem with removed chance constrained. As a consequence, a Hölder rate of 1/2 results.

Secondly, the probabiliy measures in Theorem 7 are restricted to have distribution functions in the space $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. This applies for the fixed measure μ as well as to its perturbations ν (see statement of the result in Theorem 7) Again, without such restriction no Lipschitz rate could be obtained. We refer once more to Example 2.10 in [7] (which would have to be slightly modified in the same sense as before). In this example, all assumptions of Theorem 7 are satisfied. However the perturbed measures are just Lipschitz continuous and do not belong to $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. They are constructed in such a way that the perturbed solution set $\Psi(\nu)$ grows at a Hölder rate of 1/2 away from the unperturbed solution set $\Psi(\mu)$.

Although the result in Theorem 7 is stronger than that of Corollary 3 in that it improves the Hölder rate towards a Lipschitz rate, it provides only an upper estimate whereas the estimate of Corollary 3 is two-sided by relying on the Hausdorff distance. Furthermore, even the upper estimate of Theorem 7 is slightly weaker than its one-sided counterpart in Corollary 3, since, by definition of $\|\cdot\|_b^{1,1}$ and of d_K, one has

$$\left\| F_{\nu_1} - F_{\nu_2} \right\|_b^{1,1} \leq d_K(\nu_1, \nu_2) \quad \text{for all } \nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^s),\ F_{\nu_1}, F_{\nu_2} \in \mathcal{C}_b^{1,1}(\mathbb{R}^s).$$

Of course, imposing new restrictions raises the question of which class of probability measures still meets the new assumptions. Theorem 7 requires that both the original and all the perturbed measures have distribution functions in $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$. The following proposition identifies two classes of such measures:

Proposition 8. *Let $\nu \in \mathcal{P}(\mathbb{R}^s)$.*

1. *If ν is a nondegenerate multivariate normal distribution, then $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.*
2. *If ν is the distribution of a random vector with independent components and if the 1-dimensional distributions $\nu_i \in \mathcal{P}(\mathbb{R})$ of these components have bounded and Lipschitzian densities f_{ν_i}, then $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.*

Proof. Ad 1.: Without loss of generality, one may consider standard normal distributions (zero mean and unit variances). It is well known then (e.g. [12], p. 203), that the partial derivatives of F_ν can be calculated as

$$\frac{\partial F_\nu}{\partial x_i}(x) = \tilde{F}_{\tilde{\nu}}(\tilde{x}_i) \cdot f(x_i) \quad (i = 1, \dots, s),$$

where $\tilde{F}_{\tilde{\nu}}$ is the distribution function of some nondegenerate multiv $\tilde{\nu} \in \mathcal{P}(\mathbb{R}^{s-1})$, $\tilde{x}_i \in \mathbb{R}^{s-1}$ and f is the density of the 1-dimensional standard normal distribution. Taking into account that F_ν, $\tilde{F}_{\tilde{\nu}}$ and f are bounded (say by some $M > 0$), it follows immediately that $F_\nu \in \mathcal{C}^1(\mathbb{R}^s)$ is bounded and has bounded derivative. Since $\tilde{F}_{\tilde{\nu}}$ (as a nondegenerate multivariate normal distribution function) and f are Lipschitzian on $\mathbb{R}^{s-1}$ and $\mathbb{R}$, respectively, it follows that the partial derivatives of F_ν are Lipschitzian on $\mathbb{R}^s$ (as products of functions which are bounded and Lipschitzian on $\mathbb{R}^s$). Hence, $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$.
Ad 2.: Clearly, F_ν is bounded as a distribution function. By the assumption of independence, $F_\nu = F_{\nu_1} \cdots F_{\nu_s}$, where F_{ν_i} are the marginal distributions of ν. Since the marginal densities f_{ν_i} were assumed to be Lipschitzian, the F_{ν_i} and, hence, F_ν itself are of class $\mathcal{C}^1$. The assumed boundedness of the f_{ν_i} yields that the F_{ν_i} are Lipschitzian. Furthermore,

$$\frac{\partial F_\nu}{\partial x_1} = f_{\nu_1} \cdot F_{\nu_2} \cdots F_{\nu_s}.$$

Therefore, $\frac{\partial F_\nu}{\partial x_1}$ is bounded and Lipschitzian according to the assumptions. The same argumentation applies to the other partial derivatives, whence $F_\nu \in \mathcal{C}_b^{1,1}(\mathbb{R}^s)$. □

4. Illustration of the Stability Results

In this section we illustrate the obtained stability result for a simple 2-dimensional example. We consider the problem

$$\min\{x_1 + x_2 | P(\xi_1 \le x_1, \xi_2 \le x_2) \ge 1/2\},$$

where ξ is assumed to have a distribution μ which is normal with independent components of mean zero and unit variance. Clearly, this problem satisfies the basic data assumptions (BCA). The solution set of this problem consists of a singleton $\Psi(\mu) = \{q, q\}$, where $q \approx 0.55$ is the $1/\sqrt{2}$-quantile of the 1-dimensional standard normal distribution. First, we check the assumptions of Theorem 2. Obviously, $\Psi(\mu)$ is nonempty and bounded. Next, a Slater point certainly exists, any $\hat{x}$ with $\hat{x}_1 = \hat{x}_2 > q$ satisfies $F_\mu(\hat{x}) > F_\mu(q, q) = 1/2$. Furthermore, as μ is a normal distribution with independent components, F_μ^r is strongly concave for any $r < 0$ and on any bounded, convex set (see remarks below Theorem 2). As a consequence, F_μ^r is strongly concave on some convex, open neighbourhood of $\Psi(\mu)$. Summarizing, the first three assumptions of Theorem 2 are satisfied. Finally, in our example, g is linear (in particular: convex-quadratic), $X = \mathbb{R}^2$ is trivially polyhedral and h is linear as the identity. Hence, Corollary 3 guarantees the Hausdorff Hölder continuity with rate 1/2 of the solution set mapping Ψ for any approximation $\nu \in \mathcal{P}(\mathbb{R}^s)$ of μ.

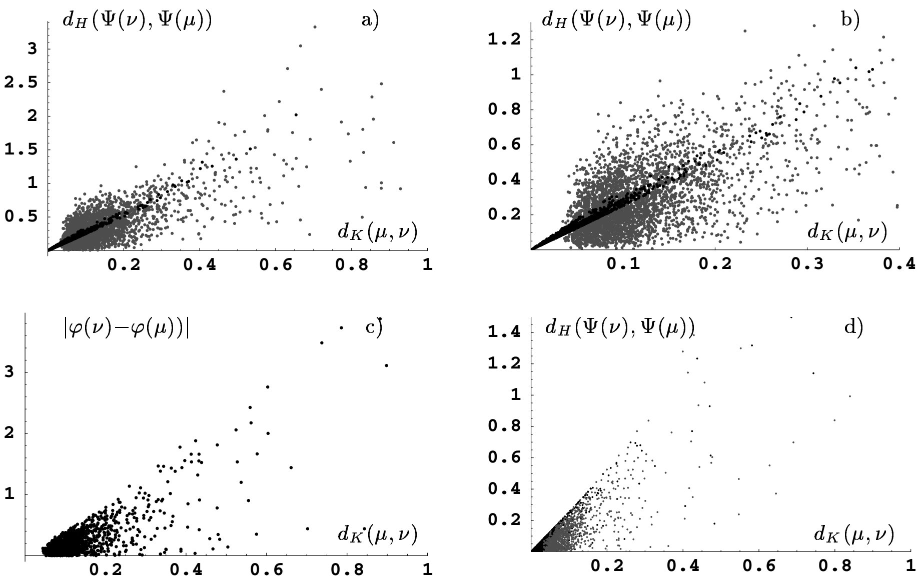

Fig. 1. Illustration of stability results for simulated data

We want to focus now on two specific approximations both of which are based on a sample $Z_1, \dots Z_N$ of observations of ξ. The empirical measure derived from this sample is defined as $\nu = N^{-1} \sum_{i=1}^{N} \delta_{Z_i}$, where δ_Z is the Dirac measure placing mass one at the point Z. The empirical measure is a suitable approximation when no information at all is available about the true measure μ. If, on the other hand, partial information about μ is given, better adapted approximations may be favorable. For instance, if we know that μ in our problem is some nondegenerate multivariate normal distribution (but do not know its parameters), then a parametric approximation defining a normal distribution with mean and (co-) variances estimated from $Z_1, \dots Z_N$ may be useful. We want to symbolize this parametric approximation by ν'. Of course, with increasing sample size N, $d_K(\nu, \mu)$ and $d_K(\nu', \mu)$ will tend to zero in a probabilistic sense, and $d_K(\nu', \mu)$ will do so even faster than $d_K(\nu, \mu)$. The issue we want to address here is convergence of the approximating solution sets, i.e., dependence of $d_H(\Psi(\nu), \Psi(\mu))$ on $d_K(\nu, \mu)$. To this aim, several thousand samples of ξ were simulated according to its distribution μ. The sample size varied up to a few hundred.

Figure 1 a) illustrates the results for the parametric (black dots) and empirical (gray dots) approximations. Clearly, in both cases the approximating solutions converge to the true solution when the approximating measure converges to the true measure. Indeed, this kind of qualitative stability is already ensured by the first two assumptions of Theorem 2 via Theorem 1. From a quantitative point of view, however, the solution sets of parametric approximations seem to converge much faster (in the worst case) than those of empiric approximation. This is particularly obvious in a region close to the origin which has been magnified in Figure 1 b). According to the diagram, there is no doubt that there exists an upper Lipschitz estimation for the parametric approximation, whereas in case of the empiric estimation increasingly large ratios between the two distances seem to be possible when $d_K(\nu, \mu)$ tends to zero. This suggests a Non-Lipschitzian relation in accordance with the observation from Example 17 that discrete approximations may lead to Hölder rate 1/2 for the stability of solution sets. At least, Corollary 3 guarantees that the corresponding cloud of points lies below some function $\alpha\sqrt{d_K(\nu, \mu)}$, where $\alpha > 0$ is sufficiently large.

As far as the parametric approximation is concerned, its Lipschitzian behavior is supported by Theorem 7. To see this, recall that both the original and the approximating measures are normal distributions, hence, their distribution functions belong to the space $\mathcal{C}_b^{1,1}(\mathbb{R}^s)$ according to Proposition 8. Furthermore, the gradient of a (nondegenerate) normal distribution function is always nonzero which yields condition 4. of Theorem 7. Finally, owing to the fact that the objective function in our example is linear, condition 5. of Theorem 7 is trivially fulfilled. It has to be noted, that Theorem 7 provides a Lipschitz result with respect to the distance $\|F_\mu - F_\nu\|_b^{1,1}$, whereas Figure 1 b) even suggests a Lipschitzian relation with respect to the stronger Kolmogorov distance $d_K(\mu, \nu)$.

As far as optimal values are concerned, Theorem 1 guarantees a Lipschitzian estimation for any approximating measure. This is observed empirically in Figure 1 c) for the example of empirical approximation (the better behaved parametric case is omitted here).

Finally, we may reduce our example to a 1-dimensional setting, i.e., to the problem

$$\min\{x | P(\xi \le x) \ge 1/2\},$$

where ξ is assumed to have a standard normal distribution μ. In this situation, the dependence of Hausdorff distances between solution sets on Kolmogorov distances between measures is seen from Figure 1 d) to be of Lipschitzian nature for both types of approximations (black dots on top of gray dots). Again, this observation is supported by our results via Proposition 4, according to which the Lipschitz rate results for any approximating measure.

5. Appendix

Lemma 9. *For $r < 0$ and $\nu \in \mathcal{P}(\mathbb{R}^s)$ it holds: If $F_\nu(y) \ge w > 0$ for all $y \in Q \subseteq \mathbb{R}^s$, then there exist constants $c, \delta > 0$ such that*

$$\left|F_\nu^r(y) - F_{\nu'}^r(y)\right| \le c d_K(\nu, \nu') \quad \forall y \in Q\, \forall \nu' \in \mathcal{P}(\mathbb{R}^s),\ d_K(\nu, \nu') < \delta.$$

Proof. Note that

$$\left|u^r - v^r\right| \le |r| \max\{u^{r-1}, v^{r-1}\} |u - v| \quad \forall u, v > 0.$$

Then, choosing $\delta := w/2$, one has

$$F_{\nu'}(y) \ge w/2 > 0 \quad \forall y \in Q\, \forall \nu' \in \mathcal{P}(\mathbb{R}^s),\ d_K(\nu, \nu') < \delta.$$

Fix c as $|r|\,(w/2)^{r-1}$. □

The next lemma provides a two-level decomposition for solutions and optimal values of the parametric problem (P_ν):

Proposition 10. *(see [5], Lemma 1) Under the assumptions of Theorem 1 let V be an open ball containing $\Psi(\mu)$. With the notations introduced in front of Theorem 2 it holds that*

1. *Y_V is convex and compact.*
2. *π is convex, finite and lower semicontinuous on Y_V.*
3. *There is some $\delta > 0$ such that for all $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$*

$$\varphi(\nu) = \inf\{\pi(y) \mid y \in Y_V, F_\nu(y) \geq p\} \tag{5}$$
$$\Psi(\nu) = \sigma(Y(\nu)) \tag{6}$$

4. *$Y : \mathcal{P}(\mathbb{R}^s) \rightrightarrows \mathbb{R}^s$ is upper semicontinuous at μ.*

The following Proposition allows separately to study quantitative stability of lower and upper level solution sets, respectively, in order to derive quantitative stability of the overall solution set:

Proposition 11. *In addition to the assumptions of Theorem 1 suppose that*

1. *Y is Hausdorff Hölder continuous with rate 1/2 at μ, i.e., there are constants $\rho, \delta > 0$ such that*

$$d_H(Y(\mu), Y(\nu)) \leq \rho d_K^{1/2}(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \delta.$$

2. *σ is Hausdorff Hölder continuous with rate κ^{-1} on Y_V, i.e., there exists $L > 0$ such that*

$$d_H(\sigma(z), \sigma(y)) \leq L d^{\kappa^{-1}}(y, z) \quad \forall z, y \in Y_V.$$

Then, Ψ is Hausdorff Hölder continuous with rate $(2\kappa)^{-1}$ at μ. More precisely, it holds that

$$d_H(\Psi(\mu), \Psi(\nu)) \leq L\rho^{\kappa^{-1}} [d_K(\mu, \nu)]^{(2\kappa)^{-1}} \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \delta.$$

Proof. For a nonempty and closed subset $Q \subseteq \mathbb{R}^s$ and $y \in \mathbb{R}^s$ denote by $proj^Q(y)$ the projection of y onto Q. Note that for $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$ and small enough δ, one has $\Psi(\nu) \neq \emptyset$ (Theorem 1) and $Y(\nu) \neq \emptyset$ by (6). Furthermore, the sets $Y(\nu)$ are closed. Indeed, by definition and by (5), they can be represented as

$$Y(\nu) = \{y \in Y_V | F_\nu(y) \geq p\} \cap \{y \in Y_V | \pi(y) \leq \varphi(\nu).$$

The first set on the right is closed due to the upper semicontinuity of distribution functions and by statement 1. of Proposition 10. The second set is closed because of the lower semicontinuity of π (statement 2. of Proposition 10). Consequently, $proj$ applies to these sets $Y(\nu)$. Recalling that $Y(\nu) \subseteq Y_V$, it follows from the assumptions and from (6), that for $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \delta$

$$\begin{aligned}
d_H(\Psi(\mu), \Psi(\nu)) &= \max\{\sup_{x\in\Psi(\mu)} d(x, \Psi(\nu)), \sup_{x'\in\Psi(\nu)} d(x', \Psi(\mu))\} \\
&= \max\{\sup_{y\in Y(\mu)} \sup_{\alpha\in\sigma(y)} d(\alpha, \sigma(Y(\nu))), \sup_{y'\in Y(\nu)} \sup_{\beta\in\sigma(y')} d(\beta, \sigma(Y(\mu)))\} \\
&\le \max\{\sup_{y\in Y(\mu)} \sup_{\alpha\in\sigma(y)} d(\alpha, \sigma(proj^{Y(\nu)}(y))), \\
&\qquad \sup_{y'\in Y(\nu)} \sup_{\beta\in\sigma(y')} d(\beta, \sigma(proj^{Y(\mu)}(y')))\} \\
&\le L \max\{\sup_{y\in Y(\mu)} d^{\kappa^{-1}}(y, proj^{Y(\nu)}(y)), \sup_{y'\in Y(\nu)} d^{\kappa^{-1}}(y', proj^{Y(\mu)}(y'))\} \\
&\le L \left[\max\{\sup_{y\in Y(\mu)} d(y, Y(\nu)), \sup_{y'\in Y(\nu)} d(y', Y(\mu))\}\right]^{\kappa^{-1}} \\
&\le L\,[d_H(Y(\mu), Y(\nu))]^{\kappa^{-1}} \\
&\le L\rho^{\kappa^{-1}}\,[d_K(\mu, \nu)]^{(2\kappa)^{-1}}.
\end{aligned}$$

□

The next proposition, which may be considered as the technical core of our analysis, provides a verifiable condition for the upper level solution set Y being Hausdorff Hölder continuous with rate 1/2. The key here is an argument of strong convexity.

Proposition 12. *Under the assumptions of Theorem 1 consider the parametric program*

$$(\widetilde{P}_\nu) \quad \min\ \{\pi(y) \mid y \in Y_V, F_\nu(y) \ge p\} \qquad (\nu \in \mathcal{P}(\mathbb{R}^s))$$

near $\mu \in \mathcal{P}(\mathbb{R}^s)$, the solution set mapping and optimal value function of which are given by Y and φ, respectively (see Prop. 10). Let the following assumption be satisfied in addition, where $r < 0$ refers to the exponent of concavity of μ

- *F_μ^r is strongly convex on some convex open neighbourhood U of $Y(\mu)$.*

Then, Y is Hausdorff Hölder continuous with rate 1/2 at μ.

Proof. Setting $b_\nu(y) := F_\nu^r(y) - p^r$ for $\nu \in \mathcal{P}(\mathbb{R}^s)$, the original problem $(\widetilde{P}_\mu)$ may be written as

$$(\widetilde{P}_\mu) \quad \min\ \{\pi(y) \mid y \in Y_V, b_\mu(y) \le 0\}.$$

As a consequence of the r- concavity of μ (where $r < 0$), F_μ^r is a convex (possibly extended-valued) function. Therefore, b_μ is a convex function finite-valued on Y_V (see definition of Y_V). Then, in view of 1. and 2. in Proposition 10, $(\widetilde{P}_\mu)$ is a convex program which satisfies the Slater condition $b_\mu(\hat{y}) < 0$ for some $\hat{y} \in Y_V$. Indeed, we may choose $x^* \in \Psi(\mu) \neq \emptyset$ (first assumption of Th. 1), hence $x^* \in X \cap V$ and $b_\mu(h(x^*)) \le 0$. Furthermore, $\hat{x} \in X$ taken from the second assumption of Theorem 1 satisfies $b_\mu(h(\hat{x})) < 0$.

With F_μ being nondecreasing as a distribution function, the composition $F_\mu^r \circ h$ is convex too due to F_μ^r being nonincreasing ($r < 0$) and to h having concave components. Therefore, $b_\mu \circ h$ is convex and, for sufficiently small $\lambda > 0$, $x_\lambda := \lambda \hat{x} + (1-\lambda)x^*$ satisfies $b_\mu(h(x_\lambda)) < 0$. Now, one may take $\hat{y} := h(x_\lambda)$.

Statement 4. in Proposition 10 and Lemma 9 guarantee that for some $c, \delta' > 0$

$$Y(\nu) \subseteq U, \ \left|b_\nu(y) - b_\mu(y)\right| \leq c\, d_K(\mu, \nu) \quad \forall y \in Y_V, \ \forall \nu \in \mathcal{P}(\mathbb{R}^s), \ d_K(\mu, \nu) < \delta'. \tag{7}$$

Finally, the additional assumption on strong convexity of F_μ^r on U means in particular that

$$b_\mu(y_1/2 + y_2/2) \leq b_\mu(y_1)/2 + b_\mu(y_2)/2 - \rho \left\|y_1 - y_2\right\|^2 \quad \forall y_1, y_2 \in U \tag{8}$$

for some $\rho > 0$. We proceed by case distinction with respect to the relation between $Y(\mu)$ and the solution set $Q := \arg\min\{\pi(y) \mid y \in Y_V\}$ of the relaxed problem $(\widetilde{P}_\mu)$ where the chance constraint $b_\mu(y) \leq 0$ is omitted.

Case 1. $Y(\mu) \cap Q = \emptyset$.

Choose some $y^* \in Y(\mu)$ (recall that $Y(\mu) \neq \emptyset$ due to $\Psi(\mu) \neq \emptyset$ and to (6)). Since π and b_μ are finite-valued on Y_V (statement 2. of Prop. 10 and $\varphi(\mu) = \pi(y^*) > -\infty$ (see (5)), the Slater condition shown above for problem $(\widetilde{P}_\mu)$ ensures the existence of a Lagrange multiplier $\lambda^* \geq 0$ such that (cf. [13], Cor. 28.2.1)

$$\pi(y^*) = \min\ \{\pi(y) + \lambda^* b_\mu(y) \mid y \in Y_V\} \text{ and } \lambda^* b_\mu(y^*) = 0. \tag{9}$$

By the case 1- assumption, one has $\lambda^* > 0$ and so $\pi + \lambda^* b_\mu$ is strongly convex on $Y_V \cap U$ due to the additional assumption in this lemma. This implies

$$\tilde{\rho} \left\|y - y^*\right\|^2 \leq \pi(y) + \lambda^* b_\mu(y) - \pi(y^*) \qquad \text{for all} \quad y \in Y_V \cap U. \tag{10}$$

for some $\tilde{\rho} > 0$ (due to $\lambda^* b_\mu(y^*) = 0$ and y^* being a minimizer in (9)). In particular, y^* is the unique minimizer of $(\widetilde{P}_\mu)$, i.e., $Y(\mu) = \{y^*\}$. For an arbitrary ν taken from (7), (10) applies. Using the results of Proposition 10 and the fact that $b_\nu(y) \leq 0$ for all $y \in Y(\nu)$ one arrives at the asserted Hölder continuity with respect to the Hausdorff distance:

$$\begin{aligned} d_H(Y(\mu), Y(\nu)) = \sup_{y \in Y(\nu)} d(y, y^*) &\leq \tilde{\rho}^{-1/2} \sup_{y \in Y(\nu)} \left[\pi(y) - \pi(y^*) + \lambda^*(F_\mu^r(y) - p^r)\right]^{1/2} \\ &\leq \tilde{\rho}^{-1/2} \sup_{y \in Y(\nu)} \left[\varphi(\nu) - \varphi(\mu) + \lambda^*(b_\mu(y) - b_\nu(y))\right]^{1/2} \\ &\leq \tilde{\rho}^{-1/2} \left[L d_K(\mu, \nu) + \lambda^* c d_K(\mu, \nu)\right]^{1/2} \\ &\leq \tilde{\rho}^{-1/2} (L + \lambda^* c)^{1/2} d_K(\mu, \nu)^{1/2} \end{aligned}$$

for all $\nu \in \mathcal{P}(\mathbb{R}^s)$, $d_K(\mu, \nu) < \min\{\delta', \delta\}$ and with $L, \delta > 0$ from Theorem 1.

Case 2. $Y(\mu) \cap Q \neq \emptyset$. In this case, $Y(\mu)$ has the simple representation

$$Y(\mu) = \{y \in Q \mid b_\mu(y) \leq 0\}. \tag{11}$$

Note that Q is closed and convex by the properties of π and Y_V stated in Proposition 10.

Case 2.1. $\exists\ \bar{y} \in Y(\mu)$, $b_\mu(\bar{y}) < 0$. Then, $\bar{y}$ is a Slater point of the constraint $b_\mu(y) \leq 0$ with respect to Q. Then, by the Robinson-Ursescu Theorem (cf. [14]), the inverse H^{-1} of the multifunction

$$H(t) := \{y \in Q | b_\mu(y) \leq t\}.$$

is metrically regular at all points $(y, 0)$ with $y \in Y(\mu)$. This amounts to the existence of neighbourhoods U_y and constants $\varepsilon_y, L_y > 0$ such that

$$d(y', H(t)) \leq L_y \max\{b_\mu(y') - t, 0\} \quad \forall t, t' \in (-\varepsilon_y, \varepsilon_y)\, \forall y' \in Q \cap U_y \tag{12}$$

Now, let $\nu \in \mathcal{P}(\mathbb{R}^s)$ be arbitrary such that $d_K(\mu, \nu) < \delta'$ with δ' from (7). If $y' \in H(-cd_K(\mu, \nu))$ (where c refers to (7)), then $y' \in Q$ and

$$b_\mu(y') \leq -cd_K(\mu, \nu) \leq \min\{0, b_\mu(y') - b_\nu(y')\}$$

by definition of H and of $d_K(\mu, \nu)$. It follows that

$$H(-cd_K(\mu, \nu)) \subseteq Y(\mu) \cap Y(\nu). \tag{13}$$

Combining (12) with (13), we obtain for all $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \min\{\delta', c^{-1}\varepsilon_y\}$:

$$\begin{aligned} \max\{d(y', Y(\mu)), d(y', Y(\nu))\} &\leq d(y', H(-cd_K(\mu, \nu))) \\ &\leq L_y \max\{b_\mu(y') + cd_K(\mu, \nu), 0\} \\ &\leq \begin{cases} L_y cd_K(\mu, \nu) & \forall y' \in Y(\mu) \cap U_y \\ 2L_y cd_K(\mu, \nu) & \forall y' \in Y(\nu) \cap U_y \end{cases}, \end{aligned}$$

where in the second estimation the relation $b_\mu(y') \leq b_\mu(y') - b_\nu(y') \leq cd_K(\mu, \nu)$ was used (see 7). Summarizing, each $y \in Y(\mu)$ is supplied with neighbourhoods U_y of y and constants $\tilde{\varepsilon}_y, \tilde{L}_y > 0$ such that

$$\begin{aligned} d(y', Y(\nu)) &\leq \tilde{L}_y d_K(\mu, \nu) \quad \forall y' \in Y(\mu) \cap U_y \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \tilde{\varepsilon}_y \\ d(y', Y(\mu)) &\leq \tilde{L}_y d_K(\mu, \nu) \quad \forall y' \in Y(\nu) \cap U_y \quad \forall \nu \in \mathcal{P}(\mathbb{R}^s),\ d_K(\mu, \nu) < \tilde{\varepsilon}_y. \end{aligned}$$

The compactness of $Y(\mu) \subseteq Y_V$ (statement 1. of Prop. 10) then allows to extract constants $\varepsilon^*, L > 0$ and an open set $\tilde{U}$ containing $Y(\mu)$ such that for all $\nu \in \mathcal{P}(\mathbb{R}^s)$ with $d_K(\mu, \nu) < \varepsilon^*$ one has

$$\begin{aligned} d(y, Y(\nu)) &\leq Ld_K(\mu, \nu) \quad \forall y \in Y(\mu) \\ d(y, Y(\mu)) &\leq Ld_K(\mu, \nu) \quad \forall y \in Y(\nu) \cap \tilde{U}. \end{aligned}$$

By upper semicontinuity of Y (statement 4. of Prop. 10), one has $Y(\nu) \subseteq \tilde{U}$ for all $\nu \in \mathcal{P}(\mathbb{R}^s)$, $d_K(\mu, \nu) < \varepsilon'$ with some $\varepsilon' > 0$. Hence, even Hausdorff Lipschitz continuity of Y at μ follows from the above inequalities: $d_H(Y(\mu), Y(\nu)) \leq Ld_K(\mu, \nu)$ for all $\nu \in \mathcal{P}(\mathbb{R}^s)$, $d_K(\mu, \nu) < \min\{\varepsilon^*, \varepsilon'\}$. This, of course, implies the asserted Hölder continuity with rate 1/2.

Case 2.2. $b_\mu(y) = 0\ \forall y \in Y(\mu)$.
The convexity of $Y(\mu)$ along with (8) yield that $Y(\mu)$ reduces to a singleton, say $Y(\mu) = \{y^*\}$. Then, $b_\mu(y^*) = 0$ and $y^* \in Q \subseteq Y_V$ by (11). For any ν satisfying (7), let $y \in Y(\nu) \subseteq U$ be arbitrary, hence $y \in Y_V$ and $b_\nu(y) \leq 0$. Put

$$\lambda' := \inf\{\lambda \geq 0 \mid b_\mu(\lambda y^* + (1-\lambda)y) \leq 0\}.$$

Then, $\lambda' \in [0, 1]$. Define $y' := \lambda' y^* + (1-\lambda')y$. Assume first that $\lambda' > 0$. Since the convex function $\alpha(\lambda) = b_\mu(\lambda y^* + (1-\lambda)y)$ is upper semicontinuous on $[0, 1]$ and continuous on $(0, 1)$, it follows that $b_\mu(y') = 0$. Since, for $\lambda' > 0$, $b_\mu(y) > 0$, one has $y' \neq y$ and $b_\mu(y/2 + y'/2) > 0$ according to the definition of y'. Then, (7) and (8) yield

$$\begin{aligned} c d_K(\mu, \nu) &\geq b_\mu(y) - b_\nu(y) \geq b_\mu(y)/2 + b_\mu(y')/2 \geq b_\mu(y/2 + y'/2) + \rho \left\| y - y' \right\|^2 \\ &\geq \rho \left\| y - y' \right\|^2, \end{aligned}$$

whence

$$\left\| y - y' \right\| \leq \sqrt{c/\rho}\sqrt{d_K(\mu, \nu)}. \tag{14}$$

In the excluded case of $\lambda' = 0$, the same inequality follows trivially from $y' = y$. Now, we want to estimate the distance between y' and y^*, hence, without loss of generality, we may assume that $y' \neq y^*$. Then, $\lambda' < 1$ and $y' \notin Q$ (if $y' \in Q$, then $y' \in Y(\mu)$ due to $b_\mu(y') = 0$ and (11), whence a contradiction to $Y(\mu) = \{y^*\}$). Now, $y \notin Q$ since $y^* \in Q$ and Q is convex (otherwise the contradiction $y' \in Q$). Consequently, $\pi(y) > \pi(y^*)$. Put, $y'' := y^*/2 + y'/2$, hence $y'' = \frac{\lambda'+1}{2} y^* + \frac{1-\lambda'}{2} y$, which is a convex combination of y^* and y. Then, $y'' \in Y_V \cap U$ due to convexity of $Y_V \cap U$. It follows that

$$\pi(y'') \leq \frac{\lambda'+1}{2}\pi(y^*) + \frac{1-\lambda'}{2}\pi(y) < \pi(y).$$

If $b_\nu(y'') \leq 0$, then a contradiction to $y \in \Psi(\nu)$ results, hence $b_\nu(y'') > 0$. Again referring to (7) and (8), it follows that

$$\begin{aligned} c d_K(\mu, \nu) &\geq b_\nu(y'') - b_\mu(y'') \geq -b_\mu(y^*/2 + y'/2) \geq -(b_\mu(y^*) + b_\mu(y'))/2 \\ &\quad + \rho \left\| y^* - y' \right\|^2 = \rho \left\| y^* - y' \right\|^2. \end{aligned}$$

Combining this with (14), one arrives at the desired estimation

$$d_H(Y(\mu), Y(\nu)) = \sup_{y \in Y(\nu)} \left\| y - y^* \right\| \leq 2\sqrt{c/\rho}\sqrt{d_K(\mu, \nu)}. \qquad \square$$

Collecting the previous results allows to prove our main theorem on Hölder rates:

Proof of Theorem 2. The first two assumptions of the Theorem serve to apply Theorem 1 in Propositions 10, 11 and 12. By virtue of the third assumption, Proposition 12 guarantees Hausdorff Hölder continuity of Y (upper level solution set) at μ with rate $1/2$. Combining this with the last assumption of the Theorem (Hölder rate for the lower level solution set), Proposition 11 provides the stated result. □

In the case of a 1-dimensional random variable, the assertion of Proposition 12 can be sharpened even without the strong convexity assumption made there:

Proposition 13. *If $s = 1$ then, under the assumptions of Theorem 1, Y is Hausdorff Lipschitz continuous (i.e., Hausdorff Hölder continuous with rate $\kappa = 1$) at μ.*

Proof. We consider the parametric program from Proposition 12 which Y is the solution mapping of:

$$(\widetilde{P}_\nu) \quad \min\ \{\pi(y) \mid y \in Y_V, F_\nu(y) \geq p\} \qquad (\nu \in \mathcal{P}(\mathbb{R}))$$

We have $Y_V = [a, b]$ for some $a, b \in \mathbb{R}$ (see 1. in Prop. 10). Choosing some $x^* \in \Psi(\mu) \subseteq X$ according to the assumption of Theorem 1 , it follows that $h(x^*) \in Y_V \neq \emptyset$, hence $a \leq b$. Since F_ν is upper semicontinuous and nondecreasing as a distribution function, one gets

$$\{y \in \mathbb{R} \mid F_\nu(y) \geq p\} = [\alpha(\nu), \infty), \quad \alpha(\nu) := \min\{y \in \mathbb{R} \mid F_\nu(y) \geq p\}.$$

Clearly, $\{\alpha(\nu)\}$ is the solution set of a parametric program of type (P_ν) (see introduction) which at the fixed measure μ satisfies the basic data assumptions (BCA) (with $g(x) = h(x) = x$ and $X = \mathbb{R}$). Since $p \in (0, 1)$ and F_μ is a distribution function, there exists some $\bar{y} \in \mathbb{R}$ with $F_\mu(\bar{y}) > p$. Now, Theorem 1 allows to derive the existence of $L, \delta > 0$ such that

$$|\alpha(\nu) - \alpha(\mu)| = |\varphi(\nu) - \varphi(\mu)| \leq L d_K(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \delta,$$

where $\varphi(\nu)$ refers to the optimal value function of the parametric problem defining $\alpha(\nu)$. Summarizing, we may rewrite $(\widetilde{P}_\nu)$ as

$$(\widetilde{P}_\nu) \quad \min\ \{\pi(y) \mid y \in [b(\nu), b]\} \qquad (\nu \in \mathcal{P}(\mathbb{R})),$$

where $b(\nu) := \max\{\alpha(\nu), a\}$ satisfies

$$|b(\nu) - b(\mu)| \leq L d_K(\mu, \nu) \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \delta. \tag{15}$$

We argue that $b(\nu) \leq b$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta}$ and some $\tilde{\delta} > 0$. This is obvious from (15) if $b(\mu) < b$. If $b(\mu) = b$, then we refer to some $\hat{y} \in Y_V$ with $F_\mu(\hat{y}) > p$ (see proof of Prop. 12). Consequently, $a = b = \hat{y}$ and $F_\mu(b) > p$. Then, $F_\nu(b) \geq p$ and, hence, $b(\nu) \leq b$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta} := F_\mu(b) - p$.

Now, π is a lower semicontinuous, convex and finite function on the nonempty intervals $[b(\nu), b] \subseteq Y_V$ (see Prop. 10). In particular, $Y(\nu) \neq \emptyset$ for all $\nu \in \mathcal{P}(\mathbb{R})$ with $d_K(\mu, \nu) < \tilde{\delta}$. Elementary calculus shows that

$$d_H(Y(\nu), Y(\mu)) \leq |b(\nu) - b(\mu)| \quad \forall \nu \in \mathcal{P}(\mathbb{R}), d_K(\mu, \nu) < \tilde{\delta}.$$

Along with (15), this yields the assertion of the Lemma. □

Proof of Proposition 4. Proposition 13 yields the Hausdorff Lipschitz continuity of Y at μ. This means that the Hölder rate equals 1, so Proposition 11 provides the stated result. □

Proposition 14. *Let μ have an s-dimensional normal distribution with independent components. Then, the logarithm of the distribution function F_μ of μ is strongly concave on any bounded, convex subset $C \subseteq \mathbb{R}^s$. As a consequence, for any $r < 0$, F_μ^r is strongly convex on C.*

Proof. We assume first the case of a 1-dimensional standard normal distribution having the distribution function Φ and the density $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. With $\Psi := \log\Phi$, one has that $\Psi' = \varphi/\Phi$ and

$$\Psi'' = \frac{-\varphi(x)\theta(x)}{\Phi^2(x)} \quad \text{with} \quad \theta(x) = \varphi(x) + x\Phi(x). \tag{16}$$

We argue that $\theta(x) > 0$ for all x. Evidently, this is true for $x > 0$, and, for $x < 0$, one gets

$$\begin{aligned}\theta(x) &= \varphi(x) + x\int_{-\infty}^{x}\varphi(\xi)d\xi = \varphi(x) + \int_{-\infty}^{x}\xi\varphi(\xi)d\xi + \int_{-\infty}^{x}(x-\xi)\varphi(\xi)d\xi \\ &= \int_{-\infty}^{x}(x-\xi)\varphi(\xi)d\xi \geq \int_{-\infty}^{2x}(-x)\varphi(\xi)d\xi = -x\Phi(2x) > 0.\end{aligned}$$

From (16), it follows that $\Psi''(x) < 0$ for all x. Now, let F_μ be the distribution function of a 1-dimensional normal distribution with mean m and variance σ^2. Then, for

$$\tilde{\Psi} := \log F_\mu = \log\Phi(\sigma^{-1}(x-m)) = \Psi(\sigma^{-1}(x-m))$$

one has that $\tilde{\Psi}''(x) = \sigma^{-2}\Psi''(x) < 0$ for all x. Therefore, on each compact interval $I \subseteq \mathbb{R}$, $\tilde{\Psi}''$ is bounded above by some negative constant, which implies that $\log F_\mu$ is strongly concave on I with some modulus $\kappa(I) > 0$: for all $x, y \in I$ and all $\lambda \in [0, 1]$, one has

$$\log F_\mu(\lambda x + (1-\lambda)y) \geq \lambda\log F_\mu(x) + (1-\lambda)\log F_\mu(y) + \kappa(I)\lambda(1-\lambda)(x-y)^2.$$

Finally, let F_μ be the distribution function of an s-dimensional normal distribution with independent components. Assume that m is the mean vector and Σ is the diagonal covariance matrix of this distribution with nonzero variances σ_i^2. Then,

$$F_\mu(x) = F_{\mu_1}(x_1)\cdots F_{\mu_s}(x_s) \quad \forall x \in \mathbb{R}^s,$$

where the F_{μ_i} are the distribution functions of 1-dimensional normal distributions with mean m_i and variance σ_i^2. Let $C \subseteq \mathbb{R}^s$ be any bounded convex subset and choose $I \subseteq \mathbb{R}$ large enough that $C \subseteq I^s$. It follows that for all $x, y \in C$ and all $\lambda \in [0, 1]$

$$\begin{aligned}\log F_\mu(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{s}\log F_{\mu_i}(\lambda x_i + (1-\lambda)y_i) \\ &\geq \sum_{i=1}^{s}(\lambda\log F_{\mu_i}(x_i) + (1-\lambda)\log F_{\mu_i}(y_i) \\ &\quad + \kappa(I)\lambda(1-\lambda)(x_i-y_i)^2) \\ &= \lambda\log F_\mu(x) + (1-\lambda)\log F_\mu(y) + \kappa(I)\lambda(1-\lambda)\|x-y\|^2,\end{aligned}$$

which is the asserted strong concavity of $\log F_\mu$ on C. The last statement of the proposition follows from the fact that continuity and strong concavity of $\log F_\mu$ on some compact convex subset of $\{x|F_\mu(x) > 0\}$ implies F_μ^r to be strongly convex on that same subset for each $r < 0$ (see [6], Prop. 4). □

Example 15. Let $m = 2$, $s = 2$ and

$$\min\{x_2 - x_1|(x_1, x_2) \in X,\ \mathbb{P}(\xi_1 \le x_1, \xi_2 \le x_2) \ge 1/4\},\ \ X = \{(x_1, x_2)|x_1 + x_2 \le 3/2\},$$

where $\xi = (\xi_1, \xi_2)$ is assumed to have a uniform distribution μ over the triangle

$$\text{conv}\{(1, 0), (0, 1), (1, 1)\}.$$

Clearly, our basic convexity assumptions (BCA) are satisfied. The distribution function of ξ is easily calculated as

$$F_\mu(x_1, x_2) = \begin{cases} \min\{1, \min\{x_1^2, x_2^2, (x_1 + x_2 - 1)^2\}\} & \text{if } x_1 \ge 0,\ x_2 \ge 0,\ x_1 + x_2 \ge 1 \\ 0 & \text{else} \end{cases}.$$

Accordingly, the constraint set becomes

$$\begin{aligned}\{x \in X|F_\mu(x) \ge 1/4\} &= \{(x_1, x_2)|x_1 + x_2 \le 3/2,\ x_1 \ge 1/2,\ x_2 \ge 1/2,\ x_1 + x_2 \ge 3/2\} \\ &= \{(x_1, x_2)|x_1 + x_2 = 3/2,\ \ x_1 \in [1/2, 1]\},\end{aligned}$$

and the (unique) solution of this problem is $(1, 1/2)$. Now, consider a sequence of perturbed measures ν_n defined by uniform distributions over the (shifted) triangles

$$\text{conv}\{(1 + n^{-1}, n^{-1}), (n^{-1}, 1 + n^{-1}), (1 + n^{-1}, 1 + n^{-1})\}.$$

Then, F_{ν_n} calculates much like F_μ but with the shifted arguments $x_1 - n^{-1}, x_2 - n^{-1}$. It follows that $\nu_n \to \mu$ in the sense of Kolmogorov distance. However, the feasible set becomes empty:

$$\{(x_1, x_2)|x_1 + x_2 \le 3/2,\ x_1 \ge 1/2 + n^{-1},\ x_2 \ge 1/2 + n^{-1},\ x_1 + x_2 \ge 3/2 + n^{-1}\} = \emptyset.$$

Hence, there are no solutions at all for this special sequence of approxiamting problems. Finally, consider a different sequence of perturbed measures. To this aim, let ν be the uniform distribution over the square $[1/2, 1]^2$ and define the sequence $\nu_n := (1 - n^{-1})\mu + n^{-1}\nu$ of probability measures. The induced distribution functions calculate as $F_{\nu_n} = (1 - n^{-1})F_\mu + n^{-1}F_\nu$, where F_μ has the explicit representation given above and

$$F_\nu(x_1, x_2) = 4 * (\max\{\min\{x_1, 1\}, 1/2\} - 1/2) * (\max\{\min\{x_2, 1\}, 1/2\} - 1/2).$$

We claim that $(3/4, 3/4)$ is the only feasible point in the perturbed constraint set

$$M_n = \{x \in X|F_{\nu_n}(x) \ge 1/4\}.$$

Indeed, one easily checks that $F_\mu(x), F_\nu(x) \le 1/4$ for all $x \in X$. Since F_{ν_n} is a strict convex combination of F_μ and F_ν, any point x of M_n must satisfy $x \in X$ and $F_\mu(x) = F_\nu(x) = 1/4$. However, the only $x \in X$ with $F_\nu(x) = 1/4$ is evidently $\bar{x} = (3/4, 3/4)$, which at the same time fulfills $F_\mu(\bar{x}) = 1/4$. As the only feasible point, $\bar{x}$ trivially coincides with the solution set Ψ_n of all the approximating problems. Consequently, the approximating solutions do not converge towards the solution $(1, 1/2)$ of the original problem.

Example 16. In problem (P), let $m = s = 2$, $X = \mathbb{R}^2$, $h(x) = x$, $g(x) = (x_1 + x_2 - \sqrt{2})^2$, $p = 0.5$ and $\mu =$ uniform distribution over the unit square $[0, 1]^2$. Evidently, these data satisfy all the basic assumptions formulated in the introduction (in particular, μ is log-concave, hence r-concave for any $r < 0$). Next we verify the assumptions of Theorem 2: since the distribution function of μ satisfies $F_\mu(x) = x_1x_2$ for all $(x_1, x_2) \in [0, 1]^2$, it follows that $\Psi(\mu) = \{(\sqrt{1/2}, \sqrt{1/2})\}$ which entails 1. in Theorem 2. It is elementary to verify that one may assume $Y(\mu) \subseteq [0, 1]^2$ (after shrinking the open ball $V \supseteq \Psi(\mu)$ used in Prop. 10). Evidently, F_μ^r is strongly convex for any $r < 0$, whence 3. With $\hat{x} := (1, 1)$, one has $F_\mu(\hat{x}) = 1 > p$, which is 2. Finally, since g is convex-quadratic, X is trivially a polyhedral set and h is linear, it follows that σ is Hausdorff Lipschitz continuous (see remarks above Corollary 3). This provides 4. with $\kappa = 1$, and, thus, Theorem 2 ensures that Ψ is Hausdorff Hölder continuous with rate $1/2$ at μ. This rate is sharp. Indeed, considering the perturbed measures $\nu_\varepsilon \in \mathcal{P}(\mathbb{R}^s)$ defined for $\varepsilon > 0$ as uniform distributions over the squares $[-\varepsilon, 1 - \varepsilon]^2$, a straightforward calculation shows that

$$\Psi(\nu_\varepsilon) = \operatorname{conv}\{(a_\varepsilon, b_\varepsilon), (b_\varepsilon, a_\varepsilon)\} \quad \text{and} \quad d_K(\mu, \nu_\varepsilon) = \varepsilon(1 + \varepsilon),$$

where $a_\varepsilon/b_\varepsilon = \sqrt{1/2} \pm \sqrt{\varepsilon(\sqrt{2} + \varepsilon)}$. Consequently,

$$d_H(\Psi(\mu), \Psi(\nu_\varepsilon)) = \sqrt{2}\sqrt{\varepsilon(\sqrt{2} + \varepsilon)} \geq \sqrt{\varepsilon(1 + \varepsilon)} = \sqrt{d_K(\mu, \nu_\varepsilon)},$$

which shows that the Hölder rate $1/2$ cannot be improved in this example.

Example 17. In the previous example, we fix an $\varepsilon > 0$ and consider the associated measure ν_ε which is a uniform distribution over the square $[-\varepsilon, 1 - \varepsilon]^2$. Certainly, ν_ε can be approximated by a discrete measure $\tilde{\nu}_\varepsilon$ (by placing an increasing number of uniformly distributed atoms in the square). In particular, $\tilde{\nu}_\varepsilon$ may be chosen such that $d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon) \leq d_K(\mu, \nu_\varepsilon)/3$. It follows that, due to $d_K(\mu, \nu_\varepsilon) = \varepsilon(1 + \varepsilon) \to 0$ (for $\varepsilon \to 0$), one also has that $d_K(\mu, \tilde{\nu}_\varepsilon)$, i.e., $\tilde{\nu}_\varepsilon$ approximates μ for $\varepsilon \to 0$. Applying the triangle inequality yields that

$$d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon) \leq d_K(\mu, \tilde{\nu}_\varepsilon)/3 + d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon)/3,$$

whence $d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon) \leq d_K(\mu, \tilde{\nu}_\varepsilon)/2$. Furthermore, one easily checks from the data in Example 16 that all the perturbed problems

$$(P_\varepsilon) \quad \min\{g(x) | F_{\nu_\varepsilon}(x) \geq 0.5\}$$

continue to satisfy the assumptions of Corollary 3: the basic data assumptions remain valid (ν_ε is a uniform distribution over a square similar to μ), the point $\hat{x} := (1, 1)$ satisfies $F_{\nu_\varepsilon}(\hat{x}) = 1 > p$ and the solution set $\Psi(\nu_\varepsilon)$ is nonempty and bounded. Finally, similar to F_μ^r, it holds that $F_{\nu_\varepsilon}^r$ is strongly convex for any $r < 0$. Now, Corollary 3 applies to ν_ε as the original measure and $\tilde{\nu}_\varepsilon$ as the perturbed measure. Since $\tilde{\nu}_\varepsilon$ may be choosen arbitrarily close to ν_ε, we may assume that

$$d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon) \leq (2L)^{-2} d_H^2(\Psi(\mu), \Psi(\nu_\varepsilon)),$$

where L refers to the constant from Corollary 3. Now, the corollary provides

$$d_H(\Psi(\tilde{\nu}_\varepsilon), \Psi(\nu_\varepsilon)) \leq L\sqrt{d_K(\tilde{\nu}_\varepsilon, \nu_\varepsilon)} \leq d_H(\Psi(\mu), \Psi(\nu_\varepsilon))/2.$$

Summarizing, one may invoke the estimation from Example 16 and exploit the triangle inequality for the Hausdorff distance to arrive at

$$\begin{aligned} d_H(\Psi(\mu), \Psi(\tilde{\nu}_\varepsilon)) &\geq d_H(\Psi(\mu), \Psi(\nu_\varepsilon)) - d_H(\Psi(\tilde{\nu}_\varepsilon), \Psi(\nu_\varepsilon)) \\ &\geq d_H(\Psi(\mu), \Psi(\nu_\varepsilon))/2 \geq \sqrt{d_K(\mu, \nu_\varepsilon)}/2 \\ &\geq \sqrt{d_K(\mu, \tilde{\nu}_\varepsilon) - d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon)}/2 \geq \sqrt{d_K(\mu, \tilde{\nu}_\varepsilon)/2}/2 \\ &= \sqrt{d_K(\mu, \tilde{\nu}_\varepsilon)}/(2\sqrt{2}). \end{aligned}$$

From Example 16 we know that $d_K(\mu, \nu_\varepsilon) = \varepsilon(1+\varepsilon) \to 0$ (for $\varepsilon \to 0$). Since $d_K(\nu_\varepsilon, \tilde{\nu}_\varepsilon) \leq d_K(\mu, \nu_\varepsilon)/3$, it follows that $\tilde{\nu}_\varepsilon$ is a discrete approximation of μ. Now, the above chain of inequalities confirms, that discrete approximations may reslut in a Hölder rate $1/2$ for stability of solutions of program (P) under the assumptions of Corollary 3.

Acknowledgement. This research was supported by the DFG Research Center "Mathematics for key technologies: modelling, simulation and optimization of real-world processes".

References

[1] Bonnans, J.F., Shapiro, A.: Nondegeneracy and quantitative stability of parametrized optimization problems with multiple solutions. SIAM J. Optim. **8**, 940–946 (1998)
[2] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York, 2000
[3] Borell, C.: Convex Sets in d-Space. Periodica Mathematica Hungarica **6**, 111–136 (1975)
[4] Gröwe, N.: Estimated stochastic programs with chance constraints. Eur. J. Oper. Res. **101**, 285–305 (1997)
[5] Henrion, R.: Qualitative stability of convex programs with probabilistic constraints. In: (V.H. Nguyen, J.-J. Strodiot and P. Tossings eds.) Optimization, Lecture Notes in Economics and Mathematical Systems, Vol. **481**, Springer, Berlin 2000, pp. 164–180
[6] Henrion, R., Römisch, W.: Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. Math. Program. **84**, 55–88 (1999)
[7] Henrion, R., Römisch, W.: Stability of solutions to chance constrained stochastic programs. In: (J. Guddat, R. Hirabayashi, H.Th. Jongen and F. Twilt eds.) Parametric Optimization and Related Topics V, Peter Lang, Frankfurt a.M. 2000, pp. 95–114
[8] Henrion, R.: Perturbation Analysis of Chance-Constrained Programs under variation of all constraint data, in K. Marti et al. (eds.): Dynamic Stochastic Optimization, Lecture Notes in Economics and Mathematical Systems, Vol. **532**, Springer, Heidelberg 2004, pp. 257–274.
[9] Kaňková, V.: A note on multifunctions in stochastic programming. In: Stochastic Programming Methods and Technical Applications (K. Marti and P. Kall eds.), Lecture Notes in Economics and Mathematical Systems Vol. **458**, Springer, Berlin 1998, pp. 154–168
[10] Klatte, D., Thiere, G.: Error bounds for solutions of linear equations and inequalities. ZOR – Math. Meth. Oper. Res. **41**, 191–214 (1995)
[11] Lepp, R.: Discrete approximation of extremum problems with chance constraints. In: Stochastic Optimization Techniques (K. Marti ed.), Lecture Notes in Economics and Mathematical Systems Vol. **513**, Springer, Berlin 2000, pp. 21–33
[12] Prekopá, A.: Stochastic Programming. Kluwer, Dordrecht, 1995
[13] Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, 1970
[14] Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin, 1997
[15] Römisch, W., Schultz, R.: Stability analysis for stochastic programs. Ann. Oper. Res. **30**, 241–266 (1991)
[16] van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press, 1998

[17] Wang, J.: Continuity of feasible solution sets of probabilistic constrained programs. J. Optim. Theory Appl. **63**, 79–89 (1989)
[18] Wets, R.J-B.: Stochastic programs with chance constraints: Generalized convexity and approximation issues. In: Generalized Convexity, Generalized Monotonicity: Recent Results (J.-P. Crouzeix, J.-E. Martínez-Legaz and M. Volle eds.), Kluwer, Dordrecht 1998, pp. 61–74