A critical note on empirical (sample average, Monte Carlo) approximation of solutions to chance constrained programs

René Henrion[⋆]

Weierstrass Institute Berlin, Mohrenstr. 39, 10117 Berlin, Germany

Abstract. The solution of chance constrained optimization problems by means of empirical approximation of the underlying multivariate distribution has recently become a popular alternative to conventional methods due to the efficient application of appropriate mixed integer programming techniques. As the complexity of required computations depends on the sample size used for approximation, exponential estimates for the precision of optimal solutions or optimal values have become a key argument for controlling the sample size. However, these exponential estimates may involve unknown constants such that the required sample size to approximate the solution of a problem may become arbitrarily large. We will illustrate this effect for Gaussian distributions.

Keywords: chance constrained programming, probabilistic constraints, empirical approximation, sample average approximation, convergence of solution sets, sample size

1 Introduction

A chance constrained optimization problem has the general form

$$\min\{g(x)|\mathbb{P}(h(x,\xi) \geq 0) \geq p, x \in C\}, \tag{1}$$

where $x \in \mathbb{R}^n$ is a decision vector, $g : \mathbb{R}^n \to \mathbb{R}$ is an objective function, ξ is an s-dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ is a Borel measurable with respect to the second argument mapping, $C \subseteq \mathbb{R}^n$ represents some abstract deterministic constraint and $p \in [0,1]$ is a fixed probability level. The random inequality system $h(x,\xi) \geq 0$ may reflect some technological constraints in engineering problems which are affected by uncertainty. Since usually a decision x has to be taken before the uncertain parameter ξ is observed, it has become a standard approach of robust modeling to define x as feasible, whenever the probability of satisfying the random inequality is at least p. This is expressed in the so-called chance constraint $\mathbb{P}(h(x,\xi) \geq 0) \geq p$. For a standard introduction to chance constrained programming we refer to the classical monograph [6] and to the more recent treatise in [7].

Recent progress in mixed integer programming techniques tailored to chance constraints has led to the idea of solving (1) by empirical approximation (or sample average approximation) of the original random vector ξ (e.g., [4, 5]). This means that an i.i.d. sample $\xi^1, \ldots, \xi^N$ of size N is drawn from the distribution of ξ and that the law $\mathbb{P} \circ \xi^{-1}$ of ξ in (1) is replaced by the empirical measure

$$N^{-1} \sum_{i=1}^{N} \delta(\xi^i),$$

where $\delta(z)$ is the Dirac measure centered on z. Doing so, the original chance constraint $\mathbb{P}(h(x, \xi) \geq 0) \geq p$ turns into its empirical counterpart

$$\# \left\{ i | h(x, \xi^i) \geq 0 \right\} \geq pN.$$

Of course this change of constraint leads to another optimization problem whose solution deviates from the solution of (1) and one has to answer two questions: do the approximating solutions converge with $N \to \infty$ and if so, how large should N be chosen in order to guarantee a given precision of the solution obtained? The first question is answered by the stability theory of chance constrained programming mainly developed in [1, 2] which applies to arbitrary approximations of the original distribution and, in particular, to empirical ones. Under some explicitly verifiable conditions, not only qualitative convergence of approximating solutions can be guaranteed but also rates for this convergence can be derived. The latter allow us to obtain exponential bounds (in terms of sample size) for the precision of solutions in case of empirical approximations. However, one has to take into account that the exponential term involves apart from the sample size N also some other constants which may depend on the conditioning of the problem and may be hard to estimate. Thus, exponential estimates of solutions do not exclude the need for a large sample size even in small dimension in order to arrive at a reasonable precision of the solution. This situation occurs in particuar if the law of the original random vector ξ has unbounded support. We will illustrate and explain this effect for a multivariate Gaussian distribution (but similar observations could be made for other classes of multivariate distributions such as log-normal or t-). In order to keep the presentation as simple as possible we restrict ourselves to the simplest yet meaningful instance of problem (1):

$$\min\{c^T x \mid \mathbb{P}(\xi \leq x) \geq p\}.$$

This means that we consider just linear objective functions, we forget about additional abstract deterministic constraints and we assume the chance constraint being in elementary separated form. Recalling the definition of the distribution function $F_\xi(x) := \mathbb{P}(\xi \leq x)$ of a random vector ξ, we may rewrite this problem as

$$\min\{c^T x \mid F_\xi(x) \geq p\} \qquad (\mathrm{P}_{\mathrm{c},\xi,\mathrm{p}}). \tag{2}$$

In order to emphasize the dependence on the problem data c, ξ and p, we label problem (2) as $(\mathrm{P}_{\mathrm{c},\xi,\mathrm{p}})$. Before coming back to the issue of empirical distributions

discussed above, we derive in the next section our main result demonstrating the difficulty of approximating a chance constrained program with unbounded support of the underlying distribution by means of distributions with bounded support. Note that, in particular, empirical measures have bounded support.

2   Main Result

We start by recalling the following well-known representation for partial derivatives of Gaussian distribution functions. We make use of the familiar notation $\xi \sim \mathcal{N}(\mu, \Sigma)$ to designate a Gaussian random vector with expectation μ and covariance matrix Σ.

Theorem 1 ([6], p. 204). *Let $\xi \sim \mathcal{N}(\mu, \Sigma)$ with some positive definite covariance matrix $\Sigma = (\sigma_{ij})$ of order (s, s). Then, the distribution function F_ξ is continuously differentiable at any $z \in \mathbb{R}^s$ and*

$$\frac{\partial F_\xi}{\partial z_j}(z) = f_{\xi_j}(z_j) \cdot F_{\tilde{\xi}(z_j)}(z_1, \ldots, z_{j-1}, z_{j+1} \ldots, z_s) \quad (j = 1, \ldots, s).$$

Here, f_{ξ_j} denotes the one-dimensional Gaussian density of the component ξ_j, $\tilde{\xi}(z_j)$ is an (s-1)-dimensional Gaussian random vector distributed according to $\tilde{\xi}(z_j) \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$, $\hat{\mu}$ results from the vector $\mu + \sigma_{jj}^{-1}(z_j - \mu_j)\sigma_j$ by deleting component j and $\hat{\Sigma}$ results from the matrix $\Sigma - \sigma_{jj}^{-1}\sigma_j\sigma_j^T$ by deleting row j and column j, where σ_j refers to column j of Σ.

Corollary 1. *In the context of the previous Theorem, one has that*

$$\frac{\partial F_\xi}{\partial z_j}(z) > 0 \quad \forall z \in \mathbb{R}^s, \ \forall j \in \{1, \ldots, s\}.$$

Proof. This follows immediately from the formula in Theorem 1 and the fact that both the density and the distribution function of a regular Gaussian distribution are strictly positive.

With each problem $(P_{c,\xi,p})$ in (2) we associate its (possibly empty) solution set

$$\Psi_{c,\xi,p} := \arg\min\{c^T x \mid \mathbb{P}(\xi \le x) \ge p\}.$$

Lemma 1. *For problem $(P_{c,\xi,p})$ in (2) assume that $c_i > 0$ for $i = 1, \ldots, s$. Then,*

$$\Psi_{c,\xi,p} \subseteq [a, b] := \{x \in \mathbb{R}^s \mid a_i \le x_i \le b_i \quad (i = 1, \ldots, s)\},$$

where, with 'supp' denoting the support of a random vector,

$$a_i := \inf\{z_i | z \in \text{supp}\,\xi\} \quad b_i := \sup\{z_i | z \in \text{supp}\,\xi\} \quad (i = 1, \ldots, s).$$

Proof. Assume that there is some $x^* \in \Psi_{c,\xi,p}$ and some i such that $x_i^* > b_i$. Then, $b_i < \infty$ and we may define $\bar{x}$ by

$$\bar{x}_i := (x_i^* + b_i)/2, \quad \bar{x}_j := x_j^* \quad (j \neq i).$$

From x^* being feasible for problem ($P_{c,\xi,p}$), we conclude that

$$p \leq \mathbb{P}(\xi \leq x^*) = \mathbb{P}(\xi \leq \bar{x}) + \mathbb{P}(\bar{x}_i \leq \xi_i \leq x_i^*, \ \xi_j \leq \bar{x}_j \quad (j \neq i)).$$

Now, since $\bar{x}_i > b_i$, it follows that $\{x \in \mathbb{R}^n \mid \bar{x}_i \leq x_i \leq x_i^*\} \cap \operatorname{supp} \xi = \emptyset$, whence

$$\mathbb{P}(\bar{x}_i \leq \xi_i \leq x_i^*, \ \xi_j \leq \bar{x}_j \quad (j \neq i)) = 0$$

and $\mathbb{P}(\xi \leq \bar{x}) \geq p$. Therefore, $\bar{x}$ too is feasible for problem ($P_{c,\xi,p}$). On the other hand, $c^T \bar{x} < c^T x^*$ due to $c_i > 0$, $\bar{x}_i < x_i^*$ and $\bar{x}_j = x_j^*$ for $j \neq i$. This contradicts the assumption $x^* \in \Psi_{c,\xi,p}$. Consequently, $x^* \leq b$ for any $x^* \in \Psi_{c,\xi,p}$. Similarly one shows that $x^* \geq a$ for any $x^* \in \Psi_{c,\xi,p}$. It follows that $\Psi_{c,\xi,p} \subseteq [a,b]$.

Now, we are in a position to state our main result:

Theorem 2. *Let $s > 1$. Assume that ξ has a regular normal distribution according to $\xi \sim \mathcal{N}(\mu, \Sigma)$ and that η is a random vector with compact support. Then, for any $p \in (0,1)$ there exists a sequence $c^{(n)} \in \mathbb{R}^n$ with $c_i^{(n)} > 0$ for $i = 1, \ldots, s$ such that $\Psi_{c^{(n)},\xi,p} \neq \emptyset$ and*

$$\inf\{\|x - y\| \mid x \in \Psi_{c^{(n)},\xi,p}, y \in \Psi_{c^{(n)},\eta,p}\} > n \quad \forall n \in \mathbb{N}. \tag{3}$$

Proof. Fix an arbitrary $p \in (0,1)$ and an arbitrary $n \in \mathbb{N}$. Since $\operatorname{supp} \eta$ is compact, we may apply Lemma 1 to η in order to derive the existence of some compact(!) rectangle $[a,b]$ such that

$$\Psi_{c,\eta,p} \subseteq [a,b] \quad \forall c \in \mathbb{R}^s : c_i > 0 \quad (i = 1, \ldots, s). \tag{4}$$

With $[a,b]$ being compact, we may choose $L_n > 0$ such that

$$\|y - z\| \geq n \quad \forall y \in [a,b], \ \forall z : \|z\| \geq L_n. \tag{5}$$

Since $s > 1$ by assumption, we may define the ratio

$$\varkappa(z) := \frac{\partial F_\xi}{\partial z_1}(z) \Big/ \frac{\partial F_\xi}{\partial z_2}(z) \quad (z \in \mathbb{R}^s). \tag{6}$$

Note that the partial derivatives of F_ξ are continuous (see Theorem 1) and strictly positive (see Corollary 1), hence $\varkappa$ is correctly defined and continuous. Consequently the quantity

$$\bar{\varkappa} := \sup\{\varkappa(z) \mid z \in \mathbb{B}(0, L_n)\} \tag{7}$$

is finite. Next, let q_p be the p-quantile of the first marginal distribution of ξ, i.e., of the distribution of the first component ξ_1. Since ξ_1 has a one-dimensional normal distribution and $p \in (0, 1)$, $q_p \in \mathbb{R}$ is uniquely defined by $\mathbb{P}(\xi_1 \leq q_p) = p$. We claim that for each $k \in \mathbb{N}$ there exists some $t_k \in \mathbb{R}$ such that

$$F_\xi\left(q_p + k^{-1}, t_k, \ldots, t_k\right) = p. \tag{8}$$

Indeed, for arbitrarily fixed k one has that

$$\lim_{\tau \to -\infty} F_\xi\left(q_p + k^{-1}, \tau, \ldots, \tau\right) = 0$$

as a general property of distribution functions and that

$$p = \mathbb{P}(\xi_1 \leq q_p) < \mathbb{P}\left(\xi_1 \leq q_p + k^{-1}\right) = \lim_{\tau \to \infty} F_\xi\left(q_p + k^{-1}, \tau, \ldots, \tau\right).$$

Now, the existence of t_k with the desired property (8) follows from continuity of F_ξ and from $p > 0$. Next we claim that $t_k \to_k \infty$. If there existed some subsequence t_{k_l} and some $r \in \mathbb{R}$ such that $t_{k_l} \leq r$ for all $l \in \mathbb{N}$, then

$$F_\xi\left(q_p + k_l^{-1}, r, \ldots, r\right) \geq F_\xi\left(q_p + k_l^{-1}, t_{k_l}, \ldots, t_{k_l}\right) = p \quad \forall l \in \mathbb{N}$$

which again by continuity of F_ξ as a function of each of its components yields the contradiction

$$p \leq F_\xi\left(q_p, r, \ldots, r\right) = \mathbb{P}(\xi_1 \leq q_p, \xi_2 \leq r, \ldots, \xi_s \leq r) < \mathbb{P}(\xi_1 \leq q_p) = p.$$

Here, the strict inequality relies on the fact that a regular Gaussian distribution has a density which is strictly positive everywhere. Hence, we have shown that $t_k \to_k \infty$. Consider the sequence

$$z^{(k)} := \left(q_p + k^{-1}, t_k, \ldots, t_k\right) \quad (k \in \mathbb{N}).$$

Then, by (8),

$$F_\xi\left(z^{(k)}\right) = p \quad (k \in \mathbb{N}). \tag{9}$$

Moreover, Theorem 1 yields that

$$\frac{\partial F_\xi}{\partial z_1}(z^{(k)}) = f_{\xi_1}(q_p + k^{-1}) \cdot F_{\tilde{\xi}\left(z_1^{(k)}\right)}(t_k, \ldots, t_k), \tag{10}$$

where f_{ξ_1} denotes the one-dimensional Gaussian density of the component ξ_1 and $\tilde{\xi}\left(z_1^{(k)}\right)$ is an (s-1)-dimensional Gaussian random vector distributed according to $\tilde{\xi}\left(z_1^{(k)}\right) \sim \mathcal{N}(\hat{\mu}^{(k)}, \hat{\Sigma})$ where $\hat{\mu}^{(k)}$ and $\hat{\Sigma}$ result from the original parameters μ and Σ, respectively, of ξ as detailed in Theorem 1. In particular,

$$\hat{\mu}^{(k)} = (\mu_2, \ldots, \mu_s) + \sigma_{11}^{-1}\left(q_p + k^{-1} - \mu_1\right)(\sigma_{21}, \ldots, \sigma_{s1})$$

and we observe that

$$\hat{\mu}^{(k)} \to_k \hat{\mu} := (\mu_2, \ldots, \mu_s) + \sigma_{11}^{-1} (q_p - \mu_1) (\sigma_{21}, \ldots, \sigma_{s1}). \tag{11}$$

Note also that in contrast to $\hat{\mu}^{(k)}$, the covariance matrix $\hat{\Sigma}$ does not depend on the index k. Now we define the centered random vector

$$\hat{\xi} := \tilde{\xi}\left(z_1^{(k)}\right) - \hat{\mu}^{(k)} \sim \mathcal{N}(0, \hat{\Sigma}),$$

whose distribution does no longer depend on the index k. Exploiting the relation

$$F_{\tilde{\xi}\left(z_1^{(k)}\right)}(t_k, \ldots, t_k) = F_{\hat{\xi}}\left(t_k - \hat{\mu}_1^{(k)}, \ldots, t_k - \hat{\mu}_{s-1}^{(k)}\right)$$

and noting that all components of the argument $\left(t_k - \hat{\mu}_1^{(k)}, \ldots, t_k - \hat{\mu}_{s-1}^{(k)}\right)$ tend to infinity due to (11) and $t_k \to_k \infty$, we conclude that

$$F_{\hat{\xi}}\left(t_k - \hat{\mu}_1^{(k)}, \ldots, t_k - \hat{\mu}_{s-1}^{(k)}\right) \to_k 1$$

because the values of the (fixed) distribution function $F_{\hat{\xi}}$ tend to one if all its components tend to infinity. This implies

$$F_{\tilde{\xi}\left(z_1^{(k)}\right)}(t_k, \ldots, t_k) \to_k 1,$$

whence (10) leads to

$$\frac{\partial F_\xi}{\partial z_1}(z^{(k)}) \to_k f_{\xi_1}(q_p) > 0 \tag{12}$$

by continuity and positivity of the density f_{ξ_1}. Similarly, the second partial derivative of F_ξ calculates from Theorem 1 as

$$\frac{\partial F_\xi}{\partial z_2}(z^{(k)}) = f_{\xi_2}(t_k) \cdot F_{\tilde{\xi}\left(z_2^{(k)}\right)}\left(q_p + k^{-1}, t_k, \ldots, t_k\right), \tag{13}$$

where f_{ξ_2} denotes the one-dimensional Gaussian density of the component ξ_2 and $\tilde{\xi}\left(z_2^{(k)}\right)$ is a certain (s-1)-dimensional Gaussian random vector. From $f_{\xi_2}(t_k) \to_k 0$ (due to $t_k \to_k \infty$) and from the fact that distribution functions are bounded between zero and one, we infer that

$$\frac{\partial F_\xi}{\partial z_2}(z^{(k)}) \to_k 0,$$

which along with (12) and (6) provides that $\varkappa(z^{(k)}) \to_k \infty$. Therefore, with our arbitrarily fixed number $n \in \mathbb{N}$ we may associate an index $k_n \in \mathbb{N}$ such that $\varkappa(z^{(k_n)}) > \bar{\varkappa}$ where $\bar{\varkappa}$ is defined in (7). Now, we assign to n the cost vector $c^{(n)} := \nabla F_\xi\left(z^{(k_n)}\right)$ for the linear objective function in problem $(P_{c^{(n)}, \xi, p})$ in (2). Then, by Corollary 1, we have that $c_i^{(n)} > 0$ for $i = 1, \ldots, s$ as required in the

statement of our theorem. Knowing that $\log F_\xi$ is a concave function (see [6]), the problem $(\mathrm{P}_{c^{(n)},\xi,\mathrm{p}})$ in (2) may be written equivalently as a convex optimization problem

$$\min\{c^{(n)T}x \mid -\log F_\xi(x) \leq -\log p\}. \qquad (\mathrm{P}_{c^{(n)},\xi,\mathrm{p}}) \qquad (14)$$

With $c^{(n)} \neq 0$, a solution x^* to this problem is equivalently characterized by the conditions

$$-\log F_\xi(x^*) = -\log p \quad \text{and} \quad c^{(n)} + \lambda \nabla\left(-\log F_\xi\right)(x^*) = 0 \quad \text{for some} \quad \lambda > 0.$$

Simplifying these leads to the equivalent conditions

$$F_\xi(x^*) = p \quad \text{and} \quad c^{(n)} = \lambda \nabla\left(F_\xi\right)(x^*) \quad \text{for some} \quad \lambda > 0. \qquad (15)$$

Now, since $c^{(n)} = \nabla F_\xi\left(z^{(k_n)}\right)$ and $F_\xi(z^{(k_n)}) = p$ by (9), we conclude that $z^{(k_n)}$ is a solution to $(\mathrm{P}_{c^{(n)},\xi,\mathrm{p}})$. This shows that $\Psi_{c^{(n)},\xi,p} \neq \emptyset$ as asserted in our theorem. Finally, we show that $\Psi_{c^{(n)},\xi,p} \cap \mathbb{B}\left(0, L_n\right) = \emptyset$ with L_n defined in (5). Assume the contrary and choose some $x^* \in \Psi_{c^{(n)},\xi,p}$ with $\|x^*\| \leq L_n$. From (6) and (15) we derive

$$\varkappa\left(x^*\right) = \frac{\partial F_\xi}{\partial z_1}(x^*) \bigg/ \frac{\partial F_\xi}{\partial z_2}(x^*) = c_1^{(n)} \bigg/ c_2^{(n)} = \frac{\partial F_\xi}{\partial z_1}(z^{(k_n)}) \bigg/ \frac{\partial F_\xi}{\partial z_2}(z^{(k_n)})$$

$$= \varkappa(z^{(k_n)}) > \bar{\varkappa}.$$

which is a contradiction with (7). Consequently, $\Psi_{c^{(n)},\xi,p} \cap \mathbb{B}\left(0, L_n\right) = \emptyset$. Now, select arbitrary $x \in \Psi_{c^{(n)},\xi,p}$ and $y \in \Psi_{c^{(n)},\eta,p}$. Then, $\|x\| > L_n$. Since also $\Psi_{c^{(n)},\eta,p} \subseteq [a, b]$ by (4), it follows from (5) that $\|x - y\| \geq n$. Since x and y were arbitrarily chosen, we end up at the final assertion (3) of our theorem.

Theorem 2 can be interpreted as follows in the context of empirical approximation upon observing that the support of empirical measures is finite, hence compact: no matter how large the sample size N for the empirical approximation of the original random vector ξ is chosen, there is always an instance of problem (2) (by choosing an appropriate cost vector c) such that the solutions between the original problem and its empirical approximation are arbitrarily far from each other. Note that relation (3) implies (and actually is much stronger than) the Hausdorff distance between both solution sets being larger than any prescribed n. Moreover, this effect of ill-conditioning is not caused by letting the probability level tend to one, because the result of the theorem holds true for any fixed p. In the following section we look at the same phenomenon from a slightly different viewpoint.

3   Exponential estimates with ill-conditioned constants

The recent literature on empirical or sample average approximation on chance constraints [4, 5] compiles several convergence results for feasible sets, optimal

values and solutions, most of them of qualitative nature (continuity, upper semi-continuity), some of them providing exponential estimates (convergence of feasible sets, lower bounds for optimal values). For the general stability theory of chance constrained programming with arbitrary approximations (not just empirical ones) and quantitative convergence results even for solution sets, we refer to [1, 2]. In this section we will appy these results to the special case of empirical approximations in order to obtain an exponential bound for the convergence (in the sense of Hausdorff distance!) of solution sets. Despite this positive result we will show then, that the existence of exponential estimates does not exclude the need for a possibly excessive sample size in the empirical approximation. We start by citing the following stability result for chance constraints whch we present here in a simplified form sufficient for our purposes

Theorem 3 ([2], Corollary 3). *In problem* $(\mathrm{P}_{c,\xi,\mathrm{p}})$ *in (2) let* $p \in (0,1)$ *and the following assumptions be satisfied:*

1. $\log F_\xi$ *is a strongly concave function.*
2. $\Psi_{c,\xi,p}$ *is nonempty and compact.*

   Then, there exist $L, \delta > 0$ *such that*

$$d_H \left(\Psi_{c,\xi,p}, \Psi_{c,\eta,p} \right) \leq L \sqrt{\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_\eta(z)|} \quad \forall \eta : \sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_\eta(z)| < \delta.$$

$$(16)$$

Here, d_H *refers to the Hausdorff distance.*

A prototype example for a problem (2) which automatically satisfies all assumptions of Theorem 3 is given by a random vector ξ having a standard Gaussian distribution. As a preparation we show the following property which is of independent interest:

Proposition 1. *In problem* $(\mathrm{P}_{c,\xi,\mathrm{p}})$ *in (2) let* $p \in (0,1)$, $c_i > 0$ *for* $i = 1, \ldots, s$ *and* $\xi \sim \mathcal{N}(0, I_s)$. *Then, the problem has a solution.*

Proof. Referring back to the proof of Theorem 2, a solution of problem $(\mathrm{P}_{c,\xi,\mathrm{p}})$ is equivalently characterized by the conditions (15) applied to c rather than $c^{(n)}$. The distribution assumption implies that for all $z \in \mathbb{R}^s$ and all $i = 1, \ldots, s$,

$$F_\xi(z) = \Phi(z_1) \cdots \Phi(z_s); \quad \frac{\partial F_\xi}{\partial z_i}(z) = f(z_i) \Phi(z_1) \cdots \Phi(z_{i-1}) \Phi(z_{i+1}) \cdots \Phi(z_s),$$

where f and Φ denote the one-dimensional standard normal density and distribution function, respectively. From (15) we derive that x^* is a solution to $(\mathrm{P}_{c,\xi,\mathrm{p}})$ if there exists some $\lambda > 0$ such that

$$\Phi(x_1^*) \cdots \Phi(x_s^*) = p; \quad \lambda \frac{f(x_i^*)}{\Phi(x_i^*)} = c_i \quad (i = 1, \ldots, s). \tag{17}$$

(recall that Φ is strictly positive). Defining $\alpha := f/\Phi$, we observe that due to $c_i > 0$ the second relation of (17) amounts to the fact that $\alpha(x_i^*)/c_i$ is constant for all i. We may assume, without loss of generality, that c_1 is the largest of the coefficients c_i. Then, x^* is a solution to $(\mathrm{P_{c,\xi,p}})$ if there exist coefficients $\rho_1, \ldots, \rho_{s-1} \geq 1$ such that

$$\Phi(x_1^*) \cdots \Phi(x_s^*) = p; \quad \alpha(x_1^*) = \rho_1 \alpha(x_2^*) = \cdots = \rho_{s-1} \alpha(x_s^*). \tag{18}$$

We recall from the properties of the one-dimensional standard normal density and distribution function that $\alpha(t) \to 0$ for $t \to \infty$. Consequently, given any $t \in \mathbb{R}$, there exist values $\beta_i(t)$ for $i = 1, \ldots, s-1$ such that $\beta_i(t) \geq t$ and

$$\alpha(t) = \rho_1 \alpha(\beta_1(t)) = \cdots = \rho_{s-1} \alpha(\beta_{s-1}(t)). \tag{19}$$

Taking into account that $\lim_{t \to -\infty} \Phi(t) = 0$ and $\Phi(\beta_i(t)) \leq 1$ on the one hand and

$$\lim_{t \to \infty} \Phi(t) = \lim_{t \to \infty} \Phi(\beta_i(t)) = 1$$

due to $\beta_i(t) \geq t$, on the other hand, we conclude that

$$\lim_{t \to -\infty(+\infty)} \Phi(t) \Phi(\beta_1(t)) \cdots \Phi(\beta_{s-1}(t)) = 0(1).$$

For continuity reasons, there exists some t^* such that

$$\Phi(t^*) \Phi(\beta_1(t^*)) \cdots \Phi(\beta_{s-1}(t^*)) = p.$$

Setting $x_1^* := t^*$ and $x_i^* := \beta_{i-1}(t^*)$ for $i = 2, \ldots, s$, one verifies via (19) that (18) is satisfied and, hence, x^* is a solution to $(\mathrm{P_{c,\xi,p}})$.

Corollary 2. *Under the assumptions of Proposition 1 the estimate (16) holds true.*

Proof. We have to check that the assumptions of Theorem 3 are satisfied. The strong concavity of the log of Gaussian distribution functions with independent components is easy to verify (see [2, Prop. 14]). As shown in Proposition 1, the solution set $\Psi_{c,\xi,p}$ is nonempty. On the other hand, there may not exist more than one solution to problem $(\mathrm{P_{c,\xi,p}})$ because in its equivalent description

$$\min\{c^T x \mid -\log F_\xi(x) \leq -\log p\} \qquad (\mathrm{P_{c,\xi,p}})$$

the inequality constraint is strongly convex according to what we have mentioned in the begining of this proof.

We emphasize that in the result of Theorem 3 the approximating random vector η can be arbitrary. In the special case that η is an empirical approximation, one may exploit exponential bounds from empirical process theory (e.g., [8]) to further interpret the obtained stability result. In order to keep the presentation

simple, we refer here to a classical inequality by Kiefer [3] stating in any dimension s the existence of constants k_1 and $k_2 < 2$ (where k_2 may be chosen arbitrarily close to 2) such that for all $\tilde{\varepsilon} > 0$ and all η^N having an empirical distribution of an i.i.d. sample of ξ with size N the following estimate applies:

$$\mathbb{P}\left(\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_{\eta^N}(z)| \geq \tilde{\varepsilon}\right) \leq k_1 \exp\left(-k_2 \tilde{\varepsilon}^2 N\right). \tag{20}$$

Since by (16) one has for all $\varepsilon > 0$ the implication

$$\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_\eta(z)| < \min\{\delta, (\varepsilon/L)^2\} \implies d_H\left(\Psi_{c,\xi,p}, \Psi_{c,\eta,p}\right) < \varepsilon,$$

it follows from (20) with $\tilde{\varepsilon} := \min\{\delta, (\varepsilon/L)^2\}$ that

$$\mathbb{P}\left(d_H\left(\Psi_{c,\xi,p}, \Psi_{c,\eta^N,p}\right) \geq \varepsilon\right) \leq k_1 \exp\left(-k_2\left(\min\{\delta, (\varepsilon/L)^2\}\right)^2 N\right). \tag{21}$$

This last relation establishes an exponential bound for the convergence of Hausdorff distance between the solution sets of the original problem and the problem approximated by a sample of size N. Obviously, the quantity

$$k_2\left(\min\{\delta, (\varepsilon/L)^2\}\right)^2$$

determines the exponential decay of the required sample size. However, for practical use, one would have to know the values or at least estimates for δ and L which is difficult or impossible in general. Then, the availability of an exponential convergence result does not exclude excessive sample sizes even in order to give a sense to (21), i.e., to ensure that the right-hand side is smaller than one as an upper probability estimate. Revisiting Theorem 3, one observes that the couple (δ, L) in (16) is not uniquely determined. Therefore, let us define the best possible coefficient of exponential decay by

$$\vartheta\left(c, \xi, p, \varepsilon\right) := \sup\left\{k_2\left(\min\{\delta, (\varepsilon/L)^2\}\right)^2 \middle| (\delta, L) \text{ satisfy (16) for } (\mathrm{P}_{c,\xi,p})\right\}.$$

Then, (21) can be formally improved to

$$\mathbb{P}\left(d_H\left(\Psi_{c,\xi,p}, \Psi_{c,\eta^N,p}\right) \geq \varepsilon\right) \leq k_1 \exp\left(-\vartheta\left(c, \xi, p, \varepsilon\right) \cdot N\right).$$

The following result demonstrates that for a given significant problem class, this coefficient of exponential decay may be arbitrarily close to zero, thus driving the required sample size to infinity.

Theorem 4. *In (2), let $s > 1$, $\xi \sim \mathcal{N}(0, I_s)$ and $p \in (0, 1)$ be arbitrarily given. Then, for any $\varepsilon > 0$ one has that*

$$\inf\left\{\vartheta\left(c, \xi, p, \varepsilon\right) | c \succ 0\right\} = 0,$$

where '$c \succ 0$' means $c_i > 0$ for $i = 1, \ldots, s$.

Proof. Denote by τ the infimum above and assume that $\tau > 0$. Then, $\vartheta\left(c, \xi, p, \varepsilon\right) \geq \tau$ for all $c \succ 0$. By definition of $\vartheta\left(c, \xi, p, \varepsilon\right)$, we infer that

$$\forall c \succ 0 \, \exists\, (\delta, L) : (\delta, L) \text{ satisfy (16) for } (\mathrm{P}_{\mathrm{c},\xi,\mathrm{p}}) \text{ and } k_2 \left(\min\{\delta, (\varepsilon/L)^2\}\right)^2 \geq \tau/2.$$

The last relation entails that

$$\delta \geq \sqrt{\frac{\tau}{2k_2}} =: \bar{\delta}, \quad L \leq \varepsilon \sqrt[4]{\frac{2k_2}{\tau}} =: \bar{L}.$$

Note that $\bar{\delta}$ and $\bar{L}$ do not depend on c. Consequently, we have shown that there exist $\bar{\delta} > 0$ and $\bar{L}$ such that

$$\forall c \succ 0 \, \exists \delta \geq \bar{\delta}, L \leq \bar{L} : \; (16) \text{ holds true for } (\mathrm{P}_{\mathrm{c},\xi,\mathrm{p}}).$$

This statement can be evidently reduced to:

$$\forall c \succ 0 : d_H\left(\Psi_{c,\xi,p}, \Psi_{c,\eta,p}\right) \leq \bar{L} \sqrt{\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_\eta(z)|} \tag{22}$$

$$\forall \eta : \sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_\eta(z)| < \bar{\delta}.$$

From (20) we infer that, for any N,

$$\mathbb{P}\left(\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_{\eta^N}(z)| < \bar{\delta}/2\right) \geq 1 - k_1 \exp\left(-k_2 \bar{\delta}^2 N/4\right),$$

where η^N has an empirical distribution of an i.i.d. sample of ξ with size N. For $N \to \infty$, the right-hand side tends to one such that the probability on the left-hand side is at least strictly positive. As a consequence, for some fixed N large enough, there exists a discrete random vector η^N with N atoms such that

$$\sup_{z \in \mathbb{R}^s} |F_\xi(z) - F_{\eta^N}(z)| < \bar{\delta}/2.$$

Then, (22) implies that

$$\forall c \succ 0 : \quad d_H\left(\Psi_{c,\xi,p}, \Psi_{c,\eta^N,p}\right) \leq \bar{L}\sqrt{\bar{\delta}/2}. \tag{23}$$

Now, since $\operatorname{supp} \eta^N$ is compact, Theorem 2 yields the existence of some $\tilde{c} \succ 0$ such that

$$\inf\{\|x - y\| \mid x \in \Psi_{\tilde{c},\xi,p}, y \in \Psi_{\tilde{c},\eta^N,p}\} > \bar{L}\sqrt{\bar{\delta}/2}.$$

This, however, is a contradiction with (23) because

$$d_H\left(A, B\right) \geq \inf\{\|x - y\| \mid x \in A, y \in B\}$$

for any closed sets A, B. Hence, $\tau = 0$, as was to be shown.

The effect of the previous theorem can already be illustrated in two dimensions:

Example 1. Consider the following 2-dimensional problem:

$$\min\{x_1 + 10^{-4}x_2 \mid \mathbb{P}(\xi_1 \le x_1, \xi_2 \le x_2) \ge 0.99\}, \quad \xi \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

By independence of components, we may rewrite the chance constraint as

$$\Phi(x_1)\Phi(x_2) \ge 0.99,$$

where Φ denotes the one-dimensional standard normal distribution function. Referring to the optimality conditions as in (15), the solution of this problem is equivalently characterized by the following three nonlinear equations in the three variables x_1, x_2, λ:

$$\Phi(x_1)\Phi(x_2) = 0.99, \quad \varphi(x)\Phi(y) = \lambda, \quad \varphi(y)\Phi(x) = 10^{-4}\lambda.$$

Here, φ is the density of the one-dimensional standard normal distribution. This system is easily solved numerically, providing a unique optimal solution $x_1^* = 2.33$, $x_2^* = 4.88$.

Now, suppose that we want to approximate this solution empirically up to a precision of $\varepsilon = 0.1$. Then, in particular, the second component of the solution to the problem with empirical approximation has to exceed the value 4.78. Now, the second components $\xi_2^1, \ldots, \xi_2^N$ of an i.i.d. sample of ξ are i.i.d. standard normal. Hence,

$$\mathbb{P}(\max_{i=1,\ldots,N} \xi_2^i \le t) = \Phi^N(t).$$

For instance, for $N = 10^6$ and $t = 4.78$, one has $\Phi^N(t) \approx 0.42$. This means that the probability of obtaining a one-digit precise solution by empirical approximation with a huge sample size like one million is less than $1 - 0.42 = 0.58$.

Certainly, the effect of the example relies on the highly unbalanced cost vector $c = (1, 10^{-4})$. Making it more reasonably balanced like $c = (1, 0.1)$, one would still need a sample size of $N \approx 6.300$ for estimating the solution of the problem with a precision of 0.1 at a reasonably high probability of 0.99. Taking into account that N corresponds to the number of binary variables required in the discrete optimization problem, this is already a considerable quantity given the trivial dimension $s = 2$ of the problem. Of course, things may be expected to become much worse in larger dimension.

References

1. R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, Math. Program. 84 (1999), 55-88.
2. R. Henrion and W. Römisch, Hölder and Lipschitz Stability of Solution Sets in Programs with Probabilistic Constraints, Math. Program. 100 (2004), 589-611.

3. J. Kiefer, On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm, Pacific J. Math. 11 (1961), 649-660.
4. J. Luedtke and S. Ahmed, A sample approximation approach for optimization with probabilistic constraints, SIAM J. Optim. 19 (2008), 674699.
5. Pagnoncelli, B., Ahmed, S. and Shapiro, A., Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications, J. Optim. Theory Appl. 142 (2009), 399-416.
6. A. Prékopa, Stochastic Programming, Kluwer, Dordrecht, 1995.
7. A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*, MPS-SIAM series on optimization 9, 2009.
8. M. Talagrand, Sharper bounds for Gaussian and empirical processes, Ann. Probab. 22 (1994), 28-76