中央研究院統計科學研究所
學 術 演 講

講 題：High-dimensional clustering by multi-faceted and outcome-guided latent class models with application to disease subtyping of omics data

演 講 人：Prof. George Tseng（曾建城 教授）

Department of Biostatistics, University of Pittsburgh, U.S.A.

時 間：2025-03-19（Wed.）10:30-12:00

地 點：Auditorium, B1F, Institute of Statistical Science；The tea reception will be held at 10:10.

備 註：Online live streaming through Cisco Webex will be available.

Abstract

Cluster analysis of high-dimensional omics data is a critical step to discover novel molecular disease subtypes in a complex disease as a first step towards precision medicine. High-dimensional omics data, however, often contain intricate and multi-faceted information, resulting in the coexistence of multiple plausible sample partitions based on different subsets of gene features (e.g., male/female characterized by sex chromosome genes, age groups by age-associated genes, and disease subtypes by disease related genes). Existence of such a multi-faceted cluster structure poses challenges to identify the desired clinically useful disease subtypes. To this end, we first propose a model-based multi-facet clustering (MFClust) method based on a mixture of Gaussian mixture models, where the former mixture achieves facet assignment for genes and the latter mixture determines cluster assignment of samples. This method simultaneously achieves detection of the gene signature and its associated sample clustering in each facet. In the second method, we develop an outcome-guided clustering (ogClust) approach by utilizing a pre-determined clinical outcome to guide the detection of the targeted facet of sample clustering that most likely have the translational potential towards precision medicine. We have extended the approach to allow multivariate mixed-type of clinical outcomes in the clustering guidance. Finally, I will discuss an on-going Bayesian outcome-guided clustering project under a consensus clustering setting to allow a unified modeling and selection of multivariate clinical outcomes, and to match with the multi-facet clustering structure in the omics data.