Distributionally robust simple integer recourse

Weijun Xie*[1] and Shabbir Ahmed†[2]

[1]Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061

[2]School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332

February 4, 2018

Dedicated to the memory of Maarten van der Vlerk

Abstract

The simple integer recourse (SIR) function of a decision variable is the expectation of the integer round-up of the shortage/surplus between a random variable with a known distribution and the decision variable. It is the integer analogue of the simple (continuous) recourse function in two-stage stochastic linear programming. Structural properties and approximations of SIR functions have been extensively studied in the seminal works of van der Vlerk and coauthors. We study a distributionally robust SIR function (DR-SIR) that considers the worst-case expectation over a given family of distributions. Under the assumption that the distribution family is specified by its mean and support, we derive a closed form analytical expression for the DR-SIR function. We also show that this nonconvex DR-SIR function can be represented using a mixed-integer second-order conic program.

*Email: wxie@vt.edu.

†Email: sahmed@isye.gatech.edu.

1 Introduction

Background

A two-stage stochastic program with simple integer recourse [8] takes the form

$$\min_{z}\{c^\top z + \mathbb{E}_{\mathbb{P}}[Q(z,\tilde{\xi})] : \ z \in Z\} \tag{1}$$

with

$$Q(z,\xi) = \min_{u,v}\{q^\top u + r^\top v : u \geq \xi - Tz, \ v \geq Tz - \xi, \ u, v \in \mathbb{Z}^m_+\} \tag{2}$$

where $z \in \mathbb{R}^d$ denotes the first-stage decision vector with associated cost vector $c \in \mathbb{R}^d$ and constraints denoted by $Z \subseteq \mathbb{R}^d$; $T \in \mathbb{R}^{n\times d}$ is a deterministic *tender* matrix; $\tilde{\xi}$ denotes an n-dimensional random vector with given probability distribution $\mathbb{P}$ and realizations denoted by ξ; finally, $u \in \mathbb{Z}^n_+$ and $v \in \mathbb{Z}^n_+$ are nonnegative integer valued second-stage decision vectors with associated nonnegative cost vectors $q \in \mathbb{R}^n_+$ and $r \in \mathbb{R}^n_+$.

The interpretation of problem (1) is as follows. The decision maker chooses the first-stage decision z before observing a realization of the random vector $\tilde{\xi}$. After a realization ξ of $\tilde{\xi}$ is observed, a penalty is paid on the integer roundup of the shortage or surplus between ξ and Tz. The overall goal is to minimize the sum of the first-stage cost and the expected second-stage penalty. Problem (1) is the integer analog of a two-stage stochastic linear program with simple recourse that generalizes the classical newsvendor model.

Introducing new variables $x = Tz$, problem (1) can be recast as

$$\min_{x,z}\left\{c^\top z + \sum_{j=1}^{n} q_j \psi_j^+(x_j) + \sum_{j=1}^{n} r_j \psi_j^-(x_j) : \ \ z \in Z, \ x = Tz\right\} \tag{3}$$

with

$$\psi_j^+(x_j) = \mathbb{E}_{\mathbb{P}}\left(\lceil(\tilde{\xi}_j - x_j)_+\rceil\right) \text{ and } \psi_j^-(x_j) = \mathbb{E}_{\mathbb{P}}\left(\lceil(x_j - \tilde{\xi}_j)_+\rceil\right) \tag{4}$$

for $j = 1, \ldots, n$. Here, $(a)_+ = \max\{0, a\}$ and $\lceil a\rceil$ is the integer round up or ceiling of a.

The univariate functions ψ_j^+ and ψ_j^- are referred to as shortage and surplus expected simple integer recourse (SIR) functions, respectively. Structural properties and convex approximations of SIR functions have been extensively studied in the seminal works of van der Vlerk and coauthors [5, 6, 7, 8]. These approximations have led to tight deterministic approximations for two-stage stochastic programs with simple integer recourse of the form (3). The SIR approximation ideas have also been extended to build deterministic convex approximations of more general integer recourse functions [11, 12, 13, 14, 15, 21].

A distributionally robust analog of the two-stage stochastic program (1) takes the form

$$\min_{z} \left\{ c^\top z + \sup_{\mathbb{P}\in\mathcal{P}} \{\mathbb{E}_{\mathbb{P}}[Q(z,\tilde{\xi})]\} : \; z \in Z \right\}, \tag{5}$$

where $\mathcal{P}$ is a specified family of probability distributions and the goal is to find a solution z that optimizes the sum of the first-stage cost and the *worst-case* expected second-stage cost over the specified family of distributions. The distributionally robust optimization setting is motivated by the need to mitigate incomplete information on the probability distribution of the random problem parameters, and goes back to the pioneering work of Scarf [16] for the newsvendor problem, and subsequently, Dupacova [24] for general stochastic linear programming. In recent years, there has been a renewed interest in distributionally robust optimization (see e.g. [1, 2, 3, 4, 9, 18, 19, 22, 25] and references therein). Another viewpoint of distributionally robust optimization is that of risk averse stochastic programming which has been extensively studied [20]. Much of the existing structural results in distributionally robust optimization and risk averse stochastic programming is restricted to the convex setting. A notable exception is [17] wherein properties of risk averse stochastic integer programming is studied. To the best of our knowledge, there are no studies on distributionally robust simple integer recourse problems.

Contributions

As a first step in the study of two-stage distributionally robust optimization with simple integer recourse (5), we consider the following restricted setting:

(i) We address the one sided (shortage) simple integer recourse case, i.e. we assume $r = 0$ in (2).

(ii) We assume that the family of probability distributions $\mathcal{P}$ is specified by a given mean $\mu \in \mathbb{R}^n$ and support $[L, U]$ with $L, U \in \mathbb{R}^n$. Then $\mathcal{P} = \times_{j=1}^n \mathcal{P}_j$ with

$$\mathcal{P}_j = \left\{ \mathbb{P}_j \in \mathcal{M}_+(\Xi_j) : \mathbb{E}_{\mathbb{P}_j}[\tilde{\xi}_j] = \mu_j, \mathbb{E}_{\mathbb{P}_j}[1] = 1 \right\}, \tag{6}$$

where $\Xi_j = [L_j, U_j]$ is the support of the ξ_j, and $\mathcal{M}_+(\Xi_j)$ represents all positive measures with support Ξ_j, for $j = 1, \ldots, n$.

Under the above assumptions, two-stage distributionally robust optimization with simple integer recourse (5) takes the form

$$\min_{x,z} \left\{ c^\top z + \sum_{j=1}^n q_j f_j(x_j) : \;\; z \in Z, \; x = Tz \right\} \tag{7}$$

where, for $j = 1, \dots, n$,

$$f_j(x_j) = \sup_{\mathbb{P}_j \in \mathcal{P}_j} \left\{ \mathbb{E}_{\mathbb{P}_j} \left(\lceil (\tilde{\xi}_j - x_j)_+ \rceil \right) \right\}. \tag{8}$$

We refer to the univariate function f_j as the distributionally robust simple integer recourse (DR-SIR) function. The remainder of the paper studies the DR-SIR function. Our main contributions are as follows.

- Under some mild assumptions on L_j, U_j, μ_j and the domain of x_j, we derive a closed form analytical expression for the DR-SIR function f_j.
- Using the derived analytical form, we provide a representation of the epigraph of DR-SIR as a mixed-integer second order conic (MISOC) program.

Our results allow for two-stage distributionally robust optimization with simple integer recourse (5) to be recast as a deterministic MISOC program, and therefore, to be amenable to standard solvers.

2 Analytical Expression

Main Result

Next, we provide an analytical formula for the DR-SIR function f_j. Throughout the rest of this paper, we make the following assumptions:

(A1) $L, U \in \mathbb{Z}_+^n$ are nonnegative integers,

(A2) $L_j + 1 \leq \mu_j \leq U_j - 1$ for each $j = 1, \dots, n$.

The integrality in Assumption (A1) is for notational convenience and can be relaxed. Assumption (A2) requires that for each $j = 1, \dots, n$, the mean of demand μ_j is not too close to the lower or upper bound, otherwise we can get a good approximation by replacing $\tilde{\xi}_j$ in (8) by its mean. The main result of this section is the following theorem.

Theorem 1. *Under Assumptions (A1)-(A2), for any $j = 1, \dots, n$, the DR-SIR function f_j can be expressed as*

$$\begin{aligned} f_j(x_j) &= \max \left\{ \mu_j - x_j + 1, \frac{(\mu_j - L_j)\lceil (U_j - x_j)_+ \rceil}{\lceil (U_j - x_j)_+ \rceil - (L_j - x_j) - 1} \right\} \\ &= \begin{cases} \mu_j - x_j + 1, & x_j \leq L_j + 1 \\ \frac{(\mu_j - L_j)\lceil (U_j - x_j)_+ \rceil}{\lceil (U_j - x_j)_+ \rceil - (L_j - x_j) - 1}, & L_j + 1 < x_j < U_j \\ 0, & U_j \leq x_j. \end{cases} \end{aligned} \tag{9}$$

Furthermore, f_j is lower semicontinuous, piecewise convex, and the possible of points of discontinuity are $\{L_j + 1, \ldots, U_j\}$.

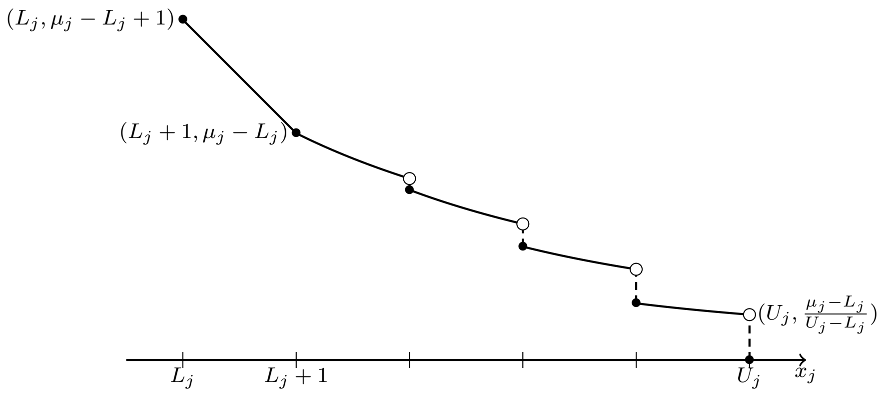

Figure 1: Illustration of function $f_j(x_j)$

Proof of Theorem 1

For notational convenience, we suppress the subscript j of $\mathcal{P}_j, L_j, U_j, x_j, \mu_j, f_j(\cdot)$. We separate the proof into six steps.

(i) Note that when $x \geq U$, we have $\tilde{\xi} \leq U \leq x$. Thus, since $f(x) = \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \mathbb{E}_{\mathbb{P}} \left(\lceil (\tilde{\xi} - x)_+ \rceil \right) \right\}$, we have $f(x) = 0$. Therefore, from now on, we assume that $x < U$.

(ii) We let $\tilde{\omega} = \tilde{\xi} - x$ for any given x. Then the inner supremum of (8) is equivalent to

$$f(x) = \sup_{\mathbb{P} \in \mathcal{P}_x} \left\{ \mathbb{E}_{\mathbb{P}} \left(\lceil (\tilde{\omega})_+ \rceil \right) \right\}, \tag{10}$$

where

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{M}_+(\Xi_x) : \mathbb{E}_{\mathbb{P}}[\tilde{\omega}] = \mu_x, \mathbb{E}_{\mathbb{P}}[1] = 1 \right\},$$

with $\mu_x = \mu - x, L_x = L - x, U_x = U - x$ and $\Xi_x = [L_x, U_x]$.

(iii) Next, due to Assumption (A2) that $L + 1 \leq \mu \leq U + 1$, the set $\mathcal{P}_x$ in (6) is nonempty for any small perturbation of μ_x, therefore, Lemma 1 in [23] implies that (10) is equivalent to the following semi-infinite minimization problem

$$\begin{aligned} f(x) = \min_{\alpha, \lambda} \quad & \alpha + \lambda \mu_x \\ \text{s.t.} \quad & \alpha + \lambda \omega \geq \lceil (\omega)_+ \rceil, \forall \omega \in [L_x, U_x]. \end{aligned}$$

By enumerating the integer points in the interval $[L_x, U_x]$, the above formulation is further reduced to

$$
\begin{aligned}
f(x) = \min_{\alpha,\lambda} \quad & \alpha + \lambda\mu_x && \text{(11a)}\\
\text{s.t.} \quad & \alpha + \lambda\omega \geq \lceil (L_x)_+ \rceil, \forall \omega \in [L_x, \lceil L_x \rceil], && \text{(11b)}\\
& \alpha + \lambda\omega \geq (k+1)_+, \forall \omega \in (k, k+1], k \in \{\lceil L_x \rceil, \ldots, \lceil U_x \rceil - 2\}, && \text{(11c)}\\
& \alpha + \lambda\omega \geq \lceil U_x \rceil, \forall \omega \in (\lceil U_x \rceil - 1, U_x]. && \text{(11d)}
\end{aligned}
$$

We observe that for any optimal solution (α^*, λ^*) of (11), we must have $\lambda^* \geq 0$. Indeed, suppose that $\lambda^* < 0$. Note that constraint (11d) for $\omega = U_x$ implies

$$\alpha^* \geq \lceil U_x \rceil - \lambda^* U_x,$$

which in turn implies that the optimal value to (11) satisfies

$$\alpha^* + \lambda^* \mu_x \geq \lceil U_x \rceil + \lambda^*(\mu - U) > \lceil U_x \rceil,$$

where the last strict inequality is due to $\lambda^* < 0$ and $\mu < U$. However, this is a contradiction, since $\lceil U_x \rceil$ is an upper bound on $f(x)$. Therefore, without loss of generality in (11), we can assume $\lambda \geq 0$, i.e.

$$
\begin{aligned}
f(x) = \min_{\alpha,\lambda} \quad & \alpha + \lambda\mu_x && \text{(12a)}\\
\text{s.t.} \quad & \alpha + \lambda\omega \geq \lceil (L_x)_+ \rceil, \forall \omega \in [L_x, \lceil L_x \rceil], && \text{(12b)}\\
& \alpha + \lambda\omega \geq (k+1)_+, \forall \omega \in (k, k+1], k \in \{\lceil L_x \rceil, \ldots, \lceil U_x \rceil - 2\}, && \text{(12c)}\\
& \alpha + \lambda\omega \geq \lceil U_x \rceil, \forall \omega \in (\lceil U_x \rceil - 1, U_x], && \text{(12d)}\\
& \lambda \geq 0. && \text{(12e)}
\end{aligned}
$$

(iv) Next, in (12c), we claim that for any $k \in \{\lceil L_x \rceil, \ldots, \lceil U_x \rceil - 2\}$, the following set

$$Q_1 = \{(\alpha, \lambda) : \alpha + \lambda\omega \geq (k+1)_+, \forall \omega \in (k, k+1], \lambda \geq 0\},$$

is equivalent to

$$Q_2 = \{(\alpha, \lambda) : \alpha + k\lambda \geq (k+1)_+, \lambda \geq 0\}.$$

It is clear that $Q_1 \subseteq Q_2$ since the constraints defining Q_2 is a subset of the ones defining Q_2 by letting

$\omega \to k$. Now suppose that there exists a $(\widehat{\alpha}, \widehat{\lambda}) \in Q_2 \setminus Q_1$, i.e. there exists an $\widehat{\omega} \in (k, k+1)$ such that $\widehat{\alpha} + \widehat{\lambda}\widehat{\omega} < (k+1)_+$. Since $\widehat{\omega} > k$, thus

$$\widehat{\alpha} + \widehat{\lambda}\widehat{\omega} \geq \widehat{\alpha} + \widehat{\lambda}k \geq (k+1)_+,$$

a contradiction. Therefore, $Q_1 = Q_2$.

By a similar argument, we can show that

$$\begin{aligned} &\{(\alpha, \lambda) : \alpha + \lambda\omega \geq \lceil (L_x)_+ \rceil, \forall \omega \in [L_x, \lceil L_x \rceil], \lambda \geq 0\} \\ &= \{(\alpha, \lambda) : \alpha + \lambda L_x \geq \lceil (L_x)_+ \rceil, \lambda \geq 0\} \end{aligned}$$

and

$$\{(\alpha, \lambda) : \alpha + \lambda\omega \geq \lceil U_x \rceil, \forall \omega \in (\lceil U_x \rceil - 1, U_x], \lambda \geq 0\} = \{(\alpha, \lambda) : \alpha + \lambda (\lceil U_x \rceil - 1) \geq \lceil U_x \rceil, \lambda \geq 0\}.$$

Hence, (12) can be reformulated as the linear program

$$\begin{aligned} f(x) = \min_{\alpha,\lambda} \quad & \alpha + \lambda\mu_x && \text{(13a)} \\ \text{s.t.} \quad & \alpha + \lambda L_x \geq \lceil (L_x)_+ \rceil, && \text{(13b)} \\ & \alpha + \lambda k \geq (k+1)_+, \forall k \in \{\lceil L_x \rceil, \ldots, \lceil U_x \rceil - 2\}, && \text{(13c)} \\ & \alpha + \lambda (\lceil U_x \rceil - 1) \geq \lceil U_x \rceil, && \text{(13d)} \\ & \lambda \geq 0. && \text{(13e)} \end{aligned}$$

(v) Note that (13) is a linear program with two variables. Thus an extreme point of the feasible region of (13) is given by the intersection of two linearly independent constraints. Since $x \leq U$, there are two cases whether the left-hand coefficient matrix is mutually row independent or not:

Case 1: $L_x > -1$ (i.e., $x < L+1$). Then as shown in Figure 2(a), there are three types of extreme points:

i. Intersecting inequality (13b) and one inequality from (13c), we obtain $\alpha^1 = \lceil (L_x)_+ \rceil - \frac{L_x}{\lceil L_x \rceil - L_x}, \lambda^1 = \frac{1}{\lceil L_x \rceil - L_x}$.

ii. Intersecting two inequalities in (13c), we obtain $\alpha^2 = 1, \lambda^2 = 1$.

iii. Intersecting (13e) and (13d) , we obtain $\alpha^3 = \lceil U_x \rceil, \lambda^3 = 0$.

Thus, we have

$$f(x) = \min_{i=1,2,3} \{g_i(x) := \alpha^i + \lambda^i \mu_x\}$$

where

$$
\begin{aligned}
g_1(x) &= \lceil (L-x)_+ \rceil - \frac{L-x}{\lceil L-x \rceil - (L-x)} + \frac{\mu - x}{\lceil L-x \rceil - (L-x)} \\
&= \lceil (L-x)_+ \rceil + \frac{\mu - L}{\lceil L-x \rceil - (L-x)} \\
g_2(x) &= \mu - x + 1 \\
g_3(x) &= \lceil U - x \rceil .
\end{aligned}
$$

Note that

$$
g_2(x) - g_3(x) = (\mu - x + 1) - \lceil U - x \rceil \leq (\mu + 1) - U \leq 0
$$

where the first inequality follows from the fact that $\lceil U - x \rceil \geq U - x$, and the last inequality is due to Assumption (A2) $U \geq \mu - 1$.

Now, for any $x \in [L-k, L-k+1)$ for $k \in \mathbb{Z}, k \geq 0$, we have $\lceil L - x \rceil = k$. Thus,

$$
\begin{aligned}
g_2(x) - g_1(x) &= (\mu - x + 1) - k - \frac{\mu - L}{k - (L-x)} \\
&= -\frac{1}{k - (L-x)} \left[k - (L-x) - 1\right] \left[k - (L-x) - (\mu - L)\right] \leq 0
\end{aligned}
$$

where the last inequality is due to $k - (L-x) < 1$ and $\mu - L \geq 1$.

Therefore, we have

$$
f(x) := g_2(x) = \mu - x + 1.
$$

Case 2: $L_x \leq -1$ (i.e., $x \geq L+1$). Then, as shown in Figure 2(a), there are two types of extreme points.

i. Intersecting (13b) and (13d), we obtain $\alpha^1 = -\frac{L_x \lceil U_x \rceil}{\lceil U_x \rceil - L_x - 1}, \lambda^1 = \frac{\lceil U_x \rceil}{\lceil U_x \rceil - L_x - 1}$.

ii. Intersecting (13e) and (13d) , we obtain $\alpha^2 = \lceil U_x \rceil, \lambda^2 = 0$.

Therefore, we have

$$
f(x) = \min_{i=1,2} \left\{ h_i(x) := \alpha^i + \lambda^i \mu_x \right\}
$$

where

$$
\begin{aligned}
h_1(x) &= -\frac{(L-x)\lceil U-x \rceil}{\lceil U-x \rceil - (L-x) - 1} + \frac{\lceil U-x \rceil}{\lceil U-x \rceil - (L-x) - 1}(\mu - x) \\
&= \frac{(\mu - L)\lceil U-x \rceil}{\lceil U-x \rceil - (L-x) - 1} \\
h_2(x) &= \lceil U - x \rceil .
\end{aligned}
$$

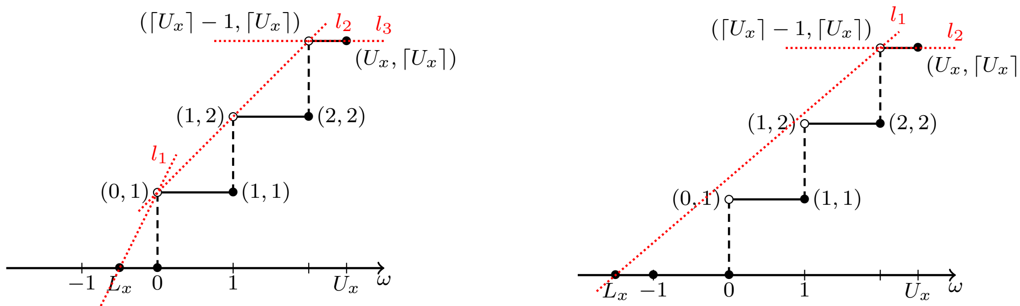

Figure 2: Two cases in step (v) in the proof of Theorem 1

We note that

$$h_1(x) - h_2(x) = \lceil U - x \rceil \left[\frac{\mu - L}{\lceil U - x \rceil - (L - x) - 1} - 1 \right] \leq \lceil U - x \rceil \left[\frac{\mu - L}{U - L - 1} - 1 \right] \leq 0.$$

where the first inequality is due to $x \leq U$ and $\lceil U - x \rceil \geq U - x$, and the last inequality is because of Assumption (A2) $\mu \leq U - 1$.

Hence,

$$f(x) = h_1(x) := \frac{(\mu - L)\lceil U - x \rceil}{\lceil U - x \rceil - (L - x) - 1}$$

whenever $L + 1 \leq x \leq U$.

(vi) Let $\phi(x) = g_2(x) - h_1(x)$. It remains to show that $g_2(x) \geq h_1(x)$ when $x < L + 1$ and $g_2(x) \leq h_1(x)$ when $x \in [L+1, U]$, i.e. to prove (1) $\inf_{x \in (\infty, L+1)}\{\phi(x)\} \geq 0$ and (2) $\max_{x \in [L+1, U]}\{\phi(x)\} \leq 0$.

Since $\phi(U) = \mu + 1 - U \leq 0$, we now suppose that $x \in [k, k+1)$, where $k \leq U - 1$ is an integer. Then $\lceil U - x \rceil = U - k$, therefore,

$$\phi(x) = \mu - x + 1 - \frac{(\mu - L)(U - k)}{U + x - k - L - 1}.$$

There are two cases:

(1) if $k \leq L$, we note that $\phi(x)$ is concave in $x \in [k, k+1)$ since its second derivative is

$$\frac{d^2\phi(x)}{dx^2} = -\frac{2(\mu - L)(U - k)}{(U + x - k - L - 1)^3} \leq 0.$$

Therefore, the minimizer of $\inf_{x \in [k, k+1)}\{\phi(x)\}$ is achieved by $x^* = k$ or $x^* = k + 1$. Note that

$$\phi(k) = \mu - k + 1 - \frac{(\mu - L)(U - k)}{U - L - 1},$$

$$\phi(k+1) = \mu - k - \frac{(\mu - L)(U-k)}{U-L}.$$

Note that both $\phi(k), \phi(k+1)$ is nonincreasing over k. Thus,

$$\inf_{k\in\mathbb{Z},k\le L} \phi(k) = \mu - L + 1 - \frac{(\mu-L)(U-L)}{U-L-1} = \frac{U-\mu}{U-L-1} \ge 0,$$
$$\inf_{k\in\mathbb{Z},k\le L} \phi(k+1) = \mu - L - \frac{(\mu-L)(U-L)}{U-L} = 0.$$

Hence, $\inf_{x\in(-\infty,L+1)}\{\phi(x)\} = \inf_{k\in\{0,1,\ldots,L\}} \inf_{x\in[k,k+1)}\{\phi(x)\} = 0$, i.e., $g_2(x) \ge h_1(x)$ when $x \in (-\infty, L+1)$.

(2) Suppose that $x \in [k, k+1)$, where $L+1 \le k \le U-1$. Then $\lceil U - x\rceil = U - k$, therefore,

$$\phi(x) = \mu - x + 1 - \frac{(\mu-L)(U-k)}{U+x-k-L-1}.$$

Note that the first derivation of $\phi(x)$ is

$$\frac{d\phi(x)}{dx} = -1 + \frac{(\mu-L)(U-k)}{(U+x-k-L-1)^2} \le -1 + \frac{\mu-L}{U+x-k-L-1} \le -1 + \frac{\mu-L}{U-L-1} \le 0,$$

where the first inequality is due to $x \ge L+1$, the second one is due to $x \ge k$ and the last one is because of $U \ge \mu+1$. Thus, $\phi(x)$ is nonincreasing over $x \in [k, k+1)$.

Therefore, the maximizer of $\max x \in [k,k+1)\{\phi(x)\}$ is achieved by $x = k$, i.e.,

$$\phi(x) \le \phi(k) = \mu - k + 1 - \frac{(\mu-L)(U-k)}{U-L-1}.$$

Note that $\phi(k)$ is also nonincreasing over $k \in [L+1, U-1]$. Hence, for any $x \in [L+1, U]$,

$$\phi(x) \le \phi(L+1) = \mu - L - (\mu - L) = 0.$$

Thus, $\max_{x\in[L+1,U]}\{\phi(x)\} = 0$, i.e., $g_2(x) \le h_1(x)$ when $x \in [L+1, U]$.

This completes the proof of Theorem 1.

Remark: Based on the above proof, the worst-case distribution is as follows:

(i) if $x_j < L_j + 1$, then the worst-case distribution is $\mathbb{P}\left\{\tilde{\xi}_j = x_j + \lceil L_j - x_j\rceil\right\} = 1 - \frac{\mu_j - x_j - \lceil L_j - x_j\rceil}{\lceil U_j - x_j\rceil - \lceil L_j - x_j\rceil - 1}$ and $\mathbb{P}\left\{\tilde{\xi}_j = x_j + \lceil U_j - x_j\rceil - 1\right\} = \frac{\mu_j - x_j - \lceil L_j - x_j\rceil}{\lceil U_j - x_j\rceil - \lceil L_j - x_j\rceil - 1}$

(ii) if $L_j + 1 \le x_j < U_j$, then the worst-case distribution is $\mathbb{P}\left\{\tilde{\xi}_j = L_j\right\} = 1 - \frac{\mu_j - L_j}{\lceil U_j - x_j\rceil - 1 - L_j + x_j}$ and

$\mathbb{P}\left\{\tilde{\xi}_j = x_j + \lceil U_j - x_j \rceil - 1\right\} = \frac{\mu_j - L_j}{\lceil U_j - x_j \rceil - 1 - L_j + x_j}$

(iii) if $x_j \geq U_j$, then one of the worst-case distributions is $\mathbb{P}\left\{\tilde{\xi}_j = \mu_j\right\} = 1$.

3 Mixed-Integer Conic Reformulation

Main Result

Next, we show that the DR-SIR function admits a mixed-integer second order conic (MISOC) representation. From Theorem 1, we know that the DR-SIR functions $\{f_j(x_j)\}$ are discontinuous, and the possible of points of discontinuity are $\{L_j + 1, \ldots, U_j\}$. Therefore, to formulate set W_j as an MISOC program, we make the following assumption:

(A3) For any first-stage solution (z, x), we have $|x_j| \leq M_j$ with positive constant $M_j \geq U_j$ for each $j = 1, \ldots, n$.

Let us first define the epigraph of $f_j(\cdot)$ for each $j = 1, \ldots, n$:

$$W_j = \{(x_j, w_j) : w_j \geq f_j(x_j), |x_j| \leq M_j\}. \tag{14}$$

The following result provides a MISOC formulation of W_j .

Theorem 2. *Under Assumptions (A1)-(A3), the set W_j in* (14) *is equivalent to*

$$W_j = \left\{ (w_j, x_j) : \begin{array}{ll} w_j \geq \mu_j - x_j + 1, & \text{(15a)} \\ \left\| \begin{bmatrix} w_j - ((i - L_j - 1)y_{ij} + u_j) \\ 2\sqrt{i(\mu_j - L_j)}y_{ij} \end{bmatrix} \right\|_2 \leq w_j + ((i - L_j - 1)y_{ij} + x_j), & \\ \forall i = 0, \ldots, U_j - L_j - 1, & \text{(15b)} \\ (\mu_j - L_j)(1 - \chi_j) \leq w_j, & \text{(15c)} \\ \sum_{i=0}^{U-L-1} y_{ij} = \chi_j, & \text{(15d)} \\ \sum_{i=0}^{U-L-1} i y_{ij} \geq U_j \chi_j - x_j, & \text{(15e)} \\ x_j = u_j + v_j, & \text{(15f)} \\ (L_j + 1)\chi_j \leq u_j \leq M_j(1 - \chi_j), & \text{(15g)} \\ |v_j| \leq M_j \chi_j, & \text{(15h)} \\ y_{ij} \in \{0, 1\}, \forall i = 0, \ldots, U - L - 1, w_j \geq 0. & \text{(15i)} \end{array} \right\}$$

for each $j = 1, \ldots, n$.

Remark: Note that in (15), there are $O(U_j - L_j)$ MISOC constraints, which could be exponential in the bit length of problem instance. Unfortunately, in general, we are not able to reduce this number since the number of discontinuities in $f_j(x_j)$ is exactly equal to $U_j - L_j - 1$ (see Figure 1 for an illustration). To represent each such discontinuous point, we have to introduce at least one binary variable and one constraint, therefore, the number of MISOC constraints is at least $\Omega(U_j - L_j)$.

Proof of Theorem 2

For notational convenience, we suppress the subscript j of $L_j, U_j, M_j, x_j, u_j, v_j, \chi_j, w_j, \mu_j, f_j(\cdot), W_j$.

By breaking down the maximum in (9) and Theorem 1, the set W is equivalent to

$$W = \left\{(w, x) : w \geq \mu - x + 1, w \geq \frac{(\mu - L)\lceil (U - x)_+ \rceil}{\lceil (U - x)_+ \rceil - (L - x) - 1}, |x| \leq M\right\}. \tag{16}$$

Now, let

$$\widehat{W} = \left\{ (w, x) : \begin{array}{ll} w \geq \mu - x + 1, & \text{(17a)} \\ i(\mu - L)y_i^2 \leq w\left[(i - 1 - L)\, y_i + u\right], \forall i = 0, \ldots, U - L - 1, & \text{(17b)} \\ (\mu - L)(1 - \chi) \leq w & \text{(17c)} \\ \sum_{i=0}^{U-L-1} y_i = \chi, & \text{(17d)} \\ \sum_{i=0}^{U-L-1} i y_i \geq U\chi - u, & \text{(17e)} \\ x = u + v, & \text{(17f)} \\ (L + 1)\chi \leq u \leq M\chi, & \text{(17g)} \\ |v| \leq M(1 - \chi), & \text{(17h)} \\ |x| \leq M, & \text{(17i)} \\ y_i \in \{0, 1\}, \forall i = 0, \ldots, U - L - 1, w \geq 0, \chi \in \{0, 1\}. & \text{(17j)} \end{array} \right\}$$

It is easily verified that the quadratic constraint (17b) is equivalent to the conic inequality (15b). Thus we need to show that $W = \widehat{W}$. We proceed in the following three steps

(i) We claim that, without loss of generality, we can assume $0 \leq x < U$ in both sets W and $\widehat{W}$. Indeed, for any (w, x) such that $w \geq 0, M \geq x \geq U$, we must have $(w, x) \in W$ and $(w, x) \in \widehat{W}$ by choosing

$\chi = 1, u = x, v = 0$ and $y_0 = 1, y_i = 0$ for any $i \neq 0$ in (17). Thus,

$$W \cap \{(w,x) : M \geq x \geq U, w \geq 0\} = \widehat{W} \cap \{(w,x) : M \geq x \geq U, w \geq 0\} = \{(w,x) : w \geq 0, U \leq x \leq M\}.$$

On the other hand, if $-M \leq x < 0$, then Theorem 1 implies that $\mu - x + 1 \geq \frac{(\mu-L)\lceil (U-x)_+ \rceil}{\lceil (U-x)_+ \rceil - (L-x) - 1}$. And by choosing $\chi = 0, u = 0, v = x$ and $y_i = 0$ for any i in (17), we have

$$\begin{aligned} W \cap \{(w,x) : -M \leq x < 0, w \geq 0\} &= \widehat{W} \cap \{(w,x) : -M \leq x < 0, w \geq 0\} \\ &= \{(w,x) : w \geq \mu - x + 1, -M \leq x < 0\}. \end{aligned}$$

Thus from now on, we assume that $0 \leq x < U$ in both set $W, \widehat{W}$.

(ii) We now show that $W \subseteq \widehat{W}$. Given $(\bar{w}, \bar{x}) \in W$, suppose that $\bar{x} \in [U-i, U-i+1)$ for some $i \in \{1, \ldots, U\}$. Then we have $\lceil U - \bar{x} \rceil = i$. There are two cases:

Case 1. if $1 \leq i \leq U - L - 1$, then $U > \bar{x} \geq L + 1$. Thus, Theorem 1 implies that $\mu - \bar{x} + 1 \leq \frac{(\mu-L)\lceil (U-\bar{x})_+ \rceil}{\lceil (U-\bar{x})_+ \rceil - (L-\bar{x}) - 1} := \frac{i(\mu-L)}{i-(L-\bar{x})-1}$. Therefore, let $\bar{\chi} = 1, \bar{u} = \bar{x}, \bar{v} = 0$, $\bar{y}_i = 1$ and $\bar{y}_{i'} = 0$ for $i' \neq i \in \{0, \ldots, U - L - 1\}$. Thus clearly, $(\bar{w}, \bar{\chi}, \bar{u}, \bar{v}, \bar{x}, \bar{y})$ satisfies (17), i.e. $(\bar{w}, \bar{x}) \in \widehat{W}$.

Case 2. if $i \geq U-L-1$, then $0 \leq \bar{x} < L+1$. Thus, Theorem 1 implies that $\mu - \bar{x} + 1 \geq \frac{(\mu-L)\lceil (U-\bar{x})_+ \rceil}{\lceil (U-\bar{x})_+ \rceil - (L-\bar{x}) - 1} := \frac{i(\mu-L)}{i-(L-\bar{x})-1}$. Therefore, let $\bar{\chi} = 0, \bar{u} = 0, \bar{v} = \bar{x}$, $\bar{y}_{i'} = 0$ for $i' \in \{0, \ldots, U - L - 1\}$. Thus clearly, $(\bar{w}, \bar{\chi}, \bar{u}, \bar{v}, \bar{x}, \bar{y})$ satisfies (17), i.e. $(\bar{w}, \bar{x}) \in \widehat{W}$.

(iii) We now show that $W \supseteq \widehat{W}$. Given $(\bar{w}, \bar{x}) \in \widehat{W}$ with $0 \leq \bar{x} < U$, there exists $(\bar{\chi}, \bar{u}, \bar{v}, \bar{y})$ such that $(\bar{w}, \bar{\chi}, \bar{u}, \bar{v}, \bar{x}, \bar{y})$ satisfies (17). Suppose that $\bar{x} \in [U-i, U-i+1)$ for some $i \in \{1, \ldots, U\}$, i.e. $\lceil U - \bar{x} \rceil = i$. There are two cases:

Case 1. If $i \geq U - L$ (i.e., $0 \leq \bar{x} \leq L+1$), then by Theorem 1, we know that $\frac{(\mu-L)\lceil U-\bar{x} \rceil}{\lceil U-\bar{x} \rceil - (L-\bar{x}) - 1} \leq \mu - \bar{x} + 1$. Therefore, (17) implies that $\bar{w} \geq \mu - \bar{x} + 1, \bar{w} \geq 0$, i.e. $(\bar{w}, \bar{x}) \in W$.

Case 2. If $i < U - L$ (i.e., $U > \bar{x} > L+1$), then by Theorem 1, we know that $\mu - L > \frac{(\mu-L)\lceil U-\bar{x} \rceil}{\lceil U-\bar{x} \rceil - (L-\bar{x}) - 1} \geq \mu - \bar{x} + 1$. Therefore, if $\bar{w} \geq \mu - L$, then $(\bar{w}, \bar{x}) \in W$; otherwise, (17c) implies that $\bar{\chi} = 1$. Since $\sum_{i'=0}^{U-L-1} i' \bar{y}_{i'} \geq U - \bar{x} > 0$ and $\sum_{i'=0}^{U-L-1} \bar{y}_{i'} = 1$, thus we must have $\sum_{i'=i}^{U-L-1} \bar{y}_{i'} = 1$. Suppose that $\bar{y}_{\widehat{i}} = 1$ for some $\widehat{i} \in \{i, \ldots, U - L - 1\}$. Then (17) implies that $\bar{w} \geq 0$ and

$$\bar{w} \geq \frac{(\mu-L)\widehat{i}}{\widehat{i} + \bar{x} - L - 1} \geq \frac{(\mu-L)i}{i + \bar{x} - L - 1} := \frac{(\mu-L)\lceil U - \bar{x} \rceil}{\lceil U - \bar{x} \rceil + \bar{x} - L - 1},$$

where the second inequality is because the function $h(t) = \frac{(\mu-L)t}{t+\bar{x}-L-1}$ is nondecreasing in t when $\bar{x} \geq L+1$. Hence, $(\bar{w}, \bar{x}) \in W$.

This completes the proof.

4 Numerical Illustration

In this section, we consider the following two-stage stochastic program with simple integer recourse for a numerical illustration:

$$\min_x \left\{ x^2 + 50\mathbb{E}_{\mathbb{P}}[\lceil(\tilde{\xi} - x)_+\rceil] : x \geq 0 \right\}. \tag{18}$$

Its distributionally robust counterpart is

$$\min_x \left\{ x^2 + 50 \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\lceil(\tilde{\xi} - x)_+\rceil] : x \geq 0 \right\} \tag{19}$$

with the ambiguity set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{M}_+([20, 80]) : \mathbb{E}_{\mathcal{P}}[\tilde{\xi}] = 50, \mathbb{E}_{\mathbb{P}}[1] = 1 \right\},$$

i.e., the random variable $\tilde{\xi}$ has mean 50 and support in $[20, 80]$. For the stochastic program (18), we assume that the random variable $\tilde{\xi}$ obeys a uniform distribution in $[20, 80]$. We solve the following sample average approximation of (18) using $N = 1000$ i.i.d. samples $\{\xi^i\}_{i\in[N]}$ from $\tilde{\xi}$:

$$\min_{x,y} \left\{ x^2 + 50 \sum_{i\in[N]} \frac{1}{N} y^i : y^i \geq \xi^i - x, y^i \in \mathbb{Z}_+, \forall i \in [N], x \geq 0 \right\}. \tag{20}$$

On the other hand, the distributionally robust model (19) can be exactly formulated as an MISOC program according to Theorem 2. We refer to the stochastic programming model (18) as SP, the distributionally robust model (19) as DRO, and the sample average approximation model (20) as SAA.

We solve both (19) and (20) using the commercial solver Gurobi with time limits set to an hour. The results are displayed in Table 1. We observe that the DRO model (19) can be solved within seconds while the SAA model (20) is relatively difficult to solve within the time limit. The value of the solution to the DRO model, i.e. $x = 25$, with respect to the objective of the SAA model is 3277.1, which is about 8.7% higher than that of the SAA solution. This indicates the expected conservativeness of the DRO approach if the assumed distribution in the stochastic model is correct.

Table 1: Comparison between models (19) and (20)

Models	Best Upper Bound	Gap	Solution Time (s)	Best x
DRO model (19)	3422	0.0%	2	25.0
SAA model (20)	3015	0.8%	3600	36.8

To compare the solution qualities if the distribution deviates from the assumed one, we suppose that the true distribution of the random variable $\tilde{\xi}$ is a truncated logistic distribution $\log(50, s)$ with support in $[20, 80]$, where s denotes the scale parameter proportional to the standard deviation. We consider $s \in \{10, 20, 30, \ldots, 200\}$ and for each s, we estimate the true objective function $x^2 + 50\mathbb{E}_{\mathbb{P}}[\lceil(\tilde{\xi} - x)_+\rceil]$ of the SP model (18) corresponding to the best solutions x of the DRO model (19) and the SAA model (20). The objective is estimated using 1000 i.i.d. samples from the true distribution. The results are shown in Figure 3, where v_R, v_S denote the estimated objective values of the DRO solution and the SAA solution, respectively. We observe that when the scale parameter s grows (i.e., the variance of random variable increases), the DRO model (19) consistently outperforms the SAA model (20), i.e. the DRO model (19) has smaller estimated objective value of SAA model. This demonstrates that the DRO model (19) immunizes against ambiguity of the probability distribution and is thus is more reliable.

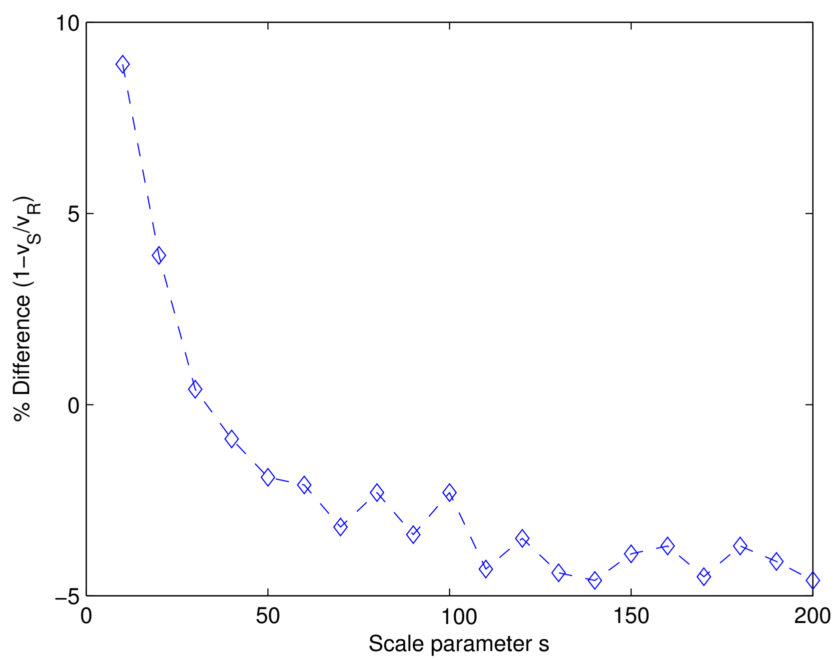

Figure 3: Comparison of best solutions from models (19) and (20)

5 Conclusions and Future Research

We derived an analytical expression for the one-sided distributionally robust simple integer recourse function when the distribution family is specified by its support and mean. We showed that the derived expression has a mixed-integer conic representation. These results allow for a two-stage distributionally robust optimization with simple integer recourse to be recast as a deterministic mixed-integer conic program, and therefore, to be amenable to standard solvers. Our numerical study shows that the proposed model can be solved efficiently and the results are more reliable with respect to distributional ambiguity than its stochastic counterpart. Two nontrivial extensions of this work include considering two-sided distributionally simple integer recourse functions of the form

$$f_j(x_j) = \sup_{\mathbb{P}_j \in \mathcal{P}_j} \left\{ \mathbb{E}_{\mathbb{P}_j} \left(q_j \lceil (\tilde{\xi}_j - x_j)_+ \rceil + r_j \lceil (x_j - \tilde{\xi}_j)_+ \rceil \right) \right\},$$

and extensions to other distribution families such as those specified by first and second moments, i.e.

$$\mathcal{P}_j = \left\{ \mathbb{P}_j \in \mathcal{M}_+(\Xi_j) : \mathbb{E}_{\mathbb{P}_j}[\tilde{\xi}_j] = \mu_j, \mathbb{E}_{\mathbb{P}_j}[\tilde{\xi}_j^2] = \sigma_j^2, \mathbb{E}_{\mathbb{P}_j}[1] = 1 \right\},$$

or other parmeters (e.g., mean-deviation [10] and lower/upper bounds on the mean).

Acknowledgements:

This research has been supported in part by the National Science Foundation grant #1633196. The authors thank the editor and two anonymous referees for constructive comments.

References

[1] G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. In *Tutorials in Operations Research*, INFORMS, pages 1–19, 2015.

[2] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[3] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming, accepted, 2017.

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

[5] W. K. Klein Haneveld, L. Stougie, and M. van der Vlerk. On the convex hull of the simple integer recourse objective function. *Annals of Operations Research*, 56:209–224, 1995.

[6] W. K. Klein Haneveld, L. Stougie, and M. van der Vlerk. An algorithm for the construction of convex hulls in simple integer recourse programming. *Annals of Operations Research*, 64:67–81, 1996.

[7] W. K. Klein Haneveld, L. Stougie, and M. van der Vlerk. Simple integer recourse models: convexity and convex approximations. *Mathematical Programming, Series B*, 108:435–473, 2006.

[8] F. V. Louveaux and M. H. van der Vlerk. Stochastic programming with simple integer recourse. *Mathematical programming*, 61(1-3):301–325, 1993.

[9] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55:98–112, 2007.

[10] K. Postek, A. Ben-Tal, D. Den Hertog, and B. Melenberg. Exact robust counterparts of ambiguous stochastic constraints under mean and dispersion information. Available at Optimization Online, 2015.

[11] W. Romeijnders, D. P. Morton, and M. H. van der Vlerk. Assessing the quality of convex approximations for two-stage totally unimodular integer recourse models. *INFORMS Journal on Computing*, 29(2):211–231, 2017.

[12] W. Romeijnders, R. Schultz, M. H. van der Vlerk, and W. K. K. Haneveld. A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization*, 26:426–447, 2016.

[13] W. Romeijnders, L. Stougie, and M. H. van der Vlerk. Approximation in two-stage stochastic integer programming. *Surveys in Operations Research and Management Science*, 19(1):17–33, 2014.

[14] W. Romeijnders, M. H. van der Vlerk, and W. K. K. Haneveld. Convex approximations for totally unimodular integer recourse models: A uniform error bound. *SIAM Journal on Optimization*, 25:130–158, 2015.

[15] W. Romeijnders, M. H. van der Vlerk, and W. K. Klein Haneveld. Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Mathematical Programming*, 157:3–46, 2016.

[16] H. Scarf. A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production,* Stanford University Press, pages 201–209, 1958.

[17] R. Schultz. Risk aversion in two-stage stochastic integer programming. In *Stochastic Programming: The State of the Art In Honor of George B. Dantzig,* G. Infanger (ed.), Springer New York, pages 165–187, 2011.

[18] A. Shapiro. Distributionally robust stochastic programming. Available at `http://www.optimization-online.org/DB_HTML/2015/12/5238.html`, 2015.

[19] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14:1237–1249, 2004.

[20] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory,* (Second edition). SIAM Publishers, 2014.

[21] M. H. Van der Vlerk. Convex approximations for a class of mixed-integer recourse models. *Annals of Operations Research*, 177:139–150, 2010.

[22] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[23] W. Xie and S. Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. SIAM Journal on Optimization, accepted, 2017.

[24] J. Zackova. On minimax solutions of stochastic linear programming problems. *Casopis pro pestovani matematiky*, 091(4):423–430, 1966.

[25] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, pages 1–32, 2013.