

PhD program in Statistics

DSS Statistics Seminar

November 3, 2023, 15:00

Webinar <https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09>

Passcode: 432940

A parsimonious family of mixtures of multivariate Poisson log-normal factor analyzers for clustering count data

Sanjeena Dang (Subedi), Ph.D.

School of Mathematics and Statistics, Carleton University, Canada.

Multivariate count data are commonly encountered in bioinformatics. Although the Poisson distribution seems a natural fit for these count data, its multivariate extension is computationally expensive. Recently, mixtures of multivariate Poisson lognormal (MPLN) models have been used to efficiently analyze these multivariate count measurements. In the MPLN model, the counts, conditional on the latent variable, are modelled using a Poisson distribution, and the latent variable comes from a multivariate Gaussian distribution. Due to this hierarchical structure, the MPLN model can account for over-dispersion as opposed to the traditional Poisson distribution and allows for correlation between the variables. The mixture of multivariate Poisson-log normal distributions for high dimensional data is extended by incorporating a factor analyzer structure in the latent space. A family of parsimonious mixtures of multivariate Poisson lognormal distributions are proposed by decomposing the covariance matrix and imposing constraints on these decompositions. The performance of the model is demonstrated using simulated and real datasets.



SAPIENZA
UNIVERSITÀ DI ROMA