MATHEMATICS OF OPERATIONS RESEARCH
Vol. 26, No. 2, May 2001, pp. 421–445
Printed in U.S.A.

RANDOM LSC FUNCTIONS: AN ERGODIC THEOREM

LISA A. KORF AND ROGER J.-B. WETS

An ergodic theorem for random lsc (lower semicontinuous) functions is obtained by relying on a "scalarization" of such functions. In the process, Kolmogorov's extension theorem for random lsc functions is established. Applications to statistical estimation problems, composite materials, and stochastic optimization problems are briefly noted.

1. Introduction. Solution procedures for stochastic programming problems, statistical estimation problems (constrained or not), stochastic optimal control problems, and other stochastic optimization problems often rely on sampling. The justification for such an approach passes through "consistency." A comprehensive, satisfying, and powerful technique is to obtain the consistency of the optimal solutions, statistical estimators, controls, etc., as a consequence of the consistency of the stochastic optimization problems themselves. To do this, as explained in §2, one can appeal to the ergodic properties of random lsc (lower semicontinuous) functions set forth in this paper.

A streamlined version of this basic ergodic theorem, see §8, can be formulated as follows: Let (X, d) be a Polish space, i.e., a complete, separable, metric space, with $\mathcal{B}$ the Borel field on X, $(\Xi, \mathcal{S}, P)$ a probability space and, for now, let us assume that $\mathcal{S}$ is P-complete. A *random lsc (lower semicontinuous) function* is then an extended real-valued function $f : \Xi \times X \to \overline{\mathbb{R}}$ such that

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

THEOREM 1.1. *Let f be a random lsc function defined on $\Xi \times X$, $\varphi : \Xi \to \Xi$ an ergodic measure preserving transformation. Then, whenever $\xi \mapsto \inf_X f(\xi, \cdot)$ is summable,*

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

DEFINITION 1.2 A sequence of functions $\{g^\nu : X \to \overline{\mathbb{R}}, \nu \in \mathbb{N}\}$ epi-converges to $g : X \to \overline{\mathbb{R}}$, written $g^\nu \to^{e} g$, if for all $x \in X$,

(i) $\liminf_\nu g^\nu(x^\nu) \geq g(x)$ for all $x^\nu \to x$; and
(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 made precise in §7. It is so named because it agrees with the set convergence of the epigraphs, cf. Attouch (1984), Aubin and Frankowska (1990), Rockafellar and Wets (1984, 1998); the *epigraph* of a function $g : X \to \overline{\mathbb{R}}$ consists of those points that lie on or above the graph of g. The assumption "$\inf_X f(\cdot, x)$ majorizes a summable function" will be considerably relaxed and "$\mathcal{S}P$-complete" will be dropped in the statement of the Ergodic Theorem 8.2.

Received February 3, 2000; revised October 19, 2000, January 22, 2001.
MSC 2000 subject classification. Primary: 60H07, 90C15, 60B11, 49J99.
OR/MS subject classification. Primary: Programming/stochastic.
Key words. Stationary processes, ergodic theorem, epi-convergence, Kolmogorov's extension theorem, stochastic optimization, Bayesian decision theory, composite materials, stochastic programming, random lower semicontinuous functions, random samples.

0364-765X/01/2602/0421/$05.00
1526-5471 electronic ISSN, © 2001, INFORMS

A first attempt at stating an ergodic theorem for stochastic optimization problems goes back to Kankova (1978). She relied on classical analytic tools and consequently has to make assumptions (strict convexity and deterministic constraints) that seriously limit the range of potential applications. A more immediate precursor to our Ergodic Theorem is the Law of Large Numbers for random lsc functions (Attouch and Wets 1990, Artstein and Wets 1995), cf. also Hess (1996), and Castaing and Ezzaki (1995) for extensions, that posits iid (*independent identically distributed*) sampling whereas here only stationarity needs to be assumed.

The argument is built on a "scalarization" of random lsc functions developed in §§3–5; a more narrowly focused presentation can be found in Korf and Wets (2000). One also needs a generalization of Kolmogorov's Extension Theorem for measures defined on a space that is not Polish, viz. the space of lsc functions endowed with epi-convergence. This is covered in §6.

2. Examples. This section is devoted to three basic examples. We limit ourselves to a brief description of the problems and indicate how the ergodicity theorem can be used to justify the concerns we might have about "consistency."

EXAMPLE 1: TIME SERIES. Classical consistency results for certain estimation procedures of the parameters of a time series can be guaranteed by means of our Ergodic Theorem for random lsc functions. But one can deal equally well with "nonclassical" situations as would be the case if we wanted to include some information one might have about certain relationships between the parameters in the estimation process. We are going to illustrate this by considering the following estimation problem. Suppose that the coefficients $a_0, a_1, \dots, a_p$ of the (linear) transfer function of the *autoregressive* (AR) model,

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + \varepsilon_t, \quad t = \dots, 0, 1, \dots,$$

are to be estimated given a number of observations, say $\eta_{1-p}, \dots, \eta_\nu$ of the process Y_t, and given that

$$a_1 \geq a_2 \geq \cdots \geq a_p,$$

i.e., the more distant the past, the smaller the contribution to Y_t. This might follow from certain physical laws or economic considerations.

Let us assume, as usual, that the ε_t are independent normally distributed random variables with mean 0 and variance σ^2. They account for the disturbances in the dynamics not captured by the transfer function. Assuming that the coefficients of the *autoregressive polynomial* $1 - a_1 r - \cdots - a_p r^p$ are such that its roots lie outside the unit circle, this AR-model has a solution $\{Y_t, t = \dots, -1, 0, 1, \dots\}$ that is *stationary*, cf. for example, Tiao (1985).

If the monotonicity condition is ignored, one could choose as estimates the solution of the following optimization problem:

$$\min_{x_0, \dots, x_p} \frac{1}{\nu} \sum_{t=1}^{\nu} \left| \eta_t - x_0 - \eta_{t-1} x_1 - \cdots - \eta_{t-p} x_p \right|^2;$$

the normalizing factor $1/\nu$ does not affect the solution, it just scales the optimal value. One would find estimates that minimize the role played by the "disturbances" (innovations), i.e., those factors not included in the (linear) transfer function; note that $\nu^{-1} \sum_{t=1}^{\nu} (Y_t - a_0 - Y_{t-1} a_1 - \cdots - Y_{t-p} a_p)^2 = \nu^{-1} \sum_{t=1}^{\nu} (\varepsilon_t)^2$.

To embed this in the general framework of §1, one defines the vector process,

$$\{X_t = (Y_{t-p}, Y_{t-p+1}, \dots, Y_t), t = \dots, -1, 0, 1, \dots\}.$$

The stationarity of this process is inherited from that of the Y_t. An observation of X_t, say ξ_t, is then a vector of the type $\xi_t = (\eta_{t-p}, \dots, \eta_{t-1}, \eta_t)$. With

$$f(\xi_t, x) = \left|\eta_t - x_0 - \eta_{t-1}x_1 - \cdots - \eta_{t-p}x_p\right|^2,$$

given the observations $\xi_1, \dots, \xi_\nu$, the optimization problem yielding the estimates takes the form

$$\min \frac{1}{\nu}\sum_{t=1}^{\nu} f(\xi_t, x).$$

With $(\Xi, \mathcal{S}, P)$, the (common) probability space on which the random variables X_t are defined, it is easy to verify that $f : \Xi \times \mathbb{R}^{p+1} \to \mathbb{R}$ is a random lsc function. By appealing to the Ergodic Theorem 8.2, the consistency of the proposed estimates $(x_0^\nu, x_1^\nu, \dots, x_p^\nu)$ can be settled by showing that almost surely,

$$(a_0, a_1, \dots, a_p) \in \operatorname{argmin} E^{\mathcal{J}} f(x) := E\{f(\boldsymbol{\xi}, x)|\mathcal{J}\},$$

where $\boldsymbol{\xi}$ is a random vector with the same distribution as $X_t = (Y_{t-p}, Y_{t-p+1}, \dots, Y_t)$, and $\mathcal{J}$ is the invariant σ-field of the stationary process $\{X_t | t = \dots, 0, 1, \dots\}$, cf. §4. Uniqueness of the solution and almost sure epi-convergence, which implies the almost sure convergence of the optimal solutions, yield

$$(x_0^\nu, x_1^\nu, \dots, x_p^\nu) \to (a_0, a_1, \dots a_p) \text{ a.s.};$$

for a detailed derivation, see §2.1 of Korf and Wets (2000).

The standard approach to obtaining the consistency of these or related estimates is usually tied to the *specific* form of the selected criterion. For example, Box and Jenkins (1970) obtain consistency for the minimizers of the *conditional likelihood function.* In contrast, the approach suggested by our Ergodic Theorem for random lsc functions is not restricted to a specific criterion function. The same arguments, up to certain calculations, can be used if rather than the ℓ^2-criterion above, one chooses, as in Bloomfield and Steiger (1983, Chapter 3), the ℓ^1-criterion,

$$\frac{1}{\nu}\sum_{t=1}^{\nu}\left|\eta_t - x_0 - \eta_{t-1}x_1 - \cdots - \eta_{t-p}x_p\right|,$$

which, by the way, is not differentiable. More significantly, one can include restrictions in the formulation of the optimization problem used to obtain the estimates. This is particularly useful, and potentially important, if the number of samples available is small. In particular, one can include monotonicity restrictions on the choice of the coefficients, similar to those analyzed in Barlow et al. (1972) and Robertson et al. (1988), by modifying the function f as follows:

$$\tilde{f}(\xi_t, x) = \begin{cases} \left|\eta_t - x_0 - \eta_{t-1}x_1 - \cdots - \eta_{t-p}x_p\right|^2 & \text{if } x_1 \geq \cdots \geq x_p, \\ \infty & \text{otherwise.} \end{cases}$$

Also for such a criterion function, the same arguments yield the consistency of the estimates

$$(\tilde{x}_0^\nu, \tilde{x}_1^\nu, \dots, \tilde{x}_p^\nu) \in \operatorname{argmin} \frac{1}{\nu}\sum_{t=1}^{\nu}\tilde{f}(\xi_t, x).$$

From an operational viewpoint, the advantage lies in having estimates that satisfy a relationship that is known to hold for the coefficients $(a_1, \dots, a_p)$, even when only a small number of samples can be collected.

The Law of Large Numbers for random lsc functions, mentioned in §1, allows us to work with an equally rich collection of criteria but because of the lack of independence between the observations, it doesn't apply in this context.

EXAMPLE 2: NONHOMOGENEOUS MATERIALS—POROUS MEDIA. Modern technology relies extensively on composite materials. In particular, this has led to the study of the properties and the behavior of random media. Given a composite material, its complex structure, whether fully known or not, typically renders the study of its behavior computationally intractable. Methods of *homogenization* are useful to deal with the complexity of the microscopic make-up of the material. These methods work by essentially replacing the material with an *averaged* one, whose properties are "close" in a certain sense to those of the original model. In particular, stochastic problems study the behavior of a material whose structure is only partially known (e.g., composed of two or more materials in a fixed proportion). *Stochastic homogenization* approximates such a problem by replacing it with a deterministic, *homogenized*, one, which basically preserves the behavior of the original material.

To illustrate, let us consider the example of a conductor occupying a region Ω in $\mathbb{R}^3$. Suppose that the conducting material is an inhomogeneous composite of two or more components, each with a different conductivity. One may model the conductivity as a random function of position, of the form $a(\boldsymbol{\xi}, x)$, where $a(\boldsymbol{\xi}, x)$ is stationary with respect to spatial location, positive and bounded. Associated with this function is the stochastic partial differential equation, describing the temperature $u(\xi, x)$ by

$$-\nabla \cdot \big(a(\boldsymbol{\xi}, x)\nabla u(\boldsymbol{\xi}, x)\big) = h(x) \quad \text{for } x \in \Omega,$$

$$u(\boldsymbol{\xi}, x) = 0 \quad \text{for } x \in \operatorname{bdry} \Omega.$$

The goal is to obtain the *homogenized equation*, which can be accomplished by computing the appropriate (deterministic) *effective coefficient*, $a(x)$, of conductivity. The resulting homogenized equation would then be given by

$$-\nabla \cdot \big(a(x)\nabla u(x)\big) = h(x) \quad \text{for } x \in \Omega,$$

$$u(x) = 0 \quad \text{for } x \in \operatorname{bdry} \Omega.$$

Taking into account that the inconsistencies of the material occur at a microscopic level, it is accepted that the behavior of the solution, u, of the homogenized problem will approximate that of the original problem if $u(x) = E\{u(\boldsymbol{\xi}, x)\}$ for $x \in \Omega$. Note that contrary to first intuition, setting $a(x) = E\{a(\boldsymbol{\xi}, x)\}$ does *not* generate a homogenized equation with the desired properties.

Procedures that have been suggested for solving such a stochastic homogenization problem rely on scaling the material by a parameter, ε, and employing methods such as asymptotic analysis to find the limiting problem, see e.g., Papanicolaou and Varadhan (1981) and Kozlov et al. (1994). However, there are many problems to which these methods cannot be applied due to the complexities and randomness of the materials being studied. The following numerical methodology based, instead, on sampling, attempts to provide a means to approach a much broader class of problems. For a full discourse, consult Attouch et al. (1999).

To cast this question in the framework of §1, let us reformulate the inhomogeneous problem in its variational form. The function $g(\xi, \cdot) : L^2 \to (-\infty, \infty]$ is to be minimized for each ξ:

$$\min_{u \in H_0^1(\Omega)} g(\xi, u) := \frac{1}{2} \int_\Omega a(\xi, x) |\nabla u|^2 dx - \langle h, u \rangle.$$

The goal is to find the homogenized functional, g^{hom} such that

$$E\{u(\boldsymbol{\xi}, \cdot)\} = \bar{u}(\cdot) \in \operatorname{argmin}\{g^{\text{hom}}(u) \,|\, u \in H_0^1(\Omega)\}.$$

Given the samples $a(\xi^1,\cdot),\dots,a(\xi^\nu,\cdot)$, it is easy to check that only exceptionally can $\bar{u}$ be an approximate solution of the optimization problem whose criterion function is the "mean" criterion function,

$$\min\left\{\frac{1}{\nu}\sum_{k=1}^{\nu} g(\xi^k,u)\,\middle|\, u\in H_0^1(\Omega)\right\}.$$

It cannot serve as an approximation for the homogenized problem; it would correspond to averaging $a(\boldsymbol{\xi},x)$, and thus there is no way in which we can assert that the solution of this problem could provide a (consistent) approximation of $\bar{u}$.

To derive g^{hom}, we rely on some facts from conjugate duality. To approximate it, we appeal to our ergodic theorem for random lsc functions.

Let (X,τ) be a Banach space and X^* its topological dual; here $X=H_0^1$ with τ the norm topology. The *conjugate* function $q^*: X^*\to\overline{\mathbb{R}}$ of a function $q: X\to\overline{\mathbb{R}}$ is

$$q^*(v):=\sup_x\left\{\langle v,x\rangle-q(x)\right\},$$

and $q^{**}=(q^*)^*$ is the *biconjugate* to q. The mapping $q\mapsto q^*$ is called the *Legendre-Fenchel transform.* The *epi-multiple* of q by $\lambda\geq 0$ is the function $\lambda * q: X\to\overline{\mathbb{R}}$ defined by

$$(\lambda * q)(x):=\lambda q(\lambda^{-1}x)\quad\text{for }\lambda>0$$

$$(0*q)(x):=\begin{cases}0 & \text{if } x=0,\, q\not\equiv\infty\\ \infty & \text{otherwise.}\end{cases}$$

For functions $p,q: X\to\overline{\mathbb{R}}$, the *epi-sum* is the function $p\#q: X\to\overline{\mathbb{R}}$, defined by

$$(p\#q)(x):=\inf_z\left\{p(z)+q(x-z)\right\}.$$

The *epi-integral* of a random lsc function $q:\Xi\times X\to\overline{\mathbb{R}}$ with respect to a probability measure P is the function $e\text{-}\int q(\xi,\cdot)*P(d\xi)$, defined by

$$\left(e\text{-}\int q(\xi,\cdot)*P(d\xi)\right)(x):=\inf_{z(\cdot)}\left\{\int_\Xi q(\xi,z(\xi))\,P(d\xi)\,\middle|\,\int_\Xi z(\xi)\,P(d\xi)=x\right\},$$

where it is understood that z is an integrable function from Ξ to X.

It is straightforward to verify that

$$\bar{u}=E\{u(\boldsymbol{\xi},\cdot)\}\in\operatorname{argmin}\left\{g^{\text{hom}}(u):=\left(e-\int g(\xi,\cdot)*P(d\xi)\right)(u)\middle| u\in H_0^1\right\},$$

and one might reasonably expect that for ν sufficiently large,

$$\bar{u}^\nu\in\operatorname{argmin}\{(\nu^{-1}*[g(\xi^1,\cdot)\#\cdots\#g(\xi^\nu,\cdot)])(u)|u\in H_0^1\}$$

would approximate $\bar{u}$.

Let us now proceed and sketch out a justification that will also suggest a way to actually carry out these calculations. The following two theorems focus on some relevant properties of the Legendre-Fenchel transform. Theorem 2.1 emphasizes, in (ii)–(v), the duality between the epi-operations above and the standard operations of addition and multiplication; for the proofs one could consult Castaing and Valadier (1977), Rockafellar and Wets (1998), and Attouch (1984). Recall that an extended real-valued function q is called *proper* if $-\infty<q\not\equiv\infty$.

THEOREM 2.1. *Let $p, q : X \to \overline{\mathbb{R}}$ be lsc, proper, and convex, where X is a reflexive Banach space. Then*

(i) $p^{**} = p$;
(ii) $(p\#q)^* = p^* + q^*$;
(iii) $(\lambda * p)^* = \lambda p^*$;
(iv) *If* $\operatorname{dom} p - \operatorname{dom} q$ *is a neighborhood of* 0, *then* $(p+q)^* = p^*\#q^*$;
(v) $(\lambda p)^* = \lambda * p^*$.

If $q : \Xi \times X \to \overline{\mathbb{R}}$ is a random lsc function, convex on X for all $\xi \in \Xi$, and such that its epi-integral with respect to the probability measure P is a (well-defined) proper, lsc, convex function, then

$$\left(e\text{-}\int_\Xi q(\xi, \cdot) * P(d\xi)\right)^* = \int_\Xi q^*(\xi, \cdot)\, P(d\xi),$$

where $q^*(\xi, \cdot)$ is the conjugate of $q(\xi, \cdot)$.

The next theorem focuses on the bicontinuity of the Legendre-Fenchel transform under epi-convergence; refer to §1. A sequence of functions, $q^\nu : X \to \overline{\mathbb{R}}$ is *equicoercive* if for all $\{x^\nu, \nu \in \mathbb{N}\}$, $\sup_\nu q^\nu(x^\nu) < \infty$ implies $\sup_\nu |x^\nu| < \infty$.

THEOREM 2.2. *Let $q^\nu : X \to \overline{\mathbb{R}}$ be lsc, proper, convex, and equicoercive, where X is a separable, reflexive Banach space. Then $q = \text{w-e-}\lim_\nu q^\nu$ if and only if $q^* = e\text{-}\lim_\nu q^{\nu *}$.*

$$\begin{array}{ccc} q^\nu & \overset{\text{w-e}}{\to} & q \\ *\downarrow & & \uparrow * \\ q^{\nu *} & \overset{\text{e}}{\to} & q^* \end{array}$$

Here, w-e-$\lim_\nu$ refers to epi-convergence with respect to the weak topology on X, whereas e-$\lim_\nu$ refers to the epi-limit with respect to the strong topology on X^*. This last theorem leads to a dual method of analyzing a limit function q. Namely, given a sequence q^ν, one may first pass to the conjugate sequence, $q^{\nu *}$, find its epi-limit q^*, then compute the conjugate again to arrive back at q.

Now let us return to the problem at hand. For $k = 1, \ldots, \nu$, let

$$u^k \in \operatorname{argmin}\{g(\xi^k, u) \,|\, u \in H_0^1\}.$$

By the definitions of epi-addition and epi-multiplication,

$$\nu^{-1}(u_1 + \cdots + u_\nu) = \bar{u}^\nu \in \operatorname{argmin}_{H_0^1} \nu^{-1} * [g(\xi^1, u)\# \cdots \# g(\xi^\nu, u)].$$

Moreover, assuming that the conditions laid out in Theorem 2.1(ii) and (iii) are satisfied, one has

$$\nu^{-1} * [g(\xi^1, \cdot)\# \cdots \# g(\xi^\nu, \cdot)] = \left(\frac{1}{\nu}\sum_{k=1}^{\nu} f(\xi^k, \cdot)\right)^*,$$

where $f(\xi^k, \cdot) = g^*(\xi^k, \cdot)$ for $k = 1, \ldots, \nu$.

As expected, when the sequence of random lsc (convex) functions $\{g(\boldsymbol{\xi}^k, \cdot), k = 1, \ldots\}$ is ergodic, so is the sequence of random lsc (convex) functions $\{f(\boldsymbol{\xi}^k, \cdot), k = 1, \ldots\}$. This allows us to apply our Ergodic Theorem to conclude that

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

where

$$Ef = \int_\Xi f(\xi, \cdot)\, P(d\xi).$$

Assuming again that the conditions laid in Theorem 2.1 are satisfied as well as g^{hom} lsc,

$$(Ef)^* = g^{\text{hom}} = \text{e-}\int g(\xi,\cdot) * P(d\xi).$$

There now remains only to appeal to Theorem 2.2 to obtain

$$\begin{aligned} g^{\text{hom}} &= \text{w-e} - \lim_{\nu} \left(\nu^{-1} * [g(\boldsymbol{\xi}^1,\cdot)\#\cdots\# g(\boldsymbol{\xi}^\nu,\cdot)]\right) \\ &= \text{w-e} - \lim_{\nu} \left(\nu^{-1}[f(\boldsymbol{\xi}^1,\cdot)+\cdots+f(\boldsymbol{\xi}^\nu,\cdot)]\right)^*. \end{aligned}$$

Following this procedure, the homogenized functional may be evaluated (numerically) and its properties can be analyzed, and this for a much larger class of problems than is possible via other suggested methods. Moreover, the convergence of minimizers of epi-convergent functions (Theorem 7.2) implies the (weak) convergence of solutions $\bar{u}^\nu$ to the homogenized solution $\bar{u}$. Similar techniques may also be employed to compute the effective coefficient $a(\cdot)$.

EXAMPLE 3: SOLUTION PROCEDURES FOR STOCHASTIC OPTIMIZATION PROBLEMS. Stochastic programming models deal with decision making under uncertainty. We consider the following simple, but quite general, formulation of a stochastic programming problem:

$$\min_{x\in\mathbb{R}^n} E\{f(\boldsymbol{\xi},x)\} = Ef(x),$$

where

- $f : \Xi \times \mathbb{R}^n \to \overline{\mathbb{R}} = [-\infty,\infty]$, $f(\xi,x)$ is the "cost" associated with a decision x when the random variable $\boldsymbol{\xi}$ takes on the value ξ;
- $\boldsymbol{\xi}$ is a $\mathbb{R}^N$-valued random variable with possible values in $\Xi \subset \mathbb{R}^N$; and
- $Ef : \mathbb{R}^n \to \overline{\mathbb{R}}$, the function to be minimized,

is defined by

$$Ef(x) = \int_\Xi f(\xi,x)\,P(d\xi).$$

The standard two-stage and multistage stochastic programs with recourse (Kall and Wallace 1994, Birge and Louveaux 1997) can be recast in this format; cf., Wets (1989) for details.

Minimizing Ef on $\mathbb{R}^n$ is basically a nonlinear programming problem, but to evaluate the function, to obtain its gradients or subgradients, requires N-dimensional integration. Except when $N \approx 1$, the N-dimensional integration can be a real challenge. Moreover, this integration should not involve an unreasonable amount of computational effort since evaluating the function $f(\cdot,x)$ might itself require extensive computations.

One way to proceed is to replace the given problem with a sampled one. Suppose $\xi^1,\ldots,\xi^\nu$ are samples of $\boldsymbol{\xi}$. One can view

$$\min \frac{1}{\nu}\sum_{k=1}^{\nu} f(\xi^k,x) \quad \text{such that } x\in\mathbb{R}^n$$

as a "sample" of the stochastic optimization problem. Hopefully, the solution x^ν of such a problem will provide an acceptable approximation of a solution $x^* \in \operatorname{argmin} Ef$. To justify such an approach, one might appeal to the Law of Large Numbers for random lsc functions (Arstein and Wets 1995), provided the samples are iid. But there are some instances when one cannot assert independence. At best, one might be able to claim stationarity. This is typically the case when dealing with applications where the uncertainty comes from the environment, cf., King et al. (1988) for an application dealing with lake eutrophication management and Salinger (1997) for an application involving the control of water reservoirs to generate hydropower. In this more general situation, one can appeal to our ergodic theorem and the convergence of the minimizers of epi-converging functions (Theorem 7.2) to claim (under certain additional conditions) that

$$x^\nu(\xi^1,\dots,\xi^\nu) \in \operatorname{argmin} \frac{1}{\nu}\sum_{k=1}^{\nu} f(\xi^k, x)$$

converge almost surely to $x^* \in \operatorname{argmin} Ef$, assuming that $\operatorname{argmin} Ef$ is nonempty.

3. Scalarization of random lsc functions. The setting in which we work is that introduced in §1: (X,d) is a Polish space with $\mathcal{B}$ the Borel field on X and $(\Xi, \mathcal{S}, P)$ a probability space, but we do not assume, as in §1, that $\mathcal{S}$ is P-complete. The definition of a random lsc function needs then to be slightly more general; the definitions coincide if $\mathcal{S}$ is P-complete, see below. Now it will be shown that in this setting, a random lsc function f is completely identified by a countable collection of extended real-valued random variables

$$f \longleftrightarrow \{\pi_{x,\rho} \mid x \in R, \rho \in \mathbb{Q}_+\}, \quad \text{where } R \text{ is a countable dense subset of } X.$$

We shall refer to such a vector $\{\pi_{x,\rho}, x \in R, \rho \in \mathbb{Q}_+\}$ as a *scalarization* of f. Vogel (1995) already recognized that scalarization, to which she refers as "the approach via pointwise convergence," could play a role in obtaining a Law of Large Numbers for random lsc functions.

We begin with the notion of a random set. A set-valued mapping $S : \Xi \rightrightarrows X$ is a *random set* if it is a *measurable* mapping, i.e., for any open set $O \subset X$,

$$\{\xi \in \Xi \mid S(\xi) \cap O \neq \varnothing\} =: S^{-1}(O) \in \mathcal{S}.$$

It is a *random closed set* if, in addition, it is *closed-valued*, i.e., for all $\xi \in \Xi$, $S(\xi)$ is closed. One can then also view S as a function from Ξ to cl-sets(X), the *(hyper)space* of closed subsets of X. The *Effrös field* on cl-sets(X) is the σ-field $\mathcal{E}(X)$ generated by all sets of the form

$$\mathcal{E}_O = \{C \in \text{cl-sets}(X) \mid C \cap O \neq \varnothing\}, \quad O \subset X, \quad \text{open};$$

cf. Effrös (1965) and Beer (1993). It is clear that the closed-valued mapping $S : \Xi \rightrightarrows X$ is measurable if and only if it is $(\mathcal{S}, \mathcal{E}(X))$-measurable when viewed as a function from $(\Xi, \mathcal{S})$ to $(\text{cl-sets}(X), \mathcal{E}(X))$.

We will need some known properties of random sets (Castaing and Valadier 1977, Himmelberg 1975, Salinetti and Wets 1986) listed here below. The first one follows directly from the definition of a random set.

PROPOSITION 3.1. *Suppose $S : \Xi \rightrightarrows X$ is a random set and for all $\xi \in \Xi$, let* $\operatorname{cl} S(\xi) = \operatorname{cl}(S(\xi))$. *Then* $\operatorname{cl} S : \Xi \rightrightarrows X$ *is a random closed set.*

PROOF. For $O \subset X$ open, clearly $O \cap S(\xi) \neq \varnothing$ if and only if $O \cap \operatorname{cl} S(\xi) \neq \varnothing$. □

That measurability is preserved under projections might not be surprising, but it is a nontrivial result:

THEOREM 3.2: MEASURABLE PROJECTION THEOREM (Castaing and Valadier 1977, Theorem III.23). *Suppose $\mathcal{S}$ is P-complete and G is an $\mathcal{S} \otimes \mathcal{B}$-measurable subset of $\Xi \times X$. Then,* $\operatorname{prj}_\Xi G \in \mathcal{S}$, *i.e., the projection of G on Ξ is $\mathcal{S}$ measurable.*

And finally, a series of measurability criteria for closed-valued mappings that will set the stage for our scalarization result.

PROPOSITION 3.3. *A closed-valued mapping $S : \Xi \rightrightarrows X$ is a random-closed set if and only if $S^{-1}(D) \in \mathcal{S}$ for all $D \in \mathcal{D}$, where $\mathcal{D}$ is any one of the following collections of subsets* *of X:*

(a) $\mathcal{D}$ = *the open balls* $\mathbb{B}^o(x,\rho) = \{x' \in X \mid d(x',x) < \rho\}$;

(b) $\mathcal{D}$ = *the open rational balls* $\mathbb{B}^o(x,\rho)$ *with* $x \in R$, *a dense countable subset of* X, *and* $\rho \in \mathbb{Q}_+$.

Moreover, if X is σ-compact or if $\mathcal{S}$ is P-complete, then $\mathcal{D}$ can also be any one of the following collection of subsets of X:

(c) $\mathcal{D}$ = *the closed sets C;*

(d) $\mathcal{D}$ = *the closed balls* $\mathbb{B}(x,\rho)$;

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

PROOF. Clearly, S measurable $\Longrightarrow$(a) $\Longrightarrow$ (b). Since X is Polish, every open set O can be written as the countable union of open rational balls: With $x^\nu \in R$ and $\rho^\nu \in \mathbb{Q}_+$, one has

$$O = \bigcup_{\nu=1}^{\infty} \mathbb{B}^o(x^\nu,\rho^\nu), \qquad S^{-1}(O) = \bigcup_{\nu=1}^{\infty} S^{-1}(\mathbb{B}^o(x^\nu,\rho^\nu)) \in \mathcal{S}.$$

Also, (c) $\Longrightarrow$ (d) $\Longrightarrow$ (e) does not need proof. An argument similar to the one above yields (e) $\Longrightarrow$ S measurable.

Now let us assume that X is σ-compact and show that S measurable $\Longrightarrow$ (c). Every closed set $C \subset X$ can now be written as the countable union of compact sets $\{B^\nu, \nu \in \mathbb{N}\}$ from which follows that $S^{-1}(C) = \bigcup_{\nu\in\mathbb{N}} S^{-1}(B^\nu)$. It now suffices to observe that for all $\nu \in \mathbb{N}$, $S^{-1}(B^\nu) \in \mathcal{S}$. Indeed, even with X just a metric space, given any nonempty, compact set $D \subset X$, define the open sets $D^\nu := \{x \in X \mid d(x,D) < 1/\nu\}$ for $\nu \in \mathbb{N}$. Since $S(\xi) \cap D \neq \varnothing$ if and only if $S(\xi) \cap D^\nu \neq \varnothing$ for all $\nu \in \mathbb{N}$, we have $S^{-1}(D) = \bigcap_\nu S^{-1}(D^\nu)$. Hence, $S^{-1}(D)$ is the intersection of a countable collection of measurable sets, and therefore is itself measurable.

Finally, let us assume that $\mathcal{S}$ is P-complete and again show that S measurable $\Longrightarrow$ (c). Let R be a countable dense subsets of X. Since S is closed-valued, $\bar{x} \in S(\xi)$ if and only if for all $\rho \in \mathbb{Q}_+$, there exists $x_\rho \in R$, such that $\bar{x} \in \mathbb{B}^o(x_\rho,\rho)$ and $S(\xi) \cap \mathbb{B}^o(x_\rho,\rho) \neq \varnothing$. This means that

$$\begin{aligned} \text{gph}S &= \{(\xi,x) \in \Xi \times X | x \in S(\xi)\} \\ &= \bigcap_{\rho\in\mathbb{Q}_+} \bigcup_{x\in R} [S^{-1}(\mathbb{B}^o(x,\rho)) \times \mathbb{B}^o(x,\rho)]. \end{aligned}$$

is a $\mathcal{S} \otimes \mathcal{B}$-measurable subset of $\Xi \times X$, where $\mathcal{B}$ is the Borel field on (X,d). Indeed, by (b), each set $S^{-1}(\mathbb{B}^o(x,\rho)) \times \mathbb{B}^o(x,\rho)$ belongs to $\mathcal{S} \times \mathcal{B}$ and gph S can be written as the countable intersection of a countable union of sets of this type. To complete the proof, one appeals to the Projection Theorem 3.2 which yields

$$S^{-1}(C) = \text{prj}_\Xi\left(\text{gph } S \cap (\Xi \times C)\right) \in \mathcal{S},$$

since gph $S \cap (\Xi \times C) \in \mathcal{S} \otimes \mathcal{B}$. $\square$

Let lsc-fcns(X) denote the space of extended real-valued, lower semicontinuous (lsc) functions from X to $\overline{\mathbb{R}}$. A *random lsc function* is a function $f : \Xi \to$ lsc-fcns(X) such that the associated *epigraphical mapping,*

$$S_f : \Xi \rightrightarrows X \times \mathbb{R} \quad \text{with} \quad S_f(\xi) := \text{epi} f(\xi,\cdot) = \{(x,\alpha) \in X \times \mathbb{R} \mid \alpha \geq f(\xi,x)\},$$

is a random closed set. Note that a necessary (but *not sufficient*) condition for a function f : $\Xi \to$ lsc-fcns(X) to be a random lsc function is that for each fixed $x \in X$, the function $\xi \mapsto f(\xi,x)$ is measurable (Rockafellar and Wets 1998, Proposition 14.28). It will sometimes be convenient to identify an lsc function $f(\xi)$ with its bivariate representation $(\xi, x) \mapsto f(\xi, x)$. If f is a random lsc functions, then $(\xi, x) \mapsto f(\xi, x)$ is $\mathcal{S} \otimes \mathcal{B}$-measurable. On the other hand, if the bivariate representation of a function $f : \Xi \to \text{lsc-fcns}(X)$ is such that $(\xi, x) \mapsto f(\xi, x)$ is $\mathcal{S} \otimes \mathcal{B}$-measurable *and* $(\Xi, \mathcal{S})$ *is P-complete*, then f is a random lsc function (Rockafellar and Wets 1998, Proposition 14.34).

The concept of a random lsc function is due to Rockafellar (1976), who introduced it in the context of the calculus of variations under the name of "normal integrand." Further properties of random lsc functions are set forth in Rockafellar and Wets (1998, Chapter 14), Vervaat (1988), and here in §4–6.

In view of the definition of a random lsc function, there is a one-to-one correspondence between the lsc functions on X and the closed subsets of $X \times \mathbb{R}$ that are epigraphs. The *Effrös field* on lsc-fcns(X), simply denoted $\mathscr{E}$, can be identified with the restrictions of the Effrös field on cl-sets$(X \times \mathbb{R})$ to the closed subsets of $X \times \mathbb{R}$ that are epigraphs. $\mathscr{E}$ not only includes all sets of the form

$$\mathscr{E}_{(O,\alpha)} := \left\{ f \in \text{lsc-fcns}(X) \mid \inf_O f < \alpha \right\}, \; O \subset X, \quad \text{open}, \quad \alpha \in \mathbb{R},$$

but also can be generated by these sets; simply observe that $\inf_O f < \alpha$ if and only if epi f misses the open set $O \times (-\infty, \alpha)$, cf. Salinetti and Wets (1986) for example.

As mentioned at the outset of this section, random lsc functions can be characterized in terms of certain random vectors (with entries in $\overline{\mathbb{R}}$), to which we refer as "scalarizations" of the random lsc functions. The following theorem equates the measurability of lsc-fcns(X)-valued random mappings (implying that these mappings are random lsc functions) with the measurability of the corresponding scalarizations. Further probabilistic properties attainable through scalarization will be uncovered in §§4–5, and exploited later in §8 in the proofs of the ergodic theorems.

THEOREM 3.4: SCALARIZATION OF RANDOM LSC FUNCTIONS. *Let* $f : \Xi \to \text{lsc-fcns}(X)$,

$$\textit{and for } D \subset X: \qquad \textit{let } \pi_D(\xi) := \inf_{x \in D} f(\xi, x).$$

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 subsets of X:

(a) $\mathcal{D} =$ *the open sets O;*

(b) $\mathcal{D} =$ *the open balls $\mathbb{B}^o(x, \rho) = \{x' \in X \mid d(x', x) < \rho\}$;*

(c) $\mathcal{D} =$ *the open rational balls $\mathbb{B}^o(x, \rho)$ with $x \in R$, a dense countable subset of X, and $\rho \in \mathbb{Q}_+$.*

Moreover, if X is σ-compact or $\mathcal{S}$ is complete, then $\mathcal{D}$ can also be any one of the following collection of subsets of X:

(d) $\mathcal{D} =$ *the closed sets C;*

(e) $\mathcal{D} =$ *the closed balls $\mathbb{B}(x, \rho)$;*

(f) $\mathcal{D} =$ *the closed rational balls $\mathbb{B}(x, \rho)$ with $x \in R$ and $\rho \in \mathbb{Q}_+$.*

PROOF. Let (m) stand for the property: f is a random lsc function. We show that

$$(\mathrm{m}) \Longrightarrow (\mathrm{a}) \Longrightarrow (\mathrm{b}) \Longrightarrow (\mathrm{c}) \Longrightarrow (\mathrm{m}).$$

Only the implications (m) $\Longrightarrow$ (a) and (c) $\Longrightarrow$ (m) need proof.
(m) $\Longrightarrow$ (a): For any open set $O \subset X$ and $\alpha \in \mathbb{R}$,

$$\begin{aligned}\pi_O^{-1}((-\infty, \alpha)) &= \{\xi \in \Xi \mid \inf_O f(\xi, \cdot) < \alpha\} \\ &= \{\xi \in \Xi \mid S_f(\xi) \cap (O \times (-\infty, \alpha)) \neq \emptyset\} = S_f^{-1}(O \times (-\infty, \alpha)) \in \mathcal{S}\end{aligned}$$

since $O \times (-\infty, \alpha)$ is open and S_f is a random closed set. In particular, it implies $\pi = \pi_X$ measurable.

(c) $\Longrightarrow$ (m): Because for each $\xi \in \Xi$, $S_f(\xi)$ is an epigraph, for any set $D \subset X$.

$$\begin{aligned}\{\xi \in \Xi \mid \pi_D(\xi) < \beta\} &= \{\xi \in \Xi \mid S_f(\xi) \cap (D \times (-\infty, \beta)) \neq \emptyset\} \\ &= \{\xi \in \Xi \mid S_f(\xi) \cap (D \times (\alpha, \beta)) \neq \emptyset\} \ \forall\ \alpha < \beta.\end{aligned}$$

It is assumed that these sets belong to $\mathcal{S}$ when D is an open rational ball. Since every open set $O \subset X \times \mathbb{R}$ can be written as the countable union of sets of the type $O^\nu = \mathbb{B}^o(x^\nu, \rho^\nu) \times (\alpha^\nu, \beta^\nu)$ and $S_f^{-1}(O^\nu) \in \mathcal{S}$, it follows that

$$S_f^{-1}(O) = \bigcup_{\nu=1}^{\infty} S_f^{-1}(O^\nu) \in \mathcal{S},$$

which, by Proposition 3.3(b), implies that S_f is measurable. It is a random closed set since the lower semicontinuity of f implies that S_f is also closed-valued.

Now, let us assume that X is σ-compact or $\mathcal{S}$ is P-complete and let us show that under either one of these additional assumptions,

$$(\mathrm{m}) \Longrightarrow (\mathrm{d}) \Longrightarrow (\mathrm{e}) \Longrightarrow (\mathrm{f}) \Longrightarrow (\mathrm{m}).$$

Again, only the implications (m) $\Longrightarrow$ (d) and (f) $\Longrightarrow$ (m) need proof.

(m) $\Longrightarrow$ (d): It has already been established that if f is a random lsc function, $\pi = \pi_X$ is measurable. Given any closed set $C \subset X$, the function g with

$$g(\xi, x) = \begin{cases} f(\xi, x) & \text{if } x \in C, \\ \infty & \text{otherwise,} \end{cases}$$

is again a random lsc function, since, in view of Proposition 3.3(c), when X is σ-compact or $\mathcal{S}$ is P-compete, for any closed set $D \subset X \times \mathbb{R}$,

$$S_g^{-1}(D) = \{\xi \in \Xi \mid S_f(\xi) \cap (C \times \mathbb{R}) \cap D \neq \emptyset\} = S_f^{-1}((C \times \mathbb{R}) \cap D) \in \mathcal{S}$$

and thus S_g is measurable and clearly closed-valued. Hence, $\xi \mapsto \inf g(\xi, \cdot) = \pi_C(\xi)$ is measurable as noted at the end of the proof that (m) $\Longrightarrow$ (a).

(f) $\Longrightarrow$ (m): It suffices to show that (f) $\Longrightarrow$ (a) since (a) $\Longrightarrow$ (m). Every open subset O of X can be written as the countable union of closed balls: With $x^\nu \in R$ and $\rho^\nu \in \mathbb{Q}_+$, one has

$$O = \bigcup_{\nu=1}^{\infty} \mathbb{B}(x^\nu, \rho^\nu), \qquad \pi_O = \inf_\nu \pi_{\mathbb{B}(x^\nu, \rho^\nu)},$$

i.e., π_O can be obtained as the infimum of a countable collection of measurable function and consequently is measurable. □

COROLLARY 3.5: COUNTABLE SCALARIZATION. *Let $f : \Xi \to$ lsc-fcns(X). For $x \in R$, a countable dense subset of X, and $\rho \in \mathbb{Q}_+$, define*

$$\pi_{x,\rho}(\xi) := \pi_{\mathbb{B}^o(x,\rho)}(\xi) = \inf_{x' \in \mathbb{B}^o(x,\rho)} f(\xi, x').$$

Then f is a random lsc function if and only if the countable collection of functions

$$\{\pi_{x,\rho} : \Xi \to \overline{\mathbb{R}} \mid x \in R, \rho \in \mathbb{Q}_+\}$$

are measurable.

When either X is σ-compact or $\mathcal{S}$ is P-complete, such a countable collection can also be obtained by replacing the open balls $\{\mathbb{B}^o(x, \rho) \mid x \in R, \rho \in \mathbb{Q}_+\}$ by their closed counterparts.

PROOF. This is just a reformulation of parts (c) and (f) of the theorem. □

4. Lower semicontinuity of conditional expectations. We still work with (X, d), a Polish space, and $(\Xi, \mathcal{S}, P)$, a probability space. We are going to show that given a random lsc function $f : \Xi \to \text{lsc-fcns}(X)$ and any σ-field $\mathcal{R} \subset \mathcal{S}$, there always is a version of the conditional expectation of f with respect to $\mathcal{R}$ that is lsc for all $\xi \in \Xi$. The proof relies heavily on the scalarization results of the previous section.

Recall that a function $g : X \to \overline{\mathbb{R}}$ is *lsc at* $\bar{x}$ if for all $x^\nu \to \bar{x}$, $\liminf_\nu g(x^\nu) \geq g(\bar{x})$. Equivalently, g is lsc at $\bar{x}$ if for some (decreasing) sequence $\{V^\nu \subset X,\ \nu \in \mathbb{N}\}$ of neighborhoods of $\bar{x}$ such that $V^\nu \supset V^{\nu+1}$ and $\bigcap_{\nu\in\mathbb{N}} V^\nu = \{\bar{x}\}$,

$$p^\nu(\bar{x}) := \inf_{x\in V^\nu} g(x) \nearrow g(\bar{x}).$$

Furthermore, g is *lsc* if and only if $\operatorname{epi} g \subset X \times \mathbb{R}$ is closed.

PROPOSITION 4.1. *Let R be a countable dense subset of X. A function $g : X \to \overline{\mathbb{R}}$ is lsc at $\bar{x}$ if and only if there exist $x^\nu \in R$ and $\rho^\nu \in \mathbb{Q}_+$, such that $x^\nu \to \bar{x}$, $\rho^\nu \searrow 0$, $d(x^\nu, \bar{x}) < \rho^\nu$ and*

$$\inf_{x\in\mathbb{B}^o(x^\nu,\rho^\nu)} g(x) =: p_{x^\nu,\rho^\nu} \nearrow g(\bar{x}).$$

PROOF. Simply observe that one can then choose a decreasing (sub)sequence,

$$\{V^\nu := \mathbb{B}^o(x^\nu, \rho^\nu), \quad x^\nu \in R, \quad \rho^\nu \in \mathbb{Q}_+\},$$

of neighborhoods of $\bar{x}$, such that $\bigcap_\nu V^\nu = \{\bar{x}\}$. And g is lsc at $\bar{x}$ if and only if $\inf_{V^\nu} g \nearrow g(\bar{x})$, as recalled above. □

The existence of an lsc-version of the conditional expectation will be proved under a technical assumption that is also required to obtain the Ergodic Theorem 8.2.

DEFINITION 4.2. A random lsc function $f : \Xi \to \text{lsc-fcns}(X)$ is locally inf-integrable if for every $x \in X$ there is a closed neighborhood V of x such that for the (scalar) function

$$\xi \mapsto \pi_V(\xi) := \inf_{x'\in V} f(\xi, x') : \quad E\{\pi_V\} > -\infty.$$

From f a random lsc function, it already follows that for any closed set V, the function $\xi \mapsto \inf_V f(\xi, \cdot)$ is $\mathcal{S}$-measurable (Rockafellar and Wets 1998, Theorem 14.37).

THEOREM 4.3. *Let $f : \Xi \to$ lsc-fcns(X) be a locally inf-integrable random lsc function and $\mathcal{R} \subset \mathcal{S}$ a σ-field. Then, there exists a version $E^{\mathcal{R}} f$ of the conditional expectation of f that is* lsc-fcns(X)*-valued, i.e., for all $\xi \in \Xi$, $E^{\mathcal{R}} f(\xi)$ is an lsc function.*

Moreover, there is an lsc-version with the following property: For all $\xi \in \Xi'$ with $\Xi' \subset \Xi$ of P-measure 1, *and $x \in X$,*

$$E^{\mathcal{R}} f(\xi, x) = \lim_{\nu\to\infty} \pi^{\mathcal{R}}_{x^\nu,\rho^\nu}(\xi),$$

where $x^\nu \to x$, $\rho^\nu \searrow 0$ and $\pi^{\mathcal{R}}_{x^\nu,\rho^\nu}$ is a version of the conditional expectation of the scalar function,

$$\pi_{x^\nu,\rho^\nu}(\xi) = \inf\{f(\xi, x') \mid x' \in \mathbb{B}^o(x^\nu, \rho^\nu)\},$$

for $\mathbb{B}^o(x^\nu, \rho^\nu)$, the open ball centered at x^ν of radius ρ^ν.

PROOF. We actually prove the second assertion which implies the first one. Let R be a countable dense subset of X and define

$$\{\pi_{x,\rho}(\xi) = \inf_{\mathbb{B}^o(x,\rho)} f(\xi) \mid x \in R, \rho \in \mathbb{Q}_+\},$$

a scalarization of f. Let $\pi^{\mathcal{R}}_{x,\rho} := E^{\mathcal{R}}\{\pi_{x,\rho}\}$ be a version of the conditional expectation of $\pi_{x,\rho}$ with respect to $\mathcal{R}$.

If $\pi_{x,\rho} \leq \pi_{x',\rho'}$ P-a.s., then $\pi^{\mathcal{R}}_{x,\rho} \leq \pi^{\mathcal{R}}_{x',\rho'}$ P-a.s. The inequality might fail on a set of measure 0. Since there are only a countable number of possible pairs (x, ρ, x', ρ'), the union of all such sets, i.e., on which the inequality does not hold, is of measure 0. So, let Ξ_0 be the subset of Ξ of P-measure 1, such that

$$\pi^{\mathcal{R}}_{x,\rho} \leq \pi^{\mathcal{R}}_{x',\rho'} \text{ on } \quad \Xi_0 \quad \text{whenever} \quad \pi_{x,\rho} \leq \pi_{x',\rho'} P\text{-a.s.}$$

Given $\bar{x} \in X$, choose $x^\nu \to \bar{x}$ with $x^\nu \in R$ and $\rho^\nu \searrow 0$ with $\rho^\nu \in \mathbb{Q}_+$, such that $\mathbb{B}^o(x^\nu, \rho^\nu)$ is a neighborhood of both x^ν and $\bar{x}$, and $\mathbb{B}^o(x^{\nu+1}, \rho^{\nu+1}) \subset \mathbb{B}^o(x^\nu, \rho^\nu)$. By the lower semicontinuity of $f(\xi, \cdot)$ at $\bar{x}$, we know from Proposition 4.1 that $\pi_{x^\nu,\rho^\nu}(\xi) \nearrow f(\xi, \bar{x})$ for all $\xi \in \Xi$. In view of the above, for all $\xi \in \Xi_0$, the sequence $\{\pi^{\mathcal{R}}_{x^\nu,\rho^\nu}(\xi),\ \nu \in \mathbb{N}\}$ is monotone increasing. For $\xi \in \Xi$, define

$$f^{\mathcal{R}}(\xi, \bar{x}) := \lim_{\nu \to \infty} \pi^{\mathcal{R}}_{x^\nu,\rho^\nu}(\xi).$$

Let us first observe that the value assigned to $f^{\mathcal{R}}(\xi, \bar{x})$ is independent of the choice of the sequences $x^\nu \to \bar{x}$ and $\rho^\nu \searrow 0$. Indeed, let $\hat{x}^\nu \to \bar{x}$ and $\hat{\rho}^\nu \searrow 0$ be another pair of sequences satisfying the conditions: $x^\nu \in R$, $\rho^\nu \in \mathbb{Q}_+$, $\mathbb{B}^o(\hat{x}^\nu, \hat{\rho}^\nu)$ is a neighborhood of both $\hat{x}^\nu$ and $\bar{x}$ and $\mathbb{B}^o(\hat{x}^{\nu+1}, \rho^{\nu+1}) \subset \mathbb{B}^o(\hat{x}^\nu, \rho^\nu)$. Because both sequences of balls are decreasing neighborhoods of $\bar{x}$, for ν sufficiently large, $\mathbb{B}^o(x^\nu, \rho^\nu) \supset \mathbb{B}^o(\hat{x}^\mu, \hat{\rho}^\mu)$ for some $\mu \geq \nu$ and vice-versa. This implies that on Ξ_0, $\pi^{\mathcal{R}}_{x^\nu,\rho^\nu} \leq \pi^{\mathcal{R}}_{\hat{x}^\mu,\hat{\rho}^\mu}$, for some $\mu \geq \nu$ and vice-versa. Thus, both sequences must have the same limit.

Since $\pi_{x^\nu,\rho^\nu} \nearrow f(\cdot, \bar{x})$, the Monotone Convergence Theorem for conditional expectations, appealing here to local inf-integrability, implies that $f^{\mathcal{R}}$ is actually a version of the conditional expectation of f.

We show next that for all $\xi \in \Xi_0$, $f^{\mathcal{R}}(\xi, \cdot)$ is lsc. Consider a sequence $x^\nu \to \bar{x}$, $x^\nu \in R$, and pick a subsequence and $\rho^\nu \searrow 0$, $\rho^\nu \in \mathbb{Q}_+$, such that $\mathbb{B}^o(x^\nu, \rho^\nu)$ is a neighborhood of both x^ν and $\bar{x}$, and $\mathbb{B}^o(x^{\nu+1}, \rho^{\nu+1}) \subset \mathbb{B}^o(x^\nu, \rho^\nu)$. Then,

$$f^{\mathcal{R}}(\cdot, x^\nu) \geq \pi^{\mathcal{R}}_{x^\nu,\rho^\nu} \text{ on } \Xi_0.$$

Taking liminf of both sides yields $\liminf_\nu f^{\mathcal{R}}(\cdot, x^\nu) \geq f^{\mathcal{R}}(\cdot, \bar{x})$. Since this holds for any such subsequence of $x^\nu \to \bar{x}$, it must hold for the sequence as well.

So far, we only considered sequences $x^\nu \to \bar{x}$ with $x^\nu \in R$. In the case of an arbitrary sequence $\bar{x}^\nu \to \bar{x}$ with $\bar{x}^\nu \in X$ note that it is always possible to find a sequence $x^\nu \to \bar{x}$ with $x^\nu \in R$, $d(\bar{x}^\nu, x^\nu)$ going sufficiently rapidly to 0 so that with an appropriate choice of $\rho^\nu \searrow 0$, for all ν, $\mathbb{B}^o(x^\nu, \rho^\nu)$ is not only a neighborhood of x^ν and $\bar{x}$ but also of $\bar{x}^\nu$. The same type of argument then yields $\liminf_\nu f^{\mathcal{R}}(\cdot, \bar{x}^\nu) \geq f^{\mathcal{R}}(\cdot, \bar{x})$. Thus, for all $\xi \in \Xi_0$, $f^{\mathcal{R}}(\xi, \cdot)$ is lsc.

Finally, for $\xi \in \Xi_0$, set $(E^{\mathcal{R}} f)(\xi) = f^{\mathcal{R}}(\xi, \cdot)$ and otherwise simply set $(E^{\mathcal{R}} f)(\xi) = \operatorname{cl} f^{\mathcal{R}}(\xi, \cdot)$, where $\operatorname{cl} f^{\mathcal{R}}(\xi, \cdot)$ is the lower semicontinuous closure of $f^{\mathcal{R}}(\xi, \cdot)$. The function $E^{\mathcal{R}} f : \Xi \to \text{lsc-fcns}(X)$ is then an lsc version of the conditional expectation of f. □

5. Probabilistic framework. The examples of §2 all rely on certain "stationary" and "ergodic" properties of random lsc functions. We will now make these properties precise with a description of a probabilistic framework for random lsc functions.

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

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

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

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

For a sequence $\{f^\nu,\ \nu \in \mathbb{N}\}$ of random lsc functions, let us denote by P^∞ the probability measure on the sequence space $(\text{lsc-fcns}(X)^\infty, \mathscr{E}^\infty)$ that is consistent with the joint distribution of the f^ν. The existence of such a measure follows from Kolmogorov's extension theorem for random lsc functions, and the fact that the Effrös field is the Borel field associated with a topology τ_{cw} on lsc-fcns(X) that makes $(\text{lsc-fcns}(X), \tau_{cw})$ a Polish space, as covered in §6.

Properties such as independence, stationarity and ergodicity of sequences of random lsc functions may now all be defined in a straightforward manner. Random lsc functions are said to be *independent* if their distributions are independent. A sequence of random lsc functions $\{f^\nu,\ \nu \in \mathbb{N}\}$ is *pairwise independent* if for any pair $k, l \in \mathbb{N}$ and $\mathscr{A}_1, \mathscr{A}_2 \in \mathscr{E}$,

$$P_{f^k,f^l}(\mathscr{A}_1, \mathscr{A}_2) = P_{f^k}(\mathscr{A}_1)P_{f^l}(\mathscr{A}_2).$$

The sequence is *independent* if for any finite subcollection, $\{f^{\nu_1}, \dots, f^{\nu_k},\ k \in \mathbb{N}\}$,

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

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 $\{f^{\nu_1}, \dots, f^{\nu_k}\}$, $k \in \mathbb{N}$, any $l \in \mathbb{N}$ and any $\mathscr{A}_1, \dots, \mathscr{A}_k \in \mathscr{E}$, one has

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

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 $A \in \mathscr{S}$, $P(\varphi^{-1}(A)) = P(A)$. If f is a random lsc function, one verifies easily that the sequence $\{f, f\circ\varphi, f\circ\varphi^2, \dots\}$ 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 $(\text{lsc-fcns}(X)^\infty, \mathscr{E}^\infty)$. Redefine the f^ν as follows:

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

i.e., the νth element of the sequence $\zeta = (\zeta^1, \zeta^2, \dots) \in \text{lsc-fcns}(X)^\infty$. 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 a new probability space. Letting $\varphi : \text{lsc-fcns}(X)^\infty \to \text{lsc-fcns}(X)^\infty$ be the shift operator,

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

and defining $f : \text{lsc-fcns}(X)^\infty \to \text{lsc-fcns}(X)$ as $f(\zeta) = \zeta^1$, one has $f(\varphi^\nu(\zeta)) = \zeta^{\nu+1}$, so that $f, f\circ\varphi, f\circ\varphi^2, \dots,$ defines the same stationary sequence on lsc-fcns$(X)^\infty$ 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 $A \in \mathscr{S}$ 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 $\mathscr{J} \subset \mathscr{E}$ 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 $\mathscr{J}$ is trivial, i.e., for all $A \in \mathscr{J}$, $P(A) \in \{0, 1\}$. For f, a random lsc function, the sequence $\{f^\nu = f(\varphi^\nu(\cdot)),\ \nu \in \mathbb{N}\}$ of random lsc functions is ergodic if the associated (measure preserving) shift operator φ on the sequence space $(\text{lsc-fcns}(X)^\infty, \mathscr{E}^\infty, P^\infty)$ is ergodic.

In the remainder of this section, we show that stationarity and ergodicity properties of sequences of random lsc functions are inherited by the sequences of corresponding scalarizations.

Thus, given a sequence of random lsc functions $\{f^\nu : \Xi \to \text{lsc-fcns}(X),\ \nu \in \mathbb{N}\}$, one can always associate, by scalarization, a corresponding sequence of vector-valued random variables

$$\{\pi^\nu_{x,\rho},\ \nu \in \mathbb{N} \mid x \in R, \rho \in \mathbb{Q}_+\}.$$

As we demonstrate next, independence, stationarity, and ergodicity properties of the sequence of the random lsc functions are inherited by these sequences of vectors generated by scalarization. To do so, we rely on Dynkin's π-λ theorem and some of its consequences, which are briefly reviewed in Theorem 5.4, for details, cf. Durrett (1991).

Instead of restricting ourselves to scalar functions obtained via minimization over balls, we derive the results for minimization over arbitrary open sets.

DEFINITION 5.3. For a set Ω, $\mathcal{P}$ is a π-system (of subsets of Ω) if $\Omega \in \mathcal{P}$ and $\mathcal{P}$ is closed under intersections. $\mathcal{L}$ is a *λ-system* (of subsets of Ω) if it satisfies: $\Omega \in \mathcal{L}$, $B\backslash A \in \mathcal{L}$ for all A, $B \in \mathcal{L}$, such that $A \subset B$, and $A \in \mathcal{L}$ whenever $A = \bigcup_n A_n$ for a nested sequence $A_1 \subset A_2 \subset \cdots$ such that $A_n \in \mathcal{L}$.

THEOREM 5.4: DYNKIN'S π-λ THEOREM. *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.*

(a) *If $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k$ are independent π-systems, then the (generated) σ-fields σ-$(\mathcal{P}_1)$, σ-$(\mathcal{P}_2)$,$\ldots$, σ-$(\mathcal{P}_k)$ are also independent.*

(b) *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then σ-$(\mathcal{P}) \subset \mathcal{L}$.*

(c) *Let $\mathcal{P}$ be a π-system, P_1 and P_2 probability measures that agree on $\mathcal{P}$. Then P_1 and P_2 also agree on σ-$(\mathcal{P})$.*

For O^ν open subsets of X and $\alpha^\nu \in \mathbb{R} \cup \{\infty\}$, let

$$\mathcal{P} := \left\{\mathcal{A} = \{f \in \text{lsc-fcns}(X) | \pi_{O^\nu} \le \alpha^\nu,\ \forall \nu \in \mathbb{N}\}\right\}.$$

Observe that $\mathcal{P}$ is a generating class for the Effrös field $\mathcal{E}$ on lsc-fcns(X). Also define, the classes of product sets: for $k \in \mathbb{N}$,

$$\mathcal{P}^k = \mathcal{P} \times \mathcal{P} \times \cdots \times \mathcal{P} \quad (k \text{ times}),$$

and observe that

$$\sigma\text{-}(\mathcal{P}^k) = \mathcal{E}^k = \mathcal{E} \otimes \cdots \otimes \mathcal{E},$$

the σ-field generated by the product of k copies of $\mathcal{E}$.

LEMMA 5.5. *For all $k \in \mathbb{N}$, $\mathcal{P}^k$ are π-systems on* (lsc-fcns(X))^k.

PROOF. For any $k \in \mathbb{N}$ and $i \in \{1, \ldots, k\}$, one has

$$\mathcal{A}_i := \{f \in \text{lsc-fcns}(X) \mid \pi_{O_i^\nu} \le \infty\, \forall \nu \in \mathbb{N}\} = \text{lsc-fcns}(X),$$

and $\mathcal{A}_i \in \mathcal{P}$ whenever $\{O^\nu,\ \nu \in \mathbb{N}\}$ are open subsets of X. So, $\prod_{i=1}^k \mathcal{A}_i =$ (lsc-fcns(X))^k. whereby (lsc-fcns(X))$^k \in \mathcal{P}^k$. Now, for arbitrary collections of open subsets of X and scalars in $\mathbb{R} \cup \{\infty\}$,$\{(O_1^\nu, \alpha_1^\nu),\ \nu \in \mathbb{N}\}$ and $\{(O_2^\nu, \alpha_2^\nu),\ \nu \in \mathbb{N}\}$, let

$$\mathcal{A}_1 := \{f \in \text{lsc-fcns}(X) \mid \pi_{O_1^\nu} \le \alpha_1^\nu,\ \forall \nu \in \mathbb{N}\} \in \mathcal{P},$$

$$\mathcal{A}_2 := \{f \in \text{lsc-fcns}(X) \mid \pi_{O_2^\nu} \le \alpha_2^\nu,\ \forall \nu \in \mathbb{N}\} \in \mathcal{P}.$$

Then,

$$\mathcal{A}_1 \cap \mathcal{A}_2 = \{f \in \text{lsc-fcns}(X) \mid \pi_{O_i^\nu} \le \alpha_i^\nu,\ \forall \nu \in \mathbb{N},\ i = 1, 2\} \in \mathcal{P},$$

whereby $\mathcal{P}$ is a π-system. For $k \in \mathbb{N}$, $i = 1, \ldots, k$, let $\mathcal{A}_i^1$, $\mathcal{A}_i^2 \in \mathcal{P}$ and $\mathcal{A}^1 := \prod_{i=1}^k \mathcal{A}_i^1$, similarly $\mathcal{A}^2 := \prod_{i=1}^k \mathcal{A}_i^2$. Then, using the fact that $\mathcal{P}$ is a π-system, we obtain

$$\mathcal{A}^1 \cap \mathcal{A}^2 = \prod_{i=1}^k (\mathcal{A}_i^1 \cap \mathcal{A}_i^2) \in \mathcal{P}^k$$

as claimed. □

THEOREM 5.6. *Let (X, d) be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and f^1, f^2 random lsc functions defined on Ξ. Then, f^1 and f^2 are identically distributed if and only if for all O^ν open subsets of X, $\alpha^\nu \in \mathbb{R} \cup \{\infty\}$, $\nu \in \mathbb{N}$:*

$$P\{\xi \in \Xi \mid \pi^1_{O^\nu}(\xi) \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\} = P\{\xi \in \Xi \mid \pi^2_{O^\nu}(\xi) \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\}.$$

PROOF. Suppose f^1 and f^2 are identically distributed. Then, for all $\mathcal{A} \in \mathcal{E}$,

$$P\{\xi \in \Xi \mid f^1(\xi, \cdot) \in \mathcal{A}\} = P\{\xi \in \Xi \mid f^2(\xi, \cdot) \in \mathcal{A}\},$$

hence this holds in particular for the sets

$$\mathcal{A} = \{f \in \text{lsc-fcns}(X) \mid \pi_{O^\nu} \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\}.$$

For the reverse direction, suppose that for all $\nu \in \mathbb{N}$, O^ν open subsets of X, $\alpha^\nu \in \mathbb{R} \cup \{\infty\}$, one has

$$P\{\xi \in \Xi \mid \pi^1_{O^\nu}(\xi) \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\} = P\{\xi \in \Xi \mid \pi^2_{O^\nu}(\xi) \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\}.$$

For $i = 1, 2$, let μ_i be the measure induced by f^i on $\mathcal{E}$, i.e., for $\mathcal{A} \in \mathcal{E}$, $\mu_i(\mathcal{A}) = P\{\xi \in \Xi \mid f^i(\xi, \cdot) \in \mathcal{A}\}$. Then, by supposition, $\mu_1 = \mu_2$ on $\mathcal{P}$. Since $\mathcal{P}$ generates the Effrös field, applying Theorem 5.4(c), yields $\mu_1 = \mu_2$ on $\mathcal{E}$. □

THEOREM 5.7. *Let (X, d) be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and $\{f^\nu,\ \nu \in \mathbb{N}\}$ a sequence of random lsc functions defined on Ξ. Then, the sequence $\{f^\nu,\ \nu \in \mathbb{N}\}$ is independent if and only if for all $k \in \mathbb{N}$, indices $\ell_1, \ldots, \ell_k$, scalars $\alpha_1^\nu, \ldots, \alpha_k^\nu \in \mathbb{R} \cup \{\infty\}$ and $O_1^\nu, \ldots, O_k^\nu$ open subsets of X, $\nu \in \mathbb{N}$,*

$$P\{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \ldots, k,\ \forall\, \nu \in \mathbb{N}\} = \prod_{i=1}^k P\{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ \forall\, \nu \in \mathbb{N}\}.$$

In particular, for any open sets $O \subset X$, the sequence $\{\pi^\nu_O,\ \nu \in \mathbb{N}\}$ is independent whenever $\{f^\nu,\ \nu \in \mathbb{N}\}$ is independent.

PROOF. Suppose $\{f^\nu,\ \nu \in \mathbb{N}\}$ is independent. Then for all $k, \nu \in \mathbb{N}$, indices $\ell_1, \ldots, \ell_k$, for all $\mathcal{A}_i \in \mathcal{E}$, $i = 1, \ldots, k$,

$$P\{\xi \in \Xi \mid f^{\ell_i}(\xi, \cdot) \in \mathcal{A}_i,\ i = 1, \ldots, k\} = \prod_{i=1}^k P\{\xi \in \Xi \mid f^{\ell_i}(\xi, \cdot) \in \mathcal{A}_i\},$$

hence this holds in particular for the sets

$$\mathcal{A}_i = \{f \in \text{lsc-fcns}(X) \mid \pi_{O_i^\nu} \le \alpha_i^\nu\ \forall\, \nu \in \mathbb{N}\}.$$

Suppose now that the asserted identity holds for all $k \in \mathbb{N}$, indices $\ell_1, \ldots, \ell_k$, scalars $\alpha_1^\nu, \ldots, \alpha_k^\nu$ in $\mathbb{R} \cup \{\infty\}$, and $O_1^\nu, \ldots, O_k^\nu$ open subsets of X, $\nu \in \mathbb{N}$. Fix $k \in \mathbb{N}$, $\ell_1, \ldots, \ell_k \in \mathbb{N}$, and let

$$\mathcal{P}_i := \text{ sets of the form } \{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu} \le \alpha_i^\nu,\ \forall\, \nu \in \mathbb{N}\}.$$

$\mathcal{P}_i$ is a π-system and σ-$(\mathcal{P}_i) = \sigma$-(f^{ℓ_i}). The independence of the σ-(f^{ℓ_i}) follows from Theorem 5.4(a) which implies that the sequence $\{f^\nu,\ \nu \in \mathbb{N}\}$ is independent. □

COROLLARY 5.8. *Let (X, d) be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and $\{f^\nu,\ \nu \in \mathbb{N}\}$ a sequence of random lsc functions defined on Ξ. Then, $\{f^\nu,\ \nu \in \mathbb{N}\}$ is iid if and only if the two following conditions are satisfied:*

(a) *For any pair f^l, f^k,*

$$P\{\xi \in \Xi \mid \pi^l_{O^\nu}(\xi) \le \alpha^\nu\ \forall\, \nu \in \mathbb{N}\} = P\{\xi \in \Xi \mid \pi^k_{O^\nu}(\xi) \le \alpha^\nu,\ \forall\, \nu \in \mathbb{N}\},$$

O^ν open subsets of X and $\alpha^\nu \in \mathbb{R} \cup \{\infty\}$, $\nu \in \mathbb{N}$;

(b) *For all $k \in \mathbb{N}$, indices $\ell_1, \dots, \ell_k$, scalars $\alpha_1^\nu, \dots, \alpha_k^\nu \in \mathbb{R} \cup \{\infty\}$ and $O_1^\nu, \dots, O_k^\nu$ open subsets of X, $\nu \in \mathbb{N}$,*

$$P\{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \dots, k,\ \forall\, \nu \in \mathbb{N}\} = \prod_{i=1}^{k} P\{\xi \in \Xi \vert \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu\ \forall\, \nu \in \mathbb{N}\}.$$

In particular, if $\{f^\nu,\ \nu \in \mathbb{N}\}$ is iid, then $\{\pi_O^\nu\}$ is iid for any O an open subset of X.

The next two theorems establish the stationarity and ergodicity of a sequence of random lsc functions through scalarization.

THEOREM 5.9. *Let (X, d) be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and $\{f^\nu,\ \nu \in \mathbb{N}\}$ a sequence of random lsc functions defined on Ξ. Then, $\{f^\nu,\ \nu \in \mathbb{N}\}$ is stationary if and only if for all $k, r \in \mathbb{N}$, indices $\ell_1, \dots, \ell_k \in \mathbb{N}$, scalars $\alpha_1^\nu, \dots, \alpha_k^\nu \in \mathbb{R} \cup \{\infty\}$ and $O_1^\nu, \dots, O_k^\nu$ open subsets of X, $\nu \in \mathbb{N}$,*

$$P\{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \dots, k,\ \forall\, \nu\} = P\{\xi \in \Xi \mid \pi^{\ell_i + r}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \dots, k,\ \forall\, \nu\}.$$

In particular, for any open sets $O \subset X$, the sequence $\{\pi_O^\nu,\ \nu \in \mathbb{N}\}$ is stationary whenever $\{f^\nu,\ \nu \in \mathbb{N}\}$ is stationary.

PROOF. Suppose $\{f^\nu,\ \nu \in \mathbb{N}\}$ is stationary. Then, for all $k, r \in \mathbb{N}$, indices $\ell_1, \dots, \ell_k$, and for all $\mathcal{A}_i \in \mathcal{E}$, $i = 1, \dots, k$,

$$P\{\xi \in \Xi \mid f^{\ell_i}(\xi, \cdot) \in \mathcal{A}_i,\ i = 1, \dots, k\} = P\{\xi \in \Xi \mid f^{\ell_i + r}(\xi, \cdot) \in \mathcal{A}_i,\ i = 1 \dots, k\},$$

hence, this holds in particular for the sets

$$\mathcal{A}_i = \{f \in \text{lsc-fcns}(X) \mid \pi_{O_i^\nu} \le \alpha_i^\nu\ \forall\, \nu \in \mathbb{N}\}.$$

Suppose now that for all $k, r \in \mathbb{N}$, indices $\ell_1, \dots, \ell_k$, scalars $\alpha_1^\nu, \dots, \alpha_k^\nu \in \mathbb{R} \cup \{\infty\}$, and $O_1^\nu, \dots, O_k^\nu$ open subsets of X, $\nu \in \mathbb{N}$,

$$P\{\xi \in \Xi \mid \pi^{\ell_i}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \dots, k,\ \forall\, \nu\} = P\{\xi \in \Xi \mid \pi^{\ell_i + r}_{O_i^\nu}(\xi) \le \alpha_i^\nu,\ i = 1, \dots, k,\ \forall\, \nu\}.$$

For fixed $k, r, \ell_i \in \mathbb{N}$, $i = 1 \dots, k$, let μ_1 be the measure induced on $\mathcal{E}^k$ by $(f^{\ell_1}, \dots, f^{\ell_k})$. Let μ_2 be the measure induced on $\mathcal{E}^k$ by $(f^{\ell_1 + r}, \dots, f^{\ell_k + r})$. By supposition, $\mu_1 = \mu_2$ on $\mathcal{P}^k$. Hence, by Theorem 5.4(c), $\mu_1 = \mu_2$ on $\mathcal{E}^k$. □

THEOREM 5.10. *If $\{f \circ \varphi^\nu,\ \nu \in \mathbb{N}\}$ is an ergodic sequence of random lsc functions, then for all open $O \subset X$, $\{\pi_O \circ \varphi^\nu,\ \nu \in \mathbb{N}\}$ is an ergodic sequence of extended real-valued random variables.*

PROOF. The shift operator, $\varphi : \text{lsc-fcns}(X)^\infty \to \text{lsc-fcns}(X)^\infty$ is ergodic, and π_O defined on $\text{lsc-fcns}(X)^\infty$ by $\pi_O(\zeta) := \inf_O \zeta_1$ is measurable. Hence, the sequence $\{\pi_O \circ \varphi^\nu\}$, is ergodic, and equivalent to the original sequence. □

6. Kolmogorov's Extension Theorem. Let us now turn to Kolmogorov's Extension Theorem for $(\text{lsc-fcns}(X)^\infty, \mathscr{E}^\infty)$. Because random lsc functions can be identified with random closed sets defined on $(\Xi, \mathscr{S}, P)$ with values in $(X \times \mathbb{R})$, let us derive the result in the following framework: Given a sequence $\{S^\nu : \Xi \rightrightarrows X,\ \nu \in \mathbb{N}\}$ of random closed sets with X a Polish space, let us denote, for every ν, the *distribution* P^ν of S^ν, i.e.,

$$P^\nu(\mathscr{A}) := P(\{\xi \in \Xi \mid S^\nu(\xi) \in \mathscr{A}\}) \text{ for } \mathscr{A} \in \mathscr{E}(X),$$

and the *joint distribution* of a finite collection $\{S^{\nu_1}, \ldots, S^{\nu_k}\}$ of random sets is then

$$P_{\nu_1,\ldots,\nu_k}(\mathscr{A}_1, \ldots, \mathscr{A}_k) := P(\{\xi \in \Xi \mid S^1(\xi) \in \mathscr{A}_1, \ldots, S^k(\xi) \in \mathscr{A}_k\})$$

for $\mathscr{A}_1, \ldots, \mathscr{A}_k \in \mathscr{E}(X)$.

Assuming it exists, let us denote by P^∞ the probability measure on the sequence space $(\text{cl-sets}(X)^\infty, \mathscr{E}(X)^\infty)$ that is consistent with the joint distribution of any finite subcollection of random closed sets $\{S^{\nu_1}, \ldots, S^{\nu_k}\}$. To assert the existence of such a measure, one usually appeals to Varadarajan's version of the Kolmogorov Extension Theorem (Parthasarathy 1967, Theorem V.5.1):

THEOREM 6.1. *Let $\{Y^\nu,\ \nu \in \mathbb{N}\}$ be random variables defined on $(\Xi, \mathscr{S}, P)$ with values in a Polish space (Z, τ) and $\mathscr{B}$ the associated Borel field on Z. Then there exists a unique measure μ^∞ on $(Z^\infty, \mathscr{B}^\infty)$ that is consistent with the joint distribution of $(Y^{\nu_1}, \ldots, Y^{\nu_k})$ for any finite subcollection of indices $\{\nu_1, \ldots, \nu_k\} \subset \mathbb{N}$.*

However, this is not quite our situation. At this stage, we do not even have a topology on cl-sets(X) that would allow us to construct the Borel field on cl-sets(X). It turns out that the Effrös field is the Borel field generated from both the Fell topology, associated with (Painlevé-Kuratowski) set convergence, and by the Choquet-Wijsman topology, associated with the pointwise convergence of the distance functions (Beer 1993). While the Fell topology is the right one for set and epi-convergence (cf. Theorem 1.1), it is the Choquet-Wijsman topology that renders $(\text{cl-sets}(X), \tau_{cw})$ a Polish space, heading us towards the validation of the Kolmogorov's Extension Theorem that is required here. We elaborate on these statements in what follows.

The *Fell topology* τ_f (Fell 1962) on cl-sets(X) is the topology generated by a subbase consisting of sets of the form

$$\mathscr{V}_O = \{C \in \text{cl-sets}(Y) \mid C \cap O \neq \varnothing\}, \quad O \subset Y, \quad O \text{ open},$$

and

$$\mathscr{V}_K = \{C \in \text{cl-sets}(Y) \mid C \cap K = \varnothing\}, \quad K \subset Y, \quad K \text{ compact}.$$

That sequential τ_f-convergence on cl-sets(X) corresponds to set convergence of a sequence of closed subsets of a first countable Hausdorff space (hence a Polish space) was first shown in Francaviglia et al. (1985), and a straightforward proof of this fact can be found in Beer (1993).

PROPOSITION 6.2. *$\mathscr{E}(X) \subset \mathscr{B}_f$, where $\mathscr{B}_f$ is the Borel field on* cl-sets(X) *when it is equipped with τ_f, the Fell topology.*

PROOF. This is an immediate consequence of the fact that the sets $\mathscr{V}_O$ in the subbase for the Fell topology generate $\mathscr{E}(X)$. □

The Choquet-Wijsman topology τ_{cw} (Wijsman 1966, and Choquet 1966) on cl-sets(X) is defined in terms of the *distance functions*: $d_C(x) := \inf_{y \in C} d(x, y)$, where d is the metric on X; by convention: $d_\varnothing(x) = +\infty$. This topology τ_{cw}, consistent with the pointwise convergence of these distance functions, is generated by a subbase consisting of sets of the form

$$\mathscr{W}_{x,\alpha,-} := \{C \in \text{cl-sets}(X) \mid d_C(x) < \alpha\}, \quad x \in X, \quad \alpha > 0,$$

and

$$\mathcal{W}_{x,\alpha,+} := \{C \in \text{cl-sets}(X) \mid d_C(x) > \alpha\}, \quad x \in X, \quad \alpha > 0.$$

One has $C^\nu \to_{cw} C$ if and only if $d_{C^\nu}(x) \to d_C(x)$ for all $x \in X$; for more about this topology, see Beer (1993). It is known that the Choquet-Wijsman topology is finer than the Fell topology (Beer 1993). This immediately implies the inclusion,

$$\mathcal{B}_f \subset \mathcal{B}_{cw},$$

where $\mathcal{B}_{cw}$ is the Borel field on cl-sets(X) when it is equipped with the Choquet-Wijsman topology.

Hess (1986) showed that the Borel field generated by the Choquet-Wijsman topology is the Effrös field, but for the hyperspace of closed *nonempty* subsets of a separable space. The empty set in cl-sets(X) corresponds, in the epigraphical setting, to the function, $f \equiv \infty$. Therefore the empty set cannot be excluded from our considerations. Nevertheless, it can easily be shown that $\mathcal{B}_{cw} = \mathcal{E}(X)$ for cl-sets(X). For example, by referring to a fact mentioned by Beer (1991): Let cl-sets$_{\neq\varnothing}(X)$ denote the hyperspace of closed subsets of X, not including the empty set.

PROPOSITION 6.3: BEER 1991. *Suppose τ_a is a topology on* cl-sets(X) *such that the Borel field on* cl-sets$_{\neq\varnothing}(X)$ *generated by the τ_a-open sets is the Effrös field on* cl-sets$_{\neq\varnothing}(X)$. *Then, $\mathcal{E}(X)$ is the Borel field on* cl-sets(X) *generated by the τ_a-open sets if and only if $\{\varnothing\}$ is a Borel subset of* cl-sets(X). *In particular, this is true if $\{\varnothing\}$ is τ_a-closed.*

THEOREM 6.4. *On* cl-sets(X), $\mathcal{E}(X) = \mathcal{B}_{cw}$.

PROOF. We know from Hess (1986) that on cl-sets$_{\neq\varnothing}(X)$, $\mathcal{E}(X)$ coincides with the Borel field generated by τ_{cw}. In view of the preceding proposition, it suffices to verify that $\{\varnothing\}$ is τ_{cw}-closed, and this follows immediately from the fact that the complement of the empty set can be written as the union of open sets in (cl-sets(X), τ_{cw}), viz.

$$\text{cl-sets}_{\neq\varnothing}(X) = \bigcup_{y\in X}\bigcup_{\alpha\geq 0}\{C \in \text{cl-sets}(X) \mid d_C(y) < \alpha\}.$$

Hence, $\{\varnothing\}$ is τ_{cw}-closed. □

So far, we have established the following string of inclusions and equalities,

$$\mathcal{E}(X) \subset \mathcal{B}_f \subset \mathcal{B}_{cw} = \mathcal{E}(X),$$

which tells us that the Borel fields for both the Fell topology and the Choquet-Wijsman topology coincide with the Effrös field.

To be able to apply Theorem 6.1, there only remains to establish that (cl-sets(X), τ_{cw}) is a Polish space. In fact, this is settled by the following result of Beer:

THEOREM 6.5: BEER 1991. *Suppose (X, d) is a Polish space, and* (cl-sets(X), τ_{cw}) *is the hyperspace of closed subsets of X equipped with the Choquet-Wijsman topology. Then, the (hyper)space* (cl-sets(X), τ_{cw}) *is Polish.*

From this theorem, $\mathscr{E}(X) = \mathscr{B}_f$, Theorem 6.1, and the fact that random lsc functions can be identified with random closed sets (whose values are closed epigraphs), one obtains:

THEOREM 6.6: KOLMOGOROV'S EXTENSION THEOREM. *Let* $\{f^\nu : (\Xi, \mathscr{S}, P) \to$ lsc-fcns$(X), \nu \in \mathbb{N}\}$, *where* (X, d) *is a Polish space, and let* $\mathscr{E}$ *be the Effrös field on* lsc-fcns(X). *For all* ν, *let* P^ν *be the distribution induced by* f^ν *on* (lsc-fcns$(X), \mathscr{E}$). *Then, there exists a unique measure* P^∞ *on* (lsc-fcns$(X)^\infty, \mathscr{E}^\infty$) *that is consistent with the family of measures* $(P^1, P^2, \dots)$, *i.e., whose finite dimensional projection yield the finite dimensional distributions.*

7. Epi-convergence. We are going to need a number of properties of epi-convergent sequences of functions. Recall that for functions $\{g, g^\nu : X \to \overline{\mathbb{R}},\ \nu \in \mathbb{N}\}$ epi-convergence, written $g^\nu \to^{\mathrm{e}} g$ means

- (i) $\forall x^\nu \to x,\quad \liminf_{\nu\to\infty} g^\nu(x^\nu) \geq g(x)$;
- (ii) $\exists \bar{x}^\nu \to x$, such that $\limsup_{\nu\to\infty} g^\nu(\bar{x}^\nu) \leq g(x)$,

for all $x \in X$, cf. Definition 1.2. When conditions (i) and (ii) are satisfied at some x, it is convenient to say that the functions g^ν epi-converge to g *at* x. Epi-convergence at a point x can also be characterized in terms of upper and lower epi-limits. Let $\mathscr{N}(x)$ denote the neighborhood system of x.

DEFINITION 7.1. The lower and upper epi-limits of a sequence of functions $\{g, g^\nu : X \to \overline{\mathbb{R}},\ \nu \in \mathbb{N}\}$ are defined by,

$$\text{e-}\liminf g^\nu(x) := \sup_{V\in\mathscr{N}(x)} \liminf_{\nu\to\infty} \inf_{y\in V} g^\nu(y),$$

$$\text{e-}\limsup g^\nu(x) := \sup_{V\in\mathscr{N}(x)} \limsup_{\nu\to\infty} \inf_{y\in V} g^\nu(y).$$

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

It follows from Definition 7.1 that the upper and lower epi-limits are always lsc. Epi-convergence of g^ν to g corresponds to the set convergence of $\operatorname{epi} g^\nu$ to $\operatorname{epi} g$ in the Fell topology (Beer 1993). It is neither implied by, nor does it imply pointwise convergence, but instead can be viewed as a one-sided uniform convergence. It is exactly what is needed to ensure the convergence of minimizers of g^ν to the minimizers of g, in the following sense.

THEOREM 7.2: ROCKAFELLAR AND WETS 1998, THEOREM 7.31. *Suppose* $\{g^\nu\}_{\nu\in\mathbb{N}}$ *is a sequence of extended real-valued lsc functions such that* $g^\nu \to^{\mathrm{e}} g$. *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* $K \subset X$, *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 $\varepsilon\text{-}\operatorname{argmin} g := \{x \in X \mid g(x) \leq \inf g + \varepsilon < \infty\}$.

Here we state only sufficient conditions for the convergence of the ε-argmin. For necessary conditions, as well as conditions for the convergence of infima, consult Rockafellar and Wets (1998, Chapter 7), and for certain extensions of these results when dealing with random lsc functions, see also Artstein and Wets (1995).

In order to obtain our almost sure epi-convergence result for random lsc functions, the following result based on the separability of X is essential. It tell us that epi-convergence needs only to be verified at the points of a countable dense set.

LEMMA 7.3: ATTOUCH AND WETS 1990. *Let* $f, g : X \to \overline{\mathbb{R}}$ *with* f *lsc. Let* $R \subset X$ *be the projection on* X *of a countable dense subset of* $\operatorname{epi} g$. *If* $f \leq g$ *on* R, *then* $f \leq g$ *on all of* X.

PROOF. The set R above exists since $X \times \mathbb{R}$ is separable. Suppose $f \le g$ on R. This is equivalent to $\{(x, \alpha) \mid \alpha \ge g(x),\ x \in R\} \subset \operatorname{epi} f$. 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 X. $\square$

8. Ergodic theorems. The proofs of the ergodic theorems rely on the following lemma, which is of independent interest. It tells us 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 vector-valued random variables obtained through scalarization. Local inf-integrability (Definition 4.2), already used in the construction of an lsc-version of the conditional expectation of a random lsc function, will be needed throughout; but this is not a significant restriction.

LEMMA 8.1. *Let (X, d) be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and $\mathcal{R} \subset \mathcal{S}$, a σ-field. Let $\{f, f^\nu : \Xi \to \text{lsc-fcns}(X) \mid \nu \in \mathbb{N}\}$ be a sequence of random lsc functions with f locally inf-integrable. For $x \in X$, $\rho \in \mathbb{R}_+$, let $\{\pi_{x,\rho}, \pi^\nu_{x,\rho}, \nu \in \mathbb{N}\}$ be a sequence of scalarizations of these random lsc functions obtained as follows*:

$$\pi_{x,\rho}(\xi) := \inf_{\mathbb{B}^\circ(x,\rho)} f(\xi, \cdot) \quad \pi^\nu_{x,\rho}(\xi) := \inf_{\mathbb{B}^\circ(x,\rho)} f^\nu(\xi, \cdot),$$
$$\pi_{x,0}(\xi) := f(\xi, x), \quad \pi^\nu_{x,0}(\xi) := f^\nu(\xi, x).$$

Suppose that for all $x \in X$, there exists $\kappa_x > 0$, such that for all $\rho \in [0, \kappa_x)$,

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

where $\pi^{\mathcal{R}}$ denotes a version of the conditional expectation of π with respect to $\mathcal{R}$.

If there exists a countable subset $R^+ \subset X \times \mathbb{R}$ that is dense in $\operatorname{epi} E^{\mathcal{R}} f(\xi)$ *P-a.s., then*

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

In particular, if the σ-field $\mathcal{R}$ is independent of the σ-field generated by f, then

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

PROOF. Let $E^{\mathcal{R}} f$ be a version of the conditional expectation of f whose values are in lsc-fcns(X) as guaranteed by Theorem 4.3 and such that for all $\xi \in \Xi_1$, a set of measure 1, and $x \in X$: $E^{\mathcal{R}} f(\xi)(x) = \lim_\nu \pi^{\mathcal{R}}_{x^\nu,\rho^\nu}(\xi)$ for $x^\nu \to x$ and $\rho^\nu \searrow 0$. Let R be a countable dense subset of X that contains the projection onto X of a countable dense set $R^+ \subset X \times \mathbb{R}$ that is dense in $\operatorname{epi} E^{\mathcal{R}} f(\xi)$ for all $\xi \in \Xi_2$, also a subset of Ξ of measure 1. Fix $x \in R$, $\rho \in [0, \kappa_x)$ and let $\Xi_{x,\rho} \subset \Xi$ be such that $P(\Xi_{x,\rho}) = 1$ and

$$\frac{1}{\nu}\sum_{k=1}^{\nu} \pi^k_{x,\rho}(\xi) \to \pi^{\mathcal{R}}_{x,\rho}(\xi) \quad \forall\, \xi \in \Xi_{x,\rho}.$$

Finally, let

$$\Xi_0 := \Xi_1 \cap \Xi_2 \cap \left[\bigcap_{x \in R} \bigcap_{\rho \in Q_+} \Xi_{x,\rho}\right].$$

Then $P(\Xi_0) = 1$, since Ξ_0 is the countable intersection of sets of measure 1.

Let us begin by showing that

$$\forall \xi \in \Xi_0 : \text{ e-}\liminf_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, \cdot) \geq E^{\mathcal{R}} f(\xi).$$

For $\xi \in \Xi_0$, $x \in X$, one has

$$\begin{aligned} \text{e-}\liminf_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x) &= \sup_{\rho>0} \liminf_{\nu\to\infty} \inf_{x' \in \mathbb{B}(x,\rho)} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x') \\ &\geq \sup_{\rho>0} \liminf_{\nu\to\infty} \frac{1}{\nu} \sum_{k=1}^{\nu} \inf_{x' \in \mathbb{B}(x,\rho)} f^k(\xi, x') \\ &\geq \sup_{l\in N} \liminf_{\nu\to\infty} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^k_{x^l,\rho^l}(\xi), \end{aligned}$$

where $x^l \to x$, $\rho^l \searrow 0$, and

$$\forall\, l \in \mathbb{N}: \; x^l \in R,\; \rho^l \in [0, \kappa_x) \cap \mathbb{Q},\; x \in \mathbb{B}(x^l, \rho^l),\; \mathbb{B}(x^{l+1}, \rho^{l+1}) \subset \mathbb{B}(x^l, \rho^l).$$

For every l, $\xi \in \Xi_0$, one has

$$\liminf_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^k_{x^l,\rho^l}(\xi) = \pi^{\mathcal{R}}_{x^l,\rho^l}(\xi)$$

by assumption. Now, observing that when $x^l \to x$ and $\rho^l \searrow 0$, $\pi^{\mathcal{R}}_{x^l,\rho^l}(\xi) \nearrow E^{\mathcal{R}} f(\xi, x)$ for $\xi \in \Xi_0 \subset \Xi_1$, taking the supremum over $l \in \mathbb{N}$ yields

$$\text{e-}\liminf_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x) \geq \sup_{l\in\mathbb{N}} \pi^{\mathcal{R}}_{x^l,\rho^l}(\xi) = E^{\mathcal{R}} f(\xi)(x).$$

Hence, for all $\xi \in \Xi_0$, e-$\liminf_\nu \frac{1}{\nu} \sum_{k=1}^{\nu} f^\nu(\xi, \cdot) \geq E^{\mathcal{R}} f(\xi)$ (on all of X).

Next, let us turn to the inequality involving the upper epi-limit. For $\xi \in \Xi_0$ and $x \in R$, if $x^\nu \equiv x$, then by assumption,

$$\limsup_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x^\nu) = \limsup_{\nu\to\infty} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^k_{x,0}(\xi) = E\pi^{\mathcal{R}}_{x,0}(\xi) = E^{\mathcal{R}} f(\xi)(x).$$

In view of Definition 7.1, this is the same as for all $x \in R$,

$$\left(\text{e-}\limsup \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, \cdot)\right)(x) \leq E^{\mathcal{R}} f(\xi)(x).$$

Using the facts that $E^{\mathcal{R}} f$ is an lsc version of the conditional expectation of f with respect to $\mathcal{R}$, and that R is a countable dense subset of X containing the projection onto X of a countable dense subset of epi $E^{\mathcal{R}} f$, one appeals to Lemma 7.3 to obtain

$$\forall \xi \in \Xi_0 : \text{ e-}\limsup_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, \cdot) \leq E^{\mathcal{R}} f(\xi) \quad P\text{-a.s. on } X.$$

Hence,

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

When the σ-field generated by f is independent of $\mathcal{R}$, $E^{\mathcal{R}}f \equiv Ef$. In that case, since X is separable and Ef is lsc, one can always find a countable dense subset R^+ of $\operatorname{epi} Ef$. Consequently,

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

as claimed. □

The condition, "there exists a countable subset $R^+ \subset X \times \mathbb{R}$ that is dense in $\operatorname{epi} E^{\mathcal{R}}f(\xi)$ P-a.s.," is certainly satisfied when $\mathcal{R}$ can be generated by countably many atoms (and sets of measure zero). Indeed, then R^+ can be chosen to be the union over all such atoms $\alpha \in \mathcal{R}$ of X_α, where X_α is a countable dense subset of X containing the projection onto X of a countable dense subset of $\operatorname{epi} E^{\mathcal{R}}f(\alpha)$, and $E^{\mathcal{R}}f(\alpha)(x) := \int_\alpha f(\xi,x)P(d\xi)/P(\alpha)$. This situation occurs when there exist a countable number of sets $A \in \mathcal{S}$ that partition Ξ, and such that $\{\varphi(\xi) \mid \xi \in A\} = A$ and $\varphi : A \to A$ is ergodic, i.e., all $\mathcal{S}$-measurable sets $B \subset A$ such that $\varphi^{-1}(B) = B$ satisfy $P(B) = P(A)$ or $P(B) = 0$. For a stationary sequence of random lsc functions, this corresponds to the ability to partition the sequence into countably many ergodic subsequences.

There are also many other important cases which satisfy the assumption of the existence of a countable dense subset R^+ of $\operatorname{epi} E^{\mathcal{R}}f(\xi)$ P-a.s.. For example, when $\operatorname{epi} E^{\mathcal{R}}f(\xi)$ is P-almost surely a *solid set*, i.e.,

$$\operatorname{cl}(\operatorname{int}(\operatorname{epi} E^{\mathcal{R}}f(\xi))) = \operatorname{epi} E^{\mathcal{R}}f(\xi) \quad P\text{-a.s.},$$

then any countable dense subset of $X \times \mathbb{R}$ could fill the role of R^+. This situation arises commonly in applications in which for each ξ, the function $x \mapsto f(\xi,x)$ is continuous on its domain (where it is finite), or more broadly when $\operatorname{epi} f(\xi,\cdot)$ is itself a solid set.

We are now all set to state and prove our ergodic theorem. The classical version of Birkhoff-Khintchine Ergodic Theorem can be found in Loève (1978), Durrett (1991, §6.2), and de Fitte (1997), for example. For our purposes, we need a version that allows for functions that are extended real-valued. A straightforward modification of Loève (1978, Theorem 33.B, Theorem 34.A) takes care of this situation.

THEOREM 8.2. *Let (X,d) be a Polish space, $(\Xi,\mathcal{S},P)$ a probability space, $\varphi : \Xi \to \Xi$ a measure preserving transformation and $\mathcal{J}$ its invariant σ-field. Let f be a locally inf-integrable random lsc function. If there exists a countable subset of $X \times \mathbb{R}$ that is dense in* $\operatorname{epi} E^{\mathcal{J}}f(\xi)$ P-a.s., *then*

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

In particular, if φ is ergodic, then

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

PROOF. For $x \in X$ and $\rho \in \mathbb{Q}_+$, let $\pi_{x,\rho} := \inf_{\mathbb{B}^o(x,\rho)} f$ be the random variables obtained through scalarization, with $\pi_{x,0} := f(\cdot,x)$. Given $x \in X$, $\rho \in \mathbb{Q}_+$, since φ is measure preserving, $\rho \in \mathbb{Q}_+$, the sequence $\{\pi_{x,\rho} \circ \varphi^\nu,\ \nu \in \mathbb{N}\}$ is stationary. By the Birkhoff-Khintchine Ergodic Theorem for all $x \in X$, $\rho \in \mathbb{Q}_+$, one obtains

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

where $\pi_{x,\rho}^{\mathcal{J}}$ is a version of the conditional expectation of $\pi_{x,\rho}$ with respect to the invariant field $\mathcal{J}$. We are now in the setting of Lemma 8.1, which immediately yields

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

If φ is ergodic, $\mathcal{J}$ is trivial, whereby $\mathcal{J}$ is independent of the σ-field generated by f, so that again by Lemma 8.1,

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

as claimed. □

Acknowledgments. Research supported in part by grants of the National Science Foundation.

References

Artstein, Z., R. J.-B. Wets. 1995. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2** 1–17.

Attouch, H. 1984. *Variational Convergence for Functions and Operators*. Applicable Mathematics Series. Pitman.

———, R. J.-B. Wets. 1990. Epigraphical processes: laws of large numbers for random lsc functions. *Sém. Anal. Convexe, Montepellier* 13.1–13.29.

Attouch, H., C. Licht, R. J.-B. Wets. 1999. Stochastic homogenization and ergodic theory. Manuscript, University of California, Davis.

Aubin, J.-P., H. Frankowska. 1990. *Set-Valued Analysis*. Birkhäuser Boston Inc., Boston, MA.

Barlow, R., D. Bartholomew, J. Bremmer, H. Brunk. 1972. *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley Series in Probability and Statistics. John Wiley & Sons, London.

G. Beer. 1991a. A polish topology for the closed subsets of a polish space. *Proc. Amer. Math. Soc.* **113** 1123–1133.

———. 1991b. Topologies on closed and closed convex sets and the Effrös measurability of set valued functions. *Sém. Anal. Convexe, Montepellier* 21–2.

———. 1993. *Topologies on Closed and Closed Convex Sets*. Kluwer Academic Publishers, Dordrecht.

Birge, J., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.

Bloomfield, P., W. L. Steiger. 1983. *Least Absolute Deviations: Theory, Applications and Algorithms*. Birkhäuser Boston Inc., Boston, MA.

Box, G. E. P., G. M. Jenkins. 1970. *Time Series Analysis Forecasting and Control*. Holden-Day.

Castaing, C., F. Ezzaki. 1995. SLLN for convex random sets and random lower semicontinuous integrands. Manuscript, Université de Montpellier.

———, M. Valadier. 1977. *Convex Analysis and Measurable Multifunctions*. Lecture Notes in Mathematics 580. Springer, Berlin.

Choquet, G. 1966. Outils topologiques et métriques de l'analyse mathematique. Lecture Notes, Centre de documentation universitaire et SEDES réunis, Paris.

de Fitte, P. R. 1997. Théorème ergodique ponctuel et lois fortes des grands nombres pour des points aléatoires d'un espace métrique á courbure négative. *Ann. Probab.* **25** 738–766.

Durrett, R. 1991. *Probability: Theory and Examples*. Wadsworth & Brooks/Cole.

Effrös, E. 1965. Convergence of closed subsets in a topological space. *Proceedings of the American Mathematical Society* **16** 929–931.

Fell, J. 1962. A Hausdorff topology for the closed subsets of a locally compact Hausdorff space. *Proceedings of the American Mathematical Society* **13** 472–476.

Francaviglia, S., A. Lechicki, S. Levi. 1985. Quasi-uniformization of hyperspaces and convergence of nets of semicontinuous multifunctions. *J. Math. Anal. Appl.* **112** 347–370.

Hess, C. 1986. *Contributions á l'étude de la mesurabilité, de la loi de probabilité, et de la convergence de multifonctions*. Thèse d'état, Université de Montpellier II.

———. 1996. Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. *Ann. Statist.* **24**(3) 1298–1315.

Himmelberg, C. J. 1975. Measurable relations. *Fundamenta Mathematicae* **87** 52–72.

Kall, P., S. Wallace. 1994. *Stochastic Programming*. Wiley, Chichester.

Kaňková, V. 1978. Optimum solution of a stochastic optimization problem with unknown parameters. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, and of the 1974 European Meeting of Statisticians, B* 239–244.

King, A., R. T. Rockafellar, L. Somlyódy, R. J.-B. Wets. 1988. Lake eutrophication management: the lake Balaton project. Y. Ermoliev, R. Wets, eds. *Numerical Techniques for Stochastic Optimization*, Springer, Berlin, 435–444.

Korf, L. A., R. J.-B. Wets. 2000a. Random lsc functions: An ergodic theorem. `http:// dochost.rz.hu-berlin.de/speps`.

———, ———. 2000b. Random lsc functions: Scalarization. `http://dochost.rz.hu-berlin.de/speps`, 2000.

Kozlov, S., O. Oleinik, V. Zhikov. 1994. *Homogenization of Differential Operators and Integral Functionals.* Springer, Berlin.

Loève, M. 1978. *Probability Theory II*, 4th ed. Springer, New York.

Papanicolaou, G. C., S. R. Varadhan. 1981. Boundary value problems with rapidly oscillating random coefficients. J. Fritz, J. L. Lebowitz, D. Szasz, eds. *Random Fields*, volume 27 of *Colloquia Mathematica Scoietatis Janos Bolyai*, North Holland, 835–873.

Parthasarathy, K. R. 1967. *Probability Measures on Metric Spaces.* Academic Press, New York.

Robertson, T., F. Wright, R. Dykstra. 1988. *Order-Restricted Statistical Inference.* Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.

Rockafellar, R. T. 1976. Integral functionals, normal integrands and measurable selections. J. Gossez, L. Waelbroeck, eds. *Nonlinear Operators and the Calculus of Variations*, Lecture Notes in Mathematics No. 543, Springer, 157–207.

———, R. J.-B. Wets. 1984. Variational systems, an introduction. G. Salinetti, ed. *Multifunctions and Integrands: Stochastic Analysis, Approximation and Optimization*, volume 1091 of *Lecture Notes in Mathematics*, Springer Verlag, Berlin, 1–54.

———, ———. 1998. *Variational Analysis*. Springer.

Salinetti, G., R. J.-B. Wets. 1986. On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Oper. Res.* **11** 385–419.

Salinger, D. 1997. *A splitting algorithm for multistage stochastic programming with application to hydropower scheduling*. Ph.d. thesis, University of Washington, Seattle.

Tiao, G. C. 1985. Autoregressive moving average models, intervention problems and outlier detection in times series. E. J. Hannan, P. R. Krishnaiah, M. M. Rao, eds. *Times Series in the Time Domain*, volume 5 of *Handbook of Statistics*, Elsevier Science Publishers, New York, 85–118.

Vervaat, W. 1988. Random upper semicontinuous functions and extremal processes. Report MS-R8801, Center for Wiskunde en Informatica, Amsterdam, The Netherlands.

Vogel, S. 1995. On stability in stochastic programming–sufficient conditions for continuous convergence and epi-convergence. TU Ilmenau, Inst. Math. Forthcoming.

Wets, R. J.-B. 1989. Stochastic programming. A. Rinnooy Kan, G. Nemhauser, M. Todd, eds. *Optimization, Handbook for Operations Research and Management Sciences*, Volume 1, North Holland, New York, 573–629.

Wijsman, R. 1966. Convergence of sequences of convex sets, cones, and functions, ii. *Trans. Amer. Math. Soc.*, **123** 32–45.

L. A. Korf: Department of Mathematics, University of Washington, Seattle, WA 98195; e-mail: korf@math.washington.edu

R. J.-B. Wets: Department of Mathematics, University of California, Davis, CA 95616. e-mail: rjbwets@ucdavis.edu