OPERATIONS RESEARCH
Vol. 55, No. 6, November–December 2007, pp. 1058–1071
ISSN 0030-364X | EISSN 1526-5463 | 07 | 5506 | 1058

DOI 10.1287/opre.1070.0441

A Robust Optimization Perspective on Stochastic Programming

Xin Chen
Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, xinchen@uiuc.edu

Melvyn Sim
NUS Business School, National University of Singapore and Singapore MIT Alliance (SMA), Singapore, dscsimm@nus.edu.sg

Peng Sun
Fuqua School of Business, Duke University, Durham, North Carolina 27708, psun@duke.edu

In this paper, we introduce an approach for constructing uncertainty sets for robust optimization using new deviation measures for random variables termed the *forward and backward deviations*. These deviation measures capture distributional asymmetry and lead to better approximations of chance constraints. Using a linear decision rule, we also propose a tractable approximation approach for solving a class of multistage chance-constrained stochastic linear optimization problems. An attractive feature of the framework is that we convert the original model into a second-order cone program, which is computationally tractable both in theory and in practice. We demonstrate the framework through an application of a project management problem with uncertain activity completion time.

Subject classifications: programming: stochastic.
Area of review: Optimization.
History: Received December 2004; revisions received June 2005, May 2006; accepted December 2006.

1. Introduction

In recent years, robust optimization has gained substantial popularity as a modeling framework for immunizing against parametric uncertainties in mathematical optimization. The first step in this direction was taken by Soyster (1973), who proposed a worst-case model for linear optimization such that constraints are satisfied under all possible perturbations of the model parameters. Recent developments in robust optimization focused on more elaborate uncertainty sets to address the issue of over-conservatism in worst-case models, as well as to maintain computational tractability of the proposed approaches (see, for example, Ben-Tal and Nemirovski 1998, 1999, 2000; El-Ghaoui et al. 1997, 1998; Goldfarb and Iyengar 2003; Bertsimas and Sim 2003, 2004a, 2004b, 2006; and Atamtürk 2006). Assuming very limited information of the underlying uncertainties, such as mean and support, the robust model can provide a solution that is feasible to the constraints with high probability, although avoiding the extreme conservatism of Soyster's worst-case model. Computational tractability of robust linear constraints is achieved by considering tractable uncertainty sets such as ellipsoids (see Ben-Tal and Nemirovski 2000) and polytopes (see Bertsimas and Sim 2004a), which yield *robust counterparts* that are second-order conic constraints and linear constraints, respectively. The methodology of robust optimization has also been applied to dynamic settings involving multiperiod optimization, in which future decisions (recourse variables) depend on the realization of the present uncertainty. Such models are generally intractable. Ben-Tal et al. (2004) proposed a tractable approach for solving fixed recourse instances using affine decision rules—recourse variables as affine functions of the uncertainty realization. Some applications of robust optimization in a dynamic environment include inventory management (Bertsimas and Thiele 2006, Ben-Tal et al. 2004) and supply contracts (Ben-Tal et al. 2005). Two important characteristics of robust linear optimization that make it practically appealing are:

(1) Robust linear optimization models are polynomial in size and in the form of linear programming (LP) or second-order cone programming (SOCP). One therefore can leverage on the state-of-the-art LP and SOCP solvers, which are becoming increasingly powerful, efficient, and robust. For instance, CPLEX 9.1 offers SOCP modeling with integrality constraints.

(2) Robust optimization requires only modest assumptions about distributions, such as a known mean and bounded support. This relieves users from having to know the probabilistic distributions of the underlying stochastic parameters, which are often unavailable.

In linear optimization, Bertsimas and Sim (2004) and Ben-Tal and Nemirovski (2000) obtain probability bounds against constraint violation by assuming independent and symmetrically bounded coefficients, while using the support information (rather than the variance or standard deviation) to derive the probability of constraint violation. The assumption of distributional symmetry, however, is limiting in many applications, such as financial modeling, in which distributions often are known to be asymmetric. In cases where the variances of the random variables are small, whereas the support of the distributions is wide, the robust solutions obtained via the above approach can also be rather conservative. The idea of guaranteeing constraint feasibility with a certain probability is closely related to the chance-constrained programming literature (Charnes and Cooper 1959, 1963). Finding exact solutions to chance-constrained problems is typically intractable. Pintér (1989) proposes various deterministic approximations of chance-constrained problems via probability inequalities such as Chebyshev's inequality, Bernstein's inequality, Hoefding's inequality, and their extensions (see also Birge and Louveaux 1997, Chapter 9.4; and Kibzun and Kan 1996). The deterministic approximations are expressed in terms of the mean, standard deviation, and/or range of the uncertainties. The resulting models are generally convex minimization problems. In this paper, we propose an approach to robust optimization that addresses asymmetric distributions. At the same time, the proposed approach may be used as a deterministic approximation of chance-constrained problems. Our goal in this paper is therefore twofold.

(1) First, we refine the framework for robust linear optimization by introducing a new uncertainty set that captures the asymmetry of the underlying random variables. For this purpose, we introduce new deviation measures associated with a random variable—namely, the forward and backward deviations—and apply them to the design of uncertainty sets. Our robust linear optimization framework generalizes previous works of Bertsimas and Sim (2004) and Ben-Tal and Nemirovski (2000).

(2) Second, we propose a tractable solution approach for a class of stochastic linear optimization problems with chance constraints. By applying the forward and backward deviations of the underlying distributions, our method provides feasible solutions for stochastic linear optimization problems. The optimal solution from our model provides an upper bound to the minimum objective value for all underlying distributions that satisfy the parameters of the deviations. One way in which our framework improves upon existing deterministic equivalent approximations of chance constraints is that we turn the model into an SOCP, which is advantageous in computation. Another attractive feature of our approach is its computational scalability for multiperiod problems. The literature on multiperiod stochastic programs with chance constraints is rather limited, which could be due to the lack of tractable methodologies.

In §2, we introduce a new uncertainty set and formulate the robust counterpart. In §3, we present new deviation measures that capture distributional asymmetry. Section 4 shows how one can integrate the new uncertainty set with the new deviation measures to obtain solutions to chance-constrained problems. In §5, we present an SOCP approximation for stochastic programming with chance constraints. Section 6 contains a summary and conclusions. In §A of the online appendix, we apply our framework to a project management problem with uncertain completion time. An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Notations. We denote a random variable, $\tilde{x}$, with the tilde sign. Boldface lowercase letters, such as $\mathbf{x}$, represent vectors, and the corresponding uppercase letters, such as $\mathbf{A}$, denote matrices.

2. Robust Formulation of a Stochastic Linear Constraint

Consider a stochastic linear constraint,

$$\tilde{\mathbf{a}}'\mathbf{x} \leqslant \tilde{b}, \tag{1}$$

where the input parameters $(\tilde{\mathbf{a}}, \tilde{b})$ are random. We assume that the uncertain data, $\widetilde{\mathbf{D}} = (\tilde{\mathbf{a}}, \tilde{b})$, has the following underlying perturbations.

Affine Data Perturbation. We represent uncertainties of the data $\widetilde{\mathbf{D}}$ as affinely dependent on a set of independent random variables, $\{\tilde{z}_j\}_{j=1:N}$, as follows:

$$\widetilde{\mathbf{D}} = \mathbf{D}^0 + \sum_{j=1}^{N} \Delta\mathbf{D}^j \tilde{z}_j,$$

where $\mathbf{D}^0$ is the nominal value of the data, and $\Delta\mathbf{D}^j$, $j \in N$, is a direction of data perturbation. We call $\tilde{z}_j$ the primitive uncertainty, which has mean zero and support in $[-\underline{z}_j, \bar{z}_j]$, $\underline{z}_j, \bar{z}_j > 0$. If N is small, we model situations involving a small collection of primitive independent uncertainties, which implies that the elements of $\widetilde{\mathbf{D}}$ are strongly dependent. If N is large, we model the case that the elements of $\widetilde{\mathbf{D}}$ are weakly dependent. In the limiting case when the number of entries in the data equals N, the elements of $\widetilde{\mathbf{D}}$ are independent. We desire a set of solutions $X(\epsilon)$ such that $\mathbf{x} \in X(\epsilon)$ is feasible for the linear constraint (1) with probability of at least $1-\epsilon$. Formally, we can describe the set $X(\epsilon)$ using the following chance-constraint representation (see Charnes and Cooper 1959):

$$X(\epsilon) = \{\mathbf{x}\colon \mathrm{P}(\tilde{\mathbf{a}}'\mathbf{x} \leqslant \tilde{b}) \geqslant 1-\epsilon\}. \tag{2}$$

The parameter ϵ in the set $X(\epsilon)$ varies the conservatism of the solution. Unfortunately, however, when $\epsilon > 0$, the set $X(\epsilon)$ is often nonconvex and computationally intractable (see Birge and Louveaux 1997). Furthermore, the evaluation of probability requires complete knowledge of data distributions, which is often an unrealistic assumption. In view of these difficulties, robust optimization offers a different approach to handling data uncertainty. Specifically, in addressing the uncertain linear constraint of (1), we represent the set of robust feasible solutions

$$X_r(\Omega) = \{\mathbf{x}\colon \mathbf{a}'\mathbf{x} \leqslant b \ \forall (\mathbf{a}, b) \in \mathcal{U}_\Omega\}, \tag{3}$$

where the uncertain set, $\mathcal{U}_\Omega$, is compact. The parameter Ω, which we refer to as the budget of uncertainty, varies the size of the uncertainty set radially from the central point, $\mathcal{U}_{\Omega=0} = (\mathbf{a}^0, b^0)$, such that $\mathcal{U}_\Omega \subseteq \mathcal{U}_{\Omega'} \subseteq \mathcal{W}$ for all $\Omega_{\max} \geqslant \Omega' \geqslant \Omega \geqslant 0$. Here the worst-case uncertainty set $\mathcal{W}$ is the convex support of the uncertain data, defined as follows:

$$\mathcal{W} = \left\{(\mathbf{a}, b)\colon \exists \mathbf{z} \in \Re^N, (\mathbf{a}, b) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\Delta\mathbf{a}^j, \Delta b^j) z_j, \ -\underline{\mathbf{z}} \leqslant \mathbf{z} \leqslant \bar{\mathbf{z}}\right\}, \tag{4}$$

which is the smallest closed convex set satisfying $\mathrm{P}((\tilde{\mathbf{a}}, \tilde{b}) \in \mathcal{W}) = 1$. Value $\Omega_{\max}$ is the worst-case budget of uncertainty, i.e., the minimum parameter Ω such that $\mathcal{U}_\Omega = \mathcal{W}$. Therefore, under affine data perturbation, the worst-case uncertainty set is a parallelotope for which the feasible solution is characterized by Soyster (1973), i.e., a very conservative approximation to $X(\epsilon)$. To derive a less conservative approximation, we need to choose the budget of uncertainty, Ω, appropriately. In designing such an uncertainty set, we want to preserve both the theoretical and practical computational tractability of the nominal problem. Furthermore, we want to guarantee the probability such that the robust solution is feasible without being overly conservative. In other words, for a reasonable choice of ϵ, such as 0.001, there exists a parameter Ω such that $X_r(\Omega) \subseteq X(\epsilon)$. Furthermore, the budget of uncertainty Ω should be substantially smaller than the worst-case budget $\Omega_{\max}$, such that the solution is potentially less conservative than the worst-case solution. For symmetric and bounded distributions, we can assume without loss of generality that the primitive uncertainties $\tilde{z}_j$ are distributed in $[-1, 1]$, that is, $\underline{\mathbf{z}} = \bar{\mathbf{z}} = \mathbf{1}$. The natural uncertainty set to consider is the intersection of a norm uncertainty set, $\mathcal{V}_\Omega$, and the worst-case support set, $\mathcal{W}$, as follows:

$$\mathcal{S}_\Omega = \underbrace{\left\{(\mathbf{a}, b)\colon \exists \mathbf{z} \in \Re^N, (\mathbf{a}, b) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\Delta\mathbf{a}^j, \Delta b^j) z_j, \|\mathbf{z}\| \leqslant \Omega\right\}}_{=\mathcal{V}_\Omega} \cap \mathcal{W}$$
$$= \left\{(\mathbf{a}, b)\colon \exists \mathbf{z} \in \Re^N, (\mathbf{a}, b) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\Delta\mathbf{a}^j, \Delta b^j) z_j, \ \|\mathbf{z}\| \leqslant \Omega, \ \|\mathbf{z}\|_\infty \leqslant 1\right\}. \tag{5}$$

As the budget of uncertainty Ω increases, the norm uncertainty set $\mathcal{V}_\Omega$ expands radially from the point $(\mathbf{a}^0, b^0)$ until it engulfs the set $\mathcal{W}$, at which point the uncertainty set $\mathcal{S}_\Omega = \mathcal{W}$. Hence, for any choice of Ω, the uncertainty set $\mathcal{S}_\Omega$ is always less conservative than the worst-case uncertainty set $\mathcal{W}$. Various choices of norms $\|\cdot\|$ are considered in robust optimization. Under the l_2 or ellipsoidal norm proposed by Ben-Tal and Nemirovski (2000), the feasible solutions to the robust counterpart of (3), in which $\mathcal{U}_\Omega = \mathcal{S}_\Omega$, is guaranteed to be feasible for the linear constraint with probability of at least $1 - \exp(-\Omega^2/2)$. The robust counterpart is a formulation with second-order cone constraints. Bertsimas and Sim (2004) consider an $l_1 \cap l_\infty$ norm of the form $\|\mathbf{z}\|_{l_1 \cap l_\infty} = \max\{(1/\sqrt{N})\|\mathbf{z}\|_1, \|\mathbf{z}\|_\infty\}$, and show that the feasibility guarantee is also $1 - \exp(-\Omega^2/2)$. The resultant robust counterpart under consideration remains a linear optimization problem of about the same size, which is practically suited for optimization over integers. However, in the worst case, this approach can be more conservative than the use of ellipsoidal norm. In both approaches, the value of Ω is relatively small. For example, for a feasibility guarantee of 99.9%, we only need to choose $\Omega = 3.72$. We note that by comparison with the worst-case uncertainty set, $\mathcal{W}$, for Ω greater than $\sqrt{N}$, the constraints $\|\mathbf{z}\|_2 \leqslant \Omega$ and $\max\{(1/\sqrt{N})\|\mathbf{z}\|_1, \|\mathbf{z}\|_\infty\} \leqslant \Omega$ are the consequence of $\mathbf{z}$ satisfying $\|\mathbf{z}\|_\infty \leqslant 1$. Hence, it is apparent that for both approaches, the budget of uncertainty Ω is substantially smaller than the worst-case budget, in which $\Omega_{\max} = \sqrt{N}$. In this paper, we restrict the vector norm $\|.\|$ to be considered in an uncertainty set as follows:

$$\|\mathbf{u}\| = \||\mathbf{u}|\|, \tag{6}$$

where $|\mathbf{u}|$ is the vector with the j component equal to $|u_j| \ \forall j \in \{1, \ldots, N\}$ and

$$\|\mathbf{u}\| \leqslant \|\mathbf{u}\|_2 \quad \forall \mathbf{u}. \tag{7}$$

We call this an absolute norm. It is easy to see that the ellipsoidal norm and the $l_1 \cap l_\infty$ norm mentioned above satisfy these properties. The dual norm $\|\cdot\|^*$ is defined as

$$\|\mathbf{u}\|^* = \max_{\|\mathbf{x}\| \leqslant 1} \mathbf{u}'\mathbf{x}.$$

We next show some basic properties of absolute norms that we subsequently will use in our development.

PROPOSITION 1. *If the norm* $\|\cdot\|$ *satisfies Equation* (6) *and Equation* (7), *then we have*

(a) $\|\mathbf{w}\|^* = \||\mathbf{w}|\|^*$.
(b) *For all* $\mathbf{v}, \mathbf{w}$ *such that* $|\mathbf{v}| \leqslant |\mathbf{w}|$, $\|\mathbf{v}\|^* \leqslant \|\mathbf{w}\|^*$.
(c) *For all* $\mathbf{v}, \mathbf{w}$ *such that* $|\mathbf{v}| \leqslant |\mathbf{w}|$, $\|\mathbf{v}\| \leqslant \|\mathbf{w}\|$.
(d) $\|\mathbf{t}\|^* \geqslant \|\mathbf{t}\|_2 \ \forall \mathbf{t}$.

PROOF. The proofs of (a), (b), and (c) are shown in Bertsimas and Sim (2006).

(d) It is well known that the dual norm of the Euclidian norm is also the Euclidian norm, i.e., it is self-dual. For all $\mathbf{t}$, observe that

$$\|\mathbf{t}\|^* = \max_{\|\mathbf{z}\| \leqslant 1} \mathbf{t}'\mathbf{z} \geqslant \max_{\|\mathbf{z}\|_2 \leqslant 1} \mathbf{t}'\mathbf{z} = \|\mathbf{t}\|_2^* = \|\mathbf{t}\|_2. \quad \square$$

To build a generalization of the uncertainty set that takes into account the primitive uncertainties being asymmetrically distributed, we first ignore the worst-case support set, $\mathscr{W}$, and define the asymmetric norm uncertainty set as follows:

$$\mathscr{A}_\Omega = \Bigg\{(\mathbf{a}, b)\colon \exists \mathbf{v}, \mathbf{w} \in \Re^N, \quad (\mathbf{a}, b) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\mathbf{\Delta a}^j, \Delta b^j)(v_j - w_j), \quad \|\mathbf{P}^{-1}\mathbf{v} + \mathbf{Q}^{-1}\mathbf{w}\| \leqslant \Omega, \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}\Bigg\}, \tag{8}$$

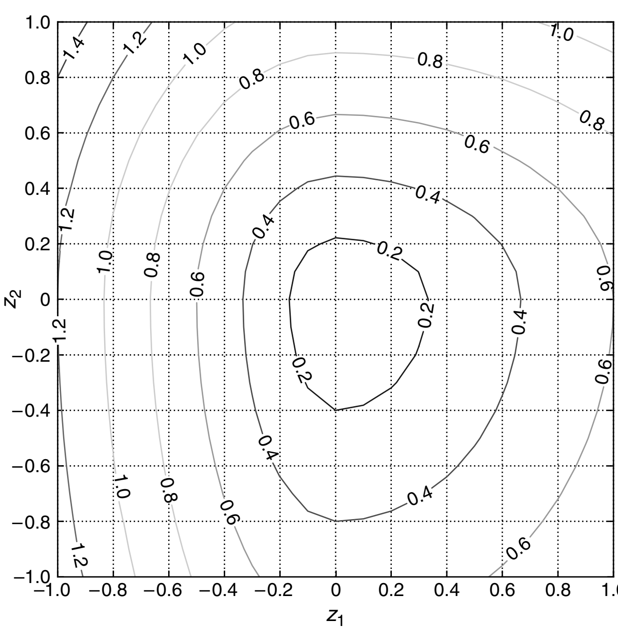
where $\mathbf{P} = \text{diag}(p_1, \ldots, p_N)$ and likewise $\mathbf{Q} = \text{diag}(q_1, \ldots, q_N)$ with $p_j, q_j > 0$, $j \in \{1, \ldots, N\}$. Figure 1 shows a sample shape of the asymmetric uncertainty set.

In the next section, we clarify how $\mathbf{P}$ and $\mathbf{Q}$ relate to the forward and backward deviations of the underlying primitive uncertainties. The following proposition shows the connection of the set $\mathscr{A}_\Omega$ with the uncertainty set described by norm $\mathscr{V}_\Omega$ defined in (5).

Proposition 2. *When $p_j = q_j = 1$ for all $j \in \{1, \ldots, N\}$, the uncertainty sets $\mathscr{A}_\Omega$ and $\mathscr{V}_\Omega$ are equivalent.*

The proof is shown in §B of the online appendix. To capture distributional asymmetries, we decompose the primitive data uncertainty $\tilde{z}$ into two random variables, $\tilde{v} = \max\{\tilde{z}, 0\}$ and $\tilde{w} = \max\{-\tilde{z}, 0\}$, such that $\tilde{z} = \tilde{v} - \tilde{w}$. The multipliers $1/p_j$ and $1/q_j$ normalize the effective perturbation contributed by both $\tilde{v}$ and $\tilde{w}$, such that the norm of the aggregated values falls within the budget of uncertainty.

Figure 1. An uncertainty set represented by $\mathscr{A}_\Omega$ as Ω varies for $N = 2$.



Because $p_j, q_j > 0$ for $\Omega > 0$, the point $(\mathbf{a}^0, b^0)$ lies in the interior of the uncertainty set $\mathscr{A}_\Omega$. Hence, we can easily evoke strong duality to obtain a computationally attractive equivalent formulation of the robust counterpart of (3), such as in linear or second-order cone optimization problems. To facilitate our exposition, we need the following proposition.

Proposition 3. *Let*

$$\begin{aligned} z^* = \max \quad & \mathbf{a}'\mathbf{v} + \mathbf{b}'\mathbf{w} \\ \text{s.t.} \quad & \|\mathbf{v} + \mathbf{w}\| \leqslant \Omega, \\ & \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}. \end{aligned} \tag{9}$$

Then, $\Omega\|\mathbf{t}\|^* = z^*$, where $t_j = \max\{a_j, b_j, 0\}$, $j \in \{1, \ldots, N\}$.

We present the proof in §C of the online appendix.

Theorem 1. *The robust counterpart of* (3) *in which $\mathscr{U}_\Omega = \mathscr{A}_\Omega$ is equivalent to*

$$\left\{ \mathbf{x}\colon \begin{array}{l} \exists \mathbf{u} \in \Re^N, h \in \Re, \\ \mathbf{a}^{0\prime}\mathbf{x} + \Omega h \leqslant b^0, \\ \|\mathbf{u}\|^* \leqslant h, \\ u_j \geqslant p_j(\mathbf{\Delta a}^{j\prime}\mathbf{x} - \Delta b^j) \quad \forall j \in \{1, \ldots, N\}, \\ u_j \geqslant -q_j(\mathbf{\Delta a}^{j\prime}\mathbf{x} - \Delta b^j) \quad \forall j \in \{1, \ldots, N\}. \end{array} \right\} \tag{10}$$

Proof. We first express the robust counterpart of (3) in which $\mathscr{U}_\Omega = \mathscr{A}_\Omega$ as follows:

$$\mathbf{a}^{0\prime}\mathbf{x} + \sum_{j=1}^N \underbrace{(\mathbf{\Delta a}^{j\prime}\mathbf{x} - \Delta b^j)}_{=y_j}(v_j - w_j) \leqslant b^0 \quad \forall \mathbf{v}, \mathbf{w} \in \Re^N, \|\mathbf{P}^{-1}\mathbf{v} + \mathbf{Q}^{-1}\mathbf{w}\| \leqslant \Omega, \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}$$

$$\Updownarrow$$

$$\mathbf{a}^{0\prime}\mathbf{x} + \max_{\substack{\{\mathbf{v}, \mathbf{w}\colon \|\mathbf{P}^{-1}\mathbf{v} + \mathbf{Q}^{-1}\mathbf{w}\| \leqslant \Omega \\ \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}\}}} (\mathbf{v} - \mathbf{w})'\mathbf{y} \leqslant b^0.$$

Observe that

$$\max_{\substack{\{\mathbf{v}, \mathbf{w}\colon \|\mathbf{P}^{-1}\mathbf{v} + \mathbf{Q}^{-1}\mathbf{w}\| \leqslant \Omega \\ \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}\}}} (\mathbf{v} - \mathbf{w})'\mathbf{y} = \max_{\substack{\{\mathbf{v}, \mathbf{w}\colon \|\mathbf{v} + \mathbf{w}\| \leqslant \Omega \\ \mathbf{v}, \mathbf{w} \geqslant \mathbf{0}\}}} (\mathbf{P}\mathbf{y})'\mathbf{v} - (\mathbf{Q}\mathbf{y})'\mathbf{w} = \Omega\|\mathbf{t}\|^*, \tag{11}$$

where $t_j = \max\{p_j y_j, -q_j y_j, 0\} = \max\{p_j y_j, -q_j y_j\}$ because $p_j, q_j > 0$ for all $j \in \{1, \ldots, N\}$. Furthermore, the equality (11) follows from the direct transformation of vectors $\mathbf{v}, \mathbf{w}$ to $\mathbf{Pv}, \mathbf{Qw}$, respectively. The last equality follows directly from Proposition 3. Hence, the equivalent formulation of the robust counterpart is

$$\mathbf{a}^{0\prime}\mathbf{x} + \Omega\|\mathbf{t}\|^* \leqslant b^0. \tag{12}$$

Finally, suppose that $\mathbf{x}$ is feasible for the robust counterpart of (3), in which $\mathscr{U}_\Omega = \mathscr{A}_\Omega$. From Equation (12), if we let $\mathbf{u}=\mathbf{t}$ and $h=\|\mathbf{t}\|^*$, constraint (10) is also feasible. Conversely, if $\mathbf{x}$ is feasible in (10), then $\mathbf{u}\geqslant\mathbf{t}$. Following Proposition 1(b), we have

$$\mathbf{a}^{0\prime}\mathbf{x}+\Omega\|\mathbf{t}\|^*\leqslant\mathbf{a}^{0\prime}\mathbf{x}+\Omega\|\mathbf{u}\|^*\leqslant\mathbf{a}^{0\prime}\mathbf{x}+\Omega h\leqslant b^0. \quad \square$$

The complete formulation and complexity class of the robust counterpart depends on the representation of the dual norm constraint, $\|\mathbf{u}\|^*\leqslant y$. In §D of the online appendix, we tabulate the common choices of absolute norms, the representation of their dual norms, and the corresponding references. The $l_1\cap l_\infty$ norm is an attractive choice if one wishes the model to remain linear and modest in size.

Incorporating the Worst-Case Support Set $\mathscr{W}$

We now incorporate the worst-case support set $\mathscr{W}$ as follows:

$$\mathscr{G}_\Omega=\mathscr{A}_\Omega\cap\mathscr{W}.$$

Because we can represent the support set of $\mathscr{W}$ equivalently as

$$\mathscr{W}=\left\{(\mathbf{a},b)\colon \exists\mathbf{v},\mathbf{w}\in\Re^N,\right.$$
$$(\mathbf{a},b)=(\mathbf{a}^0,b^0)+\sum_{j=1}^N(\boldsymbol{\Delta}\mathbf{a}^j,\Delta b^j)(v_j-w_j),$$
$$\left.-\underline{\mathbf{z}}\leqslant\mathbf{v}-\mathbf{w}\leqslant\bar{\mathbf{z}},\mathbf{w},\mathbf{v}\geqslant\mathbf{0}\right\}, \tag{13}$$

it follows that

$$\mathscr{G}_\Omega=\left\{(\mathbf{a},b)\colon \exists\mathbf{v},\mathbf{w}\in\Re^N,(\mathbf{a},b)=(\mathbf{a}^0,b^0)\right.$$
$$+\sum_{j=1}^N(\boldsymbol{\Delta}\mathbf{a}^j,\Delta b^j)(v_j-w_j),\ \|\mathbf{P}^{-1}\mathbf{v}+\mathbf{Q}^{-1}\mathbf{w}\|\leqslant\Omega,$$
$$\left.-\underline{\mathbf{z}}\leqslant\mathbf{v}-\mathbf{w}\leqslant\bar{\mathbf{z}},\mathbf{w},\mathbf{v}\geqslant\mathbf{0}\right\}. \tag{14}$$

We show in the §D of the online appendix an equivalent formulation of the corresponding robust counterpart under the generalized uncertainty set $\mathscr{G}_\Omega$.

Theorem 2. *The robust counterpart of* (3) *in which* $\mathscr{U}_\Omega=\mathscr{G}_\Omega$ *is equivalent to*

$$\left\{\mathbf{x}\colon\begin{array}{l}\exists\mathbf{u},\mathbf{r},\mathbf{s}\in\Re^N,h\in\Re,\\ \mathbf{a}^{0\prime}\mathbf{x}+\Omega h+\mathbf{r}'\bar{\mathbf{z}}+\mathbf{s}'\underline{\mathbf{z}}\leqslant b^0,\\ \|\mathbf{u}\|^*\leqslant h,\\ u_j\geqslant p_j(\boldsymbol{\Delta}\mathbf{a}^{j\prime}\mathbf{x}-\Delta b^j-r_j+s_j)\\ \quad\forall j=\{1,\ldots,N\},\\ u_j\geqslant -q_j(\boldsymbol{\Delta}\mathbf{a}^{j\prime}\mathbf{x}-\Delta b^j-r_j+s_j)\\ \quad\forall j=\{1,\ldots,N\},\\ \mathbf{u},\mathbf{r},\mathbf{s}\geqslant\mathbf{0}.\end{array}\right\} \tag{15}$$

3. Forward and Backward Deviation Measures

When random variables are incorporated in optimization models, operations are often cumbersome and computationally intractable. Moreover, in many practical problems, we often do not know the precise distributions of uncertainties. Hence, one may not be able to justify solutions based on assumed distributions. Instead of using complete distributional information, our aim is to identify and exploit some salient characteristics of the uncertainties in robust optimization models, so as to obtain nontrivial probability bounds against constraint violation. We commonly measure the variability of a random variable using the variance or the second moment, which does not capture distributional asymmetry. In this section, we introduce new deviation measures for bounded random variables that do capture distributional asymmetries. Moreover, the deviation measures applied in our proposed robust methodology guarantee the desired low probability of constraint violation. We also provide a method that calculates the deviation measures based on potentially limited knowledge of the distribution. Specifically, if one knows only the support and the mean, one can still construct the forward and backward deviation measures, albeit more conservatively. In the following, we present a specific pair of deviation measures that exist for bounded random variables. There is a more general framework of deviation measures that is useful for broader settings. We present the more general framework in §F of the online appendix.

3.1. Definitions and Properties of Forward and Backward Deviations

Let $\tilde{z}$ be a random variable and $M_{\tilde{z}}(s)=\mathrm{E}(\exp(s\tilde{z}))$ be its moment-generating function. We define the set of values associated with forward deviations of $\tilde{z}$ as follows:

$$\mathscr{P}(\tilde{z})=\left\{\alpha\colon\alpha\geqslant 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(\frac{\phi}{\alpha}\right)\leqslant\exp\left(\frac{\phi^2}{2}\right)\ \forall\phi\geqslant 0\right\}. \tag{16}$$

Likewise, for backward deviations, we define the following set:

$$\mathscr{Q}(\tilde{z})=\left\{\alpha\colon\alpha\geqslant 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(-\frac{\phi}{\alpha}\right)\leqslant\exp\left(\frac{\phi^2}{2}\right)\ \forall\phi\geqslant 0\right\}. \tag{17}$$

For completeness, we also define $\mathscr{P}(c)=\mathscr{Q}(c)=\Re_+$ for any constant c. Observe that $\mathscr{P}(\tilde{z})=\mathscr{Q}(\tilde{z})$ if $\tilde{z}$ is symmetrically distributed around its mean. For known distributions, we define the forward deviation of $\tilde{z}$ as $p^*_{\tilde{z}}=\inf\mathscr{P}(\tilde{z})$ and the backward deviation as $q^*_{\tilde{z}}=\inf\mathscr{Q}(\tilde{z})$. We note that the deviation measures defined above exist for some distributions with unbounded support, such as the normal distribution. However, some other distributions do not have finite deviation measures according to the above definition, e.g., the exponential and the gamma distributions. The following result summarizes the key properties of the deviation measures after we perform linear operations on independent random variables.

Theorem 3. *Let $\tilde{x}$ and $\tilde{y}$ be two independent random variables with zero means, such that $p_{\tilde{x}} \in \mathcal{P}(\tilde{x})$, $q_{\tilde{x}} \in \mathcal{Q}(\tilde{x})$, $p_{\tilde{y}} \in \mathcal{P}(\tilde{y})$, and $q_{\tilde{y}} \in \mathcal{Q}(\tilde{y})$.*

(a) *If $\tilde{z} = a\tilde{x}$, then*

$$(p_{\tilde{z}}, q_{\tilde{z}}) = \begin{cases} (ap_{\tilde{x}}, aq_{\tilde{x}}) & \text{if } a \geqslant 0, \\ (-aq_{\tilde{x}}, -ap_{\tilde{x}}) & \text{otherwise}, \end{cases}$$

satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$, respectively. In other words, $p_{\tilde{z}} = \max\{ap_{\tilde{x}}, -aq_x\}$ and $q_{\tilde{z}} = \max\{aq_{\tilde{x}}, -ap_x\}$.

(b) *If $\tilde{z} = \tilde{x} + \tilde{y}$, then $(p_{\tilde{z}}, q_{\tilde{z}}) = (\sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2}, \sqrt{q_{\tilde{x}}^2 + q_{\tilde{y}}^2})$ satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$.*

(c) *For all $p \geqslant p_{\tilde{x}}$ and $q \geqslant q_{\tilde{x}}$, we have $p \in \mathcal{P}(\tilde{x})$ and $q \in \mathcal{Q}(\tilde{x})$.*

(d) $P(\tilde{x} > \Omega p_{\tilde{x}}) \leqslant \exp\left(-\frac{\Omega^2}{2}\right)$

and

$$P(\tilde{x} < -\Omega q_{\tilde{x}}) \leqslant \exp\left(-\frac{\Omega^2}{2}\right).$$

Proof. (a) We can examine this condition easily from the definitions of $\mathcal{P}(\tilde{z})$ and $\mathcal{Q}(\tilde{z})$.

(b) To prove part (b), let $p_{\tilde{z}} = \sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2}$. For any $\phi \geqslant 0$,

$$\begin{aligned} \mathrm{E}\left(\exp\left(\phi\frac{\tilde{x}+\tilde{y}}{p_{\tilde{z}}}\right)\right) &= \mathrm{E}\left(\exp\left(\phi\frac{\tilde{x}}{p_{\tilde{z}}}\right)\exp\left(\phi\frac{\tilde{y}}{p_{\tilde{z}}}\right)\right) \\ &\qquad \text{(because } \tilde{x} \text{ and } \tilde{y} \text{ are independent)} \\ &= \mathrm{E}\left(\exp\left(\phi\frac{p_{\tilde{x}}}{p_{\tilde{z}}}\frac{\tilde{x}}{p_{\tilde{x}}}\right)\right)\mathrm{E}\left(\exp\left(\phi\frac{p_{\tilde{y}}}{p_{\tilde{z}}}\frac{\tilde{y}}{p_{\tilde{y}}}\right)\right) \\ &\leqslant \exp\left(\frac{\phi^2}{2}\frac{p_{\tilde{x}}^2}{p_{\tilde{z}}^2}\right)\exp\left(\frac{\phi^2}{2}\frac{p_{\tilde{y}}^2}{p_{\tilde{z}}^2}\right) \\ &= \exp\left(\frac{\phi^2}{2}\right). \end{aligned}$$

Thus, $p_{\tilde{z}} = \sqrt{p_{\tilde{x}}^2 + p_{\tilde{y}}^2} \in \mathcal{P}(\tilde{z})$. Similarly, we can show that $\sqrt{q_{\tilde{x}}^2 + q_{\tilde{y}}^2} \in \mathcal{Q}(\tilde{z})$.

(c) Observe that

$$\begin{aligned} \mathrm{E}\left(\exp\left(\phi\frac{\tilde{x}}{p}\right)\right) &= \mathrm{E}\left(\exp\left(\phi\frac{p_{\tilde{x}}}{p}\frac{\tilde{x}}{p_{\tilde{x}}}\right)\right) \\ &\leqslant \exp\left(\frac{\phi^2}{2}\frac{p_{\tilde{x}}^2}{p^2}\right) \leqslant \exp\left(\frac{\phi^2}{2}\right). \end{aligned}$$

The proof for the backward deviation is similar.

(d) Note that

$$\begin{aligned} \mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) &= \mathrm{P}\left(\frac{\Omega\tilde{x}}{p_{\tilde{x}}} > \Omega^2\right) \\ &\leqslant \frac{\mathrm{E}(\exp(\Omega\tilde{x}/p_{\tilde{x}}))}{\exp(\Omega^2)} \leqslant \exp\left(-\frac{\Omega^2}{2}\right), \end{aligned}$$

where the first inequality follows from Chebyshev's inequality. The proof of the backward deviation is the same. □

For some distributions, we can find closed-form bounds on the deviations p^* and q^*, or even the exact expressions. In particular, for a general distribution, we can show that these values are not less than the standard deviation of the distribution. Interestingly, for a normal distribution, the deviation measures p^* and q^* are identical with the standard deviation.

Proposition 4. *If the random variable $\tilde{z}$ has mean zero and standard deviation σ, then $p_{\tilde{z}}^* \geqslant \sigma$ and $q_{\tilde{z}}^* \geqslant \sigma$. In addition, if $\tilde{z}$ is normally distributed, then $p_{\tilde{z}}^* = q_{\tilde{z}}^* = \sigma$.*

Proof. Note that for any $p \in \mathcal{P}(\tilde{z})$, we have

$$\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{p}\right)\right) = 1 + \frac{1}{2}\phi^2\frac{\sigma^2}{p^2} + \sum_{k=3}^{\infty}\frac{\phi^k \mathrm{E}[\tilde{z}^k]}{p^k k!}$$

and

$$\exp\left(\frac{\phi^2}{2}\right) = 1 + \frac{\phi^2}{2} + \sum_{k=2}^{\infty}\frac{\phi^{2k}}{2^k k!}.$$

According to the definition of $\mathcal{P}(\tilde{z})$, we have $\mathrm{E}(\exp(\phi(\tilde{z}/p))) \leqslant \exp(\phi^2/2)$ for any $\phi \geqslant 0$. In particular, this inequality is true for ϕ close to zero, which implies that

$$\frac{1}{2}\phi^2\frac{\sigma^2}{p^2} \leqslant \frac{\phi^2}{2}.$$

Thus, $p \geqslant \sigma$. Similarly, for any $q \in \mathcal{Q}(\tilde{z})$, $q \geqslant \sigma$. For the normal distribution, the proof follows from the fact that

$$\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{\alpha}\right)\right) = \mathrm{E}\left(\exp\left(\phi\frac{\sigma}{\alpha}\frac{\tilde{z}}{\sigma}\right)\right) = \exp\left(\frac{\phi^2\sigma^2}{2\alpha^2}\right). \quad \square$$

For most distributions, we are unable to obtain closed-form expressions for p^* and q^*. Nevertheless, we can still determine their values numerically. For example, if $\tilde{z}$ is uniformly distributed over $[-1, 1]$, we can determine numerically that $p^* = q^* = 0.58$, which is close to the standard deviation 0.5774. In Table 1, we compare the values of p^*, q^*, and the standard deviation σ, where $\tilde{z}$ has the following parametric discrete distribution:

$$\mathrm{P}(\tilde{z} = k) = \begin{cases} \beta & \text{if } k = 1, \\ 1 - \beta & \text{if } k = -\dfrac{\beta}{1-\beta}. \end{cases} \tag{18}$$

In this example, the standard deviation is close to q^*, but underestimates the value of p^*. Hence, it is apparent that if the distribution is asymmetric, the forward and backward deviations may be different from the standard deviation.

Table 1. Numerical comparisons of different deviation measures for centered Bernoulli distributions.

β	p^*	q^*	σ	$\bar{p}$	$\bar{q}$
0.5	1	1	1	1	1
0.4	0.83	0.82	0.82	0.83	0.82
0.3	0.69	0.65	0.65	0.69	0.65
0.2	0.58	0.50	0.50	0.58	0.50
0.1	0.47	0.33	0.33	0.47	0.33
0.01	0.33	0.10	0.10	0.33	0.10

3.2. Approximation of Deviation Measures

It will be clear in the next section that we can use the values of $p^* = \inf\{\mathcal{P}(\tilde{z})\}$ and $q^* = \inf\{\mathcal{Q}(\tilde{z})\}$ in our uncertainty set to obtain the desired probability bound against constraint violation. Unfortunately, however, if the distribution of $\tilde{z}$ is not precisely known, we cannot determine the values of p^* and q^*. Under such circumstances, as long as we can determine (p, q) such that $p \in \mathcal{P}(\tilde{z})$ and $q \in \mathcal{Q}(\tilde{z})$, we can still construct the uncertainty set that achieves the probabilistic guarantees, albeit more conservatively. We first identify (p, q) for a random variable $\tilde{z}$, assuming that we only know its mean and support. We then discuss how to estimate the deviation measures from independent samples.

3.2.1. Deviation Measure Approximation from Mean and Support.

Theorem 4. *If $\tilde{z}$ has zero mean and is distributed in $[-\underline{z}, \bar{z}]$, $\underline{z}, \bar{z} > 0$, then*

$$\bar{p} = \frac{\underline{z}+\bar{z}}{2}\sqrt{g\left(\frac{\underline{z}-\bar{z}}{\underline{z}+\bar{z}}\right)} \in \mathcal{P}(\tilde{z})$$

and

$$\bar{q} = \frac{\underline{z}+\bar{z}}{2}\sqrt{g\left(\frac{\bar{z}-\underline{z}}{\underline{z}+\bar{z}}\right)} \in \mathcal{Q}(\tilde{z}),$$

where

$$g(\mu) = 2\max_{s>0} \frac{\phi_\mu(s) - \mu s}{s^2}$$

and

$$\phi_\mu(s) = \ln\left(\frac{e^s + e^{-s}}{2} + \frac{e^s - e^{-s}}{2}\mu\right).$$

Proof. We focus on the proof of the forward deviation measure. The case for the backward deviation is the same. It is clear from scaling and shifting that

$$\tilde{x} = \frac{\tilde{z} - (\bar{z} - \underline{z})/2}{(\underline{z}+\bar{z})/2} \in [-1, 1].$$

Thus, it suffices to show that

$$\sqrt{g(\mu)} \in \mathcal{P}(\tilde{x}),$$

where

$$\mu = \mathrm{E}[\tilde{x}] = \frac{\underline{z}-\bar{z}}{\underline{z}+\bar{z}} \in (-1, 1).$$

First, observe that $p \in \mathcal{P}(\tilde{x})$ if and only if

$$\ln(\mathrm{E}[\exp(s\tilde{x})]) \leqslant \mathrm{E}(\tilde{x})s + \frac{p^2}{2}s^2 \quad \forall\, s \geqslant 0. \tag{19}$$

We want to find a p such that inequality (19) holds for all possible random variables $\tilde{x}$ distributed in $[-1, 1]$ with mean μ. For this purpose, we formulate a semi-infinite linear program as follows:

$$\begin{aligned}
\max\ & \int_{-1}^{1} \exp(sx) f(x)\, dx \\
\text{s.t.}\ & \int_{-1}^{1} f(x)\, dx = 1, \\
& \int_{-1}^{1} x f(x)\, dx = \mu, \\
& f(x) \geqslant 0.
\end{aligned} \tag{20}$$

The dual of the above semi-infinite linear program is

$$\begin{aligned}
\min\ & u + v\mu \\
\text{s.t.}\ & u + vx \geqslant \exp(sx) \quad \forall\, x \in [-1, 1].
\end{aligned}$$

Because $\exp(sx) - vx$ is convex in x, the dual is equivalent to a linear program with two decision variables:

$$\begin{aligned}
\min\ & u + v\mu \\
\text{s.t.}\ & u + v \geqslant \exp(s), \\
& u - v \geqslant \exp(-s).
\end{aligned} \tag{21}$$

It is easy to check that $(u^*, v^*) = ((e^s + e^{-s})/2, (e^s - e^{-s})/2)$ is the unique extreme point of the feasible set of problem (21), and that $\mu \in (-1, 1)$. Hence, problem (21) is bounded. In particular, the unique extreme point (u^*, v^*) is the optimal solution of the problem. Therefore, $((e^s + e^{-s})/2 + (e^s - e^{-s})/2)\mu$ is the optimal objective value. By weak duality, it is an upper bound of the infinite dimensional linear program (20). Note that $\phi_\mu(0) = 0$ and $\phi'_\mu(0) = \mu$. Therefore, for any random variable $\tilde{x} \in [-1, 1]$ with mean μ, we have

$$\begin{aligned}
&\ln(\mathrm{E}[\exp(s\tilde{x})]) \leqslant \phi_\mu(s) \\
&\quad = \phi_\mu(0) + \phi'_\mu(0)s + \frac{1}{2}s^2 \frac{\phi_\mu(s) - \mu s}{\frac{1}{2}s^2} \leqslant \mu s + \frac{1}{2}s^2 g(\mu).
\end{aligned}$$

Hence, $\sqrt{g(\mu)} \in \mathcal{P}(\tilde{x})$. □

Remark 1. This theorem implies that all probability distributions with bounded support have finite forward and backward deviations. It also enables us to find valid deviation measures from the support of the distribution. In Table 1, we show the values of $\bar{p}$ and $\bar{q}$, which coincide with p^* and q^*, respectively. Indeed, one can see that $\sqrt{g(\mu)} = \inf\{\mathcal{P}(\tilde{x})\}$ for the two-point random variable $\tilde{x}$, which takes value 1 with probability $(1+\mu)/2$ and -1 with probability $(1-\mu)/2$.

REMARK 2. The function $g(\mu)$ defined in the theorem appears hard to analyze. Fortunately, the formulation can be simplified to $g(\mu) = 1-\mu^2$ for $\mu \in [0,1)$. In fact, we note that

$$\frac{\phi_\mu(s)-\mu s}{\frac{1}{2}s^2} = 2\int_0^1 \phi''_\mu(s\xi)(1-\xi)\,d\xi$$

and

$$\phi''_\mu(s) = 1-\left(\frac{\alpha(s)+\mu}{1+\alpha(s)\mu}\right)^2,$$

where $\alpha(s) = (e^s - e^{-s})/(e^s + e^{-s}) \in [0,1)$ for $s \geqslant 0$. Because for $\mu \in (-1,1)$, $\inf_{0\leqslant\alpha<1}(\alpha+\mu)/(1+\alpha\mu) = \mu$, we have for $\mu \in [0,1)$,

$$\phi''_\mu(s) \leqslant \phi''_\mu(0) = 1-\mu^2 \quad \forall\, s \geqslant 0,$$

which implies that $g(\mu) = 1-\mu^2$ for $\mu \in [0,1)$.

Unfortunately, for $\mu \in (-1,0)$, we do not have a closed-form expression for $g(\mu)$. However, we can obtain upper and lower bounds for the function $g(\mu)$. First, note that when $\mu \in (-1,0)$, we have $\phi''_\mu(s) \geqslant \phi''_\mu(0) = 1-\mu^2$ for s close to zero. Hence, $1-\mu^2$ is a lower bound for $g(\mu)$. Numerically, we observe from Figure 2 that $g(\mu) \leqslant 1 - 0.3\mu^2$. On the other hand, when μ is close to -1, the lower bound for $g(\mu)$ is tighter as follows:

$$\underline{p}^2(\mu) = \frac{(1-\mu)^2}{-2\ln((1+\mu)/2)}.$$

Indeed, because any distribution $\tilde{x}$ in $[-1,1]$ with mean μ satisfies

$$\mathrm{P}\big(\tilde{x}-\mu > \Omega\sqrt{g(\mu)}\big) \leqslant \exp(-\Omega^2/2),$$

we have that

$$\sqrt{g(\mu)} \geqslant \underline{p} = \inf\{p\colon \mathrm{P}(\tilde{x}-\mu > \Omega p) < \exp(-\Omega^2/2)\}.$$

In particular, when $\Omega = \sqrt{-2\ln((1+\mu)/2)}$, for the two-point distribution $\tilde{x}$, which takes value 1 with probability $(1+\mu)/2$ and -1 with probability $(1-\mu)/2$, we obtain $\underline{p}^2 = \underline{p}^2(\mu) = (1-\mu)^2/(-2\ln((1+\mu)/2))$. From Figure 2, we observe that as μ approaches -1, $\underline{p}^2(\mu)$ and $g(\mu)$ converge to zero at the same rate.

3.2.2. Deviation Measure Estimators from Samples. From the definition of the forward deviation measure, we can easily derive an alternative expression

$$p^* = \sup_{t>0} \frac{1}{t}\sqrt{2\ln \mathrm{E}[\exp(t(\tilde{z}-\mathrm{E}[\tilde{z}])]}.$$

When the forward deviation measure is finite, given a set of M independent samples of $\tilde{z}$, $\{\nu_1,\ldots,\nu_M\}$, with sample mean $\bar{\nu}$, we can construct an estimator as

$$\hat{p}^*_M = \sup_{t>0}\frac{1}{t}\sqrt{2\ln\frac{1}{M}\sum_{i=1}^M \exp(t(\nu_i-\bar{\nu}))}.$$

A similar estimator can be constructed for the backward deviation measure.

Although closed-form expressions of the bias and variance of the above estimator may be hard to obtain, we can empirically test the accuracy of the above estimator compared with the true value of the deviation measure. Specifically, in Figure 3 we present the empirical histogram of the deviation estimator $\hat{p}^*_M$ with samples from a standard normal distribution, with the true p^* being one.

As can be seen in Figure 3, the accuracy of the estimator increases with the sample size. The estimator seems to be

Figure 2. Function $g(\mu)$ and related bounds.

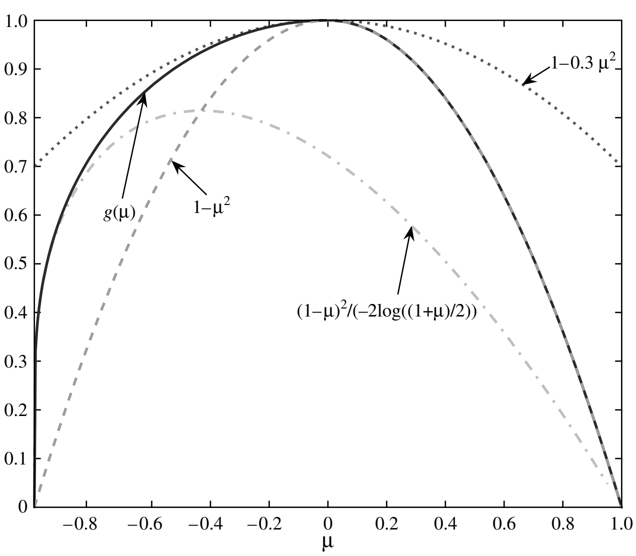


Figure 3. Empirical histogram of the deviation estimator for a standard normal distribution.

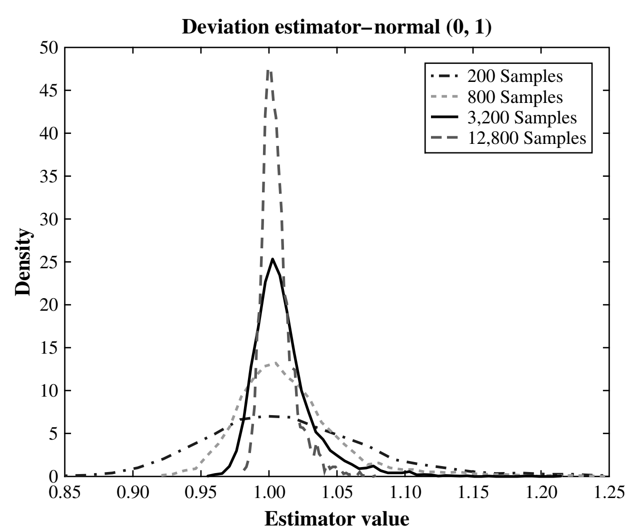


Table 2. Bias and standard deviation of deviation estimators.

M	$\hat{b}(\hat{p}_M^*)$	$\hat{\sigma}(\hat{p}_M^*)$	$1/\sqrt{M}$
100	0.0137	0.0860	0.1
400	0.0163	0.0529	0.05
1,600	0.0134	0.0331	0.025
6,400	0.0077	0.0116	0.0125

upward biased. Empirically, we generated 5,000 estimators for each sample size; the results are summarized in Table 2. From the table, we observe that both the bias ($\hat{b}(\hat{p}_M^*)$) and the standard deviation ($\hat{\sigma}(\hat{p}_M^*)$) of the estimators decrease with the increasing sample size. More specifically, the standard deviation of the estimators decreases approximately as the square root of the sample size.

4. Probability Bounds of Constraint Violation

In this section, we will show that the new deviation measures can be used to guarantee the desired level of constraint violation probability in the robust optimization framework.

Model of Data Uncertainty U.

We assume that the primitive uncertainties $\{\tilde{z}_j\}_{j=1:N}$ are independent, zero-mean random variables, with support $\tilde{z}_j \in [-\underline{z}_j, \bar{z}_j]$, and deviation measures (p_j, q_j), satisfying

$$p_j \in \mathcal{P}(\tilde{z}_j),\ q_j \in \mathcal{Q}(\tilde{z}_j) \quad \forall\, j = \{1, \ldots, N\}.$$

We consider the generalized uncertainty set $\mathcal{G}_\Omega$, which takes into account the worst-case support set $\mathcal{W}$.

Theorem 5. *Let* $\mathbf{x}$ *be feasible for the robust counterpart of* (3) *in which* $\mathcal{U}_\Omega = \mathcal{G}_\Omega$. *Then,*

$$\mathrm{P}(\tilde{\mathbf{a}}'\mathbf{x} > \tilde{b}) \leqslant \exp\left(\frac{-\Omega^2}{2}\right).$$

Proof. Note that $\mathbf{x}$ is feasible in (15), and from §E of the online appendix, we have

$$\mathbf{a}^{0\prime}\mathbf{x} + \min_{\mathbf{r},\,\mathbf{s}\geqslant\mathbf{0}}\{\Omega\|\mathbf{t}(\mathbf{r},\mathbf{s})\|^* + \mathbf{r}'\bar{\mathbf{z}} + \mathbf{s}'\underline{\mathbf{z}}\} \leqslant b^0,$$

where

$$\mathbf{t}(\mathbf{r},\mathbf{s}) = \begin{bmatrix} \max(p_1(y_1 - r_1 + s_1), -q_j(y_1 - r_1 + s_1)) \\ \vdots \\ \max(p_N(y_N - r_N + s_N), -q_j(y_N - r_N + s_N)) \end{bmatrix}$$

and $y_j = \mathbf{\Delta a}^{j\prime}\mathbf{x} - \Delta b^j$. Therefore, it follows that

$$\begin{aligned}\mathrm{P}(\tilde{\mathbf{a}}'\mathbf{x} > \tilde{b}) &= \mathrm{P}(\mathbf{a}^{0\prime}\mathbf{x} + \tilde{\mathbf{z}}'\mathbf{y} > b^0) \\ &\leqslant \mathrm{P}\Big(\tilde{\mathbf{z}}'\mathbf{y} > \min_{\mathbf{r},\,\mathbf{s}\geqslant\mathbf{0}}\{\Omega\|\mathbf{t}(\mathbf{r},\mathbf{s})\|^* + \mathbf{r}'\bar{\mathbf{z}} + \mathbf{s}'\underline{\mathbf{z}}\}\Big) \\ &\leqslant \mathrm{P}\Big(\tilde{\mathbf{z}}'\mathbf{y} > \min_{\mathbf{r},\,\mathbf{s}\geqslant\mathbf{0}}\{\Omega\|\mathbf{t}(\mathbf{r},\mathbf{s})\|_2 + \mathbf{r}'\bar{\mathbf{z}} + \mathbf{s}'\underline{\mathbf{z}}\}\Big).\end{aligned}$$

Let

$$(\mathbf{r}^*, \mathbf{s}^*) = \arg\min_{\mathbf{r},\,\mathbf{s}\geqslant\mathbf{0}}\{\Omega\|\mathbf{t}(\mathbf{r},\mathbf{s})\|_2 + \mathbf{r}'\bar{\mathbf{z}} + \mathbf{s}'\underline{\mathbf{z}}\}$$

and $\mathbf{t}^* = \mathbf{t}(\mathbf{r}^*, \mathbf{s}^*)$. Observe that because $-\underline{z}_j \leqslant \tilde{z}_j \leqslant \bar{z}_j$, we have $r_j^*\bar{z}_j \geqslant r_j^*\tilde{z}_j$ and $s_j^*\underline{z}_j \geqslant -s_j^*\tilde{z}_j$. Therefore,

$$\mathrm{P}(\tilde{\mathbf{z}}'\mathbf{y} > \Omega\|\mathbf{t}^*\|_2 + \mathbf{r}^{*\prime}\bar{\mathbf{z}} + \mathbf{s}^{*\prime}\underline{\mathbf{z}}) \leqslant \mathrm{P}(\tilde{\mathbf{z}}'(\mathbf{y} - \mathbf{r}^* + \mathbf{s}^*) > \Omega\|\mathbf{t}^*\|_2).$$

From Theorem 3(a), we have $t_j^* \in \mathcal{P}(\tilde{z}_j(y_j - r_j^* + s_j^*))$. Following Theorem 3(b), we have

$$\|\mathbf{t}^*\|_2 \in \mathcal{P}(\tilde{\mathbf{z}}'(\mathbf{y} - \mathbf{r}^* + \mathbf{s}^*)).$$

Finally, the desired probability bound follows from Theorem 3(d). □

We use the Euclidian norm as the benchmark to obtain the desired probability bound. It is possible to use other norms, such as the $l_1 \cap l_\infty$-norm, $\|\mathbf{z}\| = \max\{(1/\sqrt{N})\|\mathbf{z}\|_1, \|\mathbf{z}\|_\infty\}$, to achieve the same bound, but the approximation may not be worthwhile. Note from inequality (12) that the value $\Omega\|\mathbf{t}\|^*$ gives the desired "safety distance" against constraint violation. Because $\|\mathbf{t}\|^* \geqslant \|\mathbf{t}\|_2$, one way to compare the conservativeness of different norms is through the following worst-case ratio:

$$\gamma = \max_{\mathbf{t}\neq\mathbf{0}} \frac{\|\mathbf{t}\|^*}{\|\mathbf{t}\|_2}.$$

It turns out that for the $l_1 \cap l_\infty$ norm, $\gamma = \sqrt{\lfloor\sqrt{N}\rfloor + (\sqrt{N} - \lfloor\sqrt{N}\rfloor)^2} \approx N^{1/4}$ (Bertsimas and Sim 2004, Bertsimas et al. 2004). Hence, although the resultant model is linear and of manageable size, the choice of the polyhedral norm can yield more conservative solutions than does the Euclidian norm. In the remainder of the section, we compare the proposed approach with the worst-case approach as well as with other approximation methods of chance constraints.

4.1. Comparison with the Worst-Case Approach

Using the forward and backward deviations, the proposed robust counterpart generalizes the results of Ben-Tal and Nemirovski (2000) and Bertsimas and Sim (2004). Indeed, if $\tilde{z}_j$ has symmetrical support in $[-1, 1]$, from Theorem 4, we have $p_j = q_j = 1$. Hence, our approach provides the same robust counterparts. Our result is actually stronger because we do not require symmetric distributions to ensure the same probability bound of $\exp(-\Omega^2/2)$. The worst-case budget $\Omega_{\max}$ is at least $\sqrt{N}$, such that $\mathcal{G}_{\Omega_{\max}} = \mathcal{W}$. This can be very conservative when N is large. We generalize this result for independent primitive uncertainties, $\tilde{z}_j$, with asymmetrical support, $[-\underline{z}_j, \bar{z}_j]$.

Theorem 6. *The worst-case budget* $\Omega_{\max}$ *for the uncertainty set*

$$\mathcal{G}_\Omega = \mathcal{A}_\Omega \cap \mathcal{W}$$

satisfies

$$\Omega_{\max} \geqslant \sqrt{N}.$$

Proof. From Theorem 4, we have

$$p_j, q_j \leqslant \frac{\underline{z}_j + \bar{z}_j}{2}.$$

Hence, the set $\mathscr{A}_\Omega$ is a subset of

$$\mathscr{D}_\Omega = \left\{ (\mathbf{a}, b)\colon \exists \mathbf{z} \in \Re^N, (\mathbf{a}, b) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\mathbf{\Delta a}^j, \Delta b^j) z_j, \sqrt{\sum_{j=1}^N \frac{z_j^2}{d_j^2}} \leqslant \Omega \right\},$$

where $d_j = (\underline{z}_j + \bar{z}_j)/2$. To show that $\Omega_{\max} \geqslant \sqrt{N}$, it suffices to show that there exist $(\mathbf{a}, b) \in \mathscr{W}$, such that $(\mathbf{a}, b) \notin \mathscr{D}_\Omega \supseteq \mathscr{A}_\Omega$ for all $\Omega < \sqrt{N}$. Let

$$y_j = \begin{cases} \bar{z}_j & \text{if } \bar{z}_j \geqslant \underline{z}_j, \\ -\underline{z}_j & \text{otherwise,} \end{cases}$$

and

$$(\mathbf{a}^*, b^*) = (\mathbf{a}^0, b^0) + \sum_{j=1}^N (\mathbf{\Delta a}^j, \Delta b^j) y_j.$$

Clearly, $(\mathbf{a}^*, b^*) \in \mathscr{W}$ and $|y_j| \geqslant d_j$. Observe that

$$\sqrt{\sum_{j=1}^N \frac{y_j^2}{d_j^2}} \geqslant \sqrt{N}.$$

Hence, $(\mathbf{a}^*, b^*) \notin \mathscr{D}_\Omega \supseteq \mathscr{A}_\Omega$ for all $\Omega < \sqrt{N}$. □

Therefore, even if one knows little about the underlying distribution besides the mean and the support, this approach is potentially less conservative than the worst-case solution.

4.2. Comparison with Other Chance-Constraint Approximation Approaches

Our approach relies on an exponential bound and the relevant Chebyshev inequality to achieve an upper bound on the constraint violation probability. Various other forms of the Chebyshev inequality, such as the one-sided Chebyshev inequality, and the Bernstein inequality, have been used to derive explicit deterministic approximations of chance constraints (see, for example, Kibzun and Kan 1996, Birge and Louveaux 1997, and Pintér 1989). Those approximations usually assume that the mean, variance, and/or support are known, whereas our approach depends on the construction of the forward and backward deviations.

One important advantage of our approach is that we are able to reformulate the approximation of the chance-constrained problem as an SOCP problem. On the other hand, the forward and backward deviations have their own limitations. First of all, as mentioned before, the forward and backward deviations do not exist for some unbounded random variables. For example, the exponential distribution does not have a finite backward deviation. In some cases, we know the first two moments of the random variable, but not the support. In these cases, probability inequalities based on power moments may naturally apply, while bounds based on the forward and backward deviations could be infinite. Second, for bounded random variables, it is possible that the ratio between the deviation measure and the standard deviation is arbitrarily large. This can be seen in Table 1 by comparing the p^* column and the σ column. The implication is that approximations based on probability inequalities using the standard deviation are likely to be less conservative than approximations based on the much larger forward or backward deviations.

To overcome the limitations of the forward and backward deviations, we discuss in §F of the online appendix a general framework for constructing deviation measures, including the standard deviation, to facilitate bounding the probability of constraint violation. These deviation measures, combined with various forms of the Chebyshev inequality (see, for example, Kibzun and Kan 1996), may handle more general distributions. In addition, general deviation measures may provide less conservative approximations when the above forward and backward deviations do not exist or are too large compared with the standard deviation. In the practical settings where the forward and backward deviations are not too large compared with the standard deviation, we believe that our framework should provide a comparable or even better bound. We elaborate on this point in the following subsection.

4.3. Comparison of the Approximation Scheme Based on Forward/Backward Deviations with the Scheme Based on Standard Deviation

For any random variable $\tilde{z}$ with mean zero and standard deviation σ, forward deviation p^* and backward deviation q^*, we have the following from the one-sided Chebyshev inequality:

$$\mathrm{P}(\tilde{z} > \Lambda\sigma) \leqslant 1/(\Lambda^2 + 1), \tag{22}$$

whereas the bound provided by the forward deviation is

$$\mathrm{P}(\tilde{z} > \Omega p^*) \leqslant \exp(-\Omega^2/2). \tag{23}$$

For the same constraint violation probability, ϵ, bound (22) suggests $\Lambda = \sqrt{(1-\epsilon)/\epsilon}$, whereas bound (23) requires $\Omega = \sqrt{-2\ln(\epsilon)}$. Because the probability bounds are tight or asymptotically tight for some distributions,[1] to compare the above two bounds, we can examine the magnitudes of $\Lambda\sigma$ and Ωp^* for various distributions when ϵ approaches zero. For any distribution having the forward deviation close to the standard deviation (such as the normal distribution), we expect bound (22) to perform poorly as compared to (23). Furthermore, because p^* is finite for bounded distributions, the magnitude of $\Lambda\sigma$ will exceed Ωp^* as ϵ approaches zero. For example, in the case of the centered Bernoulli distribution defined in (18), with $\beta = 0.01$, we have $\sigma = 0.1$ and $p^* = 0.33$. Hence, $\Lambda\sigma > \Omega p^*$ for $\epsilon < 0.0099$. It is often necessary in robust optimization to protect against low-probability "disruptive events" that may result in large deviations, such as $\tilde{z}=1$ in this example. Therefore, it may be reasonable to choose $\epsilon < 0.0099 \approx P(\tilde{z}=1) = 0.01$. In this case, it would be better to use bound (23). Another disadvantage of using the standard deviation for bounding probabilities is its inability to capture distributional skewness. As is evident from the two-point distribution of (18), when β is small, the value $\Lambda\sigma$ that ensures $\mathrm{P}(\tilde{z} < -\Lambda\sigma) < \epsilon$ can be large compared to Ωq^*.

5. Stochastic Programs with Chance Constraints

Consider the following two-stage stochastic program:

$$
\begin{aligned}
Z^* = \min\ & \mathbf{c}'\mathbf{x} + \mathrm{E}(\mathbf{d}'\mathbf{y}(\tilde{\mathbf{z}})) \\
\text{s.t.}\ & \mathbf{a}_i(\tilde{\mathbf{z}})'\mathbf{x} + \mathbf{b}_i'\mathbf{y}(\tilde{\mathbf{z}}) \leqslant f_i(\tilde{\mathbf{z}}) \text{ a.e.} \quad \forall\, i \in \{1,\ldots,m\}, \\
& \mathbf{x} \in \Re^{n_1}, \\
& \mathbf{y}(\cdot) \in Y,
\end{aligned} \tag{24}
$$

where $\mathbf{x}$ corresponds to the first-stage decision vector, and $\mathbf{y}(\tilde{\mathbf{z}})$ is the recourse function from a space of measurable functions, Y, with domain $\mathscr{W}$ and range $\Re^{n_2}$.

Note that optimizing over the space of measurable functions amounts to solving an optimization problem with a potentially large or even infinite number of variables. In general, however, finding a first-stage solution, $\mathbf{x}$, such that there exists a feasible recourse for any realization of the uncertainty may be intractable (see Ben-Tal et al. 2004 and Shapiro and Nemirovski 2005). Nevertheless, in some applications of stochastic optimization, the risk of infeasibility often can be tolerated as a trade-off to improve upon the objective value. Therefore, we consider the following stochastic program with chance constraints, which have been formulated and studied in Nemirovski and Shapiro (2004) and Ergodan and Iyengar (2005):

$$
\begin{aligned}
Z^* = \min\ & \mathbf{c}'\mathbf{x} \\
\text{s.t.}\ & \mathrm{P}(\mathbf{a}_i(\tilde{\mathbf{z}})'\mathbf{x} + \mathbf{b}_i'\mathbf{y}(\tilde{\mathbf{z}}) \leqslant f_i(\tilde{\mathbf{z}})) \geqslant 1-\epsilon_i \quad \forall\, i \in \{1,\ldots,m\}, \\
& \mathbf{x} \in \Re^{n}, \\
& \mathbf{y}(\cdot) \in Y,
\end{aligned} \tag{25}
$$

where $\epsilon_i > 0$. To obtain a less conservative solution, we could vary the risk level, ϵ_i, of constraint violation, and therefore enlarge the feasible region of the decision variables, $\mathbf{x}$ and $\mathbf{y}(\cdot)$. Observe that in the above stochastic programming model, we do not include the second-stage cost. We consider such a model for two reasons. First, the second-stage cost is not necessary for many applications, including, for instance, the project management example under uncertain activity time presented in the §A of the online appendix. Second, incorporating a linear second-stage cost into model (25) with chance constraints introduces an interesting modeling issue. That is, because the decision maker is allowed to violate the constraint with certain probability without paying a penalty, he/she may do so intentionally to reduce the second-stage cost, regardless of the uncertainty outcome. To avoid this issue, in this paper we will not include the second-stage cost in the model. We refer the readers to our companion paper Chen et al. (2006) for a more general multistage stochastic programming framework.

Under the Model of Data Uncertainty U, we assume that $\tilde{z}_j \in [-\underline{z}_j, \bar{z}_j]$, $j \in \{1,\ldots,N\}$ are independent random variables with mean zero and deviation parameters (p_j, q_j), satisfying $p_j \in \mathscr{P}(\tilde{z}_j)$ and $q_j \in \mathscr{Q}(\tilde{z}_j)$. For all $i \in \{1,\ldots,m\}$, under the Affine Data Perturbation, we have

$$\mathbf{a}_i(\tilde{\mathbf{z}}) = \mathbf{a}_i^0 + \sum_{j=1}^{N} \Delta\mathbf{a}_i^j \tilde{z}_j$$

and

$$f_i(\tilde{\mathbf{z}}) = f_i^0 + \sum_{j=1}^{N} \Delta f_i^j \tilde{z}_j.$$

To design a tractable robust optimization approach for solving (25), we restrict the recourse function $\mathbf{y}(\cdot)$ to one of the linear decision rules as follows:

$$\mathbf{y}(\mathbf{z}) = \mathbf{y}^0 + \sum_{j=1}^{N} \mathbf{y}^j z_j. \tag{26}$$

Linear decision rules emerged in the early development of stochastic optimization (see Garstka and Wets 1974) and reappeared recently in the affinely adjustable robust counterpart introduced by Ben-Tal et al. (2004). The linear decision rule enables one to design a tractable robust optimization approach for finding feasible solutions in model (25) for all distributions satisfying the Model of Data Uncertainty U.

Theorem 7. *The optimal solution to the following robust counterpart,*

$$
\begin{aligned}
Z_r^* = \min\ & \mathbf{c}'\mathbf{x} \\
\text{s.t.}\ & \mathbf{a}_i^{0\prime}\mathbf{x} + \mathbf{b}_i'\mathbf{y}^0 + \Omega_i h_i + \mathbf{r}^{i\prime}\bar{\mathbf{z}} + \mathbf{s}^{i\prime}\underline{\mathbf{z}} \leqslant f_i^0 \quad \forall\, i \in \{1,\ldots,m\}, \\
& \|\mathbf{u}^i\|^* \leqslant h_i \quad \forall\, i \in \{1,\ldots,m\}, \\
& u_j^i \geqslant p_j(\Delta\mathbf{a}_i^{j\prime}\mathbf{x} + \mathbf{b}_i'\mathbf{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall\, i \in \{1,\ldots,m\},\ j \in \{1,\ldots,N\}, \\
& u_j^i \geqslant -q_j(\Delta\mathbf{a}_i^{j\prime}\mathbf{x} + \mathbf{b}_i'\mathbf{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall\, i \in \{1,\ldots,m\},\ j \in \{1,\ldots,N\}, \\
& \mathbf{x} \in \Re^n, \\
& \mathbf{y}^j \in \Re^k \quad \forall\, j \in \{0,\ldots,N\}, \\
& \mathbf{u}^i, \mathbf{r}^i, \mathbf{s}^i \in \Re_+^N,\ h_i \in \Re \quad \forall\, i \in \{1,\ldots,m\},
\end{aligned} \tag{27}
$$

where $\Omega_i = \sqrt{-2\ln(\epsilon_i)}$, *is feasible in the stochastic optimization model* (25) *for all distributions that satisfy the Model of Data Uncertainty U and* $Z_r^* \geqslant Z^*$.

PROOF. Restricting the space of recourse solutions $\mathbf{y}(\mathbf{z})$ in the form of Equation (26), we have the following problem:

$$\begin{aligned} Z_1^* = \min\ & \mathbf{c}'\mathbf{x} \\ \text{s.t. } & \mathrm{P}\biggl(\mathbf{a}_i^{0'}\mathbf{x} + \mathbf{b}_i'\mathbf{y}^0 \\ & \qquad + \sum_{j=1}^{N} (\boldsymbol{\Delta}\mathbf{a}_i^{j'}\mathbf{x} + \mathbf{b}_i'\mathbf{y}^j - \Delta f_i^j)\tilde{z}_j \leqslant f_i^0\biggr) \qquad (28) \\ & \quad \geqslant 1 - \epsilon_i \quad \forall\, i \in \{1, \ldots, m\}, \\ & \mathbf{x} \in \Re^n, \\ & \mathbf{y}^j \in \Re^k \quad \forall\, j \in \{0, \ldots, N\}, \end{aligned}$$

which gives an upper bound to model (25). Applying Theorem 5 and using Theorem 2, the feasible solution of model (27) is also feasible in model (28) for all distributions that satisfy the Model of Data Uncertainty U. Hence, $Z_r^* \geqslant Z_1^* \geqslant Z^*$. $\square$

We can easily extend the framework to T stage stochastic programs with chance constraints as follows:

$$\begin{aligned} Z^* = \min\ & \mathbf{c}'\mathbf{x} \\ \text{s.t. } & \mathrm{P}\biggl(\mathbf{a}_i(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_T)'\mathbf{x} \\ & \qquad + \sum_{t=1}^{T} \mathbf{b}_{it}'\mathbf{y}_t(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_t) \leqslant f_i(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_T)\biggr) \qquad (29) \\ & \quad \geqslant 1 - \epsilon_i \quad \forall\, i \in \{1, \ldots, m\}, \\ & \mathbf{x} \in \Re^n, \\ & \mathbf{y}_t(\mathbf{z}_1, \ldots \mathbf{z}_t) \in \Re^k \quad \forall\, t = 1, \ldots T,\ \underline{\mathbf{z}}_t \leqslant \mathbf{z}_t \leqslant \bar{\mathbf{z}}_t. \end{aligned}$$

In the multiperiod model, we assume that the underlying uncertainties, $\tilde{\mathbf{z}}_1 \in \Re^{N_1}, \ldots, \tilde{\mathbf{z}}_T \in \Re^{N_T}$, unfold progressively from the first period to the last period. The realization of the primitive uncertainty vector, $\tilde{\mathbf{z}}_t$, is only available at the tth period. Hence, under the Affine Data Perturbation, we may assume that $\tilde{\mathbf{z}}_t$ is statistically independent in different periods. With the above assumptions, we obtain

$$\mathbf{a}_i(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_T) = \mathbf{a}_i^0 + \sum_{t=1}^{T}\sum_{j=1}^{N_t} \boldsymbol{\Delta}\mathbf{a}_{it}^j \tilde{z}_t^j$$

and

$$f_i(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_T) = f_i^0 + \sum_{t=1}^{T}\sum_{j=1}^{N_t} \Delta f_{it}^j \tilde{z}_t^j.$$

To derive the robust formulation of the multiperiod model, we use the following linear decision rule for the recourse function:

$$\mathbf{y}_t(\mathbf{z}_1, \ldots, \mathbf{z}_t) = \mathbf{y}_t^0 + \sum_{\tau=1}^{t}\sum_{j=1}^{N_t} \mathbf{y}_\tau^j z_\tau^j,$$

which fulfills the nonanticipativity requirement. Essentially, the multiperiod robust model is the same as the two-period model presented above, and does not suffer from the "curse of dimensionality."

5.1. On Linear Decision Rules

The linear decision rule is the key enabling mechanism that permits scalability to multistage models. It has appeared in earlier proposals for solving stochastic optimization problems (see, for example, Charnes and Cooper 1963 and Charnes et al. 1958). However, due to its perceived limitations, the method was short-lived (see Garstka and Wets 1974). While we acknowledge the limitations of using linear decision rules, it is worth considering the arguments for using such a simple rule to achieve computational tractability. One criticism is that a purportedly feasible stochastic optimization problem may not be feasible any more if one restricts the recourse function to a linear decision rule. Indeed, hard constraints, such as $y(\tilde{\mathbf{z}}) \geqslant 0$, can nullify any benefit of linear decision rules on the recourse function, $y(\tilde{\mathbf{z}})$. As an illustration, consider the following hard constraint

$$\begin{aligned} & y(\tilde{\mathbf{z}}) \geqslant 0, \\ & y(\tilde{\mathbf{z}}) \geqslant b(\tilde{\mathbf{z}}) = b_0 + \sum_{j=1}^{N} b_j \tilde{z}_j, \end{aligned} \qquad (30)$$

where $b_j \neq 0$, and the primitive uncertainties, $\tilde{\mathbf{z}}$, have unbounded support and finite forward and backward deviations (e.g., normally distributed). It is easy to verify that a linear decision rule,

$$y(\tilde{\mathbf{z}}) = y_0 + \sum_{j=1}^{N} y_j \tilde{z}_j,$$

is not feasible for constraints (30).

On the other hand, the linear decision rule can survive under soft constraints such as

$$\begin{aligned} & \mathrm{P}(y(\tilde{\mathbf{z}}) \geqslant 0) \leqslant 1 - \epsilon, \\ & \mathrm{P}(y(\tilde{\mathbf{z}}) \geqslant b(\tilde{\mathbf{z}})) \leqslant 1 - \epsilon, \end{aligned}$$

even for very small ϵ. For example, if $p_j = q_j = 1$ and $\epsilon = 10^{-7}$, the following robust counterpart approximation of the chance constraints becomes

$$\begin{aligned} & y_0 \geqslant \Omega \|[y_1, \ldots, y_N]\|_2, \\ & y_0 - b_0 \geqslant \Omega \|[y_1 - b_1, \ldots, y_N - b_N]\|_2, \end{aligned}$$

where $\Omega = 5.68$. Because $\Omega = \sqrt{-2\ln(\epsilon)}$ is a small number even for very high reliability (ϵ very small), the space of feasible linear decision rules may not be overly constrained. Hence, the linear decision rule may remain viable if one can tolerate some risk of infeasibility in the stochastic optimization model.

Another criticism of linear decision rules is that, in general, linear decision rules are not optimal. Indeed, as pointed out by Garstka and Wets (1974), the optimal policy is given by a linear decision rule only under very restrictive assumptions. However, Shapiro and Nemirovski (2005, pp. 142–143) have stated:

> The only reason for restricting ourselves with affine decision rules[2] stems from the desire to end up with a computationally tractable problem. We do not pretend that affine decision rules approximate well the optimal ones—whether it is so or not, it depends on the problem, and we usually have no possibility to understand how good in this respect is a particular problem we should solve. The rationale behind restricting to affine decision rules is the belief that in actual applications it is better to pose a modest and achievable goal rather than an ambitious goal which we do not know how to achieve.

Further, even though linear decision rules are not optimal, they seem to perform reasonably well for some applications (see Ben-Tal et al. 2004, 2005), as will be seen in the project management example presented in §A of the online appendix.

6. Conclusions

The new deviation measures enable us to refine the descriptions of uncertainty sets by including distributional asymmetry. This in turn enables one to obtain less conservative solutions while achieving better approximation to the chance constraints. An empirical study has recently been done by Natarajan et al. (2006), where they use the approach discussed in this paper as a heuristic for minimizing the value-at-risk of a portfolio. Surprisingly, it gives superior out-of-sample performance when tested on a set of real data. We also used linear decision rules to formulate multiperiod stochastic models with chance constraints as a tractable robust counterpart. Advances of SOCP solvers make it possible to solve robust models of decent size. Using the robust optimization approach to tackle certain types of stochastic optimization problems thus can be both practically useful and computationally appealing.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. Bound (22) is tight for the centered Bernoulli distribution of (18), in which $\beta=\epsilon$. Indeed, to safeguard against the low probability event of $\tilde{z}=1$, we require Λ to be at least $1/\sigma = 1/\sqrt{\beta+\beta^2/(1-\beta)} = \sqrt{(1-\epsilon)/\epsilon}$, so that $\mathrm{P}(\tilde{z}>\Lambda\sigma)<\epsilon$. For the same two-point distribution, we verify numerically that Ωp^* converges to one, as $\beta=\epsilon$ approaches zero, suggesting that bound (23) is also asymptotically tight.
2. An affine decision rule is equivalent to a linear decision rule in our context.

Acknowledgments

The authors thank the reviewers of this paper for several insightful comments. The research of the second author was supported by SMA, NUS Risk Management Institute, NUS academic research grants R-314-000-066-122 and R-314-000-068-122.

References

Atamtürk, A. 2006. Strong formulations of robust mixed 0-1 programming. *Math. Programming* **108**(2) 235–250.

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

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions to uncertain programs. *Oper. Res. Lett.* **25** 1–13.

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* **88** 411–424.

Ben-Tal, A., B. Golany, A. Nemirovski, J. Vial. 2005. Supplier-retailer flexible commitments contracts: A robust optimization approach. *Manufacturing Service Oper. Management* **7**(3) 248–271.

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

Bertsimas, D., M. Sim. 2003. Robust discrete optimization and network flows. *Math. Programming* **98** 49–71.

Bertsimas, D., M. Sim. 2004a. Price of robustness. *Oper. Res.* **52**(1) 35–53.

Bertsimas, D., M. Sim. 2004b. Robust discrete optimization and downside risk measures. Working paper, National University of Singapore, Singapore.

Bertsimas, D., M. Sim. 2006. Tractable approximations to robust conic optimization problems. *Math. Programming* **107**(1) 5–36.

Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Oper. Res.* **54**(1) 150–168.

Bertsimas, D., D. Pachamanova, M. Sim. 2004. Robust linear optimization under general norms. *Oper. Res. Lett.* **32** 510–516.

Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.

Charnes, A., W. W. Cooper. 1959. Uncertain convex programs: Randomize solutions and confidence level. *Management Sci.* **6** 73–79.

Charnes, A., W. W. Cooper. 1963. Deterministic equivalents for optimizing and satisficing under chance constraints. *Oper. Res.* **11** 18–39.

Charnes, A., W. W. Cooper, G. Symonds. 1958. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil production. *Management Sci.* **4** 253–263.

Chen, X., M. Sim, P. Sun, J. Zhang. 2006. A linear-decision based approximation approach to stochastic programming. *Oper. Res.* Forthcoming.

El-Ghaoui, L., H. Lebret. 1997. Robust solutions to least-square problems to uncertain data matrices. *SIAM J. Matrix Anal. Appl.* **18** 1035–1064.

El-Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **9** 33–52.

Ergodan, G., G. Iyengar. 2005. On two-stage convex constrained problems. CORC technical report TR-2005-06, Columbia University, New York.

Garstka, S. J., R. J.-B. Wets. 1974. On decision rules in stochastic programming. *Math. Programming* **17**(1) 117–143.

Goldfarb, D., G. Iyengar. 2003. Robust convex quadratically constrained programs. *Math. Programming Series B* **97**(3) 495–515.

Kibzun, A., Y. Kan. 1996. *Stochastic Programming Problems with Probability and Quantile Functions*. Johns Wiley & Sons, New York.

Natarajan, K., D. Pachamanova, M. Sim. 2006. A tractable parametric approach to value-at-risk optimization. Working paper, NUS Business School, Singapore.

Nemirovski, A., A. Shapiro. 2006. Scenario approximations of chance constraints. G. Calafiore, F. Dabenne, eds. *Probabilistic and Randomized Methods for Design Under Uncertainty*. Springer-Verlag, London, UK.

Pintér, J. 1989. Deterministic approximations of probability inequalities. *ZOR - Methods Models Oper. Res.* **33** 219–239.

Prékopa, A. 1995. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, Boston.

Shapiro, A., A. Nemirovski. 2005. On complexity of stochastic programming problems. V. Jeyakumar, A. M. Rubinov, eds. *Continuous Optimization: Current Trends and Modern Applications*. Springer, Heidelberg, Germany, 111–144.

Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* **21** 1154–1157.