Operations Research

Robust Optimization with Ambiguous Stochastic Constraints Under Mean and Dispersion Information

Krzysztof Postek, Aharon Ben-Tal, Dick den Hertog, Bertrand Melenberg

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management
science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Robust Optimization with Ambiguous Stochastic Constraints Under Mean and Dispersion Information

Krzysztof Postek,[a, b] **Aharon Ben-Tal,**[b, c, d] **Dick den Hertog,**[e] **Bertrand Melenberg**[e]

[a] Econometric Institute, Erasmus University Rotterdam, 3062 PA Rotterdam, Netherlands; [b] Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa 3200003, Israel; [c] Shenkar College, Ramat Gan 5252626, Israel; [d] Tilburg University, 5037 AB Tilburg, Netherlands; [e] CentER and Department of Econometrics and Operations Research, Tilburg University, 5037 AB Tilburg, Netherlands

Contact: krzysztofp@technion.ac.il, http://orcid.org/0000-0002-4028-3725 (KP); abental@ie.technion.ac.il (AB-T); d.denhertog@tilburguniversity.edu, http://orcid.org/0000-0002-1829-855X (DdH); b.melenberg@tilburguniversity.edu (BM)

Abstract. In this paper we consider ambiguous stochastic constraints under partial information consisting of means and dispersion measures of the underlying random parameters. Whereas the past literature used the variance as the dispersion measure, here we use the mean absolute deviation from the mean (MAD). This makes it possible to use the 1972 result of Ben-Tal and Hochman (BH) in which *tight* upper and lower bounds on the expectation of a convex function of a random variable are given. First, we use these results to treat ambiguous *expected feasibility* constraints to obtain exact reformulations for both functions that are convex and concave in the components of the random variable. This approach requires, however, the independence of the random variables and, moreover, may lead to an exponential number of terms in the resulting robust counterparts. We then show how upper bounds can be constructed that alleviate the independence restriction, and require only a linear number of terms, by exploiting models in which random variables are linearly aggregated. Moreover, using the BH bounds we derive three new safe tractable approximations of *chance constraints* of increasing computational complexity and quality. In a numerical study, we demonstrate the efficiency of our methods in solving stochastic optimization problems under mean-MAD ambiguity.

Supplemental Material: The electronic companion is available at https://doi.org/10.1287/opre.2017.1688.

Keywords: robust optimization • ambiguity • stochastic programming • chance constraints

1. Introduction
1.1. Problem and Contribution

Consider an optimization problem with a constraint

$$f(\mathbf{x}, \mathbf{z}) \leqslant 0,$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the decision vector, $\mathbf{z} \in \mathbb{R}^{n_z}$ is an uncertain parameter vector, and $f(\cdot, \mathbf{z})$ is assumed to be convex for all $\mathbf{z}$. There are three principal ways to address such constraints. One of them is *robust optimization*. In this approach, $\mathcal{U}$ is a user-provided convex compact uncertainty set and the constraint is to hold for all $\mathbf{z} \in \mathcal{U}$, i.e., $\mathbf{x}$ is robust feasible if

$$\sup_{\mathbf{z} \in \mathcal{U}} f(\mathbf{x}, \mathbf{z}) \leqslant 0. \qquad (1)$$

The key issue in this approach is to reformulate (1) to an equivalent, computationally tractable form (Ben-Tal and Nemirovski 1998; Ben-Tal et al. 2009, 2015).

In the other approaches, which go under the name of *distributionally robust optimization* (DRO), $\mathbf{z}$ is a *random* parameter vector whose distribution $\mathbb{P}_\mathbf{z}$ belongs to a set $\mathcal{P}$ (the so-called *ambiguity set*). A typical example for $\mathcal{P}$ is a set of all distributions with given values of the first two moments. In such a setting, there are two principal constraint types: the *worst-case expected feasibility constraints*:

$$\sup_{\mathbb{P}_\mathbf{z} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_\mathbf{z}} f(\mathbf{x}, \mathbf{z}) \leqslant 0, \qquad (2)$$

and *chance constraints*:

$$\sup_{\mathbb{P}_\mathbf{z} \in \mathcal{P}} \mathbb{P}_\mathbf{z}(f(\mathbf{x}, \mathbf{z}) > 0) \leqslant \epsilon. \qquad (3)$$

For constraint (2) the key challenge is, for a given ambiguity set $\mathcal{P}$, to obtain a tractable exact form of the worst-case expectation, or a good upper bound. Constraint (2) is also used in the construction of safe approximation of the ambiguous chance constraint (3), where by a *safe approximation* is meant a system $\mathcal{S}$ of computationally tractable constraints, such that $\mathbf{x}$ feasible for $\mathcal{S}$ is also feasible for constraint (3).

In this paper, we consider problems with ambiguity sets consisting of distributions having given mean-dispersion measures. The literature of these types of problems started with the paper by Scarf (1958). Under mean-variance information, Scarf derived the exact worst-case expectation formula for a single-variable piecewise linear objective function used in

the newsvendor problem. Later, his result has been extended to more elaborate cases of inventory and newsvendor problems by, for example, Gallego (1992), Gallego et al. (2001), and Perakis and Roels (2008). In a paper by Popescu (2007), it has been proved that for a wide class of increasing concave utility functions the problem of maximizing the worst-case expected utility under mean-variance distributional information reduces to solving a parametric quadratic optimization problem.

In a broader context, the idea of constructing an approximation of the worst-case expectation of a given function by a discrete distribution falls into the category of bounding strategies based on distributional approximation; see Edirisinghe (2011) who provides a broad overview of results obtained in this field. Rogosinsky (1958) and Karr (1983) show that the worst-case probability distributions corresponding to the moment problems are discrete, with a number of points corresponding to the number of moment conditions. Shapiro and Kleywegt (2002) develop a duality theory for stochastic programs where the saddle points are also vectors of discrete probabilities. Dupačová (1966) and Gassmann and Ziemba (1986) give convex upper bounds on the expectation of a convex function under first-moment conditions over a polyhedral support, based on the dual of the related moment problem. Birge and Wets (1987) and Edirisinghe and Ziemba (1994a) extend this approach to distributions with unbounded support. Dulá (1992) provides a bound for the expectation of a simplicial function of a random vector using first moments and the sum of all variances. His approach is extended by Kall (1991) demonstrating that the related moment problems can be solved using nonsmooth optimization problems with linear constraints. Other notable works in this field include Frauendorfer (1988) and Edirisinghe and Ziemba (1992, 1994b).

As already mentioned, constraint (2) can also be used to construct safe approximations of chance constraint (3). This is typically achieved by constructing the worst-case expectations used in Markov or Chebyshev inequalities or the conditional value-at-risk; see, for example, Rockafellar and Uryasev (2000), Nemirovski and Shapiro (2006), and Ben-Tal et al. (2009). An alternative to this approach is to use duality results for moment problems to evaluate the worst-case expectations of characteristic functions, which began with the work of Isii (1962).

Despite numerous works, to the best of our knowledge, no *closed-form tight upper bounds* are known on the expectations of *general convex functions* under mean-variance information. Similarly, there are no exact, computationally tractable formulations of linear chance constraints under mean-variance information under the independence assumption. Surprisingly,

already in 1972, a result of Ben-Tal and Hochman (1972) (from now on referred to as BH) was available, providing tight upper and lower bounds on the expectation of a general convex $f(\mathbf{x}, \cdot)$ for the case where $\mathscr{P}$ consists of all distributions of componentwise independent $\mathbf{z}$ with known supports, means, but with another dispersion measure: the *mean absolute deviation from the mean* (*MAD*). In this setup, our contributions can be summarized as follows:

• We show how the result of BH can be used to obtain exact reformulations of worst-case expected feasibility constraints (2) for $f(\mathbf{x}, \cdot)$ *convex* in the components of $\mathbf{z}$ under mean-MAD distributional information on $\mathbf{z}$ whose components are independent. We show how the same can be achieved for $f(\mathbf{x}, \cdot)$ *concave* in the components of $\mathbf{z}$ with additional information on the probability that a component z_i is greater than or equal to its mean μ_i. In this way, we are able to obtain an entire interval for the value of the left-hand side in (2) so that it is known precisely what is the remaining ambiguity after the distributional information is accounted for. These reformulations involve a number of terms that are exponential in the dimension of $\mathbf{z}$. However, for the important special case of linearly aggregated random variables $\mathbf{a}^\top \mathbf{z}$ and $\mathbf{a}(\mathbf{x})^\top \mathbf{z}$, with $\mathbf{a}(\cdot)$ affine, we derive new, polynomial-size, upper bounds on the worst-case expectations of $f(\mathbf{x}, \mathbf{a}^\top \mathbf{z})$ and nonnegative $f(\mathbf{a}(\mathbf{x})^\top \mathbf{z})$, where $f(\mathbf{x}, \cdot)$ and $f(\cdot)$ are convex.

• Under the assumption of independent random variables z_i, we derive new tractable safe approximations of chance constraint (3) under mean-MAD information. In particular, in the mean-dispersion setting of this paper we obtain new improved Bernstein-type tractable approximations based on the *exponential polynomials*, resolving thus an unsolved case of section 4.3.6 of Ben-Tal et al. (2009).

• Examples and numerical experiments illustrating our approach show, among others, (i) the power of DRO to enhance RO solutions in the case of existence of multiple optimal RO solutions; (ii) that the solutions based on the linear aggregation techniques yield good results and circumvent the exponential growth of the number of components in the expression for the bounds; (iii) the importance of accounting for the independence or dependence of stochastic parameters for the quality of solutions to DRO problems; and (iv) the power of the exponential polynomial-based safe tractable approximation of chance constraints.

The supports, means, and MADs of the z_i's can be estimated from past data using, for example, the procedures given in the electronic companion. There, we also refer to statistical procedures for testing independence of random variables. Compared to the variance, the MAD has some advantages. First, for the MAD we have closed-form tight upper and lower bounds on

the expectations of general *convex* and *concave* functions, which are not available for the variance, and we use these for both (2) and (3). Second, compared to the variance, the MAD enjoys the following statistical properties: (i) For some distributions the MAD exists whereas the second moment does not—an example of this is the class of stable distributions or the t_2 distribution; see Ben-Tal and Hochman (1985) and the electronic companion. (ii) From the point of view of robust statistics, in certain situations involving outlier observations in the data the MAD exhibits a better *asymptotic relative efficiency* compared to the standard deviation; see Huber and Ronchetti (2009). (iii) In situations where even small deviations need to be accounted for (for example, the implementation error example considered in the electronic companion), the MAD gives a greater relative weight to small deviations, compared to the variance.

In the electronic companion we list some properties of the MAD and exact formulas for its value for several known distributions. For more references related to the MAD, we refer the reader to Gorard (2005) and El Amir (2012). The statistical properties of the MAD do not mean, however, that MAD should be the dispersion measure in every application—one must always consider in full the statistical properties of the application at hand and the computational complexity of the resulting problem formulation.

1.2. Alternative Ambiguity Setups

There are alternative ways of specifying the set $\mathscr{P}$, for example, as sets of distributions deviating from a known distribution according to a certain distance measure (see, for example, Ben-Tal et al. 2013, where ϕ-divergence distance-like measures are used). For a broad overview of types of ambiguity sets we refer the reader to Postek et al. (2016) and Hanasusanto et al. (2015). Among these alternative setups, there are some cases for which exact reformulations are possible. For the worst-case expectation constraints, examples are given in Ben-Tal et al. (2013), Wiesemann et al. (2014) and Esfahani and Kuhn (2017). Settings in which exact reformulations are possible for individual chance constraints are, for example, given in Calafiore and El Ghaoui (2006) and Jiang and Guan (2016).

When the components of the random vector $\mathbf{z}$ are not independent, there are some specific examples of the objective functions $f(\mathbf{x}, \mathbf{z})$, where exact results or approximations can be obtained; see Delage and Ye (2010), Goh and Sim (2010), and Zymler et al. (2013).

An important contribution is Wiesemann et al. (2014) who have recently introduced a class of quite general ambiguity sets for which they derive computationally tractable counterparts of worst-case expected feasibility constraints for a certain class of convex $f(\mathbf{x}, \cdot)$. In their framework, independence of the components of $\mathbf{z}$ cannot be modeled explicitly and has to be approximated, for example, by bounds on the covariance matrix of $\mathbf{z}$. Therefore, our and their approaches are not comparable in a general *convex function* setup. In the electronic companion, we illustrate the marked difference between the two approaches when $f(\mathbf{x}, \mathbf{z}) = \exp(\mathbf{x}^T \mathbf{z})$ where, without the assumption of independence, the resulting robust constraint is *strictly convex* in the uncertain parameter. In their approach, *convexity* in the uncertain parameter is required, whereas ours can also deal with the *concavity* case; see Example 2 of Section 3.3.

In the specific case of aggregated random variables discussed in Section 4, we drop the assumption of independence and, therefore, the methods can be compared. From a theoretical standpoint, in this case the advantage of our aggregation approach is the simplicity of deriving the approximate bounds for general convex functions $f(\mathbf{x}, \cdot)$ and $f(\cdot)$—they do not require an exponential number of terms. The advantage of Wiesemann et al. (2014) is the exactness of their reformulation. However, in their approach $f(\mathbf{x}, \cdot)$ has to satisfy some additional tractability conditions or an additional optimization problem needs to be solved.

1.3. Paper Structure

In Section 2, we describe the mean-MAD results of BH. In Section 3, we show how the mean-MAD results can be used in optimization problems involving stochastic constraints (2). In Section 4 we outline the results for the case of linearly aggregated random variable. Section 5 includes new results on safe tractable approximations of chance constraints (3). Section 6 includes numerical examples illustrating our approach and Section 7 concludes.

2. Bounds on the Expectation of a Convex Function of a Random Variable

2.1. Introduction

In this section we introduce the results of BH on tight upper and lower bounds on the expected value of a convex function of a componentwise independent $\mathbf{z} = (z_1, \ldots, z_{n_\mathbf{z}})^\top$. From now on, we drop the subscript $\mathbf{z}$ from $\mathbb{P}_\mathbf{z}$ and the probability distribution applies to $\mathbf{z}$. The pieces of partial distributional information on z_i's constituting the ambiguity sets in BH are the following:

(i) Support including intervals: $\operatorname{supp}(z_i) \subseteq [a_i, b_i]$, where $-\infty < a_i \leqslant b_i < \infty$, $i = 1, \ldots, n_\mathbf{z}$. BH show also that their bounds hold in cases where $a_i = -\infty$ and/or $b_i = +\infty$. We illustrate this in Remark 3. In the remainder of the paper, however, we concentrate on the bounded case, with RO applications in mind.

(ii) Means: $\mathbb{E}_\mathbb{P}(z_i) = \mu_i$.

(iii) Mean absolute deviations from the means (MAD): $\mathbb{E}_{\mathbb{P}}|z_i - \mu_i| = d_i$. The MAD is known to satisfy the bound (BH, lemma 1):

$$0 \leqslant d_i \leqslant d_{i,\max} = \frac{2(b_i - \mu_i)(\mu_i - a_i)}{b_i - a_i}, \quad i = 1, \ldots, n_{\mathbf{z}}. \quad (4)$$

(iv) Probabilities of z_i's being greater than or equal to μ_i: $\mathbb{P}(z_i \geqslant \mu_i) = \beta_i$. For example, in the case of continuous symmetric distribution of z_i we know that $\beta_i = 0.5$. This quantity is known to satisfy the bounds:

$$\frac{d_i}{2(b_i - \mu_i)} = \underline{\beta}_i \leqslant \beta_i \leqslant \overline{\beta}_i = 1 - \frac{d_i}{2(\mu_i - a_i)}, \quad i = 1, \ldots, n_{\mathbf{z}}. \quad (5)$$

Using these building blocks, we define two types of ambiguity sets $\mathscr{P}$:

• The (μ, d) *ambiguity set*, consisting of the distributions with known (i)–(iii) for each z_i:

$$\mathscr{P}_{(\mu,d)} = \{\mathbb{P}: \operatorname{supp}(z_i) \subseteq [a_i, b_i], \mathbb{E}_{\mathbb{P}}(z_i) = \mu_i,$$
$$\mathbb{E}_{\mathbb{P}}|z_i - \mu_i| = d_i, \forall i, z_i \perp\!\!\!\perp z_j, \forall i \neq j\}, \quad (6)$$

where $z_i \perp\!\!\!\perp z_j$ denotes the stochastic independence of components z_i and z_j.

• The (μ, d, β) *ambiguity set*, consisting of the distributions with known (i)–(iv) for each z_i:

$$\mathscr{P}_{(\mu,d,\beta)} = \{\mathbb{P}: \mathbb{P} \in \mathscr{P}_{(\mu,d)}, \mathbb{P}(z_i \geqslant \mu_i) = \beta_i, \forall i\}. \quad (7)$$

In the following, we present the results of BH on

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}), \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) \quad \text{and}$$

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}), \min_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}),$$

where $f: \mathbb{R}^{n_{\mathbf{z}}} \to \mathbb{R}$ is convex. We note that in the case of concave $f(\cdot)$ the upper bounds become lower bounds, and vice versa.

2.2. One-Dimensional z

We begin with the simpler and more illustrative case of one-dimensional random variable z. For that reason, we drop the subscript i.

Upper bounds. BH show the following:

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = p_1 f(a) + p_2 f(\mu) + p_3 f(b), \quad (8)$$

where

$$p_1 = \frac{d}{2(\mu - a)}, \quad p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)},$$
$$p_3 = \frac{d}{2(b - \mu)}. \quad (9)$$

Hence, the worst-case distribution is a three-point distribution on $\{a, \mu, b\}$. The same bound holds for the (μ, d, β) ambiguity set.

Remark 1. A special case of (8) and (9) is the upper bound on $f(z)$ when only the interval $[a, b]$ and the mean μ are known. Such a bound is known as the Edmundson-Madansky bound (Edmundson 1956, Madansky 1959):

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu)}} f(z) = \frac{b - \mu}{b - a} f(a) + \frac{\mu - a}{b - a} f(b) \quad \text{where}$$

$$\mathscr{P}_{(\mu)} = \{\mathbb{P}: \operatorname{supp}(z) \subseteq [a, b], \mathbb{E}_{\mathbb{P}} z = \mu\}. \quad (10)$$

Indeed, inserting the biggest possible value of MAD (see (4)) equal to $d_{\max} = 2(b - \mu)(\mu - a)/(b - a)$ into (9) yields the probability of outcome μ equal to 0. This follows from the fact that the original bound is a nondecreasing function of d, as proven in Proposition 3. □

Lower bounds. To obtain a closed-form lower bound on $\mathbb{E}_{\mathbb{P}} f(z)$, additional information is needed in the form of the parameter β. Then, it holds that

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1 - \beta)}\right). \quad (11)$$

In case β is not known, BH show

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z)$$
$$= \min_{\underline{\beta} \leqslant \beta \leqslant \overline{\beta}} \left\{ \beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1 - \beta)}\right) \right\}, \quad (12)$$

where the minimization over β is a convex problem in β and for a strictly convex $f(\cdot)$ there is a unique optimal solution.

Remark 2. In the case of no knowledge about d, the lower bound is obtained at $d^* = 0$, which corresponds to the well-known Jensen bound (Jensen 1906). □

Remark 3. In the case where $a = -\infty$ and/or $b = +\infty$, bounds can still be obtained under additional conditions, namely, that the limits $\lim_{t \to \pm\infty} f(t)/t$ exist and are finite, with the "+" corresponding to $b = +\infty$, and the "−" corresponding to $a = -\infty$. We illustrate this on the example $a \in \mathbb{R}, b = +\infty$. Assume that $\lim_{t \to +\infty} f(t)/t = \gamma$. We then have

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z)$$

$$= \lim_{b \to \infty} \left\{ \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)} \right. \right.$$
$$\left. \left. - \frac{d}{2(b - \mu)}\right) f(\mu) + \frac{d}{2(b - \mu)} f(b) \right\}$$
$$= \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)}\right) f(\mu) + \frac{d}{2} \gamma;$$

and for the lower bound we have

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} f(z) = \frac{d}{2} \gamma + f\left(\mu - \frac{d}{2}\right).$$

The lower bound for the (μ, d, β) ambiguity set is the same as (11). □

2.3. Multidimensional z

Upper bounds. For $n_{\mathbf{z}} > 1$, the worst-case probability distribution under (μ, d) information is a component-wise counterpart of (8)–(9):

$$p_1^i = \frac{d_i}{2(\mu_i - a_i)}, \quad p_2^i = 1 - \frac{d_i}{2(\mu_i - a_i)} - \frac{d_i}{2(b_i - \mu_i)}, \quad p_3^i = \frac{d_i}{2(b_i - \mu_i)}, \quad i = 1, \dots, n_{\mathbf{z}}. \tag{13}$$

The worst-case expectation of $f(\mathbf{z})$ is obtained by applying the bound (8) for each z_i, i.e., by enumerating over all $3^{n_{\mathbf{z}}}$ permutations of outcomes a_i, μ_i, b_i of components z_i. It holds then that (BH)

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \sum_{\boldsymbol{\alpha} \in \{1, 2, 3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p_{\alpha_i}^i f(\tau_{\alpha_1}^1, \dots, \tau_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}), \tag{14}$$

where

$$\tau_1^i = a_i, \quad \tau_2^i = \mu_i, \quad \tau_3^i = b_i \quad \text{for } i = 1, \dots, n_{\mathbf{z}}. \tag{15}$$

Again, the same upper bound holds for the (μ, d, β) ambiguity set.

Lower bounds. Similar to the one-dimensional case, the closed-form lower bound under (μ, d) information requires known $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{n_{\mathbf{z}}})^\top$:

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \sum_{\boldsymbol{\alpha} \in \{1, 2\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} q_{\alpha_i}^i f(v_{\alpha_1}^1, \dots, v_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}), \tag{16}$$

where $\underline{\boldsymbol{\beta}} = (\underline{\beta}_1, \dots, \underline{\beta}_{n_{\mathbf{z}}})^\top, \bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \dots, \bar{\beta}_{n_{\mathbf{z}}})^\top$ and

$$q_1^i = \beta_i, \quad q_2^i = 1 - \beta_i, \quad v_1^i = \mu_i + d_i/2\beta_i, \quad v_2^i = \mu_i - d_i/2(1 - \beta_i). \tag{17}$$

If $\boldsymbol{\beta}$ is unknown, the bound is obtained by minimization:

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \inf_{\underline{\boldsymbol{\beta}} \leqslant \boldsymbol{\beta} \leqslant \bar{\boldsymbol{\beta}}} \sum_{\boldsymbol{\alpha} \in \{1, 2\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} q_{\alpha_i}^i f(v_{\alpha_1}^1, \dots, v_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}). \tag{18}$$

In the multidimensional case, minimization over $\boldsymbol{\beta}$ is a nonconvex problem—it is only convex in β_i when other $\beta_j, j \neq i$ are fixed.

A statistical procedure for estimating the parameters μ, d, and β is provided in the electronic companion, Appendix EC.1.

3. Robust Counterparts of Expected Feasibility Constraints

3.1. Bounds on the Expectations of Convex and Concave Functions

In this section we demonstrate how the results of BH can be used to solve problems involving constraints

$$\text{Val}_{\mathbb{P}}(\mathbf{x}) = \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leqslant 0 \quad \forall \mathbb{P} \in \mathscr{P}, \tag{19}$$

where $f(\cdot, \mathbf{z})$ is either convex or concave. By the results presented in Section 2 we can obtain bounds on the value of the left-hand side (19), as stated in the following propositions, which follow straightforwardly from (14) and (18).

Proposition 1. *If $f(\mathbf{x}, \cdot)$ is convex, it holds that*

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) = g_U(\mathbf{x})$$
$$= \sum_{\boldsymbol{\alpha} \in \{1, 2, 3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p_{\alpha_i}^i f(\mathbf{x}, \tau_{\alpha_1}^1, \dots, \tau_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}), \tag{20}$$

with $p_{\alpha_i}^i, \tau_{\alpha_i}^i$ defined as in (13) and (15).

As we can see, $g_U(\cdot)$ in (20) inherits the convexity in $\mathbf{x}$ from $f(\cdot, \mathbf{z})$ and its functional form depends only on the form of $f(\cdot, \mathbf{z})$.

Proposition 2. *If $f(\mathbf{x}, \cdot)$ is concave it holds that*

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) = g_L(\mathbf{x})$$
$$= \sum_{\boldsymbol{\alpha} \in \{1, 2\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} q_{\alpha_i}^i f(\mathbf{x}, v_{\alpha_1}^1, \dots, v_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}), \tag{21}$$

with $q_{\alpha_i}^i, v_{\alpha_i}^i$ defined by (17).

Similar to the previous proposition, $g_L(\cdot)$ in (21) inherits the convexity in $\mathbf{x}$ from $f(\cdot, \mathbf{z})$.

It is important to note that to obtain $g_L(\mathbf{x})$ one needs the information on $\boldsymbol{\beta}$. For the case of convexity (concavity) of $f(\mathbf{x}, \cdot)$, a lower bound on $\mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$ is given by $g_L(\mathbf{x})$ $(g_U(\mathbf{x}))$, with the corresponding distributional assumptions. Overall, the upper and lower bound give a closed interval in which $\text{Val}_{\mathbb{P}}(\mathbf{x})$ lies.

Corollary 1. *If $f(\mathbf{x}, \cdot)$ is convex for all $\mathbf{x}$ then*

$$\text{Val}_{\mathbb{P}}(\mathbf{x}) \in [g_L(\mathbf{x}), g_U(\mathbf{x})] \quad \forall \mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}.$$

If $f(\mathbf{x}, \cdot)$ is concave for all $\mathbf{x}$ then

$$\text{Val}_{\mathbb{P}}(\mathbf{x}) \in [g_U(\mathbf{x}), g_L(\mathbf{x})] \quad \forall \mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}.$$

Using (μ, d, β) information, Corollary 1 provides an interval for the expected value of $f(\mathbf{x}, \mathbf{z})$ instead of only an upper bound. The following example illustrates the power of using the (μ, d, β) information as compared to using only information about the mean.

Example 1. Consider the expectation $\mathbb{E}_{\mathbb{P}} f(\mathbf{z})$ of the function

$$f(\mathbf{z}) = \max_{x_1, x_2}\{z_1 x_1^2 + z_2^2 x_2^2: \ 0 \leqslant x_1 \leqslant 10, \ 0 \leqslant x_2 \leqslant 5\},$$

where $\mathbf{z} = (z_1, z_2)$, $\text{supp}(z_1) \subseteq [-0.5, 0.5]$, $\text{supp}(z_2) \subseteq [0, 1]$, $\mathbb{E}_{\mathbb{P}}(z_1) = 0$, $\mathbb{E}_{\mathbb{P}}(z_2) = 0.5$, $\mathbb{E}_{\mathbb{P}}|z_1 - 0| = 0.25$, $\mathbb{E}_{\mathbb{P}}|z_2 - 0.5| = 0.25$, $\mathbb{P}(z_1 \geqslant 0) = 0.5$, $\mathbb{P}(z_2 \geqslant 0.5) = 0.5$. As

a supremum of convex functions, $f(\cdot)$ is convex in $\mathbf{z}$. If only the information on the support and the means is used, then we obtain that

$$\mathbb{E}_{\mathbb{P}} f(\mathbf{z}) \in [6.25, 37.5].$$

However, if the support, mean, MAD, β information is used, the interval shrinks significantly to

$$\mathbb{E}_{\mathbb{P}} f(\mathbf{z}) \in [20.31, 21.87],$$

where the left bound has been obtained by optimizing over β as in (18). The calculations for the (μ, d, β) case are given in the electronic companion. $\quad\square$

3.2. Dimensionality and Dependence

There are two difficulties associated with the bounds (20) and (21). One is the computational difficulty: the number of terms in (20) and (21) is exponential in $n_{\mathbf{z}}$. Second is the assumption of independence of z_i's: when the independence hypothesis is rejected, the solutions obtained using (20) and (21) might underperform significantly. In Section 4 we discuss a wide class of functions $f(\mathbf{x}, \mathbf{z})$ for which both of these difficulties are alleviated. Here, we discuss cases where these difficulties are not present or can be alleviated using existing techniques.

First, in certain applications the number $n_{\mathbf{z}}$ of random variables is small (less than 10). Second, an important special case is when $f(\mathbf{x}, \mathbf{z})$ is a sum of functions

$$f(\mathbf{x}, \mathbf{z}) = \sum_{j=1}^{n_c} f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)}),$$

where $f^{(j)}(\mathbf{x}, \cdot)$ have small numbers n_j of uncertain variables.

An important special case is the function $f(\mathbf{x}, \mathbf{z}) = \exp(\mathbf{x}^\top \mathbf{z})$. Upper bounds on moment generating functions $\mathbb{E}\exp(\mathbf{x}^\top \mathbf{z})$ are a key tool in constructing safe tractable approximations of chance constraints. As we show in Section 5, the properties of the $\exp(\cdot)$ allow for a simple, closed-form formula for its worst-case expectation under (μ, d) information and for which the number of terms is linear in $n_{\mathbf{z}}$.

In the end, if the dimensionality remains an issue, for problems with linear and piecewise linear $f(\mathbf{x}, \cdot)$ one can use, for example, the stochastic decomposition method (Higle and Sen 1996) where scenarios (in our case support points) are added iteratively until the current model is a good enough approximation of the original model. In cutting-plane methods, the verification of the ambiguous constraint can exploit the tree structure of the worst-case distribution support. In this tree structure, each outcome of z_1 leads to three (or two for the concave case) outcomes of z_2, each of these leading to another three outcomes of z_3 etc. Then one can

determine if the constraint holds already after investigating the first few layers of the tree, which may lead to a verification of much less than all $3^{n_{\mathbf{z}}}$ scenarios. Other approximate approaches are the sample average approximation (Shapiro et al. 2009) or the scenario reduction technique (Dupačová et al. 2003).

If the random uncertain vector $\mathbf{z}$ contains dependent components, it can be decomposed by means of factor analysis, for example, based on principal component analysis (see Jolliffe 2002), into linear combinations of a limited number of uncorrelated factors. For example, in a situation of portfolio optimization problem with 25 assets, it is natural to decompose them into three to four uncorrelated risk factors (see, for example, Baillie et al. 2002), whose empirical distribution provides information also about their support, means, and MADs. Even though uncorrelatedness can be weaker than independence, such a technique is often a practical solution. Additionally, factor decomposition also helps to alleviate the issue of dimensionality. In the electronic companion we refer to tests for verifying the independence of z_i's.

3.3. Use of the Bounds in Some General Applications

In this section we present three cases where the reformulations of the worst-case expected feasibility constraints presented in Section 3.1 can be used.

Average-case enhancement of RO solutions. The first application lies in finding worst-case-optimal solutions with good average-case performance to the following RO problem:

$$
\begin{aligned}
\min_{\mathbf{x}, t} \quad & t \\
\text{s.t.} \quad & \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z}) \leqslant t, \\
& \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}, \mathbf{z}) \leqslant 0, \quad i = 1, \dots, m.
\end{aligned}
\tag{22}
$$

It happens frequently that there exist multiple optimal solutions to (22); see Iancu and Trichakis (2013) and de Ruiter et al. (2016). Whereas the worst-case performance of such solutions is the same, their average-case performance may differ dramatically. For that reason, once the optimal value $\bar{t}$ for (22) is known, a second optimization step may be used to select one of the optimal solutions to provide good average-case behavior. Since the results of BH provide exact bounds on the worst-case expectations, they can be used in such a step. In the following, we describe such a two-step procedure:

1. Solve problem (22) and denote its optimal value by $\bar{t}$.

2. Solve the following problem, minimizing the worst-case expectation of the objective value, with the worst-case value of $f(\mathbf{x}, \mathbf{z})$ less than or equal to $\bar{t}$:

$$\min_{\mathbf{x}, u} \; u$$
$$\text{s.t.} \; \sup_{\mathbb{P} \in \mathscr{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leqslant u,$$
$$\sup_{z \in \mathscr{Z}} f(\mathbf{x}, \mathbf{z}) \leqslant \bar{t}, \tag{23}$$
$$\sup_{z \in \mathscr{Z}} g_i(\mathbf{x}, \mathbf{z}) \leqslant 0, \quad i = 1, \dots, m.$$

The two-step procedure is expected to select the optimal solution with good average-case performance for its focus on the worst-case expectation among the best worst-case solutions. If the uncertainty is present only in the constraints involving functions $g_i(\cdot, \cdot)$, a similar two-step approach can be designed to maximize the worst-case expected slack in the worst-case constraints in (22); see Iancu and Trichakis (2013). We note that following the theory of Iancu and Trichakis (2013), there might exist multiple optimal solutions to (23) and one may need to include another "enhancement step" to choose among them.

An alternative approach to enhancing robust solutions is to sample a number S of scenarios for $\mathbf{z}$ to find a solution that optimizes the average of the objective value over the sample.[1] This approach, however, has as a shortcoming that the outcome might depend on the choice of sample size S and the sample itself. For that reason, the DRO methods can provide a good alternative to enhancing the quality of RO solutions. In our paper, we test the application of the (μ, d) bounds to enhance average-case performance in an inventory management problem in Section 6.2.

Concave functions of uncertainty. Another important case is the constraint

$$\mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leqslant t, \quad \forall \mathbb{P} \in \mathscr{P},$$

where $f(\mathbf{x}, \mathbf{z})$ is convex in $\mathbf{x}$ and *concave* in $\mathbf{z}$. Knowing the mean of $\mathbf{z}$, a trivial upper bound on the worst-case expectation can be obtained using Jensen's inequality. However, in case of a nondegenerate random variable $\mathbf{z}$, this may be a crude bound. Also, the distributionally robust convex optimization tools of Wiesemann et al. (2014) are not applicable as they require convexity in $\mathbf{z}$. In contrast, using (μ, d, β) information on $\mathbf{z}$ it is possible to obtain a closed-form tight upper bound on $\mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$ for convex or concave $f(\mathbf{x}, \cdot)$.

The following toy example illustrates the potential striking difference between minimizing (i) the worst-case value of the objective function, and (ii) the worst-case expectation knowing the mean, with or without dispersion information.

Example 2. Consider the following problem:

$$\min_{x} \; \{x \min\{-z, -0.5z\} + 0.1x\}$$
$$\text{s.t.} \; 0 \leqslant x \leqslant 1,$$

where $z \in \mathscr{Z} = [-1, 1]$ and there are two possible ambiguity sets for z:

$$\mathscr{P}_{(\mu)} = \{\mathbb{P}: z \in \mathscr{Z}, \mathbb{E}_{\mathbb{P}} z = 0\},$$
$$\mathscr{P}_{(\mu, d, \beta)} = \{\mathbb{P}: z \in \mathscr{Z}, \mathbb{E}_{\mathbb{P}} z = 0, \; \mathbb{E}_{\mathbb{P}} |z - 0| = 0.8, \; \mathbb{P}(z \geqslant 0) = 0.5\}.$$

If the objective is to minimize the worst-case value of the objective function then we have that

$$\sup_{z \in [-1, 1]} \{x \min\{-z, -0.5z\} + 0.1x\} =$$
$$x \sup_{z \in [-1, 1]} \min\{-z, -0.5z\} + 0.1x = 0.6x;$$

and the optimal solution is given by $x^{WC} = 0$. If the objective is to minimize the worst-case expectation over $\mathscr{P}_{(\mu)}$ then we have

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu)}} \{\mathbb{E}_{\mathbb{P}} x \min\{-z, -0.5z\} + 0.1x\} =$$
$$x \sup_{\mathbb{P} \in \mathscr{P}_{(\mu)}} \{\mathbb{E}_{\mathbb{P}} \min\{-z, -0.5z\}\} + 0.1x = 0x + 0.1x = 0.1x,$$

where the third equality follows from the Jensen's inequality and the optimal solution is given by $x^{(\mu)} = 0$. However, when the extra information on the MAD and the β is used to minimize the worst-case expectation over $\mathscr{P}_{(\mu, d, \beta)}$, we obtain by the results of Section 2 that

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}} \{\mathbb{E}_{\mathbb{P}} x \min\{-z, -0.5z\} + 0.1x\}$$
$$= x \sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d, \beta)}} \{\mathbb{E}_{\mathbb{P}} \min\{-z, -0.5z\}\} + 0.1x$$
$$= 0.5 \min\{-1(-0.8), -0.5(-0.8)\}$$
$$\quad + 0.5 \min\{-1(0.8), -0.5(0.8)\}$$
$$= -0.1x;$$

and the optimal solution is given by $x^{(\mu, d, \beta)} = 1$, at the opposite end of the feasible interval compared to the previous two solutions. This shows the importance of exploiting the nonzero dispersion in the case of concavity in the uncertain parameter, which is possible using the (μ, d, β) information. □

Implementation error. The third application we consider is when the decision variables cannot be implemented with the designed value because of implementation error in the following problem:

$$\min_{\mathbf{x}, t} \; t$$
$$\text{s.t.} \; f(\mathbf{x}) \leqslant t, \tag{24}$$
$$g_i(\mathbf{x}) \leqslant 0, \quad i = 1, \dots, m.$$

In the case of the existence of an additive implementation error $\mathbf{z}$ the implemented value is $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{z}$, where $\bar{\mathbf{x}}$ is the designed value and $\mathbf{z} = (z_1, \ldots, z_{n_x})^\top$ is the error. Then, the problem becomes

$$
\begin{aligned}
\min_{\bar{\mathbf{x}}, t} \quad & t \\
\text{s.t.} \quad & \sup_{\mathbf{z} \in \mathcal{Z}} f(\bar{\mathbf{x}} + \mathbf{z}) \leq t, \\
& \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\bar{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \ldots, m.
\end{aligned}
\tag{25}
$$

Since $f(\mathbf{x})$ is convex in $\mathbf{x}$, in (25) the function $f(\bar{\mathbf{x}} + \mathbf{z})$ is convex in $\mathbf{z}$. For that reason, optimization of the worst-case value of the objective function could be difficult, as typically RO techniques rely on the constraint being concave in the uncertain parameter (see Ben-Tal et al. 2009, 2015). Therefore, optimizing the worst-case values of convex constraints under implementation error is a problem leading to computational intractability, apart from special cases such as linear constraints (see Ben-Tal et al. 2015) or (conic) quadratic constraints with simultaneously diagonalizable quadratic forms defining the constraint and the uncertainty set for the error (see Ben-Tal and den Hertog 2011).

Because of the above, it may be an alternative to optimize the worst-case expectation of the objective function, for which our DRO method applies under the corresponding distributional assumptions on $\mathbf{z}$, i.e., that the ambiguity set for the distribution of $\mathbf{z}$ is $\mathcal{P}_{(\mu, d)}$. Then, the problem becomes

$$
\begin{aligned}
\min_{\bar{\mathbf{x}}, t} \quad & t \\
\text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{x}} + \mathbf{z}) \leq t, \\
& \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\bar{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \ldots, m.
\end{aligned}
\tag{26}
$$

The first constraint in (26) is convex in $\mathbf{z}$ and one can apply the reformulation (20). For (26) to be tractable, the functions $g_i(\bar{\mathbf{x}} + \mathbf{z})$ need to be affine in $\mathbf{z}$ or need to belong to one of the special cases considered in Ben-Tal and den Hertog (2011). Similarly, one can reformulate a problem where multiplicative error occurs, i.e., where $\mathbf{x} = (\bar{x}_1 z_1, \ldots, \bar{x}_{n_x} z_{n_x})^\top$.

Convex constraints and linear decision rules. The fourth application of our DRO approach comes when the constraints of a problem are convex in z as a result of applying linear decision rules. To show how such a situation occurs, we consider a two-stage RO problem:

$$
\begin{aligned}
\min_{\mathbf{x}_1, \mathbf{x}_2, t} \quad & t \\
\text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq t, \\
& \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq 0, \quad i = 1, \ldots, m,
\end{aligned}
\tag{27}
$$

where $\mathbf{x}_1 \in \mathbb{R}^{n_{x_1}}$ is implemented before $\mathbf{z}$ is known (time 1) and $\mathbf{x}_2 \in \mathbb{R}^{n_{x_2}}$ is implemented after $\mathbf{z}$ is known

(time 2), i.e., $\mathbf{x}_2 = \mathbf{x}_2(\mathbf{z})$. In such cases, it is possible to define the time-2 decisions as a linear function $\mathbf{x}_2(\mathbf{z}) = \mathbf{v} + \mathbf{V}\mathbf{z}$ of the uncertain parameter z (see Ben-Tal et al. 2004), to provide adjustability of decisions at time 2.[2] The problem is then

$$
\begin{aligned}
\min_{\mathbf{x}_1, \mathbf{v}, \mathbf{V}, t} \quad & t \\
\text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{v} + \mathbf{V}\mathbf{z}, \mathbf{z}) \leq t, \\
& \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{v} + \mathbf{V}\mathbf{z}, \mathbf{z}) \leq 0, \quad i = 1, \ldots, m.
\end{aligned}
\tag{28}
$$

If $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z})$ is jointly convex in $(\mathbf{x}_2, \mathbf{z})$, the first constraint in (28) is also convex in $\mathbf{z}$. In such a case, a further reformulation of problem (27) can be conducted as in Section 3.3. We combine linear decision rules with (μ, d) information in the inventory problem of Sections 6.1 and 6.2.

4. Extension—Aggregated Random Vectors

4.1. Introduction

Up to now, we have been deriving exact worst-case expectations that (i) rely on the assumption of independence of the components of $\mathbf{z}$, and (ii) result in bounds involving 3^{n_z} terms. In this section, we consider practically relevant cases where both of these difficulties can be alleviated.

In many cases, uncertainty appears in a linearly aggregated way such as $y = \mathbf{a}^\top \mathbf{z}$ or $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^\top \mathbf{z}$ with $\mathbf{a}(\cdot)$ being affine. Then, instead of considering the worst-case expectation of $f(\mathbf{x}, \mathbf{a}^\top \mathbf{z})$ or $f(\mathbf{a}(\mathbf{x})^\top \mathbf{z})$ with respect to $\mathbf{z}$, it is possible to consider the worst-case expectations of $f(\mathbf{x}, y)$ with respect to the single-dimensional $y = \mathbf{a}^\top \mathbf{z}$ or $y = \mathbf{a}(\mathbf{x})^\top \mathbf{z}$ whose parameters are estimated from the information about $\mathbf{z}$. However, since the estimates of the MAD of $\mathbf{a}^\top \mathbf{z}$ and $\mathbf{a}(\mathbf{x})^\top \mathbf{z}$ will not always be exact, we need the following proposition, which says that it is sufficient to obtain an upper bound on the MAD of the single random variable to get a valid bound on the worst-case expectation.

Proposition 3. *The worst-case expectation* (8) *is a nondecreasing function of d.*

Proof. The worst-case expectation (8) is

$$
\begin{aligned}
F(d) = \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}\right) \\
\cdot f(\mu) + \frac{d}{2(b - \mu)} f(b).
\end{aligned}
$$

We claim that

$$
\begin{aligned}
F'(d) = \frac{1}{2(\mu - a)} f(a) - \left(\frac{1}{2(\mu - a)} + \frac{1}{2(b - \mu)}\right) f(\mu) \\
+ \frac{1}{2(b - \mu)} f(b) \geq 0.
\end{aligned}
$$

Indeed, multiplying the last inequality by $2(b - \mu) \cdot (\mu - a)/(b - a)$ and using

$$\mu = \frac{b - \mu}{b - a}a + \frac{\mu - a}{b - a}b$$

we obtain the inequality

$$\frac{b - \mu}{b - a}f(a) + \frac{\mu - a}{b - a}f(b) \geqslant f\left(\frac{\mu - a}{b - a}b + \frac{b - \mu}{b - a}a\right),$$

which is valid by convexity of $f(\cdot)$. $\quad\square$

In the following, we show how the aggregated approach can be used. Without loss of generality, we assume that the support of $\mathbf{z}$ is given by $\|\mathbf{z}\|_\infty \leqslant 1$ and that $-1 \leqslant \mu \leqslant 1$.

4.2. Fixed Vector a

We begin our analysis with the case where the vector $\mathbf{a}$ is not dependent on the decision variables, motivated by the following example.

Example 3. In an inventory problem the holding and backlogging costs at time $t + 1$ depend on the state of inventory x_{t+1}. If the ordering decisions $q_t(\mathbf{z}^{t-1})$ are static (nonadjustable), then

$$x_{t+1} = x_1 + \sum_{s=1}^{T}q_s - \sum_{s=1}^{T}z_s = x_1 + \sum_{s=1}^{T}q_s - \mathbf{1}^\top\mathbf{z}^t,$$

and the aggregated random variable is $y = \mathbf{1}^\top\mathbf{z}^t$ not depending on the decision variables q_t. $\quad\square$

To use the results of BH to construct the worst-case expectation of $f(\mathbf{x}, y)$, we need to extract the distributional information on $y = \mathbf{a}^\top\mathbf{z}$ from the information on $\mathbf{z}$. We have that

$$\text{supp}(\mathbf{a}^\top\mathbf{z}) = \left[\min_{\mathbf{z}}\mathbf{a}^\top\mathbf{z}, \max_{\mathbf{z}}\mathbf{a}^\top\mathbf{z}\right] = [-\|\mathbf{a}\|_1, \|\mathbf{a}\|_1] \quad (29)$$
$$\mathbb{E}_\mathbb{P}(\mathbf{a}^\top\mathbf{z}) = \mathbf{a}^\top\mu.$$

As stated in Proposition 3, any upper bound on the MAD $M(\mathbf{a}^\top\mathbf{z})$ will generate an upper bound on $\mathbb{E}_\mathbb{P}f(\mathbf{x}, y)$. In the next two propositions, we present four upper bounds $\bar{M}(\mathbf{a}^\top\mathbf{z})$ on $M(\mathbf{a}^\top\mathbf{z})$. The three bounds of Proposition 4 do not use the assumption of independence of the components of $\mathbf{z}$ and are computable in polynomial time.

Proposition 4. *If the distribution $\mathbb{P}$ of $\mathbf{z}$ satisfies*

$$\mathbb{P} \in \mathscr{P}_{(\mu, d)}^{\text{dep}} = \{\mathbb{P}: \text{supp}(z_i) \subseteq [-1, 1], \mathbb{E}_\mathbb{P}(z_i) = \mu_i,$$
$$\mathbb{E}_\mathbb{P}|z_i - \mu_i| = d_i, \forall i\},$$

then the following bounds hold for the MAD of $y = \mathbf{a}^\top\mathbf{z}$:

$$\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top\mu| \leqslant \bar{M}_1(\mathbf{a}^\top\mathbf{z}) := |\mathbf{a}|^\top\mathbf{d} \quad (30)$$

and

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d)}}\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top\mu| \leqslant \bar{M}_2(\mathbf{a}^\top\mathbf{z})$$
$$:= \min_{\phi_1, \phi_2 \geqslant 0, w, \beta, \kappa} w$$
$$\text{s.t.} \quad \mathbf{b}^\top\beta + \kappa \leqslant w,$$
$$\mathbf{c}^\top\phi_1 - \mathbf{a}^\top\mu \leqslant \kappa,$$
$$\mathbf{c}^\top\phi_2 + \mathbf{a}^\top\mu \leqslant \kappa, \quad (31)$$
$$\mathbf{C}^\top\phi_1 + \mathbf{A}^\top\beta = \mathbf{a},$$
$$\mathbf{C}^\top\phi_2 + \mathbf{A}^\top\beta = -\mathbf{a},$$
$$\mathbf{D}^\top\phi_1 + \mathbf{B}^\top\beta = 0,$$
$$\mathbf{D}^\top\phi_2 + \mathbf{B}^\top\beta = 0,$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{2n_\mathbf{z} \times n_\mathbf{z}}, \mathbf{b} \in \mathbb{R}^{2n_\mathbf{z}}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{6n_\mathbf{z} \times n_\mathbf{z}}, and \mathbf{c} \in \mathbb{R}^{6n_\mathbf{z}}$ are defined as

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mu \\ \mathbf{d} \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \\ \mathbf{I} \\ -\mathbf{I} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{I} \\ -\mathbf{I} \\ \mathbf{I} \\ -\mathbf{I} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mu \\ -\mu \\ 2 \cdot \mathbf{1} \\ \mathbf{0} \end{bmatrix}.$$

If the covariance matrix Σ of $\mathbf{z}$ is available, then another bound is

$$\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top\mu| \leqslant \bar{M}_3(\mathbf{a}^\top\mathbf{z}) := \sqrt{\mathbf{a}^\top\Sigma\mathbf{a}}. \quad\square \quad (32)$$

Proof. For the proof of (30) we have

$$\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top\mu| = \mathbb{E}_\mathbb{P}\left|\sum_{i=1}^{n_\mathbf{z}}a_iz_i - \sum_{i=1}^{n_\mathbf{z}}a_i\mu_i\right|$$
$$\leqslant \sum_{i=1}^{n_\mathbf{z}}\mathbb{E}_\mathbb{P}|a_iz_i - a_i\mu_i| = \sum|a_i|d_i = |\mathbf{a}|^\top\mathbf{d}.$$

For the proof of (32) we have

$$\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top\mu| \leqslant \sqrt{\mathbb{E}(y - \mathbf{a}^\top\mu)^2} = \sqrt{\mathbb{E}(\mathbf{a}^\top\mathbf{z} - \mathbf{a}^\top\mu)^2}$$
$$= \sqrt{\text{Var}(\mathbf{a}^\top\mathbf{z})} = \sqrt{\mathbf{a}^\top\Sigma\mathbf{a}}.$$

For the proof of bound (31) note that the maximum value of

$$\mathbb{E}_\mathbb{P}|\mathbf{a}^\top\mathbf{z} - \mathbf{a}^\top\mu|$$

is bounded above, in line with the notation of Wiesemann et al. (2014), by

$$\sup_{\mathbb{P}_{(\mathbf{z}, \mathbf{u})} \in \mathscr{P}'}\mathbb{E}_{\mathbb{P}_{(\mathbf{z}, \mathbf{u})}}\max\{\mathbf{a}^\top\mathbf{z} - \mathbf{a}^\top\mu, -\mathbf{a}^\top\mathbf{z} + \mathbf{a}^\top\mu\}, \quad (33)$$

where

$$\mathscr{P}' = \left\{\mathbb{P}: \begin{array}{l} \mathbb{E}_\mathbb{P}(\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}) = \mathbf{b} \\ \mathbb{P}((\mathbf{z}, \mathbf{u}) \in \mathscr{C}) = 1 \end{array}\right\}, \quad \mathscr{C} = \{(\mathbf{z}, \mathbf{u}): \mathbf{C}\mathbf{z} + \mathbf{D}\mathbf{u} \leqslant \mathbf{c}\},$$

where the vector $\mathbf{u} \in \mathbf{R}^{n_\mathbf{z}}$ consists of components u_i, each of which is an auxiliary analysis variable corresponding to the MAD of z_i. The first (moment condition) in the definition of $\mathscr{P}'$ should ensure that the first

moment of $\mathbf{z}$ is equal to μ and the first moment of $\mathbf{u}$ is equal to $\mathbf{d}$. We define thus $\mathbf{A}$, $\mathbf{B}$, $\mathbf{b}$ as in the theorem.

The second (support) condition in the definition of $\mathscr{P}'$ should ensure that the support of $\mathbf{z}$ is the unit box and that $\mathbf{u}$ indeed corresponds to an upper bound on the deviation of $\mathbf{u}$. We need to ensure that

$$\|\mathbf{z}\|_\infty \leqslant 1, \quad \mathbf{u} \geqslant \mathbf{z} - \mu, \quad \mathbf{u} \geqslant \mu - \mathbf{z}, \quad \mathbf{u} \geqslant 0, \quad \mathbf{u} \leqslant 2 \cdot \mathbf{1},$$

where the last condition ensures boundedness of $\mathscr{C}$, required by Wiesemann et al. (2014). We ensure these conditions by setting $\mathbf{C}$, $\mathbf{D}$, $\mathbf{c}$ as in the theorem. Wiesemann et al. (2014) prove (Theorem 1 in their paper) that under mild conditions, satisfied in our case, (33) is equivalent to the LP as stated in the theorem. □

There are cases where the MAD bound (30) is tight, for example, $\mu_i = 0$ and $d_i = d_j$ for all $i \neq j$. Thanks to bound (31), the method of Wiesemann et al. (2014) can be used to enhance our method for aggregated random variables. Building up the upper bound on the MAD of $\mathbf{a}^\top \mathbf{z}$ via (31) is preferable to (30) if solving optimization problems is not burdensome.

The fourth bound, given in Proposition 5 requires the independence of z_i's.

Proposition 5. *If $\mathbb{P} \in \mathscr{P}_{(\mu, d)}$ then*

$$\mathbb{E}_\mathbb{P}|y - \mathbf{a}^\top \mu| \leqslant \bar{M}_4(\mathbf{a}^\top \mathbf{z}) = \sum_{\alpha \in \{1,2,3\}^{n_\mathbf{z}}} \left(\prod_{i=1}^{n_\mathbf{z}} p^i_{\alpha_i} \right) |\mathbf{a}^\top \mathbf{z}(\alpha) - \mathbf{a}^\top \mu|.$$
(34)

Proof. The proposition follows from considering the ambiguity set for the probability distribution $\mathbb{P}_y$ of a single-dimensional random variable $y = \mathbf{a}^\top \mathbf{z}$:

$$\mathscr{P}_y = \{\mathbb{P}_y : \mathrm{supp}(y) \subseteq [-\|\mathbf{a}\|_1, \|\mathbf{a}\|_1], \mathbb{E}_\mathbb{P}(y) = \mathbf{a}^\top \mu,$$
$$\mathbb{E}_{\mathbb{P}_y}|y - \mathbf{a}^\top \mu| = \mathbb{E}_\mathbb{P}|\mathbf{a}^\top \mathbf{z} - \mathbf{a}^\top \mu|\}$$

and applying Proposition 3 since $\bar{M}_4(\mathbf{a}^\top \mathbf{z}) \geqslant \mathbb{E}_\mathbb{P}|\mathbf{a}^\top \mathbf{z} - \mathbf{a}^\top \mu|$. □

With respect to (34), we note that the computation can be done before the optimization problem is set up. Therefore, the optimization problem involving $f(\mathbf{x}, \mathbf{a}^\top \mathbf{z})$ is easier than would be the case if formula (20) of Section 3.1 were used.

The following proposition states that under the obtained distributional information on $\mathbf{a}^\top \mathbf{z}$ one can obtain an upper bound on the expectation of $f(\mathbf{x}, \mathbf{a}^\top \mathbf{z})$.

Proposition 6. *It holds for $i = 1, 2, 3$ that*

$$\sup_{\mathbb{P} \in \mathscr{P}^{\mathrm{dep}}_{(\mu, d)}} f(\mathbf{x}, \mathbf{a}^\top \mathbf{z}) \leqslant \frac{\bar{M}_i(\mathbf{a}^\top \mathbf{z})}{2(\mathbf{a}^\top \mu + \|\mathbf{a}\|_1)} f(\mathbf{x}, -\|\mathbf{a}\|_1)$$
$$+ \left(1 - \frac{\bar{M}_i(\mathbf{a}^\top \mathbf{z})}{2(\mathbf{a}^\top \mu + \|\mathbf{a}\|_1)} - \frac{\bar{M}_i(\mathbf{a}^\top \mathbf{z})}{2(\|\mathbf{a}\|_1 - \mathbf{a}^\top \mu)} \right) f(\mathbf{x}, \mathbf{a}^\top \mu)$$
$$+ \frac{\bar{M}_i(\mathbf{a}^\top \mathbf{z})}{2(\|\mathbf{a}\|_1 - \mathbf{a}^\top \mu)} f(\mathbf{x}, \|\mathbf{a}\|_1)$$
(35)

and

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu, d)}} f(\mathbf{x}, \mathbf{a}^\top \mathbf{z}) \leqslant \frac{\bar{M}_4(\mathbf{a}^\top \mathbf{z})}{2(\mathbf{a}^\top \mu + \|\mathbf{a}\|_1)} f(\mathbf{x}, -\|\mathbf{a}\|_1)$$
$$+ \left(1 - \frac{\bar{M}_4(\mathbf{a}^\top \mathbf{z})}{2(\mathbf{a}^\top \mu + \|\mathbf{a}\|_1)} - \frac{\bar{M}_4(\mathbf{a}^\top \mathbf{z})}{2(\|\mathbf{a}\|_1 - \mathbf{a}^\top \mu)} \right) f(\mathbf{x}, \mathbf{a}^\top \mu)$$
$$+ \frac{\bar{M}_4(\mathbf{a}^\top \mathbf{z})}{2(\|\mathbf{a}\|_1 - \mathbf{a}^\top \mu)} f(\mathbf{x}, \|\mathbf{a}\|_1),$$
(36)

where $\bar{M}_i, i = 1, \ldots, 4$ are the bounds given in Propositions 4 and 5. □

The statement follows trivially from (8) and Proposition 3. Since $\mathbf{a}$ does not depend on $\mathbf{x}$, the right-hand sides in the bounds of Proposition 6 are convex in $\mathbf{x}$.

4.3. Vector $\mathbf{a}(\mathbf{x})$ Depends on $\mathbf{x}$

We now assume that $\mathbf{a}(\mathbf{x})$ is an affine vector-valued function of $\mathbf{x}$ and the function whose worst-case expectation we seek is $f(\mathbf{a}(\mathbf{x})^\top \mathbf{z})$, where $f(\cdot)$ is convex and $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^\top \mathbf{z}$. This assumption holds, for example, for an inventory problem with linear decision rules.

Example 4. Using linear decision rules $q_{t+1}(\mathbf{z}^t) = q_{t+1,0} + \sum_{j=1}^t q_{t+1, j} z_j$ for the ordering decisions, the state of inventory at time $t + 1$ is

$$x_{t+1} = x_1 + \sum_{s=1}^t \left(q_{t,0} + \sum_{j=1}^t q_{t,j} z_j \right) - \sum_{j=1}^t z_t$$
$$= x_1 + \sum_{s=1}^t q_{t,0} + \sum_{s=1}^t \left(\sum_{j=s+1}^t q_{j,s} - 1 \right) z_s.$$

In each time period the aggregated random variable is $\sum_{s=1}^t (\sum_{j=s+1}^t q_{j,s} - 1) z_s$, which indeed depends on the decision variables. □

Similar to the previous case, one can consider the worst-case expectation of $f(y(\mathbf{x}))$ where $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^\top \mathbf{z}$. We assume that the function $f(\cdot)$ is nonnegative—if $f(\cdot)$ is not nonnegative but is bounded from below, nonnegativity can be ensured by adding a sufficiently large constant. We obtain the following bound using the MAD approximation (30), which does not require the assumption of independence of z_i's.

Proposition 7. *For $f(\cdot)$ nonnegative it holds that*

$$\sup_{\mathbb{P} \in \mathscr{P}^{\mathrm{dep}}_{(\mu, d)}} f(\mathbf{a}(\mathbf{x})^\top \mathbf{z})$$
$$\leqslant \max_i \frac{d_i}{2(1 - |\mu_i|)} (f(-\|\mathbf{a}(\mathbf{x})\|_1) + f(\|\mathbf{a}(\mathbf{x})\|_1))$$
$$+ \left(1 - \min_i \frac{d_i}{1 + |\mu_i|} \right) f(\mathbf{a}(\mathbf{x})^\top \mu). \quad \square$$
(37)

Proof. Consider the single-dimensional random variable $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^\top \mathbf{z}$. Similar to the case $y = \mathbf{a}^\top \mathbf{x}$ it holds that

$$\mathrm{supp}(\mathbf{a}(\mathbf{x})^\top \mathbf{z}) = \left[\min_{\mathbf{z}} \mathbf{a}(\mathbf{x})^\top \mathbf{z}, \max_{\mathbf{z}} \mathbf{a}(\mathbf{x})^\top \mathbf{z} \right]$$

$$= [-\|\mathbf{a}(\mathbf{x})\|_1, \|\mathbf{a}(\mathbf{x})\|_1]$$

$$\mathbb{E}_{\mathbb{P}}(\mathbf{a}(\mathbf{x})^\top \mathbf{z}) = \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu}$$

$$\mathbb{E}_{\mathbb{P}}|\mathbf{a}(\mathbf{x})^\top \mathbf{z} - \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu}| \leqslant |\mathbf{a}(\mathbf{x})|^\top \mathbf{d}.$$

Using this and Proposition 3 to obtain an upper bound on the expectation of the variable $y(\mathbf{x})$ we obtain

$$\sup_{\mathbb{P}_{y(\mathbf{x})} \in \mathscr{P}_{y(\mathbf{x})}} \mathbb{E}_{\mathbb{P}_{y(\mathbf{x})}} f(y(\mathbf{x})) \leqslant \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} f(-\|\mathbf{a}(\mathbf{x})\|_1)$$

$$+ \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\|\mathbf{a}(\mathbf{x})\|_1 - \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu})} f(\|\mathbf{a}(\mathbf{x})\|_1)$$

$$+ \left(1 - \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} - \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\|\mathbf{a}(\mathbf{x})\|_1 - \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu})} \right)$$

$$\cdot f(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu}). \tag{38}$$

Now note that it holds that

$$\frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} = \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu} + |\mathbf{a}(\mathbf{x})|^\top \mathbf{1})}$$

$$\leqslant \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(-|\mathbf{a}(\mathbf{x})|^\top |\boldsymbol{\mu}| + |\mathbf{a}(\mathbf{x})|^\top \mathbf{1})}$$

$$= \frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2|\mathbf{a}(\mathbf{x})|^\top (\mathbf{1} - |\boldsymbol{\mu}|)}$$

$$\leqslant \max_i \frac{d_i}{2(1 - |\mu_i|)}.$$

In a similar way, we obtain that

$$\frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\|\mathbf{a}(\mathbf{x})\|_1 - \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu})} \leqslant \max_i \frac{d_i}{2(1 - |\mu_i|)}$$

and

$$\frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} \geqslant \min_i \frac{d_i}{2(1 + |\mu_i|)},$$

$$\frac{|\mathbf{a}(\mathbf{x})|^\top \mathbf{d}}{2(\|\mathbf{a}(\mathbf{x})\|_1 - \mathbf{a}(\mathbf{x})^\top \boldsymbol{\mu})} \geqslant \min_i \frac{d_i}{2(1 + |\mu_i|)}.$$

Inserting these inequalities into (38) and using the nonnegativity of $f(\cdot)$ yields the result. $\square$

The quality of the bound (37) depends now on the dispersion of the terms $d_i/(2(1 - |\mu_i|))$ and $d_i/(2(1 + |\mu_i|))$ if they are equal, the bound is tight. Tractability of (37) depends on the convexity of the sum

$$f(-\|\mathbf{a}(\mathbf{x})\|_1) + f(\|\mathbf{a}(\mathbf{x})\|_1), \tag{39}$$

which turns out to be the case if $\mathbf{a}(\mathbf{x})$ is affine, as the following proposition shows.

Proposition 8. *For affine* $\mathbf{a}(\mathbf{x})$ *and convex* $f: \mathbb{R} \to \mathbb{R}$, *the function* $f(-\|\mathbf{a}(\mathbf{x})\|_1) + f(\|\mathbf{a}(\mathbf{x})\|_1)$ *is convex.*

Proof. Define $g(t) = f(t) + f(-t)$ for $t \in \mathbb{R}_+$ and $h(\mathbf{x}) = \|\mathbf{a}(\mathbf{x})\|_1$. Then we have that

$$f(-\|\mathbf{a}(\mathbf{x})\|_1) + f(\|\mathbf{a}(\mathbf{x})\|_1) = g(h(\mathbf{x})).$$

The function $g(h(\mathbf{x}))$ is convex if $g(t)$ is convex and nondecreasing and $h(\mathbf{x})$ is convex. Convexity of $h(\mathbf{x})$ is clear as it is a norm of an affine function of $\mathbf{x}$. Also, convexity of $g(\cdot)$ follows from convexity of $f(\cdot)$. We need to show that $g(\cdot)$ is nondecreasing, i.e., that

$$g(t + \alpha) \geqslant g(t), \quad \forall t \geqslant 0, \, \alpha \geqslant 0.$$

Consider the subgradients $v_1 \in \partial f(-t)$ and $v_2 \in \partial f(t)$. By properties of subgradients we have that

$$v_1 \leqslant \frac{f(t) - f(-t)}{2t} \leqslant v_2.$$

From this, it follows that for $\alpha \geqslant 0$,

$$g(t + \alpha) = f(-t - \alpha) + f(t + \alpha)$$

$$\geqslant f(-t) + \sup_{v \in \partial f(-t)} (-\alpha v) + f(t) + \sup_{v \in \partial f(t)} (\alpha v)$$

$$\geqslant f(-t) + (-\alpha v_1) + f(t) + (\alpha v_2)$$

$$= g(t) + \alpha(v_2 - v_1)$$

$$\geqslant g(t). \quad \square$$

5. Safe Approximations of Chance Constraints

5.1. Introduction
In this section we show how the results of Ben-Tal and Hochman (1972) can be used to construct safe tractable approximations of scalar chance constraints:

$$\mathbb{P}(\mathbf{a}(\mathbf{z})^\top \mathbf{x} > \mathbf{b}(\mathbf{z})) \leqslant \epsilon, \quad \forall \mathbb{P} \in \mathscr{P}_{(\mu, d)},$$

$$\text{where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0]. \tag{40}$$

For ease of exposure we assume that the components z_i have a support contained in $[-1, 1]$ and mean 0:

$$\mathscr{P}_{(\mu, d)} = \{\mathbb{P}: \mathrm{supp}(z_i) \subseteq [-1, 1], \, \mathbb{E}_{\mathbb{P}} z_i = 0, \, \mathbb{E}_{\mathbb{P}} |z_i| = d_i,$$

$$i = 1, \ldots, n_{\mathbf{z}}, \, z_i \perp\!\!\!\perp z_j, \, \forall i \neq j\}.$$

To construct the safe tractable approximations, we use the mathematical framework of Ben-Tal et al. (2009). In this framework, the key step consists of bounding from above the *moment-generating function* of z_i, $i = 1, \ldots, n_{\mathbf{z}}$:

$$\mathbb{E}_{\mathbb{P}} \exp(w z_i) = \int \exp(w z_i) \, d\mathbb{P}_i(z_i),$$

and then using the resulting bound in combination with the Markov inequality to obtain upper bounds on

the probability $\mathbb{P}(\mathbf{a}(\mathbf{z})^\top \mathbf{x} > \mathbf{b}(\mathbf{z}))$—often referred to as the *Bernstein approximation*.

A strong motivation for using the ambiguity set $\mathcal{P}_{(\mu,d)}$ is because of the fact that a tight explicit bound on $\mathbb{E}_\mathbb{P} \exp(\mathbf{w}^\top \mathbf{z})$ is obtained easily in this setting by the BH results described in Section 2. Indeed, because of the independence of $z_1, \ldots, z_{n_\mathbf{z}}$ we have

$$
\begin{aligned}
&\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_\mathbb{P} \exp(\mathbf{z}^\top \mathbf{w}) \\
&= \prod_{i=1}^{n_\mathbf{z}} \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_\mathbb{P} \exp(z_i w_i) \\
&= \prod_{i=1}^{n_\mathbf{z}} \left(\frac{d_i}{2} \exp(-w_i) + \frac{d_i}{2} \exp(w_i) + (1 - d_i) \exp(0) \right) \\
&= \prod_{i=1}^{n_\mathbf{z}} (d_i \cosh(w_i) + 1 - d_i). \quad (41)
\end{aligned}
$$

The worst-case expectation is evaluated separately for each component of $\mathbf{z}$, avoiding the computational burden of summation of $3^{n_\mathbf{z}}$ terms as in (14).

5.2. Safe Approximations—Results

We now show how (41) can be used to obtain safe approximations of (40). First, we present two simple safe approximations in order of increasing tightness, which are similar in their functional form to those obtained in Ben-Tal et al. (2009) for the case of mean-variance information. Next, we show that the (μ, d) information is particularly suitable for obtaining even tighter safe approximations, based on the exponential polynomials, which, in the sense of computational tractability, remained an unsolved case in Ben-Tal et al. (2009). The numerical experiment in Section 6.5 illustrates the power of this new approximation compared to the previous two.

The first approximation requires the use of theorem 2.4.4 of Ben-Tal et al. (2009), repeated in the electronic companion, Appendix EC.4.1.

Theorem 1. *If for a given vector* $\mathbf{x}$ *there exist* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_\mathbf{z}+1}$ *such that* $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ *satisfies the constraint system*

$$
\begin{cases}
(\mathbf{a}^i)^\top \mathbf{x} - b_i = u_i + v_i, & 0 \leqslant i \leqslant n_\mathbf{z} \\
u_0 + \sum_{i=1}^{n_\mathbf{z}} |u_i| \leqslant 0 \\
v_0 + \sqrt{2 \log(1/\epsilon)} \sqrt{\sum_{i=1}^{n_\mathbf{z}} \sigma_i^2 v_i^2} \leqslant 0,
\end{cases}
\quad (42)
$$

where

$$
\sigma_i = \sup_{t \in \mathbb{R}} \sqrt{\frac{2 \log(d_i \cosh(t) + 1 - d_i)}{t^2}}, \quad (43)
$$

then $\mathbf{x}$ *is feasible to* (40)*, that is, constraint system* (42) *is a safe approximation of* (40)*. Moreover,* (42) *is the robust counterpart of the following robust constraint:*

$$
\mathbf{a}(\mathbf{z})^\top \mathbf{x} \leqslant \mathbf{b}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{U},
$$

$$
\text{where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_\mathbf{z}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0], \quad (44)
$$

and

$$
\mathcal{U} = \left\{ \mathbf{z} \in \mathbb{R}^{n_\mathbf{z}} : \sqrt{\sum_{i=1}^{n_\mathbf{z}} \frac{z_i^2}{\sigma_i^2}} \leqslant \sqrt{2 \log(1/\epsilon)}, \right. \\
\left. -1 \leqslant z_i \leqslant 1, \ i = 1, \ldots, n_\mathbf{z} \right\}.
$$

Constraint system (42) involves only linear and second-order conic constraints, making it highly tractable even for large-dimensional problems.

The second safe approximation is tighter and relies on the somewhat more complicated mathematical machinery of Ben-Tal et al. (2009).

Theorem 2. *If there exists* $\alpha > 0$ *such that* $(\mathbf{x}, \alpha)$ *satisfies the constraint*

$$
(\mathbf{a}^0)^\top \mathbf{x} - b_0 + \alpha \log \left(\sum_{i=1}^{n_\mathbf{z}} \left(d_i \cosh \left(\frac{(\mathbf{a}^i)^\top \mathbf{x} - b_i}{\alpha} \right) + 1 - d_i \right) \right) \\
+ \alpha \log(1/\epsilon) \leqslant 0, \quad (45)
$$

then $\mathbf{x}$ *satisfies constraint* (40)*. That, is* (45) *is a safe approximation of* (40)*.*

Similar to Theorem 1, one can construct an explicit convex uncertainty set $\mathcal{U}$ for which (45) is the robust counterpart of (44) corresponding to $\mathcal{U}$. Constraint (45) is convex in $(\mathbf{x}, \alpha)$, being a sum of a linear function and $n_\mathbf{z}$ perspective functions of the convex log-sum-exp function; see Boyd and Vandenberghe (2004). For that reason, it can be handled with convex optimization algorithms such as interior point methods.

5.3. Toward Better Safe Approximations— Exponential Polynomials

Ben-Tal et al. (2009) discuss the fact that the bounds obtained using a single exponential function can still be improved by, instead of the moment-generation function, constructing the worst-case expectation of exponential polynomials:

$$
\gamma(s) = \sum_{v=0}^{L} c_v \exp\{\omega_v s\}, \quad (46)
$$

to bound the probability of constraint violation, where $c_v, \omega_v, v = 0, \ldots, L$ are complex numbers and

$\gamma(\cdot)$ is convex and nondecreasing,
$$
\gamma(s) \geqslant 0, \quad \gamma(0) \geqslant 0, \quad \gamma(s) \to 0 \quad \text{as } s \to -\infty. \quad (47)
$$

The worst-case expectation of the exponential polynomial $\gamma(s)$, similar to the worst-case expectation of the moment-generating function (41), can then be used to obtain better upper bounds on $\mathbb{P}(\mathbf{a}(\mathbf{z})^\top \mathbf{x} > \mathbf{b}(\mathbf{z}))$. In fact, the bound found in Theorem 2 is obtained using a special case of (46), where $L = 0$, $c_0 = \omega_0 = 1$. The difficulty of using general polynomials (46) lies in the (un)availability of tight, computationally tractable upper bounding function $\Psi(\mathbf{w})$ on (46):

$$\mathbb{E}_{\mathbb{P}} \gamma \left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i \right) \leqslant \Psi(\mathbf{w}), \quad \forall \mathbb{P} \in \mathscr{P}.$$

In the following, we show that under (μ, d) information, the result of BH can be easily applied in this case as well. Indeed, the corresponding supremum over $\mathscr{P}_{(\mu,d)}$ is given by

$$\Psi(\mathbf{w}) = \sup_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \gamma \left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} z_i w_i \right)$$

$$= \sum_{v=0}^{L} c_v \exp\{\omega_v w_0\} \prod_{i=1}^{n_{\mathbf{z}}} (d_i \sinh(\omega_v w_i) + 1 - d_i). \tag{48}$$

Now, we can use proposition 4.3.1 from Ben-Tal et al. (2009) to obtain the following result.

Theorem 3. *Consider an exponential polynomial $\gamma(s)$ satisfying (47), the corresponding function $\Psi(\mathbf{w})$ and the set Γ_ϵ such that*

$$\begin{aligned} \Gamma_\epsilon = \{\mathbf{x} \colon \exists\, \alpha > 0 \colon \Psi(\alpha \mathbf{w}) \leqslant \epsilon\}, \\ w_i = (\mathbf{a}^i)^\top \mathbf{x} - b_i, \quad i = 1, \dots, n_{\mathbf{z}}. \end{aligned} \tag{49}$$

Then, any $\mathbf{x} \in \mathrm{cl}\,\Gamma_\epsilon$ is also feasible for the chance constraint (40).

It is also important to note that constraint (49) is convex representable in $(\mathbf{x}, \alpha)$. Theorem 3 extends the results of Ben-Tal et al. (2009), which provides a safe approximation using only known supports and means of the components z_i.

6. Numerical Experiments
6.1. Inventory Management—Average Case Performance

Introduction. We consider an application of the (μ, d) method to minimization of the average-case cost in inventory management, where we adapt the numerical example from Ben-Tal et al. (2005) with a single product and where inventory is managed periodically over T periods. At the beginning of each period t the decision maker has an inventory of size x_t and he orders a quantity q_t for unit price c_t. The customers then place their demands z_t. The retailer's status at the beginning of the planning horizon is given through the parameter x_1 (initial inventory). Apart from the ordering cost, the

following costs are incurred over the planning horizon (i) holding cost $h_t \max\{0, x_{t+1}\}$, where h_t are the unit holding cost; (ii) shortage cost $p_t \max\{0, x_{t+1}\}$, where p_t is the unit shortage cost.

Inventory x_{T+1} left at the end of period T has a unit salvage value s. Also, one must impose $h_T - s \geqslant -p_T$ to maintain the problem's convexity. The practical interpretation of this constraint is that in the last period it is more profitable to satisfy the customer demand rather than to be left with an excessive amount of inventory. The constraints in the model include (i) balance equations linking the inventory in each period to the inventory, order quantity, and demand in the preceding period; (ii) upper and lower bounds on the order quantities in each period $L_t \leqslant q_t \leqslant U_t$; (iii) upper and lower bounds on cumulative order quantities in each period $\hat{L}_t \leqslant \sum_{\tau=1}^t q_\tau \leqslant \hat{U}_t$.

With ordering decisions $\mathbf{q}(\mathbf{z}) = (q_1, q_2(\mathbf{z}^1), \dots, q_T(\mathbf{z}^{T-1}))^\top$, where $\mathbf{z}^t = (z_1, \dots, z_t)^\top$, the objective function value for a given demand vector $\mathbf{z}$ is

$$\begin{aligned} f(\mathbf{q}(\mathbf{z}), \mathbf{z}) = -s \max\{x_{T+1}(\mathbf{z}^T), 0\} + \sum_{t=1}^{T} (c_t q_t(\mathbf{z}^{t-1}) \\ + h_t \max\{x_{t+1}(\mathbf{z}^t), 0\} + p_t \max\{-x_{t+1}(\mathbf{z}^t), 0\}). \end{aligned}$$

The optimization problem to be solved is given by the following, two-variant formulation where the minimized quantity is the worst-case value or the worst-case expectation of the objective function:

$$\begin{aligned} \min_{\mathbf{q}(\cdot), \mathbf{x}(\cdot), u} \quad & u \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathscr{P}} \mathbb{E}_{\mathbb{P}} \text{ or } \sup_{\mathbf{z} \in \mathscr{Z}} f(\mathbf{q}(\mathbf{z}), \mathbf{x}(\mathbf{z}), \mathbf{z}) \leqslant u, \\ & x_{t+1}(\mathbf{z}^t) = x_t(\mathbf{z}^{t-1}) + q_t(\mathbf{z}^{t-1}) - z_t, \quad t = 1, \dots, T \\ & \hspace{6cm} \forall \mathbf{z} \in \mathscr{Z}, \\ & L_t \leqslant q_t(\mathbf{z}^{t-1}) \leqslant U_t, \quad t = 1, \dots, T \ \forall \mathbf{z} \in \mathscr{Z}, \\ & \hat{L}_t \leqslant \sum_{\tau=1}^{t} q_\tau(\mathbf{z}^{\tau-1}) \leqslant \hat{U}_t, \quad t = 1, \dots, T \ \forall \mathbf{z} \in \mathscr{Z}, \end{aligned} \tag{50}$$

where $\mathscr{Z}$ is the uncertainty set for $\mathbf{z}$ and $\mathscr{P}$ is the ambiguity set of probability distributions with support being a subset of $\mathscr{Z}$. The objective function in (50) has the sum-of-maxima form, which typically is problematic in RO because of the difficulty of maximizing a convex function; see, for example, Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016). Models such as (50) in the multiperiod setting can lead to issues such as time (in)consistency, i.e., the question whether the chosen multiperiod strategy remains optimal at later stages for all possible outcomes of the uncertainty at the earlier stage. This phenomenon is discussed, for example, by Xin et al. (2015).

We assume that the uncertainty set $\mathscr{Z} = \mathscr{Z}_1 \times \cdots \times \mathscr{Z}_T$, where $\mathscr{Z}_t = [a_t, b_t]$, $t = 1, \dots, T$. The worst-case form of

problem (50) has to be solved by enumerating all vertices of the uncertainty set $\mathcal{Z}$. For the worst-case expectation form of (50) we assume that $\mu_t = (a_t + b_t)/2$, and that $d_t = \mathbb{E}_{\mathbb{P}}|z_t - \mu_t| = \theta(b_t - a_t)$, yielding the following ambiguity set:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P}\colon \operatorname{supp}(\mathbb{P}) \subset [\mathbf{a}, \mathbf{b}],\ \mathbb{E}_{\mathbb{P}}\mathbf{z} = \boldsymbol{\mu},$$
$$\mathbb{E}_{\mathbb{P}}|\mathbf{z} - \boldsymbol{\mu}| = \mathbf{d},\ z_i \perp\!\!\!\perp z_j\ \forall\, i \neq j\},$$

where $\mathbf{a} = (a_1,\ldots,a_T)^\top$, $\mathbf{b} = (b_1,\ldots,b_T)^\top$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_T)^\top$, and $\mathbf{d} = (d_1,\ldots,d_T)^\top$. The ordering decisions are assumed to be linear functions of the past demand: $q_{t+1}(\mathbf{z}^t) = q_{t+1,0} + \sum_{j=1}^t q_{t+1,j}z_j$ and require that $q_{t+1}(\mathbf{z}^t) \geqslant 0$ for all $\mathbf{z} \in \mathcal{Z}$, for $t = 1,\ldots,T$. We solve exactly the following two variants of problem (50):

- RO solution—the objective function in (50) is preceded by $\sup_{\mathbf{z} \in \mathcal{Z}}$, and
- (μ, d) solution–the objective function in (50) is preceded by $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}$.

We run an experiment with $T = 6$ and 50 problem instances. We set $\theta = 0.25$, corresponding to the mean absolute deviation of the uniform distribution. The ranges for the uniform sampling of parameters are given in Table 1.

Upper and Lower Bounds for the Expectation of the Objective Function.

For each inventory problem instance and the optimal solution $\bar{\mathbf{q}}(\cdot)$, we compute the following quantities:

- the worst-case expected cost under (μ, d) information: $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$;
- the best-case expected cost $\inf_{\mathbb{P} \in \mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$ with three possibilities for the skewness of the demand distribution, i.e., with $\beta_t = \beta \in \{0.25, 0.5, 0.75\}$, corresponding to left-skewness, symmetry, and right-skewness of the demand distribution in all periods, respectively.

The two values provide us with information about the interval within which the expected objective function value lies under three different assumptions on the parameter β. Additionally, for each solution we compute the worst-case cost $\sup_{\mathbf{z} \in \mathcal{Z}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$ to verify how the minimization of the worst-case expectation affects the worst-case performance of the solution.

Table 2 presents the results. As can be expected, the RO solution yields the best worst-case value of 1,950, which is far better than the (μ, d) solution, whose

Table 2. Results of the Inventory Management—Worst-Case Cost and Ranges for the Expectation of the Objective over $\mathcal{P}_{(\mu,d,\beta)}$

		Minimum cost	
Objective type	β	RO	(μ, d)
Worst-case value	—	1,950	2,384
Expectation range	0.25	[1,255, 1,280]	[1,004, 1,049]
Expectation range	0.5	[1,223, 1,280]	[970, 1,049]
Expectation range	0.75	[1,230, 1,280]	[994, 1,049]

Note. All numbers are averages.

worst-case value is 2,384. Rows 2 to 4 show that the (μ, d) solution not only yields better upper bounds on the expected value of the solution, but also leads to an improvement of the best-case expectation for all β. For example, for $\beta = 0.5$ the interval for the expected cost related to the RO solution is given by [1,255, 1,280], whereas for the (μ, d) solution it is [970, 1,049]. That means that the worst-case expectation obtained by the (μ, d) solution is better than the worst-case expectation obtained by the RO solution.

Simulation Results. Since the solutions are obtained with different objective functions, comparing their average-case performance in a "fair" way is difficult. We compare their performance using two samples of demand vectors $\hat{\mathbf{z}}$:

- uniform sample—demand scenarios $\hat{\mathbf{z}}$ are sampled from a uniform distribution on $\mathcal{Z}$;
- (μ, d) sample—demand scenarios $\hat{\mathbf{z}}$ are sampled from a distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$. That is, first, a discretized distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$ is sampled using the hit-and-run method. This method is implemented here as follows. For the $[0, 1]$-interval we construct a grid of 50 equidistant points. For a fixed (μ, d) the set of probability masses assigned to these points satisfying the μ and d values is a polytope. We sample 10 probability distributions uniformly from this polytope with the classical hit-and-run method (mixing algorithm) of Smith (1984), where we choose the starting point to be the analytic center of the polytope and we use only every 20th vector sampled with the mixing algorithm. Then, each component of the vector $\hat{\mathbf{z}}$ is sampled randomly from a randomly chosen distribution $\hat{\mathbb{P}}$.

For each instance, we sample 10^4 demand scenarios, with both of the sampling methods. Table 3 presents

Table 1. Ranges for Parameter Sampling in the Inventory Experiment

Parameter	Range	Parameter	Range
a_t	$[0, 20]$	x_1	$[20, 50]$
b_t	$[a_t, a_t + 100]$	L_t	0
c_t, p_t	$[0, 10]$	U_t	$[50, 70]$
h_t	$[0, 5]$	$\hat{L}_t$	0
s	0	$\hat{U}_t$	$0.8 \sum_{t=1}^T U_t$

Table 3. Simulation Results for the First Inventory Problem

		Cost	
Objective type	Demand sample type	RO	(μ, d)
Objective mean	Uniform sample	**1,230**	**994**
Objective standard deviation	Uniform sample	157	259
Objective mean	(μ, d) sample	**1,246**	**1,003**
Objective standard deviation	(μ, d) sample	160	265

the results. The averages of the objective function values over the two sample types over all instances are put in bold. For example (row 1), the (μ, d) solutions perform better on average in the uniform sample, with values 994 and 1,230, respectively. A similar observation holds for the (μ, d) sample (row 3). In Figure 1 we present a comparison of the empirical cumulative distribution functions of the total cost incurred by the RO and (μ, d) solutions. Roughly speaking, one can speak of a partial shift of the cumulative distribution function in case of the (μ, d) solution in the desired direction—a higher cost becomes less probable. However, it is important to remember that the RO solutions still provide a better worst-case performance, as given in Table 2. Overall though, the (μ, d) solutions appear to be preferable if the mean outcome is the priority.

6.2. Inventory Management—Enhancement of RO Solutions

With the good average-case performance of the (μ, d) solutions in the previous experiment, we investigate now whether the (μ, d) method can be used to enhance the average-case performance if there are multiple optimal RO solutions. For each of the problem instances of

the previous subsection we apply the two-step procedure of Section 3.3.

We consider three enhancement types, corresponding to three different objective functions:

- (μ, d) enhancement: $\min \sup_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{q}(\mathbf{z}), \mathbf{z})$;

- sample enhancement: $\min (1/S) \sum_{j=1}^{S} f(\mathbf{q}(\hat{\mathbf{z}}_j), \hat{\mathbf{z}}_j)$, where $\hat{\mathbf{z}}_j$ are $S = 200$ demand scenarios sampled uniformly from $\mathscr{Z}$;

- nominal enhancement: $\min f(\mathbf{q}(\mu), \mu)$ considered by Iancu and Trichakis (2013).

Table 4 presents the results. In the uniform sample (row 1) the (μ, d)-enhanced solution yields an average cost of 1,168, compared to 1,230 for the nonenhanced solution, that is 5.04% less. For the (μ, d) sample (row 3) the corresponding number is 5.93%. The nominal enhancement turns out to be slightly worse than the (μ, d) and sample enhancements; compare, for example, the means 1,180 and 1,168 for the uniform sample and the higher standard deviations of the nominal enhancement.

In Figure 2 we present the empirical CDFs of the nonenhanced and (μ, d)-enhanced solutions in a sample problem in the uniform demand sample, where we can observe again a partial shift of the CDF to the left.

Figure 1. Empirical Cumulative Distribution Functions of the Total Cost of the Solutions Under the Uniform Sample (Left) and (μ, d) Sample (Right)
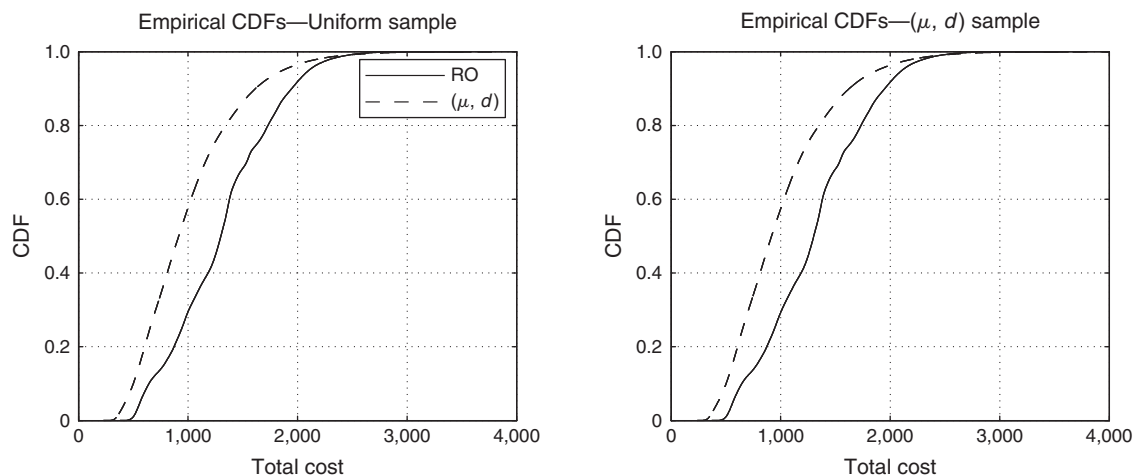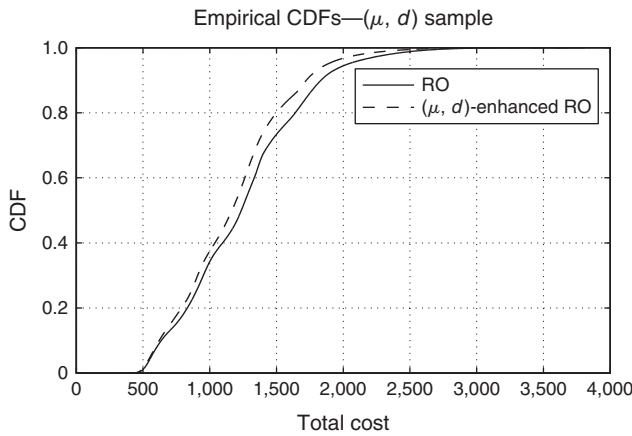

Note. Aggregated from all problem instances.

Table 4. Results of the Inventory Management—Enhancement of RO Solutions Example

		Cost			
Objective type	Enhancement type	–	(μ, d)	Sample	Nominal
Objective mean	Uniform sample	**1,230**	**1,168** (−5.04%)	1,168 (−5.04%)	1,180 (−4.06%)
Objective standard deviation	Uniform sample	157	158 (+0.63%)	156 (−0.63%)	161 (+2.54%)
Objective mean	(μ, d) sample	1,246	**1,172** (−5.93%)	1,172 (−5.93%)	1,184 (−4.97%)
Objective standard deviation	(μ, d) sample	160	161 (+0.62%)	160 (0.00%)	164 (+2.50%)

Notes. All numbers are averages. Numbers in parentheses denote the percent change compared to the initial solution with no enhancement (first column).

Figure 2. Empirical Cumulative Distribution Functions of the Total Cost in Simulation in the Uniform Demand Sample for the Nonenhanced RO Solution and the (μ, d)-Enhanced RO Solution

Note. Aggregated from all problem instances.

6.3. Inventory Experiment—Aggregation and Independent Demand

In this section we revisit our inventory experiment using the results for aggregated random vectors. The inventory experiment can be studied in this way, since the objective function is

$$\sum_{t=1}^{T}(c_t q_t(\mathbf{z}^{t-1}) + h_t \max\{x_{t+1}(\mathbf{z}^t), 0\} + p_t \max\{-x_{t+1}(\mathbf{z}^t), 0\}),$$

where we dropped the first term as we assumed $s = 0$. The objective consists of T terms

$$f_t(\mathbf{q}, \mathbf{z}) = c_t q_t(\mathbf{z}^{t-1}) + h_t \max\{x_{t+1}(\mathbf{z}^t), 0\} + p_t \max\{-x_{t+1}(\mathbf{z}^t), 0\}, \qquad (51)$$

that depend on the state of inventory x_{t+1} each. Therefore, in line with Examples 3 and 4, we can use our results for aggregated random variables to build worst-case expectations of $f_t(\mathbf{q}, \mathbf{z})$ and to use them in the optimization problem.

Since the methods of this section aim at an alleviation of the independence assumption and the piecewise linear objective function is tractable using the results of Wiesemann et al. (2014), we compare our solutions also to Wiesemann et al. (2014), whose results do not rely on the independence assumption either. We note here that to use their results to obtain an exact reformulation, we need to formulate the sum-of-maxima objective function as a maximum of linear functions, which leads to 2^T terms. For the case where the independence assumption is satisfied, we refer the reader to the electronic companion, Appendix EC.5, where the difference between their reformulations and our reformulations is illustrated.

In the following, we consider 10 solutions to the same 50 instances as in Sections 6.1 and 6.2. The first four solutions consider the ordering decisions to be static:

- S-1: using BH to evaluate the true worst-case expectation of the objective with $\mathcal{O}(3^T)$ terms in the problem formulation;
- S-2: using the aggregated random variables $y = \mathbf{1}^\top \mathbf{z}$ for which an upper bound on the MAD is computed using the independence of z_t's with (34)—$\mathcal{O}(T)$;
- S-3: using the methodology of Wiesemann et al. (2014) without the independence used—$\mathcal{O}(2^T)$; and
- S-4: using aggregated random variables for which an upper bound on the MAD is computed without using independence with (30)—$\mathcal{O}(T)$.

The other six solutions consider the ordering decisions to be linear functions of past demand:

- LDR-1: using BH to evaluate the true worst-case expectation of the objective—$\mathcal{O}(3^T)$;
- LDR-2: using the methodology of Wiesemann et al. (2014) to evaluate the exact worst-case expectation without independence—$\mathcal{O}(2^T)$;
- LDR-3: using the methodology of Wiesemann et al. (2014) and evaluating the worst-case expectation for the cost in each period separately:

$$\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \sum_{t=1}^{T} f_t(\mathbf{q}, \mathbf{z}) \leqslant \sum_{t=1}^{T} \sup_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f_t(\mathbf{q}, \mathbf{z});$$

this simplification also alleviates the complexity of the reformulation at the cost of exactness—$\mathcal{O}(T)$;

- LDR-4: using the methodology of Wiesemann et al. (2014) and using linear decision rules to bound the sum-of-maxima cost in each period,

$$f_t(\mathbf{q}, \mathbf{z}) \leqslant v_t + (\mathbf{w}^t)^\top \mathbf{z} \quad \forall \mathbf{z} \in \mathscr{Z};$$

see Gorissen and den Hertog (2013); this simplification alleviates the complexity of the reformulation at the cost of exactness—$\mathcal{O}(T)$;

- LDR-5: using methodology of Wiesemann et al. (2014) and using linear decision rules to bound the sum-of-maxima cost in each period (see Gorissen and den Hertog 2013), with the addition of making the decision rules adjustable also with respect to the auxiliary random vector $\mathbf{u}$ used in the construction of the ambiguity set (similar to the ambiguity set of the second bound of Proposition 4), as proposed in Bertsimas et al. (2016):

$$q_{t+1}(\mathbf{z}^t, \mathbf{u}^t) = q_{t+1,0} + \sum_{j=1}^{t} q_{t+1,j} z_j + \sum_{j=1}^{t} q_{t+1,t+j} u_j$$

and

$$f_t(\mathbf{q}, \mathbf{z}, \mathbf{u}) \leqslant v_t + (\mathbf{w}_z^t)^\top \mathbf{z} + (\mathbf{w}_u^t)^\top \mathbf{u} \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathscr{Z}',$$

where

$$\mathscr{Z}' = \{(\mathbf{z}, \mathbf{u}): a_t \leqslant z_t \leqslant b_t, \ u_t \geqslant z_t - \mu_t, \\ u_t \geqslant \mu_t - z_t, \ u_t \leqslant 2(b_t - \mu_t), \ \forall t\}.$$

Table 5. Results of the Inventory Management with Aggregated Random Vectors—Ranges for the Expectation of the Objective over $\mathscr{P}_{(\mu,d,\beta)}$ (with Independence Assumption in Computing the Bounds)

	Cost expectation range									
β	S-1	S-2	S-3	S-4	LDR-1	LDR-2	LDR-3	LDR-4	LDR-5	LDR-6
0.25	[1,093,1,175]	[1,124,1,205]	[1,131,1,190]	[1,131,1,190]	[1,004,1,049]	[1,058,1,093]	[1,063,1,098]	[1,154,1,192]	[1,038,1,062]	[1,058,1,092]
0.5	[1,038,1,175]	[1,065,1,205]	[1,061,1,190]	[1,061,1,190]	[970,1,049]	[1,029,1,093]	[1,033,1,098]	[1,138,1,192]	[1,005,1,062]	[1,028,1,092]
0.75	[1,079,1,175]	[1,117,1,205]	[1,106,1,190]	[1,106,1,190]	[994,1,049]	[1,047,1,093]	[1,052,1,098]	[1,156,1,192]	[1,021,1,062]	[1,047,1,092]

Notes. Results for LDR-1 are readily taken from Table 2. The upper and lower bounds of the intervals are obtained ex post using BH results, after the solutions are found. All numbers are averages.

Compared to LDR-4, the dependence on **u** helps to take the dispersion information into account, as the results will show—$\mathscr{O}(T)$; and

• LDR-6: using our approximation (37) to obtain upper bound on the worst-case expectation of each of the terms $f_t(\mathbf{q}, \mathbf{z})$, hence similar to LDR-4—$\mathscr{O}(T)$.

Table 5 presents the worst-case and best-case expectations for all 10 solutions, under the assumption of independence of z_t as in the first experiment. For the solutions with static decisions we can see that the new solutions S-2 and S-4 based on the aggregation are only slightly worse than the original solution based on BH bound with 3^T terms. For example, their worst-case expectations are 1,205 and 1,190, respectively, whereas solution S-1 yields 1,175, which makes them only 2% worse than the exact formulation.

Solutions S-3 and S-4 are identical for all instances since our aggregation technique is exact in this case, just as the method of Wiesemann et al. (2014)—this is because in each case, the mean of the uncertain demand is in the middle of the support and the proportion of the MADs of z_t's to the support intervals' widths is the same for all t; see the description of the setting in Section 6.1.

We observe that the intervals obtained by our aggregated solutions LDR-6 are very similar to the two solutions obtained using the plain methodology of Wiesemann et al. (2014); compare LDR-2 and LDR-3 versus LDR-6. This comes as a surprise since our bound (37) is just an approximation, whereas the solution LDR-2 is exact and involves 2^T terms in the problem formulation. Part of the explanation lies in the similarity of LDR-2 and LDR-3—we can see that in this problem, interchanging the supremization and summation in the objective does not change the results dramatically. For the last two solutions based on Wiesemann et al. (2014) we notice first the exceptionally bad performance of LDR-4. This is because of the linear approximation of the sum-of-maxima costs, the objective function becomes linear in the uncertain parameter and no dispersion information is used. The LDR-5 solution, based on an extra enhancement of Bertsimas et al. (2016), is able to use the dispersion information and gives the best intervals among the solutions with the size of formulation scaling as $\mathscr{O}(T)$.

6.4. Inventory Experiment—Aggregation and Dependent Demand

Much of the discussion so far has revolved around the issue of independence of the demands z_t and in the previous subsection the solutions were evaluated using demand samples in which the demands from subsequent periods were independent. One may ask, how do the solutions perform when the realized demand sample exhibits some dependence pattern?

To investigate this, we run an experiment where the demands **z** are sampled using copulas that couple a multivariate distribution function to its marginal distributions. Separating the dependence structure between random variables from their marginal distributions makes them a premier tool for simulating correlated random variables when particular marginal distributions are desired (Sklar 1959). In our case, we want the marginal distributions to come from our (μ, d) sample (results for the uniform marginals are very similar).

We use the T-dimensional Gaussian copula[3] and assume the dependence between the z_t's to follow an autocorrelative pattern where the correlations between the random variables used in the copula from periods t_1 and t_2 is equal to $\rho^{|t_1 - t_2|}$, where $\rho \in \{0.1, 0.2, \ldots, 0.9\}$, that is, we use the correlation matrix:

$$\begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 \\ \vdots & & \ddots & & & \vdots \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

For conciseness, we focus only on the LDR-1 and LDR-5 solutions.

Figure 3 presents the results, with the respective mean costs and standard deviations plotted against the correlation strength ρ. The left panel shows that as the degree of correlation among the demands increases with ρ (on the horizontal axis), the mean costs obtained by solution LDR-5 approaches the one of LDR-1 and eventually becomes smaller, with the crossing point approximately around $\rho = 0.6$. In the right panel we see that the standard deviation of the costs obtained by LDR-5 is smaller than the one of LDR-1 for all values of ρ.

Figure 3. Results of the Simulation Experiment for Inventory Solutions LDR-1 and LDR-5 with Dependent (μ, d) Demand Sample
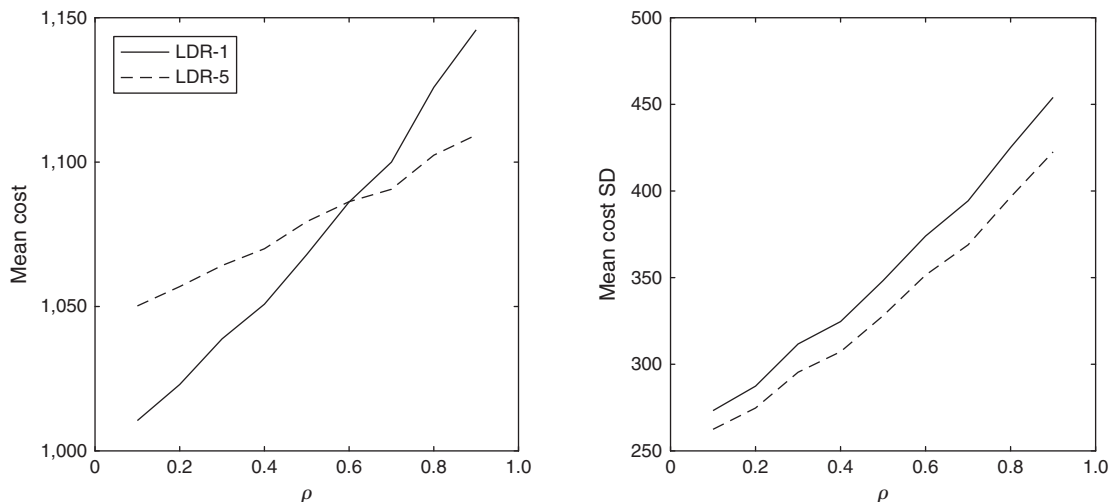
Note. The left panel involves results on the simulated mean costs and the right panel on the standard deviations.

These results provide a strong argument that the LDR-5 solution, though being constructed on the basis of approximation (37) and not assuming the independence of z_t's, might be better than the LDR-1 solution based on the full (μ, d) information in situations when the realized demand exhibits dependence among its components from different periods. More generally, this indicates that the solutions based on aggregated random vectors, while being less computationally burdensome than the "exact solutions," can yield better performance when the true random variables deviate from the independence assumption.

6.5. Chance Constraints—Power of Different Safe Tractable Approximations

We illustrate now the differences between (i) the power of the three approximations of the chance constraints, and (ii) knowing and not knowing the MAD. We consider here the following problem from Section 4.3.6.2 of Ben-Tal et al. (2009):

$$\max_{x_0} \quad x_0$$

$$\text{s.t.} \quad \sup_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{P}\left(x_0 + \sum_{i=1}^{n_z} x_i z_i > 0\right) \le \epsilon, \quad (52)$$

$$x_i = 1, \quad i = 1, \ldots, n_z.$$

We solve this problem using all three safe tractable approximations of the chance constraint, for two different cases:

- no information about d—which corresponds to setting $d_i = 1, i = 1, \ldots, n_z$ (the largest possible value for d_i; see Remark 1, page 817, about the Edmundson-Madansky bound when d is maximum possible),
- knowing that $d_i = d = 0.5, i = 1, \ldots, n_z$.

We consider three probability levels: $\epsilon \in \{10^{-1}, 10^{-2}, 10^{-3}\}$ and $n_z = 128$. Whereas safe approximations

corresponding to Theorems 1 and 2 are well defined by the theorems, we need to choose the exponential polynomial used in the approximation of Theorem 3. Similar to Ben-Tal et al. (2009), we use the polynomial

$$\gamma_{d,T}(s) = \exp(s)\chi_{c^*}(s),$$

where

$$\chi_{c^*}(s) = \sum_{v=0}^{d} (c_v^* \exp(\iota \pi v s / T) + \bar{c}_v^* \exp(-\iota \pi v s / T))$$

is an optimal solution of the best uniform approximation problem:

$$\mathbf{c}^* \in \arg\min \left\{ \max_{-T \le s \le T} |\exp(s)\chi_{\mathbf{c}}(s) - \max\{1 + s, 0\}| : \right.$$
$$\left. 0 \le \chi_{\mathbf{c}}(s) \le \chi_{\mathbf{c}}(0) = 1, \ \forall s \in \mathbb{R} \right\}$$

and $\exp(s)\chi_{\mathbf{c}}(s)$ is convex nondecreasing on $[-T, T]$, with parameter values $d = 11$ ("degree of approximation" of the function $\max\{1 + s, 0\}$), $T = 8$ ("window width" on which the function $\max\{1 + s, 0\}$ is approximated).

Table 6 presents the results. First, for all safe approximations and all security levels, one can observe a substantial value of having the information about the parameters d_i. For example, for $\epsilon = 0.01$ and the safe approximation according to Theorem 3, the optimal solution obtained without knowledge of d is -30.55, whereas the corresponding number for known $d = 0.5$ is -21.69. A similar pattern can be observed for other values of ϵ and other approximations.

Secondly, one can see the increasing power of the safe tractable approximations that use exactly the same information. For example, for $\epsilon = 10^{-3}$ and $d = 0.5$

Table 6. Maximum Values of x_0 in Problem (52), Depending on the Safe Tractable Approximation Used, Probability Bound, and the Assumptions on the Knowledge About d

		Maximum x_0		
ϵ	Safe approximation	Theorem 1	Theorem 2	Theorem 3
10^{-1}	Unknown d	−24.28	−24.21	−20.43
	$d = 0.5$	−17.16	−17.14	−14.48
10^{-2}	Unknown d	−34.34	−34.13	−30.55
	$d = 0.5$	−24.27	−24.20	−21.69
10^{-3}	Unknown d	−42.05	−41.67	−38.34
	$d = 0.5$	−29.73	−29.60	−27.25

the subsequent optimal values are −29.73, −29.60, and −27.25. For all values of ϵ and d there is a bigger difference between the second and third tractable approximation than between the first and second.

This example illustrates the extra power because of the knowledge of d, giving a strong reason to estimate this quantity in order to obtain better chance constraint approximations. Also, the difference between the quality of the safe tractable approximations of Theorems 1–3 illustrates that the exponential polynomial-based approximations are very powerful if the parameters a, b, μ, and d can be estimated with sufficient precision.

7. Summary

In this paper, we have considered two types of ambiguous stochastic constraints—expected feasibility constraints and chance constraints. In contrast to previous research, which employs the variance as a dispersion measure, we use the mean absolute deviation. This allows us to use the 1972 results of BH on tight upper and lower bounds on the expectation of a convex function of a random variable, and thus, to provide convex robust counterparts for expected feasibility constraints and to obtain safe tractable approximations of ambiguous chance constraints. Numerical examples show the proposed methodology to perform well and, in particular, to offer substantial improvements in the worst-case expected performance and the probabilistic guarantees on the constraints' feasibility. In particular, for the worst-case expected feasibility constraints, we identify an important class of functions for which we can relax the assumption of independence of random variables needed by BH, and for which we construct highly computationally tractable approximations. Numerical experiments show that these approximations yield good practical performance and can be preferred in settings where the independence assumption of the random variables does not hold.

Acknowledgments

Endnotes

[1] As a special case, one can choose only one scenario, corresponding to the nominal values of the uncertain parameters (Iancu and Trichakis 2013).

[2] One may also use other decision rules. However, we limit ourselves to the analysis of the linear case as the linear decision rules are very often a powerful enough tool; see Bertsimas et al. (2010). Moreover, the (non)convexity of the problem resulting from the application of linear decision rules is easy to verify; see Boyd and Vandenberghe (2004).

[3] We use the MATLAB function copularnd( ). As this function simulates only the CDFs of the marginal distributions, we need to convert them into the respective uniform and (μ, d) sample by inverting their distribution functions.

References

Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Oper. Res.* 64(2):474–494.

Baillie RT, Booth GG, Tse Y, Zabotina T (2002) Price discovery and common factor models. *J. Financial Markets* 5(3):309–321.

Ben-Tal A, den Hertog D (2011) Immunizing conic quadratic optimization problems against implementation errors. *Math. Programming* 149(1–2):265–299.

Ben-Tal A, Hochman E (1972) More bounds on the expectation of a convex function of a random variable. *J. Appl. Probab.* 9(4):803–812.

Ben-Tal A, Hochman E (1985) Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift Oper. Res.* 29(7):285–300.

Ben-Tal A, Nemirovski A (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.

Ben-Tal A, den Hertog D, Vial J-P (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Programming* 149(1–2):265–299.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press, Princeton, NJ).

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Math. Programming* 99(2):351–376.

Ben-Tal A, den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* 59(2):341–357.

Bertsimas D, Iancu DA, Parrilo PA (2010) Optimality of affine policies in multistage robust optimization. *Math. Oper. Res.* 35(2):363–394.

Bertsimas D, Sim M, Zhang M (2016) A practically efficient approach for solving adaptive distributionally robust linear optimization problems. Working paper, Massachusetts Institute of Technology, Cambridge. http://www.optimization-online.org/DB_FILE/2016/03/5353.pdf.

Birge JR, Wets RJ-B (1987) Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res.* 12(1):149–162.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Calafiore GC, El Ghaoui L (2006) On distributionally robust chance-constrained linear programs. *J. Optim. Theory Appl.* 130(1):1–22.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

de Ruiter F, den Hertog D, Brekelmans R (2016) The impact of the existence of multiple adjustable robust solutions. *Math. Programming* 160(1–2):531–545.

Dulá JH (1992) An upper bound on the expectation of simplicial functions of multivariate random variables. *Math. Programming* 55(1–3):69–80.

Dupačová J (1966) On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* 91(4): 423–430.

Dupačová J, Gröwe-Kuska N, Römisch W (2003) Scenario reduction in stochastic programming: An approach using probability metrics. *Math. Programming* 95(3):493–511.

Edirisinghe NCP (2011) Stochastic programming approximations using limited moment information, with application to asset allocation. *Stochastic Programming: The State of the Art in Honor of George B. Dantzig* (Springer, New York), 97–138.

Edirisinghe NCP, Ziemba WT (1992) Tight bounds for stochastic convex programs. *Oper. Res.* 40(4):660–677.

Edirisinghe NCP, Ziemba WT (1994a) Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* 19(2):314–340.

Edirisinghe NCP, Ziemba WT (1994b) Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* 19(2):292–313.

Edmundson HP (1956) Bounds on the expectation of a convex function of a random variable. *Acta Math.* 30:175–193.

El Amir EAH (2012) On uses of mean absolute deviation: Decomposition, skewness and correlation coefficients. *Metron* 70(2–3): 145–164.

Esfahani PM, Kuhn D (2017) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming*, ePub ahead of print July 7, https://doi.org/10.1007/s10107-017-1172-1.

Frauendorfer K (1988) SLP recourse problems with arbitrary multivariate distributions—The dependent case. *Math. Oper. Res.* 13(3):377–394.

Gallego G (1992) A minmax distribution free procedure for the (q, r) inventory model. *Oper. Res. Lett.* 11(1):55–60.

Gallego G, Ryan JK, Simchi-Levi D (2001) Minimax analysis for finite-horizon inventory models. *IIE Trans.* 33(10):861–874.

Gassmann HI, Ziemba WT (1986) A tight upper bound for the expectation of a convex function of a multivariate random variable. Prékopa A, Wets RJ-B, eds. *Stochastic Programming, 84 Part 1*, Mathematical Programming Studies, Vol. 27 (Springer, Berlin), 39–53.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Oper. Res.* 58(4):902–917.

Gorard S (2005) Revisiting a 90-year-old debate: The advantages of the mean deviation. *British J. Educational Stud.* 53(4):417–430.

Gorissen BL, den Hertog D (2013) Robust counterparts of inequalities containing sums of maxima of linear functions. *Eur. J. Oper. Res.* 227(1):30–43.

Hanasusanto GA, Roitch V, Kuhn D, Wiesemann W (2015) A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Programming* 151(1):35–62.

Higle JL, Sen S (1996) *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming* (Kluwer, Dordrecht, Netherlands).

Huber PJ, Ronchetti EM (2009) *Robust Statistics* (Springer, Berlin).

Iancu DA, Trichakis N (2013) Pareto efficiency in robust optimization. *Management Sci.* 60(1):130–147.

Isii K (1962) On sharpness of Tchebycheff-type inequalities. *Ann. Institute Statist. Math.* 14(1):185–197.

Jensen JLWV (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* 30(1):175–193.

Jiang R, Guan Y (2016) Data-driven chance constrained stochastic program. *Mathematical Programming* 158(1–2):291–327.

Jolliffe I (2002) *Principal Component Analysis* (Springer, New York).

Kall P (1991) An upper bound for SLP using first and total second moments. *Ann. Oper. Res.* 30(1):267–276.

Karr AF (1983) Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.* 8(1):74–85.

Madansky A (1959) Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Statist.* 30(3): 743–746.

Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. *SIAM J. Optim.* 17(4):969–996.

Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1):188–203.

Popescu I (2007) Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55(1):98–112.

Postek K, den Hertog D, Melenberg B (2016) Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Rev.* 58(4):603–650.

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2(3):21–42.

Rogosinsky WW (1958) Moments of non-negative mass. *Proc. Roy. Soc. London, Ser. A* 245(1240):1–27.

Scarf HE (1958) A min-max solution of an inventory problem. Arrow KJ, Karlin S, Scarf HE, eds. *Studies in the Mathematical Theory of Inventory and Production* (Stanford University Press, Stanford, CA), 201–209.

Shapiro A, Kleywegt AJ (2002) Minimax analysis of stochastic problems. *Optim. Methods Software* 17(3):523–542.

Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM, Philadelphia).

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8:229–231.

Smith RL (1984) Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.* 32(6):1296–1308.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Oper. Res.* 62(6):1358–1376.

Xin L, Goldberg DA, Shapiro A (2015) Time (in)consistency of multistage distributionally robust inventory models with moment constraints. Preprint arXiv:1304.3074.

Zymler S, Kuhn D, Rustem B (2013) Distributionally robust joint chance constraints with second-order moment information. *Math. Programming* 137(1–2):167–198.

Krzysztof Postek is an assistant professor of operations research at the Department of Econometrics of the Erasmus University Rotterdam. His main research interests are the applications and algorithms for solving (distributionally) robust optimization problems.

Aharon Ben-Tal is an emeritus professor of operations research at Technion and a professor at Shenkar College of Design and Engineering. His interests include continuous optimization, particularly nonsmooth and large-scale problems, conic and robust optimization, and convex and nonsmooth analysis. Recently the focus of his research has been on using robust optimization to address problems affected by uncertainty. He was awarded the 2007 EURO Gold Medal and the 2016 INFORMS Khachiyan Prize.

Dick den Hertog is a professor of operations research at Tilburg University. His research interests cover various fields in linear and nonlinear optimization, e.g., robust optimization and simulation-based optimization. He is also active in applying operations research theory in real-life applications. In 2000 he received the EURO Best Applied Paper Award, together with Peter Stehouwer (CQM). In 2013 he was a member of the team that received the INFORMS Franz Edelman Award.

Bertrand Melenberg is a professor of econometrics and finance at Tilburg University. His research area is econometrics. His research interests cover various topics such as longevity risk, empirical finance (asset pricing), robust optimization, and environmental econometrics.