

# **Distributed Learning of Finite Gaussian Mixtures**

**Jiahua Chen**

Department of Statistics, University of British Columbia  
2329 West Mall, Vancouver, BC, Canada  
jhchen@stat.ubc.ca

## **Abstract**

Advances in information technology have led to extremely large datasets that are often kept in different storage centers. Existing statistical methods must be adapted to overcome the resulting computational obstacles while retaining statistical validity and efficiency. Split-and-conquer approaches have been applied in many areas, including quantile processes, regression analysis, principal eigenspaces, and exponential families. We study split-and-conquer approaches for the distributed learning of finite Gaussian mixtures. We recommend a reduction strategy and develop an effective MM algorithm. The new estimator is shown to be consistent and retains root-n consistency under some general conditions. Experiments based on simulated and real-world data show that the proposed split-and-conquer approach has comparable statistical performance with the global estimator based on the full dataset, if the latter is feasible. It can even slightly outperform the global estimator if the model assumption does not match the real-world data. It also has better statistical and computational performance than some existing methods.