www.palgrave.com/journals

Stochastic linear optimization under partial uncertainty and incomplete information using the notion of probability multimeasure

Davide La Torre[1,2]* and Franklin Mendivil[3]

[1]*Department of Economics, Management, and Quantitative Methods, University of Milan, Via Conservatorio 7, 20122 Milan, Italy;* [2]*Department of Mathematics, Nazarbayev University, Astana, Kazakhstan; and* [3]*Department of Mathematics and Statistics, Acadia University, Wolfville, NS, Canada*

We consider a scalar stochastic linear optimization problem subject to linear constraints. We introduce the notion of deterministic equivalent formulation when the underlying probability space is equipped with a probability multimeasure. The initial problem is then transformed into a set-valued optimization problem with linear constraints. We also provide a method for estimating the expected value with respect to a probability multimeasure and prove extensions of the classical strong law of large numbers, the Glivenko–Cantelli theorem, and the central limit theorem to this setting. The notion of sampling with respect to a probability multimeasure and the definition of cumulative distribution multifunction are also discussed. Finally, we show some properties of the deterministic equivalent problem.

Journal of the Operational Research Society (2017). doi:10.1057/s41274-017-0249-9

Keywords: stochastic linear optimization; deterministic equivalent problem; set-valued optimization; probability multimeasure

1. Stochastic linear optimization

It is well known that stochastic optimization in both the scalar and vector cases plays a significant role in the analysis, modeling, design, and operation of modern systems. Stochastic optimization refers to a collection of methods for minimizing or maximizing an objective function when randomness is present and, in general, stochastic optimization methods and techniques generalize those used for deterministic problems. In recent years stochastic optimization has become an essential tool for modeling in science, engineering, business, computer science, and statistics. Applications include business and decision making, computer simulations, medicine and laboratory experiments, traffic management, signal analysis, and many others. In practical applications it is easy to find situations in which the decision maker (DM) wishes to optimize an objective which depends on some random parameters.

In financial portfolio management (see Markowitz, 1952) the use of stochastic linear optimization is well known: In fact if $r_j(\omega) \geq 0$, $j = 1 \ldots m$, are the stochastic returns of j financial investments that are depending on the event ω, the portfolio financial decision-making problem can be written as:

$$\max \sum_{j=1}^{m} r_j(\omega) x_j$$

subject to:

$$\begin{cases} \sum_{j=1}^{m} & x_j = 1 \\ x_j \geq 0, & j = 1 \ldots m. \end{cases}$$

In practical cases the DM solves the above problem by taking its deterministic equivalent version that can be formulated by taking into consideration the expected value of each investment, their related covariances, or other criteria such as dividends, liquidity, sustainability (see Markowitz, 1952; Hirschberger *et al*, 2013). More general, let $(\Omega, \mathcal{A}, P)$ be a probability space where Ω is the basic space of events, $\mathcal{A}$ is a σ-algebra and P is a probability measure. The classical formulation of a stochastic linear optimization model is as follows:

$$\max \sum_{j=1}^{m} \alpha_j(\omega) x_j \qquad (SLP)$$

subject to:

$$\begin{cases} Ax = b \\ x_j \geq 0 \quad j = 1 \ldots m. \end{cases}$$

*Correspondence: Davide La Torre, Department of Mathematics, Nazarbayev University, Astana, Kazakhstan.
E-mail: davide.latorre@unimi.it

where ω is an event in the probability space Ω, A is a deterministic matrix of coefficients, b is a deterministic vector, and x is the vector of input variables. One way to simplify and solve the above problem consists of introducing the notion of a deterministic equivalent formulation as follows:

$$\max \sum_{j=1}^{m} \mathbb{E}(\alpha_j)x_j \qquad (DEP)$$

subject to:

$$\begin{cases} Ax = b \\ x_j \geq 0, \quad j = 1 \ldots m. \end{cases}$$

The notion of deterministic equivalent formulation reduces the complexity of the initial stochastic formulation: There is a price to pay of course, and this is mainly related to the loss of information when switching from a stochastic context to a deterministic one.

Very rarely the decision maker has a complete knowledge of this probability distribution, as very often he is subject to incomplete and partial information on the probability distribution P. When such a scenario happens, the formulation of the above deterministic equivalent problem is not so straightforward. Several attempts have been made in the literature to mathematically describe this lack of complete information (Abdelaziz and Masri, 2005; Ben Abdelaziz and Masri, 2010; Bitran, 1980; Ermoliev and Gaivoronski, 1985; Dupacova, 1987; Urli and Nadeau, 1990, 2004), and all of them rely on the imposition of lower and upper bounds for the underlying probability distribution.

Here we propose an innovative approach based on our notion of *probability multimeasure*: This definition allows to formally describe the uncertainty related to the estimation of the probability associated with a certain event. The name probability multimeasure is essentially due to the fact that the probability of an event takes multiple values. Several authors have studied the main properties of this extension of the classical notion of measure including, among others, Radon–Nikodým theorems, martingales (see Artstein, 1972, 1974; Hess, 2002; Hiai, 1978). The aim of this paper is then to analyze and discuss the main properties of the deterministic equivalent problem when the probability measure is replaced by a probability multimeasure: The main difference with respect to the classical context is that now the expectation is replaced by the expected value of a random variable with respect to a probability multimeasure. We first introduce the notion of probability multimeasure and then define a deterministic equivalent problem with respect to this new object. The most important features of this model are the estimation of the expected value of coefficient. This is typically done by assuming an underlying probability distribution of events that allows to estimate the above quantities.

This paper proceeds as follows: Section 2 presents the main mathematical and statistical properties of this object. Section 3 presents the deterministic equivalent problem and studies its main properties. The last section concludes.

2. Imprecise information and the notion of probability multimeasures

In the literature several approaches are available to model the notion of uncertainty in complex systems. In many cases this is done by assuming the existence of an underlying probability measure or distribution, but there are situations where this assumption cannot be made due to the lack of data or the vagueness, imprecision, or incompleteness of the available information. Alternative techniques to describe the level of imprecise information rely on fuzzy sets and set-valued analysis. In both these two contexts the degree of uncertainty is modeled using sets: The idea is that a set can contain all possible outcomes or states of the world without specifying any particular value. Our approach to set-valued measures or multimeasure is a further attempt along this direction: We suppose that the probability associated with a certain event is no longer a number but a compact and convex subset of $\mathbb{R}^d$. We used this definition in other previous papers, mainly dealing with the notion of self-similarity and the extension of the classical Monge–Kantrorovich distance between probability measures (see Kunze *et al*, 2012; Torre and Mendivil, 2007, 2009, 2011, 2015). With respect to other definitions in the literature (see Hess, 2002; Stojaković, 2012) that are essentially based on the notion of selector, this definition allows one to introduce a parametrized family of classical probability measures that are obtained from the multimeasure through the process of scalarization via support function. This approach works well any time one has to deal with abstract integrals with respect to a probability multimeasure as it is possible to reduce the complexity of the set-valued problem to a family of scalar problems and then use classical results.

2.1. Preliminaries on compact convex sets

Let $\mathcal{K}$ denote the collection of all nonempty compact and convex subsets of $\mathbb{R}^d$ with addition and scalar multiplication ($\lambda \in \mathbb{R}$) defined as

$$A + B := \{a + b : a \in A, b \in B\} \quad \text{and} \quad \lambda A = \{\lambda a : a \in A\}.$$

For $A \in \mathcal{K}$, we say that A *is nonnegative* ($A \geq 0$) if $0 \in A$. Given $A \in \mathcal{K}$ the *support function* $\mathrm{spt}(\cdot, A) : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$\mathrm{spt}(p, A) = \sup\{p \cdot a : a \in A\}$$

and one can recover A as

$$A = \bigcap_{\|p\|=1} \{x : x \cdot p \leq \mathrm{spt}(p, A)\}. \qquad (2.1)$$

The support function satisfies the properties that, for all $\lambda \geq 0$ and $A, B \in \mathcal{K}$,

$$\begin{aligned} \mathrm{spt}(p, \lambda A + B) &= \lambda \mathrm{spt}(p, A) + \mathrm{spt}(p, B), \\ \mathrm{spt}(p, -B) &= \mathrm{spt}(-p, B). \end{aligned} \tag{2.2}$$

However, it is usually not the case that $\mathrm{spt}(p, -A) = -\mathrm{spt}(p, A)$. For any $A \in \mathcal{K}$, we define the *norm* of A as

$$\|A\| := \sup\{\|x\| : x \in A\} = \sup_{\|p\|=1} \mathrm{spt}(p, A).$$

It is easy to show that this satisfies the usual properties of a norm.

For $A, B \in \mathcal{K}$, we also have that

$$d_H(A, B) = \sup_{\|p\|=1} |\mathrm{spt}(p, A) - \mathrm{spt}(p, B)|,$$

where d_H is the *Hausdorff metric* on $\mathcal{K}$ (Beer, 1993). Using this fact and properties of the support function, it is easy to show that if $A_n \to A$ and $B_n \to B$ in the Hausdorff metric on $\mathcal{K}$ then $A_n + B_n \to A + B$.

A set $A \subset \mathbb{R}^d$ is *balanced* if $\lambda A \subseteq A$ for all $|\lambda| \leq 1$. A *unit ball* in $\mathbb{R}^d$ is any balanced $\mathbb{B} \in \mathcal{K}$ with $0 \in int(\mathbb{B})$. Any such unit ball defines a norm on $\mathbb{R}^d$ via the Minkowski functional

$$\|x\| = \sup\{\lambda \geq 0 : \lambda x \in \mathbb{B}\}.$$

Whenever we have chosen such a set $\mathbb{B}$, we will always use this induced norm on $\mathbb{R}^d$. The *dual sphere* is defined as

$$\mathbb{S}^* = \{y : \sup\{y \cdot x : x \in \mathbb{B}\} = 1\} \subset \mathbb{R}^d$$

and is also a nonempty compact set. Notice that since $\mathbb{B}$ is compact, for each $y \in \mathbb{S}^*$, there is some $x \in \mathbb{B}$ with $y \cdot x = 1$.

2.2. *Multimeasures*

We provide only basic definitions and those properties of multimeasures that we will need; for more information and proofs see Artstein (1972, 1974), Arstein and Vitale (1975), Aubin and Frankowska (1990), Hiai (1978), Kandilakis (1992), Torre and Mendivil (2011). Given a set Ω and a σ-algebra $\mathcal{A}$ on Ω a *set-valued measure* or *multimeasure* on $(\Omega, \mathcal{A})$ with values in $\mathcal{K}$ is a function $\phi : \mathcal{A} \to \mathcal{K}$ such that $\phi(\emptyset) = \{0\}$ and

$$\phi\left(\bigcup_i A_i\right) = \sum_i \phi(A_i) \tag{2.3}$$

for any sequence of disjoint sets $A_i \in \mathcal{A}$. The left side of (2.3) is the infinite Minkowski sum defined as

$$\sum_i K_i = \left\{\sum_i k_i : k_i \in K_i, \sum_i |k_i| < \infty\right\}.$$

We comment that the left side of (2.3) also converges in the Hausdorff distance on $\mathcal{K}$. The *total variation* of a multimeasure ϕ is defined in the usual way as

$$|\phi|(A) = \sup \sum_i \|\phi(A_i)\|,$$

where the supremum is taken over all finite measurable partitions of $A \in \mathcal{A}$. The set function $|\phi|$ defined in this fashion is a (nonnegative and scalar) measure on Ω. If $|\phi|(\Omega) < \infty$, then ϕ is of *bounded variation*.

We will say that a multimeasure ϕ is *nonnegative* if $\phi(A) \geq 0$ (i.e., $0 \in \phi(A)$) for all A. Nonnegative multimeasures are monotone: If $A \subseteq B$, then $\phi(A) = \{0\} + \phi(A) \subseteq \phi(B \setminus A) + \phi(A) = \phi(B)$. This makes nonnegative multimeasures a nice generalization of (nonnegative) scalar measures. If ϕ is a multimeasure and $p \in \mathbb{R}^d$, then the *scalarization* ϕ^p defined by

$$\phi^p(A) = \mathrm{spt}(p, \phi(A)) \tag{2.4}$$

is a signed measure on Ω and is a measure if ϕ is nonnegative.

One simple way to construct a multimeasure is by integrating a *multifunction density* f with respect to a measure μ:

$$\phi(A) = \int_A f(x)\, \mathrm{d}\mu(x). \tag{2.5}$$

There are several approaches to defining this integral (see Aubin and Frankowska, 1990). For our purpose we only consider $f : \Omega \to \mathcal{K}$ and so we can define the integral in (2.5) as an element of $\mathcal{K}$ via support functions using the property (see Aubin and Frankowska, 1990, Proposition 8.6.2)

$$\mathrm{spt}\left(q, \int_\Omega f(x)\, \mathrm{d}\mu(x)\right) = \int_\Omega \mathrm{spt}(q, f(x))\, \mathrm{d}\mu(x),$$

which defines the set as in (2.1). If the multifunction f is nonnegative (that is, $0 \in f(x)$ for all x), then the resulting multimeasure will also be nonnegative. In addition, if $0 \leq f(x) \leq g(x)$ and ϕ is a positive multimeasure, then (see Torre and Mendivil, 2011)

$$\int f(x)\, \mathrm{d}\phi(x) \subseteq \int g(x)\, \mathrm{d}\phi(x),$$

the convexity of the values of ϕ is crucial. For more results on set-valued analysis see Aubin and Frankowska (1990).

2.3. *Probability multimeasures*

Definition 2.1 (*probability multimeasure*) Let $\mathbb{B} \subset \mathbb{R}^d$ be a unit ball. A $\mathbb{B}$-*probability multimeasure (pmm)* on $(\Omega, \mathcal{A})$ is a nonnegative multimeasure ϕ with $\phi(\Omega) = \mathbb{B}$.

One strong motivation for this definition is that a pmm ϕ defines a parameterized family, ϕ^p for $p \in \mathbb{S}^*$, of probability measures. However, in general ϕ^p and ϕ^q are related and the relationship can be quite complicated (the main constraint on this relationship is that $p \mapsto \phi^p(A)$ is convex).

We can construct a pmm by using a density as in (2.5) and integrate against a finite measure μ. Of course, we need some conditions on f in order to define a pmm. The simplest conditions are to assume that $f(x) \in \mathcal{K}$ is balanced for each x, $\|f(x)\| \leq C$ for some C and all x, and

$$0 \in int \int_\Omega f(x)\ \mathrm{d}\mu = int(\mathbb{B}).$$

In general, it is difficult to choose a density to obtain a given $\mathbb{B}$; it is better to use the integral of the density to define $\mathbb{B}$.

An example of a finitely supported pmm is given in Section 4, so here we give a simple example of a continuous pmm.

Example 2.2 Let μ be any probability measure fully supported on the unit circle $S \subset \mathbb{R}^2$ and for each $x \in S$ let $F(x) = \{\lambda x : -1 \leq \lambda \leq 1\}$. Then (2.5) defines a pmm fully supported on the circle S as well.

In this context, a *random variable* on $(\Omega, \mathcal{A})$ is a Borel measurable function $X : \Omega \to \mathbb{R}$. The *expectation* of X with respect to a pmm ϕ is defined in the usual way as

$$\mathbb{E}_\phi(X) = \int_\Omega X(\omega)\ \mathrm{d}\phi(\omega). \tag{2.6}$$

This integral can also be constructed using support functions (that is, using the ϕ^p) and each part of the decomposition $X = X^+ - X^-$ separately (since support functions work best with nonnegative scalars); see Kandilakis (1992) for another approach. Since $0 \in \phi(A)$ for each A, it is easy to see that $0 \in \mathbb{E}_\phi(X)$ as well.

We easily obtain a version of Chebyshev's inequality in this setting.

Theorem 2.3 (Chebyshev inequality) *Suppose that $f : [0, \infty) \to [0, \infty)$ and is nondecreasing and X is is a nonnegative random variable with $\mathbb{E}_\phi(f(X)) \in \mathcal{K}$. Then for all $a \geq 0$ with $f(a) > 0$,*

$$\phi(X \geq a) \subseteq \frac{\mathbb{E}_\phi(f(X))}{f(a)}. \tag{2.7}$$

Proof We see that

$$\phi(X \geq a) = \int_{X \geq a} 1\ \mathrm{d}\phi(x) = \frac{1}{f(a)} \int_{X \geq a} f(a)\ \mathrm{d}\phi(x)$$
$$\subseteq \frac{1}{f(a)} \int_{X \geq a} f(x)\ \mathrm{d}\phi(x) \subseteq \frac{1}{f(a)} \int_\Omega f(x)\ \mathrm{d}\phi(x).$$

□

2.4. Statistical properties of probability multimeasures

In this section we provide extensions of the strong law of large numbers, the Glivenko–Cantelli theorem, and the central limit theorem. To do this, we introduce the notion of a cumulative distribution multifunction associated with a probability multimeasure.

2.4.1. Samples and the strong law of large numbers The strong law of large numbers is so fundamental that, in order to be useful, any theory of set-valued probability should have an analogous result. However, as we will see, the idea of an iid sequence of samples is fundamentally different in the set-valued case; the standard framework does not work. Recall that, given a probability measure μ on $\mathbb{R}$, a the standard construction of an iid sample from μ is any element of the infinite product space $\mathbb{R}^{\mathbb{N}}$ equipped with the infinite product measure generated by μ on each factor.

This construction does not work in the set-valued context; the construction breaks down even for the product of two multimeasures. Thus, another approach is required. We have chosen to use the path of Radon–Nikodym derivatives of a pmm with respect to a probability measure. This allows us to convert the context from that of probability multimeasures to the setting of random sets, where there is a wealth of results.

Proposition 2.4 *Any probability multimeasure is of bounded variation.*

Proof To show this, let $e_i^* \in \mathbb{S}^*$ be a basis for $(\mathbb{R}^d)^*$. Then there is a $K > 0$ so that $\|x\| \leq K \sum_i |e_i^*(x)|$ since all norms on $\mathbb{R}^d$ are equivalent. Now let $C \in \mathcal{K}$. Then

$$\|C\| = \sup_{c \in C} \|c\| \leq K \sup_{c \in C} \sum_i |e_i^*(c)| \leq K \sum_i |\mathrm{spt}(e_i^*, C)| + |\mathrm{spt}(-e_i^*, C)|.$$

Using this, for any finite measurable partition $\{A_j\}$ of A

$$\sum_j \|\phi(A_i)\| \leq \sum_j K \sum_i \phi^{-e_i^*}(A_j) + \phi^{e_i^*}(A_j)$$
$$= K \sum_i \phi^{-e_i^*}\left(\bigcup_j A_j\right) + \phi^{e_i^*}\left(\bigcup_j A_j\right)$$
$$\leq K \sum_i \phi^{-e_i^*}(\Omega) + \phi^{e_i^*}(\Omega) \leq 2dK,$$

since each ϕ^q is a probability measure. This shows that $\|\phi\| \leq 2dK$. □

Let the probability measure μ_ϕ be defined by $\mu_\phi(A) = |\phi|(A)/|\phi|(\Omega)$. Then $\phi(A) = \{0\}$ whenever $\mu_\phi(A) = 0$ (that is, $\phi \ll \mu_\phi$) and thus by the Radon–Nikodym theorem for multimeasures (see Hiai, 1978, Corollary 5.3) there is a multifunction f_ϕ with compact and convex values such that

$$\phi(A) = \int_A f_\phi(x)\ \mathrm{d}\mu_\phi(x).$$

Notice that $f_\phi : \Omega \to \mathcal{K}$ is a random set when we use the probability measure μ_ϕ on Ω. In addition, notice that $\|f_\phi(x)\| \leq |\phi|(\Omega)$ for all x.

Definition 2.5 *(i.i.d. sample)* Let ϕ be a $\mathbb{B}$-pmm on Ω and $X : \Omega \to \mathbb{R}$ be a random variable. Then by an *iid sample from* (X, ϕ) we mean an element from the product space

$$\Xi := \{(X(\omega_1)f_\phi(\omega_1), X(\omega_2)f_\phi(\omega_2), \ldots) : \omega_i \in \Omega\} \subseteq \mathcal{K}^{\mathbb{N}},$$

where we place the product measure on Ξ induced by μ_ϕ on each factor.

Unlike in the case of scalar probability, a sample needs to include some "set-valued" information along with the sample values from the random variable X. It is too much to hope that a sequence of scalar samples would allow us to recover the set-valued expectation (2.6); this is unfortunate but unavoidable.

Theorem 2.6 (Strong law of large numbers) *Suppose that* $\mathbb{E}_{\mu_\phi}(|X|) < \infty$ *and let* $x_n f_\phi(x_n)$ *be an i.i.d. sample from* (X, ϕ). *Then almost surely*

$$\lim_N \frac{1}{N} \sum_{n \leq N} x_n f_\phi(x_n) = \mathbb{E}_\phi(X),$$

where the set convergence is in the Hausdorff distance.

Proof The function $\omega \mapsto X(\omega) f_\phi(\omega)$ is a random set and $\mathbb{E}_{\mu_\phi}(\|Xf_\phi\|) < \infty$ by our assumption. Thus, by the strong law of large numbers for random sets (Arstein and Vitale, 1975; Molchanov, 2005), we have that

$$\lim_N \frac{1}{N} \sum_{n \leq N} x_n f_\phi(x_n) = \mathbb{E}_{\mu_\phi}(Xf_\phi) = \int_\Omega X(\omega) f_\phi(\omega)\, d\mu_\phi(\omega) = \mathbb{E}_\phi(X)$$

almost surely. □

2.4.2. Cumulative distribution multifunctions and the Glivenko–Cantelli Theorem For $x, y \in \mathbb{R}^m$, we define $x \leq y$ if $x_i \leq y_i$ for $i = 1, 2, \ldots, m$ and also define the set $(-\infty, x] := \{y \in \mathbb{R}^m : y \leq x\}$. Using these notions, we say that a multifunction $F : \mathbb{R}^m \to \mathcal{K}$ is *increasing* if $x \leq y$ implies $F(y) = F(x) + A$ with $A \geq 0$.

Given a pmm ϕ on $\mathbb{R}^m$ the cumulative distribution multifunction (cdmf) is defined in the usual way as

$$F_\phi(x) = \phi((-\infty, x]).$$

It is easy to see that F_ϕ is a nonnegative and increasing multifunction which is cádlág in that $F(x) = \lim_n F(x_n) = \cap_n F(x_n)$ whenever $x_n \searrow x \in \mathbb{R}^m$ and $\lim_n F(x_n) = \cup_n F(x_n)$ exists whenever $x_n \nearrow x \in \mathbb{R}^m$ (these limits also exist in the Hausdorff distance on $\mathcal{K}$).

We can also convert from a cdmf to a pmm; for simplicity we restrict attention to one-dimension.

Theorem 2.7 (a cdmf induces a pmm) *Let* $F : \mathbb{R} \to \mathcal{K}$ *be a càdlàg nonnegative increasing multifunction with* $\cap_x F(x) = \{0\}$ *and* $\cup_x F(x) = \mathbb{B}$. *Then there is a* $\mathbb{B}$-*pmm* ϕ *so that* $F(x) = \phi((-\infty, x])$.

Proof Take $a < b$. Then $F(b) = F(a) + A_a^b$, for some nonnegative $A_a^b \in \mathcal{K}$. Define $\phi((a, b]) = A_a^b$ and let $\mathcal{B}$ be the algebra generated by sets of the form $(a, b]$. Using the obvious modification of standard arguments (see for example [?, Chapter 12]), it is possible to show that ϕ defines a countably additive multimeasure on $\mathcal{B}$. In addition, ϕ^p extends to a Borel probability measure for all $p \in \mathbb{S}^*$. Thus, by (Kandilakis, 1992, Theorem 2.6) there is a multimeasure extension of ϕ to the Borel σ-algebra; this extension is clearly the desired pmm. □

Given an i.i.d. sample $x_i f_\phi(x_i)$ from (X, ϕ), we can construct the *empirical cdmf* of this sample

$$F_n(z) = \frac{1}{n} \sum_{i \leq n} f_\phi(x_i)\, \mathbb{1}_{\{z \leq x_i\}}(z). \tag{2.8}$$

Theorem 2.8 (Glivenko–Cantelli) *We have that as* $n \to \infty$,

$$\sup_{z \in \mathbb{R}} \sup_{p \in \mathbb{S}^*} |\mathrm{spt}(p, F_n(z)) - \mathrm{spt}(p, F(z))| \to 0 \quad \text{almost surely.}$$

In particular, we have $F_n(z) \to F(z)$ *in the Hausdorff distance uniformly in* z.

Proof Let $M = |\phi|(\Omega) < \infty$. Since $\|f_\phi(z)\| \leq M$ and $\|F(z)\| \leq M$ for all x, we also have $\|F_n(z)\| \leq M$ for all z. Thus, as a function of $p \in \mathbb{S}^*$, both $\mathrm{spt}(p, F(x))$ and $\mathrm{spt}(p, F_n(x))$ for all n and x are Lipschitz with factor at most M.

Let $\epsilon > 0$ be given and $q_1, q_2, q_\ell \in \mathbb{S}^*$ be such that they form an $\epsilon/(3M)$-cover of $\mathbb{S}^*$. This means that for any $q \in \mathbb{S}^*$ there is some i so that $|\mathrm{spt}(q, G) - \mathrm{spt}(q_i, G)| < \epsilon/3$ where G is any one of $F_n(x)$ or $F(x)$, for any n or x.

By the Glivenko–Cantelli theorem, for large enough n we have almost surely

$$\sup_{z \in \mathbb{R}} \sup_{1 \leq i \leq \ell} |\mathrm{spt}(q_i, F_n(z)) - \mathrm{spt}(q_i, F(z))| < \epsilon/3$$

this, and the choice of the q_i gives the desired result. □

2.4.3. Central limit theorem The theory of random sets also contains versions of many standard results from probability theory (see Molchanov, 2005; Cascales *et al*, 2007; Cressie, 1979; Puri and Ralescu, 1983; Rockafellar and Wets, 1998). One example of this is the central limit theorem. Here we briefly discuss how the CLT for random sets translates into our setting. For simplicity we restrict to nonnegative random variables X.

The standard CLT characterizes the distributional behavior of the averages $(1/n)\sum_{i \leq n}(Z_i - \mathbb{E}(Z))$. However, since there is no analogue of subtraction in the arithmetic of sets, we have to be content with analyzing the behavior of the distance between the sample average and the expected value. The appropriate distance to use is the Hausdorff distance. A *random Gaussian variable* ξ in a Banach space $\mathbb{Y}$ is a random variable with values in $\mathbb{Y}$ and such that $y^*(\xi)$ is a scalar Gaussian random variable for all $y^* \in \mathbb{Y}^*$.

Theorem 2.9 (Central limit theorem) *Suppose that* $\mathbb{E}_{\mu_\phi}(|X|^2) < \infty$ *and let* $x_n f_\phi(x_n)$ *be an i.i.d. sample from* (X, ϕ)*. Then*

$$\sqrt{n}\, d_H\Big(\frac{1}{n}\sum_{i\le n} x_i f_\phi(x_i), \mathbb{E}_\phi(X)\Big) \xrightarrow[\;]{[} \text{distribution} \sup_{p\in\mathbb{S}^*} \|\xi(p)\|,$$

where ξ *is a centered Gaussian random variable in* $C(\mathbb{S}^*)$ *with covariance structure*

$$\Gamma_X(p,q) := \mathrm{spt}(\mathbb{E}_\phi[\mathrm{spt}(Xf_\phi(X),q)],p) - \mathrm{spt}(\mathbb{E}_\phi(X),p)\,\mathrm{spt}(\mathbb{E}_\phi(X),q), \quad p,q\in\mathbb{S}^*.$$

3. The deterministic equivalent problem

The aim of this section is to present a notion of deterministic equivalent problem associated with the stochastic linear optimization model

$$\max \sum_{j=1}^m \alpha_j(\omega) x_j$$

subject to:

$$\begin{cases} Ax = b \\ x_j \ge 0 \quad j = 1\ldots m, \end{cases}$$

where $\omega \in \Omega$ (Ω is a basic space of events, $\mathcal{A}$ is a σ-algebra, and ϕ is a pmm defined on the $\mathcal{A}$). For simplicity we also assume that the feasible set is compact.

In the following, let E_j be the expected value of the random variables α_j with respect to a probability multimeasure ϕ, that is

$$E_j = \mathbb{E}(\alpha_j) = \int_\Omega \alpha_j(\omega)\, \mathrm{d}\phi(\omega).$$

Since ϕ is a postive multimeasure we have that, for all j, $E_j \in \mathcal{K}$ with $0 \in E_j$ (i.e., E_j are positive). The deterministic equivalent problem can be written as

$$\max F(x) := \sum_{j=1}^m E_j x_j \qquad (DLP)$$

and subject to:

$$\begin{cases} Ax = b \\ x_j \ge 0 \quad j = 1\ldots m. \end{cases}$$

This is a set-valued optimization problem where the objective function F takes compact and convex values. The following definition introduces the notion of ordering between elements in $\mathcal{K}$ (see also Kuroiwaa, 2003).

Definition 3.1 Given two sets $A, B \in \mathcal{K}$ we say that $A \le B$ if $A \subseteq B$.

A standard separation argument gives the following lemma.

Lemma 3.2 *Suppose* $A, B \in \mathcal{K}$*. Then* $A < B$ *iff* $\mathrm{spt}(q,A) \le \mathrm{spt}(q,B)$ *for all* q *and there is a* p *with* $\mathrm{spt}(p,A) < \mathrm{spt}(p,B)$.

Definition 3.3 We say that a point $\hat{x}$ is a *solution* to (DLP) there is no feasible y for which $F(y) > F(\hat{x})$.

Proposition 3.4 *There is at least one solution to* (*DLP*).

Proof Let K be the compact and convex feasible set for (*DLP*) and let $q_n \in \mathbb{R}^d$, with $\|q_n\| = 1$ for each n, be a countable dense set in the unit sphere in $\mathbb{R}^d$. Since F is continuous, so is each $f_n(x) = \mathrm{spt}(q_n, F(x))$.

Define $A_1 := \{x \in K : f_1(x) \ge f_1(y) \text{ for all } y \in K\}$. Since f_1 is continuous and K is compact, A_1 is compact as well (in fact, A_1 is also convex). Having defined A_n, we define $A_{n+1} = \{x \in A_n : f_{n+1}(x) \ge f_{n+1}(y) \text{ for all } y \in A_n\}$. We obviously have $\emptyset \ne A_{n+1} \subseteq A_n$ and each A_n is compact and convex, and thus, $\cap_n A_n$ is nonempty. We claim that any $\hat{x} \in \cap_n A_n$ is a solution to (*DLP*).

If not, then there is some $y \in K$ with $F(y) > F(\hat{x})$ which means that $\mathrm{spt}(q, F(y)) \ge \mathrm{spt}(q, F(\hat{x}))$ for all q and there is some p with $\mathrm{spt}(p, F(y)) > \mathrm{spt}(p, F(\hat{x}))$. This implies that $f_n(y) \ge f_n(\hat{x})$ for all n and there is some m so that $f_m(y) > f_m(\hat{x})$, which is not possible by the construction of $\hat{x}$. □

Our next result relates the solutions of (*DLP*) to the solutions of the scalarizations of (*DLP*). The proof is immediate and so we do not include it.

Proposition 3.5 *Let* $\hat{x}$ *be a solution to the optimization problem*

$$\max \sum_{j=1}^m E_j x_j$$

subject to:

$$\begin{cases} Ax = b \\ x_j \ge 0 \quad j = 1\ldots m \end{cases}$$

Then there exists $p \in \mathbb{R}^d$ *so that* $\hat{x}$ *solves the following scalarized linear optimization problem:*

$$\max \sum_{j=1}^m \mathrm{spt}(p, E_j) x_j \qquad (3.1)$$

subject to:

$$\begin{cases} Ax = b \\ x_j \geq 0 \quad j = 1 \ldots m. \end{cases}$$

Theorem 2.6 gives a way of using samples to obtain a sequence of estimates for the sets E_j in (*DLP*), which in turn lead to a sequence of problems which converge (in the appropriate sense) to (*DLP*). Our next result is a stability result and shows that almost surely solutions to these problems converge to a solution to (*DLP*).

Let $a^j_i f_\phi(a^j_i)$ be an i.i.d. sample from (α_i, ϕ) for $j = 1, 2, \ldots, m$. From these data, we can construct sequences of estimates of $E_j = \mathbb{E}_\phi(\alpha_j)$, which are given by

$$E^n_j = \frac{1}{n}\sum_{j=1}^{n} a^j_i f_\phi(a^j_i). \tag{3.2}$$

Associated with each of these collections, for $j = 1, 2, \ldots, m$, there is a (DLP) given by

$$\max F^n(x) := \sum_{j=1}^{m} E^n_j x_j \qquad (n\text{−DLP})$$

and subject to:

$$\begin{cases} Ax = b \\ x_j \geq 0 \quad j = 1 \ldots m. \end{cases}$$

Theorem 3.6 *Suppose that $\hat{x}^n$ is a solution to n-DLP for each n. Then almost surely any cluster point of $\hat{x}^n$ is a solution to (DLP).*

Proof First we note that since a.s. $E^n_j \to E_j$ in the Hausdorff metric (by Theorem 2.6) and the feasible set is compact, it is straightforward to show that a.s. $F^n \to F$ uniformly, in the Hausdorff distance, on the feasible set.

Suppose that $\hat{x}^{n_k} \to \hat{x}$ and $\hat{x}$ is not a solution to (*DLP*). Then there is some feasible y with $F(y) > F(\hat{x})$. By Lemma 3.2 this means that there is a p so that $\mathrm{spt}(p, F(y)) > \mathrm{spt}(p, F(\hat{x}))$. By the uniform convergence of F^n to F, the properties of support functions, and the definition of Hausdorff distance in terms of support functions, this means that for large enough k we have

$$[\mathrm{spt}(p, F(y)) > \mathrm{spt}(p, F^{n_k}(y)) > \mathrm{spt}(p, F^{n_k}(\hat{x}_{n_k})) > \mathrm{spt}(p, F(\hat{x})),$$

and so $F^{n_k}(y) > F^{n_k}(\hat{x}_{n_k})$ which contradicts the fact that $\hat{x}_{n_k}$ is a solution to n_k-*DLP*. □

4. Numerical examples

As illustrative examples let us consider a space of events $\Omega = \{\omega_1, \omega_2\}$ composed of only two possible states of nature, let us say ω_1 and ω_2, corresponding to economic growth and recession, respectively. Suppose that three different investments are available, and let us denote by α_1, α_2, and α_3 the corresponding returns.

Example 4.1 For our first example we take ϕ to be the multimeasure defined by $\phi(\omega_1) = [-1, 0]$ and $\phi(\omega_2) = [0, 1]$ (so that $\phi(\Omega) = [-1, 1] := \mathbb{B}$). The three random variables $\alpha_1, \alpha_2, \alpha_3 : \Omega \to \mathbb{R}$ are given by

$$\begin{aligned} &\alpha_1(\omega_1) = 1/4, \quad &\alpha_2(\omega_1) = 0, \quad &\alpha_3(\omega_1) = 1/2, \\ &\alpha_1(\omega_2) = 1/4, \quad &\alpha_2(\omega_2) = 1/2, \quad &\alpha_3(\omega_2) = 0. \end{aligned}$$

Adding the constraint $x_1 + x_2 + x_3 = 1$ completes the specification of the problem. The optimal financial portfolio allocation is obtained by solving the following stochastic linear problem

$$\max \alpha_1(\omega)x_1 + \alpha_2(\omega)x_2 + \alpha_3(\omega)x_3$$

subject to:

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_j \geq 0 \quad j = 1 \ldots 3. \end{cases}$$

We can easily see that

$$E_1 := \mathbb{E}_\phi(\alpha_1) = \frac{1}{4}[-1, 0] + \frac{1}{4}[0, 1] = [-\frac{1}{4}, \frac{1}{4}].$$

In a similar way, it is easy to see that $E_2 = [0, 1/2]$ and $E_3 = [-1/2, 0]$ and so $F(x) = \left[-\frac{1}{4}x_1 - \frac{1}{2}x_3, \frac{1}{4}x_1 + \frac{1}{2}x_2\right]$. With this information, the two scalarizations are easy to compute:

$$\mathrm{spt}(1, F(x)) = \frac{1}{4}x_1 + \frac{1}{2}x_2,$$

and

$$\mathrm{spt}(-1, F(x)) = \frac{1}{4}x_1 + \frac{1}{2}x_3.$$

The first of these is maximized when $x_1 = x_3 = 0$ and $x_2 = 1$, while the second is maximized when $x_1 = x_2 = 0$ and $x_3 = 1$. Thus, it is impossible to simultaneously maximize both. Of course, this is due to the fact that the situation is completely symmetric with respect to the two risky investments α_2 and α_3 and so no preference is really possible since they are completely equivalent.

Example 4.2 In our second example we keep the same investments (random variables $\alpha_1, \alpha_2, \alpha_3$) and constraints but we change the uncertainty given by the pmm. Take $\phi(\omega_1) = [-1/2, 0]$ and $\phi(\omega_2) = [-1/2, 1]$. Since $[-1/2, 0] \subset [-1/2, 1]$, we view ω_2 as being more probable and thus associated with less uncertainty.

In this case, $E_1 = [-1/4, 1/4]$, $E_2 = [-1/4, 1/2]$, and $E_3 = [-1/4, 0]$ and so

$$F(x) = \left[-\frac{1}{4}x_1 - \frac{1}{4}x_2 - \frac{1}{4}x_3, \frac{1}{4}x_1 + \frac{1}{2}x_2\right] = \left[-\frac{1}{4}, \frac{1}{4}x_1 + \frac{1}{2}x_2\right].$$

Again the two scalarizations are easy to compute:

$$\mathrm{spt}(1, F(x)) = \frac{1}{4}x_1 + \frac{1}{2}x_2,$$

and

$$\mathrm{spt}(-1, F(x)) = \frac{1}{4}.$$

In this case clearly it is optimal to set $x_1 = x_3 = 0$ and $x_2 = 1$. The interpretation is that while the payouts of the two risky investments α_2 and α_3 are equal, their uncertainty is not and thus α_2 is the best choice.

5. Conclusions

In this paper we have analyzed how to study a stochastic linear programming problem when the underlying space is subject to partial and incomplete information of the probability distribution and this uncertainty is modeled using the notion of a probability multimeasure. Stochastic linear optimization is a model of huge interest in financial applications as it allows to determine an optimal portfolio allocation. We have showed how this problem can be transformed into a deterministic equivalent problem that takes the form of a set-valued optimization model. We have also provided some statistical properties of probability multimeasures that can be used whenever a practical real case requires the statistical estimation of the expected value of a random variable with respect to a probability multimeasure. Finally, an illustrative example has showed how the method works practically.

Acknowledgements—The second author (FM) was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant (238549-2012).

References

Artstein Z (1972). Set-valued measures. *Transactions of the AMS* **165**:103–125.

Artstein Z (1974). On the calculus of closed set-valued functions. *Indiana Univeristy Mathematical Journal* **24**:433–441.

Arstein Z and Vitale R (1975). A strong law of large numbers for random compact sets. *The Annals of Probability* **3**(5):879–882.

Aubin JP and Frankowska H (1990). *Set-Valued Analysis*. Birkhäuser: Boston.

Ben Abdelaziz F and Masri H (2005). Stochastic programming with fuzzy linear partial information on probability distribution. *European Journal of Operational Research* **162**(3):619629.

Ben Abdelaziz F and Masri H (2010) A compromise solution for the multiobjective stochastic linear programming under partial uncertainty. *European Journal of Operational Research* **202**:5559.

Beer G (1993). *Topologies on Closed and Closed Convex Sets*. Kluwer: Netherlands.

Bitran GR (1980). Linear multiobjective problems with interval coefficient. *Management Science* **26**:694706.

Cascales B, Kadets V, and Rodrìguez J (2007). The Pettis integral for multi-valued functions via single-valued ones. *Journal of Mathematical Analysis and Applications* **332**:1–10.

Cressie N (1979). A central limit theorem for random sets. *Z. Wahrsch. Verw. Gebiete* **49**(1):37–47.

Dupacova J (1987). Stochastic programming with incomplete information: a survey of results on post optimization and sensitivity analysis. *Optimization* **18**(4):507532.

Ermoliev Y and Gaivoronski A (1985). Stochastic optimization problems with incomplete information on distribution functions. *SIAM Journal on Control and Optimization* **23**(5):697716.

Hess C (2002). Set-valued integration and set-valued probability theory: an overview. In *Handbook of Measure Theory*. vols. 1, II. North-Holland: Amsterdam.

Hiai F (1978). Radon–Nikodým theorems for set-valued measures. *Journal of Multivariate Analysis* **8**(1):96–118.

Hirschberger M, Steuer RE, Utz S, Wimmer M, and Qi Y (2013). Computing the nondominated surface in tri-criterion Portfolio selection. *Operations Research* **61**(1):169–183.

Kandilakis D (1992). On the extension of multimeasures and integration with respect to a multimeasure. *Proceedings of the AMS* **116**(1):85–92.

Kunze H, La Torre D, Mendivil F, and Vrscay ER (2012).*Fractal-Based Methods in Analysis*. Springer: New York.

Kuroiwaa D (2003). Existence theorems of set optimization with set-valued maps. *Journal of Information and Optimization Sciences* **24**(1):73–84.

La Torre D, and Mendivil F (2011). Minkowski-additive multimeasures, monotonicity and self-similarity. *Image Analysis and Stereology* **30**:135–142.

La Torre D, and Mendivil F (2015). The Monge–Kantorovich metric on multimeasures and self-similar multimeasures. *Set-Valued and Variational Analysis* **23**(2):319–331.

La Torre D, and Mendivil F (2009). Union-additive multimeasures and self-similarity. *Communications in Mathematical Analysis* **7**(2):51–61.

La Torre D, and Mendivil F (2007). Iterated function systems on multifunctions and inverse problems *Journal of Mathematical Analysis and Applications* **340**(2):1469–1479.

Markowitz H (1952). Portfolio selection. *The Journal of Finance* **7**:77–91.

Molchanov I (2005). *Theory of Random Sets*. Springer: London.

Puri M, and Ralescu D (1983). Strong law of large numbers with respect to a set-valued probability measure. *The Annals of Probability* **11**(4):1051–1054.

Rockafellar RT and Wets RJ-B (1998). *Variational Analysis*. Springer: New York.

Stojaković M (2012). Set valued probability and its connection with set valued measure. *Statistics & Probability Letters* **82**:1043–1048.

Urli B and Nadeau R. Stochastic MOLP with incomplete information: an interactive approach with recourse. *Journal Operational Research Society* **41**(12):11431152.

Urli B, and Nadeau R (2004). PROMISE/scenarios: An interactive method for multiobjective stochastic linear programming under partial uncertainty. *European Journal Operational Research* **155**:361372.

Weil W (1982). An application of the Central Limit Theorem for Banach-space valued random variables to the theory of random sets. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **60**:203–208.

Received 12 October 2016;
accepted 8 May 2017