COSMOSTATS 2023 - Bridging the Gap: Statistical Modeling of Cosmology Extremes

Tuesday, 12 September 2023 - Friday, 15 September 2023

Astrophysical Observatory at Monte Pennar, University of Padova

Book of Abstracts

DESCRIPTION

.

Modern cosmology heavily relies on statistical methods. At the same time, the recent explosion of available data, resulting from current and forthcoming astronomical sky surveys, provides unique challenges, and opportunities, to statistics. Successful collaboration grows from the ground. For this to happen, researchers, especially junior ones, need opportunities to meet each other and to exchange ideas.

The aim of this project is to promote novel research and fuel new discoveries at the interface of two areas of much current interest: the statistical theory of extreme values and the study of cosmological observables. Statistics of extremes concerns inference for rare, possibly unobserved events. In cosmology, these may refer to the cosmic microwave background radiation –the oldest light we can see –, the clustering properties of large scale structure tracers and the cosmological signatures of gravitational lensing.

This will be achieved by means of a three-day intensive residential workshop open to a limited number of highly motivated junior participants. In particular, the workshop will intertwine presentations of the state-of-the-art of both fields, roundtable discussion, peer learning and coaching activities with the intent to foster the interchange of ideas among junior and senior researchers. Networking will be facilitated by the relaxed and enriching environment of the workshop venue.

.

Participation to the workshop is upon invitation, though all session talks will be live-streamed.

.

Asiago, September 12, 2023

.

Alessandra R. Brazzale
Valérie Chavez-Demoulin
Sabino Matarrese

.

.

Department of Statistical Sciences	Department of Physics and Astronomy "G. Galileo"	Faculty of Business and Economics
University of Padova	**University of Padova**	**University of Lausanne**

Figure 1: A collaborative project within the framework of the UNIL - UNIPD 2022 joint call for projects

Contents

Statistics / 33

An Introduction to Extreme Value Theory: from Basic Results to Tail Risk Inference in Time Series

Author: Simone Padoan[1]

[1] *Bocconi University*

In this talk we introduce the basic results of univariate Extreme Value Theory (EVT) that give rise to two of the most popular statistical methods for analysing extremes used in applications: the block maxima and peaks-over-threshold. We discuss one of the most important practical aims of EVT that is the tail risk estimation. We introduce expectiles as an alternative risk measure to the commonly used value at risk. We provide a basis for inference on extreme expectiles in a general β-mixing context that encompasses ARMA and GARCH models with heavy tailed innovations. We show the utility of the proposal when the data are serially dependent with simulations and a real data application.

Statistics / 34

From Classics to Punk: Almost 50 Years of Extreme Value Analysis

Author: Stefano Rizzelli[1]

[1] *University of Padova*

The statistical theory for extremes was initiated by J. Pickands III (1975). Since then, a plethora of frequentist methods (based or not on likelihood functions) have been proposed to estimate tail probabilities. In this field, Bayesian methods were first developed in the 1990s, however they have gained popularity only recently. On the one hand, this is due to the availability of machines with increasing computing power, allowing to speed up the calculation routines required for inference. On the other hand, this is due to the increasing focus on prediction, naturally incorporated in a Bayesian setting, which offers convenient tools for probabilistic forecasting. This talk illustrates the developments in statics of extremes, spanning from estimation techniques based on empirical processes to Bayesian inferential procedures. It concludes skimming over some more recently proposed hybrid methods (empirical Bayes).

Welcome / 29

Welcome by workshop organisers

Authors: Alessandra R Brazzale[1]; Sabino Matarrese[2]; Valérie Chavez-Demoulin[3]

[1] *University of Padova - DSS*

[2] *University of Padova - DFA*

[3] *University of Lausanne*

Welcome / 30

Welcome by Chair of Dept of Physics and Astronomy "G. Galileo"

Author: Flavio Seno[1]

[1] *University of Padova - DFA*

Welcome / 31

Welcome by Chair of Dept of Statistical Sciences

Author: Giovanna Boccuzzo[1]

[1] *University of Padova*

Welcome / 32

Welcome by Dean of Faculty of Business and Economics

Author: Marianne Schmid Mast[1]

[1] *University of Lausanne*

Welcome / 36

On-site information

Author: Alessandra R Brazzale[1]

[1] *University of Padova - DSS*

Cosmology / 38

An Introduction to Some Cosmological Observables: from Inflation and Gravitational Waves to non-Gaussianity

Author: Nicola Bartolo[1]

[1] *University of Padova*

In this talk I will provide a brief overview of some key words in the standard dictionary used in Cosmology, including some of the basic statistical tools to investigate various cosmological processes (focusing in particular on the so called Inflationary phase during the Early Universe and the generation of primordial gravitational waves). I will focus on the physics behind some of these processes that might be responsible for distinct statistical signatures in cosmological observables, such as non-Gaussianity of the primordial density fields leading to structure formation in the Universe, specific aspects of primordial gravitational waves, statistical "anomalies" in the Cosmic Microwave Background. The talk is mainly intended to set up a common language between the two main audiences of the workshop.

Cosmology / 37

The non-Gaussian Universe: a Challenge in Cosmological Data Analysis

Author: Michele Liguori[1]

[1] *University of Padova*

One of the main goals of observational cosmology is the measurement of cosmological parameters, using the Cosmic Microwave Background (CMB) or the galaxy distribution 2-point function (power spectrum). This procedure would optimally extract all cosmological information if the CMB and galaxy density fields were perfectly Gaussian. Non-Gaussian features are however imprinted in these fields, both through gravitational evolution of cosmic structures and through possible non-linear interactions during the primordial inflationary process. Cosmological non-Gaussianity is therefore a powerful tool to test inflation, improve our constraints on cosmological parameters and get a better understanding of the structure formation process. Its observational and statistical study is a complex data analysis task, which I will discuss in this talk.

Poster session / 44

Efficient Disentangling γ-Ray Sources from Diffuse Background in the Sky Map

Author: Francesco Freni[1]

[1] *University of Padova*

Searching for as yet undetected γ-ray sources is a major target of the Fermi LAT Collaboration. We address the problem by clustering the directions of the high-energy photon emissions detected by the telescope onboard the Fermi spacecraft. Putative sources are identified as the excess mass of disconnected high density regions on a sphere mesh, which allows for their joint discrimination from the diffuse γ-ray background spreading over the entire area. Density is estimated nonparametrically via binned directional kernel methods. The identification is accomplished by breaking the problem into independent subregions of the sphere separated by empty bins, thus leading to a remarkable gain in efficiency.

Joint work with Giovanna Menardi (University of Padova).

Poster session / 27

Skew-Symmetric Approximations of Posterior Distributions

Author: Francesco Pozza[1]

[1] *University of Padova - DSS*

In Bayesian statistics, deterministic approximations of the posterior distribution are often the preferred choice in the case of complex models, mainly for computational reasons. A common drawback of many of these approximations is that they usually belong to the Gaussian family and, therefore, can miss important characteristics of the posterior such as asymmetry. To alleviate this issue, this work proposes the use of a new family of approximations that provide accurate results, in a wide variety of different settings, and it is based on a simple, skew-inducing, perturbation of a Gaussian density. This new methodological proposal is accompanied both by rigorous theoretical studies and

a real-world data example which confirms that our method is competitive also with respect to some state-of-the-art alternatives.

Joint work with Daniele Durante and Botond Szabo (both from Bocconi University).

Poster session / 43

A Statistical Formulation of Conjunction Assessment

Author: Soumaya Elkantassi[1]

[1] *University of Lausanne*

Satellite conjunctions involving "near misses" of space objects are becoming increasingly likely. One approach to risk analysis for them involves the computation of the collision probability, but this has been regarded as having some counterintuitive properties and its interpretation has been debated. We propose a new approach based on a simple statistical model and discusses inference on the miss distance between the two objects, both when the relative velocity can be taken as known and when its uncertainty must be taken into account. The ideas are illustrated with case studies and Monte Carlo results that show its excellent performance. More details of this work can be also found in [1].

Joint work with Anthony C. Davison (Swiss Federal Institute of Technology).
[1] Elkantassi, S.; Davison, A. C. Journal of Guidance, Control, and Dynamics 2022, 45, 2258–2274.

Poster session / 45

Locating γ-Ray Sources on the Celestial Sphere via Modal Clustering

Author: Anna Montin[1]

[1] *Alira Health*

Searching for as yet undetected γ-ray sources is a major target of the Fermi LAT Collaboration. This type of high-energy photon emission typically presents itself as a highly concentrated point-like spot in the whole sky map, which blends in with the irregularly shaped background emission spread over the entire area. The identification of high-energy emitting sources is a fundamental task to better understand the mechanisms that both create and accelerate particles emitted by celestial objects. We discuss the application of nonparametric clustering for γ-ray source detection via an adjustment of the mean-shift algorithm to the directional nature of the data. The issue of selecting the smoothing amount is addressed adaptively, by combining scientific input with optimal selection guidelines, as known from the literature. Using statistical tools from hypothesis testing and classification, we furthermore present an automatic way to skim off sound candidate sources from the γ-ray emitting diffuse background and to quantify their significance. Efficient tools to account for the computational burden required to analyse huge amounts of data are also discussed. Our method was calibrated on simulated data provided by the Fermi LAT collaboration and will be illustrated on a real Fermi LAT case-study.

Joint work with Alessandra R. Brazzale and Giovanna Menardi (both from University of Padova).

Poster session / 28

Empirical Bayes Inference for the Peaks Over a Threshold Method

Author: Pietro Scanzi[1]

[1] *Bocconi University*

The Block Maxima and Peaks over a Threshold (POT) methods are the most popular statistical procedures used to analyse univariate extremes by means of the Generalised Extreme Value (GEV) and Generalised Pareto (GP) distributions, respectively. Exploiting the three-parameters GEV family of distributions as an asymptotic approximation for the underlying data distribution when this is computed for suitably large values, one can deduce the so-called censored-likelihood inferential procedure for extremes. Unlike the POT method that uses only the threshold excedances of a sample, the censored likelihood is a valid alternative as it relies on the entire dataset. We propose a Bayesian inferential approach for the estimation of extreme events based on the GEV censored-likelihood. We show its practical utility and compare its performance with well-known competitors.

Joint work with Simone Padoan (Bocconi University) and Nicola Sartori (University of Padova).

Poster session / 35

Efficient Estimation of Mixed Effect Models via Variational Message Passing

Author: Cristian Castiglione[1]

[1] *University of Padova - DSS*

Abstract: Nowadays, approximate Bayesian methods, such as integrated nested Laplace approximation, variational Bayes, expectation propagation and stochastic variational inference, are routinely used in statistics for the estimation of complex hierarchical models. They are particularly convenient, if not necessary, when Markov chain Monte Carlo algorithms can not be employed due to memory or time constraints. In these cases, minimal assumptions on the regularity of the likelihood and the conditional conjugacy of the prior, eventually after some model transformation via data augmentation, must be imposed in order to obtain tractable computations. As an alternative, we propose a simple and efficient variational message passing procedure to approximate the posterior density function of additive and mixed regression models without requiring either differentiability or conjugacy. Generalized linear models, support vector machines, quantile, robust and sparse regression can be naturally accommodated using the proposed approach, which also allows for many generalizations to more structured model specifications and stochastic optimization schemes. Simulation studies and real data applications confirm that the proposed method enjoys increasing computational and statistical advantages over alternative gold standard methods as the dimension and complexity of the model grow.

Poster session / 48

A Modified Dividing Local Gaussian Processes Algorithm for Theoretical Particle Physics Applications

Author: Timo Braun[1]

[1] *University of Oslo*

The computational cost of individual likelihood evaluations and physics simulations is a key limiting factor for BSM global fits and other large-scale parameter scans. One approach to tackle this is to use fast, pretrained emulators for the most expensive computations. However, as the set of relevant experimental results is frequently updated, many pretrained emulators have limited reusability. Here we propose an approach for training and applying fast emulators on the fly during parameter scans, based on the Dividing Local Gaussian Processes (DLGP) algorithm of Lederer et al. During a scan, the DLGP algorithm iteratively divides the input space using an evolving binary tree, where each leaf contains a local Gaussian process (GP) emulator. Guided by the typical computational requirements of BSM global fits, we extend the DLGP approach to improve its prediction accuracy. Our modifications include the use of new covariance functions, more detailed and frequent retraining of the local GPs, and new approaches to performing the iterative division of the input space. We demonstrate our approach on data from a recent global fit by the GAMBIT Collaboration.

Joint work with Riccardo De Bin and Anders Kvellestad (both from University of Oslo)

Poster session / 47

Bayesian Clustering of Multivariate Extremes

Author: Sonia Alouini[1]

[1] *Swiss Federal Institute of Technology Lausanne*

In extreme value theory, the dependence structure between multivariate exceedances over a high threshold is fully characterised by their projections on the unit simplex. Under mild conditions, the only constraint on these angular variables is that their marginal means are equal. Their distribution functions thus form a non-parametric class within which deriving flexible and easy to use models is challenging, especially in high dimensions. Dirichlet mixtures are natural candidates to approximate such functions, but they are not necessarily valid angular distributions themselves. Previous approaches constrained the Dirichlet parameters in order to enforce the marginal mean property but the implementation of such models tends to be too slow especially in high dimensions. Instead of constraining the parameters, we let them vary freely and apply a transformation to the whole mixture in order to tilt the marginal means towards their desired values. The tilted mixtures of Dirichlet distributions are a new class of functions that are dense in the space of angular distributions and well-defined in all dimensions. We propose an MCMC procedure which is fast in all dimensions and does not require fine tuning. Furthermore, the mixture captures heterogeneity in the extremal dependence structure and allow the probabilistic clustering of observations. We demonstrate the performance of the proposed model on simulated data and show its usefulness on financial applications.

How 2's / 41

The Euclid Spectroscopic Sample: a Unique Opportunity to Test General Relativity

Author: Elena Sarpa[1]

[1] *International School for Advanced Studies*

The Euclid survey will soon deliver the most extensive and densely populated galaxy catalogue ever observed, reaching objects as distant as 10 billion light-years away from us. By precisely mapping the 3D galaxy distribution, the spectroscopic catalogue presents a unique opportunity for measuring the expansion of the Universe and the growth of structures with sub-percent precision - a requirement for a definitive answer on the validity of general relativity.
In this talk, I will guide you through the planned Euclid analysis by introducing the spectroscopic catalogue and describing three pivotal probes selected to extract cosmological information. We will

see how the signature of Baryonic acoustic oscillations and redshift-space distortions in the two-point statistics of the galaxy distribution allow us to measure the expansion of the Universe and the growth of structure, respectively, and explore the study of the morphology and N-point statistics of underdense regions (cosmic voids) as a complementary probe to the clustering studies.

Statistics / 39

Spatial Extremes

Author: Anthony C. Davison[1]

[1] *Swiss Federal Institute of Technology Lausanne*

Statistics / 40

Machine Learning for Extreme Values

Author: Sebastian Engelke[1]

[1] *University of Geneva*

Cosmology / 42

Non-Gaussian Statistics and Non-Gaussian Tensions

Author: Marco Raveri[1]

[1] *University of Genova*

I will review cosmological results obtained using non-Gaussian statistics with the Dark Energy Survey's (DES) latest data release. This will allow us to discuss statistical methods to probe discrepancies between cosmological data, without assumptions of Gaussianity in data and parameter space. I will show the worked examples of cosmological constraints from moments and peaks of the lensing mass field, as measured by DES.

Cosmology / 46

Statistical Analysis and Exploitation of Sky Maps in Cosmic Microwave Background Observations

Author: Carlo Baccigalupi[1]

[1] *International School for Advanced Studies*

We focus on the statistical properties of cosmological processes appearing in CMB anisotropies, as well on the associated aspects of Data Analysis. We consider the total intensity of the large fields observed by Planck and the associated constraints of the statistics of perturbations in the Early Universe, as well as the polarization measurements from operating and future probes looking in

particular at the curl component (B-Mode), which may be sourced by cosmological Gravitational Waves and Gravitational Lensing. We outline the landscape of observations and the production and analysis of maps. We describe how the different statistics of the observed emissions lead to specialized analysis, reduction, and even exploitation procedures, focusing on the impact of these methodologies on the Early Universe and Cosmic Acceleration

How 2's / 49

CRAN Task View: Extreme Value Analysis

Author: Anthony C. Davison[1]

[1] *Swiss Federal Institute of Technology Lausanne*

Networking / 50

Tripadvisor's list of "Things to do in Asiago"

https://www.tripadvisor.com/Attractions-g194677-Activities-oa0-Asiago_Province_of_Vicenza_Veneto.html

Networking / 51

Wanderlog's "Top 13 things to do and attractions in Asiago"

https://wanderlog.com/list/geoCategory/1587973/top-things-to-do-and-attractions-in-asiago

Networking / 52

My Corner of Italy's suggestions of "What to do in Asiago apart from eating cheese?"

https://mycornerofitaly.com/asiago/

COMMITTEES

.

Scientific Committee
- Alessandra R Brazzale (UniPD)
- Valérie Chavez-Demoulin (UniL)
- Sabino Matarrese (UniPD)

.

Local Organizing Committee
- Alessandra R Brazzale (UniPD)
- Sabino Matarrese (UniPD)
- Alessandro Siviero (OAPd)

.

Secretariat
- Elena Guzzonato (UniPD)