Optimization Methods and Software, Vol. 17, pp. 523–542

MINIMAX ANALYSIS OF STOCHASTIC PROBLEMS

ALEXANDER SHAPIRO* and ANTON KLEYWEGT

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA

(Received 31 December 2000; In final form 25 September 2001)

In practical applications of stochastic programming the involved probability distributions are never known exactly. One can try to hedge against the worst expected value resulting from a considered set of permissible distributions. This leads to a min–max formulation of the corresponding stochastic programming problem. We show that, under mild regularity conditions, such a min–max problem generates a probability distribution on the set of permissible distributions with the min–max problem being equivalent to the expected value problem with respect to the corresponding weighted distribution. We consider examples of the news vendor problem, the problem of moments and problems involving unimodal distributions. Finally, we discuss the Monte Carlo sample average approach to solving such min–max problems.

Keywords: Stochastic programming; Min–max optimization; Problem of moments; Monte Carlo sampling; Sample average approximation

1 INTRODUCTION

A large class of stochastic optimization problems can be formulated in the form of the minimization of an expected value function $\mathbb{E}_{\mu}[\phi(x, \omega)]$ over a feasible set S (see [4,8,17] and references therein). In such a formulation it is assumed that the probability distribution μ of the involved random variables is known. However, in practical applications the required distributions are never known exactly and can be

*Corresponding author. E-mail: ashapiro@isye.gatech.edu

ISSN 1055-6788 print; ISSN 1029-4937 online
DOI: 10.1080/1055678021000034008

estimated at best from historical data. Even worse, in many cases the required data are either unavailable or unreliable. In such cases the probability distribution to be used is quite often determined by subjective judgment.

An approach that can be used if the probability distribution μ of the random variables in the problem is not known, is as follows. Suppose that a family $\mathcal{A}$ of probability distributions (measures) is constructed which is viewed as the set of conceivable distributions relevant for the considered problem. Then one may hedge against the worst expected value resulting from the distributions in the set $\mathcal{A}$ by solving the min–max stochastic optimization problem

$$\min_{x \in S} \left\{ f(x) := \sup_{\mu \in \mathcal{A}} \mathbb{E}_{\mu}[\phi(x, \omega)] \right\}. \tag{1.1}$$

The min–max approach to optimization certainly is not new. It takes its origins in von Neumann's game theory and was used for a long time in the statistical decision analysis (see, e.g., [1] and references therein). Starting with work of Žáčková [21], min–max stochastic problems of the form (1.1) have been studied in [3,5,6,9,12], for example. The considered set $\mathcal{A}$, of probability distributions, can be "large", e.g., defined by some general properties such as specified moments of the distributions of $\mathcal{A}$, or can consist of a small finite number of distributions. For example, the future demands of products can be modeled with log-normal distributions with means that are increasing, constant, or decreasing functions of time, and prescribed variances. A choice of the family $\mathcal{A}$ used by various authors is discussed in [7].

An alternative to the above min–max approach is to define an *a priori* probability measure on the set $\mathcal{A}$, and hence to reduce the problem to the standard expected value formulation. This is a Bayesian type approach. For example, if the set $\mathcal{A}$ is finite, say $\mathcal{A} = \{\mu_1, \dots, \mu_k\}$, then by assigning *a priori* probabilities $p_1, \dots, p_k$ to the corresponding measures, one obtains the unique distribution $\bar{\mu} := \sum_{i=1}^k p_i \mu_i$ for which the corresponding expected value problem

$$\min_{x \in S} \mathbb{E}_{\bar{\mu}}[\phi(x, \omega)] \tag{1.2}$$

could be solved. That is, one averages over the set of possible distributions. Again, the choice of such an *a priori* distribution over $\mathcal{A}$ is often subjective. We show that under mild regularity conditions, for each min–max problem (1.1) there exists a probability distribution $p_1, \dots, p_k$ (with finite support) over $\mathcal{A}$ such that (1.1) is equivalent to the expected value problem (1.2), where $\bar{\mu} = \sum_{i=1}^k p_i \mu_i$ and each $\mu_i \in \mathcal{A}$. Although such an *a priori* distribution $p_1, \dots, p_k$, associated with the min–max problem (1.1), is not given explicitly, this suggests a relation between the min–max and Bayesian approaches.

If the considered probability distribution μ is given by Dirac[1] measure δ_a, then $\mathbb{E}_\mu[\phi(x, \omega)] = \phi(x, a)$. Therefore, in case the set $\mathcal{A} := \{\delta_a : a \in A\}$ is formed by Dirac measures, the above problem (1.3) becomes the min–max problem

$$\min_{x \in S} \max_{a \in A} \phi(x, a). \tag{1.3}$$

The set A can be viewed as a region of uncertainty of the involved data parameters and the approach associated with the min–max problem (1.3) is often referred to as the robust optimization method. It seems also that the min–max approach to stochastic programming makes the considered numerical problem better conditioned or, in other words, more robust. This, however, requires a further investigation.

2 MIN–MAX STOCHASTIC PROBLEMS

In this section we discuss theoretical properties of min–max stochastic optimization problems of the form (1.1). We assume that $\mathcal{A}$ is a nonempty (not necessarily convex) set of probability measures on measurable space $(\Omega, \mathcal{F})$, S is a subset of $\mathbb{R}^n$ and $\phi : \mathbb{R}^n \times \Omega \to \mathbb{R}$. We also assume that the set S is *closed* and *convex* and there is a convex neighborhood V of the set S such that the following conditions hold.

(A1) For all $x \in V$ and $\mu \in \mathcal{A}$ the function $\phi(x, \cdot)$ is $\mathcal{F}$-measurable and μ-integrable.

[1] Recall that Dirac measure δ_a is the measure of mass one at the point a.

(A2) For every $\omega \in \Omega$ the function $\phi(\cdot, \omega)$ is convex on V.
(A3) For all $x \in V$ the max-function $f(x)$ is finite valued, i.e., $f(x) < +\infty$.

Then it follows that the function $f(\cdot)$ is the supremum of real valued convex functions, and hence is real valued and convex on V.

Denote

$$g_\mu(x) = g(x, \mu) := \mathbb{E}_\mu[\phi(x, \omega)].$$

By the above assumptions, for every $\mu \in \mathcal{A}$ the function $g_\mu(\cdot)$ is real valued and convex on a neighborhood of S and

$$f(x) = \sup_{\mu \in \mathcal{A}} g_\mu(x).$$

Note that since $g(x, \mu)$ is linear in μ, the supremum in (1.1) does not change if the set $\mathcal{A}$ is replaced by its convex hull

$$\mathcal{C} := \operatorname{conv}(\mathcal{A}) = \left\{ \mu = \sum_{i \in I} p_i \mu_i \colon \mu_i \in \mathcal{A},\ p_i > 0, \quad \sum_{i \in I} p_i = 1,\ |I| < \infty \right\}. \tag{2.1}$$

Recall the following definition of a saddle point of the min–max problem (1.1), with $\mathcal{A}$ replaced by its convex hull $\mathcal{C}$.

Definition 2.1 It is said that $(\bar{x}, \bar{\mu}) \in S \times \mathcal{C}$ is a *saddle point* of the problem (1.1) if

$$\bar{x} \in \arg\min_{x \in S} g(x, \bar{\mu}) \quad \text{and} \quad \bar{\mu} \in \arg\max_{\mu \in \mathcal{C}} g(\bar{x}, \mu). \tag{2.2}$$

By the standard min–max theory we have that if there exists a saddle point $(\bar{x}, \bar{\mu})$, then

$$\inf_{x \in S} \sup_{\mu \in \mathcal{C}} g(x, \mu) = \sup_{\mu \in \mathcal{C}} \inf_{x \in S} g(x, \mu), \tag{2.3}$$

and $\bar{x}$ is an optimal solution of (1.1). The problem

$$\max_{\mu\in\mathcal{C}}\left\{h(\mu):=\inf_{x\in S} g(x,\mu)\right\}, \tag{2.4}$$

can be viewed as dual of (1.1). We have that the optimal value of the problem (1.1) is always greater than or equal to the optimal value of its dual (2.4). Existence of a saddle point $(\bar{x},\bar{\mu})$ ensures that problems (1.1) and (2.4) have the same optimal value and $\bar{\mu}$ is an optimal solution of the dual problem (2.4). Also, in that case the set of saddle points is given by the Cartesian product of the sets of optimal solutions of (1.1) and (2.4).

Since $g(\cdot,\bar{\mu})$ and S are convex, we have that $\bar{x}$ is a minimizer of $g(\cdot,\bar{\mu})$ over S iff

$$0\in\partial g_{\bar{\mu}}(\bar{x})+N_S(\bar{x}), \tag{2.5}$$

where $N_S(\bar{x})$ denotes the normal cone to the set S at the point $\bar{x}\in S$ and $\partial g_{\bar{\mu}}(\bar{x})$ is the subdifferential of $g_{\bar{\mu}}(\cdot)$ at $\bar{x}$ ([18]). (Note that by definition, $N_S(\bar{x})=\emptyset$ if $\bar{x}\notin S$, and hence the above condition (2.5) implies that $\bar{x}\in S$.) Denote

$$\mathcal{A}^*(x):=\arg\max_{\mu\in\mathcal{A}} g(x,\mu). \tag{2.6}$$

Since $g(x,\cdot)$ is linear, we have that

$$\operatorname{conv}\left\{\mathcal{A}^*(x)\right\}=\arg\max_{\mu\in\mathcal{C}} g(x,\mu). \tag{2.7}$$

Therefore the second inclusion in (2.2) holds iff there exist $\mu_1,\dots,\mu_k\in\mathcal{A}^*(\bar{x})$ and positive numbers $p_1,\dots,p_k$ such that $\sum_{i=1}^k p_i=1$ and $\bar{\mu}=\sum_{i=1}^k p_i\mu_i$. Moreover,

$$\partial g_{\bar{\mu}}(\bar{x}) \;=\; \partial\left(\sum_{i=1}^k p_i g_{\mu_i}\right)(\bar{x}) \;=\; \sum_{i=1}^k p_i\partial g_{\mu_i}(\bar{x}), \tag{2.8}$$

where the first equality in (2.8) follows by linearity of $g(\bar{x},\cdot)$ and the second is implied by Moreau–Rockafellar theorem. We obtain the following necessary and sufficient conditions for $(\bar{x}, \bar{\mu})$ to be a saddle point, and hence sufficient conditions for $\bar{x}$ to be optimal.

PROPOSITION 2.1 *A point $(\bar{x}, \bar{\mu})$ is a saddle point of the problem* (1.1) *iff there exist $\mu_1, \dots, \mu_k \in \mathcal{A}^*(\bar{x})$ and positive numbers $p_1, \dots, p_k$* such that $\sum_{i=1}^k p_i = 1$, $\bar{\mu} = \sum_{i=1}^k p_i \mu_i$ *and*

$$0 \in \sum_{i=1}^{k} p_i \, \partial g_{\mu_i}(\bar{x}) + N_S(\bar{x}). \tag{2.9}$$

In the last case $\bar{x}$ is an optimal solution of the problem (1.1).

Remark 2.1 Suppose that $\bar{x}$ satisfies the above sufficient conditions, i.e., there exist μ_i and p_i satisfying the assumptions of Proposition 2.1 and such that (2.9) holds. Consider the problem

$$\min_{x \in S} \{ g_{\bar{\mu}}(x) = \mathbb{E}_{\bar{\mu}}[\phi(x, \omega)] \}. \tag{2.10}$$

By (2.5) and (2.8) we have that (2.9) is a necessary and sufficient condition for $\bar{x}$ to be an optimal solution of the problem (2.10). Consequently for any saddle point $(\bar{x}, \bar{\mu})$, the point $\bar{x}$ is an optimal solution of the expected value problem (2.10). Such distribution $\bar{\mu}$ can be viewed as a *worst probability distribution* associated with the min–max problem (1.1).

Under certain regularity assumptions (discussed below) the above conditions (2.9) are also necessary for optimality of $\bar{x}$. Suppose that the following formula for the subdifferential of the max-function $f(x)$ holds

$$\partial f(\bar{x}) = \operatorname{conv} \left\{ \bigcup_{\mu \in \mathcal{A}^*(\bar{x})} \partial g_{\mu}(\bar{x}) \right\}. \tag{2.11}$$

Regularity conditions which are required for (2.11) to hold will be discussed later. Note that by (2.8) we have that

$$\operatorname{conv} \left\{ \bigcup_{\mu \in \mathcal{A}^*(\bar{x})} \partial g_{\mu}(\bar{x}) \right\} = \bigcup_{\mu \in \operatorname{conv}\{\mathcal{A}^*(\bar{x})\}} \partial g_{\mu}(\bar{x}). \tag{2.12}$$

PROPOSITION 2.2 *Let $\bar{x}$ be an optimal solution of the problem* (1.1) *and suppose that formula* (2.11) *holds. Then there exist measures $\mu_1, \dots, \mu_k \in \mathcal{A}^*(\bar{x})$, with $k \le n+1$, and positive numbers $p_1, \dots, p_k$ such that $p_1 + \cdots + p_k = 1$, and $\bar{x}$ is an optimal solution of the problem* (2.10) *with $\bar{\mu} := \sum_{i=1}^k p_i \mu_i$. Moreover, the optimal values of problems* (1.1) *and* (2.10) *are equal to each other.*

Proof Since problem (1.1) is convex and $f(x)$ is finite valued for all x in a neighborhood of $\bar{x}$, we have that the following first-order optimality condition holds at the point $\bar{x}$:

$$0 \in \partial f(\bar{x}) + N_S(\bar{x}). \tag{2.13}$$

Condition (2.13) means that there exists a subgradient $z \in \partial f(\bar{x})$ such that $-z \in N_S(\bar{x})$. Now, by formula (2.11), there exist $\mu_1, \dots, \mu_k \in \mathcal{A}^*(\bar{x})$ and $z_i \in \partial g_{\mu_i}(\bar{x})$, $i = 1, \dots, k$, such that $z = \sum_{i=1}^k p_i z_i$ for some positive numbers $p_1, \dots, p_k$ satisfying $p_1 + \cdots + p_k = 1$. Moreover, since vector z is n-dimensional, the above holds with $k \le n+1$. Define $\bar{\mu} := \sum_{i=1}^k p_i \mu_i$. Then it follows that $z \in \partial g_{\bar{\mu}}(\bar{x})$, and hence condition (2.5) follows. Since problem (2.10) is convex, it follows from condition (2.5) that $\bar{x}$ is an optimal solution of the corresponding problem (2.10).

Finally, since $g_{\bar{\mu}}(\bar{x})$ is equal to the optimal value of (2.10) and $g_{\mu_i}(\bar{x})$, $i = 1, \dots, k$, are equal to the optimal value of (1.1), the optimal values of problems (1.1) and (2.10) are the same. ■

Remark 2.2 By the above arguments the point $(\bar{x}, \bar{\mu})$, constructed in Proposition 2.2, is a saddle point of the problem (1.1). Therefore, existence of an optimal solution of (1.1) together with the corresponding formula (2.11), ensure existence of a saddle point of (1.1).

Next we discuss sufficient conditions for (2.11) to hold. Formula (2.11) certainly holds if the set $\mathcal{A}$ is finite. If $\mathcal{A}$ is infinite the situation is more subtle of course. In that case we need to equip $\mathcal{A}$ with a topology such that the following properties hold.

(B1) The set $\mathcal{A}$ is compact (in the considered topology of $\mathcal{A}$).
(B2) For every x in a neighborhood of $\bar{x}$ the function $\mu \mapsto g(x, \mu)$ is upper semicontinuous on $\mathcal{A}$.

Note that it follows from the above assumptions that the set $\mathcal{A}^*(\bar{x})$ is nonempty and compact. Note also that by assumptions (A1) and (A2) we have that for every $\mu \in \mathcal{A}$ the function $g_\mu(\cdot)$ is convex and real valued on a neighborhood of $\bar{x}$, and hence is continuous at $\bar{x}$. It follows then by a general result about subdifferentials of the supremum of a family of convex functions ([10, p. 201, Theorem 3]) that $\partial f(\bar{x})$ is equal to the topological closure (in the standard topology of $\mathbb{R}^n$) of the set given in the right hand side of (2.11). We obtain that, under the above assumptions, formula (2.11) holds if the set in the right hand side of (2.11) is closed.

Let us observe that since $f(x)$ is convex and real valued for all x in a neighborhood of $\bar{x}$, it is continuous at $\bar{x}$, and hence $\partial f(\bar{x})$ is bounded. It follows that the set inside the parentheses in the right hand side of (2.11) is bounded.

Consider the point-to-set mapping $\mathcal{G}(\mu) := \partial g_\mu(\bar{x})$, from $\mathcal{A}$ into the set of subsets of $\mathbb{R}^n$. It is said that $\mathcal{G}$ is closed if its graph

$$\mathrm{gph}(\mathcal{G}) := \{(\mu, z)\colon \mu \in \mathcal{A},\ z \in \mathcal{G}(\mu)\}$$

is a closed subset of $\mathcal{A} \times \mathbb{R}^n$. If $\mathcal{G}$ is closed, then it follows by compactness of $\mathcal{A}^*(\bar{x})$ that the set inside the parentheses in the right hand side of (2.11) is closed, and hence is compact. Since the convex hull of a compact set in $\mathbb{R}^n$ is compact, it follows then that the set in the right hand side of (2.11) is closed. By the definition of the subdifferential $\partial g_\mu(\bar{x})$ we have that

$$\mathcal{G}(\mu) = \left\{z \in \mathbb{R}^n\colon g_\mu(x) - g_\mu(\bar{x}) \geq z \cdot (x - \bar{x}),\ \forall\, x \in \mathbb{R}^n\right\}.$$

Therefore, it follows by the upper semicontinuity of $g(x, \cdot)$ that $\mathcal{G}$ is closed if, in addition, $g(\bar{x}, \cdot)$ is continuous on $\mathcal{A}$, i.e., the following condition holds.

(B3) The function $\mu \mapsto g(\bar{x}, \mu)$ is continuous on $\mathcal{A}$.

We obtain the following result.

Proposition 2.3 *Suppose that assumptions* (A1)–(A3) *and* (B1)–(B3) *are satisfied. Then formula* (2.11) *holds.*

Propositions 2.2 and 2.3 imply that if assumptions (A1)–(A3) and (B1)–(B2) hold and assumption (B3) is satisfied with respect to an optimal solution $\bar{x}$ of (1.1), then the min–max problem (1.1) is equivalent to the expected value problem (2.10). The corresponding numbers $p_1, \dots, p_k$ can be viewed as a probability measure, with the finite support $\{\mu_1, \dots, \mu_k\}$, on the set $\mathcal{A}$. Of course, the numbers p_i are not given explicitly and, in general, in order to find these numbers one needs to solve the corresponding min–max problem.

Regarding assumptions (B1)–(B3) let us observe the following. If Ω is a metric space and $\mathcal{F}$ is its Borel sigma algebra, we can equip $\mathcal{A}$ with the standard weak topology (e.g., [2]). In that case conditions for compactness of $\mathcal{A}$ are well known; by Prohorov's theorem $\mathcal{A}$ is compact if (and in case Ω is separable, only if) $\mathcal{A}$ is tight and closed (see, e.g., [2, Section 5]). Recall that it is said that $\mathcal{A}$ is tight if for any $\varepsilon > 0$ there exists a compact set $\Xi \subset \Omega$ such that $\mu(\Xi) > 1 - \varepsilon$ for every $\mu \in \mathcal{A}$. In particular, if Ω is a compact metric space equipped with its Borel sigma algebra $\mathcal{F}$, then any closed (in the weak topology) set $\mathcal{A}$ of probability measures on $(\Omega, \mathcal{F})$ is weakly compact. Also if, for some $x \in V$, the function $\phi(x, \cdot)$ is bounded and continuous (upper semicontinuous) on Ω, then the mapping $\mu \mapsto g(x, \mu)$ is continuous (upper semicontinuous) in the weak topology.

By the above discussion together with propositions 2.2 and 2.3 we obtain the following results.

THEOREM 2.1 *Let $\bar{x}$ be an optimal solution of the problem* (1.1). *Suppose that:* (i) *assumptions* (A1)–(A3) *hold,* (ii) Ω *is a metric space and $\mathcal{F}$ is its Borel sigma algebra,* (iii) *the set $\mathcal{A}$ is tight and closed* (*in the weak topology*)*,* (iv) *for every x in a neighborhood of $\bar{x}$ the function $\phi(x, \cdot)$ is bounded and upper semicontinuous on Ω and the function $\phi(\bar{x}, \cdot)$ is bounded and continuous on Ω.*

Then there exist measures $\mu_1, \dots, \mu_k \in \mathcal{A}^*(\bar{x})$, with $k \leq n+1$, and positive numbers $p_1, \dots, p_k$ such that $p_1 + \cdots + p_k = 1$, and for $\bar{\mu} := \sum_{i=1}^{k} p_i \mu_i$ the following holds: (a) $(\bar{x}, \bar{\mu})$ *is a saddle point of* (1.1)*,* (b) $\bar{\mu}$ *is an optimal solution of the dual problem* (2.4)*,* (c) *the optimal values of* (1.1) *and its dual* (2.4) *are equal to each other,* (d) $\bar{x}$ *is an optimal solution of the problem* (2.10) *and the optimal values of problems* (1.1) *and* (2.10) *are equal to each other.*

Remark 2.3 If the set $\mathcal{A}$ is *finite*, then assumptions (B1)–(B3) hold automatically. In that case assumptions (ii)–(iv) in the above theorem are superfluous.

3 EXAMPLES AND DISCUSSION

In this section we discuss several examples which demonstrate various aspects of the min–max approach.

3.1 The News Vendor Problem

The news vendor problem (see, e.g., [4, Section 1.1]) can be formulated as the problem of minimization of the function

$$g_\mu(x) := -ax + \int_0^x F(w)\, dw, \tag{3.1}$$

over $x \in \mathbb{R}_+$. Here $a \in [0, 1]$ is a parameter, $F(\cdot)$ is a cumulative distribution function (cdf) on $\mathbb{R}_+$, and μ is the probability measure corresponding to $F(\cdot)$. The cdf $F(\cdot)$ represents a probability distribution of the associated demand per day. Suppose now that such probability distribution cannot be specified exactly, and only a family $\mathcal{A}$ of relevant cumulative distribution functions on $\mathbb{R}_+$ can be specified. Consider the corresponding max-function

$$f(x) := -ax + \sup_{F \in \mathcal{A}} \int_0^x F(w)\, dw, \tag{3.2}$$

and let

$$h_F(x) := \int_0^x F(w)\, dw \quad \text{and} \quad h(x) := \sup_{F \in \mathcal{A}} h_F(x).$$

Since $0 \leq F(\cdot) \leq 1$, it follows that $h_F(\cdot)$ is Lipschitz continuous with Lipschitz constant 1 for each F, and thus $h(\cdot)$ is also Lipschitz continuous with Lipschitz constant 1. Hence there exists a function $\bar{F}(\cdot)$, which is given by the derivative of $h(\cdot)$ almost everywhere, such that $h(x) = \int_0^x \bar{F}(w)\, dw$. That is,

$$f(x) = -ax + \int_0^x \bar{F}(w)\, dw.$$

Moreover, the function $\bar{F}$ has the following properties. Since each $F \in \mathcal{A}$ is nondecreasing, it follows that each $h_F(\cdot)$ is convex, and thus $h(\cdot)$ is convex, and hence $\bar{F}(\cdot)$ is nondecreasing, $\bar{F}(w) = 0$ for all $w < 0$ and $\bar{F}(w) \to 1$ as $w \to +\infty$, and finally $\bar{F}(\cdot)$ can be chosen to be right continuous. Consequently, $\bar{F}(\cdot)$ in itself is a cumulative probability distribution function.

It follows that the corresponding min–max stochastic problem can be formulated as the news vendor problem with the cdf $\bar{F}$. Note that the cdf $\bar{F}$ does not depend on the parameter a.

3.2 Problem of Moments

Let Ω be a metric space and $\mathcal{F}$ be its Borel sigma algebra. Let $\psi_1(\omega), \ldots, \psi_m(\omega)$ be real valued measurable functions on $(\Omega, \mathcal{F})$ and $\mathcal{A}$ be the set of probability measures μ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{E}_\mu[\psi_j(\omega)] = b_j, \quad j = 1, \ldots, m, \tag{3.3}$$

for some $b_1, \ldots, b_m \in \mathbb{R}$. Then, for a fixed x, the max-problem inside the parentheses in (1.1) becomes the classical problem of moments (see, e.g., [15] and references therein for the historical background of the problem of moments). Note that here the set $\mathcal{A}$ is convex.

The following result is due to Rogosinsky [19].

Lemma 3.1 *Let μ be a probability measure on $(\Omega, \mathcal{F})$ such that $\psi_1(\omega), \ldots, \psi_m(\omega)$ are μ-integrable. Then there exists a probability measure μ' on $(\Omega, \mathcal{F})$ with a finite support of at most $m + 1$ points such that $\mathbb{E}_\mu[\psi_j(\omega)] = \mathbb{E}_{\mu'}[\psi_j(\omega)]$ for all $j = 1, \ldots, m$.*

By the above lemma it suffices to solve the corresponding max-problem with respect to measures with a finite support of at most $s = m + 1$ points. That is, the corresponding min–max problem can be written in the form

$$\min_{x \in S} \left\{ \begin{array}{ll} \max & \sum_{i=1}^s p_i \phi(x, \omega_i) \\ \text{subject to} & \sum_{i=1}^s p_i \psi_j(\omega_i) = b_j, \quad j = 1, \ldots, m \\ & \sum_{i=1}^s p_i = 1, \quad p_i \geq 0, \quad i = 1, \ldots, s \end{array} \right\}, \tag{3.4}$$

where the maximum inside the parentheses in (3.4) is taken with respect to $p \in \mathbb{R}^s_+$ and $\omega_1, \ldots, \omega_s \in \Omega$.

Let us make the following assumptions: (i) the set S is compact, (ii) assumptions (A1)–(A3) hold, (iii) the set Ω is compact, and (iv) the functions $\psi_1(\cdot), \dots, \psi_m(\cdot)$ and $\phi(x, \cdot)$, $x \in V$, are continuous on Ω.

By Prohorov's theorem it follows from the above assumption (iii) that the set of all probability measures on $(\Omega, \mathcal{F})$ is compact (in the weak topology). Moreover, since $\psi_1, \dots, \psi_m$ are continuous on Ω, the set $\mathcal{A}$ is a closed subset of the set of all probability measures, and hence is also compact. Note also that the function $f(x)$ is convex on a neighborhood of S and hence is continuous on S, and hence the corresponding min–max problem has an optimal solution. We obtain then by Theorem 2.1 that the optimal value of the min–max problem (3.4) coincides with the optimal value of its dual

$$\max_{\mu \in \mathcal{A}} \left\{ h(\mu) := \inf_{x \in S} \mathbb{E}_\mu[\phi(x, \omega)] \right\}. \tag{3.5}$$

Denote by A_s be the set of measures $\mu \in \mathcal{A}$ with a finite support of at most s points. The set A_s is a closed subset of $\mathcal{A}$, and hence is compact. Note that, unless $|\Omega| \le s$, the set A_s is not convex. With such notation, and $s = m + 1$, we can write the min–max problem (3.4) as follows

$$\min_{x \in S} \max_{\mu \in A_s} \mathbb{E}_\mu[\phi(x, \mu)]. \tag{3.6}$$

We have by Theorem 2.1 that, under the above assumptions, the min–max problem (3.6) has a saddle point $(\bar{x}, \bar{\mu})$ with measure $\bar{\mu}$ having a finite support of at most $(n + 1)(m + 1)$ points. It follows that the optimal value of (3.6) is equal to the optimal value of its dual, which is obtained by interchanging the order of min and max operators in (3.6). Consequently, we obtain that the optimal value of the dual of (3.6) is equal to the optimal value of (3.5), and hence $\bar{\mu}$ is an optimal solution of the dual problem (3.5). Let us summarize the above discussion in the following proposition.

PROPOSITION 3.1 *Suppose that the assumptions* (i)–(iv) *are satisfied. Then the following holds:* (a) *problems* (3.4) *and* (3.5) *possess optimal solutions* $\bar{x}$ *and* $\bar{\mu}$*, respectively, with* $\bar{\mu}$ *having a finite support of at most* $(n + 1)(m + 1)$ *points,* (b) $(\bar{x}, \bar{\mu})$ *is a saddle point of the corresponding min–max problem,* (c) *the optimal values of* (3.4) *and* (3.5) *are equal* *to each other,* (d) $\bar{x}$ *is an optimal solution of the expected value problem* (2.10)*, and the optimal values of* (2.10) *and* (3.4) *are equal to each other.*

In particular if $m = 0$, i.e., $\mathcal{A}$ is the set of all probability measures on $(\Omega, \mathcal{F})$, then the primal problem (3.4) is reduced to the min–max problem

$$\min_{x \in S} \max_{\omega \in \Omega} \phi(x, \omega). \tag{3.7}$$

Also by the above discussion we have that, under the assumptions (i)–(iii), it suffices to solve its dual problem (3.5) with respect to probability measures with a finite support of at most $n + 1$ points. Moreover, if Ω is a convex subset of a normed space and the function $\phi(x, \cdot)$ is concave for any $x \in S$, then the set $\arg\max_{\omega \in \Omega} \phi(x, \omega)$ is convex. Therefore, in that case it suffices to solve the dual problem (3.5) with respect to measures of mass one (Dirac measures), and hence (3.5) is equivalent to the problem which is obtained from (3.7) by interchanging the order of min and max operators.

It is also possible to show (see [11,13]) that the max-problem inside the parentheses in (3.4) has the following dual

$$\begin{array}{ll} \min_{y \in \mathbb{R}^{m+1}} & y_0 + b_1 y_1 + \cdots + b_m y_m \\ \text{subject to} & y_0 + y_1 \psi_1(\omega) + \cdots + y_m \psi_m(\omega) \geq \phi(x, \omega), \quad \forall\, \omega \in \Omega. \end{array} \tag{3.8}$$

Suppose that the set Ω is compact, that for every $x \in V$ the function $\phi(x, \cdot)$ is upper semicontinuous and bounded on Ω, and that the functions $\psi_1, \ldots, \psi_m$ are continuous on Ω. Then the optimal value of the problem inside the parentheses in (3.4) is equal to the optimal value of the problem (3.8) (e.g., [13]), and hence problem (3.4) is equivalent to the problem

$$\begin{array}{ll} \min_{x \in S,\ y \in \mathbb{R}^{m+1}} & y_0 + b_1 y_1 + \cdots + b_m y_m \\ \text{subject to} & y_0 + y_1 \psi_1(\omega) + \cdots + y_m \psi_m(\omega) \geq \phi(x, \omega), \quad \forall\, \omega \in \Omega. \end{array} \tag{3.9}$$

3.3 Unimodal Distributions

A univariate distribution is said to be *unimodal* with mode a if its cumulative (probability) distribution function is convex on the interval $(-\infty, a)$ and concave on the interval $(a, +\infty)$. At the point a the cumulative distribution function can be discontinuous. Equivalently, the distribution is unimodal if it is a mixture of the distribution of mass one at a (Dirac measure δ_a) and a distribution with density function that is nondecreasing on $(-\infty, a]$ and nonincreasing on $[a, +\infty)$. Since we can always translate a unimodal distribution with mode a to a unimodal distribution with mode 0, we assume subsequently that all considered unimodal distributions have mode 0.

By a result due to Khintchine we have that a distribution is unimodal with mode 0 iff it is the distribution of a product UZ, where U and Z are independent random variables, U is uniformly distributed on [0,1] and the distribution of Z is arbitrary. Moreover, the unimodal distribution is supported on interval $[a, b]$, with $a \leq 0 \leq b$, iff the random variable Z supported on $[a, b]$.

Suppose that $\omega = (W_1, \dots, W_k)$ is a k-dimensional random vector such that its components $W_1, \dots, W_k$ are mutually independent and each component W_i has a unimodal distribution with mode 0 and finite support $[a_i, b_i]$. That is, $\Omega = \mathsf{X}_{i=1}^{k} [a_i, b_i]$ is the Cartesian product of the intervals $[a_i, b_i]$, $\mathcal{F}$ is its Borel sigma algebra and the set $\mathcal{A}$ consists of all probability measures on $(\Omega, \mathcal{F})$ with independent components each having a unimodal distribution with mode 0. Let $U_i Z_i$ be Khintchine's representation of W_i and $U = (U_1, \dots, U_k)$, $Z = (Z_1, \dots, Z_k)$ be the corresponding random vectors. Then it follows that for any $\mu \in \mathcal{A}$,

$$\mathbb{E}_\mu[\phi(x, \omega)] = \mathbb{E}_Z\left\{\mathbb{E}_{UZ|Z}[\phi(x, \omega)]\right\} = \mathbb{E}[\psi(x, Z)],$$

where $\psi(x, z)$ is the conditional expectation of $\phi(x, \omega)$ given $Z = z$. Note that it follows from convexity of $\phi(\cdot, \omega)$ that $\psi(\cdot, z)$ is also convex.

We obtain that the primal min–max problem (1.1) can be written here as the problem of maximization of $\mathbb{E}[\psi(x, Z)]$ over all probability measures on $(\Omega, \mathcal{F})$ with independent components, and then minimization over $x \in S$. Clearly the optimal value of such min–max problem will be less than or equal to the optimal value of the min–max problem of maximization of $\mathbb{E}[\psi(x, Z)]$ over all probability measures on $(\Omega, \mathcal{F})$, and then minimization over $x \in S$. As it was shown in the previous section it suffices to solve the last max problem with respect to Dirac measures only, i.e., the last min–max problem is equivalent to the problem

$$\min_{x \in S} \max_{z \in \Omega} \psi(x, z). \tag{3.10}$$

Now any Dirac measure on Ω belongs to the set of probability measures on Ω with independent components. Therefore, we obtain that the primal min–max problem (1.1) is equivalent here to the problem (3.10), and hence to the problem

$$\min_{x \in S} \max_{\mu \in \mathcal{U}} \mathbb{E}_\mu[\phi(x, w)], \tag{3.11}$$

where $\mathcal{U}$ is the set of uniform probability distributions having supports of the form $\mathsf{X}_{i=1}^k L_i$ with either $L_i = [0, b_i']$, $b_i' \leq b_i$, or $L_i = [a_i', 0]$, $a_i \leq a_i'$. Moreover, if the set S is compact, then there exists a saddle point $(\bar{x}, \bar{\mu})$ such that $\bar{\mu}$ is a weighted sum of at most $n + 1$ uniform distributions of the above form.

Similar analysis can be performed if the component $W_1, \ldots, W_k$ are independent, and each W_i has a unimodal and symmetric distribution on interval $[-a_i, a_i]$. Note that in such case the distribution of W_i can be represented as the distribution of $U_i Z_i$, where U_i and Z_i are independent random variables, U_i is uniformly distributed on $[-1, 1]$ and Z_i has an arbitrary distribution on $[0, a_i]$.

4 MONTE CARLO SIMULATION METHOD OF SAMPLE AVERAGE APPROXIMATION

In this section we consider situations where the set $\mathcal{A}$ contains a *finite* number of probability measures. That is, let $\mathcal{A} := \{\mu_1, \ldots, \mu_k\}$ and hence the min–max problem (1.1) becomes

$$\min_{x \in S} \left\{ f(x) := \max_{i=1,\ldots,k} f_i(x) \right\}, \tag{4.1}$$

where

$$f_i(x) := \mathbb{E}_{\mu_i}[\phi(x, \omega)], \quad i = 1, \dots, k.$$

Note that problem (4.1) can be written in the following equivalent form

$$\begin{array}{ll} \min_{x \in S,\ z \in \mathbb{R}} & z \\ \text{subject to} & f_i(x) \le z, \quad i = 1, \dots, k. \end{array} \tag{4.2}$$

Since the functions $f_i(\cdot)$ and the set S are assumed to be convex and the Slater condition for the problem (4.2) always holds, it follows that if $\bar{x}$ is an optimal solution of the problem (4.1) (and hence $(\bar{x}, f(\bar{x}))$ is an optimal solution of (4.2)), then there exist Lagrange multipliers satisfying the corresponding optimality conditions (e.g., [10, p. 68]). That is, there exist $\lambda_i \ge 0$, $i = 1, \dots, k$, such that $\sum_{i=1}^k \lambda_i = 1$, $\lambda_i = 0$ if $f_i(\bar{x}) < f(\bar{x})$, and

$$\bar{x} \in \arg\min_{x \in S} \left[\sum_{i=1}^k \lambda_i f_i(x) \right].$$

These Lagrange multipliers are exactly the weights (probabilities) p_i discussed in Section 2.

We denote by Λ^* the set of all Lagrange multiplier vectors $\lambda = (\lambda_1, \dots, \lambda_k)$ satisfying the above optimality conditions. Note that Λ^* is given by the set of optimal solutions of the dual problem

$$\max_{\lambda \in \Lambda} \min_{x \in S} \sum_{i=1}^k \lambda_i f_i(x), \tag{4.3}$$

where $\Lambda := \left\{ \lambda \in \mathbb{R}^k_+ : \sum_{i=1^k} \lambda_i = 1 \right\}$, and therefore does not depend on $\bar{x}$.

Now let us discuss the sample average approximation (SAA) approach to solving the min–max problem (4.1) or its equivalent (4.2). Let $\hat{f}_i(x)$ be the sample average approximation, based on a random sample of size N_i, of the function $f_i(x)$, $i = 1, \dots, k$. That is, if $\omega^1, \ldots, \omega^{N_i} \sim \mu_i$ is a generated random sample, then $\hat{f}_i(x) := N_i^{-1} \sum_{j=1}^{N_i} \phi(x, \omega^j)$. Then problem

$$\min_{x \in S} \left\{ \hat{f}(x) := \max_{i=1,\ldots,k} \hat{f}_i(x) \right\}, \tag{4.4}$$

provides an approximation of the problem (4.1). Let $N := (N_1, \ldots, N_k)$, and $\hat{v}_N$, $\hat{S}_N$ and $\hat{\Lambda}_N$ be the optimal value, the set of optimal solutions and the set of Lagrange multiplier vectors, respectively, of the approximating problem (4.4).

It is not difficult to show that under mild regularity conditions, the SAA estimators $\hat{v}_N$, $\hat{x}_N \in \hat{S}_N$ and $\hat{\lambda}_N \in \hat{\Lambda}_N$ are consistent estimators of their counterparts of the "true" problem (4.1). This is because of the fact that if each sample average function $\hat{f}_i(x)$ converges with probability one (w.p.1) to the expected value function $f_i(x)$ uniformly on a set $U \subset \mathbb{R}^n$, then the corresponding max-function $\hat{f}(x)$ converges w.p.1 to $f(x)$ uniformly on U.

We say that the Law of Large Numbers (LLN) holds *pointwise*, on the set V, if for any given $x \in V$ and each $i \in \{1, \ldots, k\}$, $\hat{f}_i(x)$ converges with probability one to $f_i(x)$ as $N_i \to \infty$. If each sample is i.i.d., then this holds by the classical (strong) LLN.

PROPOSITION 4.1 *Suppose that the assumptions* (A1)–(A2) *hold, that the "true" problem* (4.1) *has a nonempty and bounded set S^* of optimal solutions and that the LLN holds pointwise. Then $\hat{v}_N$ converges w.p.*1 *to the optimal value of* (4.1)*, and* $\sup_{x \in \hat{S}_N} \mathrm{dist}(x, S^*) \to 0$ *and* $\sup_{\lambda \in \hat{\Lambda}_N} \mathrm{dist}(\lambda, \Lambda^*) \to 0$ *w.p.*1 *as the sample sizes N_i, $i = 1, \ldots, k$, tend to infinity.*

Proof Proof of this proposition is rather standard, we quickly outline it for the sake of completeness. Let U be a convex compact subset of V such that $S^* \subset \mathrm{int}(U)$, where $\mathrm{int}(U)$ denotes the interior of the set U. Let D be a countable dense subset of U. By the pointwise LLN we have that for any $x \in D$, each $\hat{f}_i(x)$ converges w.p.1 to $f_i(x)$, and hence $\hat{f}(x)$ converges w.p.1 to $f(x)$. Since the set D is countable, it follows that the event "$\hat{f}(x)$ converges to $f(x)$ for every $x \in D$" happens w.p.1. Since the SAA functions $\hat{f}(x)$ are convex and D is dense in U, it follows by a result from convex analysis ([18, Theorem 10.8]) that the event "$\hat{f}(x)$ converge to $f(x)$ uniformly on the compact set U" happens w.p.1.

Now let $\tilde{S}_N$ be the set of minimizers of $\hat{f}(x)$ over $x \in S \cap U$. Since $S \cap U$ is compact it follows then by the uniform convergence w.p.1 of $\hat{f}(x)$ to $f(x)$ that $\sup_{x \in \tilde{S}_N} \operatorname{dist}(x, S^*) \to 0$ w.p.1. It remains to note that because of the convexity assumption, $\hat{S}_N = \tilde{S}_N$ provided that $\tilde{S}_N \subset \operatorname{int}(U)$. The assertions follow. ∎

It is a more delicate issue to estimate rates of converges of the SAA estimators. The Central Limit Theorem type results for estimators which are obtained by solving sample average approximations of the problem (4.2) are discussed in [20, Section 6.6].

Finally let us observe that for any $\bar{\lambda} \in \Lambda$ we have the following

$$\begin{aligned}
\inf_{x \in S} f(x) &= \inf_{x \in S} \sup_{\lambda \in \Lambda} \sum_{i=1}^{k} \lambda_i f_i(x) \\
&= \inf_{x \in S} \sup_{\lambda \in \Lambda} \mathbb{E}\left[\sum_{i=1}^{k} \lambda_i \hat{f}_i(x)\right] \\
&\geq \inf_{x \in S} \mathbb{E}\left[\sum_{i=1}^{k} \bar{\lambda}_i \hat{f}_i(x)\right] \geq \mathbb{E}\left[\inf_{x \in S} \sum_{i=1}^{k} \bar{\lambda}_i \hat{f}_i(x)\right].
\end{aligned} \tag{4.5}$$

By solving the SAA problems with respect to the measure $\bar{\mu} := \sum_{i=1}^{k} \bar{\lambda}_i \mu_i$ several times, one can estimate the last expected value in (4.5) by the corresponding average of the obtained optimal values of the SAA problems. This gives a valid statistical lower bound for the optimal value of problem (4.1). For stochastic programming problems this bound was suggested by Norkin *et al.* [16] and Mak *et al.* [14]. Note that if $\bar{\lambda} \in \Lambda^*$, then

$$\inf_{x \in S} f(x) = \inf_{x \in S} \mathbb{E}\left[\sum_{i=1}^{k} \bar{\lambda}_i \hat{f}_i(x)\right].$$

Therefore, a good choice of $\bar{\lambda}$ would be to take $\bar{\lambda}$ from the set Λ^* of Lagrange multipliers. Although a "true" Lagrange multiplier vector usually is unknown, it can be consistently estimated by solving a SAA problem. Consequently, the obtained estimate can be used in the above statistical lower bound procedure.

Acknowledgments

The authors are grateful to J. Dupačová for constructive comments which helped to improve the presentation of this paper. A. Shapiro was supported, in part, by the National Science Foundation under grant DMS-0073770 and A. Kleywegt was supported, in part, by the National Science Foundation under grant DMI-9875400.

References

[1] J. Berger (1984). The robust Bayesian viewpoint. In: J. Kadane (Ed.), *Robustness of Bayesian Analyses*, pp. 63–124. Elsevier Science Publishers.
[2] P. Billingsley (1999). *Convergence of Probability Measures*, 2nd Edn. Wiley, New York.
[3] J.R. Birge and J.R.B. Wets (1986). Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Mathematical Programming Study*, **27**, 54–102.
[4] J.R. Birge and F. Louveaux (1997). *Introduction to Stochastic Programming*. Springer, New York.
[5] J. Dupačová (1980). On minimax decision rule in stochastic linear programming. In: A. Prékopa (Ed.), *Studies on Mathematical Programming*, pp. 47–60. Akadémiai Kiadó, Budapest.
[6] J. Dupačová (1987).The minimax approach to stochastic programming and an illustrative application. *Stochastics*, **20**, 73–88.
[7] J. Dupačová (2001). Stochastic programming: minimax approach. In: Ch. A. Floudas and P.M. Pardalos (Eds.), *Encyclopedia of Optimization*, Vol. V., pp. 327–330. Kluwer.
[8] Y. Ermoliev and R.J.B. Wets (Eds) (1988). *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin.
[9] A.A. Gaivoronski (1991). A numerical method for solving stochastic programming problems with moment constraints on a distribution function. *Annals of Operations Research*, **31**, 347–369.
[10] A.D. Ioffe and V.M. Tihomirov (1979). *Theory of Extremal Problems*. North-Holland Publishing Company, Amsterdam.
[11] K. Isii (1963). On sharpness of Tchebycheff type inequalities. *Ann. Inst. Stat. Math.*, **14**, 185–197.
[12] R. Jagannathan (1977). Minimax procedure for a class of linear programs under uncertainty. *Oper. Research*, **25**, 173–177.
[13] J.H.B. Kemperman (1968). The general moment problem, a geometric approach. *Annals Math. Statist.*, **39**, 93–122.
[14] W.K. Mak, D.P. Morton and R.K. Wood (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, **24**, 47–56.
[15] H.J. Landau (Ed.) (1987). Moments in Mathematics. *Proc. Sympos. Appl. Math.*, **37**. Amer. Math. Soc., Providence, RI.
[16] V.I. Norkin, G.C. Pflug, and A. Ruszczynski (1998). A branch and bound method for stochastic global optimization. *Mathematical Programming*, **83**, 425–450.
[17] A. Prékopa (1995). *Stochastic Programming*. Kluwer Academic Publishers.
[18] R.T. Rockafellar (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

[19] W.W. Rogosinsky (1958). Moments of non-negative mass. *Proc. Roy. Soc. London Ser. A*, **245**, 1–27.
[20] R.Y. Rubinstein and A. Shapiro (1993). *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. Wiley, New York, NY.
[21] J. Žáčková (1966). On minimax solutions of stochastic linear programming problems. *Čas. Pěst. Mat.*, **91**, 423–430.