Some generalizations of the Hungarian strong invariance principles in the abstract empirical process setting, with applications

Philippe Berthet

Université de Toulouse

With D.M. Mason we developped a Gaussian coupling method to provide strong approximation bounds of the empirical process by a Brownian bridge in the indexed by functions setting. The aim of this talk is to give an overview of some variants that have been investigated in the last few years, that are powerful enough to encompass or improve most known results, and then to focus on applications in various mathematical statistics contexts.

The idea behind this global approach comes from the Hungarian school : coupling for finite n an empirical process with its limiting process is a powerful and easy way of deriving weak and strong invariance principles for many statistics – viewed as transforms of the empirical measure. This is the Komlós-Major-Tusnády construction and the main viewpoint put forward in the books of Csörgő and Révész (1981), Shorack and Wellner (1986) and Csörgő and Horváth (1993). Now, in the general case the technique of proof is very different and relies on tools from probability in Banach spaces, thus it can not lead to very sharp results, however this suffices to compute rates in the Donsker type theorems stated in the survey books of Wellner and Van der Vaart (1996) and Dudley (1999).

These rates relate the structure or the entropy of the class of function to the underlying law of probability under various kinds of hypotheses. The coupling argument being based on a very general result of Zaitsev, they are clearly suboptimal in special cases, but rather accurate when few is known on the law. One of the main and very difficult problems left open is to identify the worse law of probability with respect to a given class of functions and conversely.

Indeed KMT–approximations methods have proved to be very powerful and useful in the real line setting to control very complicated statistical procedures and the message of this talk is that the generalized version of KMT, yet less precise, can now be used in a similar way to handle modern statistical procedures in high dimension and fruitfully exploit Gaussianity at finite sample size n. Important applications concern lower bounds of concentration for the empirical process, generalized U-processes, hybrid empirical processes, invariance principles for dependent data, control of the excess risk of M-estimators, boundary local empirical processes, limiting shapes of set valued M-estimators, stability of contrast minimizers, slope heuristic in model selection, asymptotic risk distributions, multiple testing procedures. All these applications are on the way and take advantage of the Brownian approximation bounds.