Bayesian Mixture Models: Foundations and Innovations

by

Professor Maria DE IORIO
National University of Singapore

Date: 28 February 2025 (Friday)
Time: 10:45am – 11:45am
Venue: Case Room 1001, LSK Business Building

Abstract

Mixture models are essential tools for analysing heterogeneous populations. From a Bayesian nonparametric perspective, we introduce a novel class of priors, the Normalised Independent Point Process, and develop both marginal and conditional algorithms for finite mixture models with a random number of components. By employing an auxiliary variable MCMC approach, we efficiently address the challenges posed by intractable posterior distributions, creating a flexible and extensible framework for Bayesian modelling.

In this talk, we highlight key extensions of this framework for the analysis of multiview data, where dependencies arise across multiple data domains. To address these complexities, we propose a probabilistic framework for conditional partial exchangeability, specifically designed for multiview and longitudinal data. This framework introduces flexible random partitions that vary across features, effectively capturing dependencies across multiple domains. Furthermore, it can be extended to link random partitions of subjects across datasets through an underlying shared partition structure, enabling information sharing and facilitating robust inference. By accounting for within-subject dependencies and modelling marginal relationships between datasets, this approach enhances both the flexibility and interpretability of clustering structures.

We also present Bayesian Distance Clustering (BDC), a hybrid approach that defines the likelihood on pairwise distances between observations rather than on the observations themselves. The novelty of BDC lies in its ability to incorporate both cohesion and repulsion terms, ensuring cluster identifiability. This method strikes a balance between computational feasibility and probabilistic interpretability, making it particularly effective for clustering large and complex datasets. We extend BDC to multiview data by introducing scalable models that maintain predictive accuracy while improving efficiency. Inspired by K-medoids, we propose a novel tessellation-based method that identifies tessellation centres, or "medoids," significantly enhancing the efficiency of the clustering process.

We validate these methods through extensive simulations and applications to real-world datasets, demonstrating their effectiveness across a variety of contexts. These advancements provide a robust and versatile toolkit for addressing modern data analysis challenges, particularly in the context of multiview and longitudinal data.

Bio

Maria De Iorio is Professor at Yong Loo Lin School of Medicine, National University of Singapore, and Principal Investigator at the Singapore Institute for Clinical Sciences (SICS), A*STAR. She obtained her undergraduate degree in Economics from Bocconi University, Italy and her MSc and PhD in Statistics from Duke University, USA. She then joined the Mathematical Genetics group of the University of Oxford as a postdoctoral research assistant. In 2003 she moved to Imperial College London, working first as a lecturer and then as a senior lecturer. In September 2011 she joined the statistics group at University College London as Reader in Statistics and she, then, became Professor of Biostatistics in 2015. She moved to Singapore in August 2018, joining Yale-NUS College, where she served as Head of Studies for the Mathematical, Computational and Statistical Sciences Major. In 2020, she obtained a joint appointment with NUS Medical School, where she moved full time in 2021. She has extensive expertise in Bayesian statistics, Bayesian Nonparametrics, Biostatistics & computational methods. She has a long track-record in modelling complex biomedical data and analysing high throughput data in genomics/metabolomics. To date she has published 107 articles in international journals.

Her research achievements have been recognised by her peers first in 2005 when she was awarded the Research Prize of the Royal Statistical Society and then in 2020 she was elected ISBA fellow, the highest recognition for a Bayesian statistician. Over the years she has served on different professional boards. She is currently the Chair of the Research Committee of the Royal Statistical Society.