A Robust Optimization Perspective on Stochastic Programming

Xin Chen[*], Melvyn Sim[†] and Peng Sun[‡]

Submitted: December 2004

Revised: May 16, 2006

Abstract

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.

[*]Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign. Email: xinchen@uiuc.edu

[†]NUS Business School, National University of Singapore. Email: dscsimm@nus.edu.sg. The research of the author is supported by NUS academic research grant R-314-000-066-122.

[‡]Fuqua School of Business, Duke University Box 90120, Durham, NC 27708, USA. Email: psun@duke.edu

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 [35] 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 in order 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 [2, 3, 4], El-Ghaoui et al. [22, 23], Iyangar and Goldfarb [26], Bertsimas and Sim [9, 10, 11, 12] and Atamtürk [1]). 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, while avoiding the extreme conservatism of the 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 [4]) and polytopes (see Bertsimas and Sim [10]), which yields *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. [5] 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 [13], Ben-Tal et al. [5]) and supply contracts (Ben-Tal et al. [6]).

Two important characteristics of robust linear optimization that make it practically appealing are

(a) Robust linear optimization models are polynomial in size and in the form of Linear Programming 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.

(b) 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 [10] and Ben-Tal and Nemirovski [4] 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 while 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 [17], [18]). Finding exact solutions to chance constrained problems is typically intractable. Pintér [31] 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 and Birge and Louveaux [14], Chapter 9.4 and Kibzun and Kan [29]). 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 [10] and Ben-Tal and Nemirovski [4].

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 Section 2, we introduce a new uncertainty set and formulate the robust counterpart. In Section 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. We present in Section 5 an SOCP approximation for stochastic programming with chance constraints. In Section 6 we apply our framework to a project management problem with uncertain completion time. Section 7 presents a summary and conclusions.

Notations We denote a random variable, $\tilde{x}$, with the tilde sign. Bold face lower case letters, such as $\boldsymbol{x}$, represent vectors and the corresponding upper case letters, such as $\boldsymbol{A}$, denote matrices.

2 Robust Formulation of a Stochastic Linear Constraint

Consider a stochastic linear constraint,

$$\tilde{\boldsymbol{a}}'\boldsymbol{x} \leq \tilde{b}, \tag{1}$$

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

Affine Data Perturbation:

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

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

where $\boldsymbol{D}^0$ is the nominal value of the data, and $\boldsymbol{\Delta D}^j$, $j \in N$, 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 $\tilde{\boldsymbol{D}}$ are strongly dependent. If N is large, we model the case that the elements of $\tilde{\boldsymbol{D}}$ are weakly dependent. In the limiting case when the number of entries in the data equals N, the elements of $\tilde{\boldsymbol{D}}$ are independent.

We desire a set of solutions $X(\epsilon)$ such that $\boldsymbol{x} \in X(\epsilon)$ 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 [17]),

$$X(\epsilon) = \left\{ \boldsymbol{x} \ : \ \mathrm{P}(\tilde{\boldsymbol{a}}'\boldsymbol{x} \leq \tilde{b}) \geq 1 - \epsilon \right\}. \tag{2}$$

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 non-convex and computationally intractable (see Birge and Louveaux [14]). 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

$$X_r(\Omega) = \{\boldsymbol{x} \ : \ \boldsymbol{a}'\boldsymbol{x} \le b \ \forall (\boldsymbol{a}, b) \in \mathcal{U}_\Omega\}, \tag{3}$$

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, $\mathcal{U}_{\Omega=0} = (\boldsymbol{a}^0, b^0)$, such that $\mathcal{U}_\Omega \subseteq \mathcal{U}_{\Omega'} \subseteq \mathcal{W}$ for all $\Omega_{\max} \ge \Omega' \ge \Omega \ge 0$. Here the worst case uncertainty set $\mathcal{W}$ is the convex support of the uncertain data, defined as follows,

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

which is the smallest closed convex set satisfying $\mathrm{P}((\tilde{\boldsymbol{a}}, \tilde{b}) \in \mathcal{W}) = 1$. 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 [35], 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, $\underline{\boldsymbol{z}} = \bar{\boldsymbol{z}} = \boldsymbol{1}$. 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.

$$\begin{aligned} \mathcal{S}_\Omega &= \underbrace{\left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j) z_j, \|\boldsymbol{z}\| \le \Omega \right\}}_{=\mathcal{V}_\Omega} \cap \mathcal{W} \\ &= \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j) z_j, \|\boldsymbol{z}\| \le \Omega, \|\boldsymbol{z}\|_\infty \le 1 \right\}. \end{aligned} \tag{5}$$

As the budget of uncertainty Ω increases, the norm uncertainty set $\mathcal{V}_\Omega$ expands radially from the point $(\boldsymbol{a}^0, b^0)$ 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 [4], 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 $1 - \exp\left(-\Omega^2/2\right)$. The robust counterpart is a formulation with second order cone constraints. Bertsimas and Sim [10] consider an $l_1 \cap l_\infty$ norm of the form $\|\boldsymbol{z}\|_{l_1 \cap l_\infty} = \max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\}$, and show that the feasibility guarantee is also $1-\exp\left(-\Omega^2/2\right)$. 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 $\|\boldsymbol{z}\|_2 \leq \Omega$ and $\max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\} \leq \Omega$ are the consequence of $\boldsymbol{z}$ satisfying, $\|\boldsymbol{z}\|_\infty \leq 1$. 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,

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

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

$$\|\boldsymbol{u}\| \leq \|\boldsymbol{u}\|_2, \forall \boldsymbol{u}. \tag{7}$$

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 $\|.\|^*$ is defined as

$$\|\boldsymbol{u}\|^* = \max_{\|\boldsymbol{x}\| \leq 1} \boldsymbol{u}'\boldsymbol{x}.$$

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

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

(a) $\|\boldsymbol{w}\|^* = \||\boldsymbol{w}|\|^*$.

(b) *For all $\boldsymbol{v}, \boldsymbol{w}$ such that $|\boldsymbol{v}| \leq |\boldsymbol{w}|$, $\|\boldsymbol{v}\|^* \leq \|\boldsymbol{w}\|^*$.*

(c) *For all $\boldsymbol{v}, \boldsymbol{w}$ such that $|\boldsymbol{v}| \leq |\boldsymbol{w}|$, $\|\boldsymbol{v}\| \leq \|\boldsymbol{w}\|$.*

(d) $\|\boldsymbol{t}\|^* \geq \|\boldsymbol{t}\|_2, \forall \boldsymbol{t}$.

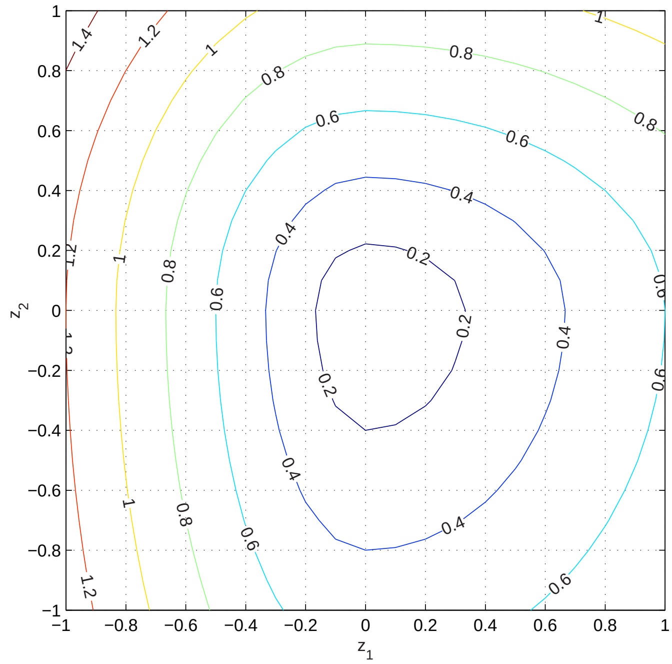

Figure 1: An uncertainty set represented by $\mathcal{A}_\Omega$ as Ω varies for $N = 2$.

Proof : The proofs of **(a)**, **(b)** and **(c)** are shown in Bertsimas and Sim [11].

(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 $\boldsymbol{t}$ observe that

$$\|\boldsymbol{t}\|^* = \max_{\|\boldsymbol{z}\|\leq 1} \boldsymbol{t}'\boldsymbol{z} \geq \max_{\|\boldsymbol{z}\|_2\leq 1} \boldsymbol{t}'\boldsymbol{z} = \|\boldsymbol{t}\|_2^* = \|\boldsymbol{t}\|_2.$$

■

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, $\mathcal{W}$, and define the asymmetric norm uncertainty set as follows,

$$\mathcal{A}_\Omega = \Big\{ (\boldsymbol{a}, b) \;:\; \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, (\mathrm{a}, b) = (\mathrm{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j)(v_j - w_j), \quad \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega, \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0} \Big\}, \tag{8}$$

where $\boldsymbol{P} = \mathrm{diag}(p_1, \ldots, p_N)$ and likewise $\boldsymbol{Q} = \mathrm{diag}(q_1, \ldots, q_N)$ 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 clafity how $\boldsymbol{P}$ and $\boldsymbol{Q}$ relate to the forward and backward deviations of the underlying primitive uncertainties. The following proposition shows the connection of the set $\mathcal{A}_\Omega$ with the uncertainty set described by norm, $\mathcal{V}_\Omega$ defined in (5).

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

The proof is shown in Appendix A.

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.

Since $p_j, q_j > 0$ for $\Omega > 0$, the point $(\boldsymbol{a}^0, b^0)$ lies in the interior of the uncertainty set $\mathcal{A}_\Omega$. 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*

$$\begin{aligned} z^* = \quad & \max \quad \boldsymbol{a}'\boldsymbol{v} + \boldsymbol{b}'\boldsymbol{w} \\ & \text{s.t.} \quad \|\boldsymbol{v} + \boldsymbol{w}\| \leq \Omega \\ & \qquad\;\; \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0}, \end{aligned} \tag{9}$$

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

We present the proof in Appendix B.

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

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

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

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

$$\Updownarrow$$

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \max_{\substack{\{\boldsymbol{v}, \boldsymbol{w} \,:\, \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \leq \Omega \\ \boldsymbol{v}, \boldsymbol{w} \geq \boldsymbol{0}\}}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \leq b^0$$

Observe that

$$
\begin{aligned}
& \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}\,:\,\|\boldsymbol{P}^{-1}\boldsymbol{v}+\boldsymbol{Q}^{-1}\boldsymbol{w}\|\leq\Omega \\ \boldsymbol{v},\boldsymbol{w}\geq\boldsymbol{0}\}}} (\boldsymbol{v}-\boldsymbol{w})'\boldsymbol{y} \\
= & \max_{\substack{\{\boldsymbol{v},\boldsymbol{w}\,:\,\|\boldsymbol{v}+\boldsymbol{w}\|\leq\Omega \\ \boldsymbol{v},\boldsymbol{w}\geq\boldsymbol{0}\}}} (\boldsymbol{P}\boldsymbol{y})'\boldsymbol{v}-(\boldsymbol{Q}\boldsymbol{y})'\boldsymbol{w} \qquad (11) \\
= & \ \Omega\|\boldsymbol{t}\|^*
\end{aligned}
$$

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

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{t}\|^* \leq b^0. \qquad (12)$$

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

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{t}\|^* \leq \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega\|\boldsymbol{u}\|^* \leq \boldsymbol{a}^{0\prime}\boldsymbol{x} + \Omega h \leq b^0.$$

■

The complete formulation and complexity class of the robust counterpart depends on the representation of the dual norm constraint, $\|\boldsymbol{u}\|^* \leq y$. In Appendix C, 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 worst case support set, $\mathcal{W}$

We now incorporate the worst case support set $\mathcal{W}$ as follows

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

Since we can represent the support set of $\mathcal{W}$ equivalently as

$$\mathcal{W} = \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, (\boldsymbol{a}, b) = (\boldsymbol{a}^0, b^0) + \sum_{j=1}^N (\boldsymbol{\Delta a}^j, \Delta b^j)(v_j - w_j), -\underline{\boldsymbol{z}} \leq \boldsymbol{v} - \boldsymbol{w} \leq \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \geq \boldsymbol{0} \right\}, \qquad (13)$$

it follows that

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

We will show an equivalent formulation of the corresponding robust counterpart under the generalized uncertainty set, $\mathcal{G}_\Omega$.

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

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

Proof : As in the exposition of Theorem 1, the robust counterpart of (3), in which $\mathcal{U}_\Omega = \mathcal{G}_\Omega$, is as follows,

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \max_{(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{C}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \le b^0$$

where

$$\mathcal{C} = \left\{ (\boldsymbol{v}, \boldsymbol{w}) \ : \ \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, -\underline{\boldsymbol{z}} \le \boldsymbol{v} - \boldsymbol{w} \le \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \ge \boldsymbol{0} \right\}$$

and $y_j = \boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j$. Since $\mathcal{C}$ is a compact convex set with nonempty interior, we can use strong duality to obtain the equivalent representation. Observe that

$$\begin{aligned} & \max_{\substack{\{\boldsymbol{v}, \boldsymbol{w} \ : \ \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, \\ -\underline{\boldsymbol{z}} \le \boldsymbol{v} - \boldsymbol{w} \le \bar{\boldsymbol{z}}, \boldsymbol{w}, \boldsymbol{v} \ge \boldsymbol{0}\}}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} \\ = \ & \min_{\boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}} \left\{ \max_{\substack{\{\boldsymbol{v}, \boldsymbol{w} : \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, \\ \boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0}\}}} (\boldsymbol{v} - \boldsymbol{w})'\boldsymbol{y} + \boldsymbol{r}'(\bar{\boldsymbol{z}} - \boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{s}'(\underline{\boldsymbol{z}} + \boldsymbol{v} - \boldsymbol{w}) \right\} \\ = \ & \min_{\boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}} \left\{ \max_{\substack{\{\boldsymbol{v}, \boldsymbol{w} : \|\boldsymbol{P}^{-1}\boldsymbol{v} + \boldsymbol{Q}^{-1}\boldsymbol{w}\| \le \Omega, \\ \boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0}\}}} (\boldsymbol{y} - \boldsymbol{r} + \boldsymbol{s})'\boldsymbol{v} - (\boldsymbol{y} - \boldsymbol{r} + \boldsymbol{s})'\boldsymbol{w} + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\} \\ = \ & \min_{\boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}} \left\{ \max_{\substack{\{\boldsymbol{v}, \boldsymbol{w} : \|\boldsymbol{v} + \boldsymbol{w}\| \le \Omega, \\ \boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0}\}}} (\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{r} + \boldsymbol{s}))'\boldsymbol{v} - (\boldsymbol{Q}(\boldsymbol{y} - \boldsymbol{r} + \boldsymbol{s}))'\boldsymbol{w} + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\} \\ = \ & \min_{\boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}} \left\{ \Omega \|\boldsymbol{t}(\boldsymbol{r}, \boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}} \right\}, \end{aligned}$$

where the first equality is due to strong Lagrangian duality (see, for instance, Bertsekas [7]) and the last inequality follows from Proposition 3, in which

$$\begin{aligned}
\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s}) &= \begin{bmatrix} \max(p_1(y_1 - r_1 + s_1), -q_j(y_1 - r_1 + s_1), 0) \\ \vdots \\ \max(p_N(y_N - r_N + s_N), -q_j(y_N - r_N + s_N), 0) \end{bmatrix} \\
&= \begin{bmatrix} \max(p_1(y_1 - r_1 + s_1), -q_j(y_1 - r_1 + s_1)) \\ \vdots \\ \max(p_N(y_N - r_N + s_N), -q_j(y_N - r_N + s_N)) \end{bmatrix}.
\end{aligned}$$

Hence, the robust counterpart is the same as

$$\boldsymbol{a}^{0\prime}\boldsymbol{x} + \min_{\boldsymbol{r},\boldsymbol{s}\geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\} \leq b^0. \tag{16}$$

Using similar arguments as in Theorem 1, we can easily show that the feasible solution of (16) is equivalent to (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 Appendix D.

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,

$$\mathcal{P}(\tilde{z}) = \left\{\alpha \ : \ \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(\frac{\phi}{\alpha}\right) \leq \exp\left(\frac{\phi^2}{2}\right) \quad \forall \phi \geq 0\right\}. \tag{17}$$

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

$$\mathcal{Q}(\tilde{z}) = \left\{\alpha \ : \ \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(-\frac{\phi}{\alpha}\right) \leq \exp\left(\frac{\phi^2}{2}\right) \quad \forall \phi \geq 0\right\}. \tag{18}$$

For completeness, we also define $\mathcal{P}(c) = \mathcal{Q}(c) = \Re_+$ for any constant c. Observe that $\mathcal{P}(\tilde{z}) = \mathcal{Q}(\tilde{z})$ if $\tilde{z}$ is symmetrically distributed around its mean. For known distributions, we define the forward deviation of $\tilde{z}$ as $p^*_{\tilde{z}} = \inf \mathcal{P}(\tilde{z})$ and the backward deviation as $q^*_{\tilde{z}} = \inf \mathcal{Q}(\tilde{z})$.

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*

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

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 $(p_{\tilde{z}}, q_{\tilde{z}}) = \left(\sqrt{p^2_{\tilde{x}} + p^2_{\tilde{y}}}, \sqrt{q^2_{\tilde{x}} + q^2_{\tilde{y}}}\right)$ satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$.*

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

(d)

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

and

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

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 $\phi \ge 0$,

$$\begin{aligned}
& \mathrm{E}\left(\exp\left(\phi\tfrac{\tilde{x}+\tilde{y}}{p_{\tilde{z}}}\right)\right) \\
= & \mathrm{E}\left(\exp\left(\phi\tfrac{\tilde{x}}{p_{\tilde{z}}}\right)\exp\left(\phi\tfrac{\tilde{y}}{p_{\tilde{z}}}\right)\right) \qquad \text{[since } \tilde{x} \text{ and } \tilde{y} \text{ are independent]} \\
= & \mathrm{E}\left(\exp\left(\phi\tfrac{p_{\tilde{x}}}{p_{\tilde{z}}}\tfrac{\tilde{x}}{p_{\tilde{x}}}\right)\right)\mathrm{E}\left(\exp\left(\phi\tfrac{p_{\tilde{y}}}{p_{\tilde{z}}}\tfrac{\tilde{y}}{p_{\tilde{y}}}\right)\right) \\
\le & \exp\left(\tfrac{\phi^2}{2}\tfrac{p_{\tilde{x}}^2}{p_{\tilde{z}}^2}\right)\exp\left(\tfrac{\phi^2}{2}\tfrac{p_{\tilde{y}}^2}{p_{\tilde{z}}^2}\right) \\
= & \exp\left(\tfrac{\phi^2}{2}\right).
\end{aligned}$$

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

$$\mathrm{E}\left(\exp\left(\phi\frac{\tilde{x}}{p}\right)\right) = \mathrm{E}\left(\exp\left(\phi\frac{p_{\tilde{x}}}{p}\frac{\tilde{x}}{p_{\tilde{x}}}\right)\right) \le \exp\left(\frac{\phi^2}{2}\frac{p_{\tilde{x}}^2}{p^2}\right) \le \exp\left(\frac{\phi^2}{2}\right).$$

The proof for the backward deviation is similar.
(d) Note that

$$\mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) = \mathrm{P}\left(\frac{\Omega\tilde{x}}{p_{\tilde{x}}} > \Omega^2\right) \le \frac{\mathrm{E}\left(\exp\left(\frac{\Omega\tilde{x}}{p_{\tilde{x}}}\right)\right)}{\exp(\Omega^2)} \le \exp\left(-\frac{\Omega^2}{2}\right),$$

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 $p_{\tilde{z}}^* \ge \sigma$ and $q_{\tilde{z}}^* \ge \sigma$. In addition, if $\tilde{z}$ is normally distributed, then $p_{\tilde{z}}^* = q_{\tilde{z}}^* = \sigma$.*

Proof : Notice 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 $\mathrm{E}\left(\exp\left(\phi\frac{\tilde{z}}{p}\right)\right) \le \exp\left(\frac{\phi^2}{2}\right)$ for any $\phi \ge 0$. In particular, this inequality is true for ϕ close to zero, which implies that

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

Thus, $p \ge \sigma$. Similarly, for any $q \in \mathcal{Q}(\tilde{z})$, $q \ge \sigma$.

For the normal distribution, the proof follows from the fact that

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

■

For most distributions, we are unable to obtain closed-form expressions for p^* and q^*. Nevertheless, we can still determine their values numerically. For instance, 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

$$\mathrm{P}(\tilde{z} = k) = \begin{cases} \beta & \text{if } k = 1 \\ 1 - \beta & \text{if } k = -\frac{\beta}{1-\beta} \end{cases}. \tag{19}$$

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.

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 can not 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.

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

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

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

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

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

$$\ln\left(\mathrm{E}\left[\exp\left(s\tilde{x}\right)\right]\right) \leq \mathrm{E}(\tilde{x})s + \frac{p^2}{2}s^2, \forall s \geq 0. \tag{20}$$

We want to find a p such that the inequality (20) 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:

$$\begin{array}{ll} \max & \int_{-1}^{1} \exp(sx)f(x)dx \\ \text{s.t.} & \int_{-1}^{1} f(x)dx = 1 \\ & \int_{-1}^{1} xf(x)dx = \mu \\ & f(x) \geq 0. \end{array} \tag{21}$$

The dual of the above semi-infinite linear program is

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

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

$$\begin{array}{ll} \min & u + v\mu \\ \text{s.t.} & u + v \geq \exp(s) \\ & u - v \geq \exp(-s). \end{array} \tag{22}$$

It is easy to check that $(u^*, v^*) = \left(\frac{e^s+e^{-s}}{2}, \frac{e^s-e^{-s}}{2}\right)$ is the unique extreme point of the feasible set of problem (22), and that $\mu \in (-1,1)$. Hence, problem (22) is bounded. In particular, the unique extreme point (u^*, v^*) is the optimal solution of the problem. Therefore, $\frac{e^s+e^{-s}}{2} + \frac{e^s-e^{-s}}{2}\mu$ is the optimal objective value. By weak duality, it is an upper bound of the infinite dimensional linear program (21).

Notice 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

$$\ln\left(\mathrm{E}\left[\exp\left(s\tilde{x}\right)\right]\right) \leq \phi_\mu(s) = \phi_\mu(0) + \phi'_\mu(0)s + \frac{1}{2}s^2 \frac{\phi_\mu(s) - \mu s}{\frac{1}{2}s^2} \leq \mu s + \frac{1}{2}s^2 g(\mu).$$

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 notice 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 $s \geq 0$. Since for $\mu \in (-1, 1)$, $\inf_{0\leq\alpha<1} \frac{\alpha+\mu}{1+\alpha\mu} = \mu$, we have for $\mu \in [0, 1)$,

$$\phi''_\mu(s) \leq \phi''_\mu(0) = 1 - \mu^2, \forall s \geq 0,$$

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, notice that when $\mu \in (-1, 0)$, we have $\phi''_\mu(s) \geq \phi''_\mu(0) = 1 - \mu^2$ for s close to 0. Hence, $1 - \mu^2$ is a lower bound for $g(\mu)$. Numerically, we observe from Figure 2 that $g(\mu) \leq 1 - 0.3\mu^2$. 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, since any distribution $\tilde{x}$ in $[-1, 1]$ with mean μ satisfies

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

we have that

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

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 $\underline{p}^2 = \underline{p}^2(\mu) = \frac{(1-\mu)^2}{-2\ln((1+\mu)/2)}$. From Figure 2, we observe that as μ approaches -1, $\underline{p}^2(\mu)$ and $g(\mu)$ converge to 0 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

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

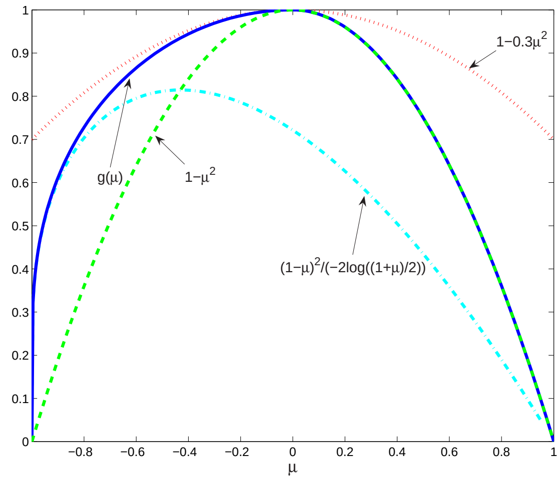

Figure 2: Function $g(\mu)$ and related bounds

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.

While closed form expressions of the bias and variance of the above estimator may be hard to obtain, we can test empirically 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 1.

As can be seen in Figure 3, the accuracy of the estimator increases with the sample size. The estimator seems to be upward biased. Empirically, we generated 5000 estimators for each sample size; the results are summarized in Table 3.2.2. 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.

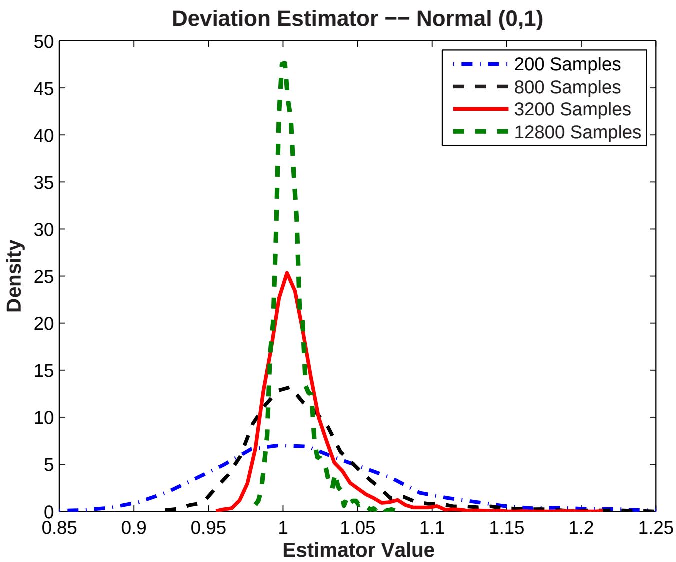

Figure 3: Empirical histogram of the deviation estimator for a standard normal distribution.

M	$\hat{b}\left(\hat{p}_M^*\right)$	$\hat{\sigma}\left(\hat{p}_M^*\right)$	$1/\sqrt{M}$
100	0.0137	0.0860	0.1
400	0.0163	0.0529	0.05
1600	0.0134	0.0331	0.025
6400	0.0077	0.0116	0.0125

Table 2: Bias and standard deviation of deviation estimators.

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,

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

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

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

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

Proof : Since $\boldsymbol{x}$ is feasible in (15), using the equivalent formulation of inequality (16), it follows that

$$\begin{aligned} & \mathrm{P}\left(\tilde{\boldsymbol{a}}'\boldsymbol{x} > \tilde{b}\right) \\ = \ & \mathrm{P}\left(\boldsymbol{a}^{0\prime}\boldsymbol{x} + \tilde{\boldsymbol{z}}'\boldsymbol{y} > b^0\right) \\ \leq \ & \mathrm{P}\left(\tilde{\boldsymbol{z}}'\boldsymbol{y} > \min_{\boldsymbol{r},\boldsymbol{s} \geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|^* + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\}\right) \\ \leq \ & \mathrm{P}\left(\tilde{\boldsymbol{z}}'\boldsymbol{y} > \min_{\boldsymbol{r},\boldsymbol{s} \geq \boldsymbol{0}} \{\Omega\|\boldsymbol{t}(\boldsymbol{r},\boldsymbol{s})\|_2 + \boldsymbol{r}'\bar{\boldsymbol{z}} + \boldsymbol{s}'\underline{\boldsymbol{z}}\}\right), \end{aligned}$$

where $y_j = \boldsymbol{\Delta a}^{j\prime}\boldsymbol{x} - \Delta b^j$ and

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

Let

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

and $\boldsymbol{t}^* = \boldsymbol{t}(\boldsymbol{r}^*, \boldsymbol{s}^*)$. Observe that since $-\underline{z}_j \leq \tilde{z}_j \leq \bar{z}_j$, we have $r_j^*\bar{z}_j \geq r_j^*\tilde{z}_j$ and $s_j^*\underline{z}_j \geq -s_j^*\tilde{z}_j$. Therefore,

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

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

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

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, $\|\boldsymbol{z}\| = \max\left\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\right\}$, to achieve the same bound, but the approximation may not be worthwhile. Notice, from the inequality (12), the value $\Omega\|\boldsymbol{t}\|^*$ gives the desired "safety distance" against constraint violation. Since $\|\boldsymbol{t}\|^* \geq \|\boldsymbol{t}\|_2$, one way to compare the conservativeness of different norms is through the following worst case ratio

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

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 [10] and Bertsimas et al. [8]). 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 [4] and Bertsimas and Sim [10]. 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, as 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

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

Proof : From Theorem 4, we have

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

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

$$\mathcal{D}_\Omega = \left\{ (\boldsymbol{a}, b) \ : \ \exists \boldsymbol{z} \in \Re^N, \quad (\mathrm{a}, b) = (\mathrm{a}^0, b^0) + \sum_{j=1}^{N} (\boldsymbol{\Delta a}^j, \Delta b^j) z_j \ , \quad \sqrt{\sum_{j=1}^{N} \frac{z_j^2}{d_j^2}} \le \Omega \right\},$$

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

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

and

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

Clearly, $(\boldsymbol{a}^*, b^*) \in \mathcal{W}$ and that $|y_j| \ge d_j$. Observe that,

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

Hence, $(\boldsymbol{a}^*, b^*) \notin \mathcal{D}_\Omega \supseteq \mathcal{A}_\Omega$ 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 instance, Kibzun and Kan [29], Birge and Louveaux [14] and Pintér [31]). Those approximations usually assume that the mean, variance and/or support are known, while 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 Appendix D 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 instance, Kibzun and Kan [29]), 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. This point is elaborated in the following subsection.

4.3 Comparison of approximation scheme based on forward/backward deviations with 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,

$$\mathrm{P}(\tilde{z} > \Lambda\sigma) \leq 1/(\Lambda^2 + 1), \tag{23}$$

while the bound provided by the forward deviation is

$$\mathrm{P}(\tilde{z} > \Omega p^*) \leq \exp\left(-\Omega^2/2\right). \tag{24}$$

For the same constraint violation probability, ϵ, bound (23) suggests $\Lambda = \sqrt{\frac{1-\epsilon}{\epsilon}}$, while bound (24) requires $\Omega = \sqrt{-2\ln(\epsilon)}$. Since 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

[1] The bound (23) is tight for the centered Bernoulli distribution of (19) 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 the bound (24) is also asymptotically tight.

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 the bound (23) to perform poorly as compared to (24). Furthermore, since 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 (19), 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 the bound (24). 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 (19), 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,

$$\begin{aligned} Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} + \mathrm{E}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} \quad & \boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}(\tilde{\boldsymbol{z}}) \le f_i(\tilde{\boldsymbol{z}}), \text{ a.e.}, \qquad \forall i \in \{1, \ldots, m\}, \\ & \boldsymbol{x} \in \Re^{n_1}, \\ & \boldsymbol{y}(\cdot) \in Y, \end{aligned} \tag{25}$$

where $\boldsymbol{x}$ corresponds to the first stage decision vector, and $\boldsymbol{y}(\tilde{\boldsymbol{z}})$ is the recourse function from a space of measurable functions, Y, with domain $\mathcal{W}$ 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, $\boldsymbol{x}$, such that there exists a feasible recourse for any realization of the uncertainty may be intractable (see Ben-Tal et al. [5] and Shapiro and Nemirovski [34]). Nevertheless, in some applications of stochastic optimization, the risk of infeasibility often can be tolerated as a tradeoff 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 [30] and Ergodan and Iyengar [25]:

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}(\tilde{\boldsymbol{z}}) \leq f_i(\tilde{\boldsymbol{z}})\right) \geq 1 - \epsilon_i \quad \forall i \in \{1, \ldots, m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}(\cdot) \in Y,
\end{aligned}
\tag{26}
$$

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, $\boldsymbol{x}$ and $\boldsymbol{y}(\cdot)$. 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 Section 6. Second, incorporating a linear second stage cost into model (26) with chance constraints introduces an interesting modeling issue. That is, since 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 the present paper we will not include the second stage cost in the model. We refer the readers to our companion paper [20] for a more general multi-stage 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 $p_j \in \mathcal{P}(\tilde{z}_j)$ and $q_j \in \mathcal{Q}(\tilde{z}_j)$. For all $i \in \{1, \ldots, m\}$, under the *Affine Data Perturbation*, we have

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

and

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

To design a tractable robust optimization approach for solving (26), we restrict the recourse function $\boldsymbol{y}(\cdot)$ to one of the *linear decision rules* as follows,

$$\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum_{j=1}^{N} \boldsymbol{y}^j z_j. \tag{27}$$

Linear decision rules emerged in the early development of stochastic optimization (see Garstka and Wets [27]) and reappeared recently in the *affinely adjustable robust counterpart* introduced by Ben-Tal et al. [5]. The linear decision rule enables one to design a tractable robust optimization approach for finding feasible solutions in the model (26) for all distributions satisfying the *Model of Data Uncertainty, U.*

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

$$
\begin{aligned}
Z_r^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{a}_i^{0\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^0 + \Omega_i h_i + \boldsymbol{r}^{i\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{i\prime}\underline{\boldsymbol{z}} \le f_i^0 \quad \forall i \in \{1,\dots,m\} \\
& \|\boldsymbol{u}^i\|^* \le h_i \quad \forall i \in \{1,\dots,m\} \\
& u_j^i \ge p_j(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall i \in \{1,\dots,m\}, j \in \{1,\dots,N\} \\
& u_j^i \ge -q_j(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j - r_j^i + s_j^i) \quad \forall i \in \{1,\dots,m\}, j \in \{1,\dots,N\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}^j \in \Re^k \quad \forall j \in \{0,\dots,N\} \\
& \boldsymbol{u}^i, \boldsymbol{r}^i, \boldsymbol{s}^i \in \Re_+^N, h_i \in \Re \quad \forall i \in \{1,\dots,m\},
\end{aligned}
\tag{28}
$$

where $\Omega_i = \sqrt{-2\ln(\epsilon_i)}$ is feasible in the stochastic optimization model (26) for all distributions that satisfy the Model of Data Uncertainty, U and $Z_r^* \ge Z^*$.

Proof : Restricting the space of recourse solutions $\boldsymbol{y}(\boldsymbol{z})$ in the form of Eq. (27), we have the following problem,

$$
\begin{aligned}
Z_1^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i^{0\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^0 + \textstyle\sum_{j=1}^N \left(\boldsymbol{\Delta a}_i^{j\prime}\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}^j - \Delta f_i^j\right)\tilde{z}_j \le f_i^0\right) \ge 1 - \epsilon_i \quad \forall i \in \{1,\dots,m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}^j \in \Re^k \quad \forall j \in \{0,\dots,N\},
\end{aligned}
\tag{29}
$$

which gives an upper bound to the model (26). Applying Theorem 5 and using Theorem 2, the feasible solution of the model (28) is also feasible in the model (29) for all distributions that satisfy the *Model of Data Uncertainty, U*. Hence, $Z_r^* \ge Z_1^* \ge Z^*$. ■

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

$$
\begin{aligned}
Z^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \mathrm{P}\left(\boldsymbol{a}_i(\tilde{\boldsymbol{z}}_1,..,\tilde{\boldsymbol{z}}_T)'\boldsymbol{x} + \textstyle\sum_{t=1}^T \boldsymbol{b}_{it}'\boldsymbol{y}_t(\tilde{\boldsymbol{z}}_1,..,\tilde{\boldsymbol{z}}_t) \le f_i(\tilde{\boldsymbol{z}}_1,..,\tilde{\boldsymbol{z}}_T)\right) \ge 1 - \epsilon_i \quad \forall i \in \{1,\dots,m\} \\
& \boldsymbol{x} \in \Re^n, \\
& \boldsymbol{y}_t(\boldsymbol{z}_1,..\boldsymbol{z}_t) \in \Re^k \quad \forall t = 1,..T, \underline{\boldsymbol{z}}_t \le \boldsymbol{z}_t \le \bar{\boldsymbol{z}}_t,
\end{aligned}
\tag{30}
$$

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

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

and

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

In order to derive the robust formulation of the multi-period model, we use the following linear decision rule for the recourse function,

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

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."

On linear decision rules

The linear decision rule is the key enabling mechanism that permits scalability to multi-stage models. It has appeared in earlier proposals for solving stochastic optimization problems (see, for instance, Charnes and Cooper [18] and Charnes et al. [19]). However, due to its perceived limitations, the method was short-lived (see Garstka and Wets [27]). 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 $y(\tilde{\boldsymbol{z}}) \geq 0$, can nullify any benefit of linear decision rules on the recourse function, $y(\tilde{\boldsymbol{z}})$. As an illustration, consider the following hard constraint

$$\begin{aligned} y(\tilde{\boldsymbol{z}}) &\geq 0 \\ y(\tilde{\boldsymbol{z}}) &\geq b(\tilde{\boldsymbol{z}}) = b_0 + \textstyle\sum_{j=1}^{N} b_j \tilde{z}_j, \end{aligned} \qquad (31)$$

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

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

is not feasible for the constraints (31).

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

$$\mathrm{P}(y(\tilde{\boldsymbol{z}}) \geq 0) \leq 1 - \epsilon$$
$$\mathrm{P}(y(\tilde{\boldsymbol{z}}) \geq b(\tilde{\boldsymbol{z}})) \leq 1 - \epsilon \ ,$$

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

$$y_0 \geq \Omega \|[y_1, \ldots, y_N]\|_2,$$
$$y_0 - b_0 \geq \Omega \|[y_1 - b_1, \ldots, y_N - b_N]\|_2,$$

where $\Omega = 5.68$. Since $\Omega = \sqrt{-2\ln(\epsilon)}$ 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 [27], the optimal policy is given by a linear decision rule only under very restrictive assumptions. However, Shapiro and Nemirovski [34] 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 [5, 6]), as will be seen in the project management example presented in the next.

6 Application Example: Project Management under Uncertain Activity Time

Project management problems can be represented by a directed graph with m arcs and n nodes. Each node on the graph represents an event marking the completion of a particular subset of activities. We

[2] An affine decision rule is equivalent to a linear decision rule in our context.

denote the set of directed arcs on the graph as $\mathcal{E}$. Hence, an arc $(i,j) \in \mathcal{E}$ is an activity that connects event i to event j. By convention, we use node 1 as the start event and the last node n as the end event.

We consider a project with several activities. Each activity has a random completion time, $\tilde{t}_{ij}$. The completion of activities must satisfy precedent constraints. For example, activity e_1 precedes activity e_2 if activity e_1 must be completed before starting activity e_2. Analysis of stochastic project management problems, such as determining the expected completion time and quantile of completion time, is notoriously difficult (Hagstrom [28]).

In our computational experiment, we assume that the random completion time, $\tilde{t}_{ij}$, depends on some additional amount of resource, $x_{ij} \in [0, \bar{x}_{ij}]$, committed to the activity as follows

$$\tilde{t}_{ij} = (1 + \tilde{z}_{ij})b_{ij} - a_{ij}x_{ij}, \tag{32}$$

where $\tilde{z}_{ij} \in [-\underline{z}_{ij}, \bar{z}_{ij}]$, $\underline{z}_{ij} \leq 1$, $(i,j) \in \mathcal{E}$ are independent random variables with means zero and deviation measures (p_{ij}, q_{ij}) satisfying $p_{ij} \in \mathcal{P}(\tilde{z}_{ij})$ and $q_{ij} \in \mathcal{Q}(\tilde{z}_{ij})$. We also assume that $\tilde{t}_{ij} \geq 0$ for all realizations of $\tilde{z}_{ij}$ and valid ranges of x_{ij}. We note that the assumption of independent activity completion times can be rather strong and difficult to verify in practice. Nevertheless, we must specify some form of independence in order to enjoy the benefit of risk pooling; otherwise, the solution would be a conservative one. We emphasize that the model easily can be extended to include linear dependencies of activity completion times, such as sharing common resources with independent failure probabilities.

Let c_{ij} be the cost of using each unit of resource for the activity on the arc (i,j). Our goal is to find a resource allocation to each activity $(i,j) \in \mathcal{E}$, such that the total project cost is minimized, while ensuring that the probability of completing the project within time T is at least $1 - \epsilon$.

6.1 Formulation of the project management problem

We first consider the "hard constrained" case, in which the project has to be finished within time T. It may be formulated as a a two stage stochastic program as follows.

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & y_n(\tilde{\boldsymbol{z}}) \leq T \\ & y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \geq (1 + \tilde{z}_{ij})b_{ij} - a_{ij}x_{ij} \quad \forall (i,j) \in \mathcal{E} \\ & y_1(\tilde{\boldsymbol{z}}) = 0 \\ & \boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}} \\ & \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\ & \boldsymbol{y}(\boldsymbol{z}) \in \Re^n \quad \forall \underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \bar{\boldsymbol{z}}. \end{array} \tag{33}$$

Variables $y_j(\boldsymbol{z})$ represent the completion time of event j, when the realization of the uncertainty is $\boldsymbol{z}$.

When we allow the total completion time to be longer than T with a small probability ϵ, we may revise Formulation (33) to the following model with a joint chance constraint.

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \mathrm{P}\left(\begin{array}{l} y_n(\tilde{\boldsymbol{z}}) \leq T \\ y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \geq (1+\tilde{z}_{ij})b_{ij} - a_{ij}x_{ij} \quad \forall (i,j) \in \mathcal{E} \\ y_1(\tilde{\boldsymbol{z}}) = 0 \end{array}\right) \geq 1-\epsilon \\ & \mathbf{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}} \\ & \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\ & \boldsymbol{y}(\boldsymbol{z}) \in \Re^n \quad \forall \underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \bar{\boldsymbol{z}}. \end{array} \tag{34}$$

Notice that for any uncertainty realization, $\boldsymbol{z}$, that satisfies all the constraints, $y_j(\boldsymbol{z})$ still represents the completion time of event j. The feasibility of the constraints, therefore, indicates that the project is completed within time T.

Problems with joint chance constraints are considerably harder to solve compared to separable single chance constraints. Fortunately, we can approximate the joint chance constraint using separate single chance constraints through Bonferroni's inequality. The following proposition provides the basis for such an approximation in the project management problem.

Proposition 5 *For any ϵ_0 and ϵ_{ij}, $\forall (i,j) \in \mathcal{E}$ in $(0,1)$ such that*

$$\epsilon_0 + \sum_{(i,j)\in\mathcal{E}} \epsilon_{ij} \leq \epsilon \ , \tag{35}$$

if there exists a measurable function $y_i(\tilde{z})$ for every node i, satisfying

$$\begin{array}{l} \mathrm{P}\left(y_n(\tilde{\boldsymbol{z}}) \leq T\right) \geq 1-\epsilon_0 \\ \mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \geq \tilde{t}_{ij}\Big) \geq 1-\epsilon_{ij} \quad \forall (i,j) \in \mathcal{E} \\ y_1(\tilde{\boldsymbol{z}}) = 0 \ , \end{array}$$

the probability that the project is completed within time T is at least $1-\epsilon$.

Proof : For any realization $\boldsymbol{z}$ of $\tilde{\boldsymbol{z}}$, the project can be finished within time period T if and only if there exists y_i for each node i such that the following inequalities are satisfied

$$\begin{array}{rcll} y_n & \leq & T & \\ y_1 & = & 0 & \\ y_j & \geq & y_i + t_{ij}(\boldsymbol{z}) & \forall (i,j) \in \mathcal{E}. \end{array}$$

Since $y_1(\tilde{\boldsymbol{z}}) = 0$, we have

$$
\begin{aligned}
& \mathrm{P}\,(\text{ Project finished within } T \text{ }) \\
\geq\ & \mathrm{P}\left(\{y_n(\tilde{\boldsymbol{z}}) \leq T\} \bigcap_{(i,j)\in\mathcal{E}} \{y_j(\tilde{\boldsymbol{z}}) \geq y_i(\tilde{\boldsymbol{z}}) + \tilde{t}_{ij}\}\right) \\
=\ & 1 - \mathrm{P}\left(\{y_n(\tilde{\boldsymbol{z}}) > T\} \bigcup_{(i,j)\in\mathcal{E}} \{y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) < \tilde{t}_{ij}\}\right) \\
\geq\ & 1 - \left(\mathrm{P}\Big(y_n(\tilde{\boldsymbol{z}}) > T\Big) + \sum_{(i,j)\in\mathcal{E}} \mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) < \tilde{t}_{ij}\Big)\right) \quad \text{(Bonferroni inequality)} \\
\geq\ & 1 - \left(\epsilon_0 + \sum_{(i,j)\in\mathcal{E}} \epsilon_{ij}\right) \\
\geq\ & 1 - \epsilon\ .
\end{aligned}
$$

■

Therefore, we have the following stochastic optimization model for the project management problem:

$$
\begin{aligned}
Z^* = \min\quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.}\quad & \mathrm{P}\,(y_n(\tilde{\boldsymbol{z}}) \leq T) \geq 1 - \epsilon_0 \\
& \mathrm{P}\Big(y_j(\tilde{\boldsymbol{z}}) - y_i(\tilde{\boldsymbol{z}}) \geq (1 + \tilde{z}_{ij})b_{ij} - a_{ij}x_{ij}\Big) \geq 1 - \epsilon_{ij} \quad \forall (i,j) \in \mathcal{E} \\
& y_1(\tilde{\boldsymbol{z}}) = 0 \\
& \boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}} \\
& \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\
& \boldsymbol{y}(\boldsymbol{z}) \in \Re^n \quad \forall \underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \bar{\boldsymbol{z}}.
\end{aligned}
\tag{36}
$$

In the above model, $y_i(\boldsymbol{z})$ corresponds to the completion time of event i whenever the uncertainty realization $\boldsymbol{z}$ is such that all the constraints are satisfied with certainty according to the feasible solution x and $y_i(\boldsymbol{z})$. Notice that Formulation (36) is an approximation of the original joint chance constrained problem. It is not the only way of approximating the joint chance constraint. Depending on specific problem structures in other applications, one may obtain less conservative approximations. The advantage of the above approach is that it is easy to construct and compute. We will test its conservativeness in the computational study. We defer the discussion of how to choose ϵ_0 and ϵ_{ij} for $(i,j) \in \mathcal{E}$ until after we provide a further, computationally tractable, approximation of Model (36).

Applying Theorem 7, we further approximate Model (36) using a Second Order Cone Program as follows

$$
\begin{aligned}
Z_r^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & y_n^0 + \Omega_0 h_0 + \boldsymbol{r}^{0\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{0\prime}\underline{\boldsymbol{z}} \le T \\
& \|\boldsymbol{u}^0\|_2 \le h_0 \\
& u_{ij}^0 \ge p_{ij}(y_n^{ij} - r_{ij}^0) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^0 \ge -q_{ij}(y_n^{ij} + s_{ij}^0) \quad \forall (i,j) \in \mathcal{E} \\
& y_j^0 - y_i^0 \ge b_{ij} - a_{ij}x_{ij} + \Omega_{ij}h_{ij} + \boldsymbol{r}^{ij\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{ij\prime}\underline{\boldsymbol{z}} \quad \forall (i,j) \in \mathcal{E} \\
& \|\boldsymbol{u}^{ij}\|_2 \le h_{ij} \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{ij} \ge p_{ij}(b_{ij} + y_i^{ij} - y_j^{ij} - r_{ij}^{ij} + s_{ij}^{ij}) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{kl} \ge p_{ij}(y_k^{ij} - y_l^{ij} - r_{ij}^{kl} + s_{ij}^{kl}) \quad \forall (i,j),(k,l) \in \mathcal{E}, (i,j) \ne (k,l) \\
& u_{ij}^{ij} \ge -q_{ij}(b_{ij} + y_i^{ij} - y_j^{ij} - r_{ij}^{ij} + s_{ij}^{ij}) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{kl} \ge -q_{ij}(y_k^{ij} - y_l^{ij} - r_{ij}^{kl} + s_{ij}^{kl}) \quad \forall (i,j),(k,l) \in \mathcal{E}, (i,j) \ne (k,l) \\
& y_1^0 = 0, y_1^{ij} = 0 \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{0} \le \boldsymbol{x} \le \bar{\boldsymbol{x}} \\
& \boldsymbol{u}^0, \boldsymbol{u}^{ij}, \boldsymbol{r}^0, \boldsymbol{r}^{ij}, \boldsymbol{s}^0, \boldsymbol{s}^{ij} \in \Re_+^{|\mathcal{E}|} \quad \forall (i,j) \in \mathcal{E} \\
& h_0, h_{ij} \in \Re \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\
& \boldsymbol{y}^0, \boldsymbol{y}^{ij} \in \Re^n \quad \forall (i,j) \in \mathcal{E}.
\end{aligned}
\tag{37}
$$

Choosing ϵ_0 and ϵ_{ij} for $(i,j) \in \mathcal{E}$ in Model (36) becomes choosing Ω_0 and Ω_{ij} for all $(i,j) \in \mathcal{E}$ in Model (37). The selection of Ω_0 and Ω_{ij}, such that the probability of completing the project in timely fashion is at least $1-\epsilon$, is not unique. One sufficient condition is

$$\exp(-\Omega_0^2/2) + \sum_{(i,j)\in\mathcal{E}} \exp(-\Omega_{ij}^2/2) \le \epsilon \ . \tag{38}$$

We suggest choosing an equal budget allocation, i.e.,

$$\Omega_0 = \Omega_{ij} = \underbrace{\sqrt{-2\ln\left(\frac{\epsilon}{|\mathcal{E}|+1}\right)}}_{=\Omega}, \quad \forall (i,j) \in \mathcal{E}. \tag{39}$$

For a particular problem instance, there could be an allocation of budgets that is less conservative than the equal budget allocation scheme. However, we claim that the improvement would be marginal. To see this, we notice that the inequality (38) implies that $\Omega_0, \Omega_{ij} \ge \sqrt{-2\ln(\epsilon)}$ for each $(i,j) \in \mathcal{E}$. The

m	$\rho(m)$
1000	1.414
10000	1.528
100000	1.633

Table 3: Conservative measure with $\epsilon = 0.001$.

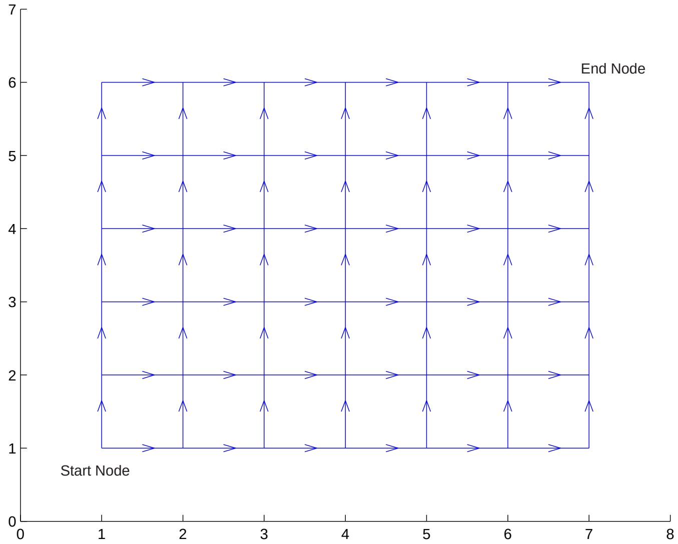

Figure 4: Project management "grid" with height, $H = 6$ and width $W = 7$.

ratio of our suggested value to the smallest possible value, $\sqrt{-2\ln(\epsilon)}$, is rather small even in problems with relatively large $|\mathcal{E}|$. As an example, we demonstrate in Table 3 that the ratio[3]

$$\rho(m) = \frac{\sqrt{-2\ln(\epsilon/m)}}{\sqrt{-2\ln(\epsilon)}}$$

grows slowly with $m \triangleq |\mathcal{E}|$.

6.2 Computation results

For our computational experiment, we create a fictitious project with the activity network in the form of a H by W grid (see Figure 4). There are a total of $H \times W$ nodes, with the first node at the bottom left corner and the the last node at the upper right corner. Each arc on the graph either directs up or

[3]Ben-Tal and Nemirovski [2] proposed a similar ratio to compare the relative size of uncertainty budgets in quantifying the conservativeness of intractable robust counterparts.

H	W	m	n	Ω	Λ	Z_r^*	Z_σ^*	Z_w^*	$\frac{Z_r^*}{Z_w^*}$
6	4	38	24	4.07	62.44	759.62	950	950	0.80
8	3	37	24	4.06	61.64	485.07	925	925	0.52
10	3	47	30	4.12	69.27	520.41	1175	1175	0.44
12	3	57	36	4.16	76.15	566.46	1425	1425	0.40
14	3	67	42	4.20	82.46	611.2	1675	1675	0.36

Table 4: Computation results for project management.

to the right. We assume that every activity on arc has independent and identical completion time. In particular, for all arcs (i, j),

$$\mathrm{P}(\tilde{z}_{ij} = z) = \begin{cases} 0.9 & \text{if } z = -25/900 \\ 0.1 & \text{if } z = 25/100 \ . \end{cases}$$

Hence, $\underline{z}_{ij} = 25/900$, $\bar{z}_{ij} = 25/100$ and we can determine numerically that the standard, forward and backward deviations are $\sigma_{ij} = 0.0833$, $p_{ij} = 0.1185$ and $q_{ij} = 0.0833$, respectively. Note that for the above two point distribution, the upper bounds of the deviation measures provided in Theorem 4 are tight. We further assume that for all activities (i, j), $a_{ij} = c_{ij} = 1$, $\bar{x}_{ij} = 25$ and $b_{ij} = 100$.

We also want high confidence (at least 99%) that the completion time of the project is no more than $100(H + W - 2)$, which is the average completion time of any path with $x_{ij} = 0$ on all arcs. Therefore, additional resources are needed to meet the desired reliability level of project completion time.

Table 4 summarizes the comparison of our approach (Z_r^*) with the worst case approach (Z_w^*) and an approach based on standard deviation (Z_σ^*).

In the worst case scenario, the delay happens to every activity, that is, $\tilde{z}_{ij} = 0.25$ and $(1 + \tilde{z}_{ij})b_{ij} = 125$ for all (i, j). In this case, each activity on arc must be assigned to the maximum resource at $x_{ij} = 25$, so that $\tilde{t}_{ij} = 100$ and no critical path (longest paths) is longer than $100(H + W - 2)$. Since there are a total of $m = H(W - 1) + W(H - 1)$ arcs on the H by W grid, the optimal objective function value according to the worst case scenario is $Z_w^* = 25m$. This value is reflected in the column Z_w^* of Table 4.

The values Z_σ^* are calculated according to the following model, derived from Bertsimas et al. ([8]), and using the linear decision rule.

$$
\begin{aligned}
Z_\sigma^* = \min \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & y_n^0 + \Lambda h_0 + \boldsymbol{r}^{0\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{0\prime}\underline{\boldsymbol{z}} \leq T \\
& \|\boldsymbol{u}^0\|_2 \leq h_0 \\
& u_{ij}^0 = \sigma_{ij}(y_n^{ij} - r_{ij}^0) \quad \forall (i,j) \in \mathcal{E} \\
& y_j^0 - y_i^0 \geq b_{ij} - a_{ij}x_{ij} + \Lambda h_{ij} + \boldsymbol{r}^{ij\prime}\bar{\boldsymbol{z}} + \boldsymbol{s}^{ij\prime}\underline{\boldsymbol{z}} \quad \forall (i,j) \in \mathcal{E} \\
& \|\boldsymbol{u}^{ij}\|_2 \leq h_{ij} \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{ij} = \sigma_{ij}(b_{ij} + y_i^{ij} - y_j^{ij} - r_{ij}^{ij} + s_{ij}^{ij}) \quad \forall (i,j) \in \mathcal{E} \\
& u_{ij}^{kl} = \sigma_{ij}(y_k^{ij} - y_l^{ij} - r_{ij}^{kl} + s_{ij}^{kl}) \quad \forall (i,j),(k,l) \in \mathcal{E}, (i,j) \neq (k,l) \\
& y_1^0 = 0, y_1^{ij} = 0 \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}} \\
& \boldsymbol{u}^0, \boldsymbol{u}^{ij}, \boldsymbol{r}^0, \boldsymbol{r}^{ij}, \boldsymbol{s}^0, \boldsymbol{s}^{ij} \in \Re_+^{|\mathcal{E}|} \quad \forall (i,j) \in \mathcal{E} \\
& h_0, h_{ij} \in \Re \quad \forall (i,j) \in \mathcal{E} \\
& \boldsymbol{x} \in \Re^{|\mathcal{E}|} \\
& \boldsymbol{y}^0, \boldsymbol{y}^{ij} \in \Re^n \quad \forall (i,j) \in \mathcal{E},
\end{aligned}
\tag{40}
$$

where

$$
\Lambda = \sqrt{\frac{1 - \frac{\epsilon}{|\mathcal{E}|+1}}{\frac{\epsilon}{|\mathcal{E}|+1}}} = \sqrt{\frac{|\mathcal{E}|+1}{\epsilon} - 1} \ .
$$

The goal is to bound the constraint violation using probability inequalities based on the standard deviation.

We solved the optimization models in Table 4 using both SDPT3 [36] and also CPLEX 9.1 on a Pentium Xenon machine, with 4 gigabytes of RAM. All the models were solved to optimality within a few minutes.

In all the cases computed in Table 4, the standard deviation based approach, Model (40), performed as conservatively as the worst case solution. This is because the parameter Λ in Model (40) is far greater than the Ω in Model (37) in these problems. At the same time, the standard deviations of the primitive uncertainties are close to the corresponding forward and backward deviation measures. Hence, it is not surprising that Model (40) performs poorly. On the other hand, the reason that Λ is so large in this model is that the number of chance constraints is close to the number of primitive uncertainties. In other settings where the number of chance constraints is much smaller and ϵ is not too small, we expect the standard deviation based approach to yield results different from the worst case approach.

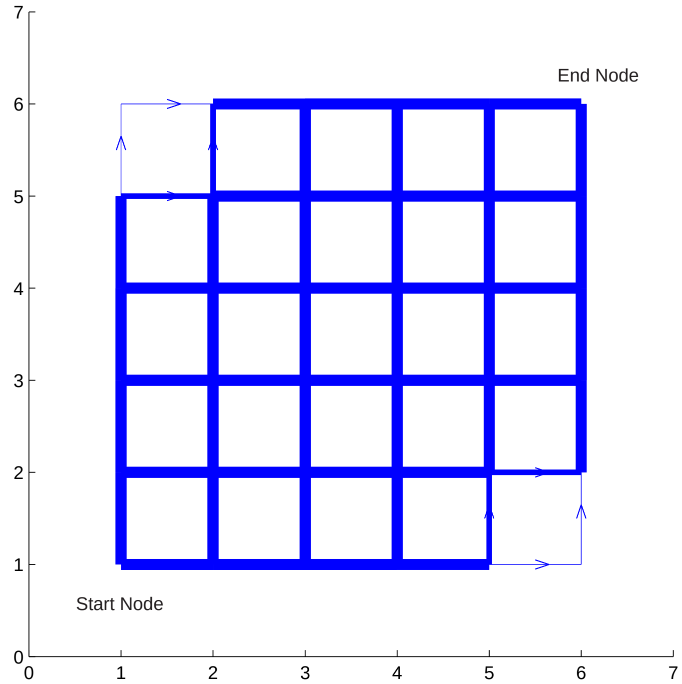

Figure 5: Project management solution for $H = 6$ and $W = 6$.

Interestingly, the relative savings from using the robust model, compared with the worst case scenario, depend on the activity network topology. In Figure 5, we illustrate the optimal resource allocation to each activity, by varying the thickness of the arcs on the grid. Except for the upper-left and lower-right activities, almost all other activities require all available resources. Since the activities are tightly connected, any delay in one activity may lead to a delay in the entire project. Hence, there is little leeway for cost savings without having to compromise the completion time.

In Figure 6, we illustrate the solution on the grid of a rectangular activity graph. Recall that the worst case solution corresponds to allocating all resources across all arcs. Figure 6 indicates that when the activity network is more rectangular, the cost savings is much more substantial. This suggests that if there are more sequential activities and fewer parallel activities, robust project management can yield better cost savings, while guaranteeing a high probability that the project will be completed on time.

6.3 Comparison with the sampling approach

Under discrete distributions, the exact solution of the two stage chance constraint problem can be formulated as an MIP model with perhaps an exponential number of binary variables. As an example, in a small, 6 by 4 grid, even for the two point distribution mentioned above, the number of binary

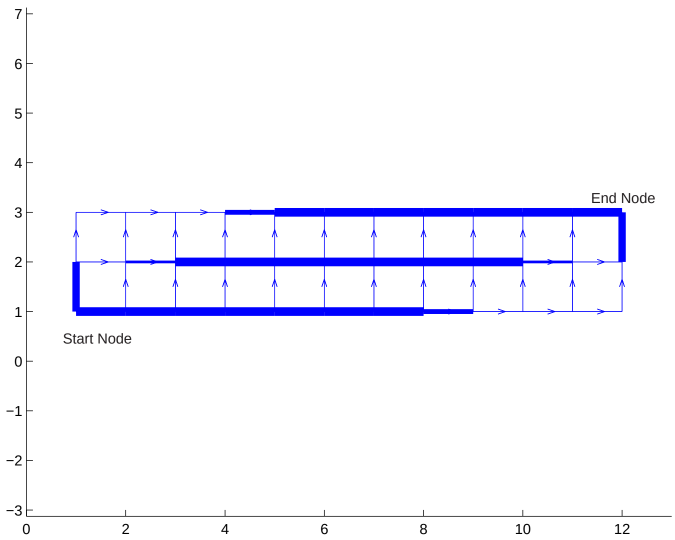

Figure 6: Project management solution for $H = 3$ and $W = 12$.

variables is $2^{38} \approx 2.75 \times 10^{11}$. As such, we could only compare our approach with the scenario based solution approach as represented in the following model, assuming that we have $k = 1, \ldots, K$ random samples of the uncertainty vector $\tilde{\boldsymbol{z}}$.

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & y_n^k \le T \\ & y_j^k - y_i^k \ge (1 + z_{ij}^k) b_{ij} - a_{ij} x_{ij} \quad \forall (i,j) \in \mathcal{E}, \; k = 1, \ldots K \\ & y_1^k = 0 \\ & \boldsymbol{0} \le \boldsymbol{x} \le \bar{\boldsymbol{x}} \;. \end{aligned} \tag{41}$$

We will refer to Problem (41) as the "sampling approach". A body of work on using the sampling approach to approximate chance constrained problems has recently appeared. Calafiore and Campi [16], for example, provide a theoretical bound on the sample size, K, such that a convex chance constraint is satisfied with high probability. They consider the following problem

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} \\ & \mathrm{P}(f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \ge 0) \ge 1 - \epsilon \\ & \boldsymbol{x} \in X, \end{aligned} \tag{42}$$

where $\boldsymbol{X} \subseteq \Re^n$ is a convex set, while $f(\boldsymbol{x}, \tilde{\boldsymbol{z}})$ is a concave function in $\boldsymbol{x}$ for any realization $\tilde{\boldsymbol{z}}$ of $\tilde{\boldsymbol{z}}$. They show that if $\boldsymbol{x}$ is feasible in

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} \\ & f(\boldsymbol{x}, \boldsymbol{z}^i) \geq 0, i = 1, \ldots, K \\ & \boldsymbol{x} \in X, \end{aligned} \tag{43}$$

where $\boldsymbol{z}^i$, $i = 1, \ldots, K$ are independent samples of $\tilde{\boldsymbol{z}}$, with the sample size

$$K \geq \frac{n}{\epsilon\delta} - 1,$$

then $\boldsymbol{x}$ is feasible for the chance constraint in Model (42) with a probability of at least $1-\delta$. By choosing $X = \{\boldsymbol{x} : \boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}}\}$ and

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{z}) = \min_{\boldsymbol{y}, w} \quad & w \\ & T - y_n \geq w \\ & y_j - y_i - (1 + z_{ij})b_{ij} - a_{ij}x_{ij} \geq w \quad \forall (i,j) \in \mathcal{E} \\ & y_1 = 0, \end{aligned}$$

we can easily extend the result to our context; if

$$K \geq \frac{|\mathcal{E}|}{\epsilon\delta} - 1 \ ,$$

an optimal solution to Model (41) is feasible for the chance constraint in Model (34) with a probability of at least $1-\delta$. Since our problem is a two stage stochastic optimization problem with chance constraints, we are unable to the use stronger complexity results, such as those in de Farias and Van Roy [21] and Calafiore and Campi [15], which analyze sampling approaches for single stage problems (see the discussion in Ergodan and Iyengar [24]).

The use of finite samples precludes comparing the sampling approach with the robust optimization approach, as the latter ensures feasibility of the chance constraints with certainty. Instead, for the sampling approach, we generate random samples of varying sizes and solve Model (41) to obtain a series of optimal solutions. After obtaining each solution, we estimated the probability of constraint violation, i.e., the completion time being longer than T, using a separate set of testing samples with $500,000$ random scenarios. For the robust approach, Model (37), we also varied the Ω level to obtain a series of optimal solutions. Then we evaluated the probabilities of constraint violation using the same set of testing samples. We also compared the objective function values yielded by the above implementations of the two approaches.

Robust approach			Sampling approach		
Ω	Objective	Probability of Constr. Violation	Samples	Objective	Probability of Constr. Violation
0	0	0.9821	100	311.11	0.0588
1.199	125.90	0.3395	200	283.33	0.0642
1.681	202.12	0.0721	500	313.89	0.0405
2.052	267.65	0.0262	1000	425.00	0.0063
2.366	325.12	0.0091	*1000*	*331.48*	*0.0275*
2.643	382.63	0.0043	2000	377.78	0.0068
2.893	444.25	0.0018	5000	416.67	0.005
3.338	560.57	0.0002	10000	488.89	0.0018

Table 5: Comparison between the robust approach and the sampling approach on a 6×4 grid network with $T = 800$.

Table 5 summarizes the comparison of the objective function values and the probabilities of constraint violation. From the table we see that while the probability of constraint violation following the robust approach monotonically decreases with increasing Ω, the same probability from the sampling approach behaves more irregularly. To highlight the point, in the table we present the sampling approach from two sets of random scenarios with the same size, 1000. We observe that they generate different objective function values and probabilities of constraint violation.

To better compare the results in Table 5, we plot the objective values and the estimated probabilities of constraint violation in Figure 7. It is clear from the figure that the efficient frontier from the robust approach dominates the frontier from the sampling approach. This observation is rather surprising since in the robust approach we introduced various levels of approximations in order to achieve tractability. We have no explanation for this finding.

Another interesting observation in Figure 7 is that the trade-off between the objective function value and the log of the constraint violation probability appears to be linear. If it could be shown through formal analysis that this trade-off is necessarily linear, it would suggest that the robust model provides a very efficient way of obtaining the efficient frontier.

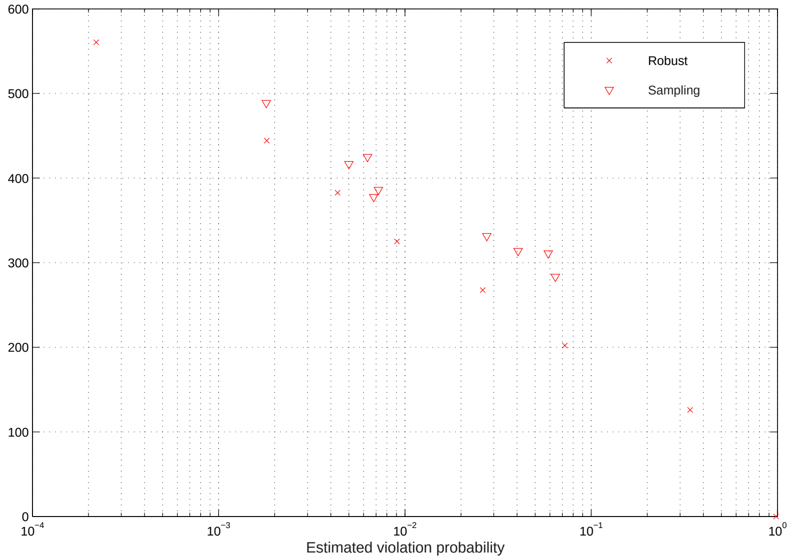

Figure 7: Comparison of the robust approach to the sampling approach. Plot for Table 5

7 Conclusions

The new deviation measures enable 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. 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.

Acknowledgements

We would like to thank the reviewers of the paper for several insightful comments.

A Proof of Proposition 2

Let

$$X = \{\boldsymbol{u} : \|\boldsymbol{u}\| \leq \Omega\}$$

and

$$Y = \left\{ \boldsymbol{u} : \exists \boldsymbol{v}, \boldsymbol{w} \in \Re^N, \boldsymbol{u} = \boldsymbol{v} - \boldsymbol{w}, \|\boldsymbol{v} + \boldsymbol{w}\| \le \Omega, \boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0} \right\}.$$

It suffices to show that $X = Y$. For every $\boldsymbol{u} \in X$, let

$$(v_j, w_j) = \begin{cases} (u_j, 0) & \text{if } u_j \ge 0 \\ (0, -u_j) & \text{if } u_j < 0 \end{cases}$$

Clearly, $\boldsymbol{v}, \boldsymbol{w} \ge \boldsymbol{0}$ and $v_j + w_j = |u_j|$ for all $j = 1, \ldots, N$. Since the norm is absolute, we have

$$\|\boldsymbol{v} + \boldsymbol{w}\| = \||\boldsymbol{u}|\| = \|\boldsymbol{u}\| \le \Omega,$$

hence, $X \subseteq Y$. Conversely, suppose $\boldsymbol{u} \in Y$ and hence, $u_j = v_j - w_j$, for all $j = 1, \ldots, N$. Clearly

$$|u_j| \le v_j + w_j.$$

Therefore, since the norm is absolute, we have

$$\|\boldsymbol{u}\| = \||\boldsymbol{u}|\| \le \|\boldsymbol{v} + \boldsymbol{w}\| \le \Omega,$$

hence $Y \subseteq X$. ■

B Proof of Proposition 3

Observe that

$$\begin{aligned} \Omega\|\boldsymbol{t}\|^* = \max \quad & \sum_{j=1}^{N} \max\{a_j, b_j, 0\} r_j \\ \text{s.t.} \quad & \|\boldsymbol{r}\| \le \Omega. \end{aligned} \tag{44}$$

Suppose $\boldsymbol{r}^*$ is an optimal solution to (44). For all $j \in \{1, \ldots, N\}$, let

$$\begin{aligned} v_j = w_j = 0 \quad & \text{if } \max\{a_j, b_j\} \le 0 \\ v_j = |r_j^*|, w_j = 0 \quad & \text{if } a_j \ge b_j, a_j > 0 \\ w_j = |r_j^*|, v_j = 0 \quad & \text{if } b_j > a_j, b_j > 0. \end{aligned}$$

Observe that $a_j v_j + b_j w_j \ge \max\{a_j, b_j, 0\} r_j^*$ and $w_j + v_j \le |r_j^*|$ $\forall j \in \{1, \ldots, N\}$. From Proposition 1(c) we have $\|\boldsymbol{v} + \boldsymbol{w}\| \le \|\boldsymbol{r}^*\| \le \Omega$, and thus $\boldsymbol{v}, \boldsymbol{w}$ are feasible in Problem (9), leading to

$$z^* \ge \sum_{j=1}^{N} (a_j v_j + b_j w_j) \ge \sum_{j=1}^{N} \max\{a_j, b_j, 0\} r_j^* = \Omega\|\boldsymbol{t}\|^*$$

Absolute Norms	$\|\boldsymbol{z}\|$	$\|\boldsymbol{u}\|^* \leq h$	**References**
l_2	$\|\boldsymbol{z}\|_2$	$\|\boldsymbol{u}\|_2 \leq h$	[4]
Scaled l_1	$\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1$	$\sqrt{N}u_j \leq h, \forall j \in \{1,\ldots,N\}$	[8]
l_∞	$\|\boldsymbol{z}\|_\infty$	$\sum_{j=1}^N u_j \leq h$	[8]
Scaled l_p, $p \geq 1$	$\min\{N^{\frac{1}{2}-\frac{1}{p}},1\}\|\boldsymbol{z}\|_p$	$\max\{N^{\frac{1}{p}-\frac{1}{2}},1\}\left(\sum_{j=1}^N u_j^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} \leq h$	[8]
$l_1 \cap l_\infty$ norm	$\max\{\frac{1}{\sqrt{N}}\|\boldsymbol{z}\|_1, \|\boldsymbol{z}\|_\infty\}$	$\sqrt{N}p + \sum_{j=1}^N s_j \leq h$ $s_j + p \geq u_j,\ \forall j \in \{1,\ldots,N\}$ $p \in \Re_+, \boldsymbol{s} \in \Re_+^N$	[8]

Table 6: Representation of the dual norm for $\boldsymbol{u} \geq \boldsymbol{0}$

Conversely, let $\boldsymbol{v}^*, \boldsymbol{w}^*$ be an optimal solution to Problem (9). Let $\boldsymbol{r} = \boldsymbol{v}^* + \boldsymbol{w}^*$. Clearly $\|\boldsymbol{r}\| \leq \Omega$ and observe that

$$r_j \max\{a_j, b_j, 0\} \geq a_j v_j^* + b_j w_j^* \quad \forall j \in \{1,\ldots,N\}.$$

Therefore, we have

$$\Omega\|\boldsymbol{t}\|^* \geq \sum_{j=1}^N \max\{a_j, b_j, 0\} r_j \geq \sum_{j=1}^N (a_j v_j^* + b_j w_j^*) = z^* \ .$$

■

C Table of Dual Norm Representation

Table 6 lists the common choices of absolute norms, the representation of their dual norm inequalities, $\|\boldsymbol{u}\|^* \leq h$, in which $\boldsymbol{u} \geq \boldsymbol{0}$, and the corresponding references.

D Deviation Measures for Asymmetric Distributions

Recently, general deviation measures were introduced and systematically analyzed in Rockafellar et al. [33]. The authors defined deviation measures on $\mathcal{L}^2(L, \mathcal{F}, Q)$, where L is the sample space, $\mathcal{F}$ is a field of sets in L and Q is a probability measure on $(L, \mathcal{F})$. Among the assumptions for a deviation measure, positive homogeneity and subadditivity play a fundamental role. Careful analysis of the forward and backward deviations defined in Section 3 reveals that these properties are also essential for Theorem 3 and Theorem 5, which suggests the possibility of adapting Rockafellar et al.'s [33] deviation measure definition to our setting.

For our purpose, we focus on a closed convex cone $\mathcal{S}$ of $\mathcal{L}^2(L, \mathcal{F}, Q)$. A deviation measure on the closed convex cone $\mathcal{S}$ is a functional $\mathcal{D} : \mathcal{S} \to \Re_+$ satisfying the following axioms.

(D1) $\mathcal{D}(\tilde{x} + c) = \mathcal{D}(\tilde{x})$ for any random variable $\tilde{x} \in \mathcal{S}$ and constant c.

(D2) $\mathcal{D}(0) = 0$ and $\mathcal{D}(\lambda\tilde{x}) = \lambda\mathcal{D}(\tilde{x})$ for any $\tilde{x} \in \mathcal{S}$ and $\lambda \geq 0$.

(D3) $\mathcal{D}(\tilde{x} + \tilde{y}) \leq \mathcal{D}(\tilde{x}) + \mathcal{D}(\tilde{y})$ for any $\tilde{x} \in \mathcal{S}$ and $\tilde{y} \in \mathcal{S}$.

(D4) $\mathcal{D}(\tilde{x}) > 0$ for any nonconstant $\tilde{x} \in \mathcal{S}$, whereas $\mathcal{D}(c) = 0$ for any constant c.

Standard deviation, semideviations and conditional value at risk are some examples of the above deviation measure. For more details, see Rockafellar et al. [33]. Since in our model, uncertain data are assumed to be affinely dependent on a set of independent random variables, (D3) can be relaxed. In particular, we assume

(D3') $\mathcal{D}(\tilde{x} + \tilde{y}) \leq \mathcal{D}(\tilde{x}) + \mathcal{D}(\tilde{y})$ for any independent $\tilde{x} \in \mathcal{S}$ and $\tilde{y} \in \mathcal{S}$.

In addition, to derive meaningful probability bounds against constraint violation, we prefer a deviation measure having one of the following properties.

(D5) For any random variable $\tilde{x} \in \mathcal{S}$ with mean zero, $P(\tilde{x} > \Omega\mathcal{D}(\tilde{x})) \leq f_{\mathcal{S}}(\Omega)$ for some function $f_{\mathcal{S}}$ depending only on $\mathcal{S}$ such that $f_{\mathcal{S}}(\Omega) \to 0$ as $\Omega \to \infty$.

(D5') For any random variable $\tilde{x} \in \mathcal{S}$ with mean zero, $P(\tilde{x} < -\Omega\mathcal{D}(\tilde{x})) \leq f_{\mathcal{S}}(\Omega)$ for some function $f_{\mathcal{S}}$ depending only on $\mathcal{S}$ such that $f_{\mathcal{S}}(\Omega) \to 0$ as $\Omega \to \infty$.

Assumption (D5) is associated with the upside or forward deviation while Assumption (D5') is associated with the downside or backward deviation. Of particular interests are functions with an exponential decay rate, i.e. $f_{\mathcal{S}}(\Omega) = O(e^{-\Omega})$ or $f_{\mathcal{S}}(\Omega) = O(e^{-\Omega^2/2})$.

Notice that from Theorem 3, the forward (backward) deviation defined in Section 3 satisfies Assumptions (D1), (D2), (D3'), (D4) and (D5) (or (D5')) with $f_{\mathcal{S}}(\Omega) = e^{-\Omega^2/2}$. In fact, Theorem 3 (b) is even stronger than (D3'). However, as we have already pointed out, the forward and backward deviations may not exist for some unbounded distributions. We now introduce different forward and backward deviation measures, which are well defined for more general distributions.

For a given $\kappa > 1$, define

$$\mathcal{P}_\kappa(\tilde{z}) = \left\{\alpha \; : \; \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(\frac{1}{\alpha}\right) \leq \kappa\right\}, \tag{45}$$

and

$$\mathcal{Q}_\kappa(\tilde{z}) = \left\{\alpha \ : \ \alpha \geq 0, M_{\tilde{z}-\mathrm{E}(\tilde{z})}\left(-\frac{1}{\alpha}\right) \leq \kappa\right\}. \tag{46}$$

Let

$$\hat{\mathcal{S}}_+ = \{\tilde{z} : M_{\tilde{z}}(s) < \infty \text{ for some } s > 0\},$$

and

$$\hat{\mathcal{S}}_- = \{\tilde{z} : M_{\tilde{z}}(s) < \infty \text{ for some } s < 0\},$$

It is clear that $\mathcal{P}_\kappa(\tilde{z})$ is nonempty if and only if $\tilde{z} \in \hat{\mathcal{S}}_+$ and $\mathcal{Q}_\kappa(\tilde{z})$ is nonempty if and only if $\tilde{z} \in \hat{\mathcal{S}}_-$. Note that an exponentially distributed random variable has finite derivation measure p_κ while not p^*. We now show that a result similar to Theorem 3 holds for $\mathcal{P}_\kappa(\tilde{z})$ and $\mathcal{Q}_\kappa(\tilde{z})$.

Theorem 8 *Let $\tilde{x}$ and $\tilde{y}$ be two independent random variables with zero means such that $p_{\tilde{x}} \in \mathcal{P}_\kappa(\tilde{x})$, $q_{\tilde{x}} \in \mathcal{Q}_\kappa(\tilde{x})$, $p_{\tilde{y}} \in \mathcal{P}_\kappa(\tilde{y})$ and $q_{\tilde{y}} \in \mathcal{Q}_\kappa(\tilde{y})$.*
(a) *If $\tilde{z} = a\tilde{x}$, then*

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

satisfy $p_{\tilde{z}} \in \mathcal{P}_\kappa(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}_\kappa(\tilde{z})$. 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 $(p_{\tilde{z}}, q_{\tilde{z}}) = (p_{\tilde{x}} + p_{\tilde{y}}, q_{\tilde{x}} + q_{\tilde{y}})$ satisfy $p_{\tilde{z}} \in \mathcal{P}(\tilde{z})$ and $q_{\tilde{z}} \in \mathcal{Q}(\tilde{z})$.*

Proof : (a) This result directly follows from the definition.
(b) We only show the result for the forward deviation measure. The result for the backward deviation measure follows from a similar argument. Notice that

$$\begin{aligned} M_{\tilde{z}}\left(\tfrac{1}{p_{\tilde{z}}}\right) &= M_{\tilde{x}}\left(\tfrac{1}{p_{\tilde{z}}}\right) M_{\tilde{y}}\left(\tfrac{1}{p_{\tilde{z}}}\right) \text{ [since } \tilde{x} \text{ and } \tilde{y} \text{ are independent]} \\ &= M_{\tilde{x}}\left(\tfrac{1}{p_{\tilde{x}}}\tfrac{p_{\tilde{x}}}{p_{\tilde{z}}}\right) M_{\tilde{y}}\left(\tfrac{1}{p_{\tilde{y}}}\tfrac{p_{\tilde{y}}}{p_{\tilde{z}}}\right) \\ &\leq \left(M_{\tilde{x}}\left(\tfrac{1}{p_{\tilde{x}}}\right)\right)^{\frac{p_{\tilde{x}}}{p_{\tilde{z}}}} \left(M_{\tilde{y}}\left(\tfrac{1}{p_{\tilde{y}}}\right)\right)^{\frac{p_{\tilde{y}}}{p_{\tilde{z}}}} \text{ [Hölder's Inequality]} \\ &\leq \kappa. \end{aligned}$$

Thus, $p_{\tilde{z}} = p_{\tilde{x}} + p_{\tilde{y}} \in \mathcal{P}(\tilde{z})$. ■

Theorem 9 *Let $p_\kappa(\tilde{z}) = \inf \mathcal{P}_\kappa(\tilde{z})$ and $q_\kappa(\tilde{z}) = \inf \mathcal{Q}_\kappa(\tilde{z})$. Then p_κ satisfies (D1), (D2), (D3'), (D4) and (D5) with $f_{\hat{\mathcal{S}}_+}(\Omega) = \kappa \exp(-\Omega)$ and q_κ satisfies (D1), (D2), (D3'), (D4) and (D5) with $f_{\hat{\mathcal{S}}_-}(\Omega) = \kappa \exp(-\Omega)$.*

Proof : We show the claim for p_κ only. The result for q_κ follows from a similar argument. It is clear that (D1), (D2) and (D4) hold for p_κ and (D3') follows from Theorem 8 **(b)**.

It remains to show (D5). Notice that if $\tilde{x} = 0$, (D5) follows trivially for all $\Omega, p_{\tilde{x}} \geq 0$. Otherwise, observe that for all $\Omega \geq 0$,

$$\mathrm{P}\left(\tilde{x} > \Omega p_{\tilde{x}}\right) = \mathrm{P}\left(\frac{\tilde{x}}{p_{\tilde{x}}} > \Omega\right) \leq \frac{\mathrm{E}\left(\exp\left(\frac{\tilde{x}}{p_{\tilde{x}}}\right)\right)}{\exp(\Omega)} \leq \kappa \exp(-\Omega),$$

where the first inequality follows from Chebyshev's inequality and the second inequality follows from the definition of $p_{\tilde{x}}$. ■

Parallel results to Theorem 4 can also be obtained. All our results hold for the deviation measures p_κ and q_κ with corresponding modifications. As previously mentioned, p_κ and q_κ can be applied to certain classes of random variables, such as exponential random variables, in which p^* and q^* are not well defined. Of course, since Theorem 8 part (b) is weaker than Theorem 3 part (b), one can expect the probability bound derived based on p^* and q^* to be stronger than that based on p_κ and q_κ, if both deviation measures are well defined. Thus, for simplicity of presentation, we focus on the deviation measures p^* and q^* in this paper.

References

[1] Atamtürk, A. (2003): Strong formulations of robust mixed 0-1 programming. *Working Paper*, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

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

[3] Ben-Tal, A., Nemirovski, A. (1999): Robust solutions to uncertain programs, *Oper. Res. Let.*, 25, 1-13.

[4] Ben-Tal, A., Nemirovski, A. (2000): Robust solutions of Linear Programming problems contaminated with uncertain data, *Math. Progr.*, 88, 411-424.

[5] Ben-Tal, A., A. Goryashko, E. Guslitzer and A. Nemirovski. (2004): Adjusting Robust Solutions of Uncertain Linear Programs. *Mathematical Programming*, 99, 351-376.

[6] Ben-Tal, A., B. Golany, A Nemirovski. and J. Vial. (2003): Supplier-Retailer Flexible Commitments Contracts: A Robust Optimization Approach. *Working Paper*, Technion.

[7] Bertsekas, D. (1999): Nonlinear Programming, *Athena Scientific*, Belmont, MA.

[8] Bertsimas, D., Pachamanova, D. and Sim, M. (2004): Robust Linear Optimization under General Norms, Operations Research Letters, 32:510516

[9] Bertsimas, D. and Sim, M. (2003): Robust Discrete Optimization and Network Flows, *Math. Progr.*, 98, 49-71.

[10] Bertsimas, D. and Sim, M. (2004): Price of Robustness, *Oper. Res.*, 52(1), 35-53.

[11] Bertsimas, D. and Sim, M. (2004): Robust Conic Opimization, *Working Paper*, MIT.

[12] Bertsimas, D. and Sim, M. (2004): Robust Discrete Opimization and Downside Risk Measures, *Working Paper*, National University of Singapore.

[13] Bertsimas, D. and A. Thiele. (2003): A Robust Optimization Approach to Supply Chain Management. Working Paper, MIT.

[14] Birge, J. R. and Louveaux, F. (1997): Introduction to Stochastic Programming. *Springer*, New York.

[15] Campi, M. C. and Calafiore, G. (2004): *Decision Making in an Uncertain Environment: the Scenario based Optimization Approach.* In: Multiple Participant Decision Making (J. Andrysek, M. Karny and J. Kracik eds.), pages 99-111, Advanced Knowledge International.

[16] Calafiore, G. and Campi, M. C. (2005): *Uncertain convex programs: randomized solutions and confidence levels.* In: Multiple Participant Decision Making (J. Andrysek, M. Karny and J. Kracik eds.), pages 99-111, Advanced Knowledge International.

[17] Charnes, A. and Cooper, W. W. (1959): Uncertain Convex Programs: Randomize solutions and confidence level, *Management Science*, 6, 73-79.

[18] Charnes, A. and Cooper, W. W. (1963): Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*, 11, 18-39.

[19] Charnes, A., Cooper, W. W. and Symonds, G. (1958): Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil production. *Management Science*, 4, 253-263.

[20] Chen, X., Sim, M., Sun, P. and Zhang, J. (2006): A Tractable Approximation of Stochastic Programming via Robust Optimization. *Working Paper*, National University of Singapore.

[21] de Farias, D. P. and Van Roy, B. (2001): On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res.*, to appear.

[22] El-Ghaoui, Lebret, H. (1997): Robust solutions to least-square problems to uncertain data matrices, *SIAM J. Matrix Anal. Appl.*, 18, 1035-1064.

[23] El-Ghaoui, L., Oustry, F., Lebret, H. (1998): Robust solutions to uncertain semidefinite programs, *SIAM J. Optim.*, 9, 33-52.

[24] Ergodan, G. and Iyengar, G. (2004): Ambiguous chance constrained problems and robust optimization, *Mathematicsl Programming Series B*, Special issue on Robust Optimization, to appear.

[25] Ergodan, G. and Iyengar, G. (2005): On two-stage convex constrained problems, *CORC Technical Report TR-2005-06*, Columbia University.

[26] Goldfarb, D. and Iyengar, G.(2003): Robust quadratically constrained programs, *Math. Prog. Ser. B*, 97(3), 495-515

[27] Garstka S. J. and Wets R.J-B. (1974): On decision rules in stochastic programming, *Math. Progr.*, 117-143.

[28] Hagstrom, J.N. (1988): Computational complexity of PERT problems. *Networks*, 18:139-147.

[29] Kibzun, A. and Kan, Y. (1996): Stochastic Programming Problems with Probability and Quantile Functions, *Johns Wiley & Sons*, New York.

[30] Nemirovski, A. and Shapiro, A. (2004): Scenario approximations of chance constraints. To appear in *Probabilistic and randomized methods for design under uncertainty.*

[31] Pintér, J. (1989): Deterministic Approximations of Probability Inequalities, *ZOR - Methods and Models of Operations Research*, 33, 219-239.

[32] Prékopa, A. (1995): Stochastic Programming. *Kluwer Academic Publishers, Dordrecht*, Boston.

[33] Rockafellar, R.T., Uryasev S. and M. Zabarankin. (2004) Generalized Deviations In Risk Analysis. Research Report 2004-4, ISE Dept., University of Florida.

[34] Shapiro, A. and Nemirovski, A. (2004): On Complexity of stochastic programming Problems, *Working Paper*, Georgia Tech, Atlanta.

[35] Soyster, A.L. (1973): Convex programming with set-inclusive constraints and applications to inexact linear programming, *Oper. Res.*, 21, 1154-1157.

[36] Toh, K.C. , Todd, M.J. , and Tütüncü, R.H. (1999): SDPT3- A Matlab Software package Programming, Optimization Methods and Software, 11, pp. 545-581.