



**SCHOOL ON LARGE SCALE PROBLEMS IN MACHINE LEARNING AND
WORKSHOP ON COMMON CONCEPTS IN
MACHINE LEARNING AND STATISTICAL PHYSICS
20 - 31 August 2012, Trieste, Italy**

**Exchangeable random partitions,
Bayesian inference and species sampling problems**

Antonio LIJOI

**Department of Economics and Business, University of Pavia and
Collegio Carlo Alberto Moncalieri, Italy**

Abstract:

Sampling from a population whose units belong to different types or species is a common feature in many applied areas. The inferential issues typically addressed in this framework include the evaluation of species richness, the design of sampling experiments and the estimation of rare species variety. Exchangeable random partitions are a useful tool for providing elegant and effective answers to these questions and the Bayesian learning procedure yields the most natural approach for predicting and estimating several quantities of interest. The talk will focus on a large and flexible class of exchangeable random partitions known as Gibbs-type models that include the Dirichlet and the two-parameter Poisson-Dirichlet processes as special cases. Given an observed basic sample of size n , one is interested in the prediction of some key aspects of the outcome from an additional sample of size m . Among them, it is worth mentioning: (i) the number of new species that will be detected in the additional sample; (ii) the number of species that will be detected with a certain frequency in the enlarged sample of size $n + m$; (iii) the discovery probability, i.e. the probability of detecting, at the $(n + m + 1)$ -th observation, species that have been observed with any given frequency in the enlarged sample of size $n + m$. The Bayesian estimators we obtain will then be applied to two real datasets. Finally, a concise illustration of asymptotic properties of these models will be sketched.

