Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 1, pp. 24-44

Finitely Additive Uniform Limit Theorems

Patrizia Berti
Universita' di Modena e Reggio-Emilia, Modena, Italy
Pietro Rigo
Universita' di Pavia, Pavia, Italy

Abstract

Some finitely additive limit theorems, which do not need a strategic setting, are proved. Let (X_n) be a sequence of random variables, $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $a_n(\cdot) = P(X_{n+1} \in \cdot \mid X_1, \ldots, X_n)$, where all probability measures (both conditional and unconditional) are assessed according to de Finetti's coherence principle. In the main result, connected with Bayesian predictive inference, conditions for $\sup_{A \in \mathcal{D}} |\mu_n(A) - a_n(A)| \to 0$ in probability are given, where $\mathcal{D}$ is any class of events. Under mild assumptions, it is also shown that $\sup_{A \in \mathcal{D}} |\mu_n(A) - \mu_m(A)| \to 0$, in probability, whenever (X_n) has stationary finite dimensional distributions. Further, asymptotic exchangeability of a certain class of sequences is proved, and this allows to extend a characterization of exchangeability due to Kallenberg (1988).

AMS (2000) *subject classification.* Primary 60A05, 60A10, 60F15.
Keywords and phrases. Coherence, empirical measure, exchangeability, finitely additive probability, predictive measure, strategic probability, uniform limit theorem.

---

1 Introduction

When dealing with sequences of random variables, three possible frameworks can be distinguished. The first (and more important) is the usual one, based on Kolmogorov's axioms, where probabilities are σ-additive, conditioning is done with respect to sub-σ-fields, and so on. It will be referred to as the *standard framework.* The second is the *strategic framework*, introduced by Dubins and Savage (1965), Dubins (1974) and developed by Purves and Sudderth (1976). Here, probabilities are allowed to be finitely additive but most of the usual regularity conditions are still available. In particular, conditional and unconditional laws are connected via disintegrability relationships. The third is the *coherent framework* where all probability assessments, both conditional and unconditional, are only asked to meet de Finetti's coherence principle. These three settings are nested, in the sense that the standard can be seen as a particular case of the strategic, which in turn is a particular case of the coherent; see Berti et al. (1997) and Purves and Sudderth (1976).

A number of limit theorems do not need the standard framework. Most examples arise in the strategic setting, where almost all classical limit theorems admit finitely additive analogs. See Chen (1977), Gangopadhyay and Rao (1997), Purves and Sudderth (1983) and Ramakrishnan (1981). Instead, very few results are available when strategicity is not assumed, a remarkable exception being Karandikar's general principle; cf. Karandikar (1982, 1988).

This paper includes some finitely additive limit theorems which hold in the coherent framework. We focus on convergence in probability and in distribution, while strong limit theorems are neglected. In fact, in the coherent setting, almost sure convergence often reduces to existence of certain extensions. A well known example is the SLLN. Let P be a finitely additive probability (f.a.p.) on the field of *finite dimensional* Borel subsets of $\mathbb{R}^\infty$. Then,

$$H = \left\{ x \in \mathbb{R}^\infty : \frac{1}{n}\sum_{i=1}^{n} x_i \text{ converges to a finite limit} \right\}$$

has inner probability 0 and outer probability 1 with respect to P. Since one is not forced to select any particular extension of P (like the σ-additive one if P is σ-additive, or the strategic one if P is strategic), H can be coherently attached any probability between 0 and 1. Other examples of this type are in Berti et al. (1997).

Our main result (Theorem 4.1 and Corollary 4.1) is the following uniform limit theorem. Let $\mathcal{D}$ be a class of subsets of a set $\mathcal{X}$, (X_n) a sequence of $\mathcal{X}$-valued random variables, $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ the empirical measure and

$$a_n(\cdot) = P(X_{n+1} \in \cdot \mid X_1, \ldots, X_n)$$

the predictive measure. Then, in order that

$$\sup_{A \in \mathcal{D}} |\mu_n(A) - a_n(A)| \to 0 \quad \text{in probability as } n \to \infty, \tag{1.1}$$

it is enough that:

(i) For large n, the "future observations" $(X_k : k > n)$ are identically distributed (at least on $\mathcal{D}$) given the past $(X_1, \ldots, X_n)$;

(ii) $\sup_{A \in \mathcal{D}} |\mu_n(A) - \mu_m(A)| \to 0$ in probability, as $n, m \to \infty$;

(iii) Certain disintegrability relationships are satisfied.

The basic condition is (i). It has been introduced, in the standard setting, in Berti et al. (2004).

Roughly speaking, (1.1) means that μ_n is a "consistent estimate" of a_n. Thus, (1.1) is potentially useful whenever the predictive measure a_n is the object of interest but cannot be calculated in closed form, mainly in Bayesian predictive inference, gambling and sequential procedures; cf. Berti et al. (2002) and Berti and Rigo (2002). We also note that (1.1) is connected with Glivenko-Cantelli-type theorems and empirical processes for dependent data.

In addition to giving conditions for (1.1), some other results are obtained. We mention Theorem 3.1, which extends a characterization of exchangeability given by Kallenberg (1988), and Theorems 4.2 and 4.3 which provide conditions for (ii). Further, Section 6 is exclusively devoted to examples.

All quoted results, being available in the coherent framework, hold in particular (usually in simpler form) in the strategic setting; see Section 5. In such a setting, among other things, strong limit theorems also make sense. In fact, in Theorem 5.1 it is shown that, for each $\delta > 0$ one has

$$\limsup_n \sup_{A \in \mathcal{D}} |\mu_n(A) - a_n(A)| \leq \delta \quad \text{a.s.,}$$

provided (i) and a strengthening of (ii) hold ((iii) is granted in the strategic framework).

2 The Coherent Framework

Throughout, Ω is a nonempty set.

2.1. Karandikar's general principle. Let P be a f.a.p. on a field $\mathcal{A}$ of subsets of Ω, X_n and X real functions on Ω, and

$$P^*(A) = \inf\{P(B) : B \in \mathcal{A}, B \supset A\}, \quad P_*(A) = 1 - P^*(A^c), \quad A \subset \Omega.$$

Say that $X_n \to X$ in probability, denoted $X_n \overset{P}{\to} X$, in case $P^*\big(|X_n - X| > \epsilon\big) \to 0$ for all $\epsilon > 0$, and that X is P-measurable in case X is the limit in probability of some sequence of $\mathcal{A}$-simple functions. Suppose now that S is a metric space, $X_n : \Omega \to S$, and ν is a f.a.p. on $\mathcal{B}(S)$. Say that X_n converges to ν in distribution in case

$$\int f(X_n)dP_0 \to \int f d\nu \quad \text{for all } f \in \mathcal{C}_b(S)$$
$$\text{and all f.a.p.'s } P_0 \text{ on the power set of } \Omega \text{ extending } P;$$

see Berti and Rigo (2004) and Karandikar (1988). If ν is regular on open sets, i.e.,

$$\nu(U) = \sup\{\nu(F) : F \text{ closed}, F \subset U\} \quad \text{for all open } U \subset S,$$

then X_n converges to ν in distribution if and only if

$$\limsup_n P^*(X_n \in F) \le \nu(F) \quad \text{for all closed } F \subset S.$$

A standard probability space is a triplet $(\Omega', \mathcal{A}', Q)$ where Ω' is a set, $\mathcal{A}'$ a σ-field on Ω' and Q a σ-additive probability on $\mathcal{A}'$. The following, useful result (cf. Karandikar, 1982, 1988) allows to turn certain statements into the standard setting.

THEOREM 2.1 (KARANDIKAR). *For each sequence (X_n) of real P-measurable functions on Ω, there are a standard probability space $(\Omega', \mathcal{A}', Q)$ and a sequence (Z_n) of real random variables on $(\Omega', \mathcal{A}', Q)$ satisfying*

$$\int f(X_1, \ldots, X_n)dP = \int f(Z_1, \ldots, Z_n)dQ$$

for all $n \ge 1$ and $f \in \mathcal{C}_b(\mathbb{R}^n)$. Moreover,

$$P_*\big((X_1, \ldots, X_n) \in B\big) = P^*\big((X_1, \ldots, X_n) \in B\big) = Q\big((Z_1, \ldots, Z_n) \in B\big)$$

for all $n \ge 1$ and $B \in \mathcal{B}(\mathbb{R}^n)$ such that $Q\big((Z_1, \ldots, Z_n) \in \partial B\big) = 0$.

2.2. The general coherent framework. Let $\mathcal{C}$ be any class of bounded functions defined on (possibly different) subsets of Ω. The members of $\mathcal{C}$ are thus of the form $X|H$, where $X : \Omega \to \mathbb{R}$ is bounded and $H \subset \Omega$, $H \neq \emptyset$. In case $H = \Omega$, we write X instead of $X|\Omega$. According to de Finetti's coherence principle, a real function E on $\mathcal{C}$ is *coherent* if, for each $n \geq 1$, $X_1|H_1, \ldots, X_n|H_n \in \mathcal{C}$ and $c_1, \ldots, c_n \in \mathbb{R}$, one has

$$\inf G|H \leq 0 \leq \sup G|H$$

where G and H are defined as

$$G = \sum_{i=1}^{n} c_i I_{H_i}\big(X_i - E(X_i \mid H_i)\big) \quad \text{and} \quad H = \bigcup_{i=1}^{n} H_i.$$

Heuristically, suppose E describes the opinions of a bookie on the elements of $\mathcal{C}$. If E is not coherent, a gambler can select a finite combination of bets (on $X_1|H_1, \ldots, X_n|H_n$ with stakes $c_1, \ldots, c_n$) which makes the bookie a sure loser. In other terms, no Dutch book against a coherent bookie is possible.

A *conditional expectation* is a coherent function E and a *conditional probability* is the restriction of a conditional expectation to indicators, i.e.,

$$P(A \mid H) = E(I_A \mid H) \quad A \subset \Omega, I_A|H \in \mathcal{C}.$$

A fundamental fact is that a coherent function E can be extended, preserving coherence, to the class of *all* functions $X|H$ with $X : \Omega \to \mathbb{R}$ bounded and H nonempty subset of Ω; see Regazzini (1985) for a proof. Moreover, if $\mathcal{A}$ is a field of subsets of Ω and $\mathcal{C} = \{I_A : A \in \mathcal{A}\}$, then E is coherent if and only if $A \mapsto E(I_A)$ is a f.a.p. on $\mathcal{A}$. Or else, if $\mathcal{C}$ is a linear space of bounded functions on Ω with $1 \in \mathcal{C}$, then E is coherent if and only if it is a linear positive functional such that $E(1) = 1$.

2.3. Sequences of random variables in the coherent framework. Consider $(\mathcal{X}, \mathcal{E})$ to be a measurable space, with $\mathcal{E}$ including the singletons, and let $X_n : \Omega \to \mathcal{X}$ for $n \geq 1$. Let us call $\mathcal{A}$ the corresponding field of finite dimensional subsets of Ω. The elements of $\mathcal{A}$ are thus of the form $\{(X_1, \ldots, X_n) \in B\}$ for some $n \geq 1$ and $B \in \mathcal{E}^n$. For $\omega \in \Omega$, define further

$$H_0(\omega) = \Omega, \quad H_n(\omega) = \{u \in \Omega : X_1(u) = X_1(\omega), \ldots, X_n(u) = X_n(\omega)\},$$

and denote by $\mathcal{L}$ the class of those bounded functions on Ω which are the uniform limit of some sequence of $\mathcal{A}$-simple functions.

For our purposes (investigating convergence in probability of quantities depending on a finite number of the X_n), a reasonable choice of $\mathcal{C}$ is

$$\mathcal{C} = \{X|H_n(\omega) : X \in \mathcal{L}, n \geq 0, \omega \in \Omega\}. \tag{2.1}$$

For such a $\mathcal{C}$, in addition, coherence admits a simple characterization. As shown in Berti et al. (1997), when $\mathcal{C}$ is given by (2.1), a real function E on $\mathcal{C}$ is coherent if and only if

(i) For each $n \geq 0$ and $\omega \in \Omega$, $E\big(\cdot \mid H_n(\omega)\big)$ is a linear positive functional on $\mathcal{L}$ such that $E\big(1 \mid H_n(\omega)\big) = E\big(I_{H_n(\omega)} \mid H_n(\omega)\big) = 1$;

(ii) $E\big(I_{H_m(\omega)} X \mid H_n(\omega)\big) = E\big(X \mid H_m(\omega)\big) E\big(I_{H_m(\omega)} \mid H_n(\omega)\big)$ whenever $X \in \mathcal{L}$, $\omega \in \Omega$ and $m > n \geq 0$.

In the sequel, with the exception of Section 5 (devoted to the strategic framework), *$\mathcal{C}$ is of the form* (2.1) *and $E : \mathcal{C} \to \mathbb{R}$ is any coherent function. We denote P the conditional probability corresponding to E, i.e.*

$$P\big(A \mid H_n(\omega)\big) = E\big(I_A \mid H_n(\omega)\big) \quad A \in \mathcal{A}, n \geq 0, \omega \in \Omega,$$

and we adopt the usual notations $E(\cdot) = E(\cdot \mid \Omega)$ and $P(\cdot) = P(\cdot \mid \Omega)$. Note that, by (i), each $P\big(\cdot \mid H_n(\omega)\big)$ is a f.a.p. on $\mathcal{A}$ with $P\big(H_n(\omega) \mid H_n(\omega)\big) = 1$.

If there is a f.a.p. P_n on the power set of Ω such that

$$E(X) = \int E\big(X \mid H_n(\omega)\big) P_n(d\omega) \quad X \in \mathcal{L},$$

then E is *disintegrable* on the partition $\{H_n(\omega) : \omega \in \Omega\}$. Generally, E can fail to be disintegrable for some (or even for all) n. This is a major difference between the coherent and the standard and strategic frameworks.

3 Asymptotic Exchangeability of Certain Sequences of Random Variables

In Berti et al. (2004), in the standard framework, the condition

$$Prob\big(X_k \in \cdot \mid X_1, \ldots, X_n\big) = Prob\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big) \text{ a.s., } k > n \geq 0 \tag{3.1}$$

is introduced. Roughly speaking (3.1) means that, at each time n, the future observations $(X_k : k > n)$ are identically distributed given the past $(X_1, \ldots, X_n)$. In the standard framework, (3.1) is equivalent to

$$(X_1, \ldots, X_n, X_k) \text{ is distributed as } (X_1, \ldots, X_n, X_{n+1}) \quad \text{for all } k > n \geq 0, \tag{3.2}$$

and it implies asymptotic exchangeability of (X_n); cf. Berti et al. (2004). Moreover, by a result of Kallenberg (1988), (X_n) is exchangeable if and only if it is stationary and meets (3.2).

Let $\mathcal{D} \subset \mathcal{E}$ and let $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ be the empirical measure. In Berti et al. (2002), under mild conditions on $(\mathcal{X}, \mathcal{E})$ and $\mathcal{D}$, it is shown that

$$\sup_{A \in \mathcal{D}} \left|\mu_n(A) - Prob\big(X_{n+1} \in A \mid X_1, \ldots, X_n\big)\right| \to 0 \quad \text{a.s.}$$

whenever (3.1) holds and μ_n converges uniformly on $\mathcal{D}$ a.s.. As pointed out in the Introduction, this result is significant whenever $Prob\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big)$ is the object of interest but cannot be evaluated in closed form, mainly in Bayesian predictive inference.

In Section 4, such a result is proved in the coherent framework, with convergence in probability in place of a.s. convergence (for the reasons explained in the Introduction). Whether the other quoted results can be extended, from the standard to the coherent setting, is investigated in the rest of this section.

In the notation of this paper, condition (3.1) becomes

$$\begin{gathered} P\big(X_k \in A \mid H_n(\omega)\big) = P\big(X_{n+1} \in A \mid H_n(\omega)\big) \\ \text{for all } k > n \geq 0, A \in \mathcal{E} \text{ and almost all } \omega \in \Omega. \end{gathered} \tag{3.3}$$

In the coherent framework, the unconditional distributions do not determine the conditional distributions. Thus, (3.2) does not imply (3.3). Conversely, (3.3) implies (3.2) if disintegrability is asked, but not in general. Instead, as we now prove, asymptotic exchangeability of sequences satisfying (3.2) and Kallenberg's result are still in force.

For $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, say that (X_n) has stationary finite dimensional distributions (f.d.d.) if, for some countable set $T \subset \mathbb{R}$, one has

$$\begin{aligned} P\big(X_1 \leq a_1, \ldots, X_n \leq a_n\big) &= P\big(X_2 \leq a_1, \ldots, X_{n+1} \leq a_n\big) \\ &\text{for all } n \geq 1 \text{ and } a_1, \ldots, a_n \in T^c. \end{aligned}$$

Likewise, (X_n) has exchangeable f.d.d. if, for some countable $T \subset \mathbb{R}$,

$$P\big(X_{j_1} \leq a_1, \ldots, X_{j_n} \leq a_n\big) = P\big(X_1 \leq a_1, \ldots, X_n \leq a_n\big)$$

for all $n \geq 1$, all permutations $(j_1, \ldots, j_n)$ of $(1, \ldots, n)$ and $a_1, \ldots, a_n \in T^c$. Finally, for applying Theorem 2.1, we ask the following condition:

$$\lim_{a \to \infty} P(|X_n| > a) = 0 \quad \text{for all } n. \tag{3.4}$$

THEOREM 3.1. *Suppose* $(\mathcal{X},\mathcal{E}) = (\mathbb{R},\mathcal{B}(\mathbb{R}))$ *and condition* (3.4) *holds. If, for some countable set* $T \subset \mathbb{R}$,

$$P\big(X_1 \le a_1,\dots,X_n \le a_n, X_k \le a\big) = P\big(X_1 \le a_1,\dots,X_n \le a_n, X_{n+1} \le a\big) \tag{3.5}$$

$$\text{for all } k > n \ge 0 \text{ and } a_1,\dots,a_n,a \in T^c,$$

then $(X_n, X_{n+1},\dots)$ *converges in distribution to* λ, *where* λ *is a* σ*-additive and exchangeable probability measure on* $\mathcal{B}(\mathbb{R}^\infty)$. *Moreover,* (X_n) *has exchangeable f.d.d. if and only if it has stationary f.d.d. and meets* (3.5) *for some countable* $T \subset \mathbb{R}$.

For proving Theorem 3.1, the following result is needed. In the sequel, if x is a point of $\mathbb{R}^\infty$ or $\mathbb{R}^m$, x_i denotes the i-th coordinate of x.

PROPOSITION 3.1. *Let* (X_n), (Z_n) *and* Q *be as in Theorem 2.1. If* $(Z_n, Z_{n+1},\dots)$ *converges in distribution to* λ, *where* λ *is a* σ*-additive probability on* $\mathcal{B}(\mathbb{R}^\infty)$, *then* $(X_n, X_{n+1},\dots)$ *converges in distribution to* λ.

PROOF. Fix $\epsilon > 0$, an open set $U \subset \mathbb{R}^\infty$, and let $\mathbb{I}$ be the class of open intervals I of the form $I = \mathbb{R}$ or $I = (a,b)$ with $Q(Z_n = a) = Q(Z_n = b) = 0$ for all n. Then, $U = \bigcup_j U_j$ where $U_j = \{x \in \mathbb{R}^\infty : x_i \in I_i^j,\, i = 1,\dots,k_j\}$ and $I_i^j \in \mathbb{I}$ for all i and j. Since λ is σ-additive, $\lambda(U) < \epsilon + \lambda\big(\bigcup_{j=1}^N U_j\big)$ for some integer N. Set $G = \bigcup_{j=1}^N U_j$ and note that $G = \{x \in \mathbb{R}^\infty : (x_1,\dots,x_k) \in G_0\}$, where $k = \max\{k_1,\dots,k_N\}$ and $G_0 = \bigcup_{j=1}^N \{(x_1,\dots,x_k) \in \mathbb{R}^k : x_i \in I_i^j,\, i = 1,\dots,k_j\}$. Since $I_i^j \in \mathbb{I}$, one obtains $Q\big((Z_{n+1},\dots,Z_{n+k}) \in \partial G_0\big) = 0$ for all n. Thus, Theorem 2.1 yields

$$\begin{aligned}
\lambda(U) - \epsilon < &\lambda(G) \le \liminf_n Q\big((Z_{n+1}, Z_{n+2},\dots) \in G\big)\\
&= \liminf_n Q\big((Z_{n+1},\dots,Z_{n+k}) \in G_0\big)\\
&= \liminf_n P_*\big((X_{n+1},\dots,X_{n+k}) \in G_0\big)\\
&\le \liminf_n P_*\big((X_{n+1}, X_{n+2},\dots) \in U\big).
\end{aligned}$$

□

PROOF OF THEOREM 3.1. Since $X_n^{-1}(\mathcal{E}) \subset \mathcal{A}$ for all n, condition (3.4) implies that the X_n are P-measurable. Therefore, there are (Z_n) and Q as in Theorem 2.1. Suppose (3.5) holds for some countable T and let

$$S = T \cup \{a \in \mathbb{R} : Q(Z_n = a) > 0 \text{ for some } n\}.$$

By (3.5) and Theorem 2.1, one obtains

$$\begin{aligned} Q\big(Z_1 \le a_1, \dots, Z_n \le a_n, Z_k \le a\big) =& P\big(X_1 \le a_1, \dots, X_n \le a_n, X_k \le a\big) \\ =& P\big(X_1 \le a_1, \dots, X_n \le a_n, X_{n+1} \le a\big) \\ =& Q\big(Z_1 \le a_1, \dots, Z_n \le a_n, Z_{n+1} \le a\big) \end{aligned}$$

for all $k > n \ge 0$ and $a_1, \dots, a_n, a \in S^c$. Since S is countable and Q σ-additive,

$$(Z_1, \dots, Z_n, Z_k) \text{ is distributed as } (Z_1, \dots, Z_n, Z_{n+1}) \quad \text{for all } k > n \ge 0.$$

By Theorem 2.5 of Berti et al. (2004), it follows that $(Z_n, Z_{n+1}, \dots)$ converges in distribution to λ, where λ is σ-additive and exchangeable. An application of Proposition 3.1 concludes the proof of the first part. As to the second one, since the "only if" part is trivial, suppose (X_n) meets (3.5) for some countable T and has stationary f.d.d. for some countable T_0. By what already proved, $(X_n, X_{n+1}, \dots)$ converges in distribution to a σ-additive and exchangeable law λ. Define

$$S = T_0 \cup \{a \in \mathbb{R} : \lambda\{x \in \mathbb{R}^\infty : x_n = a\} > 0 \text{ for some } n\}.$$

Fix $m \ge 1$ and $a_1, \dots, a_m \in S^c$. Then,

$$\begin{aligned} P\big(X_1 \le a_1, \dots, X_m \le a_m\big) =& \lim_n P\big(X_{n+1} \le a_1, \dots, X_{n+m} \le a_m\big) \\ =& \lambda\{x \in \mathbb{R}^\infty : x_1 \le a_1, \dots, x_m \le a_m\} \end{aligned}$$

where the first equality depends on (X_n) has stationary f.d.d., and the second is due to $(X_n, X_{n+1}, \dots)$ converges in distribution to λ and

$$\lambda\big(\partial\{x \in \mathbb{R}^\infty : x_1 \le a_1, \dots, x_m \le a_m\}\big) \le \sum_{i=1}^m \lambda\{x \in \mathbb{R}^\infty : x_i = a_i\} = 0.$$

Since S is countable and λ is exchangeable, it follows that (X_n) has exchangeable f.d.d.. □

4 Uniform Approximation of Predictive Measures via Empirical Measures

Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the empirical measure and

$$a_n(A, \omega) = P\big(X_{n+1} \in A \mid H_n(\omega)\big) \quad n \ge 1, A \in \mathcal{E}, \omega \in \Omega$$

the *predictive measure*. Given $\mathcal{D} \subset \mathcal{E}$, let $\|\cdot\|$ denote the sup-norm on $\mathcal{D}$. For getting

$$\|\mu_n - a_n\| = \sup_{A\in\mathcal{D}} |\mu_n(A) - a_n(A)| \xrightarrow{P} 0,$$

condition (3.3) is too strong and can be weakened into:

$$\text{For each } \epsilon > 0, \text{ there are } n_\epsilon \geq 1 \text{ and } B_\epsilon \subset \Omega \text{ such that } P^*(B_\epsilon^c) \leq \epsilon \text{ and} \tag{4.1}$$
$$\Big|P\big(X_k \in A \mid H_n(\omega)\big) - P\big(X_{n+1} \in A \mid H_n(\omega)\big)\Big| \leq \epsilon$$
$$\text{for all } k > n \geq n_\epsilon, A \in \mathcal{D} \text{ and } \omega \in B_\epsilon.$$

We are now able to state our main result.

THEOREM 4.1. *Fix* $\mathcal{D} \subset \mathcal{E}$, *define*

$$f_n(\omega) = \|\mu_n(\cdot,\omega) - \mu_{n^2}(\cdot,\omega)\| = \sup_{A\in\mathcal{D}} |\mu_n(A,\omega) - \mu_{n^2}(A,\omega)|$$

and suppose $f_n \in \mathcal{L}$ *for all* n. *If condition* (4.1) *holds and* $E\big(f_n \mid H_n(\cdot)\big) \xrightarrow{P} 0$, *then* $\|\mu_n - a_n\| \xrightarrow{P} 0$. *Moreover,* $E\big(f_n \mid H_n(\cdot)\big) \xrightarrow{P} 0$ *whenever:*

(a) $\|\mu_n - \mu_m\| \xrightarrow{P} 0$ *as* $n, m \to \infty$; *and*

(b) $E\big(f_n \mid H_n(\cdot)\big)$ *is* P*-measurable and* $E(f_n) = \int E\big(f_n \mid H_n(\omega)\big)P(d\omega)$ *for all* n.

PROOF. First note that, for all $n \geq 1$ and $A \in \mathcal{E}$, the function $\omega \mapsto \mu_n(A,\omega)$ is in $\mathcal{L}$ and $\mu_n(A,u) = \mu_n(A,\omega)$ whenever $X_i(u) = X_i(\omega)$ for $i = 1,\dots,n$. Hence, coherence implies

$$E\big(\mu_n(A) \mid H_n(\omega)\big) = \mu_n(A,\omega) \quad \text{for all } \omega \in \Omega, n \geq 1 \text{ and } A \in \mathcal{E}.$$

Fix $\epsilon > 0$ and take $n_\epsilon \geq 1$ and $B_\epsilon \subset \Omega$ according to condition (4.1). Then,

$$\|\mu_n(\cdot,\omega) - a_n(\cdot,\omega)\| = \sup_{A\in\mathcal{D}} \Big|E\big(\mu_n(A) \mid H_n(\omega)\big) - a_n(A,\omega)\Big|$$
$$\leq \sup_{A\in\mathcal{D}} \Big|E\Big(\mu_n(A) - \mu_{n^2}(A) \mid H_n(\omega)\Big)\Big| + \sup_{A\in\mathcal{D}} \Big|E\big(\mu_{n^2}(A) \mid H_n(\omega)\big) - a_n(A,\omega)\Big|$$
$$\leq E\big(f_n \mid H_n(\omega)\big) + \frac{1}{n} + \frac{1}{n^2} \sum_{i=n+1}^{n^2} \sup_{A\in\mathcal{D}} \Big|P\big(X_i \in A \mid H_n(\omega)\big) - a_n(A,\omega)\Big|$$
$$\leq E\big(f_n \mid H_n(\omega)\big) + \frac{1}{n} + \epsilon \quad \text{for all } n \geq n_\epsilon \text{ and } \omega \in B_\epsilon.$$

Since $P^*(B_\epsilon^c) \leq \epsilon$ and $E\big(f_n \mid H_n(\cdot)\big) \xrightarrow{P} 0$, one obtains

$$\begin{aligned}\limsup_n P^*\Big(\|\mu_n - a_n\| > 2\epsilon\Big) &\leq \epsilon + \limsup_n P^*\Big(B_\epsilon \cap \{\|\mu_n - a_n\| > 2\epsilon\}\Big)\\ &\leq \epsilon + \limsup_n P^*\Big(E\big(f_n \mid H_n(\cdot)\big) + \frac{1}{n} > \epsilon\Big)\\ &= \epsilon.\end{aligned}$$

This concludes the proof of the first part of the theorem. Finally, if (a)-(b) hold, then (a) implies $E(f_n) \to 0$, and thus $E\big(f_n \mid H_n(\cdot)\big) \xrightarrow{P} 0$ follows from (b). □

The preliminary condition $(f_n) \subset \mathcal{L}$ of Theorem 4.1 automatically holds as far as $\|\cdot\|$ can be defined by taking sup over a countable subclass of $\mathcal{D}$. This is actually true for most significant choices of $\mathcal{D}$, in particular when $\mathcal{D}$ is *countably determined* as defined in Berti and Rigo (2002). Condition (4.1) is a weaker version of (3.3), with essentially the same meaning (i.e., the future observations are identically distributed given the past ones). Condition (b) is a disintegrability assumption, which holds in the strategic setting, but not necessarily in the coherent one.

As to (a), we now give two sufficient conditions. The first is obtained via Karandikar's Theorem 2.1 while the second relies on elementary arguments. Also, in the first condition, the probability distribution of (X_i, X_j) is required not to assign adherent mass to the diagonal $\{(a,a) : a \in \mathbb{R}\}$, that is,

$$P(X_i = X_j) = \lim_{\epsilon \to 0} P\big(|X_i - X_j| \leq \epsilon\big) \quad \text{for all } i \neq j. \tag{4.2}$$

THEOREM 4.2. *Suppose* $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$, *and conditions* (3.4) *and* (4.2) *hold. Let* (Z_n) *and* Q *be as in Theorem 2.1 and* $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$. *Then,*

$$\|\nu_n - \nu_m\| \xrightarrow{Q} 0 \quad \Longleftrightarrow \quad \|\mu_n - \mu_m\| \xrightarrow{P} 0.$$

In particular, $\|\mu_n - \mu_m\| \xrightarrow{P} 0$ *if* (X_n) *has stationary f.d.d..*

PROOF. First note that the X_n are P-measurable (by (3.4)), and thus there are (Z_n) and Q as in Theorem 2.1. Moreover, if (X_n) has stationary f.d.d., then (Z_n) is stationary so that $\|\nu_n - \nu_m\| \xrightarrow{Q} 0$; see e.g. Proposition 5 of Berti and Rigo (2002). It remains to show that, in general, $\|\nu_n - \nu_m\| \xrightarrow{Q} 0$ if and only if $\|\mu_n - \mu_m\| \xrightarrow{P} 0$.

For fixed $m \geq 2$, let $\{B_1, \ldots, B_k\}$ be the partition of $\mathbb{R}^m$ in the sets

$$B_r = \{x \in \mathbb{R}^m : sign(x_i - x_j) = a_{i,j}(r) \text{ for all } 1 \leq i < j \leq m\} \quad r = 1, \ldots, k,$$

where $sign(u) = I_{(0,\infty)}(u) - I_{(0,\infty)}(-u)$, $u \in \mathbb{R}$, and the coefficients $a_{i,j}(r)$ take values in $\{-1, 0, 1\}$ in all possible ways. It is enough proving that

$$P\big((X_1, \ldots, X_m) \in B_r\big) = Q\big((Z_1, \ldots, Z_m) \in B_r\big) \quad r = 1, \ldots, k. \tag{4.3}$$

In fact, given $n \leq m$, define

$$\phi(x) = \sup_t \Big| \frac{1}{m} \sum_{i=1}^m I_{(-\infty,t]}(x_i) - \frac{1}{n} \sum_{i=1}^n I_{(-\infty,t]}(x_i) \Big| \quad x \in \mathbb{R}^m,$$

and note that ϕ is constant on each B_r. Thus, letting $I_\epsilon = \{r \in \{1, \ldots, k\} : \phi > \epsilon \text{ on } B_r\}$ for $\epsilon > 0$, condition (4.3) implies

$$\begin{aligned} P\big(\|\mu_n - \mu_m\| > \epsilon\big) =& P\big(\phi(X_1, \ldots, X_m) > \epsilon\big) \\ =& \sum_{r \in I_\epsilon} P\big((X_1, \ldots, X_m) \in B_r\big) \\ =& \sum_{r \in I_\epsilon} Q\big((Z_1, \ldots, Z_m) \in B_r\big) \\ =& Q\big(\phi(Z_1, \ldots, Z_m) > \epsilon\big) \\ =& Q\big(\|\nu_n - \nu_m\| > \epsilon\big). \end{aligned}$$

In order to prove (4.3), for each $\epsilon > 0$ define

$$\begin{aligned} B_r(\epsilon) = \{x \in \mathbb{R}^m : (x_i - x_j) a_{i,j}(r) > \epsilon \text{ if } a_{i,j}(r) \neq 0 \\ \text{and } |x_i - x_j| \leq \epsilon \text{ if } a_{i,j}(r) = 0, \text{ for all } 1 \leq i < j \leq m\}. \end{aligned}$$

Then, condition (4.2) yields:
Claim: For $r = 1, \ldots, k$,

$$P\big((X_1, \ldots, X_m) \in B_r\big) = \lim_{\epsilon \to 0} P\big((X_1, \ldots, X_m) \in B_r(\epsilon)\big).$$

Since (4.2) can be written as $\lim_{\epsilon \to 0} P\big(X_i \neq X_j, |X_i - X_j| \leq \epsilon\big) = 0$, the Claim is obvious for $m = 2$. For $m > 2$, it suffices considering one at time the various (i, j) with $i < j$. Let $A_1 = \{sign(X_i - X_j) = a_{i,j}(r), 1 \leq i < j \leq m, (i, j) \neq (1, 2)\}$. If $a_{1,2}(r) = -1$, say, then

$$\begin{aligned} P\big((X_1, \ldots, X_m) \in B_r\big) =& P\big(A_1, X_1 < X_2\big) \\ =& \lim_{\epsilon \to 0} P\big(A_1, X_1 < X_2, |X_1 - X_2| > \epsilon\big) \\ =& \lim_{\epsilon \to 0} P\big(A_1, X_1 + \epsilon < X_2\big). \end{aligned}$$

Having settled $(1,2)$, one more step can be taken. Let $A_2=\{sign(X_i-X_j)=a_{i,j}(r), 1\le i<j\le m, (i,j)\neq(1,2) \text{ and } (i,j)\neq(2,3)\}$. If $a_{2,3}(r)=0$, say, then

$$\begin{aligned}
&P\big((X_1,\ldots,X_m)\in B_r\big)\\
&=\lim_{\epsilon\to 0}P\big(A_2, X_2=X_3, X_1+\epsilon<X_2\big)\\
&=\lim_{\epsilon\to 0}\Big(P\big(A_2, X_1+\epsilon<X_2\big)-P\big(A_2, X_2\neq X_3, X_1+\epsilon<X_2\big)\Big)\\
&=\lim_{\epsilon\to 0}\Big(P\big(A_2, X_1+\epsilon<X_2\big)-P\big(A_2, |X_2-X_3|>\epsilon, X_1+\epsilon<X_2\big)\Big) \text{ by (4.2)}\\
&=\lim_{\epsilon\to 0}P\big(A_2, |X_2-X_3|\le\epsilon, X_1+\epsilon<X_2\big).
\end{aligned}$$

Iterating this procedure, the Claim is proved by a finite number of steps.

Finally, on noting that

$$\partial B_r(\epsilon)\subset\bigcup_{1\le i<j\le m}\{x\in\mathbb{R}^m : |x_i-x_j|=\epsilon\} \quad \text{and}$$

$$Q\big(|Z_i-Z_j|=\epsilon\big)=0 \quad \text{for all } i,j \text{ and all but countably many } \epsilon>0,$$

one obtains

$$Q\big((Z_1,\ldots,Z_m)\in\partial B_r(\epsilon)\big)=0 \quad \text{for all } r \text{ and all but countably many } \epsilon>0.$$

Hence, σ-additivity of Q, Theorem 2.1 and the Claim yield

$$\begin{aligned}
Q\big((Z_1,\ldots,Z_m)\in B_r\big)&=\lim_{\epsilon\to 0}Q\big((Z_1,\ldots,Z_m)\in B_r(\epsilon)\big)\\
&=\lim_{\epsilon\to 0}P\big((X_1,\ldots,X_m)\in B_r(\epsilon)\big)\\
&=P\big((X_1,\ldots,X_m)\in B_r\big) \quad r=1,\ldots,k.
\end{aligned}$$

□

REMARK 4.1. Condition (4.2) is granted in case $(\mathcal{X},\mathcal{E})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$, (3.4) holds and $Q(Z_i=Z_j)=0$ for $i\neq j$. In fact, for $i\neq j$,

$$\begin{aligned}
P(X_i=X_j)&\le\lim_{\epsilon\to 0}P\big(|X_i-X_j|\le\epsilon\big)\\
&=\lim_{\epsilon\to 0}Q\big(|Z_i-Z_j|\le\epsilon\big)=Q(Z_i=Z_j)=0,
\end{aligned}$$

where the first equality depends on Theorem 2.1 and the second on σ-additivity of Q.

THEOREM 4.3. *Let* $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *and* $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$. *In order to* $\|\mu_n - \mu_m\| \overset{P}{\to} 0$, *it is sufficient that* $P(X_i = X_j) = 0$ *for all* $i \neq j$ *and*

$$P\big(X_{j_1} < X_{j_2} < \ldots < X_{j_n}\big) = P\big(X_1 < X_2 < \ldots < X_n\big) \tag{4.4}$$

for all n *and all permutations* $(j_1, \ldots, j_n)$ *of* $(1, \ldots, n)$.

PROOF. Let (U_n) be an exchangeable sequence of real random variables, on some standard probability space $(\Omega_0, \mathcal{A}_0, Q_0)$, such that $Q_0(U_1 = U_2) = 0$. Given $m \geq 2$, define B_r as in the proof of Theorem 4.2. Arguing as in such a proof (see (4.3)), it is enough to see that $P\big((X_1, \ldots, X_m) \in B_r\big) = Q_0\big((U_1, \ldots, U_m) \in B_r\big)$ for all r. Indeed, by (4.4) and $P(X_i = X_j) = 0$ for $i \neq j$, either

$$P\big((X_1, \ldots, X_m) \in B_r\big) = Q_0\big((U_1, \ldots, U_m) \in B_r\big) = \frac{1}{m!}$$

(if $a_{i,j}(r) \neq 0$ for all $i < j$) or $P\big((X_1, \ldots, X_m) \in B_r\big) = Q_0\big((U_1, \ldots, U_m) \in B_r\big) = 0$ (if $a_{i,j}(r) = 0$ for some $i < j$). □

COROLLARY 4.1. *Let* $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$, *and suppose that conditions* (4.1) *and (b) of Theorem 4.1 hold. Then,* $\|\mu_n - a_n\| \overset{P}{\to} 0$ *in case* (X_n) *has stationary f.d.d. and conditions* (3.4), (4.2) *hold, or in case* $P(X_i = X_j) = 0$ *for* $i \neq j$ *and condition* (4.4) *holds.*

PROOF. Just note that $f_n \in \mathcal{L}$ for all n and apply Theorems 4.1, 4.2 and 4.3. □

5 The Strategic Framework

Some of the previous results take a simpler form when E is *strategic*, since in this case condition (b) of Theorem 4.1 (disintegrability) is granted. We briefly recall basic definitions.

Let $\Omega = \mathcal{X}^\infty$ and X_n the n-th coordinate projection on Ω, i.e.,

$$X_n(\omega) = x_n \quad \text{for all } n \geq 1 \text{ and } \omega = (x_1, x_2, \ldots) \in \mathcal{X}^\infty = \Omega.$$

Denote $\mathcal{B}$ the Borel σ-field on Ω, where $\mathcal{X}$ is given the discrete topology and Ω the corresponding product topology. Let $\mathcal{X}^*$ be the set of finite sequences of elements of $\mathcal{X}$ (including the empty sequence). A *strategy* is an indexed family $\sigma = (\sigma(p) : p \in \mathcal{X}^*)$ of f.a.p.'s on the power set of $\mathcal{X}$.

Denote σ_0 the f.a.p. corresponding to the empty sequence, i.e., $\sigma_0 = \sigma(\text{empty sequence})$. Informally, σ_0 should be viewed as the distribution of X_1, $\sigma(x)$ as the distribution of X_2 given $\{X_1 = x\}$, $\sigma(x, y)$ as the distribution of X_3 given $\{X_1 = x, X_2 = y\}$, and so on. By means of a finitely additive version of the Ionescu Tulcea theorem, a strategy σ uniquely determines a f.a.p. P_σ on $\mathcal{B}$, to be regarded as the distribution of (X_n). Such P_σ meets some nice properties, though it can fail to be countably additive. In particular, for each n and $B \subset \mathcal{X}^n$, one has

$$P_\sigma\big((X_1, \dots, X_n) \in B\big)$$
$$= \int \dots \int \sigma(x_1, \dots, x_{n-1})(Bx_1 \dots x_{n-1})\sigma(x_1, \dots, x_{n-2})(dx_{n-1}) \dots \sigma_0(dx_1)$$

where $Bx_1 \dots x_{n-1} = \{x \in \mathcal{X} : (x_1, \dots, x_{n-1}, x) \in B\}$.

We refer to Berti et al. (1997) and Section 6 of Purves and Sudderth (1976) for the connections between the present strategic framework and the coherent and standard ones, respectively. Here, we merely note that, under mild conditions on $(\mathcal{X}, \mathcal{E})$ (e.g., $\mathcal{X}$ universally measurable subset of a Polish space and $\mathcal{E} = \mathcal{B}(\mathcal{X})$), any countably additive probability P_0 on the product space $(\mathcal{X}^\infty, \mathcal{E}^\infty)$ is of the form $P_0 = P_\sigma$ for some strategy σ.

For $p \in \mathcal{X}^*$ and $z \in \mathcal{X}^*$ or $z \in \Omega$, let pz denote the sequence consisting of the elements of p followed by those of z. Starting from a strategy σ and $p \in \mathcal{X}^*$, let us define a new strategy $\sigma[p]$ as $\sigma[p](q) = \sigma(pq)$, $q \in \mathcal{X}^*$. Then, $P_{\sigma[p]}\big(Bp\big)$ can be seen as the conditional probability of $B \in \mathcal{B}$ given the past p, where $Bp = \{\omega \in \Omega : p\omega \in B\}$ and $P_{\sigma[p]}$ is the f.a.p. induced by $\sigma[p]$.

In the rest of this section, we fix a strategy σ and we let

$$E_\sigma(f) = \int f dP_\sigma \quad \text{and}$$
$$E_\sigma\big(f \mid H_n(\omega)\big) = \int f(pu)P_{\sigma[p]}(du) \quad \text{where } p = (X_1(\omega), \dots, X_n(\omega)),$$
$$\text{for all } n \geq 1, \omega \in \Omega \text{ and bounded } \mathcal{B}\text{-measurable } f : \Omega \to \mathbb{R}.$$

Note that E_σ has been defined for each f in a class much larger than $\mathcal{L}$ (even if $\mathcal{E}$ is taken to be the power set of $\mathcal{X}$). In any case, E_σ is coherent (cf. Berti et al., 1997) and

$$E_\sigma(f) = \int E_\sigma\big(f \mid H_n(\omega)\big)P_\sigma(d\omega) \quad \text{for all } n \text{ and bounded } \mathcal{B}\text{-measurable } f.$$

Since the functions $f_n = \|\mu_n - \mu_{n^2}\|$ are bounded and $\mathcal{B}$-measurable and condition (b) automatically holds, Theorem 4.1 and Corollary 4.1 reduce to:

COROLLARY 5.1. *Suppose* $E = E_\sigma$ *and denote* P *the corresponding conditional probability. In order to* $\|\mu_n - a_n\| \overset{P}{\to} 0$, *it is sufficient that conditions* (4.1) *and (a) of Theorem 4.1 hold. Moreover, in case* $\mathcal{X} = \mathbb{R}$ *and* $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$, *it is sufficient that conditions* (3.4), (4.1), (4.2) *hold and* (X_n) *has stationary f.d.d., or that conditions* (4.1), (4.4) *hold and* $P(X_i = X_j) = 0$ *for* $i \neq j$.

Once σ is assigned, P_σ is *uniquely* determined on all of $\mathcal{B}$. Thus, in the strategic setting, strong limit theorems also make sense.

THEOREM 5.1. *Suppose* $E = E_\sigma$ *and denote* P *the corresponding conditional probability. If condition* (4.1) *holds and*

$$\sup_{i,j \geq n} \|\mu_i - \mu_j\| \overset{P}{\to} 0 \quad \textit{as } n \to \infty, \tag{5.1}$$

then

$$P\big(\limsup_n \|\mu_n - a_n\| > \delta\big) = 0 \quad \textit{for all } \delta > 0.$$

PROOF. Fix $\epsilon > 0$, $k \geq 1$, and take $n_\epsilon \geq 1$ and $B_\epsilon \subset \Omega$ according to condition (4.1). For each $\omega \in B_\epsilon$ and $n \geq \max(n_\epsilon, k)$, one obtains

$$\|\mu_n(\cdot,\omega) - a_n(\cdot,\omega)\| - \frac{1}{n} - \epsilon \leq E\big(f_n \mid H_n(\omega)\big) \leq E\big(sup_{i,j\geq k}\|\mu_i - \mu_j\| \mid H_n(\omega)\big)$$

where $f_n = \|\mu_n - \mu_{n^2}\|$ and the first inequality has been proved in the proof of Theorem 4.1. Let $\phi_k = sup_{i,j\geq k}\|\mu_i - \mu_j\|$. By a version of the martingale convergence theorem available in the strategic setting (Purves and Sudderth, 1983, p. 35),

$$\limsup_n \Big| E\big(\phi_k \mid H_n(\omega)\big) - \phi_k(\omega)\Big| < \epsilon$$

for all ω in some set $B_k \in \mathcal{B}$ with $P(B_k) = 1$. Therefore,

$$\limsup_n \|\mu_n(\cdot,\omega) - a_n(\cdot,\omega)\| \leq \limsup_n E\big(\phi_k \mid H_n(\omega)\big) + \epsilon \leq \phi_k(\omega) + 2\epsilon$$

whenever $\omega \in B_\epsilon \cap B_k$. It follows that

$$\begin{aligned} P\big(\limsup_n \|\mu_n - a_n\| > \delta\big) &\leq \epsilon + P^*\big(B_\epsilon \cap B_k, \limsup_n \|\mu_n - a_n\| > \delta\big) \\ &\leq \epsilon + P\big(2\epsilon + \phi_k > \delta\big) \quad \text{for each } \delta > 0. \end{aligned}$$

Finally, fix $\delta > 0$ and take $\epsilon \in (0, \frac{\delta}{2})$. Since $\delta - 2\epsilon > 0$, condition (5.1) implies $P\big(2\epsilon + \phi_k > \delta\big) \to 0$, as $k \to \infty$, and this concludes the proof. □

In connection with Theorem 5.1, we recall that in a finitely additive (even if strategic) setting it may be that $P\big(|f| > \delta\big) = 0$, for some function f and all $\delta > 0$, while $P(f \neq 0) > 0$.

6 Examples

In all the following examples, $(\mathcal{X}, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$.

Example 6.1 (Orbits). Given $\lambda \in (0,1)$, define $h(x) = x + \lambda$ or $h(x) = x + \lambda - 1$ according to whether $x \in [0, 1-\lambda)$ or $x \in [1-\lambda, 1)$. Suppose $0 \leq X_1 < 1$, $X_n = h(X_{n-1})$ for $n \geq 2$ and X_1 has "uniform distribution on the rationals of $[0,1]$", that is, $P(X_1 \in \mathbb{Q}) = 1$ and $P(X_1 \leq t) = t$ for $t \in [0,1]$. Since $P(X_1 \leq a) = P(h(X_1) \leq a)$ for all $a \in \mathbb{R}$, (X_n) has stationary f.d.d.. For fixed $i \neq j$, either $X_i = X_j$ or $|X_i - X_j| \geq \delta_{ij}$ for some constant $\delta_{ij} > 0$. Therefore, condition (4.2) holds while (3.4) is trivially true. Thus, Theorem 4.2 yields $\|\mu_n - \mu_m\| \overset{P}{\to} 0$.

In Example 6.1, X_1 could be the proportion of white balls in an urn, where both the number of white balls and the total number of balls are unknown. If no more information is available, it is natural to assess $P(X_1 \leq t) = t$ for each $t \in [0,1]$. In the standard framework, however, this has the disturbing consequence that $P(X_1 \in \mathbb{Q}) = 0$ although X_1 takes rational values only. One merit of the coherent setting is overcoming this kind of drawback. A similar remark holds for Example 6.5 below.

Example 6.2 (Finitely additive mixtures of σ-additive i.i.d. laws). For each θ in some index set Θ, let P_θ be a σ-additive probability on the finite dimensional Borel subsets of $\mathbb{R}^\infty$ such that the coordinate projections are i.i.d. under P_θ and $P_\theta\{x \in \mathbb{R}^\infty : x_1 = t\} = 0$ for all t. Given a f.a.p. π on the power set of Θ, suppose that

$$P\big((X_1, \ldots, X_n) \in B\big) = \int P_\theta\{x \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B\} \pi(d\theta) \tag{6.1}$$

for all $n \geq 1$ and $B \in \mathcal{B}(\mathbb{R}^n)$. For instance, the distribution of the sequence

$$X_n = U + V_n$$

can be written as in (6.1) provided (V_n) is i.i.d., according to a σ-additive diffuse law, and U is independent of (V_n) with a finitely additive distribution.

In any case, since $P(X_i = X_j) = 0$ for $i \neq j$ and condition (4.4) holds, Theorem 4.3 implies $\|\mu_n - \mu_m\| \xrightarrow{P} 0$.

In the remaining examples, the X_n are the coordinate projections on $\Omega = \mathbb{R}^\infty$.

EXAMPLE 6.3 (GLIVENKO-CANTELLI THEOREM). Results on $\|\mu_n - a_n\|$ are conceived mainly for non independent sequences, but apply in particular to the independent case. Suppose (X_n) is i.i.d., that is, $E = E_\sigma$ for some strategy σ satisfying $\sigma(p) = \sigma_0$ for all finite sequences p. Define $F(t) = \sigma_0((-\infty, t])$, $t \in \mathbb{R}$, and

$$D_n = \|\mu_n - a_n\| = \sup_t \Big| \mu_n((-\infty, t]) - F(t) \Big|.$$

Since $D_n \geq \max(F(-\infty), 1 - F(+\infty))$, one has $F(-\infty) = 1 - F(+\infty) = 0$ in case $D_n \xrightarrow{P} 0$. Likewise, $D_n \xrightarrow{P} 0$ implies $F(t+) - F(t-) = \sigma_0\{t\}$ for all t. Conversely, by Corollary 5.1, if $F(-\infty) = 1 - F(+\infty) = 0$ and $F(t+) - F(t-) = \sigma_0\{t\}$ for all t, then $D_n \xrightarrow{P} 0$. In fact, (X_n) has stationary f.d.d., (4.1) trivially holds, and the assumptions on F imply (3.4) and (4.2). As to (4.2), for all $i \neq j$:

$$\begin{aligned} P(X_i = X_j) &= \int \sigma_0\{t\}\sigma_0(dt) \\ &= \int (F(t+) - F(t-))dF(t) \\ &= \lim_{\epsilon \to 0} \int (F(t+\epsilon) - F(t-\epsilon))dF(t) \\ &= \lim_{\epsilon \to 0} P\big(|X_i - X_j| \leq \epsilon\big). \end{aligned}$$

By a result in Chen and Ramakrishnan (1983), the previous conditions on F are also equivalent to $P\big(D_n \to 0\big) = 1$. To our knowledge, however, $D_n \xrightarrow{P} 0$ is not a consequence of the latter fact.

EXAMPLE 6.4 (HILL'S A_n-MODELS). According to Hill (1968, 1993), in case of vague a priori knowledge, a set of reasonable assumptions is:

(i) $P(X_i = X_j) = 0$ for all $i \neq j$;

(ii) (X_n) has exchangeable f.d.d. (with $T = \emptyset$; cf. Section 3);

(iii) $P\big(X_{n+1} \in I \mid H_n(\omega)\big) = \frac{1}{n+1}$ whenever $X_1(\omega), \ldots, X_n(\omega)$ are all distinct, I denoting anyone of the $n+1$ *open* intervals spanned by $X_1(\omega), \ldots, X_n(\omega)$.

Conditions (i)-(ii)-(iii) can be realized, in the strategic framework, by

$$\sigma_0((-\infty,t)) = \sigma_0((t,+\infty)) = \frac{1}{2} \quad \text{for all } t \in \mathbb{R}$$

$$\sigma(x_1,\ldots,x_n) = \frac{1}{n+1}\Big(\sigma_0 + \sum_{i=1}^{n} d_{x_i}\Big)$$

where d_x is any f.a.p. satisfying $d_x((x-\epsilon,x)) = d_x((x,x+\epsilon)) = \frac{1}{2}$ for all $\epsilon > 0$; see Hill (1993). Now, if $E = E_\sigma$ for such a strategy σ, then (4.1) and (4.4) are satisfied. In fact, a direct calculation shows that

$$P\big(X_k \leq t \mid H_n(\omega)\big) = P\big(X_{n+1} \leq t \mid H_n(\omega)\big)$$

$$P\big(X_{j_1} < X_{j_2} < \ldots < X_{j_n}\big) = \frac{1}{n!}$$

for all $k > n \geq 1$, $t \in \mathbb{R}$, $\omega \in \Omega$, and all permutations $(j_1,\ldots,j_n)$ of $(1,\ldots,n)$. Thus, Corollary 5.1 yields $\|\mu_n - a_n\| \overset{P}{\to} 0$.

EXAMPLE 6.5 (EXCHANGEABLE "GAUSSIAN PROCESS ON THE RATIONALS"). Suppose $E = E_\sigma$ and the strategy σ meets $\sigma(p)(\mathbb{Q}) = 1$ for all finite sequences p. Then, Theorem 2.1 of Chen (1977) yields

$$P\big(X_n \in \mathbb{Q} \text{ for all } n\big) = 1.$$

Given $\rho \in (0,1)$, suppose also that

$$\sigma_0((-\infty,t]) = \Phi(t), \quad \sigma(x_1,\ldots,x_n)((-\infty,t]) = \Phi\left(\frac{t - \rho'_{\mathbf{n}}\mathbf{\Sigma}_{\mathbf{n}}^{-\mathbf{1}}\mathbf{x}}{\big(1-\rho'_{\mathbf{n}}\mathbf{\Sigma}_{\mathbf{n}}^{-\mathbf{1}}\rho_{\mathbf{n}}\big)^{1/2}}\right)$$

for all $t, x_1,\ldots,x_n \in \mathbb{R}$, where Φ is the standard normal distribution function, $\mathbf{x}' = (x_1,\ldots,x_n)$, $\rho'_{\mathbf{n}} = (\rho,\rho,\ldots,\rho)$ and $\mathbf{\Sigma}_{\mathbf{n}}$ is the $n \times n$ matrix with diagonal elements equal to 1 and all the other elements equal to ρ. It follows that the joint distribution function of $(X_1,\ldots,X_n)$ coincides with that of a normal n-dimensional law with null mean vector and covariance matrix $\mathbf{\Sigma}_{\mathbf{n}}$. Thus, (X_n) has exchangeable f.d.d., (3.4) holds, and (4.2) follows from Remark 4.1. It is also straightforward to see that

$$P\big(X_k \leq t \mid H_n(\omega)\big) = P\big(X_{n+1} \leq t \mid H_n(\omega)\big) \text{ for all } k > n \geq 1, t \in \mathbb{R}, \omega \in \Omega.$$

Hence, condition (4.1) holds, too, and Corollary 5.1 implies $\|\mu_n - a_n\| \overset{P}{\to} 0$.

EXAMPLE 6.6 ("GAUSSIAN PROCESS ON THE RATIONALS" REVISITED). Let E_0 be a coherent function on $\mathcal{C}$. Taking E_0 as a starting point, a number of new conditional expectations can be obtained as follows: Select a proper subclass $\mathcal{C}_0 \subset \mathcal{C}$, put $E = E_0$ on $\mathcal{C}_0$, and then coherently extend E to $\mathcal{C}$.

With this program in mind, let us take $E_0 = E_\sigma$ where σ is the strategy of the previous Example 6.5. Denoting P_0 the conditional probability corresponding to E_0, let E be any coherent function on $\mathcal{C}$ such that

$$P\big(X_{j_1} < \ldots < X_{j_k} \mid H_n(\omega)\big) = P_0\big(X_{j_1} < \ldots < X_{j_k} \mid H_n(\omega)\big) \qquad (6.2)$$

$$P\big(X_k \leq \cdot \mid H_n(\omega)\big) = P_0\big(X_{n+1} \leq \cdot \mid H_n(\omega)\big) \quad \text{and} \quad E(f) = E_0(f)$$

for all $k > n \geq 0$, $\omega \in \Omega$, all permutations $(j_1, \ldots, j_k)$ of $(1, \ldots, k)$, and all functions f of the type

$$f(\omega) = P_0\big(X_{j_1} < \ldots < X_{j_k} \mid H_n(\omega)\big).$$

Roughly speaking, E only retains from E_σ the conditional previsions of the next observation, the previsions of orderings (both conditional and unconditional) and the mean of the latter. In view of Theorem 4.1, however, it is still true that $\|\mu_n - a_n\| \overset{P}{\to} 0$. Let us check conditions (4.1), (a) and (b) for E. Condition (4.1) holds by definition of E. As to (a), it suffices noting that $\|\mu_n - \mu_m\|$ is constant on sets of the form $\{X_{j_1} < \ldots < X_{j_k}\}$ where $k = \max(n, m)$; cf. proof of Theorem 4.2. Thus, $\|\mu_n - \mu_m\| \overset{P}{\to} 0$ follows from $\|\mu_n - \mu_m\| \overset{P_0}{\to} 0$ and $P\big(X_{j_1} < \ldots < X_{j_k}\big) = P_0\big(X_{j_1} < \ldots < X_{j_k}\big)$ for all permutations $(j_1, \ldots, j_k)$. Finally, we prove (b). Fix n. For each permutation $\pi = (\pi_1, \ldots, \pi_{n^2})$ of $(1, \ldots, n^2)$, denote $h_\pi(\omega) = P_0\big(X_{\pi_1} < \ldots < X_{\pi_{n^2}} \mid H_n(\omega)\big)$ and let c_π be the (constant) value of $f_n = \|\mu_n - \mu_{n^2}\|$ on the set $\{X_{\pi_1} < \ldots < X_{\pi_{n^2}}\}$. By (6.2) (applied with $n = 0$), $E(f_n) = E_0(f_n)$. By disintegrability of $E_0 = E_\sigma$, one also has $P_0\big(X_{\pi_1} < \ldots < X_{\pi_{n^2}}\big) = E_0(h_\pi)$. Hence,

$$\begin{aligned} E(f_n) = E_0(f_n) &= \sum_\pi c_\pi P_0\big(X_{\pi_1} < \ldots < X_{\pi_{n^2}}\big) \\ &= \sum_\pi c_\pi E_0(h_\pi) = \sum_\pi c_\pi E(h_\pi) = E\Big(\sum_\pi c_\pi h_\pi\Big). \end{aligned}$$

Since $\sum_\pi c_\pi h_\pi(\omega) = E\big(f_n \mid H_n(\omega)\big)$ for all ω, condition (b) holds.

References

Berti, P., Regazzini, E. and Rigo, P. (1997). Well-calibrated, coherent forecasting systems. *Theory Probab. Appl.*, **42**, 82–102.

Berti, P., Mattei, A., and Rigo, P. (2002). Uniform convergence of empirical and predictive measures. *Atti Sem. Mat. Fis. Univ. Modena*, **L**, 465–477.

Berti, P. and Rigo, P. (2002). A uniform limit theorem for predictive distributions. *Statist. Probab. Lett.*, **56**, 113–120.

Berti, P. and Rigo, P. (2004). Convergence in distribution of non measurable random elements. *Ann. Probab.*, **32**, 365–379.

Berti, P., Pratelli, L. and Rigo, P. (2004). Limit theorems for a class of identically distributed random variables. *Ann. Probab.*, **32**, 2029–2052.

Chen, R. (1977). On almost sure convergence in a finitely additive setting. *Z. Wahrsch. Verw. Gebiete*, **37**, 341–356.

Chen, R. and Ramakrishnan, S. (1983). On the Glivenko-Cantelli theorem in a finitely additive setting. *Sankhyā, Ser. A*, **45**, 161–167.

Dubins, L.E. and Savage, L.J. (1965). *How to gamble if you must: Inequalities for stochastic processes.* McGraw Hill, New York.

Dubins, L.E. (1974). On Lebesgue-like extensions of finitely additive measures. *Ann. Probab.*, **2**, 456–463.

Gangopadhyay, S. and Rao, B.V. (1997) Some finitely additive probability: Random walks. *J. Theoret. Probab.*, **10**, 643–657.

Hill, B. (1968). Posterior distribution of percentiles: Bayes theorem for sampling from a population. *J. Amer. Statist. Assoc.*, **63**, 677–691.

Hill, B. (1993). Parametric models for A_n: Splitting processes and mixtures. *J. Roy. Statist. Soc., Ser. B*, **55**, 423–433.

Kallenberg, O. (1988). Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.*, **16**, 508–534.

Karandikar, R.L. (1982). A general principle for limit theorems in finitely additive probability. *Trans. Amer. Math. Soc.*, **273**, 541–550.

Karandikar, R.L. (1988). A general principle for limit theorems in finitely additive probability: the dependent case. *J. Multivariate Anal.*, **24**, 189–206.

Purves, R.A. and Sudderth, W.D. (1976). Some finitely additive probability. *Ann. Probab.*, **4**, 259–276.

Purves, R.A. and Sudderth, W.D. (1983). Finitely additive zero-one laws. *Sankhyā, Ser. A*, **45**, 32–37.

Ramakrishnan, S. (1981). Finitely additive Markov chains. *Trans. Amer. Math. Soc.*, **265**, 247–272.

Regazzini, E. (1985). Finitely additive conditional probabilities. *Rend. Sem. Mat. Fis. Milano* **55**, 69-89. Corrections in *Rend. Sem. Mat. Fis. Milano*, **57** (1987) 599.

Patrizia Berti
Dipartimento di Matematica
Pura ed Applicata "G.Vitali"
Universita' di Modena e Reggio-Emilia
via Campi 213/B,41100 Modena,Italy
E-mail: berti.patrizia@unimore.it

Pietro Rigo
Dipartimento di Economia
Politica e Metodi Quantitativi
Universita' di Pavia
via S.Felice 5, 27100 Pavia, Italy
E-mail: prigo@eco.unipv.it

Paper received January 2005; revised January 2006.