Statistics Seminar

Prof. Patrice Bertail
MODAL’X, Universite Paris-Nanterre; Chaire Big Data, TeleComParis-Tech

Survey sampling for non-parametric statistics and big data

Abstract:

Subsampling methods as well as general sampling methods appear as natural tools to handle very large database (big data in the individual dimension) when traditional statistical methods or statistical learning algorithms fail to be implemented on too large datasets. The choice of the weights of the survey sampling scheme may reduce the loss implied by the choice of a much more smaller sampling size (according to the problem of interest).

I will first review some asymptotic results for general survey sampling based empirical processes indexed by class of functions (Bertail and Clemencon, 2017), for Poisson type and conditional Poisson (rejective) survey samplings. These results may be extended to a large class of survey sampling plans via the notion of negative association of most survey sampling plans (Bertail and Rebecq, 2017). Then, in the perspective to generalize some statistical learning tasks to sampled data, we will obtain exponential bounds for the probabilities of deviation of a sample sum from its expectation when the variables involved in the summation are obtained by sampling in a finite population according to a rejective scheme, generalizing sampling without replacement and using an appropriate normalization.

In contrast to Poisson sampling, classical deviation inequalities in the i.i.d. setting do not straightforwardly apply to sample sums related to rejective schemes due to the inherent dependence structure of the sampled points. We show here how to overcome this difficulty by combining the formulation of rejective sampling as Poisson sampling conditioned upon the sample size with the Escher transformation. In particular, the Bennet/Bernstein type bounds established highlight the effect of the asymptotic variance of the (properly standardized) sample weighted sum, and are shown to be much more accurate than those based on the negative association property.

Special Note:
Professor Bertail will be visiting UCSD from Oct. 20 to Nov 5.
His personal webpage is: http://www.parisnanterre.fr/m-patrice-bertail697327.kjsp

Host: Dimitris Politis

Tuesday, October 24, 2017
10:00 AM
AP&M 6402