INDIAN STATISTICAL INSTITUTE

Theoretical Statistics and Mathematics Unit, Kolkata

SEMINAR

Date: November 29, 2023
Time: 03:15 PM

VENUE:

L-infinity

(5th Floor, A.N. Kolmogorov Bhavan), ISI Kolkata

TITLE:

Some nonparametric tests for high-dimensional and functional data

SPEAKER:

Bilol Banerjee

Stat-Math Unit, ISI Kolkata

ABSTRACT:

The explosive development of information technology and science has enabled researchers to collect, store, process and analyze huge amounts of data. Some of these data sets have dimensions comparable to or even much larger than the sample size. Many popular classical statistical methods cannot be used for analysing such high-dimensional data sets. So, we need novel high-dimensional methods for their analysis. Sometimes we also deal with data sets, where the observations are functions. Analysis of such functional data brings new challenges to statisticians. In this Ph.D. thesis, we would like to develop and analyze some nonparametric tests for such high-dimensional and functional data.

We begin with a two-sample problem for high-dimensional data, where we test $H_0 : F = G$ against $H_1 : F \neq G$ based on two sets of independent observations $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ and $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m\}$ from two d-dimensional probability distributions F and G, respectively. First, we propose a test based on ball divergence, which is calibrated using the permutation method. We investigate the behaviour of the test in three different asymptotic regimes, i.e., the classical regime, where d is considered to be fixed while n and m diverge to infinity, the high dimension, low sample size (HDLSS) regime, where the sample sizes m, n remain fixed, while the dimension d diverge to infinity, and the random matrix regime, where d, n and m all grow to infinity at appropriate rates.

Next, we consider testing conditional independence between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ given another random vector $\mathbf{Z}$. In other words, we test the null hypothesis $H_0 : P_{\mathbf{X}|\mathbf{Z}} = P_{\mathbf{X}|\mathbf{Y},\mathbf{Z}}$ (almost surely) against the general alternative H_1, where $P_{\mathbf{X}|\mathbf{Z}}$ and $P_{\mathbf{X}|\mathbf{Y},\mathbf{Z}}$ denote the regular conditional probability distributions of $\mathbf{X} \mid \mathbf{Z}$ and $\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}$, respectively. We propose a new test statistic and use a novel bootstrap algorithm for calibrating the test. This bootstrap method guarantees the level property in the classical asymptotic regime and makes the test large sample consistent. We also consider the case, where $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ are high-dimensional random vectors and carry out our analysis in other two asymptotic regimes. For high dimensional data, we consider a new test statistic based on the characteristic function and calibrate the test under the model-$\mathbf{X}$ framework.

We also consider the problem of testing the spherical symmetry of the distribution of a random vector $\mathbf{X}$. This problem can also be viewed as testing $H_0 : \mathbf{X} \stackrel{d}{=} \|\mathbf{X}\|\mathbf{U}$ against $H_1 : \mathbf{X} \stackrel{d}{\neq} \|\mathbf{X}\|\mathbf{U}$, where $\mathbf{U}$ and $\|\mathbf{X}\|$ are independent, and $\mathbf{U}$ is uniform on $\mathcal{S}^{d-1}$, the surface of the d-dimensional unit sphere. Here, we propose a test statistic based on the characteristic functions and calibrate the test using a novel resampling method.

Next, we consider some problems involving functional data. In particular, we consider the two-sample problem and the problem of testing independence between two functional random variables. For the two-sample problem, we propose a new test statistic based on kernel mean embedding and calibrate the test using the permutation method. We extend the local asymptotic normality result using mixture alternatives for functional data and study the local asymptotic behaviour of our permutation test under such alternatives. For the other problem, we propose a new test of independence based on the d-variate Hilbert Schmidt Independence Criterion and calibrate the test using the permutation method.

ALL ARE CORDIALLY INVITED