RATE OF CONVERGENCE OF EMPIRICAL MEASURES FOR EXCHANGEABLE SEQUENCES

PATRIZIA BERTI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. Let S be a finite set, (X_n) an exchangeable sequence of S-valued random variables, and $\mu_n = (1/n)\sum_{i=1}^n \delta_{X_i}$ the empirical measure. Then, $\mu_n(B) \overset{a.s.}{\longrightarrow} \mu(B)$ for all $B \subset S$ and some (essentially unique) random probability measure μ. Denote by $\mathcal{L}(Z)$ the probability distribution of any random variable Z. Under some assumptions on $\mathcal{L}(\mu)$, it is shown that

$$\frac{a}{n} \le \rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(\mu)\big] \le \frac{b}{n} \quad \text{and} \quad \rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(a_n)\big] \le \frac{c}{n^u}$$

where ρ is the bounded Lipschitz metric and $a_n(\cdot) = P\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big)$ is the predictive measure. The constants $a,\, b,\, c > 0$ and $u \in (\frac{1}{2}, 1]$ depend on $\mathcal{L}(\mu)$ and card$\,(S)$ only.

1. Introduction

In the sequel, $(\Omega, \mathcal{A}, P)$ is a probability space and $(S, \mathcal{B})$ a measurable space. Also, $\mathcal{P}$ is the set of probability measures on $\mathcal{B}$ and Σ the σ-field on $\mathcal{P}$ generated by the maps $\nu \in \mathcal{P} \mapsto \nu(B)$ for all $B \in \mathcal{B}$. A *random probability measure* (r.p.m.) on $\mathcal{B}$ is a measurable map $\mu : (\Omega, \mathcal{A}) \to (\mathcal{P}, \Sigma)$.

Denote by $\mathcal{L}(Z)$ the probability distribution of any random variable Z. If μ is a r.p.m. on $\mathcal{B}$, thus, $\mathcal{L}(\mu)$ is the probability measure on Σ defined by

$$\mathcal{L}(\mu)(C) = P\big\{\omega : \mu(\omega) \in C\big\} \quad \text{for all } C \in \Sigma.$$

Two remarkable r.p.m.'s are as follows. Suppose $(S, \mathcal{B})$ is nice, say S a Borel subset of a Polish space and $\mathcal{B}$ the Borel σ-field on S. Fix a sequence

$$X = (X_n : n \ge 1)$$

of S-valued random variables on $(\Omega, \mathcal{A}, P)$. Then,

$$\mu_n = (1/n)\sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad a_n(\cdot) = P\big(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\big)$$

are r.p.m.'s on $\mathcal{B}$. Here, δ_x denotes the unit mass at x and a_n is meant as a regular version of the conditional distribution of X_{n+1} given $\sigma(X_1, \ldots, X_n)$. Usually, μ_n is called the *empirical measure* and a_n the *predictive measure*.

Suppose now that X is exchangeable. Then, a third significant r.p.m. on $\mathcal{B}$ is

$$\mu(\cdot) = P\big(X_1 \in \cdot \mid \tau\big)$$

2010 *Mathematics Subject Classification.* 60G09, 60B10, 60A10, 60G57, 62F15.

Key words and phrases. Empirical measure – Exchangeability – Predictive measure – Random probability measure – Rate of convergence.

where τ is the tail σ-field of X. In a sense, μ is the limit of both μ_n and a_n. In fact, for fixed $B \in \mathcal{B}$, one obtains

$$\mu_n(B) \xrightarrow{a.s.} \mu(B) \quad \text{and} \quad a_n(B) \xrightarrow{a.s.} \mu(B).$$

There are reasons, both theoretical and practical, to estimate how μ_n is close to a_n for large n. Similarly, it is useful to contrast μ_n and μ, as well as a_n and μ, for large n. We refer to [1]-[8] for such reasons. Here, we mention two different approaches for comparing r.p.m.'s.

Let α and β be r.p.m.'s on $\mathcal{B}$ and let $\mathcal{Q}$ denote the set of probability measures on Σ. To contrast α and β, one can select a (separable) distance d on $\mathcal{P}$ and focus on the random variable $\omega \mapsto d\big[\alpha(\omega), \beta(\omega)\big]$. Or else, one can evaluate $\rho\big[\mathcal{L}(\alpha),\, \mathcal{L}(\beta)\big]$ for some distance ρ on $\mathcal{Q}$.

Both the approaches make sense and are worthy to be developed. The first, perhaps more natural, has been followed in [1]-[5] and [7]. In such papers, the asymptotic behavior of the sequence $r_n\, d(\mu_n, a_n)$ is investigated for suitable constants r_n. The distance d is taken to be the bounded Lipschitz metric, the Wasserstein distance, or the uniform distance on a subclass $\mathcal{D} \subset \mathcal{B}$. The constants r_n are to determine the rate of convergence of the random variables $d(\mu_n, a_n)$. For instance, in [5, Corollary 3], it is shown that

$$\limsup_n \sqrt{\frac{n}{\log\log n}} \,\sup_{B\in\mathcal{D}} |\mu_n(B) - a_n(B)| \leq \sqrt{2 \sup_{B\in\mathcal{D}} \mu(B)\,(1-\mu(B))} \quad \text{a.s.}$$

under mild conditions on $\mathcal{D} \subset \mathcal{B}$. Since the right-hand member is finite (it is actually bounded by $1/\sqrt{2}$) it follows that

$$r_n \sup_{B\in\mathcal{D}} |\mu_n(B) - a_n(B)| \xrightarrow{a.s.} 0 \quad \text{whenever} \quad r_n \sqrt{\frac{\log\log n}{n}} \to 0.$$

The main reason for the second approach is that, in most real problems, the meaningful objects are the probability distributions of random variables, rather than the random variables themselves. In Bayesian nonparametrics, for instance, $\mathcal{L}(\mu)$ is the prior distribution, one of the basic ingredients of the problem. Thus, the rate at which $\mathcal{L}(\mu)$ can be approximated by $\mathcal{L}(\mu_n)$, or by $\mathcal{L}(a_n)$, is certainly of interest. Similarly, it is useful to estimate the distance between $\mathcal{L}(\mu_n)$ and $\mathcal{L}(a_n)$. This approach has been carried on in [9]-[10] and partially in [5].

In this paper, inspired by [10], we take the second point of view. Our goal is to compare $\mathcal{L}(\mu_n)$ with $\mathcal{L}(\mu)$ and $\mathcal{L}(\mu_n)$ with $\mathcal{L}(a_n)$. The sequence X is exchangeable and the distance ρ on $\mathcal{Q}$ is the bounded Lipschitz metric. Furthermore, as a preliminary but significant step, we focus on the special case where S is *finite*. Indeed, to our knowledge, all existing results concerning the second approach refer to $S = \{0, 1\}$; see [5], [9] and [10].

Our main results can be summarized as follows. To fix ideas, let $S = \{0, 1, \ldots, d\}$. If $\big(\mu\{1\}, \ldots, \mu\{d\}\big)$ admits a *suitable* density with respect to Lebesgue measure, then

$$\frac{a}{n} \leq \rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(\mu)\big] \leq \frac{b}{n} \quad \text{and} \quad \rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(a_n)\big] \leq \frac{c}{n^u}.$$

The constants a, b, $c > 0$ and $u \in (\frac{1}{2}, 1]$ depend on $\mathcal{L}(\mu)$ and d only.

2. Notation and basic assumptions

Let (T, d) be a metric space. Given two Borel probability measures on T, say ν and γ, the *bounded Lipschitz metric* is

$$\rho(\nu, \gamma) = \sup_f |\nu(f) - \gamma(f)|$$

where sup is over those functions $f : T \to [-1, 1]$ such that $|f(x) - f(y)| \le d(x, y)$ for all $x, y \in T$. Note that ρ agrees with the Wasserstein distance whenever T is separable and $d \le 1$.

A function $f : T \to \mathbb{R}$ is *Holder continuous* if there are constants $\delta \in (0, 1]$ and $r \in [0, \infty)$ such that $|f(x) - f(y)| \le r\, d(x, y)^\delta$ for all $x, y \in T$. In that case, δ and r are called the exponent and the Holder constant, respectively. If $\delta = 1$, f is also said to be a *Lipschitz* function.

Let $BV[0, 1]$ be the set of real functions on $[0, 1]$ with bounded variation. For $f \in BV[0, 1]$, we denote by ν_f the Borel measure on $[0, 1]$ such that

$$\nu_f[x, y] = f(y+) - f(x-) \quad \text{for all } 0 \le x \le y \le 1$$

where $f(0-) = f(0)$ and $f(1+) = f(1)$. As usual, $|\nu_f|$ is the total variation measure of ν_f. Note that, if f is absolutely continuous on $[0, 1]$, then $f \in BV[0, 1]$ and $|\nu_f|$ has density $|f'|$ with respect to Lebesgue measure.

In the remainder of this paper, S is a finite set, $\mathcal{B}$ the power set of S, and $X = (X_n : n \ge 1)$ an *exchangeable* sequence of S-valued random variables on the probability space $(\Omega, \mathcal{A}, P)$. Exchangeability means that $\big(X_{\pi_1}, \ldots, X_{\pi_n}\big)$ is distributed as $\big(X_1, \ldots, X_n\big)$ for all $n \ge 1$ and all permutations $(\pi_1, \ldots, \pi_n)$ of $(1, \ldots, n)$. Since X is exchangeable and $(S, \mathcal{B})$ is nice, de Finetti's theorem yields

$$P(X \in A) = \int \mu(\omega)^\infty(A)\, P(d\omega) \quad \text{for } A \in \mathcal{B}^\infty$$

where μ is the r.p.m. on $\mathcal{B}$ introduced in Section 1 and $\mu^\infty = \mu \times \mu \times \ldots$

For definiteness (and without loss of generality) we take S to be

$$S = \{0, 1, \ldots, d\}.$$

We also adopt the following notation

$$V_j = \limsup_n \mu_n\{j\} \quad \text{for each } j \in S,$$
$$\overline{X}_n = \big(\mu_n\{1\}, \ldots, \mu_n\{d\}\big) \quad \text{and} \quad V = \big(V_1, \ldots, V_d\big),$$
$$W_n = \big(E(V_1 \mid \mathcal{G}_n), \ldots, E(V_d \mid \mathcal{G}_n)\big) \quad \text{where } \mathcal{G}_n = \sigma(X_1, \ldots, X_n).$$

Since X is exchangeable, $\mu\{j\} = V_j$ and $a_n\{j\} = E\big\{\mu\{j\} \mid \mathcal{G}_n\big\} = E(V_j \mid \mathcal{G}_n)$ a.s. Therefore, V and W_n can be regarded as

$$V = \big(\mu\{1\}, \ldots, \mu\{d\}\big) \quad \text{and} \quad W_n = \big(a_n\{1\}, \ldots, a_n\{d\}\big) \quad \text{a.s.}$$

Recall that $\mathcal{Q}$ is the collection of probability measures on Σ. The map

$$\nu \in \mathcal{P} \mapsto \big(\nu\{1\}, \ldots, \nu\{d\}\big)$$

is a bijection from $\mathcal{P}$ onto

$$I = \big\{x \in \mathbb{R}^d : x_i \ge 0 \text{ for all } i \text{ and } \sum_{i=1}^d x_i \le 1\big\}.$$

Thus, $\mathcal{P}$ can be identified with I and $\mathcal{Q}$ with the set of Borel probabilities on I. More precisely, let $\mathcal{P}$ be equipped with the distance

$$d(\nu,\gamma)=\sqrt{\sum_{j=1}^{d}(\nu\{j\}-\gamma\{j\})^2}$$

and let ρ be the corresponding bounded Lipschitz metric on $\mathcal{Q}$. Define also

$$L=\big\{\phi\in\mathbb{R}^I:-1\leq\phi(x)\leq 1\ \text{ and }\ |\phi(x)-\phi(y)|\leq\|x-y\|\text{ for all }x,\,y\in I\big\}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. Then, it is not hard to see that

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]=\sup_{\phi\in L}\Big|E\big\{\phi(\overline{X}_n)\big\}-E\big\{\phi(V)\big\}\Big|\quad\text{and}$$

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(a_n)\big]=\sup_{\phi\in L}\Big|E\big\{\phi(\overline{X}_n)\big\}-E\big\{\phi(W_n)\big\}\Big|.$$

In the sequel, ρ is as above, namely, *ρ is the bounded Lipschitz metric on $\mathcal{Q}$.*

3. μ_n versus μ

Because of exchangeability, conditionally on V, the sequence X is i.i.d. with

$$P(X_1=j\mid V)=V_j\quad\text{a.s.}$$

In particular,

$$E\big(\mu_n\{j\}\big)=E\big\{E\big(\mu_n\{j\}\mid V\big)\big\}=E(V_j)\quad\text{and}$$

$$E\big(\mu_n^2\{j\}\big)=E\big\{E\big(\mu_n^2\{j\}\mid V\big)\big\}=E\Big\{V_j^2+\frac{V_j\,(1-V_j)}{n}\Big\}=E(V_j^2)+\frac{E\big\{V_j\,(1-V_j)\big\}}{n}.$$

Letting $\phi_j(x)=x_j^2-x_j$ for all $x\in I$, it follows that

$$E\big\{\phi_j(\overline{X}_n)\big\}-E\big\{\phi_j(V)\big\}=E\big(\mu_n^2\{j\}\big)-E\big(\mu_n\{j\}\big)-E(V_j^2)+E(V_j)=\frac{E\big\{V_j\,(1-V_j)\big\}}{n}.$$

On noting that $\phi_j\in L$ for all j, one obtains

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]\geq\max_j\Big(E\big\{\phi_j(\overline{X}_n)\big\}-E\big\{\phi_j(V)\big\}\Big)=\frac{a}{n}$$
$$\text{where}\quad a=\max_j\,E\big\{V_j\,(1-V_j)\big\}.$$

This provides a lower bound for $\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]$. In fact, a is strictly positive apart from trivial situations. The rest of this section is devoted to the search of an upper bound, possibly of order $1/n$.

We begin with $d=1$. In this case, $S=\{0,1\}$ and

$$\overline{X}_n=\mu_n\{1\}=(1/n)\sum_{i=1}^{n}X_i\quad\text{and}\quad V=V_1=\mu\{1\}\text{ a.s.}$$

Theorem 1. *Let $S=\{0,1\}$ and X exchangeable. Suppose $\mu\{1\}$ admits a density f, with respect to Lebesgue measure on $[0,1]$, and $f\in BV[0,1]$. Then,*

$$\frac{a}{n}\leq\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]\leq\frac{b}{n+1}$$

for each $n \geq 1$, where

$$b = 1 + \frac{1}{2}\int_0^1 x\,(1-x)\,|\nu_f|(dx).$$

(The measure $|\nu_f|$ has been defined in Section 2).

Proof. Note that $I = [0,1]$ and fix $\phi \in L$. Since $\phi_n \to \phi$ uniformly, for some sequence $\phi_n \in L \cap C^1$, it can be assumed $\phi \in L \cap C^1$. Define

$$\Phi(x) = \int_0^x \phi(t)\,dt \quad \text{and} \quad B_n(x) = \sum_{j=0}^n \Phi(\frac{j}{n}) \begin{pmatrix} n \\ j \end{pmatrix} x^j (1-x)^{n-j}.$$

On noting that

$$B'_{n+1}(x) = \sum_{j=0}^n \begin{pmatrix} n \\ j \end{pmatrix} (n+1) \Big\{ \Phi(\frac{j+1}{n+1}) - \Phi(\frac{j}{n+1}) \Big\} x^j (1-x)^{n-j}$$

$$\text{and} \quad \Big| (n+1) \Big\{ \Phi(\frac{j+1}{n+1}) - \Phi(\frac{j}{n+1}) \Big\} - \phi(\frac{j}{n}) \Big| \leq \frac{1}{n+1},$$

one obtains

$$\int_0^1 \Big| B'_{n+1}(x) - \sum_{j=0}^n \phi(\frac{j}{n}) \begin{pmatrix} n \\ j \end{pmatrix} x^j (1-x)^{n-j} \Big| f(x)\,dx \leq \frac{1}{n+1}.$$

Since X is exchangeable, it follows that

$$E\big\{\phi(\overline{X}_n)\big\} - E\big\{\phi(V)\big\} = \sum_{j=0}^n \phi(\frac{j}{n})\, P\big(n\,\overline{X}_n = j\big) - E\big\{\phi(V)\big\}$$

$$= \sum_{j=0}^n \phi(\frac{j}{n}) \begin{pmatrix} n \\ j \end{pmatrix} \int_0^1 x^j (1-x)^{n-j} f(x)\,dx - \int_0^1 \phi(x)\; f(x)\,dx$$

$$\leq \frac{1}{n+1} + \int_0^1 \big\{ B'_{n+1}(x) - \Phi'(x) \big\} f(x)\,dx.$$

Since $B_{n+1}(0) = \Phi(0)$ and $B_{n+1}(1) = \Phi(1)$, an integration by parts yields

$$\int_0^1 \big\{ B'_{n+1}(x) - \Phi'(x) \big\} f(x)\,dx = -\int_0^1 \big\{ B_{n+1}(x) - \Phi(x) \big\} \nu_f(dx).$$

Observe now that $B_{n+1}(x) = E_x\big\{\Phi(\overline{X}_{n+1})\big\}$, where E_x denotes expectation under the probability measure P_x which makes the sequence X i.i.d. with $P_x(X_1 = 1) = x$. Hence, since $E_x(\overline{X}_{n+1}) = x$, Taylor formula yields

$$B_{n+1}(x) - \Phi(x) = E_x\Big\{ \Phi(\overline{X}_{n+1}) - \Phi(x) \Big\} = (1/2)\, E_x\Big\{ (\overline{X}_{n+1} - x)^2 \phi'(Z_n) \Big\}$$

for some random variable Z_n. Since $|\phi'| \leq 1$ (due to $\phi \in L \cap C^1$) then

$$|B_{n+1}(x) - \Phi(x)| \leq (1/2)\, E_x\big\{ (\overline{X}_{n+1} - x)^2 \big\} = \frac{x\,(1-x)}{2\,(n+1)}.$$

Collecting all these facts together, one finally obtains

$$E\big\{\phi(\overline{X}_n)\big\} - E\big\{\phi(V)\big\} \leq \frac{1}{n+1} + \frac{1}{2(n+1)} \int_0^1 x\,(1-x)\,|\nu_f|(dx) = \frac{b}{n+1}.$$

This concludes the proof.

□

In view of Theorem 1, the rate of $\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]$ is $1/n$ provided $S=\{0,1\}$ and $\mu\{1\}$ admits a suitable density with respect to Lebesgue measure. This fact, however, is essentially known. In fact, by [10, Theorem 1.2],

$$\frac{a}{n}\le\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]\le\frac{b^*}{n}$$

for a suitable constant b^* whenever $S=\{0,1\}$ and $\mu\{1\}$ has a smooth density f such that $\int_0^1 x\,(1-x)\,|f'(x)|\,dx<\infty$. Note that, if f is absolutely continuous on $[0,1]$, then

$$\int_0^1 x\,(1-x)\,|f'(x)|\,dx=\int_0^1 x\,(1-x)\,|\nu_f|(dx).$$

Theorem 1 has been actually suggested by [10, Theorem 1.2]. With respect to the latter, however, Theorem 1 has two little merits. Its proof is remarkably shorter (and possibly more direct) and it often provides a smaller upper bound. For instance,

$$b^*=2\,b\,-\,1\,+\,\frac{3}{\sqrt{2\,\pi\,e}}\,+\,\int_0^1\big\{\,|1-2x|+x^2+(1-x)^2\big\}\,f(x)\,dx$$

in case f is absolutely continuous on $[0,1]$.

We next turn to the general case $d\ge 1$. Indeed, under an independence assumption, a version of Theorem 1 is available.

Theorem 2. *Let $S=\{0,1,\ldots,d\}$ and X exchangeable. Suppose $V_0>0$ a.s. and define*

$$R_j=\frac{V_j}{\sum_{i=0}^{j}V_i}.$$

Suppose also that $R_1,\ldots,R_d$ are independent, each R_j (with $j>0$) admits a density f_j with respect to Lebesgue measure on $[0,1]$, and $f_j\in BV[0,1]$. Then,

$$\frac{a}{n}\le\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big]\le\frac{b}{n+1}$$

for all $n\ge 1$, where

$$b=1\,+\,\sqrt{2}\,(d-1)\,+\,\frac{1}{\sqrt{2}}\sum_{j=1}^{d}\int_0^1 x\,(1-x)\,|\nu_{f_j}|(dx).$$

(The measure $|\nu_{f_j}|$ has been defined in Section 2).

Proof. Since $1/2<1/\sqrt{2}$, the theorem holds true for $d=1$. The general case follows from an induction on d. Here, we deal with $d=2$. The inductive step (from $d-1$ to d) can be processed in exactly the same way, but the notation becomes awful. Hence, the explicit calculations are omitted.

Let $d=2$ and let L^* be the set of functions $\psi:[0,1]\to[-1,1]$ such that $|\psi(x)-\psi(y)|\le|x-y|$ for all $x,\,y\in[0,1]$ (namely, L^* is the class L corresponding to $d=1$). Fix $\phi\in L$ and define

$$h(y)=\int_0^1\phi\big(x\,(1-y),\,y\big)\,f_1(x)\,dx.$$

Since R_1 is independent of R_2 and $V_2 = R_2$ a.s.,

$$E\{\phi(V)\} = E\Big\{\phi\big(R_1\,(1-V_2),\,V_2\big)\Big\} = \int_0^1\int_0^1 \phi\big(x\,(1-y),\,y\big)\,f_1(x)\,dx\,f_2(y)\,dy$$
$$= \int_0^1 h(y)\,f_2(y)\,dy = E\{h(V_2)\}.$$

On the other hand, $\phi \in L$ implies

$$|h(y)-h(z)| \le \int_0^1 |\phi\big(x\,(1-y),\,y\big) - \phi\big(x\,(1-z),\,z\big)|\,f_1(x)\,dx$$
$$\le |y-z|\int_0^1 \sqrt{x^2+1}\,f_1(x)\,dx \le \sqrt{2}\,|y-z|.$$

Hence, $h/\sqrt{2} \in L^*$. By this fact and $f_2 \in BV[0,1]$, Theorem 1 yields

$$|E\{\phi(V)\} - E\{h(\mu_n\{2\})\}| = |E\{h(V_2)\} - E\{h(\mu_n\{2\})\}|$$
$$\le \frac{\sqrt{2}}{n+1}\Big\{1 + \frac{1}{2}\int_0^1 x\,(1-x)\,|\nu_{f_2}|(dx)\Big\}.$$

Next, define

$$m_{j,k} = E\{R_1^k(1-R_1)^{n-j-k}\} = \int_0^1 x^k(1-x)^{n-j-k}\,f_1(x)\,dx$$

and note that

$$E\{V_1^kV_2^j(1-V_1-V_2)^{n-j-k}\} = E\{R_1^k(1-R_1)^{n-j-k}V_2^j(1-V_2)^{n-j}\}$$
$$= E\{R_1^k(1-R_1)^{n-j-k}\}\,E\{V_2^j(1-V_2)^{n-j}\} = m_{j,k}\,E\{V_2^j(1-V_2)^{n-j}\}.$$

Hence, $E\{\phi(\overline{X}_n)\}$ can be written as

$$E\{\phi(\overline{X}_n)\} = \sum_{j=0}^{n}\sum_{k=0}^{n-j}\phi(\frac{k}{n},\,\frac{j}{n})\,P\big(n\,\overline{X}_n = (k,\,j)\big) = \phi(0,1)\,P\big(\mu_n\{2\}=1\big) +$$
$$+\sum_{j=0}^{n-1}\sum_{k=0}^{n-j}\phi(\frac{k}{n-j}\,(1-\frac{j}{n}),\,\frac{j}{n})\binom{n}{j}\binom{n-j}{k}\,m_{j,k}\int_0^1 y^j(1-y)^{n-j}\,f_2(y)\,dy.$$

Similarly,

$$E\{h(\mu_n\{2\})\} = \sum_{j=0}^{n} h(\frac{j}{n})\,P\big(n\,\mu_n\{2\} = j\big) = \phi(0,1)\,P\big(\mu_n\{2\}=1\big) +$$
$$+\sum_{j=0}^{n-1}\binom{n}{j}\int_0^1 \phi\big(x\,(1-\frac{j}{n}),\,\frac{j}{n}\big)\,f_1(x)\,dx\int_0^1 y^j(1-y)^{n-j}\,f_2(y)\,dy.$$

For $j < n$, define

$$\phi_j(x) = \frac{n}{n-j}\Big\{\phi\big(x\,(1-\frac{j}{n}),\,\frac{j}{n}\big) - \phi\big(0,\,\frac{j}{n}\big)\Big\}.$$

Since $f_1 \in BV[0,1]$ and $\phi_j \in L^*$ for each $j < n$, Theorem 1 implies again

$$\Big|\sum_{k=0}^{n-j} \binom{n-j}{k} \phi\big(\frac{k}{n-j}(1-\frac{j}{n}),\, \frac{j}{n}\big)\, m_{j,k} - \int_0^1 \phi\big(x\,(1-\frac{j}{n}),\, \frac{j}{n}\big)\, f_1(x)\, dx\Big|$$

$$= \frac{n-j}{n}\,\Big|\sum_{k=0}^{n-j} \binom{n-j}{k} \phi_j\big(\frac{k}{n-j}\big)\, m_{j,k} - \int_0^1 \phi_j(x)\, f_1(x)\, dx\Big|$$

$$\le \frac{n-j}{n\,(n-j+1)}\,\Big\{1 + \frac{1}{2}\int_0^1 x\,(1-x)\,|\nu_{f_1}|(dx)\Big\}.$$

On noting that $\frac{n-j}{n\,(n-j+1)} \le \frac{1}{n+1}$, the previous inequality yields

$$|E\{h(\mu_n\{2\})\} - E\{\phi(\overline{X}_n)\}| \le \frac{1}{n+1}\Big\{1 + \frac{1}{2}\int_0^1 x\,(1-x)\,|\nu_{f_1}|(dx)\Big\}.$$

This concludes the proof. In fact,

$$|E\{\phi(V)\} - E\{\phi(\overline{X}_n)\}| \le |E\{\phi(V)\} - E\{h(\mu_n\{2\})\}| + |E\{h(\mu_n\{2\})\} - E\{\phi(\overline{X}_n)\}|$$

$$\le \frac{1}{n+1}\Big\{1 + \sqrt{2} + \frac{1}{\sqrt{2}}\sum_{j=1}^{2}\int_0^1 x\,(1-x)\,|\nu_{f_j}|(dx)\Big\} = \frac{b}{n+1}.$$

□

A simple real situation where the assumptions of Theorem 2 make sense is as follows.

Example 3. Let $\{Y_{n,j} : n \ge 1,\ 0 \le j < d\}$ be an array of random indicators. Define $T_n = \min\{j : Y_{n,j} = 1\}$, with the convention $\min\emptyset = d$, and $X_n = d - T_n$. Fix a random vector $R = (R_1, \ldots, R_d)$ satisfying $0 < R_j < 1$ for all j, and suppose that, conditionally on R, the $Y_{n,j}$ are independent with $P(Y_{n,j} = 1 \mid R) = R_{d-j}$ a.s. Then, by construction, $X = (X_1, X_2, \ldots)$ is exchangeable and

$$P\big(X_n = j \mid R\big) = R_j \prod_{i=j+1}^{d} (1 - R_i) \quad \text{a.s., where } R_0 = 1.$$

Thus, to realize the assumptions of Theorem 2, it suffices to take $R_1, \ldots, R_d$ independent with each R_j having a density $f_j \in BV[0,1]$.

A last remark is that Theorems 1-2 can be generalized. Roughly speaking, to get certain upper bounds (larger than those obtained so far) it suffices that $R_1, \ldots, R_d$ are independent but admit BV-densities only *locally*. Here is an example. We focus on $d = 1$ for the sake of simplicity, but analogous results are available for any $d \ge 1$.

Theorem 4. *Let $S = \{0,1\}$ and X exchangeable. Suppose that*

$$P\big(\mu\{1\} \in A \cap B\big) = \int_A f(x)\, dx \quad \textit{for each Borel set } A \subset [0,1],$$

where $B \subset [0,1]$ is a fixed Borel set and f satisfies $f \ge 0$ and $f \in BV[0,1]$. Then,

$$\frac{a}{n} \le \rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(\mu)\big] \le \frac{P\big(\mu\{1\} \notin B\big)}{2\sqrt{n}} + \frac{b}{n+1}$$

for each $n \geq 1$, with

$$b = P\big(\mu\{1\} \in B\big) \, + \, \frac{1}{2} \int_0^1 x\,(1-x)\,|\nu_f|(dx).$$

(The measure $|\nu_f|$ has been defined in Section 2).

Proof. Fix $\phi \in L \cap C^1$ and recall that $V = \mu\{1\}$ a.s. (for $d = 1$). Then,

$$E\Big\{\big(\phi(\overline{X}_n) - \phi(V)\big)\, I_{B^c}(V)\Big\} \leq E\Big\{|\overline{X}_n - V|\, I_{B^c}(V)\Big\}$$

$$= E\Big\{I_{B^c}(V)\, E\big\{|\overline{X}_n - V| \mid V\big\}\Big\} \leq E\Big\{I_{B^c}(V)\,\sqrt{E\big\{(\overline{X}_n - V)^2 \mid V\big\}}\Big\}$$

$$= E\Big\{I_{B^c}(V)\,\sqrt{\frac{V\,(1-V)}{n}}\Big\} \leq \frac{P\big(V \notin B\big)}{2\,\sqrt{n}}$$

where the last inequality depends on $V\,(1-V) \leq 1/4$. Next, note that $E\big\{g(V)\,I_B(V)\big\} = \int_0^1 g(x)\,f(x)\,dx$ for each bounded Borel function g. Hence,

$$P\big(n\,\overline{X}_n = j,\, V \in B\big) = E\Big\{I_B(V)\,P\big(n\,\overline{X}_n = j \mid V\big)\Big\}$$

$$= \binom{n}{j}\, E\Big\{V^j(1-V)^{n-j}\, I_B(V)\Big\} = \binom{n}{j} \int_0^1 x^j(1-x)^{n-j} f(x)\,dx.$$

It follows that

$$E\Big\{\big(\phi(\overline{X}_n) - \phi(V)\big)\, I_B(V)\Big\} = \sum_{j=0}^{n} \phi(\frac{j}{n})\, P\big(n\,\overline{X}_n = j,\, V \in B\big) - E\big\{\phi(V)\, I_B(V)\big\}$$

$$= \sum_{j=0}^{n} \phi(\frac{j}{n}) \binom{n}{j} \int_0^1 x^j(1-x)^{n-j} f(x)\,dx \, - \, \int_0^1 \phi(x)\, f(x)\,dx.$$

Thus, the argument exploited in the proof of Theorem 1 can be replicated. Since $\int_0^1 f(x)\,dx = P\big(V \in B\big)$, one finally obtains

$$E\Big\{\big(\phi(\overline{X}_n) - \phi(V)\big)\, I_B(V)\Big\} \leq \frac{P\big(V \in B\big)}{n+1} \, + \, \frac{1}{2(n+1)} \int_0^1 x\,(1-x)\,|\nu_f|(dx) = \frac{b}{n+1}.$$

□

The rate provided by Theorem 4, namely $n^{-1/2}$, is available without any assumption on $\mu\{1\}$; see [10, Proposition 3.1]. Sometimes, however, Theorem 4 allows to get a better rate. As highlighted by (the second part of) the next example, the idea is to take B depending on n.

Example 5. Let $B = [s, t]$, $f = 0$ on B^c and f monotone on B. Then,

$$\int_0^1 x\,(1-x)\,|\nu_f|(dx) \leq \frac{1}{2}\, \max\big\{f(s),\, f(t)\big\}$$

and Theorem 4 yields

$$\rho\big[\mathcal{L}(\mu_n),\, \mathcal{L}(\mu)\big] \leq \frac{P\big(\mu\{1\} \notin B\big)}{2\,\sqrt{n}} \, + \, \frac{P\big(\mu\{1\} \in B\big) \, + \, (1/4)\, \max\big\{f(s),\, f(t)\big\}}{n+1}.$$

Suppose in fact f increasing and $0 < s < t < 1$ (all other cases can be treated in exactly the same manner). Since $f \geq 0$ and $f(t+) = 0$,

$$|\nu_f|\{t\} = |\nu_f\{t\}| = |f(t+) - f(t-)| = f(t-).$$

Similarly, $|\nu_f|\{s\} = f(s+)$. Therefore,

$$\int_0^1 x\,(1-x)\,|\nu_f|(dx) \leq \frac{1}{4}\,|\nu_f|[s,t] = \frac{f(t-) - f(s+) + |\nu_f|\{s,t\}}{4} = \frac{f(t-)}{2} \leq \frac{f(t)}{2}.$$

This simple bound works nicely in some practical situations. For instance, in [10, Proposition 5.1], it is shown that

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big] = \mathrm{O}\big(n^{-\frac{1+u}{2}}\big)$$

if $\mu\{1\}$ has a density g of the form $g(x) = k\,\big(x - \frac{1}{2}\big)^{u-1}\,I_{(\frac{1}{2},\,\frac{3}{4})}(x)$, with $u \in (0,1)$ and k a normalizing constant. Such result can be easily proved using the bound obtained above. Fix in fact $n > 16$ and define

$$B = \Big[\frac{1}{2} + \frac{1}{\sqrt{n}}\,,\,\frac{3}{4}\Big] \quad \text{and} \quad f = g\,I_B.$$

With such B and f, one obtains

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(\mu)\big] \leq \frac{P\big(\mu\{1\} < \frac{1}{2} + \frac{1}{\sqrt{n}}\big)}{2\,\sqrt{n}} + \frac{1\,+\,(1/4)\,f(\frac{1}{2} + \frac{1}{\sqrt{n}})}{n+1} \leq q\,n^{-\frac{1+u}{2}}$$

for some constant q.

4. μ_n versus a_n

In this section, it is convenient to work on the coordinate space. Accordingly, we let $(\Omega,\mathcal{A}) = (S^\infty,\mathcal{B}^\infty)$ and we take X_n to be the n-th canonical projection on $\Omega = S^\infty$. Recall also that $\mathcal{G}_n = \sigma(X_1,\ldots,X_n)$.

Our last result provides a (sharp) upper bound for $\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(a_n)\big]$.

Theorem 6. *Let $S = \{0,1,\ldots,d\}$ and X exchangeable. Suppose $\big(\mu\{1\},\ldots,\mu\{d\}\big)$ admits a density f, with respect to Lebesgue measure on I, and f is Holder continuous. Then,*

$$\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(a_n)\big] \leq \frac{2\,d}{n+d+1} + 2\,r\,\frac{\sqrt{d}}{d!}\,\Big(\frac{d^2}{(d+1)\,(d+2)}\Big)^{\frac{1+\delta}{2}}\,\Big(\frac{1}{n}\Big)^{\frac{1+\delta}{2}}$$

for each $n \geq 1$, where $\delta \in (0,1]$ and $r \in [0,\infty)$ are the exponent and the Holder constant of f, respectively.

Proof. Let Q be the probability measure on $\mathcal{A}$ which makes X exchangeable with

$$E_Q\big(V_j \mid \mathcal{G}_n\big) = Q\big(X_{n+1} = j \mid \mathcal{G}_n\big) = \frac{1 + n\,\mu_n\{j\}}{n+d+1}, \quad Q\text{-a.s.}$$

Then,

$$\Big|\,\mu_n\{j\} - E_Q\big(V_j \mid \mathcal{G}_n\big)\,\Big| \leq \frac{1 + d\,\mu_n\{j\}}{n+d+1}.$$

Fix $n \geq 1$, $x_1, \ldots, x_n \in S$, and set $k_j = \sum_{i=1}^n \delta_j\{x_i\}$. Since V is uniformly distributed on I under Q,

$$P\big(X_1 = x_1, \ldots, X_n = x_n\big) = \int_I \big(1 - \sum_{i=1}^d u_i\big)^{k_0} \prod_{i=1}^d u_i^{k_i}\, f(u_1, \ldots, u_d)\, du_1, \ldots, du_d$$
$$= \frac{1}{d!} \int_\Omega \big(1 - \sum_{i=1}^d V_i\big)^{k_0} \prod_{i=1}^d V_i^{k_i}\, f(V)\, dQ$$
$$= \frac{1}{d!} \int_{\{X_1 = x_1, \ldots, X_n = x_n\}} f(V)\, dQ.$$

Hence, P has density $f(V)/d!$ with respect to Q. In particular,

$$E(V_j \mid \mathcal{G}_n) = \frac{E_Q\big\{V_j\, f(V) \mid \mathcal{G}_n\big\}}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} \quad \text{a.s.}$$

Further, each V_j has a beta distribution, with parameters 1 and d, under Q. Hence,

$$E_Q\big\{\|V - \overline{X}_n\|^2\big\} = \sum_{j=1}^d E_Q\big\{(V_j - \mu_n\{j\})^2\big\} = \sum_{j=1}^d E_Q\Big\{\frac{V_j\,(1 - V_j)}{n}\Big\} = \frac{d^2}{(d+1)\,(d+2)}\,\frac{1}{n}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$.

Next, define $U_n = f(V) - E_Q\big\{f(V) \mid \mathcal{G}_n\big\}$. Then,

$$E_Q\big(\mu_n\{j\}\, U_n \mid \mathcal{G}_n\big) = \mu_n\{j\}\, E_Q(U_n \mid \mathcal{G}_n) = 0, \quad Q\text{-a.s.}$$

Since $P \ll Q$, then $E_Q\big(\mu_n\{j\}\, U_n \mid \mathcal{G}_n\big) = 0$ a.s. with respect to P as well. Hence,

$$|\mu_n\{j\} - E(V_j \mid \mathcal{G}_n)| - \frac{1 + d\,\mu_n\{j\}}{n + d + 1} \leq |E_Q(V_j \mid \mathcal{G}_n) - E(V_j \mid \mathcal{G}_n)|$$
$$= \Big| E_Q(V_j \mid \mathcal{G}_n) - \frac{E_Q\big\{V_j\, f(V) \mid \mathcal{G}_n\big\}}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} \Big|$$
$$= \frac{|\,E_Q\big\{V_j\, U_n \mid \mathcal{G}_n\big\}\,|}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} = \frac{|\,E_Q\big\{(V_j - \mu_n\{j\})\, U_n \mid \mathcal{G}_n\big\}\,|}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}}$$
$$\leq \frac{E_Q\big\{|(V_j - \mu_n\{j\})\, U_n| \mid \mathcal{G}_n\big\}}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} \quad \text{a.s.}$$

Since f is Holder continuous and $\delta \leq 1$, one also obtains

$$E_Q(U_n^2) = E_Q\Big\{\Big(f(V) - f(\overline{X}_n) - E_Q\big\{f(V) - f(\overline{X}_n) \mid \mathcal{G}_n\big\}\Big)^2\Big\}$$
$$\leq 4\, E_Q\big\{(f(V) - f(\overline{X}_n))^2\big\} \leq 4\, r^2\, E_Q\big\{\|V - \overline{X}_n\|^{2\delta}\big\} \leq 4\, r^2 \Big(E_Q\big\{\|V - \overline{X}_n\|^2\big\}\Big)^\delta.$$

Finally, for $x \in \mathbb{R}^d$, define $\|x\|^* = \sum_{j=1}^d |x_j|$ and recall that $\|x\| \leq \|x\|^* \leq \sqrt{d}\, \|x\|$. Given $\phi \in L$, to conclude the proof, it suffices to note that

$$
\begin{aligned}
E\big\{\phi(\overline{X}_n)\big\} - E\big\{\phi(W_n)\big\} &\le E|\phi(\overline{X}_n) - \phi(W_n)| \\
&\le E\|\overline{X}_n - W_n\| \le E\|\overline{X}_n - W_n\|^* \\
&\le \sum_{j=1}^d E\Big\{ \frac{1 + d\,\mu_n\{j\}}{n+d+1} + \frac{E_Q\big\{|(\mu_n\{j\} - V_j)\,U_n| \mid \mathcal{G}_n\big\}}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} \Big\} \\
&= \frac{d + d\,P(X_1 \neq 0)}{n+d+1} + \sum_{j=1}^d E_Q\Big\{ \frac{f(V)}{d!}\,\frac{E_Q\big\{|(\mu_n\{j\} - V_j)\,U_n| \mid \mathcal{G}_n\big\}}{E_Q\big\{f(V) \mid \mathcal{G}_n\big\}} \Big\} \\
&\le \frac{2\,d}{n+d+1} + \frac{1}{d!}\,E_Q\Big\{|U_n|\,\|\overline{X}_n - V\|^*\Big\} \\
&\le \frac{2\,d}{n+d+1} + \frac{\sqrt{d}}{d!}\,E_Q\Big\{|U_n|\,\|\overline{X}_n - V\|\Big\} \\
&\le \frac{2\,d}{n+d+1} + \frac{\sqrt{d}}{d!}\,\sqrt{E_Q\big\{\|\overline{X}_n - V\|^2\big\}\,E_Q(U_n^2)} \\
&\le \frac{2\,d}{n+d+1} + 2\,r\,\frac{\sqrt{d}}{d!}\,\Big(E_Q\big\{\|\overline{X}_n - V\|^2\big\}\Big)^{\frac{1+\delta}{2}} \\
&= \frac{2\,d}{n+d+1} + 2\,r\,\frac{\sqrt{d}}{d!}\,\Big(\frac{d^2}{(d+1)\,(d+2)}\Big)^{\frac{1+\delta}{2}}\Big(\frac{1}{n}\Big)^{\frac{1+\delta}{2}}.
\end{aligned}
$$

□

Theorem 6 improves a previous result which covers the particular case where $d = 1$ and $\delta = 1$; see [5, Theorem 10].

The rate provided by Theorem 6 can not be improved. Take in fact $P = Q$ with Q as in the proof of Theorem 6. Then, V is uniformly distributed on I (so that f is even constant) and $a_n\{1\} = \frac{1+n\,\mu_n\{1\}}{n+d+1}$ a.s. Take also $\phi(x) = \frac{x_1^2}{2}$ for all $x \in I$. Since $\phi \in L$, one obtains

$$
\begin{aligned}
2\,(n+d+1)\,\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(a_n)\big] &\ge 2\,(n+d+1)\,\Big(E\big\{\phi(\overline{X}_n)\big\} - E\big\{\phi(W_n)\big\}\Big) \\
&= (n+d+1)\,\Big\{E\big(\mu_n\{1\}^2\big) - E\big(a_n\{1\}^2\big)\Big\} \\
&= (n+d+1)\,E\big(\mu_n\{1\}^2\big) - \frac{1 + n^2 E\big(\mu_n\{1\}^2\big) + 2\,n\,E\big(\mu_n\{1\}\big)}{n+d+1}
\end{aligned}
$$

$$
\ge \frac{2\,(d+1)\,n\,E\big(\mu_n\{1\}^2\big) - 2\,n\,E\big(\mu_n\{1\}\big) - 1}{n+d+1} \longrightarrow 2\,(d+1)\,E(V_1^2) - 2\,E(V_1) = \frac{2\,d}{(d+1)\,(d+2)}.
$$

Therefore, in this case, the rate of $\rho\big[\mathcal{L}(\mu_n),\,\mathcal{L}(a_n)\big]$ is actually $1/n$.

References

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

[2] Berti P., Crimaldi I., Pratelli L., Rigo P. (2009) Rate of convergence of predictive distributions for dependent data, *Bernoulli*, 15, 1351-1367.

[3] Berti P., Pratelli L., Rigo P. (2012) Limit theorems for empirical processes based on dependent data, *Electron. J. Probab.*, 17, 1-18.

[4] Berti P., Pratelli L., Rigo P. (2013) Exchangeable sequences driven by an absolutely continuous random measure, *Ann. Probab.*, 41, 2090-2102.

[5] Berti P., Pratelli L., Rigo P. (2016) Asymptotic predictive inference with exchangeable data, *Submitted*, currently available at http://www-dimat.unipv.it/ rigo/smps2016.pdf

[6] Cifarelli D.M., Regazzini E. (1996) De Finetti's contribution to probability and statistics, *Statist. Science*, 11, 253-282.

[7] Cifarelli D.M., Dolera E., Regazzini E. (2016) Frequentistic approximations to Bayesian prevision of exchangeable random elements, *Submitted*, arXiv:1602.01269v1.

[8] Fortini S., Ladelli L., Regazzini E. (2000) Exchangeability, predictive distributions and parametric models, *Sankhya* A, 62, 86-109.

[9] Goldstein L., Reinert G. (2013) Stein's method for the Beta distribution and the Polya-Eggenberger urn, *J. Appl. Probab.*, 50, 1187-1205.

[10] Mijoule G., Peccati G., Swan Y. (2016) On the rate of convergence in de Finetti's representation theorem, *Submitted*, arXiv:1601.06606v1.

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 address: patrizia.berti@unimore.it

Luca Pratelli, Accademia Navale, viale Italia 72, 57100 Livorno, Italy
E-mail address: pratel@mail.dm.unipi.it

Pietro Rigo (corresponding author), Dipartimento di Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy
E-mail address: pietro.rigo@unipv.it