ASYMPTOTIC PREDICTIVE INFERENCE WITH EXCHANGEABLE DATA

Patrizia Berti, Luca Pratelli, Pietro Rigo

Universita' di Modena e Reggio-Emilia
Accademia Navale di Livorno
Universita' di Pavia

Roma, September 12, 2016

Framework

- $(X_n : n \geq 1)$ sequence of random variables with values in a (**nice**) measurable space $(S, \mathcal{B})$

- $\mu_n = (1/n) \sum_{i=1}^n \delta_{X_i}$ empirical measure

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

Under some conditions, $\mu_n(B) - a_n(B) \to 0$ a.s. for each $B \in \mathcal{B}$. Thus, we fix $\mathcal{D} \subset \mathcal{B}$, and we investigate **uniform** convergence over $\mathcal{D}$, i.e., we focus on

$$\boxed{||\mu_n - a_n|| = \sup_{B \in \mathcal{D}} |\mu_n(B) - a_n(B)|}$$

Motivations

- Bayesian predictive inference
- Empirical processes for non-ergodic data
- Predictive distributions of exchangeable sequences
- Frequentistic approximations of Bayesian procedures

Assumptions and possible generalizations

- From now on, (X_n) is **exchangeable**. However, most results are still valid if (X_n) is **conditionally identically distributed**, namely,

$$\boxed{P(X_k \in \cdot \mid X_1, \ldots, X_n) = P(X_{n+1} \in \cdot \mid X_1, \ldots, X_n) \text{ a.s.}}$$

  for each $k > n \geq 0$

- Similarly, in most results, $\sigma(X_1, \ldots, X_n)$ can be replaced by any filtration $\mathcal{G}_n$ which makes (X_n) adapted

- To avoid measurability problems, $\mathcal{D}$ is countably determined, i.e.

$$\sup_{B \in \mathcal{D}} |\alpha(B) - \beta(B)| = \sup_{B \in \mathcal{D}_0} |\alpha(B) - \beta(B)|$$

  for some countable $\mathcal{D}_0 \subset \mathcal{D}$ and all probabilities α and β on $\mathcal{B}$

Goals

- To give conditions for

$$\boxed{||\mu_n - a_n|| \to 0 \text{ a.s. or in probability}}$$

- To determine the **rate of convergence**, i.e., to fix the asymptotic behavior of

$$\boxed{r_n\,||\mu_n - a_n||}$$

for suitable constants $r_n > 0$, e.g.

$$r_n = \sqrt{n} \quad \text{or} \quad r_n = \sqrt{\frac{n}{\log \log n}}$$

- Another issue, not discussed in this talk, is the limiting distribution (if any) of $r_n\,(\mu_n - a_n)$ under some distance on the space $l^\infty(\mathcal{D})$

Theorem 1

Recall that, because of exchangeability, there is a random probability measure μ on $\mathcal{B}$ such that

$$\mu_n(B) \to \mu(B) \text{ a.s. for each } B \in \mathcal{B}$$

$$\boxed{\text{If } ||\mu_n - \mu|| \to 0 \text{ a.s. then } ||\mu_n - a_n|| \to 0 \text{ a.s.}}$$

Examples: $||\mu_n - a_n|| \to 0$ a.s. provided

- $\mathcal{D}$ is a universal Glivenko-Cantelli-class
- $\mathcal{D} = \mathcal{B}$ and μ a.s. discrete (usual in Bayesian nonparametrics !)
- $S = \mathcal{R}^k$, $\mathcal{D} = \{$closed convex sets$\}$ and $\mu << \lambda$ a.s. where λ is a (non random) σ-finite product measure

Theorem 2

Fix the constants $r_n > 0$ and define $M = \sup_n r_n \, ||\mu_n - \mu||$. Define also the "exchangeable empirical process"

$$W_n = \sqrt{n} \, (\mu_n - \mu)$$

$$\boxed{\text{If } E(M) < \infty \text{ then } \limsup_n ||\mu_n - a_n|| < \infty \text{ a.s.}}$$

Examples:

- $\boxed{\sqrt{\frac{n}{\log n}} \, ||\mu_n - a_n|| \to 0 \text{ a.s.}}$ if $\sup_n E\{||W_n||^p\} < \infty$ for some $p > 2$

- $\boxed{\sqrt{\frac{n}{\log \log n}} \, ||\mu_n - a_n|| \leq \frac{1}{\sqrt{2}} \text{ a.s.}}$ if $\mathcal{D}$ is a VC-class

Theorem 3

If $\sup_n E\{||W_n||^p\} < \infty$, for some $p \geq 1$, then

$$r_n \, ||\mu_n - a_n|| \to 0 \text{ in probability}$$

whenever $\boxed{\frac{r_n}{\sqrt{n}} \to 0}$

Remark: Thus, if $\sup_n E\{||W_n||^p\} < \infty$, then

$$\sqrt{\frac{n}{\log \log n}} \, ||\mu_n - a_n|| \to 0 \text{ in probability}$$

but a.s. convergence can not be obtained by the LIL

Theorem 4

Define

$$h_n(B) = P\{X_{n+1} \in B \mid \mu_n(B)\}$$

Suppose that $\mu(B)$ has absolutely continuous distribution for each $B \in \mathcal{D}$ such that $0 < P(X_1 \in B) < 1$. Then, **under mild technical conditions**, one obtains

$$\boxed{\sqrt{n}\,||\mu_n - a_n|| \to 0 \text{ in probability}}$$

if and only if

$$\sqrt{n}\,\{a_n(B) - h_n(B)\} \to 0 \text{ in probability for each } B \in \mathcal{D}$$

Open problem: Characterize the above condition. More generally, in addition to $\mu_n(B)$, which further information is required to evaluate $a_n(B)$?