

Convergence in distribution of non measurable random elements

Convergenza in distribuzione di elementi aleatori non misurabili

Patrizia Berti

Dip. di Matematica

Università di Modena e Reggio Emilia

berti.patrizia@unimo.it

Pietro Rigo

Dip. Econ. Pol. Met. Quant.

Università di Pavia

prigo@eco.unipv.it

Riassunto: Una definizione di convergenza in distribuzione per elementi aleatori non (necessariamente) misurabili, dovuta a Hoffmann-Jorgensen, viene caratterizzata in termini di convergenza debole di misure di probabilità finitamente additive.

Keywords: Asymptotic behavior of statistical procedures, Convergence in distribution, Empirical processes, Finitely additive probabilities.

1. Introduction

In some problems, the need arises of a notion of convergence in distribution for non (necessarily) measurable random elements. Examples can be found in the theory of empirical processes. Accordingly, a few definitions of convergence in distribution have been proposed. See Dudley (1966), Pyke and Shorack (1968) and Pollard (1984).

One more definition, due to Hoffmann-Jorgensen, has been recently analyzed by van der Vaart and Wellner (1996) and taken as a basis for asymptotic statistics. In fact, most asymptotic results in statistics are just consequences of convergence of empirical processes.

Despite its utility and fortune, Hoffmann-Jorgensen's definition (HJ, in what follows) does not possess a clear, intuitive probabilistic meaning. Moreover, if uncritically used, HJ can lead to subtle and annoying technical problems.

In this paper, the probabilistic content of HJ is investigated. The main result (Theorem 1) is a characterization of HJ, as well as of one of its variants, in terms of weak convergence of finitely additive probabilities (f.a.p.'s). In this way, HJ assumes a transparent probabilistic meaning.

A last, general remark is the following. There are problems in probability and statistics which can not be solved in the usual countably additive setting, while admit a finitely additive solution. Examples can be found in conditional probability, statistical inference, stochastic integration and the first digit problem. Moreover, some widely used statistical procedures, like formal posteriors of improper priors, can be justified in a finitely additive setting only. See Heath and Sudderth (1978), Scozzafava (1984), Regazzini (1987), Berti and Rigo (1994), Dubins and Prikry (1995), Berti and Rigo (1999), Berti, Regazzini and Rigo (2001). In a sense, convergence in distribution of non measurable random elements is one more example of this type. Indeed, by next Theorem 1, HJ is basically a finitely additive definition.

2. Hoffmann-Jorgensen's definition

As far as possible, the notation is that of van der Vaart and Wellner (1996). Thus, A is a directed set, $\mathcal{D}$ a metric space, $\mathcal{D}$ the Borel σ -field on $\mathcal{D}$, and $C_b(\mathcal{D})$ the class of real bounded continuous functions on $\mathcal{D}$. For all $\alpha \in A$, $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ is a probability space and $X_\alpha : \Omega_\alpha \rightarrow \mathcal{D}$ an *arbitrary* map. Moreover, $(\Omega, \mathcal{A}, P)$ is a probability space and $X : \Omega \rightarrow \mathcal{D}$ a *measurable* map, i.e., $X^{-1}(\mathcal{D}) \subset \mathcal{A}$. In the standard theory, the X_α are measurable (i.e., $X_\alpha^{-1}(\mathcal{D}) \subset \mathcal{A}_\alpha$) and “ $X_\alpha \rightarrow X$ in distribution” means that $Ef(X_\alpha) \rightarrow Ef(X)$ for all $f \in C_b(\mathcal{D})$. The point is how to generalize the standard definition to the present case, where the X_α need not be measurable.

In order to give an answer, for each $\alpha \in A$ and $T : \Omega_\alpha \rightarrow \overline{\mathbb{R}}$, let us define the outer expectation of T as $E^*T = \inf \{EU : U \geq T, U : \Omega_\alpha \rightarrow \overline{\mathbb{R}} \text{ measurable and } EU \text{ exists}\}$. Here, EU exists if at least one of EU^+ and EU^- is finite, and in that case $EU = EU^+ - EU^-$. In this notation, Hoffmann-Jorgensen's definition, HJ, is:

$$X_\alpha \rightarrow X \text{ in distribution} \Leftrightarrow E^*f(X_\alpha) \rightarrow Ef(X) \text{ for all } f \in C_b(\mathcal{D}).$$

As noted in Section 1, HJ allows a rich and elegant theory and is very handy in applied situations. However, from a probabilistic point of view, what HJ really means is not totally clear. Indeed, using outer expectations looks something like a trick for avoiding measurability problems. A related shortcoming is that, as it stands, HJ heavily depends on the underlying probability spaces $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$. This is very disturbing, in principle, since one is obliged to assume something on such spaces. For instance, in van der Vaart and Wellner (1996), the $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$ are often taken to be the canonical spaces.

3. Two finitely additive solutions

Let us first introduce some notation. Throughout, μ is the distribution of X , i.e., $\mu(B) = P(X \in B)$ for $B \in \mathcal{D}$. For $\alpha \in A$, let $\mathcal{D}_\alpha = \{B \in \mathcal{D} : X_\alpha^{-1}(B) \in \overline{\mathcal{A}_\alpha}\}$ and $\mu_\alpha(B) = \overline{P_\alpha}(X_\alpha \in B)$ for $B \in \mathcal{D}_\alpha$, where $(\Omega_\alpha, \overline{\mathcal{A}_\alpha}, \overline{P_\alpha})$ is the completion of $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$. Roughly speaking, since X_α is not necessarily measurable, its distribution can not be defined on the whole of $\mathcal{D}$; $\mathcal{D}_\alpha$ is the sub- σ -field of $\mathcal{D}$ where such distribution, which is just μ_α , can be unambiguously defined. Clearly, $\mathcal{D}_\alpha = \mathcal{D}$ whenever X_α is measurable. Further, let $\mathcal{P}_{0,\alpha}$ and $\mathcal{P}_{1,\alpha}$ denote the following classes of f.a.p.'s:

$$\begin{aligned} \mathcal{P}_{0,\alpha} &= \{\text{those f.a.p.'s } \nu \text{ on } \mathcal{D} \text{ of the form } \nu = Q_\alpha \circ X_\alpha^{-1}, \\ &\quad \text{for some f.a.p. } Q_\alpha \text{ on the power set of } \Omega_\alpha \text{ extending } \overline{P_\alpha}\}, \\ \mathcal{P}_{1,\alpha} &= \{\text{those f.a.p.'s } \nu \text{ on } \mathcal{D} \text{ such that } \nu|_{\mathcal{D}_\alpha} = \mu_\alpha \text{ and } \nu^*(X_\alpha(\Omega_\alpha)) = 1\}. \end{aligned}$$

Here, ν^* is the usual outer measure, and the request $\nu^*(X_\alpha(\Omega_\alpha)) = 1$ is to grant that ν can be further extended in such a way that the extension is supported by $X_\alpha(\Omega_\alpha)$.

Next, there is a well established theory of weak convergence of f.a.p.'s which includes the standard theory as a particular case; see Girotto and Holzer (1993). Given f.a.p.'s β_α and β on $\mathcal{D}$, $\beta_\alpha \rightarrow \beta$ weakly whenever $\beta_\alpha f \rightarrow \beta f$ for every $f \in C_b(\mathcal{D})$. If β is σ -additive, $\beta_\alpha \rightarrow \beta$ weakly if and only if $\limsup \beta_\alpha(F) \leq \beta(F)$ for all closed $F \subset \mathcal{D}$.

To motivate our results, we now discuss a classical example: the uniform empirical process. Let $\xi_1, \xi_2, \dots$ be i.i.d. real random variables, defined on $(\Omega, \mathcal{A}, P)$, with ξ_1 uniformly distributed on $[0, 1]$. The n -th empirical process $X_n(t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n I_{(\xi_i \leq t)} - t \right)$,

$t \in [0, 1]$, can be seen as a map $X_n : \Omega \rightarrow \mathcal{D} = D[0, 1]$. Such X_n is not measurable if $D[0, 1]$ is equipped with uniform topology, it is measurable if $D[0, 1]$ is equipped with Skorohod topology, and, in the latter case, $X_n \rightarrow X$ in distribution according to the standard definition. The limit X is a Brownian bridge process which can be taken, without loss of generality, as a measurable map $X : \Omega \rightarrow C[0, 1]$. Let $\mathcal{D}_U$ denote the Borel σ -field on $D[0, 1]$ under uniform topology. Non measurability of X_n when $\mathcal{D} = \mathcal{D}_U$ implies that $\mathcal{D}_n$ is strictly smaller than $\mathcal{D}_U$. Suppose for a moment that each μ_n admits a σ -additive extension ν_n to $\mathcal{D}_U$. Then, $\nu_n \rightarrow \mu$ weakly, with respect to uniform topology, according to the standard definition. In fact, if $F \subset D[0, 1]$ is closed in the uniform topology and $\overline{F}_S$ denotes its closure in the Skorohod topology, then

$$\limsup \nu_n(F) \leq \limsup \mu_n(\overline{F}_S) \leq \mu(\overline{F}_S) = \mu(\overline{F}_S \cap C[0, 1]) \leq \mu(F).$$

In other terms, X_n *morally* converges in distribution to X with respect to uniform topology. The point is that μ_n *can not be extended to $\mathcal{D}_U$ in a σ -additive way*.

Since convergence in distribution fails only because of a lack of extendibility, although it is morally true, it is natural to involve finitely additive extensions of μ_n . Indeed, the latter are always available. Furthermore, for any metric space $\mathcal{D}$, the dual of $C_b(\mathcal{D})$ is just a space of bounded finitely additive measures. And finally, the finitely additive theory of probability, based on de Finetti's coherence principle, is well developed and understood.

In any case, calling ν_n any finitely additive extension of μ_n to $\mathcal{D}_U$, the above calculation for $\limsup \nu_n(F) \leq \mu(F)$ still works, and since μ is σ -additive it follows that $\nu_n \rightarrow \mu$ weakly. To sum up, in this basic example, $\nu_n \rightarrow \mu$ *weakly whenever ν_n is a f.a.p. on $\mathcal{D}_U$ and $\nu_n = \mu_n$ on $\mathcal{D}_n$ for all n* .

The above example and remarks suggest reformulating HJ, perhaps in a more transparent way, in terms of weak convergence of f.a.p.'s. There are at least two reasonable choices:

$$X_\alpha \rightarrow X \text{ in distribution} \Leftrightarrow \nu_\alpha \rightarrow \mu \text{ weakly whenever } \nu_\alpha \in \mathcal{P}_{1,\alpha} \text{ for each } \alpha, \quad (a)$$

$$X_\alpha \rightarrow X \text{ in distribution} \Leftrightarrow \nu_\alpha \rightarrow \mu \text{ weakly whenever } \nu_\alpha \in \mathcal{P}_{0,\alpha} \text{ for each } \alpha. \quad (b)$$

Note that, since $\mathcal{P}_{0,\alpha} \subset \mathcal{P}_{1,\alpha}$, (a) implies (b). Under both (a) and (b), the underlying ideas are: (i) convergence in distribution of the X_α should solely concern their distributions; (ii) in absence of the distributions on $\mathcal{D}$ (due to non measurability), *any* finitely additive extension ν_α to $\mathcal{D}$ of the “partial distributions” μ_α can play the same role, provided $\nu_\alpha^*(X_\alpha(\Omega_\alpha)) = 1$. It is a matter of opinion, however, how many extensions to involve: is it better to extend directly the μ_α (as in (a)), or first extend the P_α and then take the image measures (as in (b)) ?

4. Results

All proofs of the results in this section are omitted, being too long for the present “reduced-version” of the paper. For each $\alpha \in A$ and real bounded Borel measurable function f on $\mathcal{D}$, define: $E^0 f(X_\alpha) = \inf\{E\phi(X_\alpha) : \phi : \mathcal{D} \rightarrow \mathbb{R} \text{ bounded and Borel measurable, } \phi(X_\alpha) \text{ is } \overline{\mathcal{A}_\alpha}\text{-measurable, } \phi(X_\alpha) \geq f(X_\alpha)\}$.

We are now able to give our main result. It states that HJ is equivalent to (b), and that (a) can be expressed, like HJ, in terms of (suitably defined) outer expectations.

Theorem 1 $X_\alpha \rightarrow X$ in distribution according to HJ if and only if $X_\alpha \rightarrow X$ in distribution according to (b). Moreover, $X_\alpha \rightarrow X$ in distribution according to (a) if and

only if $E^0 f(X_\alpha) \rightarrow Ef(X)$ for every $f \in C_b(\mathcal{D})$.

Next result gives an equivalent condition for $\mathbb{P}_{0,\alpha} = \mathbb{P}_{1,\alpha}$. Under such condition, by Theorem 1, (a), (b) and HJ are all equivalent.

Theorem 2 *Let $\nu \in \mathbb{P}_{1,\alpha}$. Then, $\nu \in \mathbb{P}_{0,\alpha}$ if and only if $\nu(B) \leq P_\alpha^*(X_\alpha \in B) \forall B \in \mathcal{D}$.*

Theorem 2 works, for instance, for the uniform empirical process X_n of Section 3.

Corollary 3 *Suppose $(\Omega_n, \mathcal{A}_n, P_n) = ([0, 1], \mathcal{B}, \lambda)^\infty$ for all n , where $\mathcal{B}$ is the Borel σ -field on $[0, 1]$ and λ the Lebesgue measure. Let ξ_n be the n -th coordinate projection, X_n the n -th empirical process, and let $\mathcal{D} = \mathcal{D}[0, 1]$ be equipped with the uniform topology. Then, $\mathbb{P}_{0,n} = \mathbb{P}_{1,n}$ for all n .*

References

- Berti, P. and Rigo, P. (1994). Coherent inferences and improper priors, *The Annals of Statistics* **22**, 1177–1194.
- Berti, P. and Rigo, P. (1999). Sufficient conditions for the existence of disintegrations, *J. of Theoretical Probability* **12**, 75–86.
- Berti, P., Regazzini, E. and Rigo, P. (2001). Integration with respect to brownian-like processes, *submitted for publication*.
- Dubins, L. and Prikrý, K. (1995). On the existence of disintegrations, *Seminaire de Probabilites XXIX, Lectures Notes in Mathematics, Springer* **1613**, 248–259.
- Dudley, R. (1966). Weak convergence of measures on nonseparable metric spaces and empirical measures on euclidean spaces, *Illinois J. of Mathematics* **10**, 109–126.
- Giroto, B. and Holzer, S. (1993). Weak convergence of masses on normal topological spaces, *Sankhya* **37**, 364–366.
- Heath, D. and Sudderth, W. (1978). On finitely additive priors, coherence, and extended admissibility, *The Annals of Statistics* **6**, 333–345.
- Pollard, D. (1984). *Convergence of stochastic processes*, Springer.
- Pyke, R. and Shorack, G. (1968). Weak convergence of a two-sample empirical process and a new approach to chernoff-savage theorems, *The Annals of Mathematical Statistics* **39**, 755–771.
- Regazzini, E. (1987). de Finetti’s coherence and statistical inference, *The Annals of Statistics* **15**, 845–864.
- Scozzafava, R. (1984). A survey on some common misunderstandings concerning the role and meaning of finitely additive probabilities in statistical inference, *Statistica* **44**, 21–45.
- van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes*, Springer.