2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Randomized min-max optimization: the exact risk of multiple cost levels

Algo Carè, Simone Garatti, Marco C. Campi

Abstract— **In this paper, we present a theoretical result that applies to convex optimization problems in the presence of an uncertain stochastic parameter. We consider the min-max sample-based solution, i.e. the min-max solution computed over a finite sample of instances of the uncertain stochastic parameter, and the costs incurred by this solution in correspondence of the sampled parameter instances. Our goal is to evaluate the risks associated to the various costs, where the risk associated to a cost is the probability that the cost is exceeded when a new uncertainty instance is seen. The theoretical result proven in this paper is that the risks form a random vector whose probability distribution is always an ordered Dirichlet distribution, irrespective of the probability measure of the uncertain stochastic parameter. This evaluation characterizes completely the risks associated to the costs, and represents a full-fledged result on the reliability of the min-max sample-based solution.**

Index Terms— **Randomized Algorithms, Optimization, Optimization Algorithms.**

I. INTRODUCTION

In this paper, we consider uncertain optimization problems where a decision, modeled as the selection of a variable x belonging to a convex set $\mathscr{X} \subseteq \mathbb{R}^d$, has to be made so as to minimize a cost function which depends on a random uncertainty parameter too. Precisely, the uncertain parameter is a random element δ which takes values in a set Δ according to a probability measure $\mathbb{P}$, and the cost function is $f(x,\delta)$. Throughout we will assume that, for every δ, $f(x,\delta)$ is convex in x. The presence of δ in f implies that any choice of x is associated to a variety of possible costs, depending on the value assumed by δ.

A well-known rule to make a decision in this context is that of considering N "scenarios", i.e. N instances of δ, say $\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$, independent and identically distributed (i.i.d.) according to $\mathbb{P}$, and minimizing the worst-case cost over these scenarios, that is:

$$\text{min-max}_N: \quad \min_{x\in\mathscr{X}\subseteq\mathbb{R}^d} \max_{i=1,\ldots,N} f(x,\delta^{(i)}). \qquad (1)$$

Algo Carè is with the Dipartimento di Ingegneria dell'Informazione, Università di Brescia, via Branze 38, 25123 Brescia, Italia. E-mail: algo.care@ing.unibs.it, website: http://bsing.ing.unibs.it/algo.care/

Simone Garatti is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italia. E-mail: sgaratti@elet.polimi.it, website: http://home.dei.polimi.it/sgaratti/

Marco Campi is with the Dipartimento di Ingegneria dell'Informazione, Università di Brescia, via Branze 38, 25123 Brescia, Italia. E-mail: marco.campi@ing.unibs.it, website: http://bsing.ing.unibs.it/campi/

The optimal solution of (1) is indicated by x^* and can be obtained by rewriting (1) in epigraphic form as

$$\text{EPI}_N: \quad \min_{l\in\mathbb{R}, x\in\mathscr{X}\subseteq\mathbb{R}^d} l$$
$$\text{subject to: } f(x,\delta^{(i)}) \le l, \quad i=1,\ldots,N, \quad (2)$$

and then by resorting to standard numerical solvers, [5]. See Table I for some examples of min-max problems arising in various applicative contexts.

Since x^* is computed based on a finite sample of scenarios only, the main concern with x^* is that of assessing its reliability with respect to the whole set of uncertainty instances Δ, and a possible indicator to this purpose seems to be $l^* = \max_{i=1,\ldots,N} f(x^*,\delta^{(i)})$, i.e. the worst cost among the seen scenarios. l^*, however, is just an *empirical* quantity, and it is clear that l^* is meaningful only if an assessment of the risk that a new uncertainty instance δ carries a cost $f(x^*,\delta)$ greater than l^* is provided. Quantitatively, this entails to study the variable $R = \mathbb{P}\{\delta\in\Delta : f(x^*,\delta) > l^*\}$, which is called the risk associated to l^*. This problem has been extensively studied, and the literature spans a wide range that goes from the analysis of asymptotic properties (i.e. for $N\to\infty$), see e.g. [28], to finite sample characterizations both for convex, [6], [7], [9], and non-convex cost function f, [1]. This work is in the vein of the so-called *theory of the scenario approach*, [6], [7], [9], which under the assumption that $f(x,\delta)$ is convex in x provides the sharpest possible characterization of the risk R. Note that R is a random variable since it depends on x^* and l^*, which in turn depend on the random samples $\delta^{(1)},\ldots,\delta^{(N)}$. The fundamental result of [9] is that, irrespective of $\mathbb{P}$, the probability distribution function of R is always equal to or bounded by a Beta probability distribution with parameters $d+1$ and $N-d$ (recall that d is the dimension of the decision variable). Based on this result, it can be shown that R not only tends to zero with probability 1 as $N\to\infty$, but also that, for finite N, R keeps smaller than threshold $\frac{1}{N}\left(d+\ln\frac{1}{\beta}+\sqrt{2d\ln\frac{1}{\beta}}\right)$ with confidence $1-\beta$, [2]. In other words, the risk associated to the worst-case cost l^* can be kept under control thanks to the knowledge of its probabilistic behavior.

Despite the sharp theoretical result offered by the theory of the scenario approach, limiting to l^* to characterize the reliability of x^* may turn out to be conservative, and it is advisable to seek other indicators besides l^*.

The idea is to consider the whole set of costs $f(x^*,\delta^{(i)})$ associated to the various scenarios $\delta^{(1)},\delta^{(2)},\ldots,\delta^{(N)}$. In the

TABLE I

A FEW EXAMPLES OF MIN-MAX PROBLEMS.

	Interpretation of δ	Interpretation of x	Interpretation of $f(x,\delta)$	References
Linear regression theory	Data point	Coefficients weighting regressor functions	Regression error	[18], [8]
Investment theory	Asset return	Proportion of the assets in a portfolio	Investment loss	[22], [26]
Control theory	Disturbance Realization	Controller parameters	Output variance	[11], [7]

following, these costs will be indicated by $l^*_{d+1}, l^*_{d+2}, \ldots, l^*_N$, see Fig. 1.

As is clear from an intuitive point of view, $l^*_{d+1}, l^*_{d+2}, \ldots, l^*_N$ all together provide a more sensible characterization of x^* than by using l^* only. Assume for instance that the gap between the maximum cost l^* and the second greatest cost and, similarly, other gaps between costs are large. Then, one expects that a new δ carries a cost $f(x^*,\delta)$ which is significantly smaller than l^* with a high probability. On the contrary, when the values $f(x^*,\delta^{(i)})$ concentrate all around l^*, it is apparent that $f(x^*,\delta)$ will be almost always close to l^*. See also [19].

In order to put such kind of reasoning on a solid quantitative ground, the risk R_k associated to the cost level l^*_k, i.e. the probability to observe an uncertainty instance δ carrying a cost higher than l^*_k, must be evaluated simultaneously for $k = d+1, \ldots, N$. At the present time, however, the theory of the scenario approach applies to the sole l^* and does not provide any characterization of the risks associated to other cost levels. This paper aims to fill this gap by computing the *joint probability distribution* of all risks $R_{d+1}, R_{d+2}, \ldots, R_N$. Our main achievement is that, no matter what the probability measure $\mathbb{P}$ is, such joint probability distribution is a so-called *ordered Dirichlet distribution* whose parameters depend on the number of samples N and the number of decision variables d only. This result completely characterizes the variables R_k's and can be suitably employed to support decisions in many real cases even for small sizes of N.

To sum up, two kinds of quantities are central in the characterization of the reliability of x^*:

$$\begin{pmatrix} l^*_{d+1} \\ l^*_{d+2} \\ \vdots \\ l^*_N \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} R_{d+1} \\ R_{d+2} \\ \vdots \\ R_N \end{pmatrix},$$

i.e. the vectors of cost levels and of the associated risks. While the cost levels are known as soon as the optimal decision variable x^* is computed, the corresponding risks are hidden to the decision maker. Nevertheless, their joint probability distribution is known given the theory so that the risks can be kept under control. In particular, since the ordered Dirichlet distribution is thin tailed, the risks can be bounded with high confidence by dropping the tails of the probability distribution. This way, a complete characterization of the reliability of x^* is obtained.

Since the characterization of the reliability of x^* holds irrespective of the probability measure $\mathbb{P}$, this characterization of x^* can play a prominent role in *data-driven optimization* ([3]) where the designer has no knowledge of the uncertainty other than through the scenarios $\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ obtained as observations in a procedure of real-data acquisition.

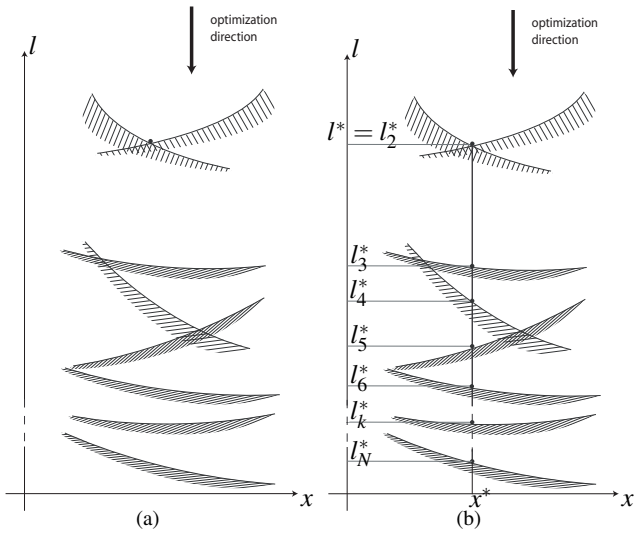

Fig. 1. On the left, a pictorial representation of the optimization problem (2), where each scenario $\delta^{(i)}$ corresponds to a constraint of the form $f(x,\delta^{(i)}) \le l$. On the right, the cost levels of x^* are put in evidence.

Structure of the Paper

In the next section our result is formally stated and some relevant aspects are discussed. Section III provides a numerical example.

II. MAIN RESULT

We first give the formal definition of cost levels and risks.

Definition 1 (Cost levels): We define the *cost levels* – or just *costs* – of the optimal decision variable x^* as $l^*_k = \max\{l \in \mathbb{R} : f(x^*,\delta^{(i)}) \ge l$ for a choice of k indexes i among $\{1,\ldots,N\}\}$, for $k = d+1,\ldots,N$. ∗

Note that the subscript k of l^*_k expresses the number of scenarios carrying a cost higher than or equal to l^*_k in correspondence of x^*.

Definition 2 (Risk): The *risk* of the cost level l^*_k of the optimal decision variable x^* is indicated with R_k, and defined as $R_k = \mathbb{P}\{\delta \in \Delta : f(x^*,\delta) > l^*_k\}$, for $k = d+1,\ldots,N$. ∗

Our main result is given under the following condition.

Condition 1: For every fixed integer m and fixed given instances $\delta^{(1)},\ldots,\delta^{(m)}$, the optimal solution to

$$\text{min-max}_m : \quad \min_{x \in \mathscr{X} \subseteq \mathbb{R}^d} \max_{i=1,\ldots,m} f(x,\delta^{(i)})$$

exists and is unique. ∗

This condition can be relaxed, see e.g. the discussion provided in [9], Section 2.1, but we here prefer to maintain it to avoid technical complications.

The exact evaluation in Theorem 1 of the joint probability distribution function is obtained under the following additional Condition 2. If Condition 2 does not hold, the second part of Theorem 1 applies.

Condition 2: $\forall x \in \mathscr{X}, \forall l \in \mathbb{R}, \quad \mathbb{P}\{\delta \in \Delta, \quad f(x,\delta) = l\} = 0.$ $*$

Condition 2 basically requires that, for every fixed x, the values of the cost functions $f(x,\delta)$ in correspondence of various δ's do not accumulate over the same point, and it is satisfied by problems whose constraints are "generically distributed".

Theorem 1: Under Conditions 1 and 2, the joint probability distribution function of $R_{d+1},\dots,R_N$ is as follows:

$$\begin{aligned}\mathbb{P}^N\{R_{d+1} \le \varepsilon_{d+1}, R_{d+2} \le \varepsilon_{d+2}, R_{d+3} \le \varepsilon_{d+3}, \dots, R_N \le \varepsilon_N\} \\ = \frac{N!}{d!}\int_0^{\varepsilon_{d+1}} r_{d+1}^d \int_0^{\varepsilon_{d+2}} \int_0^{\varepsilon_{d+3}} \cdots \int_0^{\varepsilon_N} \mathbb{1}\{0 \le r_{d+1} \le \cdots \\ \cdots \le r_N \le 1\} \mathrm{d}r_N \cdots \mathrm{d}r_{d+3} \mathrm{d}r_{d+2} \mathrm{d}r_{d+1}, \end{aligned} \tag{3}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. Moreover, under Condition 1 only, the joint probability distribution function is lower bounded by the right-hand side of (3):

$$\begin{aligned}\mathbb{P}^N\{R_{d+1} \le \varepsilon_{d+1}, R_{d+2} \le \varepsilon_{d+2}, R_{d+3} \le \varepsilon_{d+3}, \dots, R_N \le \varepsilon_N\} \\ \ge \frac{N!}{d!}\int_0^{\varepsilon_{d+1}} r_{d+1}^d \int_0^{\varepsilon_{d+2}} \int_0^{\varepsilon_{d+3}} \cdots \int_0^{\varepsilon_N} \mathbb{1}\{0 \le r_{d+1} \le \cdots \\ \cdots \le r_N \le 1\} \mathrm{d}r_N \cdots \mathrm{d}r_{d+3} \mathrm{d}r_{d+2} \mathrm{d}r_{d+1}. \end{aligned} \tag{4}$$

$*$

Proof: See [12]. The proof is also available on request from the authors. ■

It is worth noticing that the right-hand side of (3) and of (4) do not depend on $\mathbb{P}$. Moreover, since the result holds with equality for the class of problems satisfying Conditions 1 and 2 (equation (3)), the result of the theorem is tight.

A. Practical use of Theorem 1

Theorem 1 can be applied in various ways. The following two ways are especially useful in contexts where the uncertainty instances $\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)}$ come as observations obtained from a data acquisition experiment.

Post-experiment analysis

The user has collected N samples of δ and has solved problem min-max$_N$ obtaining x^* and the corresponding cost levels l_k^*, $k = d+1,\dots,N$. He fixes a confidence parameter $\beta \in (0,1)$ to a very small value, e.g. $\beta = 10^{-5}$ or even $\beta = 10^{-10}$, and determines a vector $\boldsymbol{\varepsilon} = (\varepsilon_{d+1},\dots,\varepsilon_N)$ such that the right-hand side of (4) is bigger than or equal to $1-\beta$. Then, the user can claim with high confidence $1-\beta$ that, simultaneously for $k = d+1,\dots,N$, the risk R_k of each cost level l_k^* is no larger than ε_k.

Experiment design

The user fixes a very small $\beta \in (0,1)$, e.g. $\beta = 10^{-5}$ or even $\beta = 10^{-10}$. Then he fixes a vector of m nondecreasing elements $0 \le \varepsilon_{d+1} \le \cdots \le \varepsilon_{d+m} \le 1$ corresponding to the desired upper bounds on the first m cost levels, $l_{d+1}^*,\dots,l_{d+m}^*$. Then, by letting $\varepsilon_h = 1$ for $h \ge d+m+1$, he computes the minimum number N of samples which guarantees that the right-hand side of (4) is no less than $1-\beta$. If N instances of δ are indeed extracted and problem min-max$_N$ is solved, then the obtained x^* and the corresponding cost levels are such that $R_k \le \varepsilon_k$, $k = d+1,\dots,N$, with high confidence $1-\beta$.

In both cases the user can link a cost l_k^* obtained through the optimization procedure to the value ε_k limiting its risk R_k. This provides a strong knowledge about the decision variable x^* without any further sampling effort.

B. Some theoretical and computational properties

We highlight some properties of the probability distribution function of the risks as given in (3). Some references about computational aspects are provided too.

Comments on Dirichlet distributions

The cumulative distribution function in the right-hand side of (3) is an instance of the so-called $(N-d)$-*dimensional ordered Dirichlet distribution function* with the first parameter equal to $d+1$ and other parameters equal to 1, see e.g. [30] - page 182. The corresponding probability density function is

$$\begin{aligned}&p_R(r_{d+1}, r_{d+2}, \dots, r_N) \\ &= \frac{N!}{d!} r_{d+1}^d \mathbb{1}\{0 \le r_{d+1} \le r_{d+2} \le \cdots \le r_N \le 1\}.\end{aligned}$$

By applying the following transformation to the random variables R_k's

$$\begin{aligned} D_N &= 1 - R_N \\ D_{N-1} &= R_N - R_{N-1} \\ &\vdots \\ D_{d+1} &= R_{d+2} - R_{d+1} \end{aligned}$$

the vector $D_{d+1}, D_{d+2}, \dots, D_N$ is obtained, which is distributed according to the well-known *Dirichlet distribution*, [30], [17]. Hence, the evaluation of an ordered Dirichlet distribution function is usually converted to the problem of evaluating a Dirichlet distribution function. The reader is referred to [14], [15], [13], [21] and references therein for studies on computational issues about Dirichlet distributions.

Beta distributions as marginals

By a well-known property of the Dirichlet distribution, the marginal probability distribution function of variable R_k is a Beta with parameters $(k, N-k+1)$, for each $k = d+1,\dots,N$, see [16]. That is,

$$\mathbb{P}^N\{R_k \le \varepsilon_k\} = 1 - \sum_{i=0}^{k-1} \binom{N}{i} \varepsilon_k^i (1-\varepsilon_k)^{N-i}. \tag{5}$$

This distribution function has the same form of the distribution function of the risk associated to l^* for the class of the so-called "fully-supported" problems, see [9]. Notably, the right-hand side of (5) can be easily evaluated by means of common tools, like the betainc function in MATLAB, [23], or pbeta in R, [27].

As is clear, a lower bound for the joint distribution function of $R_{d+1},\dots,R_N$ is given by the sum of the marginals, i.e.

$$\begin{aligned}\mathbb{P}^N\{R_{d+1}\le\varepsilon_{d+1},R_{d+2}\le\varepsilon_{d+2},R_{d+3}\le\varepsilon_{d+3},\dots,R_N\le\varepsilon_N\}\\ \ge 1-\sum_{k=d+1}^{N}\mathbb{P}^N\{R_k>\varepsilon_k\}\\ =1-\sum_{k=d+1}^{N}\sum_{i=0}^{k-1}\binom{N}{i}\varepsilon_k^i(1-\varepsilon_k)^{N-i},\end{aligned}\tag{6}$$

and (6) may be an acceptable approximation of (3) in some practical cases (see also Section III).

Based on (6) and following similar calculations as in [10] - Appendix B, it can be shown that, for a given $\beta\in(0,1)$, if

$$N\ge\max_{k=d+1,\dots,N}N^{(k)},\tag{7}$$

where

$$N^{(k)}:=\left\lfloor\frac{2}{\varepsilon_k}\left(k+\ln\frac{1}{\beta}\right)+\frac{4}{\varepsilon_k}\ln\left(\frac{2}{\varepsilon_k}\left(k+\ln\frac{1}{\beta}\right)\right)\right\rfloor+1$$

($\lfloor\cdot\rfloor$ denotes the integer part operator), then $\mathbb{P}^N\{R_{d+1}\le\varepsilon_{d+1},\dots,R_N\le\varepsilon_N\}\ge1-\beta$, i.e. conditions $R_k\le\varepsilon_k$, $k=d+1,\dots,N$, simultaneously hold with high confidence $1-\beta$. Although (7) may be loose, it reveals the logarithmic dependence of N on β by which it is possible to enforce *very* high confidence without affecting too much the sampling effort.

Connection with order statistics

Consider the sampling of N random variables, uniformly and independently distributed in $[0,1]$, and sort them in order of magnitude,

$$X^{(1)}\le X^{(2)}\le\cdots\le X^{(N)},$$

$X^{(i)}$ being the $i-$th smallest value. Variable $X^{(i)}$ is called the $i-$th order statistic and it is well known, [30], [25], that order statistics have joint ordered Dirichlet distribution with unitary parameters, that is $\mathbb{P}\{X^{(1)}\le\varepsilon_1,X^{(2)}\le\varepsilon_2,\dots,X^{(d+1)}\le\varepsilon_{d+1},\dots,X^{(N)}\le\varepsilon_N\}$ can be expressed as

$$\begin{aligned}N!\int_0^{\varepsilon_1}\int_0^{\varepsilon_2}\cdots\int_0^{\varepsilon_{d+1}}\cdots\int_0^{\varepsilon_N}\mathbb{1}\{0\le x_1\le\cdots\\ \cdots\le x_N\le1\}\mathrm{d}x_N\cdots\mathrm{d}x_{d+1}\cdots\mathrm{d}x_2\mathrm{d}x_1.\end{aligned}\tag{8}$$

If $\varepsilon_1=\varepsilon_2=\cdots=\varepsilon_{d+1}$, then, by integrating with respect to the first $d+1$ components, (8) becomes

$$\begin{aligned}\frac{N!}{d!}\int_0^{\varepsilon_{d+1}}x_{d+1}^d\int_0^{\varepsilon_{d+2}}\cdots\int_0^{\varepsilon_N}\mathbb{1}\{0\le x_{d+1}\le\cdots\\ \cdots\le x_N\le1\}\mathrm{d}x_N\cdots\mathrm{d}x_{d+2}\mathrm{d}x_{d+1},\end{aligned}$$

which is exactly the right-hand side of (3) in Theorem 1. In short, the computation of $\mathbb{P}^N\{R_{d+1}\le\varepsilon_{d+1},R_{d+2}\le\varepsilon_{d+2},R_{d+3}\le\varepsilon_{d+3},\dots,R_N\le\varepsilon_N\}$ can be reduced to the well known problem of computing the joint cumulative distribution function of order statistics, see e.g. [13], [21]. The freely distributed package μtoss for *R*, [4], [24], [27], contains the function jointCDF.orderedUnif which computes (8), though, because of numerical issues, it is reliable for $N\le100$ only.

Computability through Monte-Carlo methods

By virtue of the analogy with the distribution of order statistics, even Monte-Carlo methods can be employed to evaluate (3) for fixed values of $\varepsilon_{d+1},\dots,\varepsilon_N$. Indeed, one can repeat a large number of times, say M times, the following steps (C is a counter initially set to 0):

- draw a sequence of N independent samples from a uniform distribution in $[0,1]$;
- sort the sequence, i.e. compute all the order statistics from $X^{(1)}$ (the smallest value) to $X^{(N)}$ (the largest);
- evaluate the condition $X^{(i)}\le\varepsilon_i$ for $i=d+1,\dots,N$, and increment the counter C by 1 if it is satisfied for every value of the index i.

Then, $\hat{P}:=\frac{C}{M}$ is an estimate of the sought probability $P:=\mathbb{P}^N\{R_{d+1}\le\varepsilon_{d+1},R_{d+2}\le\varepsilon_{d+2},\dots,R_N\le\varepsilon_N\}$.

$\hat{P}$ and P are related by the Hoeffding's inequality (see [20], [29]), which guarantees that $P\ge\hat{P}-\gamma$ holds with confidence $1-\eta$ (e.g. $\eta=10^{-6}$) as long as the number of experiments is large enough (precisely, as long as $M\ge\frac{1}{2\gamma^2}\ln\frac{2}{\eta}$). This method becomes increasingly impractical as γ gets smaller, and more advanced randomized schemes must be considered if lowering γ under 10^{-4} is needed.

III. Numerical Example

In this section, the main result of this paper is applied to a simple min-max problem.

Problem formulation

During a campaign of data acquisition, 3000 independent vectors $\delta^{(i)}=(\delta_0^{(i)},\delta_1^{(i)},\dots,\delta_{10}^{(i)})\in\mathbb{R}^{11}$ were acquired, and a min-max regression problem was set up where the first component of the δ vector was regressed against the remaining 10 components of δ. Precisely, the following min-max problem with 10 decision variables ($d=10$) and 3000 scenarios ($N=3000$)

$$\min_{x\in[-1,1]^{10}}\max_{i=1,\dots,3000}\left|\delta_0^{(i)}-\sum_{j=1}^{10}x_j\cdot\delta_j^{(i)}\right|\tag{9}$$

was solved. The obtained solution x^* was

$$x^*=(0.74,-0.94,0.43,0.47,0.25,0.22,0.01,0.25,-0.49,0.15).$$

After computing x^*, we evaluated the cost levels l_k^* for $11\le k\le3000$ according to Definition 1:

$$l_k^*=\max\left\{l\in\mathbb{R}:\left|\delta_0^{(i)}-\sum_{j=1}^{10}x_j^*\cdot\delta_j^{(i)}\right|\le l\text{ for a choice of }k\text{ indexes }i\text{ among }\{1,\dots,3000\}\right\}.$$

This consisted in evaluating $\left|\delta_0^{(i)} - \sum_{j=1}^{10} x_j^* \cdot \delta_j^{(i)}\right|$ for the various scenarios and sorting these values in decreasing order.

Bounding the risks

The vector of risk thresholds $\boldsymbol{\varepsilon} = (\varepsilon_{d+1}, \ldots, \varepsilon_N)$ was designed according to the following rule: a parameter $\beta' \in [0,1]$ was fixed and, for each $k = d+1, \ldots, N$, $\varepsilon_k \in [0,1]$ was selected such that

$$\sum_{i=0}^{k-1} \binom{N}{i} \varepsilon_k^i (1-\varepsilon_k)^{N-i} = \frac{\beta'}{N-d}.$$

In words, the rule consists in choosing $\boldsymbol{\varepsilon}$ so that the marginal probability $\mathbb{P}^N\{R_k > \varepsilon_k\}$ is equal to $\frac{\beta'}{N-d}$ for all $k = d+1, \ldots, N$.

By (6), the adopted choice for $\boldsymbol{\varepsilon}$ entails that

$$\begin{aligned} &\mathbb{P}^N\{R_{d+1} \le \varepsilon_{d+1}, \ldots, R_N \le \varepsilon_N\} \\ &\quad \ge \quad 1 - \sum_{k=d+1}^{N} \mathbb{P}^N\{R_k > \varepsilon_k\} \\ &\quad = \quad 1 - \sum_{k=d+1}^{N} \frac{\beta'}{N-d} = 1 - \beta', \end{aligned} \tag{10}$$

i.e. the risks are simultaneously less than the elements of $\boldsymbol{\varepsilon}$ with confidence at least $1-\beta'$. For example, we have confidence 0, 0.9, 0.99 for $\beta' = 1, 10^{-1}, 10^{-2}$ respectively. A more refined evaluation is obtained through the right-hand side of (3). Using the right-hand side of (3) along with the Monte-Carlo algorithm in Section II-B, it turned out that the conditions $R_{11} \le \varepsilon_{11}, \ldots, R_{3000} \le \varepsilon_{3000}$ simultaneously hold with confidence equal to 0.98 when $\beta' = 1$, 0.997 when $\beta' = 10^{-1}$, and 0.9997 when $\beta' = 10^{-2}$.

Fig. 2 shows the values of ε_k for $\beta' = 1$, 10^{-1}, and 10^{-2}. As it is apparent, the values of ε_k are quite insensitive to the value of β' so that enforcing a high confidence only marginally impacts on the ε_k's.

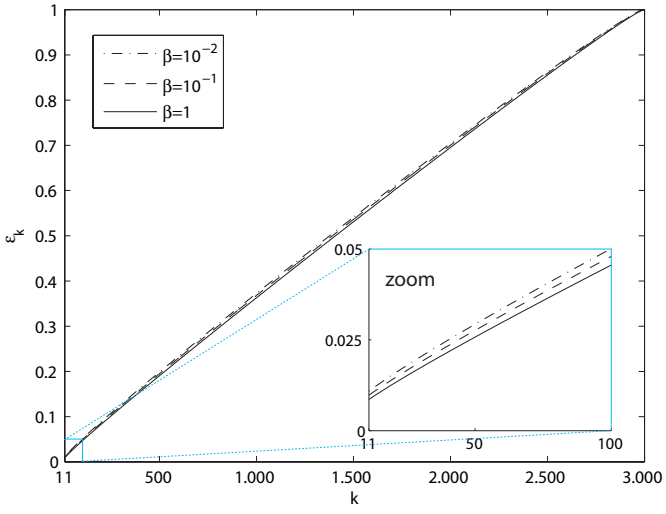

Fig. 2. Values of ε_k, $k = 11, \ldots, 3000$, for $\beta' = 1$ (solid line), $\beta' = 10^{-1}$ (dashed line), and $\beta' = 10^{-2}$ (dash-dotted line).

Costs-risks evaluation

We selected $\beta' = 10^{-2}$, so that confidence was 0.9997. Fig. 3 profiles l_k^* against ε_k. The interpretation of the graph is that each cost level l_k^* on the y-axis has a risk smaller than the ε_k on the x-axis with overall confidence 0.9997. By looking e.g. at the worst cost, we see that the risk that x^* carries a cost which is greater than $l_{11}^* = 5.55$ is just 1.09%, i.e. cost 5.55 is guaranteed for about the 99% of the possible uncertainty outcomes of δ. At the same time, the cost $l_{63}^* = 4$ is guaranteed for the 96% of the δ's, while cost $l_{229}^* = 3.01$ is guaranteed for the 90% of the δ's. By looking at the whole graph in Fig. 3, a complete characterization of the reliability of the solution x^* is obtained.

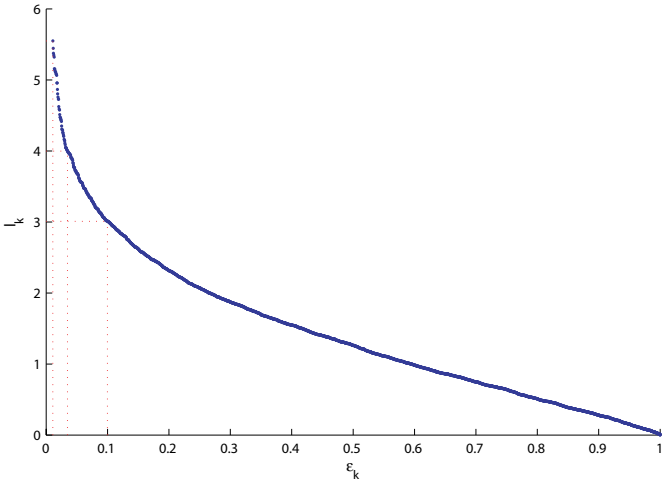

Fig. 3. Characterization of the optimal solution to problem (9). With confidence 0.9997, costs l_k^*'s (on the y-axis) have risks no higher than the corresponding ε_k's (on the x-axis).

REFERENCES

[1] T. Alamo, R. Tempo, and E.F. Camacho. Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, 54(11):2545–2559, 2009.

[2] T. Alamo, R. Tempo, and A. Luque. On the sample complexity of randomized approaches to the analysis and design under uncertainty. In the Proceedings of the *American Control Conference (ACC)*, Baltimore, Maryland, 2010.

[3] D. Bertsimas and A. Thiele. Robust and data-driven optimization: Modern decision-making under uncertainty. In *Tutorials on Operations Research*. INFORMS, 2006.

[4] G. Blanchard, T. Dickhaus, N. Hack, F. Konietschke, K. Rohmeyer, J. Rosenblatt, M. Scheer, and W. Werft. Mutoss - multiple hypothesis testing in an open software system. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*, volume 11, 2010.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] G.C. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[7] G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[8] M.C. Campi, G. Calafiore, and S. Garatti. Interval predictor models: identification and reliability. *Automatica*, 45(2):382–392, 2009.

[9] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[10] M.C. Campi and S. Garatti. Variable robustness control: principles and algorithms. In the Proceedings of the *19th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Budapest, Hungary, 2010.

[11] M.C. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149 – 157, 2009.

[12] A. Carè, S. Garatti, and M.C. Campi. Randomized min-max optimization: the exact risk of multiple cost levels. *Technical Report*, 2011.

[13] H. Finner and M. Roters. Multiple hypotheses testing and expected number of type I errors. *The Annals of Statistics*, 30(1):220–238, 2002.

[14] Ashraf Gouda and Tamás Szántai. New sampling techniques for calculation of Dirichlet probabilities. *Central European Journal of Operations Research*, 12:389–403, 2004.

[15] Ashraf Gouda and Tamás Szántai. On numerical calculation of probabilities according to Dirichlet distribution. *Annals of Operations Research*, 177:185–200, 2010.

[16] A.K. Gupta and S. Nadarajah, editors. *Handbook of Beta Distribution and Its Applications*. Mercel Dekker, Inc., 2004.

[17] R. D. Gupta and D. S. P. Richards. The history of the Dirichlet and Liouville distributions. *International Statistical Review*, 69:433–446, 2001.

[18] H.L. Harter. Minimax methods. In *Encyclopedia of Statistical Sciences*, volume 4, pages 514–516. John Wiley & Sons, 1982.

[19] Ronald Hochreiter. An evolutionary computation approach to scenario-based risk-return portfolio optimization for general risk measures. In Mario Giacobini, editor, *Applications of Evolutionary Computing*, volume 4448 of *Lecture Notes in Computer Science*, pages 199–207. Springer Berlin / Heidelberg, 2007.

[20] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[21] Koon-Shing Kwong and Yiu-Man Chan. On the evaluation of the joint distribution of order statistics. *Computational Statistics & Data Analysis*, 52(12):5091–5099, 2008.

[22] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[23] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[24] MuToss Coding Team. Package mutoss- unified multiple testing procedures, version 0.1-4. `http://cran.za.r-project.org/web/packages/mutoss/`, October 2010.

[25] H.N. Nagaraja H.A. David. *Order Statistics, 3rd Edition*. Wiley, 2003.

[26] B.K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.

[27] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. `http://www.R-project.org`.

[28] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. The society for industrial and applied mathematics and the mathematical programming society, Philadelphia, USA, 2009.

[29] R. Tempo, G. Calafiore, and F. Dabbene. *Randomized Algorithms for Analysis and Control of Uncertain Systems*. Springer, London, UK, 2005.

[30] S.S. Wilks. *Mathematical Statistics*. Wiley, New York, 1962.