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 deterministic 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@isye.gatech.edu`
`research of this author was partly supported by the Binational Science Foundation grant #2002038`
[†]Georgia Institute of Technology, Atlanta, Georgia 30332, USA, `ashapiro@isye.gatech.edu`
`research of this author was partly supported by the NSF grant DMS-0510324`

1 Introduction

Let us consider the following optimization problem

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

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 $\operatorname{Prob}(A)$ 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 [16] and Prékopa [21].

Aside of potential modelling problems with formulation (1.1) (e.g., the necessity to know the probability distribution of the random vector ξ, which in practice not always is easy), there is a serious numerical problem with chance constraints. 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 the outlined severe difficulties not always arise, and a number of efficient in "good cases" approaches to solving (1.1) were proposed in the literature, most notably the concept of *logarithmically concave distributions* due to A. Prekopa [22] – a wide family of probability distributions on $\mathbb{R}^n$ such that the associated sets of the form $\{x \in \mathbb{R}^m : \operatorname{Prob}\{\xi : Ax \ge \xi\} \ge \alpha\}$ (or, more general, of the form $\{x \in \mathbb{R}^m : \operatorname{Prob}\{\xi : (x,\xi) \in X\} \ge \alpha\}$, where X is a convex set in $\mathbb{R}^m_x \times \mathbb{R}^n_\xi$) are convex for all $\alpha \in [0,1]$. In this case, the feasible set of (1.1) is convex, provided that $F(x,\xi)$ is jointly convex in (x,ξ). Unfortunately, this case does not cover many important applications, e.g., linear programs with randomly perturbed coefficients in the constraint matrix. Indeed, a single scalar constraint

$$f(x,\xi) = a(\xi) + \sum_j a_j(\xi) x_j \le 0, \tag{1.2}$$

bi-linear in x, ξ (i.e., with affine in ξ coefficients) satisfies the joint convexity requirement only when all $a_j(\xi)$ are in fact independent of ξ. Moreover, to the best of our knowledge, the only known generic case where the chance constrained version $\operatorname{Prob}\{f(x,\xi) \le 0\} \ge 1-\alpha$ of a bi-linear constraint (1.2) has a provably convex feasible set is the case of normally distributed ξ and $\alpha \le 1/2$. There are also ways to circumvent, to some extent, the difficulties caused by non-convexity of the feasible set of a typical chance constraint (see, e.g., [23, 10]), but the problem still persists: in general, chance constrained version of a randomly perturbed constraint $F(x,\xi) \le 0$, even as simple-looking one as the bi-linear constraint (1.2), is "severely computationally intractable". Whenever this is the case, a natural course of actions is to look for *tractable approximations* of the chance constraint, i.e., for efficiently verifiable *sufficient conditions* for its validity. In addition to being sufficient, such a condition should define a convex and "computationally tractable" set in the x-space, e.g., 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

$$\min_{x \in X, u \in \mathbb{R}^s} f(x) \text{ subject to } G(x,u) \le 0 \tag{1.3}$$

is a Convex Programming program. As such it is efficiently solvable, provided that $G(x,u)$ is efficiently computable, and provides a *conservative approximation* of the chance constrained problem of interest, meaning that the projection of the feasible set of (1.3) onto the space of x-variables is contained in the feasible set of the chance constrained problem (1.1), so that an optimal solution to (1.3) is *feasible sub*optimal solution to (1.1).

A general way to build computationally tractable approximations (not necessarily conservative) 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.

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 extent 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 [13] 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

$$N \ge N^* := \text{Ceil}\left[\frac{2n}{\alpha}\log\left(\frac{12}{\alpha}\right) + \frac{2}{\alpha}\log\left(\frac{2}{\delta}\right) + 2n\right], \tag{1.4}$$

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 (1.4) 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 (1.4) (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 problems (with $\alpha = 0.01$ and $n = 200$, $\delta = 0.01$, the estimate (1.4) results in $N^* = 285,063$). Note that for a properly modified scenario approximation, "bad" dependence of N on α given by (1.4) can be replaced with

$$N = O(1)\left[\log(1/\delta) + dm^2 \log(d \log(1/\alpha))\right], \tag{1.5}$$

provided that $F(x, \xi)$ is affine in ξ and ξ has a "nice" distribution, e.g., uniform in a box, or on the vertices of a box, or normal [19].

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

$$f_0(x) + \sum_j \xi_j f_j(x) \le 0 \tag{1.6}$$

(cf., (1.2)). 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

$$f_0(x) + \Omega \left(\sum_{j=1}^{d} \sigma_j^2 f_j^2(x) \right)^{1/2} \le 0, \tag{1.7}$$

where $\Omega > 0$ is a "safety" parameter, then x violates the randomly perturbed constraint (1.6) 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

$$f_i(x, \xi) = f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x), \tag{1.8}$$

then the optimization program

$$\operatorname*{Min}_{x \in X} f(x) \ \text{ subject to } f_{i0}(x) + \Omega \left(\sum_{j=1}^{d} \sigma_j^2 f_{ij}^2(x) \right)^{1/2} \le 0, \ i = 1, ..., m, \tag{1.9}$$

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 $j \ge 1$ is either affine, or nonnegative. Another, slightly more convenient computationally, analytical approximation of randomly perturbed constraint (1.6) was proposed in [4]. Analytical approximations of more complicated chance constraints, notably a randomly perturbed conic quadratic inequality, are presented in [18]. 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 develop 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 the form (1.8) 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.

After the initial version of this paper was released, we became aware of the paper of J. Pinter [20] proposing (although not in full generality) Bernstein approximation, even in its advanced "ambiguous" form (see section 6 below). The only (but, we believe, a quite important) step ahead in what follows as compared to Pinter's paper is that with our approach, the natural scale parameter of Bernstein approximation ("h" in Pinter's paper) becomes a variable rather than an ad hoc chosen constant (as it is the case in [20]). Specifically, we manage to represent Bernstein bound in a form which is jointly convex in the original decision variables *and* the scale parameter, which allows to deal, staying all the time within the Convex Programming framework, with the bound with is point-wise optimized in the scale parameter.

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 *ambiguously chance constrained 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.

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 the 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 (1.3). 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 (1.3) 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 (1.3) 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

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

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:

($*$) $\psi(z) > \psi(0) = 1$ 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,

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

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:

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

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

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

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 $\inf\limits_{t>0}\phi(t) \leq 0$ 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 $(x, t > 0)$.

[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

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

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

$$\operatorname*{Min}_{x\in X,\, t>0} f(x) \ \text{ subject to } \ \inf_{t>0} \left[\Psi(x,t) - t\alpha\right] \le 0 \tag{2.6}$$

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). If the right derivative $\psi'_+(0)$ is zero, then $\psi(z) \ge \psi(0) = 1$ 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 $\psi(z) \ge \max\{1 + az, 0\}$ 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

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

Replacing in the left hand side $\inf_{t>0}$ with $\inf_t$, we clearly do not affect the validity of the relation; thus, we can rewrite (2.8) equivalently as

$$\inf_{t\in\mathbb{R}} \left[-t\alpha + \mathbb{E}\big[[F(x,\xi) + t]_+\big]\right] \le 0. \tag{2.9}$$

In that form the constraint is related to the concept of Conditional Value at Risk (CVaR) going back to [12]. Recall that CVaR of a random variable Z is

$$\mathrm{CVaR}_{1-\alpha}(Z) := \inf_{\tau\in\mathbb{R}} \left[\tau + \frac{1}{\alpha}\mathbb{E}[Z - \tau]_+\right]. \tag{2.10}$$

It is easily seen that $\mathrm{CVaR}_{1-\alpha}(Z)$ is a convex and monotone functional on the space of random variables with finite first moment, and that the $(1-\alpha)$-quantile ("Value at Risk")

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

of the distribution of Z is a minimizer of the right hand side in (2.10), so that it always holds that $\mathrm{CVaR}_{1-\alpha}(Z) \ge \mathrm{VaR}_{1-\alpha}(Z)$. Since the chance constraint in (1.1) is nothing but $\mathrm{VaR}_{1-\alpha}[F(x,\xi)] \le 0$, the constraint

$$\mathrm{CVaR}_{1-\alpha}[F(x,\xi)] \le 0 \tag{2.11}$$

defines a convex conservative approximation of the chance constraint. The idea of using CVaR as a convex approximation of VaR is due to Rockafellar and Uryasev [24]. Recalling the definition of CVaR, we see that the constraints (2.9) and (2.11) are equivalent to each other.

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.

There are several ways how the above construction can be extended for $m > 1$. One simple way is to replace the constraints $f_i(x,\xi) \le 0$, $i = 1, ..., m$, with one constraint $f(x,\xi) \le 0$, 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 $F(x,\xi) \le 0$ 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

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

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 $\psi(z) \ge \psi(y)$,

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

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

By ($\substack{\star\\\star}$) 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 $x, y \in \mathbb{R}^n$,

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

(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 $\text{Prob}\big\{\cup_{i=1}^m \{\xi : f_i(x,\xi) > 0\}\big\} < \alpha$. Since

$$\text{Prob}\Big\{\bigcup_{i=1}^m \{f_i(x,\xi) > 0\}\Big\} \le \sum_{i=1}^m \text{Prob}\big\{f_i(x,\xi) > 0\big\},$$

it follows that the system of constraints

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

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):

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

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 2.1 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 $F(x,\xi)$. 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 and A2, the function

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

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 $p(z) := \mathrm{Prob}\big\{z_0 + \sum_{j=1}^{d} \xi_j z_j > 0\big\}$ 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 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 $\alpha_i > 0$, $\sum_i \alpha_i \le \alpha$, and build the optimization problem

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

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 Bernstein approximation is an explicit convex program with efficiently computable constraints and as such is efficiently solvable.

Remark 3.1 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 (1.9). The justification of (1.9) 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 (1.7), 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

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

Clearly, Bernstein approximation as presented here is less conservative than (1.9), 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 $z = (z_0, ..., z_d) \in \mathbb{R}^d$, and let ξ_j be independent random variables with distributions P_j, $j = 1, ..., d$. How conservative could be Bernstein sufficient condition*
>
> $$\inf_{t>0} \left[t\log\left(\mathbb{E}\left[\exp\{t^{-1}[z_0 + \sum_{j=1}^d \xi_j z_j]\} \right] \right) - t\log\beta \right] \le 0 \tag{3.5}$$

for the validity of the chance constraint

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

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 $d = 1$, $z = (-\kappa, 1+\kappa)^T$ 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

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

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

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

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 $p(z) \le 0.8841$.

There is, however, another way to measure the conservatism, perhaps a more natural one. After appropriate updating of z_0, we can assume w.l.o.g. that all ξ_j have zero means. Now let us consider chance constraint (3.6) as a member of the parametric family

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

where $\rho > 0$ is an "uncertainty level"; the original constraint corresponds to $\rho = 1$. Typically, the validity of (3.7) for a not too large β implies that $z_0 \le 0$ (recall that ξ_j are with zero means); e.g., when ξ_j are symmetrically distributed, one has $z_0 \le 0$ whenever $\beta < 1/2$. It is technically convenient for us to impose on z the requirement $z_0 \le 0$ explicitly, independently of whether it is or is not implied by (3.7), and to define the feasible set Z_ρ of (3.7) as the set of all $z = (z_0, ..., z_d)$ *with* $z_0 \le 0$ satisfying (3.7). Now, it is immediately seen that if $z_0 \le 0$, then the larger is ρ, the larger is $p_\rho(z)$, so that the set Z_ρ shrinks as ρ grows. We can now measure the conservatism of (3.5) by the smallest $\rho = \rho_* \ge 1$ 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 [19]. 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 3.1 *Let ξ_j be distributed in segments $[-\sigma_j, \sigma_j]$ and such that $\mathbb{E}\{\xi_j\} = 0$, $\mathbb{E}\{\xi_j^2\} \geq 4\alpha^2\sigma_j^2$ with some $\alpha \in (0, 1/2]$, $1 \leq j \leq d$, and let $\beta < \widehat{\beta} = \frac{\alpha^2}{24\log(1/\alpha)}$. Then the conservatism of* (3.5) *is at most*

$$\widehat{\rho} = 1 + \frac{24\sqrt{\log(1/\alpha)}}{\alpha^2}\sqrt{2\log(1.02/\beta)}, \tag{3.8}$$

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

Proof. Passing from random variables ξ_j to the variables $\xi_j/(2\sigma_j)$ and multiplying simultaneously coefficients z_j of ξ_j in (3.7) by $2\sigma_j$, we can assume that $\sigma_j = 1/2$, $j \leq d$. Let z with $z_0 \leq 0$ be feasible for (3.7) with ρ replaced with $\widehat{\rho}$; we should prove that then 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 w.l.o.g. 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

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

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,

$$\mathrm{Prob}\left\{e^T\xi > -\widehat{\rho}^{-1}z_0\right\} \leq \beta \leq \frac{\alpha^2}{24\log(1/\alpha)} \leq 0.015 \tag{3.10}$$

(note that by evident reasons we have $\alpha \in (0, 1/2]$). From (3.10), by Lemma 7.2 in Appendix, it follows that

$$\mu := -\widehat{\rho}^{-1}z_0 \geq \frac{\alpha^2}{24\sqrt{\log(1/\alpha)}}. \tag{3.11}$$

Now let $D(y) := \min\limits_{y' \in A}\|y - y'\|_2 = \max[0, e^Ty + \widehat{\rho}^{-1}z_0] = \max[0, e^Ty - \mu]$. Since ξ_j are independent and uniformly distributed on $[-1/2, 1/2]$, from Talagrand Inequality (see, e.g., [14]) it follows that

$$\mathbb{E}\left[\exp\{\frac{D^2(\xi)}{4}\}\right] \leq \frac{1}{\mathrm{Prob}\{A\}} \leq \frac{1}{0.985}, \tag{3.12}$$

where the concluding inequality is given by (3.10). 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,

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

whence, in view of (3.12),

$$\mathbb{E}\left[\exp\{\tau[z_0+\sum_{j=1}^d \xi_j z_j]\}\right] \leq \frac{c}{0.985} \leq 1.02\exp\{-\frac{\mu^2(\widehat{\rho}-1)^2}{4}\} \leq 1.02\exp\{-\frac{\alpha^4(\widehat{\rho}-1)^2}{2304\log(1/\alpha)}\},$$

where the concluding inequality is given by (3.11). 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 when treating α as an absolute constant, the level of conservatism of Bernstein approximation as stated in (3.8) is not that disastrous – it is nearly independent of β. Note also that the "big" factor $\frac{24\sqrt{\log(1/\alpha)}}{\alpha^2}$ in (3.8) (it is at least about 80 due to $\alpha \leq 1/2$) can be reduced for appropriate distributions; e.g., for the uniform distribution on the vertices of the box $|\xi_j| \leq \sigma_j$, $j=1,...,d$, (3.8) can be improved to $\widehat{\rho}=1+4.93\sqrt{\log(1.05/\beta)}$ provided $\beta \leq 0.05$ (cf., [19, section 2]). We should add that there is another case when Bernstein approximation is "nearly accurate" – the one of normally distributed ξ_j, but this is of no interest, since here a scalar chance constraint $\mathrm{Prob}\{z_0+\sum_{j=1}^d \xi_j z_j > 0\} \leq \beta$ 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 justified sample size (1.4), 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 (1.4)), 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 $p(\bar{x}) := \mathrm{Prob}\big\{F(\bar{x},\xi) \not\leq 0\big\}$ 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

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

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 $\{\mathrm{Opt}_\mu\}_{\mu=1,...,M}$ in the nondescending order: $\mathrm{Opt}_{(1)} \leq ... \leq \mathrm{Opt}_{(M)}$ (in the statistics literature these are called the order statistics of the sample $\{\mathrm{Opt}_\mu\}_{\mu=1,...,M}$). By definition, the lower bound on the true optimal value is the random quantity $\mathrm{Opt}_{(L)}$.

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 $\mathrm{Opt}_\mu \leq f(x)$. Thus, for every $\mu \in \{1, ..., M\}$ and for every $\varepsilon > 0$ we have

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

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

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

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

Thus,

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

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

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

We have arrived at the following simple result.

Proposition 4.1 *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 $\mathrm{Opt}_{(L)}$ *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 $\mathrm{Prob}\{\mathrm{Opt}_1 > a\}$ 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:

$$\operatorname*{Max}_{x\geq 0,\, t}\, t-1 \;\text{ subject to }\; \mathrm{Prob}\left\{t > \sum_{j=0}^{n} r_j x_j\right\} \leq \alpha,\; \sum_{j=0}^{n} x_j \leq 1, \tag{P_α}$$

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 $\mathbb{E}[r_i]=1+\rho_i$, with the nominal profits ρ_i varying in $[0,0.1]$ and growing with i;

- The random variables r_i, $1 \le i \le 64$, are of the form

$$r_i = \eta_i + \sum_{\ell=1}^{8} \gamma_{i\ell}\zeta_\ell, \tag{5.1}$$

where $\eta_i \sim \mathcal{LN}(\mu_i, \sigma_i^2)$ (that is, $\log \eta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$) is the individual noise in i-th return, $\zeta_\ell \sim \mathcal{LN}(\nu_\ell, \theta_\ell^2)$ are "common factors" affecting all returns, and $\gamma_{i\ell} \ge 0$ are deterministic "influence coefficients". All "primitive" random variables (64 of η_i's and 8 of ζ_ℓ's) are independent of each other.

We used $\nu_\ell = 0$, $\theta_\ell = 0.1$, $\mu_i = \sigma_i$ (that is, the more promising is an asset at average, the more risky it is). The influence coefficients $\gamma_{i\ell}$ and the parameters μ_i were chosen in such a way that $\mathbb{E}[\sum\limits_{\ell=1}^{8} \gamma_{i\ell}\zeta_\ell] = \rho_i/2$ and $\mathbb{E}[\eta_i] = 1 + \rho_i/2$ for all i.

Processing log-normal distributions. The random returns r_i are linear combinations of independent random variables $\eta_1, ..., \eta_{64}$, $\zeta_1, ..., \zeta_8$, so that the structure of (P_α) allows for applying Bernstein approximation. The difficulty, however, is that the random variables in question are log-normal and thus the corresponding moment-generating functions are $+\infty$ outside of the origin. This difficulty can be easily circumvented, specifically, as follows. Given a log-normal random variable $\xi \sim \mathcal{LN}(\mu, \sigma^2)$, and positive "threshold probability" $\epsilon > 0$ and "resolution" $\Delta > 0$, we associate with these data a discrete random variable $\widehat{\xi}$ as follows. Let $\pi(s)$ be the $\mathcal{N}(0,1)$-Gaussian density and R be such that $\int\limits_R^\infty \pi(s)ds = \epsilon/2$; we split the segment $[-R, R]$ into bins $[a_k, a_{k+1}]$, $1 \le k < n$, of length $\sigma^{-1}\Delta$ (the last bin can be shorter) and assign the points $b_0 = 0$, $b_k = \exp\{\sigma a_k + \mu\}$, $k = 1, ..., n$, probability masses $\nu_k = \int\limits_{a_k}^{a_{k+1}} \pi(s)ds$, where $a_0 = -\infty$ and $a_{n+1} = \infty$. The variable $\widehat{\xi}$ takes the values b_k, $k = 0, ..., n$ with probabilities ν_k. Note that this random variable can be thought of as a "rounding" of $\xi \sim \mathcal{LN}(\mu, \sigma^2)$: given a realization a of ξ, we look to which one of the $n+1$ sets $[0, b_1)$, $[b_1, b_2)$,..., $[b_{n-1}, b_n)$, $[b_n, \infty)$ a belongs, and replace a with the left endpoint of this set, thus obtaining a realization $\widehat{a}$ of $\widehat{\xi}$. Note that with our choice of a_i, we always have $\widehat{a}/a \le 1$, and $\widehat{a}/a \ge \exp\{-\Delta\}$ unless $a < b_1$ or $a > b_n$; the latter can happen with probability at most ϵ. Thus, $\widehat{\xi}$ can be though of as a lower bound on ξ which with probability $\ge 1 - \epsilon$ is tight within factor $\exp\{\Delta\}$. Now let us replace in (P_α) underlying log-normal random variables $\eta_1, ..., \zeta_8$ with their roundings $\widehat{\eta}_1, ..., \widehat{\zeta}_8$. Since we "round down" and all $\gamma_{i\ell}$ are nonnegative, every feasible solution to the resulting chance constrained problem will be feasible for (P_α) as well. At the same time, the new problem is an affinely perturbed chance constrained problem with *discrete* random variables, and building its Bernstein approximation causes no problems at all. This is the scheme we used in our experiments, the parameters being $\epsilon = 1.\text{e-}6$ and $\Delta = 0.0025$. Even with that high (in fact, redundant) quality of discretization, there was no difficulty with handling the resulting discrete random variables – the average, over all 71 discrete random variables in question, number of different values taken by a variable was just ≈ 138, which made computing Bernstein bound a pretty easy task.

Tuning the approximations. Both approximations we are dealing with in our experiments – the Scenario and Bernstein one – are conservative in the sense that a solution yielded by an

Quantity	Value	Empirical risk$^{a)}$	Inferred risk$^{a)}$
Nominal optimal value$^{b)}$	0.0950	—	—
Upper bound$^{c)}$	0.0799	—	—
Bernstein optimal value (tuned)$^{d_b)}$	0.0689	0.043	0.050
Bernstein optimal value$^{d_a)}$	0.0586	0.002	0.004
Scenario optimal value (tuned)$^{e_b)}$	0.0674	0.040	0.047
Scenario optimal value$^{e_a)}$ ($N = 14,684$)	0.0557	0.001	0.003
Robust optimal value$^{f)}$	0.0000	—	—

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

approximation violates the randomly perturbed constraint in question with probability $\alpha_{\rm f}$ which is less than the required risk α (this claim is completely true for Bernstein approximation and is "true with high probability" for the Scenario one). Experiments show that the ratio $\alpha/\alpha_{\rm f}$ could be pretty large (see Table 1 below), which makes it natural to look for ways to reduce the resulting conservatism. To some extent, this indeed can be done via a simple tuning, provided that α is not too small, so that the probabilities of order of α can be measured reliably by Monte-Carlo simulations with samples of reasonable size. When tuning Bernstein approximation, we replace the required risk α by a larger quantity α_+, solve the approximation as if the required risk were α_+, and then run Monte-Carlo simulation in order to check with a desired reliability whether the actual risk $\alpha_{\rm f}$ of the resulting solution is $\leq \alpha$. We then choose the (nearly) largest possible α_+ which meets the outlined requirement and treat the associated solution as the result of our tuning. Of course, tuning can be used in the case of Scenario approximation as well, with the number of scenarios in the role of tuning parameter.

The experiments were conducted for the value of risk $\alpha = 0.05$. The reliability $1-\delta$ for the scenario approximation (see (1.4)) 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.

Explanations to Table 1: $^{a)}$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.

$^{b)}$Optimal value in the nominal problem – the one where all randomly perturbed coefficients are set to their expected values.

$^{c)}$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.

$^{d_a)}$Optimal value in Bernstein approximation (3.4) of $(P_{0.05})$.

[db])Optimal value in tuned Bernstein approximation. In our experiment, the best tuning corresponded to replacing the true value 0.05 of risk with the value 0.3.

[ea])Optimal value in the scenario approximation (P^N) of ($P_{0.05}$), the sample size N being chosen according to (1.4) (where $n = 66$, $\alpha = 0.05$ and $\delta = 0.001$).

[eb])Optimal value in tuned scenario approximation. In our experiment, the best tuning corresponded to reducing the number of scenarios with its theoretical value 14,684 to 550.

[f])Optimal value 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 $\alpha = 0$. In our case, the robust optimal value is 0, meaning that there is no way to make guaranteed profit, so that the best, in the worst-case setting, policy is to not to invest into "non-money" assets at all.

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 (2.2% for tuned approximations).

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.

C. It should be added that Bernstein approximation 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 (1.4) 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 results in the profits 0.0500 for $\alpha = 0.5\%$ and 0.0445 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 as 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 The case of ambiguous chance constraints

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; for results in this direction, see [11, 29, 25, 26, 13] and references therein. When applying the minimax approach to chance constrained problems, one assumes that the distribution P of random vector ξ in (1.1) belongs to a given in advance family $\mathfrak{P}$ of probability distributions supported on a (closed) set $\Xi \subset \mathbb{R}^d$ and replaces the chance constraint in (1.1) with its worst-case, over $P \in \mathfrak{P}$, version, thus arriving at the *ambiguously chance constrained* problem

$$\operatorname*{Min}_{x\in X} f(x) \ \text{ subject to } \ \operatorname{Prob}_P\big\{F(x,\xi)\le 0\big\} \ge 1-\alpha, \ \ \forall\, P\in\mathfrak{P}, \tag{6.1}$$

where Prob_P is the P-probability of the corresponding event.

Of course, we can replace the probability constraints in (6.1) with one constraint by taking the minimum of $\operatorname{Prob}_P\big\{F(x,\xi)\le 0\big\}$ 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:

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

where $A_x := \{\xi\in\Xi : F(x,\xi)\not\le 0\}$. The "worst-case-distribution" (or minimax) stochastic programming problems were considered in a number of publications (e.g., [11, 29]). When applied to chance constraints, such worst-case-distribution problems are called *ambiguous* chance constrained problems (see [13] 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

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

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:

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

Here γ_1 and γ_2 are constants such that $0\le\gamma_1\le 1\le\gamma_2$, 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 $P_1(A)\le P_2(A)$ 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 $\kappa := (\gamma_2-1)/\gamma_2$,

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

(cf., [25, 26]).

By the definition (6.4) of $\mathfrak{P}$ we have that $\mathbb{E}_P\left[\mathbb{1}_{A_x}\right] \le \gamma_2 P^*(A_x)$ 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 $\gamma_2 P^*(A_x) + \gamma_1(1 - P^*(A_x)) \le 1$, i.e., iff $P^*(A_x) \le \frac{1-\gamma_1}{\gamma_2-\gamma_1}$. We obtain the following.

> If $\alpha \le (1-\gamma_1)/(\gamma_2-\gamma_1)$, 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

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

where $\zeta' = dP/dP^*$ denotes the density of P with respect to P^* (cf., [26]). 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 $t + 2ct(1-t) \le \alpha$ for $t = P^*(A_x)$, we obtain that $P^*(A_x) \le \varphi(\alpha)$, 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 $\mathfrak{P}$. Our current goal is to develop Bernstein-type approximation of the constraint in (6.1). As before, we restrict ourselves with problems where the "bodies" of the constraints are affine in ξ:

$$\operatorname*{Min}_{x\in X} f(x) \ \text{ s.t. } \ \inf_{P\in\mathfrak{P}} \mathrm{Prob}_P\left\{\xi : f_{i0}(x) + \sum_{j=1}^{d} \xi_j f_{ij}(x) \le 0,\ i = 1, ..., m\right\} \ge 1 - \alpha. \tag{6.9}$$

6.1 Assumptions and construction

Assumptions. From now on, we make the following assumptions about the "data" of (6.9):

B1. The family $\mathfrak{P}$ of possible distributions of ξ 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 $\mathcal{M}$ weakly 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 and bounded on the axis function f), then $\{P_j\}_{j=1}^d \in \mathcal{M}$ (weak 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 $\Delta := \bigcup\limits_{j=1}^{d} D_j = \{s_1, ..., s_K\}$, and let $\mathcal{M}$ be a closed and convex set of matrices $P = [p_{kj}]_{\substack{1\leq k\leq K\\ 1\leq j\leq d}}$ with nonnegative entries such that $\sum\limits_k p_{kj} = 1$ 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 $Q = \{Q_j\}_{j=1}^d$.

B2. 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. 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

$$
\begin{array}{rcl}
\Phi(z,Q) & := & \log\left(\mathbb{E}_{Q_1\times\ldots\times Q_d}\left[\exp\{z_0+\sum\limits_{j=1}^d \xi_j z_j\}\right]\right)\\
& = & z_0+\sum\limits_{j=1}^d \log\left(\int \exp\{z_j s\}dQ_j(s)\right),\ \ Q=\{Q_j\}_{j=1}^d\in\mathcal{M},\\
\widehat{\Phi}(z) & := & \max\limits_{Q\in\mathcal{M}}\Phi(z,Q).
\end{array}
\tag{6.10}
$$

By B1, $\Phi(z,Q)$ 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 (6.10) 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 B2 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

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

and we arrive at the implication:

$$
\begin{array}{ll}
P(\beta): & \left\{\forall Q\in\mathcal{M}:\inf\limits_{t>0}\left[\Theta_Q(z,t)-t\log\beta\right]\leq 0\right\}\\
& \qquad\qquad \text{implies that}\\
Q(\beta): & \sup\limits_{P\in\mathfrak{P}}\mathrm{Prob}_P\left\{z_0+\sum\limits_{j=1}^d \xi_j z_j>0\right\}\leq\beta.
\end{array}
\tag{6.11}
$$

We are about to replace (6.11) with an equivalent and more convenient computationally implication:

$$
\begin{array}{ll}
\widehat{P}(\beta): & \left\{\inf\limits_{t>0}\left[\widehat{\Theta}(z,t)-t\log\beta\right]\leq 0\right\}\\
& \qquad\qquad \text{implies that}\\
Q(\beta): & \sup\limits_{P\in\mathfrak{P}}\mathrm{Prob}_P\left\{z_0+\sum\limits_{j=1}^d \xi_j z_j>0\right\}\leq\beta.
\end{array}
\tag{6.12}
$$

The advantage of (6.12) as compared to (6.11) 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 (6.12) 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 (6.12) and the equivalence of (6.11) and (6.12) are given by the following lemma.

Lemma 6.1 *Let $0 < \beta < 1$. Then the following holds:*

$$\widehat{P}(\beta) \text{ if and only if } P(\beta). \tag{6.13}$$

Proof. Implication $\Rightarrow$ in (6.13) is evident, since $\widehat{\Theta}(z,t) = \max\limits_{Q\in\mathcal{M}} \Theta_Q(z,t)$. Note that this implication combines with (6.11) to imply the validity of (6.12).

Now let us prove the implication $\Leftarrow$ in (6.13). 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

$$\min_{1\leq i\leq N} \theta(t_i,Q) \leq \gamma, \ \ \forall Q \in \mathcal{M}. \tag{6.14}$$

Since θ is concave and continuous in $Q \in \mathcal{M}$ and $\mathcal{M}$ is convex, (6.14) implies that

$$\exists \lambda^* \in \Delta_N := \{\lambda \in \mathbb{R}^N_+ : \sum_i \lambda_i = 1\} : \sum_i \lambda_i^* \theta(t_i,Q) \leq \gamma, \ \ \forall Q \in \mathcal{M}. \tag{6.15}$$

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 $\sum\limits_i \lambda_i\theta(t_i,Q) \leq \gamma$; 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 (6.14), the equality – by von Neumann Lemma; the required point in $\bigcap\limits_i \Lambda_{Q_i}$ is $\operatorname*{argmin}\limits_{\lambda\in\Delta_N} \max\limits_{Q\in Q_N} \sum\limits_{i=1}^N \lambda_i\theta(t_i,Q)$).

Since θ is convex in $t > 0$, setting $t_\gamma = \sum\limits_i \lambda_i^* t_i$ we get from (6.15) that $\Theta_Q(t_\gamma,z) - t_\gamma\log\beta \equiv \theta(t_\gamma,Q) \leq \sum\limits_i \lambda_i^*\theta(t_i,Q) \leq \gamma$ for all $Q \in \mathcal{M}$, whence $\widehat{\Theta}(t_\gamma,z) - t_\gamma\log\beta \equiv \max\limits_{Q\in\mathcal{M}} \Theta_Q(t_\gamma,z) - t_\gamma\log\beta \leq \gamma$. Since t_γ is positive by construction and $\gamma > 0$ is arbitrary, we conclude that $\inf\limits_{t>0}\left[\widehat{\Theta}(t_\gamma,z) - t_\gamma\log\beta\right] \leq 0$, so that $\widehat{P}(\beta)$ is valid. ■

Putting things together, we arrive at the following result.

Theorem 6.1 *Assume that the ambiguously chance constrained problem* (6.9) *satisfies Assumptions* B1 – B2*, and let α_i, $i = 1, ..., m$, be positive reals such that $\sum_i \alpha_i \le \alpha$. Then the program*

$$
\begin{array}{c}
\operatorname*{Min}_{x\in X} f(x) \ \text{ s.t. } \ \inf_{t>0} \underbrace{\left[f_{i0}(x) + t\widehat{\Psi}(t^{-1}z^i[x]) - t\log\alpha_i\right]}_{g_i(x,t)} \le 0,\ i = 1, ..., m, \\
z^i[x] \;=\; (f_{i1}(x), ..., f_{id}(x)),\ \ \widehat{\Psi}(z) = \max_{\{Q_j\}_{j=1}^d \in \mathcal{M}} \sum_{j=1}^d \log\left(\int \exp\{z_j s\} dQ_j(s)\right)
\end{array}
\tag{6.16}
$$

is a conservative approximation of problem (6.9)*: 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 X is 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

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

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 (6.10)). The inner functions $f_{i0}(x)$, $f_{ij}(x)$, $j \ge 1$, 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 (6.16) is indeed a convex program. Further, if x is feasible for (6.16), then $x \in X$, and for every i the predicate $\widehat{P}(\alpha_i)$ corresponding to $z = (f_{i0}(x), f_{i1}(x), ..., f_{id}(x))$ is valid, which, by (6.12), implies that

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

Since $\sum_i \alpha_i \le \alpha$, x is feasible for (6.9). ■

Remark 6.1 Assumption B1 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. In order not to care for nonnegativity of ξ_j's associated with non-affine $f_{ij}(\cdot)$, we assume from now on that all functions f_{ij}, $j = 1, ..., d$, in (6.9) 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 may assume w.l.o.g. 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

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

consequently, approximation (6.16) becomes

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

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

$$\operatorname*{Min}_{x\in X} f(x) \ \text{ s.t. } f_{i0}(x)+\sum_{j=1}^{d}|f_{ij}(x)|\le 0,\ i=1,...,m. \tag{6.17}$$

As it could be expected, in the situation in question, Bernstein approximation recovers the *Robust Counterpart* of the original uncertain problem [3], which in our case is the semi-infinite optimization program:

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

It is clear that in the extreme case we are considering the approximation is *exactly equivalent* to the chance constrained problem (6.9). A relatively good news about Bernstein approximation (6.17) is that in our example it is no more conservative than Robust Counterpart. It is immediately seen that this is a general fact: whenever Bernstein approximation (6.16) 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, both the chance constrained problem (6.9) 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 may further assume w.l.o.g. that ξ_j vary in $[-1,1]$ and have known expectations μ_j, $|\mu_j|\le 1$. 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 $|\mu|\le 1$, then

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

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

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

[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.

We therefore have

$$
\begin{array}{rcl}
\widehat{\Phi}(z) & := & \max\limits_{P\in\mathfrak{P}} \log\left(\mathbb{E}_P\left\{\exp\{z_0+\sum\limits_{j=1}^d \xi_j z_j\}\right\}\right)\\
& = & z_0+\sum\limits_{j=1}^d \log\left(\cosh(z_j)+\mu_j\sinh(z_j)\right) \le \widetilde{\Phi}(z) := z_0+\sum\limits_{j=1}^d\left[\mu_j z_j+\frac{z_j^2}{2}\right],\\
\widehat{\Theta}(z,t) & := & t\widehat{\Phi}(t^{-1}z) = z_0+\sum\limits_{j=1}^d t\log\left(\cosh(t^{-1}z_j)+\mu_j\sinh(t^{-1}z_j)\right)\\
& \le & \widetilde{\Theta}(z,t) = z_0+\sum\limits_{j=1}^d \mu_j z_j+\frac{1}{2t}\sum\limits_{j=1}^d z_j^2.
\end{array}
\tag{6.18}
$$

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 (6.16) of ambiguous chance constrained problem in question as the convex program

$$
\begin{array}{c}
\mathop{\mathrm{Min}}\limits_{x\in X}\left\{f(x): G(z^i[x],\alpha_i)\le 0,\ i=1,...,m\right\},\\
z^i[x]=(f_{i0}(x),f_{i1}(x),...,f_{id}(x))^T
\end{array}
\tag{6.19}
$$

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 (6.19) 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 (6.18). From the concluding inequality in (6.18) it follows that

$$
\begin{array}{rcl}
G(z,\beta) & := & \inf\limits_{t>0}\left[\widehat{\Theta}(z,t)-t\log\beta\right]\\
& \le & G_*(z,\beta) := \inf\limits_{t>0}\left[z_0+\sum\limits_{j=1}^d\mu_j z_j+\frac{1}{2t}\sum\limits_{j=1}^d z_j^2-t\log\beta\right]\\
& = & z_0+\sum\limits_{j=1}^d\mu_j z_j+\sqrt{2\log(1/\beta)}\left(\sum\limits_{j=1}^d z_j^2\right)^{1/2}.
\end{array}
\tag{6.20}
$$

It follows that the convex optimization program

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

is an approximation (more conservative than Bernstein one) of the ambiguous chance constrained problem (6.9), 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 (1.9)) and Remark 3.1.

[3)] It should be stressed that this bounding is completely irrelevant as far as numerical processing of (6.19) is concerned; the only purpose of the exercise to follow is to link our approach with some previously known constructions.

Comparing (6.17) and (6.19) – (6.20), 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 (6.17) are pointwise $\geq$ their counterparts in (6.19), so that (6.19) always is less conservative than (6.17). 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 (6.17) 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 (6.19), first, allows for $h_i(x)$ to belong to $V_1(x)$ (recall that (6.19) is less conservative than (6.17)) 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 (6.20) and take into account that $\mu_j \equiv 0$); by convexity, it follows that i-th constraint in (6.19) allows for $h_i(x)$ to belong to the set $V_{1,2}(x) = \mathrm{Conv}\{V_1(x) \cup V_2(x)\} \supset V_1(x)$. 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 (6.19) is "much less restricting" than its counterpart in (6.17). 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 (6.19) becomes progressively ("by factor $\sqrt{d}$") less conservative than (6.17).

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 comprised of distributions with support in a given finite segment (which we w.l.o.g. can assume to be $[-1,1]$). 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 notation, we present the expressions for $\exp\{\Lambda^{\Pi}(t)\}$ rather than for Λ^{Π} itself.

Example 3: "light tail" families. In previous examples, all distributions from Π were supported on a fixed finite segment. Now consider the case when Π is comprised of Borel probability distributions P on the axis such that $\mathbb{E}_P[\exp\{|x|^r/r\}] \leq \exp\{\sigma^r/r\}$, where $r \in (1,\infty)$ and $\sigma \in (0,\infty)$ are given

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

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

[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 $|\mu| \leq 1$, $\sigma^2 \leq 1$, $\mu^2 \leq \sigma^2$.

parameters. In this case, precise computations of $\Lambda^{\Pi}(t)$ seems to be difficult, but we can point out a tight convex upper bound on $\Lambda^{\Pi}(\cdot)$, specifically,

$$\Lambda^{\Pi}(t) \leq \left\{ \begin{array}{ll} \sigma|t|, & |t| \leq \sigma^{r-1} \\ \frac{\sigma^r}{r} + \frac{|t|^{r_*}}{r_*}, & |t| \geq \sigma^{r-1} \end{array} \right., \; r_* = \frac{r}{r-1}. \tag{6.21}$$

This bound coincides with $\lambda_{\Pi}(t)$ when $|t| \leq \sigma^{r-1}$ and coincides with $\Lambda^{\Pi}(t)$ within additive constant $-\log(1 - \exp\{-\sigma^r/r\})$ when $|t| \geq \sigma^{r-1}$.

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 6.1 are severely violated, but Bernstein approximation scheme still works.

Example 4: 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 $i = 0, 1, \ldots$ 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 B1: convexity in the space of parameters has small in common with convexity in the space of distributions. For example, when the mean 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 6.1. 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 is always the case for the logarithmic moment generating function, and is concave in λ, which is a 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 6.1 and (6.16) to conclude that the optimization problem

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

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).

Acknowledgement. We express our gratitude to Yuri Kan who brought to our attention the paper of Pinter [20].

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] Dentcheva, D., Prekopa, A., Ruszczynski, A., "Concavity and efficient points of discrete distributions in probabilistic programming", *Math. Prog.* **89** (2000), 55-77

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

[12] Klein Haneveld, W.K., *Duality in Stochastic Linear and Dynamic Programming*, Lecture Notes in Economics and Mathematical systems 274, Springer, 1986.

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

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

[15] 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.

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

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

[18] 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.

[19] 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.

[20] Pinter, J., Deterministic approximations of probability inequalities, *ZOR Methods and Models of Operations Research, Series Theory*, 33 (1989), 219239.

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

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

[23] Prékopa, A., Vizvari, B., Badics, T., Programming under probabilistic constraint with discrete random variables – in: L. Grandinetti et al. (Eds.), *New Trends in Mathematical Programming*, Kluwer, 1997, 235-257.

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

[25] 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.

[26] 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.

[27] 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.

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

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

Appendix

Lemma 7.2 *Let $\sigma > 0$, and let ξ_j, $j = 1, ..., d$, be independent random variables taking values in $[-1/2, 1/2]$ such that* $\mathrm{Mean}\{\xi_j\} = 0$ *and* $\mathrm{Mean}\{\xi_j^2\} \geq \sigma^2$ *for all j. Assume that a half-space $\{x \in \mathbb{R}^d : e^T x \leq \mu\}$ with e normalized by $\|e\|_2 = 1$ is such that*

$$\mathrm{Prob}\left\{\xi : e^T\xi > \mu\right\} \leq \epsilon := \frac{\sigma^2}{24\log(1/\sigma)}. \tag{7.22}$$

Then

$$\mu \geq h := \frac{\sigma^2}{24\sqrt{\log(1/\sigma)}}. \tag{7.23}$$

Proof. Let $\zeta = \sum_j e_j\xi_j$.

1^0. Observe that

$$\forall\gamma : \mathrm{Mean}\left\{\exp\{\gamma\zeta\}\right\} \leq \exp\{\gamma^2/8\} \tag{7.24}$$

Indeed, $|\xi_j| \leq 1/2$ implies that $\exp\{\gamma e_j\xi_j\} \leq \cosh(\gamma e_j/2) + \sinh(\gamma e_j/2)\xi_j$ (note that the left hand side in this inequality is convex in ξ_j, while the right hand side is linear in ξ_j, so that the inequality is valid for $|\xi_j| \leq 1/2$ if and only if it is valid for $\xi_j = \pm 1/2$, which indeed is the case). Taking expectations of both sides and invoking $\mathrm{Mean}\{\xi_j\} = 0$, we get $\mathrm{Mean}\left\{\exp\{\gamma e_j\xi_j\}\right\} \leq \cosh(\gamma e_j/2) \leq \exp\{\gamma^2 e_j^2/8\}$ (the concluding inequality is straightforward). We now have

$$\mathrm{Mean}\left\{\exp\{\gamma\zeta\}\right\} = \prod_{j=1}^{n} \mathrm{Mean}\left\{\exp\{\gamma e_j\xi_j\}\right\} \leq \prod_{j=1}^{n} \exp\{\gamma^2 e_j^2/8\} = \exp\{\gamma^2/8\},$$

as required in (7.24).

2^0. From (7.24), by Tschebyshev inequality, it follows that when $T > 0$, we have $\mathrm{Prob}\{\zeta \geq T\} \leq \inf\limits_{\gamma>0} \exp\{-\gamma T\}\exp\{\gamma^2/8\} = \exp\{-2T^2\}$, and similarly for $\mathrm{Prob}\{\zeta \leq -T\}$. Thus,

$$T \geq 0 \Rightarrow P(T) := \mathrm{Prob}\{\zeta \geq T\} \leq \exp\{-2T^2\},\ \mathrm{Prob}\{\zeta \leq -T\} \leq \exp\{-2T^2\}. \tag{7.25}$$

3^0. Assume, on the contrary to what should be proved, that $\mu < h$, and let

$$J_+ = \mathrm{Mean}\,\{\zeta_+ := \max[0,\zeta]\} = \int\limits_0^\infty t(-dP(t))dt.$$

Since $\mu < h$ and $\mathrm{Prob}\{\zeta > \mu\} \leq \epsilon$, invoking (7.25) we have

$$t \geq h \Rightarrow P(t) \leq \min[\epsilon, \exp\{-2t^2\}],$$

whence, setting $H = \sqrt{\log(1/\epsilon)/2}$, so that $\epsilon = \exp\{-2H^2\}$:

$$\begin{array}{l} J_+ \leq h + \int\limits_h^\infty t(-dP(t)) = h + hP(h) + \int\limits_h^\infty P(t)dt \leq h + h\epsilon + \int\limits_h^\infty \min[\epsilon, \exp\{-2t^2\}]dt \\ = h + h\epsilon + \int\limits_h^H \epsilon dt + \int\limits_H^\infty \exp\{-2t^2\}dt \leq h + \epsilon H + \int\limits_H^\infty \exp\{-2t^2\}\frac{4t}{4H}dt = h + \epsilon H + \frac{1}{4H}\exp\{-2H^2\}. \end{array}$$

Recalling the definition of H and noting that $h = \epsilon H\sqrt{2}$ we conclude that

$$J_+ \leq \epsilon\left[\sqrt{2}H + H + \frac{1}{4H}\right].$$

Now, we clearly have $\sigma \leq \frac{1}{2}$, whence $\epsilon \leq \frac{1}{96\log 2} \leq \frac{1}{66.5}$ and therefore $H = \sqrt{\log(1/\epsilon)/2} \geq 1.448$. We conclude that

$$J_+ \leq \left[\sqrt{2} + 1 + 1/(4H^2)\right]\epsilon H \leq 1.792\epsilon\sqrt{\log(1/\epsilon)}; \tag{7.26}$$

since ζ is with zero mean along with all ξ_j, we conclude that

$$J_- : \mathrm{Mean}\,\{\zeta_- := \max[0,-\zeta]\} \leq 1.792\epsilon\sqrt{\log(1/\epsilon)}. \tag{7.27}$$

4^0. Now, for every $T > 0$ we have

$$\begin{array}{l} \sigma_+^2 := \mathrm{Mean}\{\zeta_+^2\} = \int\limits_0^\infty t^2(-dP(t)) \leq T\int\limits_0^T t(-dP(t)) + \int\limits_T^\infty t^2(-dP(t)) \leq TJ_+ + \int\limits_T^\infty t^2(-dP(t)) \\ = TJ_+ + T^2P(T) + \int\limits_T^\infty 2tP(t)dt \leq TJ_+ + T^2\exp\{-2T^2\} + \int\limits_T^\infty 2t\exp\{-2t^2\}dt \\ = TJ_+ + [T^2 + 1/2]\exp\{-2T^2\} \end{array}$$

and similarly for $\sigma_-^2 := \mathrm{Mean}\{\zeta_-^2\}$ with replacing J_+ with J_-. Bounding $J_\pm$ via (7.26), (7.27), we arrive at the relation

$$\sigma^2 \leq \mathrm{Mean}\{\zeta\}^2 = \sigma_+^2 + \sigma_-^2 \leq 3.584T\epsilon\sqrt{\log(1/\epsilon)} + (2T^2+1)\exp\{-2T^2\}\ \forall T > 0 \tag{7.28}$$

(the very first inequality is given by $\mathrm{Mean}\{\zeta^2\} = \sum\limits_j e_j^2\mathrm{Mean}\{\xi_j^2\} \geq \sum\limits_j e_j^2\sigma^2 = \sigma^2$). Substituting $T = \sqrt{\log(1/\epsilon)/2}$, we get

$$\sigma^2 \leq \left[(3.584/\sqrt{2} + 1)\log(1/\epsilon) + 1\right]\epsilon \leq 3.78\epsilon\log(1/\epsilon);$$

it is immediately seen that when $\sigma \leq 1/2$ and ϵ is as defined in Lemma, the latter inequality in fact fails to be true, which is the desired contradiction. ■

Justification of Example 3. It suffices to verify the inequality (6.21) when $t \geq 0$. Let $P \in \Pi$. First of all, we have $\frac{|x|^r}{r} + \frac{t^{r_*}}{r_*} - tx \geq 0$ for all x, whence $\int \exp\{tx\}dP(x) \leq \int \exp\{\frac{|x|^r}{r} + \frac{t^{r_*}}{r_*}\}dP(x) \leq \exp\{\frac{\sigma^r}{r} + \frac{t^{r_*}}{r_*}\}$; thus, bound (6.21) holds true when $t \geq \sigma^{r-1}$. Now let us prove that the bound holds true when $0 \leq t \leq \sigma^{r-1}$. Since it clearly holds true when $t = 0$ and is linear in t, while $\Lambda^{\Pi}(t)$ is convex in t, it suffices to verify that the bound holds true when $t = \sigma^{r-1}$. This we already know, since with $t = \sigma^{r-1}$ we have $\frac{\sigma^r}{r} + \frac{t^{r_*}}{r_*} = t\sigma$. Observe that when $0 \leq t \leq \sigma^{r-1}$ our upper bound coincides with $\Lambda^{\Pi}(t)$ – look what happens when P assigns mass 1 to the point $x = \sigma$. Finally, let $t > \sigma^{r-1}$, and let P be the distribution which assigns the mass $\mu = \lambda \exp\{\frac{\sigma^r - t^{r_*}}{r}\}$ to the point t^{r_*-1} and the mass $1 - \mu$ to the point 0; here $\lambda = \frac{1-\exp\{-\sigma^r/r\}}{1-\exp\{-t^{r_*}/r\}}$. Since $t \geq \sigma^{r-1}$, we have $t^{r_*} \geq \sigma^r$, so that $\lambda \leq 1$ and $\mu \in [0,1]$; thus, P indeed is a probability distribution. An immediate computation shows that $\int \exp\{|x|^r/r\}dP(x) = \exp\{\sigma^r/r\}$, so that $P \in \Pi$. We now have $\int \exp\{tx\}dP(x) \geq \mu \exp\{t^{r_*}\} = \lambda \exp\{\sigma^r/r + t^{r_*}/r_*\}$, so that $\Lambda^{\Pi}(t) \geq \frac{\sigma^r}{r} + \frac{t^{r_*}}{r_*} - \log\lambda \geq \frac{\sigma^r}{r} + \frac{t^{r_*}}{r_*} - \log(1 - \exp\{-\sigma^r/r\})$. ■

Response to Referees

A. We are extremely grateful to the Referees for their effort and valuable comments. We did our best to revise the paper along the lines suggested by the Referees. The major changes are as follows:

1. In Introduction, we added references recommended by Referee 1, and reference recommended by Referee 3, same as shortened and streamlined the discussion of intrinsic difficulties associated with chance constraints, mentioned the convexity results known for chance constrained problems, and made some other changes.

2. Following one of comments of Referee 3, we replaced Proposition 1 with a much stronger statement which does not assume neither symmetry, nor uniformity of the distributions in question.

3. We changed the model considered in numerical example, as suggested by Referees 1 and 3; now the returns are log-normal. As suggested by Referee 3, we "lifted" the explanation of tuning to the main body of the paper.

4. In Section 6, we have eliminated the "mixed uncertainty model" and now focus solely on Bernstein approximation of ambiguous chance constraints.

B. There is a number of comments which we either did not understand, or did not agree with. The comments of the first type are as follows:

1. Both Referee 1 and Referee 3 claim that when processing a chance constraint

$$p(x) \equiv \mathrm{Prob}\{\underbrace{x_0 + \sum_{i=1}^{n} x_i \xi_i}_{F(x,\xi)} > 0\} \leq \alpha \qquad (*)$$

   with ξ_i uniformly distributed in $[0,1]$ and independent of each other, there is no necessity in approximation: the constraint defines a convex set in the space of x's, provided that $\alpha \leq 1/2$. We, however, do believe that approximations make perfect sense here: the situation in question "as it is" is in fact intractable, with intractability coming not from lack of convexity, but from *impossibility to compute* $p(x)$ *efficiently*. Indeed, a well-known result of L. Khachiyan states that computing the volume of a convex polytope, even as simple as the intersection of a half-space and the unit box (note that $p(x)$ is exactly the volume of such an intersection) is an NP-hard problem. In particular, *unless* P=NP, *there is no algorithm which, given on input rational data* x_i *and* $\epsilon > 0$, *can approximate* $p(x)$ *within accuracy* ϵ *in time polynomial in* $\log(1/\epsilon)$ *and the total bit length of* x_i*'s.* (Since the original paper of Khachiyan was published in USSR and thus is not readily available, we attach to this response a simple proof of the cited statement.) We see that "closed form formula for the convolution of uniform distributions" mentioned by Referee 1 (and any other formula, for this matter) does not help in processing chance constraints with uniform distributions, and there is nothing strange in it: not every "closed form formula" allows for efficient computation... We should add that in contrast to $p(x)$ itself, the Bernstein upper bound on this quantity is efficiently computable.

2. Referee 2 states that when ξ_i in $(*)$ are discrete random variables taking 2 values each, then the calculation of $\tau + \frac{1}{\alpha}\mathbb{E}[[F(x,\xi) - \tau]_+]$ requires solving a finite linear program. We do not understand why this is relevant: the same arguments as in the proof of Khachiyan's theorem demonstrate that when ξ_i are independent and take values ± 1 with probabilities $1/2$, the quantity in question does *not* admit efficient computation. "Finite" is by far not the same as efficient... By similar reasons we do not understand the statement of Referee 2 "However, if the ξ_j take only finitely many values, the situation is quite different: Independence seems not to be important any more and other algorithms can work".

3. We do not understand the remark of Referee 1 "...since the number of random variables in the model is large, the probability distribution of the sum can be fairly accurately approximated by a normal distribution...". It would be true if we were speaking about *sum*, while in fact we are speaking about a *weighted sum* $F(x,\xi)$, with the weights of the random variables being functions of our design variables. It well may happen that while the weighted sum in $(*)$ contains hundreds or thousands of terms, at a "point of interest" (e.g., at the optimal solution obtained when replacing actual random variables with Gaussian ones) the sum is dominated by few terms, like 1 or 5; in this case, there is no reason to trust in Gaussian approximation...

Essentially, the only comment we did understand and do not agree with, is eliminating the section on Bernstein bound for ambiguous chance constraints (Referee 3). We believe extending the Bernstein bound on the case of ambiguous chance constraints is natural and useful. Indeed, the major advantage of Bernstein bound as compared to much more universal Scenario approximation is the possibility to process the "high reliability" case of α like 1.e-3, 1.e-4 and less. To require that high reliability makes sense only when we believe that the accuracy to which we know the underlying probability distribution allows to make valid conclusions on probabilities of the outlined order of magnitude, while in reality that accurate knowledge could be problematic. Thus, in our opinion, high reliability chance constraints in many cases make sense only "taken together with ambiguity".

We should add that while we did eliminate the story about mixed uncertainty, the only reason was to make the paper shorter and more "streamlined", and not the arguments of Referee 3 against this uncertainty model. We do believe that "mixed uncertainty" is meaningful, e.g., it seems to be the best way to account for small inaccuracies in the parameters of distributions (e.g., means) which, we believe, cannot be avoided in reality.

Attachment

Consider the following problem:

(U) Given a positive integer n, rational x and rational a_i, $i = 1, ..., n$, and $\epsilon > 0$, compute within accuracy ϵ the probability

$$p(a) = \text{Prob}\left\{-x + \sum_{i=1}^{n} a_i \xi_i > 0\right\},$$

where ξ_i are uniformly distributed on $[0,1]$ and independent of each other random variables.

Theorem 7.2 *Unless* P=NP*, there is no way to compute* $p(a)$ *within accuracy* ϵ *efficiently, that is, in time polynomial in the total bit length* $\text{bitlength}(x) + \sum_i \text{bitlength}(a_i)$ *of the data* x, a *and in* $\log(1/\epsilon)$*.*

Proof. 1^0. We start with the well-known fact: the problem

Stones: Given a positive integer integer n and n positive integers $a_1, ..., a_n$, check whether the equation $\sum_i x_i a_i = 0$ has a solution with $x_i \in \{-1; 1\}$

is NP-complete. We now make the following observation: Problem *Stones** which differs from *Stones* by the only restriction that all given integers a_i are odd also is NP-complete.

Indeed, let us reduce *Stones* to *Stones**. To this end, given data $\{a_i\}$ of an instance of *Stones*, let us set $n_+ = 2n$ and

$$a_i^+ = \begin{cases} 4na_i + 1, & i = 1, ..., n \\ 1, & i = n+1, ..., 2n \end{cases}$$

and consider the *Stones** problem with the data $\{a_i^+\}_{i=1}^{n_+}$. We claim that the equations

$$\begin{array}{ll} (a) & \sum\limits_{i=1}^{n} a_i x_i = 0 \\ (b) & \sum\limits_{i=1}^{n_+} a_i^+ x_i = 0 \end{array}$$

simultaneously have or have not ± 1-solutions, so that *Stones* indeed is polynomially reducible to *Stones**. Indeed, if (b) has a solution $\{\bar{x}_i \in \{-1; 1\}\}_{i=1}^{2n}$, then

$$4n\left(\sum_{i=1}^{n} \bar{x}_i a_i\right) = -\sum_{i=1}^{2n} \bar{x}_i;$$

since the absolute value of the right hand side is $\leq 2n$ and $\sum\limits_{i=1}^{n} \bar{x}_i a_i$ is an integer, the equality may take place only when $\sum\limits_{i=1}^{n} \bar{x}_i a_i = 0$, that is, (a) is solvable. Vice versa, let $\{\bar{x}_i \in \{-1,1\}\}_{i=1}^n$ be a solution to (a). Setting $\bar{x}_{n+i} = -\bar{x}_i$, $i = 1,...,n$, we get $\bar{x}_i \in \{-1,1\}$ for all $i \leq 2n$ and

$$\sum_{i=1}^{2n} a_i^+ \bar{x}_i = 4n \underbrace{\sum_{i=1}^{n} a_i \bar{x}_i}_{=0} + \sum_{i=1}^{2n} \bar{x}_i = 0,$$

where the last equality is due to $x_{n+i} = -x_i$, $i = 1,...,n$. Thus, (b) has a solution. $\square$

2^0. In view of the just established NP-completeness of *Stones**, in order to prove Theorem it suffices to demonstrate that if $p(a)$ can be computed within accuracy ϵ in time polynomial in $\mathrm{bitlength}(a) + \sum\limits_{i=1}^{n} \mathrm{bitlength}(a_i)$ and $\log(1/\epsilon)$, then *Stones** is polynomially solvable. To this end, assume that $p(a)$ indeed admits efficient in the aforementioned sense computation. Given the data n, $a_1, ..., a_n$ of an instance of *Stones**, let L be the bit length of the data. Observe that the function

$$p_a(x) = \mathrm{Prob}\left\{\sum_{i=1}^{n} a_i \xi_i > x\right\}$$

in an open interval $(k, k+1)$, k being an integer, is a polynomial of degree at most $n-1$, and the $(n-1)$-st derivative of this polynomial is, up to the factor A^{-1}, $A = \prod\limits_{i=1}^{n} a_i$, the quantity

$$\psi_k = \int\limits_{k+1/2}^{\infty} (\delta(s-a_1) - \delta(s)) * (\delta(s-a_2) - \delta(s)) * ... * (\delta(s-a_n) - \delta(s)) ds,$$

where $\delta(\cdot)$ is the Dirac δ-function and $*$ stands for convolution. In other words, the quantity $\psi_k = Ap_a^{(n-1)}(x)$ for $x \in (k, k+1)$ is independent of x and can be described as follows.

We form the set $\mathcal{L}$ of all integers ℓ which can be represented as $a_{i_1} + a_{i_2} + ... + a_{i_\nu}$, where $0 \leq \nu \leq n$ and $1 \leq i_1 < i_2 < ... < i_\nu \leq n$. Every such a representation of $\ell \in \mathcal{L}$ is assigned weight 1, if $n - \nu$ is even, and weight -1, if $n - \nu$ is odd; $\ell \in \mathcal{L}$ itself is assigned the weight $w(\ell)$ which is the total weight of all its representations. Finally, $\psi_k \equiv Ap_a^{(n-1)}(x)$, $x \in (k, k+1)$, is the total weight of those $\ell \in \mathcal{L}$ which are $\geq k+1$, let this total weight be $W(k+1)$.

Now, we have assumed that there exists a possibility of efficient computing $p_a(x)$ at a rational x within accuracy ϵ, in the sense that the complexity of the computation is polynomial in L, the bit length of x and in $\log(1/\epsilon)$. Invoking finite differences and taking into account that $n \leq L$, we conclude that there is a possibility to compute ψ_k within accuracy ϵ with complexity $\mathrm{Poly}(L, \log(1/\epsilon), \mathrm{bitlength}(k))$ (that is, polynomial in L, $\log(1/\epsilon)$ and the bit length of the integer k). Now note that for every $\ell \in L$ one has $|w(\ell)| \geq 1$. Indeed, since all a_i, $i \geq 1$, are odd, a number ℓ which can be represented as $a_{i_1} + ... + a_{i_\nu}$ with $1 \leq i_1 < ... < i_\nu \leq n$ "remembers" whether ν is even or odd: ℓ is even iff ν is so. With this observation, $|w(\ell)|$ is exactly the number of representations of ℓ as $a_{i_1} + ... + a_{i_\nu}$. Recalling that $\psi_k = \sum\limits_{\ell \in \mathcal{L}, \ell \geq k+1} w(\ell)$, we conclude that *for a positive integer k, the quantities ψ_k and ψ_{k+1} are either equal to each other (this is the case when $k+1 \notin \mathcal{L}$), or differ by at least 1 (this is the case when $k+1 \in \mathcal{L}$).* In particular, we can check whether the system $\sum_{i=1}^{n} a_i x_i = 0$ has or has not a ± 1 solution, or, which is the same, whether the number $k_* = \frac{1}{2}\sum_{i=1}^{n} a_i$ belongs or does not belong to $\mathcal{L}$, as follows. We compute k_* and check whether it is integer; if not, the required answer is "No". Otherwise we compute ψ_{k_*-1} and ψ_{k_*} within accuracy 0.1; by the above, $k_* \in \mathcal{L}$ iff the results of these computations differ by at least 0.5. It remains to note that $\mathrm{bitlength}(k_*) \leq \mathrm{Poly}(L)$, so that the complexity of our test is polynomial in L. Thus, efficient computability $p_a(\cdot)$ would imply polynomial time solvability of NP-complete problem *Stones**. ■