

東海大學統計學系學術演講

時 間：115 年 9 月 29 日（星期二下午 14:10~15:00）

地 點：管理學院新大樓 M240

主講人：張源俊 教授（中央研究院 統計科學研究所）

講 題：PCA-Guided Quantile Sampling

Abstract

In large-scale statistical modeling, reducing data size through subsampling is essential for balancing computational efficiency and statistical accuracy. We propose a new method, Principal Component Analysis guided Quantile Sampling (PCA-QS), which projects data onto principal components and applies quantile-based sampling to construct subsets that spread retained observations across aggregate cutoff-count profiles derived from leading-PC quantile ranks. Across synthetic and real-world datasets, PCA-QS preserves distributional structure competitively with uniform random sampling, leverage score sampling, and coresets methods while remaining computationally efficient. In our experiments, its largest advantages occur when a few leading principal components capture dominant structure in the data; its behavior approaches that of simple random sampling when the covariance spectrum is relatively flat, so its relative advantage is not uniform across datasets, discrepancy measures, and numbers of retained components. The method is intended as a general-purpose, structure-aware subsampling framework; the experiments delimit when its PCA-guided stratification is most useful and when its behavior approaches that of SRS.

This work is co-authored with Dr. Hui-Mean Foo, Institute of Statistical Science, Academia Sinica

敬請公告，歡迎參加！

* 爾後演講公告將於本系網站刊載 *

<http://stat.thu.edu.tw/>