

ANALYSIS

S
SECRET
SECRET

THE UNIVERSITY of EDINBURGH
Centre for Statistics



Contents

EVA2021 28/06 - 02/07 Edinburgh 2

1	Invited and contributed talks	3
2	Posters	134

1 Invited and contributed talks

Valid sequential inference on probability forecast performance

Alexander Henzi

University of Bern

Abstract

Probability forecasts for binary events play a central role in many applications. Their quality is commonly assessed with proper scoring rules, which assign forecasts a numerical score such that a correct forecast achieves a minimal expected score. In this paper, we construct e -values for testing the statistical significance of score differences of competing forecasts in sequential settings. E -values have been proposed as an alternative to p -values for hypothesis testing, and they can easily be transformed into conservative p -values by taking the multiplicative inverse. The e -values proposed in this article are valid in finite samples without any assumptions on the data generating processes. They also allow optional stopping, so a forecast user may decide to interrupt evaluation taking into account the available data at any time and still draw statistically valid inference, which is generally not true for classical p -value based tests. In a case study on postprocessing of precipitation forecasts, state-of-the-art forecasts dominance tests and e -values lead to the same conclusions.

Time and Session: Friday 11.30–13.00, CS Prediction and validation for extremes

Regression-type analysis for block maxima on block maxima

Alina Kumukova*, Miguel de Carvalho and Goncalo dos Reis

*University of Edinburgh, MIGSAA

Abstract

In this talk we introduce a regression-type model for the situation where both the response and covariates are extreme. The proposed approach is designed for the setting where both the response and covariates are themselves block maxima, and thus contrarily to standard regression methods it takes into account the key fact that the limiting distribution of suitably standardized componentwise maxima is an extreme value copula. An important target in the proposed framework is the regression manifold, which consists of a family of regression lines obeying the latter asymptotic result. To learn about the proposed model from data, we employ a Bernstein polynomial prior on the space of angular densities which leads to an induced prior on the space of regression manifolds. Numerical studies suggest a good performance of the proposed methods, and a finance real-data illustration reveals interesting aspects on the comovements of extreme losses between two leading stock markets.

Time and Session: Thursday 17.30–18.45, CS Regression techniques (II)

Convergence of extreme values of Poisson point processes at small times

Ana Ferreira*, Boris Buchmann and Ross A. Maller

*Instituto Superior Tecnico University of Lisbon

Abstract

We study the behaviour of large values of extremal processes at small times, obtaining an analogue of the Fisher-Tippett-Gnedenko Theorem. Thus, necessary and sufficient conditions for local convergence of such maxima, linearly normalised, to the Fréchet or Gumbel distributions, are established. Weibull distributions are not possible limits in this situation. Moreover, assuming second order regular variation, we prove local asymptotic normality for intermediate order statistics, and derive explicit formulae for the normalising constants for tempered stable processes. We adapt Hill's estimator of the tail index to the small time setting and establish its asymptotic normality under second order regular variation conditions, illustrating this with simulations. Applications to the fine structure of asset returns processes, possibly with infinite variation, are indicated.

Research partially supported by ARC DP160104737, FCT: UIDB/04621/2020, UIDP/04621/2020, SFRH/BSAB/142912/2018.

Time and Session: Friday 11.30–13.00, CS Extremes of stochastic processes (II)

Extreme value theory for spatial random fields, with application to a Lévy-driven field

Anders Rønn-Nielsen

Copenhagen Business School

Abstract

First, we consider a stationary random field indexed by an increasing sequence of discrete sets of spatial grid points. Under certain mixing and anti-clustering conditions combined with a very broad assumption on how the sequence of spatial index sets increases, we obtain an extremal result that relates a normalized version of the distribution of the maximum of the field over the index sets to the tail distribution of the individual variables. Furthermore, we identify the limiting distribution as an extreme value distribution. Secondly, we consider a continuous, infinitely divisible spatial random field given as an integral of a kernel function with respect to a Lévy basis with a levy measure that is either of the following: convolution equivalent; regularly varying; subexponential and in the Gumbel domain of attraction. When observing the supremum of this field over an increasing sequence of (continuous) index sets, we obtain an extreme value theorem for the distribution of this supremum. The proof relies on discretization and a conditional version of the technique applied in the first part, as we condition on the high activity and light-tailed part of the field.

Time and Session: Tuesday 11.30–13.00, CS Spatial extremes (I)

Functional strong laws of large numbers for Euler characteristic processes of extreme sample clouds

Andrew Thomas* and Takashi Owada

*Purdue University

Abstract

In recent years, significant progress has been made in understanding the stochastic topology of noise. In particular, researchers have looked at how topological features behave when they are based off an increasing number of random points in Euclidean space lying at ever greater distances from the origin. In this talk, we will look at how the Euler characteristic of a filtration of random geometric simplicial complexes behaves when the points that generate them come from two distinct families of extreme value distributions. We will demonstrate a functional strong law of large numbers (FSLN) for the Euler characteristic process—a summary of how this extremal topology evolves—in each distributional context.

Time and Session: Thursday 16.00–17.15, IS Long memory processes and non-standard EVT

Extreme values in negative curvature

Anish Gosh

Tata Institute of Fundamental Research

Abstract

This talk is meant to be an invitation for probabilists to enter into the world of ergodic theory in negative curvature. I will describe the geodesic flow on Riemannian manifolds of negative curvature via examples. These flows form a natural playing ground for ergodic theorists, probabilists and geometers, and have been extensively studied. I will then recall some prominent results on the asymptotic behaviour of geodesics including a dynamical phenomenon called the *shrinking target property*. Extreme Value theory in this rich setting has not yet been fully developed and I will end by stating some partial results and open problems.

Time and Session: Tuesday 11.30–13.00, IS Extremes random structures (branching and dynamics, geometry)

A 2×2 random switching model and its dual risk model

Anita Behme

TU Dresden

Abstract

A special case of two coupled M/G/1-queues is considered, where two servers are exposed to two types of jobs that are distributed among the servers via a random switch. In this model, the asymptotic behavior of the workload buffer exceedance probabilities for the two single servers/both servers together/one (unspecified) server is determined. Hereby, we distinguish between jobs that are either (multivariate) subexponential or (multivariate) regularly varying or light-tailed. The results are derived via the dual risk model of the studied coupled M/G/1-queues for which the asymptotic behavior of different ruin probabilities is determined.

Time and Session: Thursday 11.30–13.00, CS Insurance

Non-asymptotic bounds for probability weighted moment estimators

Anna Ben-Hamou*, Philippe Naveau and Maud Thomas

*Sorbonne University (LPSM)

Abstract

In hydrology and other applied fields, Probability weighted moments (PWM) have been frequently used to estimate the parameters of classical extreme value distributions ([de Haan and Ferreira 2006](#)). This method-of-moment technique can be applied when second moments are finite, a reasonable assumption in hydrology. Two advantages of PWM estimators are their ease of implementation and their close connection to the well-studied class of U-statistics. Consequently, precise asymptotic properties can be deduced. In practice, sample sizes are always finite and, depending on the application at hand, the sample length can be small, e.g. a sample of only 30 years of maxima of daily precipitation is quiet common in some regions of the globe. In such a context, asymptotic theory is on a shaky ground and it is desirable to get non-asymptotic bounds.

Deriving such bounds from off-the-shelf techniques (Chernoff method) requires exponential moment assumptions, which are unrealistic in many settings. To bypass this hurdle, we propose a new estimator for PWM, inspired by the median-of-means framework of [Devroye et al. \(2016\)](#). This estimator is then shown to satisfy a sub-Gaussian inequality, with only second moment assumptions. This allows us to derive non-asymptotic bounds for the estimation of the parameters of extreme value distributions, and of extreme quantiles.

References

- de Haan L. and Ferreira A., 2006. *Extreme Value Theory: An Introduction* Springer, New York
- Devroye L., Lerasle M., Lugosi G. and Oliveira R. I., 2016. *Annals of Statistics* **44**, 2695–2725

Time and Session: Monday 16.00–17.15, CS Inference and robust extremes

Scoring probabilistic forecasts with a focus on extremes

Anne-Laure Fougères

Institut Camille Jordan, Université Claude Bernard - Lyon 1

Abstract

The framework of this talk is the evaluation of probabilistic forecasts performances. We will in particular discuss the meteorological context and the situation where extreme events are of main interest.

Time and Session: Thursday 17.30–18.45, IS Forecasting, metrics, evaluations and scoring of extremes

Extreme expectile regression: theory and applications

Antoine Usseglio-Carleve

ENSAI Rennes/Toulouse School of Economics

Abstract

Expectiles and quantiles can both be defined as the solution of minimization problems. Contrary to quantiles though, expectiles are determined by tail expectations rather than tail probabilities, and define a coherent risk measure. For these two reasons in particular, expectiles have recently started to be considered as serious candidates to become standard tools in actuarial and financial risk management. However, expectiles and their sample versions do not benefit from a simple explicit form, making their analysis significantly harder than that of quantiles and order statistics. This difficulty is compounded when one wishes to integrate auxiliary information about the phenomenon of interest through a finite dimensional covariate, in which case the problem becomes the estimation of conditional expectiles. In this work, we firstly construct nonparametric kernel estimators of extreme conditional expectiles. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions, and realize that such a nonparametric approach is no longer suitable when the covariate dimension is too large. We thus also build a general theory for the estimation of extreme conditional expectiles in heteroscedastic regression models with heavy-tailed noise; our approach is supported by general results of independent interest on residual-based extreme value estimators in heavy-tailed regression models, and is intended to cope with covariates having a large but fixed dimension. We demonstrate how our results can be applied to a wide class of important examples, among which linear models, single-index models as well as ARMA and GARCH time series models. Our estimators are showcased on a numerical simulation study and on real sets of actuarial and financial data.

Time and Session: Friday 10.00–11.15, IS Insurance

Tail index regression-adjusted functional covariate

Anwar Alabdulathem

University of Edinburgh

Abstract

In this talk I will propose a statistical model the aims to assess how the extreme values of a random variable can change with a functional covariate. The proposed model can be understood as a regression model for heavy-tailed response and where the covariate is a random function. To learn about the proposed method a likelihood-based estimator is defined by solving an approximation to a certain variational calculus problem of integral. Simulation data are used to evaluate the performance of the proposed methods. Applications of the proposed methods are envisioned finance, in order to understand how the risk of an extreme loss in the stock market may change according to a certain functional covariate.

Time and Session: Thursday 17.30–18.45, CS Regression techniques (II)

Extreme value estimation of the conditional risk premium in reinsurance

Armelle Guillou*, Yuri Goegebeur and Jing Qin

*Strasbourg University and CNRS

Abstract

In the talk, we study the estimation of reinsurance premiums when the claim size is observed together with additional information in the form of random covariates. Using extreme value arguments, we propose an estimator for the risk premium conditional on a value for the covariate, and derive its asymptotic properties, after suitable normalization. The finite sample behavior is evaluated with a simulation experiment, and we apply the methodology to a dataset of automobile insurance claims from Australia.

Time and Session: Thursday 11.30–13.00, CS Insurance

A sparse Gaussian scale mixture process for modeling short-range extremal dependence and long-range independence

Arnab Hazra, Raphaël Huser and David Bolin

Abstract

Various natural phenomena, such as precipitation, generally exhibit extremal dependence at short distances only, while spatial dependence usually fades away as the distance between sites increases arbitrarily. However, the available methods in the literature for spatial extremes, which are based on max-stable or Pareto processes or comparatively less computationally demanding *sub-asymptotic* models based on location and/or scale mixtures, generally assume that spatial extremal dependence persists across the entire spatial domain. In this paper, we develop a novel Bayesian Gaussian scale mixture model, where the Gaussian process component is driven through a stochastic partial differential equation (SPDE) that yields a sparse precision matrix, and the random scale component is modeled as a low-rank Pareto-tailed or Weibull-tailed spatial process determined by compactly supported basis functions. We show that our model is tail-stationary, and we demonstrate that it can capture a wide range of extremal dependence structures as a function of distance. The sparse structure of our spatial model allows fast Bayesian computations, even in high spatial dimensions. In our application, we fit our model to analyze heavy monsoon rainfall data in Bangladesh. Our inference approach relies on a well-designed Markov chain Monte Carlo (MCMC) algorithm. A cross-validation study indicates that the proposed model outperforms some natural alternatives, and that the model provides a good fit to the rainfall data. Finally, we use the fitted model to draw inferences on long-term return levels for marginal rainfall at each site, and for spatial aggregates.

Time and Session: Tuesday 11.30–13.00, CS Spatial extremes (I)

On the disjoint and sliding block maxima method for piecewise stationary time series

Axel Bücher* and Leandra Zanger

*Heinrich-Heine-Universität Düsseldorf

Abstract

Modeling univariate block maxima by the generalized extreme value distribution constitutes one of the most widely applied approaches in extreme value statistics. Next to maximum likelihood, estimation based on matching probability weighted moments provides the most common estimation strategy, in particular for negative shape parameters. Within a traditional sampling scheme and a new sampling scheme involving certain piecewise stationarities, it is shown that the latter estimator may be improved by calculating block maxima in an overlapping way. Irrespective of the serial dependence, the estimation variance is shown to be smaller for the new estimator, while the bias often stays the same or varies comparably little. The results are illustrated by Monte Carlo simulation experiments and are applied to a common situation involving temperature extremes in a changing climate.

Time and Session: Friday 16.00–17.15, IS Asymptotic statistics (inc. empirical processes)

Estimation of extreme conditional quantiles with coupling method.

Benjamin Bobbia

Université de Cergy-Pontoise

Abstract

The aim of this work is the estimation of extreme quantile of a real random variable conditionally to a covariate in dimension d . In the framework of proportional tails, inspired of the heteroscedastics extremes, we develop a coupling approach based on Wasserstein distance to provide asymptotic results about quantile and tail estimators. Those coupling arguments rely on the control of the Wasserstein distance between empirical measures and sharp bounds between the law of exceedances and the Pareto distribution. This approach also provide a validation procedure for proportional tails model.

Time and Session: Thursday 11.30-13.00, CS Prediction and validation for extremes

Modeling first arrival of migratory birds using a hierarchical max-infinitely divisible process

Benjamin Shaby

Colorado State University

Abstract

Humans have recorded the arrival dates of migratory birds for millennia, searching for trends and patterns. As the first arrival among individuals in a species is the realized tail of the probability distribution of arrivals, the appropriate statistical framework with which to analyze such events is extreme value theory. Here, for the first time, we apply formal extreme value techniques to the dynamics of bird migrations. We study the annual first arrivals Magnolia Warblers using modern tools from the statistical field of extreme value analysis. Using observations from the eBird database, we model the spatial distribution of Magnolia Warbler arrivals as a max-infinitely divisible process, which allows us to spatially interpolate observed annual arrivals in a probabilistically coherent way, and to project arrival dynamics into the future by conditioning on climatic variables.

Time and Session: Tuesday 17.30–18.45, CS Applications of extremes (II)

Novel diagnostic and uncertainty characterisation tools for multivariate return curves

Callum Barltrop

Lancaster University

Abstract

Risk measures that represent extremal probabilities are often used as summary statistics during statistical inference involving extreme value theory. In the univariate setting, arguably the most common measure is a return level; this is the value exceeded by a random variable with some user-specified probability p . However, little consideration has been given to extending this measure to the general multivariate setting. In this talk, we discuss one possible extension which we term a multivariate return curve. Similarly to return levels, these curves are defined, for a given probability p , to be all values of a multivariate random variable for which the joint survival probability is equal to p . For particularly small probabilities, these curves can be used to quantify the risk from extreme multivariate events and are often considered to be the natural multivariate extension to a return level. Moreover, for applications where the risk from combinations of two (or more) variables is considered important, such as the analysis of off-shore structures, these curves may allow resources to be better allocated than if the extremes of the variables were considered separately. We also introduce novel diagnostic and uncertainty characterisation tools for return curve estimation and apply these methods to compare curve estimates from a range of multivariate extreme value models. Furthermore, we demonstrate our approach using a case study and illustrate some of the potential applications of return curves.

Time and Session: Monday 16.00–17.15, CS Multivariate extremes

Distributed inference for tail empirical and quantile processes

Chen Zhou

Erasmus University Rotterdam

Abstract

value statistics using more observations drawn from the tail of an underlying distribution. However, if such data sets are saved in multiple machines and cannot be combined into one oracle sample due to privacy reasons, it poses computational challenges for computing the estimate using an oracle estimator based on the oracle sample. Such a situation is regarded as the distributed inference setup. To overcome this problem, distributed inference often considers a divide-and-conquer (DC) algorithm: first compute the estimate using observations on each machine, then transmit the estimates from all machines to the central machine, and eventually take the average of the estimates on the central machine. If the final distributed estimator possesses the same asymptotic behavior as the hypothetical oracle estimator based on the oracle sample, then it is regarded as satisfying the oracle property.

Estimators in extreme value statistics do not satisfy the oracle property per se. In this paper, we introduce a set of tools regarding the asymptotic behavior of the tail empirical and quantile processes under the distributed inference setup. Using these tools, one may establish the oracle property for most extreme value estimators under proper sufficient conditions. As an example, we show the oracle property for the probability weighted moment estimator.

Time and Session: Monday 11.30–13.00, CS Machine Learning for extremes

Extremal lifetimes of persistent loops and holes

Christian Hirsch* and N. Chenavier

*University of Groningen

Abstract

Persistent homology captures the appearances and disappearances of topological features such as loops and holes when growing disks centered at a Poisson point process. We study extreme values for the lifetimes of features dying in bounded components and with birth resp. death time bounded away from the threshold for continuum percolation. First, we describe the scaling of the minimal lifetimes for general feature dimensions, and of the maximal lifetimes for holes in the Čech complex. Then, we proceed to a more refined analysis and establish Poisson approximation for large lifetimes of holes and for small lifetimes of loops. Finally, we also study the scaling of minimal lifetimes in the Vietoris-Rips setting and point to a surprising difference to the Čech complex.

Time and Session: Thursday 16.00–17.15, IS Long memory processes and non-standard EVT

Hill random forests

Christian Robert

ENSAE

Abstract

In extreme value statistics, the tail index is used to measure the occurrence and the intensity of extreme events. In many applied fields, the tail behavior of such events depends on explanatory variables. This article proposes an ensemble learning method for tail index regression which is called Hill random forests and combines Hill's approach on tail index estimation (Hill (1975)) with the aggregation of randomized decision trees based on the gamma deviance. We prove a consistency result when the tail index function is a multiplicative function.

Time and Session: 11.30–13.00, CS Machine Learning for extremes

Simulating flood event sets using extremal principal components

Christian Rohrbeck* and Daniel Cooley

*University of Bath

Abstract

Catastrophe models are an important tool to estimate the impact of natural hazards. Such models are, for instance, applied by insurance companies to predict the financial capital required to cover potential future payouts. One component of these models is a simulated hazard event set, which represents, for instance, a collection of floods over a long period, e.g., 1,000 or 10,000 years.

In this talk, we introduce a new approach in extreme value analysis to generate hazard event sets in various settings. The methodology is illustrated using a set of 45 river flow gauges in northern England and southern Scotland, an area which regularly experiences severe flood events. We start by utilising the methodology by Cooley and Thibaud (2019) to study extremal dependence across river flow gauges. We then introduce our new framework which uses dimension-reduction techniques, in combination with the concept of extremal principal components introduced by Cooley and Thibaud (2019). Specifically, we approximate the joint extreme value distribution by defining a model which fully captures large-scale spatial structures in the extremes and provides a reasonable fit for the local dynamics. Our approach enables us to simulate large hazard event sets in a few seconds, and we will discuss/highlight the agreement of the characteristics of the observed and simulated extreme events.

Time and Session: Thursday 10.00–11.15, CS Flood risk

Gradient boosting for extreme quantile regression

Clément Dombry

Universite Bourgogne Franche-Comte

Abstract

Extreme quantile regression provides estimates of conditional quantiles outside the range of the data. Classical methods such as quantile random forests perform poorly in such cases since data in the tail region are too scarce. Extreme value theory motivates to approximate the conditional distribution above a high threshold by a generalized Pareto distribution with covariate dependent parameters. This model allows for extrapolation beyond the range of observed values and estimation of conditional extreme quantiles. We propose a gradient boosting procedure to estimate a conditional generalized Pareto distribution by minimizing its deviance. Cross-validation is used for the choice of tuning parameters such as the number of trees and the tree depths. We discuss diagnostic plots such as variable importance and partial dependence plots, which help to interpret the fitted models. In simulation studies we show that our gradient boosting procedure outperforms classical methods from quantile regression and extreme value theory, especially for high-dimensional predictor spaces and complex parameter response surfaces. An application to statistical post-processing of weather forecasts with precipitation data in the Netherlands is proposed.

Time and Session: Thursday 17.30-18.45, IS Forecasting, metrics, evaluations and scoring of extremes

Sharpening our view of extreme heat

Colin Raymond

NASA Jet Propulsion Laboratory

Abstract

Increasing extreme heat is a global challenge with fundamentally local characteristics: its hotspots occur at scales too small for most climate models to capture and are often missed by patchy observational networks. Simultaneous extreme heat and humidity—the most direct correlate of heat stress—further involves multivariate conditional dependence and requires the consideration of additional physical constraints, some of which are under-observed and consequently can result in unanticipated statistical behavior. I illustrate these issues with examples from recent analyses (my own and others') and describe efforts that have been made at addressing them. I also briefly remark on the difficult but valuable objective of aligning such analyses with the design of systems that modulate the ultimate impacts of climate extremes.

Time and Session: Monday 16.00–17.15, IS Climate extremes

Distributionally robust tail bounds based on Wasserstein distance and f -divergence

Corina Birghila, Maximilian Aigner and Sebastian Engelke

University of Waterloo

Abstract

In this work, we provide robust bounds on the index of tail heaviness in the context of model misspecification. They are defined as the optimal value when computing the worst-case tail behaviour over all models within some neighbourhood of the reference model. The choice of the discrepancy between the models used to build this neighbourhood plays a crucial role in assessing the heaviness of the asymptotic bounds. To this end, we evaluate the robust tail behaviour in ambiguity sets based on the Wasserstein distance and Csiszar f -divergence and obtain explicit expressions for the corresponding asymptotic bounds. Numerical simulations are provided to emphasize the difference between these bounds and the importance of the choice of discrepancy measure.

Time and Session: Monday 17.30–18.45, Best student paper (II)

Weather and climate risk in power systems with renewables

David Brayshaw

Reading University

Abstract

The growing use of weather-dependent renewable power is changing the way electricity systems operate. The traditional paradigm where large power plants are managed to meet variations in electricity demand is being replaced by a situation in which both demand and supply are strongly influenced by weather. This has profound consequences for power systems where supply-demand balance must be maintained in real time and the need for high-quality meteorological information—particularly regarding extreme events—has never been greater. Recent developments in weather and climate science offer opportunities for addressing power system risk but also raise new challenges for how we understand, model and quantify meteorological impacts in systems as complex as electricity networks.

This talk will review some of the challenges associated in power system planning and operations from a meteorological perspective, particularly focussing on the tension between computational tractability and the robust characterisation of weather and climate variability. The role of extreme weather events and climate variations in determining optimal power system design and operation will be discussed and, specifically, the challenges associated with ensuring adequate climate-sampling highlighted. Recent results from a novel implementation of *Importance Subsampling* are presented as a potential solution.

Time and Session: Thursday 10.00–11.15, IS Extremes of energy systems

Extremal quantile treatment effects for heavy-tailed distributions

David Deuber

ETH Zurich

Abstract

Causal inference for rare events has important applications in many fields such medicine, climate science and finance. We introduce an extremal quantile treatment effect as the difference of extreme quantiles of the potential outcome distributions. Estimation of this effect is based on extrapolation results from extreme value theory in combination with a new counterfactual Hill estimator that uses propensity scores as adjustment. We establish the asymptotic theory of this estimator and propose a variance estimation procedure that allows for valid statistical inference. Our method is applied to analyze the effect of college education on high wages.

Time and Session: Monday 17.30–18.45, IS Graphical modelling

Modelling the athletics long jump performance: an approach with the r largest order statistics

Domingos Silva* and Frederico Caeiro

*Centro de Investigação em Matemática e Aplicações, Universidade de Évora

Abstract

The r largest order statistics method is an important extension of the block maxima approach since it allows us to use more information from the data. However, the choice of r is not easy. If r is too large, bias can occur, and if r is too small, the variance of the estimator can be very high. So, we must deal with a bias-variance trade-off. In the present study, we use the largest order statistics of athletics long jump, Men and Women, between 1980 and 2020. In addition to the tail inference, we will analyse the wind effect on the athlete's performance.

Time and Session: Monday 10.00–11.15, CS Applications of extremes (I)

Extremal random forests

Edossa Merga Terefe

University of Geneva

Abstract

Methods from statistics and machine learning for quantile regression fail in cases where the quantile of interest is so extreme that only a few or no training data points exceed it. To overcome this problem, asymptotic results from extreme value theory can be used to extrapolate beyond the range of the data, and several approaches exist that use linear regression, kernel methods, or generalized additive models. All of these methods break down if the predictor space has more than a few dimensions or if the regression function of extreme quantiles is complex. We propose a method for extreme quantile regression that combines the flexibility of random forests with the theory of extrapolation. Our extremal random forest (ERF) estimates the parameters of a generalized Pareto distribution, conditional on the predictor vector, by maximizing a weighted likelihood, where the localizing weights are extracted from a quantile random forest. We further penalize the shape parameter in this likelihood to regularize its variability in the predictor space. A range of simulation setups shows that our ERF outperforms both the classical machine learning methods for quantile regression and the existing regression approaches from extreme value theory.

Time and Session: Tuesday 16.00–17.15, Best student paper (III)

 Extreme value analysis for mixture models with heavy-tailed impurity

Ekaterina Morozova* and Vladimir Panov

*HSE University

Abstract

While there exists a well-established theory for the asymptotic behaviour of maxima of the i.i.d. sequences, very few results are available for the triangular arrays, when the distribution can change over time. Typically, the papers on this issue deal with convergence to the Gumbel law, see [Anderson et al. \(1997\)](#), [Dkengne et al. \(2016\)](#), or twice-differentiable distribution, see [Freitas and Hüsler \(2003\)](#).

The current research is available in [Panov and Morozova \(2021\)](#). It contributes to the aforementioned problem by providing the extreme value analysis for the mixture models with varying parameters, which can be viewed as triangular arrays. In particular, we consider the case of the heavy-tailed impurity, which appears when one of the components has a heavy-tailed distribution, and the corresponding mixing parameter tends to zero as the number of observations grows. We analyse two ways of modelling this impurity, namely, by the non-truncated regularly varying law and its upper-truncated version with an increasing truncation level. The set of possible limit distributions for maxima turns out to be much more diverse than in the classical setting, especially for a mixture with the truncated component, where it includes four discontinuous laws. In the latter case, the resulting limit depends on the asymptotic behaviour of the truncation point, which is shown to be related to the truncation regimes introduced in [Chakrabarty and Samorodnitsky \(2012\)](#).

For practical purposes we describe the procedure of the application of the considered model to the analysis of financial returns. In this context, our research develops the idea that their distribution is somewhere *in between* the exponential and power laws, see [Malevergne et al. \(2005\)](#).

References

- Panov, V. and Morozova E., 2021. Extreme value analysis for mixture models with heavy-tailed impurity preprint *arXiv:2103.07689*
- Anderson C.W., Coles S.G. and Hüsler J., 1997. Maxima of Poisson-like variables and related triangular arrays *Ann. Appl. Probab.* **7**, 953–971
- Chakrabarty A. and Samorodnitsky G., 2012. Understanding heavy tails in a bounded world or, is a truncated heavy tail heavy or not? *Stochastic Models* **28**, 109–143
- Dkengne, P.S., Eckert, N., and Naveau, P., 2016. A limiting distribution for maxima of discrete stationary triangular arrays with an application to risk due to avalanches *Extremes* **19**, 25–40
- Freitas A. V. and Hüsler J., 2003. Condition for the convergence of maxima of random triangular arrays *Extremes* **6**, 381–394
- Malevergne Y., Pisarenko V., and Sornette D., 2005. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance* **5**, 379–401

Extreme sea level estimation: accounting for seasonality

Eleanor D'Arcy

STOR-i Centre for Doctoral Training, Lancaster University

Abstract

Storm surges pose an increasing risk to coastline communities. These events, combined with high tide, can result in coastal flooding. To reduce the impact of storm surges, an accurate estimate of coastal flood risk is necessary. Specifically, estimates are required for the return level of sea levels (still water). This estimate is used as an input to determine the height for a coastal defence, such as a sea wall. The return level estimation requires statistical analysis based on extreme value theory, as we need to know about the frequency of events that are more extreme than those previously observed. Large storm surges exhibit seasonality, they are typically at their worst in the winter and least extreme in the summer. We focus on the skew surge: the difference between the observed and predicted high water within a tidal cycle. As well as an annual cyclic trend for seasonality, we investigate a linear trend in the mean skew surge and any residual trend in each season by sharing information spatially. The seasonal pattern of skew surge differs from that of the tide, whose seasonality is driven astronomically, resulting in tidal peaks at the spring and autumn equinoxes. Hence, the worst levels of the two components of still water level are likely to peak at different times in the year, and so statistical methods that treat them as independent variables are likely to over-estimate return levels. Our work aims to model how the distribution of skew surges changes over a year and we combine our results with the known seasonality of tides to derive estimates of still water level return levels.

Time and Session: Thursday 10.00–11.15, CS Flood risk

Continuous simulation of storm processes

Emilie Chautru, Marine Demangeot, Rémi Carnec and Christian Lantuéjoul

MINES ParisTech - PSL University

Abstract

Storm processes constitute prototype models for spatial extremes. They are classically simulated on a finite number of points within a given domain. We propose a new algorithm that allows to perform such a task everywhere, not just anywhere, in continuous domains like hyperrectangles or balls, in arbitrary dimension. This consists in generating basic ingredients that can subsequently be used to assign a value at any and every point of the simulation field. Therefore, the resolution of a single simulation can be refined indefinitely; this is particularly appropriate to investigate the geometrical properties of storm processes. Particular attention is paid to efficiency: by introducing and exploiting the notion of domain of influence of each storm, the running time is considerably reduced. Besides, most parts of the algorithm are designed to be parallelizable.

Time and Session: Thursday 16.00–17.15, CS Spatial extremes (II)

Robust estimation of matrices and the consequences in matrix completion

Emilien Joly

CIMAT

Abstract

Robust estimation has known a second wave of attention in the last decade. One explanation resides in the need for estimators insensitive to possibly numerous outliers in active research fields as machine learning. Recently, a paper of Minsker (2018) has converted the concentration results of one dimensional robust estimators into concentration inequalities for the mean of random matrices. We will explain the techniques used for defining those robust estimators and give two consequences for robust matrix completion in the presence of highly corrupted data.

Time and Session: Thursday 16.00–17.15, CS Spatial Extremes (II)

A geometric investigation into the tail dependence of vine copulas

Emma Simpson*, Jennifer L. Wadsworth and Jonathan Tawn

*University College London

Abstract

Vine copulas are a type of multivariate dependence model, composed of a collection of bivariate copulas that are combined according to a specific underlying graphical structure. Their flexibility and practicality in moderate and high dimensions have contributed to the popularity of vine copulas, but relatively little attention has been paid to their extremal properties. To address this issue, we present results on the tail dependence properties of some of the most widely studied vine copula classes. We focus our study on the coefficient of tail dependence and the asymptotic shape of the sample cloud, which we calculate using the geometric approach of Nolde (2014). We offer new insights by presenting results for trivariate vine copulas constructed from asymptotically dependent and asymptotically independent bivariate copulas, focusing on bivariate extreme value and inverted extreme value copulas, with additional detail provided for logistic and inverted logistic examples.

Time and Session: Thursday 17.30–18.45, CS Dependence modelling

Transform MCMC schemes for sampling intractable factor copula models

Emmanuel Gobet*, Cyril Benezet and Rodrigo Targino

*Ecole Polytechnique

Abstract

In financial risk management, modelling dependency within a random vector is crucial and a standard approach is the use of a copula model. A flexible family of copulas, known as the factor copulas, is formed by the copulas extracted from factor models. Sampling from a factor copula is equivalent to sampling from the factor model and applying the cumulative distribution function (c.d.f.) to each component of the sample. Nonetheless, in many models of interest the c.d.f.'s are not explicitly known. In this talk I'll present theoretical and numerical properties of a transform Markov chain Monte Carlo (MCMC) scheme developed to efficiently compute expectations conditional to rare events in which the unconditional distribution is given by an intractable factor copula.

Time and Session: Friday 16.00–17.15, IS Rare event simulation

Increasing probability of record-shattering climate extremes

Erich Fischer

ETH Zurich

Abstract

Recent climate extremes have broken long-standing records by large margins. Such extremes unprecedented in the observational period often have substantial impacts due to a tendency to adapt to the highest intensities, and no higher, experienced during a lifetime. Here we show climate models project not only more intense extremes but also events that break previous records by much larger margins. These record-shattering extremes, nearly impossible in the absence of warming, are likely to occur in the coming decades. We provide an analytical solution to show that the changes in probability of record-shattering extremes simulated by numerical climate models is well understood and can be accurately reproduced in a purely statistical approach. The increase in record-shattering extremes can be reproduced with an analytical solution assuming normally distributed random uncorrelated variables or assuming a non-stationary GEV distribution with time-varying location and scale parameter for the random variables. Based on this analytical solution we demonstrate that the probability of occurrence of record-shattering extremes depends on warming rate, rather than global warming level, and is thus emission pathway-dependent. In high-emission scenarios, week-long heat extremes that break records by three or more standard deviations are 2–7 times more probable in 2021–2050 and 3–21 times more probable in 2051–2080, compared to the last three decades. In 2051–2080, such events are estimated to occur about every 6–37 years somewhere in the northern mid-latitudes.

Time and Session: Monday 16.00–17.15, IS Climate extremes

Multivariate regular variation in marked Hawkes processes

Fabien Baeriswyl*, Valérie Chavez-Demoulin and Olivier Wintenberger

*Université de Lausanne and Sorbonne Université

Abstract

In this talk, we discuss the tail asymptotics of the clusters of marked Hawkes processes, that are self-exciting point processes describing situations in which points tend to cluster together. These models are notably used in insurance (e.g. to model claim arrivals in view of computing ruin probabilities), in epidemiology to model the spread of a disease or, more recently, in modelling tweets popularity (by considering the cascade of retweets) on Twitter. In particular, we emphasise how (multivariate) regular variation transfers from the immigrant and offspring processes as well as from the mark distribution to the distribution of the cluster itself, generalising a result due to Basrak, Wintenberger and Žugec (2019). We illustrate how the individual components of the process (i.e. the variables governing both the average number of children of an event and its mark) impact the model.

Time and Session: Thursday 11.30–13.00, CS Prediction and validation for extremes

Evolution of groups at high risk of death from COVID-19 using hospital data

Felix Cheyssou

Sorbonne University

Abstract

In France, death rates due to COVID-19 halved between the beginning and the end of the first wave of the pandemic. While this diminution can be explained by better knowledge of the disease, better care for the patients, and a slight increase in the proportion of young patients during the second half of the wave, it is not clear whether it is homogeneous for all age groups. In this talk, we focus on the estimation of death rates for groups at risk from COVID-19. Rather than arbitrarily building the groups at risk from patients' demographic data, they are estimated using classification trees (CART). To study the temporal evolution of groups at risk from COVID-19, we introduce statistical tools for the comparison of CART trees derived from the theory of robust estimation. This allows us to propose a statistical test to determine changes through time in the death rates for each demographic group. Finally, we illustrate this method on the first wave of the pandemic.

Time and Session: Tuesday 10.00–11.15, IS Public health, epidemiology, life sciences and life lengths

Exchangeable min-id sequences: characterization, exponent measures and non-decreasing id-processes

Florian Brück

Technical University Munich

Abstract

We establish a correspondence between exchangeable sequences of random variables whose finite-dimensional distributions are min- (or max-) infinitely divisible and non-negative, non-decreasing, infinitely divisible stochastic processes. The exponent measure of a min-id sequence is shown to be the sum of a very simple *drift measure* and a mixture of product probability measures, which corresponds uniquely to the Lévy measure of a non-decreasing infinitely divisible process. The latter is shown to be supported on non-negative and non-decreasing functions. Our results provide an analytic umbrella which embeds the de Finetti subfamilies of many classes of multivariate distributions, such as exogenous shock models, exponential and geometric laws with lack-of-memory property, min-stable multivariate exponential and extreme-value distributions, as well as reciprocal Archimedean copulas with completely monotone generator and Archimedean copulas with log-completely monotone generator.

Time and Session: Monday 11.30-13.00, Best student paper (I)

Total positivity in graphical extremes

Frank Röttger*, Sebastian Engelke and Piotr Zwiernik

Université de Genève

Abstract

Engelke and Hitz (2020) introduce a general theory for conditional independence and graphical models for extremes. Similarly to Gaussian graphical models, they show that for Hüsler–Reiss graphical models, conditional independencies are encoded through a transformation of the corresponding parameter matrix Γ .

Multivariate total positivity of order 2 (MTP_2) is a strong form of positive dependence that induces many interesting properties in graphical modeling. A multivariate Gaussian distribution is MTP_2 when its precision matrix is an M-matrix, i.e., when all the non-diagonal entries in the inverse covariance matrix are non-positive. We show that a Hüsler–Reiss distribution is MTP_2 if and only if the inverse Farris transform of the parameter matrix Γ is the inverse of a diagonally dominant M-matrix. Furthermore, we prove that an extremal tree model is MTP_2 if and only if its bivariate marginals are MTP_2 . This implies that all Hüsler–Reiss tree models are MTP_2 , while in comparison Gaussian tree models are only MTP_2 if their covariance matrices are non-negative.

Similar to Lauritzen et al. (2019) we construct a coordinate descent algorithm to find a pseudo maximum likelihood estimator under the MTP_2 constraint and show that it enforces a sparse extremal graphical model. We apply this method to simulated and real data.

References

- Engelke S., and Hitz A. H., 2020. Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 871–932
- Lauritzen S., Uhler C., and Zwiernik P., 2019. Maximum likelihood estimation in Gaussian models under total positivity. *Ann. Statist.* **47**, 1835–1863.

Time and Session: Friday 10.00–11.15, CS Graphical models

A comparison of generalized and extended Hill estimators

Frederico Caeiro*, Ivanilda Cabral Maria Ivette Gomes

*CMA, Universidade NOVA de Lisboa, Portugal

Abstract

For heavy tails, classical extreme value index estimators, such as the Hill estimator (Hill, 1975), usually have a strong bias. Consequently, new interesting alternative estimators have appeared in the literature. In this work we compare several generalized and extended Hill estimators of the extreme value index. The comparison is performed not only with the Hill, but also with an asymptotically unbiased Hill estimator. The comparison study is performed asymptotically, under a third-order framework, and for finite samples, through a Monte Carlo simulation study.

Time and Session: Friday 11.30–13.00, CS Univariate tail estimation

Causal inference with spatio-temporal data

Georgia Papadogeorgou

University of Florida

Abstract

Classic causal inference framework is not directly applicable when the treatment and outcome variables are generated by spatio-temporal processes with an infinite number of possible event locations at each point in time. We take up the challenge of extending the potential outcomes framework to these settings by formulating the treatment point process as stochastic intervention. Our causal estimands include the expected number of outcome events in a specified area of interest under a particular stochastic treatment assignment strategy. We develop an estimation technique that applies the inverse probability of treatment weighting method to spatially-smoothed outcome surfaces. We demonstrate that the proposed estimator is consistent and asymptotically normal as the number of time period approaches infinity using the true or an estimated propensity score surface. A primary advantage of our methodology is its ability to avoid structural assumptions about spatial spillover and temporal carryover effects.

Time and Session: Tuesday 16.00–17.15, IS Causal inference

Extremile regression

Gilles Stupfler

ENSAI CREST

Abstract

Regression extremiles define a least squares analogue of regression quantiles. They are determined by weighted expectations rather than tail probabilities. Of special interest is their intuitive meaning in terms of expected minima and maxima. Their use appears naturally in risk management where, in contrast to quantiles, they fulfill the coherency axiom and take the severity of tail losses into account. In addition, they are comonotonically additive and belong to both the families of spectral risk measures and concave distortion risk measures. We provide the first detailed study exploring the estimation of regression extremiles in the presence of covariates. In particular we extend extremile regression far into the tails of heavy-tailed distributions. Extrapolated estimators are constructed and their asymptotic theory is developed. Real data applications will be discussed to conclude the talk.

Time and Session: Tuesday 11.30–13.00, CS Regression techniques (I)

Modeling clusters of extreme events over short periods

Gloria Buriticá

LPSM

Abstract

We study the extreme dependence structure of stationary regularly varying time series. The classical approach for modeling extremes in this setting is based on the identification of short periods with several exceedances also known as clusters of extremes. The main summary of the clustering phenomenon is given by the extremal index. We suggest studying a general version of clustering where we consider short periods for which a modulus function applied to the vector of consecutive days exceeds a high threshold. Then for a convenient choice of modulus, the estimation of the extremal index can be avoided. We deduce an inference method for computing extreme returns levels that avoids the computation of the extremal index. We show applications for precipitation data in France.

Time and Session: Monday 11.30–13.00, Best student paper (I)

A new link function for frequentist beta regression

Gloria Gheno

Ronin Institute

Abstract

The beta regression is used to analyze variables which affect variables whose value is included in the unit interval. For the regressions with binary data, the literature has debated the problem of incorrect link functions and therefore new links have been proposed, such as gev (generalized extreme value). For the mean of the beta regression, instead, the traditional link functions used for binary regressions, i.e. logit, probit and complementary log-log, are used. In my previous work I proposed a new non-monotone function for the beta regression in the Bayesian context. In this paper I propose, instead, a different link function for the mean parameter of a beta regression but in frequentist domain, a modification which also requires the creation of an algorithm created by me ad hoc for parameter estimation. Both link functions have as their particular cases the logit function, representing a traditional symmetric link function, and the gev function, proposed for binary data due to its asymmetry. These two link functions, proposed by me, have the advantage that they can also be non-monotone, unlike those until now present in the literature. In this work the parameters are estimated maximizing the likelihood function, using a version, which I modified, of the genetic algorithm so as to give greater relevance to the traditional link functions than the others. The algorithm divides the dataset into training sets and test sets to evaluate the model and choose the best one. I compare my method with those already present in the literature, in which the researcher decides a priori the link function, using simulated data, in order to establish which of the 2 methods is closest to the true values. My method is better because it is able to correctly determine the link function with which the data are simulated and to estimate the parameters with less error. Its use increases the understanding of the relationships among variables.

Time and Session: 17.30–18.45, CS Regression techniques (II)

Scaling limits of branching random walks and branching stable processes

Hairuo Yang

Institute of Mathematics, University of Zurich

Abstract

Branching-stable processes have recently appeared as counterparts of stable subordinators, when addition of real variables is replaced by branching mechanism for point processes. Here, we are interested in their domains of attraction and describe explicit conditions for a branching random walk to converge after a proper magnification to a branching-stable process. This contrasts with deep results that have been obtained during the last decade on the asymptotic behavior of branching random walks and which involve either shifting without rescaling, or demagnification.

Time and Session: Tuesday 11.30–13.00, IS Extremes random structures (branching and dynamics, geometry)

Using spatial extreme statistics to provide climate uplifts for flood risk management.

Hayley Fowler*, Steven Chan, Elizabeth Kendon, Benjamin Youngman,
Giorgia Fosser, Christopher Short, Simon Tucker and Murray Dale

*Newcastle University

Abstract

Managing the changing risk of flash floods in urban drainage infrastructure commonly uses a simple percentage climate uplift to the derived design guidelines. However, much of this guidance globally is based on projections from coarse-resolution global or regional climate models with parameterised convection, which underestimate the intensity of short-duration rainfall extremes and their future intensity increases. Here we use a 12-member 2.2km convection-permitting climate model (CPM) ensemble, UKCP Local, at a resolution consistent with weather forecasting. CPMs improve the representation of precipitation extremes; thus, UKCP Local is particularly useful for water stakeholders (water utilities and flood risk management professionals) in managing changing hazards and risks for adaptation planning in wastewater and flood risk management. Diagnosing return level ‘uplifts’ for such high-resolution model simulations is difficult due to both their spatial-temporal variability and correlation across space and time. We develop a spatial extreme statistical model that incorporates spatial-temporal variability and correlation of precipitation extremes to provide robust estimates of uplifts for high return levels across all of the UK for months and seasons of interest. The limitations of the climate uplift approach, and potential improvements, will also be discussed.

Time and Session: Monday 10.00–11.15, IS Natural hazards and impacts

Frequency increase in extreme rainfall events in the Northeastern USA with stable intensity distribution

Helga Kristin Olafsdottir

Gothenburg University and Chalmers University of Technology

Abstract

For infrastructure planning, information of both intensities and frequencies of individual extreme rainfall events are of importance. Climate change brings an additional complexity, since extreme daily rainfalls can be increased by the individual extreme rainfalls becoming more frequent, more intense, or both. We develop a new statistical extreme value model, the PGEV model, to predict to what extent these three scenarios will occur. The model combines the generalized extreme value (GEV) distribution for annual maxima series and generalized Pareto (GP) distribution for exceedances over threshold for the partial duration series. This allows the use of high quality annual maxima series data instead of less well checked daily data to estimate trends in intensity and frequency separately.

The method is applied to annual maxima data from the NOAA Atlas 14, Volume 10. For the majority of stations in the Northeastern USA, the frequency of extreme rainfall events increases with increased mean temperature, without much evidence of trends in the distribution of the intensities of individual extreme daily rainfall events. Naturally, increasing trends in frequency also increase the yearly or 10-yearly risks of very extreme rainfall events.

The median of the frequency trends corresponds to extreme rainfalls becoming 83% more frequent for each centigrade degree of temperature increase, but there is an important local variation in the trends. For many stations, the frequency increase exceeds 150% per centigrade temperature increase. The analysis is extended to three other large areas in the contiguous USA, showing a similar but weaker effect.

Time and Session: Monday 17.30–18.45, CS Climate extremes (I)

Bootstrap for block-based extreme value statistics

Holger Drees

University of Hamburg

Abstract

Let $(X_t)_{1 \leq t \leq n}$ be a stationary \mathbb{R}^d -valued time series. In recent years, the asymptotic behavior of extreme value statistics of disjoint blocks $(X_{(i-1)l_n+j}, 1 \leq i \leq [n/l_n])$, or sliding blocks $(X_{i-1+j}, 1 \leq j \leq l_n, 1 \leq i \leq n-l_n+1)$, has been thoroughly studied. Usually, such limit theorems do not directly lend themselves to the construction of confidence intervals, because the influence of the underlying serial dependence structure on the limit distribution is too intricate.

Taking up an approach by [Drees and Neblung \(2021\)](#), we present an abstract bootstrap result which applies both to statistics based on disjoint blocks and on sliding blocks. However, the multiplier bootstrap of sliding blocks requires to select an additional larger block length as a tuning parameter, which complicates the construction of confidence sets in practice.

References

Drees H. and Neblung S., 2021. Asymptotics for sliding blocks estimators of rare events *Bernoulli* **27**, 1239–1269.

Time and Session: Thursday 11.30–13.00, IS Time series

Real-time prediction of severe influenza epidemics using multivariate generalized Pareto modelling

Holger Rootzén* and Maud Thomas

*Chalmers University of Technology

Abstract

Real-time prediction of severe influenza epidemics using multivariate generalized Pareto modelling

Each year, seasonal influenza epidemics put high loads on health care systems and cause hundreds of thousands of deaths worldwide. A main concern for resource planning is the risk of exceptionally severe epidemics. This talk describes how multivariate generalized Pareto models can be used for real-time prediction of the risk that an ongoing epidemic will be exceptionally severe, and for real-time detection of anomalous and potentially dangerous epidemics. The methods are applied to data from the French sentinelles influenza surveillance system. The talk also describes a strategy based on standardized Brier scores, precision-recall curves and average precision scores for assessing the quality of the predictions. Extreme value statistics has a large and rather unexploited potential for aiding healthcare and drug development, and the talk will also mention one more such opportunity: evaluation of results from clinical trials.

Time and Session: Tuesday 10.00–11.15, IS Public health, epidemiology, life sciences and life lengths

Palm theory for extremes of stationary regularly varying time series and random fields

Hrvoje Planinić

University of Zagreb

Abstract

The tail process $(Y_i)_{i \in \mathbb{Z}^d}$ of a stationary regularly varying time series or random field $\mathbf{X} = (X_i)_{i \in \mathbb{Z}^d}$ represents the asymptotic local distribution of \mathbf{X} as seen from its typical exceedance over a threshold u as $u \rightarrow \infty$. Motivated by the standard Palm theory, we show that every tail process satisfies an invariance property called exceedance-stationarity and that this property, together with the polar decomposition of the tail process, characterizes the class of all tail processes. We then restrict to the case when $Y_i \rightarrow 0$ as $|i| \rightarrow \infty$ and establish a couple of Palm-like dualities between the tail process and the so-called anchored tail process which, under suitable conditions, represents the asymptotic distribution of a typical cluster of extremes of \mathbf{X} . The main message is that the distribution of the tail process is biased towards clusters with more exceedances. Finally, we use these results to determine the distribution of the typical cluster of extremes for moving average processes with random coefficients.

Time and Session: Monday 17.30–18.45, Best student paper (II)

Precise large deviations for m dependent subexponential sequences

Igor Rodionov* and Thomas Mikosch

*Trapeznikov Institute of Control Sciences of RAS

Abstract

In this talk, we study precise large deviations for the partial sums of a stationary m -dependent sequence with a subexponential marginal distribution. Our main focus is on distributions which either have a regularly varying or a lognormal-type tail. We apply the results to prove limit theory for the maxima of the entries of the large sample covariance matrices.

Time and Session: Friday 16.00–17.15, CS Time series

Modeling seasonal variations of extreme rainfall on different time scales in Germany

Jana Ulrich*, Felix S. Fauer, Henning W. Rust

*Freie Universität Berlin

Abstract

To estimate the properties of extreme precipitation events on different time scales, intensity-duration-frequency (IDF) curves are a well-known tool in hydrology. The idea of implementing the dependence of precipitation intensity on duration directly into the parameters of an extreme value distribution was proposed by [Koutsoyiannis et al. \(1998\)](#). Following this, we use a duration-dependent generalized extreme value distribution (d-GEV) in the framework of the block maxima approach. When modeling precipitation extremes, this approach is typically used with an annual block size. Although large amounts of data are lost this way, choosing a smaller block size requires a more complex model that accounts for intra-annual variations. Yet, at least for planning or adaptation of hydrological structures, these variations are mostly not even relevant.

Nevertheless, it has been shown that within the block maxima approach monthly block sizes can be sufficient to model extreme precipitation in the mid-latitudes [e.g., [Rust et al. 2009](#)]. Likewise, several studies have demonstrated that it is possible to model the intra-annual variations using smooth periodic functions as covariates for the parameters of the GEV [e.g., [Coles \(2001\)](#), [Rust et al. \(2009\)](#), [Fischer et al. \(2019\)](#)]. In addition, information on the occurrence probabilities in different months may well be needed by stakeholders in, for example, agriculture or water storage.

However, monthly block sizes have not yet been considered for modeling the IDF relationship. Still, particularly for this application, the use of monthly maxima may be relevant, because extreme precipitation events related to different time scales typically occur in different seasons due to different mechanisms causing them. In Germany, convective extreme events with a duration of minutes to a few hours occur almost exclusively in the summer months, while long-lasting extreme events are scattered throughout the year or even occur more frequently in autumn and winter months, depending on the location.

In this study, we model monthly precipitation maxima at different stations in Germany for a wide range of durations from one minute to 5 days using a d-GEV with seasonally varying parameters. On the one hand, this allows us to compare the characteristics of the IDF curves of different months: Since in summer months the curves are steeper and have higher intensities for short durations than in the rest of the year, at some stations the curves (for a given quantile) for different months cross. The meteorological interpretation of this crossing is that the season at which a certain extreme event is most likely to occur shifts for longer durations (usually towards autumn or winter). On the other hand, we can also investigate the influence of intra-annual variations on the distribution of annual maxima of different durations: We find that the annual IDF curves that result from modeling monthly maxima diverge from the assumption of scale invariance, causing a flattening in the slope of the IDF curves for long durations.

References

- Coles S., 2001. *An Introduction to Statistical Modeling of Extreme Values* Springer, New York
- Fischer M., Rust H. W., and Ulbrich U., 2019. A spatial and seasonal climatology of extreme precipitation return-levels: a case study. *Spat. Stat.* **34**, 100275
- Koutsoyiannis D., Kozonis D. and Manetas A., 1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships *J. Hydrol.* **206**, 118–135
- Rust H. W., Maraun D., and Osborn T. J., 2009. Modelling seasonality in extreme precipitation *Eur. Phys. J. Spec. Top.* **174**, 99–111

Time and Session: Monday 17.30–18.45, CS Climate extremes (I)

Extremal dependence properties and representations for spatial extremes

Jennifer L. Wadsworth* and Natalia Nolde

*Lancaster University

Abstract

We examine the differences between spatial extreme value representations when focusing on the joint tail versus taking the approach of conditioning on a single location being extreme. As a case study, we explore the model proposed in Huser and Wadsworth (2019), and elucidate why parameter changes that result in different strengths of joint tail dependence do not alter the limiting conditional extremes representation. This is illustrated geometrically via the shape of limit sets for the finite-dimensional margins.

Time and Session: Monday 11.30–13.00, IS Spatial extremes

Transformed-linear combination of regularly varying random variables and linear prediction for extremes.

Jeongjin Lee

Colorado State University

Abstract

We derive the best transformed-linear predictor for extremes on the positive orthant within the vector space which is a set of transformed-linear combinations of regularly varying random variables with finite coefficients. The optimized weights for the transformed-linear predictor are represented by the tail pairwise dependence matrix (TPDM) as analogous to those for the BLUP are expressed by a covariance matrix in Gaussian cases. We construct uncertainty quantification using the polar geometry of regular variation utilizing a completely positive decomposition of the TPDM. We produce a 95% joint probability region and 95% conditional prediction intervals given large values of the transformed-linear predictor. We apply our method to the NO₂ pollution data in Washington, DC and to daily returns of 30 industry portfolios and obtain good coverage rates.

Time and Session: Thursday 17.30–18.45, CS Regression techniques (II)

Probabilistic forecasting of regional net-load with conditional extremes

Jethro Browell

University of Strathclyde

Abstract

The increasing penetration of embedded renewables makes forecasting net-load, consumption less embedded generation, a significant and growing challenge. This talk will present a framework for producing probabilistic forecasts of net-load with particular attention given to the tails of predictive distributions, which are required for managing risk associated with low-probability events. Only small volumes of data are available in the tails, by definition, so estimation of predictive models and forecast evaluation requires special attention. We propose a solution based on a best-in-class load forecasting methodology adapted for net-load, and model the tails of predictive distributions with the generalized Pareto distribution, allowing its parameters to vary smoothly as functions of covariates. The resulting forecasts are shown to be calibrated and sharper than those produced with unconditional tail distributions. In a use-case inspired evaluation exercise based on reserve setting, the conditional tails are shown to reduce the overall volume of reserve required to manage a given risk. Furthermore, they identify periods of high risk not captured by other methods. The proposed method therefore enables user to both reduce costs and avoid excess risk.

Time and Session: Thursday 10.00–11.15, IS Extremes of energy systems

Graphical models for infinite measures with applications to extremes and Lévy processes

Jevgenijs Ivanovs*, Sebastian Engelke and Kirstin Strokorb

*Aarhus University

Abstract

Measures exploding at the origin are fundamental in the study of extremes and stochastic processes. For example, they appear in construction of max-infinitely divisible distributions and Lévy processes. We define a notion of conditional independence for such measures and establish a number of its equivalent characterizations. In particular, it includes the recent notion of extremal conditional independence for multivariate Pareto distributions without a density assumption. Furthermore, structural max-linear models can be put into this framework as well.

Time and Session: Monday 17.30–18.45, IS Graphical modelling

Extreme U-statistics

Jochem Oorschot*, Johan Segers and Chen Zhou

*Erasmus University Rotterdam

Abstract

One-sample U-statistics generalize the notion of averages and lead to minimum-variance unbiased estimators. [Segers \(2001\)](#) considered U- for sequences of location-scale invariant kernels whose degree grows with the sample size and which depend only on the largest of their arguments. By ensuring that the kernel's expectation converges to a function of the extreme value index, it is possible to construct consistent estimators of the latter using such extreme U-statistics. The all-block maxima in [Oorschot and Zhou \(2020\)](#) follow a similar logic, by recording the maxima of all sub-samples of observations of a given size. In this paper we explore the asymptotic normality of general extreme-U-statistics. The theory draws largely from the one of U-statistics by considering kernels applied to blocks of intermediate length $m(n)$ that grow as a function of the sample size n .

References

- Segers, J., 2001. Extremes of a random sample: limit theorems and statistical applications. PhD Thesis. , KU Leuven.
- Oorschot, J. and Zhou, C., 2020. All block maxima method for estimating the extreme value index. preprint *arXiv:2010.15950*.

Time and Session: Monday 16.00-17.15, CS Inference and robust extremes

Properties and consistency of QTree in max-linear models under observational noise

Johannes Buck*, Ngoc Mai Tran and Claudia Klüppelberg

*Technical University of Munich

Abstract

Recently, we proposed the QTree algorithm for causal inference in river networks. Motivated by max-linear models, the algorithm utilizes the quantile gap between differences of random variables to unravel causal relationships and achieves almost perfect recovery of the upper Danube basin, clearly outperforming existing methods. In this talk, we present the qualitative features of the estimator. Assuming that the data comes from a max-linear Bayesian network with additive noise, we prove that QTree is consistent given some mild conditions on the tail of the noise distribution.

Time and Session: Thursday 10.00–11.15, CS Causal inference

Extremes of interpoint distances of high-dimensional random vectors

Johannes Heiny

Ruhr University Bochum

Abstract

We study point processes of dependent random variables and derive optimal conditions for their convergence. Our results are applied to finding the asymptotic distributions of the extremes of interpoint distances and U-statistics.

Time and Session: Friday 11.30–13.00, CS Extremes of stochastic processes (II)

Empirical tail copulas for functional data

John Einmahl* and Johan Segers

*Tilburg University

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). An application for testing tail copula stationarity is presented. 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.

Time and Session: Tuesday 17.30–18.45 , IS Multivariate extremes (sparsity, high-dimensional, copulas, anomaly detection)

Robust sparse reconstruction

John Nolan

American University

Abstract

Traditional compressive sensing (CS) assumes light-tailed models for the underlying signal and/or noise. This assumption is not met in the case of highly impulsive environments, where non-Gaussian processes arise. In this case, traditional sparse reconstruction methods perform poorly, since they are incapable of suppressing the effects of heavy-tailed sampling noise. This talk will describe the use of heavy tailed stable distributions to design a robust algorithm for sparse signal reconstruction from linear random measurements corrupted by infinite-variance additive noise. We demonstrate the improved reconstruction performance of the proposed algorithm when compared against standard CS techniques for a broad range of impulsive environments.

Time and Session: Tuesday 17.30–18.45 , IS Multivariate extremes (sparsity, high-dimensional, copulas, anomaly detection)

Using scoring functions to evaluate point process forecasts

Jonas Brehmer

Heidelberg Institute for Theoretical Studies

Abstract

Many decision makers in industry or public institutions rely on forecasts of certain random quantities when choosing among alternative actions and assessing the associated risks. If relevant events cannot be measured at fixed points, but occur randomly in space and/or time, point processes arise as natural models in many applications, e.g. epidemiology, seismology, or quantitative criminology. Such models are often used to create point process-based forecasts, for instance (conditional) intensities, or higher order summary statistics such as pair correlations. Usually, decision makers will face a number of different predictions concerning these quantities, making a comparison of their accuracy crucial for well-founded decisions. A principled approach to comparative forecast evaluation relies on scoring or loss functions, which assign a real number to each pair of forecast and realized observation of a random variable. We establish some general results which transfer the idea of scoring functions to the point process setting. This leads to a novel comparative evaluation method for point process-based forecasts and provides a new perspective on several existing ones. Since our approach focuses on relative performance, it can be used as a model selection technique which complements existing goodness-of-fit tests.

Time and Session: Thursday 17.30–18.45, IS Forecasting, metrics, evaluations and scoring of extremes

Can causal discovery benefit from extreme values?

Jonas Peters

University of Copenhagen

Abstract

In causal discovery, the goal is to learn causal structure from observational data. In this talk, we show two examples of methods that can potentially benefit from extreme values. (1) In linear models, the graph structure becomes identifiable if the noise variables have a non-Gaussian distribution. (2) For heterogeneous data sets, in which the causal mechanism for a target variable remains the same for all observations, it has been suggested to learn causal structures by searching for invariant models. We briefly introduce such methods and provide intuition for why it may help if the distributions are heavy-tailed. No prior knowledge about causality is required.

Time and Session: Tuesday 16.00–17.15, IS Causal inference

A bias-reduced version of the Weissman extreme quantile estimator

Jonathan El Methni* and Stéphane Girard

*Université de Paris

Abstract

Weissman extrapolation device for estimating extreme quantiles is based on two estimators: an order statistic to estimate an intermediate quantile and an estimator of the tail index. The common practice is to select the same intermediate sequence for both estimators. In this work, we show how an adapted choice of two different intermediate sequences leads to a reduction of the asymptotic bias associated with the resulting Weissman estimator. Our approach is compared to other bias reduced estimators of extreme quantiles both on simulated and real data.

Time and Session: Friday 11.30–13.00, CS Univariate tail estimation

Extended generalized Pareto for subasymptotic tail analysis with an application to heatwave intensities

Jonathan Jalbert

Polytechnique Montréal

Abstract

For environmental applications, tail distribution plays a major role as it is often associated with extreme risks. A classical approach to characterize the tails is to model the exceedances above a sufficiently high threshold with the generalized Pareto distribution. For some applications, the choice of the threshold is difficult and the asymptotic conditions may not be satisfied. Extended generalized Pareto models can then be used when the threshold is not high enough. However, the existing extensions have an infinite or zero density at the origin, which makes them unsuitable for modeling the exceedances. A new extension of the extended generalized Pareto distribution is developed to model the exceedances above a subasymptotic threshold having a finite density and not zero at the threshold. The proposed extension provides a better estimate of extreme quantiles than existing models, especially for small samples. Finally, this new distribution is used to model heatwave intensities.

Time and Session: Monday 16.00–17.15, CS Inference and robust extremes

Jonathan Koh

EPFL

Abstract

Accurate spatiotemporal modeling of conditions leading to moderate and large wildfires provides better understanding of mechanisms driving fire-prone ecosystems and improves risk management. We study daily summer wildfire data for the French Mediterranean basin during 1995–2018. We jointly model occurrence intensity and wildfire sizes by combining extreme-value theory and point processes within a Bayesian hierarchical model. The occurrence component models wildfire ignitions as a spatiotemporal log-Gaussian Cox process. Burnt areas are numerical marks attached to points, and we consider fires with marks exceeding a high threshold as extreme. The size component is a two-component mixture varying in space and time that jointly models moderate and extreme fires. We capture non-linear influence of covariates (Fire Weather Index, forest cover) through component-specific smooth functions, which may vary with season. To reveal common drivers of different aspects of wildfire activity, we share random effects between model components to improve interpretability and parsimony without compromising predictive skill. Stratified subsampling of zero counts is implemented to cope with large observation vectors. We compare and validate models through predictive scores and visual diagnostics. Our methodology provides a holistic approach to explain and predict the drivers of wildfire activity and their associated uncertainties.

Time and Session: Monday 17.30–18.45, CS Climate extremes (I)

Modelling the extremes of spatial aggregates of precipitation using conditional methods

Jordan Richards

Lancaster University

Abstract

Fluvial flooding is not caused by high intensity rainfall at a single location, rather it is caused by the extremes of precipitation events aggregated over spatial catchment areas. Accurate modelling of the tail behaviour of such events can help to mitigate the financial aspects associated with floods, especially if river defences are built within specification to withstand an n -year event of this kind. Within an extreme value analysis framework, univariate methods for estimating the size of these n -year events are well studied and cemented in asymptotic theory. To complement these techniques, we develop a high-resolution spatial model for extreme precipitation by providing a fully spatial extension of the conditional approach for modelling multivariate extremes. We simulate realistic precipitation fields from this model and use univariate techniques to make inference about the extremal behaviour of aggregates over specified spatial domains. The challenge of zero precipitation data is overcome and further applications of the model are discussed. The model is fit to data from a convection permitting forecast model within the 2018 UK Climate Projections (UKCP18). To complement these techniques, we develop a high-resolution spatial model for extreme precipitation by providing a fully spatial extension of the conditional approach for modelling multivariate extremes. We simulate realistic precipitation fields from this model and use univariate techniques to make inference about the extremal behaviour of aggregates over specified spatial domains. The challenge of zero precipitation data is overcome and further applications of the model are discussed. The model is fit to data from a convection permitting forecast model within the 2018 UK Climate Projections (UKCP18).

Time and Session: Monday 17.30–18.45, CS Climate extremes (I)

Nonparametric changepoint detection tests based on the breaking of records

Jorge Castillo-Mateo

University of Zaragoza

Abstract

Three new nonparametric statistics are introduced for the changepoint detection problem. They are based on functions of the numbers of upper or/and lower records which occur in a series. The asymptotic Kolmogorov distribution of the test statistics is obtained from a characterization based on Wiener processes and Brownian bridges. A Monte Carlo study of size, power and changepoint estimate is developed. The methods are illustrated by the analysis of climate change through the record occurrence of temperature time series. Record statistics are not a common approach to extreme value analysis, but this work shows some of their potential interests.

Time and Session: Tuesday 10.00–11.15, CS Climate extremes (II)

Evaluation of binary classifiers for extremes

Juliette Legrand*, Philippe Naveau and Marco Oesting

*LSCE

Abstract

Machine learning classification methods usually assume that all possible classes are sufficiently present within the training set. Due to their inherent rarities, extreme events are always under-represented and classifiers tailored for predicting extremes need to be carefully designed to handle this under-representation. In this paper, we address the question of how to assess and compare classifiers with respect to their capacity to capture extreme occurrences. This is also related to the topic of scoring rules used in forecasting literature. In this context, we propose and study a risk function adapted to extremal classifiers. The inferential properties of our empirical risk estimator are derived under the framework of multivariate regular variation. As an example, we study in detail the special class of linear classifiers and show that the optimisation of our risk function leads to a consistent solution. A simulation study compares different classifiers and indicates their performance with respect to our risk function. To conclude, we apply our framework to the analysis of extreme river discharges in the Danube river basin. The application compares different predictive algorithms and test their capacity at forecasting river discharges from other river stations. As a by-product, we identify the explanatory variables that contribute the most to extremal behavior.

Time and Session: Thursday 10.00–11.15, CS Flood risk

Detection of causality in time series using extreme values

Juraj Bodík

Charles University, Prague

Abstract

We deal with the following problem: Let us have two stationary (possibly nonlinear) time series with heavy-tailed marginal distributions. We want to detect whether there is some Granger causality present. Even more, we want to determine the minimal lag, i.e. the time how much it takes for information to travel from one time series to another. We will examine the asymmetry in extremes between the cause and effect, and present a statistic that can estimate such asymmetries. The basis of the idea stands by the so-called causal tail coefficient for time series, which in some way represents the behaviour in extremes of one series conditioned on the presence of an extreme in the other.

Time and Session: Thursday 10.00–11.15, CS Causal inference

Bayesian semiparametric modeling of jointly heteroscedastic extremes

Karla Vianey Ramirez* and Miguel de Carvalho

*University of Edinburgh

Abstract

We introduce a Bayesian semiparametric model for learning about the magnitude and frequency of joint extreme values. The joint scedasis function for joint extremes is here devised as a function that carries information on the frequency of joint extremes over time. We develop Bayesian estimators for the two parameters in the model—the joint scedasis function and the coefficient of tail dependence; to learn about the joint scedasis function we resort to finite mixtures of Polya trees, as they can be used to define a flexible prior in the space of scedasis functions. The simulation study shows that the proposed methods are able to recover the true magnitude and frequency of joint extremes in a variety of simulation scenarios. An application of the proposed methodology to the so-called FAANG (Facebook, Apple, Amazon, Netflix and Alphabet's Google) stocks reveals some interesting insights on the dynamics governing their joint extreme losses over time.

Time and Session: Tuesday 16.00–17.15, CS Bayesian extremes

Two-dimensional ruin for Brownian motions with drift dependent on initial capital

Konrad Krystecki

University of Wrocław

Abstract

Denote by $(W_1(s), W_2(t)), s, t \geq 0$ a bivariate Brownian motion with standard Brownian motion marginals and constant correlation $\rho \in (-1, 1)$. In this contribution we study the asymptotics following ruin probability model

$$p_{\alpha, \beta, \rho}(c, a) = \mathbb{P}\left\{ \sup_{t \in [0, T]} W_1(t) - c_1 u^\alpha t > u^\beta, W_2(t) - c_2 u^\alpha t > a u^\beta \right\}, \quad \text{as } u \rightarrow \infty.$$

Constants $c_1, c_2, \alpha \in \mathbf{R}, \beta \geq 0$ model the behaviour of the drift function, which is dependent on the initial capitals $u, au, a \leq 1$. Moreover, we give the connection and give asymptotics for the two-dimensional many-source ruin model.

$$\mathbb{P}\left\{ \sup_{t \in [0, T]} \sum_{i=1}^N B_i(t) - c_1 t > N, \sum_{i=1}^N B'_i(t) - c_2 t > N \right\},$$

.

$$\mathbb{P}\left\{ \sup_{t \in [0, T]} \sum_{i=1}^N B_i(t) - c_1 t > N, \sum_{i=1}^N B'_i(t) - c_2 t > N \right\},$$

where B_i, B'_i are standard Brownian motions, $\text{cov}(B_i(t), B'_i(t)) = \rho t$, as $N \rightarrow \infty$.

Time and Session: Tuesday 17.30–18.45, CS Extremes of stochastic processes (I)

Bounds on the expected supremum of fractional Brownian motion with drift

Krzysztof Bisewski

University of Lausanne

Abstract

Expected supremum of fractional Brownian motion with Hurst parameter $H \in (0, 1]$ is one of the most fundamental quantities in theory of extremes of Gaussian processes, and yet its value is known only in two cases (when its Hurst parameter equals $1/2$ or 1). In this talk, we review the most up-to-date lower and upper bounds for this quantity for $H \in (0, 1]$. Second, we introduce a new representation of fractional Brownian motion, which enables us to derive a novel lower bound. Numerical experiments suggest that our lower bound is close to the ground truth when $H \in (0, 1]$

Time and Session: Monday 10.00–11.15, IS Extremes of stochastic processes (ambit, Gaussian)

Extremes of vector-valued Gaussian processes

Krzysztof Dębicki*, Enkelejd Hashorva and Longmin Wang

*University of Wrocław

Abstract

We develop the uniform double-sum method for the vector-valued setting obtaining the exact asymptotics of the exceedance probabilities for both stationary and non-stationary vector-valued Gaussian processes.

We apply our findings to two classes of processes, namely the operator fractional Brownian motion and the operator fractional Ornstein-Uhlenbeck process.

Time and Session: Monday 10.00–11.15, IS Extremes of stochastic processes (ambit, Gaussian)

Running supremum of Brownian motion in dimension 2: exact and asymptotic results

Krzysztof Kępczyński

University of Wrocław

Abstract

This paper investigates

$$\pi_T(a_1, a_2) = \mathbb{P} \left(\sup_{t \in [0, T]} (\sigma_1 B(t) - c_1 t) a_1, \sup_{t \in [0, T]} (\sigma_2 B(t) - c_2 t) a_2 \right),$$

where $\{B(t) : t \geq 0\}$ is a standard Brownian motion, with $T > 0, \sigma_1, \sigma_2, c_1, c_2 \in \mathbb{R}$. We derive explicit formula for the probability $\pi_T(a_1, a_2)$ and find its asymptotic behavior both in the so called many-source and high-threshold regimes.

Time and Session: Friday 11.30–13.00, CS Extremes of stochastic processes (II)

Extrema of multi-dimensional Gaussian processes over random intervals

Lanpeng Ji* and Xiaofan Peng

*University of Leeds

Abstract

Consider an n -dimensional Gaussian process where each coordinate is a Gaussian process with stationary increments and with a (positive or negative) linear drift, and all the coordinates are assumed to be independent.

We shall discuss the joint tail asymptotics of the extrema of this multi-dimensional Gaussian process over random intervals, say $[0, T_1], [0, T_2], \dots, [0, T_n]$, respectively, where $T = (T_1, T_2, \dots, T_n)$ is a regularly varying random vector which is independent of the Gaussian processes. Our result shows that the structure of the joint tail asymptotics is determined by the signs of the drifts.

As a relevant application, we shall discuss a multi-dimensional regenerative model, which is a process with a random alternating environment, where an independent multi-dimensional fractional Brownian motion (fBm) with drift is assigned at each environment alternating time. We derive the corresponding ruin probability by analysing a related perturbed random walk.

Time and Session: Tuesday 17.30–18.45, CS Extremes of stochastic processes (I)

Statistical learning of extreme spatio-temporal events with an application to global terror attacks

Lekha Patel

Sandia National Laboratories

Abstract

Extreme events with potential deadly outcomes, such as those organized by terror groups, are highly unpredictable in nature and an imminent threat to society. In particular, quantifying the likelihood of a terror attack occurring in an arbitrary space-time region and its relative societal risk, would facilitate strengthened measures ensuring our national security. This talk focuses on describing our statistical modeling efforts to do so and presenting results from our approach using the open-source Global Terrorism Database (GTD). We first present a semi-parametric self-exciting model of attacks whose inhomogeneous baseline intensity is written a function of typical covariates found in the region of interest. Its triggering intensity is succinctly modeled with a Gaussian Process prior distribution, able to flexibly capture intricate spatio-temporal dependencies between an arbitrary attack and previous terror events. By inferring parameters of this model, we highlight specific space-time areas in which attacks are likely to occur. Furthermore, by measuring the outcome of an attack in terms of the number of casualties it produces, we then incorporate an extreme valued probability distribution of casualties, able to determine high-risk attacks in these regions. Our framework will last be utilized to study historic terror attacks in central and southern Asia.

Time and Session: Tuesday 17.30–18.45, CS Applications of extremes (II)

Spatial scale-aware tail dependence modeling for high-dimensional spatial extremes

Likun Zhang

Lawrence Berkeley National Laboratory

Abstract

Extreme events over large spatial domains like the contiguous United States may exhibit highly heterogeneous tail dependence characteristics, yet most existing spatial extremes models yield only one dependence class over the entire spatial domain. To accurately characterize ‘storm dependence’ in analysis of extreme events, we propose a mixture component model that achieves flexible dependence properties and allows truly high-dimensional inference for extremes of spatial processes. We modify the popular random scale construction that multiplies a Gaussian random field by a single radial variable; that is, we add non-stationarity to the Gaussian process while allowing the radial variable to vary smoothly across space. As the level of extremeness increases, this single model exhibits both long-range asymptotic independence and short-range weakening dependence strength that leads to either asymptotic dependence or independence. Under the assumption of local stationarity, we make inference on the model parameters using local Bayesian hierarchical models, and run adaptive Metropolis algorithms concurrently via parallelization. Then, after conducting posterior inference locally, the mixture component representation of the model coherently ties the local posteriors together to obtain a globally nonstationary model

Time and Session: Thursday 16.00–17.15, CS Spatial Extremes (II)

Discrete dependent extremes

Linda Mhalla

University of Lausanne

Abstract

With view towards the current COVID-19 global epidemic, the study of time series of disease incidence is becoming increasingly important. In particular, we are interested in the temporal dynamics of the extremes of the number of fatalities during a pandemic. We propose constructions for time-dependent discrete generalized Pareto random variables, and specifically a hierarchical model for high discrete threshold exceedances by embedding a latent process with Gamma margins for the success probability of a geometric distribution. Based on pairwise-likelihood estimation of these models for the COVID-19 data in the US, we discuss the extremal dependence in the state-wise death toll while accounting for the inherent marginal non-stationarity through covariates.

Time and Session: Tuesday 10.00–11.15, IS Public health, epidemiology, life sciences and life lengths

Distributed inference for extreme value index

LiuJun Chen

Abstract

We investigate a divide-and-conquer algorithm for estimating the extreme value index when data are stored in multiple machines. The oracle property of such an algorithm based on extreme value methods is not guaranteed by the general theory of distributed inference. We propose a distributed Hill estimator, establish its asymptotic theories and provide sufficient, sometimes also necessary, condition, under which the oracle property holds. In applications, estimates based on the distributed Hill estimator can be sensitive to the choice of the number of the exceedance ratios used in each machine. Even with choosing the number at a low level, a high asymptotic bias may arise. We overcome this potential drawback by designing a bias correction procedure for the distributed Hill estimator, adhere to the setup of distributed inference. The asymptotically unbiased distributed estimator we obtained, on the one hand, is applicable for the distributed data, on the other hand, inherits all known advantages of bias correction methods in extreme value statistics.

Time and Session: Monday 11.30–13.00, CS Machine Learning for extremes

Modelling panels of extremes

Luca Trapin

University of Bologna

Abstract

Estimation of the parameters of the generalized extreme value distribution is notoriously challenging due to the small number of observations that are usually available in applications. Regression techniques are commonly used to pool information and to capture cross-sectional heterogeneity or non-stationarity in the data. We propose the notion of an extreme value panel consisting of several groups of individuals or gauging stations. We propose a new algorithm that jointly assigns the individuals to the latent groups and estimates the model parameters of the regression models inside each group. An extensive simulation study shows that this method efficiently recovers the underlying group structure without prior information. We apply our approach to risk assessment for financial, hydrological and climate data. All of these applications can be formulated as a panel problem and we show that our method provides improved estimates of extreme quantiles and helps to answer important domain-specific questions.

Time and Session: Tuesday 11.30–13.00, CS Regression techniques (I)

Informative selection mechanisms for extreme value analyses

Léo Belzile*, Anthony Davison, Holger Rootzén, Dmitrii Zholud and Jutta Gampe.

*HEC Montréal

Abstract

The sampling scheme is typically ignored in statistical analyses as it is assumed that the way in which data were collected is independent of the phenomenon under study. However, in many instances, the selection mechanism is informative and ignoring it leads to biased inference. We study the impact of allowing for the selection using two examples, starting with rainfall records subject to a stopping rule. We also consider lifelengths of semi-supercentenarian, trying to assess whether the support of the lifetime distribution is bounded or not.

Time and Session: Thursday 16.00–17.15, IS Inferential issues

Box-Cox estimation of parameters of extreme events

Lígia Henriques-Rodrigues*, Maria Ivette Gomes and

*University of Évora

Abstract

The reduction of bias of the Hill estimator has been extensively addressed in the literature of extreme value theory. Several techniques have been used to achieve such reduction of bias, either by removing the main component of the bias of the Hill estimator of the extreme value index (EVI) or by constructing new estimators based on generalized means or norms that generalize the Hill estimator. In statistical literature the Box-Cox transformations are used to make the data more suitable for statistical analysis. Using the regular variation theory, Teugles and Vanrolen, in 2004, studied the effect of the Box-Cox transformations in the speed of convergence of the second order condition and proposed the Box-Cox Hill estimator. In this work we are going to address the choice and estimation of the power and shift parameters of the Box-Cox transformation, not only for the EVI-estimation, but also for the estimation of other parameters of extreme events. We shall prove the consistency and asymptotic normality of these estimators and their performance for finite samples is illustrated through a small-scale Monte-Carlo simulation study.

Time and Session: Friday 11.30–13.00, CS Univariate tail estimation

Statistical inference for decomposable Hüsler-Reiss graphical models

Manuel Hentschel* and Sebastian Engelke

*University of Mannheim

Abstract

In [Engelke and Hitz \(2020\)](#), the authors introduce extremal graphical models for multivariate Pareto distributions. They define an analogue to the classical Markov property and show a Hammersley-Clifford theorem for the class of connected, decomposable graphs. Furthermore, they introduce methods for statistical inference on block graphs, a subclass of the much more general class of decomposable graphs.

In this work, we study properties of Hüsler-Reiss graphical models on general decomposable graphs. In particular, we prove a matrix completion theorem that shows that it is sufficient to specify the distribution on the lower-dimensional cliques of the graph and explicitly construct the unique completion of its variogram matrix. We further show how this can be translated into effective statistical inference.

References

Engelke S. and Hitz, A. S., 2020. Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 871–932.

Time and Session: Friday 10.00–11.15, CS Graphical models

Time scale determines the spatial patterns and extents of compound hot-dry events: an assessment using a multi-site multi-variable weather generator

Manuela Brunner

University of Freiburg

Abstract

Compound hot and dry weather and climate events can lead to severe societal impacts across a range of their time scales and spatial extents. Still, these two characteristics have received little attention despite the growing interest in climate change impacts on compound events. Here, we investigate how event time scale relates to spatial patterns of compound hot-dry events, the spatial extent of compound events, and the importance of temperature and precipitation as drivers of compound event occurrence using a gridded data set for the United States. To study such rare spatial and multivariate events, we introduce a multi-site multi-variable weather generator (PRSim.weather), which enables simulation of a large number of spatial compound hot-dry events. PRSim.weather combines an empirical spatio-temporal model based on the wavelet transform and phase randomization with two flexible distributions for temperature and precipitation. Our model evaluation shows that the stochastic model realistically simulates distributional and temporal autocorrelation characteristics of temperature and precipitation at single sites, dependencies between the two variables temperature and precipitation, spatial correlation patterns, and spatial heat and drought indicators and their co-occurrence probabilities. Our compound event analysis demonstrates that (1) compound hot-dry events are most likely in the Northwestern and Southeastern United States independent of the analysis time scale and this likelihood decreases with increasing time scale, (2) the spatial extent and time scale of compound events are strongly related with sub-seasonal events showing the largest extents, and (3) the importance of temperature and precipitation as drivers of compound events varies with the time scale of the analysis, where temperature and precipitation are most important at short and seasonal time scales, respectively. We conclude that time scale is an important choice in compound event assessments and suggest that climate change impact assessments should evaluate joint events using several instead of a single time scale to understand fully how their compound event characteristics may evolve.

Time and Session: Monday 16.00–17.15, IS Climate extremes

Long range dependence in the tails

Marco Oesting*, Vitalii Makogin, Albert Rapp and Evgeny Spodarev.

*University of Stuttgart

Abstract

The presence or absence of long memory in time series are known to have major effects on the asymptotic properties of statistical estimators. While classical notions of long range dependence typically rely on characteristics of the bulk of the distribution such as covariances, many common estimators in extreme value statistics are based on observations in the tails only. Motivated by this fact, in this talk, we propose to separately study long memory in the extremes as given by the tail process of a regularly varying times series and the corresponding max-stable analogue. Based on a recent definition of long range dependence by Kulik and Spodarev that is invariant under marginal transformations, we provide a necessary and sufficient criterion for long range dependence of max-stable time series in terms of the pairwise extremal coefficient function. We present statistical applications of this characterization and show its effect on limit theorems. Furthermore, we discuss the extension of the results to processes in the max-domain of attraction.

Time and Session: Thursday 11.30–13.00, IS Time series

A few progresses in statistics of extremes through the use of generalized means

Maria Ivette Gomes

Centro de Estatística e Aplicações

Abstract

Statistics of extremes, either univariate or multivariate, have been recently faced with many challenges, especially the ones related to topics like risk modelling of big data and robustness of the methodologies that enable to understand the complexity of extreme events in the most diverse areas of applications. In statistical extreme value theory (EVT), generalized means (GMs) have recently been used with success in the estimation of a positive extreme value index (EVI). Due to the specificity of the Weibull tail coefficient (WTC), its relevance and its deep link to a positive EVI, we shall now make use of GMs in the estimation of the WTC, also crucial for an adequate risk assessment. Regarding the mean-of-order- p estimation, we could always find a value of p (negative, contrarily to what happens with the mean-of-order- p EVI- estimation), such that, for adequate values of the threshold, there is a reduction in mean square error, as well as in bias. The lack of efficiency of the mean-of-order- p WTC-estimators for positive p , and of the mean-of-order- p EVI-estimators for negative p , together with results related to the robustness of the mean-of-order- p EVI-estimators associated with $p = -1$, deserves a further discussion of the topic ‘robustness versus efficiency’.

Time and Session: Monday 16.00–17.15, CS Inference and robust extremes

Capturing elements of weather-related risks in a climate change context

Marie Ekstrom

Cardiff University

Abstract

Climate risk and resilience are two terms that go hand-in-hand. The first must be quantified and characterised so that the latter can be achieved for a business, a community, an infrastructure asset or an ecosystem function or service. The requirement to assess climate risk is now well established in the United Kingdom, an integral part of e.g., planning/regulation policy, service agreements, or adhering to industry standards. Current terminology and praxis used to assess climate risk is strongly shaped by experiences in the disaster risk management community. The hazard, or extreme weather event, sits squarely in the centre of the risk framework, as the need to assess risk only arise when we expect the occurrence of a hazard. The full appreciation, or relevance, of the hazard becomes apparent only when evaluated together with its companion elements: exposure and vulnerability. The former often (in a climate change context) referring to the spatial footprint of the hazard, and the latter characteristics of the receptor that act to enhance or reduce the impact of the hazard. In climate change science, several disciplines act to evolve our understanding of each element.

This multi-disciplinary space, in combination with the adoption of a risk framework evolved primarily for disaster management makes for a rich learning environment, but it is a difficult space for practitioners to navigate because there is much discussion on how to quantify and capture each of those three risk elements for multi-decadal time frames. In this talk we will explore, with examples, some of these difficulties, highlighting the need for innovation in how climate risk can be meaningfully captured in risk assessments and associated metrics.

Time and Session: Monday 10.00–11.15, IS Natural hazards and impacts

Estimation of the extremal coefficient function based on a single observation

Marine Demangeot*, Emilie Chautru and Anne Sabourin

*MINES ParisTech - PSL University, Centre de Géosciences

Abstract

The extremal coefficient function is a bivariate measure of spatial dependence for stationary max-stable processes [Schlather and Tawn \(2003\)](#). It is usually estimated from time series, when the spatial object under study is observed through time (e.g. extreme precipitations, extreme temperatures, high concentrations of pollution in the air). However, in some cases, such types of data cannot be accessed: only one or just a few records are made available. This is the case, for instance, in mining resources estimation, soil contamination evaluation or any other applications where the phenomenon of interest either varies too slowly across time to hope for a decent time series, or is too expensive to sample from. This situation is rarely addressed in the spatial extremes community, contrary to geostatistics, which typically deals with such issues. A basic geostatistical tool is the so-called variogram [J.-P. Chilès and Delfiner P., \(2012\)](#), which is also a bivariate measure of spatial dependence. Considering the indicator variogram, above some threshold, of a stationary max-stable random field, we propose a new nonparametric estimator of the extremal coefficient function based on the variogram's Nadaraya-Watson estimator. The latter has been studied by [Garcia-Soidán et al. \(2004\)](#) and [Garcia-Soidán \(2019\)](#); from their work, we derive asymptotic properties of our estimator when it is computed from a single spatial set of observations. Namely, under some assumptions, we show that it is consistent and asymptotically normal. These results are illustrated by numerical experiments and a comparison with the well-known F-madogram based estimator [Cooley et al. \(2006\)](#) is performed. An application on a real dataset is also presented.

References

- Schlather M. and Tawn J. A., 2019. A dependence measure for multivariate and spatial extreme values: properties and inference *Biometrika* **90**, 139–156
- Chilès J. P. and Delfiner P., 2012. *Geostatistics: Modeling Spatial Uncertainty* (second edition) Wiley series in probability and statistics
- Cooley D., Naveau P. and Poncet P., 2006. Variograms for spatial max-stable random fields in *Lecture Notes in Statistics: Dependence in Probability and Statistics by Bertail P., Doukhan P. and Soulier P.*
- Garcia-Soidán, P. H. and Febrero-Bande M. and González-Manteiga W., 2004. Nonparametric kernel estimation of an isotropic variogram *Journal of Statistical Planning and Inference* **121**, 65–92
- Garcia-Soidán, P. H., 2019. Asymptotic normality of the Nadaraya–Watson semivariogram estimators *Biometrika* **16**, 479–503

Time and Session: Thursday 16.00–17.15, CS Spatial Extremes (II)

Detecting changes in daily precipitation extremes over the contiguous United States

Mark Risser

Lawrence Berkeley National Laboratory

Abstract

In spite of the diverse literature on spatial extreme value analysis, characterizing trends in extremes of an environmental process like daily precipitation for a large network of monitoring stations over a heterogeneous spatial domain remains a challenging statistical problem. Here, we compare and contrast two methods that are scalable to high-dimensional, heterogeneous spatial data sets, namely conditional independence-based GEV methods that utilize bootstrapping and hierarchical Bayesian methods for block maxima. Both approaches are used to detect long-term trends in precipitation extremes over the contiguous United States for a large network of several thousand weather stations. Locally, in spite of significant noise, we are able to detect statistically significant trends in seasonal precipitation extremes, generally resulting in larger and more frequent extreme events, although there are also important areas where the opposite is true.

Time and Session: Thursday 16.00–17.15, IS Inferential issues

Phase-type distributions for insurance pricing

Martin Bladt

UNIL

Abstract

We demonstrate through the use of matrix calculus a transparent analysis of inhomogeneous Markov models for life insurance where transition matrices commute. The resulting formulae are intuitive matrix generalizations of their single-state counterparts, and the absorption times are matrix versions of well-known scalar distributions. A further advantage of this approach is that it allows extending the analysis to the non-Markovian case where sojourns are Mittag-Leffler distributed, and where the absorption times are fractional phase-type distributed. Considering deterministic time transforms gives rise to fractional inhomogeneous phase-type distributions as absorption times. The latter underlying processes are an example of a regime where not only the present but also the history of a policyholder influences its future evolution. The sub-exponential nature of stable distributions translates into the multi-state insurance model as a random longevity risk at any given state of the chain.

Time and Session: Friday 10.00–11.15, IS Insurance

Asymptotic analysis of sampling probabilities and backward simulation algorithms for
coalescent models

Martina Favero* and Henrik Hult

*Stockholm University

Abstract

Observing a certain configuration of genetic material in a large sample is an event with a small probability which can be estimated by importance sampling based on backward simulation of coalescent processes. In this talk, a novel framework for the analysis of asymptotic properties of these backward simulation algorithms is presented. We start by showing that the sampling probabilities under the Kingman coalescent decay polynomially in the sample size. Then we present a weak convergence result for a sequence of Markov chains related to the coalescent, including associated cost chains. Finally we illustrate how these results can be applied to analyse asymptotic properties of backward sampling algorithms, in particular the asymptotic behaviour of importance sampling weights.

Time and Session: Friday 16.00–17.15, IS Rare event simulation

Conex-connect: learning patterns in extremal brain connectivity from multi-channel EEG data

Matheus Guerrero*, Raphaël Huser and Hernando Ombao

*KAUST

Abstract

Epilepsy is a chronic neurological disorder affecting more than 50 million people globally. An epileptic seizure acts like a temporary shock to the neuronal system, disrupting normal electrical activity in the brain. Epilepsy is frequently diagnosed with electroencephalograms (EEGs). Current methods study only the time-varying spectra and coherence but do not directly model changes in extreme behavior. To overcome this limitation, we propose a new approach to characterize brain connectivity based on the joint tail (i.e., extreme) behavior of the EEGs. Our proposed method, the conditional extremal dependence for brain connectivity (Conex-Connect), is a pioneering approach that links the association between extreme values of higher oscillations at a reference channel with the other brain network channels. Using the Conex-Connect method, we discover changes in the extremal dependence driven by the activity at the foci of the epileptic seizure. Our model-based approach reveals that, pre-seizure, the dependence is notably stable for all channels when conditioning on extreme values of the focal seizure area. By contrast, the dependence between channels is weaker during the seizure, and dependence patterns are more *chaotic*. Moreover, in terms of spectral decomposition, we find that extreme values of the high-frequency Gamma-band are the most relevant features to explain the conditional extremal dependence of brain connectivity.

Time and Session: Monday 16.00–17.15, CS Multivariate extremes

Large degrees in scale-free inhomogeneous random graphs

Matthias Schulte* and Chinmoy Bhattacharjee

*Hamburg University of Technology

Abstract

We consider a class of random graphs whose construction involves weights and whose degree distributions follow power-laws. Examples are some long-range percolation models, the random connection model with weights, the Norros-Reittu model and the Chung-Lu model. For such random graphs we study the maximum degree in a growing observation window and show that its limiting distribution is Fréchet. More generally, we establish that the point process of large degrees converges in distribution to an inhomogeneous Poisson process on the positive half-line. An important statistical question is to estimate the tail exponent of the degree distribution. Here we prove consistency of the Hill estimator.

Time and Session: Friday 11.30–13.00, IS Networks

Robust statistical learning

Matthieu Lerasle

CNRS/Crest/Ensaë

Abstract

I will present two tools that allow to obtain strong deviation properties of estimators when data are both heavy tailed and possibly contaminated. The first tool originally developed by Lugosi and Mendelson leads to a deviation inequality for median of means processes. The second tool is an homogeneity lemma that simplifies the more classical peeling argument when, for example, the loss is convex. I will illustrate the extent of these tools in various statistical learning problems.

Time and Session: Tuesday 10.00–11.15, IS Machine learning (theory, inc. tail-adapted loss functions, concentration inequalities)

Extreme partial least-squares regression

Meryem Bousebata*, Geoffroy Enjolras and Stéphane Girard

*Université Grenoble Alpes

Abstract

In this communication, we propose a new approach, called Extreme-PLS, for dimension reduction in regression and adapted to distribution tails. The goal is to find linear combinations of predictors that best explain the extreme values of the response variable by maximizing the associated covariance. This adaptation of the PLS estimator to the extreme-value framework is achieved in the context of a non-linear inverse regression model. In practice, it allows to quantify the effect of the covariates on the extreme values of the response variable in a simple and interpretable way. Moreover, it should yield improved results for most estimators dealing with conditional extreme values thanks to the dimension reduction achieved in the projection step. From the theoretical point of view, the asymptotic normality of the Extreme-PLS estimator is established under a heavy tail assumption but without recourse to linearity nor independence assumptions. The performance of the method is assessed on simulated data. Finally, the Extreme-PLS approach is used to analyse the influence of various parameters on extreme cereal yields collected on French farms. For further details see the preprint: <https://hal.inria.fr/hal-03165399/document>

Time and Session: Tuesday 11.30–13.00, CS Regression techniques (I)

On the approximation of extreme quantiles with ReLU neural networks

Michaël Allouche*, Stéphane Girard and Emmanuel Gobet

*Ecole Polytechnique

Abstract

Feedforward neural networks based on rectified linear units (ReLU) cannot efficiently approximate quantile functions which are not bounded in the Fréchet maximum domain of attraction. We thus propose a new parametrization for the generator of a generative adversarial network (GAN) adapted to this framework of heavy-tailed distributions. We provide an analysis of the uniform error between the extreme quantile and its GAN approximation. It appears that the rate of convergence of the error is mainly driven by the second-order parameter of the data distribution. The above results are illustrated on simulated data and real financial data.

Time and Session: Monday 11.30–13.00, CS Machine Learning for extremes

Concentration and asymptotic normality of the empirical variogram, with application to structure learning

Michaël Lalancette

University of Toronto

Abstract

Multivariate peaks-over-threshold methods often seek to fit a multivariate Pareto distribution to a subset of the data where at least one measured variable is large. An important property of multivariate Pareto distributions is a certain variogram matrix. For instance, if the data is attracted to a Hüsler–Reiss distribution, its extremal dependence structure is fully characterized by this matrix. In this talk we introduce novel theory for the empirical variogram, a natural nonparametric estimator of the variogram, both in the finite sample and asymptotic regimes.

We first present concentration inequalities that guarantee uniform consistency of the empirical variogram, while allowing the dimension of the matrix to grow superpolynomially in the sample size. We further showcase how the bounds may be used to justify different model selection techniques for graphical extreme value theory in high dimension. Examples include a minimum weight spanning tree and a graphical LASSO procedure.

In fixed dimension, we also characterize the asymptotic distribution of the empirical variogram. For this purpose, we prove weighted uniform weak convergence of a certain tail empirical process on infinite L-shaped sets, a result of its own interest.

Time and Session: Friday 16.00–17.15, IS Asymptotic statistics for extremes (inc. empirical processes)

Poisson approximation in the Poisson hyperplane mosaic

Moritz Otto

University of Magdeburg

Abstract

We consider point processes of centers of large cells in the Poisson hyperplane mosaic. As a first result, we shall discuss cells with large inradius in arbitrary dimension and derive a Poisson approximation result for the point process of their centers. Our argument uses an appropriate coupling with a Palm version of this process. In the proof we are facing the difficulty that sets that are arbitrarily far apart from each other can be hit by the same hyperplane. In a second step we will generalize our result to other size functionals (e.g. volume, surface area). Finally we shall discuss implications for the distributions of maximal cells.

Time and Session: Tuesday 17.30–18.45, CS Extremes of stochastic processes (I)

Tails and clusters of random sums and maxima and their relation to graphical models

Natalia Markovich

Abstract

Recent results of author for sums and maxima of non-stationary random length sequences of regularly varying distributed random variables (r.v.s) are presented. A doubly-indexed array of regularly varying r.v.s in which the *row index* corresponds to time, and the *column index* corresponds to the level is considered. Each *column* series is assumed to be stationary distributed with some tail and extremal indices. We focus on sums and maxima of weighted row sequences of random length. In [Markovich and Rodionov \(2020\)](#) it is assumed that there is a unique series with minimum tail index. The novelty is that *column* series are arbitrary mutually dependent. Conditions when tail and extremal indices of mentioned sums and maxima are the same are obtained. The random length is assumed to be regularly varying with a lighter tail than r.v.s in the series. In [Markovich \(2021\)](#) d *column* series with minimum tail index are assumed. All elements in pairs of the most heavy-tailed *column* sequences have to show the same mutual dependence. Then the sums and maxima have the same minimum tail index. If $d > 1$ is fixed and d *column* series are mutually independent, and independent of the rest of *column* series, then the sums and maxima have the same extremal index. If the independence

does not hold, then their extremal indices may not exist. If d is a bounded discrete r.v., then the extremal indices of the sums and maxima sequences do not exist. Theorems in [Markovich and Rodionov \(2020\)](#), [Markovich \(2021\)](#) are valid if there are non-zero elements in each *row* sequence corresponding to *column* sequences with minimum tail index. Particularly, if the most heavy-tailed column is unique and at least one of its elements is equal to zero, then the sums and maxima are non-stationary. This property plays a role for graphs. The sums and maxima may be associated with graphical model that allows us to obtain tail and extremal indices of PageRank (PR) and the Max-Linear Model (MLM), Markovich (2021b). The latter are used as node influence measures in complex networks. The *row* sequences serve as coordinates of points in \mathbb{R}^d . Then the PR of a newly attached node (a webpage) is determined as sums of PRs of node parents taken as the row elements. The MLM of the node is determined by maximum of the row elements. It is shown that the tail and extremal indices of the PR and MLM of attached nodes are determined by the graph community with minimum tail index. The PR of a webpage is often considered as the solution to the fixed-point problem (see, [Jelenkovic and Olvera-Cravioto \(2010\)](#), [Volkovich and Litvak \(2010\)](#) among others), where PRs of in-coming nodes are assumed to be iid. We allow these PRs to be dependent and non-stationary which is more plausible for real networks.

References

- Markovich N. M. and Rodionov I. V., 2020. Maxima and sums of non-stationary random length sequences *Extremes* **23**, 451–464
- Markovich N. M., 2021. Extremes of Sums and Maxima with Application to Random Networks *5th International Conference on Stochastic Methods*, Moscow
- Markovich N. M., 2021. Extremal properties of evolving networks: local dependence and heavy tails *Working paper*
- Volkovich Y. V. and Litvak N., 2010. Asymptotic analysis for personalized web search *Adv. Appl. Prob.* **42**, 577–604
- Jelenkovic P. R. and Olvera-Cravioto M., 2010. Information ranking and power laws on trees *Adv. Appl. Prob.* **42**, 1057–1093

Time and Session: Friday 10.00–11.15, CS Graphical models

Transformed-linear models for time series extremes

Nehali Mhatre

Colorado State University

Abstract

In order to capture the dependence in the upper tail of a time series, we develop non-negative regularly-varying time series models that are constructed similarly to classical non-extreme ARMA models. We first investigate consistency requirements among the finite-dimensional collections of the elements of a regularly-varying time series. We define the tail pairwise dependence function (TPDF) to quantify the extremal dependence between two elements of the regularly-varying time series, and use the TPDF to define the concept of weak tail stationarity for regularly-varying time series. To develop our non-negative regularly-varying ARMA-like time series models, we use transformed-linear operations. We show existence and stationarity of these models and develop their properties, such as the model TPDF's. Motivated by investigating conditions conducive to the spread of wildfires, we fit models to hourly windspeed data and find that the fitted transformed-linear models produce better estimates of upper tail quantities than alternative models.

Time and Session: 16.00–17.15, CS Time series

Causal inference for extremes on river networks

Ngoc Tran

UT Austin

Abstract

Causal inference for extreme aims to discover cause and effect relation between large observed values of random variables.

This is a fundamental problem to in many applications, from finance, engineering risks, nutrition to hydrology, to name a few. Unique challenges to extreme values are lack of data and lack of model smoothness due to the max operator. Existing methods in extreme value statistics for dimensions $d \geq 3$ are limited due to an intractable likelihood, while techniques for learning Bayesian networks require a large amount of data to fit nonlinear models. This talk showcases the max-linear model and new algorithms for fitting them. Our method performs well on real data, recovering a directed graph for both the Danube and the Lower Colorado with high accuracy purely through extreme measurements. We also compare our method to state-of-the-art algorithms for causal inference for nonlinear models, and outline open problems in hydrology, extreme value statistics and machine learning.

Time and Session: Monday 17.30–18.45, IS Graphical modelling

A Weissman-type estimator of the conditional marginal expected shortfall

Nguyen Ho*, Yuri Goegebeur, Armelle Guillou and Jing Qin

*University of Southern Denmark

Abstract

The marginal expected shortfall is an important risk measure in finance and actuarial science, which has been extended recently to the case where the random variables of main interest are observed together with a covariate. This leads to the concept of conditional marginal expected shortfall for which an estimator is proposed allowing extrapolation outside the data range. The main asymptotic properties of this estimator have been established, using empirical processes arguments combined with the multivariate extreme value theory. The finite sample behavior of the proposed estimator is evaluated with a simulation experiment, and the practical applicability is illustrated on vehicle insurance customer data.

Time and Session: Thursday 11.30–13.00, CS Insurance

Semi-parametric estimation of multivariate extreme expectiles

Nicholas Beck

HEC Montréal

Abstract

This paper focuses on semi-parametric estimation of multivariate expectiles for extreme levels of risk. Multivariate expectiles and their extremes have been the focus of plentiful research in recent years. In particular, it has been noted that due to the difficulty in estimating these values for elevated levels of risk, an alternative formulation of the underlying optimization problem would be necessary. However, in such a scenario, estimators have only been provided for the limiting cases of tail dependence: independence and comonotonicity. In this paper, we extend the estimation of multivariate extreme expectiles (MEEs) by providing a consistent estimation scheme for random vectors with any arbitrary dependence structure. Specifically, we show that if the upper tail dependence function, tail index, and tail ratio can be consistently estimated, then one would be able to accurately estimate MEEs. The finite-sample performance of this methodology is illustrated using both simulated and real data.

Time and Session: Monday 16.00–17.15, CS Multivariate extremes

Causal discovery in heavy-tailed models

Nicola Gnecco

University of Geneva

Abstract

Causal questions are omnipresent in many scientific problems. While much progress has been made in the analysis of causal relationships between random variables, these methods are not well suited if the causal mechanisms only manifest themselves in extremes. This work aims to connect the two fields of causal inference and extreme value theory. We define the causal tail coefficient that captures asymmetries in the extremal dependence of two random variables. In the population case, the causal tail coefficient is shown to reveal the causal structure if the distribution follows a linear structural causal model. This holds even in the presence of latent common causes that have the same tail index as the observed variables. Based on a consistent estimator of the causal tail coefficient, we propose a computationally highly efficient algorithm that estimates the causal structure. We prove that our method consistently recovers the causal order and we compare it to other well-established and non-extremal approaches in causal discovery on synthetic and real data. The code is available as an open-access R package.

Time and Session: Tuesday 16.00–17.15, IS Causal inference

Multivariate sparse clustering for extremes

Nicolas Meyer

University of Copenhagen

Abstract

Studying the tail dependence of multivariate extremes is a major challenge in extreme value analysis. Under a regular variation assumption, the dependence structure of the positive extremes is characterized by a measure, the spectral measure, defined on the positive orthant of the unit sphere. This measure gathers information on the localization of large events and has often a sparse support since such events do not simultaneously occur in all directions. However, it is defined via weak convergence which does not provide a natural way to capture this sparsity structure. In this talk, we introduce the notion of sparse regular variation which allows to better learn the tail structure of a random vector X . We use this concept in a statistical framework and provide a procedure which captures clusters of extremal coordinates of X . This approach also includes the identification of a threshold above which the values taken by X are considered as extreme. It leads to an efficient algorithm called MUSCLE. We illustrate our method on numerical experiments and on wind speed data in Ireland.

Time and Session: Thursday 16.00–17.15, IS Sparsity in high-dimensional extremes

Causal modelling of heavy-tailed variables and confounders

Olivier Pasche*, Valérie Chavez-Demoulin and Anthony Davison

*University of Geneva

Abstract

Identifying causation is central to understanding the world around us, and the field of causal inference has developed massively in recent decades. In many situations, causal mechanisms manifest themselves only in extreme events or simplify in the tails of distributions. Standard methods of causal inference are ill-suited for the study of such phenomena, and recent research has begun to forge links between causality and extreme value theory. This talk addresses a central challenge: the presence of confounders, which can make it hard or impossible to correctly infer causal relationships. We propose a method that removes or reduces the unwanted effect of a known confounder on an extremal causal analysis, by considering it as a covariate in the modelling, and then present a statistical test for direct causality between two observed extremal variables. This enables causal discovery and inference for a greater variety of situations, as confounders are present in many, if not all, real-world situations. The methodology is illustrated on discharge data from stations in the Rhine and Aare catchments in Switzerland.

Time and Session: Tuesday 16.00-17.15, Best student paper (III)

Threshold selection for cluster inference based on large deviation principles.

Olivier Wintenberger

LPSM, Sorbonne University

Abstract

In the setting of regularly varying time series, a cluster of exceedances is a short period for which the supremum norm exceeds a high threshold. We propose to study a generalization of this notion considering short periods, or blocks, with norm $\ell^p(\mathbb{R}^d)$ above a large threshold. We derive large deviation principles of blocks and apply these results to improve cluster inference. We focus on block estimators and show they are consistent when we use large quantiles from the sample of ℓ^p -norm over blocks as threshold levels. Our results lead to a threshold selection method for cluster inference.

Time and Session: Thursday 11.30–13.00, IS Time series

Concentration inequalities for NA random variables, applications to survey sampling

Patrice Bertail*, Stephan Cl  mencon, Guyonvarch Y. and Noiry N.

*Universit   Paris-Nanterre

Abstract

The purpose of this talk is first to present some Hoeffding and Bernstein bounds for negatively associated (NA) random variables and to apply them to some specific time series and to some NA survey sampling plans. We then show how it is possible to get concentration inequalities for empirical processes indexed by classes of functions of NA variables, under quite standard conditions. Applications to some specific survey sampling plans as well as connections to permutation tests are presented.

Time and Session: Tuesday 10.00–11.15, IS Machine learning (theory, inc. tail-adapted loss functions, concentration inequalities)

Modeling and exact simulation of non-stationary temperature maxima with max-infinitely divisible processes

Peng Zhong, Raphaël Huser and Thomas Opitz

KAUST

Abstract

The modeling of spatio-temporal trends in temperature extremes can help better understand the structure and frequency of heatwaves in a changing climate. Here, we study annual temperature maxima over Southern Europe using a century-spanning dataset observed at 44 monitoring stations. Extending the spectral representation of max-stable processes, our modeling framework relies on a novel construction of max-infinitely divisible (max-id) processes, which include covariates to capture spatio-temporal non-stationarities. Our new model keeps a popular max-stable process on the boundary of the parameter space, while flexibly capturing weakening extremal dependence at increasing quantile levels and asymptotic independence. This is achieved by linking the overall magnitude of a spatial event to its spatial correlation range, in such a way that more extreme events become less spatially dependent, thus more localized. Our model reveals salient features of the spatio-temporal variability of European temperature extremes, and it clearly outperforms natural alternative models. As risk assessment of spatial extreme events is often reported in terms of long-term return levels or return periods for spatial aggregates, which generally cannot be computed in explicit form for the proposed max-id processes, it is natural to rely on simulation to estimate such risk measures. If time allows, we will also describe novel exact simulation algorithms for general classes of max-id processes that are both fast and highly accurate.

Time and Session: Tuesday 11.30–13.00, CS Spatial extremes (I)

On maximal claim size for marked Hawkes processes

Petra Žugec, Bojan Basrak and Nikolina Milinčević

University of Zagreb

Abstract

We use the theory of branching and point processes to study asymptotic distribution of the maximal claim size for marked Hawkes processes, in which marks determine the size and other characteristics of the individual claims and influence arrival rate of the future claims. We present sufficient conditions under which the maximal claim size tends in distribution to a max-stable random variable.

Time and Session: Thursday 11.30–13.00, CS Insurance

A large deviations analysis of piecewise deterministic Markov processes for MCMC

Pierre Nyquist*, Joris Bierkens and Mikola Schlottke

*KTH Royal Institute of Technology

Abstract

Rare-event sampling is a hindrance to efficiently sample from Gibbs measures, especially in the settings of high dimension or low temperature, where the prevalence of (deep) local minima causes standard algorithms to converge slowly, at times rendering them useless. The standard tool for sampling from such measures is Markov chain Monte Carlo (MCMC) methods and constructing samplers that can overcome this difficulty related to rare-event sampling is an active area of research. In this talk I will discuss using tools related to rare events of the underlying empirical measure, specifically large deviation results, to analyse MCMC methods. After a general discussion of this approach, I will focus on MCMC algorithms based on piecewise deterministic Markov processes, a class of methods currently receiving a lot of attention due to their potential benefits, e.g. in the setting of large data sets.

Time and Session: Friday 16.00–17.15, IS Rare event simulation

Tails in networks: a tale of finding the right slope.

Pim van der Hoorn

Eindhoven University of Technology

Abstract

Network science has long been fascinated with studying the tail of the distribution of the degrees of nodes in a network. Early observations that this tail often seemed to behave as a regularly varying function with exponent between 1 and 2, have driven research in inferring these tail-exponents in network data and studying how they impact other network properties. It was however only recently that this problem was properly placed within the framework of extreme value theory. This link has allowed for well-established tools to be applied for analyzing degree distributions and made the problem of tail inference of degrees more mathematically grounded. However, degrees do not have a monopoly on regularly varying behavior in networks. There are many degree dependent structural measures of networks which seem to have regularly varying tails as well. An example of this is the clustering function, which measures for each value k the average fraction of triangles in which nodes with degree k participate. When the degrees have a regularly varying distribution, the tail of this functions decays with the degree as a regularly varying function as well. Here the exponent often depends on specific model parameters and can exhibit interesting phase-transitions. More interestingly, these exponents also exhibit some form of universality. Meaning that they behave similarly for a wide variety of different models. In this talk I will discuss both occurrences of regularly varying tails in networks. I will start by highlighting some interesting developments and new insights regarding the inference of tail-exponents in degree distributions. After that I will introduce several degree-dependent measures and their corresponding tail behavior. My goal is to open a scientific discussion on how to properly analyze such tails and understanding the relation between these and the tails of the degrees.

Time and Session: Friday 11.30–13.00, IS Networks

K-regular self-similar fragmentation process

Piotr Dyszewski*, Nina Gantert, Samuel G. G. Johnston, Joscha Prochno and Dominik Schmid

*Technische Universität München

Abstract

We study the asymptotics of the k -regular self-similar fragmentation process. For $\alpha > 0$ and an integer $k \geq 2$, this is the Markov process $(I_t)_{t \geq 0}$ in which each I_t is a union of open subsets of $[0, 1)$, and independently each subinterval of I_t of size u breaks into k equally sized pieces at rate u^α . Let k^{-N_t} and k^{-n_t} be the respective sizes of the largest and smallest fragments in I_t . By relating $(I_t)_{t \geq 0}$ to a branching random walk, we find that there exist explicit deterministic functions $g(t)$ and $h(t)$ such that $|n_t - h(t)| \leq 1$ and $|N_t - g(t)| \leq 1$ for all sufficiently large t .

Time and Session: Monday 10.00–11.15, IS Natural hazards and impacts

On the rate of concentration of maxima in Gaussian arrays

Rafail Kartsioukas

University of Michigan - Ann Arbor

Abstract

It has recently been established that the concentration of maxima phenomenon is the key to solving the exact sparse support recovery problem in high dimensions. This phenomenon, known also as relative stability, has been little studied in the context of dependence. Here, we obtain bounds on the rate of concentration of maxima in Gaussian triangular arrays. These results are used to establish sufficient conditions for the uniform relative stability of functions of Gaussian arrays, leading to new models that exhibit phase transitions in the exact support recovery problem. Finally, the optimal rate of concentration for Gaussian arrays is studied under general assumptions implied by the classic Berman condition.

Time and Session: Monday 17.30–18.45, Best student paper (II)

Estimation of cluster functionals for heavy tailed time series

Rafal Kulik

University of Ottawa

Abstract

In this talk we will consider multivariate regularly varying time series. Using probabilistic tools developed recently, we will study estimators of so-called cluster functionals. These cluster functionals quantify extremal clustering. We will present asymptotic theory for disjoint blocks, sliding blocks and runs estimators in the peak-over-threshold framework. In particular, we will show that all classes of estimators have the same asymptotic variance. This is in contrast to the situation when block maxima method is implemented.

Time and Session: Friday 16.00–17.15, IS Asymptotic statistics for extremes (inc. empirical processes)

Sub-asymptotic models for functional peaks-over-threshold modelling

Raphaël de Fondeville

Abstract

Functional peaks-over-threshold analysis has recently been introduced as the generalization to functions of the classical, and widely applied, univariate methodology. In this framework, functional exceedances are defined through a real valued functional, called risk functional, allowing to characterize and study complex extreme events. In practice, functional peaks-over-threshold analysis is attractive for its versatility in terms of risk definition, and the existence of models for which inference procedures are tractable in high-dimensions. These models, due to their asymptotic nature, are however likely to lack flexibility when applied to limited quantity of data. Indeed, when analysing events of increasing levels of intensity, decreasing trends in dependence have been observed in multiple environmental applications. To accommodate for such phenomenon, we propose several sub-asymptotic models based on log-Gaussian random functions that can be estimated using existing inference techniques and we give their regime of asymptotic dependence. This work offers to analysts realistic models that can be estimated in high-dimension with the freedom to choose between regimes of extremal dependence depending on their aversion to risk. We illustrate the models attractiveness and flexibility on several case studies.

Time and Session: Monday 11.30–13.00, IS Spatial extremes

Modeling and estimation of extreme Red Sea surface temperature hotspots

Raphaël Huser* and Arnab Hazra

*KAUST

Abstract

Modeling, estimation and prediction of spatial extremes is key for risk assessment in a wide range of geo-environmental, geo-physical, and climate science applications. In this work, we propose a flexible approach for modeling and estimating extreme sea surface temperature (SST) hotspots, i.e., high threshold exceedance regions, for the whole Red Sea, a vital region of high biodiversity. Our proposed model is a semiparametric Bayesian spatial mixed-effects linear model with a flexible mean structure to capture spatially-varying trend and seasonality, while the residual spatial variability is modeled through a Dirichlet process mixture of low-rank spatial Student- t processes to efficiently handle high dimensional data with strong tail dependence. With our model, the bulk of the SST residuals influence tail inference and hotspot estimation only moderately, while our approach can automatically identify spatial extreme events without any arbitrary threshold selection. Posterior inference can be drawn efficiently through Gibbs sampling. Moreover, we show how hotspots can be estimated from the fitted model, and how to make high-resolution projections until the year 2100, based on the Representative Concentration Pathways 4.5 and 8.5. Our results show that the estimated 95% credible region for joint high threshold exceedances include large areas covering major endangered coral reefs in the southern Red Sea.

Time and Session: Tuesday 10.00–11.15, CS Climate extremes (II)

Extreme value theory and chess ratings

Richard Smith*, Weiji Ma, Nikos Bosse, Jose Camacho Collados and David Smerdon

*University of North Carolina, Chapel Hill

Abstract

The Elo rating system is well established as a method of evaluating competitive chess players. However, since the early days of the system, women have been severely underrepresented at the top levels of the game. The *participation rate* hypothesis suggests that this can be explained in terms of the relative numbers of male and female players—the underlying populations are similar, but men prevail at the top level simply because there are so many more of them. I shall present two approaches to assessing this hypothesis, one based on permutation tests for the null hypothesis that the male and female populations have the same underlying distribution, the other directly using extreme value theory. Preliminary results suggest that the data are quite consistent with the participation rate hypothesis in some countries, but not in others, and we are exploring the potential reasons for this in more detail.

Time and Session: Monday 10.00–11.15, CS Applications of extremes (I)

A flexible Bayesian framework for modeling extreme spatial threshold exceedances using product mixtures of random fields

Rishikesh Yadav* and Raphaël Huser

*KAUST

Abstract

In this work, we develop a constructive modeling framework for extreme spatial threshold exceedances, based on general product mixtures of random fields possessing light or heavy-tailed margins and various spatial dependence characteristics, which are suitably designed to provide high flexibility in the tail and at pre-asymptotic levels. Our proposed model is akin to a recently proposed Gamma-Gamma model, but it possesses a higher degree of flexibility in its joint tail structure, capturing strong dependence more easily. Thanks to the model's hierarchical formulation, inference may be conveniently performed based on Markov chain Monte Carlo (MCMC) methods, and we demonstrate how the stochastic gradient Langevin dynamics (SGLD) algorithm can be exploited to fit our proposed model very efficiently in relatively big spatio-temporal dimensions despite the large number of latent parameters, while simultaneously censoring non-threshold exceedances and performing spatial prediction at multiple sites. We explore the theoretical properties of the proposed model, and illustrate the proposed methodology by simulation and application to daily precipitation data from Eastern Spain measured at about 100 stations over the period 2011-2020

Time and Session: Tuesday 16.00–17.15, CS Bayesian extremes

Universally limited lifespans despite individual heterogeneity

Ross Maller

Australian National University

Abstract

Despite extensive analysis of biological and actuarial data over many years, and many suggested underlying mechanisms, there is still no agreement as to whether human lifetimes have a finite natural limit, or are potentially unbounded. Huang, Maller and Ning (Ins. Math. Econ. 2020) constructed lifetables separately for a total of 231,129 Netherlands females and 73,788 males aged 65+ born in 1-year cohorts from 1893 to 1908. They fitted to each cohort a model consisting of a Gompertz distribution up till a data determined threshold age, followed by a Pareto distribution for greater ages. In each of the 16 cohorts the best fitting Pareto had an estimated finite right endpoint which, however, varied over a range of around 110-118 years for the different cohorts. This leads us to propose a general population model which embodies a fine distinction: our model incorporates heterogeneity in life-lengths, such that lifetimes are finite individually, but unbounded overall. With this we discuss and are able to reconcile contradictions between some previous studies.

Time and Session: Monday 10.00–11.15, CS Applications of extremes (I)

Annual maximum precipitation in Indonesia linked to climate variability: extreme value analysis

Saat Mubarrok and Chan Joo Jang

Ocean Circulation Research Center, Korea Institute of Ocean Science and Technology (KIOST), Busan, South Korea - Integrated Ocean Sciences, University of Science and Technology, Daejeon, South Korea - Program Study of Geophysics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

Abstract

Heavy rainfall frequently occurs in tropical climate countries, leading to a major disaster such as flood and landslide, especially in Indonesia. With climate change, the future change projection is important to reduce the impact of extreme rainfall. In this study, we investigated inter-annual variability of rainfall extremes and its relation with climate variability in Indonesia using a generalized extreme value (GEV) distribution based on daily rainfall data recorded from 1985 to 2014 at ten meteorological stations across the country. Maximum likelihood estimation was used to find the parameters of GEV distribution. We applied four GEV models whose location parameter consider climate variability, El Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), Madden-Julian Oscillation (MJO), and choose the best fit model using corrected Akaike Information Criterion (AICc) and likelihood ratio test. We used possible time-varying non-stationary behavior of the location parameter only and keep the scale and the shape parameter constant. Overall, trend-free prewhitening (TFPW) Mann-Kendall test shows a significant trend of annual maxima only in Surabaya station with averaged country trend is about 29.5 mm/day over 30 years period. Furthermore, two stations, Waingapu and Luwuk, covariate significantly with ENSO, while the other two stations, Perak and Jakarta, covariate insignificantly to IOD. Conversely, the MJO signal in annual maxima was less prominent in all stations and thus does not improve the stationary GEV model. Additionally, the estimated shape parameter at seven stations classified as the Gumbel distribution, but Jakarta, Perak, and Masamba classified as the Fréchet distribution. It concludes that the annual precipitation maxima in Indonesia are well described by the Gumbel distribution rather than the Fréchet or the Weibull distribution.

Time and Session: Thursday 10.00–11.15, CS Flood risk

Bi-factor and second-order copula models for item response data

Sayed H. Kadhem* and Aristidis K. Nikoloulopoulos

*University of East Anglia

Abstract

In the context of item response data, latent maxima, minima and means can arise depending on how a respondent considers specific items. An item might make the respondent think about past events. The case of a latent maxima/minima can occur if the participant's response is based on a best or worst case. For different dependent items based on latent maxima or minima, multivariate extreme value and copula theory can be used to select suitable distributions for the latent variables. Copulas that arise from extreme value theory have more probability in one joint tail (upper or lower) than expected with a multivariate normal (MVN) distribution. Even, in the case where the item responses are based on discretizations of latent variables that are means, then it is possible that there can be more probability in both the joint upper and joint lower tail, compared with MVN distributed latent variables. This happens if the respondents consist of a *mixture* population (e.g., different locations or genders). From the theory of elliptical distributions and copulas, it is known that the multivariate Student- t distribution as a scale mixture of MVN has more dependence in the tails. We propose copula extensions for bi-factor and second-order models. The construction of the models exploits the use of bivariate copulas that link the observed variables to the common and group-specific factors. Our general models include the Gaussian bi-factor and second-order models as special cases, have interpretations of latent maxima/minima (in comparison with latent means) and can lead to more probability in the joint upper or lower tail compared with the Gaussian bi-factor and second-order models. Details on maximum likelihood estimation of parameters for the bi-factor and second-order copula models are given, as well as model selection and goodness-of-fit techniques. Our general methodology is demonstrated with an extensive simulation study and illustrated for the Toronto Alexithymia Scale.

References

Kadhem S. H. and Nikoloulopoulos A. K., 2021. Bi-factor and second-order copula models for item response data preprint *arXiv:2102.10660*.

Time and Session: Thursday 17.30–18.45, CS Dependence modelling

Extremal graphical lasso and high-dimensional extremes

Sebastian Engelke*, Michaël Lalancette and Stanislav Volgushev.

*University of Geneva

Abstract

Statistical inference for the extremal graphical models introduced in Engelke and Hitz (2020, JRSSB) is so far restricted to simple structures called block graphs. We develop an extremal graphical lasso that can be used estimate in a data-driven way the structure in general Hüsler–Reiss graphical models. We propose an efficient algorithm and prove that it recovers the underlying graph structure consistently even for growing dimension d . This enables the use of the extremal graphical lasso in high-dimensional settings where the sample size n is comparable or larger than d . In extremes, such settings are of particular interest since the effective sample size, namely the number k of exceedances, is much smaller than n .

Time and Session: IS Multivariate extremes (sparsity, high-dimensional, copulas, anomaly detection)

Cluster based estimator for the spectral tail process

Sebastian Neblung, Holger Drees and Anja Janßen

University of Hamburg

Abstract

A new type of estimator for the spectral tail process of a regularly varying time series is introduced. The approach is based on the time change formula and a equivalent characterizing invariance property of the spectral tail process introduced by [Janßen \(2019\)](#). Using a projection technique, this property is fully exploited to define the new estimator. We state uniform asymptotic normality of this projection based estimator using sliding blocks limit theory of [Drees and Neblung \(2021\)](#). Results of a simulation study illustrate that the new procedure provides a more stable performance than previously proposed estimators. In particular, the new estimator often has a smaller variance.

References

- Drees H. and Neblung S., 2021. Asymptotics for sliding blocks estimators of rare events *Bernoulli* **27**, 1239–1269.
- Drees H., Janßen A. and Neblung S., 2021. Cluster based inference for extremes of time series preprint *arXiv:2103.08512*.
- Janßen A., 2019. Spectral tail processes and max-stable approximations of multivariate regularly varying time series *Stoch. Proc. Applic.* **129**, 1993–2009

Time and Session: Monday 11.30–13.00, Best student paper (I)

Implications of bottom-up framing for climate impact assessments

Seth Westra

The University of Adelaide

Abstract

There has been considerable recent interest in ‘bottom-up’ (or ‘scenario neutral’) framings of the climate impact assessment process, which places primacy on system values (and/or performance metrics and thresholds), stress testing and identification of alternative options. Applications of this framing have shown that: (1) there is rarely a single clearly defined and agreed upon ‘system’, with different stakeholders emphasising different values and in turn providing alternate system delineations, and moreover (2) system delineations invariably span multiple spatial and temporal scales and consider multiple atmospheric variables, meaning that (3) almost all climate stressors are ‘compound’, and that (4) most systems are vulnerable to combinations of extreme and non-extreme climatic (and indeed also non-climatic forcings). Implications for extreme value statistical methods are then discussed, arguing that a flexible and accessible ‘library’ of methods urgently needed to minimise transaction costs for integrating extreme value theory into bottom-up assessments.

Time and Session: Monday 10.00–11.15, IS Natural hazards and impacts

Extremes of the spatial impact of heat waves.

Shrijita Bhattacharya

Michigan State University

Abstract

A region is said to be under a heat wave spell if its temperature remains above a given threshold for a prolonged span of time. Heat waves are often accompanied with some of the serious consequences like forest fires, droughts, impeded agricultural productivity and high mortality rates. As such, there has been an increasing amount of research to allow for the quantification, prediction and inference on heat waves. In this paper, we develop a statistical framework for modeling the areal impact of heat waves. A time series is constructed by considering the area of US under profound heat wave activity over moving windows of time. We extensively make use of the Pickands-Balkema-de Haan theorem to model the extremes of this time series using the generalized Pareto distribution (GPD). As a main contribution, we estimate the out-of-sample return levels for this time series as a function of covariates like the season (time of the year) and El Niño Southern Oscillation (ENSO) index. Sensitivity analysis of the proposed method with respect to factors like intensity level, duration and geographical location of heat wave events have been well studied. The proposed methodology has been evaluated in the context of United States Historical Climatology Network (USHCN) daily temperature records collected over a span of 100 years across the continental US.

Time and Session: Tuesday 10.00–11.15, CS Climate extremes (II)

New representations of Hermite processes

Shuyang Bai

University of Georgia

Abstract

Hermite processes are a class of self-similar processes with stationary increments. They often arise in limit theorems under long-range dependence. We introduce new representations of Hermite processes with multiple Wiener-Ito integrals, whose integrands involve the local time of intersecting stationary stable regenerative sets.

Time and Session: Thursday 16.00–17.15, IS Long memory processes and non-standard EVT

Modelling extreme sub-daily precipitation with the blended generalized extreme value distribution

Silius M. Vandeskog

Norwegian University of Science and Technology

Abstract

Short-term extreme precipitation can cause flash floods, large economic losses and immense destruction of infrastructure. In order to prepare for future extreme precipitation, a Bayesian hierarchical model is applied for estimating return levels for the yearly maxima of sub-daily precipitation in Norway.

A modified version of the generalized extreme value (GEV) distribution, called the blended GEV (bGEV) distribution, is used as the model for yearly maxima of sub-daily precipitation. Inference with the GEV distribution is known to be difficult, partly because its support depends on its parameters. The bGEV distribution has the right tail of a Fréchet distribution and the left tail of a Gumbel distribution, resulting in a constant support that allows for more stable inference. Fast inference is performed using integrated nested Laplace approximations (INLA).

We propose a new model for block maxima that borrows strength from the peaks over threshold methodology by linking the scale parameter of a block maximum to the standard deviation of observations larger than some threshold. The new model is found to outperform the standard block maxima model when fitted to the yearly maxima of sub-daily precipitation in Norway. Evaluation is performed with the threshold weighted continuous ranked probability score (twCRPS), where the weight function only focuses on large quantiles.

Time and Session: Tuesday 11.30–13.00, CS Spatial extremes (I)

Joint inference on extreme expectiles for multivariate heavy-tailed distributions

Simone Padoan

Bocconi University

Abstract

The notion of expectiles, originally introduced in the context of testing for homoscedasticity and conditional symmetry of the error distribution in linear regression, induces a law-invariant, coherent and elicitable risk measure that has received a significant amount of attention in actuarial and financial risk management contexts. A number of recent papers have focused on the behaviour and estimation of extreme expectile-based risk measures and their potential for risk management. Joint inference of several extreme expectiles has however been left untouched; in fact, even the inference of a marginal extreme expectile turns out to be a difficult problem in finite samples. We investigate the simultaneous estimation of several extreme marginal expectiles of a random vector with heavy-tailed marginal distributions. This is done in a general extremal dependence model where the emphasis is on pairwise dependence between the margins. We use our results to derive accurate confidence regions for extreme expectiles, as well as a test for the equality of several extreme expectiles. Our methods are showcased in a finite-sample simulation study and on real financial data.

Time and Session: Friday 10.00–11.15, IS Insurance

On modeling tail dependence via t-copula

Siyang Tao

Ball State University

Abstract

The tail dependence coefficient is a bivariate measure of dependence in the tail, and the tail dependence matrix (TDM) is a bidimensional array of these coefficients corresponding to a random vector. The TDM serves as a parsimonious measure of multivariate tail dependence akin to the correlation matrix in the context of dependence. The set of all TDMs corresponding to d -dimensional random vectors is a convex polytope with an intricate description that is known only for d less than or equal to six. In particular, both the numbers of facets and vertices of the set of all TDMs grow exponentially in d . We posit that the richness of the subset of the set of all TDMs that can be accommodated by a copula family is a practically important feature in its choice for modeling in the presence of tail dependence. In this talk, our focus is on the t -copula family that is a popular choice, especially in the presence of tail dependence, in many areas of statistical modeling including risk management and financial econometrics. We discuss some geometric properties of the subset of the set of all TDMs that is supported by the t -copula family and provide an efficient algorithm to determine the t -copula that best captures the tail dependence specified using a target TDM.

Time and Session: Thursday 17.30–18.45, CS Dependence modelling

Modelling the extremes of bivariate mixture distributions with application to oceanographic data

Stan Tendijck

Lancaster University

Abstract

There currently exist a variety of statistical methods for modelling bivariate extremes. However, when the dependence between variables is driven by more than one latent process, these methods are likely to fail to give reliable inferences. We consider situations in which the observed dependence at extreme levels is a mixture of a possibly unknown number of much simpler bivariate distributions. For such structures we demonstrate the limitations of existing methods and we propose two new methods: an extension of the Heffernan–Tawn conditional extreme value model to allow for mixtures and an extremal quantile-regression approach. The two methods are examined in a simulation study and are applied to oceanographic data. We discuss extensions including a subasymptotic version of the proposed model, which has the potential to give more efficient results by incorporating data that are less extreme. Both of the new methods outperform existing approaches when mixtures are present. The quantile-regression model performs better than the extension of the Heffernan–Tawn conditional extreme value model when the number of mixture components is unknown, and the methods are comparable when this number is known.

Time and Session: Thursday 17.30–18.45, CS Dependence modelling

Tree structure learning for extremes

Stanislav Volgushev* and Natalia Nolde

*University of Toronto

Abstract

We examine the differences between spatial extreme value representations when focusing on the joint tail versus taking the approach of conditioning on a single location being extreme. As a case study, we explore the model proposed in Huser and Wadsworth (2019), and elucidate why parameter changes that result in different strengths of joint tail dependence do not alter the limiting conditional extremes representation. This is illustrated geometrically via the shape of limit sets for the finite-dimensional margins.

Time and Session: Thursday 16.00–17.15, IS Sparsity in high-dimensional extremes

Exceedance probability estimation: some experience on bias correction and confidence intervals

Stefan Aulbach* and Jan Erik Stellet

*Robert Bosch GmbH

Abstract

In applications, common tasks are to estimate the probability that a future observation will exceed a certain high threshold and, additionally, to state how *confident* one is in that estimate – expressed as a confidence interval. Particularly in the case that the threshold lies beyond the sample maximum, generalized Pareto distributions (GPDs) are a natural tool in the extrapolation of the probability in question. An advantage of a maximum likelihood estimation is that it can provide both quantities – the point estimate and the confidence interval – simultaneously, e.g., via the profile likelihood function. However, it is a common problem that, mostly due to the limited availability of data, a bias occurs in the estimation results. In this talk, we briefly revisit the analytical bias correction formulas for the GPD-parameters derived by Giles et al. (2016), which are based on results by Cox and Snell (1968) and Cordeiro and Klein (1994), and apply them in the estimation of the exceedance probability and the corresponding confidence bounds. Afterwards, we discuss the current state of a multivariate extension of the presented approach and compare two variants in the context of a bivariate logistic GPD model based on simulation results.

Time and Session: Thursday 16.00–17.15, IS Inferential issues

Consistency of Bayesian and empirical Bayesian inference on multivariate max-stable distributions

Stefano Rizzelli

Catholic University of the Sacred Heart, Milan

Abstract

Predicting extreme events is important in many applications in risk analysis. Under mild assumptions, extreme-value theory justifies modelling extremes by max-stable distributions. The Bayesian approach provides a natural framework for statistical prediction. Although various Bayesian inferential procedures have been proposed in the literature of univariate extremes and some for multivariate extremes, the study of their asymptotic properties has been left largely untouched. In this contribution we focus on a semiparatic Bayesian method for estimating max-stable distributions in arbitrary dimension. We establish consistency of the pertaining posterior distributions for fairly general, well-specified max-stable models, whose margins can be short-, light- or heavy-tailed. We then extend our consistency results to the case where the data come from a distribution lying in a neighbourhood of a max-stable one, resorting to data-dependent prior distributions.

Time and Session: Tuesday 16.00–17.15, CS Bayesian extremes

Extremes of Markov random fields on block graphs

Stefka Asenova* and Johan Segers

*UCLouvain

Abstract

We study the joint occurrence of large values of a Markov random field or undirected graphical model associated to a block graph. Such graphs containing trees as special cases, we aim to generalize recent results for extremes of Markov trees. Every pair of nodes in a block graph is connected by a unique shortest path. These paths are shown to determine the limiting distribution of the properly rescaled random field given that a fixed variable exceeds a high threshold. When the said limit is a Hüsler–Reiss multivariate Pareto distribution, the local Markov property of the original field induces a particular structure on the limiting parameter matrix. Thanks to these algebraic relations, the parameters are still identifiable even if some variables are latent. The precise identifiability criterion turns out to be close to the one for Markov trees. Still on block graphs, we study an additive factor graphical model with respect to a directed acyclic graph. The model has the same form as a max-linear Bayesian network but with maxima replaced by sums. The factor variables are assumed to be heavy-tailed and absolutely continuous, properties which then transfer to the joint distribution. We give a sufficient condition on the graph structure for such a Bayesian network to also satisfy the global Markov property with respect to the undirected graph. The latter property then guarantees that the joint extremes of the model have the special form induced by the shortest (undirected) paths connecting nodes. The identifiability of the parameters of the limit distribution in case some variables are latent is discussed as well.

Time and Session: Tuesday 16.00-17.15, Best student paper (III)

The PCA-based control charts for monitoring multiple-stream processes

Su-Fen Yang

National Chengchi University

Abstract

The cosmetic bottle cover weights are collected from a filling machine with eight identical filling heads every 4 hours. The cover weights are a critical quality variable of the bottles. The weights produced from every filling head should be identical, independent, and satisfy the requested specification limits of customers. However, the variation of the weights from each filling head is large, and the weights distributions from each filling machine are correlated but not normal and identical. Hence, the multiple processes of the bottle weights are the kind of multiple-stream processes. There are some in-control and out-of-control data of the bottle weights are collected from the multiple-stream processes, and the data are analyzed. To monitor whether the processes are in-control or out-of-control, the principal component analysis (PCA)-based EWMA mean and covariance control charts are proposed to demonstrate the in-control samples and monitor the out-of-control samples. We found that the proposed PCA-based EWMA mean and covariance control charts perform well.

Time and Session: Monday 10.00–11.15, CS Applications of extremes (I)

Extremal dependence as given by the tail pairwise dependence matrix in precipitation and temperature data

Svenja Szemkus

Rheinische Friedrich-Wilhelms-Universität Bonn

Abstract

A better understanding of the dynamics and impacts of extreme weather events and their changes due to climate change is the subject of the ClimXtreme project (climxtreme.net) funded by the German Federal Ministry of Education and Research. The CoDEx project is investigating how data compression techniques can contribute to a better description and understanding of extremes. Various unsupervised learning approaches, such as clustering or principal component analysis, focusing on extremes have been developed recently. We use principal component analysis to study the spatial (co-)occurrence during extreme weather events such as heavy precipitation, heat waves or droughts. The focus on extreme events is done by using the tail pairwise dependence matrix (TPDM), proposed by Cooley and Thibaud (2019) as an analogue to the covariance matrix for extremes. Since the simultaneous occurrence of precipitation deficits and high temperature played an important role, especially in heat waves, we explore how Cooley and Thibaut's concept can be used in this regard. We propose an estimation of the TPDM based on pairwise dependencies of two variables. A singular value decomposition gives us insight into the spatial co-occurrence of extreme spatial patterns, contributing to the understanding of so-called compound events. We use daily precipitation and temperature data both at observation stations and in regional reanalyses in Germany and Europe, extract spatial patterns, and investigate historical events using this method.

Time and Session: Thursday 16.00–17.15, CS Spatial Extremes (II)

Convergence of persistence diagram in the subcritical regime

Takashi Owada

Purdue University

Abstract

The objective of this paper is to examine the asymptotic behavior of persistence diagrams associated with Čech filtration. A persistence diagram is a graphical descriptor of a topological and algebraic structure of geometric objects. We consider Čech filtration over a scaled random sample $r_n^{-1}\mathcal{X}_n = \{r_n^{-1}X_1, \dots, r_n^{-1}X_n\}$, such that $r_n \rightarrow 0$ as $n \rightarrow \infty$. We treat persistence diagrams as a point process and establish their limit theorems in the subcritical regime: $nr_n^d \rightarrow 0$, $n \rightarrow \infty$. In this setting, we show that the asymptotics of the k th persistence diagram depends on the limit value of the sequence $n^{k+2}r_n^{d(k+1)}$. If $n^{k+2}r_n^{d(k+1)} \rightarrow \infty$, the scaled persistence diagram converges to a deterministic Radon measure almost surely in the vague metric. If r_n decays faster so that $n^{k+2}r_n^{d(k+1)} \rightarrow c \in (0, \infty)$, the persistence diagram weakly converges to a limiting point process without normalization. Finally, if $n^{k+2}r_n^{d(k+1)} \rightarrow 0$, the sequence of probability distributions of a persistence diagram should be normalized, and the resulting convergence will be treated in terms of the \mathcal{M}_0 topology.

Time and Session: Tuesday 17.30–18.45, CS Extremes of stochastic processes (I)

Scoring and validation of dynamic probability forecast

Thibault Modeste

Institut Camille Jordan

Abstract

Forecast and its evaluation are major task in statistic. In real applications, forecast often take the form of a dynamic process evolving over time and this sequential point of view must be taken into account. We propose and discuss a minimal framework for dynamic probability forecast and its evaluation. Proper scoring rules are a crucial concept for probability forecast evaluation and we show, under minimal assumptions, that they can still be used in the dynamic framework because they are minimized, in the sense of the long term average score, by the ideal forecast. Another strategy for forecast evaluation is calibration theory based on the probability integral transform. Here ideal forecast is characterized by conditional calibration and we present some new tests based on regression trees that we compare to the ones proposed by Straehl and Ziegel (EJS 2017) in the framework of cross-calibration.

Time and Session: Thursday 11.30–13.00, CS Prediction and validation for extremes

Stochastic geometry of Gaussian mixture processes and spatial extreme-value analysis

Thomas Opitz*, Anne Estrade and Elena di Bernardino

*BioSP, INRAE

Abstract

Excursion sets of spatial stochastic processes above a high threshold allow characterizing extreme event episodes. The distributions of geometric characteristics of these sets, such as their area, perimeter or Euler characteristic, provide information about the structure of spatial clusters of extreme values. The estimation of these geometric characteristics has strong potential to provide new insights in climatological, environmental and ecological applications. In this work, we consider the setting of independent replicates of Gaussian processes with a random variable embedded for the mean or the variance of the process, leading to Gaussian location or scale mixture processes, respectively. Members of this class of processes have been used for flexible modeling of spatial data. Moreover, standard models in spatial extreme-value theory for threshold exceedances above a high threshold, known as generalized Pareto processes, arise as limits of such constructions when the threshold tends to infinity. Using classical results from the literature for the stochastic geometry of excursion sets in Gaussian processes, we extend the Gaussian theory to Gaussian mixture processes and their limits. In particular, we derive expectation formulas for geometric summaries. Finally, we illustrate estimation of such properties on simulated and real data.

Time and Session: Monday 11.30–13.00, IS Spatial extremes

Reciprocity and large degree dependence in a preferential attachment model

Tiandong Wang

Texas AM University

Abstract

Empirical studies show that online social networks have not only in- and out-degree distributions with Pareto-like tails but also a high proportion of reciprocal edges. A classical directed preferential attachment (PA) model generates in- and out-degree distribution with power-law tails, but theoretical properties of the reciprocity feature in this model have not yet been studied. We derive the asymptotic results on the number of reciprocal edges between two fixed nodes, as well as the proportion of reciprocal edges in the entire PA network. We see that with certain choices of parameters, the proportion of reciprocal edges in a directed PA network is close to 0, which differs from the empirical observation. This points out one potential problem of fitting a classical PA model to a given network dataset with high reciprocity and indicates alternative models need to be considered. We then discuss one possibility and study the dependence structure between large in- and out-degrees.

Time and Session: Tuesday 17.30–18.45, CS Applications of extremes (II)

Multivariate extremes for correlated Brownian motions with drift

Tomasz Rolski

Mathematical Institute, Wrocław University

Abstract

Let $(B_1(t) - \mu_1 t, \dots, B_d(t) - \mu_d t)$ be a correlated, d -dimensional Brownian motion with drift. In contrast with one dimensional theory, when by an extreme event is meant a crossing of a high (or small) level, it is not obvious what is a counterpart for multi-dimensional cases. In this talk we will discuss two approaches.

In [Dębicki et al. \(2020\)](#) we consider

$$P(u) = \mathbb{P}\{\exists_{t \geq 0} \forall_{i=1}^d (B_i(t) - \mu_i t \alpha_i u)\}, \text{ as } u \rightarrow \infty.$$

In [Dębicki et al. \(2019\)](#) another concept of an extreme is proposed, that is a coordinate-wise crossing of high barrier

$$P(u) = \mathbb{P}\{\forall_{i=1}^d \exists_{t_i \geq 0} (B_i(t_i) - \mu_i t_i \alpha_i u)\} \text{ as } u \rightarrow \infty,$$

In the second case the marginals $X_i = \sup_{t \geq 0} B_i(t) - \mu_i t$ are exponentially distributed, but properties of such a multivariate exponential distribution are not known. We give asymptotics for $P(u)$ as $u \rightarrow \infty$, although in the second case for $d = 2$ only.

A general theory for extremes of Gaussian processes is outlined in [Dębicki et al. \(2010\)](#).

References

- Dębicki, K., Kosinski, K. Mandjes, M. and Rolski, T., 2010. Extremes of multidimensional Gaussian process *Stoch. Proc. Applic.* **120**, 2289–2301
- Dębicki K., Ji L. and Rolski T., 2020. Exact asymptotics of component-wise extrema of two-dimensional Brownian motion *Extremes* **23**, 569–602
- Dębicki K., Ji L. and Rolski T., 2019. Logarithmic Asymptotics for Probability of Component-Wise Ruin in a Two-Dimensional Brownian Model *Risks* **7**, 83
- Dębicki K., Hashorva E., Ji L., 2020. Extremal behavior of hitting a cone by correlated Brownian motion with drift *Stoch. Proc. Applic.* **128**, 4171–4296

Time and Session: Monday 10.00–11.15, IS Extremes of stochastic processes (ambit, Gaussian)

Tail probabilities of random linear functions of regularly varying random vectors

Vicky Fasen-Hartmann

Karlsruhe Institute of Technology

Abstract

We provide a new extension of Breiman's Theorem on computing tail probabilities of a product of random variables to a multivariate setting. In particular, we give a characterization of regular variation on cones in $[0, \infty)^d$ under random linear transformations. This allows us to compute probabilities of a variety of tail events, which classical multivariate regularly varying models would report to be asymptotically negligible. We illustrate our findings with applications to risk assessment in financial systems and reinsurance markets under a bipartite network structure.

Time and Session: Friday 11.30–13.00, IS Networks

Multifidelity Monte Carlo estimation for extremes

Vladas Pipiras

University of North Carolina, Chapel Hill

Abstract

When modeling a random phenomenon, data are often available from multiple sources, or models, of varying fidelity, those with higher fidelity carrying higher costs. Multifidelity Monte Carlo (MFMC) methods offer tools that allow combining the data from multiple sources for better estimation of quantities for high-fidelity models. With a few exceptions though, much of the focus of the MFMC literature has been on characterizing uncertainty related to averages, in the context of non-rare problems where data are available to estimate these averages directly. In this work, we extend some MFMC methods to estimation of extremal quantities, for example, probabilities of rare events, possibly those that have not been observed in high-fidelity data. The suggested approaches are based on statistical extreme value theory, applied to simultaneously extreme observations from low-fidelity and high-fidelity models. The ideas are illustrated with synthetic data examples and the application to extreme ship motions.

Time and Session: Tuesday 17.30–18.45, CS Applications of extremes (II)

Spherical clustering in detection of groups of concomitant extremes

Vladimir Fomichov* and Jevgenijs Ivanovs

*Aarhus University

Abstract

There is growing empirical evidence that spherical k-means clustering performs well in identification of groups of concomitant extremes in high dimensions, thereby leading to sparse models.

We provide first theoretical results supporting this approach, but also identify some pitfalls. Furthermore, we show that an alternative cost function may be more appropriate for identification of concomitant extremes, and it results in a novel spherical k-principal-components clustering algorithm. Our main result establishes a broadly satisfied sufficient condition guaranteeing the success of this method, albeit in a rather basic setting.

Finally, we illustrate in simulations that k-principal-components outperforms k-means in the difficult case of weak asymptotic dependence within the groups.

Time and Session: Thursday 16.00–17.15, IS Sparsity in high-dimensional extremes

Marine heatwaves in Korean waters: seasonal and regional differences

Wonkeun Choi and Chan Joo Jang

Korea Institute of Ocean Science and Technology

Abstract

Marine heatwaves (MHWs) are phenomena in which the sea surface temperature (SST) extremely increases over thousands of kilometers for days to months. MHWs can cause substantial socio-economic damages along with the impacts in marine ecosystem such as a decrease in catch production and species diversity, and also contribute to increases in the occurrences of mass mortality of farming fishes and harmful algae blooms. Korean waters are regions where MHWs occur due to a combination of various factors such as atmospheric phenomena and ocean influences, and each sea area has different characteristics and seasonal fluctuations. In this study, we aim to investigate the characteristics of MHWs in the Korean waters focusing on their seasonal and regional differences, using OISST (Optimum Interpolation Sea Surface Temperature) data for the recent 37 years (1982-2018). The MHWs in the East Sea was generally 1.5 times stronger than in the Yellow Sea and the East China Sea; but exceptionally, in spring, the MHW in the Yellow Sea was the strongest. In summer, MHWs were as strong as about 3°C throughout the East/Japan Sea, but in fall and winter, distributions of MHWs intensity(3°C) were relatively concentrated on the polar front in the East/Japan Sea. In particular, in the fall, the influence of ocean current is dominant on the occurrence of MHWs in the East/Japan Sea. Our findings suggest that MHWs over Korean waters are different depending on the season and region, and also ocean variability could have a larger contribution than atmosphere for initiation of extensive MHWs. We will also present return level and return period of MHWs calculated by extreme value analysis.

Time and Session: Tuesday 10.00–11.15, CS Climate extremes (II)

Extreme conditional quantiles for panel data model with individual effects

Xuan Leng, Yanxi Hou and Yinggang Zhou

Xiamen University

Abstract

Panel quantile regression models play an important role in real applications of finance, econometrics, insurance and risk management. However, direct estimates of the extreme conditional quantiles may lead unstable results due to data sparsity on the tail regions. Moreover, the presence of individual effects complicates the inference for extreme quantiles and a study on their theoretical properties is necessary. This paper proposes a two-stage method to estimate/predict the extreme conditional quantiles where an intermediate quantile is first estimated according to panel regression models and the extrapolation of the intermediate quantile to an extreme quantile is carried out in the second stage. Under a set of second-order regular variation conditions of heteroscedastic extremes, we establish the asymptotic theories for the two-stage prediction while its finite sample performance is demonstrated and compared to the direct prediction by simulations. Finally, we apply the two-stage method to the macroeconomic and housing price data, and find strong evidence of housing bubbles and common economic factors as well as the cross country heterogeneity.

Time and Session: Tuesday 11.30–13.00, CS Regression techniques (I)

Long-range clustering for extremes

Yizao Wang*, Olivier Durieu, and Gennady Samorodnitsky

*University of Cincinnati

Abstract

In this talk I will go over a few recent examples of stationary sequences, of which the extremes form long-range clustering. The phenomena of clustering of extremes have been extensively investigated in the literature since 1980s. However, for most examples the extremal clustering occurs only locally. That is, the locations of extremes within each cluster are bounded and shrink to a single point at the macroscopic level after normalization. For long-range clustering, on the other hand, the locations of extremes within each cluster are unbounded, and they can be further characterized by a random closed set in the scaling limit.

The talk will be of expository nature and aim at providing an overview of recent results on long-range clustering of extremes arising from two classes of models: the Karlin models and the stable-regenerative models.

Time and Session: 16.00–17.15, CS Time series

The use of extreme value theory for forecasting long-term substation maximum electricity demand

Yun Li

Business Intelligence Analytics, Western Power

Abstract

Substation annual maximum electricity demand events are extreme, as customers respond to infrequent and extreme weather. Despite the extreme nature of annual maximum demand, the statistical theory of extreme values has only rarely, if ever, been applied. To support long term planning, utilities typically complete energy consumption and maximum demand forecasts, which are often conducted separately through two different process, leading to inconsistent trends and messages. To address these shortcomings, a point process approach from extreme value theory is proposed to model substation maximum demand as a function of trends in three common factors already required by utilities including customer count, average demand, and installed photovoltaic capacity. The point process model can be parameterized as a nonstationary generalized extreme value distribution with location and scale parameters dependent on the trends of these factors. As the generalized extreme value distribution governs the behaviors of block maxima (annual maximum demand) with forecast trends of three common factors, substation maximum demand can be estimated as per quantiles required by planning standards. Therefore, the proposed approach is not only realistic and flexible to forecast maximum demand but also ensures consistent outcomes and messaging between the two outputs from energy consumption and maximum demand forecasts.

Time and Session: Thursday 10.00–11.15, IS Extremes of energy systems

Robust estimation of the conditional stable tail dependence function

Yuri Goegebeur, Armelle Guillou, and Jing Qin

University of Southern Denmark

Abstract

We consider the estimation of the stable tail dependence function when the variables of main interest are recorded along with a random covariate. In particular we focus on the development of a robust estimator, obtained by locally applying the minimum density power divergence criterion to suitably transformed observations. Under classical regularity conditions, we derive the finite dimensional weak convergence of the estimator, after proper normalisation. The performance of our estimator in terms of efficiency and robustness is illustrated through a small simulation study.

Time and Session: Monday 16.00–17.15, CS Inference and robust extremes

Extremes of subexponential processes under moderate long memory

Zaoli Chen

Cornell University

Abstract

We consider extremes in a stationary sequence with subexponential tails in the maximum domain of attraction of the Gumbel distribution. We establish extremal limit theorems and the limits have Gumbel distribution. Our results demonstrate that the *heuristic of a single big jump* could break down even under moderate long-range dependence. Additionally, as tails range from lognormal distributions to semiexponential distributions, the extremal clusterings become finer. When the tails are light enough, a sharp phase transition that involves Cox processes will appear.

Time and Session: 16.00–17.15, CS Time series

Modeling spatial extremes using normal mean-variance mixtures

Zhongwei Zhang*, Raphaël Huser, Thomas Opitz, Jennifer L. Wadsworth

*KAUST

Abstract

Classical models for multivariate or spatial extremes are mainly based upon the asymptotically justified max-stable or generalized Pareto processes. These models are suitable when asymptotic dependence is present, i.e., the joint tail decays at the same rate as the marginal tail. However, recent environmental data applications suggest that asymptotic independence is equally important and, unfortunately, existing spatial models in this setting that are both flexible and can be fitted efficiently are scarce. Here, we propose a new spatial copula model based on the generalized hyperbolic distribution, which is a specific normal mean-variance mixture and is very popular in financial modeling. The tail properties of this distribution have been studied in the literature, but with contradictory results. It turns out that the proofs from the literature contain mistakes. We here give a corrected theoretical description of its tail dependence structure and then exploit the model to analyze a simulated dataset from the inverted Brown–Resnick process, hindcast significant wave height data in the North Sea, and wind gust data in the state of Oklahoma, USA. We demonstrate that our proposed model is flexible enough to capture the dependence structure not only in the tail but also in the bulk.

Time and Session: Thursday 16.00–17.15, CS Spatial Extremes (II)

2 Posters

Mixed moment estimator for inference on space-time extremes

Jessica Silva Lomba*, Maria Isabel Fraga Alves and Cláudia Neves

*Centro de Estatística e Aplicações, University of Lisbon

Abstract

Extreme meteorological phenomena such as heavy precipitation seem to be growing more severe and frequent due to accelerating climate change. Fewer statistical methods for tackling extreme events are more topical right now than those encompassing inference for spatio-temporal processes with a non-stationary feature. In this work, we take the semi-parametric approach to inference for heteroscedastic extremes. We look at how one can pool information from data collected at irregularly spaced locations so as to enable the mixed moment estimator of the extreme value index (Fraga Alves et al. 2009) to be used in the presence of space-time heteroscedasticity. The latter refers to a non-monotonic trend in the frequency of extremes, which does not necessarily imply a change in the dynamics of the underlying physical process generating the data. Asymptotic properties of the mixed moment estimator are developed in this non-i.i.d. setting, which accommodates spatial dependence in relation to those extreme values larger than a sufficiently high threshold. Building on the foundational works by Einmahl et al. (2021), we aim to develop further the second order approximation to both tail empirical and tail quantile processes, making explicit a crucial component for capturing the asymptotic bias in the estimation of the extreme value index, thus paving the way for wider implementation of bias reduction techniques. Application of the extended mixed moment estimator is illustrated through a large batch of reanalysis precipitation data from several locations across the UK.

Acknowledgments: J. Silva Lomba and M.I. Fraga Alves' research is supported by Fundação para a Ciência e a Tecnologia, I.P. through PhD grant SFRH/BD/130764/2017 (J.S.L.) and project UIDB/00006/2020. C. Neves gratefully acknowledges support from EPSRC-UKRI Innovation Fellowship grant EP/S001263/1.

References

- Einmahl J. H. J., Ferreira A., de Haan L., Neves C., Zhou C., 2021. Spatial dependence and space-time trend in extreme events *Annals of Statistics*. To appear.
- Fraga Alves M. I., Gomes M. I., de Haan L., Neves C., 2009. Mixed moment estimator and location invariant alternatives. *Extremes* **12**, 149–185

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Estimating bivariate return curves for non-stationary processes

Callum Barltrop

Lancaster University

Abstract

A multivariate return curve provides a logical extension to return levels in the multivariate setting. This risk measure is defined, for a given probability p , to be all values of a multivariate random variable for which the joint survival probability is equal to p . In the environmental context, such curves have a variety of applications, including safeguarding off-shore constructions, flood risk management, and structural engineering. However, environmental variables often exhibit non-stationarity and temporal dependence, making the analysis of these types of data set more challenging. Moreover, while a range of methods are available for dealing with these features in the univariate context, such problems are seldom considered in multivariate modelling approaches and if they are accounted for, it is through inclusion in models for the univariate margins rather than in higher dimensional structures. For this poster, we present novel tools for incorporating both non-stationarity and temporal dependence in the construction of a multivariate return curve. We illustrate how these features alter both the definition and uncertainty associated with this risk measure. We further demonstrate our approaches using data for a coastal location from the UK Climate Projections (UKCP18).

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Feature clustering for support identification in extreme regions

Hamid Jalalzai

INRIA

Abstract

Understanding the complex structure of multivariate extremes is a major challenge in various fields from portfolio monitoring and environmental risk management to insurance. In the framework of multivariate extreme value theory, a common characterization of extremes' dependence structure is the angular measure. It is a suitable measure to work in extreme regions as it provides meaningful insights concerning the subregions where extremes tend to concentrate their mass. The presentation focuses on a novel optimization-based approach to assess the dependence structure of extremes. This support identification scheme rewrites as estimating clusters of features which best capture the support of extremes. The dimension reduction technique we provide is applied to statistical learning tasks such as feature clustering and anomaly detection. Numerical experiments provide strong empirical evidence of the relevance of our approach.

Time and Session: Poster 14.00–16.00

A non-parametric test for comparison of extremal dependence

Sonia Alouini

EPFL

Abstract

We propose a simple and efficient test of whether two bivariate random samples have the same extremal dependence based on the Kolmogorov–Smirnov distance between non-parametric estimates of their angular measures. The sample angular measure is estimated by maximizing the empirical likelihood, thus enforcing the moment constraints while avoiding parametric assumptions. In general, the shape of the extremal regions on which the data are observed might result in an unwanted tilting of empirically estimated measures. The choice of the proper angular transformations cancels this effect and allows us to compare samples observed in different extremal regions. Simulations show that this test has better power and control of the error rate than the standard Kolmogorov–Smirnov test. Our results are used to compare the relations between wildfire risk indices and extreme windspeed in different locations across South California.

Time and Session: Poster 14.00–16.00

Models and inference for spatial extremes based on tree-based multivariate Pareto distributions

Daniela Cisneros* and Raphaël Huser

*KAUST

Abstract

Multivariate Pareto distributions have been widely used for modeling the simultaneous occurrence of extreme events over a high threshold. However, existing likelihood-based inference approaches are computationally prohibitive in high dimensions due to the need to censor observations in the bulk of the distribution. In this work, we construct mixture models for spatial extremes by exploiting the sparse conditional independence structure of multivariate Pareto distributions of Hüsler-Reiss type defined on trees. Such models have a simplified likelihood function, which can be computed more efficiently, thus opening the door to higher-dimensional extreme-value problems. We illustrate our methodology by simulation and application to the McArthur Forest Fire Danger Index (FFDI) in Australia.

Time and Session: Poster 14.00–16.00

A novel high quantile estimator for light to moderately-heavy tailed distributions with application to aviation safety

Graham Davies

University of Reading

Abstract

When required to extrapolate to very rare return levels, standard extreme value models and their typified statistical inference techniques, such as maximum likelihood, can lead to a large amount of uncertainty involving parameter estimates. In some situations, the quality of estimated values for very high quantiles rapidly erodes in the sense that estimates start to exceed plausible levels, which can reduce end-user confidence in results from extreme value theory. This is of particular relevance when dealing with moderately-heavy tailed distributions, for example those arising in connection to the modelling of extreme rainfall. We introduce a semi-parametric estimator designed specifically for harnessing high quantiles that emanate from distributions belonging to the domain of attraction of an extreme value distribution with associated positive but near zero extreme value index. For an underlying distribution with finite right endpoint, irrespective of which domain of attraction it is found, the proposed estimator resolves to the general endpoint estimator (Fraga Alves and Neves, *Statistica Sinica*, 2014). We show consistency, develop asymptotic distribution and go on to demonstrate good finite-sample performance of this proposed estimator. Finally, application of the high quantile estimator is illustrated in relation to taxiway deviations for Boeing 747 aircraft measured at two international airports.

Time and Session: Poster 14.00–16.00

Flexible modeling of multivariate spatial extremes

Yan Gong* and Raphaël Huser

KAUST*

Abstract

We propose a novel multi-factor copula model for multivariate spatial extremes, which is designed to capture different combinations of marginal and cross-extremal dependence structures within and across different spatial random fields. Our model can capture all possible distinct combinations of extremal dependence structures within each individual spatial process, while allowing for flexible cross-process extremal dependence structures, for both the upper and lower tails. The model may be seen as a multi-factor copula model, and we perform Bayesian inference using a Markov chain Monte Carlo algorithm based on carefully designed block proposals with an adaptive step size. We apply our model to study the lower and upper extremal dependence structures of daily maximum Air Temperature (TMAX) and daily minimum Air Temperature (TMIN) from a region of South-Eastern US.

Time and Session: Poster 14.00–16.00

Simulating spatially realistic extreme temperature events in Ireland.

D Healy

Hamilton Institute, Maynooth University

Abstract

We investigate a method to simulate spatially realistic extreme temperature events over the island of Ireland. Magnitudes of observed spatial events are measured with a carefully chosen cost-function, the events whose cost exceeds some high threshold are known to converge to an r -Pareto process. Our analysis is based on synoptic data observations courtesy of Met Éireann and MIDAS. We generate a rich database of 25,000 extreme temperature events over Ireland and show how one could use these simulations to answer different questions about the probabilities of extreme events of interest. For illustrative purposes we calculate the probability and return period of experiencing temperatures above 30 degrees at Dublin, Belfast and Cork city simultaneously.

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

A new class of estimators for residual dependence index and its application on tropical rainfall

Jennifer Israelsson

University of Reading

Abstract

The simultaneous occurrence of extremes from different sources or locations is of great interest in many areas such as wind-energy production and flood predictions. The extent and magnitude of a flooding event can greatly increase if the extreme rain event affects several locations around the same river basin. Through multivariate extreme value analysis, we can estimate the probability of two variables or locations begin extreme at the same time by determining their asymptotic dependence structure. This work focuses on a new class of estimators for residual dependence index, also known as coefficient of tail dependence, which includes the widely used Hill estimator for the index of regular variation. The new class of estimators builds on the *power mean-of-order-p* estimator, by introducing a second power parameter to allow for greater flexibility in the estimator parametrisation. Asymptotic normality is devised, alongside a reduced bias version and its asymptotic distribution. The finite sample performance is evaluated via simulations from several bivariate copula models with varying residual dependence index. We will standardise marginal distributions to unit Fréchet with an appropriate location-shift, which we will show improves the bias reduction.

Finally the estimator performance on real world data will be presented using daily tropical rainfall data collected over Ghana, using station pairs located within close proximity or far apart.

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Autoregressive conditional accelerated Fréchet model for decoupling systemic risk into endogenous and exogenous competing risks

Jingyu Ji, Deyuan Li and Zhengjun Zhang

Abstract

Identifying systemic risk patterns in social, political, economic, financial, market, regional, global, environmental, transportation, epidemiological, material, chemical, and physical systems and their impacts is the key to risk management. This work integrates the newly introduced new extreme value theory for maxima of maxima and a new time series benchmark model of autoregressive conditional Fréchet (AcF) for modeling systemic risk into a new autoregressive conditional accelerated Fréchet (AcAF) model for decoupling systemic risk into endogenous and exogenous competing risks. In the paper, the focus is paid on market risk and systemic financial risk. Nevertheless, the AcAF model can be applied to all systems above and beyond. The AcAF model and its resulting competing risks provide clear endogenous and exogenous competing risk patterns of market risks and reveal the causes of the financial crisis, which was not detected using the existing models. The probabilistic properties of stationarity and ergodicity of the AcAF model are established. Statistical inference is developed through conditional maximum likelihood estimation. The consistency and asymptotic normality of the estimators are derived. Simulated numerical examples are used to demonstrate the efficiency of the proposed estimators. Empirical studies of time series of maxima of maxima of stock returns in SP 500, intra-day returns of high-frequency trading stocks, and intra-day returns of cryptocurrency trading show the superior performance of the proposed model in terms of the identified risk patterns being informative with greater interpretability, enhancing the understanding of the systemic risks of a market and their causes, and making better risk management possible.

Time and Session: Poster 14.00–16.00

Assessing the skill of a max-stable process model for modeling extreme rainfall events for different seasons in Germany

Oscar E. Jurado* and Henning W. Rust

*Freie Universität Berlin

Abstract

When modeling extreme rainfall events in a spatial setting, accounting for the dependence that exists between neighboring rain gauges allows the model to use the information more efficiently, by *borrowing strength across locations*. A commonly used model for this kind of problem involves a mixture of trend surface models (a GLM using spatial covariates) for the overall marginal estimation and a max-stable process for capturing the residual spatial dependence.

This residual spatial dependence arises from the fact that individual extreme events can affect multiple rain gauges simultaneously, inducing a local spatial dependence for nearby stations. This spatial effect is, in turn, heavily dependent on the physical properties of the data-generating process (e.g., local convective storms in Summer are spatially very different from frontal precipitation in Winter).

In this study, we aim to better understand how modeling the spatial residual dependence impacts the skill of an extreme rainfall model that predicts return levels for rare events (e.g., the 100-year return level). We do this by using monthly block maxima to model extreme rainfall for two different seasons in Germany with differing storm-generating mechanisms: winter and summer.

We compare the predictive performance of different residual spatial dependence models for each season using the cross-validated quantile score, a measure of out-of-sample accuracy. All the models are Brown-Resnick max-stable process, using response surfaces with spatial covariates.

Furthermore, we compare the skill of our residual dependence model with a trend surface model that assumes conditional independence between sites, to assess how much skill is gained when accounting for the residual spatial dependence.

Time and Session: Poster 14.00–16.00

Appropriate statistical methods to assess COVID-19 data

Hulya Kocyigit

University of Georgia

Abstract

Since December 8, 2019, a mystery epidemic infection with a novel coronavirus recognized as COVID-19 has circulated quickly from Wuhan to all world. Scientists have made many evaluations using various statistical approaches throughout the past year to understand the spread of COVID-19. Infectious diseases are influenced by a variety of influences, including weather. This paper aims to recommend suitable methods by testing and comparing Pearson, Spearman, and Kendall tests based on COVID-19 and weather data sets. Thus, this study emphasizes the importance of selecting correct statistical methods to assess relevant results based on the COVID-19. Besides, this paper supply assessment of the COVID-19 case in Turkey in various circumstances, i.e., evaluation relies on the different regions and months of weather conditions. The findings of the study could help authorities and researchers devise concrete measures.

Time and Session: Poster 14.00–16.00

Estimating an extreme Bayesian network via scalings

Mario Krali

EPFL

Abstract

A recursive max-linear vector models causal dependence between its components by expressing each node variable as a max-linear function of its parental nodes in a directed acyclic graph and some exogenous innovation. Motivated by extreme value theory, innovations are assumed to have regularly varying distribution tails. We propose a scaling technique in order to determine a causal order of the node variables. All dependence parameters are then estimated from the estimated scalings. Furthermore, we prove asymptotic normality of the estimated scalings and dependence parameters based on asymptotic normality of the empirical spectral measure. Finally, we apply our structure learning and estimation algorithm to financial data and food dietary interview data.

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Spatial clustering of heavy precipitation over Switzerland

Le Gall Philomène* and Philippe Naveau

*Université Grenoble Alpes

Abstract

Rainfall are subject to local orography features and their intensities can be highly variable. In this context, identifying climatically coherent regions can greatly help interpreting rainfall patterns and improve the inference of return levels. In practice, partitioning a region of interest into homogeneous sub-regions is a delicate statistical task, especially in regards to modeling heavy rainfall features.

In this work, our main goal is to propose and study a fast and efficient clustering algorithm. Compared to classical regional frequency analysis techniques, a key aspect is that our algorithm does not rely on the a priori choice of covariates. In addition, it takes into account dependence between sites. The proposed numerical scheme is only based on the precipitation dataset at hand, considering bloc maxima. In terms of inference, our approach builds on a normalized version of the F-madogram.

While being in compliance with extreme value theory, we do not impose a parametric form on rainfall distributions. By construction, our clustering method combines the scale invariance principle of any classical regional frequency analysis with the spatial dependence.

The performance of our clustering algorithm is assessed on an experimental design based on the multivariate generalized extreme value distribution. Sensitivity to the number of clusters is carefully analyzed.

We apply our clustering algorithm on Switzerland weekly maxima precipitation measured at 191 sites. The found homogeneous regions are consistent with local orography. This work was supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), the Swiss Federal Office for Environment (FOEN), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the Federal Office for Civil Protection (FOCP), and the Federal Office of Meteorology and Climatology, MeteoSwiss, through the project EXAR ("Evaluation of extreme Flooding Events within the Aare-Rhine hydrological system in Switzerland").

Part of this work was supported by the DAMOCLES-COST-ACTION on compound events and the French national program: FRAISE-LEFE/INSU, ANR-Melody and ANR-TRex.

Time and Session: Poster 14.00–16.00

Modelling extreme sub-daily precipitation with the blended generalized extreme value distribution

Silius M. Vandeskog

Norwegian University of Science and Technology

Abstract

Short-term extreme precipitation can cause flash floods, large economic losses and immense destruction of infrastructure. In order to prepare for future extreme precipitation, a Bayesian hierarchical model is applied for estimating return levels for the yearly maxima of sub-daily precipitation in Norway.

A modified version of the generalized extreme value (GEV) distribution, called the blended GEV (bGEV) distribution, is used as the model for yearly maxima of sub-daily precipitation. Inference with the GEV distribution is known to be difficult, partly because its support depends on its parameters. The bGEV distribution has the right tail of a Fréchet distribution and the left tail of a Gumbel distribution, resulting in a constant support that allows for more stable inference. Fast inference is performed using integrated nested Laplace approximations (INLA).

We propose a new model for block maxima that borrows strength from the peaks over threshold methodology by linking the scale parameter of a block maximum to the standard deviation of observations larger than some threshold. The new model is found to outperform the standard block maxima model when fitted to the yearly maxima of sub-daily precipitation in Norway. Evaluation is performed with the threshold weighted continuous ranked probability score (twCRPS), where the weight function only focuses on large quantiles.

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

The value of regularisation and model robustness in the context of climate extremes

Joel Zeder

Institute for Atmospheric and Climate Science, ETH Zurich

Abstract

Single-model initial condition large ensembles—a multitude of climate simulations based on the same model with a minor perturbation in the initial conditions—provide novel opportunities to study the physical drivers of large-scale climate extremes. The severity and conditional probability of extreme events, such as heatwaves, can be estimated with a non-stationary generalized extreme value distribution GEV that accounts for variability in the climate system.

In this work we explore the potential of applying established and recently developed methods of statistical learning theory to address the challenge of building a robust model and selecting relevant covariates from (often highly correlated) climatic fields. Based on millennial simulations of stationary and transient climate we use regularisation to obtain an objective subset of explanatory physical predictors (ranging from large-scale thermodynamic to local short-term dynamic climate variables) determining model parameters, and the application of an anchor regression approach ([Rothenhäusler et al. 2021](#)) ensures model stability also under decadal internal climate variability.

References

Rothenhäusler, D., Meinshausen, N., Bühlmann, P., Peters, J., 2021. Anchor regression: heterogeneous data meet causality. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **83**, 215-246

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Unsupervised threshold selection in POT modelling: a comparative study

Baitshephi Mashabe*, Georgios Aivaliotis, Leonid Bogachev and Lanpeng Ji

*University of Leeds

Abstract

Developing efficient methods for choosing an adequate threshold in a peaks-over-threshold (POT) modelling of extreme values is of paramount importance for the faithfulness of the generalized Pareto distribution (GPD) approximation. The suitable threshold should be high enough for the GPD to be applicable but not too high in order to retain sufficiently many threshold exceedances in the data. The problem is to have evidence-based methods to decide how high is *high enough*. Furthermore, automated (unsupervised) methods may be needed if multiple data samples must be processed simultaneously (e.g., for environmental observations on a spatial grid) or if online processing of streamed data is required.

Traditional methods use graphical diagnostics by leveraging threshold stability of GPD (i.e., the linear dependence of the scale parameter on the threshold), that is, by looking for a linear (straight line) part in the mean exceedances plot (Coles, 2001; Northrop Coleman, 2014). By design, such approaches are subjective and not easily automated (Scarrott MacDonald, 2012). A more objective method allowing for an automated (unsupervised) implementation was proposed by Thompson et al. (2009). The idea is to estimate the scale and shape parameters via the maximum likelihood (ML) method on a grid of thresholds within a trial range and to test for their joint normality, using that the distribution of such estimates should be approximately normal as long as the chosen thresholds comply with the GPD regime. If normality is rejected, then the trial range is modified until acceptance. More specifically, by taking suitable linear combinations of the ML-estimates at neighbouring thresholds, the hypothetical normal distribution will have zero mean. However, tests for normality used in Thompson et al. (2009) tacitly assume that the entries are (at least approximately) mutually independent and have the same standard deviation, which is not quite correct due to the nested nature of estimates at increasing thresholds.

In the present study, we propose to modify the above method by focusing on the shape parameter of the GPD, which should stay constant under a proper choice of the threshold range. The potential advantage is that only this parameter needs to be estimated, although heteroscedasticity and lack of full independence may still be an issue. We compare both approaches in a simulation study by generating multiple samples from a given parent distribution and then testing for normality (e.g., using the Shapiro–Wilk test), also monitoring if the rejection rate is consistent with the chosen significance level. At the next stage, we validate the proposed *optimal* range of thresholds by testing the goodness-of-fit with the projected GPD (e.g., by the Anderson–Darling test).

In addition, to attempt to mitigate the lack of independence, we have also piloted an alternative approach by using a simple autoregressive model and then testing for normality of residuals as described above. We have created an automated code in R to implement each of the four models' procedure above to identify the suitable range of thresholds. Our preliminary results suggest that the four methods work reasonably well and fairly consistent across by producing similar threshold ranges. However, we have observed that the new models based

only on the shape parameter are close to one another but produce marginally but consistently broader ranges than the other two models that involve both scale and shape parameter. This may be beneficial in practical applications due to faster estimation and retaining more data for the POT analysis.

Time and Session: Poster blitz 13.00–14.00, Poster 14.00–16.00

Author Index

- Alexander Henzi, 3
 Alina Kumukova, 3
 Ana Ferreira, 4
 Anders Rønn-Nielsen, 5
 Andrew Thomas, 6
 Anish Ghosh, 7
 Anita Behme, 8
 Anna Ben-Hamou, 8
 Anne-Laure Fougères, 9
 Antoine Usseglio-Carleve, 10
 Anwar Alabdulathem, 11
 Armelle Guillou, 12
 Arnab Hazra, 12
 Axel Bücher, 13
 Baitshephi Mashabe, 149
 Benjamin Bobbia, 14
 Benjamin Shaby, 15
 Callum Bartrop, 15, 135
 Chen Zhou, 16
 Christian Hirsch, 17
 Christian Robert, 18
 Christian Rohrbeck, 19
 Clément Dombry, 19
 Colin Raymond, 20
 Corina Birghila, 21
 D Healy, 140
 Daniela Cisneros, 137
 David Brayshaw, 22
 David Deuber, 22
 Domingos Silva, 23
 Edossa Merga Terefe, 23
 Ekaterina Morozova, 24
 Eleanor D’Arcy, 26
 Emilie Chautru, 26
 Emilien Joly, 27
 Emma Simpson, 28
 Emmanuel Gobet, 28
 Erich Fischer, 29
 Fabien Baeriswyl, 30
 Felix Cheysson, 31
 Florian Brück, 32
 Frank Röttger, 33
 Frederico Caeiro, 34
 Georgia Papadogeorgou, 35
 Gilles Stupfler, 36
 Gloria Buriticá, 36
 Gloria Gheno, 37
 Graham Davies, 138
 Hairuo Yang, 37
 Hamid Jalalzai, 135
 Hayley Fowler, 38
 Helga Kristin Olafsdottir, 39
 Holger Drees, 40
 Holger Rootzén, 41
 Hrvoje Planinić, 42
 Hulya Kocyigit, 143
 Igor Rodionov, 43
 Jana Ulrich, 44
 Jennifer Israelsson, 141
 Jenny Wadsworth, 45
 Jeongjin Lee, 45
 Jessica Silva Lomba, 134
 Jethro Browell, 46
 Jevgenijs Ivanovs, 47
 Jingyu Ji, 141
 Jochem Oorschot, 48
 Joel Zeder, 147
 Johannes Buck, 49
 Johannes Heiny, 50
 John Einmahl, 50
 John Nolan, 51
 Jonas Brehmer, 52
 Jonas Peters, 53
 Jonathan El Methni, 54
 Jonathan Jalbert, 54
 Jonathan Koh, 55
 Jordan Richards, 56
 Jorge Castillo-Mateo, 57
 Juliette Legrand, 58
 Juraj Bodík, 59
 Karla Vianey Ramirez, 60
 Konrad Krystecki, 61
 Krzysztof Bisewski, 61
 Krzysztof Debicki, 62
 Krzysztof Kępczyński, 62
 Lanpeng Ji, 63
 Le Gall Philomène, 145
 Lekha Patel, 64
 Likun Zhang, 65
 Linda Mhalla, 66
 Liujun Chen, 67
 Luca Trapin, 68
 Léo Belzile, 69
 Lígia Henriques-Rodrigues, 70
 Manuel Hentschel, 70
 Manuela Brunner, 71
 Marco Oesting, 72
 Maria Ivette Gomes, 73
 Marie Ekstrom, 74
 Marine Demangeot, 75
 Mario Krali, 144
 Mark Risser, 77
 Martin Bladt, 77
 Martina Favero, 78
 Matheus Guerrero, 79
 Matthias Schulte, 80
 Matthieu Lerasle, 81
 Meryem Bousebata, 81
 Michaël Allouche, 82
 Michaël Lalancette, 83
 Moritz Otto, 84
 Natalia Markovich, 85
 Nehali Mhatre, 86
 Ngoc Tran, 87
 Nguyen Ho, 88
 Nicholas Beck, 89
 Nicola Gnecco, 90

Nicolas Meyer, 90	Sayed H. Kadhem, 104	Thibault Modeste, 120
Olivier Pasche, 91	Sebastian Engelke, 105	Thomas Opitz, 121
Olivier Wintenberger, 92	Sebastian Neblung, 106	Tiandong Wang, 122
Oscar E. Jurado, 142	Seth Westra, 107	Tomasz Rolski, 123
Patrice Bertail, 93	Shrijita Bhattacharya, 108	Vicky Fasen-Hartmann, 124
Peng Zhong, 93	Shuyang Bai, 109	Vladas Pipiras, 125
Petra Žugec, 94	Silius M. Vandeskog, 110, 146	Vladimir Fomichov, 126
Pierre Nyquist, 95	Simone Padoan, 110	Wonkeun Choi, 126
Pim van der Hoorn, 95	Siyang Tao, 111	Xuan Leng, 127
Piotr Dyszewski, 96	Sonia Alouini, 136	Yan Gong, 139
Rafail Kartsioukas, 97	Stan Tendijck, 112	Yizao Wang, 128
Rafal Kulik, 98	Stanislav Volgushev, 113	Yun Li, 129
Raphaël de Fondeville, 98	Stefan Aulbach, 114	Yuri Goegebeur, 130
Raphaël Huser, 99	Stefano Rizzelli, 115	Zaoli Chen, 131
Richard Smith, 100	Stefka Asenova, 116	Zhongwei Zhang, 131
Rishikesh Yadav, 101	Su-Fen Yang, 117	
Ross Maller, 102	Svenja Szemkus, 118	
Saat Mubarrok, 103	Takashi Owada, 119	

