Welcome seminar for new faculty

Kernel-based quadratic distances: a unified framework for goodness-of-fit testing in the QuadratiK package

A seminar by Giovanni Saraceno

University of Padua - ITA

Thursday 7 Nov 2024 | 2:30 p.m.
Room Benvenuti
Department of Statistical Sciences

In the statistical literature, as well as in artificial intelligence and machine learning, discrepancy measures between probability distributions play a key role in developing goodness-of-fit metrics. Kernel-based methods have gained popularity due to their ability to handle high-dimensional data.
Here, we focus on kernel-based quadratic distances (KBQDs), which leverage non-negative definite kernels, to present a unified framework for two-sample and k-sample goodness-of-fit tests, for which we define the concept of matrix distance. We explore the asymptotic distribution of the proposed tests under the null hypothesis, along with practical considerations for their computation. The KBQD tests exhibit high power for contiguous alternatives, heavy tailed distributions and in higher dimensions.
The innovative framework can be easily utilized via the QuadratiK package, accessible in both R and Python, which provides an extensive suite of tools for assessing goodness-of-fit and performing clustering through kernel-based quadratic distances. The package includes one-sample, two-sample, and k-sample tests, ensuring a robust and efficient way to evaluate distribution fit. Additionally, QuadratiK offers tests for uniformity on d-dimensional spheres using Poisson kernel densities and includes algorithms to generate random samples from this density.
One of the features of the software is a clustering algorithm designed for spherical data, utilizing a mixture of Poisson kernel-based densities on the sphere. This is complemented by a range of graphical functions, empowering users to validate, visualize, and interpret clustering outcomes with ease.

Università degli Studi di Padova

Dipartimento di Scienze Statistiche