Department of Decision Sciences

Statistics Seminar

Model-based clustering via NRMI mixtures

Raffaele Argiento
Istituto di Matematica Applicata e Tecnologie Informatiche IMATI

Thursday, 14th May
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

Mixtures of parametric densities arise as the natural statistical tool when dealing with model-based clustering problems, and a very flexible class of such mixtures is obtained when the mixing measure is a random probability measure, possibly a.s. discrete. In this talk, I will consider mixtures when the mixing probability measure is within a large class of random probability measures, that is normalized random measures with independent increments (NRMI).
However, the computational effort to compute the relevant posteriors in this case can be very burdensome, since MCMC schemes are complicated by the presence of an infinite number of parameters.
Here, we propose a truncation method to approximate the mixing measure and simplify the computations. Since a NRMI is a random discrete measure where the weights are obtained by normalization of the points of a Poisson process, we discard those larger than a threshold epsilon.
Hence, the number of parameters becomes finite, so that an efficient blocked Gibbs sampler to simulate from the posterior is built.
Furthermore we assign a prior distribution for epsilon and, consequently, data can "drive" the degree of approximation through the posterior.
To show the performance of our algorithm and the flexibility of the model, we will briefly illustrate two case studies via NRMI mixtures: the first concerns the reconstruction of Italian seismogenetic sources, while the second deals with an association study between a disease and SNPs data.

Department of Decision Sciences

Via Röntgen 1 - 20136Milano

Tel. 02 5836.5632
Fax 02 5836.5630