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:

\[
\mathbb{E} \left( \sup_{n \geq 1} b_n d_{[S]}(p_0, \hat{p}_n) \right)^p \leq C_p(p_0) \quad \text{(for some } p \geq 1) \tag{1}
\]

\[
\limsup_{n \to \infty} b_n d_{[S]}(p_0, \hat{p}_n) \leq Y(p_0) \quad \mathcal{P} - \text{a.s.} \tag{2}
\]

Here: \( S \) denotes some topologically nice metric space (e.g., Polish); \([S]\) and \( d_{[S]}\) stand for the space of all probability measures (p.m.'s) on \( S \) and a suitable distance on \([S]\), respectively; \( p_0 \) is a fixed p.m. on \( S \); \( \hat{p}_n \) is a random p.m. on \( S \), depending the first \( n \)-segment \((\xi_1, \ldots, \xi_n)\) of an i.i.d. sequence \( \{\xi_i\}_{i \geq 1} \) from \( p_0 \) (like in the original Glivenko-Cantelli setting, where \( \hat{p}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i} \)); \( \{b_n\}_{n \geq 1} \) is a non-random diverging sequence, which quantifies the rate of convergence; \( C_p(p_0) \) and \( Y(p_0) \) are two non-random numbers, explicitly depending on \( p_0 \).

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 \( C(p_0) \) and \( Y(p_0) \) are given in the following cases of current interest:

<table>
<thead>
<tr>
<th>( S )</th>
<th>( d_{[S]} )</th>
<th>( p_0 )</th>
<th>( \hat{p}_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{R}^d )</td>
<td>( p )-Wasserstein</td>
<td>any p.m. on ( \mathbb{R}^d ) with moment restriction</td>
<td>( \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i} )</td>
</tr>
<tr>
<td>( \mathbb{N} )</td>
<td>Aly-Silvey index</td>
<td>any p.m. on ( \mathbb{N} ) with regularity restriction</td>
<td>( \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i} )</td>
</tr>
<tr>
<td>( \mathbb{R}^d )</td>
<td>Kullback-Leibler</td>
<td>Gaussian p.m. on ( \mathbb{R}^d )</td>
<td>Gaussian with estimated (MLE) mean and covariance</td>
</tr>
<tr>
<td>( \mathbb{R}^d )</td>
<td>Kullback-Leibler</td>
<td>exponential family on ( \mathbb{R}^d )</td>
<td>exponential with estimated (MLE) parameter</td>
</tr>
</tbody>
</table>
any separable metric space with Lipschitz embedding in $L^p(\Omega, \mathcal{F}, \mu)$ | 1-Wasserstein | any p.m. on $S$ with regularity restriction | $\frac{1}{p} \sum_{i=1}^n \delta_{\xi_i}$

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 $\pi(\cdot | \xi_1, \ldots, \xi_n) := \mathcal{P}[\hat{\pi} \in \cdot | \xi_1, \ldots, \xi_n]$ and the predictive distribution, here written as $q_m(\cdot | \xi_1, \ldots, \xi_n) := \mathcal{P}[\frac{1}{m} \sum_{i=1}^m \delta_{\xi_{i+m}} \in \cdot | \xi_1, \ldots, \xi_n]$, where $\hat{\pi}$ stands for the random probability measure satisfying the de Finetti representation $\mathcal{P}[\{\xi_i\}_{i \geq 1} \in \cdot | \hat{\pi}] = \hat{\pi}^\infty(\cdot)$. Then, we show that, if $E[C_p(\hat{\pi})] < +\infty$ with the same $C_p$ as in (1), a martingale argument due to Blackwell and Dubins [1] yields, with probability one,

$$\limsup_{n \to \infty} b_n d_{[S]}^{(W_p)}(q_m(\cdot | \xi_1, \ldots, \xi_n), \hat{\pi}^\infty_n \circ \tilde{\epsilon}_{n,m}^{-1}) \leq \limsup_{n \to \infty} b_n d_{[S]}^{(W_p)}(\pi(\cdot | \xi_1, \ldots, \xi_n), \delta_{\hat{\pi}_n}) \leq Y(\hat{\pi})$$

with $b_n$ and $Y$ as in (1)-(2), $d_{[S]}^{(W_p)}$ denoting the $p$-Wasserstein distance on the space of all p.m.’s on $([S], d_{[S]})$ and $\hat{\pi}_n^\infty \circ \tilde{\epsilon}_{n,m}^{-1}(\cdot)$ standing for $\hat{\pi}_n^\infty((x_1, x_2, \ldots) \in S^\infty | \frac{1}{m} \sum_{i=1}^m \delta_{\xi_i} \in \cdot)$. 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


