NEW FRONTIERS IN STATISTICS OF EXTREMES

Programme and Book of Abstracts

https://workshopnfsextremes2020.weebly.com/

February 6–7, 2020

Universidade de Lisboa, Portugal

About this workshop: This workshop will gather experts on Statistics of Extremes with the aims of breaking new ground on the statistical modeling of extreme events, and of shaping new frontiers for future research. The event is organized under the umbrella of the research project *Data Fusion and Calibration Methods for Spatial Risk Analysis* from Fundação para a Ciência e Tecnologia (FCT), Portugal.

Programme

- **Thursday 6** (Morning: P. de Zea Bermudez; Afternoon M. de Carvalho)
  - 8:30–8:50: Registration.
  - 8:50–9:00: Welcome Session.
  - 9:00–9:55: Laurens de Haan (Universidade de Lisboa and CEAUL and Erasmus University Rotterdam).
  - 10:00–10:55: Ana Ferreira (IST—Universidade de Lisboa, CEMAT and CEAUL).
  - 11:00–11:30: Coffee Break.
  - 11:30–12:25: Anna Kiriliouk (University of Namur).
  - 12:30–14:00: Lunch (not included).
  - 14:00–14:55: Alan Gelfand (Duke University).
  - 15:00–15:25: Junho Lee (University of Edinburgh).
  - 15:30–16:00: Coffee Break.
  - 16:00–16:55: Soraia Pereira (Universidade de Lisboa and CEAUL).
  - 17:00–17:55: Daniela Castro-Camilo (University of Glasgow).
- **Friday 7** (Morning: K. Feridun Turkman; Afternoon P. Pereira)
  - 9:00–9:55: John Einmahl (Tilburg University).
  - 10:00–10:55: Paulo M. M. Rodrigues (Banco de Portugal and NOVA School of Business and Economics).
  - 11:00–11:30: Coffee Break.
  - 11:30–12:25: Miguel de Carvalho (University of Edinburgh).
  - 12:30–14:00: Lunch (not included).
  - 14:00–14:55: Ioannis Papastathopoulos (University of Edinburgh).
  - 15:00–15:25: Jessica Silva Lomba (NOVA School of Business and Economics and CEAUL).
  - 15:30–16:00: Coffee Break.
  - 16:00–16:55: M. Ivette Gomes (Universidade de Lisboa and CEAUL).
  - 17:00–17:45: Discussion. (Moderator: M. de Carvalho)

Thursday 6

Spatial Dependence and Space-time Trend in Extreme Events (Part I)

Laurens de Haan

Universidade de Lisboa, Portugal and Erasmus University Rotterdam, Netherlands

ldehaan@ese.eur.nl

Abstract: The statistical theory of extremes is extended to observations that are non-stationary and not independent. The non-stationarity over time and space is controlled via the scedasis (tail scale) in the marginal distributions. Spatial dependence stems from multivariate extreme value theory. We establish asymptotic theory for both the weighted sequential tail empirical process and the weighted tail quantile process based on all observations, taken over time and space. The results yield two statistical tests for homoscedasticity in the tail, one in space and one in time. Our leading example and application is rainfall in Northern Germany.

†**Time slot**: 9:00–9:55.

Spatial Dependence and Space-time Trend in Extreme Events (Part II)

Ana Ferreira

Instituto Superior Técnico da Universidade de Lisboa, CEMAT and CEAUL, Portugal

anafh@tecnico.ulisboa.pt

Abstract: Continuing the same setting as in "Spatial Dependence and Space-time Trend in Extreme Events (Part I)" by Laurens de Haan: In the first part of the talk we show that the common extreme value index can be estimated via a pseudo-maximum likelihood procedure based on pooling all (non-stationary and dependent) observations. In the second part of the talk, we discuss results after applying the new statistical tests for evaluating time-trends and spatial inhomogeneities, to daily accumulated precipitation observed at 64 weather stations in north-western Germany. It is considered the time period from 1931 to 2014 (84 years), separately for northern hemispheric 'winter' (November–March) and 'summer' (May–September). The results reveal a clear seasonal differentiation on precipitation extremal behaviour regarding intensity, frequency, trend and spatial dependence. After the removal of non-stationary effects, extremal dependence over space is still evaluated through variogram analysis.

†**Time slot**: 10:00–10:55.

Climate Extreme Event Attribution using Multivariate Peaks-over-thresholds Modeling and Counterfactual Theory

Anna Kiriliouk

University of Namur, Belgium

anna.kiriliouk@unamur.be

Abstract: Numerical climate models are key tools to quantify the relative contribution of potential anthropogenic causes (e.g., the current increase in greenhouse gases) on high impact atmospheric variables like heavy rainfall. These so-called climate extreme event attribution problems are particularly challenging in a multivariate context, that is, when the atmospheric variables are measured on a possibly high-dimensional grid. In this paper, we leverage two statistical theories to assess causality in the context of multivariate extreme event attribution. As we consider an event to be extreme when at least one of the components of the vector of interest is large, extreme-value theory justifies, in an asymptotical sense, a multivariate generalized Pareto distribution to model joint extremes. Under this class of distributions, we derive and study probabilities of necessary and sufficient causation as defined by the counterfactual theory of Pearl. Our approach is then applied to weekly winter maxima precipitation outputs of the French CNRM from the recent CMIP6 experiment. Joint work with Philippe Naveau.

†**Time slot**: 11:30–12:25.

Long-term Spatial Modeling for Characteristics of Extreme Heat Events

Alan Gelfand

Duke University, USA

alan@stat.duke.edu

Abstract: There is increasing evidence that global warming manifests itself in more frequent warm days and that heat waves will become more frequent. At present, a formal definition of a heat wave is not agreed upon in the literature. To avoid this debate, we consider extreme heat events which are well defined at local scales. Specifically, an extreme heat event (EHE) is defined as a run of consecutive days above a specified local threshold. Characteristics of EHEs are of primary interest, such as incidence and duration, as well as the magnitude of the average exceeedance and maximum exceedance above the threshold during the EHE. Using approximately 60 years of time series consisting of daily maximum temperature data collected at 18 locations in and around Aragon in Spain, we develop a spatio-temporal model to study the behavior of EHEs over time. The model enables prediction of EHE behavior in terms of the foregoing characteristics at unobserved locations within the region. Specifically, our approach employs a two-state space-time model for EHEs with local thresholds to fit to the daily maximum temperature data where one state defines above threshold temperatures and the other below threshold temperatures. We show that our model is able to recover the EHE characteristics of interest and, in terms of out-of-sample prediction, outperforms a customary autoregressive model that ignores thresholds. †**Time slot**: 14:00–14:55.

Bayesian Modelling of Time-varying Extremal Dependence in International Stock Markets

Junho Lee

University of Edinburgh, UK

s1634014@sms.ed.ac.uk

Abstract: In this talk I will propose a Bayesian time-varying model to capture the dynamics of joint extremes over time. The proposed model relies on dual non-parametric time-varying extremal dependence measures which can be used assess the strength of dependence of joint extremes over time under the settings of asymptotic dependence and asymptotic independence. The proposed Bayesian smoothing method is motivated by the need of examining how the occurrence and magnitude of extreme values may change along with a covariate. The model is built using a Pareto-type specification for the tail of the response, with covariates being modelled via generalized additive model (GAM) and Bayesian P-splines. Finally, details on computational implementation will be discussed over the talk, along with a set of illustrations on simulated and real data. We will cover the main results from a Monte Carlo experiment. †**Time slot**: 15:00–15:55.

An Extreme Value Bayesian Lasso for the Conditional Bulk and Tail

Soraia Pereira

Universidade de Lisboa and CEAUL, Portugal

sapereira@fc.ul.pt

Abstract: In this talk, I will introduce a novel regression model for the conditional bulk and conditional tail of an heavy-tailed response. The proposed model can be used to learn the effect of covariates on a heavy-tailed response via a Lasso-type specification based on a Lagrangian restriction. Our model can be used to track if some covariates are significant for the bulk, but not for the tail—and vice-versa; in addition to this, the proposed model bypasses the need for conditional threshold selection in an extreme value theory framework. We assess the finite-sample performance of the proposed methods through a simulation study that reveals that our method recovers the true conditional distribution over a variety of simulation scenarios, along with being accurate on variable selection. Rainfall data are used to showcase how the proposed method can learn to distinguish between key drivers of moderate rainfall, against those of extreme rainfall. Joint work with S. Pereira, P. Pereira, and P. de Zea Bermudez. †**Time slot**: 16:00–16:55.

Extreme Wind Speed Probabilistic Forecasting with INLA

Daniela Castro-Camilo

University of Glasgow, UK

Daniela.CastroCamilo@glasgow.ac.uk

Abstract: Renewable sources of energy such as wind power have become a sustainable alternative to fossil fuel-based energy. However, the uncertainty and fluctuation of the wind speed derived from its intermittent nature bring a great threat to the wind power production stability, and to the wind turbines themselves. Lately, much work has been done on developing models to forecast average wind speed values, yet surprisingly little has focused on proposing models to accurately forecast extreme wind speeds, which can damage the turbines. In this work, we develop a flexible spliced Gamma-Generalized Pareto model to forecast extreme and non-extreme wind speeds simultaneously. Our model belongs to the class of latent Gaussian models, for which inference is conveniently performed based on the integrated nested Laplace approximation method. Considering a flexible additive regression structure, we propose two models for the latent linear predictor to capture the spatio-temporal dynamics of wind speeds. Our models are fast to fit and can describe both the bulk and the tail of the wind speed distribution while producing short-term extreme and non-extreme wind speed probabilistic forecasts. †**Time slot**: 17:00–17:55.

Friday 7

Empirical Tail Copulas for Functional Data

John H. J. Einmahl

Tilburg University, Netherlands

J.H.J.Einmahl@tilburguniversity.edu

Abstract: For multivariate distributions in the domain of attraction of a max-stable distribution, the tail copula and the stable tail dependence function are equivalent ways to capture the dependence in the upper tail. The empirical versions of these functions are rank-based estimators whose inflated estimation errors are known to converge weakly to a Gaussian process that is similar in structure to the weak limit of the empirical copula process. We extend this multivariate result to continuous functional data by establishing the asymptotic normality of the estimators of the tail copula, uniformly over all finite subsets of at most D points (D fixed). As a special case we obtain the uniform asymptotic normality of all estimated upper tail dependence coefficients. The main tool for deriving the result is the uniform asymptotic normality of all the D-variate tail empirical processes. The proof of the main result is non-standard. Joint work with Johan Segers. †**Time slot**: 9:00–9:55.

Measuring Income Inequality under Right Censoring

Paulo M. M. Rodrigues

Banco de Portugal and NOVA School of Business and Economics, Portugal

pmrodrigues@bportugal.pt

Abstract: This paper investigates potential changes that may have occurred in the probability mass of the right tail of the wage distribution over the last two decades through the analysis of changes in the corresponding tail index. In specific, a new tail index estimator is introduced which explicitly allows for right tail censoring (topcoding) of the data, which is a feature of the widely used current population survey (CPS), as well as other surveys. The CPS income data has historically been topcoded for confidentiality reasons, and ignoring this fact may lead to an understatement of inequality and of its evolution over time. Thus, having a tail index estimator that explicitly accounts for this feature is of importance to better understand and compute the tail index dynamics in the censored right tail of a distribution. The contribution of this paper is threefold. One is the introduction of a new conditional tail index estimator that explicitly handles the top coding problem, and an evaluation of its finite sample performance and comparison with competing methods; it is shown that the factor values for ad hoc adjustment of topcoded earnings changed over time and across individual characteristics; and finally an in depth empirical analysis of the dynamics of the wage distribution's right tail using the public-use CPS database from 1992 to 2017 is also provided.

†**Time slot**: 10:00–10:55.

Bayesian Conditionally Heteroscedastic Extremes

Miguel de Carvalho

University of Edinburgh, UK

Miguel.deCarvalho@ed.ac.uk

Abstract: In this talk, I will introduce a Bayesian semiparametric model for heteroscedastic extremes. The proposed model is based on an tail index regression and on a proportional tails model, and can be used for assessing how the magnitude and frequency of the extreme values can change along with a covariate. We start with the unconditional setting for estimating the tail index and the scedasis function and show that the proposed inference methods for the scedasis density—based on a Bernstein–Dirichlet prior—perform well in Monte Carlo simulation studies, are exact apart from Monte Carlo error, and have full support on the space of all continuous scedasis functions. We then extend the proposed methods to the conditional setting using dependent Bernstein-Dirichlet process. We resort to the proposed methodologies to examine an extreme currency demand in Portugal. The signatures of the fitted scedasis densities of extreme currency demand—over different denominations—reveal some interesting insights on the dynamics governing currency demand during periods of economic stress. Joint work with Junho Lee and with António Rua.

†**Time slot**: 11:30–12:25.

Single-Variable Conditional Extremes on Decomposable Graphical Models

Ioannis Papastathopoulos

University of Edinburgh, UK

i.papastathopoulos@ed.ac.uk

Abstract: Decomposable graphical models form a flexible and tractable class of statistical models that facilitate inference in high-dimensions. The rationale is that the underlying dependence structure of a random vector is described by a set of lower dimensional distributions which are easier to infer in practice. For risk assessment, it is the extreme events of random vectors that are of most practical concern. We illustrate how recent theory for the limiting joint behaviour of Markov chains, conditioned on extreme states, can be lifted to obtain extreme value limit theorems on decomposable graphs. We present statistical extreme value models and discuss likelihood based procedures to infer how extremes propagate on graphs.

†**Time slot**: 14:00–14:55.

Size-over-Threshold: The Myth of the Kraken

Jessica Silva Lomba

NOVA School of Business and Economics and CEAUL, Portugal

jslomba@fc.ul.pt

Abstract: The *Architeuthis* sp. giant squid's maximum size has been a controversial enigma amongst ecologists, given they live at large depths and are rarely seen or captured. To our knowledge, extreme value analysis (EVA) has not yet been considered for this problem. Being one of the largest invertebrates, their only known predator is the sperm whale, but the recovered eaten beaks suggest this predation occurs mostly while squids are immature. We use a data set of giant squid lengths, compiled by Paxton (2016, *Journal of Zoology*), to estimate the beast's maximum size under the EVA umbrella, while illustrating a novel automatic approach for threshold selection, built on Hosking's L-moments, for peaks-over-threshold (POT) analysis Silva Lomba and Fraga Alves (2019, arXiv: 1905.08726). The presented technique, aiming at efficiency for large/simultaneous batch analysis, without compromise for performance on small samples, and elimination of subjective judgment, is heuristic in nature. We evaluate its performance by simulation for various i.i.d. scenarios, including misspecification and data quantization, with some state-of-the-art comparisons. Alternative automatic methodologies, based on L-moments' asymptotics, were considered. We conclude on the plausibility of Kraken-sized giant squids by estimation of return levels and of the right endpoint of fitted Generalized Pareto distributions, with a look at competitive techniques. Joint work with Maria Isabel Fraga Alves.

§Research partially supported by Fundação para a Ciência e a Tecnologia (FCT) (SFRH/BD/130764/2017, UIDB/00006/2020).

†**Time slot**: 15:00–15:25.

Non-Normal Generalized Means in Statistical EVT

M. Ivette Gomes

Universidade de Lisboa and CEAUL, Portugal

migomes@fc.ul.pt

Abstract: The Hill estimator, the most popular extreme value index (EVI) estimator in a heavy-tailed framework, can be regarded as the logarithm of the geometric mean, i.e. the power mean of order $p = 0$ of a certain set of statistics. And instead of such a geometric mean, we can more generally consider the mean of order p (MOp), of those statistics, with p any real number, and to use it in order to build a class of MOp EVI-estimators. Similar procedures can be used for the estimation of other parameters of extreme events, like the value-at-risk, the expected shortfall and the tail dependence coecient. The normal asymptotic behaviour of MOp EVI-estimators has already been obtained for $p < 1/(2\xi)$, with consistency achieved for $p < 1/\xi$, where ξ denotes the EVI. We shall now consider the non-regular case, $1/(2\xi) \leq p \leq 1/\xi$, a situation in which a non-normal sum-stable law is obtained. An algorithm for an 'almost degenerate' EVI-estimation is suggested and studied for finite samples, through a large-scale simulation study. Also, an application to simulated and real data is performed.

§Research partially supported by Fundação para a Ciência e Tecnologia (FCT).

†**Time slot**: 16:00–16:55.

Sponsors and Partners

CMStatistics Team, Statistics of Extremes & Applications

Organizing and Scientific Committee

Chairs

- Patrícia de Zea Bermudez, Universidade de Lisboa.
- Miguel de Carvalho, University of Edinburgh.

Members

- M. A. Amaral Turkman, Universidade de Lisboa.
- L. de Haan, Universidade de Lisboa and Erasmus University Rotterdam.
- M. González Scotto, Universidade de Lisboa.
- P. Pereira, Instituto Politécnico de Setúbal.
- S. Pereira, Universidade de Lisboa.
- K. F. Turkman, Universidade de Lisboa.

Financial support: FCT Portugal UID/MAT/00006/2019 and PTDC/MAT-STA/28649/2017.