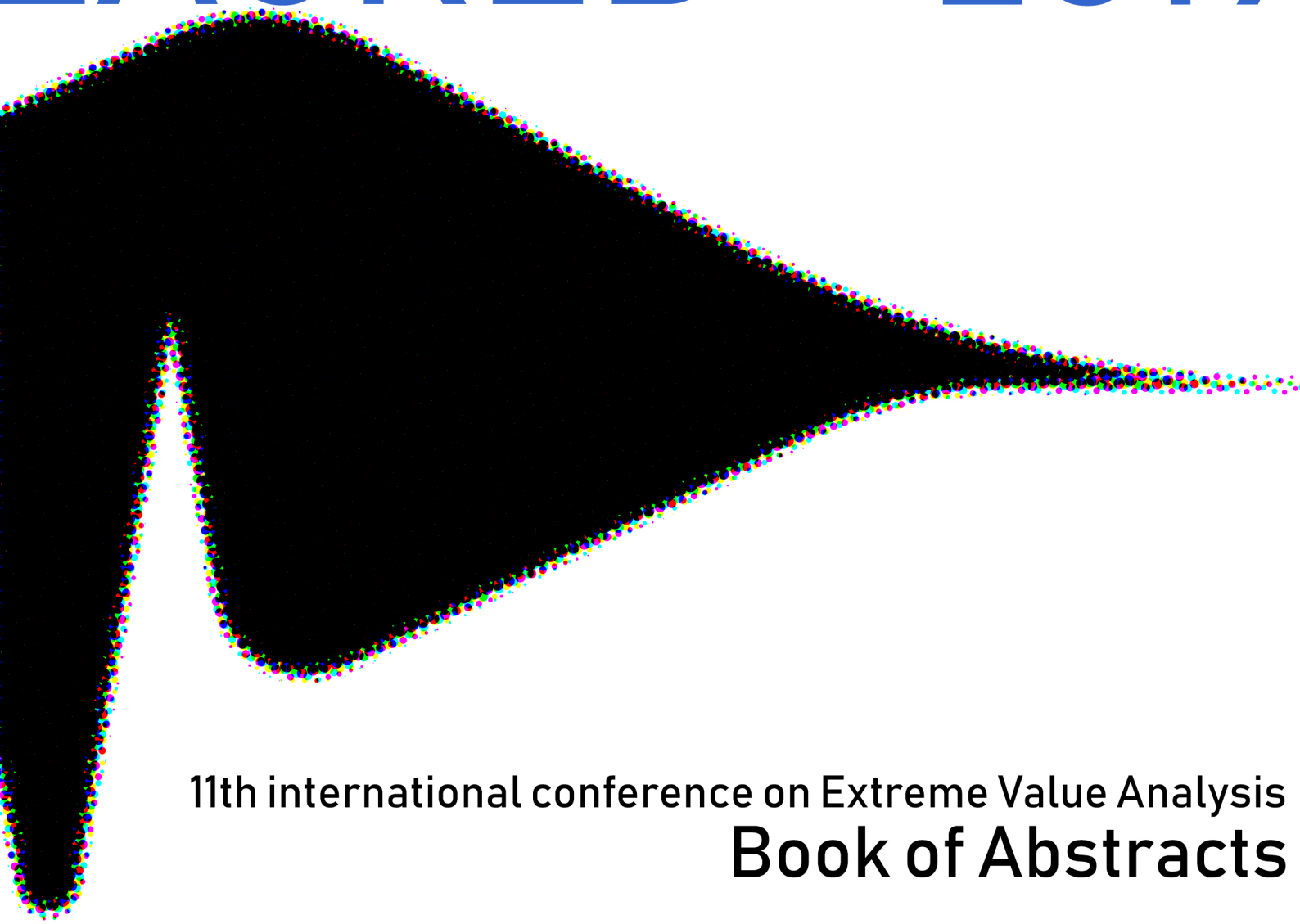
EVA

ZAGREB 2019



11th international conference on Extreme Value Analysis

Book of Abstracts

Contents

Invited and contributed talks

Improved estimation of the extreme value index using related variables

Hanan Ahmed[1], John Einmhal[2]
Tilburg University, Netherlands[1,2]
hegahmed@uvt.nl[1], J.H.J.Einmahl@uvt.nl[2]

Heavy tailed phenomena are naturally analyzed by extreme value statistics. A crucial step in such an analysis is the estimation of the extreme value index, which describes the tail heaviness of the underlying probability distribution. We consider the situation where we have next to the n observations of interest another $n + m$ observations of one or more related variables, like, e.g., financial losses due to earthquakes and the related amounts of energy released, for a longer period than that of the losses. Based on such a data set, we present an adapted version of the Hill estimator that shows greatly improved behavior and we establish the asymptotic normality of this estimator. For this adaptation the tail dependence between the variable of interest and the related variable(s) plays an important role. A simulation study confirms the substantially improved performance of our adapted estimator relative to the Hill estimator. We also present an application to the aforementioned earthquake losses.

Session, time and room:
Tail index estimation
Monday, 10:30 – 10:50 (A101)

A test for equality of expected proportional shortfall

Maximilian Aigner[1], Valérie Chavez-Demoulin[2], Armelle Guillou[3]
Faculté des Hautes Etudes Commerciales, Université de Lausanne, Switzerland[1,2], Université de Strasbourg, France[3]
maximilian.aigner@unil.ch[1], valerie.chavez@unil.ch[2],
armelle.guillou@math.unistra.fr[3]

We construct a test for equality of expected proportional shortfall (EPS) in the context of multivariate β-mixing time series. Using an unbiased estimator of the tail index γ, we propose an estimator for the EPS and derive convergence results. The corresponding test statistic is self-normalised to obtain an asymptotic pivot distribution.

Unlike other risk measures, the EPS allows the comparison of risk across series of different types or units. On the other hand, compared to the tail index which also shares this property, it is arguably a more interpretable measure of risk for practitioners. In a simulation study, we assess the power of the test for equal EPS and we also compare tests based on tail indices and value at risk for our unbiased γ estimator with those based on the Hill estimator. We illustrate the usefulness of our test on an application to U. S. Transportation Security Administration (TSA) insurance claims.

Session, time and room:
Testing and inference
Thursday, 11:15 – 11:35 (A102)

Multivariate models connected with random sums and maxima

Marek Arendarczyk[1]
Mathematical Institute, University of Wroclaw, Poland[1]
marendar@math.uni.wroc.pl[1]

We present recent results concerning a stochastic model for the sum X and the maximum Y of dependent components with Pareto II distributions. Our results include explicit forms of the probability density and cumulative distribution functions of (X, Y), as well as marginal and conditional distributions, moments and related parameters, parameter estimation, and stochastic representations. We also derive bivariate mixed tail conditional expectations, $\mathbb{E}(X|Y > y)$ and $\mathbb{E}(Y|X > x)$, which provide useful and practical risk measures in addition to the univariate TCEs recently obtained for this model. Finally, we present real data examples from finance, illustrating the applicability of the model.

The talk is based on joint work with Tomasz J. Kozubowski and Anna K. Panorska (University of Nevada, Reno).

References

[1] Arendarczyk, M., Kozubowski, T.J., Panorska, A.K. The joint distribution of the sum and maximum of dependent Pareto risks. *J. Multivariate Anal.*, 167, 136 – 156, 2018.

[2] Arendarczyk, M., Kozubowski, T.J., Panorska, A.K. The joint distribution of the sum and the maximum of heterogeneous exponential random variables. *Statist. Probab. Lett.*, 139, 10 – 19, 2018.

[3] Arendarczyk, M., Kozubowski, T.J., Panorska, A.K. A bivariate distribution with Lomax and geometric margins. *J. Korean Statist. Soc.*, 47(4), 405 – 422, 2018.

Session, time and room:
Conditional extremes
Thursday, 14:00 – 14:20 (A101)

On aggregation of subcritical Galton–Watson branching processes with regularly varying immigration

Mátyás Barczy[1], Fanni K. Nedényi[2], Gyula Pap[3]
University of Szeged, Hungary[1,2,3]
barczy@math.u-szeged.hu[1], nfanni@math.u-szeged.hu[2], papgy@math.u-szeged.hu[3]

Let $(X_k)_{k\geq 0}$ be a strongly stationary subcritical Galton–Watson branching process with regularly varying immigration:

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{k,j} + \varepsilon_k, \qquad k \geq 1,$$

where

- $\{X_0, \xi_{k,j}, \varepsilon_k : k, j \geq 1\}$ are independent non-negative integer-valued random variables,
- $\{\xi_{k,j} : k, j \geq 1\}$ and $\{\varepsilon_k : k \geq 1\}$ consist of identically distributed random variables, respectively,
- $\mathrm{E}(\xi_{1,1}) \in [0, 1)$ (subcritical case), and ε_1 is regularly varying with index $\alpha \in (0, 2)$, i.e.,
$$\lim_{x\to\infty} \frac{\mathrm{P}(\varepsilon_1 > qx)}{\mathrm{P}(\varepsilon_1 > x)} = q^{-\alpha} \qquad \forall\, q > 0,$$
- in case of $\alpha \in [1, 2)$, we suppose additionally that $\mathrm{E}(\xi_{1,1}^2) < \infty$,
- the law of X_0 is chosen as the unique stationary distribution of $(X_k)_{k\geq 0}$.

Due to Basrak et al. [1], the strongly stationary process $(X_k)_{k\geq 0}$ is jointly regularly varying with index α (i.e., all its finite dimensional distributions are regularly varying with index α).

Limit theorems will be presented for the aggregated stochastic process $\sum_{j=1}^{N} \sum_{k=1}^{\lfloor nt \rfloor} X_k^{(j)}$, $t \geq 0$, after appropriate centering and scaling, where $(X_k^{(j)})_{k\geq 0}$, $j \geq 1$, are independent copies of $(X_k)_{k\geq 0}$. We will provide limit theorems where N and n converge to infinity in an iterated manner, meaning that first N, and then n converges to infinity, or vica versa. In both cases, the limit process turns out to be the same α-stable process if $\alpha \in (0, 1) \cup (1, 2)$, and the same deterministic line with slope 1 if $\alpha = 1$.

References

[1] Bojan Basrak, Rafał Kulik, and Zbigniew Palmowski, Heavy-tailed branching process with immigration, *Stochastic Models*, 29(4), 413–434, 2013

Session, time and room:
Limit theorems
Monday, 14:00 – 14:20 (A101)

Inference for extreme values under threshold-based stopping rules

Anna Maria Barlow[1], Chris Sherlock[2], Jonathan Tawn[3]
Lancaster University, United Kingdom[1,2,3]
a.m.barlow@lancaster.ac.uk[1], c.sherlock@lancaster.ac.uk[2], j.tawn@lancaster.ac.uk[3]

There is a propensity for an extreme value analyses to be conducted as a consequence of the occurrence of a large flooding event. This timing of the analysis introduces bias and poor coverage probabilities into the associated risk assessments and leads subsequently to inefficient flood protection schemes. We explore these problems through studying stochastic stopping criteria and propose new likelihood-based inferences that mitigate against these difficulties. Our methods are illustrated through the analysis of the river Lune, following it experiencing the UK's largest ever measured flow event in 2015. We show that without accounting for this stopping feature there would be substantial over-design in response to the event.

Session, time and room:
Flood risks
Thursday, 11:15 – 11:35 (003)

Predicting extreme surges from sparse data using a copula-based hierarchical Bayesian spatial model

Nicholas Beck[1], Christian Genest[2], Jonathan Jalbert[3], Mélina Mailhot[4]
McGill University, Canada[1,2], Polytechnique Montréal, Canada[3], Concordia University, Canada[4]
nicholas.beck@mail.mcgill.ca[1], christian.genest@mcgill.ca[2],
jonathan.jalbert@polymtl.ca[3], melina.mailhot@concordia.ca[4]

In this presentation, a hierarchical Bayesian model is proposed to quantify the magnitude of extreme surges on the Atlantic Coast of Canada with limited data. At the data level, generalized extreme-value distributions (GEVs) are fitted to surges derived from water levels measured at 21 buoys along the coast. Defining $\boldsymbol{Z} = (Z_1, \ldots, Z_{21})$ to be the vector of observed annual maximum surges across the domain, we set, for location $\boldsymbol{s}_i$,

$$Z_i \sim \text{GEV}\left(\mu_i, \sigma_i, \xi\right).$$

In such setups, it is often assumed that the GEVs will be conditionally independent given the parameters. However, as discussed by Tawn et al. [1], the consideration of events at a site-by-site basis, independent of the surrounding area, is overly simplistic when impact across the entire domain is of interest, as could be the case in insurance applications. To this end, a Student's t-copula is introduced at the data level to connect the marginal distributions. This will allow for inference to be done across the entire coast, factoring in the interactions between neighboring buoys. Additionally, the t-copula is asymptotically dependent and will be useful for handling missing observations in our data set.

At the process level, the location and log-scale parameters of the GEVs are linked together through Gaussian fields whose mean and variance are driven by atmospheric sea-level pressure and the distance between stations, respectively. For location parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{21})$ and log-scale parameters $\boldsymbol{\Phi} = (\ln(\sigma_1), \ldots, \ln(\sigma_{21}))$, we set

$$\boldsymbol{\mu} \sim \mathcal{N}_{21}(\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\mu}}, \tau^2_{\boldsymbol{\mu}}\, \Sigma_{\boldsymbol{\mu}}) \quad \text{and} \quad \boldsymbol{\Phi} \sim \mathcal{N}_{21}(\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\Phi}}, \tau^2_{\boldsymbol{\Phi}}\, \Sigma_{\boldsymbol{\Phi}}),$$

where $\mathbf{X}$ contains the pressure information, $\Sigma_{\boldsymbol{\mu}}$ and $\Sigma_{\boldsymbol{\Phi}}$ are the distance driven correlation matrices and $\tau^2_{\boldsymbol{\mu}}$ and $\tau^2_{\boldsymbol{\Phi}}$ represent the variances of $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$, respectively. Inclusion of this latent spatial process will allow for information sharing across the original stations, strengthening fit. With this we were able to easily fit GEVs at nearly 1400 unmonitored locations along the coast, expanding the possibilities for inference.

In this sense, the latent Gaussian fields can be seen to quantify the relationships between the general behavior of the individual buoys while the copula quantifies the relationship between individual events. An analogy for this distinction would be the difference between climate and weather.

Model fit is done using a Gibbs sampling scheme with a Metropolis–Hastings accept/reject step. Several models are compared using Watanabe–Akaike information criterion. Finally, it is shown how the extreme surges derived from the model can be combined with the tidal process to predict potentially catastrophic water levels. With this information, further exploration can be done on the risks of flooding across this large domain.

References

[1] Tawn, J.A., Shooter, R., Towe, R. and Lamb, R. Modelling spatial extreme events with environmental applications. *Spatial Statistics*, **28**, 39–58. (2018)

Session, time and room:
Best student paper I
Monday, 16:15 – 16:35 (A002)

Handling censoring mechanisms is large claim modelling

Jan Beirlant[1]
KU Leuven, Belgium[1]
jan.beirlant@kuleuven.be[1]

In actuarial practice censoring mechanisms can appear in different ways.

In long-tailed insurance business claim development can take extremely long such that the claim sizes of recent large claims are often not known. The percentage of censoring can then be quite high leading to challenging extreme value problems. Actuarial experts however build reserves based on the specificity of each open claim which can be used to improve the estimation based on the available payment information.

On the other hand, claim payments are also influenced by the management of an insurance such that for claims above a certain threshold a more rigorous treatment results in tail models for the largest payments which are lighter than the models for medium size claims which often exhibit Pareto behaviour. The observed payments can then be modelled through a random right censoring mechanism where a heavy tail is tempered by a lighter tail.

We review some recently developed estimation methods for those circumstances and present some practical examples.

Session, time and room:

IS - Risk analysis in insurance (Peng)
Friday, 11:15 – 11:45 (A002)

Likelihood inference for univariate extremes: higher-order asymptotics

Léo Belzile[1]
EPFL, Switzerland[1]
leo.belzile@epfl.ch[1]

We study the frequentist properties of coverage intervals derived using a tangent exponential model approximation to the profile likelihood for generalized extreme value and generalized Pareto distributions. A simulation study conducted using domain of attraction conditions shows good properties of higher order approximations and lower error rate than profile-based confidence intervals. We consider the problem of estimating the limit for human lifetime based on survival of Italian supercentenarians.

Session, time and room:
Tail inference
Thursday, 9:00 – 9:20 (A101)

Extremes of branching random walk

Ayan Bhattacharya[1]
CWI, Amsterdam, Netherlands[1]
ayanbhattacharya.isi@gmail.com[1]

We consider discrete time branching random walk on real line where the displacements have regularly varying tail. Using the one large jump asymptotics, we derive large deviation for the extremal processes associated to the suitably scaled positions of particles in the n-th generation where the genealogical tree satisfies Kesten-Stigum condition. The large deviation limiting measure in this case is identified in terms of the cluster Poisson point process obtained in the underlying weak limit of the point processes. As a consequence of this, we derive large deviation for the rightmost particle in the nth generation giving the heavy-tailed analogue of recent work by Gantert and Höfelsauer(2018).

Session, time and room:
IS - Extremes of branching walks and free fields (Roy)
Monday, 10:30 – 11:00 (A001)

General regular variation and extremes

Nicholas Bingham[1]
Imperial College London, United Kingdom[1]
n.bingham@imperial.ac.uk[1]

The talk is a summary, with particular reference to extremes, of N. H. Bingham and A. J. Ostaszewski, General regular variation, Popa groups and quantifier weakening, arXiv:1901.05996. As with earlier aspects of regular variation (RV), much rests on functional equations (FE) (the Cauchy FE for Karamata (K) and Bojanic-Karamata/de Haan (BKdH) RV, the *Gołąb-Schinzel* FE for Beurling (B) and general (G) RV). Contents: 1. Karamata RV. 2. Bojanic-Karamata/de Haan RV. 3. Beurling RV. 4. Popa groups: Algebra. 5. Popa groups: Topology. 6. General RV. 7. Beurling's Tauberian theorem. 8. Extremes.

Session, time and room:
IS ST - Regular variation: history, ideas and people (Basrak)
Friday, 11:50 – 12:20 (A001)

Distributionally robust bounds for tail indices

Maximilian Aigner[1], Corina Birghila[2], Sebastian Engelke[3]
University of Lausanne, Switzerland[1], University of Vienna, Austria[2], University of Geneva, Switzerland[3]
maximilian.aigner@unil.ch[1], corina.birghila@univie.ac.at[2], sebastian.engelke@unige.ch[3]

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 behavior over all models within some neighborhood of the reference model. The choice of 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 Csiszár 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.

Session, time and room:
Tail inference
Thursday, 9:50 – 10:10 (A101)

Coupling method for extreme quantile regression

Benjamin Bobbia[1], Clément Dombry[2], Davit Varron[3]
Université de Bourgogne-Franche-Comté, Laboratoire de mathématiques de Besançon, France[1,2,3]
benjamin.bobbia@univ-fcomte.fr[1], dombry.clement@univ-fcomte.fr[2],
davit.varron@univ-fcomte.fr[3]

We revisit the classical extreme value theory with the help of tools from optimal transport theory. This approach relies on coupling argument between the distribution of exceedances above high threshold and limit generalized Pareto distribution. We use the Wasserstein distance to metricize the convergence in distribution of random measures. This methodology is used to prove the asymptotic normality of several estimators of the extremes values index.

Our method allows us to study a general framework for extreme quantile regression: Take $Y \in \mathbb{R}$ and $X \in \mathbb{R}^d$ and a sample of i.i.d copies of (X, Y). We are interested in estimation of the conditional quantile of Y given $X = x$ of order $1 - \alpha$ for a very small $\alpha > 0$.

We consider the proportional tail model where Y has an heavy tail $\bar{F}$ with extreme value index $\gamma > 0$ and the conditional tails $\bar{F}_x$ are asymptotically equivalent to $\sigma(x)\bar{F}$. We propose estimators for γ and the integral of σ and prove their consistency and asymptotic normality. This work is strongly connected to the statistics of heteroscedastic extremes developed by Einmahl, de Haan and Zhou (2016).

Session, time and room:
Conditional extremes
Thursday, 15:15 – 15:35 (A101)

Why scoring functions cannot assess tail properties

Jonas Brehmer[1]
University of Mannheim, Germany[1]
jbrehmer@mail.uni-mannheim.de[1]

A much debated question in forecast evaluation is the sound assessment of the relative performance of several competing forecasts. Important tools to perform this task for point forecasts are *scoring functions*. If a functional, i.e. a statistical property, of a distribution is a minimizer of the expected score, the scoring function is *consistent* for this functional. The functional is called *elicitable* if it defines the unique minimizer for any distribution.

Elicitability is a desirable property for comparative forecast evaluation and it enables regression and M-estimation. Although many widely used functionals like the mean or quantiles are elicitable, there are major exceptions such as the variance, the mode, and the prominent financial risk measure Expected Shortfall. As a result, the idea of *elicitation complexity* of a functional was revisited and refined. Loosely speaking, a functional has elicitation complexity at most k, if there is a k-dimensional elicitable intermediate functional from which it can be computed.

Motivated by the growing interest in properties of distribution tails, we investigate whether consistent scoring functions can be used to compare point forecasts of such properties. To do this, we introduce the class of *max-functionals*, which contains key characteristics from extreme value theory, in particular the extreme value index. We demonstrate

that max-functionals fail to be elicitable and using mild regularity assumptions, we go on to show that they have infinite elicitation complexity. Therefore, we then consider a different forecast setting, in which the whole distribution function is reported, and then evaluated using proper scoring rules. We show that one cannot expect a satisfying comparison of max-functional values from this method.

Our findings highlight that the use of scoring functions or scoring rules is impossible or has serious drawbacks when it comes to important tail properties such as max-functionals. We conclude that caution should be exercised in forecast evaluation if relevant information is encoded by such functionals.

Session, time and room:
Prediction
Tuesday, 16:40 – 17:00 (A102)

Multiple block sizes and overlapping blocks for multivariate time series extremes

Axel Bücher[1], Stanislav Volgushev[2], Nan Zou[3]
Heinrich-Heine Universität Düsseldorf, Germany[1], University of Toronto, Canada[2,3]
axel.buecher@hhu.de[1], stanislav.volgushev@utoronto.ca[2], nan.zou@utoronto.ca[3]

Block maxima methods constitute a fundamental part of the statistical toolbox in extreme value analysis. However, most of the corresponding theory is derived under the simplifying assumption that block maxima are independent observations from a genuine extreme value distribution. In practice however, block sizes are finite and observations from different blocks are dependent. Theory respecting the latter complications is not well developed, and, in the multivariate case, has only recently been established for disjoint blocks of a single block size. We show that using overlapping blocks instead of disjoint blocks leads to a uniform improvement in the asymptotic variance of the multivariate empirical distribution function of rescaled block maxima and any smooth functionals thereof (such as the empirical copula), without any sacrifice in the asymptotic bias. We further derive functional central limit theorems for multivariate empirical distribution functions and empirical copulas that are uniform in the block size parameter, which seems to be the first result of this kind for estimators based on block maxima in general. The theory allows for various aggregation schemes over multiple block sizes, leading to substantial improvements over the single block length case and opens the door to further methodology developments. In particular, we consider bias correction procedures that can improve the convergence rates of extreme-value estimators and shed some new light on estimation of the second-order parameter when the main purpose is bias correction.

Session, time and room:
IS - Time series extremes (Kulik)
Thursday, 14:00 – 14:30 (A002)

Ridge Regression Estimators for the Extreme Value Index

Sven Buitendag[1]
Stellenbosch University, South Africa[1]
svenbuitendag@gmail.com[1]

We consider bias reduced estimators of the extreme value index (EVI) in case of Pareto-type distributions and under all max-domains of attraction. To this purpose we revisit the regression approach started in [1] and in [2] for the case of a positive EVI, and in [3] for real-valued EVI.

We generalize these approaches using ridge regression exploiting the mathematical fact that the bias tends to 0 when the number of top data points used in the estimation is decreased. The penalty parameter is selected by minimizing the asymptotic mean squared error of the proposed estimator.

References

[1] Feuerverger, A. & Hall, P. Estimating a tail exponent by modelling departure from a pareto distribution. Annals of Statistics, 1991

[2] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G. Tail index estimation and an exponential regression model. Extremes, 1999

[3] Beirlant, J., Dierckx, G., Guillou, A. Estimation of the extreme-value index and generalized Quantile plots. Bernoulli, 2005

Session, time and room:
Best student paper II
Tuesday, 9:00 – 9:20 (A101)

A nonparametric estimator of the extremal index

Juan Juan Cai[1], Andrea Krajina[2]
Delft University of Technology, Netherlands[1], RABO Bank, Netherlands[2]
j.j.cai@tudelft.nl[1], akrajina@gmail.com[2]

This paper studies the estimation of an extremal index θ. We build a connection between θ and the stable tail dependence function, which is also used to formulate the $D^{(d)}(u_n)$ condition, a typical mixing condition for studying the extremal index. We establish an inference procedure for estimating d and a non-parametric estimator of θ depending on d. The performance of the estimations is confirmed by the obtained asymptotic normality and a simulation study. We also apply our method to a real data set on the hourly wind speed measured at Schiphol airport.

Session, time and room:
Statistics for time series
Tuesday, 9:25 – 9:45 (A002)

Extra-Parametrized Extreme Value Copula : Extension to a Spatial Framework

Julie Carreau[1], Gwladys Toulemonde[2]

Institut de Recherche pour le Développement, France[1], Institut Montpellierain Alexander Grothendiek, Inria project-team LEMON, France[2]

julie.carreau@ird.fr[1], gwladys.toulemonde@umontpellier.fr[2]

When hazard assessment is based on quantities involving several locations, knowledge on spatial dependence of extreme events is necessary. While the Generalized Extreme Value distribution can approximate univariate marginal distributions of maxima of variables, extreme-value copulas can be used to describe their dependence structure, as justified by the theory of multivariate extremes. However, most high-dimensional copulas are too simplistic for a spatial application. Recently, a class of flexible extreme-value copulas was put forward by combining two Gumbel extreme-value copulas with a weight parameter in the unit hyper-cube. In a multisite study, the copula dimension being the number of sites, this extra-parametrized approach, called *XGumbel*, quickly becomes over-parametrized. In addition, interpolation is not straightforward. Motivated by an application on annual maxima of daily precipitation totals in a relatively large region in the French Mediterranean, we aim to extend this approach to a spatial framework. The spatial XGumbel is defined thanks to a spatial mapping for the extra-parameters that changes randomly from year to year. As both log-likelihood and pairwise log-likelihood are intractable, we implement a version of the Approximate Bayesian Computation (ABC) scheme for the inference. The suitability and workability of the proposed spatial XGumbel model and of the ABC inference scheme are assessed on the precipitation data from our study area. Further analysis is performed by estimating probabilities involving three stations jointly.

Session, time and room:
Copula based methods
Tuesday, 9:00 – 9:20 (A102)

A spliced Gamma-Generalized Pareto model for short-term extreme wind speed probabilistic forecasting

Daniela Castro-Camilo[1], Raphaël Huser[2], Håvard Rue[3]

University of Glasgow, United Kingdom[1], King Abdullah University of Science and Technology, Saudi Arabia[2,3]

daniela.castro.camilo@gmail.com[1], raphael.huser@kaust.edu.sa[2], haavard.rue@kaust.edu.sa[3]

Renewable sources of energy such as wind power have become a sustainable alternative to fossil fuel-based energy. However, the uncertainty and fluctuation of the wind speed derived from its intermittent nature bring a great threat to the wind power production stability, and to the wind turbines themselves. Lately, much work has been done on developing models to forecast average wind speed values, yet surprisingly little has focused on proposing models to accurately forecast extreme wind speeds, which can damage the turbines. In this work, we develop a flexible spliced Gamma-Generalized Pareto model to forecast extreme and non-extreme wind speeds simultaneously. Our model belongs to

the class of latent Gaussian models, for which inference is conveniently performed based on the integrated nested Laplace approximation method. Considering a flexible additive regression structure, we propose two models for the latent linear predictor to capture the spatio-temporal dynamics of wind speeds. Our models are fast to fit and can describe both the bulk and the tail of the wind speed distribution while producing short-term extreme and non-extreme wind speed probabilistic forecasts.

Session, time and room:
Environmental extremes
Monday, 14:00 – 14:20 (A002)

Modeling clustered subasymptotic extremes

Simon Chatelain[1], Anne-Laure Fougères[2], Johanna G. Nešlehová[3], Samuel Perreault[4]
McGill University, Canada[1,3], Université Lyon 1, France[2], Univesité Laval, Canada[4]
simon.chatelain@mail.mcgill.ca[1], fougeres@math.univ-lyon1.fr[2],
johanna.neslehova@mcgill.ca[3], samuel.perreault.3@ulavala.ca[4]

There is a growing literature that addresses the need to adapt clustering techniques to the context of extremes (e.g. [1]), to bridge asymptotic dependence and independence (e.g. [2], [3]), and to account for extremes at subasymptotic levels (e.g. [4]).

This talk contributes to this ongoing research by presenting a hierarchical approach to capture the dependence between clustered subasymptotic extremes. The proposed model uses dependent distortions $R_1, \ldots, R_N$ to account for the dependence between N clusters, each of which represents an asymptotically dependent regime characterized by an Archimax copula. The tail behavior of the model is elicited; as will be seen, it allows for both asymptotic independence and dependence between clusters. We present a clustering algorithm driven by the stochastic representation of the model and its tail properties to learn the cluster structure from data, and explain how the model can be fitted using semi-parametric techniques as developed in [5]. An application to rainfall data over France is used to illustrate the procedures.

References

[1] E. Bernard, P. Naveau, M. Vrac, O. Mestre, Clustering of maxima: Spatial dependencies among heavy rainfall in France, *Journal of Climate*, 26(20), 7929-7937, 2013

[2] J. L. Wadsworth, and J. A. Tawn, A. C. Davison, D. M. Elton, Modelling across extremal dependence classes, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1), 149-175, 2017

[3] R. Huser, T. Opitz, E. Thibaud, Modelling across extremal dependence classes, *Spatial Statistics*, 21, 166-186, 2017

[4] P. Naveau, R. Huser, P. Ribereau, A. Hannart, Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, *Water Resources Research*, 52(4), 2753-2769, 2016

[5] S. Chatelain, A.-L. Fougères, J.G. Nešlehová. Inference for Archimax copulas, *The Annals of Statistics*, 47, in press, 2019

Session, time and room:
Copula based methods
Tuesday, 9:25 – 9:45 (A102)

Extremes of Long Range Dependent Stable Random Fields

Zaoli Chen[1]
Cornell University, United States[1]
zc288@cornell.edu[1]

Random sup-measures(RSM) reflect the extremal behaviors and dependent structures of random fields. We establish limit theorems of RSM for some long range dependent stable random fields. We show that only for certain range of parameters RSM limits have Frechet distributions.

Session, time and room:
IS ST - Infinitely divisible models and their extremes (Samorodnitsky)
Friday, 9:00 – 9:30 (A001)

The maximal degree in a Poisson-Delaunay graph

Gilles Bonnet[1], Nicolas Chenavier[2]
Ruhr-Universität, Bochum, Germany[1], Université du Littoral Côte d'Opale, France[2]
gilles.bonnet@ruhr-uni-bochum.de[1], nicolas.chenavier@univ-littoral.fr[2]

A Delaunay triangulation associated with a locally finite subset χ in $\mathbf{R}^2$ is a triangulation $DT(\chi)$ such that no point in χ belongs to the interior of the circumdisk of any triangle in $DT(\chi)$. This model is the key ingredient of the first algorithm for computing the minimal spanning tree and is extensively used in various domains, such as medical image segmentation and finite element methods to build meshes.

In this talk, we consider the case where $\chi = \eta$ is a homogeneous Poisson point process in $\mathbf{R}^2$. We investigate the maximal degree Δ_n of the so-called Delaunay graph associated with η (consisting of the set of edges of triangles) observed in the window $W_n = [-n, n]^2$, namely

$$\Delta_n = \max_{x \in \eta \cap W_n} d_\eta(x),$$

where $d_\eta(x)$ is the degree of any point x, i.e. the number of edges passing through x. As n goes to infinity, we show that Δ_n is concentrated on two consecutive integers. We also provide the exact order of Δ_n.

Session, time and room:
IS - Topological and geometric extremes (Owada)
Thursday, 14:00 – 14:30 (A001)

Kernel estimation of local tail-event correlation

Ji-Eun Choi[1], Dong Wan Shin[2]
Ewha womans university, Korea, Republic of[1,2]
cje6479@gmail.com[1], shindw@ewha.ac.kr[2]

Tail event correlation coefficient is considered to measure local tail dependence between a pair of time series which the usual dynamic correlation is unable to capture. A nonparametric kernel estimation method is employed to estimate the local tail event correlation coefficient. Limiting normality is established for the local estimator. Confidence interval and equality test of the tail event correlations at specific time points are developed. Finite sample validity is demonstrated for the proposed statistical inferences to have reasonable coverage, size and power performances. A real data analysis finds that left-tail correlations between a financial system stock return and financial institution stock return become significantly stronger during a period before the out-break of the global financial crisis, while the tail correlations at median level do not reveal such increased tendency.

Session, time and room:
Statistics for time series
Tuesday, 9:50 – 10:10 (A002)

Extreme value analysis of dry spells with Bayesian inference

Ksenija Cindrić Kalin[1], Zoran Pasarić[2]
Croatian Meteorological and Hydrological Service, Croatia[1], Department of Geophysics, Faculty of Science, University of Zagreb, Croatia[2]
cindric@cirus.dhz.hr[1], zpasaric@gfz.hr[2]

Dry spells are defined as consecutive sequences of days with total daily precipitation amount lower than some fixed, predefined limit (taken here to be 1, 5 and 10 mm/day). In order to obtain reliable estimates of dry-spell return values for different return periods we employ the extreme value distributions: the generalized extreme value distribution is applied to the annual maxima data (AM-GEV) together with the Gumbel distribution, while the generalized Pareto distribution is applied to the peak-over-threshold data (POT-GP). Methods devised to estimate all three GP-parameters are compared to those dealing with only the shape and scale parameters. All inference work is done within the Bayesian paradigm. In this way it was straightforward to take into account that dry spell durations are observed as discrete values, obtained by rounding the underlying true durations to the whole number of days. The results confirmed precautionary estimations of the GEV model applied to AM data in comparison with a simpler Gumbel model which is often employed in practice. Regarding the POT-GP modelling, it is suggested that all three GP parameters, thus including the location parameter, should be assessed when dealing with rounded data (such as dry spells). The location parameter estimates should be compatible with the theoretical value of - 0.5. The Bayesian approach revealed a good stability of parameters, starting from rather low thresholds, and also pointed to a large uncertainty that can occur when very large thresholds are applied. The results are used to analyze the dry spells in Croatia. In order to identify spatial patterns of drought risk, return values of

dry spells for different return periods (T = 2, 5, 10, 25, 50 and 100 years) are estimated for large number of stations with precipitation data spanning the period 1961-2015.

Session, time and room:
Bayesian methods for environmental extremes
Monday, 10:30 – 10:50 (A002)

Extrema of Gaussian random interfaces

Alessandra Cipriani[1], Biltu Dan[2], Rajat Subhra Hazra[3]
TU Delft, Netherlands[1], Indian Statistical Institute, Kolkata, India[2,3]
A.Cipriani@tudelft.nl[1], biltudanmath@gmail.com[2], rajatmaths@gmail.com[3]

Random interfaces arise naturally as separating surfaces between two different thermodynamic phases or states of matter. When one views them as fields of random heights, it is natural to investigate the behavior of their rescaled maxima. In this talk we will review the current state of the art concerning the limiting behavior of extrema of some Gaussian models that play a major role in statistical mechanics: the discrete Gaussian free field (DGFF), the membrane model (MM), and the $(\nabla + \Delta)$-model, which represents a mixture between DGFF and MM. We will present their similarities and differences, in particular how a range of various approaches is needed to study their extremes. Techniques we will employ range from the theory of partial differential equations, numerical schemes and the Stein–Chen method.

Session, time and room:
IS - Extremes of branching walks and free fields (Roy)
Monday, 11:05 – 11:35 (A001)

A Lorenz View of the Bivariate Pickands Dependence Function

Pasquale Cirillo[1], Andrea Fontanari[2], Juerg Huesler[3]
Delft University of Technology, Netherlands[1,2], University of Bern, Switzerland[3]
P.Cirillo@tudelft.nl[1], A.Fontanari@tudelft.nl[2], juerg.huesler@stat.unibe.ch[3]

We propose a novel way of representing and characterizing the Pickands dependence function in the bivariate framework [1], using the geometric and probabilistic interpretations of the Lorenz curve, a well know transform in wealth inequality studies [2]. We notice that the Pickands dependence function is nothing but a Lorenz curve in a particular coordinate system studied by Corrado Gini about a century ago [3]. Once this connection is established, we can use the representation of the Lorenz curve as integral of the quantile function of positive random variables to obtain a simple way for generating Pickands dependence functions, as well as for characterizing the relative measure generating functions. This new approach also allows to import in the Pickands framework all the set of concentration measures usually found in inequality studies, giving them brand new interpretations in terms of extremal dependence.

References

[1] J.A. Tawn, Bivariate extreme value theory: models and estimation, *Biometrika* 75: 397-415, 1988.

[2] M.O. Lorenz, Methods of measuring the concentration of wealth, *Publications of the American Statistical Association* 9: 209-219, 1905.

[3] C. Gini, Intorno alle curve di concentrazione, *Metron* 9: 3-76, 1932.

Session, time and room:
Tail dependence
Thursday, 11:15 – 11:35 (A001)

Rare event simulation for recursive sequences governed by random matrices

Jeffrey Collamore[1]
University of Copenhagen, Denmark[1]
collamore@math.ku.dk[1]

Random recursive sequences arise in a wide class of problems in applied probability, including integrated insurance/financial models and the GARCH financial processes, as well as certain non-standard risk models. In these contexts, it is often of interest to compute the tail probabilities of the process under stationarity. In a one-dimensional setting, a variety of techniques have been developed for simulating these "rare event" probabilities (cf. [1], [2]).

The goal of this talk will be to extend these methods to a multidimensional setting. We are particularly interested in developing rare event simulation methods for the "matrix" recursive sequence

$$V_n = A_n V_{n-1} + B_n, \quad n = 1, 2, \ldots,$$

where $\{(A_n, B_n) : n = 1, 2, \ldots\}$ are i.i.d. copies of (A, B), A is a $d \times d$ random matrix and B a random vector in $\mathbb{R}^d$, both with nonnegative entries, and the matrix A is "contracting" in the sense of [4], thus guaranteeing a unique stationary distribution for $\{V_n\}$. Such models appear, e.g., in the study of the general GARCH(p,q) financial time series model and the study of multi-type branching processes in random environments.

To develop an efficient algorithm in this multivariate setting, we build upon our earlier work in [2], where an analogous algorithm was developed in the one-dimensional setting, and in [3], where related rare event probabilities and paths are studied, from a theoretical perspective, for the multivariate recursive sequence $\{V_n\}$. Specifically, we utilize the regenerative structure of $\{V_n\}$, together with a generalization of the "dual" change of measure in [2], suitably adapted to random matrices, to develop an efficient algorithm. We also describe the computational complexity by explicitly analyzing the running time of the algorithm. (Based on joint work with A.N. Vidyashakar, G. Diao, and X. Xie.)

References

[1] J. Blanchet, H. Lam, B. Zwart. Efficient rare-event simulation for perpetuities. *Stochastic Process. Appl.* **122** 3361-3392.

[2] J.F. Collamore, G. Diao, A.N. Vidyashankar, A.N. (2014). Rare event simulation for processes generated via stochastic fixed point equations. *Ann. Appl. Probab.* **24** 2143-2175 (2014).

[3] J.F. Collamore, S. Mentemeier. Large excursions and conditioned laws for recursive sequences generated by random matrices. *Ann. Probab.* **46** 2064-2120 (2018).

[4] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207-248 (1973).

Session, time and room:
IS ST - Simulation of rare events and extremes (Zwart)
Tuesday, 9:00 – 9:30 (A001)

Decomposition of Extremal Dependence and Applications

Daniel Cooley[1]
Colorado State University, United States[1]
cooleyd@stat.colostate.edu[1]

Principal components (alternatively empirical orthogonal functions) are widely used to study modes of dependence in multivariate data. However PCA/EOFs arise by performing an eigendecomposition of the covariance matrix, and therefore are not well-suited for describing extremal dependence. Characterizing extreme dependence in high dimensions is difficult for most existing multivariate extreme modeling frameworks.

This talk will briefly review a recently proposed method for investigating extremal dependence. The method summarizes extremal dependence via the tail pairwise dependence matrix (TPDM), which can be seen as an extreme analog to the covariance matrix. An eigendecomposition of the TPDM results in an ordered orthonormal basis through which the modes of extremal dependence can be studied.

This talk will largely focus on applications of this method. Applying these methods to US precipitation data recorded at 1100 stations, we investigate relationships between the extremes of the time series of basis coefficients and climatological drivers such as the El Nino/La Nina oscillation. We also will present results from investigating US temperature data, and preliminary work in modeling UK flood data.

Session, time and room:
IS - Extremes and machine learning (Sabourin)
Monday, 16:15 – 16:45 (003)

M-depth and multivariate M-quantiles

Abdelaati Daouia[1], Davy Paindaveine[2]
University of Toulouse Capitole, France[1], Université libre de Bruxelles, Belgium[2]
abdelaati.daouia@tse-fr.eu[1], dpaindav@ulb.ac.be[2]

Despite the renewed interest in the concept of expectiles in fields such as econometrics and risk management, expectile regression—or, more generally, M-quantile regression—unfortunately remains limited to single-output problems. To improve on this, we introduce hyperplane-valued multivariate M-quantiles that show strong advantages, for instance in terms of equivariance, over the various point-valued multivariate M-quantiles available in the literature. Like their competitors, our multivariate M-quantiles are directional in nature and provide centrality regions when all directions are considered. These regions define a new statistical depth, the halfspace M-depth, whose deepest point, in the expectile case, is the mean vector. Remarkably, the halfspace M-depth can alternatively be obtained by substituting, in the celebrated Tukey halfspace depth, M-quantile outlyingness for standard quantile outlyingness, which supports a posteriori the claim that our multivariate M-quantile concept is the natural one. We investigate thoroughly the properties of the proposed multivariate M-quantiles, of halfspace M-depth, and of the corresponding regions. Since our original motivation was to define multiple-output expectile regression methods, we further focus on the expectile case. We show in particular that expectile depth is smoother than the Tukey depth and enjoys interesting monotonicity properties that are extremely promising for computational purposes. Unlike their quantile analogs, the proposed multivariate expectiles also satisfy the axioms of multivariate risk measures. Finally, we show that our multivariate expectiles indeed allow performing multiple-output expectile regression, which is illustrated on both simulated and real data.

Session, time and room:
IS - Advances in statistics of multivariate extremes (Einmahl)
Monday, 16:15 – 16:45 (A001)

Heavy-tails in a distributionally robust newsvendor model

Bikramjit Das[1], Anulekha Dhara[2], Karthik Natarajan[3]
Singapore University of Technology and Design, Singapore[1,3], University of Michigan, United States[2]
bikram@sutd.edu.sg[1], anulekha.dhara@gmail.com[2], karthik_natarajan@sutd.edu.sg[3]

A solution to the newsvendor problem (a classical inventory problem) with ambiguity in the demand distribution was provided by [1]. The optimal order quantity in this problem is computed for the worst possible distribution from a set of demand distributions that is characterized by partial information, such as moments. A simple observation indicates that the optimal order quantity with knowledge of just first two moments of the distribution is also optimal for a censored t-distribution with infinite variance. In this paper, we generalize this *heavy-tail optimality* property of the distributionally robust newsvendor model to the case when information on the first and the α- moment is known for any $\alpha > 1$. Unlike the mean-variance setting, this problem does not appear to be solvable in a simple closed

form manner. We show that for high critical values, the optimal order quantity for the distributionally robust newsvendor is also optimal for a regularly varying distribution with parameter indexed by α. We also provide generalizations to a multivariate set-up.

References

[1] H. Scarf, A min-max solution of an inventory problem, in *Studies in The Mathematical Theory of Inventory and Production*, 19958.

Session, time and room:
New areas for EVT
Tuesday, 11:15 – 11:35 (003)

Practical issues in the statistics of environmental extremes

Anthony Davison[1], Jonathan Tawn[2]
EPFL, Switzerland[1], Lancaster University, United Kingdom[2]
anthony.davison@epfl.ch[1], j.tawn@lancaster.ac.uk[2]

The extreme value statistician has a large tool-box of methods at hand, with well-established techniques covering times series of univariate, multivariate and spatial processes. In these talks we will aim to describe the challenges of using and adapting these methods when confronted with problems involved environmental processes involving data on sea-levels, rainfall, river flooding, high winds and heatwaves. Some of these challenges also arise in other domains of application of extreme-value theory, but they are probably most often encountered in environmental problems. We will aim to bring our combined experience of 50+ years of working with applications and collaborators from different domains to illustrate both the resulting methodological and inferential challenges and the opportunities to make a real impact with research on extreme value analysis.

Session, time and room:
Overview talk
Wednesday, 8:45 – 10:10 (Academy of Music)

Similarity-Based Clustering for Heteroscedastic Extremes

Miguel de Carvalho[1], Raphaël Huser[2], Rodrigo Rubio[3]

The University of Edinburgh / CEAUL, University of Lisbon, United Kingdom[1], King Abdullah University of Science and Technology (KAUST), Saudi Arabia[2], Pontificia Universidad Catolica de Chile, Chile[3]

Miguel.deCarvalho@ed.ac.uk[1], raphael.huser@kaust.edu.sa[2], rerubio@mat.uc.cl[3]

Statistical modelling of the *magnitude* and the *frequency* of extreme losses in a stock market is essential for institutional investors, professional money managers, and traders. In this paper, we develop statistical methods of similarity-based clustering for heteroscedastic extremes, which allow us to group stocks according to their extreme-value index and sceda-sis function, i.e., the magnitude and frequency of extreme losses, respectively. Clustering is performed here in a product-space and a tuning parameter is used to control whether more emphasis should be put on the latter or the former. The proposed approach also allows for clustering stocks with similar risk loss patterns, by identifying affinities in time-varying value-at-risk functions. This provides a practical tool to gain more insight into stocks synchronization during periods of stress, and to capture the level of similarity of stocks over such periods. Our analysis reveals an interesting mismatch between the magnitude and frequency of extreme losses on the London Stock Exchange and the corresponding economic sectors of the affected stocks. The analysis further suggests that the dynamics governing the comovement of extreme losses in the exchange contains information on the business cycle.

Session, time and room:
Statistics for time series
Tuesday, 10:15 – 10:35 (A002)

Functional Peaks-over-threshold Analysis and its Applications

Anthony Davison[1], Raphaël de Fondeville[2]

EPFL, Switzerland[1,2]

anthony.davison@epfl.ch[1], raphael.de-fondeville@epfl.ch[2]

Estimating the risk of single occurrences of natural hazards has become important in recent decades, but up until now it has been largely limited to re-using catalogs of historical events, which usually do not exceed 40 to 50 years in length, and to numerical models, which require heavy computation and are often unreliable for extrapolation. Extreme value theory provides statistical methods for estimating the frequency of past extreme events as well as for extrapolating beyond observed severities, but natural hazards cannot be modelled using only univariate results.

We present an extension of peaks-over-threshold analysis to functions which allows one to define complex extreme events as special types of exceedances, and then obtain their limit distribution for increasingly high thresholds, namely the generalized r-Pareto process. We focus on a specific model based on log-Gaussian random functions using classical covariance structures to characterize extremal dependence. Then, we describe a stochastic weather generator for extreme events, capable of quantifying the recurrence of

past events as well as generating completely new ones. The methodology is applied to several natural hazards such as windstorms and heavy rainfall.

Session, time and room:
IS - Spatial extremes (Dombry)
Monday, 15:10 – 15:40 (003)

Combining multivariate and univariate regular variation

Laurens de Haan[1]
Erasmus unversity Rotterdam and University of Lisbon, Netherlands[1]
ldehaan@ese.eur.nl[1]

We start with a reminder of how regular variation and (statistics of) extreme value theory are connected, in a one-dimensional context as well as in a multivariate framework. Then I want to mention an old result (de Haan and Resnick 1993) that allows one to get all the results starting from one condition. If time permits, we can discuss the way one can fit a generalized Pareto distribution to the tail of a multivariate probability distribution under regular variation conditions.

Session, time and room:
IS ST - Regular variation: history, ideas and people (Basrak)
Friday, 13:00 – 13:30 (A001)

Hidden regular variation for point processes and the single/multiple large point heuristic

Clement Dombry[1], Charles Tillier[2], Olivier Wintenberger[3]
Universite de Franche-Comte, France[1], Telecom Paristech, France[2], Sorbonne Universite, France[3]
clement.dombry@univ-fcomte.fr[1], charles.tillier@gmail.com[2],
olivier.wintenberger@upmc.fr[3]

We consider regular variation for marked point processes with independent heavy-tailed marks and prove a single large point heuristic: the limit measure is concentrated on the cone of point measures with one single point. We then investigate successive hidden regular variation removing the cone of point measures with at most k points, $k \geq 1$, and prove a multiple large points phenomenon: the limit measure is concentrated on the cone of point measures with $k+1$ points. We show how these results imply hidden regular variation in Skorokhod space of the associated risk process, in connection with the single/multiple large jump heuristic from Rhee et al. [1]. Finally, we provide an application to risk theory in a reinsurance model where the k largest claims are covered and we study the asymptotic behavior of the residual risk.

References

[1] C.-H. Rhee, J. Blanchet and B. Zwart, B, Sample path large deviations for heavy-tailed Levy processes and random walk, *Annals of Probab.*, 2019.

Session, time and room:
IS ST - Regular variation: generalisations and recent advances (Molchanov & Basrak)
Thursday, 9:00 – 9:30 (003)

Disjoint and moving block methods: towards a unifying framework

Holger Drees[1], Sebastian Neblung[2]
University of Hamburg, Department of Mathematics, Germany[1,2]
hdrees_2000@yahoo.de[1], sebastian.neblung@uni-hamburg.de[2]

Consider a stationary time series $(X_t)_{1\le t\le n}$ of length n. Extreme value estimators are often constructed from blockwise defined statistics. Then one may either split the time series in disjoint blocks $(X_t)_{(i-1)l+1\le t\le il}$, $1 \le i \le [n/l]$, of length l, or one may use moving blocks $(X_t)_{i\le t\le i+l-1}$, $1 \le i \le n-l+1$. For specific statistics, it has been observed that the latter approach leads to a better asymptotic behavior; see, e.g., [1] and [2].

We will discuss a general unifying framework in which both types of statistics can be analyzed asymptotically.

References

[1] A. Bücher and J. Segers, Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.*, 12, 1098–1125, 2018.

[2] C.Y. Robert, J. Segers and A.T. Ferro. A sliding blocks estimator for the extremal index. *Electron. J. Statist.*, 3, 993–1020, 2009

Session, time and room:
IS - Time series extremes (Kulik)
Thursday, 14:35 – 15:05 (A002)

Random sup-measures with long-range dependence

Olivier Durieu[1]
University of Tours, France[1]
olivier.durieu@univ-tours.fr[1]

We introduce a family of self-similar and translation-invariant random sup-measures with long-range dependence. These sup-measures are limit objects in a model based on random partitions inspired by an infinite urn scheme previously studied by Karlin (1967). They reflect the long-range dependence nature of the original process and characterize how the locations of extremes appear as long-range clusters represented by random closed sets. A limit theorem for the corresponding point-process convergence is established.
This is a joint work with Yizao Wang (University of Cincinnati, USA).

Session, time and room:
IS ST - Infinitely divisible models and their extremes (Samorodnitsky)
Friday, 9:35 – 10:05 (A001)

Extreme value statistics for high-dimensional data

John Einmahl[1], Yi He[2]
Tilburg University, Netherlands[1], Monash University, Australia[2]
j.h.j.einmahl@uvt.nl[1], yi.he2@monash.edu[2]

We propose a novel statistical formulation of the empirical power laws widely observed in high-dimensional data sets. Our approach extends classical extreme value theory to specifying the behavior of the empirical distribution of data with a complex dependence structure and possibly different marginal distributions, for a diverging number of dimensions. The main assumption is that in the intermediate tail the empirical distribution function approaches some heavy tailed distribution function that is in the max-domain of attraction. In this setup the Hill estimator consistently estimates the extreme value index and extreme quantiles are consistently estimated, on a log-scale. We discuss several model examples that satisfy our conditions. We also consider applications to finance.

Session, time and room:
High-dimensional data
Thursday, 16:15 – 16:35 (003)

Graphical Models, Sparsity and Structure Learning for Extremes

Sebastian Engelke[1], Adrien Hitz[2], Stanislav Volgushev[3]

University of Geneva, Switzerland[1], Materialize.X, United Kingdom[2], University of Toronto, Canada[3]

sebastian.engelke@unige.ch[1], Adrien.hitz@gmail.com[2], stanislav.volgushev@utoronto.ca[3]

Conditional independence, graphical models and sparsity are key notions for parsimonious models in high dimensions and for learning structural relationships in the data. The theory of multivariate and spatial extremes describes the risk of rare events through asymptotically justified limit models such as max-stable and multivariate Pareto distributions. Statistical modeling in this field has been limited to moderate dimensions so far, owing to complicated likelihoods and a lack of understanding of the underlying probabilistic structures.

We introduce a new notion of conditional independence for a multivariate Pareto distributions $\mathbf{Y} \in \mathbb{R}^d$, denoted by $\perp\!\!\!\perp_e$. For a graph $\mathcal{G} = (V, E)$ with nodes $V = \{1, \ldots, d\}$, we define an extremal graphical model on $\mathcal{G}$ as the distribution of $\mathbf{Y}$ satisfying the pairwise Markov property

$$Y_i \perp\!\!\!\perp_e Y_j \mid \mathbf{Y}_{\setminus\{i,j\}}, \qquad \text{for all } (i,j) \notin E.$$

This allows to study sparsity for extremes and to build new parametric models in a modular way. Statistical inference for such sparse models can be simplified to lower-dimensional margins. We define the extremal variogram, a new summary statistics that turns out to be a tree metric and therefore allows to efficiently learn an underlying tree structure through Prim's algorithm. For a popular parametric class of multivariate Pareto distributions we show that, similarly to the Gaussian case, the sparsity pattern of a general graphical model can be easily read of from suitable inverse covariance matrices. This enables the definition of an extremal graphical lasso that enforces sparsity in the dependence structure. We illustrate the results with an application to flood risk assessment on the Danube river.

The talk is partly based on [1].

References

[1] S. Engelke and A. Hitz, Graphical Models for Extremes, *https://arxiv.org/abs/1812.01734*, 2018.

Session, time and room:
IS - Extremes on graphs (Klueppelberg)
Friday, 9:00 – 9:30 (A002)

Local Estimation under the mixture cure model with insufficient follow-up.

Mikael Escobar-Bach[1], Ingrid van Keilegom[2]
Université d'Angers, France[1], KU Leuven, Belgium[2]
mikael.escobar-bach@univ-angers.fr[1], ingrid.vankeilegom@kuleuven.be[2]

One major topic in survival analysis is related to the modelling of the cure rate, i.e. the proportion of subjects that will never experience the event of interest. However, most estimation methods proposed so far in the literature do not handle the case of insufficient follow-up, that is when the right end point of the support of the censoring time is strictly less than that of the survival time of the susceptible subjects. In this talk, we propose a new approach to estimate the conditional cure rate and survival time distribution function using extrapolation techniques from the area of the extreme value theory. We establish the asymptotic normality of the proposed estimators, and show how the estimator works by means of a simulation study. We also illustrate its practical applicability through a short analysis of data on the survival of colon cancer patients.

Session, time and room:
Inference for censored data
Friday, 9:00 – 9:20 (A101)

Diagnosing concurrent drivers of weather extremes: application to hot and cold days in North America

Davide Faranda[1], Gabriele Messori[2], Pascal Yiou[3]
CNRS - IPSL - LSCE, France[1], Uppsala University & Meteorological Institut Stockholm University, Sweden[2], CEA - IPSL - LSCE, Croatia[3]
davide.faranda@cea.fr[1], gabriele.messori@misu.su.se[2], pascal.yiou@cea.fr[3]

A fundamental challenge in climate science is decomposing the concurrent drivers of weather extremes. This would provide insights into the drivers of individual extreme events as well as possible future changes in extreme event frequencies under greenhouse forcing. In the present work, we exploit recent results from dynamical systems theory to study the co-variation and recurrence statistics of different atmospheric fields. Specifically, we present a methodology to quantify the recurrences of bivariate fields, the repeated co-occurrences of distinct univariate fields, and the dependence between two fields. The dependence is defined by a coupling parameter, which varies according to the chosen fields, season, and domain and can be understood in terms of the underlying physics of the atmosphere. For suitably chosen fields, such as near-surface temperature and sea level pressure, this approach enables to decompose the different drivers of weather extremes. Using the above metrics, we study hot or cold days over North America. We first identify states where temperature extremes are strongly and weakly coupled to the large-scale atmospheric circulation, and then elucidate the interplay between coupling and the occurrence of temperature extremes.

Session, time and room:
IS - Extremes and climate physics (Fougeres)
Tuesday, 9:00 – 9:30 (003)

Spatial dependence and space-time trend in extreme events

John Einmahl[1], Ana Ferreira[2], Laurens de Haan[3], Cláudia Neves[4], Chen Zhou[5]
Tilburg University, Netherlands[1], IST University of Lisbon, Portugal[2], Erasmus University Rotterdam and CEAUL, Portugal[3], Reading University, United Kingdom[4], Erasmus University Rotterdam, Netherlands[5]
J.H.J.Einmahl@uvt.nl[1], anafh@tecnico.ulisboa.pt[2], ldehaan@ese.eur.nl[3], c.neves@reading.ac.uk[4], zhou@ese.eur.nl[5]

Consider a sequence of stochastic processes observed independently in time with no-time-changing extremal spatial dependence. Under heteroscedastic conditions on the marginal distribution functions, and maximum domain of attraction conditions with constant extreme value index, we provide two statistical tests to evaluate abnormal frequencies of extremal events over space, and over time.

We give a weighted approximation of the normalized intermediate order statistics from all marginal observations, from which a variety of asymptotic properties such as convergence results for estimators of extreme value parameters can be established, including those from maximum likelihood.

Results will be illustrated with rainfall data from Northwest Germany.

Session, time and room:
IS - Spatial extremes (Dombry)
Monday, 14:00 – 14:30 (003)

Use of multivariate extreme value theory for anomalies detection in time series data

Vincent Feuillard[1]
Airbus Central R&T, AI Research / Accor Advanced Analytics & AI, France[1]
vincent.feuillard@gmail.com[1]

In the context of Prognosis and Health Management of aircrafts, anomalies detection is a key tool. For instance, this helps in the identification of the degradation of a monitored system. One difficulty for this task is that the collected data from most of Airbus platforms mainly correspond to time series. Some dedicated algorithms have to be developed for this particular kind of data. In this talk, we investigate novel statistical techniques borrowing concepts and tools from machine-learning and multivariate extreme value analysis both at the same time. First, we use a Functional Principal Component Analysis for representing the data. We then consider anomalies as extreme events in this representation and we study their dependence structure. We use a multivariate extreme value based algorithm to detect this structure under a sparse assumption. In addition, the representation of the dependence structure in the extremes induces a specific notion of (dis-)similarity among anomalies. This paves the way for helping the interpretation and finding similarities in the anomalies (ex: same system suspicious behaviour).

Session, time and room:
IS - Extremes and machine learning (Sabourin)
Monday, 17:25 – 17:55 (003)

Lorenz Generators for Archimedean Copulas, with a Characterization of Tail Dependence

Pasquale Cirillo[1], Andrea Fontanari[2], Kees Oosterlee[3]
TU Delft, Netherlands[1,2], CWI, Netherlands[3]
p.cirillo@tudelft.nl[1], a.fontanari@tudelft.nl[2], c.w.oosterlee@cwi.nl[3]

We introduce a new generating mechanism for non-strict Archimedean copulas via the Lorenz curve of a positive random variable. Lorenz curves have been extensively studied in economics and statistics, as well as in insurance and finance, to characterize wealth inequality and to derive quantities like the Value-at-Risk. In our approach, such a curve is seen as an integral transform generating increasing convex functions in the unit square. We show that many of the properties of the copulas we generate, from tail dependence and stochastic ordering to their Kendall distribution function and the size of the singular part, depend on simple features of the random variable associated to the Lorenz transform. We show for example that by selecting random variables with lower bound at zero it is possible to generate copulas with near asymptotic dependence, while the log-concavity of their cumulative distribution function plays a major role in the generation of tail dependence. Finally, we discuss how an alchemy of the Lorenz curves can be used as a general framework to build multi-parametric families of copulas.

Session, time and room:
Copula based methods
Tuesday, 9:50 – 10:10 (A102)

Extremes in the climate system – definition, description and prediction of extreme weather and climate events

Petra Friederichs[1]
University of Bonn, Germany[1]
pfried@uni-bonn.de[1]

The talk gives an overview of meteorological extremes and discusses challenges with respect to the definition, description and prediction of weather and climate extremes. The complexity of the topic issues from the complexity of the climate system as well as the very different perspectives approaching weather and climate extremes. Its description and prediction requires a rigorous statistical component, and thus the talk takes a look at the role of extreme value statistics therein.

The climate system and its subsystems are considered as a high-dimensional, non-linear, open physicochemical system far away from thermodynamic equilibrium. As such they are

predestined to generate large fluctuations and exhibit very different types of extremes occurring on a large variety of scales in space and time. Physical constraints such as energy or momentum conservation shape the characteristics of extremes and their changes in time. However, the definition of extremes not only relies on physical parameters, but extreme weather and climate events generally are related to and thus often defined through their large socio-economic impacts, and hence their high relevance in nature and society. Further, climate change is mainly perceived in the (supposed or real) increasing number of extreme events.

Session, time and room:
IS - Extremes and climate physics (Fougeres)
Tuesday, 9:35 – 10:05 (003)

Contributions to Measures of Extremal Dependence

Timo Fuller[1]
Wuerzburg University, Germany[1]
timo.fuller@uni-wuerzburg.de[1]

A big motivation to do multivariate extreme value theory are scenarios, where the joint appearance of extreme events is more likely than anticipated with regular models. Any attempt to quantify this phenomenon ends with something that falls under the broad term 'measure of extremal dependence'.

The talk starts with some theory on random vectors in the max-domain of attraction of a max-stable random vector. After that the contributions fall into two categories: Contributions on what those quantities they can be used for - including a very generatal point on the importance of the Huessler–Reiss, distribution, the finite dimensional version of the Brown–Resnick process. And contributions on how those quantities can be estimated - including a very general point on how univariate extreme value theory contributes to our estimation efforts in the multivariate case.

Session, time and room:
OCS - Extremal dependence modelling (Padoan)
Tuesday, 11:15 – 11:35 (A101)

Concentration of maxima and the fundamental limits of exact support recovery in high dimensions

Zheng Gao[1], Stilian Stoev[2]
University of Michigan, United States[1,2]
gaozheng@umich.edu[1], sstoev@umich.edu[2]

We study the estimation of the support (set of non-zero components) of a high dimensional signal observed with additive and dependent noise. With the usual parameterization of the size of the support set and the signal magnitude, we characterize a phase-transition phenomenon akin to the Ingster's signal detection boundary. Namely, when the signal is above the so-called strong classification boundary, thresholding estimators achieve asymptotically perfect support recovery. This is so under arbitrary error dependence assumptions, provided that the marginal error distribution has rapidly varying tails. Conversely, under mild dependence conditions on the noise, no thresholding estimators can achieve perfect support recovery if the signal is below the boundary. For log-concave error densities, the thresholding estimators are shown to be optimal and hence the strong classification boundary is universal.

The phase-transition results stem from a certain *concentration of maxima* phenomenon known as relative stability. We provide a complete characterization of the relative stability phenomenon for Gaussian triangular arrays in terms their correlation structure. The proof uses classic Sudakov-Fernique and Slepian lemma arguments along with a curious application of Ramsey's coloring theorem.

Session, time and room:
Best student paper III
Tuesday, 16:40 – 17:00 (A101)

Nonparametric confidence interval for conditional quantiles with large-dimensional covariates

Laurent Gardes[1]
Université de Strasbourg, France[1]
gardes@unistra.fr[1]

The first contribution of this work is the definition of a nonparametric confidence interval for conditional quantile when the quantile level depends on the sample size. When this level converges to 0 or 1 as the sample size increases, the conditional quantile is said to be extreme. The method used to build this interval is based on ordered statistics selected by a nearest neighbor approach. For large-dimensional covariates, the coverage probability on this confidence interval can be very different from the nominal one. Our second contribution is the proposition of a new procedure to find the best distance to use for the selection of the nearest neighbors in order to obtain a better coverage probability. This procedure is based on the Tail Conditional Independence assumption introduced by (Gardes, *Extreme*, pp. 57–95, **18(3)**, 2018).

Session, time and room:

IS - Extremes and machine learning (Sabourin)
Monday, 16:50 – 17:20 (003)

A new link function to predict the extreme values

Gloria Gheno[1]
Innovative data analysis, Italy[1]
gloriagheno@libero.it[1]

In the field of marketing it is fundamental to analyze which are the factors that influence the forecast of the launch of a new product with particular characteristics. Therefore, it is essential to understand the differences between a new product and one already on the market, but only improved. The consumer, generally due to a lack of information, can forecast the launch of a new improved product but hardly that of a new product, if not widely announced. Therefore, the prediction about the satisfaction obtainable from the new product is less certain for the producer. Consequently, the rare values are the most interesting for the producing company because they determine very specific customer targets. In order to remain competitive and eliminate additional costs, it is useful to study how potential customers perceive the probability of the exit of new or improved products. The possession of previous versions of the product, the information leaked about the innovations proposed by the market and the personal characteristics of the potential customer are some of the variables which influence the forecast. To realize this analysis, I propose a Beta regression associated with a new link function, consisting of a part linked to the generalized extreme value link, proposed in the literature initially for binary variables, and of a new part created specifically to eliminate the monotonicity of the link function. This new method is applied to a real dataset and, to evaluate the real goodness, I compare it with other methods already present in the literature using specific statistical tests. The new link function, built to solve the problem of the extreme values, can also be applied in other situations where the monotonicity of the link function is not certain.

Session, time and room:
Product quality
Thursday, 9:00 – 9:20 (A102)

Causal inference in heavy-tailed models

Sebastian Engelke[1], Nicola Gnecco[2], Nicolai Meinshausen[3], Jonas Peters[4]
University of Geneva, Switzerland[1,2], ETH Zurich, Switzerland[3], University of Copenhagen, Denmark[4]
sebastian.engelke@unige.ch[1], nicola.gnecco@unige.ch[2],
meinshausen@stat.math.ethz.ch[3], jonas.peters@math.ku.dk[4]

We study linear structural equation models on directed acyclic graphs (DAGs) where the noise variables have heavy-tailed distributions. Intuitively, a causal effect between two variables should be strongly visible in their extreme values. We therefore define a coefficient that captures asymmetries in the bivariate extremal dependence. In the population case, the causal coefficient reveals the causal structure even in the presence of confounders that have the same tail index as the observed variables. For observational data, the matrix containing empirical estimates of the coefficients for all pairs of nodes in the DAG is the basis for a causal discovery algorithm that consistently recovers the causal order. We compare our method to other well-established approaches in causal inference such as the PC algorithm and LiNGAM. This is joint work with Sebastian Engelke, Nicolai Meinshausen and Jonas Peters.

Session, time and room:
Best student paper I
Monday, 16:40 – 17:00 (A002)

Robust nonparametric estimation of the conditional tail dependence coefficient

Yuri Goegebeur[1], Armelle Guillou[2], Nguyen Khanh Le Ho[3], Jing Qin[4]
University of Southern Denmark, Denmark[1,3,4], Strasbourg University, France[2]
yuri.goegebeur@imada.sdu.dk[1], armelle.guillou@math.unistra.fr[2], ho@imada.sdu.dk[3],
qin@imada.sdu.dk[4]

We consider robust and nonparametric estimation of the coefficient of tail dependence in presence of random covariates. The estimator is obtained by fitting the extended Pareto distribution locally to properly transformed bivariate observations using the minimum density power divergence criterion. We establish convergence in probability and asymptotic normality of the proposed estimator under some regularity conditions, taking into account that the unknown marginal conditional distribution functions are estimated nonparametrically. The finite sample performance is evaluated with a small simulation experiment, and the practical applicability of the method is illustrated on a dataset of air pollution measurements.

Session, time and room:
Tail dependence
Thursday, 11:40 – 12:00 (A001)

Improvements on robust and corrected-bias estimation of the extremal index

M. Ivette Gomes[1], Cristina Miranda[2], Manuela Souto de Miranda[3]
CEAUL and DEIO, Universidade de Lisboa, Portugal[1], Universidade de Aveiro and CIDMA, Portugal[2,3]
ivette.gomes@fc.ul.pt[1], cristina.miranda@ua.pt[2], manuela.souto@ua.pt[3]

Statistical methods of analysis of extremes of dependent data are crucial in the most diverse fields of application. And a recurrent issue that appears when there is dependency is the reliable estimation of the *extremal index* (EI), a parameter related to the clustering of large events. The most popular EI-estimators, like the blocks' [3] and sliding blocks' [5] EI-estimators are very sensitive to anomalous cluster sizes and/or to the number of upper order statistics involved in the estimation. Hence the need for robust versions of such EI-estimators, the main topic under discussion, together with the use of robust versions of slight variations of the generalized jackknife bias-corrected EI-estimators in [2]. Also, different generalized means have been recently used in a successful estimation of the extreme value index, among other parameters of univariate extreme events (see [1] and [4], among other articles on the topic), and will now be discussed for an EI-estimation.

References

[1] F. Caeiro, M.I. Gomes, J. Beirlant and T. de Wet, Mean-of-order p reduced-bias extreme value index estimation under a third-order framework, *Extremes* **19**:4, 561–589, 2016.

[2] M.I. Gomes, A. Hall and C. Miranda, Subsampling techniques and the Jackknife methodology in the estimation of the extremal index, *Comput. Statist. & Data Anal.* **52**:4, 2022–2041, 2008.

[3] T. Hsing, Estimating the parameters of rare events, *Stoch. Proc. and Appl.* **37**, 117–139, 1991.

[4] H. Penalva, M.I. Gomes, F. Caeiro and M.M. Neves, A couple of non reduced bias generalized means in extreme value theory: an asymptotic comparison, *Revstat: Statistical J*, in press. Available at: `https://www.ine.pt/revstat/pdf/Acoupleofnonreducedbiasgeneralizedmeansinextremevaluetheory.pdf`, 2018.

[5] C.Y. Robert, J. Segers and C.A.T. Ferro, Sliding blocks estimator for the extremal index, *Electr. J. of Statist.* **3**, 993–1020, 2009.

Session, time and room:
OCS - Inferences on extremal index (Markovich)
Tuesday, 14:00 – 14:20 (A101)

Asymmetric Extremal Dependence Modeling, with Application to Cryptocurrency Market Data

Yan Gong[1], Raphaël Huser[2]
King Abdullah University of Science and Technology (KAUST), Saudi Arabia[1,2]
yan.gong@kaust.edu.sa[1], raphael.huser@kaust.edu.sa[2]

In order to jointly assess the asymmetric risks related to both low and high extremal events, we develop a flexible copula model that is able to distinctively capture asymptotic dependence or independence in its lower and upper tails. Our proposed model is parsimonious and smoothly bridges both extremal dependence classes in the interior of the parameter space. The inference can be performed using the full likelihood or various types of censored likelihood approaches. Specifically, we implement three different censoring schemes that are designed to provide a good fit in the lower and upper joint tail regions and we investigate their relative efficiency in an extensive simulation study. We illustrate our methodology by studying the dependence strength that governs extreme log-returns in cryptocurrency market data. Our analysis shows that our model provides a better fit than alternative copula models, and reveals that the joint lower tail of Bitcoin and Ethereum cryptocurrencies has become increasingly dependent in the past two years, hence suggesting that big simultaneous losses are now more frequent than before.

Session, time and room:
Copula based methods
Tuesday, 10:15 – 10:35 (A102)

Robust estimation of the Pickands dependence function under random right censoring

Yuri Goegebeur[1], Armelle Guillou[2], Jing Qin[3]
University of Southern Denmark, Denmark[1,3], Strasbourg University and CNRS, France[2]
Yuri.Goegebeur@imada.sdu.dk[1], armelle.guillou@math.unistra.fr[2], qin@imada.sdu.dk[3]

We consider robust nonparametric estimation of the Pickands dependence function under random right censoring. The estimator is obtained by applying the minimum density power divergence criterion to properly transformed bivariate observations. The asymptotic properties are investigated by making use of results for Kaplan-Meier integrals. We investigate the finite sample properties of the proposed estimator with a simulation experiment and illustrate its practical applicability on a dataset of insurance indemnity losses.

Session, time and room:
IS - Advances in statistics of multivariate extremes (Einmahl)
Monday, 16:50 – 17:20 (A001)

Computing return levels in the context of a changing climate: how to deal with uncertainty and non-stationarity?

Elyse Fournier[1], Alexis Hannart[2], David Huard[3], Luc Perreault[4], Gabriel Rondeau-Genesse[5], Richard Turcotte[6]

Ouranos, Canada[1,2,3,5], Hydro-Québec, Canada[4], Ministère de l'Environnement et Lutte contre les changements climatiques, Canada[6]

fournier.elyse@ouranos.ca[1], hannart.alexis@ouranos.ca[2], huard.david@ouranos.ca[3], perreault.luc@ireq.ca[4], Rondeau-Genesse.Gabriel@ouranos.ca[5], richard.turcotte2@mddep.gouv.qc.ca[6]

In Southeastern Canada, climate change has two opposite effects on Spring flood hazard: an increase in Spring rainfall and a decrease in snow cover. The intensity and even the sign of the combined effect varies across space and time, and is uncertain due to observational and climate modeling imprecisions. The provincial government of Québec requests an update of the maps of flood hazard in use for regulatory purpose, which takes climate change into account. Likewise, hydroelectric power companies in Québec wish to reevaluate the compliance of their infrastructures in light of potentially changing flood return levels. Such practical requests, like many similar ones in the context of decision-making for adaptation to climate change, critically raises two theoretical questions: (i) how to deal with uncertainty – for instance, is the concept of posterior predictive suitable? (ii) how to deal with non-stationarity in the definition and computation of return levels associated to a return period? This talk will present a few ideas and results to address these two questions.

Session, time and room:
IS - Detection and attribution of climate change (Cooley)
Thursday, 14:00 – 14:30 (003)

A semiparametric Bayesian model for spatiotemporal extremes

Arnab Hazra[1], Brian Reich[2], Benjamin Shaby[3], Ana-Maria Staicu[4]

King Abdullah University of Science and Technology, Saudi Arabia[1], North Carolina State University, United States[2,4], Pennsylvania State University, United States[3]

arnab.hazra@kaust.edu.sa[1], bjreich@ncsu.edu[2], bshaby@psu.edu[3], astaicu@ncsu.edu[4]

In this paper, we consider a Dirichlet process mixture of spatial skew-t processes that can flexibly model the extremes as well as the bulk, with separate parameters controlling spatial dependence in these two parts of the distribution. The proposed model has nonstationary mean and covariance structure and also nonzero spatial asymptotic dependence. Max-stable processes are theoretically justified model for station-wise block maximums or threshold exceedances in the spatial extremes literature. Considering a high threshold leads to somewhat arbitrary decision about what counts as extreme, and more importantly, it disallows the possibility that events that are large but deemed insufficiently extreme can enter into the analysis at all. Probabilistic clustering of the extreme observations and allowing extremal dependence for the cluster of extremes is a solution that is explored here. Inference is drawn based on Markov chain Monte Carlo sampling. A simulation study demonstrates that the proposed model has better spatial prediction performance

compared to some competing models. We develop spatial maps of extreme Fosberg Fire Weather Index (FFWI), a fire threat index and discuss the wildfire risk throughout the Santa Ana region of California.

Session, time and room:
Bayesian methods for environmental extremes
Monday, 10:55 – 11:15 (A002)

Statistical Inference for a Relative Risk Measure

Yi He[1], Yanxi Hou[2], Liang Peng[3], Jiliang Sheng[4]
Monash University, Australia[1], Fudan University, China[2], Georgia State University, United States[3], Jiangxi University of Finance and Economics, China[4]
yi.he2@monash.edu[1], oishihou@yahoo.com[2], lpeng@gsu.edu[3], shengjiliang@163.com[4]

For monitoring systemic risk from regulators' point of view, this paper proposes a relative risk measure, which is sensitive to the market comovement. The asymptotic normality of a nonparametric estimator and its smoothed version is established when the observations are independent. In order to effectively construct an interval without complicated asymptotic variance estimation, a jackknife empirical likelihood inference procedure based on the smoothed nonparametric estimation is provided with a Wilks type of result in case of independent observations. When data follow from AR-GARCH models, the relative risk measure with respect to the errors becomes useful and so we propose a corresponding nonparametric estimator. A simulation study and real-life data analysis show that the proposed relative risk measure is useful in monitoring systemic risk.

Session, time and room:
IS - Advances in statistics of multivariate extremes (Einmahl)
Monday, 17:25 – 17:55 (A001)

How to model clustered and bursty extremes?

Katharina Hees[1]
TU Dortmund, Germany[1]
hees@statistik.tu-dortmund.de[1]

Not only the strength of extreme weather events but also the questions: "How often have we to fear its appearance?" or "How likely is it that there will be an extreme weather event in a fixed period?" are of great importance. The traditional approach is to assume that a Poisson process can model the occurrence of the extreme events, which means that the times between subsequent exceedances are exponentially distributed. The underlying time series often displays short-term dependence, which leads to a temporal clustering of the extreme events. Nowadays, inference is then commonly based on the identification of independent clusters by some declustering mechanism, assuming the occurrence of clusters to follow a Poisson process with exponential recurrence times in between subsequent

clusters. However, not only the assumption of serial correlation in the underlying time series, but also heavy tailed inter-arrival times or dependencies in the arrivals can lead to clustering of extremes and hence to different distributions of the return times. In this talk, I will give a short overview of different approaches, talk about how to model the inter-arrival times between extreme weather events and present ideas how to choose an appropriate model.

Session, time and room:
Models with clusters
Thursday, 16:15 – 16:35 (A002)

Extreme value analysis of high-dimensional Kendall's Tau and Spearman's Rho correlation matrices

Holger Dette[1], Johannes Heiny[2], Thomas Mikosch[3], Jianfeng Yao[4]

Ruhr University Bochum, Germany[1,2], University of Copenhagen, Denmark[3], University of Hong Kong, China[4]

holger.dette@rub.de[1], johannes.heiny@rub.de[2], mikosch@math.ku.dk[3], jeffyao@hku.hk[4]

In many modern applications it is essential to accurately assess the dependence structure of a multivariate time series. If the dimension p and the number of observations n from this time series are of a similar order, the sample (auto)covariance matrices will typically be misleading. Even in the null case, i.e. when the components of the time series are iid, it is well-known that the sample covariance matrix S poorly estimates the population covariance matrix.

Aside from the high dimension, the marginal distributions of the components present another major challenge for an accurate assessment of the dependence. In the literature, one typically assumes a finite fourth moment since otherwise the largest eigenvalue of the sample covariance matrix would tend to infinity when n and p increase. This moment assumption, however, excludes heavy-tailed time series from the analysis. The theory for the eigenvalues and eigenvectors of the sample autocovariance matrices stemming from such time series is quite different from the classical Marchenko–Pastur theory which applies in the light-tailed case.

To gain insight into the dependence structure of possibly heavy-tailed time series, we provide asymptotic theory for the sample Pearson, Spearman and Kendall correlation matrices. Under heavy tails the Pearson correlation matrix R turns out to be more robust than the sample covariance matrix S. In the case of time series with infinite $(2-\epsilon)$-moments, a new class of Marchenko–Pastur type laws appears as limiting spectral distributions of R.

Being distribution-free measures, the sample Spearman rho correlation matrix $\mathcal{S}$ and the Kendall tau correlation matrix K capture the dependence of the underlying time series without being affected by the marginal distributions of its components. We conduct an extreme value analysis to derive the limiting distributions of the largest and smallest off-diagonal entries of $S, R, \mathcal{S}$ and K under a wide range of growth rates on the dimension p. In particular, the largest entries of all 4 matrices turn out to be Gumbel distributed. Based on this precise asymptotic result, we identify the optimal threshold levels to construct asymptotically unbiased (in spectral norm) estimators of the population matrices $E[S], E[R], E[\mathcal{S}]$ and $E[K]$.

Session, time and room:
High-dimensional data
Thursday, 16:40 – 17:00 (003)

Focusing on regions of interest in forecast evaluation

Hajo Holzmann[1], Bernhard Klar[2], Tobias Zwingmann[3]
Marburg University, Germany[1,3], KIT, Germany[2]
holzmann@mathematik.uni-marburg.de[1], Bernhard.Klar@kit.edu[2],
zwingmann@Mathematik.Uni-Marburg.de[3]

Often, interest in forecast evaluation focuses on certain regions of the whole potential range of the outcome, and forecasts should mainly be ranked according to their performance within these regions. A prime example is risk management, which relies on forecasts of risk measures such as the value-at-risk or the expected shortfall and hence requires appropriate loss distribution forecasts in the tails. Further examples include weather forecasts with a focus on extreme conditions, or forecasts of environmental variables such as ozone with a focus on concentration levels with adverse health effects. In this talk we briefly review scoring functions and scoring rules, and present some novel asymptotic results for quantile and expectile processes. Then we show how weighted scoring rules can be used to rank several potentially misspecified forecasts objectively with the region of interest in mind. We introduce desirable properties of weighted scoring rules and present general construction principles based on conditional densities or distributions and on scoring rules for probability forecasts. In our empirical application to log-return time series all forecasts seem to be slightly misspecified, as is often unavoidable in practice, and no method performs best overall. However, using weighted scoring functions the best method for predicting losses can be identified, which is hence the method of choice for the purpose of risk management.

References

[1] Holzmann, H., Klar, B. Focusing on regions of interest in forecast evaluation. Annals of Applied Statistics 2017, 11, 2404-2431.

[2] Zwingmann, T., Holzmann, H. Weak convergence of quantile and expectile processes under general assumptions. 2019

Session, time and room:
IS - Predicting extremes (Segers)
Monday, 11:05 – 11:35 (003)

Sub-asymptotic modeling of spatial extremes based on max-infinitely divisible processes

Raphaël Huser[1]
King Abdullah University of Science and Technology (KAUST), Saudi Arabia[1]
raphael.huser@kaust.edu.sa[1]

Understanding the spatial extent of extreme precipitation is necessary for determining flood risk and adequately designing infrastructure (e.g., stormwater pipes) to withstand such hazards. While environmental phenomena typically exhibit weakening spatial dependence at increasingly extreme levels, limiting max-stable process models for block maxima have a rigid dependence structure that does not capture this type of behavior. In order to model block maxima at sub-asymptotic regimes, we suggest using models from a broader family of max-infinitely divisible (max-id) processes, which retain appealing properties reflecting the specific dependence structure of maxima, while allowing for weakening spatial dependence at increasingly high levels. In this talk, I will first present general construction principles for max-id processes and discuss how flexible asymptotically independent max-id models may be designed. I will then propose a Bayesian (conditionally) max-id process, whose likelihood function admits a hierarchical representation in terms of random effects, and which scales well to large datasets. The proposed model is constructed using flexible random basis functions that are estimated from the data, allowing for straightforward inspection of the predominant spatial patterns of extremes. In addition, the described process possesses (conditional) max-stability as a special case, making inference on the asymptotic dependence class possible. We apply our model to extreme precipitation in eastern North America, and show that the proposed model adequately captures the extremal behavior of the data. This talk is based on joint work with Thomas Opitz, Emeric Thibaud, Gregory P. Bopp, and Benjamin A. Shaby.

Session, time and room:
IS - Sub-asymptotic spatial extremes (Opitz)
Tuesday, 14:00 – 14:30 (A001)

Lévy processes: Failure to observe threshold exceedance over a dense grid.

Krzysztof Bisewski[1], Jevgenijs Ivanovs[2]
CWI, Netherlands[1], Aarhus University, Denmark[2]
K.L.Bisewski@cwi.nl[1], Jevgenijs.ivanovs@math.au.dk[2]

For a general Lévy process X on a finite time interval, consider the probability of X exceeding some threshold $x > 0$ while being below x at the points of a uniform grid. We find exact asymptotics of this probability as the number of grid points tends to infinity. This result crucially depends on the small-time behaviour of the given Lévy process and, in particular, on the limit process when zooming-in. No advanced knowledge of Lévy processes will be assumed in this talk.

Session, time and room:

IS - Model assessment in spatial extremes (Engelke)
Thursday, 9:00 – 9:30 (A002)

Estimation of the tail index by the method of block quantiles

Adam Jakubowski[1], Wojciech Rejchel[2], Patryk Truszczyński[3]
Nicolaus Copernicus University, Poland[1,2,3]
adjakubo@mat.umk.pl[1], iggypop@mat.umk.pl[2], 502121@doktorant.umk.pl[3]

We use the ideas of [1] to construct an estimator of the tail index, which is based on the block quantiles. The estimator performs very well in simulations and is much more stable comparing to the commonly used Hill estimator. Consistency as well as asymptotic normality are obtained under particular regimes.

References

[1] P. Doukhan, A. Jakubowski and G. Lang, Phantom distribution functions for some stationary sequences, *Extremes*, Vol. 18 (2015) , 697-725.

Session, time and room:
Tail index estimation
Monday, 10:55 – 11:15 (A101)

On Binary Classification in Extreme Regions

Stéphan Clémençon[1], Hamid Jalalzai[2], Anne Sabourin[3]
Telecom Paristech, France[1,2,3]
stephan.clemencon@telecom-paristech.fr[1], hamid.jalalzai@gmail.com[2],
anne.sabourin@telecom-paristech.fr[3]

In a wide variety of applications involving anomaly detection (e.g. buzzes in social network data, frauds, system failures), extreme observations play a key role because anomalies often correspond to large observations. The key issue is then to distinguish between large observation from the normal class and large observations from the anomaly class. This task can thus be formulated as a binary classification problem in extreme regions. However, extreme observations generally contribute in a negligible manner to the (empirical) error, simply because of their rarity. As a consequence, empirical risk minimizers generally perform very poorly in extreme regions. This paper develops a general framework for classification of extreme values. Precisely, under non-parametric heavy-tail assumptions, we propose a natural and asymptotic notion of risk accounting for predictive performance in extreme regions. We prove that minimizers of an empirical version of this dedicated risk lead to classification rules with good generalization capacity, by means of maximal deviation inequalities in low probability regions. Numerical experiments illustrate the relevance of the approach developed.

Session, time and room:
Statistical learning methods
Tuesday, 16:15 – 16:35 (003)

Interpolation of extreme precipitation of multiple durations in Eastern Canada

Christian Genest[1], Jonathan Jalbert[2], Luc Perreault[3]
McGill University, Canada[1], Polytechnique Montréal, Canada[2], Institut de recherche d'Hydro-Québec, Canada[3]
christian.genest@mcgill.ca[1], jonathan.jalbert@polymtl.ca[2], perreault.luc@ireq.ca[3]

Intensity-duration-frequency (IDF) curves for extreme precipitation events are widely used to design civil infrastructures like sewers and dikes. IDF curves consists in a graphical representation precipitation return levels (from 2 to 100 years) for different durations of precipitation (from 5min to 24h). In Canada, those curves are provided by Environment Canada for every meteorological station under their authority. An important issue raised by the Canadian Standards Association concerns the sparsity of the stations network. For example, in the province of Québec, the large majority of the stations are located in the south, where most of the inhabitant lives but many important hydrological structures are located in the North. Generally, for designing purposes, the IDF curve from the closest station is used, even if it is located hundreds of kilometers away. The uncertainty according to the spatial variability of extreme precipitation is then not accounted for.

The goal of the present work consists in properly interpolating the extreme precipitation of several durations in order to compute the IDF curves everywhere in Eastern Canada. In order to gather physical information where no observation is available, a reconstruction of the historical meteorology is used as the covariate for interpolating extreme precipitation characteristics. Such a covariate is included in a hierarchical Bayesian spatial for the extreme precipitation. This spatial model is especially suited for the covariate gridded structure, hence enabling fast and precise computations using Gaussian Markov random fields. This model provides reliable IDF curves over the whole spatial domain as it will be demonstrated by the cross-validation results. Hydro-Québec, which is the principal electricity producer in Québec, is now considering these IDF curves to revise their design criteria of different hydraulic structures and other assets managed by the company.

Session, time and room:
IS ST - Hydrology and extremes (Naveau & Rootzen)
Friday, 9:00 – 9:30 (003)

Scoring Predictions at Extreme Quantiles

Axel Gandy[1], Kaushik Jana[2], Almut Veraart[3]
Imperial College London, United Kingdom[1,3], Imperial College London and the Alan Turing Institute, United Kingdom[2]
a.gandy@imperial.ac.uk[1], k.jana@imperial.ac.uk[2], a.veraart@imperial.ac.uk[3]

Prediction of extreme quantiles lying outside the range of historical data is of interest in numerous applications. Extreme value theory provides various competing models through different assumptions on tail behaviors of the underlying distribution. Assessment of competing estimates based on their predictive performance is a useful way to select the best estimate in a given situation. However, due to extreme nature of this inference problem, it might be very well possible that, the predicted quantiles are not seen in the historical records. Therefore, making it challenging to validate the prediction with its realization. In this article, we propose a non-parametric scoring method for high quantile estimates by predicting a sequence of equally extremal quantiles on different parts of the data. We then use cross validation approach and covariate information for scoring the competing estimates. The performance of the scoring method is assessed in a simulation study that mimics the practical application at hand. This methodology is applied to motivating netflow data available from the Los Alamos National Laboratory.

Session, time and room:
Prediction
Tuesday, 16:15 – 16:35 (A102)

Regular variation and complexity reduction of extremes

Anja Janßen[1], Phyllis Wan[2]
KTH Royal Institute of Technology, Stockholm, Sweden[1], Erasmus University Rotterdam, Netherlands[2]
anjaj@kth.se[1], wan@ese.eur.nl[2]

Multivariate regular variation and its implied properties belong to the most important tools for the analysis of the extremal behavior of random vectors. The key result of the homogeneity of the limiting measure ν, for which $P(\mathbf{X}/t \in \cdot)/P(\|\mathbf{X}\| > t) \xrightarrow{v} \nu(\cdot)$, allows us to describe the extremal dependence structure of a random vector by focussing on the spectral measure S on $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, for which $P(\mathbf{X}/\|\mathbf{X}\| \in \cdot \mid \|\mathbf{X}\| > t) \xrightarrow{w} S(\cdot)$.

One can interpret this as a dimension reduction result, allowing us to describe the extremal dependence structure of a d-dimensional vector by a $(d-1)$-dimensional quantity. However, as the value of d grows, further methods might be needed to reduce the complexity of the spectral measure for practical applications. In this talk, we present a clustering approach that was recently introduced in [1]. We show how a spherical k-means clustering on an empirical counterpart of S can help us to identify "extremal prototypes" and derive consistency of our estimators. It is demonstrated that the procedure can be seen as an alternative to max-linear models if one assumes that the spectral measure is concentrated around few points of mass. Finally, we illustrate the procedure with the help of several data examples.

References

[1] A. Janßen and P. Wan, k-means clustering of extremes, *arXiv preprint*, arXiv:1904.02970, 2019.

Session, time and room:
IS ST - Regular variation: generalisations and recent advances (Molchanov & Basrak)
Thursday, 9:35 – 10:05 (003)

Extremal Markov chains driven by the Kendall convolution

Marek Arendarczyk[1], Barbara Jasiulis-Gołdyn[2], Jolanta Misiewicz[3], Karolina Naskręt[4], Edward Omey[5], Mateusz Staniak[6]

University of Wrocław, Poland[1,2,4], Warsaw University of Technology, Poland[3,6], KU Leuven, Belgium[5]

marendar@math.uni.wroc.pl[1], jasiulis@math.uni.wroc.pl[2], j.misiewicz@mini.pw.edu.pl[3], karolina.lukaszewicz@math.uni.wroc.pl[4], edward.omey@kuleuven.be[5], m.staniak@mini.pw.edu.pl[6]

We consider the Kendall convolution algebra world $(\mathcal{P}_+, \triangle_\alpha)$ such that the Kendall result of two deltas is continuous time probability measure $\delta_1 \triangle_\alpha \delta_1 = \pi_{2\alpha}$, where $\pi_{2\alpha}(dx) = 2\alpha x^{-2\alpha-1}\mathbf{1}_{[1,\infty)}(x)dx$ is density of the Pareto distribution.

The main goals of this talk are focused on the renewal theory ([4]), some aspects of the fluctuation theory ([5]) and asymptotic properties ([1]) for the extremal Markov sequences of the Kendall type (called also Kendall random walks [3]), which are discrete time Lévy processes under the Kendall convolution ([2]), defined in the following way:

$$X_0 = 1, \quad X_1 = Y_1, \quad X_{n+1} = M_{n+1}\left[\mathbf{1}(\xi_n > \varrho_{n+1}) + \theta_{n+1}\mathbf{1}(\xi_n < \varrho_{n+1})\right],$$

where

$$M_{n+1} = \max\{X_n, Y_{n+1}\}, \quad \varrho_{n+1} = \frac{\min\{X_n, Y_{n+1}\}^\alpha}{\max\{X_n, Y_{n+1}\}^\alpha},$$

and sequences $(Y_k) \sim i.i.d.(\nu)$, $(\xi_k) \sim i.i.d.(U([0,1]))$, $(\theta_k) \sim i.i.d.$ $(\pi_{2\alpha})$ are independent.

We show that one dimensional distributions of Kendall random walks are regularly varying ([1]) and prove finite-dimensional convergence for continuous time stochastic processes based on Kendall random walks. We construct renewal processes based on the extremal Markov sequences and investigate limit behaviour of the corresponding renewal function (for details see [4]). Limit Theorems for Kendall random walks and corresponding renewal processes ([1], [4]) will be showed using regular variation and Williamsom transform.

This work is a part of project "First order Kendall maximal autoregressive processes and their applications", which is carried out within the POWROTY/REINTEGRATION programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

References

[1] M. Arendarczyk, B.H. Jasiulis-Gołdyn, E.A.M. Omey, Asymptotic properties of extremal Markov chains of Kendall type, submitted, 2019.

[2] M. Borowiecka-Olszewska, B.H. Jasiulis-Gołdyn, J.K. Misiewicz, J. Rosiński, Lévy processes and stochastic integral in the sense of generalized convolution, *Bernoulli*, 21(4), 2513-2551, 2015.

[3] B.H. Jasiulis-Gołdyn, Kendall random walks, *Probab. Math. Stat.*, **36**(1), 165-185, 2016.

[4] B.H. Jasiulis-Gołdyn, K. Naskręt, J.K. Misiewicz, E.A.M. Omey, Renewal theory for extremal Markov sequences of the Kendall type, submitted, arXiv: https://arxiv.org/pdf/1803.11090.pdf, 2018.

[5] B.H. Jasiulis-Gołdyn, M. Staniak, Fluctuations of Markov chains of the Kendall type, submitted, https://arxiv.org/pdf/1902.00576.pdf, 2019.

Session, time and room:
Limit theorems
Monday, 14:25 – 14:45 (A101)

General linear combination of log-exceedances as tail index estimator

Jelena Jocković[1], Pavle Mladenović[2]
Faculty of Mathematics, University of Belgrade, Serbia[1,2]
jjocko@matf.bg.ac.rs[1], paja@matf.bg.ac.rs[2]

We investigate the class of estimators of the tail index $\gamma > 0$ defined as:

$$\gamma^*_{n,k} = \sum_{j=1}^{k} c_j \ln \frac{X_{n,n-j+1}}{X_{n,n-k}}$$

where

$$c_j \geq 0 \text{ for } j \in \{1, 2, \ldots, k\}, \text{ and, optionally, } c_1 + c_2 + \cdots + c_k = 1.$$

The class contains well known Hill estimator, and some other estimators proposed in the literature. We provide theoretical result related to asymptotic mean squared error of such estimators and, based on this result, indicate several members of the class with good asymptotic properties. Finally, we investigate finite-sample properties of the proposed estimators through detailed simulation study.

Session, time and room:
Tail index estimation
Monday, 11:20 – 11:40 (A101)

Efficient simulation of tail probabilities for subexponential sums with dependent random weights

Nicholas Johnson[1]
University of Queensland, Australia[1]
n.c.johnson@uq.net.au[1]

This paper deals with the problem of efficiently estimating extreme tail probabilities for the sum of n dependent and non-identically-distributed random variables $H_1, \ldots, H_n$ where each summand $H_i = \Theta_i X_i$ is the product of a subexponential random variable X_i and a random weight Θ_i. Motivated by applications to insurance problems, the sequence of random weights $\Theta_1, \ldots, \Theta_n$ considered here has a recursive dependence structure driven by an underlying sequence of iid subexponential random variables $R_1, \ldots, R_n$. A novel Asmussen–Kroese-type estimator for the tail probabilities is presented, along with numerical results which suggest the estimator enjoys good efficiency properties.

Session, time and room:
Computationally intensive methods
Thursday, 11:15 – 11:35 (A101)

Extreme rank estimator of the slopes

Jana Jureckova[1]
Institute of Information and Automation, and Charles University, Prague, Czech Republic[1]
jurecko@karlin.mff.cuni.cz[1]

We consider the family of R-estimators of the slope components $\boldsymbol{\beta}$ with nuisance intercept β_0 in the linear regression model $Y_i = \beta_0 + \mathbf{x}_{ni}^\top \boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n$ with independent errors $e_1, \ldots, e_n$, identically distributed according to an unknown distribution function F. The R-estimators are based on Hájek's rank scores $a_n(i, \alpha),\ i = 1, \ldots, n$, defined for each $0 \le \alpha \le 1$ as

$$a_n(i,\alpha) = \begin{cases} 0 & \ldots \quad i \le n\alpha \\ i - n\alpha & \ldots \quad n\alpha \le i \le n\alpha + 1 \\ 1 & \ldots \quad i \ge n\alpha + 1 \end{cases} .$$

Denote the matrix $\mathbf{Q}_n = \sum_{i=1}^n (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)^\top$. Then the process

$$\mathbf{T}_n = \left\{ \mathbf{T}_n(\alpha) = \mathbf{Q}_n^{-1/2} \sum_{i=1}^n (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) a_n(R_{ni}, \alpha) : \ \ 0 \le \alpha \le 1 \right\}$$

with R_{ni} being the ranks of $Y_1, \ldots, Y_n$, converges weakly to the vector $\mathbf{W}_p^*$ of p independent Brownian bridges, under some conditions on F and on the regressors $\mathbf{x}_{ni}$. Similarly, under the local alternative $Y_i \sim F(y - \mathbf{x}_{ni}^\top \mathbf{b}),\ i = 1, \ldots, n$

$$\left\{ \mathbf{T}_n(\alpha) - f(F^{-1}(\alpha)) \mathbf{Q}_n^{1/2} \mathbf{b} \ : \ 0 \le \alpha \le 1 \right\} \xrightarrow{\mathcal{D}} \mathbf{W}_p^* \quad \text{as} \ \ n \to \infty.$$

The estimator $\widehat{\boldsymbol{\beta}}_{n\alpha}$ of $\boldsymbol{\beta}$, defined as the minimizer of the Jaeckel rank dispersion, is asymptotically equivalent to a specific rank statistic, in the sense that

$$f(F^{-1}(\alpha))\mathbf{Q}_n^{1/2}(\widehat{\boldsymbol{\beta}}_{n\alpha} - \boldsymbol{\beta}) = \mathbf{Q}_n^{-1/2} \sum_{i=1}^{n} (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) I[R_{ni} \geq n\alpha] + o_P\left(\|\mathbf{Q}_n^{-1/2}\|\right)$$

and the process $\left\{\mathbf{Q}_n^{1/2} f(F^{-1}(\alpha))(\widehat{\boldsymbol{\beta}}_{n\alpha} - \boldsymbol{\beta})\right\}$ converges to the Brownian bridge. This has an impact of the behavior of $\widehat{\boldsymbol{\beta}}_{n\alpha}$, which can be consistent even under $\alpha \uparrow 1$.

Session, time and room:
Limit theorems II
Monday, 16:15 – 16:35 (A101)

Value-at-Risk estimation: A novel GARCH-EVT approach dealing with bias and heteroscedasticity

Hibiki Kaibuchi[1], Yoshinori Kawasaki[2]
SOKENDAI The Graduate University of Advanced Studies, Japan[1], The Institute of Statistical Mathematics, Japan[2]
hibiki.rhett.kaibuchi@gmail.com[1], kawasaki@ism.ac.jp[2]

Although Extreme Value Theory (EVT) is a popular tool for evaluating tail events in many scientific fields such as hydrology and engineering, it has not yet emerged as a dominating tool in financial risk management, i.e. VaR estimation. This is due to the time-varying volatility of financial time series. In order to overcome this problem, McNeil and Frey [3] introduced the GARCH-EVT approach. It is a two-step procedure consisting of GARCH filtering and Peaks-Over-Thresholds method. It should be noted that one drawback of this methodology is that the correction of bias is not thoroughly considered. In this research, we propose a new way, as far as we aware, to estimate conditional VaR considering both bias correction and volatility background based on standard GARCH-EVT approach. For that, we: (i) pre-whiten the financial time series with GARCH(1,1) model for forecasting volatility supported by Jalal and Rockinger [2]; (ii) apply the semi-parametric bias-corrected tail estimators from de Haan *et al.* [1] to standardized normal residuals from the GARCH analysis. The results are illustrated on simulated data and on a financial real dataset such as Dow Jones and FTSE 100. According to our results, the superiority of our approach over the standard GARCH-EVT and EVT without filtering is clearly evidenced when appropriate threshold, that is approximately top 5% to 15% of observations, is selected.

References

[1] de Haan, L., Mercadier, C. and Zhou, C., Adapting extreme value statistics to financial time series: dealing with bias and serial dependence, *Finance and Stochastics*, 2016

[2] Jalal, A. and Rockinger, M, Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data, *Journal of Empirical Finance*, 2008

[3] McNeil, A.J. and Frey, R., Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance*, 2015

Session, time and room:
Financial extremes II
Tuesday, 11:15 – 11:35 (A002)

Omega bankruptcy for different Lévy models

Adam Kaszubowski[1]
University of Wroclaw, Poland[1]
adam.kaszubowski@math.uni.wroc.pl[1]

In this talk, we consider the so-called Omega bankruptcy model, which can be seen as an alternative to the classical approach to ruin of insurance company. In contrast to the classical model, we allow the process to go below level zero, however not further than some fixed level $-d < 0$. In addition, when the process is below zero it can be killed with some intensity function ω. We will show the relations between the Omega model and classical ruin for two important Lévy models, i.e. we consider the Crámer-Lundberg process and the Markov modulated Brownian motion. Furthermore, we will present the results to the dividends problem in this model for more general class of one-sided Markov additive processes. Presentation is based on the [1] and [2].

References

[1] I. Czarna, A. Kaszubowski, S. Li, Z. Palmowski, Fluctuation identities for omega-killed Markov additive processes and dividend problem, *Submitted for publication*, 2018

[2] A. Kaszubowski Omega bankruptcy for different Lévy models, *Silesian Statistical Review*, 2019

Session, time and room:
Ruin probability
Tuesday, 11:15 – 11:35 (A102)

Darling–Erdős theorem for Lévy processes at zero

Peter Kevei[1], David Mason[2]
University of Szeged, Hungary[1], University of Delaware, United States[2]
kevei@math.u-szeged.hu[1], davidm@udel.edu[2]

Let $(X_t)_{t\geq 0}$ be a Lévy process without normal component such that its spectrally positive part $(X_t^+)_{t\geq 0}$ belongs to the domain of attraction of a nonnormal stable law, i.e. $a(t)^{-1}(X_t^+ - c(t))$ converges to a stable law as $t \downarrow 0$ with some centering and norming sequence. Let $\overline{X}_t = \sup_{s\leq t} X_s$ denote the supremum, $m_t = \sup_{s\leq t} \Delta X_s$ the largest jump, and define

$$Y_t = \sup_{t\leq s\leq 1} \frac{\overline{X}_s}{a(s)}, \quad M_t = \sup_{t\leq s\leq 1} \frac{m_s}{a(s)}.$$

We show that (under mild extra assumptions) the scaled maximum process converges to a Fréchet distribution; more precisely for all $x > 0$

$$\lim_{t\downarrow 0} \mathbf{P}\left(M_t\ (-\log t)^{-1/\alpha} \leq x\right) = e^{-x^{-\alpha}}.$$

Assuming that the spectrally negative part does not dominate the process, we show that the limit remains true for the supremum process Y.

The corresponding results for sums of iid random variables with finite second moment was proved by Erdős and Darling [2], while for random variables in the normal domain of attraction of a stable law by Bertoin [1].

References

[1] J. Bertoin. Darling-Erdős theorems for normalized sums of i.i.d. variables close to a stable law. *Ann. Probab.*, 26(2):832–852, 1998.

[2] D. A. Darling, P. Erdős, A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.*, 23, 143–155, 1956.

Session, time and room:
Limit theorems II
Monday, 16:40 – 17:00 (A101)

Consistency of the Hill estimator for time series observed with measurement errors

Mihyun Kim[1], Piotr Kokoszka[2]
Colorado State University, United States[1,2]
mhkim@colostate.edu[1], piotr.kokoszka@colostate.edu[2]

We investigate the asymptotic and finite sample behavior of the Hill estimator applied to data contaminated by measurement or other errors. We show that for all discrete time stochastic models used in practice, whose marginal distributions are regularly varying, the Hill estimator is consistent. Essentially, the only assumption on the errors is that they have lighter tails than the underlying unobservable process. The asymptotic justification however depends on the specific class of models assumed for the underlying unobservable process. We show by means of a simulation study that the asymptotic robustness of the Hill estimator is clearly manifested in finite samples. We further illustrate this robustness by a numerical study of the interarrival times of anomalies in a backbone internet network, the Internet2 in the United States; the anomalies arrival times are computed with a roundoff error.

Session, time and room:
Data contamination
Thursday, 16:15 – 16:35 (A001)

Climate event attribution using multivariate peaks-over-thresholds modelling

Anna Kiriliouk[1], Philippe Naveau[2]
University of Namur, Belgium[1], Laboratoire des Sciences du Climat et l'Environnement (CNRS), France[2]
anna.kiriliouk@unamur.be[1], naveau@lsce.ipsl.fr[2]

Quantifying the human influence on climate change and identifying potential causes is a highly relevant research area which is often referred to as detection and attribution. A common approach is to compare the probability of an extreme event in the factual world to the probability of an extreme event in a counterfactual world, i.e., a world that might have been if no humans would have existed. The event probabilities can be calculated using large scale climate model runs that simulate the evolution of the climate with and without anthropogenic forcings. The Fraction of Attributable Risk (FAR) is defined as the relative ratio of event probabilities in the factual and in the counterfactual world. Estimating the FAR will allow us to quantify the extent to which human activities have increased the risk of occurrence of an extreme event. We propose a model for the FAR based on the multivariate generalized Pareto distribution, i.e., the asymptotic distribution of suitably normalized exceedances over a high threshold. The model is used to quantify the increased risk of an extreme rainfall event in central Europe.

Session, time and room:
IS - Detection and attribution of climate change (Cooley)

Thursday, 14:35 – 15:05 (003)

Risk-sharing of regularly varying claims in bipartite networks

Claudia Klueppelberg[1]
Technical University of Munich, Germany[1]
cklu@ma.tum.de[1]

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 complete 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 present relevant applications to risk assessment in insurance markets under a bipartite network structure.

Session, time and room:
IS - Risk analysis in insurance (Peng)
Friday, 11:50 – 12:20 (A002)

Peak-over-Threshold Estimators for Spectral Tail Processes: Random vs Deterministic Thresholds

Holger Drees[1], Miran Knežević[2]
University of Hamburg, Germany[1,2]
holger.drees@uni-hamburg.de[1], miran.knezevic@uni-hamburg.de[2]

The extreme value dependence of regularly varying stationary time series can be described by the spectral tail process. Drees et al. [2] proposed estimators of the marginal distributions of this process based on exceedances over high deterministic thresholds and analyzed their asymptotic behavior. In practice, however, versions of the estimators are applied which use exceedances over random thresholds like intermediate order statistics. These modified estimators are also asymptotically normal. In a small simulation study, we compare the finite sample behavior of these estimators based on random versus deterministic thresholds.

This talk is based on a joint work with Holger Drees.

References

[1] H. Drees and M. Knežević, Peak-over-Threshold Estimators for Spectral Tail Processes: Random vs Deterministic Thresholds, *arXiv:1901.05501*, 2019

[2] H. Drees, J. Segers and M. Warchoł, Statistics for tail processes of markov chains, *Extremes 18(3)*, 2015

Session, time and room:
Statistics for time series
Tuesday, 9:00 – 9:20 (A002)

Infinitesimal perturbation analysis for risk measures based on the Smith max-stable random field

Erwan Koch[1], Christian Y. Robert[2]
EPFL (Chair of Statistics), Switzerland[1], ISFA Université Lyon 1, France[2]
erwan.koch@epfl.ch[1], christian.robert@univ-lyon1.fr[2]

When using risk or dependence measures based on a given underlying model, it is essential to be able to quantify the sensitivity or robustness of these measures with respect to the model parameters. In this paper, we consider an underlying model which is very popular in spatial extremes, the Smith max-stable random field. We study the sensitivity properties of risk or dependence measures based on the values of this field at a finite number of locations. Max-stable fields play a key role, e.g., in the modelling of natural disasters. As their multivariate density is generally not available for more than three locations, the Likelihood Ratio Method cannot be used to estimate the derivatives of the risk measures with respect to the model parameters. Thus, we focus on a pathwise method, the Infinitesimal Perturbation Analysis (IPA). We provide a convenient and tractable sufficient condition for performing IPA, which is intricate to obtain because of the very structure of max-stable fields involving pointwise maxima over an infinite number of random functions. IPA enables the consistent estimation of the considered measures' derivatives with respect to the parameters characterizing the spatial dependence. We carry out a simulation study which shows that the approach performs well in various configurations.

Session, time and room:
Max-stable fields
Friday, 12:05 – 12:25 (A101)

Trends in the extremes of environments associated with severe US thunderstorms

Anthony C. Davison[1], Erwan Koch[2], Jonathan Koh[3]
EPFL, Switzerland[1,2,3]
anthony.davison@epfl.ch[1], erwan.koch@epfl.ch[2], jonathan.koh@epfl.ch[3]

Severe thunderstorms can have devastating impacts. Concurrently high values of convective available potential energy (CAPE) and storm relative helicity (SRH) are conducive to hazardous convective weather (tornadoes, hail and damaging wind) associated with severe thunderstorms. Hence, it is highly relevant to have probabilistic models for both variables' extremes that use relevant covariate information to account appropriately for their spatial and temporal dependence. We consider a large area of the contiguous United States over the period 1979–2015 and use statistical extreme value models and appropriate multiple testing procedures to identify trends in the extremes. Various features of the data motivate a two-step approach. In the first step, we show that there is a significant time trend in the extremes for CAPE maxima in April, May and June, and for SRH maxima in April and May. These increases in CAPE are also relevant for rainfall extremes and are expected in a warmer climate but have not previously been reported. We show that the El Niño-Southern Oscillation (ENSO) explains variation in the extremes of PROD and SRH in February. Our results imply that the risk of severe thunderstorms in April and May is increasing in parts of the US where this risk was already high and that the storm risk in February tends to be higher over much of the region during La Niña years. In the second step, we focus on the local spatial extremal dependence structure and thus model the pointwise maxima using max-stable random fields. We focus especially on the Brown–Resnick field with a power variogram and investigate the effect of time and ENSO on its parameters. Allowing for multiple testing, preliminary results show that the smoothness parameter for SRH in February increases with ENSO, implying that extremes of SRH are more localized during La Niña years for this month. The range and smoothness parameters for both CAPE and SRH are lower in the spring and summer months, so the corresponding extremes are more localized during these seasons. This is consistent with known meteorological mechanisms and may have implications for risk assessment.

Session, time and room:
IS - Extremes and climate physics (Fougeres)
Tuesday, 10:10 – 10:40 (003)

Functional limit theorems for linear processes

Danijel Krizmanić[1]
University of Rijeka, Croatia[1]
dkrizmanic@math.uniri.hr[1]

For linear processes with independent and identically distributed heavy-tailed innovations we present some new functional limit theorems for partial sum stochastic processes and joint partial sum and maxima processes. The convergence takes place in the space of càdlàg functions with one of the Skorohod's topologies M_1 and M_2. More details on results presented in the talk can be found in [1] and [2].

References

[1] D. Krizmanić, Joint functional convergence of partial sums and maxima for linear processes, *Lith. Math. J.* **58**, 457-479, 2018.

[2] D. Krizmanić, Functional convergence for moving averages with heavy tails and random coefficients, preprint: arXiv:1808.07023, 2018.

Session, time and room:
Limit theorems II
Monday, 17:05 – 17:25 (A101)

Double finite-time ruin probability for correlated Brownian motions

Krzysztof Dębicki[1], Enkelejd Hashorva[2], Konrad Krystecki[3]
University of Wrocław, Poland[1], University of Lausanne, Switzerland[2], University of Wrocław/University of Lausanne, Poland[3]
Krzysztof.Debicki@math.uni.wroc.pl[1], Enkelejd.Hashorva@unil.ch[2], kkrystecki@math.uni.wroc.pl[3]

We focus on deriving the asymptotics of suprema of correlated Brownian motions with drift on finite time interval, i.e. we analyze the following probability

$$P(\sup_{s,t\in[0,1]} W_1(s) - c_1 s > u, W_2(t) - c_2 t > au)$$

as $u \to \infty$. We present the exact asymptotics of the probability above and study the influence of the dependence between a and the correlation of W_1 and W_2 on the results.

Session, time and room:
Ruin probability
Tuesday, 11:40 – 12:00 (A102)

An M-Estimator for Tail Dependence and Independence

Sebastian Engelke[1], Michaël Lalancette[2], Stanislav Volgushev[3]
University of Geneva, Switzerland[1], University of Toronto, Canada[2,3]
sebastian.engelke@unige.ch[1], lalancette@utstat.toronto.edu[2],
stanislav.volgushev@utoronto.ca[3]

This paper considers the problem of estimating the shape of the joint tail of two (possibly dependent) random variables X and Y. If F_1 and F_2 are the marginal distribution functions of X and Y, respectively, the joint tail is characterized by the *survival tail copula function*

$$c(x,y) = \lim_{t\to 0} q(t)^{-1}\mathbb{P}(F_1(X) \geq 1 - tx, F_2(Y) \geq 1 - ty), \tag{1}$$

for a suitable normalizing function q that makes the limit non trivial. Assuming n independent copies of (X,Y), a nonparametric estimator of c was proposed by [2]. We prove consistency and asymptotic normality of their estimator under significantly weaker assumptions. In particular, we do not require any kind of differentiability assumption on c for the asymptotic normality. In the case where X and Y are asymptotically independent, we prove asymptotic normality in the topology of uniform convergence for locally bounded functions. Our result is a major generalization of [2]. In the case of asymptotic dependence, where uniform weak convergence is known to fail without proper differentiability of c, the estimator is asymptotically normal in the weaker topology of hypi-convergence (see [1]).

Assuming that c belongs to a given parametric family $\{c_\theta\}$, we further propose an M-estimator that is based on the nonparametric estimator of c. A similar method was proposed by [3] to estimate the stable tail dependence function, but we modify it to account for the unknown scaling function q (in the case of the stable tail dependence function, the scaling is known). Our procedure allows the simultaneous estimation of θ and $q(t)$, for any small value of t. In view of eq. (1), this leads to a natural estimate for the probability of joint exceedances of high thresholds. All our estimators are shown to be consistent and asymptotically normal.

Finally, the usefulness of the method is illustrated through applications to flexible dependence models, allowing both asymptotic dependence and independence.

References

[1] A. Bücher, J. Segers and S. Volgushev. When Uniform Weak Convergence Fails: Empirical Processes for Dependence Functions and Residuals via Epi- and Hypographs. *Ann. Stat.*, 2014.

[2] G. Draisma, H. Drees, A. Ferreira and L. de Haan. Bivariate Tail Estimation: Dependence in Asymptotic Independence. *Bernoulli*, 2004.

[3] J.H.J Einmahl, A. Krajina and J. Segers. An M-Estimator for Tail Dependence in Arbitrary Dimensions. *Ann. Stat.*, 2012.

Session, time and room:
Tail dependence
Thursday, 12:05 – 12:25 (A001)

Markov properties of max-linear graphical models

Steffen Lauritzen[1]
University of Copenhagen, Denmark[1]
lauritzen@math.ku.dk[1]

Recursive max-linear models was introduced in [1] as structural equation models with max-linear relations corresponding to directed acyclic graphs. These become Bayesian networks which do not admit densities w.r.t. product measures so standard Markov theory for such networks is not directly applicable. In addition, the max-linearity present special challenges for identiying conditional independence relations as the models are rarely faithful and exhibit additional conditional independence relations due to extreme events cascading almost deterministically through the network. The lecture is based on joint work with C. Amendola, N. Gissibl, N. Tran, and C. Klüppelberg

References

[1] N. Gissibl and C. Klüppelberg, Max-linear models on directed acyclic graphs. *Bernoulli*, 2018

Session, time and room:
IS - Extremes on graphs (Klueppelberg)
Friday, 9:35 – 10:05 (A002)

Bayesian Semi-parametric Modelling of Heteroscedastic Extremes

Miguel de Carvalho[1], Junho Lee[2], António Rua[3]
University of Edinburgh, United Kingdom[1,2], Banco de Portugal, Portugal[3]
Miguel.deCarvalho@ed.ac.uk[1], j.lee-63@sms.ed.ac.uk[2], arua@bportugal.pt[3]

We develop Bayesian semi-parametric inference methods for heteroscedastic extremes. The proposed model is based on an extreme value index regression and on a proportional tails model and can be used for assessing how the magnitude and frequency of the extreme values can change along with a covariate. We start with the unconditional setting for estimating the extreme value index and the scedasis function and show that the proposed inference methods for the scedasis density—based on a Bernstein–Dirichlet prior—perform well in Monte Carlo simulation studies, are exact apart from Monte Carlo error, and have full support on the space of all continuous scedasis functions. We then extend the proposed methods to the conditional setting using dependent Bernstein-Dirichlet process.

We resort to the proposed methodologies to examine an extreme currency demand in Portugal. The signatures of the fitted scedasis densities of extreme currency demand—over different denominations—reveal some interesting insights on the dynamics governing currency demand during periods of economic stress.

Session, time and room:
Finance and economics
Thursday, 11:15 – 11:35 (A002)

Bias correction for the maximum likelihood estimator of the extreme value index

Xuan Leng[1], Chen Zhou[2]
Xiamen University, China[1], Erasmus University Rotterdam, Netherlands[2]
xleng0522@gmail.com[1], zhou@ese.eur.nl[2]

This paper conducts bias correction for the maximum likelihood estimator (MLE) of the extreme value index. Compared to the original MLE, the bias-corrected estimator allows for using a larger fraction of observations in tail region for estimation, which results in a lower asymptotic variance. The bias correction is achieved by subtracting the asymptotic bias from the original MLE, which is estimated by a two-step approach. We prove the asymptotic behavior of the proposed bias-corrected estimator. Extensive simulations show the superiority of the bias-corrected estimator compared to existing estimators of the extreme value index. We apply the bias-corrected MLE to test whether human life length is unlimited.

Session, time and room:
Tail index estimation
Monday, 11:45 – 12:05 (A101)

Forecaster's Dilemma: Extreme Events and Forecast Evaluation

Tilmann Gneiting[1], Sebastian Lerch[2], Francesco Ravazzolo[3], Thordis Thorarinsdottir[4]
Karlsruhe Institute of Technology and Heidelberg Institute for Theoretical Studies, Croatia[1], Karlsruhe Institute of Technology and Heidelberg Institute for Theoretical Studies, Germany[2], Free University of Bozen - Bolzano, Italy[3], Norwegian Computing Center, Norway[4]
tilmann.gneiting@h-its.org[1], Sebastian.Lerch@h-its.org[2], Francesco.Ravazzolo@unibz.it[3], thordis@nr.no[4]

In public discussions of the quality of forecasts, attention typically focuses on the predictive performance in cases of extreme events. However, the restriction of conventional forecast evaluation methods to subsets of extreme observations has unexpected and undesired effects, and is bound to discredit skillful forecasts when the signal-to-noise ratio in the data generating process is low. Conditioning on outcomes is incompatible with the theoretical assumptions of established forecast evaluation methods, thereby confronting forecasters with what we refer to as the forecaster's dilemma. For probabilistic forecasts, proper weighted scoring rules have been proposed as decision-theoretically justifiable alternatives for forecast evaluation with an emphasis on extreme events. Using theoretical arguments, simulation experiments and a real data study on probabilistic forecasts of U.S. inflation and gross domestic product growth, we illustrate and discuss the forecaster's dilemma along with potential remedies.

References

[1] Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F. and Gneiting, T., Forecaster's Dilemma: Extreme Events and Forecast Evaluation, *Statistical Science*, 32, 106–127, 2017

Session, time and room:
IS - Predicting extremes (Segers)
Monday, 10:30 – 11:00 (003)

Extreme quantile estimation for single index model

Deyuan Li[1], Huixia Judy Wang[2], Wen Xu[3]
Fudan University, China[1,3], George Washington University, United States[2]
deyuanli@fudan.edu.cn[1], judywang@gwu.edu[2], 18110690007@fudan.edu.cn[3]

Single Index model is a flexible semiparametric regression model and it reduces the dimension of the covariates. In this paper, we consider the estimation of the extreme conditional quantiles for the single index model and developed a so-called three-step estimator. We first obtain a misspecified root-n estimator of the index parameter vector under the linear quantile regression. Secondly, we apply a local polynomial regression technique to estimate intermediate conditional quantiles. Finally, we extrapolate these estimates to tails by extreme value theory. We show the asymptotic properties of the provided estimator and study its performance for finite sample by simulation. A real application to the NMMAPS dataset of LA is also provided.

Session, time and room:
Conditional extremes
Thursday, 14:25 – 14:45 (A101)

Asymptotics of Multivariate Conditional Risk Measures for Gaussian Risks

Chengxiu Ling[1]
Department of Mathematics Sciences, Xi'an Jiaotong-Liverpool University, China[1]
chengxiu.ling@xjtlu.edu.cn[1]

This paper investigates accurate approximations of marginal moment excess, marginal conditional tail moment and marginal moment shortfall for multivariate Gaussian system risks. Based on the dimension reduction property via the quadratic programming problem, the super-exponential and polynomial convergence speeds are specified. Two interesting questions involved in risk management are well addressed, namely the minimal additional risk capital injection to avoid infinite risk contagion and a sufficient and necessary condition to alternate the convergence speeds. Numerical study and typical examples are given to illustrate the efficiency of our findings. Due to the flexible moment order, additional applications may involve in risk management, including tail mean-variance portfolio and multivariate conditional risk measures of tail covariance, tail skewness with dependence and extremal risk contagion under consideration.

Session, time and room:

Financial extremes II
Tuesday, 11:40 – 12:00 (A002)

Integral range and extremal coefficient of stationary max-stable random fields

Chautru Emilie[1], Demangeot Marine[2]
MINES ParisTech, PSL University, Centre de géosciences, France[1,2]
emilie.chautru@mines-paristech.fr[1], marine.demangeot@mines-paristech.fr[2]

In many geostatistical applications (soil contamination evaluation, mining resources estimation), the physical phenomenon under study is interpreted as a particular realization of a stationary random field with a finite expectation. Can the expectation be estimated starting from a single realization ? Part of the answer is brought by the concept of integral range. Intimately related to the ergodic and mixing properties, it is a geostatistical object that characterizes the statistical fluctuations of the random field at large scale [1]. When the latter is max-stable, we show that its extremal coefficient function [2] is closely related to the integral range of the indicator function of the random field above different thresholds. This approach allows to retrieve and complete results established by Erwan Koch in a spatial risk context [3]. It thus illustrates the relevance of geostatistical tools to enrich the extreme value analysis, especially for the question of estimating with a single realization of the spatial process.

References

[1] C. Lantuéjoul, Ergodicity and integral range, *Journal of Microscopy*, 1991

[2] M. Schlather and J.A. Tawn, A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, 2003

[3] E. Koch, Spatial risk measures and applications to max-stable processes, *Extremes*, 2017

Session, time and room:
Max-stable fields
Friday, 11:40 – 12:00 (A101)

Cluster properties of non-stationary random length sequences

Natalia Markovich[1]
V.A.Trapeznikov Insitute of Control Sciences of Russian Academy of Sciences, Russian Federation[1]
nat.markovich@gmail.com[1]

We study non-stationary random length sequences of regularly varying distributed random variables. The results from [1] that are valid for non-stationary random sequences with deterministic length are extended to random length sequences. Namely, we deal with a doubly-indexed array of regularly varying r.v.s $Y_{n,i}$ in which the "row index" n corresponds to time, and the "column index" i corresponds to the level. On the same probability space, one assumes the existence of a sequence of non-negative integer-valued r.v.s $\{N_n : n \geq 1\}$ such that $EN_n < \infty$ for each n. The "column" sequences $\{N_n, Y_{n,i} : i \geq 1\}$ are stationary in n. In each column there are no assumptions on the dependence structure (in i), but the tail asymptotic of the marginal distributions of the N_n's and the $Y_{n,i}$'s (for each fixed i) are assumed. In first two theorems $Y_{n,i}$'s are indexed so that $Y_{n,1}$ always has the heaviest (marginal) tail. In the third theorem all $Y_{n,i}$'s are mutually independent and have the same heaviness of tail. Then we consider

$$\begin{aligned} Y_n^*(z, N_n) &= \max(z_1 Y_{n,1}, .., z_{N_n} Y_{n,N_n}), \\ Y_n(z, N_n) &= z_1 Y_{n,1} + .. + z_{N_n} Y_{n,N_n} \end{aligned}$$

for positive constants $z_1, z_2, ...$ and discuss tail and extremal indexes for these sequences. We obtain conditions when both sums $Y_n(z, N_n)$ and maxima $Y_n^*(z, N_n)$ of these sequences have the same tail and extremal indices. The results are based on [2].

References

[1] A. A. Goldaeva, Indices of multivariate recurrent stochastic sequences. *In book ed. Shiryaev A. N.: Modern problem of mathematics and mechanics* VIII(3), 42-51, Moscow State University ISBN 978-—5—-211—-05652—-7 (2013) (in Russian)

[2] N. M. Markovich, I. V. Rodionov, Maxima and sums of non-stationary random length sequences, Submitted paper, 2019.

Session, time and room:
OCS - Inferences on extremal index (Markovich)
Tuesday, 14:25 – 14:45 (A101)

About life and some results of Jovan Karamata (from today perspective)

Miodrag Mateljević[1]
University of Belgrade, Faculty of Mathematics, Serbia[1]
miodrag@matf.bg.ac.rs[1]

We plan to discuss the following:
1. Short biography of Jovan Karamata;
2. Results related to Regular Variation;
3. On Karamata's proof of the Landau-Ingham Tauberian theorem (the Prime Number Theorem);
4. Inequalities of Karamata, Schur and Muirhead related to convex function, and some application including Karamata's Mean Value Theorem, Some Consequences and Some Stability Results.

Session, time and room:
IS ST - Regular variation: history, ideas and people (Basrak)
Friday, 12:25 – 12:55 (A001)

Spectral measure and projection onto the simplex

Nicolas Meyer[1], Olivier Wintenberger[2]
LPSM, Sorbonne Université, France[1,2]
nicolas.meyer@upmc.fr[1], olivier.wintenberger@upmc.fr[2]

Regular variation provides a convenient theoretical framework to study large events. In the multivariate setting, the dependence structure of the extremes is characterized by the spectral measure, a measure defined on the positive orthant of the unit sphere. This measure is based on the weak convergence of the projected vector $\mathbf{X}/|\mathbf{X}| \mid |\mathbf{X}| > t$ when $t \to \infty$. Unfortunately, this definition does not provide a natural estimator of its support.

In this talk, we propose a new method based on the ℓ^2-projection π onto the simplex to study the spectral measure. We first prove several results that shows that this projection allows to better capture the dependence structure of extreme events. Indeed, it highlights the possible sparsity of the spectral measure's support. This significantly reduces the dimension. Then, we show that, conditional on the event $|\mathbf{X}| > t$, there is an equivalence between the convergence of $\pi(\mathbf{X}/t)$ and the one of $\mathbf{X}/|\mathbf{X}|$, and we give the explicit relationship between both limit measures. This leads to a new representation of the spectral measure. Based on these results, we use Dirichlet models on the simplex to illustrate to what extent the projection modifies the spectral measure. We illustrate our method on different applications.

Session, time and room:
Multivariate Extremes
Tuesday, 16:15 – 16:35 (A001)

Causal mechanism of extreme river discharges in the upper Danube basin network

Valérie Chavez-Demoulin[1], Debbie J. Dupuis[2], Linda Mhalla[3]
HEC Lausanne, Switzerland[1], HEC Montreal, Canada[2,3]
valerie.chavez@unil.ch[1], debbie.dupuis@hec.ca[2], linda.mhalla@hec.ca[3]

Extreme hydrological events in the Danube river basin have tragic consequences for human populations, living aquatic organisms, and the economic activity. One often characterizes the joint structure of the extreme events using the theory of multivariate and spatial extremes and its asymptotically justified models. There is interest however in cascading extreme events and whether one event causes another. In this paper, we argue that an improved understanding of the mechanism underlying severe events is achieved by combining extreme value modelling and causal discovery. We construct a causal inference method relying on the notion of Kolmogorov complexity of extreme conditional quantiles. Tail quantities are derived using multivariate extreme value models and causal-induced asymmetries in the data are explored through the minimum description length principle. Owing to the developed methodology, we uncover causal relations between summer extreme river discharges in the upper Danube basin and find significant causal links between the Danube and its Alpine tributary Lech.

Session, time and room:
Flood risks
Thursday, 11:40 – 12:00 (003)

A sequential design for the estimation of extreme quantile for small samples based on exceedances

Michel Broniatowski[1], Emilie Miranda[2]
Sorbonne Université, France[1,2]
michel.broniatowski@upmc.fr[1], emilie.miranda@upmc.fr[2]

Let X be a r.v in $\mathbb{R}_+$, with cumulative distribution function F. The goal is to estimate q_α, the α-quantile of its distribution, based on a sample of n observations, with α much smaller than $1/n$. The only information at hand for this inference does not consist on a sample $(X_1, .., X_n)$ of independent copies of X, but rather on indicators of events of the form $X_i > x_i$ where the thresholds x_i can be chosen by the experimenter. The problem turns therefore in a sequential choice of those x_i's which define the experimental design for the estimation of q_α.

This study is motivated by an industrial challenge: The estimation of extreme failure quantiles for a structure or a material. In this framework, the variable of interest is the resistance of the material and can not be directly observed. The information available is drawn from experimental trials and consists in indicators of exceedances over the stress level tested.

We thus propose a sequential design method to estimate a failure quantile of order α, where the relevant information given by the trials is reduced to an indicator of whether the specimen has failed at the tested stress level. This difficulty is overcome by using a splitting

approach, decomposing the target probability level α into a product of probabilities of conditional events of higher order. The method consists in gradually targeting the tail of the distribution and sampling under truncated distributions. The model is GEV, and sequential estimation of its parameters involves an improved maximum likelihood procedure for binary data, due to the large uncertainty associated with such a restricted information.

The proposed methodology is supported by simulation results and comparisons with similar estimations based on complete observations.

References

[1] Balkema A. A. and De Haan L., *Residual Life Time at Great Age*. Ann. Prob., 2(5), 762-804(1974).

[2] De Valk C., *Approximation of high quantiles from intermediate quantiles*, Extremes, 4, 661-684,2016.

[3] Dixon W. J. and Mood A. M., *A Method for Obtaining and Analyzing Sensitivity Data*, Journal of the American Statistical Association, 43, 109-126, 1948.

[4] Dixon W. J., *The Up-and-Down Method for Small Samples*,

Session, time and room:
Tail inference
Thursday, 9:25 – 9:45 (A101)

Empirical Estimator of Conditional Tail Moment in the case of censored data

Rassoul Abdelaziz[1], Ould Rouis Hamid[2], Laidi Mohamed[3]

National High School of Hydraulics, Blida, Algeria[1], University of Blida 1, Blida,, Algeria[2], National High School of Technology, Algiers, Algeria[3]

a.rassoul@ensh.dz[1], houldrouis@hotmail.com[2], mohamed.laidi@enst.dz[3]

The Conditional Tail Moment is defined as the moment of order $a \geq 0$ of the loss distribution above the upper α-quantile where $\alpha \in (0, 1)$. In this paper, we provide and explore the properties of an empirical estimator of this risk measure in the case of censored data, which it permits to give a censored empirical estimator for all risk measures based on conditional moments such as Value-at-Risk, Conditional Tail Expectation, Conditional Value-at-Risk or Conditional Tail Variance. We explore the asymptotic properties of this empirical estimator, in particular, his consistency and we construct a confidence interval for it. we finish our work by simulation studies that illustrate the performance of this estimator and we discuss the obtained results.

Session, time and room:
Inference for censored data
Friday, 9:25 – 9:45 (A101)

Forecasting heavy and low rainfall data by coupling random forests and extreme value theory

Anne-Laure Fougères[1], Olivier Mestre[2], Philippe Naveau[3], Maxime Taillardat[4]

Lyon University, France[1], Meteo France, France[2,4], CNRS-LSCE-IPSL, France[3]

fougeres@math.univ-lyon1.fr[1], olivier.mestre@meteo.fr[2],
philippe.naveau@lsce.ipsl.fr[3], maxime.taillardat@meteo.fr[4]

Nowadays, weather centers operationally produce ensemble forecasts of different atmospheric variables, especially extreme weather events. Still, to satisfy a wide range of end-users, rainfall ensemble forecasts have to be skillful for both low precipitation and extreme events. In this context, two statistical questions are paramount for practitioners: how to improve the forecast of heavy rainfall and how to measure the quality of produced forecasts. To deal with the first question, we introduce local statistical post-processing methods based on Quantile Regression Forests and Gradient Forests with a semi-parametric extension for heavy-tailed distributions. These hybrid methods make use of the forest-based outputs to fit a parametric distribution that is suitable to model jointly low, medium, and heavy rainfall intensities [2]. Our goal in terms of applications is to improve ensemble quality and value for all rainfall intensities. The proposed methods are applied to daily 51-h forecasts of 6-h accumulated precipitation from 2012 to 2015 over France using the Meteo-France ensemble prediction system called PEARP. They are verified with a cross validation strategy and compete favorably with state-of- the-art methods like Analog Ensemble or Ensemble Model Output Statistics. In addition to improvements in overall performance, hybrid forest-based procedures produced the largest skill improvements for forecasting heavy rainfall events [1].

If time allowed, the second question concerning the scoring of extremes forecasts will be briefly touched upon. In particular, a few properties related to the classical score (CRPS) will be discussed with respect to its capacities to judge extremal forecasts [3].

References

[1] M. Taillardat, A-L. Fougères, P. Naveau, and O. Mestre. Forest-based and semi-parametric methods for the postprocessing of rainfall ensemble forecasting. *Weather and Forecasting*, in press, 2019.

[2] P. Naveau, R. Huser, P. Ribereau, and A. Hannart. Modelling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 10:1002, 2016.

[3] M. Taillardat, R. de Fondeville, A-L. Fougères, and P. Naveau. Evaluation of the CRPS Distribution for extremes. *ArXiv*, 2019.

Session, time and room:
IS - Predicting extremes (Segers)
Monday, 11:40 – 12:10 (003)

Trend detection in extreme value models

Zuzana Hübnerová[1], László Németh[2], András Zempléni[3]
Brno University of Technology, Institute of Mathematics, Brno, Czech Republic[1], Eötvös Loránd University, Institute of Mathematics, Department of Probability and Statistics, Budapest, Hungary[2,3]
zuzana.hubnerova@gmail.com[1], nmthlszl@gmail.com[2], zempleni@caesar.elte.hu[3]

Classical extreme value theory has numerous results for limit distribution of block maxima (Fisher-Tippet), threshold exceedances (POT model) or multivariate data. However, all theorems assume independent, identically distributed (iid.) variables. Unfortunately, in real life problems the distribution of extremes changes due to natural/environmental effects quite quickly. We investigate two models, aiming to detect a linear trend in extreme values.

The first model is based on classical linear regression. After fitting a trend to the data, we suppose the distribution of residuals as GEV distribution. Using e.g. a likelihood ratio test one may decide, if the linear trend is significant compared to the iid. GEV model.

The second method expects trend in the parameters of the fitted GEV distribution. Likelihood ratio test could detect significant trend in the location parameter as it is parallel to the previous approach.

By using bootstrap simulations, we investigated the power and the false detection rate of the models depending on the sample size and other parameters like the GEV distribution's shape. We applied the methods to real life data - monthly maxima of fire weather index in selected stations in British Columbia. Results show that in some cases, where the sample size was larger we could identify significant trends by both methods.

Session, time and room:
Peaks over threshold
Tuesday, 14:25 – 14:45 (A002)

Testing randomly right-censored extremes

Claudia Neves[1]
University of Reading, United Kingdom[1]
c.neves@reading.ac.uk[1]

Statistical inference for extremes of a randomly right-censored sample is a recent and rapidly developing field. It has found applications in medical statistics as well as in actuarial mathematics. Despite the wealth of published literature this field, all methodologies share the same underpinning assumption of which a consequence is that the extremes of the censored variable may be recovered by adapted extreme value techniques. Invariably, no actual means for checking this assumption have been provided. This paper aims to fill in this gap by providing novel and easily applicable statistical methods for validating this hypothesis in practice. We propose two complementary testing procedures for evaluating the impact of censoring on extreme observations. Our procedures are rooted in a detailed analysis of the concept of tail censoring probability and its link to the feasibility of the estimation problem. The consistency and asymptotic power of our testing methods are ascertained under mild conditions and the finite-sample performance of the methodology

is evaluated through simulations, in which we study important settings where the proposed tests can successfully discriminate varying degrees of censoring. Finally, the devised statistical tests are showcased on two sets of real data, one from survival analysis and the other from demography.

This is joint work with Gilles Stupfler, University of Nottingham, UK.

Session, time and room:
Testing and inference
Thursday, 12:05 – 12:25 (A102)

Generalized logistic extreme value distributions

Anne-Laure Fougeres[1], Cecile Mercadier[2], John Nolan[3]
University of Lyon, France[1,2], American University, United States[3]
fougeres@math.univ-lyon1.fr[1], mercadier@math.univ-lyon1.fr[2], jpnolan@american.edu[3]

Multivariate Fréchet laws are a class of extreme value distributions that exhibit heavy tails and directional dependence controlled by an angular measure. Multivariate generalized logistic laws are a recently described sub-class that are dense in a certain sense. It is shown that these laws are related to positive multivariate sum stable laws, which gives a way to simulate from these laws. The corresponding angular measure density is described, and expressions for the density of the distribution are given.

Session, time and room:
Multivariate Extremes
Tuesday, 16:40 – 17:00 (A001)

An extreme value approach to CoVaR estimation

Natalia Nolde[1], Chen Zhou[2], Menglin Zhou[3]
University of British Columbia, Canada[1,3], Erasmus University, Netherlands[2]
natalia@stat.ubc.ca[1], zhou@ese.eur.nl[2], menglin.zhou@stat.ubc.ca[3]

The financial crisis of 2007-2009 highlighted the importance of systemic financial risk and its potential to destabilize the global economy. Accurate assessment of systemic risk would enable regulators to identify systemically important financial institutions and introduce suitable policies to mitigate the risk to the system coming from such institutions. For individual institutions, on the other hand, it is their vulnerability to extremal market movements that should be monitored and mitigated. Both of the above situations require a multivariate measure of risk that captures co-movements between a financial system (or market) and individual financial institutions. One popular measure of systemic risk is CoVaR. A methodology is proposed to estimate CoVaR semi-parametrically within the classical framework of multivariate extreme value theory. By exploring the definition of CoVaR, we combine parametric modelling of the tail dependence function to address the

issue of data sparsity in the joint tail regions and semi-parametric univariate tail estimation techniques. The performance of the methodology is illustrated via simulation studies and real data examples.

Session, time and room:
IS - Financial extremes (Zhou)
Thursday, 9:00 – 9:30 (A001)

An efficient semiparametric maxima estimator of the extremal index

Paul Northrop[1]
University College London, United Kingdom[1]
p.northrop@ucl.ac.uk[1]

The extremal index θ, a measure of the degree of clustering in the extremes of a stationary process, plays an important role in extreme value analyses. We consider semiparametric estimators of θ that are based on the approximate relationship between the distribution of block maxima and the marginal distribution of a process. These estimators depend only on selecting a suitably large block length and are substantially more efficient than their parametric counterparts. Efficiency is further improved with the use of maxima over sliding blocks, rather than over disjoint blocks. We discuss the estimator introduced in [1] and a modified version of this studied by [2]. [2] provide asymptotic theory to enable bias-adjustment and estimation of uncertainty to be performed. We compare the performance of these estimators to each other, and to other types of estimators, using simulation.

References

[1] P.J. Northrop, An efficient semiparametric maxima estimator of the extremal index, *Extremes*, 2015

[2] B. Berghaus and A. Bücher, Weak convergence of a pseudo maximum likelihood estimator for extremal index, *The Annals of Statistics*, 2018

Session, time and room:
OCS - Inferences on extremal index (Markovich)
Tuesday, 14:50 – 15:10 (A101)

Rare-event simulation for Gibbs measures: Properties and applications of the infinite swapping algorithm

Pierre Nyquist[1]
KTH Royal Institute of Technology, Sweden[1]
pierren@kth.se[1]

Rare-event sampling is a hindrance to efficiently sampling 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.

Infinite swapping is a Monte Carlo method, originally developed by Doll, Dupuis and co-authors, designed to overcome the problem of rare-event sampling, and thus speed up convergence, in the setting of computing integrals with respect to Gibbs measures. In subsequent work the properties of infinite swapping have been studied further using a combination of empirical measure large deviations and stochastic ergodic control problems, revealing explicit information on the convergence properties of the sampling scheme [2].

This talk will focus on the properties of infinite swapping and (non-standard) applications in quantum dynamics [1] and machine learning [3]. Given time we will also discuss briefly the use of empirical measure large deviations to study convergence rates for MCMC methods.

The talk is based on joint work with Paul Dupuis, Jim Doll, Henrik Hult and Carl Ringqvist.

References

[1] J. D. Doll, P. Dupuis, P. Nyquist. Thermodynamic integration methods, infinite swapping and the calculation of generalized averages. *J. Chem. Phys.*, 2017.

[2] J. D. Doll, P. Dupuis, P. Nyquist. A large deviation analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. *Appl. Math. Optim.*, 2018.

[3] H. Hult, P. Nyquist, C. Ringqvist. The infinite swapping algorithm for restricted Boltzmann machines. *Preprint*, 2019.

Session, time and room:
IS ST - Simulation of rare events and extremes (Zwart)
Tuesday, 9:35 – 10:05 (A001)

Ordinal Patterns in Spatio-Temporal Extremes

Marco Oesting[1], Alexander Schnurr[2]
University of Siegen, Germany[1,2]
oesting@mathematik.uni-siegen.de[1], schnurr@mathematik.uni-siegen.de[2]

In many applications in environmental sciences, extreme events exhibit a complex spatial-temporal structure which needs to be described in a compressed way. To this end, we propose to investigate the temporal course of various characteristics such as the magnitude of the event, its centroid or the size of the affected area. A special focus is put on the distribution of ordinal patterns of the characteristics within clusters of extremes. We verify the existence of the corresponding limit distributions in the framework of regular variation, develop non-parametric estimators and show their asymptotic normality under appropriate mixing conditions.

Session, time and room:
IS - Spatial extremes (Dombry)
Monday, 14:35 – 15:05 (003)

Changes in Frequency and Magnitude of Extreme Rainfalls in the Northeastern USA

David Bolin[1], Helga Kristin Olafsdottir[2], Holger Rootzen[3]
Chalmers University of Technology and the University of Gothenburg, Sweden[1,2,3]
david.bolin@chalmers.se[1], khelga@chalmers.se[2], hrootzen@chalmers.se[3]

Analysis of historical time series of extreme precipitation is traditionally done under the assumption of stationarity. With recent awareness of climate changes, there is reason to introduce non-stationarity in rainfall models. For the annual maxima series this can be done by making the parameters of the Generalized Extreme Value (GEV) distribution depend on time, and for the partial duration series by letting the intensity of the Poisson process of threshold exceedances and the parameters of the General Pareto (GP) distribution for the sizes of threshold excesses be time-dependent. Here we study a NOAA data set on rainfall in the Northeastern USA which contains daily rainfall measurements from twenty stations covering approximately the period 1900 - 2014. These stations were chosen since preliminary analysis gave indication of non-stationarity. Partial duration series provide data of higher resolution and thus a more direct path of understanding the behavior of individual rainstorms, but instead annual maxima measurements are often better checked and of higher quality, and more widely accessible. Using the connection between GEV model for annual maxima and the Poisson-GP model for daily threshold exceedances – a variant of "Langbein's formula" – we estimated the parameters of the Poisson process and of the GP distribution from the annual maxima series. The results indicate that the frequency of extreme rainfalls is increasing, but that their sizes have not changed. The behavior inferred from the analysis of annual data was compared with the results using the daily data.

Session, time and room:

Modelling changes in environmental extremes
Tuesday, 11:15 – 11:35 (A001)

The All Block Maxima Estimator

Jochem Oorschot[1], Chen Zhou[2]

Erasmus University Rotterdam, Department of Econometrics & Tinbergen Institute, Netherlands[1], Erasmus University Rotterdam, Department of Econometrics & Tinbergen Institute & De Nederlandsche Bank, Netherlands[2]

oorschot@ese.eur.nl[1], zhou@ese.eur.nl[2]

A classical approach in extreme value analysis is to fit a sample of block maxima to the Generalized Extreme Value (GEV) distribution — the block maxima (BM) method. A standard way to obtain block maxima is by dividing the sample into disjoint blocks. [1] shows that one may improve the disjoint block approach by randomly sampling blocks from the data. We take this idea to its extreme by considering all potential blocks — the All Block Maxima (ABM) estimator. Unlike the classical BM method, the ABM estimator is permutation invariant. The ABM method entails weighting each of the higher order statistics by the potential number of blocks in which it could have been a block maxima. For distributions in the Frechét Domain of Attraction, these weights resemble the Radon-Nikodym derivative between the Frechét and the Pareto distribution. The key element in establishing the asymptotic theory of the ABM estimator is the tail empirical process based on higher order statistics with weights.

References

[1] S. Wager, Subsampling extremes: From block maxima to smooth tail estimation, *Journal of Multivariate Analysis*, 2014

Session, time and room:
Tail inference
Thursday, 10:15 – 10:35 (A101)

Theory and practice of Gaussian-based models for spatial extremes

Thomas Opitz[1]
Biostatistics and Spatial Processes, INRA Avignon, France[1]
thomas.opitz@inra.fr[1]

Gaussian processes can be considered as the workhorse of spatial statistics, but they are known to be asymptotically independent and therefore lack flexibility for capturing varied extremal dependence behavior. However, by embedding a random variable for the variance or for the mean of a Gaussian process, we can obtain a large variety of joint tail decay rates covering asymptotic independence and asymptotic dependence, and in some cases nontrivial and tractable max-stable limits arise. In this talk with focus on spatial modeling, I will first review dependence summaries (tail correlation function, coefficient of tail dependence function, etc.) and limit processes of such constructions. Conditional distributions, especially expressions for conditional mean functions and conditional variance functions, will be discussed. I will show how such dependence summaries and conditional distributions can be used in practice for statistical inference based on original event data, even with high-dimensional datasets. In particular, tools are presented to estimate dependence parameters and to check the goodness-of-fit of dependence models possessing spatial stationarity.

References

[1] R. Huser, T. Opitz, E. Thibaud. Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, 2017

[2] S. Engelke, T. Opitz, and J. Wadsworth. Extremal dependence of random scale constructions. *arXiv preprint arXiv:1803.04221*, 2018

[3] P. Krupskii, R. Huser, M.G. Genton. Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 2018

Session, time and room:
IS - Model assessment in spatial extremes (Engelke)
Thursday, 9:35 – 10:05 (A002)

Poisson process approximation of thinnings of stationary point processes

Moritz Otto[1]
Karlsruhe Institute of Technology, Germany[1]
moritz.otto@kit.edu[1]

We consider stationary point processes that arise as a dependent thinning from a stationary Poisson process. Based on the assumption that the thinning depends only locally on the underlying process, we derive a result for Poisson process approximation of an appropriate scaling of the thinned process. In its proof we construct an adequate coupling between the thinned process and a Palm version of itself. We discuss implications of our result for the theory of extremes of random spatial structures and present an application in the context of random geometric graphs.

Session, time and room:
Limit theorems
Monday, 14:50 – 15:10 (A101)

Statistical modelling of time-changing joint extremes

Karla Vianey Palacios Ramirez[1]
University of Edinburgh, United Kingdom[1]
s1875879@sms.ed.ac.uk[1]

In this talk, I will introduce a statistical method for modelling the frequency and magnitude of joint extremes over time. The synchronous scedasis function for bivariate extremes here introduced as a function that carries information on the frequency of joint extremes over time. Our methodology can be regarded as a Bayesian semiparametric extension of the model proposed by Einmahl et al (2016, JRSS B) for the bivariate setting. Bayesian semiparametric inference is proposed to estimate the scedasis function based on mixture of Polya trees. Finally, I will touch on some aspects of changepoint analysis on the frequency of the extremes over time. I will resort to the proposed methodologies to examine the question: "What is the dynamics governing joint extreme losses of FAANG (Facebook, Apple, Amazon, Netflix and Alphabet's Google) stocks over time?" Data from the NASDAQ stock market is used to conduct the proposed inquiry.

Session, time and room:
Finance and economics
Thursday, 12:05 – 12:25 (A002)

New results on the extremes of Gaussian processes with application to the construction of confidence bands for densities

Valentin Konakov[1], Vladimir Panov[2], Vladimir Piterbarg[3]
Higher School of Economics, Russian Federation[1,2], Lomonosov Moscow State University, Russian Federation[3]
vkonakov@hse.ru[1], vpanov@hse.ru[2], piter@mech.math.msu.su[3]

Typically, the construction of confidence sets in density estimation is based on the so-called SBR-type ("Smirnov-Bickel-Rosenblatt") limit theorems, which yield the asymptotic behaviour of

$$\mathcal{D}_n = \sup_{u\in\mathbb{R}} \Big| \frac{\hat{p}_n(u) - p(u)}{\sqrt{p(u)}} \Big|,$$

where $\hat{p}_n$ is an estimate of unknown density function p. Almost all known results of this type deal with the case when $\hat{p}_n(x)$ is either a kernel density estimate or certain wavelet projection density estimates (for instance, based on the Haar wavelets or the Battle-Lemarie wavelets). In fact, the proofs of all SBR-type theorems employ the idea to find a Gaussian process with distribution function which is (in some sense) close to the distribution function of $\mathcal{D}_n$ [2]. For some estimates $\hat{p}_n$ (e.g., for kernel density estimates) one can find a stationary process with this property, but it is more common that the process is non-stationary, of the so-called cyclostationary type [3]. To the best of our knowledge, the SBR-type limit theorems for other projection estimates, which are not related to stationary or cyclostationary Gaussian processes, are not known in the literature, see [2], [4].

In this talk, I would like to present several new results on the asymptotic behaviour of the supremum of nonstationary Gaussian processes, which yield the construction of confidence bands using some other density estimates. For instance, our setup covers the case of the projection to the Legendre polynomials. We show that the constructed confidence sets are honest at polynomial rate to a broad class of probability density functions.

References

[1] Giné, E., and Koltchinskii, V., and Sakhanenko, L. (2004) *Kernel density es- timators: convergence in distribution for weighted sup-norms.* Probability Theory and Related Fields 130(2):167 - 198.

[2] Giné, E. and Nickl, R. (2016) *Mathematical foundations of infinite- dimensional statistical models*, vol 40. Cambridge University Press.

[3] Hüsler, J., Piterbarg, V. and Seleznjev, O. (2003) *On convergence of the uniform norms for Gaussian processes and linear approximation problems.* The Annals of Applied Probability 13(4):1615-1653.

[4] Chernozhukov, V., Chetverikov, D., and Kato, K.(2014) *Anti- concentration and honest, adaptive confidence bands.* The Annals of Statistics 42(5):1787-1818

Session, time and room:
Gaussian processes
Tuesday, 16:15 – 16:35 (A002)

Statistical inference of subcritical strongly stationary Galton–Watson processes with regularly varying immigration

Mátyás Barczy[1], Bojan Basrak[2], Péter Kevei[3], Gyula Pap[4], Hrvoje Planinić[5]
University of Szeged, Hungary[1,3,4], University of Zagreb, Croatia[2,5]
barczy@math.u-szeged.hu[1], bbasrak@math.hr[2], kevei@math.u-szeged.hu[3],
papgy@math.u-szeged.hu[4], hrvoje.planinic@math.hr[5]

Let $(X_k)_{k\geq 0}$ be a Galton–Watson branching process with immigration given by

$$X_i = \sum_{j=1}^{X_{i-1}} A_j^{(i)} + B_i, \qquad i \geq 1,$$

where $\{A, A_j^{(i)} : i, j \geq 1\}$ are i.i.d. nonnegative integer-valued random variables independent of another i.i.d. sequence $\{B, B_i : i \geq 1\}$ of nonnegative integer-valued random variables. We assume that $\mu_A := \mathrm{E}(A) \in (0,1)$, $\mathrm{Var}(A) \in (0,\infty)$, B is regularly varying with tail index $\alpha \in (1,2)$, X_0 is independent of $\{A_j^{(i)}, B_i : i, j \geq 1\}$ and the law of X_0 is the unique stationary distribution of $(X_k)_{k\geq 0}$. Then the conditional least squares estimator of μ_A based on the observations $X_0, X_1, \ldots, X_n$ has the form

$$\widehat{\mu_A}^{(n)} := \frac{\sum_{i=1}^n X_{i-1}(X_i - \mu_B)}{\sum_{i=1}^n X_{i-1}^2}$$

on the set $\bigl\{\sum_{i=1}^n X_{i-1}^2 > 0\bigr\}$, where $\mu_B := \mathrm{E}(B)$. We prove that for any sequence $(a_n)_{n\geq 1}$ of positive numbers with $n\mathrm{P}(X_0 > a_n) \to 1$ as $n \to \infty$, we have

$$a_n^{1/2}\bigl(\widehat{\mu_A}^{(n)} - \mu_A\bigr) \xrightarrow{d} \frac{V^{(2)}}{V^{(1)}} \qquad \text{as } n \to \infty,$$

where $V^{(1)}$ is an $\alpha/2$-stable positive random variable, $V^{(2)}$ is a symmetric $2\alpha/3$-stable random variable, $V^{(1)}$ and $V^{(2)}$ are dependent, and $(V^{(1)}, V^{(2)})$ has an operator stable distribution. We can write

$$\widehat{\mu_A}^{(n)} - \mu_A = \frac{\sum_{i=1}^n X_{i-1} M_i}{\sum_{i=1}^n X_{i-1}^2}, \qquad n \geq 1,$$

where

$$M_i := X_i - \mu_A X_{i-1} - \mu_B = \sum_{j=1}^{X_{i-1}} (A_j^{(i)} - \mu_A) + (B_i - \mu_B), \qquad i \geq 1.$$

Our analysis relies on the fact that one can determine the weak limit of the point processes

$$\sum_{\{j\in\{1,\ldots,n\}\,:\,X_j>0\}} \delta_{\left(\frac{X_j}{a_n}, \frac{M_{j+1}}{\sqrt{X_j}}\right)}$$

as $n \to \infty$ on a suitable space of point measures on $(0,\infty)\times\mathbb{R}$. Our approach here is not standard, since we define vague convergence via bounded Borel sets of $(0,\infty)\times\mathbb{R}$ furnished with a somewhat unusual metric (called a localizing structure on $(0,\infty)\times\mathbb{R}$ by Kallenberg [1]) resulting that the bounded Borel sets are those being bounded away from the line $\{0\}\times\mathbb{R}$ instead of being bounded away from the point $(0,0)$.

References

[1] O. Kallenberg, *Random measures, theory and applications*, Springer, Cham, 2017.

Session, time and room:
Limit theorems II
Monday, 17:30 – 17:50 (A101)

Applying Generalized Pareto Distribution to Rounded Data

Ksenija Cindrić Kalin[1], Zoran Pasarić[2]
Meteorological and Hydrological Service, Grič 3, 1000 Zagreb, Croatia[1], Department of Geophysics, Faculty of Science, University of Zagreb, Horvatovac b.b, Zagreb, Croatia[2]
cindric@cirus.dhz.hr[1], zpasaric@gfz.hr[2]

Under broad assumptions a Generalized Pareto (GP) distribution is an asymptotically correct model for excesses over any fixed, sufficiently high threshold. Being based on excesses, the GP distribution better utilizes the available data than does the Generalized Extreme Value distribution, which is based on block maxima. This is the reason why the peak-over-threshold (POT) approach combined with the GP model have gained considerable popularity over the last decades. Thus, the methods for fitting the parameters of GP distribution are well developed and understood. However, when these methods (in particular the maximum likelihood method and the method of L-moments) are applied to rounded data unexpected difficulties can arise. The problem is revealed when the GP distribution is fitted to some real dry-spell duration series and it is further analysed by performing systematic Monte Carlo simulations. The results depend on (1) the actual amount of rounding as determined by the actual data range (measured by the GP scale parameter, σ) vs. rounding increment (Δx), combined with (2) application of a certain (sufficiently high) threshold and working with the series of excesses instead of the original series. For a moderate amount of rounding (e.g. $\sigma/\Delta x \geq 4$), which is commonly met in practice (at least regarding the dry-spell data), and no threshold is applied, the classical methods work reasonably well. If cutting at the threshold is applied to rounded data, which actually can not be avoided when dealing with a GP distribution, then classical methods applied in a standard way can lead to very erroneous estimates, even if the rounding itself is moderate. In this case, it is necessary to adjust the theoretical location parameter for the series of excesses. The other possibility is to add an appropriate uniform noise to the rounded data (the so-called jittering). In essence this reverses the process of rounding, and, therefore the common methods can be applied as usual. Finally, if rounding is too coarse (e.g. $\sigma/\Delta x \sim 1$), then none of the above recipes would work, and, thus, specific methods for rounded data must be employed.

Session, time and room:
Data contamination
Thursday, 16:40 – 17:00 (A001)

Heavy-Tailed Processes in Degradation Analysis

Chien-Yu Peng[1]
Academia Sinica, Taiwan[1]
chienyu@stat.sinica.edu.tw[1]

Degradation data are widely analyzed using stochastic processes, and Gaussian processes are particularly common. In this article, we propose a statistical model using a heavy-tailed process to assess the lifetime information of highly reliable products. This model is statistically plausible and demonstrates substantially improved fit when applied to real data. A computationally accurate approach is proposed to calculate the first-passage-time density function of the heavy-tailed process; related properties are investigated as well. In addition, several case applications are performed to demonstrate the advantages of the proposed process.

Session, time and room:
Product quality
Thursday, 9:25 – 9:45 (A102)

A LASSO-type model for the bulk and tail of a heavy-tailed response

Miguel de Carvalho[1], Patrícia de Zea Bermudez[2], Soraia Pereira[3], Paula Pereira[4]
CEAUL and School of Mathematics, University of Edinburgh, United Kingdom[1], CEAUL and Faculdade de Ciências, Universidade de Lisboa, Portugal[2,3], ESTSetúbal, Instituto Politécnico de Setúbal and CEAUL, Portugal[4]
Miguel.deCarvalho@ed.ac.uk[1], pcbermudez@fc.ul.pt[2], soraia.gpereira@gmail.com[3], paula.pereira@estsetubal.ips.pt[4]

As widely known, in an extreme value framework, interest focuses on modelling the most extreme observations—disregarding the central part of the distribution; commonly, the effort centers on modelling the tail of the distribution by the generalized Pareto distribution, in a Peaks over threshold framework. Yet, in most practical situations it would be desirable to model both the bulk of the data along with the extreme values.

In this talk, I will introduce a novel regression model for the bulk and the tail of a heavy-tailed response. Our regression model builds over the extended generalized Pareto distribution, as recently proposed by [1]. The proposed model allows us to learn the effect of covariates on a heavy-tailed response via a LASSO-type specification conducted via a Lagrangian restriction. The performance of the proposed approach will be assessed through a simulation study, and the method will be applied to a real data set. One ultimate purpose is to use the methodology in a calibration setup.

References

[1] P. Naveau, R. Huser, P. Ribereau, and A. Hannart, Modeling Jointly Low, Moderate, and Heavy Rainfall Intensities without a Threshold Selection, *Water Resources Research*, 52, 2753-2769, 2016.

Session, time and room:
Computationally intensive methods
Thursday, 11:40 – 12:00 (A101)

Value-at-risk modeling using the peaks-over-threshold method with a non-stationary threshold

Jan Picek[1]
Technical University of Liberec, Czech Republic[1]
jan.picek@tul.cz[1]

The contributions presents a methodology for estimating high quantiles of distributions of financial losses based on peaks-over-threshold (POT) analysis with a time-dependent threshold. When a significant trend is present in the data, no fixed threshold in the POT models is suitable over longer period of time. We use of a time-dependent threshold based on the quantile regression methodology because also the limiting distribution of normalied excesses of a regression quantile thresholds is the Generalized Pareto. The choice of regression model is based on the regression rank score test. The extreme value models are applied to estimate value-at-risk.

Session, time and room:
Peaks over threshold
Tuesday, 14:50 – 15:10 (A002)

Pitfalls of data-driven peaks-over-threshold analysis: perspectives from extreme ship motions

Vladas Pipiras[1]
University of North Carolina at Chapel Hill, United States[1]
pipiras@email.unc.edu[1]

A popular peaks-over-threshold (PoT) method of Extreme Value Theory to quantify the probabilities of rare events is examined here on data generated from a nonlinear random oscillator model, describing a qualitative behavior of rolling of a ship in irregular seas. The restoring force in the oscillator model has a softening shape associated with the ship rolling application, and the response is also made bounded, so as to eliminate the possibility of "capsizing." As a result, the tail of the resulting probability density function of the response undergoes three regimes: the Gaussian core, the heavy tail and the short bounded tail. By considering several scenarios where data are available in one but not another regime, it is shown that the PoT method can produce unsatisfactory results. Some refined methods from Extreme Value Theory, for example, those based on mixture models, are also examined, but without much success. It is thus argued that a data-driven application of the PoT method may fail, if the physical aspects of the system under study are not taken into account.

Session, time and room:
Peaks over threshold
Tuesday, 15:15 – 15:35 (A002)

Anchoring the tail process

Bojan Basrak[1], Hrvoje Planinić[2]
University of Zagreb, Croatia[1,2]
bbasrak@math.hr[1], hrvoje.planinic@math.hr[2]

The tail process of a stationary time series, if it exists, describes the local behavior of the series around a typical exceedance over a large threshold. We intuitively explain that, for a large class of weakly dependent time series, the distribution of the tail process can be viewed as a length–biased version of the asymptotic distribution of a typical cluster of extremes. In order to obtain the latter distribution from the tail process, one needs to debias, and we formalize this idea by introducing the so–called concept of anchoring. We further exploit this relationship to give an another view on the properties of the tail process, such as the time–change formula. If time permits, we discuss the so–called complete point process convergence result.

Session, time and room:
IS - Time series extremes (Kulik)
Thursday, 15:10 – 15:40 (A002)

Some asymptotic results of the conditionally Gaussian processes

Goran Popivoda[1], Siniša Stamatović[2]
University of Montenegro, Faculty of Science and Mathematics, Montenegro[1,2]
goranp@ucg.ac.me[1], sstamatovic@ucg.ac.me[2]

We will obtain the exact asymptotic behaviour of probability $\mathbf{P}\left(\sup_{t\in[0,T]} X_i(t) > u\right)$, as $u \to \infty$, $i = 1, 2$, where $X_1(t) = \xi(t) + \eta(t)$, $\xi(t)$ is a Gaussian zero mean stationary process, and $\eta(t)$ another random process being independent of $\xi(t)$, and $X_2(t) = \chi_n(t) + \eta(t)$, $\chi_n(t)$ is a chi-process with n degrees of freedom generated by n independent copies of process ξ; $n > 1$ and $T > 0$.

Session, time and room:
Gaussian processes
Tuesday, 16:40 – 17:00 (A002)

Characterization results for exponential and Pareto claims in the classical risk model

Georgios Psarrakos[1]
University of Piraeus, Greece[1]
gpsarr@unipi.gr[1]

In many situations, the computation or the numerical estimation of ruin probability is a challenging problem. In this talk, we give some characterization results in the classical risk model. In particular, we consider the distribution of the deficit at the time of ruin and we provide some distributional properties in the case where the claim size distribution has a light or heavy tailed distribution, such as exponential or Pareto claims. Examples are given to illustrate our results.

This work has been partly supported by the University of Piraeus Research Center.

References

[1] Albrecher, H. and Kortschak, D. (2009). On ruin probability and aggregate claim representations for Pareto claim size distributions. *Insurance: Mathematics and Economics* 45, 362–373.

[2] Gerber, H.U., Goovaerts, M.J. and Kaas, R. (1987). On the probability and severity of ruin. *ASTIN Bulletin* 17, 151–163.

[3] Ramsay, C.M. (2003). A solution to the ruin problem for Pareto distributions. *Insurance: Mathematics and Economics* 33, 109–116.

Session, time and room:
Ruin probability
Tuesday, 12:05 – 12:25 (A102)

Analysis and Modeling of the drougth by the use of the max stable processes in the north-east of Algeria

Abdelaziz Rassoul[1], Omar Tami[2]
Higher National School of Hydraulics, Algeria[1], University of Blida 1, Algeria[2]
a.rassoul@ensh.dz[1], tamiblida@yahoo.fr[2]

The main aims of this paper is to study of the extreme rainfall events which manifestly exhibit a spatial dependence, we present a spatial modelisation of the extremes rainfall by a max stable processes, this study is an application in some bassins in the north-east of Algeria, we justificate the best model by some indicateur statistics as bias and rmse.

Next, we investigate the proposed model to evaluate their ability to predict drought and the quantitative value of drought indices, the standardized precipitation index (SPI), and the standardized precipitation evapotranspiration index (SPEI). The SPI/SPEI values may contain a one/three/six-month dry and a one/three/six-month wet period in short-term periods, and this causes instability. For this reason, 4 models for SPI/SPEI (12 months)

were trained and tested by these methods, respectively.

Session, time and room:
Environmental extremes
Monday, 14:25 – 14:45 (A002)

Extremal dependence modelling for aggregated data

Enkelejd Hashorva[1], Simone Padoan[2], Stefano Rizzelli[3]
Department of Actuarial Science, UNIL, Switzerland[1], Department of Decision Sciences, Bocconi University, Italy[2], Institute of Mathematics, EPFL, Switzerland[3]
enkelejd.hashorva@unil.ch[1], simone.padoan@unibocconi.it[2], stefano.rizzelli@epfl.ch[3]

Extreme value theory has been succesfully applied to numerous statistical problems, where it is of interest to model the probability that multiple extreme events take place simultaneously. Yet, when aggregated data are considered, few results are available, especially in high dimensions. Data aggregation is unavoidable in the statistical anlysis of several complex phenomena (e.g. internet traffic flows), due to the vast amount of information produced. Herein, we consider aggregated data in the form of maxima computed over a random number N of observations. We characterize their extremal dependence structure, for different tail behaviours of N. Exploiting an inversion method, we also provide a semi-parametric procedure to infer the extremal dependence structure of unagraggated (unobservable) data and establish its asymptotic theory.

Session, time and room:
OCS - Extremal dependence modelling (Padoan)
Tuesday, 11:40 – 12:00 (A101)

Power variations for a class of Brown-Resnick processes

Christian Robert[1]
UCBL Université Claude Bernard Lyon 1, France[1]
christian.robert@univ-lyon1.fr[1]

We consider the class of simple Brown-Resnick max-stable processes whose spectral processes are continuous exponential martingales. We develop the asymptotic theory for the realized power variations of these max-stable processes, that is, sums of powers of absolute increments. We consider an infill asymptotic setting, where the sampling frequency converges to zero while the time span remains fixed. More specifically we obtain biased central limit theorems whose bias depend on the local times of the differences between the logarithms of the underlying spectral processes. We also discuss the estimation of the integral of the extreme value index function for such a class of max-stable processes by considering the normalized total variation.

Session, time and room:
Max-stable fields
Friday, 11:15 – 11:35 (A101)

Discrepancy method for extremal index estimation

Natalia Markovich[1], Igor Rodionov[2]
V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russian Federation[1,2]
nat.markovich@gmail.com[1], vecsell@gmail.com[2]

Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables (r.v.s) with a cumulative distribution function (cdf) $F(x)$, denote $M_n = \max\{X_1, ..., X_n\}$. Recall, that the stationary sequence $\{X_n\}_{n\geq 1}$ is said to have the extremal index (EI) $\theta \in [0, 1]$ if for each $0 < \tau < \infty$ there is a sequence of real numbers $u_n = u_n(\tau)$ such that $\lim_{n\to\infty} n(1 - F(u_n)) = \tau$ and $\lim_{n\to\infty} P\{M_n \leq u_n\} = e^{-\tau\theta}$ hold.

The problem is that nonparametric estimators of EI require usually the choice of a threshold parameter u and/or a declustering parameter b or r, [1]. Our aim is to propose a nonparametric tool to find one parameter, e.g., u, b or r, by samples of moderate sizes. For this purpose, we extend the so-called discrepancy method, [2], for EI estimation in case of heavy-tailed distributions. The idea of this method is to find an unknown parameter h of the cdf as a solution of the discrepancy equation $\rho(F_h, \widehat{F}_n) = \delta$, where $\widehat{F}_n$ is an empirical cdf of a sample and $\rho(\cdot, \cdot)$ is a metric in the space of cdf's.

Following [3], define $T_1(u) = \min\{j \geq 1 : M_{1,j} \leq u, X_{j+1} > u | X_1 > u\}$, where $M_{1,j} = \max\{X_2, ..., X_j\}$, $M_{1,1} = -\infty$. Observations of $T_1(u_n)$ normalized by the tail function $\{Y_i = \overline{F}(u_n)T_1(u_n)_i\}$, $i = 1, ..., L$, $L = L(u_n)$, are such that $P\{\overline{F}(u_n)T_1(u_n) > t\} \to \theta \exp(-\theta t) =: 1 - G(t, \theta), t > 0$, as $n \to \infty$ under a specific mixing condition, [3].

In our work we propose a normalization of the von Mises-Smirnov (M-S) statistic ω_n^2 and use it as a metric ρ in the discrepancy equation to select the optimal value of a parameter. Denote for some estimator $\hat{\theta}(u)$ of EI

$$\omega_k^2(\hat{\theta}(u)) = \sum_{i=0}^{k-1} \left(\frac{G(Y_{(L-i)}, \hat{\theta}(u)) - G(Y_{(L-k)}, \hat{\theta}(u))}{1 - G(Y_{(L-k)}, \hat{\theta}(u))} - \frac{k - i - 0.5}{k} \right)^2 + \frac{1}{12k}. \qquad (2)$$

A value of u can be found as a solution of the discrepancy equation with regard to any consistent nonparametric estimator of EI. The calculation of (2) by the entire sample may lead to the lack of a solution of the discrepancy equation regarding u. The selection of k and δ remains a problem. To overcome this problem we prove that the limit distribution of $\omega_k^2(\hat{\theta}(u))$ is the same as for the M-S statistic. Then its quantiles may be used as δ.

Theorem 1 Let $L - k \to \infty$, $k \to \infty$ and the estimator of EI $\widehat{\theta} = \widehat{\theta}_{m_n}$ be such that $\sqrt{m_n}(\widehat{\theta}_{m_n} - \theta) \to^d \zeta$ as $n \to \infty$, where the r.v. ζ has a non-degenerate cdf F. Assume that the sequence m_n is such that $k/m_n = o(1)$ and $(\ln L)^2/m_n = o(1)$ as $n \to \infty$. Then

$$\hat{\omega}_k^2(\widehat{\theta}_{m_n}) \to^d \xi \sim A_1$$

holds, where A_1 is the limit distribution function of the M-S statistic.

References

[1] J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, *Statistics of Extremes: Theory and Applications*, Wiley, 2004

[2] V. N. Vapnik, N.M. Markovich, A.R. Stefanyuk, Rate of convergence in L_2 of the projection estimator of the distribution density. *Automat. Rem. Contr.*, 1992

[3] C.A.T. Ferro, J. Segers, Inference for Clusters of Extreme Values. *J. R. Statist. Soc. B*, 2003

Session, time and room:
OCS - Inferences on extremal index (Markovich)
Tuesday, 15:15 – 15:35 (A101)

Bayesian spatial clustering of extremal behaviour for hydrological variables

Christian Rohrbeck[1], Jonathan Tawn[2]
Lancaster University, United Kingdom[1,2]
c.rohrbeck@lancaster.ac.uk[1], j.tawn@lancaster.ac.uk[2]

To address the need for efficient inference for a range of hydrological extreme value problems, spatial pooling of information is the standard approach for marginal tail estimation. We propose the first extreme value spatial clustering methods which account for both the similarity of the marginal tails and the spatial dependence structure of the data to determine the appropriate level of pooling. Spatial dependence is incorporated in two ways: to determine the cluster selection and to account for dependence of the data over sites within a cluster when making the marginal inference. We introduce a statistical model for the pairwise extremal dependence which incorporates distance between sites, and accommodates our belief that sites within the same cluster tend to exhibit a higher degree of dependence than sites in different clusters. We use a Bayesian framework which learns about both the number of clusters and their spatial structure, and that enables the inference of site-specific marginal distributions of extremes to incorporate uncertainty in the clustering allocation. The approach is illustrated using simulations and an analysis of daily precipitation levels in Norway.

Session, time and room:
Bayesian methods for environmental extremes
Monday, 11:20 – 11:40 (A002)

Generalized Pareto models associated with stable mixtures

David Bolin[1], Holger Rootzén[2], Anna Rudvik[3]
Chalmers, Sweden[1,2,3]
david.bolin@chalmers.se[1], hrootzen@chalmers.se[2], Anna.Rudvik@astrazeneca.com[3]

Stable mixture generalized extreme value (GEV) models provide analogues of components of variance models, time series models, and spatial and continuous parameter models for, say, annual maxima. However, often, e.g. in flood risk estimation, it is extremes which occur at the same time as modeled by multivariate generalized Pareto (GP) distributions, and not annual maxima which can occur at different times for different components that are at the center of interest. In this talk we derive explicit representations of the GP distributions associated with the stable mixture GEV distributions. These distributions inherit the appealing properties of the GEV distributions, and additionally allow for simple explicit formulas for densities, and hence for easier ML estimation than competing GP models. In particular we will study GP distributions associated with the popular Reich and Shaby (2012) GEV spatial model. An important challenge is that extremes, say rainstorms, may hit a some, but not all, spatial locations, so that some, but not all, components are extreme. For GP modelling this means that models must be able to accommodate cases where some of the components have mass on their lower boundary, which may be finite or equal to minus infinity.

References

[1] B. J. Reich and B. A. Shaby, A hierarchical max-stable spatial model for extreme precipitation, *Ann. Appl. Statist. 6, 1430-1451*, 2012

Session, time and room:
Peaks over threshold
Tuesday, 14:00 – 14:20 (A002)

Extremal path variation asymptotics of Lévy processes with applications

Jan Rosiński[1]
University of Tennessee, United States[1]
jrosinsk@utk.edu[1]

We extend the celebrated S.J. Taylor's result on the exact asymptotic estimate of path variation of Brownian to the general class of Lévy processes. In particular, if a Lévy process has no Gaussian part, we show that its paths belong to a Wiener class, whose elements have well understood extremal variational behavior. Series expansions of such processes converge pathwise in the strong ϕ-variation norm, which makes them useful in applications, such as to construct approximate solutions to random differential equations driven by Lévy processes.

This talk is based on a joint work with Andreas Basse-O'Connor and Jørgen Hoffmann-Jørgensen.

Session, time and room:
IS ST - Infinitely divisible models and their extremes (Samorodnitsky)
Friday, 10:10 – 10:40 (A001)

How to tell a tale of two tails?

Parthanil Roy[1]
Indian Statistical Institute, Bangalore, India[1]
parthanil.roy@gmail.com[1]

A branching random walk is a system of growing particles that starts from one particle at the origin with each particle branching and moving independently of the others after unit time. In this talk, we shall discuss how the tails of progeny and displacement distributions determine the extremal properties of branching random walks. In particular, we have been able to verify two related conjectures of Eric Brunet and Bernard Derrida in a few cases that were open before.

This talk is based on a joint work with Souvik Ray (M. Stat dissertation work at Indian Statistical Institute), Rajat Subhra Hazra (Indian Statistical Institute) and Philippe Soulier (University of Paris Nanterre).

Session, time and room:
IS - Extremes of branching walks and free fields (Roy)
Monday, 11:40 – 12:10 (A001)

Robust quantile estimation under bivariate extreme value models

Sojung Kim[1], Kyoung-Kuk Kim[2], Heelang Ryu[3]
Leibniz Universität Hannover, Germany[1], KAIST, Korea, Republic of[2,3]
Sojung@stochastik.uni-hannover.de[1], catenoid@kaist.edu[2], happylang@kaist.ac.kr[3]

Risk quantification in extreme events has been a crucial but challenging issue. Extreme value theory (EVT) has been developed to provide sufficient theoretical instruments in that regard. This paper deals with the problem of estimating quantiles for bivariate Generalized Extreme Value (GEV) distributions in a robust manner. In particular, we focus on the misspecification error of a dependence structure as it is often erroneously estimated with little data in extreme events. This paper, on the one hand, follows the stream of literature on robust risk measurement, and on the other hand, advances the robust analysis of dependence in EVT. We provide explicit bounds for tail dependence functions, based on which target quantiles can be easily computed.

Session, time and room:
Best student paper II
Tuesday, 10:15 – 10:35 (A101)

Principal Component Analysis for Multivariate Extremes

Holger Drees[1], Anne Sabourin[2]
University of Hamburg, Germany[1], Telecom ParisTech, France[2]
holger.drees@uni-hamburg.de[1], anne.sabourin@telecom-paristech.fr[2]

It is well known that the first order behavior of multivariate regularly varying random vectors above large radial thresholds is determined by the homogeneous limit measure. If, for a high dimensional vector, the support of this measure is concentrated on a lower dimensional subspace, identifying this subspace and thus reducing the dimension will facilitate a refined statistical analysis. In this work we consider applying standard Principal Component Analysis (PCA) to a rescaled version of radial thresholded observations.

Within the statistical learning framework of empirical risk minimization, our main focus is to analyze the squared reconstruction error for the exceedances over large radial thresholds. We prove that the empirical risk converges to the true risk, uniformly over all projection subspaces. As a consequence, the best projection subspace is shown to converge in probability to the optimal one, in terms of the Hausdorff distance between their intersections with the unit sphere. In addition, if the exceedances are re-scaled to the unit ball, we obtain finite sample uniform guarantees to the reconstruction error pertaining to the estimated projection subspace. Numerical experiments illustrate the relevance of the proposed framework for practical purposes.

Session, time and room:
IS - Mixtures of dependence types (Wadsworth)
Monday, 14:00 – 14:30 (A001)

Limit theorems for the Euler characteristic and Betti number for the Costa-Farber complexes

Takashi Owada[1], Gennady Samorodnitsky[2], Gugan Thoppe[3]
Purdue University, United States[1], Cornell University, United States[2], Duke University, United States[3]
owada@purdue.edu[1], gs18@cornell.edu[2], gugan.thoppe@gmail.com[3]

Costa-Farber complexes are random simplical complexes of a cascade structure, in the sense that for a simplex to be alive all of its faces must be independently alive. In their dynamical version the faces alternate between ON and OFF states. We show that such complexes have a critical dimension, and establish a functional central limit theorem both for the Euler characteristic of the random complex and for the Betti number of the critical dimension.

Session, time and room:
IS - Topological and geometric extremes (Owada)
Thursday, 14:35 – 15:05 (A001)

Modelling changes in the extremal dependence of temperature maxima

Kate Saunders[1]
TU Delft, Netherlands[1]
katerobinsonsaunders@gmail.com[1]

We know climate change to be influencing the distribution of temperature extremes. Given this, there has been much research on how the marginal parameters of univariate distributions, such as the generalised extreme value distribution, are changing temporally due to climate change. However temperature extremes, as with many environmental processes, are inherently spatio-temporal. Therefore to truly understand how temperature extrema are changing with our changing climate, it is necessary to consider how the spatial dependence between temperature extrema may also be changing. In this talk, we present some preliminary results exploring whether the bivariate dependence between spatial temperature extremes is changing. We consider both observational records and future projections of temperature.

Session, time and room:
Modelling changes in environmental extremes
Tuesday, 11:40 – 12:00 (A001)

William Feller and some related extremes

Rene Schilling[1]
TU Dresden, Department of Mathematical Stochastics, Germany[1]
rene.schilling@tu-dresden.de[1]

We will give a survey on William Feller's life, his work related to extreme values and his scientific afterlife.

Session, time and room:
IS ST - Regular variation: history, ideas and people (Basrak)
Friday, 11:15 – 11:45 (A001)

Sampling Sup-Normalized Spectral Functions for Brown–Resnick Processes

Marco Oesting[1], Claudia Schillings[2], Martin Schlather[3]
Universität Siegen, Germany[1], Universität Mannheim, Germany[2,3]
oesting@mathematik.uni-siegen.de[1], c.schillings@uni-mannheim.de[2],
schlather@math.uni-mannheim.de[3]

Sup-normalized spectral functions form building blocks of max-stable and Pareto processes and therefore play an important role in modeling spatial extremes. For one of the most popular examples, the Brown–Resnick process, simulation is not straightforward. In this paper, we generalize two approaches for simulation via Markov Chain Monte Carlo methods and rejection sampling by introducing new classes of proposal densities. In both cases, we provide an optimal choice of the proposal density with respect to sampling efficiency. The performance of the procedures is demonstrated in an example.

Session, time and room:
Max-stable fields
Friday, 12:30 – 12:50 (A101)

Limit theorems for heavy-tailed Boolean models

Matthias Schulte[1]
University of Bern, Institute of Mathematical Statistics and Actuarial Science, Switzerland[1]
matthias.schulte@stat.unibe.ch[1]

The Boolean model Z is obtained as union of all grains of a stationary Poisson process of compact convex sets in $\mathbb{R}^d$. For a geometric functional ψ such as volume or surface area and a compact convex set W one is interested in the behaviour of $\psi(Z \cap W)$. For increasing inradius of the observation window W it is known that $\psi(Z \cap W)$ converges, after rescaling, in distribution to a standard Gaussian random variable if the second moments of the intrinsic volumes of the typical grain are finite. The focus of this talk is on a class of heavy-tailed Boolean models where the latter condition is violated. For this situation distributional limit theorems with alpha-stable limiting distributions are derived.

Session, time and room:
IS - Topological and geometric extremes (Owada)
Thursday, 15:10 – 15:40 (A001)

Tails of optimal transport plans for regularly varying probability measures

Cees de Valk[1], Johan Segers[2]
Koninklijk Nederlands Meteorologisch Instituut, Netherlands[1], Université catholique de Louvain, Belgium[2]
cees.de.valk@knmi.nl[1], johan.segers@uclouvain.be[2]

For the basic case of L_2 optimal transport between two probability measures on a Euclidean space, the regularity of the coupling measure and the transport map in the tail regions of these measures is studied. For this purpose, Robert McCann's classical existence and uniqueness results are extended to a class of possibly infinite measures, finite outside neighbourhoods of the origin. For convergent sequences of pairs of such measures, the stability of the multivalued transport maps is considered, and a useful notion of locally uniform convergence of these maps is verified under light assumptions. Applied to regularly varying probability measures, these general results imply the existence of tail limits of the transport plan and the coupling measure, these objects exhibiting distinct types of homogeneity.

Session, time and room:
IS ST - Regular variation: generalisations and recent advances (Molchanov & Basrak)
Thursday, 10:10 – 10:40 (003)

A model-based analogue model for assessing flood risk in future climates

Benjamin Shaby[1]
Penn State University, United States[1]
bshaby@psu.edu[1]

The hazard of pluvial flooding is largely influenced by the spatial and temporal dependence characteristics of precipitation. The risk of flooding is amplified when extreme precipitation possesses strong spatial dependence due to catchment factors like topography that increase runoff accumulation. Temporal dependence can also increase flood risk as storm water drainage systems operating at capacity can be overwhelmed by extreme precipitation occurring over multiple days. While transformed Gaussian processes are common choices for modeling precipitation, their weak tail dependence may lead to underestimation of flooding risk. Extreme value models like the generalized Pareto processes for threshold exceedances and max-stable models are attractive alternatives, but are difficult to fit when the number of observation sites is large, and are of little use for modeling the bulk of the distribution, which may also be of interest for water management planners. While the atmospheric dynamics governing precipitation are complex and difficult to fully incorporate into a parsimonious statistical model, non-mechanistic analogue methods that approximate those dynamics have proven to be promising approaches to capturing the temporal dependence of precipitation. In this paper, we present a Bayesian analogue approach to modeling precipitation, where the changing spatial dependence across varying intensities is captured using a mixture of spatial student-t processes that can accommodate both strong and weak tail dependence.

Session, time and room:
Flood risks
Thursday, 12:05 – 12:25 (003)

Heavy-Tailed Degree Distributions in Networks: Some History, Some Controversies, Some Prospects

Cosma Shalizi[1]
Carnegie Mellon University, United States[1]
cshalizi@cmu.edu[1]

The "degree" of a node in a network is the number of links it has; the "degree distribution" of a network is the probability mass function of its node degrees. It's long been known that many networks have right-skewed degree distributions, and since around 2000, there's been a lot of excitement about the idea that many degree distributions follow a power law ("scale-free networks"). If true, this would tell us a lot about dynamics on networks (including non-Gaussian collective behavior), and about how networks can form. Controversy over whether it *is* true has driven development of (part of) network data analysis, sometimes by borrowing ideas from extreme-value statistics.

Session, time and room:
IS - Heavy-tails and networks (Davis)
Tuesday, 14:00 – 14:30 (003)

Conditional spatial extremes for North Sea storm severity

Robert Shooter[1]
Lancaster University, United Kingdom[1]
r.shooter@lancaster.ac.uk[1]

Quantifying extreme ocean environments is important for safe and reliable construction and operation of offshore and coastal infrastructure. Extreme value analysis provides a framework within which the marginal and dependence characteristics of extreme ocean environments can be estimated and joint inferences corresponding to very long periods of observation made in the presence of non-stationarity with respect to covariates. We present an approach to the modelling of the dependence structure of significant wave height in North Sea basins motivated by the conditional extremes model proposed by [1], for which the spatial extension is described by [3]. This spatial model has the advantage over other spatial extremes approaches in that the dependence structure that can be modelled is flexible (i.e., the model permits both asymptotic dependence and asymptotic independence), and is conceptually straightforward. As detailed in [2], we apply the spatial conditional extremes model using simple functional forms for the conditional extremes model parameters to directional transects of storm peak data in the North Sea, as well as proposing how this may be used to aid inference of the dependence structure in more computationally-challenging examples.

References

[1] J.E. Heffernan and J.A. Tawn, A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society Series B*, 2004

[2] R. Shooter and E. Ross and J. Tawn and P. Jonathan, On spatial conditional extremes for ocean storm severity, *Enivronmetrics*, 2019

[3] J.L. Wadsworth and J.A. Tawn, Spatial conditional extremes, *In submission*, 2018

Session, time and room:
Environmental extremes
Monday, 14:50 – 15:10 (A002)

L-moments for automatic threshold selection in extreme value analysis

Maria Isabel Fraga Alves[1], Jessica Silva Lomba[2]
CEAUL and DEIO, Faculty of Sciences, University of Lisbon, Portugal[1,2]
isabel.alves@fc.ul.pt[1], jslomba@fc.ul.pt[2]

In extreme value analysis, sensitivity of inference to the definition of extreme event is a paramount issue, directly translated to the need of an appropriate choice of level in threshold-based methodologies. This choice must be established prior to any formal inference. Under the peaks-over-threshold (POT) approach, fitting a Generalized Pareto distribution to observations above a threshold that is too low yields large bias in estimation (the approximation may not hold). On the other hand, selecting only very high observations results in reduced sample size, which increases variability. As such, a suitable level that addresses this trade-off is required.

Methodologies established in the literature face recurrent challenges: inherent subjectivity of tools such as diagnostics plots, or high computational intensity of processes, making them impractical for performance of batch analysis – see examples in [1], [2], [3].

We suggest a truly automated method for threshold detection, aiming at time efficiency and elimination of subjective judgment. Based on the well established theory of L-moments presented by Hosking [4], this versatile technique can handle batch processing of large collections of extremes data, while presenting good performance on small samples.

L-moments are more robust than their classical counterpart to the presence of very extreme values in a sample and do not present sample size-related bounds, making them appropriate in this context. We take a heuristic approach to automatize a known diagnostics tool, the L-moment ratio diagram, commonly used in Hydrology and Regional Frequency Analysis to discriminate between candidate distributions for regional data.

The technique's performance is evaluated in a large simulation study and illustrated with significant wave height data sets from the literature. We find that it compares favorably to other state-of-the-art methods regarding the choice of threshold, associated parameter estimation and the ultimate goal of return level estimation.

The authors were partially supported by Fundação para a Ciência e a Tecnologia, I.P. under grant SFRH/BD/130764/2017 (J.S.L.) and through project UID/MAT/00006/2019 (J.S.L. and M.I.F.A.).

References

[1] P. J. Northrop and C. L. Coleman, Improved threshold diagnostic plots for extreme value analyses, *Extremes*, 17, 289–303, 2014

[2] J. Lee, Y. Fan and S. A. Sisson, Bayesian threshold selection for extremal models using measures of surprise, *Computational Statistics and Data Analysis*, 85, 84–99, 2015

[3] B. Bader, J. Yan and X. Zhang, Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate, *The Annals of Applied Statistics*, 12(1), 310–329, 2018

[4] J. R. M. Hosking, The theory of probability weighted moments, *Research Report RC12210 – IBM Research*, 1986

Session, time and room:
Computationally intensive methods
Thursday, 12:05 – 12:25 (A101)

New measures of dependence for multivariate extremes

Emma Simpson[1], Jonathan Tawn[2], Jennifer Wadsworth[3]
Lancaster University, United Kingdom[1,2,3]
e.simpson3@lancaster.ac.uk[1], j.tawn@lancaster.ac.uk[2], j.wadsworth@lancaster.ac.uk[3]

When studying multivariate extremes, it is useful to consider the tail dependence properties of the variables. The cases of asymptotic dependence and asymptotic independence are often examined, but do not fully cover situations where only subsets of variables can be simultaneously large, while the others are of smaller order. Under a regular variation assumption, we present a new set of indices that reveal aspects of the extremal dependence structure not available through any existing measures of dependence. We discuss theoretical properties of these indices and provide examples for some common distributions. We also use these measures to develop methodology for determining the dependence structure of multivariate extremes, and apply this to UK river flow data, estimating the probabilities of different subsets of sites being large simultaneously.

Session, time and room:
IS - Mixtures of dependence types (Wadsworth)
Monday, 14:35 – 15:05 (A001)

Aleatory and Epistemic Uncertainty in Extreme Distributions

Andrzej Kijko[1], Ansie Smit[2], Alfred Stein[3]

University of Pretoria Natural Hazard Centre, Department of Geology, South Africa[1], University of Pretoria Natural Hazard Centre, Department of Geology, Department of Statistics, South Africa[2], Department of Statistics, University of Pretoria, South Africa, and Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, Netherlands[3]

andrzej.kijko@up.ac.za[1], ansie.smit@up.ac.za[2], a.stein@utwente.nl[3]

This study presents a versatile methodology that accounts for aleatory and epistemic uncertainty in extreme distributions when modelling natural hazards. On extreme natural phenomena we come across three types of data: prehistoric data, historic data and instrumentally recorded data. Each of these contains different types of uncertainty. Our methodology is developed to deal with these three data types in an integrated way. It accounts for incomplete datasets, uncertainty associated with the observed event sizes, the applied distributions, and with the validity of occurrence of events in the dataset. We addressed these types of uncertainty using convolution and mixture distributions, and weighing the likelihood function. Likelihood functions combining the different data types allow for applying both maximum likelihood estimation (MLE) and Bayesian inference (BI) of the parameters. We tested the methodology on a synthetic earthquake dataset, with various combinations of uncertainty investigated. Estimates of the parameters yielded markedly different results, with BI providing more precise estimates than MLE. This in turn has a large effect on estimates of the return periods of natural hazards.

Session, time and room:
Testing and inference
Thursday, 11:40 – 12:00 (A102)

Influence of Climate Change on Extreme Weather Events

Ken Kunkel[1], Richard Smith[2]

North Carolina Institute for Climate Studies and North Carolina State University, United States[1], University of North Carolina, United States[2]

kekunkel@ncsu.edu[1], rls@email.unc.edu[2]

Extreme weather events have huge impact on society, human health and the economy. The widespread perception that extreme events have become both more frequent and more extreme in recent years has led to extensive discussion about the connection with human-induced climate change and the danger of even further increases in extreme events in the future. This talk will survey these topics with particular reference to the very extreme precipitation events in Houston (Texas, USA) associated with Hurricane Harvey in 2017. We construct a new dataset consisting of extreme precipitation events over the entire Gulf of Mexico region from 1949-2018; Harvey is by far the most extreme event over this time frame but an analysis of other high precipitation events shows a clear association with increases in both sea surface temperature and atmospheric carbon dioxide. Models are developed for the relative risk of a Harvey-type event in a world subject to anthropogenic climate forcings compared with one that is not, and for the projected increases in probabilities

of such events in the future. The statistical methods are based on extreme value theory, using Bayesian methods to assess uncertainties, and we address various challenges in the combination of observational and climate model data.

Session, time and room:
IS - Detection and attribution of climate change (Cooley)
Thursday, 15:10 – 15:40 (003)

On maxima of stationary fields: limiting probabilities and the extremal index

Natalia Soja-Kukieła[1]
Nicolaus Copernicus University, Poland[1]
natas@mat.umk.pl[1]

Let $\{X_{\mathbf{n}} : \mathbf{n} \in \mathbb{Z}^d\}$ be a weakly dependent stationary random field with maxima $M_A := \max\{X_{\mathbf{i}} : \mathbf{i} \in A\}$ for nonempty finite $A \subset \mathbb{Z}^d$ and $M_{\mathbf{n}} := \max\{X_{\mathbf{i}} : \mathbf{1} \leq \mathbf{i} \leq \mathbf{n}\}$ for $\mathbf{n} \in \mathbb{N}^d$. In a general setting we will prove that

$$P(M_{(n,n,\ldots,n)} \leq v_n) = \exp(-n^d P(X_{\mathbf{0}} > v_n, M_{A_n} \leq v_n)) + o(1),$$

for some increasing sequence of sets A_n of size $o(n^d)$, which is a multidimensional counterpart of the main theorem of O'Brien's paper [2]. We will show that for a class of fields satisfying a local mixing condition, including m-dependent ones, the above result holds with a constant finite A replacing A_n. We will present new formulas for the extremal index of the field $\{X_{\mathbf{n}}\}$. The formulas will be compared with the ones established by Jakubowski and Soja-Kukieła [1].

References

[1] A. Jakubowski, N. Soja-Kukieła, Managing local dependencies in asymptotic theory for maxima of stationary random fields, *Extremes*, 2018

[2] G. O'Brien, Extreme values for stationary and Markov sequences, *Ann. Probab.*, 1987

[3] N. Soja-Kukieła, On maxima of stationary fields, available at arXiv:1810.04496, 2018

Session, time and room:
Limit theorems
Monday, 15:15 – 15:35 (A101)

New Weighted Tail Index Estimator Under Random Censorship

Brahim Brahimi[1], Djamel Meraghni[2], Abdelhakim Necir[3], Louiza Soltane[4]
Laboratory of Applied Mathematics, University Mohamed Khider of Biskra, Algeria[1,2,3,4]
brah.brahim@gmail.com[1], djmeraghni@yahoo.com[2], necirabdelhakim@yahoo.fr[3],
louiza_stat@yahoo.com[4]

A tail empirical process for heavy-tailed and right-censored data is introduced and its Gaussian approximation is established. In this context, a (weighted) new Hill-type estimator for positive extreme value index is proposed and its consistency and asymptotic normality are proved by means of the aforementioned process in the framework of second-order conditions of regular variation.

References

[1] J. Beirlant, A. Guillou, G. Dierckx, and A. Fils-Villetard, Estimation of the extreme value index and extreme quantiles under random censoring, *Extremes 10, 151-174*, 2007.

[2] J. Beirlant, G. Maribe and A. Verster, Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications, *Insurance Math. Econom., 78, 114-122*, 2018.

[3] B. Brahimi, D. Meraghni, and A. Necir, Approximations to the tail index estimator of a heavy-tailed distribution under random censoring and application, *Math. Methods Statist. 24, 266-279*, 2015.

[4] J. H. J. Einmahl, A. Fils-Villetard and A. Guillou, Statistics of extremes under random censoring, *Bernoulli 14, 207-227*, 2008.

[5] B. M. Hill, A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3, 1163-1174*, 1975.

[6] L. de Haan, and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer, 2006.

[7] E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc. 53, 457-481* , 1958.

[8] J. Worms, and R. Worms, New estimators of the extreme value index under random right censoring, for heavy-tailed distributions, *Extremes 17, 337-358*, 2014.

Session, time and room:
Inference for censored data
Friday, 9:50 – 10:10 (A101)

Max-linear models in random environment

Claudia Klüppelberg[1], Ercan Sönmez[2]
Technical University of Munich, Germany[1], University of Klagenfurt, Austria[2]
cklu@tum.de[1], ercan.soenmez@aau.at[2]

We extend previous work of max-linear models on finite directed acyclic graphs to infinite graphs as well as random graphs, and investigate their relations to classical percolation theory. We show that the critical probability of percolation on the oriented square lattice graph describes a phase transition in the obtained model. Focus is on the dependence introduced by this graph into the max-linear model.

Session, time and room:
IS - Extremes on graphs (Klueppelberg)
Friday, 10:10 – 10:40 (A002)

Applying EVT in finance: climate tail risk and regime shifts in tail behavior

Stefan Straetmans[1]
School of Business and Economics, Maastricht University, Netherlands[1]
s.straetmans@maastrichtuniversity.nl[1]

There seems to be consensus that climate change is characterized by 'more extreme' global weather conditions over time. Following the methodology of Kelly and Jiang (2014), we estimate the tail index of the cross sectional country distribution for temperature and rainfall (211 countries). Repeating this procedure on an annual basis (1900-2015), we produce a time varying common tail risk factor that reflect global extreme weather conditions. Next we investigate (i) whether this factor has predictive power for future stock returns and (ii) whether it is priced in the cross section of stock returns. We find that climate risk predicts stock market returns and outperforms alternative predictors over longer horizons. Another application of EVT into finance concerns the potential existence of regimes in the unconditional tail behavior (tail index, scaling constant, tail dependence structure). The existence of regimes constitutes an additional source of bias in tail index estimators like the Hill statistic. Thus, identifying regime-dependent estimates can mitigate that bias effect. We also discuss why taking account of regime switching may be relevant for risk management, financial regulatory reform or the diversification of portfolio tail risk.

Stefan Straetmans thanks Lisa Kortlang for research assistance on the climate risk project.

Session, time and room:
IS - Financial extremes (Zhou)
Thursday, 10:10 – 10:40 (A001)

A comparative tour through the simulation algorithms for max-stable processes

Marco Oesting[1], Kirstin Strokorb[2]
University of Siegen, Germany[1], Cardiff University, United Kingdom[2]
oesting@mathematik.uni-siegen.de[1], StrokorbK@cardiff.ac.uk[2]

Being the max-analogue of stable stochastic processes, max-stable processes form one of the fundamental classes of stochastic processes. In the 80s and early 90s, they have mainly been studied from a probabilistic angle. Since the early 2000s, methods for statistical inference have been developed and, in parallel, suitable models for subclasses of max-stable processes have been introduced. With these flexible models and tools at hand, max-stable processes have become attractive for practitioners from various areas, in particular from environmental sciences, to adequately assess the risk associated with the spatio-temporal extent of extreme events. Their prevalence as a benchmark is justified by the fact that they arise as the only possible location-scale-limits of maxima of stochastic processes.

However, a serious drawback is that most probabilistic properties of max-stable processes are analytically intractable. Therefore, simulation is often a necessary part of inference of certain characteristics, in particular for future spatial risk assessment. Meanwhile, starting from a general idea coined by [7], a number of approaches to the simulation of max-stable processes have emerged: They include both approximate ([4], [6]) and exact ([1], [2], [3], [5]) simulation procedures, some of them focusing on the particularly difficult problem of simulating within the subclass of Brown-Resnick processes.

In this talk I will give an overview over the existing procedures and put them into perspective of one another and make comparisons with respect to their properties making use of some new theoretical results. A particular emphasis will be given to the important and popular subclass of Brown-Resnick processes, which are max-stable processes that are associated to Gaussian processes.

References

[1] A. B. Dieker and T. Mikosch, Exact simulation of Brown-Resnick random fields at a finite number of locations, *Extremes* **18**(2), 301–314, 2015

[2] C. Dombry, S. Engelke and M. Oesting, Exact simulation of max-stable processes, *Biometrika* **103**(2), 303–317, 2016

[3] Z. Liu, J. H. Blanchet, A. B. Dieker and T. Mikosch, On logarithmically optimal exact simulation of max-stable and related random fields on a compact set, *Bernoulli*, To appear

[4] M. Oesting, Z. Kabluchko, Z. and M. Schlather, Simulation of Brown-Resnick processes, *Extremes* **15**(1), 89–107, 2012

[5] M. Oesting, M. Schlather and C. Zhou, Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation, *Bernoulli* **24**(2), 1497–1530, 2018

[6] M. Oesting and K. Strokorb, Efficient simulation of Brown–Resnick processes based on variance reduction of Gaussian processes, *Adv. in Appl. Probab.* **50**(4), 1155–1175, 2018

[7] M. Schlather, Models for stationary max-stable random fields, *Extremes* **5**(1), 33–44, 2002

Session, time and room:
IS ST - Simulation of rare events and extremes (Zwart)
Tuesday, 10:10 – 10:40 (A001)

Persistence of Averages in Markov Switching Models: A Large Deviations Approach

Michael Stutzer[1]
University of Colorado, United States[1]
michael.stutzer@colorado.edu[1]

The behavior of sample averages, or functions of them, is important in quantitative research. For example, economic and financial time-series, including securities returns, often have sample (i.e. time) average distributions whose tails are thicker (i.e. more persistent) than those associated with normal distributions of the same mean and variance, yet still decay exponentially. Markov Switching processes exhibit this behavior. Unlike Extreme Value Theory, the Large Deviations Rate Function provides a method to quantify the rate of Exponential Decay (and hence persistence of averages) relative to alternative models, and is a simply computed and useful tool for comparing alternative models on this dimension. This rate function is calculated and, for the first time, applied to Markov Switching modeled time series of loan defaults and portfolio returns, emphasizing heretofore unexplored applications to quantitative investment.

Session, time and room:
IS - Financial extremes (Zhou)
Thursday, 9:35 – 10:05 (A001)

Risk Assessment for Product Quality Variation

Yang Su-Fen[1]
National Chengchi University, Taiwan[1]
health887459@gmail.com[1]

Almost all industries strive to offer superior products quality. The quality and loss of products are crucial factors among competitive businesses in global market. Firms widely employ a loss function to assess the risk caused by poor quality of products. From Taguchi's philosophy, the quality of products can be measured by the deviation from the specified target value. In practice, there are many situations where the distributions of the quality variables are not normal. This paper develops a loss function-based control chart to assess the risk of the deviation of quality variable from the target value considering the quality variable with a non-normal distribution. Numerical results show the performance of the risk assessment using the proposed loss function-based control chart.

Session, time and room:
Product quality
Thursday, 9:50 – 10:10 (A102)

Tail Dependence Structures from the Viewpoint of Tail Dependence Matrices

Nariankadu D. Shyamalkumar[1], Siyang Tao[2]
The University of Iowa, United States[1,2]
shyamal-kumar@uiowa.edu[1], siyang-tao@uiowa.edu[2]

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 measure of multivariate tail dependence. It is known that the set of TDMs corresponding to d-dimensional random vectors is a polytope with exponential in d number of facets and vertices. In this talk, we will compare the algorithms in order to test the membership of TDMs for any arbitrary matrix. We will show that if we restrict attention to parametric classes, then the set of TDMs can have much simpler descriptions. This is encouraging because imposing such restrictions in higher dimensions is a common way to escape the curse of dimensionality. Additionally, we will discuss some results that describe the subset of TDMs generated by some popular copula families; in some cases this subset is shown to be a surprisingly small part of the whole space of TDMs. This suggests another dimension along which to evaluate copula families for practical use.

Session, time and room:
Best student paper II
Tuesday, 9:25 – 9:45 (A101)

Cyber claims analysis through Generalized Pareto Regression Trees

Sébastien Farkas[1], Olivier Lopez[2], Maud Thomas[3]
LPSM, Sorbonne Université, France[1,2,3]
sebastien.farkas@sorbonne-universite.fr[1], olivier.lopez@sorbonne-universite.fr[2],
maud.thomas@sorbonne-universite.fr[3]

In this talk, we propose a methodology to analyze the heterogeneity of extreme events according to some covariates using regression trees. Our aim is to identify some homogeneous groups of extreme episodes that can be classified through differents characteristics. Regression trees [1] are good candidates to perform this task, since they allow both regression and classification at the same time. We adapt the construction of Maximum Likelihood Regression Trees (MLRT) introduced by [4] with a Generalized Pareto likelihood as splitting criterion. We propose an application to cyber claims: the heterogeneity of cyber claims databases is caused by the evolution of the risk but also by the evolution in the quality of data and of sources of information through time. We consider a public database, already studied by [3], which is considered as a benchmark for cyber events analysis. Indeed, [3] and [5] analysed to study the data breaches phenomena and [2] proposed an actuarial analysis with a pricing application. Using regression trees, we investigate the heterogeneity of the reported data breaches.

References

[1] Breiman, L., Friedman, J.H., Olshenn R.A., and Stone C.J. (1984), "Classification and regression trees." *Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Sotfware, Belmont, CA*.

[2] Edwards B., Hofmeyr S. and Forrest S. (2016), "Hype and heavy tails: A closer look at data breaches." *Journal of Cybersecurity*, vol. **2** (2057-2085), pp. 3-14.

[3] Eling M. and Loperfido N. (2017), "Data breaches: Goodness of fit, pricing, and risk measurement." *Insurance: Mathematics and Economics*, vol. **75** (0167-6687), pp. 126-136.

[4] Su X., Wang M. and Fan J. (2004), "Maximum Likelihood Regression Trees." *Journal of Computational and Graphical Statistics*, vol. **13** (1061-8600), pp. 586-598.

[5] Wheatley S., Maillart T. and Sornette D. (2016), "The extreme risk of personal data breaches and the erosion of privacy." *European Physical Journal B*, vol. **89** (1434-6036), pp. 7.

Session, time and room:
Statistical learning methods
Tuesday, 16:40 – 17:00 (003)

The effects of uncertainty on design flood estimation

Thordis Thorarinsdottir[1]
Norwegian Computing Center, Norway[1]
thordis@nr.no[1]

Statistical flood frequency analysis is commonly performed based on a set of annual maximum discharge values which are derived from stage measurements via a stage-discharge rating curve model. In such an analysis, several elements are influenced by uncertainties. In particular, design flood estimation techniques often ignore the uncertainty in the underlying rating curve model, treating the discharge values as precise observations. Using data from eight gauging stations in Norway, we investigate the effect of curve and sample uncertainty on design flood estimation by combining results from a Bayesian multi-segment rating curve model and a Bayesian flood frequency analysis. We find that sample uncertainty is the main contributor to the design flood estimation uncertainty. However, under extrapolation of the rating curve, the uncertainty bounds for both the rating curve model and the flood frequency analysis are highly skewed. Ignoring these features may underestimate the potential risk of flooding.

Session, time and room:
IS ST - Hydrology and extremes (Naveau & Rootzen)
Friday, 9:35 – 10:05 (003)

Hierarchical space-time modeling of asymptotically independent exceedances for hourly precipitations in southern France

Jean-Noel Bacro[1], Carlo Gaetan[2], Thomas Opitz[3], Gwladys Toulemonde[4]
Université de Montpellier, France[1], Universita Ca' Foscari di Venezia, Italy[2], INRA, Avignon, France[3], Université de Montpellier, Inria, France[4]
jean-noel.bacro@umontpellier.fr[1], gaetan@unive.it[2], thomas.opitz@inra.fr[3], gwladys.toulemonde@umontpellier.fr[4]

The statistical modeling of space-time extremes in environmental applications is a valuable approach to understand complex dependencies in observed data and to generate realistic scenarios for impact models. Motivated by hourly rainfall data in Southern France presenting asymptotic independence, we propose a novel hierarchical model for high threshold exceedances leading to asymptotic independence in space and time. Our approach is based on representing a generalized Pareto distribution as a Gamma mixture of an exponential distribution, enabling us to keep marginal distributions which are coherent with univariate extreme value theory. The key idea is to use a kernel convolution of a space-time Gamma random process based on influence zones defined as cylinders with an ellipsoidal basis to generate anisotropic spatio-temporal dependence in exceedances. Statistical inference is based on a composite likelihood for the observed censored excesses. The practical usefulness of our model is illustrated on the previously mentioned hourly precipitation data set from a region in Southern France where it improves on an alternative censored Gaussian space-time random field model.

Session, time and room:

IS - Sub-asymptotic spatial extremes (Opitz)
Tuesday, 14:35 – 15:05 (A001)

Liquidity tail risk in the wake of the financial crisis: Evidence from the U.S. stock market

Debbie Dupuis[1], Luca Trapin[2]
HEC Montréal, Canada[1], Università Cattolica del Sacro Cuore Milano, Italy[2]
debbie.dupuis@hec.ca[1], luca.trapin@unicatt.it[2]

Since the last financial crisis, market participants have perceived a deterioration in market liquidity as a consequence of the tightening regulatory constraints. Inspecting common liquidity metrics, recent research in finance however does not find any substantial reduction in liquidity compared to the pre-crisis levels. Going beyond the level of liquidity, this paper investigates liquidity tail risk, studying possible changes in the likelihood of extreme illiquidity events over the post-crisis period. To do that, we define a novel state-space extreme value model and build a robust score-driven filter and smoother of the latent states. Fitting the model to several highly liquid stocks in the S&P500 reveals that tail liquidity risk has increased in recent years. While we find insufficient evidence that this increase is solely attributable to a structural change in the dynamics of liquidity provision due to the post-crisis regulatory restructuring, the substantial increase in liquidity tail risk should prompt policymakers and market participants to take mitigating action.

Session, time and room:
Financial extremes II
Tuesday, 12:05 – 12:25 (A002)

Lamination design of solar panels based on extreme value models

Chih-Chun Tsai[1]
Department of Mathematics, Tamkang University, Taiwan[1]
chihchuntsai@gmail.com[1]

Solar power is renewable and inexhaustible energy sources and has become the clear choice in the world. Solar panels testing procedures are performed to deliver efficient performance and long-lasting durability in development stage. One of the keys to the longevity of solar panels is the lamination process, which encapsulates solar cells while attaching front glass and back sheets. The degree of crosslinking for ethylene vinyl acetate (EVA) sheets affected the quality of the solar panels significantly. The degree of crosslinking for EVA sheets can be acquired by using the extraction method to measure the gel content of EVA sheets. Motivated by lamination data on solar panels, the extreme value model was used to describe the relationship between the degree of cross-linking for EVA sheets and lamination time. Under the specification limits of the degree of cross-linking for EVA sheets, the optimal lamination time of solar panels will be derived. Finally, we addressed

the problem in optimum allocation of the total sample that measured on each lamination time.

Session, time and room:
Industrial applications
Thursday, 16:15 – 16:35 (A101)

Optimizing a new declustering approach for extreme limited time series of high resolution with an application to wind speed design values

George Kallos[1], Platon Patlakas[2], Christos Stathopoulos[3], Christos Tsalis[4]
National and Kapodistrian University of Athens – School of Science-Faculty of Physics Athens, Greece[1,2,3,4]
kallos@mg.uoa.gr[1], platon@mg.uoa.gr[2], chrisstath@mg.uoa.gr[3], tsalischristos@gmail.com[4]

It is commonly accepted in the theory of extreme value analysis, the handling of exceedances over a predefined threshold (POT), fitting a Generalized Pareto distributions (GPD). When the available sample of extremes is limited and requirements for accommodating more events is fundamental, the most acceptable models found in literature usually provide a resample were the independence criteria is often questioned. In this study, we examine the performance of a new method (DeCAUn) resampling irregularly de-clustered wind speed values, originated by a physical de-clustering model approach (DeCA), using wind speed series with high resolution, over the North Sea, Atlantic and the Mediterranean. The proposed method will be compelling against the classical threshold extreme value analysis models for limited time series such as 10 and 15 years, revealing robust estimations in terms of quantiles. A resample through a non-rectangular kernel function, commonly called slotting or slotting autocorrelation approach is considered, transcending the subjective judgment of the optimum threshold in order to obtain approximately independent observations. The Gausian Kernel function is examined over various bandwidth values, optimizing the proposed de-clustering method. Quantile estimations of all models are presented and compared with those based on classical block maxima (BM) approach.

Session, time and room:
Environmental extremes
Monday, 15:15 – 15:35 (A002)

Field Return Rate Prediction within Warranty Period Based on Laboratory Testing Data

Sheng-T Tseng[1]
National Tsing-Hua University, Taiwan[1]
sttseng@stat.nthu.edu.tw[1]

Warranty policy is a useful tool for manufacturers to attract potential customers and to compete with other companies. Therefore, achieving good prediction for the product's field return rate during the warranty period is an important task not only for better understanding of product reliability but also for successful warranty management. Recently, Tseng, Hsu and Lin (2016) proposed a hierarchical failure model to incorporate all failure information from multiple products, under which the empirical Bayes inference is applied for parameter estimation. This approach, however, may suffer from an over-fitting problem and the difficulty of explaining the physical relationship between laboratory and field data. To overcome this difficulty, this paper first constructs a limiting failure probability model for the field failure data via an extreme-value distribution. Second, we summarizes the parameters in the laboratory data in terms of a logistic distribution. Then, a simple prediction model for the product's field return rate can be successfully constructed. From a real case study, it demonstrates that the proposed procedure (by using only one type of laboratory data) can even outperform the approach of existing literature. This finding provides us a very useful information to design an efficient laboratory testing plan for the purpose of field failure rate prediction.

Session, time and room:
Industrial applications
Thursday, 16:40 – 17:00 (A101)

Nonparametric extreme conditional expectile estimation

Stéphane Girard[1], Gilles Stupfler[2], Antoine Usseglio-Carleve[3]
Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France[1,3], University of Nottingham, School of Mathematical Sciences, Nottingham NG7 2RD, United Kingdom[2]
stephane.girard@inria.fr[1], Gilles.Stupfler@nottingham.ac.uk[2], antoine.usseglio-carleve@inria.fr[3]

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 talk, we propose nonparametric estimators of extreme conditional

expectiles based on kernel smoothing techniques. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions. Applications to simulated and real data are provided.

References

[1] Daouia, A., Girard, S., and Stupfler, G., Estimation of tail risk based on extreme expectiles, *Journal of the Royal Statistical Society: Series B*, 2018

[2] Usseglio-Carleve, A., Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors, *Electronic Journal of Statistics*, 2018

Session, time and room:
Best student paper II
Tuesday, 9:50 – 10:10 (A101)

Challenges for extreme-value theory in network science: estimating tail exponents of degree distributions

Dmitri Krioukov[1], Remco van der Hofstad[2], Pim van der Hoorn[3], Ivan Voitalov[4]
Northeastern University, United States[1,3,4], University of Technology Eindhoven, Netherlands[2]
d.krioukov@northeastern.edu[1], r.w.v.d.hofstad@tue.nl[2],
w.vanderhoorn@northeastern.edu[3], i.voitalov@northeastern.edu[4]

One of the pillars of network science is the observation that many real-world networks have a "power-law" degree distribution, meaning that the fraction $f(k)$ of nodes that have degree k satisfies $f(k) \approx k^{-\gamma}$. The interpretation of the $\approx$ sign varies from paper to paper, but the main intuition has always been that $f(k)$ plotted on a log-log scale "looks like" a straight line with slope $-\gamma$. The presence of such degree distributions and in particular the value of the tail exponent γ has many important implications, related for instance to the behavior of infection spreading in networks. Hence, estimating γ is an important part of network analysis. Unfortunately, current approaches have been unsatisfactory and lacking proper rigor. The best example for this is the fact that the "state-of-the-art" procedure in network science [1] uses an estimation scheme (which only looks for pure Zeta distributions) that until very recently [2] was not even proven to be consistent.

The problem of tail-inference is well studied, and at the same time still a very active field in extreme-value theory and statistics. We utilized results in this field in our recent work [3], to properly study tail-exponents in networks. Here we replaced $f(k) \approx k^{-\gamma}$ with the notion of regularly-varying distributions, i.e distributions whose CDF's satisfy $F(k) = \ell(k)k^{-(\gamma-1)}$ with $\ell(k)$ slowly-varying at ∞. These distributions define the MDA of Fréchet, relating the extreme-value index ξ to γ by $\xi = \frac{1}{\gamma-1}$. Techniques for estimating ξ can thus be used to estimate γ. We demonstrate the strength of this approach by implementing a double-bootstrap procedure for Hill's, Moment's and Kernel estimator and applying it to 105 different real-world networks. The results provide the first robust evidence that a significant proportion of networks have regularly-varying degree distributions, many with $2 < \gamma < 3$. Our work clearly demonstrates the potential extreme-value theory can offer for network science. Unfortunately, many network scientists are completely unaware of this.

On the other hand, many problems in degree distributions analysis in network science would pose interesting theoretical questions and challenges in extreme-value theory. In this talk I will explain these problems and challenges, and translate them to the framework of extreme-value theory. Important examples include, consistency results for non i.i.d. data, second-order condition tests, detection of exponential cut-offs in distributions and identifying if distributions belong to certain subclasses of long-tailed (but not regularly-varying) distributions. We believe addressing these problems will lead to fruitful and influential cross-fertilization between the two scientific fields, and stimulate exciting collaborations.

References

[1] Aaron Clauset et al. *Power-law distributions in empirical data*, [Siam Rev. 2009 `https://doi.org/10.1137/070710111`]

[2] Holger Drees et al. *On a minimum distance procedure for threshold selection in tail analysis*, [arXiv:1811.06433, `https://arxiv.org/abs/1811.06433`]

[3] Ivan Voitalov et al. *Scale-free Networks Well Done* (2018), [arXiv:1811.02071, `https://arxiv.org/abs/1811.02071`]

Session, time and room:
New areas for EVT
Tuesday, 12:05 – 12:25 (003)

Extreme Value Theory for Open Set Classification - GPD and GEV Classifiers

Sebastian Engelke[1], Edoardo Vignotto[2]

University of Geneva, Switzerland[1,2]

`sebastian.engelke@unige.ch`[1], `edoardo.vignotto@unige.ch`[2]

Classification tasks usually assume that all possible classes are present during the training phase. This is restrictive if the algorithm is used over a long time and possibly encounters samples from unknown classes. It is therefore fundamental to develop algorithms able to distinguish between known and unknown new data. In the last few years, extreme value theory has become an important tool in multivariate statistics and machine learning. This is due to the fact that the extreme features, rather than the average ones, are often most important for discriminating between different objects. The recently introduced extreme value machine, a classifier motivated by extreme value theory, addresses this problem and achieves competitive performance in specific cases. This algorithm strongly relies on the knowledge about the known classes to infer about the unknowns and gives a non-justified premium to known classes far from all the others. We show that the extreme value machine can thus fail when the geometries of known and unknown classes differ, even if the recognition problem is fairly simple. To overcome these limitations, we propose two new algorithms for open set classification that share similar foundations but are more robust in such cases. We exploit the intuition that test points that are extremely far from the training classes are more likely to be unknown objects. We derive asymptotic results motivated by univariate extreme value theory that make this intuition precise. The effectiveness of

our classifiers is illustrated in simulations and on real data sets. Finally, we state their main strengths and some possible future improvements.

Session, time and room:
Best student paper I
Monday, 17:30 – 17:50 (A002)

Extremal variograms: definitions, estimation and applications to graphical models for extremes

Sebastian Engelke[1], Stanislav Volgushev[2]
University of Geneva, Switzerland[1], University of Toronto, Canada[2]
sebastian.engelke@unige.ch[1], stanislav.volgushev@utoronto.ca[2]

In this talk we introduce the notion of extremal variograms, which provide a new way of summarizing extremal dependence. We introduce rank-based estimation procedures and comment on conditions ensuring consistency and asymptotic normality.

From a statistical point of view, we show that extremal variograms have a distinct advantage compared to other ways of summarizing extremal dependence - zeroes of a simple transformation of the variogram matrix imply conditional independence structures in certain classes of multivariate Pareto distributions. This is similar to multivariate normal distributions where the precision matrix carries information about conditional independence.

Session, time and room:
IS - Model assessment in spatial extremes (Engelke)
Thursday, 10:10 – 10:40 (A002)

Spatial conditional extremes

Jonathan Tawn[1], Jenny Wadsworth[2]
Lancaster University, United Kingdom[1,2]
j.tawn@lancaster.ac.uk[1], j.wadsworth@lancaster.ac.uk[2]

Motivated by the need for flexible spatial dependence models that can be fitted in high dimensions, we present an approach to spatial extreme value theory based on the conditional multivariate extreme value model. Such an approach involves determining the behaviour of the random field when conditioning on a single site being extreme. This permits dependence representations that are much more flexible than those obtained through conditioning on the supremum being large, but there is usually no natural conditioning site. To overcome this, we suggest a joint inference scheme based on all observation locations, and implement an importance sampling algorithm to provide spatial realizations and estimates of quantities conditioning upon the process being extreme at any of one of an arbitrary set of locations.

Session, time and room:
IS - Sub-asymptotic spatial extremes (Opitz)
Tuesday, 15:10 – 15:40 (A001)

An Extreme Value Analysis of the Urban Skyline

Jonathan Auerbach[1], Phyllis Wan[2]
Columbia University, United States[1], Erasmus University Rotterdam, Netherlands[2]
jla2167@columbia.edu[1], wan@ese.eur.nl[2]

The world's urban population is expected to grow fifty percent by the year 2050 and exceed six billion. The major challenges confronting cities, such as sustainability, safety, and equality, will depend on the infrastructure developed to accommodate the increase. Urban planners have long debated the consequences of vertical expansion – the concentration of residents by constructing tall buildings – over horizontal expansion – the dispersal of residents by extending urban boundaries. Yet relatively little work has modeled the vertical expansion of cities if present trends continue and quantified the likelihood and therefore urgency of these consequences.

We regard tall buildings as random exceedances over a threshold and use extreme value analysis to characterize the skyscrapers that will dominate the urban skyline in 2050 if present trends continue. We find that typical skyscrapers will not be noticeably taller, but their number will see a growth of eight percent a year, far outpacing the expected urban population growth of two percent a year. We predict that the tallest skyscraper will likely exceed one thousand meters, but not one mile, despite several grand visions of mile-highs in the architectural history. If a mile-high skyscraper is constructed, it will hold fewer occupants than those previously designed. We predict roughly three-quarters the number of floors of the Mile-High Tower, two-thirds of Next Tokyo's Sky Mile Tower, and half the floors of Frank Lloyd Wright's The Illinois – three prominent mile-high plans. However, we anticipate the relationship between floor and height will vary considerably across cities.

Session, time and room:
New areas for EVT
Tuesday, 11:40 – 12:00 (003)

Degree Growth Rates and Index Estimation in a Directed Preferential Attachment Model

Sidney Resnick[1], Tiandong Wang[2]
Cornell University, United States[1,2]
sir1@cornell.edu[1], tw398@cornell.edu[2]

Preferential attachment is widely used to model power-law behavior of degree distributions in both directed and undirected networks. In a directed preferential attachment model, despite the well-known marginal power-law degree distributions, not much investigation has been done on the joint behavior of the in- and out-degree growth. Also, statistical estimates of the marginal tail exponent of the power-law degree distribution often use the Hill estimator as one of the key summary statistics, even though no theoretical justification has been given. We focus on the convergence of the joint empirical measure for in- and out-degrees and prove the consistency of the Hill estimator. To do this, we first derive the asymptotic behavior of the joint degree sequences by embedding the in- and out-degrees of a fixed node into a pair of switched birth processes with immigration and then establish the convergence of the joint tail empirical measure. From these steps, the consistency of the Hill estimators is obtained. Then we also discuss an estimation procedure that is widely adopted in social network analyses.

Session, time and room:
IS - Heavy-tails and networks (Davis)
Tuesday, 14:35 – 15:05 (003)

Conditional Tail Independence in Archimedean Copula Models

Michael Falk[1], Simone Padoan[2], Florian Wisheckel[3]
University of Würzburg, Germany[1,3], Bocconi University of Milan, Italy[2]
michael.falk@uni-wuerzburg.de[1], simone.padoan@unibocconi.it[2],
florian.wisheckel@uni-wuerzburg.de[3]

Consider a random vector U, whose distribution function coincides in its upper tail with that of an Archimedean copula. We report the fact that the conditional distribution of U, conditional on one of its components, has under a mild condition on the generator function independent upper tails, no matter what the unconditional tail behavior is. We also extended this to Archimax copulas and illustrate our findings with simulated data.

Session, time and room:
OCS - Extremal dependence modelling (Padoan)
Tuesday, 12:05 – 12:25 (A101)

Record events attribution in a climate change context

Philippe Naveau[1], Julien Worms[2]
LSCE/IPSL , CEA-CNRS-UVSQ, Paris Saclay University, France[1], University of Versailles - Paris Saclay, France[2]
philippe.naveau@lsce.ipsl.fr[1], julien.worms@uvsq.fr[2]

The topic of climate change attribution (or at least, changes in the distribution of some particular atmospheric variable), with respect to the increase of green house gases, is a particularly relevant subject in the domain of environmental statistics.

The attribution process generally implies the confrontation of (i) an observed dataset, issued from the so-called « factual » world (including the present or recent times, in other words the « world as-is ») to (ii) a dataset issued either from historical data or from numerical models (simulations of the phenomenon at stake, under a framework where anthropogenic forcings are absent ; this is the so-called « counter-factual » world, the « world that might have been »). Most of the time, this confrontation leads to the computation of some indicator, the most-well known of them being the Fraction of Attributable Risk (FAR). Computation of this FAR indicator generally requires modelling of both datasets and estimation of very small probabilities (the FAR is related to the ratio of those probabilities), and therefore likely calibration of tuning parameters (choice of a threshold for instance), with the inherent problems and statistical uncertainty which come along with their choice.

In a previous study, a strong hypothesis was that the EVT shape parameter was equal in both worlds, and this led to simple and fast computations of an alternative version of FAR for records, far(r), which did not imply separate modelling of the two worlds. The presented work is a continuation of this study, and the method is now flexible enough so that differences in shape parameters (between the factual and counter-factual underlying distributions), rather than only on position or scale parameters, can be accounted for.

Applications in mind are extreme rainfall attribution for instance , although the principle is general and can be applied to temperature, windspeed, or waveheight datasets. Theoretical results are presented for the asymptotic validity of proposed confidence intervals for the far(r) statistic, with both a nonparametric approach and a semi-parametric approach (merits and flaws of both approaches will be discussed if time permits). The subject of the goodness-of-fit of the model to real datasets will also be addressed.

Session, time and room:
Modelling changes in environmental extremes
Tuesday, 12:05 – 12:25 (A001)

Estimation of extremes for Weibull-tail distributions in the presence of random censoring

Rym Worms[1], Julien Worms[2]
Université Paris-Est. UPEC. Laboratoire d'analyse et de mathématiques appliquées (CNRS UMR 8050), France[1], Université Paris-Saclay. Université de Versailles-Saint-Quentin-En-Yvelines. Laboratoire de Mathématiques de Versailles (CNRS UMR 8100), France[2]
rym.worms@u-pec.fr[1], julien.worms@uvsq.fr[2]

The Weibull-tail class of distributions is a sub-class of the Gumbel extreme domain of attraction, and it has caught the attention of a number of researchers in the last decade, particularly concerning the estimation of the so-called Weibull-tail coefficient. In this talk, we propose an estimator of this Weibull-tail coefficient when the Weibull-tail distribution of interest is censored from the right by another Weibull-tail distribution: to the best of our knowledge, this is the first one proposed in this context. A corresponding estimator of extreme quantiles is also proposed. In both mild censoring and heavy censoring (in the tail) settings, asymptotic normality of these estimators and their finite sample behavior are presented.

Session, time and room:
Inference for censored data
Friday, 10:15 – 10:35 (A101)

Using Flood Physics to Learn About Flood Statistics in a Changing World

Daniel Wright[1]
University of Wisconsin-Madison, United States[1]
danielb.wright@wisc.edu[1]

Floods are the most common natural disasters worldwide, and amongst the most damaging. They can also exhibit remarkable variety in terms of the physical processes that cause them—including multiple types of spatially and temporally complex rainstorms, groundwater, snowmelt, ice jams, and debris flows, to name a few. Though well-known to physical hydrologists, these processes are often underappreciated in statistical hydrology, which tends to focus instead on data-driven approaches to "flood frequency analysis" (FFA; the determination of the probability distributions of floods) and for assessing flood nonstationarity. Clearly, however, "real world" distributions of floods and how they are changing arise from real world physical processes.

In this presentation, we argue that FFA can benefit from deeper consideration of the physical processes that cause floods, and from decades of progress on physically-based understanding and numerical simulation of floods. We introduce the concept of "process-based FFA," which though not new, has received relatively little research attention compared with more conventional approaches. We present a three-step framework for process-based FFA: 1. generating of large numbers of rainfall scenarios by coupling "stochastic storm transposition" with high-resolution rainfall fields; 2. using a physics-based numerical model to create a database of cyclostationary state variables including soil moisture and snowpack; 3. resolving the large numbers of combinations of rainfall scenarios and watershed states using

numerical simulation. This framework allows us to reconstruct flood frequencies across entire river networks. We show that these flood frequency reconstructions can be comparable in accuracy to more conventional statistical approaches, and can provide insight into how changes in physical drivers are leading to nonstationarity in flood frequency.

We also discuss two emerging topics that process-based FFA enables. The first is the value of high-resolution convection-resolving regional climate models (RCMs) to help understand possible flood frequency trends in a warming climate. Though subject to biases, these RCMs nonetheless point to important changes in the impacts of organized convective storms on flooding at multiple scales. The second emerging topic is the merging of process-based and statistically-based FFA. This approach leverages physically-based models and multiple observational sources including rainfall remote sensing and long flood records to address a longstanding issue in FFA—the inability to accurately estimate higher-order distribution moments based on flood records alone.

Session, time and room:
IS ST - Hydrology and extremes (Naveau & Rootzen)
Friday, 10:10 – 10:40 (003)

Flexible sub-asymptotic modeling of threshold exceedances using hierarchical mixture models

Raphaël Huser[1], Thomas Opitz[2], Rishikesh Yadav[3]
King Abdullah University of Science and Technology (KAUST), Saudi Arabia[1,3], French National Institute of Agronomic Research (INRA), France[2]
raphael.huser@kaust.edu.sa[1], thomas.opitz@inra.fr[2], RISHIKESH.YADAV@KAUST.EDU.SA[3]

In this work, we develop new flexible univariate tail models for light-tailed and heavy-tailed data, which extend a hierarchical representation of the generalized Pareto (GP) limit for univariate threshold exceedances. Our proposed models, which are constructed by treating the rate parameter of a baseline distribution as a latent variable, can accommodate departure from asymptotic threshold stability in finite samples while keeping the asymptotic GP distribution as a special (or boundary) case. Thanks to their greater flexibility at sub-asymptotic levels, our new models provide robust results at lower thresholds while estimating the tail behavior with lower variability, and may even be used to model the tails and the bulk regions jointly without losing much flexibility for either of them. By embedding our univariate models into a Bayesian hierarchical framework, we can naturally incorporate spatial dependence at the level of the latent process, while assuming that the observed data are independent conditional on this spatially structured latent process. In this Bayesian framework, we design penalized complexity priors for crucial model parameters in order to shrink our flexible models toward a simpler reference GP distribution. We show how to perform Bayesian inference based on threshold exceedances by censoring observations below their corresponding threshold value. In order to fit our models in (moderately) high dimensions, we implement a Markov chain Monte Carlo algorithm that exploits the hierarchical formulation in terms of the latent variables and the conditional independence assumption, allowing us to avoid expensive numerical evaluations of multivariate integrals in censored likelihood expressions by reducing them to univariate censored expressions which can be calculated independently of each other. To accelerate the mixing

of the Markov chains, we construct efficient proposal distributions based on the Metropolis-adjusted Langevin algorithm (MALA), and we develop an adaptive scheme to calibrate the MALA tuning parameters for obtaining appropriate acceptance rates. We demonstrate our new methodology using an extensive simulation study, and by application to air pollution data measured at a large number of monitoring stations in Western United States.

Session, time and room:
Bayesian methods for environmental extremes
Monday, 11:45 – 12:05 (A002)

Diversification and systemic risk shift

Fan Yang[1], Chen Zhou[2]
University of Waterloo, Canada[1], Erasmus University Rotterdam, Netherlands[2]
fan.yang@uwaterloo.ca[1], zhou@ese.eur.nl[2]

The stylized fact that diversification is always preferred has been shown to be reversed for extremely heavy-tailed risks. We show that even for truncated extremely heavy-tailed risks, e.g. catastrophe insurance risks, diversification may still be sub-optimal. By increasing the number of assets, the diversification can transit from sub-optimal to optimal. Such diversification is indeed optimal at the individual level. At the market level, we show that the systemic risk, however, may be increased.

Session, time and room:
IS - Risk analysis in insurance (Peng)
Friday, 12:25 – 12:55 (A002)

Simultaneous confidence bands for extremal quantile regression with splines

Takuma Yoshida[1]
Kagoshima University, Japan[1]
yoshida@sci.kagoshima-u.ac.jp[1]

This paper investigates the simultaneous confidence bands for extremal quantile regression using spline method. For the intermediate order quantile, we construct the spline estimator in a conventional quantile regression framework. The extreme order quantile estimator is obtained by extrapolating the spline estimator for the intermediate order quantile. We establish the asymptotic normality of the spline estimator and the extrapolated estimator for the intermediate and extreme order quantiles. By applying the volume of tube formula to the above two estimators, we construct the simultaneous confidence bands of the conditional quantile for the intermediate and extreme order quantiles. To confirm the performance of the proposed confidence bands, the Monte Carlo simulation and real data example are illustrated.

Session, time and room:
Conditional extremes
Thursday, 14:50 – 15:10 (A101)

Gumbel and Fréchet convergence of the maxima of independent random walks

Jorge Yslas Altamirano[1]
University of Copenhagen, Denmark[1]
jorge.yslas1@gmail.com[1]

We consider point process convergence for sequences of iid random walks. The objective is to derive asymptotic theory for the largest extremes of these random walks. We give sufficient conditions for convergence of properly normalized and centered maximum random walks to the Gumbel or the Fréchet distributions. The proofs heavily depend on subexponential and normal approximations to large deviation probabilities for sums of independent random variables. We also consider the extremes of blocks of a random walk and discuss some extensions of the results.

This talk is based on a joint work with Thomas Mikosch (preprint arXiv: 1904.04607).

Session, time and room:
Best student paper III
Tuesday, 16:15 – 16:35 (A101)

Hierarchical Scale Mixtures for Flexible Spatial Modeling

Benjamin Shaby[1], Likun Zhang[2]
Pennsylvania State University, United States[1,2]
bshaby@psu.edu[1], lfz5044@psu.edu[2]

Max-stable processes has been very helpful in capturing the joint spatial properties of extremes in addition to their marginal variation. However, when modeling environmental processes, it mistakenly identifies the vanishing spatial dependence as genuine asymptotic dependence. Also, weakening dependence does not necessarily imply asymptotic independence, therefore well-known models like the inverted max-stable process will not do well in inferring the asymptotic properties. In our work, we present a class of spatial processes which can be described by a small number of parameters and encompass both asymptotic dependence classes. Additionally, we add a measurement error to the established model to capture any possible random effects. The model is data-driven to transit smoothly between asymptotic dependence and independence. Inference is feasible in relatively large dimensions using adaptive metropolis algorithm. The model is applied to a FFWI data set collected from the Great Plains that exhibit decaying dependence structure.

Session, time and room:

Best student paper I
Monday, 17:05 – 17:25 (A002)

Testing the multivariate regular variation model

John Einmahl[1], Fan Yang[2], Chen Zhou[3]
Tilburg University, Netherlands[1], University of Waterloo, Canada[2], Erasmus University Rotterdam, Netherlands[3]
j.h.j.einmahl@uvt.nl[1], fan.yang@uwaterloo.ca[2], zhou@ese.eur.nl[3]

In this paper, we propose a test for the multivariate regular variation model. The test is based on testing whether the extreme value indices of the radial component conditional on the angular component falling in different subsets are the same. Combining the test on the constancy across different conditional extreme value indices with testing the regular variation of the radial component, we obtain the test for testing multivariate regular variation. Simulation studies demonstrate the good performance of the proposed tests. We apply this test to examine two data sets used in previous studies that are assumed to follow the multivariate regular variation model.

Session, time and room:
IS - Mixtures of dependence types (Wadsworth)
Monday, 15:10 – 15:40 (A001)

Using Stochastic Forecasting Models to Quantify Uncertainty in Superannuation

Monica Chen[1], Bonsoo Koo[2], Alec Stephenson[3], Zili Zhu[4]
RiskLab at Data61 of Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia[1,3,4], Department of Actuaries, Monash University, Australia[2]
Wen.Chen@Data61.CSIRO.au[1], Bonsoo.Koo@Monash.Edu[2], Alec.Stephenson@Data61.CSIRO.au[3], Zili.Zhu@CSIRO.au[4]

This paper is primarily about quantifying uncertainties and in particular, about studying if and how extreme economic scenarios and events should be incorporated for their impact on superannuation investment outcomes and retirement adequacy planning. Currently, the common practice of the superannuation management industry in Australia is to assume a constant or "reasonable" investment returns (such as bootstrapping historical returns) for superannuation assets. Retirement expenditure is also assumed to follow a "reasonable" pattern, though healthcare cost can create a large degree of uncertainty. It is for sensitivity analysis when different hypothetical assumptions about possible future economic scenarios are used to study potentially alternative outcomes for both the accumulation and decumulation phases of superannuation. We will first briefly describe a stochastic cascading economic scenario generator (ESG) used for forecasting long-term economic outcomes in the SUPA (Simulation of Uncertainty for Pension Analysis) model developed at CSIRO. The stochastic forecasting model can generate extreme economic

outcomes, particularly for long-term future such as 20 to 30 years. The stochastically generated economic scenarios are then used to simulate superannuation outcomes during both the accumulation and decumulation phases. By using such a stochastic forecasting model, we are able to analyse the uncertainty and risks of making important financial decisions when the future impact of such decisions will only be known in 30 to 40 years. For such long term forecasting, extreme economic scenarios are normally viewed in the industry as unlikely and therefore are often not considered. In this paper, we will canvass the idea of incorporating extreme economic scenarios when quantifying uncertainty in superannuation analysis. In the paper, we will present some examples of using commonly adopted investment strategies and withdrawal rules for expenditure to demonstrate the variability in outcomes, such as the superannuation balance at retirement, the expected ruin years and uncertain retirement incomes. We will use 5 and 20 percentiles of the worst-case scenarios to quantify the uncertainties in the simulated superannuation outcomes, and discuss the benefit of using such a stochastic ESG to generate long term forecast of future economic environment, and the impact of extreme economic outcomes on the superannuation decumulation phase (the retirement phase).

Session, time and room:
Finance and economics
Thursday, 11:40 – 12:00 (A002)

Why are blackouts in power grids heavy-tailed?

Bert Zwart[1]
CWI and TU/e, Netherlands[1]
bert.zwart@cwi.nl[1]

Blackouts in power grids are among the most disruptive events in our society. They occur through a variety of ways. The intricate interaction of various physical (weather, electricity) and man-made (economics, human errors) features make the propagation of cascading failures hard to predict.

US data suggests that blackout sizes follow a power (pun intended) law. The origin of this law is often attributed to self-organized criticality by physicists and engineers. We will provide arguments for an alternative hypothesis, armed with data, a theorem, and a case study.

Session, time and room:
IS - Heavy-tails and networks (Davis)
Tuesday, 15:10 – 15:40 (003)

Extremes of claim sizes for marked Poisson cluster processes

Bojan Basrak[1], Petra Žugec[2]

Department of Mathematics, University of Zagreb, Croatia[1], Faculty of Organization and Informatics, University of Zagreb, Croatia[2]

bbasrak@math.hr[1], petra.zugec@foi.hr[2]

We use the theory of point processes to study asymptotic distribution of the maximal claim size for marked Poisson cluster models, in which marks determine the size and other characteristics of the individual claims and potentially 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. We apply these results on three Poisson cluster models with the marked Hawkes processes as our key example.

Session, time and room:
Models with clusters
Thursday, 16:40 – 17:00 (A002)

Posters - Tuesday, 17:15 - 19:30

Functional Covariate-Adjusted Extremal Dependence

Anwar Alabdulathem[1], Miguel de Carvalho[2], Manuele Leonelli[3]
University of Edinburgh, United Kingdom[1,2], University of Glasgow, United Kingdom[3]
anwar.alabdulathem1988@gmail.com[1], Miguel.deCarvalho@ed.ac.uk[2],
manuele.leonelli@glasgow.ac.uk[3]

In this paper we propose a method that tracks how the dependence between the extreme values of a random vector may change conditionally on a random function. Our model can be regarded as a functional covariate regression model, tailored for situations where there is the need of assessing how extremal dependence changes according to a random function. The main target of interest is what we define as the angular manifold, which is a family of angular densities indexed by a functional covariate. The methods are motivated by the need of evaluating how the dependence between extreme losses in two stock markets (e.g. NYSE and NASDAQ) changes according to the shape of a certain random curve (e.g. Daily Treasury Yield Curve). To estimate the family of angular densities on the angular manifold, we follow a similar line of attack as a popular approach for extending the Nadaraya–Watson estimator to the functional context (Ferraty and Vieu [2]). Our estimator can be regarded as a Ferraty–Vieu version of that of Castro et al. [1], and the simulation study suggests that the proposed methods perform well in wealth of simulation scenarios.

References

[1] D. Castro, M. de Carvalho and J. L. Wadsworth, Time-Varying Extreme Value Dependence with Application to Leading European Stock Markets, *Annals of Applied Statistics*, 12, 283–309, 2018

[2] F. Ferraty and P. Vieu, The Functional Nonparametric Model and Application to Spectrometric Data, *Computational Statistics*, 17, 545–564, 2002

Predicting of Value-at-risk and expected shortfall of a GPD using distorted band classes

Jose P. Arias-Nicolás[1], Maribel Parra[2], Eva L. Sanjuán[3]
Universidad de Extremadura, Spain[1,2,3]
jparias@unex.es[1], mipa@unex.es[2], etlopez@unex.es[3]

The objective of extreme value analysis is to model and measure tail events that occur with small probability, using only extreme values above some high threshold rather than using all of the data. It is well known that, when we consider the values of the sample space above a certain value, called threshold, the limit distribution function is a Generalized Pareto Distribution (GPD). Therefore, such distribution has been widely employed to model exceedances in several fields such as hydrology, finance, insurance and environmental science

Two measures that we find most useful and reliable for describing the tail of the distribution are value-at-risk and expected shortfall. We compute the range of the parameters of the GPD, and these two risk measures of the predictive density under a Bayesian framework

Robust Bayesian analysis, also called Bayesian sensitivity analysis, aims to quantify and interpret the uncertainty induced by the partial knowledge of one of the three elements in Bayesian analysis (prior, likelihood and loss). Studies mainly focus on computing the range of some quantities of interest when the prior distribution varies in a class. We use the band distorted class defining by Arias-Nicolás et al in 2016 to compute the range of the parameters of a Generalized Pareto Distribution (GPD). The two risk measures, value-at-risk and expected shortfall, are constructed based on the Bayesian estimation results

The distorted band class can be considered as a "neighborhood" band of a prior distribution of reference, whose upper and lower bound distributions are the distorted distributions (concave and convex respectively). Due to the fact that this class is based on the election of two distorted functions, we can provide many possible bands just considering different concave and convex distortion functions. Of course, the choice of those functions cannot be arbitrary and should represent the uncertainty about the prior belief in each problem.

The most important property is that the likelihood ratio order we are using in this class is preserved after the application the posterior belief. In the same way, we can see something similar for predictive distributions. Therefore, both the value at risk and the conditional value at risk for the predictive distributions are ordered too. This means that, the range of these measures can be compute determining only value at risk and conditional value at risk of the posterior predictive distortion distributions. If we compute this range with respect to the value and the conditional value at risk of reference prior, we'll have a relative sensitive measure. We show that computations of these sensitivity measures should be as easy as possible, possibly looking for the extremal distributions generating the class.

Due to the fact that the distorted band depends on the election of the reference prior distribution (also on the distorted functions considered), we ilustrate with some numerical examples, how this choice affects the calculation of the measures considered.

Hüsler–Reiss Markov model for studying tail dependence in probabilistic graphical trees

Stefka Asenova[1], Johan Segers[2]
Université Catholique de Louvain, Belgium[1,2]
stefka.asenova@uclouvain.be[1], johan.segers@uclouvain.be[2]

Tail dependence modeling in multivariate extremes involves high dimensional functions which often become intractable analytically and numerically as well. Conditional independence, also called graphical models because of their representability as a graph, offer the possibility of dimension reduction. A particular case are tree graphical models where joint distributions can be factorized in bivariate functions.

We present a so called Hüsler–Reiss Markov (HRM) tree, a graphical model suitable for studying tail dependence of a graphical model on a tree. Practical interest in such structures may arise when dealing with variables measured on a river network, such as water level, flow or chemical concentration. The model is a special case of a regularly varying Markov tree in [5].

The logarithm of the excesses over a threshold of a HRM tree have multivariate normal distribution as limiting distribution, conditional on one of the variables exceeding a high threshold. The limiting distribution represents a Gaussian Graphical Model too. In the very frequent in reality situation when some of the variables are not observable on the tree the model is still identifiable if the nodes with missing data have at least three neighbors.

For estimation of a HRM tree we propose three types of estimators: method of moments [2], maximum likelihood [1] and one based on the stable tail dependence function [4]. In each of them the variable k, the number of upper order statistics considered, plays a key role. In simulation experiment it is observed that the estimators have the expected properties: the bias increasing with k and the variance is decreasing with k.

We apply the estimators to real data on Danube network in South-Eastern Germany (dataset used in [3]) and observe higher levels of tail dependence for flow connected variables. We also find good agreement between model based and empirical extremal coefficients which is a sign that the HRM tree fits well that particular data.

References

[1] S. Lauritzen, *Graphical Models*, Oxford University Press, 1996

[2] Engelke,S. and Malinowski, A. and Kabluchko, Z. and Schlather, M., Estimation of Hüsler-Reiss distributions and Brown-Resnick processes, *Journal of the Royal Statistical Society. B*, 2015

[3] Asadi, P. and Davison, A. and Engelke, S., Extremes on river networks, *The Annals of Applied Statistics 9*, 2015

[4] Einmahl, J. and Kiriliouk, A. and Segers, J., A Continuous Updating Weighted Least Squares Estimator of Tail Dependence in High Dimensions, *Extremes*, 2017

[5] Segers, Johan, One- versus multi- component regular variation and extremes of Markov trees, Applied Probability Trust, 2019

Extremal clustering in non-stationary sequences

Graeme Auld[1], Ioannis Papastathopoulos[2]
University of Edinburgh, United Kingdom[1,2]
G.R.Auld@sms.ed.ac.uk[1], i.papastathopoulos@ed.ac.uk[2]

It is well known that the extreme values of strictly stationary sequences differ from those of independent and identically distributed sequences in that extremal clustering may occur. The extent to which extremal clustering may occur is measured by the extremal index, $\theta \in [0, 1]$. In this poster we consider non-stationary sequences subject to suitable long range dependence restrictions. We consider the case of non-independent but identically distributed random variables. We find that the limiting distribution of appropriately normalized sample maxima depends on a parameter $\bar{\theta} \in [0, 1]$ that measures the average extremal clustering of the sequence and generalizes the result of [1] for the extremal index. Subsequently we derive the asymptotic distribution for the times between consecutive extreme observations which allows for the estimation of this new parameter.

References

[1] G. O'Brien, Extreme values for stationary and Markov sequences, *Annals of Probability*, 1987

Randomization test of independence based on tail behaviour of two processes

Darko Brborović[1]
OTP Invest, Croatia[1]
darko.brborovic1@gmail.com[1]

Randomization test for testing independence of two discrete random processes is presented. Test is based on the tail behaviour of the processes that may be IID or m-dependent. In essence, the test is based on statistic that counts joint exceedances of referent processes over some threshold. Given randomization test is non-parametric and fairly flexible with minimal assumptions.

Simulation study is presented in short and an application in finance. Idea of the application is to identify some time series that shows no central dependency (measured by correlation, for example), but contains some information in their tails that indicate dependency among time series we would like to analyse. When we look at return data from financial markets it often happens that correlation among two assets/asset classes is low in most of the time but rises in times of crisis. This test helps to identify such circumstances.

Short tails and the extended generalized Pareto distribution

Patricia de Zea Bermudez[1], Paula Pereira[2], Soraia Pereira[3], Miguel de Carvalho[4]
CEAUL and Faculdade de Ciências, Universidade de Lisboa, Portugal[1,3], ESTSetúbal, Instituto Politécnico de Setúbal and CEAUL, Portugal[2], School of Mathematics, University of Edinburgh, United Kingdom[4]
pcbermudez@fc.ul.pt[1], paula.pereira@estsetubal.ips.pt[2], soraia.gpereira@gmail.com[3], Miguel.deCarvalho@ed.ac.uk[4]

In a variety of situations there is a need for combining real data with simulated data (say, obtained from a climate model). One of the usual problems of real data, especially in an environment context, is the presence of missing observations. Simulated data overcomes this problem. However, additional problems may arise such as a misalignment between data sets. In this framework, marginal features of real and simulated data may differ substantially in the bulk and/or in the tail of the underlying distributions. One possibility is to bring both data sets to the same scale, using some calibration method. The purpose is to calibrate the (complete) simulated data in order to adequately match the features of the (incomplete) observed data.

Generally, the major differences between the underlying distributions are observed in the tails and consequently the modeling should rely uniquely on the extreme observations of the sample. One of the most common methods used in Extreme Value Theory is the peaks of threshold (POT) which consists on fitting a generalized Pareto distribution (GPD) to the tail. The success of the POT heavily depends on the choice of the threshold. In order to overcome this issue, [1] proposed using the extended GPD (EGPD) which has the benefit of enabling the use of the entire data set and avoiding the difficult problem of choosing an adequate threshold. [1] use the EGPD for modelling rainfall data, which typically has an exponential or a moderate right tail. In this work, we adapt the method by [1] so that it can also cover short-tailed distributions (with upper finite endpoint), develop a version of the method tailored for calibration, which can be used for both short and heavy tailed distributions, and present an application to a windspeed data set which exhibits a short-tail. The data consists on daily maximum wind speeds, recorded in 40 stations in Portugal and simulated wind speeds obtained using a WRF-Weather Research and Forecast-model.

References

[1] P. Naveau, R. Huser, P. Ribereau, and A. Hannart, Modeling Jointly Low, Moderate, and Heavy Rainfall Intensities without a Threshold Selection, *Water Resources Research*, 52, 2753-2769, 2016.

Modelling dependency effect to extreme value distributions with application to extreme wind speed at Port Elizabeth, South Africa: a frequentist and Bayesian approaches

Legesse Kassa Debusho[1], Tadele Akeba Diriba[2]
University of South Africa, South Africa[1,2]
debuslk@unisa.ac.za[1], diribtd@unisa.ac.za[2]

The dependency effect to extreme value distributions (EVDs) using the frequentist and Bayesian approaches have been used to analyse the extremes of annual and daily maximum wind speed at Port Elizabeth, South Africa. In the frequentist approach, the parameters of EVDs were estimated using maximum likelihood, whereas in the Bayesian approach the Markov Chain Monte Carlo technique with the Metropolis-Hastings algorithm was used. The results show that the EVDs fitted considering the dependency and seasonality effects with in the data series provide apparent benefits in terms of improved precision in estimation of the parameters as well as return levels of the distributions. The paper also discusses a method to construct informative priors empirically using historical data of the underlying process from other weather stations. The results from the Bayesian analysis show that posterior inference might be affected by the choice of priors used to formulate the informative priors. The Bayesian approach provides satisfactory estimation strategy in terms of precision compared to the frequentist approach, accounting for uncertainty in parameters and return levels estimation.

A simulation comparison study of the methods of the threshold selection for the heavy-tailed distributions

Redhouane Frihi[1]
Department of Mathematics, Faculty of Sciences, University of Blida, Algeria[1]
rfrihi@yahoo.fr[1]

The modeling of heavy tailed distribution is strongly linked to a real number called the queues index or extreme values index (EVI) and whose value governs the thickness of the tails. The estimation of this index, which is paramount in the modeling process, is largely dependent on the number of extreme statistics observed. This number determines the values, which are really extreme. In other words, it allows to obtain the threshold u of observations start to become extreme. The question that has been often addressed in practical applications of extreme value theory is the choice of the threshold parameter u. Overly high threshold results in little information to be used and will lead to high variance and unreliable result. On the other hand excessively low threshold leads to large bias in parameters estimation of generalized Pareto distribution (GPD). Thus, choosing a proper threshold implies a balance between the bias and the variance.The latter is generally determined using various approaches, such as graphical, parametric, semi-parametric and nonparametric methods.

In order to investigate the influence of threshold on EVI estimation in POT (peaks-over-threshold) method, some parent distributions are generated by the monte Carlo method and the extreme values at different thresholds are esimated from differentes methods and compared with the theoretical value.

Spatial extremes modelling with an application to extreme precipitation

Dora Gomes[1], Délia Gouveia-Reis[2], Manuela Neves[3]

Faculdade de Ciências e Tecnologia and Centro de Matemática e Aplicações, Universidade Nova de Lisboa, Portugal[1], Faculdade de Ciências Exactas e da Engenharia, Universidade da Madeira and Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal[2], Instituto Superior de Agronomia and Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal[3]

dsrp@fct.unl.pt[1], delia.reis@staff.uma.pt[2], manela@isa.ulisboa.pt[3]

Extreme value analysis whenever the observations are located in space and/or time has been an area of increasing interest in the last decade. For that analysis it is necessary to measure the extremal spatial dependence and to take it into account for modelling and predicting the extreme behaviour of spatially located environmental phenomena.
This work aims to present an application of procedures that constitute recent research in modelling of spatial extremes. It intends to continue the preliminary study presented in Neves and Prata Gomes [1] and applied to the North of Portugal. In this work those procedures will be applied to a set of geo-referenced data, referring to monthly maximum values, of daily precipitation recorded between 1970 and 2016, at a few relevant stations on the island of Madeira, see a first approach in Prata Gomes et al. [2], identified by previous studies as located in zones of extremes of precipitation and where special care is needed to prevent risks to human lives.

References

[1] Neves, M. Manuela and Prata Gomes, D. Geostatistics for spatial extremes. A case study of maximum annual rainfall in Portugal. *Procedia Environmental Sciences, Spatial Statistics 2011, Elsevier*, Vol. 7, 246-251, 2011.

[2] Prata Gomes, D., Gouveis-Reis, D. and Neves, M. Manuela. "Modelação espacial aplicada a dados de precipitação extrema na ilha da Madeira". *Proceedings of the meeting II Encontro Galaico-Português de Biometria*, 257-260, 2016.

Performance of tests based on the area under the ROC curve for multireader diagnostic data

Yi-Ting Hwang[1], Ming-Hsueh Lee[2]

Department of Statistics National Taipei University, Taiwan[1,2]

hwangyt@gm.ntpu.edu.tw[1], oh7993@gmail.com[2]

The ROC (Receiver operating characteristic) curve is a measure for evaluating diagnostic reagents. The AUC (area under the ROC curve) is one of the most popular indices to evaluate the accuracy of the diagnostic reagents. X-ray films or MRI scans are often used to evaluate the status of cancer patients. Nevertheless, the outcome of these instruments are determined by professional physicians. To ensure the accuracy of the instrument and avoid the possibility of reading bias from physicians, the outcome of these instruments is evaluated by several professional physicians. Owing to such a study design, the outcome of the data is correlated. The evaluations of accuracy between two instruments have to take into account of correlations. This paper evaluates several existing tests such as the Wald

type test and the test obtained from the random effect model. Furthermore, the Jackknife method suggested by [1] is also used to construct a test. Monte Carlo simulations are then used to evaluate the performance and the robustness of these tests.

References

[1] Hinkley, D. V. (1976). Robust jackknife correlation. Biostatistics Technical Report, California.

Drift in transaction-level models for asset price: evidence from a decade of high-frequency returns of Dow30 stocks

Agnieszka Jach[1]

Hanken School of Economics, Finland[1]

agnieszka.jach@hanken.fi[1]

We apply modelling and estimation framework of [1] to high-frequency Dow 30-stock data between 2008-2017. The underlying univariate model for the log price $y = \{y(t), t \geq 0\}$ (pure-jump and transaction-level) is in continuous time, is driven by point processes and embeds two components, efficient and microstructure. The model is capable of mimicking all stylized facts about transaction-level data; in particular, it allows for long memory and heavy tails in inter-trade durations. The flexibility of the model is partly reflected in a wide range of limiting distributions for the (centered and scaled) log price, and this in turn affects statistical inference for the long-term trend/growth rate μ^* based on the average return, obtained from n calendar-time returns. To make the inference for the growth rate under various scenarios possible, [1] define a self-normalized statistic $T_{2,n}$, whose distribution is non-degenerate irrespective of which scenario is actually realized. Such self-normalization gets rid of the nuisance parameter from the sample statistic, but not from the limiting distribution, so the sampling distribution of $T_{2,n}$ is approximated nonparametrically via subsampling. When this estimation paradigm is employed on 1-min high-frequency excess returns of Dow 30 stocks, quarter-by-quarter ($\approx$ 24,375 observations per stock per quarter), the null hypothesis of $\mu^* = 0$ (lack of equity premium) is less frequently rejected for stock-quarter cases where the traditional t_n-statistic (asymptotically standard normal) rejects the null. This is in line with the findings of [1], but some unexpected results come to light as well.

References

[1] W. Cao, C. Hurvich, P. Soulier, Drift in transaction-level asset price models, *Journal of Time Series Analysis*, 2017

Future changes in the mean and extreme sea surface winds over the East Asian marginal seas projected by regional climate models

Wonkeun Choi[1], Chan Joo Jang[2], Cheol-Ho Kim[3]
Ocean Circulation and Climate Research Center, Korea Institute of Ocean Science and Technology, Korea, Republic of[1,2,3]
dnjsrms@kiost.ac.kr[1], cjjang@kiost.ac.kr[2], chkim@kiost.ac.kr[3]

Sea surface wind, especially extreme winds, plays a crucial role in coastal safety including flooding and coastal erosion mainly by changing sea level and ocean wind waves, and the regional characteristics of sea surface wind can be influenced by global warming. In this study, we investigate future changes in sea surface wind (both in mean and extreme winds) due to global warming in the East Asian marginal seas (EAMS) projected by four different regional climate models contributing to CORDEX-EA (COordinated Regional climate Downscaling Experiment-East Asia) downscaling project, by using daily mean winds for the present (historical runs for 1980-2004) and the future climate (RCP4.5 scenario run for 2025-2049). In general, the climatological-mean wind speed is projected to decrease up to 30 % in winter, but to increase up to 40 % in summer, indicating seasonally different large-scale atmospheric responses to global warming in EAMS. The prevailing wind direction, a critical factor in coastal erosion in EAMS, is projected to be significantly changed in both winter and summer, contributing to the changes in the mean wind speed. Extreme winds, defined as an annual maximum of the wind speed, show negligible changes with considerably large intermodel differences. Our findings suggest that projected future changes in sea surface wind over EAMS can be considerably different depending on seasons and models.

Modeling dependence among different rainfall intensity accumulation durations: A copula approach.

Oscar E. Jurado[1], Henning W. Rust[2], Jana Ulrich[3]
Freie Universität Berlin, Germany[1,2,3]
jurado@zedat.fu-berlin.de[1], henning.rust@zedat.fu-berlin.de[2], janaulrich@met.fu-berlin.de[3]

Intensity-Duration-Frequency (IDF) curves describe the relation between the exceedance probabilities for intensities as a function of the duration of the event. Recently, some studies have explored the modeling of the relationship between intensity and duration using methods from multivariate extreme value statistics, such as Bayesian estimation [2] or max-stable processes ([3]; [1]). A common feature of those studies is the modeling of the dependence between rainfall intensity for different accumulation durations. This dependence is usually integrated into the composite likelihood term in order to fit the desired model to the data. Some of the methods, used to model the dependence, include parametric approaches and the use of different dependence coefficients.

The main focus of this work is to further investigate the properties of the dependence between different sub-daily durations for individual weather stations in a specified region. In this work, the authors propose the use of a duration non-euclidean space, following the steps of [1], on which different methods developed for spatial statistics can be used to model this dependence. A Copula-based approach was chosen to model the dependence, as their

use in the context of spatial extremes has been justified by several previous studies. We use data from sub-daily weather stations for a specific region in Germany. Precipitation data for each station is processed and divided using both the Block Maxima approach, and also by different accumulation durations, going hourly from 1 to 24 hours. Each intensity (accumulated using the different durations) is fitted to a univariate Generalized Extreme Value distribution and transformed to unit Frechet. A bivariate extreme value copula (EVC) is fitted to the resulting different pairs of intensities. The resulting dependence coefficient and extremal index are used to investigate how dependence varies among the different duration pairs, with special attention to the progression of dependence when the distance along with the duration-axis increases. This analysis is performed for each individual station of the chosen region, and a comparison is performed between stations to study how the dependence also changes when looking at different points.

References

[1] H. Tyralis, A. Langousis, Estimation of intensity–duration–frequency curves using max-stable processes, *Stochastic Environmental Research and Risk Assessment*, 2019.

[2] A. Muller, J.N. Bacro, M. Lang, Bayesian comparison of different rainfall depth-duration-frequency relationships, *Stochastic Environmental Research and Risk Assessment*, 2008.

[3] A. Stephenson, E. Lehmann, A. Phatak, A max-stable process model for rainfall extremes at different accumulation durations. *Weather and Climate Extremes*, 2016.

Stochastic model of meteorological time series to study the wind chill temperature index

Nina Kargapolova[1]

Institute of Computational Mathematics and Mathematical Geophysics SB RAS & Novosibirsk State University, Russian Federation[1]

nkargapolova@gmail.com[1]

In this paper a numerical stochastic model of the joint time series of air temperature, relative humidity and wind speed modulus is proposed. The model takes into account the diurnal and seasonal variation of the real meteorological processes. Simulated trajectories of the joint time series are used to study various statistical properties of the time series of the Wind Chill Temperature Index (WCTI) that characterizes the thermal effects of cold humid air and wind on human beings. The results of the study of the properties of rare periods of time with dangerous/anomalous values of the WCTI are presented.

This work was partly financially supported by the Russian Foundation for Basic Research (grant No 18-01-00149-a), RFBR and the Government of the Novosibirsk region according to the research project No 19-41-543001-r_mol_a.

Drift change detection for multidimensional force of mortality models

Michał Krawiec[1]
University of Wrocław, Poland[1]
michal.krzysztof.krawiec@gmail.com[1]

In my talk I will present optimal drift change detection problem applied to the analysis of changes in mortality. From the theoretical point of view, mathematical methods used are based on the Bayesian approach considered e.g. by A. N. Shiryaev. The method is to study when a certain statistics begins to take values that exceed a certain threshold. We will focus on the optimal detection for different kinds of multidimensional Lévy processes, starting with the Brownian motion and then extending the model e.g. by adding jumps or random post-change parameter. Regarding applications, we will consider force of mortality of the Polish population. We will construct a numerical procedure that detects when changes of drift in observed process become significant in the sense of our model. Such methods of optimal detection may be relevant in insurance models – concerning life tables and longevity – as well as applied to other fields in which detection of trend changes is important. The talk is based on joint works with Zbigniew Palmowski and Łukasz Płociniczak.

Statistical Downscaling of Air Temperature Extremes for Georgia Using RCMES Package

Teimuraz Davitashvili[1], Nato Kutaladze[2], Ramaz Kvatadze[3], Lia Megrelidze[4], George Mikuchadze[5], Inga Samkharadze[6]

Ilia Vekua Institute of Applied Mathematics of Ivane Javakhishvili Tbilisi State University, Georgia[1], The National Environmental Agency, Georgia[2,4], Georgian Research and Educational Networking Association GRENA, Georgia[3,5], Ivane Javakhishvili Tbilisi State University, Georgia[6]

tedavitashvili@gmail.com[1], cwlam08@gmail.com[2], ramaz@grena.ge[3], l_megrelidze@hotmail.com[4], gmikuchadze@gmail.com[5], inga.samkharadze562@ens.tsu.edu.ge[6]

Characterizing present climate conditions and providing future climate projections at a regional scale is an extremely difficult task as it involves additional uncertainties while reducing a spatial scale of Global Climate Models (GCMs) simulated climate parameters. Decreasing in spatial accuracy of GCMs simulated climate variables occurs from continental to local scale using statistical downscaling (SD) or dynamical downscaling (DD) techniques. Statistical downscaling methods by itself is divided into three groups: multiple linear regression, nonlinear regression and stochastic weather generators, which mostly are used in different sectoral impact studies. In this study monthly maximums and minimums of 2-meter air temperature from three GCMs of CMIP5 database has been statistically downscaled using RCMES package, with four different methods (delta addition, delta correction, quantile mapping; asynchronous linear regression) for 27 selected meteorological stations on the territory of Georgia. The extreme values downscaling methods have been trained for the 1961-1985 and validated for the 1986-2010 period. Some statistical parameters have been calculated by applying R statistics environment to compare observed and simulated time series and to evaluate temporal and spatial goodness of each method. Downscaling model driven by the validation study was used for future extremes (Tmin

and Tmax) time series construction for the 2021-2070 period under RCP4.5 and RCP8.5 scenarios. Temperatures time series have been constructed from a multi-model ensemble, with mean and spread. The Future change tendencies have been assessed in comparison to 1986–2010 period. Validation of statistical downscaling methods shows that all of the methods have some advantages and disadvantages on the temporal and spatial scale. The metrics used for model performance evaluation varies from station to station, year to year, and season to season.

Uniform limit theorems for estimators of point process boundaries

Ludovic Menneteau[1]
Université de Montpellier, France[1]
ludovic.menneteau@umontpellier.fr[1]

The problem of estimating a set S given a finite sample of random points drawn from its interior arises in several domains such as clustering, outliers detection, image analysis and econometrics . In many cases, the unknown support can be written as the hypograph of an unknown function f where E is an arbitrary set (typically a subset of R^d). Then, the problem boils down to estimating f. In this work, we investigate uniform limit properties for an estimator of empirical and Poisson point processes boundaries obtained by smoothing some bias corrected extreme values of the process. The proofs are based on Berry-Esseen type estimates for large deviations of the extremes.

References

[1] S. Girard and L. Menneteau, Central limit theorems for smoothed extreme value estimates of Poisson point processes boundaries, *Journal of Statistical Planning and Inference*, 2005.

[2] S. Girard and L. Menneteau, Smoothed extreme value estimators of non-uniform point processes boundaries with application to star-shaped supports estimation, *Communication in Statistics - Theory and Methods*, 2008.

[3] L. Menneteau, Multidimensional limit theorems for smoothed extreme value estimates of point processes boundaries. *ESAIM Probability and Statistics*, 2008.

Projected changes of extreme rainfall in the province of Quebec

Jonathan Jalbert[1], Éloïse Nolet-Gravel[2], Luc Perreault[3]
Department of Mathematics and Industrial Engineering, Polytechnique Montreal, Canada[1,2], Hydro-Québec's research institute, Canada[3]
jonathan.jalbert@polymtl.ca[1], eloise.nolet-gravel@polymtl.ca[2], perreault.luc@ireq.ca[3]

According to the latest report of the Intergovernmental Panel on Climate Change (IPCC), the province of Quebec is part of the regions where extreme rainfall events are most likely to increase substantially if the global warming goes from 1.5 to 2 °C. With the intensification of climate change in the last years, the natural variability of intense precipitations has increases and the probabilistic behavior of climatic extremes has been modified, meaning that Intensity-Duration-Frequency (IDF) curves are likely to evolve with time. IDF curves are the principal tool for designing infrastructures exposed to climate hazards and for defining flood plains. Therefore, the risk of failure of an infrastructure designed with the actual curves should evolve according to climate change. That is not currently taken into account in the IDF curves provided by Environment and Climate Change Canada (ECCC), which are the most used IDF curves in the country. It becomes primordial to include their effects in the estimation of IDF curves to ensure infrastructures a certain security level throughout their useful life. The proposed project consists of including the effects of climate change in the IDF rainfall curves. To update these curves, a spatial and non-stationary model will be developed for the extreme precipitations simulated by a climate model. The developed methodology will take into account the spatial and non-stationary aspects of intense precipitations to develop IDF curves in future climate.

Practical MCMC for Bayesian analysis of air pollution data

Man-Suk Oh[1], Eun Sug Park[2], Chi Kyoung Park[3]
Ewha Womans University, Korea, Republic of[1,3], Texas Transportation Institute, United States[2]
mansuk55@gmail.com[1], e-park@tamu.edu[2], ck930310@gmail.com[3]

Analysis of air pollution data has received great attention over the past decade. Multivariate receptor modelling aims to find significant pollution sources and their contributions from mixtures of chemical species in the ambient air. Bayesian analysis of receptor models is advantageous in measuring uncertainty in model selection and aparmeter estimation. However, MCMC algorithm required for Bayesian analysis is complicated due to parameter constrains needed to handle non-identifiability problem. Here we present a practical MCMC tools for Bayesian analysis of multivariate receptor models using R and JAGS software which may enable practitioners and researchers who are not expert in MCMC or program coding can easily implement Bayesian analysis of multivariate receptor models.

Some asymptotic results on the maxima and minima of non-stationary random fields with extremal index

Zacarias Panga Pedro[1], Luísa Pereira[2]
University of Beira Interior, Portugal[1,2]
panga.pedro@ubi.pt[1], lpereira@ubi.pt[2]

Let $\{Z(x)\}$ be a non-stationary random field, defined over discrete subsets of R^2, satisfying a coordinatewise-mixing condition and a local dependence condition which restricts the local path behavior of high and low values. We show that the maximum and minimum of the random field are asymptotically independent even when there is clustering of low and high values, thus extending the result of [1] that they are asymptotically independent when the random field has extremal index equals to unity. In the end the almost sure limit theorem on the joint version of maxima and minima is proved.

References

[1] Z. Panga and L. Pereira, *Statistics and Probability Letters, 137, 124-134*, 2018

An adaptive EVI-estimation through computational procedures

M. Ivette Gomes[1], M. Manuela Neves[2], Helena Penalva[3]
Faculdade de Ciências, Universidade de Lisboa, Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal[1], Instituto Superior de Agronomia, Universidade de Lisboa, Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal[2], Escola Superior de Ciências Empresariais, Instituto Politécnico de Setúbal, Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal[3]
ivette.gomes@fc.ul.pt[1], manela@isa.ulisboa.pt[2], helena.penalva@esce.ips.pt[3]

In statistical extreme value analysis, the estimation of the extreme value index (EVI), a key parameter in *extreme value theory*, can be performed in a semi-parametric approach, where the EVI- estimators are based on the largest k order statistics in the sample. How to choose this number k is a critical issue and still remains a challenge in the statistical analysis of extremes values (see, among others, Caeiro and Gomes, 2015). Regarding the EVI estimation, we consider here a recent class of semi-parametric estimators of a positive EVI, parametrized in a tuning parameter $p > 0$, the class of *Lehmer's mean of order p* (L_p) EVI-estimators (see Penalva, 2017, and Penalva *et al.*, 2018). The aim of this work is to review and compare some procedures based on bootstrap and heuristics methodologies for an adaptive EVI estimation, choosing adequate values of k and p. These procedures will be applied to a real data set.

References

[1] Caeiro, F. and Gomes, M.I. (2015). Threshold selection in extreme value analysis. In Dipak Dey and Jun Yan (eds.), *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman-Hall/CRC, pp. 69–87.

[2] Penalva, H., Caeiro, F., Gomes, M.I. and Neves, M. (2016). *An Efficient Naive Generalization of the Hill Estimator—Discrepancy between Asymptotic and Finite Sample Behaviour*. Notas e Comunicações CEAUL 02/2016.

[3] Penalva, H., Gomes, M.I., Caeiro, F., and Neves, M.M. (2018). A couple of non reduced bias generalized means in extreme value theory: an asymptotic comparison. *Revstat*. Available at: `https://www.ine.pt/revstat/pdf/Acoupleofnonreducedbiasgeneralizedmeansinextremevaluetheory.pdf`

Aggregation of Spatial Extremes

Simon Brown[1], Jordan Richards[2], Jonathan Tawn[3], Jennifer Wadsworth[4]
MET Office, United Kingdom[1], STOR-i, Lancaster University, United Kingdom[2,3,4]
simon.brown@metoffice.gov.uk[1], j.flett@lancaster.ac.uk[2], j.tawn@lancaster.ac.uk[3], j.wadsworth@lancaster.ac.uk[4]

Fluvial flooding is caused by excessive rainfall sustained over extended periods of time and over spatial catchment areas. Although methodology for modelling excessive, or extreme, rainfall events is extensive and well researched, the same cannot be said about how the extremal properties of spatial and temporal aggregations of rainfall are related. We hope to rectify this by developing a methodology for modelling extremes at different spatio-temporal scales and which incorporates a wide range of dependence structures.
Research on modelling aggregated spatial extremes is ongoing, but here we present a simpler model for extreme values of sums of multivariate (dependent) variables. Marginally these variables are assumed to have GPD tails and we focus on exploring how properties of the copula influence the tail properties of the aggregate. The implications of our theoretical results for statistical purposes will be discussed.

A model of a queue where everyone has to wait

Mariia Senina[1]
Kuban State University, Russian Federation[1]
senina.maria@mail.ru[1]

It is 7:20am on Monday morning and there are already a group of people huddled together in the street outside 10 State Clinic in Krasnodar. Some of them have been here since 7:00am as they know that by 7:30am, when the clinic opens its doors, the queue will reach down the street. Each of them wants to be the first. But each of them will wait.

We made an attempt to study this model - where everyone is waiting - and to find out what time should patient come to the Clinic to spend in the queue as less time as possible.

We have been observed the queue for 3 days from 7:00 (0 time) till 8:30 (ψ time). State Clinic opens every day at 7.30 (τ time) in the morning and stops working at 8:30 in the morning. So, it works for 60 minutes. Average service time is about $\mu = 1$ min per customer. So, if customers arrive from a source of size N and are served by one server

at a service facility according to FIFO discipline, where the requests arrive according to a Poisson process with rate λ(t), then $\alpha_1 = M\,[\mathrm{X}(0;\tau)] = \int\limits_0^{\tau} \lambda(t)dt$ - so many consumers will be in the queue before opening time. Everyone will wait at least for 30 minutes (because the first customer stood in the queue at 7:00). With our figures $\lambda(t) = 0{,}0003t^2 + 0{,}0475\mathrm{t}$ - 0,0755 ($R^2 = 0{,}7569$).

After opening time $\alpha_2 = M\,[\mathrm{X}(\tau;\,\psi)] = \int\limits_{\tau}^{\psi} \gamma(t)dt$ - so many consumers will be in the queue additionally after opening time. With our figures $\gamma(t) = 0{,}0009t^2$ - 0,0869t + 2,0976 ($R^2 = 0{,}68$).

So, at the opening time the number of costomers in the queue is:

$\alpha_1 = \int\limits_0^{30} (0,0003t^2 + 0,0475t - 0,0755)\,dt \approx 22.$

By the time when the Clinic has to be closed (from 7:30 till 8:30) the number of customers $\alpha_2 = \int\limits_0^{60} (0,0009t^2 - 0,0869t + 2,0976)\,dt \approx 34.$

So, in average service point is free for about 60 - (22 + 34) = 4 minutes and the probability that at the moment $\left(\psi - \tau - \mu\big(\int\limits_0^{\tau} \lambda(t)dt + \int\limits_{\tau}^{\mu\int_0^{\tau}\lambda(t)dt} \gamma(t)dt\big);\psi\right)$ there is no additional customer is:

$\mathrm{P}\big(0(56\text{ - }60)\big) = e^{-\int_{56}^{60}(0{,}0009t^2-0{,}0869t+2{,}0976)\,dt} \approx 0,74.$

So, it is advantageous for the customer to come at the service point almost at the end of working time. In this case he will spend in the queue as less time as possible.

References

[1] J. Patrick, L. Puterman Reducing Wait Times through Operational Research: Optimizing the Use of Surge Capacity, *Health Policy*, 2008

A new estimation method for CoVaR based on three regime bivariate normal distribution

Ji-Eun Choi[1], Dong Wan Shin[2]

Ewha Womans University, Korea, Republic of[1,2]

cje6479@naver.com[1], shindw1234@naver.com[2]

We propose a new estimation method of co-value-at-risk, CoVaR, a financial system risk measure conditional on an institution in a financial distress. The new method is based on a three regime bivariate normal (3RN) distribution which is composed of three bivariate normal distributions with asymmetric variance matrices for the right-tail, left-tail and mid-part corresponding to the return of an institution. The distribution captures explicitly the asymmetric correlation of system return and institution return: usually stronger for bad times than for good times. The 3RN distribution allows simple evaluations of the CoVaR taking full advantage of asymmetric correlation. An implementation for the quasi

maximum likelihood estimator (QMLE) is provided. Good performances of the proposed method are shown in a real data analysis for stock price data sets consisting of one financial system and four financial institutions: the US S&P 500 index, Bank of America corporation, JP Morgan Chase & Co., Goldman Sachs Group, Inc. and Citigroup Inc.

Characterizations based on regression assumptions of order statistics from overlapping samples

Nan-Cheng Su[1]
National Taipei University, Taiwan[1]
sunc@gm.ntpu.edu.tw[1]

In this work we are interested in the joint distribution of two order statistics from overlapping samples. First we study the question to what extent conditional expectation of one of such order statistic given another determines the parent distribution. In particular, we provide new characterizations by linearity of regression of an order statistic from the extended sample given the one from the original sample. To describe the correct parent distribution it is convenient to use quantile density functions. In several other cases of regressions of order statistics we study new results regarding uniqueness of the distribution in the sample.

References

[1] F. Lopez-Blazquez, N.C. Su and J. Wesolowski. Order statistics from overlapping samples: bivariate densities and regression properties. *To appear in Statistics*, 2019

Confidence Intervals Based on L-Moments for Population Quantiles of the GP and GEV Distributions with Application to Precipitation Data

Tereza Šimková[1]
Technical University of Liberec, Czech Republic[1]
tereza.simkova@tul.cz[1]

Moments, such as mean, variance, skewness and kurtosis, are traditionally used to describe features of a univariate probability distribution. Hosking [3] unified and popularised an alternative approach to conventional moments which is now widely used both in data analysis and statistical inference. This approach uses quantities called L-moments which are based on linear combinations of order statistics. The main L-moments' advantage in comparison to the conventional moments is their existence of all orders under only a finite mean assumption.

The traditional techniques for parameter estimation are the method of moments and maximum likelihood method. L-moments are nowadays used for parameter estimation as an alternative to these estimation methods. The L-moments based estimates are obtained in a similar way as in the moments method, which means the population L-moments are

equated to their corresponding sample counterparts. Hosking [3] prefers L-moments to the conventional moments because they are linear combinations and therefore they are more robust to the presence of outliers and less subject to bias in parameter estimation in small samples. Several studies also shown that the L-moments method also outperforms the maximum likelihood method in some cases [2], [6], [7].

The generalized Pareto (GP) and generalized extreme-value (GEV) distributions are distributions which have been used in modeling natural extreme events in hydrology, climatology, meteorology and other areas [1], [4], [5].

The aim is to derive asymptotic confidence intervals for high population quantiles of the GP and GEV distributions by several methods, namely the L-moments, moments, and maximum likelihood methods. The asymptotic confidence intervals are compared to each other and they are also compared to confidence intervals obtained by bootstrap techniques.

References

[1] M.M. de Oliveira, N.F. Ebecken, J.L. de Oliveira, E. Gilleland, Generalized extreme wind speed distributions in South America over the Atlantic Ocean region, *Theoretical and Applied Climatology*, 2011

[2] P. Delicado, M.N. Goria, A small sample comparison of maximum likelihood, moments and L-moments methods for asymmetric exponential power distribution, *Computational Statistics and Data Analysis*, 2008

[3] J.R.M Hosking, L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of Royal Statistical Society (Series B)*, 1990

[4] R.W. Katz, M.B. Parlange, P. Naveau, Statistics of extremes in hydrology, *Advances in Water Resources*, 2002

[5] J. Kyselý, J. Picek, Regional growth curves and improved design value estimates of extreme precipitation events in Czech Republic, *Climate Research*, 2007

[6] E.S. Martins, J.R. Stedinger, Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resources Research*, 2000

[7] T. Šimková, J. Picek, A comparison of L-, LQ-, TL-moment and maximum likelihood high quantile estimates of the GPD and GEV distribution, *Communications in Statistics - Simulation and Computation*, 2017

Extremes of a type of locally stationary Gaussian random fields with applications to Shepp statistics

Zhongquan Tan[1]
Jiaxing University, China[1]
tzq728@163.com[1]

Let $\{Z(\tau,s),(\tau,s)\in[a,b]\times[0,T]\}$ with some positive constants a,b,T be a centered Gaussian random field with variance function $\sigma^2(\tau,s)$ satisfying $\sigma^2(\tau,s)=\sigma^2(\tau)$. We firstly derive the exact tail asymptotics for the maximum $M_H(T)=\max_{(\tau,s)\in[a,b]\times[0,T]}Z(\tau,s)/\sigma(\tau)$ up crossing some level u with any fixed $0<a<b<\infty$ and $T>0$; and we further derive the extreme limit law for $M_H(T)$. As applications of the main results, we derive the exact tail asymptotics and the extreme limit law for Shepp statistics with stationary Gaussian process, fractional Brownian motion and Gaussian integrated process as input.

Consistent Intensity-Duration-Frequency Relations with Spatial Covariates

Madlen Fischer[1], Oscar E. Jurado[2], Alexander Pasternack[3], Henning W. Rust[4], Jana Ulrich[5]
Freie Universität Berlin, Germany[1,2,3,4,5]
madlen.fischer@met.fu-berlin.de[1], jurado@zedat.fu-berlin.de[2], alexander.pasternack@met.fu-berlin.de[3], henning.rust@fu-berlin.de[4], jana.ulrich@met.fu-berlin.de[5]

Intensity-Duration-Frequency (IDF) Curves are a popular tool in Hydrology for estimating the properties of extreme precipitation events. They describe the relationship between rainfall intensity and durations for a given non-exceedance probability (or frequency). A frequently used method to obtain these curves is based on a hierarchy of two separate statistical models: First the exeedance probabilities for fixed durations are described with extreme value statistics. Afterwards the dependency of fixed exceedance probabilities on duration is described with a regression model. This two step approach can lead to physically inconsistent results, such as higher intensities occurring with higher probability ('crossing of quantiles') particularly for longer durations,.

A consistent estimation of the IDF-relationship can be obtained by simultaneously modeling the probability distributions for all durations. To this end a duration-dependent generalized extreme value distribution (d-GEV, after [1]) is used. Additionally, this approach allows the integration of spatial covariates to describe the variability of the d-GEV parameters in space. Therefore we can model the distribution of extreme precipitation for a range of durations and a certain area in one step. The advantages are parameter reduction and a more efficient use of the available data. We investigate under which conditions this approach leads to an improvement compared to the single site approach.

References

[1] D. Koutsoyiannis et al. , A mathematical framework for studying rainfall intensity-duration-frequency relationships, *Journal of Hydrology*, 1998

A multiscaling intensity–duration–frequency model for extreme precipitation

Hans Van de Vyver[1]

Royal Meteorological Institute of Belgium, Belgium[1]

hvijver@meteo.be[1]

Rainfall intensity–duration–frequency (IDF) curves are a standard tool in urban water resources engineering and management. They express how return levels of extreme rainfall intensity vary with duration. The simple scaling property of extreme rainfall intensity, with respect to duration, determines the form of IDF relationships. It is supposed that the annual maximum intensity follows the generalized extreme value (GEV) distribution.

As well known, for simple scaling processes, the location parameter and scale parameter of the GEV distribution obey a power law with the same exponent. Although, the simple scaling hypothesis is commonly used as a suitable working assumption, the multiscaling approach provides a more general framework. We present a new IDF relationship that has been formulated on the basis of the multiscaling property. It turns out that the GEV parameters (location and scale) have a different scaling exponent.

Next, we apply a Bayesian framework to estimate the multiscaling GEV model and to choose the most appropriate model. Firstly, we collect annual maximum intensity data over a relevant range of rainfall durations. Secondly, we define an approximate likelihood, the "independence" likelihood, in which the correlations have been ignored between maximum intensity data of different durations. Finally, we apply Bayesian inference to obtain the adjusted posterior, which accounts for likelihood misspecification.

It is shown that the model performance increases when using the multiscaling approach. The new model for IDF curves reproduces the data very well and has a reasonable degree of complexity without overfitting on the data.

On the estimation of life table functions with multi-state factors of HIV population

Gurprit Grover[1], Manoj Kumar Varshney[2]

Department of statistics, University Of Delhi, India[1], Department of statistics, Hindu College, University Of Delhi, India[2]

gurpritgrover@yahoo.com[1], mkvarshney_du@yahoo.com[2]

This paper deals with the size of the HIV population over the time that starts with a cohort of fresh HIV and subsequently altered because of fresh HIV births, HIV deaths and transmission from HIV to AIDS; While the birth is an incremental force, death and transmission to AIDS is a decremental force in nature. Our objective is to estimate the size of HIV population over the time taking into consideration of all the incremental and decremental factors. For the estimation of HIV births and deaths use of Kendall's (1948) generalized birth and death process has been taken as a model with the birth and death intensities following Weibull distribution. For the estimation of births and deaths, data have been simulated on the basis of information of US official statistics (UNAIDS/WHO-2002 and UNAIDS-2006). The probabilities for transmission from HIV to AIDS have been extracted from the results of Munoz and Xu (1996) by further smoothing the rates for the

intermediate years in the pattern of log-normal distribution. Finally, a life table has been constructed to exhibit the expectation of life in the state of HIV from 0-20 years.

References

[1] Kendall D.G. (1948): A generalized birth and death Process, *Annals of Mathematical Statistics*, Vol.19, page 1-15.

[2] Munoz Alvaro and XU Jane (1996): Models for Incubation period of AIDS and variations according to Age and Period, *Statistics in Medicine*, Vol.15, page 2459-73

[3] UNAIDS/WHO (2002): Global Statistical Information and Tables Epidemic-Annex, Country Profiles (http://www.journals.org/statistics.php#globalstats).

[4] UNAIDS(2006): Report on Global Epidemic Annex2, Country Profiles (http://www.journals.org/statistics.php#globalstats.)

Presenting authors index