Welcome seminar for new faculty

Of mice and music: finite-infinite shared atoms nested priors for the segmentation of large-scale grouped data

A seminar by Francesco Denti

University of Padua - ITA

Thursday 7 Nov 2024 | 3:15 p.m.
Room Benvenuti
Department of Statistical Sciences

Over the last few years, the Bayesian community has dedicated increased attention to mixture priors inducing nested random partitions.
Models based on these priors allow the estimation of a two-layered partition over grouped data: across groups and across observations.
We focus on nested models based on shared observational atoms, which permit observational clusters to spread across all the distributional groups.
We introduce a novel finite-infinite nested model to overcome the high correlation between random mixing measures imposed a priori by fully nonparametric common atoms models.
This specification also enables the development of fast algorithms for posterior inference. Indeed, the tractability of the proposed prior grants the derivation of tailored mean-field variational inference algorithms, which scale up the applicability of Bayesian nested mixture models to large datasets. Such a computational strategy is highly efficient, and the accuracy of the posterior density estimate and the estimated partition is comparable with a standard Gibbs sampler algorithm.
To showcase the applicability of the proposed framework, we illustrate how this prior can be embedded in more complex models motivated by real-world problems.
In particular, we introduce and compare two models: one devised for clustering Spotify artists based on the characteristics of their songs and the other created to segment large mass-spectrometry imaging matrix data.

Università degli Studi di Padova

Dipartimento di Scienze Statistiche