Frequentistic approximations to Bayesian prevision of exchangeable random elements

Eugenio Regazzini

Università degli Studi di Pavia

Abstract. It is well-known that reasoning by induction is completely justified from a probabilistic viewpoint when, for example, observations are seen as forming a sequence of random elements with exchangeable distribution ρ. Indeed, for such a sequence, distance between the conditional (predictive) law of m future observations, given the empirical distribution $\tilde{\mathfrak{e}}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{\xi}_i}$ of the n past observations $\tilde{\boldsymbol{\xi}}^{(n)} := (\tilde{\xi}_1, \dots, \tilde{\xi}_n)$, and the m-fold product of $\tilde{\mathfrak{e}}_n$ converges to zero as $n \to +\infty$, for each $m \in \mathbb{N}$ with ρ-probability one, provided that the distance at issue stems from a metrization of weak convergence of probability measures. In this talk, based on a joint work with D.M. Cifarelli and E. Dolera, one studies the rapidity of convergence to zero of the aforesaid distance, in order to establish how great should be n in order to consider frequentistic inferences—here represented by $\tilde{\mathfrak{e}}_n^m := \underbrace{\tilde{\mathfrak{e}}_n \otimes \cdots \otimes \tilde{\mathfrak{e}}_n}_{m-\text{times}}$—as acceptable substitutes of Bayesian inferences, derived, for instance, from the predictive distribution for m future observations. In more precise form, after setting d_m for the above-mentioned distance and $p_m(\tilde{\boldsymbol{\xi}}^{(n)})$ for the predictive distribution, our study aims at determining a positive sequence divergent to infinity, say $\{b_n\}_{n\geq 1}$, together with a suitable positive constant L such that, for every $\varepsilon, \eta > 0$, there exists an index $n_0 = n_0(\varepsilon, \eta) \in \mathbb{N}$ for which

$$\rho\left(\left\{\max_{\nu \leq n \leq \nu+k} b_n \mathrm{d}_m(p_m(\tilde{\boldsymbol{\xi}}^{(n)}), \tilde{\mathfrak{e}}_n^m) \leq L + \varepsilon\right\}\right) \geq 1 - \eta$$

holds true for every $\nu \geq n_0$ and for every $k \in \mathbb{N}$. Analogous statements are valid also in parametric settings.

If one thinks of $\mathrm{d}_m(p_m(\tilde{\boldsymbol{\xi}}^{(n)}), \tilde{\mathfrak{e}}_n^m)$ as an expression of loss function, then $\{1/b_n\}_{n\geq 1}$ can be seen as a sequence of costs to be borne in order to guarantee an overall coverage of losses, in the sense that the sequence of ratios of the loss to $1/b_n$ turns out to be bounded by a

constant, uniformly with respect to n from a suitable value of n, with ρ-probability close to one. So, the rapidity of the decay of the random loss to zero can be quantified by means of the rapidity of convergence to zero of the numerical sequence $\{1/b_n\}_{n\geq 1}$.

References

[1] Cifarelli, D.M., Dolera, E. and Regazzini, E. (2016). Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Int. J. Approx. Reason*: to appear. ArXiv:1602.01269v1