SIAM J. OPTIM.
Vol. 14, No. 4, pp. 1237–1249

ON A CLASS OF MINIMAX STOCHASTIC PROGRAMS*

ALEXANDER SHAPIRO† AND SHABBIR AHMED†

Abstract. For a particular class of minimax stochastic programming models, we show that the problem can be equivalently reformulated into a standard stochastic programming problem. This permits the direct use of standard decomposition and sampling methods developed for stochastic programming. We also show that this class of minimax stochastic programs is closely related to a large family of mean-risk stochastic programs where risk is measured in terms of deviations from a quantile.

Key words. worst case distribution, problem of moments, Lagrangian duality, mean-risk stochastic programs, deviation from a quantile

AMS subject classifications. 90C15, 90C47

DOI. 10.1137/S1052623403434012

1. Introduction.

A wide variety of decision problems under uncertainty involve optimization of an expectation functional. An abstract formulation for such stochastic programming problems is

$$\operatorname*{Min}_{x \in X} \mathbb{E}_P[F(x,\omega)], \tag{1.1}$$

where $X \subseteq \mathbb{R}^n$ is the set of feasible decisions, $F : \mathbb{R}^n \times \Omega \mapsto \mathbb{R}$ is the objective function, and P is a probability measure (distribution) on the space Ω equipped with a sigma algebra $\mathcal{F}$. The stochastic program (1.1) has been studied in great detail, and significant theoretical and computational progress has been achieved (see, e.g., [18] and references therein).

In the stochastic program (1.1) the expectation is taken with respect to the probability distribution P which is assumed to be *known*. However, in practical applications, such a distribution is not known precisely and has to be estimated from data or constructed using subjective judgments. Often, the available information is insufficient to identify a unique distribution. In the absence of full information on the underlying distribution, an alternative approach is as follows. Suppose a set $\mathcal{P}$ of possible probability distributions for the uncertain parameters is known; then it is natural to optimize the expectation functional (1.1) corresponding to the "worst" distribution in $\mathcal{P}$. This leads to the following minimax stochastic program:

$$\operatorname*{Min}_{x \in X} \left\{ f(x) := \sup_{P \in \mathcal{P}} \mathbb{E}_P[F(x,\omega)] \right\}. \tag{1.2}$$

Theoretical properties of minimax stochastic programs have been studied in a number of publications. In that respect we can mention pioneering works of Žáčková [22] and Dupačová [3, 4]. Duality properties of minimax stochastic programs were thoroughly studied in Klein Haneveld [10]; for more recent publications see [19] and references therein. These problems have also received considerable attention in the

*Received by the editors August 29, 2003; accepted for publication (in revised form) February 7, 2004; published electronically August 4, 2004.
http://www.siam.org/journals/siopt/14-4/43401.html

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 (ashapiro@isye.gatech.edu, sahmed@isye.gatech.edu). The research of the second author was supported by the National Science Foundation under grant DMI-0133943.

context of bounding and approximating stochastic programs [1, 7, 9]. A number of authors have proposed numerical methods for minimax stochastic programs. Ermoliev, Gaivoronski, and Nedeva [5] proposed a method based on the stochastic quasi-gradient algorithm and generalized linear programming. A similar approach along with computational experience is reported in [6]. Breton and El Hachem [2] developed algorithms based on bundle methods and subgradient optimization. Riis and Andersen [16] proposed a cutting plane algorithm. Takriti and Ahmed [21] considered minimax stochastic programs with binary decision variables arising in power auctioning applications, and developed a branch-and-cut scheme. All of the above numerical methods require explicit solution of the inner optimization problem $\sup_{P\in\mathcal{P}} \mathbb{E}_P[F(x,\omega)]$ corresponding to the candidate solution x in each iteration. Consequently, such approaches are inapplicable in situations where calculation of the respective expectations numerically is infeasible because the set Ω although finite is prohibitively large, or possibly infinite.

In this paper, we show that a fairly general class of minimax stochastic programs can be equivalently reformulated into standard stochastic programs (involving optimization of expectation functionals). This permits a direct application of powerful decomposition and sampling methods that have been developed for standard stochastic programs in order to solve large-scale minimax stochastic programs. Furthermore, the considered class of minimax stochastic programs is shown to subsume a large family of mean-risk stochastic programs, where the risk is measured in terms of deviations from a quantile.

2. The problem of moments. In this section we discuss a variant of the problem of moments. This will provide us with basic tools for the subsequent analysis of minimax stochastic programs.

Let us denote by $\mathcal{X}$ the (linear) space of all finite signed measures on $(\Omega, \mathcal{F})$. We say that a measure $\mu \in \mathcal{X}$ is nonnegative, and write $\mu \succeq 0$, if $\mu(A) \geq 0$ for any $A \in \mathcal{F}$. For two measures $\mu_1, \mu_2 \in \mathcal{X}$ we write $\mu_2 \succeq \mu_1$ if $\mu_2 - \mu_1 \succeq 0$. That is, $\mu_2 \succeq \mu_1$ if $\mu_2(A) \geq \mu_1(A)$ for any $A \in \mathcal{F}$. It is said that $\mu \in \mathcal{X}$ is a *probability* measure if $\mu \succeq 0$ and $\mu(\Omega) = 1$. For given nonnegative measures $\mu_1, \mu_2 \in \mathcal{X}$ consider the set

$$\mathcal{M} := \{\mu \in \mathcal{X} : \mu_1 \preceq \mu \preceq \mu_2\}. \tag{2.1}$$

Let $\varphi_i(\omega)$, $i = 0, \dots, q$, be real valued measurable functions on $(\Omega, \mathcal{F})$ and $b_i \in \mathbb{R}$, $i = 1, \dots, q$, be given numbers. Consider the problem

$$\begin{array}{ll} \text{Max}_{P\in\mathcal{M}} & \int_\Omega \varphi_0(\omega) dP(\omega) \\ \text{subject to} & \int_\Omega dP(\omega) = 1, \\ & \int_\Omega \varphi_i(\omega) dP(\omega) = b_i, \quad i = 1, \dots, r, \\ & \int_\Omega \varphi_i(\omega) dP(\omega) \leq b_i, \quad i = r+1, \dots, q. \end{array} \tag{2.2}$$

In the above problem, the first constraint implies that the optimization is performed over probability measures, the next two constraints represent moment restrictions, and the set $\mathcal{M}$ represents upper and lower bounds on the considered measures. If the constraint $P \in \mathcal{M}$ is replaced by the constraint $P \succeq 0$, then the above problem (2.2) becomes the classical problem of moments (see, e.g., [13], [20], and references therein). As we shall see, however, the introduction of lower and upper bounds on the considered measures makes the above problem more suitable for an application to minimax stochastic programming.

We make the following assumptions throughout this section:

(A1) The functions $\varphi_i(\omega)$, $i = 0, \ldots, q$, are μ_2-integrable; i.e.,

$$\int_\Omega |\varphi_i(\omega)| d\mu_2(\omega) < \infty, \quad i = 0, \ldots, q.$$

(A2) The feasible set of problem (2.2) is nonempty, and, moreover, there exists a probability measure $P^* \in \mathcal{M}$ satisfying the equality constraints as well as the inequality constraints as equalities, i.e.,

$$\int_\Omega \varphi_i(\omega) dP^*(\omega) = b_i, \quad i = 1, \ldots, q.$$

Assumption (A1) implies that $\varphi_i(\omega)$, $i = 0, \ldots, q$, are P-integrable with respect to all measures $P \in \mathcal{M}$, and hence problem (2.2) is well defined. By assumption (A2), we can make the following change of variables $P = P^* + \mu$, and hence write problem (2.2) in the form

$$\begin{array}{ll} \text{Max}_{\mu \in \mathcal{M}^*} & \int_\Omega \varphi_0(\omega) dP^*(\omega) + \int_\Omega \varphi_0(\omega) d\mu(\omega) \\ \text{subject to} & \int_\Omega d\mu(\omega) = 0, \\ & \int_\Omega \varphi_i(\omega) d\mu(\omega) = 0, \quad i = 1, \ldots, r, \\ & \int_\Omega \varphi_i(\omega) d\mu(\omega) \le 0, \quad i = r+1, \ldots, q, \end{array} \tag{2.3}$$

where

$$\mathcal{M}^* := \left\{ \mu \in \mathcal{X} : \mu_1^* \preceq \mu \preceq \mu_2^* \right\} \tag{2.4}$$

with $\mu_1^* := \mu_1 - P^*$ and $\mu_2^* := \mu_2 - P^*$.

The Lagrangian of problem (2.3) is

$$L(\mu, \lambda) := \int_\Omega \varphi_0(\omega) dP^*(\omega) + \int_\Omega \mathcal{L}_\lambda(\omega) d\mu(\omega), \tag{2.5}$$

where

$$\mathcal{L}_\lambda(\omega) := \varphi_0(\omega) - \lambda_0 - \sum_{i=1}^q \lambda_i \varphi_i(\omega), \tag{2.6}$$

and the (Lagrangian) dual of (2.3) is

$$\begin{array}{ll} \text{Min}_{\lambda \in \mathbb{R}^{q+1}} & \left\{ \psi(\lambda) := \sup_{\mu \in \mathcal{M}^*} L(\mu, \lambda) \right\} \\ \text{subject to} & \lambda_i \ge 0, \quad i = r+1, \ldots, q. \end{array} \tag{2.7}$$

It is straightforward to see that

$$\psi(\lambda) = \int_\Omega \varphi_0(\omega) dP^*(\omega) + \int_\Omega [\mathcal{L}_\lambda(\omega)]_+ d\mu_2^*(\omega) - \int_\Omega [-\mathcal{L}_\lambda(\omega)]_+ d\mu_1^*(\omega), \tag{2.8}$$

where $[a]_+ := \max\{a, 0\}$.

By the standard theory of Lagrangian duality we have that the optimal value of problem (2.3) is always less than or equal to the optimal value of its dual (2.7). It is possible to give various regularity conditions (constraint qualifications) ensuring that the optimal values of problem (2.3) and its dual (2.7) are equal to each other, i.e., that there is no duality gap between problems (2.3) and (2.7). For example, we have (by the theory of conjugate duality [17]) that there is no duality gap between (2.3) and (2.7), and the set of optimal solutions of the dual problem is nonempty and bounded, if and only if the following assumption holds:

(A3) The optimal value of (2.2) is finite, and there exists a feasible solution to (2.2) for all sufficiently small perturbations of the right-hand sides of the (equality and inequality) constraints.

We may refer to [10] (and references therein) for a discussion of constraint qualifications ensuring the "no duality gap" property in the problem of moments.

By the above discussion we have the following result.

PROPOSITION 2.1. *Suppose that the assumptions* (A1)–(A3) *hold. Then problems* (2.2) *and* (2.3) *are equivalent and there is no duality gap between problem* (2.3) *and its dual* (2.7).

Remark 1. The preceding analysis simplifies considerably if the set Ω is finite, say, $\Omega := \{\omega_1, \dots, \omega_K\}$. Then a measure $P \in \mathcal{X}$ can be identified with a vector $p = (p_1, \dots, p_K) \in \mathbb{R}^K$. We have, of course, that $P \succeq 0$ if and only if $p_k \geq 0$, $k = 1, \dots, K$. The set $\mathcal{M}$ can be written in the form

$$\mathcal{M} = \left\{p \in \mathbb{R}^K : \mu_k^1 \leq p_k \leq \mu_k^2,\ k = 1, \dots, K\right\}$$

for some numbers $\mu_k^2 \geq \mu_k^1 \geq 0$, $k = 1, \dots, K$, and problems (2.2) and (2.3) become linear programming problems. In that case the optimal values of problem (2.2) (problem (2.3)) and its dual (2.7) are equal to each other by the standard linear programming duality without a need for constraint qualifications, and the assumption (A3) is superfluous.

Let us now consider, further, a specific case of (2.2), where

$$\mathcal{M} := \left\{\mu \in \mathcal{X} : (1 - \varepsilon_1)P^* \preceq \mu \preceq (1 + \varepsilon_2)P^*\right\}; \tag{2.9}$$

i.e., $\mu_1 = (1 - \varepsilon_1)P^*$ and $\mu_2 = (1 + \varepsilon_2)P^*$ for some reference probability measure P^* satisfying assumption (A2) and numbers $\varepsilon_1 \in [0, 1]$, $\varepsilon_2 \geq 0$. In that case the dual problem (2.7) takes the form

$$\begin{array}{ll} \text{Min}_{\lambda \in \mathbb{R}^{q+1}} & \mathbb{E}_{P^*}\Big\{\varphi_0(\omega) + \eta_{\varepsilon_1, \varepsilon_2}\left[\mathcal{L}_\lambda(\omega)\right]\Big\} \\ \text{subject to} & \lambda_i \geq 0, \quad i = r+1, \dots, q, \end{array} \tag{2.10}$$

where $\mathcal{L}_\lambda(\omega)$ is defined in (2.6) and

$$\eta_{\varepsilon_1, \varepsilon_2}[a] := \begin{cases} -\varepsilon_1 a & \text{if } a \leq 0, \\ \ \ \varepsilon_2 a & \text{if } a > 0. \end{cases} \tag{2.11}$$

Note that the function $\eta_{\varepsilon_1, \varepsilon_2}[\cdot]$ is convex piecewise linear and $\mathcal{L}_\lambda(\omega)$ is affine in λ for every $\omega \in \Omega$. Consequently the objective function of (2.10) is convex in λ. Thus, the problem of moments (2.2) has been reformulated as a convex stochastic programming problem (involving optimization of the expectation functional) of the form (1.1).

3. A class of minimax stochastic programs. We consider a specific class of minimax stochastic programming problems of the form

$$\operatorname*{Min}_{x \in X} f(x), \tag{3.1}$$

where $f(x)$ is the optimal value of the problem

$$\begin{array}{ll} \text{Max}_{P \in \mathcal{M}} & \int_\Omega F(x, \omega) dP(\omega) \\ \text{subject to} & \int_\Omega dP(\omega) = 1, \\ & \int_\Omega \varphi_i(\omega) dP(\omega) = b_i, \quad i = 1, \dots, r, \\ & \int_\Omega \varphi_i(\omega) dP(\omega) \leq b_i, \quad i = r+1, \dots, q, \end{array} \tag{3.2}$$

and $\mathcal{M}$ is defined as in (2.9). Of course, this is a particular form of the minimax stochastic programming problem (1.2) with the set $\mathcal{P}$ formed by probability measures $P \in \mathcal{M}$ satisfying the corresponding moment constraints.

We assume that the set X is nonempty and assumptions (A1)–(A3) of section 2 hold for the functions $\varphi_i(\cdot)$, $i = 1, \dots, q$, and $\varphi_0(\cdot) := F(x, \cdot)$ for all $x \in X$. By the analysis of section 2 (see Proposition 2.1 and dual formulation (2.10)) we then have that the minimax problem (3.1) is equivalent to the stochastic programming problem:

$$\begin{array}{lll} \text{Min}_{(x,\lambda)\in\mathbb{R}^{n+q+1}} & \mathbb{E}_{P^*}[H(x,\lambda,\omega)] \\ \text{subject to} & x \in X \ \text{ and } \ \lambda_i \geq 0, \quad i = r+1, \dots, q, \end{array} \tag{3.3}$$

where

$$H(x,\lambda,\omega) := F(x,\omega) + \eta_{\varepsilon_1,\varepsilon_2}\left[F(x,\omega) - \lambda_0 - \sum_{i=1}^{q} \lambda_i \varphi_i(\omega)\right]. \tag{3.4}$$

Note that by reformulating the minimax problem (3.1) into problem (3.3), which is a standard stochastic program involving optimization of an expectation functional, we avoid explicit solution of the inner maximization problem with respect to the probability measures. The reformulation, however, introduces $q + 1$ additional variables.

For problems with a prohibitively large (or possibly infinite) support Ω, a simple but effective approach to attacking (3.3) is the *sample average approximation* (SAA) method. The basic idea of this approach is to replace the expectation functional in the objective by a sample average function and to solve the corresponding SAA problem. Depending on the structure of the objective function $F(x, \omega)$ and hence $H(x, \lambda, \omega)$, a number of existing stochastic programming algorithms can be applied to solve the obtained SAA problem. Under mild assumptions, the SAA method has been shown to have attractive convergence properties. For example, a solution to the SAA problem quickly converges to a solution to the true problem as the sample size N is increased. Furthermore, by repeated solutions of the SAA problem, statistical confidence intervals on the quality of the corresponding SAA solutions can be obtained. Detailed discussion of the SAA method can be found in [18, Chapter 6] and references therein.

3.1. Stochastic programs with convex objectives. In this section, we consider minimax stochastic programs (3.1) corresponding to stochastic programs where the objective function is convex. Note that if the function $F(\cdot, \omega)$ is convex for every $\omega \in \Omega$, then the function $f(\cdot)$, defined as the optimal value of (3.2), is given by the maximum of convex functions and hence is convex. Not surprisingly, the reformulation preserves convexity.

PROPOSITION 3.1. *Suppose that the function $F(\cdot, \omega)$ is convex for every $\omega \in \Omega$. Then for any $\varepsilon_1 \in [0, 1]$ and $\varepsilon_2 \geq 0$ and every $\omega \in \Omega$, the function $H(\cdot, \cdot, \omega)$ is convex and*

$$\partial H(x,\lambda,\omega) = \begin{cases} (1-\varepsilon_1)\partial F(x,\omega) \times \{\varepsilon_1 \varphi(\omega)\} & \text{if } N(x,\lambda,\omega) < 0, \\ (1+\varepsilon_2)\partial F(x,\omega) \times \{-\varepsilon_2 \varphi(\omega)\} & \text{if } N(x,\lambda,\omega) > 0, \\ \cup_{\tau\in[-\varepsilon_1,\varepsilon_2]}(1+\tau)\partial F(x,\omega) \times \{-\tau\varphi(\omega)\} & \text{if } N(x,\lambda,\omega) = 0, \end{cases} \tag{3.5}$$

where the subdifferentials $\partial H(x, \lambda, \omega)$ and $\partial F(x, \omega)$ are taken with respect to (x, λ) and x, respectively, and

$$N(x,\lambda,\omega) := F(x,\omega) - \lambda_0 - \sum_{i=1}^{q} \lambda_i \varphi_i(\omega), \quad \varphi(\omega) := (1, \varphi_1(\omega), \dots, \varphi_q(\omega)).$$

Proof. Consider function $\psi(z) := z + \eta_{\varepsilon_1,\varepsilon_2}[z]$. We can write

$$H(x,\lambda,\omega) = \psi\big(N(x,\lambda,\omega)\big) + \lambda_0 + \sum_{i=1}^{q} \lambda_i \varphi_i(\omega).$$

For any $\omega \in \Omega$, the function $N(\cdot,\cdot,\omega)$ is convex, and for $\varepsilon_1 \in [0,1]$ and $\varepsilon_2 \geq 0$, the function $\psi(\cdot)$ is monotonically nondecreasing and convex. Convexity of $H(\cdot,\cdot,\omega)$ then follows. The subdifferential formula (3.5) is obtained by the chain rule. ∎

Let us now consider instances of (3.3) with a finite set of realizations of ω:

$$\text{(3.6)} \qquad \begin{array}{ll} \text{Min}_{(x,\lambda)\in\mathbb{R}^{n+q+1}} & \left\{ h(x,\lambda) := \sum_{k=1}^{K} p_k^* H(x,\lambda,\omega_k) \right\} \\ \text{subject to} & x \in X \;\text{ and }\; \lambda_i \geq 0, \quad i = r+1,\dots,q, \end{array}$$

where $\Omega = \{\omega_1,\dots,w_K\}$ and $P^* = (p_1^*,\dots,p_K^*)$. The above problem can either correspond to a problem with finite support of ω or may be obtained by sampling as in the SAA method. Problem (3.6) has a nonsmooth convex objective function, and often can be solved by using cutting plane or bundle type methods that use subgradient information (see, e.g., [8]). By the Moreau–Rockafellar theorem we have that

$$\text{(3.7)} \qquad \partial h(x,\lambda) = \sum_{k=1}^{K} p_k^* \partial H(x,\lambda,\omega_k),$$

where all subdifferentials are taken with respect to (x,λ). Together with (3.5) this gives a formula for a subgradient of $h(\cdot,\cdot)$, given subgradient information for $F(\cdot,\omega)$.

3.2. Two-stage stochastic programs. A wide variety of stochastic programs correspond to optimization of the expected value of a future optimization problem. That is, let $F(x,\omega)$ be defined as the optimal value function

$$\text{(3.8)} \qquad F(x,\omega) := \underset{y\in Y(x,\omega)}{\text{Min}} G_0(x,y,\omega),$$

where

$$\text{(3.9)} \qquad Y(x,\omega) := \{y \in Y : G_i(x,y,\omega) \leq 0,\; i = 1,\dots,m\},$$

Y is a nonempty subset of a finite dimensional vector space, and $G_i(x,y,\omega)$, $i = 0,\dots,m$, are real valued functions. Problem (1.1), with $F(x,\omega)$ given in the form (3.8), is referred to as a two-stage stochastic program, where the first-stage variables x are decided prior to the realization of the uncertain parameters, and the second-stage variables y are decided after the uncertainties are revealed. The following result shows that a minimax problem corresponding to a two-stage stochastic program is itself a two-stage stochastic program.

PROPOSITION 3.2. *If $F(x,\omega)$ is defined as in* (3.8)*, then the function $H(x,\lambda,\omega)$, defined in* (3.4)*, is given by*

$$\text{(3.10)} \qquad H(x,\lambda,\omega) = \inf_{y\in Y(x,\omega)} \mathcal{G}(x,\lambda,y,\omega),$$

where

$$\text{(3.11)} \qquad \mathcal{G}(x,\lambda,y,\omega) := G_0(x,y,\omega) + \eta_{\varepsilon_1,\varepsilon_2}\left[G_0(x,y,\omega) - \lambda_0 - \sum_{i=1}^{q} \lambda_i \varphi_i(\omega) \right].$$

Proof. The result follows by noting that

$$\mathcal{G}(x,\lambda,y,\omega) = \psi\left(G_0(x,y,\omega) - \lambda_0 - \sum_{i=1}^{q}\lambda_i\varphi_i(\omega)\right) + \lambda_0 + \sum_{i=1}^{q}\lambda_i\varphi_i(\omega),$$

and the function $\psi(z) := z + \eta_{\varepsilon_1,\varepsilon_2}[z]$ is monotonically nondecreasing for $\varepsilon_1 \le 1$ and $\varepsilon_2 \ge 0$. ◻

By the above result, if the set $\Omega := \{\omega_1,\dots,\omega_K\}$ is finite, then the reformulated minimax problem (3.3) can be written as one large-scale optimization problem:

$$\begin{array}{ll} \operatorname{Min}_{x,\lambda,y_1,\dots,y_K} & \sum_{k=1}^{K} p_k^* \mathcal{G}(x,\lambda,y_k,\omega_k) \\ \text{subject to} & y_k \in Y(x,\omega_k), \quad k=1,\dots,K, \\ & x \in X,\ \lambda_i \ge 0, \quad i=r+1,\dots,q. \end{array} \tag{3.12}$$

A particularly important case of two-stage stochastic programs are the two-stage stochastic (mixed-integer) linear programs, where $F(x,\omega) := V(x,\xi(\omega))$ and $V(x,\xi)$ is given by the optimal value of the problem:

$$\begin{array}{ll} \operatorname{Min}_y & c^Tx + q^Ty, \\ \text{subject to} & Wy = h - Tx, \quad y \in Y. \end{array} \tag{3.13}$$

Here $\xi := (q,W,h,T)$ represents the uncertain (random) parameters of problem (3.13), and X and Y are defined by linear constraints (and possibly with integrality restrictions). By applying standard linear programming modelling principles to the piecewise linear function $\eta_{\varepsilon_1,\varepsilon_2}$, we obtain that $H(x,\lambda,\xi(\omega))$ is given by the optimal value of the problem:

$$\begin{array}{ll} \operatorname{Min}_{y,u^+,u^-} & c^Tx + q^Ty + \varepsilon_1 u^- + \varepsilon_2 u^+ \\ \text{subject to} & Wy = h - Tx, \\ & u^+ - u^- = c^Tx + q^Ty - \varphi^T\lambda, \\ & y \in Y,\ u^+ \ge 0, \quad u^- \ge 0, \end{array} \tag{3.14}$$

where $\varphi := (1,\varphi_1(\omega),\dots,\varphi_q(\omega))^T$. As before, if the set $\Omega := \{\omega_1,\dots,\omega_K\}$ is finite, then the reformulated minimax problem (3.3) can be written as one large-scale mixed-integer linear program:

$$\begin{array}{ll} \operatorname{Min}_{x,\lambda,y,u^+,u^-} & c^Tx + \sum_{k=1}^{K} p_k^*\left(q_k^Ty_k + \varepsilon_1 u_k^- + \varepsilon_2 u_k^+\right) \\ \text{subject to} & W_ky_k = h_k - T_kx, \quad k=1,\dots,K, \\ & u_k^+ - u_k^- = c^Tx + q_k^Ty_k - \varphi_k^T\lambda, \quad k=1,\dots,K, \\ & y_k \in Y,\ u_k^+ \ge 0,\ u_k^- \ge 0, \quad k=1,\dots,K, \\ & x \in X. \end{array} \tag{3.15}$$

The optimization model stated above has a block-separable structure which can, in principle, be exploited by existing decomposition algorithms for stochastic (integer) programs. In particular, if Y does not have any integrality restrictions, then the L-shaped (or Benders) decomposition algorithm and its variants can be immediately applied (see, e.g., [18, Chapter 3]).

4. Connection to a class of mean-risk models. Note that the stochastic program (1.1) is risk-neutral in the sense that it is concerned with the optimization of an expectation objective. To extend the stochastic programming framework to a risk-averse setting, one can adopt the *mean-risk* framework advocated by Markowitz and further developed by many others. In this setting the model (1.1) is extended to

$$\operatorname*{Min}_{x\in X}\ \mathbb{E}[F(x,\omega)]+\gamma\mathcal{R}[F(x,\omega)], \tag{4.1}$$

where $\mathcal{R}[Z]$ is a dispersion statistic of the random variable Z used as a measure of risk, and γ is a weighting parameter to trade-off mean with risk. Classically, the variance statistic has been used as the risk-measure. However, it is known that many typical dispersion statistics, including variance, may cause the mean-risk model (4.1) to provide inferior solutions. That is, an optimal solution to the mean-risk model may be stochastically dominated by another feasible solution. Recently, Ogryczak and Ruszczyński [15] have identified a number of statistics which, when used as the risk-measure $\mathcal{R}[\cdot]$ in (4.1), guarantee that the mean-risk solutions are consistent with stochastic dominance theory. One such dispersion statistic is

$$h_\alpha[Z]:=\mathbb{E}\big\{\alpha[Z-\kappa_\alpha]_+ +(1-\alpha)[\kappa_\alpha-Z]_+\big\}, \tag{4.2}$$

where $0\le\alpha\le1$ and $\kappa_\alpha=\kappa_\alpha(Z)$ denotes the α-quantile of the distribution of Z. Recall that κ_α is said to be an α-quantile of the distribution of Z if $Pr(Z<\kappa_\alpha)\le\alpha\le Pr(Z\le\kappa_\alpha)$, and the set of α-quantiles forms the interval $[a,b]$ with $a:=\inf\{z:Pr(Z\le z)\ge\alpha\}$ and $b:=\sup\{z:Pr(Z\ge z)\le\alpha\}$. In particular, if $\alpha=\frac{1}{2}$, then $\kappa_\alpha(Z)$ becomes the median of the distribution of Z and

$$h_\alpha[Z]=\tfrac{1}{2}\mathbb{E}\left|Z-\kappa_{1/2}\right|,$$

and it represents half of the mean absolute deviation from the median.

In [15], it is shown that mean-risk models (4.1), with $\mathcal{R}[\cdot]:=h_\alpha[\cdot]$ and $\gamma\in[0,1]$, provide solutions that are consistent with stochastic dominance theory. In the following, we show that minimax models (3.3) provide a new insight into mean-risk models (4.1).

Consider functions $\mathcal{L}_\lambda(\omega)$ and $\eta_{\varepsilon_1,\varepsilon_2}[a]$, defined in (2.6) and (2.11), respectively. These functions can be written in the form

$$\mathcal{L}_\lambda(\omega)=Z(\omega)-\lambda_0\ \text{ and }\ \eta_{\varepsilon_1,\varepsilon_2}[a]=(\varepsilon_1+\varepsilon_2)\left(\alpha[a]_+ +(1-\alpha)[-a]_+\right),$$

where $Z(\omega):=\varphi_0(\omega)-\sum_{i=1}^q\lambda_i\varphi_i(\omega)$ and $\alpha:=\varepsilon_2/(\varepsilon_1+\varepsilon_2)$, and hence

$$\eta_{\varepsilon_1,\varepsilon_2}\left[\mathcal{L}_\lambda(\omega)\right]=(\varepsilon_1+\varepsilon_2)\left(\alpha[Z(\omega)-\lambda_0]_+ +(1-\alpha)[\lambda_0-Z(\omega)]_+\right). \tag{4.3}$$

We obtain that for fixed λ_i, $i=1,\dots,q$, and positive ε_1 and ε_2, a minimizer $\bar\lambda_0$ of $\mathbb{E}_{P^*}\left\{\eta_{\varepsilon_1,\varepsilon_2}\left[\mathcal{L}_\lambda(\omega)\right]\right\}$ over $\lambda_0\in\mathbb{R}$ is given by an α-quantile of the distribution of the random variable $Z(\omega)$, defined on the probability space $(\Omega,\mathcal{F},P^*)$. In particular, if $\varepsilon_1=\varepsilon_2$, then $\bar\lambda_0$ is the median of the distribution of Z. It follows that if ε_1 and ε_2 are positive, then the minimum of the expectation in (3.3), with respect to $\lambda_0\in\mathbb{R}$, is attained at an α-quantile of the distribution of $F(x,\omega)-\sum_{i=1}^q\lambda_i\varphi_i(\omega)$ with respect to the probability measure P^*. In particular, if the moment constraints are not present in (3.2), i.e., $q=0$, then problem (3.3) can be written as follows:

$$\operatorname*{Min}_{x\in X}\mathbb{E}_{P^*}\left[F(x,\omega)\right]+(\varepsilon_1+\varepsilon_2)h_\alpha[F(x,\omega)], \tag{4.4}$$

where h_α is defined as in (4.2). The above discussion leads to the following result.

PROPOSITION 4.1. *The mean-risk model* (4.1) *with* $\mathcal{R}[\cdot] := h_\alpha[\cdot]$ *is equivalent to the minimax model* (3.3) *with* $\varepsilon_1 = \gamma(1-\alpha)$, $\varepsilon_2 = \alpha\gamma$, *and* $q = 0$.

The additional term $(\varepsilon_1 + \varepsilon_2)h_\alpha[F(x,\omega)]$, which appears in (4.4), can be interpreted as a regularization term. We conclude this section by discussing the effect of such regularization.

Consider the case when the function $F(\cdot,\omega)$ is convex and piecewise linear for all $\omega \in \Omega$. This is the case, for example, when $F(x,\omega)$ is the value function of the second-stage linear program (3.13) without integrality restrictions. Consider the stochastic programming problem (with respect to the reference probability distribution P^*)

$$\operatorname*{Min}_{x\in X} \mathbb{E}_{P^*}[F(x,\omega)] \tag{4.5}$$

and the corresponding mean-risk or minimax model (4.4). Suppose that X is polyhedral, the support Ω of ω is finite, and both problems (4.4) and (4.5) have finite optimal solutions. Then from the discussion at the end of section 3, the problems (4.4) and (4.5) can be stated as linear programs. Let S_0 and $S_{\varepsilon 1,\varepsilon 2}$ denote the sets of optimal solutions of (4.5) and (4.4), respectively. Then by standard theory of linear programming, we have that, for all $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ sufficiently small, the inclusion $S_{\varepsilon_1,\varepsilon_2} \subset S_0$ holds. Consequently, the term $(\varepsilon_1+\varepsilon_2)h_\alpha[F(x,\omega)]$ has the effect of regularizing the solution set of the stochastic program (4.5). We further illustrate this regularization with an example.

Example 1. Consider the function $F(x,\omega) := |\omega - x|$, $x, w \in \mathbb{R}$, with ω having the reference distribution $P^*(\omega = -1) = p_1^*$ and $P^*(\omega = 1) = p_2^*$ for some $p_1^* > 0$, $p_2^* > 0$, $p_1^* + p_2^* = 1$. We then have that

$$\mathbb{E}_{P^*}[F(x,\omega)] = p_1^*|1+x| + p_2^*|1-x|.$$

Let us first discuss the case where $p_1^* = p_2^* = \frac{1}{2}$. Then the set S_0 of optimal solutions of the stochastic program (4.5) is given by the interval $[-1,1]$. For $\varepsilon_2 > \varepsilon_1$ and $\varepsilon_1 \in (0,1)$, the corresponding α-quantile $\kappa_\alpha(F(x,\omega))$, with $\alpha := \varepsilon_2/(\varepsilon_1+\varepsilon_2)$, is equal to the largest of the numbers $|1-x|$ and $|1+x|$, and for $\varepsilon_2 = \varepsilon_1$ the set of α-quantiles is given by the interval with the end points $|1-x|$ and $|1+x|$. It follows that, for $\varepsilon_2 \geq \varepsilon_1$, the mean-risk (or minimax) objective function in problem (4.4),

$$f(x) := \mathbb{E}_{P^*}\big[F(x,\omega)\big] + (\varepsilon_1+\varepsilon_2)h_\alpha[F(x,\omega)],$$

is given by

$$f(x) = \begin{cases} \frac{1}{2}(1-\varepsilon_1)|1-x| + \frac{1}{2}(1+\varepsilon_1)|1+x| & \text{if } x \geq 0, \\ \frac{1}{2}(1+\varepsilon_1)|1-x| + \frac{1}{2}(1-\varepsilon_1)|1+x| & \text{if } x < 0. \end{cases}$$

Consequently, $S_{\varepsilon_1,\varepsilon_2} = \{0\}$. Note that for $x = 0$, the random variable $F(x,\omega)$ has minimal expected value and variance zero (with respect to the reference distribution P^*). Therefore it is not surprising that $x = 0$ is the unique optimal solution of the mean-risk or minimax problem (4.4) for any $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > 0$.

Suppose now that $p_2^* > p_1^*$. In that case $S_0 = \{1\}$. Suppose, further, that $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 \geq \varepsilon_1$, and hence $\alpha \geq \frac{1}{2}$. Then for $x \geq 0$ the corresponding α-quantile $\kappa_\alpha(F(x,\omega))$ is equal to $|1-x|$ if $\alpha < p_2^*$, $\kappa_\alpha(F(x,\omega)) = 1+x$ if $\alpha > p_2^*$, and $\kappa_\alpha(x)$ can be any point on the interval $\big[|1-x|, 1+x\big]$ if $\alpha = p_2^*$. Consequently, for $\alpha \leq p_2^*$ and $x \geq 0$,

$$f(x) = (p_1^* + \varepsilon_2 p_1^*)(1+x) + (p_2^* - \varepsilon_2 p_1^*)|1-x|.$$

It follows then that $S_{\varepsilon_1,\varepsilon_2} = \{1\}$ if and only if $p_1^* + \varepsilon_2 p_1^* < p_2^* - \varepsilon_2 p_1^*$. Since $\alpha \le p_2^*$ means that $\varepsilon_2 \le (p_2^*/p_1^*)\varepsilon_1$, we have that for ε_2 in the interval $[\varepsilon_1, (p_2^*/p_1^*)\varepsilon_1]$, the set $S_{\varepsilon_1,\varepsilon_2}$ coincides with S_0 if and only if $\varepsilon_2 < (p_2^*/p_1^* - 1)/2$. For ε_2 in this interval we can view $\bar{\varepsilon}_2 := (p_2^*/p_1^* - 1)/2$ as the breaking value of the parameter ε_2; i.e., for ε_2 bigger than $\bar{\varepsilon}_2$ an optimal solution of the minimax problem moves away from the optimal solution of the reference problem.

Suppose now that $p_2^* > p_1^*$ and $\alpha \ge p_2^*$. Then for $x \ge 0$,

$$f(x) = (p_1^* + \varepsilon_1 p_2^*)(1 + x) + (p_2^* - \varepsilon_1 p_2^*)|1 - x|.$$

In that case the breaking value of ε_1, for $\varepsilon_1 \le (p_1^*/p_2^*)\varepsilon_2$, is $\bar{\varepsilon}_1 := (1 - p_1^*/p_2^*)/2$.

5. Numerical results. In this section we describe some numerical experiments with the proposed minimax stochastic programming model. We consider minimax extensions of two-stage stochastic linear programs with finite support of the random problem parameters. We assume that $q = 0$ (i.e., that the moment constraints are not present in the model) since, in this case, the minimax problems are equivalent to mean-risk extensions of the stochastic programs, where risk is measured in terms of quantile deviations.

Recall that, owing to the finiteness of the support, the minimax problems reduce to the specially structured linear programs (3.15). We use an ℓ_∞–trust-region based decomposition algorithm for solving the resulting linear programs. The method along with its theoretical convergence properties is described in [12]. The algorithm has been implemented in ANSI C with the GNU Linear Programming Kit (GLPK) [14] library routines to solve linear programming subproblems. All computations have been carried out on a Linux workstation with dual 2.4 GHz Intel Xeon processors and 2 GB RAM.

The stochastic linear programming test problems in our experiments are derived from those used in [11]. We consider the problems `LandS`, `gbd`, `20term`, and `storm`. Data for these instances are available from the website http://www.cs.wisc.edu/∼swright/stochastic/sampling. These problems involve extremely large numbers of scenarios (joint realizations of the uncertain problem parameters). Consequently, for each problem, we consider three instances each with 1000 sampled scenarios. The reference distribution P^* for these instances corresponds to equal weights assigned to each sampled scenario.

Recall that a minimax model with parameters ε_1 and ε_2 is equivalent to a mean-risk model (involving quantile deviations) with parameters $\gamma := \varepsilon_1 + \varepsilon_2$ and $\alpha := \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$. We consider α values of 0.5, 0.7, and 0.9, and ε_1 values of 0.1, 0.3, 0.5, 0.7, and 0.9. Note that the values of the parameters ε_2 and γ are uniquely determined by $\varepsilon_2 = \alpha\varepsilon_1/(1-\alpha)$ and $\gamma = \varepsilon_1/(1-\alpha)$. Note also that some combinations of ε_1 and α are such that $\gamma > 1$, and consequently the resulting solutions are not guaranteed to be consistent with stochastic dominance.

First, for each problem, the reference stochastic programming models (with $\varepsilon_1 = \varepsilon_2 = 0$) corresponding to all three generated instances were solved. Next, the minimax stochastic programming models for the various ε_1-α combinations were solved for all instances. Various dispersion statistics corresponding to the optimal solutions (from the different models) with respect to the reference distribution P^* were computed. Table 5.1 presents the results for the reference stochastic program corresponding to the four problems. The first six rows of the table display various cost-statistics corresponding to the optimal solution with respect to P^*. The presented data is the average over the three instances. The terms "Abs Med-Dev," "Abs Dev," "Std Dev," "Abs SemiDev," and "Std SemiDev" stand for the statistics mean absolute deviation from the median, mean absolute deviation, standard deviation, absolute semideviation, and standard semideviation, respectively. The last two rows of the table display the average (over the three instances) number of iterations and CPU seconds required. Tables 5.2–5.4 present the results for the problem `LandS`. Each table in this set corresponds to a particular α value, and each column in a table corresponds to a particular ε_1 value. The statistics are organized in the rows as in Table 5.1. Similar results are available from the authors for the problems `gbd`, `20term`, and `storm`. In Table 5.5, we present the statistics corresponding to $\alpha = 0.7$ and $\varepsilon_1 = 0.5$ for these three problems.

For a fixed level of α, increasing ε_1 corresponds to increasing the allowed perturbation of the reference distribution in the minimax model, and to increasing the weight γ for the risk term in the mean-risk model. Consequently, we observe from the tables that this leads to solutions with higher expected costs. We also observe that the value of some of the dispersion statistics decreases, indicating a reduction in risk. Similar behavior occurs upon increasing α corresponding to a fixed level of ε_1.

A surprising observation from the numerical results is that the considered problem instances are very robust with respect to perturbations of the reference distribution P^*. Even with large perturbations of the reference distribution, the perturbations of the optimal objective function values are relatively small.

A final observation from the tables is the large variability of computational effort for the various ε_1-α combinations. This can be somewhat explained by the regularization nature of the minimax (or mean-risk) objective function as discussed in section 4. For certain ε_1-α combinations, the piecewise linear objective function may become very sharp, resulting in faster convergence of the algorithm.

TABLE 5.1
Statistics corresponding to the reference stochastic program.

	`LandS`	`gbd`	`20term`	`storm`
Expected cost	225.52	1655.54	254147.15	15498557.91
Abs Med-Dev	46.63	502.01	10022.59	304941.12
Abs Dev	46.95	539.63	10145.86	313915.60
Std Dev	59.26	715.33	12079.76	371207.13
Abs SemiDev	23.47	269.81	5072.93	156957.80
Std SemiDev	44.55	605.01	8824.36	261756.11
Iterations	47.33	57.33	275.33	5000.00
CPU seconds	0.67	0.67	32.33	2309.33

TABLE 5.2
Statistics for problem `LandS` *with* $\alpha = 0.5$.

	$\varepsilon_1 = 0.1$	$\varepsilon_1 = 0.3$	$\varepsilon_1 = 0.5$	$\varepsilon_1 = 0.7$	$\varepsilon_1 = 0.9$
Expected cost	225.57	225.74	225.99	226.39	226.95
Abs Med-Dev	45.91	45.03	44.41	43.74	43.04
Abs Dev	46.24	45.38	44.70	44.15	43.47
Std Dev	58.28	57.16	56.41	55.63	54.84
Abs SemiDev	23.12	22.69	22.39	22.08	21.73
Std SemiDev	43.78	42.97	42.48	41.97	41.45
Iterations	3357.33	3357.00	75.00	70.00	67.33
CPU seconds	196.33	195.33	1.00	1.00	1.00

TABLE 5.3
Statistics for problem `LandS` *with* $\alpha = 0.7$.

	$\varepsilon_1 = 0.1$	$\varepsilon_1 = 0.3$	$\varepsilon_1 = 0.5$	$\varepsilon_1 = 0.7$	$\varepsilon_1 = 0.9$
Expected cost	225.603	225.86	226.31	226.92	227.73
Abs Med-Dev	45.69	44.76	43.91	43.14	42.32
Abs Dev	46.01	45.11	44.28	43.54	42.76
Std Dev	57.94	56.79	55.79	54.92	54.01
Abs SemiDev	23.00	22.55	22.14	21.77	21.38
Std SemiDev	43.47	42.69	42.02	41.44	40.85
Iterations	5000.00	72.67	64.67	70.67	68.00
CPU seconds	293.00	1.33	1.00	1.00	1.00

TABLE 5.4
Statistics for problem `LandS` *with* $\alpha = 0.9$.

	$\varepsilon_1 = 0.1$	$\varepsilon_1 = 0.3$	$\varepsilon_1 = 0.5$	$\varepsilon_1 = 0.7$	$\varepsilon_1 = 0.9$
Expected cost	225.66	226.23	227.16	228.24	228.72
Abs Med-Dev	45.44	44.06	42.93	42.06	41.76
Abs Dev	45.77	44.45	43.36	42.49	42.17
Std Dev	57.61	55.95	54.64	53.62	53.26
Abs SemiDev	22.88	22.23	21.68	21.25	21.09
Std SemiDev	43.21	42.13	41.27	40.54	40.28
Iterations	65.67	63.33	59.67	60.00	1700.33
CPU seconds	1.00	1.00	1.00	1.00	95.67

TABLE 5.5
Statistics for problems `gbd`, `20term`, *and* `storm` *with* $\alpha = 0.7$ *and* $\varepsilon_1 = 0.5$.

	`gbd`	`20term`	`storm`
Expected cost	1663.67	254545.40	15499225.25
Abs Med-Dev	483.94	9220.59	303585.26
Abs Dev	523.96	9360.18	312532.81
Std Dev	702.31	11002.47	369731.47
Abs SemiDev	261.98	4680.09	156266.40
Std SemiDev	598.71	7767.96	260501.73
Iterations	71.67	281.33	1718.33
CPU seconds	1.00	34.00	807.33

Acknowledgment. The authors are indebted to two anonymous referees for constructive comments which helped to improve the manuscript.

REFERENCES

[1] J. R. BIRGE AND R. J.-B. WETS, *Computing bounds for stochastic programming problems by means of a generalized moment problem*, Math. Oper. Res., 12 (1987), pp. 149–162.

[2] M. BRETON AND S. EL HACHEM, *Algorithms for the solution of stochastic dynamic minimax problems*, Comput. Optim. Appl., 4 (1995), pp. 317–345.

[3] J. DUPAČOVÁ, *Minimax stochastic programs with nonseparable penalties*, in Optimization Techniques (Proc. Ninth IFIP Conf., Warsaw, 1979), Part 1, Lecture Notes in Control and Inform. Sci. 22, Springer-Verlag, Berlin, 1980, pp. 157–163.

[4] J. DUPAČOVÁ, *The minimax approach to stochastic programming and an illustrative application*, Stochastics, 20 (1987), pp. 73–88.

[5] YU. ERMOLIEV, A. GAIVORONSKI, AND C. NEDEVA, *Stochastic optimization problems with incomplete information on distribution functions*, SIAM J. Control Optim., 23 (1985), pp. 697–716.

[6] A. A. GAIVORONSKI, *A numerical method for solving stochastic programming problems with moment constraints on a distribution function*, Ann. Oper. Res., 31 (1991), pp. 347–370.

[7] H. Gassmann and W. T. Ziemba, *A tight upper bound for the expectation of a convex function of a multi-variate random variable*, in Stochastic Programming 84.I, Math. Programming Stud. 27, North–Holland, Amsterdam, 1986, pp. 39–53.
[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms.* II, Springer-Verlag, Berlin, 1996.
[9] P. Kall, *An upper bound for SLP using first and total second moments*, Ann. Oper. Res., 30 (1991), pp. 670–682.
[10] W. K. Klein Haneveld, *Duality in Stochastic Linear and Dynamic Programming*, Lecture Notes in Econom. and Math. Systems 274, Springer-Verlag, Berlin, 1986.
[11] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behavior for sampling methods for stochastic programming*, Ann. Oper. Res., to appear.
[12] J. Linderoth and S. Wright, *Decomposition algorithms for stochastic programming on a computational grid*, Comput. Optim. Appl., 24 (2003), pp. 207–250.
[13] H. J. Landau, ed., *Moments in Mathematics*, Proc. Sympos. Appl. Math. 37, AMS, Providence, RI, 1987.
[14] A. Makhorin, *GNU Linear Progamming Kit, Reference Manual*, Version 3.2.3, http://www.gnu.org/software/glpk/glpk.html, 2002.
[15] W. Ogryczak and A. Ruszczyński, *Dual stochastic dominance and related mean-risk models*, SIAM J. Optim., 13 (2002), pp. 60–78.
[16] M. Riis and K. A. Andersen, *Applying the Minimax Criterion in Stochastic Recourse Programs*, Technical report 2002/4, Department of Operations Research, University of Aarhus, Aarhus, Denmark, 2002.
[17] R. T. Rockafellar, *Conjugate Duality and Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics 16, SIAM, Philadelphia, 1974.
[18] A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Handbooks in Operations Research and Management Science 10, North–Holland, Amsterdam, 2003.
[19] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic programs*, Optim. Methods Softw., 17 (2002), pp. 523–542.
[20] J. E. Smith, *Generalized Chebychev inequalities: Theory and applications in decision analysis*, Oper. Res., 43 (1995), pp. 807–825.
[21] S. Takriti and S. Ahmed, *Managing Short-Term Electricity Contracts under Uncertainty: A Minimax Approach*, Technical report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 2002.
[22] J. Žáčková, *On minimax solution of stochastic linear programming problems*, Časopis Pěst. Mat., 91 (1966), pp. 423–430.