Convex Approximations of Chance Constrained Programs

Arkadi Nemirovski[*]   Alexander Shapiro [†]

Abstract

We consider a chance constrained problem, where one seeks to minimize a convex objective over solutions satisfying, with a given (close to one) probability, a system of randomly perturbed convex constraints. Our goal is to build a computationally tractable approximation of this (typically intractable) problem, i.e., an explicitly given convex optimization program with the feasible set contained in the one of the chance constrained problem. We construct a general class of such convex conservative approximations of the corresponding chance constrained problem. Moreover, under the assumptions that the constraints are affine in the perturbations and the entries in the perturbation vector are independent of each other random variables, we build a large deviations type approximation, referred to as 'Bernstein approximation', of the chance constrained problem. This approximation is convex, and thus efficiently solvable. We propose a simulation-based scheme for bounding the optimal value in the chance constrained problem and report numerical experiments aimed at comparing the Bernstein and well-known scenario approximation approaches. Finally, we extend our construction to the case of ambiguously chance constrained problems, where the random perturbations are independent with the collection of distributions known to belong to a given convex compact set rather than to be known exactly, while the chance constraint should be satisfied for every distribution given by this set.

Key words: stochastic programming, chance constraints, convex programming, Monte Carlo sampling, scenario generation, large deviations bounds, ambiguous chance constrained programming.

---
[*]Technion – Israel Institute of Technology, Haifa 32000, Israel, `nemirovs@ie.technion.ac.il`
[†]Georgia Institute of Technology, Atlanta, Georgia 30332, USA, `ashapiro@isye.gatech.edu`

1 Introduction

Let us consider the following optimization problem

$$\underset{x \in X}{\text{Min}}\, f(x) \ \text{ subject to } \ \text{Prob}\big\{F(x,\xi) \le 0\big\} \ge 1 - \alpha. \tag{1.1}$$

Here ξ is a random vector with probability distribution P supported on a set $\Xi \subset \mathbb{R}^d$, $X \subset \mathbb{R}^n$ is a nonempty convex set, $\alpha \in (0,1)$, $f : \mathbb{R}^n \to \mathbb{R}$ is a real valued convex function, $F = (f_1, ..., f_m) : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$, and $\text{Prob}(A)$ denotes probability of an event A. Probability constraints of the form appearing in (1.1) arise naturally in various applications and are called *chance* (or probabilistic) constraints. Such constraints can be viewed as a compromise with the requirement of enforcing the constraints $F(x,\xi) \le 0$ for *all* values $\xi \in \Xi$ of the uncertain data vector, which could be too costly or even impossible. Chance constrained optimization problems were introduced in Charnes et al [7], Miller and Wagner [14] and Prékopa [18].

There are two basic problems with the above formulation (1.1). The first is a modeling problem. That is, in order to evaluate, for a given $x \in X$, probability of the event "$F(x,\xi) \le 0$" we need to know the probability distribution of random vector ξ. (In the sequel we denote by ξ random data vector as well as its realization; which one of these two meanings will be used in a particular situation will be clear form the context.) In practical applications the involved probability distributions are never known exactly and could be estimated, may be from historical data, at best. This raises the question of how such ambiguity with respect to a chosen probability distribution affects robustness of an optimal solution of the corresponding chance constrained problem. Also the choice of the significance level (reliability parameter, or *risk*) α often is quite arbitrary, i.e., why we prefer to take $\alpha = 5\%$ over say $\alpha = 2\%$, or even more dramatically of $\alpha = 10^{-4}$ over $\alpha = 10^{-6}$? In the absence of knowledge of the underlying probability distribution, an exact specification of α becomes questionable especially for very small values of α when a high reliability is required.

The second problem is numerical. Although some approaches to solving (1.1) were suggested in the literature, they typically are applicable either in rather specific situations or when the number of involved variables is relatively small (cf., [19]). An explanation of this is that generically, with minor exceptions, chance constrained problems are "severely computationally intractable". The reason is twofold: first, typically ξ is multi-dimensional, and in this case it is usually difficult even to check whether or not a given chance constraint is satisfied at a given point x, there are no ways to compute efficiently the corresponding probabilities to high accuracy (the latter should be of order of α, and α can be really small). Typically, the only way to estimate the probability for a chance constraint to be violated at a given point is to use Monte-Carlo simulation, and this becomes too costly when α is small. The second, and more severe difficulty with chance constraints "as they are", is that even with nice, say affine in x and in ξ, functions $F(x,\xi)$ the feasible set of a chance constraint usually is nonconvex, which makes optimization under this constraint highly problematic. It should be stressed that both outlined difficulties are "rules rather than exceptions". The only known to us generic case when these difficulties do not arise is the chance constrained version of a single linear inequality with normally distributed random coefficients and $\alpha \le 1/2$.

Since chance constrained problems "as they are" are computationally intractable, a natural way to process such problems is to look for their *tractable approximations*, i.e., for efficiently verifiable *sufficient conditions* for the validity of the chance constraint in (1.1). In addition to being sufficient, these conditions should define a convex and "computationally tractable" set in the x-space, e.g.,

1

should be represented by a system of convex inequalities $G(x, u) \le 0$ in x and, perhaps, in additional variables $u \in \mathbb{R}^s$. Whenever this is the case, the problem

$$\min_{x \in X, u} f(x) \text{ subject to } G(x, u) \le 0 \tag{1.2}$$

is a Convex Programming program. As such it is efficiently solvable, provided that $G(x, u)$ is efficiently computable, and provides a "safe approximation" of the chance constrained problem of interest – the x-component of a feasible solution to (1.2) is feasible for (1.1).

A general way to build computationally tractable approximations of chance constrained problems is offered by the *scenario approach* based on Monte Carlo sampling techniques. That is, one generates a sample $\xi^1, ..., \xi^N$ of N (independent) realizations of the random vector ξ and approximates (1.1) with the problem

$$\min_{x \in X} f(x) \text{ subject to } F(x, \xi^\nu) \le 0, \nu = 1, ..., N. \tag{P^N}$$

The main advantage of this approach is its generality, it imposes no restrictions on the distribution of ξ and on how the data enter the constraints. In order to build (P^N) there is no need even to know what is the distribution of ξ, all we need is to be able to sample from this distribution. Last, but not least, is the "tractability status" of the approximation. The approximation (P^N) is efficiently solvable, provided that the function $F(x, \xi)$ is componentwise convex in x and is efficiently computable, and the sample size N is not too large. We do not know exactly to whom the scenario approach to handling uncertain constraints should be attributed. To the best of our knowledge, it originates from [15], where, however, it was presented as an ad hoc remedy to struggle with data uncertainty in optimization, with no analysis and no links to chance constraints.

An important theoretical question related to the scenario approximation is the following. The approximation itself is random, and its solution may not satisfy the chance constraints. The question is, how large should be the sample size N in order to ensure, with probability at least $1 - \delta$, that the optimal solution to (P^N) is feasible for the problem of interest (1.1). To some extend this question was resolved in recent papers of Calafiore and Campi [5, 6] and de Farias and Van Roy [9]. Their results were then extended in [11] to a more complicated case of *ambiguous* chance constraints (that is, the case when the "true" distribution of ξ is assumed to belong to a given family of distributions rather than to be known exactly, while the samples are drawn from a specified reference distribution). The answer to the outlined question, as given in [6], is, that if $F(x, \xi)$ is componentwise convex in x, then, under mild additional conditions, with the sample size N satisfying

$$N \ge N^* := \text{Ceil}\left[\frac{2n}{\alpha} \log\left(\frac{12}{\alpha}\right) + \frac{2}{\alpha} \log\left(\frac{2}{\delta}\right) + 2n\right], \tag{1.3}$$

the optimal solution to (P^N) is, with probability at least $1 - \delta$, feasible for the chance constrained problem (1.1). A remarkable feature of this result is that it, similarly to the scenario approximation itself, is completely distribution-free.

Aside from the conservativeness (which is a common drawback of all approximations), an intrinsic drawback of the scenario approximation based on (1.3) is that, as it is easily seen, the sample size N should be at least inverse proportional to the risk α and thus could be impractically large when the risk is small. Moreover, the sample size as given by (1.3) (and by all other known results of this type) grows linearly with n, which makes it difficult to apply the approach already to medium-size

2

problems (with $\alpha = 0.01$ and $n = 200$, $\delta = 0.01$, the estimate (1.3) results in $N^* = 285,063$). Note that for a properly modified scenario approximation, "bad" dependence of N on α given by (1.3) can be replaced with

$$N = O(1) \left[\log(1/\delta) + dm^2 \log(d \log(1/\alpha)) \right], \tag{1.4}$$

provided that $F(x, \xi)$ is affine in ξ and ξ has a "nice" distribution, e.g., uniform in a box or normal [17].

An alternative to the scenario approximation is an approximation based on "analytical" upper bounding of the probability for the randomly perturbed constraint $F(x, \xi) \leq 0$ to be violated. The simplest approximation scheme of this type was proposed in [2] for the case of a single affine in ξ inequality

$$f_0(x) + \sum_j \xi_j f_j(x) \leq 0. \tag{1.5}$$

Assuming that ξ_j are independent of each other random variables with zero means varying in segments $[-\sigma_i, \sigma_i]$, it is easy to see that if x satisfies the constraint

$$f_0(x) + \Omega \left(\sum_{j=1}^{d} \sigma_j^2 f_j^2(x) \right)^{1/2} \leq 0, \tag{1.6}$$

where $\Omega > 0$ is a "safety" parameter, then x violates the randomly perturbed constraint (1.5) with probability at most $\exp\left\{ -\kappa \Omega^2 \right\}$, where $\kappa > 0$ is an absolute constant (as we shall see in Section 6, one can take $\kappa = 1/2$). It follows that if all components $f_i(x, \xi)$ are of the form

$$f_i(x, \xi) = f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x), \tag{1.7}$$

then the optimization program

$$\underset{x \in X}{\text{Min}} \, f(x) \text{ subject to } f_{i0}(x) + \Omega \left(\sum_{j=1}^{d} \sigma_j^2 f_{ij}^2(x) \right)^{1/2} \leq 0, \, i = 1, ..., m, \tag{1.8}$$

with $\Omega := \sqrt{2 \log(m\alpha^{-1})}$, is an approximation of the chance constrained problem (1.1). This approximation is convex, provided that all $f_{ij}(x)$ are convex and every one of the functions $f_{ij}(x)$ with $j \geq 1$ is either affine, or nonnegative. Another, slightly more convenient computationally, analytical approximation of randomly perturbed constraint (1.5) was proposed in [4]. Analytical approximations of more complicated chance constraints, notably a randomly perturbed conic quadratic inequality, are presented in [16]. An advantage of the "analytical" approach as compared to the scenario one is that the resulting approximations are deterministic convex problems with sizes independent of the required value of risk (reliability) α, so that these approximations remain practical also in the case of very small values of α. On the negative side, building an analytical approximation requires structural assumptions on $F(x, \xi)$ and on the stochastic nature of ξ (in all known constructions of this type, ξ_j should be independent of each other and possess "nice" distributions).

In this paper, we propose a new class of analytical approximations of chance constraints, referred to as *Bernstein* approximations. Our major assumptions are that the components of $F(x, \xi)$ are of

3

the form (1.7) with convex $f_{ij}(x)$, and ξ_j are independent of each other and possess distributions with efficiently computable moment generating functions. Besides this, we assume that for every $j \geq 1$ for which not all of the functions $f_{ij}(x)$, $i = 1, ..., m$, are affine, the corresponding random variable ξ_j is nonnegative. Under these assumptions, the approximation we propose is an explicit convex program.

The rest of the paper is organized as follows. In section 2 we introduce a class of convex conservative approximations of (1.1). Bernstein approximation of (1.1) is derived and discussed in section 3. In section 4, we propose a simple simulation-based scheme for bounding the true optimal value in (1.1), which allows to evaluate numerically the quality (that is, the conservatism) of various approximations. In section 5, we report some preliminary numerical experiments with Bernstein approximation. Our numerical results demonstrate that this approximation compares favorably with the scenario one. In concluding section 6, we extend Bernstein approximation to the case of *mixed uncertainty model*, where the tuple of distributions of (mutually independent) components ξ_j of ξ is assumed to belong to a given convex compact set rather than to be known exactly and, on the top of it, the constraints are affected by additional "uncertain but bounded" perturbations of non-stochastic nature (cf., [11], where similar extensions of the scenario approach are considered).

2 Convex approximations of chance constrained problems

In this section we discuss convex approximations of chance constrained problems of the form (1.1). As it was mentioned in Introduction, chance constrained problems, even simple-looking, typically are computationally intractable. A natural way to overcome, to some extent, this difficulty is to replace chance constraint problem (1.1) with a *tractable approximation*. That is, with an efficiently solvable problem of the form (1.2). To this end we require the function $G(x, u)$ to be *convex* in (x, u). We also would like the constraints $G(x, u) \leq 0$ to be *conservative*, in the sense that if for $x \in X$ and u it holds that $G(x, u) \leq 0$, then $\text{Prob}\{F(x, \xi) \leq 0\} \geq 1 - \alpha$. Thus, feasible solutions to (1.2) induce feasible solutions to (1.1), so that the optimal solution of the approximation is a feasible sub-optimal solution of the problem of interest. If these two conditions hold, we refer to (1.2) as a *convex conservative* approximation of the true problem (1.1). Our goal in this section is to construct a special class of convex conservative approximations.

Let us consider first the scalar case of $m = 1$, i.e., $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$. Then the probabilistic (chance) constraint of problem (1.1) is equivalent to the constraint

$$p(x) := \text{Prob}\{F(x, \xi) > 0\} \leq \alpha. \tag{2.1}$$

By $\mathbb{1}_A$ we denote the indicator function of a set A, i.e., $\mathbb{1}_A(z) = 1$ if $z \in A$ and $\mathbb{1}_A(z) = 0$ if $z \notin A$.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a *nonnegative valued, nondecreasing, convex* function satisfying the following property:

($*$) $\psi(z) > \psi(0) \geq 1$ for any $z > 0$.

We refer to function $\psi(z)$ satisfying the above properties as a (one dimensional) *generating function*. It follows from ($*$) that for $t > 0$ and random variable Z,

$$\mathbb{E}[\psi(tZ)] \geq \mathbb{E}\left[\mathbb{1}_{[0, +\infty)}(tZ)\right] = \text{Prob}\{tZ \geq 0\} = \text{Prob}\{Z \geq 0\}.$$

4

By taking $Z = F(x, \xi)$ and changing t to t^{-1}, we obtain that

$$p(x) \leq \mathbb{E}\left[\psi\left(t^{-1}F(x, \xi)\right)\right] \tag{2.2}$$

holds for all x and $t > 0$. Denote

$$\Psi(x, t) := t\,\mathbb{E}\left[\psi\left(t^{-1}F(x, \xi)\right)\right]. \tag{2.3}$$

We obtain that if there exists $t > 0$ such that $\Psi(x, t) \leq t\alpha$, then $p(x) \leq \alpha$. In fact this observation can be strengthened to:

$$\inf_{t>0}\left[\Psi(x, t) - t\alpha\right] \leq 0 \quad \text{implies} \quad p(x) \leq \alpha. \tag{2.4}$$

Indeed, let us fix x and set $\phi(t) := \Psi(x, t) - t\alpha$, $Z := F(x, \xi)$. It may happen (case (A)) that $\text{Prob}\{Z > 0\} > 0$. Then there exist $a, b > 0$ such that $\text{Prob}\{Z \geq a\} \geq b$, whence

$$\Psi(x, t) = t\,\mathbb{E}\left[\psi(t^{-1}F(x, \xi))\right] \geq tb\psi(t^{-1}a) \geq tb\left[\psi(0) + \frac{\psi(a) - \psi(0)}{t}\right]$$

provided that $0 < t < 1$ (we have taken into account that $\psi(\cdot)$ is convex). Since $\psi(a) > \psi(0)$, we conclude that

$$\Psi(x, t) \geq \gamma := b(\psi(a) - \psi(0)) > 0, \text{ for } 0 < t < 1,$$

and hence $\liminf_{t \to +0} \phi(t) > 0$. Further, we have

$$\lim\inf_{t \to \infty} \mathbb{E}\left[\psi(t^{-1}Z)\right] \geq \psi(0) \geq 1,$$

and hence $\liminf_{t \to \infty} \phi(t) = \infty$ due to $\alpha \in (0, 1)$. Finally, $\phi(t)$ is clearly lower semicontinuous in $t > 0$. We conclude that if (A) is the case, then $\inf_{t>0} \phi(t) \leq 0$ if and only if there exists $t > 0$ such that $\phi(t) \leq 0$, and in this case, as we already know, $p(x)$ indeed is $\leq \alpha$. And if (A) is not the case, then the conclusion in (2.4) is trivially true, so that (2.4) is true.

We see that the inequality

$$\inf_{t>0}\left[\Psi(x, t) - t\alpha\right] \leq 0 \tag{2.5}$$

is a conservative approximation of (2.1) – whenever (2.5) is true, so is (2.1). Moreover, assume that for every $\xi \in \Xi$ the function $F(\cdot, \xi)$ is convex. Then $G(x, t) := \Psi(x, t) - t\alpha$ is convex[1] in $(x, t > 0)$. Furthermore, since $\psi(\cdot)$ is nondecreasing and $F(\cdot, \xi)$ is convex, it follows that $(x, t) \mapsto t\psi(t^{-1}F(x, \xi))$ is convex. This, in turn implies convexity of the expected value function $\Psi(x, t)$, and hence convexity of $G(x, t)$.

We obtain, under the assumption that X, $f(\cdot)$ and $F(\cdot, \xi)$ are convex, that

$$\min_{x \in X, t > 0} f(x) \quad \text{subject to} \quad \inf_{t>0}\left[\Psi(x, t) - t\alpha\right] \leq 0 \tag{2.6}$$

---

[1] We have used the well-known fact that if $f(x)$ is convex, so is the function $g(x, t) = tf(t^{-1}x)$, $t > 0$. Indeed, given x', x'', $\lambda \in (0, 1)$ and $t', t'' > 0$ and setting $t = \lambda t' + (1 - \lambda)t''$, $x = \lambda x' + (1 - \lambda)x''$, we have

$$\begin{aligned}
\lambda t' f(x'/t') + (1 - \lambda)t'' f(x''/t'') &= t\left[\frac{\lambda t'}{t}f(x'/t') + \frac{(1-\lambda)t''}{t}f(x''/t'')\right] \\
&\geq tf\left(\frac{t'\lambda}{t}\frac{x'}{t'} + \frac{(1-\lambda)t''}{t}\frac{x''}{t''}\right) = tf(x/t).
\end{aligned}$$

5

gives a *convex* conservative approximation of the chance constrained problem (1.1).

Clearly the above construction depends on a choice of the generating function $\psi(z)$. This raises the question of what would be a "best" choice of $\psi(z)$. If we consider this question from the point of view of a better (tighter) approximation of the corresponding chance constraints, then the smaller is $\psi(\cdot)$, the better is bound (2.2). Now if $\psi(0) > 1$, then the outlined bounds can be improved by replacing $\psi(\cdot)$ with $\psi(\cdot)/\psi(0)$. Thus we may assume that $\psi(0) = 1$. Further, if the right derivative $\psi'_+(0)$ is zero, then $\psi(z) \geq \psi(0) = 1$ for all $z \in \mathbb{R}$, and the above construction produces trivial bounds. Therefore we may assume that $a := \psi'_+(0) > 0$. Since $\psi(0) = 1$ and $\psi(\cdot)$ is convex and nonnegative, we conclude that $\psi(z) \geq \max\{1 + az, 0\}$ for all z, so that the upper bounds (2.2) can be only improved when replacing $\psi(z)$ with the function $\hat{\psi}(z) := \max\{1 + az, 0\}$, which also is a generating function. But the bounds produced by the latter function are, up to scaling $z \leftarrow z/a$, the same as those produced by the function

$$\psi^*(z) := [1 + z]_+, \tag{2.7}$$

where $[a]_+ := \max\{a, 0\}$. That is, from the point of view of the most accurate approximation, the best choice of the generating function ψ is the piecewise linear function ψ^* defined in (2.7).

For the generating function ψ^* defined in (2.7) the approximate constraint (2.5) takes the form

$$\inf_{t>0} \left[\mathbb{E}\big[[F(x,\xi) + t]_+ \big] - t\alpha \right] \leq 0. \tag{2.8}$$

In that form it is related to the concept of Conditional Value at Risk (CVaR) due to Rockafellar and Uryasev [20]. Recall that CVaR of a random variable Z is

$$\mathrm{CVaR}_{1-\alpha}(Z) := \inf_{\tau \in \mathbb{R}} \left[\tau + \frac{1}{\alpha} \mathbb{E}[Z - \tau]_+ \right]. \tag{2.9}$$

Equivalently, $\mathrm{CVaR}_{1-\alpha}(Z) = \mathbb{E}\left[Z \big| Z > \mathrm{VaR}_{1-\alpha}(Z) \right]$, where

$$\mathrm{VaR}_{1-\alpha}(Z) := \inf\left[t : \mathrm{Prob}(Z \leq t) \geq 1 - \alpha \right]$$

is the corresponding quantile (also called Value at Risk) of the distribution of Z. Note that the minimizer of the right hand side of (2.9) is $\tau^* = \mathrm{VaR}_{1-\alpha}(Z)$. Therefore, it always holds that $\mathrm{CVaR}_{1-\alpha}(Z) \geq \mathrm{VaR}_{1-\alpha}(Z)$. It is also known that $\mathrm{CVaR}_{1-\alpha}(Z)$ is convex and nondecreasing in Z. Clearly the chance constraint of (1.1) can be written as $\mathrm{VaR}_{1-\alpha}[F(x,\xi)] \leq 0$. Therefore, the constraint

$$\mathrm{CVaR}_{1-\alpha}[F(x,\xi)] \leq 0 \tag{2.10}$$

defines a convex conservative approximation of the chance constrained problem (1.1). The idea of using CVaR as a convex approximation of VaR is due to Rockafellar and Uryasev [20]. Note that since $\alpha > 0$, the validity of (2.8) is independent of whether the infimum in the left hand side is taken over $t > 0$ or over $t \in \mathbb{R}$. Therefore, by the above discussion, the constraints (2.8) and (2.10) are equivalent to each other.

Remark 1 One of possible drawbacks of using the "optimal" generating function ψ^* (as compared with the exponential $\psi(z) := e^z$, which we will discuss in the next section) in the above approximation scheme is that it is unclear how to compute efficiently the corresponding function $\Psi(x,t)$ even

in the simple case $F(x,\xi) := g_0(x) + \sum_{j=1}^d \xi_j g_j(x)$ of affine in ξ function $F(x,\xi)$ and independent of each other random variables ξ_j with known and simple distributions. It is unclear how serious is this obstacle in the latter case. Note that $[s]_+$ can be written in the following integral form

$$[s]_+ = \tfrac{1}{2}s + \frac{1}{\pi}\int_0^\infty \frac{1 - \cos(\omega s)}{\omega^2} d\omega.$$

Therefore, we can write

$$\mathbb{E}\left[1 + t^{-1}F(x,\xi)\right]_+ = \tfrac{1}{2}\left(1 + t^{-1}\mathbb{E}[F(x,\xi)]\right) + \frac{1}{\pi}\int_0^\infty \frac{1 - h(\omega)}{\omega^2}d\omega, \tag{2.11}$$

where $h(\omega)$ is the real part of the complex quantity $\mathbb{E}\left[\exp\left\{i\omega\left(t^{-1}F(x,\xi) + 1\right)\right\}\right]$ (here i is the imaginary unit). In the case in question we can easily compute $h(\omega)$ for every ω, since

$$\mathbb{E}\left[\exp\{i\omega F(x,\xi)\}\right] = \exp\left\{ig_0(x)\right\}\prod_{j=1}^d \mathbb{E}\left[\exp\left\{i\omega g_j(x)\xi_j\right\}\right],$$

so that the computation reduces to taking d one-dimensional expectations with respect to simple distributions. The difficulty with the outlined integral representation is that the right hand side of (2.11) contains an oscillating integral which should be computed, also for large values of t^{-1}, within accuracy of order of α. We failed to build a routine capable to compute this integral both fast and accurate; perhaps others will be able to resolve this problem, thus achieving the limits of performance of the approach we are discussing.

There are several ways how the above construction can be extended for $m > 1$. One simple way is to replace the constraints $f_i(x,\xi) \leq 0$, $i = 1, ..., m$, with one constraint $f(x,\xi) \leq 0$, say by taking $f(x,\xi) := \max\{f_1(x,\xi), ..., f_m(x,\xi)\}$. Note, however, that this may destroy a simple, e.g., affine in ξ, structure of the constraint mapping $F(x,\xi)$. An alternative approach is the following.

Consider a closed convex cone $K \subseteq \mathbb{R}_+^m$ and the corresponding partial order $\succeq_K$, i.e., $z \succeq_K y$ iff $z - y \in K$. Of course, for the nonnegative orthant cone $K := \mathbb{R}_+^m$ the constraint $F(x,\xi) \leq 0$ means that $F(x,\xi) \preceq_K 0$. We can also consider some other convex closed cones and define constraints in that form. The corresponding chance constraint can be written in the form

$$p(x) := \mathrm{Prob}\big\{F(x,\xi) \notin -K\big\} < \alpha. \tag{2.12}$$

Let $\psi : \mathbb{R}^m \to \mathbb{R}$ be a nonnegative valued, convex function such that:

($\star$) ψ is K-monotone, i.e., if $z \succeq_K y$, then $\psi(z) \geq \psi(y)$,

($^\star_\star$) $\psi(z) > \psi(0) \geq 1$ for every $z \in \mathbb{R}^m \setminus (-K)$.

We refer to function $\psi(z)$ satisfying these properties as a K-generating function.

By ($^\star_\star$) we have that $\mathbb{E}[\psi(F(x,\xi))]$ provides an upper bound for $p(x)$, and the corresponding inequality of the form (2.2) holds. Suppose, further, that for every $\xi \in \Xi$ the mapping $F(\cdot,\xi)$ is K-convex, i.e., for any $t \in [0,1]$ and $x, y \in \mathbb{R}^n$,

$$tF(x,\xi) + (1 - t)F(y,\xi) \succeq_K F(tx + (1 - t)y, \xi).$$

7

(Note that for $K = \mathbb{R}^m_+$, K-convexity means that $F(\cdot, \xi)$ is componentwise convex.) Then for $\Psi(x, t) := t\, \mathbb{E}[\psi(t^{-1}F(x, \xi))]$, problem of the form (2.6) gives a convex conservative approximation of the chance constrained problem (1.1).

In such construction for $m > 1$, there is no "best" choice of the K-generating function $\psi(z)$. A natural choice in the case of $K = \mathbb{R}^m_+$ could be

$$\hat{\psi}(z) := \max_{1 \le i \le m}[1 + a_i z_i]_+, \tag{2.13}$$

where $a_i > 0$ are "scale parameters".

Yet there is another possible extension of the above approximation scheme for $m > 1$. Let $\alpha_1, ..., \alpha_m$ be positive numbers such that $\alpha_1 + ... + \alpha_m \le \alpha$. The chance constraint of (1.1) is equivalent to $\mathrm{Prob}\{\cup_{i=1}^m \{\xi : f_i(x, \xi) > 0\}\} < \alpha$. Since

$$\mathrm{Prob}\{\cup_{i=1}^m \{f_i(x, \xi) > 0\}\} \le \sum_{i=1}^m \mathrm{Prob}\{f_i(x, \xi) > 0\},$$

it follows that the system of constraints

$$\mathrm{Prob}\{f_i(x, \xi) > 0\} \le \alpha_i, \quad i = 1, ..., m, \tag{2.14}$$

is more conservative then the original chance constraint. We can apply now the one-dimensional construction to each individual constraint of (2.14) to obtain the following convex conservative approximation of the chance constrained problem (1.1):

$$\mathop{\mathrm{Min}}_{x \in X} f(x) \ \text{ subject to } \ \inf_{t > 0}[\Psi_i(x, t) - t\alpha_i] \le 0, \ i = 1, ..., m, \tag{2.15}$$

where $\Psi_i(x, t) := t\mathbb{E}\left[\psi_i(t^{-1}f_i(x, \xi))\right]$, and each $\psi_i(\cdot)$, $i = 1, ..., m$, is a one-dimensional generating function.

Remark 2 An open question related to the approximation (2.15) is how to choose α_i. It would be very attractive to treat these quantities in (2.15) as design variables (subject to the constraints $\alpha_i > 0$ and $\sum_i \alpha_i \le \alpha$) rather than as parameters. Unfortunately, such an attempt destroys the convexity of (2.15) and thus makes the approximation seemingly intractable. The simplest way to resolve the issue in question is to set

$$\alpha_i := \alpha/m, \ i = 1, ..., m. \tag{2.16}$$

3 Bernstein approximation

One of drawbacks of using the piecewise linear generating functions of the form (2.7) (or (2.13)), is that the corresponding constraint function may be difficult to compute even for relatively simple functions $F(x, \xi)$ (compare with Remark 1). In this section we consider the (one-dimensional) generating function $\psi(z) := e^z$. For such choice of the generating function, constructions of the previous section are closely related to the classical Large Deviations theory (cf., [8]).

We assume in this section that:

A1. The components ξ_j, $j = 1, ..., d$, of the random vector ξ are independent of each other random variables.

We denote by P_j the probability distribution of ξ_j, supported on $\Xi_j \subset \mathbb{R}$ (so that the support of the distribution P of ξ is $\Xi = \Xi_1 \times ... \times \Xi_d$), by

$$M_j(t) := \mathbb{E}\left[e^{t\xi_j}\right] = \int \exp(tz) dP_j(z)$$

the moment generating function, and by $\Lambda_j(t) := \log M_j(t)$ the logarithmic moment generating function of ξ_j.

A2. The moment generating functions $M_j(t)$, $j = 1, ..., d$, are finite valued for all $t \in \mathbb{R}$ and are efficiently computable.

In fact, we could allow for the moment generating functions to be finite valued just in a neighborhood of $t = 0$. We make the stronger assumption of requiring the moment generating functions to be finite valued for all t in order to simplify the presentation.

A3. The components $f_i(x, \xi)$ in the constraint mapping $F(x, \xi)$ are affine in ξ:

$$f_i(x, \xi) = f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x), \quad i = 1, ..., m, \tag{3.1}$$

and the functions $f_{ij}(x)$, $j = 0, 1, ..., d$, are well-defined and convex on X. Besides this, for every $j \geq 1$ such that $\Xi_j \not\subset \mathbb{R}_+$, all functions $f_{ij}(x)$, $i = 1, ..., m$, are affine. In addition, the objective $f(x)$ in (1.1) is well-defined and convex on X.

In the sequel, we refer to problem (1.1) satisfying the assumptions A1 - A3 as to *affinely perturbed convex* chance constrained problem.

Let $z = (z_0, z_1, ..., z_d) \in \mathbb{R}^{d+1}$. By A1, the function

$$\Phi(z) := \log\left(\mathbb{E}\left[\exp\left\{z_0 + \sum_{j=1}^{d} \xi_j z_j\right\}\right]\right) = z_0 + \sum_{j=1}^{d} \Lambda_j(z_j)$$

is well-defined and continuous in z. Besides this, it is convex (since, as it is well-known, the logarithmic moment generating functions are so). Moreover, $\Phi(z)$ is monotone in z_0 and in every z_j with $j \in J := \{j \geq 1 : \Xi_j \subset \mathbb{R}_+\}$. Finally, one clearly has for $t > 0$ and $p(z) := \text{Prob}\{z_0 + \sum_{j=1}^{d} \xi_j z_j > 0\}$ that

$$\Phi(t^{-1}z) \geq \log p(z).$$

Consequently, for every $\beta \in (0, 1)$,

$$\exists t > 0 : t\Phi(t^{-1}z) - t\log\beta \leq 0 \text{ implies } p(z) \leq \beta.$$

Similarly to the reasoning which led us to (2.4), the latter implication can be strengthened to:

$$\inf_{t>0}\left[t\Phi(t^{-1}z) - t\log\beta\right] \leq 0 \text{ implies } p(z) \leq \beta. \tag{3.2}$$

Now consider an affine chance constrained problem with real-valued constraint mapping

$$F(x, \xi) = g_0(x) + \sum_{j=1}^{d} \xi_j g_j(x).$$

By (3.2), the problem

$$\operatorname*{Min}_{x \in X} f(x) \text{ subject to } \inf_{t>0} \left[g_0(x) + \sum_{j=1}^{d} t \Lambda_j(t^{-1} g_j(x)) - t \log \alpha \right] \leq 0 \tag{3.3}$$

is a conservative approximation of the chance constrained problem (1.1). In fact this approximation is convex. Indeed, the function

$$G(z, t) := t \Phi(t^{-1} z) - t \log \beta$$

is convex in $(z, t > 0)$ (since $\Phi(z)$ is convex) and is monotone in z_0 and every z_j with $j \in J$, while, by A3, all $g_j(x)$, $j = 0, 1, ..., d$, are convex in $x \in X$, and all $g_j(x)$ with $j \geq 1$ such that $j \notin J$, are affine. It follows that the function $G(g_0(x), ..., g_d(x), t)$ is convex in $(x \in X, t > 0)$, whence the constraint in (3.3) is convex; the objective is convex by A3, and X was once for ever assumed to be convex when formulating (1.1). Thus, (3.3) is a *convex conservative* approximation of an affinely perturbed chance constrained problem with $m = 1$, as claimed.

We can extend the outlined construction to the multivariate case of $m > 1$ in a way similar to the construction of problem (2.15). That is, given an affinely perturbed chance constrained problem (1.1), (3.1), we choose $\alpha_i > 0$, $\sum_i \alpha_i \leq \alpha$, and build the optimization problem

$$\begin{aligned} \operatorname*{Min}_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \inf_{t>0} \left[f_{i0}(x) + \sum_{j=1}^{d} t \Lambda_j(t^{-1} f_{ij}(x)) - t \log \alpha_i \right] \leq 0, \ i = 1, ..., m. \end{aligned} \tag{3.4}$$

Similarly to the case of $m = 1$, this problem is a convex conservative approximation of (1.1). We refer to (3.4) as the *Bernstein* approximation of (1.1). The reason is that this construction is based on the ideas used by S.N. Bernstein when deriving his famous inequalities for probabilities of large deviations of sums of independent random variables.

An advantage of Bernstein approximation over the one discussed in the previous section, is that under assumptions A1 – A3 the Bernstein approximation is an explicit convex program with efficiently computable constraints and as such is efficiently solvable.

Remark 3 A somehow less accurate version of Bernstein approximation was in fact proposed in [2] for the situation where the random variables ξ_j are independent with zero mean and supported on segments $[-\sigma_i, \sigma_i]$. We have cited this result in Introduction, see (1.8). The justification of (1.8) is based on a straightforward bounding from above (going back to Bernstein) of the associated logarithmic moment generating function and demonstrating that if x satisfies (1.6), then the resulting (conservative) version of the corresponding probability bound, as applied to $z = (f_{i0}(x), f_{i1}(x), ..., f_{id}(x))$, implies that

$$\operatorname{Prob} \left\{ f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x) > 0 \right\} \leq \exp\{-\kappa \Omega^2\}.$$

Clearly, the Bernstein approximation as presented here is less conservative than (1.8), since it is based on the corresponding "true" function rather than on its upper bound given solely by the expected values and the sizes of supports of ξ_j.

3.1 How conservative is Bernstein approximation?

The question posed in the title of this section reduces to the following one:

(?) *Let $z = (z_0, ..., z_d) \in \mathbb{R}^{d+1}$, and let ξ_j be independent random variables with distributions P_j, $j = 1, ..., d$. How conservative could be the Bernstein sufficient condition*

$$\inf_{t>0} \left[t \log \left(\mathbb{E} \left[\exp\{t^{-1}[z_0 + \sum_{j=1}^{d} \xi_j z_j]\} \right] \right) - t \ln \beta \right] \leq 0 \qquad (3.5)$$

for the validity of the chance constraint

$$p(z) := \mathrm{Prob} \left\{ z_0 + \sum_{j=1}^{d} \xi_j z_j > 0 \right\} \leq \beta. \qquad (3.6)$$

To answer this question, we should of course decide how we measure the conservatism. The simplest way to do it would be to ask: How small should be the actual value of $p(z)$ in order for the sufficient condition to hold true? The answer to this question, in general, can be very pessimistic. To give an example, consider the case when $d = 1$, $z = (-\kappa, 1 + \kappa)$ and ξ takes two values, 0 and 1, with probabilities $1 - \epsilon$ and ϵ, respectively, where $\kappa > 0$ and $\epsilon \in (0, 1)$ are parameters. In this case, $p(z) = \epsilon$. On the other hand, (3.5) in our case clearly reads

$$\exists \tau > 0 : (1 - \epsilon) \exp\{-\kappa\tau\} + \epsilon \exp\{\tau\} \leq \beta,$$

or, after minimizing the left hand side in $\tau > 0$,

$$(1 + \kappa)(\kappa^{-1}\epsilon)^{\frac{\kappa}{1+\kappa}}(1 - \epsilon)^{\frac{1}{1+\kappa}} \leq \beta.$$

We see that if $\epsilon << 1$, then, in order to satisfy the latter condition, β should be of order of $\epsilon^{\frac{\kappa}{1+\kappa}} = (p(z))^{\frac{\kappa}{1+\kappa}}$, that is, the conservatism of (3.5) grows dramatically as κ approaches 0. For example, with $\kappa = 0.1$ and $p(z) \equiv \epsilon = 0.01$, (3.5) is able to certify only that $p(z) \leq 0.8841$.

There is, however, another way to measure the conservatism, perhaps a more natural one. Assume, as it is the case in many applications, that all ξ_j are symmetrically distributed around their expected values. By an appropriate updating of z_0, we can assume without loss of generality that all these expected values are 0, so that ξ_j are independent of each other and symmetrically distributed around 0. Now let us consider chance constraint (3.6) as a member of the parametric family

$$p_\rho(z) := \mathrm{Prob} \left\{ z_0 + \rho \sum_{j=1}^{d} \xi_j z_j > 0 \right\} \leq \beta, \qquad (3.7)$$

where $\rho > 0$ is an "uncertainty level"; the original constraint corresponds to $\rho = 1$. It is immediately seen that if $\beta < 1/2$, then the larger is ρ, the larger is $p_\rho(z)$, so that the feasible set Z_ρ of (3.7)

in the z-space shrinks as ρ grows. We can now measure the conservatism of (3.5) by the smallest $\rho = \rho_* \geq 1$ such that the feasible set of (3.5) is in-between the feasible set Z_1 of the "true" chance constraint (3.6) and the feasible set Z_ρ of "ρ times strengthened" version of (3.6). With this approach, the less is ρ_*, the less conservative is the approximation; in particular, the "ideal" situation $\rho_* = 1$ corresponds to no conservatism at all. Note that this approach to measuring the conservatism of (3.5) as an approximation of (3.6) is borrowed from the methodology of Robust Optimization, see, e.g., [3], and was applied to chance constrained problems in [17]. The approach is of the same spirit as the previous one: in both cases, we include the constraint to be approximated in a parametric family and ask by how much should we "strengthen" the original constraint in order for the feasible set of the approximation to be in-between the feasible sets of the constraint of interest and its strengthened version. With the former approach, the family is comprised of all constraints of the form (3.6) with the right hand side β in the role of the parameter; with the latter, the parameterization is via the "noise intensity" ρ.

It turns out that with the second approach to measuring conservatism, Bernstein approximation seems to be not that bad. Here is a result in this direction.

Proposition 1 *Let ξ_j be uniformly distributed in segments $[-\sigma_j, \sigma_j]$, $1 \leq j \leq d$, and let $\beta < 0.05$. Then the conservatism of* (3.5) *is at most*

$$\widehat{\rho} = 1 + 12\sqrt{2\log(1/(0.95\beta))}, \tag{3.8}$$

that is, the feasible set of (3.5) *contains the feasible set of the chance constraint* (3.7) *corresponding to $\rho = \widehat{\rho}$.*

Proof. Without loss of generality we can assume that ξ_j are uniformly distributed in $[-1/2, 1/2]$. Let z be feasible for (3.7) with $\rho = \widehat{\rho}$; we should prove that z is feasible for (3.5). There is nothing to prove when $z_1 = ... = z_d = 0$; assuming that the latter is not the case, we can without loss of generality normalize z to get $z_1^2 + ... + z_d^2 = 1$. Setting $e = (z_1, ..., z_d)^T \in \mathbb{R}^d$, consider the set

$$A = \left\{ y \in \mathbb{R}^d : \widehat{\rho}^{-1} z_0 + e^T y \leq 0 \right\};$$

note that

$$\mathrm{Prob}\{\xi \in A\} \equiv \mathrm{Prob}\left\{ z_0 + \widehat{\rho} \sum_{j=1}^d \xi_j z_j \leq 0 \right\} \geq 1 - \beta. \tag{3.9}$$

In particular,

$$\mathrm{Prob}\left\{ |e^T \xi| > -\widehat{\rho}^{-1} z_0 \right\} = 2\mathrm{Prob}\left\{ z_0 + \widehat{\rho} \sum_{j=1}^d \xi_j z_j > 0 \right\} \leq 2\beta \leq 0.1,$$

whence $\mathrm{Prob}\left\{ |e^T \xi| \leq -\widehat{\rho}^{-1} z_0 \right\} \geq 0.9$. From the latter inequality, by Lemma 2 in [17], it follows that

$$\mu := -\widehat{\rho}^{-1} z_0 \geq \frac{1}{6}. \tag{3.10}$$

12

Now let $D(y) := \min_{y' \in A} \|y - y'\|_2 = \max[0, e^T y + \widehat{\rho}^{-1} z_0] = \max[0, e^T y - \mu]$. Since ξ_j are independent and uniformly distributed on $[-1/2, 1/2]$, from Talagrand Inequality (see, e.g., [12]) it follows that

$$\mathbb{E}\left[\exp\{\frac{D^2(\xi)}{4}\}\right] \leq \frac{1}{\text{Prob}\{A\}} \leq \frac{1}{0.95} \tag{3.11}$$

where the concluding inequality is given by (3.9). Now let

$$\tau = \frac{\mu(\widehat{\rho} - 1)}{2}, \ c = \exp\{-\tau\mu(\widehat{\rho} - 1) + \tau^2\} = \exp\{-\frac{\mu^2(\widehat{\rho} - 1)^2}{4}\}.$$

Then, as it is immediately seen,

$$\exp\{\tau[z_0 + e^T y]\} \leq c \exp\{\frac{D^2(y)}{4}\} \ \forall y \in \mathbb{R}^d,$$

whence, in view of (3.11),

$$\mathbb{E}\left[\exp\{\tau[z_0 + \sum_{j=1}^d \xi_j z_j]\}\right] \leq \frac{c}{0.95} = \frac{1}{0.95}\exp\{-\frac{\mu^2(\widehat{\rho} - 1)^2}{4}\} \leq \frac{1}{0.95}\exp\{-\frac{(\widehat{\rho} - 1)^2}{144}\},$$

where the concluding inequality is given by (3.10). Invoking (3.8), we arrive at

$$\mathbb{E}\left[\exp\{\tau[z_0 + \sum_{j=1}^d \xi_j z_j]\}\right] \leq \beta,$$

so that the expression under $\inf_{t>0}$ in the left hand side of (3.5) is nonpositive when $t = \tau^{-1}$. Thus, z is feasible for (3.5). ■

Note that qualitatively speaking, the level of conservatism of the Bernstein approximation as stated in (3.8) is not that disastrous – it is nearly independent of β. Note also that the assumption that ξ_j are uniformly distributed plays no crucial role, and the results similar to the one of Proposition 1 can be obtained for many other symmetric distributions with bounded support (cf., [17, section 2]). We should add that there is another case when the Bernstein approximation is "nearly accurate" – the one of normally distributed ξ_j, but this is of no interest, since here a scalar chance constraint $\text{Prob}\left\{z_0 + \sum_{j=1}^d \xi_j z_j > 0\right\} \leq \beta$ with $\beta < 1/2$ is exactly equivalent to an explicit convex constraint and thus requires no approximation at all.

4 Upper and lower bounds

In general, the approximation-based approach to processing chance constrained problems requires mechanisms for: (i) measuring the actual risk (reliability) associated with the resulting solution, and (ii) bounding from below the true optimal value Opt^* of the chance constraint problem (1.1). Task (i) corresponds to the case when the approximation is not necessarily conservative, as it is the case, e.g., with the scenario approximation. With the latter, even applied with the theoretically

13

justified sample size (1.3), there still is a chance $1 - \delta$ that the resulting solution $\bar{x}$ does not satisfy the chance constraint, and we would like to check whether the solution indeed is feasible for (1.1). Task (ii) is relevant to basically all approximations, since they usually are conservative ("for sure", as Bernstein approximation, or "with probability close to 1", as the scenario approximation with sample size (1.3)), and a lower bound on Opt* allows to quantify this conservatism.

A straightforward way to measure the actual risk of a given candidate solution $\bar{x} \in X$ is to use Monte Carlo sampling. That is, a sample $\xi^1, ..., \xi^{N'}$ of N' realizations of random vector ξ is generated and the probability $p(\bar{x}) := \text{Prob}\{F(\bar{x}, \xi) \not\leq 0\}$ is estimated as Δ / N', where Δ is the number of times the constraint $F(\bar{x}, \xi^\nu) \leq 0$, $\nu = 1, ..., N'$, is violated. A more reliable upper bound on $p(\bar{x})$ is the random quantity

$$\widehat{\alpha} := \max_{\gamma \in [0,1]} \left\{ \gamma : \sum_{r=0}^{\Delta} \binom{N'}{r} \gamma^r (1 - \gamma)^{N'-r} \geq \delta \right\},$$

where $1 - \delta$ is the required confidence level. The quantity $\widehat{\alpha}$ is, with probability at least $1 - \delta$, an upper bound on $p(\bar{x})$, so that if our experiment results in $\widehat{\alpha} \leq \alpha$, we may be sure, "up to probability of bad sampling $\leq \delta$", that $\bar{x}$ is feasible for (1.1) and $f(\bar{x})$ is an upper bound on Opt*. Since the outlined procedure involves only the calculation of quantities $F(\bar{x}, \xi^\nu)$, it can be performed with a large sample size N', and hence feasibility of $\bar{x}$ can be evaluated with a high reliability, provided that α is not too small (otherwise the procedure would require an unrealistically large sample size).

It is more tricky to bound Opt* from below. Here we propose a bounding scheme as follows. Let us choose three positive integers M, N, L, with $L \leq M$, and let us generate M independent samples $\xi^{1,\mu}, ..., \xi^{N,\mu}$, $\mu = 1, ..., M$, each of size N, of random vector ξ. For each sample we solve the associated optimization problem

$$\text{Min}_{x \in X} f(x) \text{ subject to } F(x, \xi^{\nu,\mu}) \leq 0, \nu = 1, ..., N, \tag{4.1}$$

and hence calculate its optimal value Opt_μ.

We compute the quantities Opt_μ, $\mu = 1, ..., M$, by treating the infeasibility and unboundedness according to the standard optimization conventions: the optimal value of an infeasible optimization problem is $+\infty$, while for a feasible and unbounded from below problem it is $-\infty$. We then rearrange the resulting quantities $\{\text{Opt}_\mu\}_{\mu=1,...,M}$ in the nondescending order: $\text{Opt}_{(1)} \leq ... \leq \text{Opt}_{(M)}$ (in the statistics literature these are called the order statistics of the sample $\{\text{Opt}_\mu\}_{\mu=1,...,M}$). By definition, the lower bound on the true optimal value is the random quantity $\text{Opt}_{(L)}$.

Let us analyze the resulting bounding procedure. Let x be a feasible point of the true problem (1.1). Then x is feasible for problem (4.1) with probability at least $\theta_N = (1 - \alpha)^N$. When x is feasible for (4.1), we of course have $\text{Opt}_\mu \leq f(x)$. Thus, for every $\mu \in \{1, ..., M\}$ and for every $\varepsilon > 0$ we have

$$\theta := \text{Prob}\{\text{Opt}_\mu \leq \text{Opt}^* + \varepsilon\} \geq \theta_N.$$

Now, in the case of $\text{Opt}_{(L)} > \text{Opt}^* + \varepsilon$, the corresponding realization of the random sequence $\text{Opt}_1, ..., \text{Opt}_M$ contains less than L elements which are less than or equal to $\text{Opt}^* + \varepsilon$. Since the elements of the sequence are independent, the probability $\rho(\theta, M, L)$ of the latter event is

$$\rho(\theta, M, L) = \sum_{r=0}^{L-1} \binom{M}{r} \theta^r (1 - \theta)^{M-r}.$$

14

Since $\theta \geq \theta_N$, we have that $\rho(\theta, M, L) \leq \rho(\theta_N, M, L)$.

Thus,

$$\text{Prob}\left\{\text{Opt}_{(L)} > \text{Opt}^* + \varepsilon\right\} \leq \rho(\theta_N, M, L).$$

Since the resulting inequality is valid for all $\varepsilon > 0$, we arrive at the bound

$$\text{Prob}\left\{\text{Opt}_{(L)} > \text{Opt}^*\right\} \leq \sum_{r=0}^{L-1} \binom{M}{r} (1-\alpha)^{Nr} \left[1 - (1-\alpha)^N\right]^{M-r}. \tag{4.2}$$

We have arrived at the following simple result.

Proposition 2 *Given $\delta \in (0,1)$, let us choose positive integers M,N,L in such a way that*

$$\sum_{r=0}^{L-1} \binom{M}{r} (1-\alpha)^{Nr} \left[1 - (1-\alpha)^N\right]^{M-r} \leq \delta. \tag{4.3}$$

Then with probability at least $1 - \delta$, the random quantity $\text{Opt}_{(L)}$ gives a lower bound for the true optimal value Opt^*.

The question arising in connection with the outlined bounding scheme is how to choose M, N, L. Given a desired reliability $1 - \delta$ of the bound *and M and N*, it is easy to specify L: this should be just the largest $L > 0$ satisfying condition (4.3). (if no $L > 0$ satisfying (4.3) exists, the lower bound, by definition, is $-\infty$). We end up with a question of how to choose M and N. For N given, the larger is M, the better. For given N, the "ideal" bound yielded by our scheme as M tends to infinity, is the lower θ_N-quantile of the true distribution of the random variable Opt_1. The larger is M, the better can we estimate this quantile from a sample of M independent realizations of this random variable. In reality, however, M is bounded by the computational effort required to solve M problems (4.1). Note that the effort per problem is larger the larger is the sample size N. We have no definite idea how to choose N. As N grows, the distribution of Opt_1 "goes up" in the sense that $\text{Prob}\{\text{Opt}_1 > a\}$ increases for every a. As a result, every lower θ-quantile of this distribution also increases. If our bound were the lower θ-quantile of the distribution of Opt_1, it would grow (that is, improve) with N. Unfortunately, our bound is the (empirical estimate of) the lower θ_N-quantile of the distribution in question, with θ_N decreasing as N grows, and this decrease shifts the bound down. For the time being, we do not know how to balance these two opposite trends, except for a trivial way to test several values of N and to choose the best (the largest) of the resulting bounds. To keep reliability δ by testing k different values of N, would require to strengthen reliability of every one of the tests, e.g., in accordance with the Bonferroni inequality, by replacing δ in the right hand side of (4.3) with δ/k.

5 Numerical illustration

We are about to present the results of an illustrative experiment. Our major goal is to compare Bernstein approximations with the scenario approach (see Introduction).

Test problem: optimizing Value at Risk. The toy test problem we are about to consider is the following. There are $n + 1$ assets $0, 1, ..., n$ with random returns. The problem is to distribute \$ 1 between the assets in order to maximize the upper $(1 - \alpha)$-th quantile of the total profit (that is, the total return of the resulting portfolio minus the initial capital of \$ 1). The corresponding model is the chance constrained Linear Programming problem:

$$\max_{x \geq 0, \, t} t - 1 \ \text{ subject to } \ \text{Prob}\left\{ t > \sum_{j=0}^{n} r_j x_j \right\} \leq \alpha, \ \sum_{j=0}^{n} x_j \leq 1, \qquad (P_\alpha)$$

where x_j is the capital invested in asset j, and r_j is the return of this asset.

The data we used in our experiment are as follows:

There are $n + 1 = 65$ assets; asset #0 ("money") has deterministic return $r_0 \equiv 1$, while the returns r_i of the remaining 64 "true" assets are random variables with expectations $\mathbb{E}[r_i] = 1 + \rho_i$, where the nominal profits ρ_i vary in $[0, 0.1]$ and grow with i;

- Every r_i, $i \geq 1$, varies in the range $[1 + (1 - 3\theta)\rho_i, 1 + (1 + 3\theta)\rho_i]$, where θ is "uncertainty level", with some degree of dependence between different returns.

The detailed description of the distribution of returns is as follows.
• The true assets are split into two halves: "solid assets" $i = 1, ..., 32$ and "bubbles" $i = 33, ..., 64$. Recall that the expected profits $\rho_i = \mathbb{E}[r_i] - 1$, grow with i, so that solid assets are, at average, less profitable than the bubbles;
• The actual values of the returns are

$$r_i = 1 + \rho_i + \theta[2\eta_i + \sum_{\kappa=1}^{4} \zeta_\kappa P_{i\kappa}]\rho_i, \qquad (5.1)$$

where $\eta_i \sim U(-1, 1)$ is the individual noise in i-th return, and $\zeta_\kappa \sim U(-1, 1)$ are "common factors" affecting all the returns. All "primitive" random variables (64 of η_i's and 4 of ζ_κ's) are independent of each other.
• The "influence coefficients" $P_{i\kappa}$ in (5.1) are deterministic constants positive for $1 \leq i \leq 32$ ("solid assets") and negative for $i > 32$ ("bubbles"), so that common factors have positive correlations with returns of solid assets and negative correlations with returns of bubbles.
• $P_{i\kappa}$ are chosen in such a way that $\sum_\kappa |P_{i\kappa}| = 1$; thus, the total contribution of all common factors into a particular r_i varies within $\pm \theta \rho_i$, while the individual noise in the return varies within $\pm 2\theta \rho_i$.

The experiments were conducted for the value of risk $\alpha = 0.05$ and the uncertainty levels $\theta = 0.2$, $\theta = 0.5$ and $\theta = 1.0$. The reliability $1 - \delta$ for the scenario approximation (see (1.3)) was set to 0.999. Similarly, the reliability of all other simulation-based inferences (like those on actual risks of various solutions, bound on the true optimal value in the chance constrained problem, etc.) was set to 0.999. The results are presented in Table 1; the reader should be aware that we work with a maximization problem, so that the larger is the value of the objective yielded by a method, the better, what was before a lower bound on the optimal value in the chance constrained problem becomes an upper bound, etc.

16

	Quantity	Value	Empirical risk[a]	Inferred risk[a]
$\theta = 0.2$	Nominal optimal value[b]	0.0975	—	—
	Upper bound[c]	0.0673	—	—
	Bernstein optimal value (tuned)[db]	0.0620	0.043	0.0498
	Bernstein optimal value[da]	0.0588	0.008	0.0107
	Scenario optimal value (tuned)[eb]	0.0602	0.044	0.0504
	Scenario optimal value[ea] ($N = 14,900$)	0.0559	0.002	0.0038
	Robust optimal value[f]	0.0435	—	—
$\theta = 0.5$	Nominal optimal value	0.0975	—	—
	Upper bound	0.0621	—	—
	Bernstein optimal value (tuned)	0.0519	0.039	0.0450
	Bernstein optimal value	0.0470	0.006	0.0094
	Scenario optimal value (tuned)	0.0498	0.040	0.0463
	Scenario optimal value ($N = 14,900$)	0.0441	0.003	0.0054
	Robust optimal value	0.0000[g]	—	—
$\theta = 1.0$	Nominal optimal value	0.0975	—	—
	Upper bound	0.0573	—	—
	Bernstein optimal value (tuned)	0.0406	0.043	0.0493
	Bernstein optimal value	0.0330	0.007	0.0102
	Scenario optimal value (tuned)	0.0369	0.039	0.0456
	Scenario optimal value ($N = 14,900$)	0.0287	0.004	0.0059
	Robust optimal value	0.0000	—	—

Table 1. Results of experiments with the Value at Risk model.

Explanations: [a] Empirical risk makes sense only with respect to the optimal values yielded by various methods and is the empirical frequency estimate, taken over 10,000 simulations, of the probability p of violating the randomly perturbed constraint in $(P_{0.05})$ at the solution yielded by the method. Inferred risk is the 0.999-reliable upper bound on p, as inferred from the same 10,000 simulations.

[b] Nominal optimal value is the optimal value in the nominal problem – the one where all randomly perturbed coefficients are set to their expected values. It is immediately seen that for symmetrically distributed ξ (as is the case in our test problem), the nominal optimal value is an upper (we are in the case of maximization!) bound on the "chance constrained" optimal value, provided that $\alpha < 0.5$.

[c] See section 4; since $(P_{0.05})$ is a maximization problem, the corresponding construction yields an upper bound on the optimal value in $(P_{0.05})$. The reliability of the bound is 0.999.

[da] Bernstein optimal value is the optimal value in Bernstein approximation (3.4) of $(P_{0.05})$.

[db] Tuned Bernstein optimal value is the optimal value in "tuned" Bernstein approximation. Tuning, aimed to overcome, to some extent, the intrinsic conservativeness of the approximation as given by (3.4), goes via replacing in (3.4), (2.16) the required value α of risk by a larger value, α^+, still resulting in a solution satisfying the actual chance constraint (the latter is verified by simulations, the reliability of the verification being 0.999). Under this restriction on α^+, we try to make α^+ as large as possible, since the larger is α^+, the better is the optimal value in the corresponding Bernstein approximation.

[ea] Scenario optimal value is the optimal value in the scenario approximation (P^N) of $(P_{0.05})$, the sample

size N being chosen according to (1.3) (where $n = 66$, $\alpha = 0.05$ and $\delta = 0.001$).

$^{e_b)}$Tuned scenario optimal value is the optimal value in "tuned" scenario approximation, where the tuning goes via appropriate reduction of the sample size N in (P^N) as compared to the one given by (1.3), with the same purpose and in the same spirit as in the case of tuned Bernstein approximation.

$^{f)}$The robust optimal value is the one given by Robust Optimization (under mild regularity assumptions, which hold true in the case of (P), this is the same as the optimal value in (P_α) in the case of $\alpha = 0$).

$^{g)}$For uncertainty level $\theta = 0.5$ and larger, there already is no guaranteed way to get positive profit, and the worst-case optimal policy is to put everything in money, thus avoiding losses and getting zero profit.

Discussion. **A.** As far as the objective value is concerned, Bernstein approximation outperforms the (non-tuned) scenario approximation; the same is true for the tuned versions of the procedures (this is consisted with all other numerical experiments we have run, including those for test problems of different structure). The differences, although not large, are not negligible: for tuned approximations, they are 3.0% for $\theta = 0.2$, 4.2% for $\theta = 0.5$, and 10.0% for $\theta = 1.0$.

B. Additional good news about Bernstein approximation is that even with tuning, this still is an implementable routine: the solution and the optimal value in (3.4), (2.16) are well-defined functions of α, and the resulting value of the objective is the better the larger is α. Consequently, tuning becomes an easy-to-implement routine, a kind of bisection: we solve (3.4), (2.16) for certain value of α and check the actual risk of the resulting solution; if it is worse then necessary, we decrease α in (3.4), otherwise increase it. In contrast to this, the optimal value and the optimal solution of scenario approximation with a given sample size are random. For not too large sample sizes, the variability of these random entities is high, which makes tuning difficult. To get an impression of this phenomenon, here are the results for several realizations of the scenario approximation at the uncertainty level $\theta = 1$; the sample sizes correspond to the final phase of tuning.

##	N	Optimal value	Empirical risk	Inferred risk
1	900	0.0066	0.000	0.0007
2	900	0.0380	0.045	0.0517
3	900	0.0384	0.052	0.0595
4	900	0.0017	0.000	0.0007
5	900	0.0379	0.052	0.0591
6	900	0.0393	0.065	0.0726
7	900	0.0933	1.000	1.0000
8	1000	0.0377	0.048	0.0546
9	1000	0.0369	0.039	0.0456
10	1000	0.0378	0.054	0.0609

We see that with sample size 900, one can get "nearly whatever", with profits ranging from something very close to 0 to nearly the nominal optimal value, and with similarly wide range of risks; sample size 1000 may produce more risky solutions than sample size 900, etc. As the result, seeking for good and reliable solution by tuning the scenario approximation becomes a kind of lottery.

C. For large sample sizes, the results yielded by scenario approximation are more stable, and at the same time more conservative (compare the results for non-tuned Bernstein and scenario approximations in Table 1).

18

D. It should be added that Bernstein approximation (at least in its non-tuned form) remains practical in the case of very small risks α and/or high design dimension, that is, in situations where the scenario approximation requires samples of unrealistic sizes. To get an impression of the numbers, assume that we want α as small 0.5% or even 0.1%, while the reliability $1 - \delta$ of our conclusions (which in previous experiments was set to 0.999) is now increased to 0.9999. In this case the scenario approximation becomes completely impractical. Indeed, the theoretically valid sample size given by (1.3) becomes 209,571 for $\alpha = 0.5\%$ and 1,259,771 for $\alpha = 0.1\%$, which is a bit too much. Using smaller sample sizes plus tuning also is problematic, since it becomes too complicated to test the risk of candidate solutions by simulation. For example, with $\alpha = 0.005$ and $\alpha = 0.001$, it takes over 100,000 simulations to conclude, with reliability 0.9999, that a given candidate solution which in fact is feasible for $(P_{0.9\alpha})$ is feasible for (P_α).

• At the same time, Bernstein approximation with no tuning is 100% reliable, remains of the same complexity independently of how small is α, and at the uncertainty level 0.5 results in the profits as follows:

    • 0.0421 for $\alpha = 0.5\%$

    • 0.0395 for $\alpha = 0.1\%$.

This is not that bad, given that the robust optimal value in our situation is 0.

The bottom line, as suggested by the experiments (and us such, not conclusive yet) is as follows:

> The scenario approximation has no advantages whatsoever as compared to the Bernstein one, *provided the latter is applicable* (that is, that we are in the case of affinely perturbed convex chance constrained problem with known and simple enough distributions of ξ_j).

6 Extensions to the case of ambiguous chance constraints and mixed uncertainty models

As it was mentioned in Introduction, one of the basic problems with the formulation of chance constrained problem (1.1) is that it assumes an exact knowledge of the underlying probability distribution P of ξ. Therefore it appears natural to consider "robust" or minimax versions of the chance constrained problems. That is, we assume that a plausible family $\mathfrak{P}$ of probability distributions, supported on a (closed) set $\Xi \subset \mathbb{R}^d$, can be identified, and that we only know that the true distribution P belongs to $\mathfrak{P}$. This leads to the following formulation of chance constrained problems:

$$\underset{x \in X}{\text{Min}}\, f(x) \quad \text{subject to} \quad \text{Prob}_P\big\{F(x,\xi) \leq 0\big\} \geq 1 - \alpha, \quad \forall P \in \mathfrak{P}, \tag{6.1}$$

where the notation Prob_P means that the probability of the corresponding event is taken with respect to the distribution P of ξ.

Of course, we can replace the probability constraints in (6.1) with one constraint by taking the minimum of $\text{Prob}_P\big\{F(x,\xi) \leq 0\big\}$ with respect to $P \in \mathfrak{P}$. That is, problem (6.1) is constrained with respect to a "worst" distribution of the considered family $\mathfrak{P}$. We can also write the probability constraints of (6.1) in the following form:

$$\sup_{P \in \mathfrak{P}} \mathbb{E}_P\left[\mathbb{1}_{A_x}\right] \leq \alpha, \tag{6.2}$$

19

where $A_x := \{\xi \in \Xi : F(x, \xi) \not\preceq 0\}$. The "worst-case-distribution" (or minimax) stochastic programming problems were considered in a number of publications (e.g., [10, 25]). When applied to chance constraints, such worst-case-distribution problems are called *ambiguous* chance constrained problems (see [11] and references therein).

For some families of distributions the maximum in the left hand side of (6.2) can be calculated explicitly. With every family $\mathfrak{P}$ of probability distributions is associated the function

$$\rho(Z) := \sup_{P \in \mathfrak{P}} \mathbb{E}_P [Z] \tag{6.3}$$

defined on a space of real-valued random variables Z. Formula (6.3) describes a dual representation of so-called *coherent risk measures* introduced by Artzner et al [1]. Consider now the following family:

$$\mathfrak{P} := \{P : \gamma_1 P^* \preceq P \preceq \gamma_2 P^*, \; P(\Xi) = 1\}. \tag{6.4}$$

Here γ_1 and γ_2 are constants such that $0 \leq \gamma_1 \leq 1 \leq \gamma_2$, P^* is a (reference) probability distribution on Ξ and the notation $P_1 \preceq P_2$ means that for two (not necessarily probability) Borel measures P_1 and P_2 on Ξ it holds that $P_1(A) \leq P_2(A)$ for any Borel set $A \subset \Xi$. The constraint $P(\Xi) = 1$ in (6.3) is written to ensure that P is a probability measure. This family $\mathfrak{P}$ defines a coherent risk measure, which can be written in the following equivalent form

$$\rho(Z) = \mathbb{E}[Z] + \inf_{\tau \in \mathbb{R}} \mathbb{E}\left[(1 - \gamma_1)[\tau - Z]_+ + (\gamma_2 - 1)[Z - \tau]_+\right], \tag{6.5}$$

where all expectations are taken with respect to the reference distribution P^*. In particular, for $\gamma_1 = 0$ and $\kappa := (\gamma_2 - 1)/\gamma_2$,

$$\rho(Z) = \mathrm{CVaR}_\kappa[Z]$$

(cf., [21, 22]).

By the definition (6.4) of $\mathfrak{P}$ we have that $\mathbb{E}_P[\mathbb{1}_{A_x}] \leq \gamma_2 P^*(A_x)$ for any $P \in \mathfrak{P}$, with the equality holding if $P(A_x) = \gamma_2 P^*(A_x)$. Since $P(\Xi) = 1$, this can be achieved iff $\gamma_2 P^*(A_x) + \gamma_1(1 - P^*(A_x)) \leq 1$, i.e., iff $P^*(A_x) \leq \frac{1 - \gamma_1}{\gamma_2 - \gamma_1}$. We obtain the following.

> If $\alpha \leq (1 - \gamma_1)/(\gamma_2 - \gamma_1)$, then the ambiguous chance constrained problem (6.1) with $\mathfrak{P}$ given by (6.4) is equivalent to the chance constrained problem (1.1) with respect to the reference distribution P^* and with rescaled risk $\alpha \leftarrow \alpha^* := \alpha/\gamma_2$.

Another popular example of a coherent risk measure is the mean-upper-absolute semideviation

$$\rho(Z) := \mathbb{E}[Z] + c\mathbb{E}\left(\left[Z - \mathbb{E}[Z]\right]_+\right), \tag{6.6}$$

where $c \in [0, 1]$ is a constant and the expectations are taken with respect to a reference distribution P^*. It has the dual representation (6.3) with the corresponding family

$$\mathfrak{P} = \{\zeta' : \zeta' = 1 + \zeta - \mathbb{E}[\zeta], \; \|\zeta\|_\infty \leq c\}, \tag{6.7}$$

where $\zeta' = dP/dP^*$ denotes the density of P with respect to P^* (cf., [22]). By using the definition (6.6) it is straightforward to calculate that

$$\rho\left(\mathbb{1}_{A_x}\right) = P^*(A_x) + 2cP^*(A_x)(1 - P^*(A_x)). \tag{6.8}$$

By solving the quadratic inequality $t + 2ct(1-t) \leq \alpha$ for $t = P^*(A_x)$, we obtain that $P^*(A_x) \leq \varphi(\alpha)$, where

$$\varphi(\alpha) := \frac{1 + 2c - \sqrt{1 + 4c(1 - 2\alpha) + 4c^2}}{4c}$$

for $c \in (0, 1]$, and $\varphi(\alpha) = \alpha$ if $c = 0$. (Note that for $\alpha \in (0, 1)$ and $c \in (0, 1]$, it always holds that $\varphi(\alpha) \in (0, \alpha)$.) We obtain the following.

> The ambiguous chance constrained problem (6.1) with $\mathfrak{P}$ given by (6.7) is equivalent to the chance constrained problem (1.1) with respect to the reference distribution P^* and with rescaled reliability parameter $\alpha \leftarrow \alpha^* := \varphi(\alpha)$.

Of course, such explicit reduction of the ambiguous chance constrained problem (6.1) to the regular chance constrained problem (1.1) is possible only for some specific families $\mathfrak{P}$. In the next section we discuss a different approach which can be viewed as a combination of chance constraints and robust approaches.

6.1 Mixed uncertainty model

Below we extend Bernstein approximation to the case of a more general chance constrained problem than the affinely perturbed convex one introduced in section 3.

Consider an uncertain optimization problem where our goal is to minimize a convex objective $f(x)$ over a part cut off a given convex set X by a system of constraints (cf. (1.7))

$$f_{i0}(x) + \sum_{j=1}^{d} \zeta_j f_{ij}(x) \leq 0, \ i = 1, ..., m, \tag{6.9}$$

where ζ_j are uncertain coefficients. In contrast to the situation of section 3, where the coefficients were assumed to be independent random variables with known distributions, now we intend to consider the *mixed* uncertainty model where "the nature" generates vector $\zeta = (\zeta_1, ..., \zeta_d)$ of uncertain coefficients in two stages:

- At the first stage, a realization $\bar{\xi}$ of d-dimensional random vector ξ is generated;

- At the second stage, a point η from a given uncertainty set $\mathcal{U} \subset \mathbb{R}^d$ is picked, and ζ is formed as $\zeta = \bar{\xi} + \eta$.

Note that we do *not* assume that the distribution P of the random component ξ in ζ is known exactly. All that we assume is that P belongs to a known family $\mathfrak{P}$ of probability distributions on $\mathbb{R}^d$. As about the uncertainty set $\mathcal{U}$, it is assumed to be a known in advance nonempty compact set.

Let us understand what could be a natural definition of a "chance constrained" version of the uncertain problem in question. Let $\bar{x} \in X$ be a fixed candidate solution to the problem. To pass from problem with uncertainty-affected constraints (6.9) to its usual chance constrained version (corresponding to the case when $\mathcal{U} = \{0\}$ and the distribution of $\zeta \equiv \xi$ is known), we proceed as follows.

(a) We treat a realization $\bar{\xi}$ of vector ξ of random perturbations as "compatible" with $\bar{x}$, if $\bar{x}$ satisfies all the constraints (6.9) with the vectors of coefficients ζ corresponding to $\bar{\xi}$ (in the situation in question, the latter merely means that $\zeta = \bar{\xi}$).

(b) We claim $\bar{x}$ to be feasible for the chance constrained version of our uncertain problem, if the probability mass of realizations $\bar{\xi}$ compatible with $\bar{x}$ is at least $1 - \alpha$.

Let us follow the same line of reasoning in our current situation. First of all, we should understand when a particular realization $\bar{\xi}$ of ξ is compatible with a given candidate solution $\bar{x}$. Exactly as above, let us interpret the compatibility as the fact that $\bar{x}$ is feasible for the system of constraints (6.9), whenever ζ corresponds to $\bar{\xi}$, that is, whenever $\zeta = \bar{\xi} + \eta$ with $\eta \in \mathcal{U}$ (since we assume no random mechanism for picking η from $\mathcal{U}$, this worst-case-oriented version of compatibility between $\bar{x}$ and $\bar{\xi}$ seems to be the only meaningful one). With our interpretation of compatibility, it is immediately seen that $\bar{x}$ and $\bar{\xi}$ are compatible if and only if $\bar{x}$ satisfies the system of constraints

$$\max_{\eta \in \mathcal{U}} \left\{ f_{i0}(x) + \sum_{j=1}^{d} \left(\bar{\xi}_j + \eta_j \right) f_{ij}(x) \right\} \leq 0, \, i = 1, ..., m, \tag{6.10}$$

or, which is exactly the same, if and only if $\bar{x}$ is feasible for the system of constraints

$$\widehat{f}_{i0}(x) + \sum_{j=1}^{d} \bar{\xi}_j f_{ij}(x) \leq 0, \, i = 1, ..., m, \tag{6.11}$$

where

$$\widehat{f}_{i0}(x) := f_{i0}(x) + \max_{\eta \in \mathcal{U}} \sum_{j=1}^{d} \eta_j f_{ij}(x). \tag{6.12}$$

Now, if we knew the true distribution P of ξ, we, same as in (b), could qualify $\bar{x}$ as a feasible solution of the chance constrained version of our uncertain problem exactly when

$$\mathrm{Prob}_P \left\{ \widehat{f}_{i0}(\bar{x}) + \sum_{j=1}^{d} \xi_j f_{ij}(\bar{x}) \leq 0, \, i = 1, ..., m \right\} \geq 1 - \alpha.$$

The only meaningful way to extend the latter step to the case when all our knowledge of P is that $P \in \mathfrak{P}$ is to invoke again the worst-case approach, that is, to require from $\bar{x}$ to satisfy the latter inequality for every $P \in \mathfrak{P}$. We arrive at the following *definition* of a chance constrained version of uncertain problem with mixed uncertainty:

$$\min_{x \in X} f(x) \text{ s.t. } \inf_{P \in \mathfrak{P}} \mathrm{Prob}_P \left\{ \xi : \widehat{f}_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x) \leq 0, \, i = 1, ..., m \right\} \geq 1 - \alpha, \tag{6.13}$$

where $\widehat{f}_{i0}(x)$ is defined by (6.12).

Remark 4 The construction we have presented is a straightforward mixture of two ingredients known from the literature. The first ingredient is "chance constrained optimization with ambiguous chance constraints" (see, e.g., [11]), that is, the chance constrained problem with not known exactly distribution of uncertain parameters, where the probabilities with respect to (unknown) "true" distribution of the uncertain data are replaced with their worst case values at a given set of "potentially true" distributions. The second ingredient is Robust Optimization (see, e.g., [3]); this

22

is the case with no random component in the uncertain data (formally, $\mathfrak{P} := \{P\}$ is a singleton, and P is the trivial distribution on $\mathbb{R}^d$ – unit mass at the origin). In this case, all probabilities disappear, and (6.13) becomes the problem

$$\min_{x \in X} f(x) \text{ s.t. } \widehat{f}_{i0}(x) \le 0, \ i = 1, ..., m,$$

called the Robust Counterpart of the original uncertain problem.

6.2 Bernstein approximation of a chance constrained problem with mixed uncertainty model

We are about to build Bernstein approximation of problem (6.13).

Assumptions. From now on, we make the following assumptions about the "data" of (6.13):

B1. Uncertainty set $\mathcal{U}$ is given in advance and is convex, compact, nonempty and "computationally tractable". In our context, the latter means that we can optimize efficiently linear objectives over $\mathcal{U}$ (for example, $\mathcal{U}$ can be a nonempty compact polyhedral set: $\{\eta : \exists u : A\eta + Bu + r \le 0\}$).

B2. The family $\mathfrak{P}$ of possible distributions of the random component ξ in ζ is as follows. Let D_j, $j = 1, ..., d$, be nonempty compact subsets of the axis, and $\mathcal{M}$ be a nonempty set of tuples $\{P_j\}_{j=1}^d$, where P_j are Borel probability measures on D_j. We assume that

• whenever $\{P_j\}_{j=1}^d$, $\{P'_j\}_{j=1}^d$ are two elements from $\mathcal{M}$, so is $\{\lambda P_j + (1-\lambda)P'_j\}_{j=1}^d$, $\lambda \in [0,1]$ (convexity), and

• whenever a sequence $\{P_j^t\}_{j=1}^d$, $t = 1, 2, ...$ of elements of $\mathcal{M}$ w^*-converges to $\{P_j\}_{j=1}^d$ (meaning that $\int f(s)dP_j^t(s) \to \int f(s)dP_j(s)$ as $t \to \infty$ for every j and every continuous on the axis function f with compact support), then $\{P_j\}_{j=1}^d \in \mathcal{M}$ (w^*-closedness).

We assume that $\mathfrak{P}$ is comprised of all product distributions $P = P_1 \times ... \times P_d$ on $\mathbb{R}^d$ with the tuple of marginals $\{P_j\}_{j=1}^d$ running through a given set $\mathcal{M}$ with the outlined properties.
¿From now on, we equip the set $\mathcal{M}$ underlying, via the outlined construction, the set $\mathfrak{P}$ in question with the w^*-topology; it is well known that under the above assumptions this topology is yielded by an appropriate metrics on $\mathcal{M}$, and that with this metrics, $\mathcal{M}$ is a compact metric space.

The simplest example of a set $\mathfrak{P}$ of the outlined structure is as follows. Let D_j be finite subsets of $\mathbb{R}$, let $\Delta := \bigcup_{j=1}^d D_j = \{s_1, ..., s_K\}$, and let $\mathcal{M}$ be a closed and convex set of matrices $P = [p_{kj}]_{\substack{1 \le k \le K \\ 1 \le j \le d}}$ with nonnegative entries such that $\sum_k p_{kj} = 1$ for all j and $p_{kj} = 0$ whenever $s_k \notin D_j$. For every $P \in \mathcal{M}$, j-th column P_j of P can be naturally identified with a probability distribution on D_j; the set $\mathfrak{P}$ generated by $\mathcal{M}$ is comprised of all product distributions $P_1 \times ... \times P_d$ coming from matrices $P \in \mathcal{M}$.

¿From now on, we denote a generic element of $\mathcal{M}$ by $Q = \{Q_j\}_{j=1}^d$.

23

B3. The objective $f(x)$ and all functions $f_{ij}(x)$, $i = 1,...,m$, $j = 0, 1,...,d$ are convex and well-defined on X. Moreover, let

$$J = \{j : 1 \leq j \leq d, \text{ not all functions } f_{ij}, i = 1,...,m, \text{ are affine}\}.$$

We assume that whenever $j \in J$, the quantities ξ_j and η_j "always are nonnegative", that is, for every $j \in J$

— j-th marginal distribution of every $P \in \mathfrak{P}$ is supported on the nonnegative ray, and

— all points $\eta \in \mathcal{U}$ satisfy $\eta_j \geq 0$

(cf. assumption A3 in section 3).

Building Bernstein approximation. Let

$$g(r) = \max_{\eta \in \mathcal{U}} \eta^T r,$$

so that $g(\cdot)$ is a well-defined convex function by B1. Now, let J be the set of those indices j, $1 \leq j \leq d$, for which not all of the functions $f_{ij}(x)$, $i = 1,...,m$, are affine. By B3, the function g is nondecreasing in every one of the coordinates r_j with $j \in J$. From these observations and B1, B3 it follows that the functions $h_i(x)$ (and then $\widehat{f}_{i0}(x)$) given by (6.11) are convex and well-defined on X.

For $P = P_1 \times ... \times P_d$, let $\widehat{P}$ be the tuple $\{P_j\}_{j=1}^d$, so that when P runs trough $\mathcal{P}$, $\widehat{P}$ runs through $\mathcal{M}$.

Let

$$
\begin{aligned}
\Phi(z, Q) &:= \log \left(\mathbb{E}_{Q_1 \times ... \times Q_d} \left[\exp\{z_0 + \sum_{j=1}^d \xi_j z_j\} \right] \right) \\
&= z_0 + \sum_{j=1}^d \log \left(\int \exp\{z_j s\} dQ_j(s) \right), \quad Q = \{Q_j\}_{j=1}^d \in \mathcal{M}, \\
\widehat{\Phi}(z) &:= \max_{Q \in \mathcal{M}} \Phi(z, Q).
\end{aligned}
\tag{6.14}
$$

By B2, $\Phi(z, Q)$ is well-defined and continuous function of $(z, Q) \in \mathbb{R}^{d+1} \times \mathcal{M}$ (recall that $\mathcal{M}$ is equipped with w^*-topology). From (6.14) it is also evident that $\Phi(z, Q)$ is convex in $z \in \mathbb{R}^{d+1}$ and concave in $Q \in \mathcal{M}$. From these observations and the compactness of $\mathcal{M}$ it follows that $\widehat{\Phi}(z)$ is well defined everywhere and is convex. Finally, from B3 it follows that $\Phi(z, Q)$ (and therefore $\widehat{\Phi}(z)$) is nondecreasing in z_0 and in every z_j with $j \in J$.

Now let

$$\Theta_Q(z, t) = t\Phi_Q(t^{-1}z), \ \widehat{\Theta}(z, t) = t\widehat{\Phi}(t^{-1}z),$$

so that $\Theta_Q(z, t)$ and $\widehat{\Theta}(z, t)$ are well defined convex functions in the domain $t > 0$. Same as in section 3, for every $\beta \in (0, 1)$ and every $z \in \mathbb{R}^{d+1}$ we have

$$\inf_{t > 0} \left[\Theta_{\widehat{P}}(z, t) - t\log\beta \right] \leq 0 \text{ implies } \text{Prob}_P \left\{ z_0 + \sum_{j=1}^d \xi_j z_j > 0 \right\} \leq \beta,$$

24

and we arrive at the implication:

$$P(\beta): \qquad \left\{\forall Q \in \mathcal{M} : \inf_{t>0} \left[\Theta_Q(z,t) - t \log \beta\right] \leq 0\right\}$$

$$\text{implies that}$$

$$Q(\beta): \qquad \sup_{P \in \mathfrak{P}} \text{Prob}_P \left\{z_0 + \sum_{j=1}^d \xi_j z_j > 0\right\} \leq \beta. \tag{6.15}$$

We are about to replace (6.15) with an equivalent and more convenient computationally implication:

$$\widehat{P}(\beta): \qquad \left\{\inf_{t>0} \left[\widehat{\Theta}(z,t) - t \log \beta\right] \leq 0\right\}$$

$$\text{implies that}$$

$$Q(\beta): \qquad \sup_{P \in \mathfrak{P}} \text{Prob}_P \left\{z_0 + \sum_{j=1}^d \xi_j z_j > 0\right\} \leq \beta. \tag{6.16}$$

The advantage of (6.16) as compared to (6.15) is that the premise in the latter implication is semi-infinite: to verify its validity, we should check certain condition for every $Q \in \mathcal{M}$. In contrast to this, the premise in (6.16) requires checking validity of a univariate convex inequality, which can be done by bisection, provided that the function $\widehat{\Theta}$ is efficiently computable. The latter condition is equivalent to efficient computability of the function $\widehat{\Phi}(z)$, which indeed is the case when $\mathcal{M}$ is not too complicated (e.g., is finite-dimensional and computationally tractable).

The validity of (6.16) and the equivalence of (6.15) and (6.16) are given by the following lemma.

Lemma 1 *Let $0 < \beta < 1$. Then the following holds:*

$$\widehat{P}(\beta) \text{ if and only if } P(\beta). \tag{6.17}$$

Proof. Implication $\Rightarrow$ in (6.17) is evident, since $\widehat{\Theta}(z,t) = \max_{Q \in \mathcal{M}} \Theta_Q(z,t)$. Note that this implication combines with (6.15) to imply the validity of (6.16).

Now let us prove the implication $\Leftarrow$ in (6.17). This is a straightforward consequence of the fact that $\Theta_Q(z,t)$ is concave in Q and convex in $t > 0$; for the sake of completeness, we present the corresponding standard reasoning.

As we remember, $\Phi(z,Q)$ is continuous and concave in $Q \in \mathcal{M}$; since $\Theta_Q(z,t) = t\Phi(t^{-1}z,Q)$, the function $\Theta_Q(z,t)$ is continuous in $(t > 0, Q \in \mathcal{M})$ and concave in Q; the fact that this function is convex in $t > 0$ is already known to us. Now let $P(\beta)$ be valid, and let us prove the validity of $\widehat{P}(\beta)$. Let us fix z and set $\theta(t,Q) = \Theta_Q(z,t) - t \log \beta$, and let $\gamma > 0$. By $P(\beta)$, for every $Q \in \mathcal{M}$ there exists $t_Q > 0$ such that $\theta(t,Q) < \gamma$. Since $\theta(t,Q)$ is continuous in $Q \in \mathcal{M}$, there exists a neighborhood (in $\mathcal{M}$) V_Q of the point Q such that $\theta(t_Q, Q') \leq \gamma \, \forall Q' \in V_Q$. Since $\mathcal{M}$ is a compact set, there exist finitely many points $Q^i \in \mathcal{M}$ such that the corresponding neighborhoods V_{Q^i} cover the entire $\mathcal{M}$. In other words, there exist finitely many positive reals $t_1, ..., t_N$ such that

$$\min_{1 \leq i \leq N} \theta(t_i, Q) \leq \gamma, \quad \forall Q \in \mathcal{M}. \tag{6.18}$$

Since θ is concave and continuous in $Q \in \mathcal{M}$ and $\mathcal{M}$ is convex, (6.18) implies that

$$\exists \lambda^* \in \Delta_N := \left\{\lambda \in \mathbb{R}_+^N : \sum_i \lambda_i = 1\right\} : \sum_i \lambda_i^* \theta(t_i, Q) \leq \gamma, \quad \forall Q \in \mathcal{M}. \tag{6.19}$$

25

The latter conclusion is a standard fact of Convex Analysis. For the sake of a reader uncomfortable with possible infinite dimension of $\mathcal{M}$, here is a derivation of this fact from the standard von Neumann lemma. For $Q \in \mathcal{M}$, let Λ_Q be the set of those $\lambda \in \Delta_N$ for which $\sum_i \lambda_i \theta(t_i, Q) \leq \gamma$; the set Λ_Q clearly is a closed subset of the finite-dimensional compact Δ_N. All we need is to prove that all these sets have a point in common (such a point can be taken as λ^*), and to this end it suffices to prove that all sets Λ_Q from a finite family $\Lambda_{Q_1}, ..., \Lambda_{Q_M}$, $Q_j \in \mathcal{M}$, have a point in common. But the latter is readily given by the von Neumann Lemma as applied to the convex hull Q_N of the points Q_j, $j = 1, ..., M$ (which is a finite-dimensional convex compact set):

$$\gamma \geq \max_{Q \in Q_N} \min_{\lambda \in \Delta_N} \sum_{i=1}^N \lambda_i \theta(t_i, Q) = \min_{\lambda \in \Delta_N} \max_{Q \in Q_N} \sum_{i=1}^N \lambda_i \theta(t_i, Q)$$

(the inequality is given by (6.18), the equality – by von Neumann Lemma; the required point in $\bigcap_i \Lambda_{Q_i}$ is $\operatorname*{argmin}_{\lambda \in \Delta_N} \max_{Q \in Q_N} \sum_{i=1}^N \lambda_i \theta(t_i, Q)$).

Since θ is convex in $t > 0$, setting $t_\gamma = \sum_i \lambda_i^* t_i$ we get from (6.19) that $\Theta_Q(t_\gamma, z) - t_\gamma \log \beta \equiv \theta(t_\gamma, Q) \leq \sum_i \lambda_i^* \theta(t_i, Q) \leq \gamma$ for all $Q \in \mathcal{M}$, whence $\widehat{\Theta}(t_\gamma, z) - t_\gamma \log \beta \equiv \max_{Q \in \mathcal{M}} \Theta_Q(t_\gamma, z) - t_\gamma \log \beta \leq \gamma$. Since t_γ is positive by construction and $\gamma > 0$ is arbitrary, we conclude that $\inf_{t>0} \left[\widehat{\Theta}(t_\gamma, z) - t_\gamma \log \beta \right] \leq q0$, so that $\widehat{P}(\beta)$ is valid. ∎

Putting things together, we arrive at the following result.

Theorem 1 *Assume that chance constrained problem* (6.13) *with mixed uncertainty satisfies Assumptions* B1 – B3, *and let* α_i, $i = 1, ..., m$, *be positive reals such that* $\sum_i \alpha_i \leq \alpha$. *Then the program*

$$\operatorname*{Min}_{x \in X} f(x) \text{ s.t. } \underbrace{\inf_{t > 0} \left[\widehat{f}_{i0}(x) + t \widehat{\Psi}(t^{-1} z^i[x]) - t \log \alpha_i \right]}_{g_i(x,t)} \leq 0, \ i = 1, ..., m,$$

$$z^i[x] = (f_{i1}(x), ..., f_{id}(x)), \ \widehat{f}_{i0} = f_{i0}(x) + \max_{\eta \in \mathcal{U}} \eta^T z^i[x], \tag{6.20}$$

$$\widehat{\Psi}(z) = \max_{\{Q_j\}_{j=1}^d \in \mathcal{M}} \sum_{j=1}^d \log \left(\int \exp\{z_j s\} dQ_j(s) \right)$$

is a conservative approximation of problem (6.13): *every feasible solution to the approximation is feasible for the chance constrained problem. This approximation is a convex program and is efficiently solvable, provided that all f_{ij} and $\widehat{\Psi}$ are efficiently computable, and $\mathcal{U}$ and X are computationally tractable.*

Proof. Function $g_i(x, t)$ is obtained from the function $\theta_i(z, t) := \widehat{\Theta}(z, t) - t \log \alpha_i$ by the substitution

$$(z, t) \leftarrow ((\widehat{f}_{i0}(x), f_{i1}(x), ..., f_{id}(x)), t).$$

The outer function $\theta_i(z, t)$ is convex and nondecreasing in z_0 and every z_j with $j \in J$ (see remarks following (6.14)). The inner functions $\widehat{f}_{i0}(x)$, $f_{ij}(x)$, $j \geq 1$, are convex on X, and functions $f_{ij}(x)$ with $0 < j \notin J$ are affine. It follows that $g_i(x, t)$ is convex in $(t > 0, x \in X)$, so that (6.20) is indeed

26

a convex program. Further, if x is feasible for (6.20), then $x \in X$, and for every i the predicate $\widehat{P}(\alpha_i)$ corresponding to $z = (\widehat{f}_{i0}(x), f_{i1}(x), ..., f_{id}(x))$ is valid, which, by (6.16), implies that

$$\sup_{P \in \mathcal{P}} \mathrm{Prob}_P \left\{ \widehat{f}_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x) > 0 \right\} \le \alpha_i.$$

Since $\sum\limits_i \alpha_i \le \alpha$, x is feasible for (6.13). ∎

Remark 5 Assumption B2 requires, among other things, from all distributions $P \in \mathfrak{P}$ to be supported on a common *compact* set $D_1 \times ... \times D_d$. This requirement can be straightforwardly relaxed to the requirement for all $P \in \mathfrak{P}$ to have "uniformly light tails": there exists a function $\gamma(t)$, $t > 0$, such that $\exp\{\alpha t\}\gamma(t) \to 0$ as $t \to \infty$ for all α, and for every $Q = \{Q_j\} \in \mathcal{M}$, every j and every $t > 0$ one has $Q_j(\{s : |s| \ge t\}) \le \gamma(t)$.

Examples. We have seen that the presence of "uncertain but bounded" component $\eta \in \mathcal{U}$ in the vector of uncertain coefficients in (6.9) does not affect the structure of the chance constrained problem (6.13) (to take into account the η-component of the uncertainty, we should just augment properly the "deterministic parts" $f_{i0}(x)$ of the constraints). By this reason, in the examples to follow we focus on the case of "purely (ambiguous) stochastic" uncertainty, where $\mathcal{U} = \{0\}$. Besides this, in order not to care of nonnegativity of ξ_j associated with non-affine $f_{ij}(\cdot)$, we assume from now on that all functions f_{ij}, $j = 1, ..., d$, are affine.

Example 1: range information on ξ_j. Assume that all we know about the distributions of ξ is that ξ_j take values in given finite segments (and, as always, that $\xi_1, ..., \xi_d$ are independent). By shifting and scaling $f_{ij}(x)$, we can assume without loss of generality that ξ_j are independent and take values in $[-1, 1]$. This corresponds to the case where $\mathcal{M}$ is the set of all d-element tuples of Borel probability distributions supported on $[-1, 1]$. Denoting by Π the set of all Borel probability measures on $[-1, 1]$, we have

$$\begin{aligned}
\widehat{\Phi}(z) &= z_0 + \sum_{j=1}^{d} \max_{P_j \in \Pi} \log\left(\int \exp\{z_j s\} dP_j(s)\right) = z_0 + \sum_{j=1}^{d} |z_j|, \\
\widehat{\Theta}(z, t) &= t\widehat{\Phi}(t^{-1}z) = z_0 + \sum_{j=1}^{d} |z_j|;
\end{aligned}$$

consequently, approximation (6.20) becomes

$$\mathop{\mathrm{Min}}_{x \in X} f(x) \ \text{ s.t. } \ \inf_{t > 0} \left[f_{i0}(x) + \sum_{j=1}^{d} |f_{ij}(x)| - t \log \alpha_i \right] \le 0, \ i = 1, ..., m,$$

or, which is the same due to $\alpha_i \le 1$,

$$\mathop{\mathrm{Min}}_{x \in X} f(x) \ \text{ s.t. } \ f_{i0}(x) + \sum_{j=1}^{d} |f_{ij}(x)| \le 0, \ i = 1, ..., m. \tag{6.21}$$

27

As it could be expected, in the situation in question, Bernstein approximation recovers the *Robust Counterpart* of the original uncertain problem, which, in the case of uncertain constraints (6.9) and trivial uncertainty set $\mathcal{U}$ is the semi-infinite optimization program:

$$\underset{x \in X}{\text{Min}} \, f(x) \text{ s.t. } f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x) \leq 0, \; \forall i, \; \forall \xi \in \bigcup_{P \in \mathfrak{P}} \text{supp}(P). \quad \text{(RC)}$$

It is clear that in the extreme case we are considering the approximation is *exactly equivalent* to the chance constrained problem (6.13). A relatively good news about Bernstein approximation (6.21) is that it in our example it is no more conservative than Robust Counterpart. It is immediately seen that this is a general fact: whenever Bernstein approximation (6.20) is well defined, its feasible set contains the feasible set of (RC).

We see that when all our knowledge on uncertainty is the ranges of ξ_j (and the fact that they are independent), both the chance constrained problem (6.13) itself and its Bernstein approximation become the completely worst-case oriented Robust Counterpart. The situation changes dramatically when we add something to the knowledge of ranges, for example, assume that we know the expected values of ξ_j.

Example 2: ranges and expectations of ξ_j are known. Assume that we know that ξ_j are independent, take values in known finite segments and have known expectations. Same as in Example 1, we can further assume without loss of generality that ξ_j vary in $[-1, 1]$ and have known expectations μ_j, $|\mu_j| \leq 1$. We are in the situation where $\mathcal{M}$ is the set of all tuples $\{Q_j\}_{j=1}^{d}$ with Q_j belonging to the family Π_{μ_j} of all Borel probability distributions on $[-1, 1]$ with expectation μ_j, $j = 1, ..., d$, and $\mathfrak{P}$ is the set of all product distributions on $\mathbb{R}^d$ with the collection of marginal distributions belonging to $\mathcal{M}$. It is easy to see that when $|\mu| \leq 1$, then

$$\Lambda_\mu(t) := \max_{Q \in \Pi_\mu} \log \left(\int \exp\{ts\} dQ(s) \right) = \log(\cosh(t) + \mu \sinh(t)) \; \text{[2]}$$

and that $\Lambda_\mu(0) = 0$, $\Lambda'_\mu(0) = \mu$ and $\Lambda''_\mu(t) \leq 1$ for all t, whence

$$\Lambda_\mu(s) \leq \mu s + \frac{s^2}{2}, \; \forall s.$$

We therefore have

$$
\begin{aligned}
\widehat{\Phi}(z) &:= \max_{P \in \mathfrak{P}} \log \left(\mathbb{E}_P \left\{ \exp\{z_0 + \sum_{j=1}^{d} \xi_j z_j\} \right\} \right) \\
&= z_0 + \sum_{j=1}^{d} \log \left(\cosh(z_j) + \mu_j \sinh(z_j) \right) \leq \widetilde{\Phi}(z) := z_0 + \sum_{j=1}^{d} \left[\mu_j z_j + \frac{z_j^2}{2} \right], \\
\widehat{\Theta}(z, t) &:= t \widehat{\Phi}(t^{-1} z) = z_0 + \sum_{j=1}^{d} t \log \left(\cosh(t^{-1} z_j) + \mu_j \sinh(t^{-1} z_j) \right) \\
&\leq \widetilde{\Theta}(z, t) = z_0 + \sum_{j=1}^{d} \mu_j z_j + \frac{1}{2t} \sum_{j=1}^{d} z_j^2.
\end{aligned}
\quad (6.22)
$$

---

[2] Here is the verification: let $\lambda = \sinh(t)$ and $g(s) = \exp\{ts\} - \lambda s$. This function is convex and therefore takes its maximum on $[-1, 1]$ at an endpoint; it is immediately seen that this maximum is $g(1) = g(-1) = \cosh(t)$. It follows that when $Q \in \Pi_\mu$, one has $\int \exp\{ts\} dQ(s) = \int g(s) dQ(s) + \lambda \mu = \cosh(t) + \mu \sinh(t)$. The resulting upper bound on $\int \exp\{ts\} dQ(s)$ is achieved when Q is two-point distribution with mass $\frac{1+\mu}{2}$ at 1 and mass $\frac{1-\mu}{2}$ at -1.

28

To proceed, we were supposed to compute the functions

$$G(z,\beta) := \inf_{t>0} \left[\widehat{\Theta}(z,t) - t\log\beta \right]$$

and to write down Bernstein approximation (6.20) of ambiguous chance constrained problem in question as the convex program

$$\operatorname*{Min}_{x \in X} \left\{ f(x) : G(z^i[x], \alpha_i) \le 0, \ i = 1,...,m \right\},$$
$$z^i[x] = (f_{i0}(x), f_{i1}(x), ..., f_{id}(x))^T \tag{6.23}$$

where $\alpha_i > 0$ are chosen to satisfy $\sum\limits_i \alpha_i \le \alpha$. While computing $G(z,\beta)$ and its derivatives in z_j numerically (which is all we need in order to solve convex program (6.23) numerically) is easy, a closed form analytic expression for this function seems to be impossible. What we can do analytically, is to bound G from above[3], exploiting the simple upper bound on $\widehat{\Theta}$ presented in (6.22). From the concluding inequality in (6.22) it follows that

$$\begin{aligned}
G(z,\beta) &:= \inf_{t>0} \left[\widehat{\Theta}(z,t) - t\log\beta \right] \\
&\le G_*(z,\beta) := \inf_{t>0} \left[z_0 + \sum_{j=1}^d \mu_j z_j + \frac{1}{2t} \sum_{j=1}^d z_j^2 - t\log\beta \right] \\
&= z_0 + \sum_{j=1}^d \mu_j z_j + \sqrt{2\log(1/\beta)} \left(\sum_{j=1}^d z_j^2 \right)^{1/2}.
\end{aligned} \tag{6.24}$$

It follows that the convex optimization program

$$\operatorname*{Min}_{x \in X} \left\{ f(x) : \begin{array}{c} f_{i0}(x) + \sum\limits_{j=1}^d \mu_j f_{ij}(x) \\[2mm] + \sqrt{2\log(1/\alpha_i)} \left(\sum\limits_{j=1}^d f_{ij}^2(x) \right)^{1/2} \le 0, \ i = 1,...,m \end{array} \right\} \qquad \left[\sum_i \alpha_i \le \alpha \right]$$

is an approximation (more conservative than Bernstein one) of the ambiguous chance constrained problem (6.13), where the independent of each other random perturbations ξ_j are known to vary in $[-1,1]$ and possess expected values μ_j. As could be expected, we have recovered (a slightly refined version of) the results of [2] mentioned in Introduction (see (1.8)) and Remark 3.

Comparing (6.21) and (6.23) – (6.24), we clearly see how valuable could be the information on expectations of ξ_j, provided that ξ_j are independent (this is the only case we are considering). First of all, from the origin of $G(z,\beta)$ it follows that the left hand sides of constraints in (6.21) are pointwise $\ge$ their counterparts in (6.23), so that (6.23) always is less conservative than (6.21). To see how large could be the corresponding "gap", consider the case when all ξ_j have zero means ($\mu_j = 0$ for all j). In this case, i-th constraint in (6.21) requires from the vector $h_i(x) := (f_{i1}(x), ..., f_{id}(x))^T$ to belong to the centered at the origin $\|\cdot\|_1$-ball of radius $\rho(x) = -f_{i0}(x)$, let this ball be called $V_1(x)$. i-th constraint in (6.23), first, allows for $h_i(x)$

---

[3] It should be stressed that this bounding is completely irrelevant as far as numerical processing of (6.23) is concerned; the only purpose of the exercise to follow is to link our approach with some previously known constructions.

to belong to $V_1(x)$ (recall that (6.23) is less conservative than (6.21)) and, second, allows for this vector to belong to the centered at the origin $\|\cdot\|_2$-ball $V_2(x)$ of the radius $\kappa^{-1}\rho(x)$, where $\kappa = \sqrt{2\log(1/\alpha_i)}$ (see (6.24) and take into account that $\mu_j \equiv 0$); by convexity, it follows that i-th constraint in (6.23) allows for $h_i(x)$ to belong to the set $V_{1,2}(x) = \text{Conv}\{V_1(x) \cup V_2(x)\} \supset V_1(x)$. When d is not small, the set $V_{1,2}(x)$ is not merely larger, it is "much larger" than $V_1(x)$, and, consequently, i-th constraint in (6.23) is "much less restricting" than its counterpart in (6.21). To get an impression of what "much larger" means, note that the distance from the origin to the boundary of $V_2(x)$ along every direction is $\kappa^{-1}\rho(x)$; the distance to the boundary of $V_{1,2}(x)$ can only be larger. At the same time, the distance from the origin to the boundary of $V_1(x)$ along a randomly chosen direction is, with probability approaching 1 as $d \to \infty$, at most $\sqrt{\pi/2}(1+\delta)d^{-1/2}$ for every fixed $\delta > 0$. Thus, the ratio of the distances, taken along a randomly chosen direction, from the origin to the boundaries of $V_{1,2}(x)$ and of $V_1(x)$ is always ≥ 1, and with probability approaching 1 as $d \to \infty$, is at least $\frac{(1-\delta)\sqrt{2d/\pi}}{\kappa}$ for every $\delta > 0$; in this sense $V_{1,2}$ is "at average" nearly $\frac{\sqrt{2d/\pi}}{\kappa}$ times larger in linear sizes than $V_1(x)$. Now, for all practical purposes κ is a moderate constant[4]; thus, we can say that as d grows, approximation (6.23) becomes progressively ("by factor $\sqrt{d}$") less conservative than (6.21).

Coming back to our examples, observe that if $\mathcal{M} = \Pi_1 \times ... \times \Pi_d$ where Π_j is a given set in the space of Borel probability distributions on the axis, we have

$$\widehat{\Phi}(z) = z_0 + \sum_{j=1}^{d} \max_{Q \in \Pi_j} \log\left(\int \exp\{z_j s\} dQ(s)\right),$$

and therefore computation of $\widehat{\Phi}(z)$ (which is all we need in order to build Bernstein approximation) reduces to computing the functions $\Lambda^\Pi(t) \equiv \max_{Q \in \Pi} \log\left(\int \exp\{ts\} dQ(s)\right)$ for $\Pi = \Pi_1, ..., \Pi_d$. In Table 2, we present explicit expressions for $\Lambda^\Pi(\cdot)$ for a number of interesting sets Π. In the table Mean[Q], Var[Q] stand for the mean $\int s\, dQ(s)$ and the second moment $\int s^2 dQ(s)$ of distribution Q; to save

---
[4] With $\alpha_i = \alpha/m$, even risk as small as $\alpha = 1.e-12$ and the number of constraints as large as $m = 10,000,000$ result in $\kappa \leq 9.4$.

notation, we present the expressions for $\exp\{\Lambda^\Pi(t)\}$ rather than for Λ^Π itself.

| Π	$\exp\{\Lambda^\Pi(t)\}$
$\{Q : \mathrm{supp}(Q) \subset [-1,1]\}$	$\exp\{|t|\}$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1] \\ Q \text{ is symmetric} \end{array}\right\}$	$\cosh(t)$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1], Q \text{ is} \\ \text{unimodal w.r.t. } 0^{a)} \end{array}\right\}$	$\dfrac{\exp\{|t|\} - 1}{|t|}$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1], Q \text{ is unimodal} \\ \text{w.r.t. } 0 \text{ and symmetric} \end{array}\right\}$	$\dfrac{\sinh(t)}{t}$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1] \\ \mathrm{Mean}[Q] = \mu \end{array}\right\}$	$\cosh(t) + \mu\sinh(t)$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1] \\ \mu_- \leq \mathrm{Mean}[Q] \leq \mu_+ \end{array}\right\}$	$\cosh(t) + \max\left[\mu_-\sinh(t), \mu_+\sinh(t)\right]$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1] \\ \mathrm{Mean}[Q] = 0 \\ \mathrm{Var}[Q] \leq \sigma^2 \end{array}\right\}$	$\dfrac{\exp\{-|t|\sigma^2\} + \sigma^2\exp\{|t|\}}{1 + \sigma^2}$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1], Q \text{ is} \\ \text{symmetric}, \mathrm{Var}[Q] \leq \sigma^2 \end{array}\right\}$	$\sigma^2\cosh(t) + (1 - \sigma^2)$
$\left\{Q : \begin{array}{c} \mathrm{supp}(Q) \subset [-1,1] \\ \mathrm{Mean}[Q] = \mu \\ \mathrm{Var}[Q] \leq \sigma^2 \end{array}\right\}$	$\left\{\begin{array}{ll} \dfrac{(1-\mu)^2\exp\{t\frac{\mu-\sigma^2}{1-2\mu+\sigma^2}\} + (\sigma^2-\mu^2)\exp\{t\}}{1-2\mu+\sigma^2}, & t \geq 0 \\ \dfrac{(1+\mu)^2\exp\{t\frac{\mu+\sigma^2}{1+2\mu+\sigma^2}\} + (\sigma^2-\mu^2)\exp\{-t\}}{1+2\mu+\sigma^2}, & t \leq 0 \end{array}\right.$

$^{a)}$ Q is unimodal w.r.t. 0, if Q is the sum of two measures: a mass at 0 and a measure with density $p(s)$ which is nondecreasing when $t \leq 0$ and nonincreasing when $t \geq 0$

Table 2: $\exp\{\Lambda^\Pi(\cdot)\}$ for several families Π of univariate distributions.

the parameters μ, σ^2 are subject to natural restrictions $|\mu| \leq 1$, $\sigma^2 \leq 1$, $\mu^2 \leq \sigma^2$.

Fig. 1 illustrates the contents of Table 2.

We could proceed in the same fashion, adding more a priori information on the distribution of ξ; until this information becomes too complicated for numerical processing, it can be "digested" by Bernstein approximation. Instead of moving in this direction, we prefer to present example of another sort, where the assumptions underlying Theorem 1 are severely violated, but Bernstein approximation scheme still works.

Example 3: parametric uncertainty. Assume that we know a priori that some of ξ_j are normal, and the remaining ones are Poisson; however, we do not know exactly the parameters of the distributions. Specifically, let us parameterize normal distribution by its mean and variance (note: variance, not standard deviation!), and Poisson one – by its natural parameter λ (so that the probability for the corresponding random variable to attain value $i = 0, 1, ...$ is $\frac{\lambda^i}{i!}\exp\{-\lambda\}$). Let us arrange parameters of the d distributions in question in a vector ω, and assume that our a priori knowledge is that ω belongs to a known in advance convex compact set Ω. We assume also that the latter set is "realizable" in the sense that every point $\omega \in \Omega$ indeed represents a collection of distributions of the outlined type; specifically, the coordinates of $\omega \in \Omega$ which represent variances of normal distributions and the parameters of the Poisson distributions are positive. Note that our a priori knowledge is incompatible with assumption B2: convexity in the space of parameters has small in common with convexity in the space of distributions. For example, when the mean

31
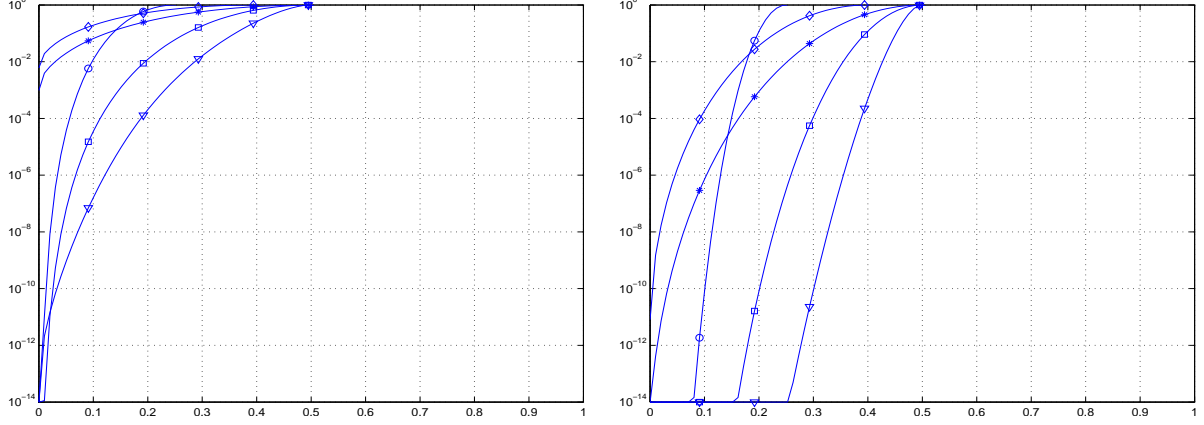
The plots summarize the results of the following numerical experiment. We choose somehow a $(d+1)$-dimensional vector $z = (z_0, ..., z_d)^T$, normalize it by shifting and scaling to ensure that

$$\max_{\xi: \|\xi\|_\infty \leq 1} [z_0 + \sum_{j=1}^{d} \xi_j z_j] = 0, \quad \min_{\xi: \|\xi\|_\infty \leq 1} [z_0 + \sum_{j=1}^{d} \xi_j z_j] = -1,$$

and consider the family of shifts $z^s = (z_0 + s, z_1, z_2, ..., z_d)^T$ of the vector. Our goal is to bound from above, via Bernstein bounding scheme, the quantity $p(s) = \max_{P_j \in \Pi, 1 \leq j \leq d} \text{Prob}_{P_1 \times ... \times P_d} \{\chi_s(\xi) > 0\}$, where $\chi_s(\xi) = z_0 + s + \sum_{j=1}^{d} \xi_j z_j$ and Π is a given family of probability distributions on $[-1, 1]$. The corresponding Bernstein bound is $\exp\left\{\inf_{t>0}\left[t^{-1}(z_0 + s) + \sum_{j=1}^{d} \Lambda^\Pi(t^{-1} z_j)\right]\right\}$.

We vary the shift s in $(0, 1)$; note that with our normalization of z, $\chi_s(\xi)$ is nonpositive on the box $[-1, 1]^d$ iff $s \leq 0$ and is nonnegative on the box iff $s \geq 1$. The plots display, for several choices of Π, the Bernstein bound on $p(s)$ as a function of $s \in (0, 1)$. The left plot corresponds to $d \equiv \dim \xi = 10$, the right plot - to $d = 50$. To reduce the size of the plots along the y-axis, the bounds were truncated below at the level 10^{-14}. The families Π underlying the bounds are as follows:

curve	Π
$\triangle$	$\{Q : \text{supp}(Q) \subset [-1, 1], \text{Mean}[Q] = 0, \text{Var}[Q] \leq 0.04\}$
$\square$	$\{Q : \text{supp}(Q) \subset [-1, 1], Q \text{ is symmetric and unimodal w.r.t. } 0\}$
o	$\{Q : \text{supp}(Q) \subset [-1, 1], Q \text{ is unimodal w.r.t. } 0\}$
$*$	$\{Q : \text{supp}(Q) \subset [-1, 1], Q \text{ is symmetric}\}^{a)}$
$\diamond$	$\{Q : \text{supp}(Q) \subset [-1, 1], |\text{Mean}[Q]| \leq 0.2\}$

[a)] the same bound corresponds to $\Pi = \{Q : \text{supp}(Q) \subset [-1, 1], \text{Mean}[Q] = 0\}$

The bound corresponding to $\Pi = \{Q : \text{supp}(Q) \subset [-1, 1]\}$ in the interval $-1 < s < 1$ is $\equiv 1$. We see how valuable additional information on the distributions of ξ_j can be: the right plot says that when all we know about ξ_j, aside of their independence, is that they are with zero means (or that they are symmetrically distributed), the guaranteed upper bound on the probability for $\chi_{0.3}(\xi)$ to be positive is 0.0437. When we know in addition that the standard deviations of ξ_j are ≤ 0.2, the bound improves to 2.3e-11 !

Figure 1: Dependence of the Bernstein bound on a priori information.

32

of a normal distribution with unit variance runs through a given segment, the distribution itself moves along a complicated curve. We, however, can try to use the same approach which led us to Theorem 1. Observe that when P_j is the Poisson distribution with parameter λ, we have

$$\log\left(\int \exp\{rs\}dP_j(s)\right) = \log\left(\sum_{i=0}^{\infty} \frac{(\lambda e^r)^i}{i!}\exp\{-\lambda\}\right) = \log(\exp\{\lambda e^r - \lambda\}) = \lambda\exp\{r\} - \lambda;$$

the resulting function is continuous, convex in r, as it always is the case for the logarithmic moment generating function, and is concave in λ, which is pure luck. We are equally lucky with the normal distribution P_j with mean μ and variance ν:

$$\log\left(\int \exp\{rs\}dP_j(s)\right) = \log\left(\frac{1}{\sqrt{2\pi\nu}}\int \exp\{rs - \frac{(s-\mu)^2}{2\nu}\}ds\right) = r\mu + \frac{r^2\nu}{2},$$

and the result again is continuous, convex in r and concave in (μ, ν). It follows that if P^ω is the joint distribution of the sequence of d normal/Poisson independent random variables ξ_j, the vector of parameters of the marginal distributions being ω, then, for every vector $z \in \mathbb{R}^{d+1}$, the function

$$\Phi_\omega(z) = \log\left(\mathbb{E}_{P^\omega}\left[\exp\{z_0 + \sum_{j=1}^{d}\xi_j z_j\}\right]\right)$$

is given by a simple explicit expression, is continuous in $(z \in \mathbb{R}^{d+1}, \omega \in \Omega)$, and is convex in z and concave (in fact even affine) in ω. We now can use the reasoning which led us to Theorem 1 and (6.20) to conclude that the optimization problem

$$\operatorname*{Min}_{x\in X} f(x) \text{ s.t. } \inf_{t>0}\left[t\widehat{\Phi}(t^{-1}z^i[x]) - t\log\alpha_i\right] \le 0, \ i = 1, ..., m,$$
$$\widehat{\Phi}(z) = \max_{\omega\in\Omega}\Phi_\omega(z), \ z^i[x] = (f_{i0}(x), f_{i1}(x), ..., f_{id}(x))$$

is an approximation of the ambiguous chance constrained problem under consideration, provided that $\alpha_i \in (0,1)$ are such that $\sum_i \alpha_i \le \alpha$. This approximation is convex, provided that all functions f_{ij} are convex and well defined on X and the functions f_{ij} with j's corresponding to normally distributed components in ξ are affine. Finally, our approximation is computationally tractable, provided that $\widehat{\Phi}(\cdot)$ is efficiently computable (which indeed is the case when Ω is computationally tractable).

References

[1] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D., Coherent measures of risk, *Mathematical Finance*, 9 (1999) 203–228.

[2] Ben-Tal, A., and Nemirovski, A. Robust solutions of Linear Programming problems contaminated with uncertain data, *Mathematical Programming*, 88 (2000), 411-424.

[3] Ben-Tal, A., and Nemirovski, A., Robust Optimization — Methodology and Applications, *Mathematical Programming Series B*, 92 (2002), 453-480.

[4] Bertsimas, D. and Sim, M., Price of Robustness, *Operations Research*, 52 (2004), 35–53.

[5] Calafiore, G., Campi, M.C., Uncertain convex programs: randomized solutions and confidence levels, *Mathematical Programming*, to appear.

[6] Calafiore, G., Campi, M.C., Decision making in an uncertain environment: the scenario-based optimization approach, Working paper, 2004.

[7] Charnes, A., Cooper, W.W. and G.H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, *Management Science*, 4 (1958), 235–263.

[8] Dembo, A. and Zeitouni, O., *Large Deviations Techniques and Applications*, Springer-Verlag, New York, NY, 1998.

[9] de Farias, D.P., Van Roy, B., On constraint sampling in the linear programming approach to approximate dynamic programming, *Mathematics of Operations Research*, 29 (2004), 462–478.

[10] Dupačová, J., The minimax approach to stochastic programming and an illustrative application, *Stochastics*, 20 (1987), 73–88.

[11] Ergodan, G., and Iyengar, G., Ambiguous chance constrained problems and robust optimization, *Mathematicsl Programming Series B*, Special issue on Robust Optimization, to appear.

[12] Johnson, W.B. and Schechtman, G., Remarks on Talagrand's deviation inequality for Rademacher functions, *Springer Lecture Notes on Mathematics*, 1470 (1991), 72–77.

[13] Kleywegt, A. J., Shapiro, A. and Homem-De-Mello, T., The sample average approximation method for stochastic discrete optimization, *SIAM Journal of Optimization*, 12 (2001), 479–502.

[14] Miller, L.B. and Wagner, H., Chance-constrained programming with joint constraints, *Operations Research*, 13 (1965), . 930–945.

[15] Mulvey, J.M., Vanderbei, R.J., Zenios, S.A., Robust optimization of large-scale systems, *Operations Research*, 43 (1995), 264–281.

[16] Nemirovski, A., On tractable approximations of randomly perturbed convex constraints, *Proceedings of the 42nd IEEE Conference on Decision and Control Maui, Hawaii USA, December 2003*, 2419-2422.

[17] Nemirovski, A. and Shapiro, A., Scenario Approximations of Chance Constraints, to appear in: "Probabilistic and Randomized Methods for Design under Uncertainty", Calafiore and Campi (Eds.), Springer, London.

[18] Prékopa, A., On probabilistic constrained programming, in: *Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton, pp. 113–138, 1970.

[19] Prékopa, A., *Stochastic Programming*, Kluwer, Dordrecht, Boston, 1995.

[20] Rockafellar, R.T. and Uryasev, S.P., Optimization of conditional value-at-risk, *The Journal of Risk*, 2 (2000), 21-41.

[21] Rockafellar, R.T., Uryasev, S., Zabarankin, M., "Deviation measures in risk analysis and optimization", Research Report 2002-7, Department of Industrial and Systems Engineering, University of Florida.

[22] Rusczyński, A. and Shapiro, A., Optimization of Risk Measures, to appear in: "Probabilistic and Randomized Methods for Design under Uncertainty", Calafiore and Campi (Eds.), Springer, London.

[23] Shapiro, A., Monte Carlo sampling methods, in: A. Rusczyński and A. Shapiro (editors), *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, North-Holland, 2003.

[24] Soyster, A.L., Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming, *Operations Research*, 21 (1973), 1154–1157.

[25] Žáčková, J., On minimax solutions of stochastic linear programming problems, *Čas. Pěst. Mat.*, 91 (1966), 423–430.