An Ergodic Theorem for Stochastic Programming Problems*

Lisa A. Korf[1] and Roger J-B Wets[2]

[1] IBM T.J. Watson Research Center, Yorktown Heights, NY
[2] Department of Mathematics, University of California, Davis

Abstract. To justify the use of sampling to solve stochastic programming problems one usually relies on a law of large numbers for random lsc (lower semicontinuous) functions when the samples come from independent, identical experiments. If the samples come from a stationary process, one can appeal to the ergodic theorem proved here. The proof relies on the 'scalarization' of random lsc functions.

1 Introduction

Stochastic programming models can be viewed as extensions of linear and nonlinear programming models to accommodate situations in which only information of a probabilistic nature is available about some of the parameters of the problem. The following formulation includes both the *stochastic programming with recourse models* and the *stochastic programming with chance constraints models* :

$$
\text{(1)} \qquad \begin{aligned} \min \quad & E\{f_0(\boldsymbol{\xi}, x)\} \\ \text{so that} \quad & E\{f_i(\boldsymbol{\xi}, x)\} \le 0, \quad i = 1, \dots, m, \\ & x \in \mathbb{R}^n \end{aligned}
$$

where

- $\boldsymbol{\xi}$ is a random vector with support $\Xi \subset \mathbb{R}^N$,
- P is a probability distribution function on $\mathbb{R}^N$,
- $f_0 : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}} = [-\infty, \infty]$,
- $f_i : \mathbb{R}^n \times \Xi \to \mathbb{R}, \quad i = 1, \dots, m$,
- for $i = 0, \dots, m$: $Ef_i(x) := E\{f_i(\boldsymbol{\xi}, x)\} = \int_\Xi f_i(\xi, x) dP(\xi)$ is assumed finite unless $\{\xi \,|\, f_0(\xi, x) = \infty\}$ has positive probability and then $Ef_0(x) = \infty$.

Let's also assume that the feasibility set

$$
S = \{x \in \mathbb{R}^n \,|\, Ef_i(x) \le 0, \ i = 1, \dots m\} \cap \{x \,|\, Ef_0(x) < \infty\}
$$

is nonempty. We are led to include the possibility that f_0 and Ef_0 take on the value ∞ to allow for the presence of *induced constraints* as will be explained shortly.

* Research supported in part by a grant of the National Science Foundation

The two-stage version of a stochastic program with recourse reads:

$$\min \quad q_1(x) + E\{Q(\boldsymbol{\xi}, x)\} \text{ so that } f_i(x) \le 0, \quad i = 1, \dots, m,$$

where

$$Q(\xi, x) = \inf_y \{q_2(\xi, y) \,|\, y \in S_2(\xi, x) \subset \mathbb{R}^{n_2}\}.$$

Immediate costs $q_1(x)$ as well as future (recourse) costs $EQ(x) = E\{Q(\boldsymbol{\xi}, x)\}$ must be taken into account in the search for an optimal decision. In terms of our canonical problem, f_0 is simply $q_1(x) + Q(\xi, x)$ and the $Ef_i \equiv f_i$ since these constraints don't depend on ξ. If $S_2(\xi, x) = \emptyset$, i.e., no feasible recourse is available in this situation, then $Q(\xi, x) = \infty$. $P\{\xi \in \Xi \,|\, Q(\xi, x) = \infty\} > 0$ means that there is a positive probability that no recourse will be available if x is chosen as the first stage decision. The 'induced' constraints restrict the choice of x to those for which, with probability 1, there will be a feasible recourse. Multistage recourse models can be 'reduced' to two-stage problems and consequently also fit our general framework, for example, cf. [1,2].

Reliability considerations lead to the inclusion of chance constraints in the formulation of the stochastic programming problem. Usually, they are expressed in the following probabilistic terms:

$$P\{\xi \in \Xi \,|\, g_k(\xi, x) \le 0,\ k = 1, \dots, q\} \ge \alpha$$

with $\alpha \in (0, 1]$ the reliability level, or they may include constraints on the moments of certain quantities such as

$$E\{g_k(\boldsymbol{\xi}, x)\} + \beta[\operatorname{var} g_k(\boldsymbol{\xi}, x)]^{1/2} \le 0$$

with β a positive constant. To bring the probabilistic constraints in concordance with the canonical form (1), define f_i as follows:

$$f_i(\xi, x) = \begin{cases} \alpha - 1 & \text{if } g_k(\xi, x) \le 0,\ k = 1, \dots, q \\ \alpha & \text{otherwise.} \end{cases}$$

Similarly, the constraint on the moments involves the sum of two functions that are both expectation functionals, since $\operatorname{var} g_k(\boldsymbol{\xi}, x)(x) = E\{g_k(\boldsymbol{\xi}, x) - Eg_k(x)\}^2$.

Let's finally observe that standard nonlinear programming problems are included as special cases of problems of type (1) since one could have $f_i(\xi, x) \equiv g_i(x)$, a function that doesn't depend on ξ, and then $Ef_i(x) = g_i(x)$. At the same time, stochastic programs of the type (1) can also be viewed as a particular class of nonlinear programming problems. Indeed, one can rewrite (1) as follows:

$$\text{(2)} \qquad \min Ef_0(x) \text{ so that } Ef_i(x) \le 0, \qquad i = 1, \dots, m,\ x \in \mathbb{R}^n.$$

The only difference is that one makes explicit the fact that that the evaluation of some, or all, of the functions Ef_i, $i = 0, \dots, m$, might require the calculation of a (multi-dimensional) integral. This is why our concerns need to go much beyond identifying the properties (linearity, convexity, differentiability) of the *expectation functionals* Ef_i and leaving the task of solving (2) to the appropriate nonlinear programming package. The major obstacle to proceeding in this manner comes precisely from the fact that evaluating Ef_i at any given x, or calculating its (sub)gradient at this x, may be a much more onerous task than solving a typical nonlinear programming problem.

Except for some special cases when the integral $\int_{\Xi} f_i(\xi, x)\, P(d\xi)$ is one-dimensional, or can be expressed as a sum of one-dimensional integrals, to evaluate this integral one must generally rely on approximation schemes with P replaced by a discrete measure P^ν obtained either from some partitioning of the sample space or as the empirical measure derived from a sample of the random quantities. In this latter instance, one needs to justify that the solution derived with the empirical measure P^ν is, at least in a probabilistic sense, an approximate solution. This paper deals with such a justification without making the usual assumption that the sample points have been obtained from independent experiments. The examples in Sect. 3 and Sect. 4 provide some of the motivation for relaxing the independence assumption, but it is also what is required to obtain the consistency of M-estimates for the parameters of regression models involving constraints coming from *a priori* information, cf. [3, Sect. 2].

2 Ergodic Theorem

A comprehensive and powerful technique to obtain the 'consistency' of the optimal solutions of the approximating problems is to actually prove that the approximating stochastic optimization problems themselves are 'consistent'. And to do this, as explained in Sect. 3 and Sect. 4, one can appeal to a general ergodic theorem for random lsc (lower semicontinuous) functions that can be formulated as follows: Let $\mathcal{B}$ be the Borel field on $\mathbb{R}^n$, $(\Xi, \mathcal{S}, P)$ a probability space with $\mathcal{S}$ P-complete; the P-completeness assumption is harmless for the application we have in mind. A *random lsc (lower semicontinuous) function* is then an extended real-valued function $f : \Xi \times \mathbb{R}^n \to \overline{\mathbb{R}}$ such that

(i) the function $(\xi, x) \mapsto f(\xi, x)$ is $\mathcal{S} \otimes \mathcal{B}$-measurable;
(ii) for every $\xi \in \Xi$, the function $x \mapsto f(\xi, x)$ is lsc.

Theorem 2.1 (Ergodic Theorem). *Let f be a random lsc function defined on $\Xi \times \mathbb{R}^n$, $\varphi : \Xi \to \Xi$ an ergodic measure preserving transformation. Then, whenever $\xi \mapsto \inf_{\mathbb{R}^n} f(\xi, \cdot)$ is summable,*

$$\frac{1}{\nu} \sum_{k=1}^{\nu} f(\varphi^k(\boldsymbol{\xi}), \cdot) \xrightarrow{e} Ef, \quad P\text{-a.s.},$$

where $\xrightarrow{e}$ stands for epi-convergence.

The immediate precursors of this theorem are the laws of large numbers for random lsc functions [4–6], that all posit iid (independent identically distributed) sampling cf. also [7,8] for further extensions. Here only stationarity is assumed; the argument relies on a 'scalarization' of random lsc functions developed in Sect. 6. The proof of the ergodic theorem can be found in Sect. 7.

Definition 2.2. *A sequence of functions* $\{g^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}, \nu \in \mathbb{N}\}$ *epi-converges to* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$*, written* $g^\nu \xrightarrow{e} g$*, if for all* $x \in \mathbb{R}^n$*,*

(i) $\liminf_\nu g^\nu(x^\nu) \geq g(x)$ *for all* $x^\nu \to x$*;*

(ii) $\limsup_\nu g^\nu(x^\nu) \leq g(x)$ *for some* $x^\nu \to x$*.*

Epi-convergence entails the convergence of the minimizers of the g^ν to those of g as is made precise below; cf. [9–11] for more about epi-convergence, theory and applications. Epi-convergence at a point x can also be characterized in terms of lower and upper epi-limits.

Definition 2.3. *For a sequence of functions* $\{g, g^\nu : \mathbb{R}^n \to \overline{\mathbb{R}},\ \nu \in \mathbb{N}\}$*, the lower and upper epi-limits are:*

$$\text{e-}\liminf g^\nu(x) := \sup_{\rho>0} \liminf_{\nu\to\infty} \inf_{y\in \mathbf{B}(x,\rho)} g^\nu(y),$$

$$\text{e-}\limsup g^\nu(x) := \sup_{\rho>0} \limsup_{\nu\to\infty} \inf_{y\in \mathbf{B}(x,\rho)} g^\nu(y).$$

If $\text{e-}\limsup g^\nu = \text{e-}\liminf g^\nu = g$*, then* $g =: \text{e-}\lim g^\nu$ *is the epi-limit of the sequence* $\{g^\nu\}_{\nu\in \mathbf{N}}$*.*

It follows immediately from Definition 2.3 that the lower and upper epi-limits are always lsc; epi-convergence of g^ν to g corresponds to the set convergence of $\operatorname{epi} g^\nu$ to $\operatorname{epi} g$. It is neither implied by, nor does it imply pointwise convergence, but instead can be viewed as a one-sided uniform convergence. But, it's exactly what is needed to ensure the convergence of minimizers of g^ν to the minimizers of g, in the following sense.

Theorem 2.4 [11, Chapter 7]**.** *Suppose that* $\{g^\nu\}_{\nu\in\mathbf{N}}$ *is a sequence of extended real-valued lsc functions such that* $g^\nu \xrightarrow{e} g$*. Then every cluster point of* $\operatorname{argmin} g^\nu$ *is an element of* $\operatorname{argmin} g$*. Moreover, if* $\operatorname{argmin} g$ *is nonempty, and there exists a compact* $K \subset \mathbb{R}^n$ *such that* $\operatorname{dom} g^\nu \subset K$*, then*

$$\operatorname{argmin} g = \bigcap_{\varepsilon>0} \operatorname{liminf}_\nu(\varepsilon\text{-}\operatorname{argmin} g^\nu),$$

where $\varepsilon\text{-}\operatorname{argmin} g := \{x \in \mathbb{R}^n \mid g(x) \leq \inf g + \varepsilon < \infty\}$*.*

Here we state only sufficient conditions for convergence of the ε-argmin. For necessary conditions, consult [11, Chapter 7], and one can refer to [6] for some extensions.

In order to obtain our almost sure epi-convergence result for random lsc functions, the following two results based on the separability of $\mathbb{R}^n$ are essential. They tell us that epi-convergence needs only to be verified at the points in a countable dense set.

Lemma 2.5 [5]. *Let $f,\ g : I\!R^n \to \overline{I\!R}$ with f lsc. Let $R \subset I\!R^n$ be the projection on $I\!R^n$ of a countable dense subset of* epi g*. If $f \le g$ on R, then $f \le g$ on all of $I\!R^n$.*

Proof. The set R above exists since $I\!R^n \times I\!R$ is separable. Suppose $f \le g$ on R. This is equivalent to $\{(x,\alpha) \,|\, \alpha \ge g(x),\ x \in R\} \subset \operatorname{epi} f$. Since f is lsc, $\operatorname{epi} f$ is closed. Taking closures on both sides yields $\operatorname{epi} g \subset \operatorname{epi} f$, which is equivalent to $f \le g$ on $I\!R^n$. □

Lemma 2.6 [5]. *Let $\{g^\nu\}_{\nu \in I\!N}$ be a sequence of extended real-valued functions defined on $I\!R^n$ and $g : I\!R^n \to \overline{I\!R}$ an lsc function. Let $R \subset I\!R^n$ be the union of R_1 and R_2, the projections onto $I\!R^n$ of a countable dense subset of* epi g *and a countable dense subset of* e-lim inf g^ν *respectively. Then $g =$* e-lim g^ν *on R implies $g =$* e-lim g^ν *on $I\!R^n$.*

Proof. In order to show that $g = \text{e-lim}\, g^\nu$, the following must hold.

$$\text{e-lim sup}\, g^\nu \le g \le \text{e-lim inf}\, g^\nu.$$

Since these are inequalities between lsc functions, we can use Lemma 2.5 to prove each inequality by having the first one satisfied on the countable set, R_1, and the second satisfied on the countable set, R_2. Then both inequalities hold on $I\!R^n$ if they are satisfied on $R := R_1 \cup R_2$. □

3 Stochastic Programs with Recourse

Consider again the two-stage stochastic program with recourse:

$$\min \quad q_1(x) + E\{Q(\boldsymbol{\xi}, x)\} \text{ so that } \ f_i(x) \le 0, \quad i = 1, \ldots, m,$$

with

$$Q(\xi, x) = \inf_y \left\{ q_2(\xi, y) \,\middle|\, y \in S_2(\xi, x) \subset I\!R^{n_2} \right\}.$$

Let's assume that this program has an optimal solution and let's denote it by x^*.

Let $\xi^1, \ldots, \xi^\nu$ be a sample of size ν of the random quantities $\boldsymbol{\xi}$ and let P^ν be the empirical measure obtained by assigning probability $1/\nu$ to each one of these sample points. Replacing P by P^ν leads us to the stochastic program:

$$\min \quad q_1(x) + \frac{1}{\nu} \sum_{k=1}^{\nu} Q(\xi^k, x) \text{ so that } \ f_i(x) \le 0, \quad i = 1, \ldots, m,$$

where for $k = 1, \dots, \nu$,

$$Q(\xi^k, x) = \inf_y \{q_2(\xi^k, y) \,|\, y \in S_2(\xi^k, x) \subset \mathbb{R}^{n_2}\}.$$

This can also be written as:

$$\begin{aligned} \min \quad & q_1(x) + \frac{1}{\nu}\sum_{k=1}^{\nu} q_2(\xi^k, y^k) \\ \text{so that} \quad & f_i(x) \le 0, \quad i = 1, \dots, m, \\ & y^k \in S_2(\xi^k, x), \quad k = 1, \dots, \nu. \end{aligned}$$

If ν is not too large, this problem can be solved by an appropriate linear or nonlinear programming package. Let x^ν be the x-component of the solution of this optimization problem. Because P^ν depends on the sample, it actually is a random measure, and consequently x^ν itself is a random variable. Proving consistency consists in showing that x^ν converges to x^* with probability 1.

The answer to this question is provided by Theorem 2.4 and the Ergodic Theorem 2.1 if the samples are iid or more generally, are generated from an ergodic process:

$$\{\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \dots, \boldsymbol{\xi}^\nu, \dots\},$$

and the following function f is a random lsc function:

$$f(\xi, x) = \begin{cases} q_1(x) + Q(\xi, x) & \text{if } f_i(x) \le 0, \quad i = 1, \dots, m, \\ \infty & \text{otherwise.} \end{cases}$$

For example, this will be the case under the following assumptions:

- q_1, f_i, $i = 1, \dots, m$ are lsc functions;
- $f_2(\xi, x, y) = \begin{cases} q_2(\xi, x, y) & \text{if } y \in S_2(\xi, x), \\ \infty & \text{otherwise,} \end{cases}$ is a random lsc function;
- for all (ξ, x): $y \mapsto f_2(\xi, x, y)$ is inf-compact (lsc with bounded level sets).

A proof could be constructed on the basis of Proposition 1.18 (about epigraphical projections) and Theorem 14.37 (about the measurability of optimal values) in [11].

The need to go beyond the iid case comes from situations when the sample is obtained from a time series. In such situations the samples aren't iid but usually the process is ergodic. This is typically the case when dealing with applications where the uncertainty comes from the environment, cf. [12] for an application dealing with lake eutrophication management and [13] for an application involving the control of water reservoirs to generate hydropower.

4 Stochastic Programs with Chance Constraints

As we have seen in Sect. 1, in a stochastic program with probabilistic constraints the expectation functional appears in the constraints in the following form: $E\{f(\boldsymbol{\xi}, x)\} \leq 0$ where

$$f(\xi, x) = \begin{cases} \alpha - 1 & \text{if } g_k(\xi, x) \leq 0, \ k = 1, \dots, q \\ \alpha & \text{otherwise.} \end{cases}$$

For an introduction to stochastic programs with chance constraints one could consult [14], a number of applications are described in [15] and a comprehensive treatment can be found in [16]. To obtain 'consistency' we follow an approach similar to that in [17].

We need the following result about the convergence of level sets of an epi-convergent sequence of functions; the *inner* and *outer limits* of a sequence $\{C^\nu\}_{\nu \in \mathbb{N}}$ of subsets of $\mathbb{R}^n$ are defined as follows:

$$\liminf_{\nu \to \infty} C^\nu = \{x = \lim_{\nu \to \infty} x^\nu \mid x^\nu \in C^\nu \text{ eventually } \}$$
$$\limsup_{\nu \to \infty} C^\nu = \{x = \lim_{k \to \infty} x^{\nu_k} \mid x^{\nu_k} \in C^{\nu_k}, \ k \in \mathbb{N}\}$$

The *limit* of the sequence exists if the outer and inner limit sets are equal:

$$\lim_{\nu \to \infty} C^\nu := \limsup_{\nu \to \infty} C^\nu = \liminf_{\nu \to \infty} C^\nu .$$

Proposition 4.1 [11, Proposition 7.7]. *For functions g^ν and g on $\mathbb{R}^n$, one has:*

(a) $g \leq \text{e-}\liminf g^\nu$ *if and only if* $\limsup_\nu (\text{lev}_{\leq \alpha^\nu} g^\nu) \subset \text{lev}_{\leq \alpha} g$ *for all sequences* $\alpha^\nu \to \alpha$;

(b) $g \geq \text{e-}\limsup g^\nu$ *if and only if* $\liminf_\nu (\text{lev}_{\leq \alpha^\nu} g^\nu) \supset \text{lev}_{\leq \alpha} g$ *for some sequence* $\alpha^\nu \to \alpha$, *in which case such a sequence can be chosen with* $\alpha^\nu \downarrow \alpha$.

(c) $g = \text{e-}\lim g^\nu$ *if and only if both conditions hold.*

Theorem 2.4, about the convergence of the minimizers of epi-convergent functions, and the Ergodic Theorem 2.1 combined with Proposition 4.1 yield the following:

Theorem 4.2. *Let's consider the following stochastic program with chance constraints:*

$$\min f_0(x) \ \text{ so that } \ P\{\xi \in \Xi \mid g_k(\xi, x) \leq 0, \ k = 1, \dots, q\} \geq \alpha \qquad (P)$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ is continuous, the functions g_k are lsc on $\Xi \times \mathbb{R}^n$ and $\alpha \in (0, 1]$. Let $\xi^1, \xi^2, \dots$ be random samples of $\boldsymbol{\xi}$, P^ν the empirical measure associated with $\xi^1, \xi^2, \dots, \xi^\nu$, and consider the following stochastic programs:

$$\min f_0(x) \ \text{ so that } \ P^\nu\{\xi \in \Xi \mid g_k(\xi, x) \leq 0, \ k = 1, \dots, q\} \geq \alpha^\nu, \qquad (P^\nu)$$

with $\alpha^\nu \to \alpha$. Suppose that for all ν,

$$x^\nu \in S^\nu = \Big\{x \,\Big|\, P^\nu\big(\{\xi \mid g_k(\xi, x) \leq 0, \ k = 1, \dots, q\}\big) \geq \alpha^\nu\Big\},$$

then almost surely every cluster point of the sequence $\{x^\nu\}_{\nu \in N}$ *is a feasible solution of the stochastic program with chance constraints* (P).

Moreover there exists a sequence $\alpha^\nu \uparrow \alpha$ *such that for* x^ν *optimal solutions of* (P^ν), *every cluster point of the sequence* $\{x^\nu\}_{\nu \in N}$ *is actually an optimal solution of* (P).

Proof. The assumptions immediately imply that

$$f(\xi, x) = \begin{cases} -1 & \text{if } g_k(\xi, x) \le 0,\ k = 1, \dots, q \\ 0 & \text{otherwise,} \end{cases}$$

is a random lsc function; one can also appeal to [11, Theorem 14.31]. Also, $Ef(x) := \int_\Xi f(\xi, x)\, P(d\xi)$ and $E^\nu f(x) := \int_\Xi f(\xi, x)\, P^\nu(d\xi)$. The Ergodic Theorem 2.1 implies that $Ef = \text{e-lim}\, E^\nu f$ almost surely. In turn, this yields via Proposition 4.1, that

$$\limsup_\nu (\text{lev}_{\le -\alpha^\nu} E^\nu f) \subset \text{lev}_{\le -\alpha} Ef$$

which means that whenever $\bar{x}$ is a cluster point of a sequence of points $\{x^\nu\}_{\nu \in N}$ with $x^\nu \in \text{lev}_{\le -\alpha^\nu} E^\nu f$, then $\bar{x} \in \text{lev}_{\le -\alpha} Ef$.

Proposition 4.1 also guarantees the existence of a sequence $\alpha^\nu \uparrow \alpha$ such that

$$\text{lev}_{\le -\alpha} Ef =: C = \lim_\nu C^\nu \ \text{ with } \ C^\nu := \text{lev}_{\le -\alpha^\nu} E^\nu f.$$

And thus $\delta_C = \text{e-lim}\, \delta_{C^\nu}$ where δ_C is the indicator function of the set C. It easy to verify that

$$f_0 + \delta_C = \text{e-lim}(f_0 + \delta_{C^\nu}).$$

Finally, Theorem 2.4 tells us: if $x^\nu \in \text{argmin}(f_0 + \delta_{C^\nu})$ and the x^ν cluster at some point $\bar{x}$, then this cluster point $\bar{x} \in \text{argmin}(f_0 + \delta_C)$, i.e., solves (P). □

In the case of a constraint involving bounds on moments or on the variance, the argument is similar.

5 Probabilistic Framework

Let lsc-fcns($\mathbb{R}^n$) denote the space of lsc (lower semicontinuous) extended real-valued functions defined on $\mathbb{R}^n$, and $(\Xi, \mathcal{S}, P)$ a probability space; we assume that $\mathcal{S}$ is P-complete. We adapt the standard probabilistic framework to lsc-fcns($\mathbb{R}^n$)-valued random variables.

Here we adopt a slightly different viewpoint of random lsc functions; we think of a *random lsc (lower semicontinuous) function* as a function $f : \Xi \to$ lsc-fcns($\mathbb{R}^n$) such that the associated bivariate function $(\xi, x) \mapsto f(\xi, x)$ is jointly measurable, i.e., $\mathcal{S} \otimes \mathcal{B}$-measurable where $\mathcal{B}$ is the Borel field on $\mathbb{R}^n$. It's convenient to identify an lsc function $f(\xi)$ with its bivariate representation so we write $f(\xi, x)$ instead of $f(\xi)(x)$ for the value of $f(\xi)$ at x. This brings us back to the framework introduced in Sect. 2. The concept of a random lsc function goes back to the work of Rockafellar in the Calculus of Variations where it comes up in the form of a 'normal integrand;' see [11, Chapter 14] for a systematic exposition.

To every random lsc function f one associates its *distribution* P_f defined by

$$P_f(\mathcal{A}) := P(\{\xi \in \Xi \mid f(\xi, \cdot) \in \mathcal{A}\}) \quad \text{for } \mathcal{A} \in \mathcal{F};$$

here $\mathcal{F}$ is the σ-field defined on lsc-fcns($\mathbb{R}^n$). Among the measurable sets in $\mathcal{F}$ are all those of the following form $\{f \in \text{lsc-fcns}(\mathbb{R}^n) \mid \inf_D f(\xi, \cdot) \leq \alpha\}$ for D any open or closed subset of $\mathbb{R}^n$; a short argument can be constructed from the joint measurability of f in (ξ, x), the lower semicontinuity in x and the separability of $\mathbb{R}^n$, or one could rely on the more comprehensive approach found in [11, Chapter 14].

Two random lsc functions, f and g, are *identically distributed* if for all $\mathcal{A} \in \mathcal{F}$, $P_f(\mathcal{A}) = P_g(\mathcal{A})$. The *joint distribution* of a finite collection $\{f^1, \ldots, f^k\}$ of random lsc functions is given, for $\mathcal{A}_1, \ldots, \mathcal{A}_k \in \mathcal{F}$, by

$$P_{\{f^1,\ldots,f^k\}}(\mathcal{A}_1, \ldots, \mathcal{A}_k) := P\big(\{\xi \in \Xi \mid f^1(\xi, \cdot) \in \mathcal{A}_1, \ldots, f^k(\xi, \cdot) \in \mathcal{A}_k\}\big).$$

For a sequence $\{f^\nu,\, \nu \in \mathbb{N}\}$ of random lsc functions, let's denote by P^∞ the probability measure on the sequence space $(\text{lsc-fcns}(\mathbb{R}^n)^\infty, \mathcal{F}^\infty)$ that is consistent with the joint distribution of the f^ν; that such a measure exists follows from Kolmogorov's Extension Theorem.

Random lsc functions are said to be *independent* if their distributions are independent. A sequence $\{f^1, f^2, \ldots\}$ is said to be *independent* if for any finite subcollection, $\{f^{\nu_1}, \ldots, f^{\nu_k},\, k \in \mathbb{N}\}$,

$$P_{\{f^{\nu_1},\ldots,f^{\nu_k}\}}(\mathcal{A}_1, \ldots, \mathcal{A}_k) = \Pi_{i=1}^k P_{f^{\nu_i}}(\mathcal{A}_i) \text{ for any sets } \mathcal{A}_1, \ldots, \mathcal{A}_k \in \mathcal{F}.$$

Definition 5.1 (iid and stationarity). *A sequence, $\{f^\nu,\, \nu \in \mathbb{N}\}$ of random lsc functions is iid (independent and identically distributed) if it is independent and for any $k, l \in \mathbb{N}$, f^k and f^l are identically distributed. The sequence is stationary if its joint distributions are invariant under shifts in the sequence, more precisely, for any finite subcollection $\{f^{\nu_1}, \ldots, f^{\nu_k}\}$, $k \in \mathbb{N}$, any $l \in \mathbb{N}$ and any $\mathcal{A}_1, \ldots, \mathcal{A}_k \in \mathcal{F}$, one has*

$$P_{\{f^{\nu_1},\ldots,f^{\nu_k}\}}(\mathcal{A}_1, \ldots, \mathcal{A}_k) = P_{\{f^{\nu_1+l},\ldots,f^{\nu_k+l}\}}(\mathcal{A}_1, \ldots, \mathcal{A}_k).$$

Stationarity can also be characterized in terms of a measure preserving transformation. Recall that a function $\varphi : \Xi \to \Xi$ is *measure preserving* if for all $A \in \mathcal{S}$, $P(\varphi^{-1}(A)) = P(A)$. If f is a random lsc function, one verifies easily that the sequence $\{f, f\circ\varphi, f\circ\varphi^2, \ldots\}$ is stationary. In fact, every stationary

sequence of random lsc functions can be redefined in terms of a (single) random lsc function and a measure preserving transformation:
Say $\{f^\nu,\, \nu \in \mathbb{N}\}$ is a stationary sequence of random lsc functions and P^∞ the measure induced on $(\text{lsc-fcns}(\mathbb{R}^n)^\infty, \mathcal{F}^\infty)$. Redefine the f^ν as follows:

$$f^\nu : \text{lsc-fcns}(\mathbb{R}^n)^\infty \to \text{lsc-fcns}(\mathbb{R}^n) \text{ with } f^\nu(\zeta) := \zeta^\nu,$$

i.e., the ν-th element of the sequence $\zeta \in \text{lsc-fcns}(\mathbb{R}^n)^\infty$. The new sequence $\{f^\nu\}_{\nu\in\mathbb{N}}$ is stationary and has the same joint distributions as the original one, but now with respect to the new probability space. Letting φ : $\text{lsc-fcns}(\mathbb{R}^n)^\infty \to \text{lsc-fcns}(\mathbb{R}^n)^\infty$ be the shift operator,

$$\varphi(\zeta^1, \zeta^2, \ldots) := (\zeta^2, \zeta^3, \ldots),$$

and defining $f : \text{lsc-fcns}(\mathbb{R}^n)^\infty \to \text{lsc-fcns}(\mathbb{R}^n)$ as $f(\zeta) = \zeta^1$, one has that $f\big(\varphi^\nu(\zeta)\big) = \zeta^{\nu+1}$, so that $f, f\circ\varphi, f\circ\varphi^2, \ldots$, defines the same stationary sequence on $\text{lsc-fcns}(\mathbb{R}^n)^\infty$ with respect to the measure preserving shift transformation φ; it is easy to check that φ is measure preserving.

If $\varphi : \Xi \to \Xi$ is measure preserving, then $A \in \mathcal{S}$ is an *invariant event* if $\varphi^{-1}(A) = A$ almost surely, i.e., in terms of the symmetric difference, $P(\varphi^{-1}(A) \triangle A) = 0$.

Definition 5.2 (ergodicity). *Let $\mathcal{I}$ denote the σ-field of invariant events and call it the invariant σ-field. A measure preserving map $\varphi : \Xi \to \Xi$ is ergodic if $\mathcal{I}$ is trivial, i.e., for all $A \in \mathcal{I}$, $P(A) \in \{0, 1\}$. A sequence $\{f^\nu,\, \nu \in \mathbb{N}\}$ of random lsc functions is ergodic if the associated (measure preserving) shift operator φ on the sequence space* $(\text{lsc-fcns}(\mathbb{R}^n)^\infty, \mathcal{F}^\infty, P^\infty)$ *is ergodic.*

6 Scalarization of Random Lsc Functions

The framework of reference is still that of Sect. 5. In this section, it will be shown that a random lsc function f is completely identified by a countable collection of extended real-valued random variables

$$f \longleftrightarrow \big\{\pi_{x,\rho} \,\big|\, x \in R, \rho \in \mathbb{Q}_+\big\} \text{ where } R \text{ is a countable dense subset of } \mathbb{R}^n.$$

We refer to such an identification as a *scalarization* of the random lsc function f; results about the scalarizations of random lsc functions with values in $\text{lsc-fcns}(X)$, for X a Polish space, appear in [18]. When $X = \mathbb{R}^n$, it's not necessary to assume $\mathcal{S}$ is P-complete, and this version can also be found in [11, Theorem 14.40]:

Theorem 6.1 (scalarization). *Let $f : \Xi \to \text{lsc-fcns}(\mathbb{R}^n)$,*

$$\textit{and for } D \subset \mathbb{R}^n : \qquad \textit{let } \pi_D(\xi) := \inf\nolimits_{x\in D} f(\xi, x).$$

Then f is a random lsc function if and only if for all $D \in \mathcal{D}$, π_D is measurable where $\mathcal{D}$ is any one of the following collection of sets:

(a) $\mathcal{D}$ = *the open sets* $O \subset \mathbb{R}^n$;

(b) $\mathcal{D}$ = *the closed sets* $C \subset \mathbb{R}^n$;

(c) $\mathcal{D}$ = *the closed balls* $\mathbb{B}(x, \rho) \subset \mathbb{R}^n$;

(d) $\mathcal{D}$ = *the closed rational balls* $\mathbb{B}(x, \rho) \subset \mathbb{R}^n$ *with* $x \in R$, R *a countable dense subset of* $\mathbb{R}^n$ *and* $\rho \in \mathbb{Q}_+$;

Corollary 6.2 (countable scalarization). *Let* $f : \Xi \to \text{lsc-fcns}(\mathbb{R}^n)$. *For* $x \in R$, *a countable dense subset of* $\mathbb{R}^n$, *and* $\rho \in \mathbb{Q}_+$, *define*

$$\pi_{x,\rho}(\xi) := \pi_{\boldsymbol{B}(x,\rho)}(\xi) = \inf\nolimits_{y \in \boldsymbol{B}(x,\rho)} f(\xi, y).$$

Then f is a random lsc function if and only if the random variables in the countable collection

$$\left\{\pi_{x,\rho} : \Xi \to \overline{\mathbb{R}} \,\middle|\, x \in R, \rho \in \mathbb{Q}_+\right\}$$

are measurable.

Proof. This is just a reformulation of part (d) of the theorem. □

To each sequence of random lsc functions $\left\{f^\nu : \Xi \to \text{lsc-fcns}(X), \nu \in \mathbb{N}\right\}$ we can associate, by scalarization, a sequence of vector-valued random variables

$$\left\{\pi^\nu_{x,\rho}, \nu \in \mathbb{N} \,\middle|\, x \in R, \rho \in \mathbb{Q}_+\right\}.$$

Independence, stationarity and ergodicity properties of the sequence of the random lsc functions are inherited by these vectors generated through scalarization. Here we are only interested in the ergodicity properties of these scalarizations.

Proposition 6.3. *If $\{f \circ \varphi^\nu\}$ is an ergodic sequence of random lsc functions, then $\{\pi_{x,\rho} \circ \varphi^\nu\}$ is an ergodic sequence of random variables for all $x \in R$, $\rho \in \mathbb{Q}_+$.*

Proof. The shift operator, $\varphi : \text{lsc-fcns}(X)^\infty \to \text{lsc-fcns}(X)^\infty$ is ergodic, and $\pi_{x,\rho}$ defined on $\text{lsc-fcns}(X)^\infty$ by $\pi_{x,\rho}(\zeta) := \inf_{\boldsymbol{B}(x,\rho)} \zeta_1$ is measurable. Therefore the sequence, $\{\pi_{x,\rho} \circ \varphi^\nu\}$ is ergodic, and equivalent to the original sequence. □

7 Proof of the Ergodic Theorem

The proof of the Ergodic Theorem relies on the following theorem of independent interest. It says that to verify the almost sure epi-convergence of the empirical means of a sequence of random lsc functions, it suffices to check the almost sure convergence of the empirical means of the corresponding scalarizations.

Theorem 7.1. *Let $\{f, f^\nu, \nu \in I\!N\}$ be a sequence of random lsc functions on $I\!R^n$, and let R be a countable dense subset of $I\!R^n$ that contains the projection onto $I\!R^n$ of a countable dense subset of* $\operatorname{epi} Ef$*. For $x \in R$, $\rho \in \mathbb{Q}_+$, let $\pi^\nu_{x,\rho} := \inf_{y\in \boldsymbol{B}(x,\rho)} f^\nu(\cdot, y)$ and $\pi_{x,\rho} := \inf_{y\in \boldsymbol{B}(x,\rho)} f(\cdot, y)$. Suppose that for all $x \in R$, $\rho \in \mathbb{Q}_+$,*

$$\frac{1}{\nu}\sum_{k=1}^{\nu} \pi^k_{x,\rho}(\xi) \to E\pi_{x,\rho} \quad P\text{-a.s.}$$

with $E\pi_{x,\rho} := E\{\pi_{x,\rho}(\boldsymbol{\xi})\}$. Then, whenever $\xi \mapsto \inf_{I\!R^n} f(\xi, \cdot)$ is summable,

$$\frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, \cdot) \xrightarrow{e} Ef \quad P\text{-a.s.}$$

Proof. Fix $x \in R$, $\rho \in \mathbb{Q}_+$. Let $\Xi_{x,\rho} \subset \Xi$ be such that $P(\Xi_{x,\rho}) = 1$ and

$$\frac{1}{\nu}\sum_{k=1}^{\nu} \pi^k_{x,\rho}(\xi) \to E\pi_{x,\rho}$$

for all $\xi \in \Xi_{x,\rho}$. Let $\Xi_R := \bigcap_{x\in R}\bigcap_{\rho\in \mathbb{Q}_+} \Xi_{x,\rho}$. Then $P(\Xi_R) = 1$, since Ξ_R is obtained from a countable intersection of sets of measure one.

To show that for all $\xi \in \Xi_R$, $\operatorname{e-lim\,inf} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, \cdot) \geq Ef$ on $I\!R^n$, let $\xi \in \Xi_R$, $x \in I\!R^n$. Then

$$\begin{aligned}
\operatorname{e-lim\,inf} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, x) &= \sup_{\rho>0} \liminf_{\nu} \inf_{y\in \boldsymbol{B}(x,\rho)} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, y) \\
&\geq \sup_{\rho>0} \liminf_{\nu} \frac{1}{\nu}\sum_{k=1}^{\nu} \inf_{y\in \boldsymbol{B}(x,\rho)} f^k(\xi, y) \\
&\geq \sup_{\ell} \liminf_{\nu} \frac{1}{\nu}\sum_{k=1}^{\nu} \pi^k_{x^\ell,\rho^\ell}(\xi),
\end{aligned}$$

where for all $\ell \in I\!N$, $x^\ell \in R$, $\rho^\ell \in \mathbb{Q}_+$, $x^\ell \to x$, $\rho^\ell \downarrow 0$, $x \in \operatorname{int} I\!B(x^\ell, \rho^\ell)$, and $I\!B(x^{\ell+1}, \rho^{\ell+1}) \subset I\!B(x^\ell, \rho^\ell)$. For each ℓ, $\xi \in \Xi_R$, from the assumptions, one has

$$\liminf{}_\nu \frac{1}{\nu}\sum_{k=1}^{\nu} \pi^k_{x^\ell,\rho^\ell}(\xi) = E\pi_{x^\ell,\rho^\ell}.$$

Continuing, we obtain

$$\operatorname{e-lim\,inf} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, x) \geq \sup_{\ell} E\pi_{x^\ell,\rho^\ell} = Ef(x)$$

by the Monotone Convergence Theorem and the lower semicontinuity of $x \mapsto f(\xi, x)$ for all ξ. Hence for all $\xi \in \Xi_R$, e-lim inf $\frac{1}{\nu}\sum_{k=1}^{\nu} f^{\nu}(\xi,\cdot) \geq Ef$ on $\mathbb{R}^n$.

For the lim sup inequality, observe that for $\xi \in \Xi_R$, $x \in R$, if $x^{\nu} \equiv x$, then by assumption

$$\limsup_{\nu} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi, x^{\nu}) = \limsup_{\nu} \frac{1}{\nu}\sum_{k=1}^{\nu} \pi_{x,0}^k(\xi) = E\pi_{x,0} = Ef.$$

Using the facts that Ef is lsc (by Fatou's Lemma) and R contains the projection on $\mathbb{R}^n$ of a countable dense subset of epi Ef, along with Lemma 2.5 and the fact that

$$\text{e-}\limsup \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi,\cdot) \leq \text{limsup}_{\nu} \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi,\cdot) \leq Ef \quad \text{on } R,$$

it follows that

$$\text{e-}\limsup \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi,\cdot) \leq Ef \quad \text{on } \mathbb{R}^n.$$

In summary, it has been shown that $P(\Xi_R) = 1$, and for all $\xi \in \Xi_R$,

$$\text{e-}\limsup \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi,\cdot) \leq Ef \leq \text{e-}\liminf \frac{1}{\nu}\sum_{k=1}^{\nu} f^k(\xi,\cdot)$$

on $\mathbb{R}^n$ as claimed. □

We are now ready to prove the Ergodic Theorem.

Theorem 7.2. *Let $(\Xi, \mathcal{S}, P)$ be a probability space, $\varphi : \Xi \to \Xi$ an ergodic measure preserving transformation, and f a random lsc function on $\mathbb{R}^n$. Then, whenever $\xi \mapsto \inf_{\mathbb{R}^n} f(\xi,\cdot)$ is summable,*

$$\frac{1}{\nu}\sum_{k=1}^{\nu} f\big(\varphi^k(\xi),\cdot\big) \xrightarrow{e} Ef \quad P\text{-a.s.}$$

Proof. Let $\{\pi_{x,\rho} \,|\, x \in R, \rho \in \mathbb{Q}_+\}$ denote the scalarization of f with R the projection of a countable dense subset of epi Ef on $\mathbb{R}^n$. Since φ is measure preserving and ergodic, we obtain that for all $x \in R$, $\rho \in \mathbb{Q}_+$, the sequence, $\{\pi_{x,\rho} \circ \varphi^{\nu}\}_{\nu \in \mathbb{N}}$ is also ergodic by Proposition 6.3. Hence by the classical Birkhoff-Khintchine Ergodic Theorem [19] with a straightforward extension to include functions which take on the value $+\infty$, for all $x \in R$, $\rho \in \mathbb{Q}_+$, we obtain,

$$\frac{1}{\nu}\sum_{k=1}^{\nu} \pi_{x,\rho}\big(\varphi^k(\xi)\big) \to E\pi_{x,\rho} \quad P\text{-a.s.}$$

Appealing to Theorem 7.1 it follows immediately that

$$\frac{1}{\nu}\sum_{k=1}^{\nu} f(\varphi^k(\xi),\cdot) \xrightarrow{e} Ef \quad P\text{-a.s.},$$

as claimed. □

References

1. Wets R. J-B (1989) *Stochastic Programming*, Handbook for Operations Research and Management sciences # 1, Elsevier Science (North Holland)
2. Birge J.R. & Louveaux F. (1997) *Introduction to Stochastic Programming*, Springer
3. Korf L.A. & Wets R. J-B (1999) "Ergodic limit laws for stochastic optimization problems," Manuscript, University of California, Davis
4. King A.J. & Wets R. J-B (1990) "Epi-consistency of convex stochastic programs," *Stochastics and Stochastics Reports* 34:83–92
5. Attouch H. & Wets R. J-B (1990) "Epigraphical processes: laws of large numbers for random lsc functions," *Séminaire d'Analyse Convexe, Montpellier*
6. Artstein Z. & Wets R. J-B (1995) "Consistency of minimizers and the SLLN for stochastic programs," *J. of Convex Analysis* 2:1–17
7. Hess C. (1991) "Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator," Cahiers no. 9121, Cérémade, Paris
8. Castaing C. & Ezzaki F. (1995) "SLLN for convex random sets and random lower semicontinuous integrands," Manuscript, Université de Montpellier
9. Attouch H. (1984) *Variational Convergence for Functions and Operators*, Applicable Mathematics Series, Pitman
10. Aubin J.-P. & Frankowska H. (1990) *Set-Valued Analysis*, Birkhäuser
11. Rockafellar R.T. & Wets R. J-B (1997) *Variational Analysis*, Grundlehren der mathematischen Wissenschaften 317, Springer-Verlag
12. King A.J., Rockafellar R.T., Somlyódy L. & Wets R. J-B (1988) "Lake eutrophication management: the Lake Balaton project," In: Ermoliev Y., Wets R. (Eds.) *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, 435–444
13. Salinger D. (1997) "A Splitting Algorithm for Multistage Stochastic Programming with Application to Hydropower Scheduling," Ph. D. Thesis, University of Washington, Seattle
14. Kall P. & Wallace S. (1994) *Stochastic Programming*, Wiley
15. Kibzun A.I. & Kan Y.S. (1996) *Stochastic Programming Problems with Probability and Quantile Functions*, J. Wiley & Sons
16. Prékopa A. (1995) *Stochastic Programming*, Kluwer
17. Wets R. J-B (1998) "Stochastic programs with chance constraints: generalized convexity and approximation issues," In: Crouzeix J.-P., Martinez-Legaz J.E., Volle M. (Eds.) *Generalized Convexity, Generalized Monotonicity: Recent Results*, Kluwer, 61–74
18. Korf L.A. & Wets R. J-B (1999) "Random lsc functions: scalarization," Manuscript, University of California, Davis
19. Durrett R. (1991) *Probability: Theory and Examples*, Wadsworth and Brooks

Index

Index