RANDOM LSC FUNCTIONS: AN ERGODIC THEOREM *

Lisa A. Korf
Department of Mathematics
University of Washington, Seattle
korf@math.washington.edu

Roger J-B Wets
Department of Mathematics
University of California, Davis
rjbwets@ucdavis.edu

Abstract. An ergodic theorem for random lsc functions is obtained by relying on a (novel) ‘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.

Keywords: 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.

AMS Classification: 60H07, 90C15, 60B11, 49J99

Date: January 22, 2000

*Research supported in part by grants of the National Science Foundation
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. And to do this, as explained in §2, one can appeal to the ergodicity properties of random *lsc* (lower semicontinuous) functions set forth in this paper.

A streamlined version of this basic ergodic theorem, see §6, 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’s assume that \(\mathcal{S}\) is \(P\)-complete. A random *lsc* (lower semicontinuous) function is then an extended real-valued function \(f: \Xi \times X \to \overline{\mathcal{M}}\) such that

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

**Theorem 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,

\[
\frac{1}{\nu} \sum_{k=1}^{\nu} f(\varphi^k(\xi),\cdot) \Rightarrow E f, \quad \text{a.s.}
\]

**Definition 1.2.** A sequence of functions \(\{g^\nu : X \to \overline{\mathcal{M}}, \nu \in \mathbb{N}\}\) epi-converges to \(g : X \to \overline{\mathcal{M}}\), written \(g^\nu \Rightarrow g\), if for all \(x \in X\),

- (i) \(\lim \inf_{\nu} g^\nu(x^\nu) \geq g(x)\) for all \(x^\nu \to x\);
- (ii) \(\lim \sup_{\nu} g^\nu(x^\nu) \leq g(x)\) for some \(x^\nu \to x\).

Epi-convergence entails the convergence of the minimizers of the \(g^\nu\) to those of \(g\) as made precise in §5. It is so named because it agrees with the set convergence of the epigraphs, cf. [2, 5, 29, 28]; the epigraph of a functions \(g : X \to \overline{\mathcal{M}}\) 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 6.2.

The immediate precursor to these results is a law of large numbers for random lsc functions [4, 1], cf. also [19, 11] for extensions, that posits iid (independent identically distributed) sampling whereas now only stationarity will be assumed. The argument
here relies on a ‘scalarization’ of random Ic functions developed in [22] and summarized in §3 and §4, which renders the argument more transparent and more effective.

2. Examples

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

Example 2.1 Time series.

Classical consistency results for certain estimation procedures of the parameters of a time series can be guaranteed by means of this ergodic theorem for random Ic functions. But one can equally well deal with ‘nonclassical’ situations as would be the case if we wanted to include in the estimation process some information one might have about certain relationships between the parameters. This is a subject that warrants a separate development. Here, we illustrate the approach by considering the following autoregressive (AR) model: for \( t = \ldots, 0, 1, \ldots \), let

\[
Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + \varepsilon_t
\]

where \( a_0, a_1, \ldots, a_p \) are the coefficients of the (linear) transfer function and the \( \varepsilon_t \) are independent normally distributed random variables with mean 0 and variance \( \sigma^2 \) and account for the disturbances in the dynamics not captured by the transfer function. Let’s assume that the coefficients of the autoregressive polynomial \( 1 - a_1 l - \cdots - a_p l^p \) are such that its roots lie outside the unit circle. This AR-model has a solution \( \{ Y_t, t = \ldots, -1, 0, 1, \ldots \} \) that is stationary, cf. for example [31].

If the coefficients \( a_0, a_1, \ldots, a_p \) are unknown but some observations, say \( \eta_{1-p}, \ldots, \eta_{p} \), of the process \( Y_t \) are available, one could estimate these coefficients by solving the following optimization problem:

\[
\min_{x_0, \ldots, x_p} \frac{1}{\nu} \sum_{t=1}^{\nu} |\eta_t - x_0 - \eta_{t-1} x_1 - \cdots - \eta_{t-p} x_p|^2,
\]

the normalizing factor \( 1/\nu \) doesn’t affect the solution but scales the optimal value. The objective pursued is to find estimates for \( a_0, \ldots, a_p \) that minimize the role played by ‘disturbances’ (innovations), i.e., those factors not included in the (linear) transfer function; note that \( \nu^{-1} \sum_{t=1}^{\nu} (Y_t - a_0 - Y_{t-1} a_1 - \cdots - Y_{t-p} a_p)^2 = \nu^{-1} \sum_{t=1}^{\nu} (\varepsilon_t)^2 \). The optimal solution of this problem,

\[
x_0^\nu, x_1^\nu, \ldots, x_p^\nu,
\]
would provide the estimates for the unknown coefficients \( a_0, a_1, \ldots, a_p \). Note that the optimal solution is unique due to the strict convexity of the function being minimized. Consistency, as usual, would imply that almost surely these estimates converge to the ‘true’ values of the unknown coefficients.

To embed this in the general framework of §1, let’s define the vector process:

\[
\{ X_t = (Y_{t-p}, Y_{t-p+1}, \ldots, Y_t), \quad t = \ldots, -1, 0, 1, \ldots \}.
\]

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

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

given the observations \( \xi_1, \ldots, \xi_\nu \), the optimization problem yielding the estimates takes the form:

\[
\min \frac{1}{\nu} \sum_{i=1}^{\nu} f(\xi_i, x).
\]

With \( (\Xi, \mathcal{S}, \mathbf{P}) \) the (common) probability space on which the random variables \( X_t \) are defined, it’s easy to verify that \( f : \Xi \times \mathbb{R}^{p+1} \to \mathbb{R} \) is a random l.s.c function. By appealing to the Ergodic Theorem 6.2, the consistency of the proposed estimates \( (x_0^\nu, x_1^\nu, \ldots, x_p^\nu) \) can be settled by showing that almost surely,

\[
(a_0, a_1, \ldots, a_p) \in \arg \min E^T f(x) := E \{ f(\xi, x) | \mathcal{I} \},
\]

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

\[
(x_0^\nu, x_1^\nu, \ldots, x_p^\nu) \to (a_0, a_1, \ldots, a_p) \quad \text{a.s.}
\]

We carry out the calculation in the (simple) case \( p = 1 \). Then

\[
Y_t = a_0 + a_1Y_{t-1} + \varepsilon_t, \quad t = \ldots, -1, 0, 1, \ldots,
\]

where, as before, the random variables \( \varepsilon_t \) are normally distributed with mean 0 and variance \( \sigma^2 \), and

\[
f(\xi, x) = |\xi_t - x_0 - \xi_0x_1|^2,
\]

with \( \xi = (\xi_0, \xi_1) \). The random vector \( \xi = (\xi_0, \xi_1) \) has the same distribution as \( (Y_{t-1}, Y_t) \). Assuming that \( a_1 \in (0, 1) \), the process \( \{ Y_t \}_{t \in \mathbb{Z}} \) is not just stationary but actually ergodic; one can check that the covariance of \( (Y_t, Y_{t+k}) \) goes to 0 as \( k \to \infty \).
To compute the distribution of \((Y_{t-1}, Y_t)\), let’s begin by observing that by recursion, for \(J = 1, 2, \ldots\),
\[
Y_t = a_0 \sum_{j=0}^{J} (a_1)^j + \sum_{j=0}^{J} (a_1)^j \varepsilon_{t-j} + (a_1)^{J+1} Y_{t-1}.
\]
Normality and independence of the \(\varepsilon_i\) imply that \(\sum_{j=0}^{J}(a_1)^j \varepsilon_{t-j}\) is normally distributed with mean 0 and variance \(\sum_{j=0}^{J}(a_1)^{2j}\sigma^2\). As \(J \to \infty\),
\[
\begin{cases}
(a_1)^{J+1} \to 0, \\
\sum_{j=0}^{J}(a_1)^j \to (1 - a_1)^{-1},
\end{cases}
\]
and \(\sum_{j=0}^{J}(a_1)^j \varepsilon_{t-j}\) converges to a random variable, normally distributed with mean 0 and variance \((1 - a_1^2)^{-1}\sigma^2\). From this, and
\[
(Y_{t-1}, Y_t) = (0, a_0) + (Y_{t-1}, \varepsilon_t) \begin{pmatrix} 1 & a_1 \\ 0 & 1 \end{pmatrix},
\]
it follows immediately that
\[
(Y_{t-1}, Y_t) \text{ is normal with mean } \frac{a_0}{1 - a_1}(1, 1) \text{ and covariance } \frac{\sigma^2}{1 - a_1^2} \begin{pmatrix} 1 & a_1 \\ a_1 & 1 \end{pmatrix}.
\]
This enables us to compute \(E^2 f = Ef\) (by ergodicity):
\[
Ef(X) = E \{ \xi_1 - x_0 - \xi_0 x_1 |^2 \} = E \{ (\xi_1 - \alpha) - x_0 + \alpha(1 + x_1) - x_1 (\xi_0 - \alpha) |^2 \}.
\]
Setting \(\alpha = E\{\xi_0\} = E\{\xi_1\} = a_0(1 - a_1)^{-1}\) and with \(\beta = \sigma^2(1 - a_1^2)^{-1}\), one obtains the quadratic form:
\[
Ef(X) = \gamma - 2\langle p, x \rangle + \langle Q x, x \rangle
\]
where
\[
\gamma = \alpha^2 + \beta, \quad p = (\alpha, \alpha^2 + a_1 \beta), \quad Q = \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 + \beta \end{pmatrix}.
\]
Because \(Q\) is positive definite, \(Ef\) is strictly convex and one verifies easily that
\[
x^* = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = Q^{-1} p
\]
is the unique optimal solution and \(Ef(x^*) = \sigma^2\). Almost sure epi-convergence of the optimization problems generating the estimates to the limit problem, \(\min Ef\), guarantees the almost sure convergence of the estimates \((x^*_0, x^*_1)\) to \((a_0, a_1)\). Moreover, the convergence of the infima also tells us that
\[
\inf \frac{1}{\nu} \sum_{t=1}^{\nu} f(\xi_t, \cdot) \to \inf Ef = \sigma^2, \ a.s.
\]
Thus, the infima provide consistent statistical estimates for the variance of the disturbances \( \varepsilon_i \) when this variance also needs to be estimated.

Of course, the consistency of these estimates can also be derived by other means, for example refer to the arguments used in [10] to prove the consistency of the estimates obtained by minimizing the ‘conditional likelihood function’. However, whereas such proofs usually apply to a particular criterion, the approach here is quite general. For example, except for the calculation of the function \(Ef\), nothing needs to be changed in the arguments if \(f\) is replaced by

\[
g(\xi, x) = |\xi - x_0 - \xi_0x_1|,
\]

i.e., when the \(\ell^1\)-norm, rather than the \(\ell^2\)-norm, is used to measure the ‘errors’.

More significantly, certain restrictions can be included in the optimization problems used to derive the estimates. This is particularly useful, and potentially important, if the number of samples available is small. Returning to the more general case when \(p \geq 1\), on the basis of certain physical laws or economic considerations, one might know that

\[
a_1 \geq a_2 \geq \cdots \geq a_p,
\]

i.e., the more distant the past, the smaller the contribution to \(Y\). In such a situation, that would be quite common, the function \(f\) would be modified as follows:

\[
\tilde{f}(\xi, x) = \left\{ \begin{array}{ll}
|\eta_0 - x_0 - \eta_1x_1 - \cdots - \eta_p x_p| & \text{if } x_1 \geq \cdots \geq x_p, \\
0 & \text{otherwise}.
\end{array} \right.
\]

Exactly the same arguments yield the consistency of the estimates

\[
(x^\nu_0, x^\nu_1, \ldots, x^\nu_p) \in \arg\min \sum_{l=1}^{\nu} \tilde{f}(\xi, x).
\]

The advantage lies in having estimates that satisfy a relationship that is known to hold for the coefficients \((a_1, \ldots, a_p)\) even when only a small number of samples is available.

**Example 2.2 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

5
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’s consider the example of a conductor occupying a region \( \Omega \) 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 \( a(\xi, x) \), where \( a(\xi, x) \) 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

\[
-\nabla \cdot (a(\xi, x) \nabla u(\xi, x)) = h(x) \text{ for } x \in \Omega, \\
u(\xi, x) = 0 \text{ for } x \in \partial \Omega.
\]

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

\[
-\nabla \cdot (a(x) \nabla u(x)) = h(x) \text{ for } x \in \Omega, \\
u(x) = 0 \text{ for } x \in \partial \Omega.
\]

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 \( u(x) = E_u\{\xi_x\} \) for \( x \in \Omega \). Note that contrary to first intuition, setting \( a(x) = E\{a(\xi, x)\} \) 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, \( \varepsilon \), and employing methods such as asymptotic analysis to find the limiting problem, see e.g., Papanicolaou and Varadhan [25], and Kozlov, Oleinik and Zhikov [23]. 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, Licht and Wets [3].

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

\[
\min_{u \in H_0^1(\Omega)} g(\xi, u) := \frac{1}{2} \int_{\Omega} a(\xi, x) |\nabla u|^2 dx - \langle h, u \rangle.
\]
The goal is to find the homogenized functional, $g_{\text{hom}}$, such that

$$E\{u(\xi, \cdot)\} = \bar{u}(\cdot) \in \text{argmin} \left\{ g_{\text{hom}}(u) \mid u \in H^1_0(\Omega) \right\}. $$

Given the samples $a(\xi^1, \cdot), \ldots, a(\xi^\nu, \cdot)$, it’s 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:

$$\min \left\{ \frac{1}{\nu} \sum_{k=1}^\nu g(\xi^k, u) \mid u \in H^1_0(\Omega) \right\}. $$

It can’t serve as an approximation for the homogenized problem; it would correspond to averaging $a(\xi, x)$ 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 $g_{\text{hom}}$, we rely on some facts from conjugate duality. To approximate it, we appeal to our ergodic theorem for random lsc functions.

Let $(X, \tau)$ be a Banach space and $X^\ast$ its topological dual; here $X = H^1_0$ with $\tau$ the norm topology. The *conjugate* function $q^\ast : X^\ast \to \overline{\mathbb{R}}$ of a function $q : X \to \overline{\mathbb{R}}$ is

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

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

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

$$(0 \cdot q)(x) := \begin{cases} 0 & \text{if } x = 0, q \neq \infty \\ \infty & \text{otherwise}. \end{cases}$$

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

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

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 $e\cdot f q(\xi, \cdot) \cdot P(d\xi)$ defined by

$$\left( e\cdot f q(\xi, \cdot) \cdot P(d\xi) \right)(x) := \inf_{z(\cdot)} \left\{ \int_\Xi q(\xi, z(\xi)) P(d\xi) \mid \int_\Xi z(\xi) P(d\xi) = x \right\}$$

where it’s understood that $z$ is an integrable function from $\Xi$ to $X$.

It’s straightforward to verify that

$$\bar{u} = E\{u(\xi, \cdot)\} \in \text{argmin} \left\{ g_{\text{hom}}(u) := \left( e\cdot f g(\xi, \cdot) \cdot P(d\xi) \right)(u) \mid u \in H^1_0 \right\}$$

7
and one might reasonably expect that for $\nu$ sufficiently large,

$$\tilde{a}^\nu \in \text{argmin} \left\{ (\nu^{-1} g(\xi^1, \cdot) \ast \cdots \ast g(\xi^\nu, \cdot)) (u) \mid u \in H_0^1 \right\}$$

would approximate $\tilde{a}$.

Let’s now proceed and sketch out a justification which 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 [12], Rockafellar and Wets [29] and Attouch [2]. Recall that an extended real-valued function $q$ is called proper if $-\infty < q \neq \infty$.

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

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

If $q : \Xi \times X \to \overline{M}$ 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

$$\left( \left( \mathbb{E} \int_\Xi q(\xi, \cdot) P(d\xi) \right)^* \right)^* = \int_\Xi q^*(\xi, \cdot) P(d\xi)$$

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, $q^\nu : X \to \overline{M}$ is equicoercive if for all $\{x^\nu, \nu \in \mathbb{N}\}$, $\sup_{x^\nu} q^\nu (x^\nu) < \infty$ implies $\sup_{x^\nu} |x^\nu| < \infty$.

**Theorem 2.2.** Let $q^\nu : X \to \overline{M}$ 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 $q^* = \text{e-lim}_\nu q^{\nu*}$.

$$q^\nu \overset{\text{w-e-lim}}{\rightarrow} q$$

$$* \downarrow \uparrow *$$

$$q^{\nu*} \overset{\text{e-lim}}{\rightarrow} q^*$$
Here, \( w\text{-e-lim}_{\nu} \) refers to epi-convergence with respect to the weak topology on \( X \) whereas \( e\text{-lim}_{\nu} \) 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 \( q^\nu \), one may first pass to the conjugate sequence, \( q^{\nu^*} \), find its epi-limit \( q^* \), then compute the conjugate again to arrive back at \( q \).

Let's now return to the problem at hand. For \( k = 1, \ldots, \nu \), let

\[ u^k \in \text{argmin} \{ g(\xi^k, u) \mid u \in H^1_0 \}. \]

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

\[ \nu^{-1}(u_1 + \cdots + u_\nu) = \tilde{a}^\nu \in \text{argmin}_{H^1_0} \nu^{-1} \cdot \left[ g(\xi^1, u) \cdot \cdots \cdot g(\xi^\nu, u) \right]. \]

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

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

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

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

\[ \frac{1}{\nu} \sum_{k=1}^{\nu} f(\xi^k, \cdot) \rightharpoonup E f, \quad \text{P-a.s.,} \]

where

\[ E f = \int_{\xi} f(\xi, \cdot) \, P(d\xi). \]

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

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

There now remains only to appeal to Theorem 2.2 to obtain

\[ g^{\text{hom}} = w\text{-e-lim}_{\nu} \left( \nu^{-1} \cdot \left[ g(\xi^1, \cdot) \cdot \cdots \cdot g(\xi^\nu, \cdot) \right] \right) \]

\[ = w\text{-e-lim}_{\nu} \left( \nu^{-1} \left( f(\xi^1, \cdot) + \cdots + f(\xi^\nu, \cdot) \right) \right)^*. \]

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 5.2) implies the (weak) convergence of solutions \( \tilde{a}^\nu \) to the homogenized solution \( \tilde{a} \). Similar techniques may also be employed to compute the effective coefficient \( a(\cdot) \).
Example 2.3 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:

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

where

- $f : \Xi \times \mathbb{R}^n \to \mathbb{R} = [-\infty, \infty]$, $f(\xi, x)$ is the 'cost' associated with a decision $x$ when the random variable $\xi$ takes on the value $\xi$;

- $\xi$ is a $\mathbb{R}^N$-valued random variable with possible values in $\Xi \subset \mathbb{R}^N$;

- $Ef : \mathbb{R}^n \to \mathbb{R}$, the function to be minimized,

is defined by

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

The standard two stage stochastic programs with recourse [20, 9] can be recast in this format. Indeed, let’s consider the following version of the two stage model:

$$\min_{x \in C_1} f_{10}(x) + Ef\left\{ \min_{x_2 \in C_2} \{ f_{20}(\xi, x_2) \mid f_{21}(\xi, x_2) \leq 0, i = 1, \ldots, m_2, \} \right\}$$

such that $f_{1i}(x) \leq 0$, $i = 1, \ldots, m_1$,

and define the function $f$ as follows:

$$f(\xi, x) = \left\{ \begin{array}{ll}
    f_{10}(x) + Q(\xi, x) & \text{if } x \in C_1, f_{1i}(x) \leq 0 \text{ for } i = 1, \ldots, m_1, \\
    \infty & \text{otherwise}
\end{array} \right.$$ 

where

$$Q(\xi, x) = \inf_{x_2 \in C_2} \{ f_{20}(\xi, x_2) \mid f_{21}(\xi, x_2) \leq 0, i = 1, \ldots, m_2 \}.$$ 

One can subject multistage models to a similar reduction and make them fit this framework.

Minimizing $Ef$ on $\mathbb{R}^n$ is basically a nonlinear programming problem. Standard solution procedures rely on calculating the gradient, at least approximately, and carrying out line minimization when a direction of descent has been identified, and this requires repeated evaluations of $Ef$. The calculation of gradients, or of subgradients in the nondifferentiable case, as well as function evaluations require $N$-dimensional integration. Except when $N \approx 1$, the $N$-dimensional integration can be a real challenge! Moreover, this integration shouldn’t involve an unreasonable amount of computational effort since
evaluating the function \(f(\cdot, x)\) itself might require extensive computations; refer to the two stage model described above.

One way to proceed is to replace the given problem with a sampled one. Suppose \(\xi^1, \ldots, \xi^\nu\) are samples of \(\xi\). One can view
\[
\min \frac{1}{\nu} \sum_{k=1}^{\nu} f(\xi^k, x) \quad \text{such that} \quad x \in \mathbb{R}^n
\]
as a ‘sample’ of the stochastic optimization problem. Hopefully, the solution \(x^\nu\) of such a problem will provide an acceptable approximation of a solution \(x^* \in \text{argmin} \ E f\). In the case of iid samples, or more generally if the samples are generated from a stationary process,
\[
\{\xi^1, \xi^2, \ldots, \xi^\nu, \ldots\},
\]
one can appeal to our ergodic theorem and the convergence of the minimizers of epi-converging functions (Theorem 5.2) to claim (under certain additional conditions) that
\[
x^\nu(\xi^1, \ldots, \xi^\nu) \in \text{argmin} \frac{1}{\nu} \sum_{k=1}^{\nu} f(\xi^k, x)
\]
converge almost surely to \(x^*\). If the samples are obtained from a time series then they usually aren’t iid but they usually come from an ergodic process. For instance, this is typically the case when dealing with applications where the uncertainty comes from the environment, cf. [21] for an application dealing with lake eutrophication management and [30] for an application involving the control of water reservoirs to generate hydropower.

3. Random lsc functions

The setting in which we work is that introduced in \(\S\)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 don’t assume, as in \(\S\)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.

We begin with the notion of a random closed set. A set-valued mapping \(S : \Xi \to X\) is a random closed set if it is closed-valued (for all \(\xi \in \Xi\), \(S(\xi)\) is closed) and measurable, i.e., for any open set \(O \subset X\),
\[
\{\xi \in \Xi \mid S(\xi) \cap O \neq \emptyset\} =: S^{-1}(O) \in \mathcal{S}.
\]
One can also view \(S\) as a function from \(\Xi\) to \(\text{cl-sets}(X)\), the (hyper)space of closed subsets of \(X\). The Effros field on \(\text{cl-sets}(X)\) is the \(\sigma\)-field \(\mathcal{E}(X)\) generated by all sets of the form
\[
\mathcal{E}_O = \{C \in \text{cl-sets}(X) \mid C \cap O \neq \emptyset\}, \ O \subset X, \ \text{open};
\]
cf. [16, 8]. It’s clear that the closed-valued mapping \( S : \Xi \to X \) is measurable if only if it’s \((\mathcal{S}, \mathcal{E}(X))\)-measurable when viewed as a function from \((\Xi, \mathcal{S})\) to \((\text{cl-sets}(X), \mathcal{E}(X))\).

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

\[
S_f : \Xi \to X \times \mathbb{R} \quad \text{with} \quad S_f(\xi) := \text{epi} f(\xi, \cdot) = \{(x, \alpha) \in X \times \mathbb{R} | \alpha \geq f(\xi, x)\}
\]

is a random closed set. Note that a necessary (but not sufficient) condition for a function \( f : \Xi \to \text{lsc-fcns}(X) \) to be a random lsc function is that for each fixed \( x \in X \), the function \( \xi \mapsto f(\xi, x) \) is measurable [29, Proposition 14.28]. It will sometimes be convenient to identify a 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 [29, Proposition 14.34].

The concept of a random lsc function is due to Rockafellar [27] 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 [29], [32], [22] and here in §4.

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 \textit{Effros field} on \( \text{lsc-fcns}(X) \), denoted simply \( \mathcal{E} \), can be identified with the restrictions of the Effros field on \( \text{cl-sets}(X \times \mathbb{R}) \) to the closed subsets of \( X \times \mathbb{R} \) that are epigraphs. \( \mathcal{E} \) not only includes all sets of the form

\[
\mathcal{E}_{(0, \alpha)} := \{ f \in \text{lsc-fcns}(X) | \inf_{O} f < \alpha \}, \quad O \subset X, \text{ open, } \quad \alpha \in \mathbb{R},
\]

in fact, it can be generated by these sets; simply observe that \( \inf_{O} f < \alpha \) if and only if \( S_f \) misses the open set \( O \times (-\infty, \alpha) \).

Random lsc functions can be characterized in terms of certain random vectors (with entries in \( \mathbb{R} \)). We refer to such vectors as ‘scalarizations’ of the random lsc functions. The following theorem equates the measurability of \( \text{lsc-fcns}(X) \)-valued random mappings (implying that these mappings are random lsc functions) with the measurability of the corresponding scalarizations. The proof can be found in [22]. Further probabilistic properties attainable through scalarization will be uncovered in §4, and exploited later in §6 in the proofs of the ergodic theorems.

\textbf{Theorem 3.1} (scalarization of random lsc functions). Let \( f : \Xi \to \text{lsc-fcns}(X) \),

\[
\text{and for } D \subset X: \quad \text{let } \pi_D(\xi) := \inf_{x \in D} f(\xi, x).
\]
Then \( f \) is a random lsc function if and only if for all \( D \in \mathcal{D} \), \( \pi_D \) is measurable where \( \mathcal{D} \) is any one of the following collection of subsets of \( X \):

(a) \( \mathcal{D} = \) the open sets \( \mathcal{O} \);
(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 \( \rho \in \mathbb{Q}_+ \).

Through scalarization, one may derive various (important) properties of random lsc functions [22, §3 & 4]. The existence of a lsc version of the conditional expectation of a random lsc function will be used in the sequel.

**Assumption 3.2.** A random lsc function \( f : \Xi \to \mathcal{Lsc}-\mathcal{fcsns}(X) \) is locally inf-integrable if for every \( x \in X \) there is a closed neighborhood \( V \) of \( x \) such that for the (scalar) function

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

**Proposition 3.3** [22, Theorem 4.3]. Let \( f : \Xi \to \mathcal{Lsc}-\mathcal{fcsns}(X) \) be a locally inf-integrable random lsc function and \( \mathcal{R} \subset \mathcal{S} \) a \( \sigma \)-field. Then, there exists a version \( E^\mathcal{R} f \) of the conditional expectation of \( f \) that is \( \mathcal{Lsc}-\mathcal{fcsns}(X) \)-valued.

Moreover, there is a lsc-version with the following property: for all \( \xi \in \Xi' \) with \( \Xi' \subset \Xi \) of \( P \)-measure 1, and \( x \in X \),

\[
E^\mathcal{R} f(\xi, x) = \lim_{\nu \to \infty} \pi_{x', \rho'}^\mathcal{R}(\xi)
\]

where \( x'' \to x, \rho'' \searrow 0 \) and \( \pi_{x', \rho'}^\mathcal{R} \) is a version of the conditional expectation of the scalar function

\[
\pi_{x', \rho'}^\mathcal{R}(\xi) = \inf \{ f(\xi, x') \mid x' \in \mathbb{B}^o(x'', \rho'') \}
\]

for \( \mathbb{B}^o(x'', \rho'') \) the open ball centered at \( x'' \) and of radius \( \rho'' \).

**Proof.** The first part of the statement is just Theorem 4.3 of [22]. The second assertion follows directly from its proof. \( \square \)

4. **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. 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

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

For a sequence $\{f^\nu, \nu \in \mathbb{N}\}$ of random lsc functions, let's denote by $P^\infty$ the probability measure on the sequence space $(\text{lsc-fcns}(X)^\infty, \mathcal{E}^\infty)$ that is consistent with the joint distribution of the $f^\nu$. As explained in the second part of this section, the existence of such a measure follows from Kolmogorov's extension theorem for random lsc functions and the fact that the Effros field is the Borel field associated with a topology $\tau_{cw}$ on lsc-fcns(X) that makes $(\text{lsc-fcns}(X), \tau_{cw})$ a Polish space, as detailed at the end of this section, see Theorem 4.10.

Properties such as independence, stationarity and ergodicity of sequences of random lsc functions may now all be defined in a straightforward manner. We restrict our attention to stationarity and ergodicity for the purposes of this paper; for a somewhat more complete treatment, consult [22, §3].

**Definition 4.1** (stationarity). A sequence, $\{f^\nu, \nu \in \mathbb{N}\}$ of random lsc functions is stationary if its joint distributions are invariant under shifts in the sequence, more precisely, for any finite subcollection $\{f^{\nu_1}, \ldots, f^{\nu_n}\}$, $n \in \mathbb{N}$, any $k \in \mathbb{N}$ and any $\mathcal{A}_1, \ldots, \mathcal{A}_n \in \mathcal{E}$, one has

$$P_{\{f^{\nu_1}, \ldots, f^{\nu_n}\}}(\mathcal{A}_1, \ldots, \mathcal{A}_n) = P_{\{f^{\nu_1+k}, \ldots, f^{\nu_n+k}\}}(\mathcal{A}_1, \ldots, \mathcal{A}_n).$$

Stationarity can also be characterized in terms of a measure preserving transformation. Recall that, a function $\varphi : \Xi \rightarrow \Xi$ is measure preserving if for all $A \in \mathcal{S}$, $P(\varphi^{-1}(A)) = P(A)$. If $f$ is a random lsc function, one verifies easily that that the sequence $\{f, f \circ \varphi, f \circ \varphi^2, \ldots\}$ is stationary. If $\varphi : \Xi \rightarrow \Xi$ is measure preserving, then $A \in \mathcal{S}$ is an invariant event if $\varphi^{-1}(A) = A$ almost surely, i.e., in terms of the symmetric difference, $P(\varphi^{-1}(A) \Delta A) = 0$.

**Definition 4.2** (ergodicity). Let $\mathcal{I}$ denote the $\sigma$-field of invariant events and call it the invariant $\sigma$-field. A measure preserving map $\varphi : \Xi \rightarrow \Xi$ is ergodic if $\mathcal{I}$ is trivial, i.e., for all $A \in \mathcal{I}$, $P(A) \in \{0, 1\}$. For $f$ a random lsc function, a sequence $\{f^\nu, \nu \in \mathbb{N}\}$ of random lsc functions defined by $f^\nu(\xi) := f(\varphi^\nu(\xi))$ is ergodic if the associated measure preserving map $\varphi$ is ergodic.

The stationarity and ergodicity properties of sequences of random lsc functions are inherited by the sequences of corresponding scalarizations. The proofs of these and related results may be found in [22].
Theorem 4.3. 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 $\Xi$. If $\{f^\nu, \nu \in \mathbb{N}\}$ is stationary, then for any open sets $O \subset X$, the sequence of random variables $\{\pi^\nu_O, \nu \in \mathbb{N}\}$ is stationary.

Theorem 4.4. If $\{f^\nu\}^{\nu \in \mathbb{N}}$ is an ergodic sequence of random lsc functions, then for all open $O \subset X$, $\{\pi^\nu_O\}^{\nu \in \mathbb{N}}$ is an ergodic sequence of random variables.

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

$$
P^\nu(A) := P(\{\xi \in \Xi | S^\nu(\xi) \in A\}) \quad \text{for } A \in \mathcal{E}$$

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

$$
P_{\nu_1,\ldots,\nu_k}(A_1, \ldots, A_k) := P(\{\xi \in \Xi | S^{\nu_1}(\xi) \in A_1, \ldots, S^{\nu_k}(\xi) \in A_k\})$$

for $A_1, \ldots, A_k \in \mathcal{E}$.

Assuming it exists, let’s denote by $P^\infty$ the probability measure on the sequence space $(\text{cl-sets}(X)^\infty, \mathcal{E}^\infty)$ 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 Varadarajian’s version of the Kolmogorov Extension Theorem, [26, Theorem V.5.1]:

Theorem 4.5. Let $\{Y^\nu, \nu \in \mathbb{N}\}$ be random variables defined on $(\Xi, \mathcal{S}, P)$ with values in a Polish space $(Z, \tau)$ and $\mathcal{B}$ the associated Borel field on $Z$. Then there exists a unique measure $\mu^\infty$ on $(Z^\infty, \mathcal{B}^\infty)$ 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 isn’t quite our situation. At this stage we don’t 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 [8]. While the Fell topology is the right one for set and epi-convergence (cf. Theorem 1.1), it’s the Choquet-Wijsman topology that renders (cl-sets$(X), \tau_{cw}$) a Polish space, heading us towards the validation of the Kolmogorov’s Extension Theorem that’s required here. We elaborate on these statements in what follows.
The Fell topology $\tau_f$ [17] on cl-sets($X$) is the topology generated by a subbase consisting of sets of the form

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

and

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

That sequential $\tau_f$-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 [18], and a straightforward proof of this fact can be found in [8].

**Proposition 4.6.** $\mathcal{E} \subset \mathcal{B}_f$ where $\mathcal{B}_f$ is the Borel field on cl-sets($X$) when it’s equipped with $\tau_f$, the Fell topology.

**Proof.** This is an immediate consequence of the fact that the sets $\mathcal{V}_O$ in the subbase for the Fell topology generate $\mathcal{E}$. \hfill \Box

The Choquet-Wijsman topology $\tau_{cw}$ [33, 13] on cl-sets($X$) is defined in terms of the distance functions $d_C(x) := \inf_{y \in C} d(x, y)$, with $d$ is the metric on $X$; by convention: $d_\emptyset(x) = +\infty$. This topology, denoted $\tau_{cw}$, consistent with the pointwise convergence of these distance functions, is generated by a subbase consisting of sets of the form

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

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 $C^\circ \xrightarrow{\text{cw}} C$ if and only if $d_{C^\circ}(x) \to d_C(x)$ for all $x \in X$; for more about this topology, see [8]. It is known that the Choquet-Wijsman topology is finer than the Fell topology [8]. 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’s equipped with the Choquet-Wijsman topology.

Hess [19] showed that the Borel field generated by the Choquet-Wijsman topology is the Effros 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 can’t be excluded from our considerations. Nevertheless, it can easily be shown that $\mathcal{B}_{cw} = \mathcal{E}$ for cl-sets($X$). For example, by referring to a fact mentioned by Beer [7]: Let cl-sets$_{\neq \emptyset}(X)$ denote the hyperspace of closed subsets of $X$, not including the empty set.
Proposition 4.7 [7]. Suppose $\tau$ is a topology on cl-sets(X) such that the Borel field on cl-sets$\neq\emptyset$(X) generated by the $\tau$-open sets is the Effrös field on cl-sets$\neq\emptyset$(X). Then, $E$ is the Borel field on cl-sets(X) generated by the $\tau$-open sets if and only if $\{\emptyset\}$ is a Borel subset of cl-sets(X). In particular, this is true if $\{\emptyset\}$ is $\tau$-closed.

Theorem 4.8. On cl-sets(X), $E = B_{\text{cw}}$.

Proof. We know from Hess [19] that on cl-sets$\neq\emptyset$(X), $E$ coincides with the Borel field generated by $\tau_{\text{cw}}$. In view of the preceding proposition, it suffices to verify that $\{\emptyset\}$ is $\tau_{\text{cw}}$-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), $\tau_{\text{cw}}$), viz.

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

Hence, $\{\emptyset\}$ is $\tau_{\text{cw}}$-closed.

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

$$E \subseteq B_f \subseteq B_{\text{cw}} = E,$$

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 4.5, there only remains to establish that (cl-sets(X), $\tau_{\text{cw}}$) is a Polish space. In fact, this is settled by the following result of Beer:

Theorem 4.9 [6]. Suppose (X, d) is a Polish space and (cl-sets(X), $\tau_{\text{cw}}$) is the hyperspace of closed subsets of X equipped with the Choquet-Wijsman topology. Then, the (hyper)space (cl-sets(X), $\tau_{\text{cw}}$) is Polish.

From this theorem, $E = B_f$, Theorem 4.5 and the fact that random lsc functions can be identified with random closed sets (whose values are closed epigraphs), one obtains:

Theorem 4.10 (Kolmogorov’s Extension Theorem). Let $\{f^\nu : (\Xi, S, P) \to \text{lsc-fcns}(X), \nu \in N\}$ where (X, d) is a Polish space and let $E$ be the Effrös field on lsc-fcns(X). For all $\nu$, let $P^\nu$ be the distribution induced by $f^\nu$ on (lsc-fcns(X), $E$). Then, there exists a unique measure $P^\infty$ on (lsc-fcns(X)$^\infty$, $E^\infty$) that is consistent with the family of measures $(P^1, P^2, \ldots)$, i.e., whose finite dimensional projection yield the finite dimensional distributions.
5. Epi-convergence

We are going to need a number of properties of epi-convergent sequences of functions. Recall that for functions \( \{g, g^\nu : X \to \mathcal{M}, \nu \in \mathbb{N} \} \) epi-convergence, written \( g^\nu \rightharpoonup g \) means

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

for all \( x \in X \). When conditions (i) and (ii) are satisfied at some \( x \), it’s convenient to say that the functions \( g^\nu \) 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 \( \mathcal{N}(x) \) denote the neighborhood system of \( x \).

**Definition 5.1.** The lower and upper epi-limits of a sequence of functions \( \{g, g^\nu : X \to \mathcal{M}, \nu \in \mathbb{N} \} \) are defined by,

\[
e\liminf g^\nu(x) := \sup_{V \in \mathcal{N}(x)} \liminf_{\nu \to \infty} g^\nu (y),
\]

\[
e\limsup g^\nu(x) := \sup_{V \in \mathcal{N}(x)} \limsup_{\nu \to \infty} g^\nu (y).
\]

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

It follows from Definition 5.1 that the upper and lower epi-limits are always lsc. Epi-convergence of \( g^\nu \) to \( g \) corresponds to the set convergence of epi \( g^\nu \) to epi \( g \) in the Fell topology. 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 \( g^\nu \) to the minimizers of \( g \), in the following sense.

**Theorem 5.2** [29, Theorem 7.31]. Suppose \( \{g^\nu\}_{\nu \in \mathbb{N}} \) is a sequence of extended real-valued lsc functions such that \( g^\nu \rightharpoonup g \). Then every cluster point of \( \argmin g^\nu \) is an element of \( \argmin g \). Moreover, if \( \argmin g \) is nonempty, and there exists a compact \( K \subset X \) such that \( \text{dom } g^\nu \subset K \), then

\[
\argmin g = \bigcap_{\varepsilon > 0} \liminf_{\nu \to \infty} (\varepsilon - \argmin g^\nu),
\]

where \( \varepsilon - \argmin g := \{x \in X | g(x) \leq \inf g + \varepsilon < \infty \} \).

Here we state only sufficient conditions for the convergence of the \( \varepsilon \)-\( \argmin \). For necessary conditions, as well as conditions for the convergence of infima, consult [29,
Chapter 7], and for certain extensions of these results when dealing with random lsc functions, see also [1].

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 tells us that epi-convergence needs only to be verified at the points of a countable dense set.

**Lemma 5.3** [4]. Let $f, g : X \to \mathbb{R}$ with $f$ lsc. Let $R \subset X$ be the projection on $X$ of a countable dense subset of 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 $f \leq g$ on $R$. This is equivalent to $\{(x, \alpha) \mid \alpha \geq g(x), x \in R\} \subset \text{epi } f$. Since $f$ is lsc, epi $f$ is closed. Taking closures on both sides yields epi $g \subset$ epi $f$, which is equivalent to $f \leq g$ on $X$. \( \square \)

### 6. 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. Assumption 3.2 of local inf-integrability, already used in the construction of a lsc-version of the conditional expectation of a random lsc function will be needed throughout. But this is not a significant restriction.

**Lemma 6.1.** Let $(X, d)$ be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, and $\mathcal{R} \subset \mathcal{S}$, a $\sigma$-field. Let $\{f, f^\nu : \Xi \to \text{lsc-lcns}(X) \mid \nu \in \mathbb{N}\}$ be a sequence of random lsc functions with $f$ locally inf-integrable. For $x \in X$, $\rho \in \mathbb{R}_+$, let $\{\pi_{x, \rho}, \pi_{x, \rho}^\nu \mid \nu \in \mathbb{N}\}$ be a sequence of scalarizations of these random lsc functions obtained as follows:

\[
\pi_{x, \rho} := \inf_{H^\nu(x, \rho)} f, \quad \pi_{x, \rho}^\nu := \inf_{H^\nu(x, \rho)} f^\nu, \\
\pi_{x, 0} := f(\cdot, x), \quad \pi_{x, 0}^\nu := f^\nu(\cdot, x).
\]

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

\[
\frac{1}{\nu} \sum_{k=1}^{\nu} \pi_{x, \rho}^k(\xi) \to \pi_{x, \rho}^\mathcal{R}(\xi) \quad P\text{-a.s.},
\]

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

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

\[
\frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, \cdot) \Rightarrow E^\mathcal{R} f(\xi) \quad P\text{-a.s.}
\]
In particular, if $\mathcal{R}$ is independent of the $\sigma$-field generated by $f$, then

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

**Proof.** Let $E^R f$ be a version of the conditional expectation of $f$ whose values are in lsc-FCN(X) as guaranteed by Theorem 3.3 and such that for all $\xi \in \Xi_1$, a set of measure 1, and $x \in X$: $E^R f(\xi)(x) = \lim_{\rho \to 0} \pi^R_{x', \rho'}(\xi)$ for $x' \to x$ and $\rho' \searrow 0$. Let $R$ be a countable dense subset of $X$ that contains the projection onto $X$ of a countable dense set $R^+ \subset \text{epi} E^R f(\xi)$ for all $\xi \in \Xi_2$, also a subset of $\Xi$ 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

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

Finally, let

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

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

Let’s begin by showing that

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

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

$$\text{e-lim inf}_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x) = \sup_{\rho > 0} \lim_{\nu \to \infty} \inf_{x' \in B(x, \rho)} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x')$$

$$\geq \sup_{\rho > 0} \lim_{\nu \to \infty} \inf_{x' \in B(x, \rho)} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^k_{x', \rho'}(\xi),$$

where $x' \to x$, $\rho' \searrow 0$ and

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

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

$$\lim_{\nu} \inf_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^k_{x^l, \rho^l}(\xi) = \pi^R_{x^l, \rho^l}(\xi)$$

20
by assumption. Now, observing that when $x \rightarrow x$ and $\rho \rightarrow 0$, $\pi^R_x(\xi) \wedge E^R f(\xi, x)$ for $\xi \in \Xi_0 \subset \Xi_1$, taking the supremum over $l \in \mathbb{N}$ yields

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

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

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

$$\limsup_{\nu} \frac{1}{\nu} \sum_{k=1}^{\nu} f^k(\xi, x^*) = \limsup_{\nu \rightarrow \infty} \frac{1}{\nu} \sum_{k=1}^{\nu} \pi^R_{x^*}(\xi) = E^R f(\xi)(x).$$

In view of Definition 5.1, this is the same as for all $x \in R$,

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

Using the facts that $E^R f$ 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 $\text{epi } E^R f$, one appeals to Lemma 5.3 to obtain:

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

Hence

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

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

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

as claimed.

The condition ‘there exists a countable subset $R^+ \subset X \times R$ that is dense in $\text{epi } E^R f(\xi)$ 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 $X_\alpha$, where $X_\alpha$ is a countable dense subset of $X$ containing the projection onto $X$ of a countable dense subset of $\text{epi } E^R f(\alpha)$, and $E^R f(\alpha)(x) := \int_\alpha f(\xi, x)P(\,d\xi)/P(\alpha)$. This situation occurs when there exist a countable number of sets.
$A \in \mathcal{S}$ that partition $\Xi$ 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 $\text{epi} E^R f(\xi)$ $P$-a.s.. For example, when $\text{epi} E^R f(\xi)$ is $P$-almost surely a solid set, i.e.,

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

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 $\xi$, the function $x \mapsto f(\xi, x)$ is continuous on its domain (where it is finite), or more broadly when $\text{epi} f(\xi, \cdot)$ 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 [24, 15, §6.2] and [14], for example. For our purposes, we need a version that allows for functions that are extended real-valued. A straightforward modification of [24, Theorem 33.B, Theorem 34.A] takes care of this situation.

**Theorem 6.2.** Let $(X, d)$ be a Polish space, $(\Xi, \mathcal{S}, P)$ a probability space, $\varphi : \Xi \to \Xi$ a measure preserving transformation and $\mathcal{I}$ its invariant $\sigma$-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 $\text{epi} E^T f(\xi)$ $P$-a.s., then

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

In particular, if $\varphi$ is ergodic, then

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

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

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

where $\pi_{x, \rho}^T$ is a version of the conditional expectation of $\pi_{x, \rho}$ with respect to the invariant field $\mathcal{I}$. We are now in the setting of Lemma 6.1, which immediately yields

$$\frac{1}{\nu} \sum_{k=1}^{\nu} f(\varphi^k(\xi), \cdot) \Rightarrow E^T f(\xi) \quad P\text{-a.s.}$$
If $\varphi$ is ergodic, $\mathcal{I}$ is trivial, whereby $\mathcal{I}$ is independent of the $\sigma$-field generated by $f$, so that again by Lemma 6.1,

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

as claimed. $\blacksquare$

References


