Astro@Stats 2017: Sino-Italian Workshop on Astrostatistics

Friday, 8 September 2017 - Friday, 8 September 2017

Department of Statistical Sciences

Book of Abstracts

Contents

Welcome / 0

Welcome by Vice Rector for International Relations

Prof. Alessandro Paccagnella, Vice Rector for International Relations

Welcome / 1

Welcome by Chair of the Department of Statistical Sciences

Prof. Tommaso di Fonzo, Chair of the Department of Statistical Sciences

Welcome / 2

Welcome by Chair of the Department of Physics and Astronomy

Prof. Francesca Soramel, Chair of the Department of Physics and Astronomy

J.C.F. Gauss / 3

Investigating the Cosmic Web with Topological Data Analysis

Author: Jessi Cisewski-Kehe[1]

[1] *Yale University*

Data exhibiting complicated spatial structures are common in many areas of science (e.g. cosmology, biology), but can be difficult to analyze. Persistent homology is a popular approach within the area of Topological Data Analysis (TDA) that offers a new way to represent, visualize, and interpret complex data by extracting topological features, which can be used to infer properties of the underlying structures. In particular, TDA may be useful for analyzing the large-scale structure (LSS) of the Universe, which is an intricate and spatially complex web of matter. In order to understand the physics of the Universe, theoretical and computational cosmologists develop large-scale simulations that allow for visualizing and analyzing the LSS under varying physical assumptions. Each point in the 3D data set represents a galaxy or a cluster of galaxies, and topological summaries ("persistent diagrams") can be obtained summarizing the different ordered holes in the data (e.g. connected components, loops, voids).

The topological summaries are interesting and informative descriptors of the Universe on their own, but hypothesis tests using the topological summaries would provide a way to make more rigorous comparisons of LSS under different theoretical models. For example, the received cosmological model has cold dark matter (CDM); however, while the case is strong for CDM, there are some observational inconsistencies with this theory. Another possibility is warm dark matter (WDM). It is of interest to see if a CDM Universe and WDM Universe produce LSS that is topologically distinct.

We present several possible test statistics for two-sample hypothesis tests using the topological summaries, carry out a simulation study to investigate the suitableness of the proposed test statistics

using simulated data from a variation of the Voronoi foam model, and finally we apply the proposed inference framework to WDM vs. CDM cosmological simulation data.

J.C.F. Gauss / 4

Quasars: the Beacons of the Universe and the Challenge for Machine Learning

Author: Maria Süveges[1]

[1] *Max Planck Institute for Astronomy Heidelberg*

J.C.F. Gauss / 5

Photometric Classification of Supernova with a Biased Training Set

Author: David van Dyk[1]

[1] *Imperial College London*

The expansion of the Universe can be studied via a Bayesian model that relates the difference between the apparent and intrinsic brightnesses of objects to their distance which in turn depends on parameters that describe this expansion. While apparent brightness can be readily measured, intrinsic brightness can only be obtained for certain objects. Type Ia Supernova (SNIa) occur when material accreting onto a white dwarf triggers a powerful supernova explosion. Because this occurs only in a particular physical scenario, we can estimate the intrinsic brightness of SNIa. To take advantage of this, however, SNIa must be precisely classified using (low resolution) photometric data and a biased training set. We use Gaussian Processes to account for irregular observation times and diffusion maps to identify features for a random forest classifier. To account for bias in the overall training set, we use propensity scores to form homogeneous groups where the training subsets are more representative. Finally we enrich the training sets by probabilistically generating synthetic data. In this way we are able to identify SNIa nearly as well as we would with an unbiased training set.

G. Galilei / 6

New Statistical Analyses for the Neutrino Mass Hierarchy Determination

Author: Luca Stanco[1]

Co-author: Emilio Ciuffoli [2]

[1] *INFN Padova*

[2] *Chinese Academy of Sciences*

We preliminary focus on some aspects of the statistical analyses in particle physics.
Then, two new analyses on the neutrino mass hierachy (MA) will be presented.
The first focuses on the determination of MA with neutrino from accelerator beams, while the second one with antineutrino from reactor plants.

G. Galilei / **7**

Extraordinary Claims: the 0.000029% Solution

Author: Tommaso Dorigo[1]

[1] *University of Padova and INFN Padova*

Extraordinary claims require extraordinary evidence. The p < 0.000029% criterion used in HEP and astro-HEP does not appear adequate to cover all experimental situations. The seminar will start with a short history of anomalies found in HEP data and their resolution, and then focus on the statistical problem of defining a proper discovery level for new phenomena and on the non-trivial issues it entails. The seminar will be in part based on this recent article: Extraordinary claims: the 0.000029% solution

G. Galilei / **8**

Nonparametric Semi-Supervised Classification with Application to Signal Detection in High Energy Physics

Author: Alessandro Casa[1]

Co-author: Giovanna Menardi [1]

[1] *University of Padova*

Since the early Sixties, the Standard Model has represented the state of the art in High Energy Physics. It describes how the fundamental particles interact with each others and with the forces between them, giving rise to the matter in the universe. Despite its empirical confirmations, there are indications that the Standard Model does itself not complete our understanding of the universe. Model independent search aims to explain the shortcomings of this theory by empirically looking for any possible signal which behaves as a deviation from the background process, representing, in turn, the known physics.

From a statistical perspective, this problem can be in principle formulated within an unsupervised framework of clustering. However, while the the signal, if present, is unknown, the background process is always present and well-known, so that a virtually infinite sample of data can be simulated from the latter process with Montecarlo techinques. Hence, available data have two different sources: an unlabelled sample which might include observations from both the processes, and an additional labelled, sample from the background only. A semisupervised approach can be particularly suitable in this context, for discriminating the two class labels; semisupervised classification techniques lie between unsupervised and supervised ones, sharing some characteristics of both the approaches. In this work we propose a procedure where additional information, available on the background, is integrated within a nonparametric clustering framework to detect deviations from known physics. Also, we propose a variable selection procedure that allows to work on a reduced subspace.

The effectiveness of the whole methodology is shown via its application on a set of data related to a simulated experiment of a proton-proton collision.

孔子 / **9**

TBA

Author: Junhui Fan[1]

[1] *Guangzhou University Center for Astrophysics*

孔子 / **10**

New Statistical Analyses for the Neutrino Mass Hierarchy Determination

Author: Emilio Ciuffoli[1]

Co-author: Luca Stanco [2]

[1] *Chinese Academy of Sciences*

[2] *INFN Padova*

We preliminary focus on some aspects of the statistical analyses in particle physics.
Then, two new analyses on the neutrino mass hierachy (MA) will be presented.
The first focuses on the determination of MA with neutrino from accelerator beams, while the second one with antineutrino from reactor plants.

孔子 / **12**

State-Space Modeling and Identification of Large-Dimensional Multivariate Time-Series

Author: Giorgio Picci[1]

[1] *University of Padova*

I shall first discuss traditional ARMA models and the difficulties one has to face for the estimation of such models when dealing with high-dimensional multivariate time series. I shall then illustrate State Space and Dynamic Factor Analysis models which have been recently introduced in the literature. Estimation techniques which are reliable and numerically robust exist which have been used successfully in many applications.

A. Dent / **13**

Covariance-Consistent Vector ARMA Modeling

Author: Zhu Bin[1]

[1] *University of Padova*

A. Dent / **14**

Spatial and Time Clustering of the High-Energy Photons Collected by the Fermi LAT

Author: Denise Costantin[1]

[1] *Guangzhou University*

A. Dent / 15

Finite Dirichlet Mixture Modeling for Classification and Detection of New Classes of Variable Stars

Author: Prince John[1]

[1] *University of Padova*

A. Dent / 16

Model Independent Searches for New Physics via Parametric Anomaly Detection

Author: Grzegorz M Kotkowski[1]

[1] *University of Padova*

A. Dent / 17

Line-Profile Variations in Radial-Velocity Measurements Using a Skew Normal Distribution

Author: Umberto Simola[1]

[1] *University of Padova & Yale University*

A. Dent / 18

Searching for Gamma Ray Sources in the Extra-Galactic Space: A Statistical Analysis of the Fermi LAT Data

Author: Andrea Sottosanti[1]

[1] *University of Padova*