Abstracts

Novel Compression Algorithms Using Multi-Stage Methods

Swarnali Banerjee, Loyola University Chicago

Our growing research demands bigger datasets which also makes compression algorithms the need of the hour. Since most of the data sizes are dynamic, sequential and multi-stage algorithms are very useful to construct such algorithms. For the past few years our research has focused on constructing multi-stage methods for data compression algorithms using Principal Component Analysis. For data arriving sequentially, however, several online PCA methodologies exist which focus only on compression risk minimization. Our work focusses on minimizing the cost-compression risk which considers the compression loss and the sampling cost using a multistage PCA procedures. While our previously published work focusses on fixed dimensions, our current work explores these algorithms under "growing dimensions". Under certain regularity conditions, these procedures are shown to be first-order and second order efficient and risk efficient. The proposed algorithms are novel, fast and efficient on various types of data as will be seen in illustrations.

Sequential Estimation and Boundary Correction for the Overlap Coefficient

Zhe Wang, Denison University

Overlap quantifies the shared area between two probability distributions and serves as an intuitive measure of similarity between populations. The Overlap Coefficient (OVL) provides a numerical value for this shared region, ranging from 0 to 1. In this presentation, we develop both fixed-sample and sequential estimation methodologies for constructing fixed-size confidence intervals of the OVL. Special attention is given to challenges that arise when the true OVL lies near the boundaries (0 or 1), where traditional normal approximations often fail. We discuss variance-stabilizing transformations and boundary corrections for improving the performance of sequential stopping rules. Simulation studies demonstrate the efficiency and accuracy of the proposed methods across small, moderate, and large OVL settings, followed by applications to real-world data examples.

Penalized Sequential Estimation for Recurrent Events with High-Dimensional Covariates

Laura Dumitrescu, Fairfield University

The study of recurrent events, such as hospital readmissions or stock market volatility spikes, is important in biostatistics and finance but poses statistical challenges due to complex dependencies and censoring. Under a semiparametric framework for modeling gap times between events, we employ a penalized sequential estimating equations approach to handle high-dimensional covariates. The procedure yields estimators with attractive large sample properties and we illustrate how the rate of divergence in the number of parameters affects the convergence rate of penalized estimators when the number of covariates grows to infinity as the number of clusters increases. The proposed method has the advantage of simultaneously estimating parameters while selecting important variables and accounting for the within-cluster correlation.

Sequential Sampling for Ellipsoidal Confidence Regions for Multivariate Means

Swathi Venkatesan, Fairfield University

Traditional fixed-size confidence region (FSCR) methods for estimating the mean of a multivariate normal distribution often fix the region's maximum diameter in advance, without regard to the quality of available data. We propose a new approach that incorporates data quality into determining the region's size. Starting with a modified FSCR method where the structure of the variance-covariance matrix is known, we introduce a minimum risk FSCR (MRFSCR) framework inspired by point estimation methods that balance estimation accuracy and sampling cost. We develop a unified multistage sampling strategy to construct these regions, ensuring desirable asymptotic properties. The methodology is illustrated through practical sampling strategies, simulation studies, and real-data examples.