Annals of Operations Research 30 (1991) 157–168

GLIVENKO–CANTELLI TYPE THEOREMS: AN APPLICATION OF THE CONVERGENCE THEORY OF STOCHASTIC SUPREMA *

Gabriella SALINETTI
Università "La Sapienza", 00100 Roma, Italy

Roger J.-B. WETS
University of California, Davis, CA 95616, USA

The uniform convergence of empirical processes on certain classes of sets follows from the convergence theory for random lower semicontinuous functions studied in the context of stochastic optimization. In the process, a richer class of sets for which one can prove this type of result is exhibited.

Keywords: Glivenko–Cantelli lemma, narrow (weak, weak*) convergence.

1. Introduction and problem setting

Let $\{X^\nu, \nu = 1, 2, \dots\}$ be an iid (independent and identically distributed) sequence of random variables defined on a probability space $(\Omega, \mathscr{A}, \mu)$ with values in a locally compact, separable metric space E and P the distribution induced by the X^ν, $\nu = 1, \dots$, on $\mathscr{B}$, the Borel field on E. Typically, the X^ν model the observation process of a statistical experiment.

The *empirical random measure* $P^\nu : \mathscr{B} \to [0, 1]$ associated with the first ν random variables $X^1, \dots, X^\nu$ is

$$P^\nu(B, \omega) = \frac{1}{\nu} \sum_{i=1}^{\nu} I_{\{X^i(\omega) \in B\}}(\omega), \quad \forall B \in \mathscr{B}.$$

with I_D the indicator function of the set D.

In the case the X^ν are real-valued random variables, the classical Glivenko–Cantelli theorem asserts the μ-almost sure uniform convergence of the empirical distribution functions P^ν to P, i.e.,

$$\sup_{\alpha \in \mathbb{R}} |P^\nu((-\infty, \alpha], \cdot) - P(-\infty, \alpha]| \to 0 \quad \mu - \text{a.s.}$$

* Research supported in part by grants from Ministero Publica Istruzione and the National Science Foundation.

Theorems of the Glivenko–Cantelli type [3,5,10] are concerned with the almost sure (a.s.) uniform convergence of the random measures $\{P^\nu, \nu \in \mathbb{N}\}$ to the distribution P for a given subclass of sets $\mathscr{C}$ of $\mathscr{B}$. More precisely, they assert that for certain classes of sets $\mathscr{C} \subset \mathscr{B}$, for μ-almost all ω in Ω,

$$\sup_{C \in \mathscr{C}} |P^\nu(C, \omega) - P(C)| \to 0 \quad \text{as } \nu \to \infty. \tag{1.1}$$

An assertion of this type certainly demands that for all $C \in \mathscr{C}$ and $\omega \in \Omega \setminus N$,

$$P^\nu(C, \omega) \to P(C). \tag{1.2}$$

Thus, a "minimal" requirement, is that the P^ν a.s. converge narrowly (weakly) to P and, that $\mathscr{C}$ is a subset of the continuity set of P, cont $P := \{B \in \mathscr{B} \mid \mu(\text{bdry } B) = 0\}$.

Almost sure uniform convergence has been proved for particular subclasses $\mathscr{C} \subset$ cont P by relying mostly on the geometrical properties of the class $\mathscr{C}$. We are going to show that it is possible to obtain these results as special cases of a general theorem that is *topological* in nature, viz., as a consequence of a certain compactness of $\mathscr{C}$. A key step in the derivation is to identify probability measures (defined on $\mathscr{B}$) with their restriction to the space $\mathscr{F}$ of closed subsets of E. We know from earlier results [8] that such (restricted) functions are upper semicontinuous on $\mathscr{F}$ with respect to the topology of set-convergence. From this point of view, stating a Glivenko–Cantelli type theorem boils down to finding conditions that guarantee that a certain sequence of (random) upper semicontinuous functions converges uniformly (almost surely). This is elaborated in section 3 and the implications for empirical processes are collected in section 4. Section 2 is a compilation of facts about set-convergence and the (hypo-) epi-convergence of functions.

2. Preliminaries

Let $\mathscr{F} = \mathscr{F}(E)$ be the class of closed subsets of E. A sequence $\{F^\nu \in \mathscr{F}, \nu \in \mathbb{N}\}$ (topologically) *converges* to the (closed) set F if

$$\limsup_{\nu \to \infty} F^\nu \subset F \subset \liminf_{\nu \to \infty} F^\nu,$$

with

$$\liminf_{\nu} F^\nu := \Big\{x \in E \mid x \text{ limit point of } \{x^\nu\}_{\nu=1}^{\infty},$$
$$x^\nu \in F^\nu \text{ for all but finitely many } \nu\Big\},$$
$$\limsup_{\nu} F^\nu := \Big\{x \in E \mid x \text{ cluster point of } \{x^\nu\}_{\nu=1}^{\infty},$$
$$x^\nu \in F^\nu \text{ for infinitely many } \nu\Big\}.$$

It is well know that this convergence induces a topology $\mathscr{T}$ on $\mathscr{F}$. With $\mathscr{G}$, the class of open subsets of E, and $\mathscr{K}$, the class of compact subsets of E, and for any $D \subset E$,

$$\mathscr{F}^D := \{ F \in \mathscr{F} \mid F \cap D = \emptyset \}$$

and

$$\mathscr{F}_D := \{ F \in \mathscr{F} \mid F \cap D \neq \emptyset \},$$

the sets

$$\left\{ \mathscr{F}^K \cap \mathscr{F}_{G_1} \cap \cdots \cap \mathscr{F}_{G_s} \mid K \in \mathscr{K},\ G_l \in \mathscr{G},\ l = 1, \ldots, s,\ s \text{ finite} \right\}$$

determine a *base* for the topology $\mathscr{T}$. The topological space $(\mathscr{F}, \mathscr{T})$ is separated (Hausdorff), has a countable base and is compact, and consequently metrizable, e.g., see [1,2].

A sequence of functions $\{ f^\nu : E \to \overline{\mathbb{R}},\ \nu \in \mathbb{N} \}$ *hypo-converges* to $f : E \to \overline{\mathbb{R}}$, $f^\nu \xrightarrow{h} f$, if for every x

$$f(x) \le \liminf_{\nu \to \infty} f^\nu(x^\nu) \text{ for some } x^\nu \to x,$$

$$f(x) \ge \limsup_{\nu \to \infty} f^\nu(x^\nu) \text{ for all } x^\nu \to x.$$

The functions f^ν *hypo-converge* to f *on* $C \subset E$ if the f^ν hypo-converge relative to C. Hypo-convergence of the functions f^ν to f implies and is implied by the identities:

$$\limsup_{\nu \to \infty} \text{hypo } f^\nu = \text{hypo } f = \liminf_{\nu \to \infty} \text{hypo } f^\nu,$$

with hypo $g = \{ (x, \alpha) \mid \alpha \le f(x),\ x \in E \} \subset E \times \mathbb{R}$, the *hypograph* of the function g.

Hypo-convergence provides a "minimal" framework that guarantees the convergence of the suprema. It is easy to verify that if $C \subset E$ is compact, then

$$f^\nu \xrightarrow{h} f \Rightarrow \sup_{x \in C} f^\nu(x) \to \sup_{x \in C} f. \tag{2.1}$$

The next proposition encapsulates the flavor of our approach.

PROPOSITION 2.1

A sequence of extended real-valued functions $\{ f^\nu,\ \nu \in \mathbb{N} \}$ defined on a metric space E converge uniformly to a function f on a compact $C \subset E$ if and only if the sequence of functions $\{ | f^\nu - f |,\ \nu \in \mathbb{N} \}$ hypo-converges on C to the function that is identically 0 on C.

Proof

It is immediate that the uniform convergence of the sequence $\{ f^\nu,\ \nu \in \mathbb{N} \}$ to a function f on a compact subset $C \subset E$, i.e., $\sup_{x \in C} | f(x) - f^\nu(x) | \to 0$ implies the hypo-convergence on C of the sequence $\{ | f - f^\nu |,\ \nu \in \mathbb{N} \}$ to the function that is identically 0 on C. For the converse, observe that $|f-f^\nu| \xrightarrow{h} 0$ on C, means that for all $x \in C$ and $\epsilon > 0$, there exists $V_{x,\epsilon}$, a neighborhood of x, and $\nu_\epsilon(x)$ such that for all $x' \in V_{x,\epsilon}$ and all $\nu \geq \nu_\epsilon(x)$, $|f^\nu(x') - f(x')| \leq \epsilon$. With x varying over C, the $V_{x,\epsilon}$ determine an open cover of C. Since C is compact, this open cover admits a finite open cover, say $\{V_{x^k,\epsilon}, k \in K, K \text{ finite}\}$. Let $\nu_\epsilon = \max_k \nu_\epsilon(x^k)$. It follows that for all $x \in C$ and all $\nu \geq \nu_\epsilon$, $|f^\nu(x) - f(x)| \leq \epsilon$. □

When E is a locally compact separable metric space hypo-convergence induces a topology, the *hypo-topology*, on the space of upper semicontinuous (usc) functions [1,2]. This provides the "functional" framework for the study of statistical processes with usc realizations, in particular for processes whose realizations are empirical distributions. Indeed, a stochastic process $\{g(x, \cdot), x \in E\}$ defined on a complete probability space $(\Omega, \mathscr{A}, \mu)$ with usc realizations $g(\cdot, \omega)$ and measurable (i.e., $(x, \omega) \mapsto g(x, \omega)$ is $\mathscr{B} \otimes \mathscr{A}$-measurable) is an *upper semicontinuous random function*, i.e., a random function that is usc with respect to x and *hypo-measurable* (measurable with respect to the Borel field generated by the hypo-topology on the space of usc functions) [8, 9, theorem 6.1].

Almost sure hypo-convergence of the stochastic processes $\{(g^\nu(x, \cdot), x \in E), \nu \in \mathbb{N}\}$ to the process $\{g(x, \cdot), x \in E\}$, denoted $g^\nu \xrightarrow{h} g$ *a.s.*, means that there exists a μ-null set N such that

$$g^\nu(\cdot, \omega) \xrightarrow{h} g(\cdot, \omega) \quad \forall \omega \in \Omega \setminus N.$$

The next proposition provides a characterization of a.s. hypo-convergence in terms of the convergence of the suprema. An earlier proof, given in [6], was based on a characterization of the a.s. convergence of measurable set-valued mappings, a more direct proof follows.

PROPOSITION 2.2 [6]

Let $\{g; g^\nu, \nu \in \mathbb{N}\}$ be a family of hypo-measurable stochastic processes with usc realizations and values in E a separable metric space. Let $\mathscr{U}$ be a base of open sets for the topology on E. Then, the g^ν a.s. hypo-converge to g if and only if for all open sets B in $\mathscr{U}$ and all $\alpha \in \mathbb{R}$ there exists a subset $N(B, \alpha)$ of Ω with $\mu(N(B, \alpha)) = 0$ such that

$$\text{(i)} \quad \left\{\omega \mid \sup_{x \in Bi} g(x, \omega) > \alpha\right\} \subset \left\{\omega \mid \sup_{x \in B} g^\nu(x, \omega) > \alpha, \text{ a.a.}\right\} \cup N(B, \alpha),$$

$$\left(\text{(ii)} \quad \left\{\omega \mid \sup_{x \in \text{cl}\, B} g(x, \omega) < \alpha\right\} \subset \left\{\omega \mid \sup_{x \in \text{cl}\, B} g^\nu(x, \omega) < \alpha \text{ a.a.}\right\} \cup N(B, \alpha)\right.$$

where cl B is the closure of B in E and "a.a." stands for *almost always* and means that the inequality is satisfied for all but a finite number of ν's.

Proof

Suppose that

$$g^\nu(\cdot, \omega) \xrightarrow{h} g(\cdot, \omega) \quad \forall \omega \in \Omega \setminus N, \ \mu(N) = 0.$$

Let B be an open ball and $\alpha \in \mathbb{R}$. Let $\bar{\omega}$ be such that $\sup_{x \in B} g(x, \bar{\omega}) > \alpha$. Then, there exists $\bar{x} \in B$ such that $g(\bar{x}, \bar{\omega}) > \alpha$. By hypo-convergence (see above) with $\omega = \bar{\omega}$ there exists $x^\nu \to \bar{x}$ with $g(\bar{x}, \bar{\omega}) \leq \liminf_\nu g^\nu(x^\nu, \bar{\omega})$. Moreover, for ν sufficiently large, $x^\nu \in B$ since B is open. Thus, for all ν sufficiently large, we have that $\sup_{x \in B} g^\nu(x, \bar{\omega}) \geq g^\nu(x^\nu, \bar{\omega}) > \alpha$ and this yields (i).

To show (ii) let $\bar{\omega}$ be such that $\sup_{x \in \mathrm{cl}\, B} g(x, \bar{\omega}) < \alpha$. Because E is a separable metric space, we can always find a countable base of E of relatively compact open balls. Arguing by contradiction, passing to a subsequence if necessary, assume that $\sup_{x \in \mathrm{cl}\, B} g^\nu(x, \bar{\omega}) \geq \alpha$ for all ν (in the subsequence). Since $\mathrm{cl}\, B$ is compact and $g^\nu(\cdot, \bar{\omega})$ is usc, this yields a sequence $\{x^\nu\}_\nu \subset \mathrm{cl}\, B$ with $g^\nu(x^\nu, \bar{\omega}) = \sup_{x \in \bar{B}} g^\nu(x, \bar{\omega}) \geq \alpha$. Since all $\{x^\nu\}_\nu$ belong to $\mathrm{cl}\, B$, they cluster at some point $\bar{x} \in \mathrm{cl}\, B$. From hypo-convergence we have

$$\alpha > \sup_{x \in \mathrm{cl}\, B} g(x, \bar{\omega}) \geq g(x, \bar{\omega}) \geq \limsup_{\nu \to \infty} g^\nu(x^\nu, \bar{\omega}) \geq \alpha;$$

this is a contradiction. And thus (ii) must hold.

To show the "if"-part, let $N(B_i, \alpha_j)$ be the sets of null μ-measure that appear in (i) and (ii) with B_i varying over a countable base (for E) of open balls and the α_j in a countable dense subset A of $\mathbb{R}$. Set $N = \cup_{i,j} N(B_i, \alpha_j)$. Note that $\mu(N) = 0$. We show that if (i) and (ii) are satisfied, then for all $\omega \in \Omega \setminus N$, $g^\nu(\cdot, \omega) \xrightarrow{\mathrm{h}} g(\cdot, \omega)$.

Let $\bar{\omega} \in \Omega \setminus N$ and $\bar{x} \in E$. Let $\{B_s(\bar{x}),\ s = 1, \ldots\}$ be a countable fundamental system of neighborhoods of $\bar{x}$ and let $\{\alpha_s\}_s$ be a sequence in A converging to $g(\bar{x}, \bar{\omega})$ with $\alpha_s < g(\bar{x}, \bar{\omega})$ for all s. For each s, $g(\bar{x}, \bar{\omega}) > \alpha_s$ implies $\sup_{x \in B_s} g^\nu(x, \bar{\omega}) > \alpha_s$ for ν sufficiently large, say $\nu > \nu_s$. Hence, for all $\nu > \nu_s$ there exists $x_s^\nu \in B_s$ with $g_\nu(x_s^\nu, \bar{\omega}) > \alpha_s$. Choosing $\nu_s \uparrow \infty$ as $s \uparrow \infty$, we can generate a sequence $\{x^\nu\}$ such that

$$x^\nu \in B_s \quad \text{and} \quad g_\nu(x^\nu, \bar{\omega}) > \alpha_s \quad \text{for } \nu_s < \nu \leq \nu_{s+1}.$$

It follows that $x^\nu \to \bar{x}$ and $\liminf_\nu g_\nu(x^\nu, \bar{\omega}) \geq \lim_s \alpha_s = g(\bar{x}, \bar{\omega})$. We have just proved that for each $\omega \in \Omega \setminus N$ the lim inf-condition for hypo-convergence is satisfied.

We obtain the lim sup-condition for hypo-convergence from (ii). Let $\bar{\omega} \in \Omega \setminus N$, $\bar{x} \in E$, $x^\nu \to x$. Let $\{\alpha_s\}_s$ be a sequence in A converging to $g(\bar{x}, \bar{\omega})$ with $\alpha_s > g(\bar{x}, \bar{\omega})$ for all s. For each s, since $g(\cdot, \bar{\omega})$ is usc at $\bar{x}$ there exists $\mathrm{cl}\, B_s$ such that $\sup_{x \in \mathrm{cl}\, B_s} g(x, \bar{\omega}) < \alpha_s$. By (ii) we also have that $\sup_{x \in \mathrm{cl}\, B} g^\nu(x, \bar{\omega}) < \alpha_s$ for ν sufficiently large, say $\nu > \nu_s$. Also, $g^\nu(x^\nu, \bar{\omega}) < \alpha_s$ for ν sufficiently large, say $\nu > \nu_s$. It follows that $\limsup_\nu g^\nu(x^\nu, \bar{\omega}) \leq \alpha_s$. The argument repeated for every s yields $\limsup_\nu g^\nu(x^\nu, \bar{\omega}) \leq \lim_s \alpha_s = g(\bar{x}, \bar{\omega})$ and completes the proof. □

3. Uniform convergence of probability measures and a.s. uniform convergence of random measures

For a probability measure P on $\mathscr{B}$, the restriction D on $\mathscr{F}-$ with $D(F) = P(F)$ for all $F \in \mathscr{F}-$ is an usc function on the topological space $(\mathscr{F}, \mathscr{T})$ [7]. Moreover, for the family of probability measures $\{P;\ P^\nu,\ \nu \in \mathbb{N}\}$ narrow convergence $P^\nu \xrightarrow{n} P$ is equivalent to the hypo-convergence of the (usc-)restrictions $\{D;\ D^\nu,\ \nu \in \mathbb{N}\}$ [7],

$$P^\nu \xrightarrow{n} P \text{ if and only if } D^\nu \xrightarrow{h} D.$$

The uniform convergence of the probability measures P^ν to P on a subset $\mathscr{C}$ of $\mathscr{F}$ can thus be expressed in terms of the convergence of suprema, namely,

$$\sup_{C \in \mathscr{C}} | D^\nu(C) - D(C) | \to 0.$$

In view of proposition 2.1, a necessary and sufficient condition for this to hold, is that the function $| D^\nu - D |$ hypo-converges on $\mathscr{C}$ to the null function. Because $| D^\nu - D | \geq 0$, from the definition of hypo-convergence, it follows that this is equivalent to demanding that for all $C \in \mathscr{C}$, and every sequence C^ν $\mathscr{T}$-converging to C

$$\limsup_{\nu \to \infty} | D^\nu(C^\nu) - D(C^\nu) | = 0. \tag{3.1}$$

Hypo-convergence of the D^ν to D is not enough to guarantee (3.1), but it will do if the set $\mathscr{C}$ is a "bicompact" subset of the continuity set, cont D = cont $P \cap \mathscr{F}$, of D; note that cont $D = \{ F \in \mathscr{F} \mid D(\text{bdry } F) = 0\}$.

DEFINITION 3.1

A subset $\mathscr{C}$ of $\mathscr{F}$ is *bicompact* if it is compact with respect to $\mathscr{T}$ and if for any set $C \in \mathscr{C}$ and any sequence $\{C^\nu,\ \nu \in \mathbb{N}\}$ with the C^ν $\mathscr{T}$-converging to C, the complements of the interior of the C^ν also $\mathscr{T}$-converge to the complement of the interior of C, i.e., cpl(int C^ν) $\to$ cpl(int C).

We call such a set ($\mathscr{T}$-)bicompact because the second condition corresponds to the compactness of $\mathscr{C}' = \{\text{int } C,\ C \in \mathscr{C}\}$ with respect to the topology $\mathscr{T}'$ on the open subsets of E induced by the $\mathscr{T}$-convergence of their complements. Classes of sets that are bicompact have been identified in [4]. They include, for example, convex sets (assuming E is a linear space), and any collection of sets that can be obtained as the (lower) level sets $\text{lev}_\alpha f := \{ x \mid f(x) \leq \alpha\}$ of a continuous function.

PROPOSITION 3.2

Suppose $\mathscr{C} \subset$ cont D is bicompact and $D^\nu \xrightarrow{h} D$. Then $| D^\nu - D | \xrightarrow{h} 0$ on $\mathscr{C}$.

Proof

From $D^\nu \xrightarrow{h} D$, it follows that for any F and sequence $F^\nu \to F$, $\limsup_\nu D^\nu(F^\nu) \leq D(F)$. In turn, this implies that for any open set G and any sequence G^ν of open sets such that cpl $G^\nu \to$ cpl G,

$$P(G) \leq \liminf_{\nu \in \mathbb{N}} P^\nu(G^\nu).$$

Because $\mathscr{C}$ is bicompact, for the closed sets F and F^ν, and with $G = \text{int } F$, $G^\nu = \text{int } F^\nu$, we have that $\text{cpl } G^\nu \to \text{cpl } G$, and thus

$$P(\text{int } F) \le \liminf_{\nu \to \infty} P^\nu(\text{int } F^\nu). \tag{3.2}$$

By (a) upper semicontinuity of D at cpl(int F), (b) cpl(int F^ν) $\to$ cpl(int F), and (c) $\mathscr{C} \subset \text{cont } D$, we have

$$\begin{aligned}\limsup_{\nu \to \infty} D(F^\nu) \le D(F) = P(\text{int } F) = 1 - P(\text{cpl}(\text{int } F))\\ = 1 - D(\text{cpl}(\text{int } F)) \le 1 - \limsup_{\nu \to \infty} D(\text{cpl}(\text{int } F^\nu))\\ \le \liminf_{\nu \to \infty} P(\text{int } F^\nu) \le \liminf_{\nu \to \infty} D(F^\nu).\end{aligned}$$

For any $\epsilon > 0$ and ν sufficiently large, it follows that

$$D^\nu(F^\nu) - D(F^\nu) = D^\nu(F^\nu) - D(F) + D(F) - D(F^\nu) < \epsilon.$$

Moreover, because D is usc at F, relation (3.2) and $F \in \mathscr{C} \subset \text{cont } D$, for ν sufficiently large we have

$$\begin{aligned}D^\nu(F^\nu) - D(F^\nu) > P^\nu(\text{int } F^\nu) - D(F) - \epsilon/2\\ > P(\text{int } F) - D(F) - \epsilon = -\epsilon.\end{aligned}$$

The last two strings of inequalities hold for every $\epsilon > 0$, they imply that $\limsup_\nu | D^\nu(F^\nu) - D(F^\nu) | = 0$ and this completes the proof. □

By relying on the correspondence between P's and D's ([7]), we can rephrase this result in terms of the probability measures P^ν.

PROPOSITION 3.3

Let $\{P; P^\nu, \nu \in \mathbb{N}\}$ be a family of probability measures defined on $\mathscr{B}$. Suppose that the P^ν converge narrowly to P. Then the P^ν converge uniformly to P on any bicompact subset of $\mathscr{F} \cap \text{cont } P$.

The extension of this result to the a.s.-convergence of random probability measures is immediate. Given a probability space $(\Omega, \mathscr{A}, \mu)$, the stochastic process $\{P^\nu(B, \cdot), B \in \mathscr{B}\}$ whose realizations are probability measures defined on $\mathscr{B}$, is called a *random probability measure*. Because every random probability measure P is uniquely defined by its restriction to the closed sets (again [7]), we can identify such a stochastic process with one that involves the corresponding (usc) functions: $\{D(F, \cdot), F \in \mathscr{F}\}$. We note that in view of our earlier observations, this is a measurable process with usc realizations.

For a family $\{P; P^\nu, \nu \in \mathbb{N}\}$ of random probability measures, *almost sure narrow convergence* means that there exists a set N of μ-measure 0 such that for all $\omega \in \Omega \setminus N$:

$$P^\nu(\cdot, \omega) \xrightarrow{\text{n}} P(\cdot, \omega),$$

or, equivalently

$$D^\nu(\cdot, \omega) \xrightarrow{\text{h}} D(\cdot, \omega), \quad \forall \omega \in \Omega \setminus N.$$

We now rely on proposition 3.3 to conclude that if $P^\nu(\cdot, \omega) \xrightarrow{\text{n}} P(\cdot, \omega)$ a.s. they also a.s. converge uniformly on every bicompact subset $\mathscr{C}$ of $\mathscr{F}$ that is contained in cont P.

PROPOSITION 3.4

Let $\{P^\nu, \nu \in \mathbb{N}\}$ be a sequence of random probability measures, and P a probability measure, all defined on $\mathscr{B}$. Suppose that $P^\nu(\cdot, \omega) \xrightarrow{\text{n}} P(\cdot)$ a.s. Then, they converge a.s. uniformly on every bicompact subset of $\mathscr{F} \cap$ cont P.

4. Glivenko–Cantelli type results

In the statistical framework described in section 1, it follows from the strong law of large numbers that for iid observations $\{X^\nu, \nu \in \mathbb{N}\}$, the sequence of random empirical measures $\{P^\nu, \nu \in \mathbb{N}\}$ converge narrowly to the distribution P, i.e.,

$$P^\nu(B, \cdot) \to P(B), \quad \forall B \in \mathscr{B}, \tag{4.1}$$

i.e., for every $B \in \mathscr{B}$ there exists a μ-null subset of Ω, say N_B such that

$$P^\nu(B, \omega) \to P(B), \quad \forall \omega \in \Omega \setminus N_B.$$

For the corresponding random usc functions, the restrictions of the P^ν to $\mathscr{F}$, we have that

$$D^\nu(F, \omega) \to D(F), \quad \forall \omega \in \Omega \setminus N_F, \forall F \in \mathscr{F}.$$

We thus have a.s.-"pointwise" convergence of the stochastic processes D^ν (indexed by F) to the "constant" valued stochastic processes $\{D(F, \cdot), F \in \mathscr{F}\}$ with $D(F, \omega) = D(F)$ for all $\omega \in \Omega$.

However, as pointed out in section 1, to obtain a.s. uniform convergence, a minimal requirement is the a.s. convergence of the empirical random probability measure, or equivalently, the a.s. hypo-convergence of the corresponding random usc functions D^ν to D. In general, for measurable stochastic processes with usc realizations, the a.s. convergence (as stochastic processes) and the a.s. hypo-convergence are not equivalent, neither implies the other [8, section 3]. But for random probability measure, their specific properties (monotonicity) allow us to show that in fact, a.s. convergence (in the classical sense of stochastic processes) is enough to ensure a.s. hypo-convergence.

Before we get to this, let us identify maximal elements for the sets in the (countable) base

$$\mathscr{S} := \left\{ \mathscr{F}^K \cap \mathscr{F}_{G_1} \cap \cdots \cap \mathscr{F}_{G_s} \mid K \in \mathscr{K}, G_l \in \mathscr{G}, l = 1, \ldots, s, s \text{ finite} \right\}$$

for the topology $\mathscr{T}$; recall that $\mathscr{K}$ and $\mathscr{G}$ consist of the compact and open subsets of E. Let $\mathscr{H}$ be a nonempty subset of $\mathscr{S}$ with $\mathrm{cl}\,\mathscr{H}$ its $\mathscr{T}$-closure; note that $\mathscr{H}$ is nonempty if for $l=1,\dots,s$, $G_l \not\subset K$. We are going to show that $\mathrm{cl}(\mathrm{cpl}\ K) =: F'$ is a maximal element with respect to inclusion in $\mathrm{cl}\,\mathscr{H}$, i.e., $F' \in \mathrm{cl}\,\mathscr{H}$ and $F \subset F'$ for all $F \in \mathrm{cl}\,\mathscr{H}$. Pick $\epsilon > 0$ such that for $l=1,\dots,s$, $G_l \not\subset K_\epsilon := \{x \in E \mid \mathrm{dist}(x, K) < \epsilon\}$. Now choose $\epsilon^\nu \downarrow 0$ such that for all ν, $\epsilon_\nu < \epsilon$ and let $F^\nu = \mathrm{cpl}\ K_{\epsilon_\nu}$. It is obvious that the $F^\nu \in \mathscr{H}$ and that $F^\nu \to F'$. Thus $F' \in \mathrm{cl}\,\mathscr{H}$. Now, for every $F \in \mathscr{H}$, we have that $F \cap K = \emptyset$ and consequently, $F \subset \mathrm{cl}(\mathrm{cpl}\ K) = F'$. Moreover, for every $F \in \mathrm{cl}\,\mathscr{H}$ there exist $F^\nu \to F$ with $F^\nu \in \mathscr{H}$ so that for all ν, $F^\nu \subset F'$. It follows that $F = \lim_\nu F^\nu \subset F'$.

THEOREM 4.1

Suppose $\{D;\ D^\nu,\ \nu \in \mathbb{N}\}$ is a family of random usc functions defined on $\mathscr{F}$ obtained by restricting probability measures $\{P;\ P^\nu,\ \nu \in \mathbb{N}\}$ (defined on $\mathscr{B}$) to $\mathscr{F}$. If

$$D^\nu(F,\cdot) \to D(F,\cdot) \text{ a.s.,} \quad \forall F \in \mathscr{F},$$

then there exists a set N of μ-measure 0 such that

$$D^\nu(\cdot,\omega) \xrightarrow{\mathrm{h}} D(\cdot,\omega), \quad \forall \omega \in \Omega \setminus N.$$

Proof

In view of proposition 2.2, it suffices to check if the inclusions (i) and (ii) of proposition 2.2 are satisfied for every open set $\mathscr{H}$ in the base

$$\mathscr{S} := \left\{ \mathscr{F}^K \cap \mathscr{F}_{G_1} \cap \cdots \cap \mathscr{F}_{G_s} \mid K \in \mathscr{K},\ G_l \in \mathscr{G},\ l=1,\dots,s,\ s \text{ finite} \right\};$$

for the topology $\mathscr{T}$ (on $\mathscr{F}$), and all $\alpha \in \mathbb{R}$. More precisely, we have to show that there exists a set $N(\mathscr{H},\alpha)$ of μ-measure null such that

$$\text{(i}'\text{)} \quad \left\{\omega \mid \sup_{F \in \mathscr{H}} D(F,\omega) > \alpha\right\} \subset \left\{\omega \mid \sup_{F \in \mathscr{H}} D^\nu(F,\omega) > \alpha \text{ a.a.}\right\} \cup N(\mathscr{H},\alpha),$$

$$\text{(ii}'\text{)} \quad \left\{\omega \mid \sup_{F \in \mathrm{cl}\,\mathscr{H}} D(F,\omega) < \alpha\right\} \subset \left\{\omega \mid \sup_{F \in \mathrm{cl}\,\mathscr{H}} D^\nu(F,\omega) < \alpha \text{ a.a.}\right\} \cup N(\mathscr{H},\alpha),$$

where a.a. stands, as in proposition 2.2, for "almost always".

Let $F_{\mathscr{H}}$ be a maximal element in $\mathrm{cl}\,\mathscr{H}$ with respect to inclusion ($\subset$). Because D and the D^ν are usc on $\mathscr{F}$, they attain their maximum on every closed subset of the compact space $\mathscr{F}$, and thus

$$D(F_{\mathscr{H}},\omega) = \sup_{F \in \mathrm{cl}\,\mathscr{H}} D(F,\omega)$$

and

$$D^\nu(F_{\mathscr{H}},\omega) = \sup_{F \in \mathrm{cl}\,\mathscr{H}} D^\nu(F,\omega),$$

so that (ii′) follows from the pointwise a.s. convergence of the D^ν to D at $F_{\mathscr{H}}$.

Let $\mathscr{H} = \mathscr{F}^K \cap \mathscr{F}_{G_1} \cap \cdots \cap \mathscr{F}_{G_s}$ be a nonempty element of the countable base $\mathscr{S}$. In the remarks that precede this theorem, we noted that $\mathscr{H} \neq \emptyset$ implies the existence of a sequence of strictly positive numbers $\{\epsilon_l, l \in \mathbb{N}\}$ monotonically decreasing to 0 such that for all l, the set

$$\mathscr{H}_l = \mathscr{F}^{K+\epsilon_l B} \cap \mathscr{F}_{G_1} \cap \cdots \cap \mathscr{F}_{G_s}$$

is nonempty. The sequence of sets $\{\mathscr{H}_l, l \in \mathbb{N}\}$ is contained in $\mathscr{H}$ and $\mathscr{H} = \cup_l \mathscr{H}_l$. Since for all l, $\sup_{\mathscr{H}} D^\nu(\cdot, \omega)$ is always larger than $\sup_{\mathscr{H}_l} D^\nu(\cdot, \omega)$, we have

$$\begin{aligned}\Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha\Big\} &= \Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \ \sup_{F \in \mathscr{H}} D^\nu(F, \omega) > \alpha\Big\} \\ &\quad \cup \Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \sup_{F \in \mathscr{H}} D^\nu(F, \omega) \le \alpha\Big\} \\ &\subset \Big\{\omega \mid \sup_{F \in \mathscr{H}} D^\nu(F, \omega) > \alpha\Big\} \\ &\quad \cup \Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \ \sup_{F \in \mathscr{H}_l} D^\nu(F, \omega) \le \alpha\Big\}.\end{aligned}$$

Because $(\mathscr{F}, \mathscr{T})$ is a compact space and the functions D^ν are usc, thus $\sup_{\mathscr{H}_l} D^\nu(\cdot, \omega) = \sup_{\text{cl}\,\mathscr{H}_l} D^\nu(\cdot, \omega)$ and this last supremum is attained at $A_l := \text{cl cpl}(K + \epsilon_l B)$; see the remarks that precede the theorem. Thus

$$\begin{aligned}\Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha\Big\} &\subset \Big\{\omega \mid \sup_{F \in \mathscr{H}} D^\nu(F, \omega) > \alpha\Big\} \\ &\quad \cup \Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \ D^\nu(A_l, \omega) \le \alpha\Big\}.\end{aligned}$$

Because this holds for every ν, it follows that

$$\begin{aligned}\Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha\Big\} &\subset \Big\{\omega \mid \sup_{F \in \mathscr{H}} D^\nu(F, \omega) > \alpha \text{ a.a.}\Big\} \\ &\quad \cup \Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \ D^\nu(A_l, \omega) \le \alpha \text{ a.a.}\Big\}.\end{aligned}$$

But this last set can only be a set of measure zero. Indeed, let

$$N(\mathscr{H}, \alpha) = \cup_l \{\omega \mid D^\nu(A_l, \omega) \nrightarrow D(A_l, \omega)\}.$$

This set is of μ-measure null; that follows from pointwise convergence and the fact that there are only countably many sets A_l. If ω does not belong to $N(\mathscr{H}, \alpha)$ and $\sup_{\mathscr{H}} D(\cdot, \omega) > \alpha$ that means that there exists $F \in \mathscr{H}$ such that $D(F, \omega) > \alpha$. For l sufficiently large, $F \subset A_l$ and from pointwise convergence, and the exclusion of $N(\mathscr{H}, \alpha)$, it follows that ω cannot belong to

$$\Big\{\omega \mid \sup_{F \in \mathscr{H}} D(F, \omega) > \alpha, \ D^\nu(A_l, \omega) \le \alpha \text{ a.a.}\Big\}.$$

And hence

$$\left\{\omega \mid \sup_{F\in\mathscr{K}} D(F,\omega) > \alpha\right\} \subset \left\{\omega \mid \sup_{F\in\mathscr{K}} D^{\nu}(F,\omega) > \alpha \text{ a.a.}\right\} \cup N(\mathscr{K},\alpha),$$

which completes the proof of (i′). □

We can now apply this result in the "Glivenko–Cantelli" framework.

THEOREM 4.2

For any sequence of iid random variables $\{X^{\nu}, \nu\in\mathbb{N}\}$ defined on a probability space $(\Omega, \mathscr{A}, \mu)$ with values in E, the empirical random measures,

$$P^{\nu}(B,\omega) = \frac{1}{\nu}\sum_{i=1}^{\nu} 1_{[X^{i}(\omega)\in B]}(\omega), \quad \forall B\in\mathscr{B},$$

a.s. converge narrowly to the common distribution P of these random variables. Moreover, they a.s. converge uniformly on every class of closed sets contained in cont P that is $\mathscr{T}$-bicompact.

Proof

An immediate consequence of the preceding theorem and proposition 3.4. □

The approach that we followed directs our attention to the fact that to obtain the a.s. uniform convergence of empirical measures there are two basic ingredients that enter into play. First, a condition is needed to ensure the a.s. narrow convergence. This role is played here by the iid condition. A second condition is needed to guarantee the passage from a.s. narrow convergence to uniform convergence. This is a condition that must guarantee that the class of sets under scrutiny has a certain property. We have seen here that bicompactness (with respect to the topology of set convergence) is a "natural" requirement.

References

[1] H. Attouch, *Variational Convergence for Functions and Operations* (Pitman, London, 1984).

[2] S. Dolecki, G. Salinetti and R.J.-B. Wets, Convergence of functions: equi-semicontinuity, Trans. AMS 276 (1983) 409–429.

[3] P. Gänssler and W. Stute, *Seminar on Empirical Processes* (Birkhäuser, Basel, 1987), DMV Seminar, vol. 9.

[4] R. Luccgetti, G. Salinetti, R.J.-B. Wets, Uniform convergence of probability measures: topological criteria, Manuscript, Department of Mathematics, University of California, Davis, USA (November 1988).

[5] D. Pollard, *Convergence of Stochastic Processes* (Springer, Berlin, 1984).

[6] G. Salinetti, Funzione aleatoree semicontinue e misurie aleatorie, Rendiconti del Seminario Matematico di Milano 57 (1987).

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

[8] G. Salinetti and R.J.-B. Wets, On the hypo-convergence of probability measures, *Optimization and Related Fields, Proc.*, Erice 1984, *Lecture Notes in Mathematics 1190*, eds. R. Conti, E. De Giorgi and F. Giannessi (Springer, Berlin, 1986) pp. 371–395.

[9] G. Salinetti and R.J.-B. Wets, Random semicontinuous functions, in: *Nonlinear Stochastic Processes*, ed. L.M. Ricciardi (Manchester University Press, Manchester, 1988); also, Working Paper, I.I.A.S.A. WP-86-47, Laxenburg, Austria (August 1986).

[10] G.R. Shorack and J. Wellner, *Empirical Processes and Applications to Statistics* (Wiley, New York, 1986).