Uniform rates of Glivenko-Cantelli convergence and their use in Bayesian inference

Emanuele Dolera

Università degli Studi di Pavia, Italy

joint work with Eugenio Regazzini

Università degli Studi di Pavia, Italy

Abstract

The first part of the talk is focused on the validity of the following two upper bounds:

$$\mathsf{E}\Big[\Big(\sup_{n\geq 1} b_n \mathrm{d}_{[\mathbb{S}]}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n)\Big)^p\Big] \quad \leq \quad C_p(\mathfrak{p}_0) \qquad \text{(for some } p \geq 1) \tag{1}$$

$$\limsup_{n\to\infty} b_n \mathrm{d}_{[\mathbb{S}]}(\mathfrak{p}_0, \hat{\mathfrak{p}}_n) \quad \leq \quad Y(\mathfrak{p}_0) \qquad \mathsf{P}-\text{a.s.} \tag{2}$$

Here: $\mathbb{S}$ denotes some topologically nice metric space (e.g., Polish); $[\mathbb{S}]$ and $\mathrm{d}_{[\mathbb{S}]}$ stand for the space of all probability measures (p.m.'s) on $\mathbb{S}$ and a suitable distance on $[\mathbb{S}]$, respectively; $\mathfrak{p}_0$ is a fixed p.m. on $\mathbb{S}$; $\hat{\mathfrak{p}}_n$ is a random p.m. on $\mathbb{S}$, depending the first n-segment $(\xi_1, \dots, \xi_n)$ of an i.i.d. sequence $\{\xi_i\}_{i\geq 1}$ from $\mathfrak{p}_0$ (like in the original Glivenko-Cantelli setting, where $\hat{\mathfrak{p}}_n = \frac{1}{n}\sum_{i=1}^n \delta_{\xi_i}$); $\{b_n\}_{n\geq 1}$ is a non-random diverging sequence, which quantifies the rate of convergence; $C_p(\mathfrak{p}_0)$ and $Y(\mathfrak{p}_0)$ are two non-random numbers, explicitly depending on $\mathfrak{p}_0$. The bound (1)—which is what we call a *uniform (mean) Glivenko-Cantelli* bound—is the first example in which $\sup_{n\geq 1}$ figures inside the expectation, the literature having investigated, till now, only non-uniform bounds. See, e.g., [4, 5]. Detailed statements with explicit $C(\mathfrak{p}_0)$ and $Y(\mathfrak{p}_0)$ are given in the following cases of current interest:

$\mathbb{S}$	$\mathrm{d}_{[\mathbb{S}]}$	$\mathfrak{p}_0$	$\hat{\mathfrak{p}}_n$
$\mathbb{R}^d$	p-Wasserstein	any p.m. on $\mathbb{R}^d$ with moment restriction	$\frac{1}{n}\sum_{i=1}^n \delta_{\xi_i}$
$\mathbb{N}$	Aly-Silvey index	any p.m. on $\mathbb{N}$ with regularity restriction	$\frac{1}{n}\sum_{i=1}^n \delta_{\xi_i}$
$\mathbb{R}^d$	Kullback-Leibler	Gaussian p.m. on $\mathbb{R}^d$	Gaussian with estimated (MLE) mean and covariance
$\mathbb{R}^d$	Kullback-Leibler	exponential family on $\mathbb{R}^d$	exponential with estimated (MLE) parameter

$\mathbb{S}$	$\mathrm{d}_{[\mathbb{S}]}$	$\mathfrak{p}_0$	$\hat{\mathfrak{p}}_n$
any separable metric space with Lipschitz embedding in $L^p(\Omega, \mathcal{F}, \mu)$	1-Wasserstein	any p.m. on $\mathbb{S}$ with regularity restriction	$\frac{1}{n}\sum_{i=1}^n \delta_{\xi_i}$

In the second part, of more statistical flavor, we assume that $\{\xi_i\}_{i\geq 1}$ is an *exchangeable* sequence (of observations) and we study the approximation of basic elements of the Bayesian inference, such as the *posterior* distribution $\pi(\cdot \mid \xi_1, \dots, \xi_n) := \mathsf{P}[\tilde{\mathfrak{p}} \in \cdot \mid \xi_1, \dots, \xi_n]$ and the *predictive* distribution, here written as $q_m(\cdot \mid \xi_1, \dots, \xi_n) := \mathsf{P}[\frac{1}{m}\sum_{i=1}^m \delta_{\xi_{i+n}} \in \cdot \mid \xi_1, \dots, \xi_n]$, where $\tilde{\mathfrak{p}}$ stands for the random probability measure satisfying the de Finetti representation $\mathsf{P}[\{\xi_i\}_{i\geq 1} \in \cdot \mid \tilde{\mathfrak{p}}] = \tilde{\mathfrak{p}}^\infty(\cdot)$. Then, we show that, if $\mathsf{E}[C_p(\tilde{\mathfrak{p}})] < +\infty$ with the same C_p as in (1), a martingale argument due to Blackwell and Dubins [1] yields, with probability one,

$$\limsup_{n\to\infty} b_n \mathrm{d}_{[[\mathbb{S}]]}^{(W_p)}\big(q_m(\cdot \mid \xi_1, \dots, \xi_n), \hat{\mathfrak{p}}_n^\infty \circ \tilde{\mathfrak{e}}_{n,m}^{-1}\big) \leq \limsup_{n\to\infty} b_n \mathrm{d}_{[[\mathbb{S}]]}^{(W_p)}\big(\pi(\cdot \mid \xi_1, \dots, \xi_n), \delta_{\hat{\mathfrak{p}}_n}\big) \leq Y(\tilde{\mathfrak{p}})$$

with b_n and Y as in (1)-(2), $\mathrm{d}_{[[\mathbb{S}]]}^{(W_p)}$ denoting the p-Wasserstein distance on the space of all p.m.'s on $([\mathbb{S}], \mathrm{d}_{[\mathbb{S}]})$ and $\hat{\mathfrak{p}}_n^\infty \circ \tilde{\mathfrak{e}}_{n,m}^{-1}(\cdot)$ standing for $\hat{\mathfrak{p}}_n^\infty\big(\{(x_1, x_2, \dots) \in \mathbb{S}^\infty \mid \frac{1}{m}\sum_{i=1}^m \delta_{x_i} \in \cdot\}\big)$. We conclude by showing how such abstract results constitute the basis for a "Bayesian theory of consistency", meaning that frequentist procedures are now seen as approximations of orthodox ones "with the glasses of the prior". See [2, 3]. In particular, we obtain a rate of approximation of orthodox Bayesian inferences by means of some empirical Bayes procedures.

References

[1] Blackwell, D. and Dubins, L.E. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33**, 882-886.

[2] Cifarelli, D.M., Dolera, E. and Regazzini, E. (2016). Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Internat. J. Approx. Reason.* **78**, 138-152.

[3] Dolera, E. and Regazzini, E. (2018). Uniform rates of the Glivenko-Cantelli convergence and their use in approximating Bayesian inferences. *Arxiv*:1712.07361v2

[4] Dudley, R.M. (1969). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.* **40**, 40-50.

[5] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162**, 707-738.