Mathematisches Forschungsinstitut Oberwolfach

Report No. 38/2024

DOI: 10.4171/OWR/2024/38

Mathematics, Statistics, and Geometry of Extreme Events in High Dimensions

Organized by
Richard Davis, New York
Rafal Kulik, Ottawa
Anne Sabourin, Paris
Stilian Stoev, Ann Arbor

19 August – 23 August 2024

Abstract. The workshop brought together researchers contributing to various recent topics in Extreme Value Theory. Discussions and talks included recent probabilistic development in the theory of regular variation, advances in multivariate representations compatible with sparsity structures, statistical inference in both high dimensional and time series frameworks, and novel applications and emerging directions that leverage recent advances in deep learning.

Mathematics Subject Classification (2020): 60-06, 60G70, 62-06.

Introduction by the Organizers

The unifying theme of the workshop was the study of extreme value phenomena from a variety of mathematical perspectives. The workshop gathered more than forty participants from all around the globe, with various backgrounds.

The study of extremes in probability and statistics has a rich history, which can be traced back to the works of Fisher and Tippett (1928) and Gnedenko (1943). From a theoretical standpoint, it is motivated by natural and fundamental questions on the limiting behavior of the maxima of independent random variables, vectors, and stochastic processes. A number of theoretical results have culminated into a rich probabilistic and statistical framework that now has a wide range of applications to science, engineering, insurance, and finance. The prototypical problem that one can solve using the classical extreme value theory tools is the

prediction of extreme events beyond the range of the historically available data. This can be done in a principled way by using the so-called de Haan-Balkema-Pickands theorem when dealing with independent data. The theory and methods extend to the realm of stochastic processes and dependent data through the methods of block-maxima or peaks-over-threshold, among others. This has led to the adoption of the classical extreme value methods as the gold standard in many areas.

The modern developments in the field, however, are motivated by a wide range of practical as well as theoretical challenges, which are, in broad strokes, related to the notion of "big data" as well as the emerging paradigm of "big models" from the field of machine learning.

In applications, the availability of large-scale spatial and spatio-temporal data, for example, in modern weather or environmental measurements and climate models challenges the computational as well as the modeling capabilities of the traditional methodology. This naturally leads to the consideration of large-scale computing architectures and large-scale high-dimensional models. These tectonic paradigm shifts naturally challenge a number of standard theoretical assumptions such as independence/weak dependence, stationarity, and other regularity conditions adopted in the classical theory. As a result, new and emerging directions of research in the field have been actively pursued, many of which were presented during the workshop.

The workshop featured 30 talks from a balanced and representative group of researchers including established senior experts, a number of active early to mid-career researchers as well, as many doctoral students and several junior faculty.

We now provide a brief overview of the variety of novel ideas introduced and discussed during the workshop, structuring the presentation into (possibly overlapping) general themes in an attempt to facilitate reading. Of course the following description cannot be, and is not exhaustive and the reader is referred to the list of extended abstracts for a complete account.

Probability theory and geometry of extremes. Two talks discussed a general unifying treatment of the fundamental notion of multivariate regular variation for random elements taking values in Banach or more generally Souslin spaces (Basrak and Molchanov). An application to limit theorems for extremal processes in the Gromov-Hausdorff topology was presented (Jin). Geometric representations of multivariate tail dependence based on limit sets arising from light-tailed sample clouds were reviewed. They were used to develop novel semi-parametric statistical inference methodology (Wadsworth). Advances on tail dependence and stochastic dominance for multivariate max-stable distributions were presented (Strokorb). Large deviations theory for k-nearest neighbor balls was developed (Owada). Connections between extremes, topological data analysis, and stochastic geometry were reviewed and new limit theorems for sparse high-dimensional geometric graphs were established (Hirsch).

Multivariate extremes: Graphical models, tail (in)dependence, and dimension reduction. The state-of-the art of the active area of graphical models

for multivariate extremes was reviewed, followed by a novel, unifying approach to asymptotic conditional independence in relation with an underlying Lévy measure (Engelke). Possibly cyclic graphical structures in max-linear models were introduced (Krali). An approach to principal component analysis for multivariate extremes was considered based on projecting the angular measure onto a hyperplane (Wan). Further approaches to principal component analysis for multivariate max-stable distributions were considered from the perspective of solving an optimization problem (Janßen). New ideas on defining inner-product structures tailored for the study of multivariate extremes were introduced (Cooley). Dimension reduction frameworks for multivariate extremes were presented, leveraging Machine Learning tools such as Kernel PCA (Samorodnitsky) and Spectral Clustering (Avella Medina). A definition of mutual asymptotic independence among multivariate heavy-tailed variables was proposed and applied to risk measures in financial networks (Fasen-Hartmann).

Statistical inference and extremes in high dimensions. A multiple testing procedure regarding equality of tail indices was presented. Its asymptotic behavior in the high-dimensional asymptotics regime where the dimension increases with the sample size was established (Zhou). In relation to graphical models mentioned above, a penalized estimation of the precision matrix in Hüsler-Reiss model allowing consistent recovery of a sparsity pattern was presented (Volgushev). The question of dimension selection for sparse regular variation was also studied were an ad-hoc information criterion was introduced for this purpose and shown to be weakly consistent (Butsch). A methodology for the automatic selection of the number of components in cluster analysis for extremes was presented and studied theoretically (Bai). A Bayesian spectral inference methodology was introduced, where the angular measure is modeled via a mixture of tilted Dirichlet distributions – well-suited for efficient Monte Carlo inference (Davison). A novel Bayesian approach to inference and conditional sampling from irregularly observed spatial models of extremes was developed (Thannheimer).

Extremes in Time Series. In the context of stationary, heavy-tailed time series, a novel circular bootstrap method for sliding block maxima statistics was presented and justified with asymptotic theory (Bücher). Clusters-over-threshold estimators together with their asymptotic properties were presented (Wintenberger). Long-range dependence for max-stable time series was studied from the perspective of tail dependence and asymptotic theory for extremogram-type estimators was established (Oesting). A framework for optimal prediction of extreme events from the perspective of Neyman-Pearson classification was presented and applied to extreme event prediction in autoregressive and FARIMA time series (Verma).

Novel applications and emerging directions. An overview of a framework for amortized inference via Bayes Neural Networks was presented and applied to large-scale spatial extremes, where computational challenges associated with likelihood methods have been a long-standing challenge (Huser). Asymptotic theory for neural estimators was established (Hentschel). A novel multi-fidelity framework

for the optimal way to combine simulations from different types of large-scale computer experiments arising in the context of extreme ship motions was presented (Pipiras). The importance of using high-quality data and the theoretical challenges in dealing with non-stationarity due to climate change were addressed in the context of forecasting rainfall extremes using scoring rules. The notion of design life period was proposed as an alternative to return period for the purpose of communicating about extremes to engineers and the public (Rootzén).

Acknowledgement: The workshop organizers, on behalf of all participants, would like to thank the staff and the leadership of MFO for the hospitality and for providing a stimulating and inspiring environment for research.

The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, "US Junior Oberwolfach Fellows".

Workshop: Mathematics, Statistics, and Geometry of Extreme Events in High Dimensions

Table of Contents

Abstracts

Kernel PCA and its application to analysis of multivariate extremes

Gennady Samorodnitsky
(joint work with Marco Avella Medina, Richard Davis)

The principal goals of this paper are twofold: 1) provide general insights into kernel PCA showing that it can effectively identify clusters of preimages when the data consists of a discrete signal with added noise, and 2) apply kernel PCA for describing dependence structure of multivariate extremes. Kernel PCA has been motivated as a tool for denoising and clustering of the approximate preimages. The idea is that such structure should be captured by the first principal components in the corresponding function space. We provide some simple insights that naturally lead to clustered preimages when the underlying data comes from a discrete signal corrupted by noise. Specifically, we use the Davis-Kahan theory to give a perturbation bound on the performance of preimages that quantifies the impact of noise in clustering a discrete signal. We then propose kernel PCA as a method for analyzing the dependence structure of multivariate extremes and demonstrate that it can be a powerful tool for clustering and dimension reduction. In this case, kernel PCA is applied only to the extremal part of the sample, i.e., the angular part of random vectors for which the radius exceeds a large threshold. More specifically, we focus on the asymptotic dependence of multivariate extremes characterized by the angular or spectral measure in extreme value theory and provide a careful analysis in the case where the extremes are generated from a linear factor model. We give theoretical analysis of the ingredients in the Davis-Kahan perturbation bounds by leveraging their asymptotic distribution. Our theoretical findings are complemented with numerical experiments illustrating the finite sample performance of our methods.

References

[1] M. Avella Medina, R.A. Davis, and G. Samorodnitsky, *Spectral learning of multivariate extremes*, Journal of Machine Learning Research **25** (2024), 1–36.

[2] B. Basrak, R.A. Davis, and T. Mikosch, *A characterization of multivariate regular variation*, Annals of Applied Probability **12** (2002), 908–920.

[3] G. Blanchard, O. Bousquet, and L. Zwald, *Statistical properties of kernel principal component analysis*, Machine Learning **66** (2007), 259–294.

[4] M. Braun, J. Buhmann, and K.-R. Muller, *On relevant dimensions in kernel feature spaces*, Journal of Machine Learning Research **9** (2008), 1875–1908.

[5] S. Bubeck, *Convex Optimization: Algorithms and Complexity*, Foundations and Trends in Machine Learning **8** (2015), 231–357.

[6] E. Chautru, *Dimension reduction in multivariate extreme value analysis*, Electronic Journal of Statistics **9** (2015), 383–418.

[7] S. Clemencon, H. Jalalzai, S. Lhaut, A. Sabourin, and J. Segers, *Concentration bounds for the empirical angular measure with statistical learning applications*, Bernoulli **29** (2023), 2797–2827.

[8] D. Cooley and E. Thibaud, *Decompositions of dependence for high-dimensional extremes*, Biometrika **106** (2019), 587–604.

[9] D. Deuber, J. Li, S. Engelke, and M. Maathuis, *Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions*, Journal of the American Statistical Association (to appear).

[10] H. Drees and A. Sabourin, *Principal component analysis for multivariate extremes*, Electronic Journal of Statistics **15** (2021), 908–943.

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

[12] S. Engelke and J. Ivanovs, *Sparse structures for multivariate extremes*, Annual Review of Statistics and Its Application **8** (2021), 241–270.

[13] S. Engelke, M. Lalancette, and S. Volgushev, *Learning extremal graphical models in high dimensions*, arXiv preprint, arXiv:2111.00840, 2022.

[14] V. Fomichov and J. Ivanovs, *Spherical clustering in detection of groups of concomitant extremes*, Biometrika **110** (2023), 135–153.

[15] N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke, *Causal discovery in heavy-tailed models*, The Annals of Statistics **49** (2021), 1755–1778.

[16] N. Goix, A. Sabourin, and S. Clémençon, *Learning the dependence structure of rare events: a non- asymptotic study*, Proceedings of The 28th Conference on Learning Theory, PMLR **40** (2015), 843–860.

[17] N. Goix, A. Sabourin, and S. Clémençon, *Sparse representation of multivariate extremes with applications to anomaly detection*, Journal of Multivariate Analysis **161** (2017), 12–31.

[18] T. Hofmann, B. Schölkopf, and A. J. Smola, *Kernel methods in machine learning*, The Annals of Statistics **36** (2008), 1171–1220.

[19] P. Honeine and C. Richard, *Preimage problem in kernel-based machine learning*, IEEE Signal Processing Magazine **28** (2011), 77–88.

[20] H. Jalalzai and R. Leluc, *Feature Clustering for Support Identification in Extreme Regions*, Proceedings of the 38th International Conference on Machine Learning, PMLR **139** (2021), 4733–4743.

[21] A. Janssen and P. Wan, *k-means clustering of extremes*, Electronic Journal of Statistics **14** (2020), 1211–1233.

[22] K. Kim, K. Jung, and H. Kim, *Face recognition using kernel principal component analysis*, IEEE Signal Processing Letters **9** (2002), 40–42.

[23] N. Meyer and O. Wintenberger, *Sparse regular variation*, Advances in Applied Probability **53** (2021), 1115–1148.

[24] N. Meyer and O. Wintenberger, *Multivariate sparse clustering for extremes*, Journal of the American Statistical Association (2023), 1–12.

[25] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, *Kernel PCA and de-noising in feature spaces*, Advances in Neural Information Processing Systems **11** (1998).

[26] S.I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York, 2007.

[27] S.I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Springer, New York, 1987.

[28] C. Rohrbeck and D. Cooley, *Simulating flood event sets using extremal principal components*, The Annals of Applied Statistics **17** (2023), 1333–1352.

[29] R. Rosipal, M. Girolami, L. Trejo, and A. Cichocki, *Kernel PCA for feature extraction and de-noising in nonlinear regression*, Neural Computing and Applications **10** (2001), 231–243.

[30] F. Röttger, S. Engelke, and P. Zwiernik, *Total positivity in multivariate extremes*, The Annals of Statistics **51** (2023), 962–1004.

[31] B. Schölkopf, A. Smola, and K.-R. Muller, *Kernel principal component analysis*, International Conference on Artificial Neural Networks (1997), 583–588.

[32] B. Schölkopf, A. Smola, and K.-R. Muller, *Nonlinear component analysis as a kernel eigenvalue problem, Neural computation*, Neural Computation **10** (1998), 1299–1319.

[33] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.

[34] J. Shawe-Taylor, C. Williams, N. Cristianini, and J. Kandola, *On the eigenspectrum of the gram matrix and the generalization error of kernel-PCA*, IEEE Transactions on Information Theory **51** (2005), 2510–2522.
[35] E. Simpson, J. Wadsworth, and J. Tawn, *Determining the dependence structure of multi variate extremes*, Biometrika **107** (2020), 513–532.
[36] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer, New York, 2008.
[37] D. Tyler, *Statistical analysis for the angular central Gaussian distribution on the sphere*, Biometrika **74** (1987), 579–589.
[38] J. van Zanten and A. van der Vaart, *Reproducing kernel Hilbert spaces of Gaussian priors*, in Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh, Institute of Mathematical Statistics, 2008, 200–222.
[39] Y. Yu, T. Wang, and R. Samworth, *A useful variant of the Davis-Kahan theorem for statisticians*, Biometrika **102** (2015), 315–323.
[40] L. Zwald and G. Blanchard, *On the convergence of eigenspaces in kernel principal component analysis*, Advances in Neural Information Processing Systems **18** (2005).

An approach to dimension reduction via max-stable principal components

Anja Janssen
(joint work with Felix Reinbott)

Principal component analysis (PCA) is one of the most popular dimension reduction techniques in statistics and is especially powerful when a multivariate distribution is concentrated near a lower-dimensional subspace. Multivariate extreme value distributions have turned out to provide challenges for the application of PCA since their constraint support impedes the detection of lower-dimensional structures and heavy-tails can imply that second moments do not exist, thereby preventing the application of classical variance-based techniques for PCA.

One way to motivate the usual PCA is via a regression problem: We search for an optimal linear subspace of dimension $p < d$ to project a d-variate random vector X upon, where optimality is measured in terms of the second moment of the projection error. It is thus equivalent to finding a solution to the regression problem

$$\min_{(B,W)\in\mathbb{R}^{d\times p}\times\mathbb{R}^{p\times d}} E\|BWX - X\|_2^2, \tag{1}$$

i.e. in finding an optimal matrix W which maps our random vector X to a p-dimensional subspace and its optimal counterpart B which reconstructs X from this lower dimensional representation.

However, this approach to PCA is only suitable if our random vector X has finite second moments and it is only meaningful if we are interested in the bulk of observations, not the extremes. Our aim is thus to modify the approach in (1) by first finding a suitable class of distributions for the random vector X and then to fix a suitable distance measure between X and its lower dimensional reconstruction.

We achieve this by first settling on the class of max-stable random vectors which are suitable limit distributions for maxima and for which a convenient representation in terms of spectral functions holds, see [4]. Based on this representation, [2]

and [3] introduced a metric ρ on max-stable distributions, called "spectral distance" in [5], which metricizes convergence in probability. Max-stable distributions are known to be closed under max-linear transformations, which we denote by $\diamond$, see [1] for a general theory of max-linear algebra.

With these three components, i.e. max-stable random vectors, max-linear transformations of them and the spectral distance, we adapt the optimization problem in (1) to the extremal setting by rephrasing it as

$$\min_{(B,W)\in\mathbb{R}^{d\times p}\times\mathbb{R}^{p\times d}} \rho\left(B\diamond W\diamond X, X\right), \tag{2}$$

for a max-stable random vector X.

We are able to characterize the case in which a perfect lower-dimensional reconstruction is possible.

Theorem 1. *Let X be a d-variate simple max-stable random vector and let $p \leq d$. Then the following statements are equivalent.*

(i) *There exists a pair of matrices $(B^*, W^*) \in [0,\infty)^{d\times p}\times[0,\infty)^{p\times d}$ such that*

$$\rho(B^*\diamond W^*\diamond X, X) = 0.$$

(ii) *After possibly permuting the entries of X, there exists a p-variate simple max-stable random vector Y and a matrix $\Lambda\in[0,\infty)^{(d-p)\times p}$, such that*

$$X = \begin{pmatrix}\mathrm{id}_p\\ \Lambda\end{pmatrix}\diamond Y. \tag{3}$$

Furthermore, if (3) holds, we can choose (B^*, W^*) as

$$B^* = \begin{pmatrix}\mathrm{id}_p\\ \Lambda\end{pmatrix}, \quad W^* = \begin{pmatrix}\mathrm{id}_p, & \mathbf{0}_{p\times(d-p)}\end{pmatrix}.$$

It should be noted that the optimization problem in (2) is not convex, nor can it be solved iteratively for growing values of p, but we apply a sequential least squares quadratic programming algorithm which works well in simulations. More details can be found in [7], see also the R-package [6].

References

[1] P. Butkovič, *Max-linear Systems: Theory and Algorithms*, Springer (2010).

[2] R. A. Davis & S.I. Resnick, *Basic properties and prediction of max-ARMA processes*, Adv. Appl. Probab. **21** (1989), 781–803.

[3] R. A. Davis & S.I. Resnick, *Prediction of stationary max-stable processes*, Ann. Appl. Probab. **3** (1993), 497–525.

[4] L. de Haan, *A Spectral Representation for Max-stable Processes*, Ann. Probab. **12** (1984), 1194–1204.

[5] A. Janssen, S. Neblung & S. Stoev, *Tail-dependence, exceedance sets, and metric embeddings*, Extremes **26** (2023), 747–785.

[6] F. Reinbott, *maxstablePCA: Apply a PCA Like Procedure Suited for Multivariate Extreme Value Distributions*, R package version 0.0.1. https://CRAN.R-project.org/package=maxstablePCA (2024)

[7] F. Reinbott & A. Janssen, *Principal component analysis for max-stable distributions*, Arxiv-Preprint 2408.10650 (2024)

Bayesian inference for functional extreme events defined via partially unobserved processes

Max Thannheimer
(joint work with Marco Oesting)

In order to describe the extremal behaviour of some stochastic process X a generalized peaks-over-threshold approach can be used, allowing us to consider single extreme events. These can be flexibly defined as exceedances of a risk functional ℓ, such as a spatial average, applied to X. Inference for the resulting limit process, the so-called ℓ-Pareto process, requires the evaluation of $\ell(X)$ and thus the knowledge of the whole process X. In practical applications, we face the challenge that observations of X are only available at single sites. To overcome this issue, we propose a two-step MCMC-algorithm in a Bayesian framework. In a first step, we sample from X conditionally on the observations in order to evaluate which observations lead to ℓ-exceedances. In a second step, we use these exceedances to sample from the posterior distribution of the parameters of the limiting ℓ-Pareto process. Alternating these steps results in a full Bayesian sampling algorithm for the extremes of X.

1. Basics

Let $S \subset \mathbb{R}^d$ be compact and $C(S)^+$ the space of non-negative continuous real-valued functions $f : S \to [0, \infty)$ equipped with norm $\|f\|_\infty := \sup_{s\in S} |f(S)|$ and the σ-algebra generated by cylinder sets. Consider a sample-continuous process $X = \{X(s) : s \in S\}$ which is in the max-domain of some sample-continuous max-stable process $Z = \{Z(s) : s \in S\}$ of α-Fréchet margins for some $\alpha > 0$.

By [1], the process Z allows for the spectral representation

$$Z(s) \overset{d}{=} \bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} W_k(s), \quad s \in S,$$

where $\{\Gamma_k\}_{k\in\mathbb{N}}$ are the arrival times of a Poisson process on $(0, \infty)$ with unit intensity and W_k are independent copies of a non-negative sample-continuous stochastic process $W = \{W(s) : s \in S\}$, the so-called *spectral process*, satisfying $\mathbb{E}(W(s)^\alpha) = 1$ for all $s \in S$.

ℓ-normalized spectral representation. Let $\ell : C(S)^+ \to \mathbb{R}$ be a non-trivial homogeneous continuous functional, the so-called *risk functional*. We define extreme events as via risk functional exceedances, i.e., realizations of X with a risk functional evaluation above a high threshold u. Provided that $\ell(W) > 0$ almost surely, there exists a non-negative sample-continuous stochastic process $W^{(\ell)} = \{W^{(\ell)}(s), s \in S\}$ such that

$$Z(s) \overset{d}{=} \sqrt[\alpha]{c_\ell} \bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \frac{W_k^{(\ell)}(s)}{\ell(W_k^{(\ell)})}, \quad s \in S, \tag{1}$$

for independent copies $W_1^{(\ell)}, W_2^{(\ell)}, \dots$ of $W^{(\ell)}$ and

$$c_\ell = \int_{C(S)^+} \ell(w)^\alpha \mathrm{d}\mathbb{P}_W(w).$$

The law of $W^{(\ell)}$ is given by

$$\mathbb{P}(W^{(\ell)} \in A) = \frac{1}{c_\ell} \int_{C(S)^+} \ell(w)^\alpha 1_{\{w \in A\}} \mathrm{d}\mathbb{P}_W(w), \quad A \in \mathcal{B}\left(\mathcal{C}(S)^+\right), \tag{2}$$

see [3].

By decomposing X into a ℓ-normalized part $\frac{X}{\ell(X)}$ and an intensity part $\ell(X)$ we have for threshold $u \to \infty$ that

$$\lim_{u\to\infty} \mathbb{P}\left(\frac{X}{\ell(X)} \in B, \ell(X) > ru \,\middle|\, \ell(X) > u\right) = \mathbb{P}\left(\frac{W^{(\ell)}}{\ell(W^{(\ell)})} \in B\right) r^{-\alpha}. \tag{3}$$

We use this limit theorem for Bayesian inference with an MCMC approach. In this work, we assume the shape parameter $\alpha > 0$ to be known and introduce a parameterized *Brown-Resnick process model* for the spectral process $W =: W_\theta$ of the form

$$W(s) = \exp\left(G(s) - G(s_0) - \frac{1}{2}\mathrm{Var}(G(s) - G(s_0))\right), \quad s \in S,$$

where $G = G_\theta$ is a parameterized centered Gaussian process with covariance given as

$$\mathrm{Cov}(G(s), G(\tilde{s})) := c||s||^\beta + c||\tilde{s}||^\beta - c||s - \tilde{s}||^\beta, \quad c > 0,\ \beta \in (0, 2).$$

The Gaussian process G_θ is a two dimensional fractional Brownian motion and can therefore be simulated very efficiently on a regular grid via circulant embedding [2]. Conditionally on being risk functional exceedances, we treat observations at sites $\mathbf{s} = (s_0, \dots, s_n)$ as observations of the spectral process using (3) via

$$\frac{X(\mathbf{s})}{X(s_0)} = \frac{X(\mathbf{s})/\ell(X)}{X(s_0)/\ell(X)} \overset{d}{\approx} \frac{W^{(\ell)}(\mathbf{s})/\ell(W^{(\ell)})}{W^{(\ell)}(s_0)/\ell(W^{(\ell)})} = W^{(\ell)}(\mathbf{s}).$$

2. Algorithm

To account for an unobservable risk functional we introduce a conditional simulation step in our MCMC procedure. In a first step we repetitively simulate our process $W^{(\ell)}$ on a fine grid conditional on our observations. With this fine grid simulations we are able to evaluate the risk functional and decide which observations lead to risk functional exceedances so that we can use them as data in the second step to estimate our parameter θ in a conventional MCMC algorithm. The current implementation can be found online at
`https://github.com/maxthannheimer/FunctionalParetoMCMC`.

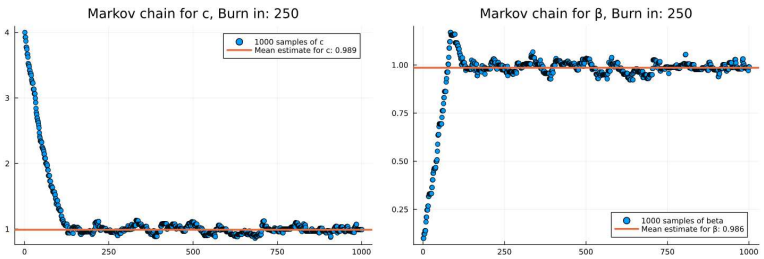


FIGURE 1. Simulate Brown-Resnick-Pareto process with two parameters, $\alpha = 1$, $n = 20$ stations and $N = 400$ fine grid points. 1000 MCMC steps, 250 burn-in, mean of remaining 750.

2.1. **First step.** Via the following MCMC approach we construct a Markov chain with stationary distribution

$$f_{W^{(\ell)}(\mathbf{t})|W^{(\ell)}(\mathbf{s})=\mathbf{x}/x_0}(\mathbf{z}) \propto \ell^{\alpha}(\mathbf{z}) f_{W(\mathbf{t})|W(\mathbf{s})=\mathbf{x}/x_0}(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^N :$$

For fixed parameter θ independently from the current fine grid simulation $\mathbf{w}$ of $W^{(\ell)}$, propose a new $\mathbf{w}'$ from the conditional density $f_{W(\mathbf{t})|W(\mathbf{s})=\mathbf{x}/x_0}$ of the spectral process W_θ, which is just a conditional log Gaussian simulation. The proposal is then accepted with the acceptance rate

$$a(\mathbf{w}, \mathbf{w}') = \min\left\{\frac{\ell^{\alpha}(\mathbf{w})}{\ell^{\alpha}(\mathbf{w}')}, 1\right\}.$$

2.2. **Second step.** Via the following MCMC approach we construct a Markov chain with stationary distribution

$$f_\theta(\theta \mid \mathbf{x}/x_0) \propto f_{W(\mathbf{s})}(\mathbf{x}/x_0 \mid \theta) \frac{1}{c_{\ell,\theta}} \mathbb{E}_\theta\left(\ell(W)^{\alpha} \mid W(\mathbf{s}) = \mathbf{x}/x_0\right) \cdot \pi(\theta) :$$

Given the current parameter θ, we propose a new parameter θ' from a suitable proposal distribution $q(\theta, \cdot)$. We accept the proposal with acceptance rate

$$\min\left\{\frac{q(\theta', \theta) {c_{\ell,\theta'}}^{-1} f_{W(\mathbf{s})}(\mathbf{x}/x_0 \mid \theta') \mathbb{E}_{\theta'}\left(\ell(W)^{\alpha} \mid W(\mathbf{s}) = \mathbf{x}/x_0\right) \pi(\theta')}{q(\theta, \theta') {c_{\ell,\theta}}^{-1} f_{W(\mathbf{s})}(\mathbf{x}/x_0 \mid \theta) \mathbb{E}_\theta\left(\ell(W)^{\alpha} \mid W(\mathbf{s}) = \mathbf{x}/x_0\right) \pi(\theta)}, 1\right\}.$$

Estimators for $c_{\ell,\theta} = \mathbb{E}(\ell(W)^{\alpha})$ and $\mathbb{E}_\theta\left(\ell(W)^{\alpha} \mid W(\mathbf{s}) = \mathbf{x}/x_0\right)$ are available.

REFERENCES

[1] L. de Haan, *A spectral representation for max-stable processes*, Ann. Probab **12(4)** (1984), 1194–1204.

[2] C. Dietrich, G. Newsam, *Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix*, SIAM Journal on Scientific Computing **18(4)** (1997), 1088–1107.

[3] C. Dombry, M. Ribatet, *Functional regular variations, Pareto processes and peaks over threshold*, Statistics and its Interface, **8(1)** (2015), 9–17.

On the optimal prediction of extreme events in heavy-tailed time series with applications to solar flare forecasting

Victor Verma
(joint work with Stilian Stoev, Yang Chen)

The prediction of extreme events in time series is a fundamental problem arising in many applications. In our recent preprint [1], we introduce a framework for studying optimal prediction of these events. In this framework, there is an unobserved response variable Y and an observed covariate vector X, and the objective is to predict whether $Y > F_Y^{\leftarrow}(p)$, where $p \in (0,1)$ is close to one and $F_Y^{\leftarrow}(p)$ is the pth quantile of Y. We use a predictor of the form $\{g(X) > \tau\}$, where g and τ are some Borel function and threshold, respectively, that we must choose. We search among predictors $\{g(X) > \tau\}$ that are *calibrated* - those for which $\mathbb{P}(g(X) > \tau) = 1 - p = \mathbb{P}(Y > F_Y^{\leftarrow}(p))$, i.e., those that raise alarms as frequently as the extreme event occurs. Using quantiles as thresholds, we consider $\{g(X) > F_{g(X)}^{\leftarrow}(p)\}$ to be optimal in the class of calibrated predictors if for any other calibrated predictor $\{k(X) > F_{k(X)}^{\leftarrow}(p)\}$,

$$\mathbb{P}(Y > F_Y^{\leftarrow}(p) \mid g(X) > F_{g(X)}^{\leftarrow}(p)) \geq \mathbb{P}(Y > F_Y^{\leftarrow}(p) \mid k(X) > F_{k(X)}^{\leftarrow}(p));$$

in other words, the *precision* of $\{g(X) > F_{g(X)}^{\leftarrow}(p)\}$ needs to be at least as large as the precision of any other calibrated predictor.

Within our framework, the optimal predictor turns out to be based on a ratio of two conditional densities. The theorem below is a simplified version of Theorem 2.1 in [1]; for the sake of concision, we give simplified versions of our results here.

Theorem 1 (Theorem 2.1 in [1])**.** *Fix $p \in (0,1)$ and let f_1 and f_0 be the conditional densities:*

$$X \mid \{Y > F_Y^{\leftarrow}(p)\} \sim f_1 \quad \textit{and} \quad X \mid \{Y \leq F_Y^{\leftarrow}(p)\} \sim f_0.$$

Define the density ratio

$$r(x) := \frac{f_1(x)}{f_0(x)};$$

then $\{r(X) > F_{r(X)}^{\leftarrow}(p)\}$ is an optimal predictor of $\{Y > F_Y^{\leftarrow}(p)\}$.

There is a sizable literature on density ratio estimation (see [2]), but even using methods from this literature, the density ratio r in Theorem 1 may be hard to estimate accurately. Fortunately, in some situations, optimal prediction does not require estimation of the density ratio. One important example is the situation in which Y is related to X via the model

$$Y = g(X) + \sigma(X)\epsilon,$$

where ϵ is independent of X and σ is a positive function. For this model, we have the following result:

Theorem 2 (Theorem 2.2 in [1]). *For all $p \in (0,1)$, an optimal predictor of $\{Y > F_Y^{\leftarrow}(p)\}$ is $\{k(X) > F_{k(X)}^{\leftarrow}(p)\}$, where*

$$k(X) = \frac{g(X) - F_Y^{\leftarrow}(p)}{\sigma(X)}.$$

Denote the precision $\mathbb{P}(Y > F_Y^{\leftarrow}(p) \mid g(X) > F_{g(X)}^{\leftarrow}(p))$ by $\lambda_p(Y, g(X))$. Letting $p \uparrow 1$, we obtain a tail dependence coefficient, assuming the limit exists:

$$\lambda(Y, g(X)) = \lim_{p \uparrow 1} \lambda_p(Y, g(X)).$$

If it exists, we call $\lambda(Y, g(X))$ the *extremal precision* of $g(X)$. We can consider optimal prediction in the extremal regime $p \uparrow 1$. We say that $g(X)$ is an *extremally optimal* predictor of Y if

$$\lambda(Y, g(X)) = \limsup_{p \uparrow 1} \sup_h \lambda_p(Y, h(X)).$$

Theorem 2.3 in [1] gives conditions under which $g(X)$ is extremally optimal when $Y = g(X) + \epsilon$, where ϵ is independent of X.

Theorem 2 can be used to derive optimal predictors for certain time series models. Consider a stationary, causal autoregressive (AR) model of order d:

$$Y_t = \sum_{j=1}^{d} \phi_j Y_{t-j} + \epsilon_t, \ t \in \mathbb{Z}. \tag{1}$$

Set $\phi := (\phi_1, \ldots, \phi_d)^\top$. Let $\{e_1, \ldots, e_d\}$ be the standard basis of $\mathbb{R}^d$; set $\Phi := (\phi \quad e_1 \quad \cdots \quad e_{d-1}) \in \mathbb{R}^{d \times d}$ and $\phi(h) := \Phi^h e_1$. The optimal predictor for model (1) can be computed using $\phi(h)$:

Theorem 3 (Theorem 3.2 in [1]). *The optimal predictor of $\{Y_{t+h} > F_Y^{\leftarrow}(p)\}$ based on $Y_{t:(t-d+1)} := (Y_t, \ldots, Y_{t-d+1})^\top$ is $\{\widehat{Y}_{t+h}(\phi) > F_{\widehat{Y}_{t+h}(\phi)}^{\leftarrow}(p)\}$, where*

$$\widehat{Y}_{t+h}(\phi) := \phi(h)^\top Y_{t:(t-d+1)}.$$

From data, we can compute a consistent estimator $\widehat{\phi}$ of ϕ using a method like LAD estimation [3]; we plug $\widehat{\phi}$ into the expressions for $\phi(h)$ and $\widehat{Y}_{t+h}(\phi)$ to get

$$\widehat{Y}_{t+h}(\widehat{\phi}) := \widehat{\phi}(h)^\top Y_{t:(t-d+1)}. \tag{2}$$

An estimate $\widehat{q}$ of the pth quantile of $\widehat{Y}_{t+h}(\widehat{\phi})$ can also be computed from data using, e.g., a sample quantile of expressions of the form (2). The approximation $\widehat{Y}_{t+h}(\widehat{\phi})$ to the optimal predictor $\widehat{Y}_{t+h}(\phi)$ has two nice asymptotic properties:

Theorem 4 (Theorem 3.3 in [1]). *Under certain conditions, including $\widehat{\phi} \xrightarrow{\mathbb{P}} \phi$, the predictor $\{\widehat{Y}_{t+h}(\widehat{\phi}) \geq \widehat{q}\}$ is:*

(i) Asymptotically calibrated, i.e., as the sample size $n \to \infty$,

$$\mathbb{P}(\widehat{Y}_{t+h}(\widehat{\phi}) \geq \widehat{q}) \to 1 - p = \mathbb{P}(Y_{t+h} \geq F_Y^{\leftarrow}(p))$$

(ii) Asymptotically optimal, i.e., as $n \to \infty$,

$$\mathbb{P}(Y_{t+h} \geq F_Y^{\leftarrow}(p) \mid \widehat{Y}_{t+h}(\widehat{\phi}) \geq \widehat{q}) \to \mathbb{P}(Y_{t+h} \geq F_Y^{\leftarrow}(p) \mid \widehat{Y}_{t+h}(\phi) \geq F^{\leftarrow}_{\widehat{Y}_{t+h}(\phi)}(p)).$$

Now consider a stationary, invertible moving average (MA) model of order ∞:

$$Y_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j},\ t \in \mathbb{Z}.$$

We have the following characterization of the optimal predictor:

Theorem 5 (Theorem 3.1 in [1])**.** *The optimal predictor of* $\{Y_{t+h} > F_Y^{\leftarrow}(p)\}$ *based on* $Y_s, -\infty < s \leq t$, *is* $\{\widehat{Y}_{t+h}^{(opt)} > F^{\leftarrow}_{\widehat{Y}_{t+h}^{(opt)}}(p)\}$, *where* $\widehat{Y}_{t+h}^{(opt)} = \sum_{j=0}^{\infty} a_{j+h}\epsilon_{t-j}$.

Assume that the ϵ_t's are iid and regularly varying with tail index α and extremal skewness p_ϵ, with $\mathbb{E}\epsilon_t = 0$ when $\alpha > 1$, and that $\sum_{j=0}^{\infty} |a_j|^\delta < \infty$ for some $\delta \in (0, \alpha) \cap (0, 2]$; these conditions are from [4]. Then a simple formula for the extremal precision of $\widehat{Y}_{t+h}^{(opt)}$ can be obtained:

Theorem 6 (Theorem 3.4 in [1])**.** *Suppose that* $\{Y_t\}$ *is invertible. For* $h \geq 0$, *define*

$$\kappa_+(a_j) := p_\epsilon \mathbb{I}(a_j > 0) + (1 - p_\epsilon)\mathbb{I}(a_j < 0) \text{ and } \eta_+(a, h) := \sum_{j=0}^{\infty} \kappa_+(a_{j+h})|a_{j+h}|^\alpha.$$

For every $h \geq 1$, *the extremal precision of* $\widehat{Y}_{t+h}^{(opt)}$ *is* $\lambda(Y_{t+h}, \widehat{Y}_{t+h}^{(opt)}) = \frac{\eta_+(a,h)}{\eta_+(a,0)}$.

Finally, the following curious result says that optimal prediction in the extremal regime can sometimes be done with just the most recent observation:

Proposition 7 (Proposition 3.2 in [1])**.** *Suppose that for some* $h \geq 1$, $\eta_+(a, h) > 0$ *and the sequence* $\{a_j\}$ *is lag-h absolute decreasing, i.e.,* $|a_{j+h}| \leq |a_j|$, *for all* $j \geq 0$. *Then, if* $a_{j+h}a_j \geq 0$ *for all* $j \geq 0$, *we have* $\lambda(Y_{t+h}, Y_t) = \lambda(Y_{t+h}, \widehat{Y}_{t+h}^{(opt)})$.

Our project was motivated by the problem of predicting strong solar flares, which can be done by predicting when a quantity called the solar X-ray flux will surpass a high threshold. See Section 1 of [1] for background on this problem and Section 5 of the same preprint for results from our attempt to predict flares using predictors motivated by Theorems 3 and 5.

References

[1] V. Verma, S. Stoev, and Y. Chen, *On the optimal prediction of extreme events in heavy-tailed time series with applications to solar flare forecasting*, arXiv preprint, arXiv:2407.11887, 2024.

[2] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density Ratio Estimation in Machine Learning*, Cambridge University Press, New York, 2012.

[3] R. Davis, K. Knight, and J. Liu, *M-estimation for autoregressions with infinite variance*, Stochastic Processes and their Applications **40** (1992), 145–180.

[4] R. Kulik and P. Soulier, *Heavy-Tailed Time Series*, Springer, New York, 2020.

Information criteria for the number of directions of extremes in high-dimensional data

Lucas Butsch
(joint work with Vicky Fasen-Hartmann)

In multivariate extreme value analysis, estimating the extremal dependence structure is a challenging task, especially in the context of high-dimensional data. Therefore, a common approach is to reduce the dimensionality by considering only the directions in which extreme values occur. The underlying models are assumed to be sparse regularly varying, recently introduced by Meyer and Wintenberger [4]. An $\mathbb{R}_+^d$-valued random vector $\mathbf{X}$ is called *sparse regularly varying*, if a $\mathbb{S}_+^{d-1}$-valued random vector $\mathbf{Z}$ and a non degenerate random variable Y exist such that as $t \to \infty$,

$$\mathbb{P}\left(\|\mathbf{X}\|/t > y, \pi\left(\mathbf{X}/t\right) \in A \mid \|\mathbf{X}\| > t\right) \longrightarrow \mathbb{P}(Y > y, \mathbf{Z} \in A),$$

for all $y > 0$ and all Borel sets $A \subset \mathbb{S}_+^{d-1} := \{\mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\|_1 = 1\}$ with $\mathbb{P}(\mathbf{Z} \in \partial A) = 0$, where π is the Euclidean projection onto $\mathbb{S}_+^{d-1}$. Sparse regular variation has the advantage of capturing the sparsity structure in which extreme events occur better than multivariate regular variation. Therefore, we use the concept of sparse regular variation and assume that the support of $\boldsymbol{Z}$ is sparse and concentrates on a subsimplex of $\mathbb{S}_+^{d-1}$.

In order to estimate the extremal directions, we define $C_\beta := \{\boldsymbol{x} \in \mathbb{S}_+^{d-1} : x_i > 0 \text{ for } i \in \beta, x_i = 0 \text{ for } i \notin \beta\} \subseteq \mathbb{S}_+^{d-1}$ for $\beta \in \mathcal{P}_d^* := \mathcal{P}_d \setminus \emptyset$, where $\mathcal{P}_d$ is the power set of $\{1, \dots d\}$, which builds a disjoint partition of $\mathbb{S}_+^{d-1}$. From the knowledge of $\mathbb{P}(\boldsymbol{Z} \in C_\beta) > 0$ for all $\beta \in \mathcal{P}_d^*$ conclusions about the direction β can be made. The set of relevant directions is defined as $\mathcal{S}(\boldsymbol{Z}) := \{\beta \in \mathcal{P}_d^* : \mathbb{P}(\boldsymbol{Z} \in C_\beta) > 0\}$ and $s^* := |\mathcal{S}(\boldsymbol{Z})|$ is the number of relevant directions. A major challenge for the estimation of the extreme directions is that the empirical estimators of the probabilities $\mathbb{P}(\boldsymbol{Z} \in C_\beta)$, $\beta \in \mathcal{P}_d^*$, detect more extremal directions than there are true extremal directions. Suppose $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ is a sequence of i.i.d. sparse regularly varying random vectors, $\|\boldsymbol{X}_{(1,n)}\| \geq \cdots \geq \|\boldsymbol{X}_{(n,n)}\|$ is the order statistic of $\|\boldsymbol{X}_1\|, \dots, \|\boldsymbol{X}_n\|$ and the number of extreme observations used for the estimations is denoted by $k_n \in \mathbb{N}$, whereas we assume that $k_n \to \infty$ as $n \to \infty$. Suppose that there exists a sequence of high thresholds $u_n > 0$ for $n \in \mathbb{N}$ such that $k_n/n \sim \mathbb{P}(\|\boldsymbol{X}\| > u_n)$ and $u_n \to \infty$ as $n \to \infty$. We consider the estimators of Meyer and Wintenberger [5]

$$\mathcal{T}_n(\beta) := \textstyle\sum_{j=1}^n \mathbf{1}\left\{\pi\left(\boldsymbol{X}_j/u_n\right) \in C_\beta, \|\boldsymbol{X}_j\| > u_n\right\}$$

and

$$T_n(\beta, k_n) := \textstyle\sum_{j=1}^n \mathbf{1}\left\{\pi\left(\boldsymbol{X}_j/\|\boldsymbol{X}_{(k_n+1)}\|\right) \in C_\beta, \|\boldsymbol{X}_j\| > \|\boldsymbol{X}_{(k_n+1)}\|\right\}$$

of the probability

$$\mathbb{P}(\boldsymbol{Z} \in C_\beta) = \lim_{n\to\infty} \mathbb{P}(\pi(\boldsymbol{X}/u_n) \in C_\beta \mid \|\boldsymbol{X}\| > u_n) =: \lim_{n\to\infty} p_n(\beta).$$

We enumerate the $\beta \in \mathcal{P}_d^*$ such that

$$\mathbb{P}(\boldsymbol{Z} \in C_{\beta_1}) \geq \ldots \geq \mathbb{P}(\boldsymbol{Z} \in C_{\beta_{s^*}}) > \mathbb{P}(\boldsymbol{Z} \in C_{\beta_{s^*+1}}) = \ldots = \mathbb{P}(\boldsymbol{Z} \in C_{\beta_{2^d-1}}) = 0.$$

Then we define

$$\boldsymbol{T}_n(k_n) := (T_n(\beta_1, k_n), \ldots, T_n(\beta_r, k_n))^\top$$

and $\boldsymbol{p}_n, \boldsymbol{\mathcal{T}}_n$, analogously, where $r := |\{\beta \in \mathcal{P}_d^* : T_n(C_\beta, k_n) > 0\}|$.

To detect the non extremal directions and hence, to estimate s^*, the idea of Meyer and Wintenberger [5] is now to fit the model $M_{k_n}^s$ defined for any $s \in \{1, \ldots, r\}$ by a multinomial distribution from the class $\{\mathrm{Mult}(k_n, \widetilde{\boldsymbol{p}}^s) : \widetilde{\boldsymbol{p}}^s \in \Theta_s\}$ where $\Theta_s := \{\widetilde{\boldsymbol{p}}^s = (\widetilde{p}_1^s, \ldots, \widetilde{p}_r^s) \in (0,1)^r : \widetilde{p}_1^s \geq \ldots \geq \widetilde{p}_r^s, \sum_{j=1}^r \widetilde{p}_j^s = 1, \widetilde{p}_{s+1}^s = \ldots = \widetilde{p}_r^s =: \widetilde{\rho}_s\}$ with likelihood function

$$L_{M_{k_n}^s}(\widetilde{\boldsymbol{p}}^s | \boldsymbol{T}_n(k_n)) = k_n! \, (\textstyle\prod_{j=1}^r T_{n,j}(k_n)!)^{-1} \prod_{j=1}^s (\widetilde{p}_j^s)^{T_{n,j}(k_n)} \prod_{j=s+1}^r (\widetilde{\rho}^s)^{T_{n,j}(k_n)}$$

and MLE $\widehat{\boldsymbol{p}}_n^s$ with entries $\widehat{p}_{n,j}^s = T_{n,j}(k_n)/k_n$ for $j = 1, \ldots, s$ and $\widehat{\rho}_n^s = \sum_{j=s+1}^r T_{n,j}(k_n)/((r-s)k_n)$. Now, an information criterion aims to find the true extremal directions $\beta_1, \ldots, \beta_{s^*}$ by considering which Model $M_{k_n}^s$ for $s \in \{1, \ldots, r\}$ best fits the distribution of $\boldsymbol{T}_n(k_n)$ and subsequently derives an estimator for s^*. In the following, we present our results for the Akaike information criterion (AIC) of Meyer and Wintenberger [5] and the derivation of the Bayesian information criterion (BIC) and the quasi Akaike information criterion QAIC. Furthermore, we derived a mean-squared error based information criterion (MUSIC), which is not addressed here. Note that assumptions are necessary for the following results, but will be skipped for the sake of brevity.

AIC: The AIC of Meyer and Wintenberger [5] is defined as $\mathrm{AIC}_{k_n}(s) := -\log L_{M_{k_n}^s}(\widehat{\boldsymbol{p}}_n^s \,|\, \boldsymbol{T}_n(k_n)) + s$ for $s = 1, \ldots, r$, which is motivated by minimizing the expected Kullback-Leibler (KL) divergence between the true distribution of $\boldsymbol{T}_n(k_n)$ and the model $M_{k_n}^s$. Due to our results $\lim_{n\to\infty} \mathbb{P}(\mathrm{AIC}_{k_n}(s) > \mathrm{AIC}_{k_n}(s^*)) < 1$ for $s > s^*$ and $\lim_{n\to\infty} \mathbb{P}(\mathrm{AIC}_{k_n}(s) > \mathrm{AIC}_{k_n}(s^*)) = 1$ for $s < s^*$, follows that the AIC is not a weakly consistent information criterion because asymptotically exists a positive probability of overestimating s^*.

BIC: The underlying idea of the BIC is that given the data $\boldsymbol{T}_n(k_n)$, the model $M_{k_n}^s$ with the highest posterior probability $\mathbb{P}(M_{k_n}^s | \boldsymbol{T}_n(k_n))$ for $s = 1, \ldots, r$ is chosen. Building up on an asymptotic upper bound of $-2 \log \mathbb{P}(M_{k_n}^s | \boldsymbol{T}_n(k_n))$ similar to Cavanaugh and Neath [3], we introduce the *Bayesian information criterion concerning the upper bound* (BICU) for $s = 1, \ldots, r$ as

$$\mathrm{BICU}_{k_n}(s) := -2 \log L_{M_{k_n}^s}(\widehat{\boldsymbol{p}}_n^s \,|\, \boldsymbol{T}_n(k_n)) + 2s \log(k_n) + s \log(r/[2\pi(r-s)]).$$

Motivated by a lower bound of $\nabla^2 \log L_{M_{k_n}^s}$ we further define the *Bayesian information criterion concerning the lower bound* (BICL) for $s = 1, \ldots, r$ as

$$\mathrm{BICL}_{k_n}(s) := -2 \log L_{M_{k_n}^s}(\widehat{\boldsymbol{p}}_n^s \,|\, \boldsymbol{T}_n(k_n)) + s \log(k_n) + s \log(k_n/[2\pi T_{n,1}(k_n)]).$$

For $s \neq s^*$ we derived that

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{BICU}_{k_n}(s) > \mathrm{BICU}_{k_n}(s^*)) = 1$$

as well as $\lim_{n\to\infty} \mathbb{P}(\mathrm{BICL}_{k_n}(s) > \mathrm{BICL}_{k_n}(s^*)) = 1$. Hence, BICU and BICL are weakly consistent and select asymptotically with probability 1 the true Model $M^{s^*}_{k_n}$.

QAIC: The QAIC, which is an abbreviation for quasi Akaike information criterion, is derived similar to the Akaike information criterion (cf. Akaike [1]), but instead of working with the likelihood function of a multinomial distribution as in Meyer and Wintenberger [5] we use the likelihood function of a Gaussian distribution. The reason for this approach is that due to a modified version of Theorem 1 in Meyer and Wintenberger [5] the asymptotic behavior as $n \to \infty$,

$$\sqrt{k_n}\,\mathrm{diag}(\boldsymbol{p}_n^*)^{-1/2}\,(\boldsymbol{\mathcal{T}}_n/k_n - \boldsymbol{p}_n^*) \xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}_r, \boldsymbol{I}_r),$$

holds, where $\boldsymbol{p}_n^* := (p_{n,1}, \ldots, p_{n,s^*}, \rho_n, \ldots, \rho_n)^\top \in \mathbb{R}^r$ with $\rho_n := \sum_{j=s^*+1}^r p_{n,j}/(r-s^*)$. Then the idea is to choose the model which minimizes the expected Kullback-Leibler divergence of the true distribution $\mathbb{P}_{\boldsymbol{\mathcal{T}}_n}$ of $\boldsymbol{\mathcal{T}}_n$ and the normal distribution $\mathcal{N}_r(k_n\widetilde{\boldsymbol{p}}^s, k_n\,\mathrm{diag}(\widetilde{\boldsymbol{p}}^s))$, $\widetilde{\boldsymbol{p}}^s = (\widetilde{p}_1^s, \ldots, \widetilde{p}_s^s, \widetilde{\rho}^s, \ldots, \widetilde{\rho}^s) \in \mathbb{R}_+^r$. For an i.i.d. copy $\widetilde{\boldsymbol{\mathcal{T}}}_n$ of $\boldsymbol{\mathcal{T}}_n$ and $\widehat{\boldsymbol{p}}_n^s(\widetilde{\boldsymbol{\mathcal{T}}}_n)$ defined analogously to $\widehat{\boldsymbol{p}_n^s}$ we approximate the Kullback-Leibler divergence by

$$\mathbb{E}\left[\mathrm{KL}(\mathbb{P}_{\boldsymbol{\mathcal{T}}_n}, \mathcal{N}_r(k_n\widetilde{\boldsymbol{p}}^s, k_n\,\mathrm{diag}(\widetilde{\boldsymbol{p}}^s)))|_{\widetilde{\boldsymbol{p}}^s = \widehat{\boldsymbol{p}}_n^s(\widetilde{\boldsymbol{\mathcal{T}}}_n)}\right].$$

Due to a convergence in distribution of the argument in the previous expectation to a centered random variable we define the *quasi Akaike information criterion* (QAIC) for $s = 1, \ldots, r$ as

$$\mathrm{QAIC}_{k_n}(s) := r\log(2\pi) + r\log(k_n) + \textstyle\sum_{j=1}^s \log(\widehat{p}_{n,j}^s) + (r-s)\log(\widehat{\rho}_n^s) + r + s.$$

Further, we showed that $\lim_{n\to\infty} \mathbb{P}(\mathrm{QAIC}_{k_n}(s) - \mathrm{QAIC}_{k_n}(s^*) > 0) = 1$ for $s \neq s^*$ and therefore the QAIC is weakly consistent.

The performance of the different information criteria were successfully tested in a simulation study.

References

[1] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Automatic Control **AC-19** (1974), 716–723.

[2] L. Butsch & V. Fasen-Hartmann, *Information criteria for the number of directions of extremes in high-dimensional data*, In preparation (2024+).

[3] J.E. Cavanaugh & A.A. Neath, *Generalizing the derivation of the Schwarz information criterion*, Comm. Statist. Theory Methods **28** (1999), 49–66.

[4] N. Meyer & O. Wintenberger, *Sparse regular variation*, Adv. in Appl. Probab. **53** (2021), 1115–1148.

[5] N. Meyer & O. Wintenberger, *Multivariate Sparse Clustering for Extremes*, J. Amer. Statist. Assoc. (2023), 1–12.

Toward inner product spaces of regularly varying random variables

Dan Cooley
(joint work with Kenneth Broadhead)

We begin by describing tail equivalent regularly varying random variables; that is, those which share a normalizing function $b(s) \to \infty$ as $s \to \infty$ leading to a nontrivial limit: $\lim_{s\to\infty} b(s)P(|X| > sx) = kx^{-\alpha}$. An advantage of using a common normalizing function is that it allows one to compare the 'scales' k of the random variables. Extending to random vectors, one definition of multivariate regular variation is $b(s)P(\boldsymbol{X}/s \in \cdot) \xrightarrow{v} \nu_{\boldsymbol{X}}(\cdot)$, where '$v$' denote vague convergence (Resnick, 2007). The limiting measure is a product measure in polar coordinates: $\nu_{\boldsymbol{X}}(dr \times d\boldsymbol{w}) = \alpha r^{-\alpha-1} dH_{\boldsymbol{X}}(\boldsymbol{w})$, where $H_{\boldsymbol{X}}$ is the angular measure, which contains all information for dependence in the limiting measure as well as information about the scales of the components X_i's.

As estimating $H_{\boldsymbol{X}}$ is difficult in high dimensions, we turn our attention to pairwise summaries of tail dependence, the extremal dependence measure (EDM) of Larsson and Resnick (2012), and the tail pairwise dependence measure ($\Sigma_{\boldsymbol{X}}$, TPDM) of Cooley and Thibaud (2019), a matrix of EDM's. The TPDM is positive definite and symmetric. We show the scale the components of $\boldsymbol{X}$ is given by the diagonal elements of the TPDM if $\alpha = 2$, and total mass is given by the diagonal sum if $H_{\boldsymbol{X}}$ is defined by the L_2 norm. We show that TPDM's values are invariant to the dimension of the vector being considered. Finally, we show that the EDM is linear if $\alpha = 2$.

As the EDM/TPDM have the properties of an inner product, we propose a vector space for regularly varying random variables. Beginning with a vector space of consistent regularly varying random variables, we define equivalence classes via the null space of random variables such that $\lim_{s\to\infty} b(s)P(|X| > sx) = 0$. The EDM serves as an inner product for the vector space of equivalence classes. We show via norm convergence, the space $\mathbb{W}$ (with $\alpha = 2$) is complete, thus is a Hilbert space. The implication of this work is that we can utilize the projection theorem to bridge regular-variation-based extremes modeling to traditional methods in statistics.

This developing work is being done jointly with Kenneth Broadhead, CSU.

References

[1] Cooley, D. and Thibaud. E, (2019) Decompositions of dependence for high-dimensional extremes, *Biometrika* 106: 587–604.
[2] Larsson, M. and Resnick, S.I., (2012) Extremal dependence measure and extremogram: the regularly varying case, *Extremes* 15:231–256.
[3] Resnick, S.I., (2007) Heavy-tail phenomena: probabilistic and statistical modeling, Springer, New York.

Piecewise-linear modelling of multivariate geometric extremes

Jennifer Wadsworth
(joint work with Ryan Campbell)

A geometric representation for multivariate extremes, based on the shapes of scaled sample clouds in light-tailed margins and their so-called limit sets [1, 2, 4], has recently been shown to connect several existing extremal dependence concepts [6]. The scaled sample cloud of independent and identically distributed random vectors with common light-tailed (von Mises) margins $\boldsymbol{X}_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ is defined as

$$N_n = \left\{ \frac{\boldsymbol{X}_1}{r_n}, \ldots, \frac{\boldsymbol{X}_n}{r_n} \right\},$$

where a suitable scaling sequence r_n is the $1 - 1/n$ marginal quantile. We assume that this random set converges in probability onto a limit set $G := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq 1\}$, where the 1-homogeneous function g is the *gauge function*. The shape of G is affected by both the margins and dependence of $\boldsymbol{X}$, so we specify to standard exponential margins. We further assume the sufficient, though not necessary, condition for convergence on to G that $\boldsymbol{X}$ has a Lebesgue density $f_{\boldsymbol{X}}$, and that $-\log f_{\boldsymbol{X}}(t\boldsymbol{x})/t \to g(\boldsymbol{x})$, for continuous g [1].

Statistical inference related to this framework is still in its infancy. We outline a framework for inference on the limit set shape, presenting both parametric methods from [9] and a novel semi-parametric approach. In [9] it is shown that the Lebesgue density of the pseudo-polar transformed variables $R = \sum_{j=1}^{d} X_j$, $\boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\| \in \mathcal{S}_{d-1} = \{\boldsymbol{w} : \|\boldsymbol{w}\| = 1\}$ often satisfies

$$f_{R|\boldsymbol{W}}(r \mid \boldsymbol{w}) \propto r^{d-1} \exp\{-r g(\boldsymbol{w})\}[1 + o(1)], \qquad r \to \infty, \tag{1}$$

motivating a truncated gamma model with rate parameter $g(\boldsymbol{w})$ for large values of the conditional variable $R \mid \boldsymbol{W} = \boldsymbol{w}$. In [9] inference was outlined via likelihood based on (1), using parametric models for the gauge function: $g(\boldsymbol{w}; \boldsymbol{\phi})$, for a parameter vector $\boldsymbol{\phi}$. However, parametric gauge functions will often be too inflexible to capture the true shape of the limit set G. We therefore outline a technique for semi-parametric estimation, defining the gauge function in a piecewise-linear manner. Semi-parametric estimation of limit sets has also been considered by [5, 8] for dimension $d = 2$ and [7] for up to $d = 3$. Our proposed piecewise-linear approach is simple to define and calculate, making estimation possible for $d > 3$. Specifically, given a set of reference angles $\boldsymbol{w}_1^\star, \ldots, \boldsymbol{w}_N^\star \in \mathcal{S}_{d-1}$, we assign a length θ_i to each angle, which specifies the L_1 distance from the origin to the boundary of the unit level set of g: $\theta_i = 1/g(\boldsymbol{w}_i^\star)$. The gauge function at other points of $\mathcal{S}_{d-1}$ is defined by piecewise-linear interpolation between these points. For $d \geq 3$ this necessitates a triangulation on which to define (hyper-)planes; we adopt a Delaunay triangulation [3]. To adequately capture the shape of G, the number of reference angles, and hence parameters, N, may be large, so we add an L_2 regularisation penalty on local differences in gradient to a likelihood based on (1).

Beyond limit set estimation, our general approach provides a new class of asymptotically motivated statistical models for the tails of multivariate distributions, and such models can accommodate any combination of simultaneous or non-simultaneous extremes through suitably-specified parametric or piecewise-linear forms for the limit set shape. Extrapolation further into the tail of the distribution is possible via simulation from the fitted model, and probability estimates are feasible in regions where other frameworks for multivariate extreme value analysis struggle.

References

[1] A. Balkema and N. Nolde *Asymptotic independence for unimodal densities*, Advances in Applied Probability **42** (2010), 411–432

[2] R. A. Davis, E. Mulrow and S. I. Resnick *Almost Sure Limit Sets of Random Samples in R^d*, Advances in Applied Probability **20** (1988), 573–599.

[3] B. Delaunay *Sur la sphère vide*, Bulletin de l'Académie des Sciences de l'URSS, Classe des Sciences Mathématiques et Naturelles **6** (1934), 793–800.

[4] K. Kinoshita and S. I. Resnick *Convergence of Scaled Random Samples in R^d*, Annals of Probability **19** (1991), 1640–1663

[5] R. Majumder, B. A. Shaby, B. Reich, and D. Cooley, *Semiparametric estimation of the shape of the limiting bivariate point cloud*, arXiv preprint arXiv:2306.13257 (2023)

[6] N. Nolde and J. L. Wadsworth, *Linking representations for multivariate extremes via a limit set*, Advances in Applied Probability **54** (2022), 688–717.

[7] I. Papastathopoulos, L. de Monte, R. Campbell and H. Rue, *Statistical inference for radially-stable generalized Pareto distributions and return level-sets in geometric extremes*, arXiv preprint arXiv:2310.06130 (2023)

[8] E. S. Simpson and J. A. Tawn, *Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties*, arXiv preprint arXiv:2207.02626 (2022)

[9] J. L. Wadsworth and R. Campbell, *Statistical inference for multivariate extremes via a geometric approach*, Journal of the Royal Statistical Society Series B: Statistical Methodology (to appear) (2024).

Distinguish Forms of Asymptotic Dependence for Heavy Tailed Data

Tiandong Wang

(joint work with Sidney I. Resnick)

In multivariate heavy tail estimation, the support of the limit measure provides information on the dependence structure of the random vector with the heavy tail distribution [7]. However, even in simple circumstances in $\mathbb{R}^2_+$, the positive quadrant in two dimensions, exploratory methods such as scatter, diamond or density estimation plots may have trouble distinguishing between cases:

- *Full dependence* where the limit measure concentrates on a ray of the form $\{(x, y) \in \mathbb{R}^2_+ : y/x = m > 0\}$;
- *Strong dependence* where the support of the limit measure is a proper *connected* subcone of $\mathbb{R}^2_+$ of the form $\{(x, y) \in \mathbb{R}^2_+ : y/x \in [m_l, m_u]\}$;

- *Weak dependence* where the support of the limit measure is all of $\mathbb{R}_+^2$; and
- *Asymptotic independence* where the limit measure concentrates on the axes $\mathbb{R}_+ \times \{0\} \cup \{0\} \times \mathbb{R}_+$.

Estimation and visualization techniques reasonably detect lack of connectedness in the second bullet so we downplay that possibility. However, exploratory techniques can struggle to distinguish the bulleted cases, the most obvious reason being the requirement that data be thresholded according to the distance from the origin. Also, there is current interest in the topic of extremal clustering [1–6] which has some philosophical connections to the current work but is less focused on identifying correct asymptotic models.

We give two test statistics D_n and T_n which help distinguish full vs not-full asymptotic dependence and show the statistics are asymptotically normal but with different asymptotic variances, depending on the case.

Let $\{\mathbf{Z}_i = (X_i, Y_i) : i \geq 1\}$ be iid, set $R_i := X_i + Y_i$, and define $\mathbf{Z}_i^* = (X_i^*, Y_i^*)$ to be the vector such that $X_i^* + Y_i^*$ is the i-th largest order statistic of $\{R_i : 1 \leq i \leq n\}$, which we denote $R_{(i)}$. Consider the following hypotheses: for fixed and known (for now) $0 < a \leq b < 1$,

$$H_0^{(1)} : S([a,b]) = 1, \qquad H_a^{(1)} : S([a,b]) < 1. \tag{1}$$

We first propose a test statistic for the test in (1). Define the statistic

$$\begin{aligned} D_n :=& \frac{1}{k(n)} \sum_{i=1}^{k(n)} \left(1 + \frac{d\big(\mathbf{Z}_i^*, \mathbb{C}_{a,b}\big)}{R_{(k(n))}}\right) \log \frac{R_{(i)}}{R_{(k(n))}} \\ =& H_{k(n),n} + \frac{1}{k(n)} \sum_{i=1}^{k_n} \left(\frac{d\big(\mathbf{Z}_i^*, \mathbb{C}_{a,b}\big)}{R_{(k(n))}}\right) \log \frac{R_{(i)}}{R_{(k(n))}} \end{aligned} \tag{2}$$

where $H_{k(n),n}$ is the Hill estimator of $1/\alpha$ applied to $\{R_i, 1 \leq i \leq n\}$ based on $k(n)$ upper order statistics, and $\{k(n)\}$ is an intermediate sequence (i.e. $k(n) \to \infty$, $n/k(n) \to \infty$, $n \to \infty$). Under $H_0^{(1)}$ as given in (1), we have

$$\sqrt{k(n)}(D_n - 1/\alpha) \Rightarrow \frac{1}{\alpha} N(0,1). \tag{3}$$

Now consider the hypothesis test as formed in Step 2:

$$H_0^{(2)} : S(\{\theta_0\}) = 1 \qquad H_a^{(2)} : S([0,1] \setminus \{\theta_0\}) > 0. \tag{4}$$

where $\theta_0 \in [a,b]$, and to capitalize on hidden regular variation resulting from 2RV, we need the assumption that $[a,b]$ is a proper subset of $[0,1]$. Since $\theta_0 \in [a,b]$ and D_n is unable to distinguish between the two hypotheses in (4), we propose another test statistic. Let $\Theta_i^* := X_i^*/(X_i^* + Y_i^*)$ be the concomitant of $R_{(i)}$, and define

$$T_n := \frac{\sum_{i=1}^{k(n)} \Theta_i^* \log \frac{R_{(i)}}{R_{(k(n))}}}{\sum_{i=1}^{k(n)} \Theta_i^*}. \tag{5}$$

Assume $H_0^{(2)}$ holds and the angular measure $S(\cdot) = \epsilon_{\theta_0}(\cdot)$, $\theta_0 \in (0,1)$. Suppose $b_0(t) \in RV_{1/(\alpha(1+\rho))}$ and $\alpha_0 = \alpha(1+\rho)$. Let $\{k(n)\}$ be an intermediate sequence.

Then for $W(\cdot)$ a standard Brownian motion we have

$$\sqrt{k(n)}\left(T_n - \frac{1}{\alpha}\right) \Rightarrow \frac{1}{\alpha}\left(\int_0^1 W(s)\frac{\mathrm{d}s}{s} - W(1)\right) \stackrel{d}{=} \frac{1}{\alpha}W(1) \sim N(0, 1/\alpha^2). \tag{6}$$

Assume strong dependence exists such that $\mathrm{supp}S(\cdot) = [a, b]$. Suppose $A(t) \in RV_{-\rho}$, $\rho > 0$, so $b_0(t) \in RV_{1/(\alpha(1+\rho))}$ and $\alpha_0 = \alpha(1+\rho)$. As before, $\{k(n)\}$ is an intermediate sequence. Define

$$\mu := \int_a^b xS(\mathrm{d}x), \qquad \sigma^2 := \int_a^b (x-\mu)^2 S(\mathrm{d}x),$$

and under strong dependence assumption $H_a^{(2)}$, we have

$$\begin{aligned}\sqrt{k(n)}\left(T_n - \frac{1}{\alpha}\right) &\Rightarrow \frac{(1+\sigma^2/\mu^2)^{1/2}}{\alpha}\left(\int_0^1 \frac{W(s)}{s}\mathrm{d}s - W(1)\right) \\ &\stackrel{d}{=} \frac{(1+\sigma^2/\mu^2)^{1/2}}{\alpha}W(1) \sim N\big(0, \frac{1}{\alpha^2}(1+\sigma^2/\mu^2)\big).\end{aligned} \tag{7}$$

We then apply the proposed testing procedure to real data. We download the daily adjusted stock prices of Chevron (CVX), Exxon (XOM) and Apple (AAPL) during the time period from January 04, 2016 to December 30, 2022. To remove the possible serial dependence of stock returns, we compute the log returns of these three stocks using their every-other-day prices. The acf plots show little serial dependence for all three stocks. This leads to a reduced dataset of $n = 880$ observations for each stock. After applying the tests, we conclude that the absolute returns of CVX and XOM show full asymptotic dependence, whereas the absolute returns of CVX and AAPL show weak asymptotic dependence.

References

[1] M. Avella-Medina, R.A. Davis, and G. Samorodnitsky (2024). Spectral Learning of multivariate extremes. *Journal of Machine Learning Research 25*(124), 1–36.

[2] R.A. Davis, L. Fernandes, and K. Fokianos (2023). Clustering multivariate time series using energy distance. *Journal of Time Series Analysis 44*(5-6), 487–504.

[3] H. Drees, A. Janßen, and S. Neblung (2021). Cluster based inference for extremes of time series. *Stochastic Process. Appl. 142*, 1–33.

[4] H. Drees and A. Sabourin (2021). Principal component analysis for multivariate extremes. *Electron. J. Stat. 15*(1), 908–943.

[5] V. Fomichov and J. Ivanovs (2023). Spherical clustering in detection of groups of concomitant extremes. *Biometrika 110*(1), 135–153.

[6] A. Janßen and P. Wan (2020). k-means clustering of extremes. *Electron. J. Stat. 14*(1), 1211–1233.

[7] J. Lehtomaa and S. Resnick (2020). Asymptotic independence and support detection techniques for heavy-tailed multivariate data. *Insurance: Mathematics and Economics 93*, 262–277.

Regular Variation of General Random Elements I+II

Bojan Basrak and Ilya Molchanov
(joint work with Nikolina Milincevic)

By now the concept of regularly varying random variables is well understood, along with its multivariate extension for random vectors, see, for example, [7, 15, 16]. Starting from the regular variation in the space of continuous functions on $[0, 1]$ studied by [8], these already classical constructions have been extended for random elements in metric spaces in [11], [14] and [17].

Different approaches to the extension of regular variation and vague convergence to general metric spaces appeared in the literature. One is based on the concept of modulus, and was particularly created with the so-called star-shaped metric spaces in mind, see [17] and also [3] who pushed this machinery beyond the star-shaped spaces framework. Another approach is based on an abstract concept of boundedness (or bornology), see [2] and [13], who develop an approach to the vague convergence promoted in [12]. As one of our auxiliary, but potentially quite practical results, we show that the two approaches to regular variation on a general topological space $\mathbb{X}$ are equivalent under mild topological assumptions. Both approaches are built on two components: a scaling operator $T_t : \mathbb{X} \to \mathbb{X}$ (an action of the group $(0, \infty)$ on $\mathbb{X}$) and a bornology $\mathcal{X}$ (a family of the so called bounded sets). For an $\mathbb{X}$-valued random element ξ to be regularly varying we assume that one can find a sequence (a_n), $a_n \to \infty$, and a nontrivial measure μ such that

$$n\mathbb{P}(T_{1/a_n}\xi \in A) \to \mu(A)\,,$$

for any $A \in \mathcal{X}$ with $\mu(\partial A) = 0$. Note that there exist earlier generalisations of regular variation made by mimicking the classical constructions in Euclidean spaces, i.e., by *constructing* $\mathcal{X}$ after excluding a zero point or a closed cone C from the underlying separable metric space (see [8], [11] and [14]). However, it is not often recognised that the resulting theory depends not only on the excluded cone, but also on the concept of boundedness (which is traditionally referred to as ideal in set theory) one constructs on the residual space remaining after this exclusion. In other words, along with what is excluded it is important to know how the exclusion is done. We illustrate this phenomenon with examples, e.g., involving the very classical situation of the joint regular variation of two independent Pareto random variables.

It should be noted that generalisations of the regular variation concept in general spaces (including the overview of these results in [13]) have been mostly derived by mimicking the classical constructions in Euclidean spaces. In particular, it has been always imposed that the underlying space is metric separable, not to say Polish. In this work we show that the metrisability of the carrier space is not necessary for the study of the tail behaviour of measures and so in this way we disentangle the roles of metric (which is only necessary to determine the topology) and scaling in studies of regular variation. Examples of non-metrisable space are abundant, e.g., the space of continuous functions with pointwise convergence (the

C_p-space, see [18]), or infinite-dimensional Banach spaces with weak topologies, or the space of linear functionals on Banach spaces with a weak-star topology. Further important examples of non-Polish spaces arise as continuous images of Polish spaces, which are (not necessarily metrisable) Souslin spaces. Even in the context of metric spaces, separability is violated in many natural examples, e.g. for the space of bounded real-valued sequences with the uniform distance or the space of functions of bounded variation.

There are many motives for generalising regular variation and the related notion of vague convergence: i) eliminating irrelevant parts of the theory seems appealing from the mathematical point of view; ii) relying on general concepts of scaling, ideal (boundedness) and topology simplifies applications of the theory in various settings; and iii) many intriguing examples do not fit standard theory and some insights are only revealed after the generalisation. Although, the standard definition of regular variation involves convergence of sequences of measures or measures indexed by a real number, we show that all the results concerning the vague convergence can be formulated in non-sequential topological spaces.

There are several further reasons to explore the regular variation property in non-Polish spaces. For instance in functional spaces, standard concepts of regular variation rely on the use of the uniform metric [8] or its variants like the Skorokhod metric in [3]. Putting a weaker topology on such a space allows for more flexibility in choosing continuous functions with values in this space, and so makes it possible to use continuous mapping arguments. On the other hand, establishing the regular variation property on a standard space with a nontraditional stronger topology allows for a larger class of continuous mappings on that space.

In this work we build the general theory of regular variation on topological spaces starting from the first principles, relying on the idea of a measurable group action on a topological space. For the treatise of weak convergence of bounded measures on topological spaces, we refer to [4, 5]. In the following we extend this to unbounded measures and define their *vague convergence*. This substantially generalises the setting of [1]. The central role is played by a general definition of the vague convergence with respect to an ideal of sets, which generalises the recently introduced useful concepts of boundedness (or bornology) on the space, see [2] and [13]. To minimise the topological assumptions, we work with measures on the Baire σ-algebra on the carrier space. If this space is perfectly normal, then the Baire and Borel σ-algebras coincide and all results become applicable in the Borel setting. Since we do not require that the underlying space is sequential, in order to ensure broader applicability of the defined vague convergence, we work with convergence of nets even despite the fact that the regular variation is defined for measures parametrised by a real number. In view of this, some results require continuity of measures with respect to nets (instead of conventional sequences), such measures are usually called τ-additive. For instance, this is guaranteed if the carrier space is hereditary Lindelöf, meaning that each open covering of any open set admits a countable subcovering.

The essential ingredients needed to study tail behaviour are scaling and functions homogeneous for this scaling (moduli or gauge functions). Such functions provide information how large is an element of the space. Homogeneous measures (also called tail measures) arise as limiting objects characterising tail behaviour of regularly varying measures. Note in this relation that homogeneous measures naturally arise as Lévy measures of stable laws on semigroups, see [6,10]. Generic representation results for homogeneous measures on measurable spaces (without any topological assumptions) have been obtained in [10] and further rediscovered in [9]. In difference to [10], we work here not at the limit (being a homogeneous measure), but in the under limit setting where measures are not necessarily homogeneous but regularly varying. We systematically work with regular variation of a positive exponent α. The case of negative α is easily obtained by redefining the scaling or by a continuous mapping argument.

While our standing assumption is the Baire measurability of the scaling operation, additional properties, like continuity of the scaling and the existence of a continuous modulus make it possible to prove useful equivalent characterisations of the regular variation property, which then becomes easier to check. This is also the case in other works, notably in [3] where the continuity assumption is tacitly added from Section 3 as a standing assumption to handle the vague convergence.

As explained in [2], the definition of the vague convergence refers to a family of sets, which are designated as bounded. These sets build a bornology (or boundedness) on the carrier space. A very similar idea, under the name of localisation, is used in [12] in the context of Polish spaces. In general, this family of sets forms an ideal (the dual to a filter), which becomes a bornology if it covers the whole space. The systematic use of ideals simplifies and generalises results concerning the vague convergence, leads to a useful formulation of a continuous mapping theorem and makes it possible to view this ideal as a variable parameter in the definition of the regular variation. It also makes it possible to keep the ground space fixed and, instead of excluding some cones from it as common in [14], work with different ideals on the same space. As our examples show, it is important not only what is excluded from the space, but also how bounded sets approach the excluded region. We show that this effect appears even on the Euclidean plane. In difference to all previous works, we include into consideration also ideals which might contain some scaling-invariant elements. This is particularly essential in working with regular variation on the space of sets, since scaling-invariant closed sets are abundant.

A constructive definition of an ideal employs continuous nonnegative homogeneous functions on the carrier space $\mathbb{X}$. The simplest case arises when the space does not contain scaling invariant elements and admits a strictly positive homogeneous function $\tau : \mathbb{X} \to (0, \infty)$. In this case such a function is called a proper modulus and is used to generate a bornology using its upper level sets

$$\{x \in \mathbb{X} : \tau(x) > u\}, \quad u > 0,$$

see [17]. We show in natural examples (e.g., looking at $\mathbb{R}^\infty$) that such a function does not necessarily exist. This was observed in [3] who worked with a countable family of moduli which are homogeneous functions, possibly vanishing on some

elements of the carrier space. We follow this approach and identify main features which explain this situation and clarify relationships between ideals and homogeneous functions. We present a result, which establishes the equivalence between the existence of a continuous modulus and topological properties of the ideal.

Our general approach leads also to simplification of proofs of many well-known results for Polish spaces, most notably, the equivalence between the regular variation property and the existence of a spectral measure which arises from the polar decomposition. In the setting of metric spaces, the roles of separability and completeness assumptions are highlighted in each particular result. Furthermore, we present a construction of the product of ideals, which applies to obtain the hidden regular variation phenomenon and which does not employ any metric on the underlying space. It is shown that in case of a metric space we recover the conventional construction based on excluding a neighbourhood of a given cone.

References

[1] M. Barczy and G. Pap. Portmanteau theorem for unbounded measures. *Statist. Probab. Lett.*, 76(17):1831–1835, 2006.

[2] B. Basrak and H. Planinić. A note on vague convergence of measures. *Statist. Probab. Lett.*, 153:180–186, 2019.

[3] M. Bladt, E. Hashorva, and G. Shevchenko. Tail measures and regular variation. *Electron. J. Probab.*, 27:Paper No. 64, 43, 2022.

[4] V. I. Bogachev. *Measure Theory. Vol. II*. Springer-Verlag, Berlin, 2007.

[5] V. I. Bogachev. *Weak Convergence of Measures*. American Mathematical Society, Providence, RI, 2018.

[6] Y. Davydov, I. Molchanov, and S. Zuyev. Strictly stable distributions on convex cones. *Electron. J. Probab.*, 13:no. 11, 259–321, 2008.

[7] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, NY, 1 edition, 2010.

[8] L. de Haan and T. Lin. On convergence toward an extreme value distribution in $C[0, 1]$. *Ann. Probab.*, 29(1):467–483, 2001.

[9] C. Dombry, E. Hashorva, and P. Soulier. Tail measure and spectral tail process of regularly varying time series. *Ann. Appl. Probab.*, 28(6):3884–3921, 2018.

[10] S. N. Evans and I. Molchanov. Polar decomposition of scale-homogeneous measures with application to Lévy measures of strictly stable laws. *J. Theoret. Probab.*, 31(3):1303–1321, 2018.

[11] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006.

[12] O. Kallenberg. *Random Measures, Theory and Applications*. Springer, 2017.

[13] R. Kulik and P. Soulier. *Heavy-Tailed Time Series*. Springer, New York, 2020.

[14] F. Lindskog, S. I. Resnick, and J. Roy. Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Probab. Surv.*, 11:270–314, 2014.

[15] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York, 1987.

[16] S. I. Resnick. *Heavy-Tail Phenomena*. Springer, New York, 2007.

[17] J. Segers, Y. Zhao, and T. Meinguet. Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes*, 20(3):539–566, 2017.

[18] V. V. Tkachuk. *A C_p-Theory Problem Book. Topological and Function Spaces*. Springer, New York, 2011.

Stochastic orderings among multivariate extreme value distributions

Kirstin Strokorb
(joint work with Michela Corradini)

1. Introduction

Research on stochastic orderings covers several decades as documented for instance in the textbook treatments of [1] and [2]. Related notions of positive dependence in Extremes have been established for some time, e.g. it is well-known that multivariate max-infinitely divisible distributions, exhibit positive association [3, 4]. Recently, an extremal version of the popular MTP2 property [5,6] has been linked to graphical modelling, sparsity and implicit regularisation in multivariate extreme value models [7].

In this work [8] we focus on the order of positive quadrant dependence (PQD order, also termed concordance order) among multivariate max-stable distributions, which is linked to the comparative behaviour of minima and maxima of the components of max-stable vectors. We show that PQD-order holds for two max-stable distributions if and only if it holds for the corresponding exponent measures. Among marginally closed parametric models the popular asymmetric Dirichlet family and the Hüsler-Reiß family turn out to be PQD-ordered according to the natural order within their parameter spaces.

2. Prerequisites

Since the PQD-order compares dependence structures, we may assume without loss of generality that the equal margins of the considered multivariate distributions are *standard Fréchet*, i.e. they have cumulative distribution function (cdf) $\Phi_1(x) = \exp\{-1/x\}$, $x > 0$. Such distributions are often termed *simple max-stable* distributions, and various representations are known. For instance, the cdf of any d-variate simple max-stable random vector $\boldsymbol{X}$ can be expressed as

$$\mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}) = \exp\left\{ -\Lambda\Big(\big\{y \in \mathbb{R}^d \,:\, y_i > x_i \text{ for some } i \in \{1,\dots,d\}\big\}\Big)\right\}$$

for $\boldsymbol{x} \in (0,\infty]^d$, where Λ is a (-1)-homogeneous measure on $[0,\infty)^d \setminus \{\boldsymbol{0}\}$, termed *exponent measure* [4].

For two d-variate random vectors $\boldsymbol{X}, \widetilde{\boldsymbol{X}}$ (or their corresponding distribution functions G and $\widetilde{G}$), we say that

- $\boldsymbol{X}$ is UO-*smaller* than $\widetilde{\boldsymbol{X}}$, denoted $\boldsymbol{X} \le_{\mathrm{uo}} \widetilde{\boldsymbol{X}}$, if $\mathbb{P}(\boldsymbol{X} \in U) \le \mathbb{P}(\widetilde{\boldsymbol{X}} \in U)$ for all upper orthants $U \subset \mathbb{R}^d$;
- $\boldsymbol{X}$ is LO-*smaller* than $\widetilde{\boldsymbol{X}}$, denoted $\boldsymbol{X} \le_{\mathrm{lo}} \widetilde{\boldsymbol{X}}$, if $\mathbb{P}(\boldsymbol{X} \in L^c) \le \mathbb{P}(\widetilde{\boldsymbol{X}} \in L^c)$ for all lower orthants $L \subset \mathbb{R}^d$.
- $\boldsymbol{X}$ is PQD-*smaller* than $\widetilde{\boldsymbol{X}}$, denoted $\boldsymbol{X} \le_{\mathrm{PQD}} \widetilde{\boldsymbol{X}}$, if $\boldsymbol{X} \le_{\mathrm{uo}} \widetilde{\boldsymbol{X}}$ and $\boldsymbol{X} \ge_{\mathrm{lo}} \widetilde{\boldsymbol{X}}$.

When defining corresponding notions of multivariate orders for exponent measures of simple max-stable random vectors, we need to take into account that exponent measures explode at the origin. Hence, it is natural to adapt the above definitions by restricting the survival-type test sets of upper orthants U or complements of lower orthants L^c to those, which are bounded away from the origin.

3. Main results

While results regarding lower orthants are straightforward and included for the sake of completeness, the key novelty lies in understanding the upper orthant order better in this context, and thereby the comparative behaviour of minima of max-stable random vectors. The implication "$\Lambda \leq_{\mathrm{uo}} \widetilde{\Lambda} \implies G \leq_{\mathrm{uo}} \widetilde{G}$" in the following theorem is non-trivial.

Theorem 1. *Let G and $\widetilde{G}$ be d-variate simple max-stable distributions with exponent measures Λ and $\widetilde{\Lambda}$, respectively. Then*

(a) $G \leq_{\mathrm{lo}} \widetilde{G} \iff \Lambda \leq_{\mathrm{lo}} \widetilde{\Lambda}$;

(b) $G \leq_{\mathrm{uo}} \widetilde{G} \iff \Lambda \leq_{\mathrm{uo}} \widetilde{\Lambda}$;

(c) If $d = 2$, then $G \leq_{\mathrm{PQD}} \widetilde{G} \iff G \leq_{\mathrm{uo}} \widetilde{G} \iff G \geq_{\mathrm{lo}} \widetilde{G}$.

The assumption $d = 2$ is essential in the last part. In higher dimensions these equivalences are no longer true. [8] includes counterexamples of so-called *Choquet-* or *Tawn-Molchanov* type, whose exponent measures lie on rays through indicator vectors. Generally, the stochastic ordering of Choquet max-stable random vectors can be easily understood in terms of their parametrizations either via *extremal coefficients* (which encodes lower orthant order) or via *tail dependence coefficients* (which encodes upper orthant order). This helps to reveal that a variety of phenomena may occur when comparing max-stable stochastic models. Still, we obtain the following corollaries from Theorem 1.

Corollary 2. *Let G and $\widetilde{G}$ be d-variate simple max-stable distributions with exponent measures Λ and $\widetilde{\Lambda}$, then*

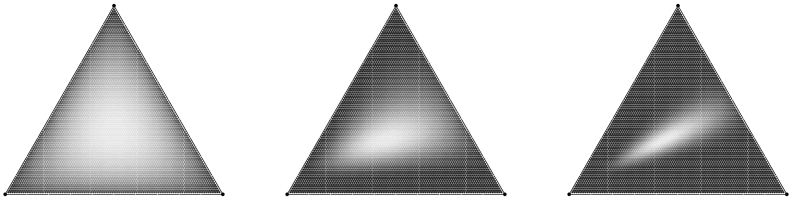
$$G \leq_{\mathrm{PQD}} \widetilde{G} \quad \iff \quad \Lambda \leq_{\mathrm{PQD}} \widetilde{\Lambda}.$$

Corollary 3. *Let G_{indep}, G_{dep}, G and G^* be d-dimensional simple max-stable distributions, where G_{indep} and G_{dep} represent the models with fully independent and fully dependent components, respectively, while G^* is the unique simple max-stable distribution of Choquet/Tawn-Molchanov type sharing an identical set of extremal coefficients with the generic max-stable distribution G. Then*

$$G_{\mathrm{indep}} \leq_{\mathrm{PQD}} G^* \leq_{\mathrm{PQD}} G \leq_{\mathrm{PQD}} G_{\mathrm{dep}}.$$

Moreover, two popular parametric families of multivariate extreme value distributions that are closed under taking marginal distributions, are the max-stable Dirichlet [9] and Hüsler-Reiß [10] distributions. These turn out to be PQD-ordered according to the natural order in their respective parameter spaces [8]. For the

Hüsler-Reiß model this holds true even for the supermodular order. Figure 1 illustrates how the PQD-order for the Dirichlet model corresponds to differences in the concentration of mass of its angular measure.



$\boldsymbol{\alpha} = (1.5, 1.5, 1.5)$ $\boldsymbol{\alpha} = (1.5, 3, 12)$ $\boldsymbol{\alpha} = (1.5, 12, 96)$

FIGURE 1. Heat maps of angular densities of the Dirichlet model; brighter colours represent larger values. The corresponding max-stable distributions are PQD-ordered, increasing from left to right.

REFERENCES

[1] M. Shaked and J. G. Shanthikumar (2007), *Stochastic Orders*, Springer Series in Statistics, Springer, New York.

[2] A. Müller and D. Stoyan (2002), *Comparison Methods for Stochastic Models and Risks*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester.

[3] A. W. Marshall and I. Olkin (1983), *Domains of attraction of multivariate extreme value distributions*, Ann. Probab. **11**(1), 168–177.

[4] S. I. Resnick (1987), *Extreme Values, Regular Variation, and Point Processes*, Vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*, Springer-Verlag, New York.

[5] S. Karlin and Y. Rinott, Y. (1980), *Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions*, J. Multivariate Anal. **10**(4), 467–498.

[6] S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, P. Zwiernik (2017), *Total positivity in Markov structures*, Ann. Statist. **45**(3), 1152–1184.

[7] F. Röttger, S. Engelke and P. Zwiernik (2023), *Total positivity in multivariate extremes*, Ann. Statist. **51**(3), 962–1004.

[8] M. Corradini and K. Strokorb (2024), *Stochastic ordering in multivariate extremes*, Extremes **27**, 357–396.

[9] S. G. Coles and J. A. Tawn (1991), *Modelling extreme multivariate events*, J. Roy. Statist. Soc. Ser. B Stat. Methodol. **53**(2), 377–392.

[10] J. Hüsler and R.-D. Reiß (1989), *Maxima of normal random vectors: between independence and complete dependence*, Statist. Probab. Lett. **7**(4), 283–286.

Convergence of Extremal Processes in Spaces of Growing Dimension

Bochen Jin
(joint work with Ilya Molchanov)

Consider a random walk $S_k^{(d)}$, $k \geq 0$, in d-dimensional Euclidean space with square integrable centred increments such that the expected square norm of the increment is one. The values of this random walk for $k = 0, \dots, n$ normalised by $\sqrt{n}$ are considered as a finite metric space $\mathcal{Z}_n^{(d)}$ which is embedded in $\mathbb{R}^d$ with the induced metric. Under condition of uniform smallness type on the d components of the *increment*, Kabluchko and Marynych (2022) proved that, as n and d go to infinity in any regime, then the metric space $\mathcal{Z}_n^{(d)}$ converges in probability in the Gromov–Hausdorff metric to the Wiener spiral. The latter space is the space of all indicators $\mathbf{1}_{[0,t]}$, $t \in [0,1]$, embedded in $L^2([0,1])$, equivalently, the interval $[0,1]$ with the metric $r(t,s) = \sqrt{|t-s|}$.

In their subsequent preprint, Kabluchko and Marynych (2023) showed that in the heavy-tailed case with $\alpha \in (0,1)$, the limiting metric space is random and is derived from an infinite-dimensional version of a subordinator (called a crinckled subordinator), assuming certain condition on the joint growth regime of n and d.

In the talk, we study an extremal version of this setting, where the random walk is replaced by a sequence of successive maxima and the underlying metric in $\mathbb{R}^d$ is taken to be ℓ_∞.

Let

$$X_i^{(d)} := (X_{i1}^{(d)}, \dots, X_{id}^{(d)}), \quad i \in \mathbb{N},$$

be a sequence of independent and identically distributed random vectors in $\mathbb{R}_+^d$. Define the sequence of consecutive componentwise maxima of these random vectors by letting $M_0^{(d)} := 0$ and

$$M_i^{(d)} := (M_{i1}^{(d)}, \dots, M_{id}^{(d)}), \quad i \in \mathbb{N},$$

where

$$M_{ij}^{(d)} = \max(X_{1j}^{(d)}, \dots, X_{ij}^{(d)}).$$

For two vectors $x, y \in \mathbb{R}_+^d$, denote by $x \vee y$ the vector built of the componentwise maxima of x and y and by $x \wedge y$ of the componentwise minima. Then

$$M_i^{(d)} = X_1^{(d)} \vee \ldots \vee X_i^{(d)}, \quad i \in \mathbb{N}.$$

Let c_{00} denote the space of sequences $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}_+^\infty$ with only finite number of non-zero components and the ℓ_∞-norm, that is,

$$\|x - y\|_\infty = \sup_{n\in\mathbb{N}} |x_i - y_i|.$$

Furthermore, let $\tilde{c}_{00} = (c_{00}/\sim)$ be the quotient space of c_{00} up to all permutations of the nonzero components. Fix an $m \in \mathbb{N}$ and define a family $\mathcal{S}_m$ of subsets A of $\tilde{c}_{00}$ such that there exists an $s > 0$, so that each $x \in A$ is equivalent to a sequence y with $\min(y_1, \dots, y_m) \geq s$.

If $m = 1$, then $\mathcal{S}_1$ is the family of sets A such that $\|x\|_\infty \geq s$ for some $s > 0$ and all $x \in A$.

Let $n = n(d)$ depend on d so that $n(d) \to \infty$ as $d \to \infty$. Assume that there exists a sequence $a(n, d) \in (0, \infty)$ (we usually suppress the dependence on d and write $a(n) = a(n, d)$) such that

$$n\mathbf{P}\left\{(a(n))^{-1} X^{(d)} \in \cdot\right\} \xrightarrow{v} \nu(\cdot) \quad \text{as } n, d \to \infty. \tag{1}$$

The vague convergence is in $\tilde{c}_{00}$ equipped with the ideal (boundedness) $\mathcal{S}_m$. By definition, the limiting measure ν is finite on $\mathcal{S}_m$.

Furthermore, assume that the asymptotic orthogonality condition holds, that is, for all $\varepsilon > 0$,

$$n^2\mathbf{P}\left\{\|X_1^{(d)} \wedge X_2^{(d)}\|_\infty \geq \varepsilon a(n)\right\} \to 0 \quad \text{as } n, d \to \infty, \tag{2}$$

which can be relaxed to

$$\mathbf{P}\left\{\|(X_1^{(d)} \vee \ldots \vee X_n^{(d)}) \wedge (\widetilde{X}_1^{(d)} \vee \ldots \vee \widetilde{X}_n^{(d)})\|_\infty \geq \varepsilon a(n)\right\} \to 0 \qquad \text{as } d \to \infty,$$

where $(\widetilde{X}_i^{(d)})_{i\in\mathbb{N}}$ are independent copies of $X^{(d)}$ and independent with $(X_i^{(d)})_{i\in\mathbb{N}}$.

Denote

$$\mathcal{M}_n^{(d)} := \left\{M_1^{(d)}, \ldots, M_n^{(d)}\right\}.$$

We endow $\mathcal{M}_n^{(d)}$ with the ℓ_∞ metric, so that $(\mathcal{M}_n^{(d)}, \|\cdot\|_\infty)$ becomes a random metric space.

Let $\eta := \sum_k \delta_{(x_k, y_k)}$ be the Poisson process on $[0, \infty) \times \tilde{c}_{00}$ with the intensity measure being the product of the Lebesgue measure and ν. Define a random closed subset of ℓ_∞

$$\mathcal{R}(T) := \Big\{ \bigtimes_{k: x_k \leq t} y_k, t \in [0, T]\Big\}.$$

Theorem 1. *Fix $T > 0$. Assume that (1) and (2) are fulfilled. Then,*

$$\left(a(n)^{-1}\mathcal{M}_{\lfloor nT\rfloor}^{(d)}, \|\cdot\|_\infty\right) \Longrightarrow (\mathcal{R}(T), \|\cdot\|_\infty) \quad \text{as } d \to \infty,$$

in distribution with respect to the topology generated by the Gromov–Hausdorff distance.

Lemma 2. *For all $x, y \in \mathbb{R}_+^d$, we have*

$$\Big|\|x \vee y - x\|_\infty - \|y\|_\infty\Big| \leq \|x \wedge y\|_\infty.$$

If $x = x_1 \vee \ldots \vee x_l$ and $y = y_1 \vee \ldots \vee y_m$ for $x_1, \ldots, x_l, y_1, \ldots, y_m \in \mathbb{R}_+^d$, then

$$\Big|\|x \vee y - x\|_\infty - \|y\|_\infty\Big| \leq \max_{j=1,\ldots,l, k=1,\ldots,m} \|x_j \wedge y_k\|_\infty.$$

Example 3. Let $X_1^{(d)} := d^{-1/\alpha}(\xi_1, \ldots, \xi_d)$, where $\xi_1, \ldots, \xi_d$ are independent copies of a random variable ξ with power tail $\mathbf{P}\{\xi > t\} \sim t^{-\alpha}$. Let $a(n) = n^{1/m\alpha}$.

Denote the k-th smallest component of $(\xi_1, \ldots, \xi_d)$ by $\xi_{(k)}$, that is $\xi_{(1)} \leq \ldots \leq \xi_{(d)}$.

For any fixed $m \geq 1$, by Section 2.2 in [1],

$$\begin{aligned} n\mathbf{P}\Big\{\xi_{(d)} \geq t_1 d^{1/\alpha} n^{1/m\alpha}, \dots, \xi_{(d-m+1)} \geq t_m d^{1/\alpha} n^{1/m\alpha}\Big\} \\ = nd(d-1)\dots(d-m+1) \int_{t_1 d^{1/\alpha} n^{1/m\alpha}}^{\infty} (1-x_1^{-\alpha})^{d-m} \alpha x_1^{-\alpha-1} dx_1 \\ \int_{t_2 d^{1/\alpha} n^{1/m\alpha}}^{\infty} \alpha x_2^{-\alpha-1} dx_2 \dots \int_{t_m d^{1/\alpha} n^{1/m\alpha}}^{\infty} \alpha x_m^{-\alpha-1} dx_m \\ \sim nd(d-1)\dots(d-m+1)(t_1 d^{1/\alpha} n^{1/m\alpha})^{-\alpha} \dots (t_m d^{1/\alpha} n^{1/m\alpha})^{-\alpha}, \end{aligned}$$

which converges to $t_1^{-\alpha} \dots t_m^{-\alpha}$ as $n, d \to \infty$. Then (1) holds with $\nu = \theta_\alpha^{\otimes m}$, where $\theta_\alpha((t,\infty)) = t^{-\alpha}$, $t > 0$.

Let ξ' be the independent copy of ξ. Then,

$$\begin{aligned} n^2 \mathbf{P}\Big\{\|X_1^{(d)} \wedge X_2^{(d)}\|_\infty \geq \varepsilon n^{1/m\alpha}\Big\} &= n^2\Big(1 - \Big(\mathbf{P}\Big\{\xi \wedge \xi' < \varepsilon n^{1/m\alpha} d^{1/\alpha}\Big\}\Big)^d\Big) \\ &= n^2\Big(1 - \big(1 - (\mathbf{P}\Big\{\xi \geq \varepsilon n^{1/m\alpha} d^{1/\alpha}\Big\})^2\big)^d\Big) \\ &\sim n^2 d (\mathbf{P}\Big\{\xi \geq \varepsilon n^{1/m\alpha} d^{1/\alpha}\Big\})^2 = n^{2-2/m} d^{-1} \varepsilon^{-2\alpha}, \end{aligned}$$

which converges to 0 as $d \to \infty$ if $m = 1$ or $n = o(d^{m/2(m-1)})$ if $m \geq 2$.

References

[1] H. A. David and H. N. Nagaraja, *Order statistics*, third edition, Wiley Series in Probability and Statistics, Wiley-Intersci., Hoboken, NJ, 2003; MR1994955

Large deviations for the volume of k-nearest neighbor balls

Takashi Owada

(joint work with Christian Hirsch, Taegyu Kang, Moritz Otto, Christoph Thäle)

The main theme of this work is to develop the large deviations theory for the point process associated with the Euclidean volume of k-nearest neighbor balls. We consider the unit cube $[0,1]^d$ equipped with the toroidal metric

$$\operatorname{dist}(x,y) = \min_{z \in \mathbb{Z}^d} \|x - y + z\|,$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. Then, the unit cube $[0,1]^d$ is considered as a flat torus with a periodic boundary. Let $B_r(x) = \big\{y \in [0,1]^d : \operatorname{dist}(x,y) \leq r\big\}$, $r > 0$, be the closed ball in $[0,1]^d$ of radius r centered at $x \in [0,1]^d$. Given a point measure ω in $[0,1]^d$ and a point $x \in [0,1]^d$, and a fixed integer $k \geq 1$, define

$$R_k(x,\omega) := \inf\big\{r > 0 : \omega\big(B_r(x) \setminus \{x\}\big) \geq k\big\}$$

to be the *k-nearest neighbor distance* of x; it gives a radius r for which $B_r(x) \setminus \{x\}$ contains exactly k points of ω with one of those points lying on the boundary of $B_r(x)$.

Given a homogeneous Poisson point process $\mathcal{P}_n$ on $[0,1]^d$ with intensity n, we are interested in the stochastic behavior of the point process

$$L_{k,n} := \sum_{X \in \mathcal{P}_n} \delta_{(X, n\kappa_d R_k(X,\mathcal{P}_n)^d - a_n)}, \tag{1}$$

where $\delta_{(x,y)}$ is the Dirac measure at $(x,y) \in [0,1]^d \times \mathbb{R}$, and κ_d is volume of the unit ball in $\mathbb{R}^d$, so that $\kappa_d R_k(X,\mathcal{P}_n)^d$ represents the volume of a k-nearest neighbor ball centered at $X \in \mathcal{P}_n$.

The primary objective of this paper is to provide comprehensive results on the asymptotics of the process (1), from the viewpoints of large deviations. We consider two distinct scenarios with respect to a divergence speed of the centering term (a_n). The first scenario examined is that

$$a_n \to \infty, \quad a_n - \log n - (k-1)\log\log n \to -\infty, \quad \text{as } n \to \infty.$$

In this case, (a_n) grows to infinity more slowly than the regime for Poisson convergence of (1). Intrinsically, there appear *infinitely many* k-nearest neighbor balls as $n \to \infty$, whose volume are approximately a_n/n up to the scale.

Define $E_0 := [0,1]^d \times (s_0, \infty]$, where s_0 is a fixed real number, and let $M_+(E_0)$ of finite non-negative Radon measures on E_0. Further, we define $b_n = n a_n^{k-1} e^{-a_n}$, $n \geq 1$. To state our LDP more precisely, we introduce the measure

$$\tau_k(du) := \frac{e^{-u}}{(k-1)!}\, 1\{u \geq s_0\} du.$$

Theorem 1. *The sequence $(L_{k,n}/b_n)_{n\geq 1}$ satisfies an LDP on $M_+(E_0)$ in the weak topology, with speed b_n and rate function $H_k(\cdot \,|\, Leb \otimes \tau_k)$, where Leb is the Lebesgue measure and*

$$H_k(\rho \,|\, Leb \otimes \tau_k) := \int_{E_0} \log\Big\{\frac{d\rho(x,u)}{d\,(Leb \otimes \tau_k)}\Big\}\rho(dx,du) - \rho(E_0) + (Leb \otimes \tau_k)(E_0).$$

In the second scenario of this work, we consider the centering term (a_n) satisfying

$$a_n - \log n - (k-1)\log\log n \to \infty, \quad a_n = o(n), \quad \text{as } n \to \infty.$$

Then, (a_n) tends to infinity more rapidly than the Poisson regime, so that the k-nearest neighbor balls centered around $\mathcal{P}_n$, whose volume are approximately a_n/n up to the scale, are even less likely to occur.

We again introduce the sequence $b_n = n a_n^{k-1} e^{-a_n}$ as before. Our objective is to investigate large deviations for the sequence $(P \circ L_{k,n}^{-1})_{n\geq 1}$ of probability distributions of $(L_{k,n})_{n\geq 1}$ on the space $M_p(E)$, where $E := [0,1]^d \times (-\infty, \infty]$. A main challenge is that the space $M_p(E)$ is not locally compact, and therefore, the vague topology would not be applicable for the convergence of such probability distributions. To overcome this difficulty, we adopt the notion of $\mathcal{M}_0$-*topology*. The main feature of $\mathcal{M}_0$-topology is that the corresponding test functions are continuous and bounded real-valued functions on $M_p(E)$ that vanish in the neighborhood of the origin. For the space $M_p(E)$, one can take the null measure $\emptyset$ as its origin. Let $B_{\emptyset,r}$ denote an open ball of radius $r > 0$ centered at $\emptyset$ in the vague

metric. Denote by $\mathcal{M}_0 = \mathcal{M}_0\big(M_p(E)\big)$ the space of Borel measures on $M_p(E)$, the restriction of which to $M_p(E) \setminus B_{\emptyset,r}$ is finite for all $r > 0$. Moreover, define $\mathcal{C}_0 = \mathcal{C}_0\big(M_p(E)\big)$ to be the space of continuous and bounded real-valued functions on $M_p(E)$ that vanish in a neighborhood of $\emptyset$. Given $\xi_n, \xi \in \mathcal{M}_0$, we say that ξ_n converges to ξ in the $\mathcal{M}_0$-topology, denoted as $\xi_n \to \xi$ in $\mathcal{M}_0$, if it holds that $\int_{M_p(E)} g(\eta)\xi_n(d\eta) \to \int_{M_p(E)} g(\eta)\xi(d\eta)$ for all $g \in \mathcal{C}_0$.

Theorem 2. *In the above setting, as $n \to \infty$,*

$$b_n^{-1} P(L_{k,n} \in \cdot) \to \xi_k, \quad \text{in } \mathcal{M}_0,$$

where

$$\xi_k(\cdot) := \frac{1}{(k-1)!} \int_E 1\{\delta_{(x,u)} \in \cdot\}\, e^{-u} dx du.$$

Subsequently, we switch our attention to the large deviations in the half-space model, which we identify by $\mathbb{H}^d$ with the product space $\mathbb{H}^d = \mathbb{R}^{d-1} \times (0, \infty)$. The Riemannian metric on $\mathbb{H}^d$ is determined by

$$ds^2 = \frac{dx_1^2 + \ldots + dx_{d-1}^2 + dy^2}{y^2}, \quad (x_1, \ldots, x_{d-1}) \in \mathbb{R}^{d-1},\ y \in (0, \infty).$$

Then, the hyperbolic volume is defined by

$$|B|_{\mathrm{hyp}} := \int_B y^{-d} dx_1 \cdots dx_{d-1} dy, \quad B \subset \mathbb{H}^d.$$

Let $\mathcal{P}$ be a Poisson point process with the hyperbolic volume measure as its intensity measure. We study large k-nearest neighbor balls whose centers are located in the restriction of $\mathcal{P}$ to the family of sampling windows $W_\lambda := [0,1]^{d-1} \times [e^{-\lambda}, \infty)$, $\lambda > 0$. Note that the hyperbolic volume of W_λ is given by $|W_\lambda|_{\mathrm{hyp}} = \int_{e^{-\lambda}}^{\infty} y^{-d}\, dy = (d-1)^{-1} e^{(d-1)\lambda}$.

Subsequently, we define the point process

$$\xi_{k,\lambda} := \sum_{x \in \mathcal{P}_{W_\lambda}} \delta_{|B_{R_k(x)}(x)|_{\mathrm{hyp}} - v_\lambda},$$

where

$$R_k(x, \mathcal{P}) := \inf\{r \geq 0 : \mathcal{P}(B_r(x)) \geq k+1\}$$

and $(v_\lambda)_{\lambda > 0}$ is a threshold sequence satisfying

$$v_\lambda - (d-1)\lambda - (k-1)\log\lambda \to -\infty, \qquad \lambda \to \infty.$$

In particular, the expected number of exceedances in the window W_λ is of order $u_\lambda := |W_\lambda|_{\mathrm{hyp}} e^{-v_\lambda} v_\lambda^{k-1}$. Let $E_0 = [s_0, \infty)$ for a fixed constant $s_0 \in \mathbb{R}$.

Theorem 3. *Let $k \geq 1$. Then, the family of random measures $(\xi_{k,\lambda}/u_\lambda)_{\lambda > 0}$ satisfies an LDP on $\mathcal{M}_+(E_0)$ with speed u_λ and rate function $H(\cdot\,|\tau_k)$, where $\tau_k(du) = e^{-u}/(k-1)!du$, $u > s_0$ and*

$$H(\rho|\tau_k) := \int_{E_0} \log \frac{d\rho}{d\tau_k}(x)\rho(dx) - \rho(E_0) + \tau_k(E_0).$$

Limit theory of sparse random geometric graphs in high dimensions

Christian Hirsch
(joint work with Gilles Bonnet, Daniel Rosen, and Daniel Willhalm)

1. General overview

In this talk, we study topological and geometric functionals of high-dimensional ℓ_∞-random geometric graph in a sparse regime, where the expected number of neighbors decays exponentially in the dimension. More precisely, we establish moment asymptotics, central limit theorems and Poisson approximation theorems for certain functionals that are additive under disjoint unions of graphs. For instance, this includes simplex counts and Betti numbers of the Rips complex, as well as general sub graph counts of the random geometric graph. We also present multiadditive extensions that cover the case of persistent Betti numbers of the Rips complex.

In the following short extended abstract, we describe very briefly the assumptions on the point process and the considered additive functionals. For the precise limit results, we refer the reader to the published work [1].

2. Point-process and graph model

We start with $\mathcal{P}_d$, which is a Poisson point process in $\mathbb{R}^d$ with some intensity λ_d^d. Later, we will let $d \to \infty$ and λ_d to 0. In other words, the intensity will decrease exponentially fast in d. Furthermore, W_d will be a cubical observation window in $\mathbb{R}^d$, which we will equip with periodic boundary conditions.

After that we construct the Rips-complex on $\mathcal{P}_d$, which is a simplicial complex. More precisely, this is the Rips-complex associated with the ℓ_∞-Gilbert graph on $\mathcal{P}_d$ at a threshold $t^{1/d}$ with $t \in [0,1]$. That is, two points x, y are connected by an edge if and only if $|x-y|_\infty \le t^{1/d}$. This Rips complex will henceforth be denoted by $\mathsf{Rips}_d(t)$. Here, we note that it is more common to consider the Gilbert graph with the Euclidean distance than with the ℓ_∞-distance. However, we found that in high dimensions, is far more convenient since then a high-dimensional distance constraint breaks down to a series of one-dimensional constraints. The sparsity assumption is illustrated in the Figure 1.

Finally, we also set $\mathsf{Rips}_d^*(t) := \bigcup_{G \in \mathsf{Comp}_d(t)} G$, where

$$\mathsf{Comp}_d(t) := \{G \subseteq \mathsf{Rips}_d(t) \colon G \text{ is a component centered in } W_d\}$$

3. Limit theory of additive functionals

In the main results of the talk, we study the asymptotic behavior of univariate nonnegative functionals on the random ℓ_∞-GG in high dimensions. That is, we investigate $A_{d,t} := a(\mathsf{Rips}_d^*(t))$ as a stochastic process on $[0,1]$ in the limit $d \to \infty$. To describe precisely the variance-scaling in our CLT, we introduce additional terminology. The key property here is additivity.

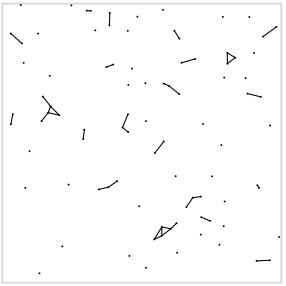


FIGURE 1. Illustration of sparse Gilbert graph

More precisely, we fix a nonnegative functional a on isomorphism classes of abstract graphs, which is *additive* in the sense that $a(R) = \sum_C a(C)$, where the sum extends over all connected components C of R. An important for such an additive functional is the triangle count illustrated in the Figure 2.

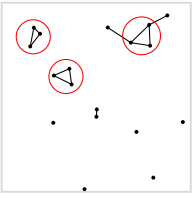


FIGURE 2. Illustration of the triangle count

Before any CLT can be established, we need to discuss the expectation scaling of the functional $A_{d,t}$. To make this precise, for $k \geq 1$ let $\mathcal{A}_k$ denote the family of connected graphs on $\{0, \dots, k\}$ with $a(G) > 0$. Then, we set $k_0 := \min\{k \geq 0 : \mathcal{A}_k \neq \emptyset\}$ and $\rho_d := |W_d| \lambda_d^{(k_0+1)d} \max_{G \in \mathcal{A}_{k_0}} v(G)^d$, where

$$v(G) := \int_{\mathbb{R}^{k_0}} 1_{\{G \subseteq \{\{i,j\} \subseteq \{0,\dots,k_0\} : |u_i - u_j| \leq 1\}\}} \mathrm{d}(u_1, \dots, u_{k_0})$$

Loosely speaking, $v(G)$ encodes the combinatorial likelihood of realizing a certain equivalence class. Then, the scaling of ρ_d can be understood as saying that in the exponentially sparse regime, only the most likely configuration determines the expectation asymptotics. We now state this observation as a theorem. Moreover, this result also contains information on the asymptotic covariances.

Theorem 1. *Assume that a is additive and that $a(G) \in e^{O(|G|)}$. Then, for $0 \leq t \leq t' \leq 1$,*

- $E[A_{d,t}] \sim \frac{\rho_d t^{k_0}}{(k_0+1)!} \sum_{G \in \mathcal{A}'_{k_0}} a(G)$
- $\mathsf{Cov}(A_{d,t'}, A_{d,t}) \sim \frac{\rho_d t^{k_0}}{((k_0+1)!)^2} \sum_{G \in \mathcal{A}'_{k_0}} a(G)^2$,

where $\mathcal{A}'_{k_0} := \{G : v(G) = \max_{G' \in \mathcal{A}_{k_0}} v(G')\}$

We now state in loose terms the main results of the talk, namely the normal and Poisson approximation for the functional $A_{d,t}$. To make this precise, we first introduce some notation. Namely,

- $r(t) := \frac{t^{k_0}}{((k_0+1)!)^2} \sum_{G \in \mathcal{A}'_{k_0}} a(G)^2$
- $(B_t)_{t \le 1}$ denotes a standard Brownian motion
- $(N_t^{(G)})_{t \le 1}$ denotes a Poisson process with intensity $K k_0((k_0+1)!)^{-1} t^{k_0-1} \mathrm{d}t$

Now, we state the announced result.

Theorem 2. *Assume that $a = a_+ - a_-$ with $a_\pm$ additive, increasing and that $a_\pm(G) \in e^{O(|G|)}$.*

(1) If $\rho_d^{1/d} \to \infty$: $(\rho_d^{-1/2}(A_{d,t} - \mathbb{E}[A_{d,t}]))_{t \le 1} \overset{fidi}{\Rightarrow} (B_{r(t)})_{t \le 1}$

(2) If $\rho_d^{1/d} \to \infty$ & $\lambda_d \rho_d^{2/(3d)} \to 0$: $(\rho_d^{-1/2}(A_{d,t} - \mathbb{E}[A_{d,t}]))_t \overset{Skorokhod}{\Longrightarrow} (B_{r(t)})_t$

(3) If $\rho_d \to K$: $(A_{d,t})_{t \le 1} \overset{Skorokhod}{\Longrightarrow} \left(\sum_{G \in \mathcal{A}'_{k_0}} N_t^{(G)} a(G) \right)_{t \le 1}$

References

[1] G. Bonnet, C. Hirsch, D. Rosen, and D. Willhalm, *Limit theory of sparse random geometric graphs in high dimensions*, Stochastic Process. Appl. **163** (2023), 203–236.

Long-Range Dependence in the Tails

Marco Oesting
(joint work with Albert Rapp, Ioan Scheffel)

The phenomenon of long-range dependence is known to have an impact on the asymptotic properties of statistical estimators in time series. In case of a stationary time series $\{X_t\}_{t \in \mathbb{Z}}$ with finite second moments $\sigma^2 = \mathbb{E}(X_0^2) < \infty$, most notions of long- and short-range dependence are defined in terms of the covariance function $\gamma(h) = \mathrm{Cov}(X_h, X_0)$. Then, short-range dependence corresponds to the case that $\gamma(h) \to 0$ sufficiently fast as $h \to \infty$, while long-range dependence means that $\gamma(h)$ decays rather slowly. For statistical estimators such as the arithmetic mean, these definitions typically result in a phase transition: The estimator converges to the quantity of interest with the same rate of convergence as in the i.i.d. case only if the time series is short-range dependent, see [6].

In extreme value analysis, where one considers the tail of the distribution, such a notion of short-/long-range dependence does not seem to be appropriate as it is based on the bulk of the distribution and requires the existence of moments. Therefore, the study of long-range dependence in extreme value analysis is still lacking a unified framework.

1. Long-Range Dependence for Max-Stable Time Series

A promising approach has recently been taken by Kulik and Spodarev [3] who define long- and short-range dependence in the following way.

Definition 1. A real-valued stationary time series $X = \{X_t\}_{t\in\mathbb{Z}}$ is called short-range dependent if

$$\sum_{t\in T} \int_{\mathbb{R}} \int_{\mathbb{R}} \Big| \mathrm{Cov}(1\{X(0) > u\}, 1\{X(t) > v\}) \Big| \, \mu(\mathrm{d}u)\, \mu(\mathrm{d}v) < \infty \tag{1}$$

for any finite measure μ on the real line. Otherwise, i.e. if there exists a finite measure μ such that the sum in inequality (1) is infinite, X is called long-range dependent.

Note that this definition does not require the existence of moments. Furthermore, it is invariant under marginal transformations. These two properties make the definition particularly attractive for extremes.

As an example, we can study this definition for stationary max-stable time series $X = \{X_t\}_{t\in\mathbb{Z}}$ with α-Fréchet margins [2]. For these processes, the upper tail dependence coefficient

$$\chi(h) = \lim_{u\to\infty} \mathbb{P}(X_h > u \mid X_0 > u), \quad h \in \mathbb{Z},$$

takes a similar role as the covariance in the classical setting. In particular, in [4], we show that the time series X is short-range dependent in the sense of Definition 1 if and only if

$$\sum_{h\in\mathbb{Z}} \chi(h) < \infty.$$

2. Consequences on Estimators for the Extremogram

In order to investigate consequences of long-range dependence for estimators, we study the extremogram [1] defined via

$$\rho_B(T) = \lim_{u\to\infty} \mathbb{P}((X_t)_{t\in T} \in uB \mid X_0 > u) \in [0, 1],$$

where $T \subset \mathbb{Z}_0$ is a finite index set and $B \subset [0,\infty]^{|T|}$ is a continuity set bounded away from 0. Here, we follow [1] and consider the ratio estimator

$$\widehat{\rho}_{B,n}(T) = \frac{\widehat{\mathbb{P}}_{n,u_n}(B)}{\widehat{\mathbb{P}}_{n,u_n}((0,\infty)^{|T|})}$$

where

$$\widehat{\mathbb{P}}_{n,u_n}(B) = \frac{1}{n} \sum\nolimits_{j=0}^{n} 1\{X_j > u_n, (X_{j+t})_{t\in T} \in u_n B\}.$$

In [5], we notice some phase transition for this estimator under short and long-range dependence, respectively. More precisely, we prove the following theorem.

Theorem 2. *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a max-stable stationary time series with α-Fréchet margins and let $\{u_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be a sequence such that $u_n \to \infty$ and $n\mathbb{P}(X_0 > u_n) \to \infty$ as $n \to \infty$. Then, we obtain the following results:*

- *If X is short-range dependent in the sense of Definition 1, then*

$$\mathrm{Var}\left[\sqrt{n\mathbb{P}(X_0 > u_n)}\widehat{\rho}_{B,n}(T)\right] \to \sigma^2 < \infty.$$

- *If X is long-range dependent in the sense of Definition 1, then*

$$\mathrm{Var}\left[\sqrt{n\mathbb{P}(X_0 > u_n)}\widehat{\mathbb{P}}_{n,u_n}((0,\infty)^{|T|})\right] = \infty.$$

In the case of short-range dependence, we can further show asymptotic normality, while the asymptotic distribution in the long-range dependent case remains an open problem. The results can also be extended from the max-stable case to the more general case of positively associated regularly varying time series if convergence to the tail process is sufficiently fast.

References

[1] R. A. Davis, and T. Mikosch, *The extremogram: A correlogram for extreme events*, Bernoulli **15**, 977–1009.

[2] L. de Haan, *A spectral representation for max-stable processes*, The Annals of Probability **12**, 1194–1204.

[3] R. Kulik, and E. Spodarev, *Long range dependence of heavy-tailed random functions*, Journal of Applied Probability **58**, 569–593.

[4] V. Makogin, M. Oesting, A. Rapp, and E.Spodarev, *Long range dependence for stable random processes*, Journal of Time Series Analysis **42**, 161–185.

[5] M. Oesting, and A. Rapp, *Long memory of max-stable time series as phase transition: asymptotic behaviour of tail dependence estimators*, Electronic Journal of Statistics **17**, 3316–3336.

[6] G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*, Springer, 2016.

Mathematics, statistics, and geometry of extremal graphical models

Sebastian Engelke

(joint work with Jevgenijs Ivanovs, Kirstin Strokorb, Jakob D. Thøstesen, Manuel Hentschel, Johan Segers, Stanislav Volgushev)

Recent methods in extreme value theory concentrate on detecting and modeling sparsity in possibly high dimensions; see [2] for a review. One branch of research is concerned with Markov structures on graphs. In this talk we present an overview over recent developments on graphical modeling of the extreme values of a random vector. We point out connections to the theory of infinitely divisible distributions and Lévy processes.

We first recall some elementary properties of max-infinitely divisible (max-i.d.) and (sum-)infinitely divisible (sum-i.d.) distributions, which are the only limits of triangular arrays of normalized, componentwise sums and maxima of independent

random vectors. A d-dimensional random vector Z with standard Fréchet margins is max-i.d. if and only if its distribution function can be written as

$$\mathbb{P}(Z \leq x) = \exp\big[-\Lambda(\mathbb{R}_+^d \setminus [0,x])\big], \qquad x \in \mathbb{R}_+^d,$$

where Λ is the exponent measure, a possibly infinite Borel measure on the space $\mathbb{R}_+^d \setminus \{0\}$ with $\Lambda(A) < \infty$ for all Borel sets A bounded away from the origin.

Similarly, for an infinitely divisible random vector Z, the characteristic function is given $\theta \in \mathbb{R}^d$ by

$$\mathbb{E}e^{\mathrm{i}\theta^\top \mathrm{Z}} = \exp\left[-\frac{1}{2}\theta^\top \Sigma\theta + \mathrm{i}\gamma^\top\theta + \int_{\mathbb{R}^\mathrm{d}\setminus\{0\}} \big(\mathrm{e}^{\mathrm{i}\theta^\top \mathrm{x}} - 1 - \mathrm{i}\theta^\top \mathrm{x}\mathrm{I}\{\|\mathrm{x}\|_2 \leq 1\}\big)\Lambda(\mathrm{dx})\right],$$

where $\gamma \in \mathbb{R}^d$ is a drift parameter and Σ is the symmetric, non-negative definite $d \times d$ covariance matrix of the Gaussian component. The Borel measure Λ on $\mathbb{R}^d \setminus \{0\}$ is the so-called *Lévy measure*, which satisfies $\Lambda(A) < \infty$ for all Borel sets A bounded away from the origin together with $\int_{\|x\|_2 \in (0,1)} \|x\|_2^2 \Lambda(\mathrm{d}x) < \infty$.

The measure Λ appears in both distributions classes and turns out to be crucial to define sparsity, conditional independence and graphical models for these models. Since Λ is typically exploding at the origin and therefore an infinite measure, classical conditional independence is not applicable. This is for instance the case when Λ is homogeneous of order -1, which correspond to the important classes of max-stable and sum-stable (or α-stable) distributions. The crucial idea of the next definition from [6] is to require this property only on test sets with product form and finite measure:

$$\mathcal{R}(\Lambda) = \left\{R = \times_{v\in V} R_v \,:\, R_v \in \mathcal{B}(\mathbb{R}),\, \Lambda(R) > 0,\, 0_V \notin \overline{R}\right\},$$

where $\mathcal{B}(\mathcal{R})$ are the Borel sets in $\mathbb{R}$.

Definition 1. For disjoint sets $A, B, C \subset V = \{1, \ldots, d\}$ that form a partition of V, we say that Λ admits *conditional independence* of A and B given C, denoted by

$$A \perp B \mid C \ [\Lambda],$$

if we have the classical conditional independence

$$Y_A \perp\!\!\!\perp Y_B \mid Y_C \qquad \text{for} \quad Y \sim \mathbb{P}_R \quad \text{for all} \quad R \in \mathcal{R}(\Lambda). \tag{1}$$

This is trivially true for A or B being empty, and for $C = \emptyset$ we say that Λ admits *independence* of A and B, and write

$$A \perp B \ [\Lambda].$$

This definition turns out to be very natural. For instance, under light assumptions, it satisfies the semi-graphoid axioms in [14] and is equivalent to the factorization of the density (if it exists) of Λ. Importantly, this Λ-conditional independence has probabilistic interpretations for both max-i.d. and sum-i.d. distributions. For the maxima case, we need to define the so-called multivariate

Pareto distribution associated with Λ by $Y \sim \mathbb{P}_Y(\cdot) = \Lambda(\cdot)/\Lambda(\mathbb{R}_+^d \setminus [0,1]^d)$. We then have the equivalence

$$A \perp B \mid C\,[\Lambda] \quad \Leftrightarrow \quad Y_A \perp_e Y_B \mid Y_C,$$

where $\perp_e$ is the extremal conditional independence introduced in [8]; see also [16] for a similar notion for extremes of Markov trees. The corresponding extremal graphical models on undirected graphs $G = (V, E)$ have recently attracted large attention and methods for graph learning [3, 7, 9, 12, 13, 17] and statistical inference have been developed [1, 11, 15]. Moreover, the Λ-conditional independence generalizes this notion and answers several discussion contributions in the article. For instance, the corresponding Λ-graphical models can be disconnected, one does not need to require densities and the max-linear models on directed acyclic graphs in [10] arise as special cases. Moreover, it is possible to model asymptotic independence; for more details see [6]. For a recent review on statistics of extremal graphical models see [4].

For the sum-i.d. case, we first note that for any Lévy measure Λ there is an associated d-dimensional Lévy process $\mathbf{X} = (X(t) : t \geq 0)$. This process can be decomposed as an independent sum $\mathbf{X} = \mathbf{W} + \mathbf{J}$, where $\mathbf{W}$ is a Brownian motion with covariance Σ and drift $\gamma \in \mathbb{R}$, and the jump process $\mathbf{J}$ is described by the Lévy measure Λ. We show in [5] that

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \quad \Leftrightarrow \quad \mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \text{ and } \mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{W}_C,$$

that is, the characterization of conditional independence of the process $\mathbf{X}$ can be separated into the corresponding statements for the Brownian and jump parts. Conditional independence of the Brownian part is easily characterized by the precision matrix Σ^{-1}. For the jump part, it turns out that

$$\mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \quad \Leftrightarrow \quad A \perp B \mid C\,[\Lambda],$$

that is, our Λ-conditional independence has a very natural interpretation for the sample paths of Lévy processes. This opens the door to a new, unexplored area of graphical models for Lévy processes.

References

[1] S. Asenova, G. Mazo, and J. Segers, *Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables*, Extremes **24** (2021), no. 3, 461–500.

[2] S. Engelke and J. Ivanovs, *Sparse structures for multivariate extremes*, Annu. Rev. Stat. Appl. **8** (2021), 241–270.

[3] S. Engelke, M. Lalancette, and S. Volgushev, *Learning extremal graphical structures in high dimensions*, 2022. Available from `https://arxiv.org/abs/2111.00840`.

[4] S. Engelke, M. Hentschel, M. Lalancette, and F. Röttger, *Graphical models for multivariate extremes*, 2024. Available from `https://arxiv.org/abs/2402.02187`.

[5] S. Engelke, J. Ivanovs, and J.D. Thøstesen, *Lévy graphical models*, In preparation (2024).

[6] S. Engelke, J. Ivanovs, and K. Strokorb, *Graphical models for infinite measures with applications to extremes*, 2022. Available from `https://arxiv.org/abs/2211.15769`.

[7] S. Engelke and S. Volgushev, *Structure learning for extremal tree models*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **84** (2022), 2055–2087.

[8] S. Engelke and A. S. Hitz, *Graphical models for extremes (with discussion)*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **82** (2020), no. 4, 871–932.
[9] S. Engelke and A. Taeb, *Extremal graphical modeling with latent variables*, 2024. Available from `https://arxiv.org/abs/2403.09604`.
[10] N. Gissibl and C. Klüppelberg, *Max-linear models on directed acyclic graphs*, Bernoulli **24** (2018), no. 4A, 2693–2720.
[11] M. Hentschel, S. Engelke, and J. Segers, *Statistical inference for Hüsler–Reiss graphical models through matrix completions*, J. Amer. Statist. Assoc., to appear (2024).
[12] S. Hu, Z. Peng, and J. Segers, *Modelling multivariate extreme value distributions via Markov trees*, 2022. Available from `https://arxiv.org/abs/2208.02627`.
[13] J. Lederer and M. Oesting, *Extremes in High Dimensions: Methods and Scalable Algorithms*, 2023. Available from `https://arxiv.org/abs/2303.04258`.
[14] S. Lauritzen, *Graphical models*, Oxford University Press, New York, **17** (1996), x+298.
[15] F. Röttger, S. Engelke, and P. Zwiernik, *Total positivity in multivariate extremes*, Ann. Statist. **51** (2023), no. 3, 962–1004.
[16] J. Segers, *One- versus multi-component regular variation and extremes of Markov trees*, Adv. in Appl. Probab. **52** (2020), no. 3, 855–878.
[17] P. Wan and C. Zhou, *Graphical lasso for extremes*, 2023. Available from `https://arxiv.org/abs/2307.15004`.

Order Selection for Clustering Multivariate Extremes

Shuyang Bai
(joint work with Shiyuan Deng, He Tang)

Multivariate extreme value theory examines the concurrent extreme values of multiple variables. After standardizing the marginal distributions of the variables, one studies the angular distribution of extreme samples (data points with the largest norms). This distribution, under the multivariate maximum domain of attraction assumption, approximates a limit distribution on the unit sphere, referred to as *spectral (or angular) measure*.

High dimensionality is a notable challenge due to the small sample sizes of extreme values. As a parsimonious model to reduce complexity, we consider a discrete spectral measure with a finite number of atoms. Several parametric models, such as heavy-tailed max-linear and sum-linear models, have this type of spectral measure. Several recent works [1–3] employed clustering algorithms on the spectral measure sphere to summarize angular distributions. The connection between discrete spectral measures and spherical clustering is evident: each spectral atom serves as a cluster center, with extreme samples forming clusters around these atoms. [2] and [3] confirmed the consistent recovery of the discrete spectral measure using such algorithms. So far, existing works have assumed a known number of atoms (or clusters), which we term the *order*. In [1–3], ad hoc methods such as elbow and scree plots were used to determine the order in real data, relying on vague visual interpretation and lacking theoretical support.

In this talk, we further explore clustering-based estimation of multivariate extreme models with a discrete spectral measure. The main contribution involves the development of an order selection method that, on the theoretical side, consistently recovers the true order, and on the practical side, enjoys intuitive and

simple implementation. Our method is based on a variant of the well-known Silhouette method [4]. In particular, we introduce an additional penalty term to the so-called simplified average silhouette width, which discourages small cluster sizes and small dissimilarities between cluster centers. The optimal order is chosen by visualizing the bending of the penalized average silhouette width curve (as a function of the order selected) when the tuning parameter of the penalty term increases. As a consequence, this method consistently estimates the order of a max-linear factor model, for which a usual information-based method is not applicable. Simulation studies demonstrate the bias-correcting effect of the penalty introduced. The method is also illustrated on a river discharge data set for stations located throughout the US. The order selected by our method matches the geographical context of these stations.

References

[1] Fomichov, V. & Ivanovs, J. Spherical clustering in detection of groups of concomitant extremes. *Biometrika*. **110**, 135-153 (2023)

[2] Janßen, A. & Wan, P. k-means clustering of extremes. *Electronic Journal Of Statistics*. **14**, 1211-1233 (2020)

[3] Medina, M., Davis, R. & Samorodnitsky, G. Spectral learning of multivariate extremes. *Journal Of Machine Learning Research*. **25**, 1-36 (2024)

[4] Rousseeuw, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal Of Computational And Applied Mathematics*. **20** pp. 53-65 (1987)

Fast amortized neural inference methods for spatial extremes in high dimensions

Raphaël Huser

(joint work with Jordan Richards, Matthew Sainsbury-Dale, Andrew Zammit-Mangion)

Important advancements in statistical deep-learning [5] have led to the development of flexible modeling approaches [7, 8], and fast and statistically-efficient inference methods for high-dimensional spatiotemporal processes [11]. These have attracted major attention recently, in part due to the ever-increasing generation of massive and high-resolution datasets (through remote sensing, climate models, high-frequency financial data, etc.), but also because the statistical models needed to analyze such datasets are becoming increasingly complex. Historically, both frequentist and Bayesian likelihood-based approaches have been developed to fit such models to big datasets and, in particular, a variety of methods based on low-rank or sparse model approximations have played a key role. However, these classical likelihood-based methods are impractical when the likelihood function is unavailable or computationally-intractable, like with popular spatial extremes models [1, 3, 6], or when large models must be fitted repeatedly in an operational or online manner. A variety of likelihood-free methods, such as Approximate Bayesian Computation (ABC; see, e.g., [10]), have been introduced to bypass this

computational bottleneck. However, these approximate inference methods are often suboptimal in practice. For example, traditional ABC relies on the choice of "suitable" summary statistics and a tolerance level used to assess whether simulated data are "close enough" to the observed data; not surprisingly, ABC can be very sensitive to these choices, especially in high dimensions. ABC has recently been combined with the Random Forests algorithm to help select appropriate summary statistics, but its performance remains heavily dependent on the tolerance level and the prior specification. More recently, neural Bayes estimators have shown an enormous potential for overcoming these limitations and enabling extremely fast and optimal inference with complex models in large-scale scientific applications including max-stable processes used to model spatially-indexed block maxima [2,4,9]. A neural estimator is simply a neural network (with data as input and parameters as output), that can be computed simply as follows:

(i) Sample parameters from a prior Ω;
(ii) Given the above parameter sets, simulate data from the model;
(iii) Choose a suitable neural network architecture to construct the neural estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}(\cdot)$;
(iv) Train the neural network (i.e., estimate $\boldsymbol{\gamma}$) to recover the true sampled parameters by minimizing a user-defined loss function L (e.g., the absolute-error or squared-error loss) averaged across all sampled parameter-data samples.

As shown in [9], the resulting neural estimator minimizes a Monte Carlo approximation of the *Bayes risk*, defined as $r_\Omega(\hat{\boldsymbol{\theta}}(\cdot)) = \int_\Theta \int_{\mathcal{S}^n} L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{Z}))p(\boldsymbol{Z} \mid \boldsymbol{\theta})\mathrm{d}\boldsymbol{Z}\mathrm{d}\Omega(\boldsymbol{\theta})$, where $\boldsymbol{Z} = (\boldsymbol{Z}_1', \dots, \boldsymbol{Z}_n')'$ are n data replicates, $\mathcal{S}$ is the sample space and Θ is the parameter space. Neural estimators can thus be easily understood through the lens of statistical decision theory, and we refer to them as *neural Bayes estimators*. Note that the above method is likelihood-free, and yet, it can accurately approximate Bayes estimators (thus—approximately—enjoying their optimality and large-sample properties). As neural networks can be evaluated very quickly on GPUs, neural Bayes estimators are thus extremely fast to evaluate (post-training), likelihood-free, and amenable to rapid bootstrap-based uncertainty quantification. Therefore, they are ideal to use with spatial extremes models in high dimensions, where estimation is often a computational bottleneck. Moreover, this inference approach is very general in the sense that it only loosely depends on the model being fitted; for example, it is not restricted to spatial models. Finally, it is *amortized* since the same neural estimator can be used repeatedly on new data *without having to retrain the neural network*, provided the new data structure conforms with the chosen neural network architecture used to construct $\hat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}(\cdot)$. This leads to significant speed-ups.

However, to model extremal peaks-over-threshold data using popular spatial extremes models such as r-Pareto processes or a variety of subasymptotic random scale mixture models, one needs to design a neural network architecture that can handle censored observations. This is indeed required as spatial extremes can sometimes include marginally non-extreme observations whose inclusion in

the inference procedure is known to bias the estimation of the extremal dependence structure. In this work, we detail how censored peaks-over-threshold data can be handled by neural Bayes estimators. This is achieved in two steps: first, after standardizing the data to an appropriate marginal scale, we mask the data that are to be censored (i.e., the low values) by replacing them with an arbitrary constant below the censoring level; second, we augment these masked data with the vector of threshold exceedance indicators, as usually done with traditional censored likelihood inference. Finally, the neural network is trained by providing these augmented and masked data as an input, arranged as a multidimensional array with two outer dimensions. The resulting estimator is shown through extensive simulations to be extremely fast to evaluate (orders of magnitude faster than state-of-the-art likelihood-based approaches), scalable to higher dimensions and different sample sizes, and applicable to varying censoring levels, while properly downweighting the influence of low observations. The method is illustrated by application to air pollution extremes over the whole Arabic peninsula where more than a hundred thousand extreme-value models are fitted locally to high values of PM2.5 data in unprecedented dimensions.

References

[1] Castruccio, S., Huser, R. and Genton, M. G. (2016) High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics* **25**(4), 1212–1229.

[2] Gerber, F. and Nychka, D. (2021) Fast covariance parameter estimation of spatial Gaussian process models using neural networks. *Stat* **10**(1), e382.

[3] Huser, R. and Wadsworth, J. L. (2022) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics* **14**, e1537.

[4] Lenzi, A., Bessac, J., Rudi, J. and Stein, M. L. (2023) Neural networks for parameter estimation in intractable models. *Computational Statistics & Data Analysis* **185**, 107762.

[5] Nalisnick, E., Smyth, P. and Tran, D. (2023) A brief tour of deep learning from a statistical perspective. *Annual Review of Statistics and Its Application* **10**, 219–246.

[6] Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**(489), 263–277.

[7] Richards, J. and Huser, R. (2022) Regression modelling of spatiotemporal extreme U.S. wildfires via partially-interpretable neural networks. arXiv preprint 2208.07581.

[8] Richards, J., Huser, R., Bevacqua, E. and Zscheischler, J. (2023a) Insights into the drivers and spatio-temporal trends of extreme Mediterranean wildfires with statistical deep-learning. *Artificial Intelligence for the Earth Systems* **2**(e220095).

[9] Sainsbury-Dale, M., Zammit-Mangion, A. and Huser, R. (2024) Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician* **78**, 1–14

[10] Sisson, S. A., Fan, Y. and Beaumont, M. (eds) (2020) *Handbook of Approximate Bayesian Computation*. Boca Raton, FL: Chapman & Hall/CRC.

[11] Wikle, C. and Zammit-Mangion, A. (2023) Statistical deep learning for spatial and spatiotemporal data. *Annual Review of Statistics and Its Application* **10**, 247–270.

Measuring risk contagion in financial networks: The effect of a Gaussian copula

Vicky Fasen-Hartmann
(joint work with Bikramjit Das)

The stability of a complex financial system may be assessed by measuring risk contagion between various financial institutions with relatively high exposure. We consider a financial network model using a bipartite graph of financial institutions (e.g., banks, investment companies, insurance firms) on one side and financial assets on the other side. Following empirical evidence, returns from such risky assets are modeled by heavy-tailed distributions, whereas their joint dependence is characterized by copula models exhibiting a variety of tail dependence behavior. In this talk we restrict our attention to the Gaussian copula, a popular model for pairwise asymptotic independence, although the results hold for general dependence structures.

In the study of extremes, the presence of *asymptotic independence* between variables in a model signifies that extreme events across multiple variables are probably less likely to occur together. Although the concept is well-understood in a bivariate context, it remains relatively unexplored when addressing the nuances of joint occurrence of extremes in higher dimensions. First we discuss the classical notion of (pairwise) asymptotic independence and then propose a notion of *mutual asymptotic independence* to capture the distinctiveness of tail behavior in dimensions larger than two. A straightforward consequence of Sibuya [4] is that the popular Gaussian copula exhibits pairwise asymptotic independence. For mutual asymptotic independence, we provide explicit conditions on the correlation matrix of the Gaussian copula characterizing mutual asymptotic independence,. Even in dimension $d = 3$ it is straightforward to construct a Gaussian copula exhibiting mutual asymptotic independence but as well to construct an example which is not mutually asymptotically independent. In Das and Fasen-Hartmann [2] these differences are illustrated using examples of various copulas not only restricting to Gaussian copulas

Next, based on Das and Fasen-Hartmann [3] we compute multivariate tail risk probabilities of the risky assets where the marginal risks are heavy-tailed and the dependence structure is a Gaussian copula. The marginal heavy-tailed risks are modeled by tail-equivalent Pareto-like tails which lead to a few interesting consequences. As the threshold increases, we note that the rate of decay of probabilities of tail sets varies depending on the type of tail sets considered and the Gaussian correlation matrix. We are able to give the explicit tail rates for rectangular sets and characterize completely the multivariate regular variation of these models on special Euclidean subcones which allow tail risk estimations for more general sets.

As measure for risk contagion, we consider CoVaR, a popular measure of risk contagion and study its asymptotic behavior under broad model assumptions, where one asset has regularly varying tail. We further propose the Extreme CoVaR Index (ECI) for capturing the strength of risk contagion between risk entities,

which is particularly useful for models exhibiting asymptotic independence. The results are illustrated by providing precise expressions when the dependence of the assets is modeled by a Gaussian copula.

Finally, we are able to give the preciese tail rates, CoVaR behavior and the ECI in a bipartite network when the dependence of the assets is modeled by a Gaussian models. Extensions to other different kind of portfolios and other dependence structures is possible as well and given in Das and Fasen-Hartmann [3].

References

[1] B. Das and V. Fasen-Hartmann, *Measuring risk contagion in financial networks* (2024) *https://arxiv.org/abs/2309.15511*.

[2] B. Das and V. Fasen-Hartmann, *On asymptotic independence in higher dimensions* (2024), *https://arxiv.org/abs/2406.19186*.

[3] B. Das and V. Fasen-Hartmann, *On heavy-tailed risks under Gaussian copula: the effects of marginal transformation*, J. Multivariate Anal. **202** (2024): Paper No. 105310, 17.

[4] M. Sibuya, *Bivariate extreme statistics*, Ann. Inst. Stat. Math. **11** (1960): 195–210.

Characterizing extremal dependence on a hyperplane

Phyllis Wan

A general framework of modeling extremes is the *peak-over-threshold* framework, in which one considers the distribution of observations conditional on being over a high threshold. In the univariate case, this framework has been well-studied and widely used. The observations exceeding a high threshold can be modelled by the class of generalized Pareto distributions, parametrized by a scale parameter and a shape parameter. This allows for straightforward statistical inference using likelihood techniques. For an overview, see e.g. [1].

Analyzing multivariate extremes requires simultaneous considerations of the marginal tails and the extremal dependence. The former can be approached by applying univariate techniques, while the latter can be separated from the former by standardizing the marginals of the data. Even so, modeling the extremal dependence remains a core problem in extreme value analysis as its structure may be complex and cannot be summarized by finite-dimensional models.

There are two common approaches in the literature to geometrically characterize the tail dependence of a random vector $\mathbf{X}$.

- **Angular measure** Θ: Standardize the marginals of $\mathbf{X}$ to the standard Pareto distribution. Then for a pre-specified norm $\|\cdot\|$, the projection of $\mathbf{X}$ onto the unit sphere $\{\mathbf{v}|\|\mathbf{v}\| = 1\}$, conditional on the norm of $\mathbf{X}$ being large, converges to

$$\left.\frac{\mathbf{X}}{r}\right|_{\|\mathbf{X}\|>r} \xrightarrow{d} R \cdot \Theta, \quad \text{as } r \to \infty, \tag{1}$$

  where Θ is a random vector on the positive unit sphere $\{\mathbf{v} \in [0,\infty)^d | \|\mathbf{v}\| = 1\}$ and R is a standard Pareto variable independent of Θ. Here Θ is called the *angular measure* or the *spectral measure*. For an overview, see e.g. [2].

- **Spectral random vector S**: Standardize the marginals of $\mathbf{X}$ to the standard Exponential distribution. Then the exceedances of $\mathbf{X}$, whose maximum component is large, converges to

$$\mathbf{X} - r \cdot \mathbf{1}|_{\max(\mathbf{X})>r} \xrightarrow{d} \mathbf{Z}, \quad \text{as } r \to \infty, \tag{2}$$

where $\mathbf{Z}$ has the stochastic representation

$$\mathbf{Z} := E \cdot \mathbf{1} + \mathbf{S},$$

such that $\mathbf{S}$ is a random vector on the irregular support $\{\mathbf{v} \in \mathbb{R}^d | \max(\mathbf{v}) = 0\}$ and E is a standard exponential random variable independent of $\mathbf{S}$. Here $\mathbf{S}$ is said to be the *spectral random vector*. See [3] and [4].

The two characterizations are connected as (2) is equivalent to (1) using the L_∞-norm. Both Θ and $\mathbf{S}$ can be used to characterize the extremal dependence structure. However, notice that the supports of Θ and $\mathbf{S}$ are nonlinear and induces intrinsic dependence between the dimensions. This poses nontrivial constraints for the construction of statistical models and their inference.

In this paper, I consider a different representation of the extremal dependence. Namely, I examine the exceedances of $\mathbf{X}$ condition on the component mean $\bar{X}$ being large and investigate whether or not

$$\mathbf{X} - r \cdot \mathbf{1}|_{\bar{X}>r} \xrightarrow{d} \mathbf{Z}^*, \quad \text{as } r \to \infty.$$

In the case where the tail of $\mathbf{X}$ has nondegenerate extremal dependence structure, the limiting distribution $\mathbf{Z}^*$ can be represented as

$$\mathbf{Z}^* := E \cdot \mathbf{1} + \mathbf{V} + \boldsymbol{\mu}_{\mathbf{V}},$$

such that

- $\mathbf{V}$ is characterized by the class of centered random vector on the hyperplane $\mathbf{1}^\perp := \{\mathbf{v} | \mathbf{v}^T \mathbf{1} = 0\}$ with the moment condition $E[e^{\max(\mathbf{V})}] < \infty$;
- $\boldsymbol{\mu}_{\mathbf{V}}$ is a location vector determined by the distribution of $\mathbf{V}$;
- E is a standard exponential random variable independent of $\mathbf{V}$.

In particular, for any positive semidefinite covariance matrix Σ that satisfies $\Sigma \cdot \mathbf{1} = \mathbf{0}$, $\mathbf{V} \sim N(\mathbf{0}, \Sigma)$ characterizes the widely used Hüsler-Reiss models [5]. This justifies the popular claim that the Hüsler-Reiss family serves as the counterpart of Gaussian family in the modeling of extremes.

The canonical spectral random vector $\mathbf{V}$ resides on a linear vector space. This translates the statistical analyses on multivariate extremes to that on a linear vector space, enabling straightforward adaptation of existing statistical techniques. As an example, I illustrate the application of principal component analysis to achieve a lower-dimensional approximation of tail observations.

References

[1] S. Coles (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag London.

[2] S.I. Resnick (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Verlag New York.

[3] H. Rootzén and N. Tajvidi (2006), *Multivariate generalized Pareto distributions*, Bernoulli, 12(5): 917–930.
[4] H. Rootzén, J. Segers and J.L. Wadsworth (2006), *Multivariate generalized Pareto distributions: Parametrizations, representations, and properties*, Journal of Multivariate Analysis, 165:117–131.
[5] J. Hüsler and R-D. Reiss (1989), *Maxima of normal random vectors: Between independence and complete dependence*, Statistics & Probability Letters, 7: 283–286.

Penalized estimation of Hüsler-Reiss precision matrices

Stanislav Volgushev
(joint work with Michaël Lalancette, Alexander Ryabchenko, Sebastian Engelke)

Hüsler-Reiss distributions form a popular class of multivariate Pareto distributions. They can be shown to arise as limits in the multivariate peak over threshold setting and are thus canonical models for the dependence structure of extremes.

As shown in [4], Hüsler-Reiss distributions can be conveniently parametrized by the so-called Hüsler-Reiss precision matrix Θ. The class of valid Hüsler-Reiss precision matrices is given by

$$\{\Theta \in \mathbb{R}^{d\times d} : \Theta = \Theta^\top, v^\top \Theta v \geq 0, \Theta 1_d = 0, \mathrm{rank}(\Theta) = d-1\},$$

where $1_d \in \mathbb{R}^d$ is the vector with a 1 in every entry.

A particularly attractive feature of the HR precision matrix Θ lies in the following fact: if Y, follows a Hüsler-Reiss distribution with parameter matrix Θ, then the zero pattern of Θ encodes the extremal conditional independence structure of Y. Here, conditional independence is understood in the sense of [1] since Y is not supported on a product space and classical conditional independence fails. This conditional independence structure can be conveniently encoded in a graph $G = (V, E)$ where the vertices $V = \{1, \ldots, d\}$ correspond to entries of Y and the absence of an edge encodes conditional independence relationships, see [5] for background on probabilistic graphical models.

Estimating Θ is challenging, especially in high dimensions. The entries of Θ cannot be expressed through moments, rather Θ is the Moore-Penrose pseudo-inverse of a certain matrix Σ with entries that correspond to certain logarithmic moments of Y. This is akin to the problem of estimating a precision matrix of a multivariate normal distribution. An additional challenge arises because we typically do not observe realizations form the Hüsler-Reiss distribution itself but rather only data whose extreme values are approximately Hüsler-Reiss.

Existing methodology for estimating Hüsler-Reiss precision matrices has focused on estimating the sparsity pattern of Θ [2], estimating the entries of Θ is the sparsity pattern is known through matrix completions [4], or does not produce valid precision matrix estimates without further pre-processing [6] because the zero row-sum and positive semi-definiteness constraints fail.

In this talk we present, for the first time, an approach to penalized estimation of Θ that automatically leads to valid Hüsler-Reiss precision matrices. The basic idea is akin to the graphical lasso [3]. However, a naive extension of the approach in [3]

is shown to fail both in simulations and through a theoretical analysis. We then propose a remedy that is based on adaptive weighting. The resulting estimator is shown to consistently estimate the correct sparsity pattern even in growing dimensions. Our theoretical analysis also extends the consistency of the matrix completion problem in [4] in growing dimensions.

References

[1] Engelke S. & Hitz A.S. (2020) *Graphical models for extremes (with discussion).* Journal of the Royal Statistical Society: Series B, 82: 871-932

[2] Engelke S., Lalancette M. & Volgushev S. *Learning extremal graphical structures in high dimensions* https://arxiv.org/abs/2111.00840

[3] Friedman, J., Hastie, T., & Tibshirani, R. (2008). *Sparse inverse covariance estimation with the graphical lasso.* Biostatistics, 9(3), 432-441.

[4] Hentschel M., Engelke S. & Segers J. (2024) *Statistical inference for Hüsler-Reiss graphical models through matrix completions.* Journal of the American Statistical Association, to appear

[5] Lauritzen, S. L. (1996). Graphical models (Vol. 17). Clarendon Press.

[6] Wan, P. & Zhou, C. (2023) *Graphical Lasso for Extremes* https://arxiv.org/abs/2307.15004

Asymptotic Theory for Neural Estimators

Manuel Hentschel

(joint work with Almut Rödder, Sebastian Engelke)

Neural estimators, as trained neural network functions, represent a powerful tool in parameter estimation. They operate by taking a sample from a parameterized distribution as input and producing an estimate of the parameters as output. Notably, they have demonstrated exceptional performance in diverse applications, such as extreme value theory and covariance estimation in spatial statistics [Lenzi et al., 2021, Gerber and Nychka, 2021, Sainsbury-Dale et al., 2023]. These examples are of particular interest due to the inapplicability of conventional techniques such as Maximum Likelihood estimation which are often too computationally expensive in practice. In this project, we aim to provide a theoretical foundation for the use of neural estimators in parameter estimation, and provide asymptotic results for their performance. To this end, we decompose the error of a neural estimator into different components, and employ results from Bayes literature [Vaart, 1998], neural network approximation theory [Cybenko, 1989], and statistical learning theory [Xu and Mannor, 2012] to show a modified notion of consistence for neural estimators.

References

[Cybenko, 1989] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

[Gerber and Nychka, 2021] Gerber, F. and Nychka, D. (2021). Fast covariance parameter estimation of spatial Gaussian process models using neural networks. *Stat*, 10(1):e382.

[Lenzi et al., 2021] Lenzi, A., Bessac, J., Rudi, J., and Stein, M. L. (2021). Neural networks for parameter estimation in intractable models.

[Sainsbury-Dale et al., 2023] Sainsbury-Dale, M., Zammit-Mangion, A., and Huser, R. (2023). Likelihood-free parameter estimation with neural bayes estimators. *The American Statistician*, pages 1–23.

[Vaart, 1998] Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[Xu and Mannor, 2012] Xu, H. and Mannor, S. (2012). Robustness and generalization. *Machine learning*, 86:391–423.

Heavy-tailed directed graphical models

Mario Krali

The interacting complex mechanisms that drive rare events present a daunting area of multivariate dependence modelling of extremes. In this work we attempt to fully describe the joint extremal dependence in high dimensions. Inspired by structural equation models [7], we model extremal dependence on a directed, but not necessarily acyclic, graphical structure $\mathcal{D} = (V, E)$, whereby every node variable is represented as

$$X_i = \bigvee_{j \in \mathrm{pa}(i)} c_{ij} X_k \vee c_{ii} Z_i, \quad i \in V, \tag{1}$$

and we refer to the model as the max-linear structural equation model (MSEM). We refer to $|V| = d$ as the dimension of the graph and write $\mathrm{pa}(i)$ and $\mathrm{an}(i)$ to denote, respectively, the parents and the ancestors of node i. We call the Z_i's innovations, and require them to be asymptotically independent [8], which marks a departure from the classical assumption of exogeneity, or independence, of the innovations — known as noise or error variables — in non-extremal statistics.

Despite the similarities between the MSEM and the recursive max-linear model (RMLM) [4], which is supported on a directed acyclic graph (DAG), the presence of cycles in the graphical structure of (1) provides a generalization over DAGs, and allows for more complex types of causal dependencies such as two-way causality.

The first problem we encounter is the representation of $\boldsymbol{X} = (X_1, \ldots, X_d)$ as a max-linear (or max-factor) model, namely,

$$X_i = \vee_{j \in \mathrm{an}(i)} a_{ij} Z_j, \quad i \in V. \tag{2}$$

The coefficients a_{ij} serve as path weights between the node variables and provide causal interpretations for the dependencies between the nodes [4]; these coefficients are arranged into the coefficient matrix $A \in \mathbb{R}_+^{d \times d}$. Conditions for the existence of such solution (via innovations) can be found in [2], and are translated to the max-times setup by [1]. However, existence of a solution does not necessarily guarantee identifiability from the matrix A.

For DAG structures the matrix A can be arranged in upper-triangular form, which allows the identification of causal relations between the node variables, and enables the association of the nodes to the respective innovation indices. Contrary to DAGs, this is in general not possible if cycles are present in directed graphs. For instance, the coefficient matrix A of the MSEM supported on the directed graph

in Figure 1 has $d^2 = 9$ non-zero entries which cannot be linked to the innovations simply by looking at the zero-pattern of the coefficient matrix A.

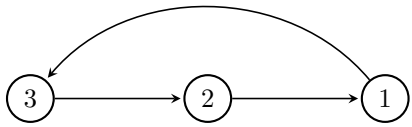


FIGURE 1. Directed graph supporting a three-dimensional MSEM.

For a vector $\boldsymbol{X}$ supported on the directed graph in Figure 1, we can write $\boldsymbol{X} = A_L \times_{\max} \boldsymbol{Z}$ and $\boldsymbol{X} = A_R \times_{\max} \boldsymbol{Z}'$ for column-permuted matrices given by

$$A_L = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{and} \quad A_R = \begin{bmatrix} a_{11} & a_{13} & a_{12} \\ a_{21} & a_{23} & a_{22} \\ a_{31} & a_{33} & a_{32} \end{bmatrix},$$

and after a simple relabelling of the innovation vector, namely, $\boldsymbol{Z} = (Z_1, Z_2, Z_3)$ and $\boldsymbol{Z}' = (Z_1, Z_3, Z_2)$. The operation $\times_{\max}$ is similar to matrix multiplication, but replaces summation with the maximum operator.

Without loss of generality we assume that the rows of A have unit, or standardized, α-th norm. Under the condition that the cyclic path weights are strictly less than one, we show that the standardized matrix A has a particular structure which helps establish identifiability.

Theorem 1. *For an MSEM $\boldsymbol{X}$ with standardized coefficient matrix A and with cyclic path weights strictly less one, it holds that $a_{ii} > a_{ji}$ for all $i, j \in V$.*

This result has important consequences in extreme value statistics, because it enables the construction of a procedure that combines the extremal dependence measure [3] and the extremal scalings [5, 6] in order to compute the coefficient matrix A, which gives a complete description of the extremal behaviour of the vector $\boldsymbol{X}$.

REFERENCES

[1] A. Amendola, C. Klüppelberg, S. Lauritzen and N. Tran *Conditional independence in max-linear Bayesian networks*, Annals of Applied Probability, **31**, (2022), 1–45.
[2] P. Butkovič *Max-linear Systems: Theory and Algorithms*, Springer, London, (2010).
[3] D. Cooley and E. Thibaud *Decompositions of dependence for high-dimensional extremes*, Biometrika, **106(3)** (2019), 587–604.
[4] N. Gissibl and C. Klüppelberg *Max-linear models on directed acyclic graphs*, Bernoulli, **24(4a)** (2018), 2693–2720.
[5] C. Klüppelberg and M. Krali *Estimating an extreme Bayesian network via scalings*, Journal of Multivariate Analysis, **181(1)** (2021).
[6] M. Krali, A. C. Davison, and C. Klüppelberg *Heavy-tailed max-linear structural equation models in networks with hidden nodes*, Arxiv Preprint, arXiv:2306.15356, (2023).

[7] J. Pearl *Causality: Models, Reasoning, and Inference*, second edition, Cambridge University Press, Cambridge, (2009).
[8] S. I. Resnick *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York, (2007).

Bootstrapping disjoint, sliding and circular block maxima

Axel Bücher
(joint work with Torben Staud)

Suppose $(X_t)_{t\in\mathbb{Z}}$ is a stationary time series such that affinely standardized maxima $a_r^{-1}\{\max(X_1,\dots,X_r)-b_r\}$ over stretches of size r weakly converge to the generalized extreme value (GEV) distribution with shape parameter γ, for $r\to\infty$. For an observed stretch of size n, the classical block maxima method is concerned with measurable functions φ_m applied to the sample of disjoint block maxima, that is, with

$$\varphi_m(\max(X_1,\dots,X_r),\max(X_{r+1},\dots,X_{2r}),\dots,\max(X_{(m-1)r+1},\dots,X_{mr})),$$

where $m=n/r$ (assumed to be an integer for simplicity) denotes the number of disjoint blocks. If $r=r_n\to\infty$, the asymptotic behavior of such expressions is typically driven by standardized empirical means $H_n=\frac{1}{\sqrt{m}}\sum_{i=1}^m\{h(Z_{i,r})-\mathbb{E}[h(Z_{i,r})]\}$, where $Z_{i,r}=a_r^{-1}\{\max(X_{(i-1)r+1},\dots,X_{ir})-b_r\}$ and where h is a measurable function. A suitable application of the central limit theorem implies asymptotic normality of H_n with asymptotic variance $\sigma^2=\mathrm{Var}(h(Z))$, where Z is GEV-distributed with cdf G, say.

As an alternative, instead of considering disjoint block maxima, one may consider measurable functions applied to the sample of sliding block maxima, that is,

$$\varphi_{n-r+1}(\max(X_1,\dots,X_r),\max(X_2,\dots,X_{r+1}),\dots,\max(X_{n-r+1},\dots,X_n));$$

note that the respective sample is of size $n-r+1$. The asymptotic behavior is then typically driven by $H'_n=\sqrt{m}\frac{1}{n-r+1}\sum_{i=1}^{n-r+1}\{h(Z'_{i,r})-\mathbb{E}[h(Z'_{i,r})]\}$, where the ith standardized sliding block maximum is given by $Z'_{i,r}=a_r^{-1}\{\max(X_i,\dots,X_{i+r-1})-b_r\}$. It has recently been found that H'_n is asymptotically normal as well with asymptotic variance given by

$$(\sigma')^2=2\int_0^1\mathrm{Cov}(h(Z_{1,\xi}),h(Z_{2,\xi}))\mathrm{d}\xi,$$

where $(Z_{1,\xi},Z_{2,\xi})$ has joint cdf $H(x,y)=G(x)^\xi G(y)^\xi G(x\wedge y)^{1-\xi}$, see [1,3], among others. Remarkably, $(\sigma')^2\le\sigma^2$ [3], with the inequality typically being strict.

Statistical inference based on disjoint or sliding block maxima typically requires suitable estimators for σ^2 or $(\sigma')^2$, respectively. Given the complicated expression for $(\sigma')^2$, this is an intricate problem, with standard plug-in methods only being feasible if explicit calculations are possible. As an alternative, one may rely on bootstrap approximations [2]. We show that naive block-bootstrap approaches are inconsistent, and provide a consistent alternative based on resampling circular

block maxima. The latter are obtained by calculating sliding maxima on small blocks of blocks (for instance, twice the block size r) after repeating the first $r-1$ observations in each such block of blocks. As a by-product, we show that the circular block maxima method has the same asymptotic variance as the sliding block maxima method.

References

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

[2] L. de Haan, C. Zhou, *Bootstrapping Extreme Value Estimators*, Journal of the American Statistical Association, **119:545** (2024), 382-393.

[3] N. Zou, S. Volgushev, A. Bücher, *Multiple block sizes and overlapping blocks for multivariate time series extremes*, Ann. Statist. **49(1)** (2012), 295–320.

Parametric multi-fidelity estimation with applications to extremes

Vladas Pipiras

(joint work with Minji Kim, Brendan Brown)

In a multi-fidelity (MF, for short) setting, data come from two sources: the high-fidelity source associated with a random variable $X^{(1)}$ and the low-fidelity one associated with $X^{(2)}$. Given the same (random) conditions, $X^{(1)}$ and $X^{(2)}$ measure the same quantity and are possibly dependent as a random vector $(X^{(1)}, X^{(2)})$. Suppose $(X_i^{(1)}, X_i^{(2)})$, $i = 1, \dots, n$, are i.i.d. observations of $(X^{(1)}, X^{(2)})$. Furthermore, a computational cost to produce $X^{(2)}$ is thought to be considerably smaller (compared to $X^{(1)}$), at the expense of precision. It is less costly to observe additional copies of $X^{(2)}$, say i.i.d. copies $X_i^{(2)}$, $i = n+1, \dots, n+m$. The MF data in question is thus

$$(X_1^{(1)}, X_1^{(2)}), \dots, (X_n^{(1)}, X_n^{(2)}), X_{n+1}^{(2)}, \dots, X_{n+m}^{(2)}. \tag{1}$$

In contrast, the baseline will consist only of the high-fidelity observations

$$X_1^{(1)}, \dots, X_n^{(1)}. \tag{2}$$

Suppose, in addition, that the distributions of $X^{(1)}$ and $X^{(2)}$ are specified parametrically: the (vector) parameter θ_j specifies the distribution of $X^{(j)}$, $j = 1, 2$, with additional (vector) parameter η, thought to capture the dependence between $X^{(1)}$ and $X^{(2)}$, combining into $(\theta_1, \theta_2, \eta)$ to specify the joint distribution of $(X^{(1)}, X^{(2)})$. An example are univariate and bivariate generalized extreme value distributions for $X^{(1)}$ and $X^{(2)}$ representing block maxima. The question addressed in this work is how the parameter θ_1 of the high-fidelity variable $X^{(1)}$ can be estimated more efficiently when using the MF data (1), compared to the baseline when only the high-fidelity data (2) are used.

Three types of MF estimators are considered for estimating θ_1 with the MF data (1): the moment MF estimator, the maximum-likelihood (ML, for short) MF estimator and the marginal maximum-likelihood MF estimator. The *moment*

estimation applies to the situation when θ_1 can be expressed in terms of the means of functions of the high-fidelity variable $X^{(1)}$, say, $\theta_1 = \mathbb{E}h(X^{(1)})$ when θ_1 is scalar and h is a (known) function. In this case, the moment estimator of θ_1 is written as

$$\widehat{\theta}_{1,mom} = \overline{(h(X^{(1)})}_n + \alpha \cdot (\overline{(h(X^{(2)})}_{n+m} - \overline{(h(X^{(2)})}_n), \tag{3}$$

where $\alpha \in \mathbb{R}$ is suitably chosen and $\overline{Y}_k = \frac{1}{k}\sum_{i=1}^{k} Y_i$. Moment-type MF estimators were considered in at least the following contexts: control variates methods (e.g. [1]), multi-fidelity methods (e.g. [2]), semi-supervised learning (e.g. [3]). The *ML estimation* is a standard parametric estimation of model parameters, including θ_1, from the MF data (1), and can be considered most efficient. In contrast to the moment estimation, the ML estimation requires a joint parametric specification and may be sensitive to misspecification of the dependence between $X^{(1)}$ and $X^{(2)}$. Placing in-between the previous two methods, the *marginal ML estimation* considers, say, in the univariate case and when the marginal parametric families are the same,

$$\widehat{\theta}_{1,mml} = (\widehat{\theta}_{1,ml})_n + \beta \cdot ((\widehat{\theta}_{2,ml})_{n+m} - (\widehat{\theta}_{2,ml})_n), \tag{4}$$

where $\beta \in \mathbb{R}$ is suitably chosen and $(\widehat{\theta}_{j,ml})_k$, $j = 1, 2$, denotes the ML estimator computed from the data $X_i^{(j)}$, $i = 1, \ldots, k$. When $X^{(1)}$ and $X^{(2)}$ are independent and no efficiency gain is expected from the low-fidelity data, (4) with $\beta = 0$ leads to the ML estimator $(\widehat{\theta}_{1,ml})_n$ based on the high-fidelity data alone. When $X^{(1)} = X^{(2)}$ are completely dependent, (4) with $\beta = 1$ leads to the ML estimator $(\widehat{\theta}_{2,ml})_{n+m}$ based on the low-fidelity (and hence high-fidelity when $X^{(1)} = X^{(2)}$) data.

Apart from introducing the three types of MF estimators above, this work compares their efficiency and that of several estimators based on the baseline data (2) for several parametric distribution families, including the bivariate Gaussian distribution, the bivariate Gumbel distribution and several forms of the bivariate Bernoulli distribution. In addition, an application is provided concerning quantification of occurrences of extreme ship motions generated by two computer codes of varying fidelity.

References

[1] B.L. Nelson, *On control variate estimators*, Computers & Operations Research **14** (1987), 219–225.

[2] B. Peherstorfer, K. Willcox and M. Gunzburger, *Optimal model management for multifidelity Monte Carlo estimation*, SIAM Journal on Scientific Computing **38** (2016), A3163–A3194.

[3] A. Zhang, L. D. Brown, T. T. Cai, *Semi-supervised inference: General theory and estimation of means*, The Annals of Statistics **47** (2019), 2538–2566.

Flexible mixture modelling for sample extremes

Anthony Davison
(joint work with Sonia Alouini)

Finite mixture models date back at least to [1] and are widely used in statistical data analysis owing to the their natural interpretation in terms of sub-populations of a larger population. The basic mixture density for a vector random variable W can be written as

$$f(w) = \sum_{k=1}^{K} \eta_k f_k(w),$$

where the component density functions f_k are interpreted in terms of distinct sub-populations that appear in non-negative proportions η_k, where $\eta_1 + \cdots + \eta_K = 1$. In applications the f_k typically depend on parameters θ_k that must be estimated, and the number of components K may be known from the context or may be unknown. In the latter case mixtures provide flexible semi-parametric models for density functions, and they have been intensively studied over many years, partly because of their importance for applications and partly because of their non-standard theoretical properties. A recent reference is [2], and earlier books include [3–5].

A key object in the statistics of multivariate extremes is the *angular distribution*, which determines the limiting dependence structure between the elements of a vector $X = (X_1, \ldots, X_D)$ of non-negative real numbers via its so-called pseudo-angles $W = (W_1, \ldots, W_D)$, where $W = X/\|X\|_1$ lies in the simplex $\mathcal{S}_{D-1} = \{(w_1, \ldots, w_D) : w_1, \ldots, w_D \geq 0, w_1 + \cdots + w_D = 1\}$. The angular distribution satisfies only the marginal constraints $D\mathrm{E}(W_d) = 1$ for each d, and thus appears to be ripe for estimation by non- or semi-parametric methods. This idea was exploited by [6], who used a reversible jump Markov chain Monte Carlo algorithm to fit a mixture of Dirichlet distributions in a Bayesian approach. Their work showed that the approach was feasible, but the algorithm was rather slow because the marginal constraints were explicitly enforced on the parameters, leading to complex Markov chain proposals. This issue was somewhat mitigated in later work [7] but remains an obstacle to widespread use of the approach in applications.

This talk described a different approach to fitting mixtures of Dirichlet distributions suggested by Theorem 2 of [8], which describes how any distribution on $\mathcal{S}_{D-1}$ that has positive means can be 'tilted' to satisfy the marginal constraints. So far as we are aware, this result has only previously been used to construct the so-called Dirichlet or Coles–Tawn model for bivariate and trivariate extremes, but it also allows the efficient fitting of mixtures of Dirichlet distributions that satisfy the marginal constraints, using a Markov chain Monte Carlo algorithm. We show that the resulting class of mixture distributions is topologically dense in the class of angular distributions, discuss approaches to hard and soft clustering of extreme events, and illustrate the power of the approach by an application to air pollution data in which the resulting clusters of pollution events have clear interpretations.

References

[1] K. Pearson (1894) *Contribution to the mathematical theory of evolution*, Philosophical Transactions of the Royal Society, series A **185**, 71–110.

[2] S. Frühwirth-Schnatter and G. Celeux and C. P. Robert (eds) (2019) *Handbook of Mixture Analysis*, Chapman & Hall/CRC: Boca Raton, FL.

[3] D. M. Titterington and A. F. M. Smith and U. E. Makov (1985) *Statistical Analysis of Finite Mixture Distributions*, Wiley: New York.

[4] B. G. Lindsay (1995) *Mixture Models: Theory, Geometry, and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Institute for Mathematical Statistics: Hayward, CA.

[5] G. J. McLachlan and D. Peel (2000) *Finite Mixture Models*, Wiley: New York.

[6] M.-O. Boldi and A. C. Davison (2007) *A mixture model for multivariate extremes*, Journal of the Royal Statistical Society, Series B (Statistical Methodology), **69**, 217–229.

[7] A. Sabourin and P. Naveau (2014) Bayesian Dirichlet mixture model for multivariate extremes: A reparametrization, *Computational Statistics and Data Analysis*, **71**, 542–567.

[8] S. Coles and J. A. Tawn (1991) *Modeling extreme multivariate events*, Journal of the Royal Statistical Society, Series B (Statistical Methodology), **53**, 377–392.

On the Clusters Over Threshold method

Olivier Wintenberger

(joint work with Gloria Buriticá)

Extremes in stationary regularly varying time series tend to occur in clusters, manifesting as short periods characterized by multiple large observations. These blocks are significant as they encapsulate critical information about the extreme behavior of the series. In this study, we delve into the analysis of cluster statistics, which are tools designed to summarize the behavior of functions acting on these extremal blocks. Among the key examples of such cluster statistics is the extremal index.

The purpose of our research is twofold. First, we establish the asymptotic normality of Clusters Over Threshold estimators based on sequences of consecutive observations with large ℓ^α-norms, where $\alpha > 0$ is the tail index. This provides a theoretical foundation for understanding the distributional properties of these estimators as the sample size increases. Second, we rigorously verify the conditions necessary for these results to hold in linear models. We prove that the asymptotic variance of traditional cluster-based index estimators is, in fact, zero, as was initially conjectured by Hsing in [1]. More generally the asymptotic variance admits an explicit expression and is simple to estimate.

To substantiate our theoretical findings, we provide an application that illustrates the practical implications of our results. We use the Clusters Over Threshold method for estimating the extremal indices of daily precipitations among different regions. Thereby we demonstrate the applicability and robustness in real-world data analysis of our method.

References

[1] T. Hsing, *On some estimates based on sample behavior near high level excursions*, Probab. Theory Related Fields **95** (1993), 331–356.

Spectral learning of multivariate extremes

Marco Avella Medina

(joint work with Richard A. Davis, Gennady Samorodnitsky)

Spectral clustering and related techniques are very popular and have found success in various applications in machine learning and statistics [1]. The central idea of spectral clustering is to use the eigenvectors of the graph Laplacian matrix constructed from an affinity graph between sample points in order to find clusters in the data. Typically these are obtained by a K-means algorithm that take these graph Laplacian eigenvectors as input. We follow this same principle but use as input to our algorithm the angular parts of the observations whose norms exceed a certain large threshold i.e., a standard spectral clustering algorithm is applied to the graph built over the angular parts of these extreme observations. Because of the nature of the extreme events that we study, we leverage tools from multivariate extreme value theory for analyzing the theoretical properties of our spectral clustering algorithm. In particular, we use multivariate regular variation as a modeling tool since it is closely connected to asymptotic characterizations of multivariate extreme value distributions [2]. The basic idea is that a d-dimensional random vector $\mathbf{X}$ is regularly varying if the distribution of the angular part $\mathbf{X}/\|\mathbf{X}\|$ stabilizes (i.e., converges in distribution) as the radial part $\|\mathbf{X}\|$ becomes large and that the radial part has Pareto-like tails. The dependence structure is then governed by the asymptotic distribution of the limiting angular part. In this work, we consider clustering the angular parts of observations with large radii.

Extremal spectral clustering algorithm. Assume that N_n observations $\mathbf{X}_i$ (with i in some set $\mathcal{V}_n$ of cardinality N_n) are in the extremal part of the sample. Associated with each $i \in \mathcal{V}_n$, is the angular component of the observation $\mathbf{X}_i/\|\mathbf{X}_i\|$. We consider mutual *$k$-Nearest Neighbor graphs* constructed as follows. A node $i_1 \in \mathcal{V}_n$ is connected to a node $i_2 \in \mathcal{V}_n$ if the point on the unit sphere corresponding to i_2 is among the k-nearest neighbors of the point corresponding to i_1, and i_1 is among the nearest neighbors of i_2. We work with weighted graphs, where we assign to the edges a weight equal to the distance between the points on the unit sphere defining the nodes. More specifically, we will take as input to our algorithm the *weighted adjacency matrix* $\mathbf{W} = [w_{i_1 i_2}]_{i_1, i_2 \in \mathcal{V}_n}$ and

$$w_{i_1 i_2} = \begin{cases} d\big(\mathbf{X}_{i_1}/\|\mathbf{X}_{i_1}\|, \mathbf{X}_{i_2}/\|\mathbf{X}_{i_2}\|\big) & \text{if } i_1 \text{ and } i_2 \text{ are connected,} \\ 0, & \text{if } i_1 \text{ and } i_2 \text{ are not connected.} \end{cases}$$

The degree of a node $i \in \mathcal{V}_n$ is defined as $d_i = \sum_{j \in \mathcal{V}_n} w_{ij}$. The degree matrix D is defined as the diagonal matrix with diagonal elements $[d_i]_{i \in \mathcal{V}_n}$ and the

normalized symmetric graph Laplacian matrix is defined as

$$L = I - D^{-1/2} W D^{-1/2},$$

where I is the identity matrix. The spectral clustering algorithm of [3] proceeds as follows:

(1) Compute the first m eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ of L (i.e., the eigenvectors corresponding to the m smallest eigenvalues of L) and define an $N_n \times m$ matrix U using these eigenvectors.
(2) Form an $N_n \times m$ matrix V by normalizing the rows of U to have unit norm.
(3) Treating each of the N_n rows of V as a vector in $\mathbb{R}^m$, cluster them into m clusters $C_1, \ldots, C_m$ using the K-means algorithm.
(4) Assign the original points $\mathbf{X}_i$ to cluster C_j if and only if row i of the matrix V was assigned to cluster C_j.

Linear factor model and main result. We now introduce the generative model that we study in this work. Let $\mathbf{X}$ be a d-dimensional random vector defined by the following linear factor model

$$\mathbf{X} = A\mathbf{Z}\,,$$

where $A = [a_{ij}]_{i=1,\ldots d; j=1,\ldots p}$ is a $d \times p$ matrix of nonnegative elements and $\mathbf{Z}$ is a p-dimensional random vector of factors consisting of independent and identically distributed random variables, that are either nonnegative or symmetric, and have asymptotically Pareto tails, i.e., $\mathbb{P}(Z_1 > z) \sim cz^{-\alpha}$, as $z \to \infty$ for some $\alpha > 0$ and $c > 0$. Note that we write $f(x) \sim g(x)$ as $x \to \infty$ to mean that $\lim_{x\to\infty} f(x)/g(x) = 1$. It is easy to show (see, for example, [4], Proposition A.1) that $\mathbf{X}$ is a multivariate regularly varying random vector satisfying

$$\lim_{x\to\infty} \mathbb{P}\left(\frac{\mathbf{X}}{\|\mathbf{X}\|} \in \cdot \mid \|\mathbf{X}\| > x, \right) \Rightarrow \Gamma(\cdot)\,,$$

where $\Rightarrow$ denotes weak convergence on the unit sphere $\mathbb{S}^{d-1}$, Γ is a discrete probability measure on $\mathbb{S}^{d-1}$ that, in the symmetric case, puts mass $\|\mathbf{a}^{(k)}\|^\alpha / w$ at $\mathbf{a}^{(k)}/\|\mathbf{a}^{(k)}\|$ for $k = 1, \ldots, p$, where $\mathbf{a}^{(k)} = (a_{1k}, a_{2k}, \ldots, a_{dk})^\top$, is the k^{th} column of A and $w = \sum_{k=1}^{p} \|\mathbf{a}^{(k)}\|^\alpha$.

Our main result is to show that with probability tending to one as $n \to \infty$, the extremal k_n-NN graph obtained from a sample drawn from the linear factor model will have exactly $m \leq p$ connected components corresponding to the m distinct asymptotic point masses of the spectral measure $\Gamma(\cdot)$ of the model. In other words, the extremal k_n-NN graph consistently identifies the underlying clusters through its connected components. This together with known results about spectral clustering (for instance Proposition 2 in [1]) show that our proposed extremal spectral clustering algorithm consistently identifies the clusters of extremes generated by the linear factor model. The main result follows from the proposition and theorem stated below. The full version of our work is [5].

We will need some additional notation in order to state our results rigorously. For $n = 1, 2, \ldots$, we define the set of indexes corresponding to extreme observations

$$\mathcal{I}_n = \big\{i = 1, \ldots, n : \|\mathbf{X}_i\| > u_n\big\},$$

and denote its cardinality by $N_n = \text{card}(\mathcal{I}_n)$. We will assume the sequence (u_n) satisfies $n^{-(\alpha+2)/(\alpha(\alpha+3))}u_n \to \infty$. Note that we may choose a further sequence (h_n) such that

$$h_n \to \infty,\ h_n = o(u_n),\ h_n = o\big(u_n^{(\alpha+1)/2}n^{-1/2}\big),\ n^{-1/\alpha}u_n h_n \to \infty \tag{1}$$

as $n \to \infty$. Indeed, the choice $h_n = u_n^{(\alpha-1)/4}n^{(2-\alpha)/(4\alpha)}$ works for this purpose.

Lemma 1. *Let (h_n) be a sequence satisfying (1) and consider the event*

$$A_n = \big\{\textit{for any } i \in \mathcal{I}_n \textit{ at most one of } Z_{im},\ m = 1, \ldots, p \textit{ exceeds } h_n\big\}.$$

Then $\mathbb{P}(A_n) \to 1$ as $n \to \infty$.

We now define the sets of indexes corresponding to extremes generated by each of the individual factors i.e. we define for $j = 1, \ldots, p$

$$\mathcal{I}_n^{(j)} = \{i = 1, \ldots, n : \|\mathbf{X}_i\| > u_n,\ Z_{ij} > u_n/a^*\},$$

where $a^* = d^{1/2}\max\{a_{mj},\ m = 1, \ldots, d;\ j = 1, \ldots, p\}$. We enumerate $\mathbf{X}_i/\|\mathbf{X}_i\|$, $i \in \mathcal{I}_n$ as $\mathbf{Y}_i$, $i = 1, \ldots, N_n$, a sample on $\mathbb{S}^{d-1}$ of random size N_n. For each $j = 1, \ldots, p$, we enumerate $\mathbf{X}_i/\|\mathbf{X}_i\|$, $i \in \mathcal{I}_n^{(j)}$ as $\mathbf{Y}_i^{(j)}$, $i = 1, \ldots, N_n^{(j)}$, a sample on $\mathbb{S}^{d-1}$ of random size $N_n^{(j)}$.

Proposition 2. *Suppose that $k_n = o\big(nu_n^{-\alpha}\big)$ as $n \to \infty$. Then there is a sequence $(B_{n,1})$ of events with $\mathbb{P}(B_{n,1}) \to 1$ as $n \to \infty$ such that, for all n large enough, on the event $B_{n,1}$, any two points $\mathbf{Y}_{i_1}^{(j_1)}$ and $\mathbf{Y}_{i_2}^{(j_2)}$, $i_1 = 1, \ldots, N_n^{(j_1)}$, $i_2 = 1, \ldots, N_n^{(j_2)}$ will belong to two different connected components of the k_n-NN graph if $\mathbf{c}_{j_1} \neq \mathbf{c}_{j_2}$.*

Theorem 3. *Assume that the generic noise variable Z is symmetric and has a probability density function f_Z such that $f_Z(z)$ is bounded away from 0 on compact intervals and bounded from above, and such that $C^{-1}z^{-(\alpha+1)} \le f_Z(z) \le Cz^{-(\alpha+1)}$, $\alpha > 1$, for all $z \ge z_0$, some $C \ge 1$. Suppose that $d = 2$. Then, if $k_n > A\log n$ with large enough $A > 0$, there is a sequence $(A_{n,2})$ of events with $\mathbb{P}(A_{n,2}) \to 1$ as $n \to \infty$ such that, for all n large enough, on the event $A_{n,2}$, any two points $\mathbf{Y}_{i_1}^{(j)}$ and $\mathbf{Y}_{i_2}^{(j)}$, $i_1 = 1, \ldots, N_n^{(j)}$, $i_2 = 1, \ldots, N_n^{(j)}$ will belong to the same connected component of the k_n-NN graph for any $j = 1, \ldots, p$.*

References

[1] U. Von Luxburg, *A tutorial on spectral clustering*, Statistics and Computing (2007), **17** p. 395–416.

[2] S. I. Resnick, *Extreme values, regular variation and point processes*, Springer, New York, 2007.

[3] A. Ng, M. Jordan and Y. Weiss, *On spectral clustering: analysis and an algorithm*, NeurIPS 2001.

[4] B. Basrak, R.A.. Davis and T. Mikosch, *A characterization of multivariate regular variation*, Annals of Applied Probability 2002.
[5] M. Avella Medina, R.A. Davis and G. Samorodnitsky, *Spectral learning of multivariate extremes*, Journal of Machine Learning Research (2024) **25**, p. 1–36.

High dimensional inference for extreme value indices

Chen Zhou
(joint work with Liujun Chen)

To simplify models on multivariate extremes, in various domain science, it is commonly assumed that all marginal distributions share the same extreme value index. This assumption is foundational to several theoretical models, for example, the multivariate regular variation model proposed and applied in [1–3]. In addition, this assumption is also adopted by spatial extremes models applied to meteorological extremes, see, e.g. [4–6]. Such a maintained assumption, the equal extreme value indices hypothesis, needs to be tested.

The classical method for testing the equal extreme value indices hypothesis is via Wald-type tests, by combining the estimates of the extreme value indices for all dimensions. These tests enjoy favorable properties when the dimensionality of the data is low, see, e.g., [7] and [8]. However, Wald-type tests exhibit unsatisfactory performance in high dimensional scenarios.

Testing the equal extreme value indices hypothesis under a high dimensional setting is therefore an important validation step before applying existing models with this maintained assumption to high dimensional data. In the field of high dimensional statistics, it is known that traditional statistical methods, originally designed in a low dimensional context, often prove inadequate when applied to high dimensional data. For instance, the existing literature on the multivariate mean tests provide new testing methods in high dimensional settings; see, e.g., [10]. We refer interested readers to [11] for a recent review of the mean test problem in high dimensional settings.

The "dimensionality curse" is more of a concern in extreme value statistics than in classical statistical problems such as the mean test. Denote the dimensionality of the data as p and the sample size of the data as n. High dimensional statistics consider situations where $p = p(n) \to \infty$ as $n \to \infty$, sometimes allowing for $p = O(\log n)$. In extreme value statistics, the effective sample size, i.e. the number of observations used for estimation, is often much lower than n. For instance, in the peak-over-threshold approach, often the top k observations are used. Theoretically, it is often required that $k := k(n)$ satisfies $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. When considering the "dimensionality curse", it is about comparing p with the effective sample size k. In other words, even if $p < n$, which is not an ultra high dimension problem in the classical sense, one may still have that $k/p \to 0$ as $n \to \infty$, making it an ultra high dimension problem in extreme value statistics.

In this paper, we propose a novel testing procedure for comparing extreme value indices in a high dimensional setting. Our testing problems are formulated

as follows. Consider identically and independently distributed (i.i.d.) observations $\boldsymbol{X}_1 = (X_1^{(1)}, \ldots, X_1^{(p)}), \ldots, \boldsymbol{X}_n = (X_n^{(1)}, \ldots, X_n^{(p)})$ drawn from a multivariate distribution function F with marginal distributions $F_1, \ldots, F_p$. For all $j = 1, \ldots, p$, assume that the distribution F_j is heavy-tailed, i.e., there exist extreme value indices $\gamma_j > 0$ such that,

$$\lim_{t\to\infty} \frac{1 - F_j(tx)}{1 - F_j(t)} = x^{-1/\gamma_j}, \quad j = 1, \ldots, p.$$

The first goal of this paper is to test the null hypothesis

$$H_0 : \gamma_j = \gamma_j^0 \quad \text{for all } j = 1, \ldots, p,$$

where $\boldsymbol{\gamma}^0 = \left(\gamma_1^0, \ldots, \gamma_p^0\right)^\top$ is a pre-specified positive vector. Additionally, we can extend our test procedure to test whether the extreme value indices are identical across p random variables, that is,

$$H_0^* : \gamma_1 = \cdots = \gamma_p.$$

Our novel testing procedure inspired by [10], with two major differences. Firstly, our analysis addresses a characteristic of the tail of marginal distributions, which differs largely from moderate level characteristics such as the mean. Secondly, our test procedure is based on estimating all marginal extreme value indices using the Hill estimator in [14]. Unlike the sample mean, this estimator involves averaging the logarithms of order statistics, which are neither independent nor identically distributed. This complexity calls for novel proofs in establishing the asymptotic theory of the test statistic in a high dimensional setting. To the best of our knowledge, there are no existing methods tailored to address the testing problem associated with hypotheses H_0 or H_0^* within a high dimensional setting, i.e., $p \to \infty$ as $n \to \infty$. The present paper reports on a first attempt with providing a few technical tools that can be used in future research.

We focus on testing the null hypothesis H_0. For $\gamma_j > 0$, an efficient estimator for the extreme value index γ_j is the Hill estimator [14]. For each dimension $j \in \{1, \ldots, p\}$, let k_j be an intermediate sequence $k_j = k_j(n)$ such that $k_j \to \infty$ and $k_j/n \to 0$ as $n \to \infty$. The Hill estimator is then defined as

$$\widehat{\gamma}_j(k_j) := \frac{1}{k_j} \sum_{i=1}^{k_j} \log \frac{X_{n-i+1,n}^{(j)}}{X_{n-k_j,n}^{(j)}}, \tag{1}$$

where $X_{1,n}^{(j)} \le \cdots \le X_{n,n}^{(j)}$ are the order statistics of $\left\{X_1^{(j)}, \ldots, X_n^{(j)}\right\}$. We introduce the test statistic

$$\boldsymbol{T}(k_1, \ldots, k_p) = \max_{1\le j\le p} k_j \left(\frac{\widehat{\gamma}_j(k_j)}{\gamma_j^0} - 1\right)^2.$$

We establish the asymptotic theory of the test statistic $\boldsymbol{T}(k_1, \ldots, k_p)$ under H_0 in the following theorem.

Theorem 1. *Assume suitable conditions regarding the marginal behavior and (weak) cross-sectional dependence. Under the null hypothesis H_0, as $n \to \infty$, for any $x \in \mathbb{R}$,*

$$\Pr(\boldsymbol{T}(k_1, \ldots, k_p) - 2\log p + \log(\log p) \leq x) \to \exp\left\{-\frac{1}{\sqrt{\pi}}\exp(-x/2)\right\}.$$

Theorem 1 demonstrates that the test statistic $\boldsymbol{T}(k_1, \ldots, k_p)$, upon appropriate normalization, converges to a Gumbel distribution, also recognized as the Type-I extreme value distribution. The limiting distribution in our theory is the same as that of the high dimensional mean test statistic in [10]. Intuitively, this follows from the fact that the test statistic $\boldsymbol{T}(k_1, \ldots, k_p)$ is a maximum of p estimation errors which are asymptotically normally distributed. Hence, obtaining the Gumbel distribution as a limit is in line with the classical extreme value theorem [15, 16], despite that the p estimation errors are neither exactly normally distributed, nor independent.

References

[1] Resnick, S. I. (2008). *Extreme values, regular variation, and point processes*. Springer Science & Business Media.

[2] Cai, J., Einmahl, J. H. J., & de Haan, L. F. M. (2011). Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics*, 39(3), 1803-1826.

[3] Mainik, G., Mitov, G., & Rüschendorf, L. (2015). Portfolio optimization for heavy-tailed assets: Extreme Risk Index vs. Markowitz. *Journal of Empirical Finance*, 32, 115-134.

[4] Buishand, T. A., de Haan, L., & Zhou, C. (2008). On Spatial Extremes: With Application to a Rainfall Problem. *The Annals of Applied Statistics*, 2(2), 624-642.

[5] Fuentes, M., Henry, J., & Reich, B. (2013). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes*, 16, 75-101.

[6] Hector, E. C., & Reich, B. J. (2023). Distributed Inference for Spatial Extremes Modeling in High Dimensions. *Journal of the American Statistical Association*, 1-12.

[7] Kinsvater, P., Fried, R., & Lilienthal, J. (2016). Regional frequency analysis based on extreme value copulas. *Environmetrics*, 27(2), 103-115.

[8] Daouia, A., Padoan, S. A., & Stupfler, G. (2024). Optimal weighted pooling for inference about the tail index and extreme quantiles. *Bernoulli*, 20(2).

[9] Kiriliouk, A., et al. & Zhou, C. (2022). Estimating the probability of multivariate "failure sets". *Journal of Multivariate Analysis*.

[10] Cai, T. T., Liu, W., & Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(2), 349-372.

[11] Huang, Y., Li, C., Li, R., & Yang, S. (2022). An overview of tests on high-dimensional means. *Journal of Multivariate Analysis*, 188, 104813.

[12] Engelke, S., & Volgushev, S. (2022). Structure learning for extremal tree models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5), 2055-2087.

[13] Wan, P., & Zhou, C. (2023). Graphical lasso for extremes. *arXiv preprint arXiv:2307.15004*.

[14] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 1163-1174.

[15] Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Proceedings of the Cambridge Philosophical Society* (Vol. 24, pp. 180-190).

[16] Gnedenko, B. V. (1943). On the distribution of the maximum of a sample. *The Annals of Mathematics*, 44(3), 423-453.

Rainfall trends in the USA: data quality and prediction skill

Holger Rootzén
(joint work with Helga Olafsdottir, David Bolin, Richard Katz)

Is climate change making individual extreme rainfall events in northeastern United States more frequent, more intense, or both? This paper discusses three aspects of this question. In [2] we focused on data quality and developed a new method which makes it possible to use high quality NOAA annual maxima series to answer the question. Alternatively, the Peaks over Threshold method could be applied directly to the daily GHCN data – but these are of lower quality, and then results could also be of lower quality.

The second aspect is the choice of statistical model. The standard method is to use goodness of fit tests, and we have done this in [2]. However an important alternative which is broadly used in climate and weather science is to use scoring rules. In [1] we constructed new scoring rules tailored to extremes and used these and existing scoring rules to evaluate the model which we use to estimate trends in extreme rainfalls in Norhteastern USA. Both our new scoring rules and existing techniques ranked the method we used to analyze the rainfall data highest. But for other problems different scoring rules may choose different models: this is e.g. the case for a data set on air particle matter concentration in the Piedmont region in Italy.

Even very good and relevant statistics is of little use if it is not communicated to scientists and the public in an understandable and usable way. This is the third – and important – topic of the talk: in the past the concepts of return levels and return periods have been standard and important tools for engineering design. But these concepts do not apply to a changing climate, whether local or global. And further, nobody except hydrologists and statisticians can understand return levels and return periods – and it is hard for them too. Hydrologists in the USA have started to understand this. But, hydrologists in Europe and statisticians from all around the world have not. In [3] we introduce the concept of Design Life Level is defined as an upper quantile (e.g. 5%) of the distribution of the maximum value of the hydrological variable (e.g. rainfall intensity) over the design life period. This concept – or related ones – should be used in hydrology and climate science.

Our overall results are that extreme daily rainfall events in the USA are becoming more frequent but that there is little evidence of increasing trends in the distribution of sizes of individual extreme daily rainfall events. The trends are strongest in the northeastern United States where for many measuring stations the frequency increase exceeds 150% for each 1°C of average temperature increase. Our aim with this work is to inform infrastructure planning, both for protection against high-impact catastrophes and for local planning of roads and sewers.

References

[1] H. Olafsdottir, H. Rootzén and D. Bolin, *Locally tail-scale invariant scoring rules for evaluation of extreme value forecasts*, Int. J. Forecasting **40** (2024), 1701–1720.

[2] H. Olafsdottir, H. Rootzén and D. Bolin, *Extreme rainfalls in the Northeastern USA become more frequent with rising temperatures, but their size distribution remains stable.*, Journal of climate **34** (2021), 8863–8877.
[3] H. Rootzén and R. Katz, *Design Life Level: Quantifying risk in changing climate* Water Resources Research **49** (2013), 5964–5972, doi:10.1002/wrcr.20425.

Reporter: Rafal Kulik

Participants

Prof. Dr. Marco Avella Medina
Department of Statistics
Columbia University
2990 Broadway
New York, NY 10027
UNITED STATES

Fabien Baeriswyl
Sorbonne Université,
Laboratoire de Probabilités, Statistique et Modélisation
&
Université de Lausanne,
Département des Opérations
4 Place Jussieu
P.O. Box 7012
75205 Paris
FRANCE

Dr. Shuyang Bai
Department of Statistics
University of Georgia
310 Herty Dr
Athens, GA 30602
UNITED STATES

Dr. Bojan Basrak
Department of Mathematics
University of Zagreb
Bijenicka 30
10000 Zagreb
CROATIA

Prof. Dr. Axel Bücher
Fakultät für Mathematik
Ruhr-Universität Bochum
44780 Bochum
GERMANY

Dr. Gloria Buritica
Research Center for Statistics
Université de Geneve
Uni Mail - 5239
Boulevard du Pont-d'Arve
1211 Genève 4
SWITZERLAND

Lucas Butsch
Karlsruher Institut für Technologie (KIT)
Institut für Stochastik
Englerstraße 2
76131 Karlsruhe
GERMANY

Dr. Dan Cooley
Department of Statistics
Colorado State University
Campus Route 1877
Fort Collins, CO 80523-1877
UNITED STATES

Prof. Dr. Richard A. Davis
Department of Statistics
Columbia University, MC 4690
1255 Amsterdam Avenue
New York, NY 10027
UNITED STATES

Prof. Dr. Anthony Davison
EPFL SB MATH-STAT
Station 8
1015 Lausanne
SWITZERLAND

Dr. Holger Drees
Department Mathematik
Universität Hamburg
Bundesstr. 55
20146 Hamburg
GERMANY

Prof. Dr. Sebastian Engelke
Research Center for Statistics
Geneva School of Economics and Management
40, Boulevard du Pont-d'Arve
1211 Genève 4
SWITZERLAND

Vicky Fasen-Hartmann
Karlsruher Institut für Technologie (KIT)
Institut für Stochastik
Englerstraße 2
76131 Karlsruhe
GERMANY

Bartosz Glowacki
Department of Mathematics and Statistics
University of Ottawa
Ottawa ON K1N 6N5
CANADA

Manuel Hentschel
Research Center for Statistics
Université de Geneve
Bd du Pont-d'Arve 40
1205 Genève
SWITZERLAND

Assoc.-Prof. Dr. Christian Hirsch
Matematisk Institut
Aarhus Universitet
Ny Munkegade 118
8000 Aarhus C
DENMARK

Dr. Raphaël Huser
King Abdullah University of Science and Technology (KAUST)
Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division
Statistics Program
Thuwal 23955-6900
SAUDI ARABIA

Dr. Anja Janßen
Fakultät für Mathematik
Otto-von-Guericke-Universität Magdeburg
Universitätsplatz 2
39106 Magdeburg
GERMANY

Bochen Jin
Mathematisches Institut
Universität Bern
Sidlerstr. 5
3012 Bern
SWITZERLAND

Prof. Dr. Claudia Klüppelberg
Lehrstuhl für Mathematische Statistik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München
GERMANY

Mario Krali
EPFL SB MATH-STAT
Station 8
1015 Lausanne
SWITZERLAND

Prof. Dr. Rafal Kulik
Department of Mathematics and Statistics
University of Ottawa
Ottawa ON K1N 6N5
CANADA

Alexandre Mansire
Université Paris Cité
Laboratoire MAP5
UMR 8145
45 Rue des Saints-Pères
75270 Paris 6
FRANCE

Prof. Dr. Thomas Mikosch
Department of Mathematics
University of Copenhagen
Universitetsparken 5
2100 København
DENMARK

Prof. Dr. Ilya Molchanov
Institut f. Mathematische Statistik
Universität Bern
Alpeneggstrasse 22
3012 Bern
SWITZERLAND

Prof. Dr. Natalia Nolde
Department of Statistics
University of British Columbia
2207 Main Mall
Vancouver BC V6T 1Z4
CANADA

Jun.-Prof. Dr. Marco Oesting
Institut für Stochastik und
Anwendungen
Universität Stuttgart
Allmandring 5b
70569 Stuttgart
GERMANY

Dr. Takashi Owada
Department of Statistics
Purdue University
West Lafayette, IN 47907-1395
UNITED STATES

Prof. Dr. Vladas Pipiras
Department of Statistics and
Operations Research
UNC – Chapel Hill
Hanes Hall
P.O. Box 3260
Chapel Hill, NC 27599-3260
UNITED STATES

Prof. Dr. Holger L. Rootzen
Dept. of Mathematical Sciences
Chalmers University of Technology
41296 Göteborg
SWEDEN

Prof. Dr. Anne Sabourin
Université Paris Cité, Laboratoire MAP5
45 rue des Saints-Pères
75006 Paris
FRANCE

Prof. Dr. Gennady Samorodnitsky
School of Operations Research and
Information Engineering
Cornell University
Rhodes Hall
Ithaca, NY 14853
UNITED STATES

Prof. Dr. Johan Segers
LIDAM/ISBA
UCLouvain
Voie du Roman Pays, 20
1348 Louvain-la-Neuve
BELGIUM

Prof. Dr. Stilian A. Stoev
Department of Statistics
University of Michigan
272 West Hall
1085 South University
Ann Arbor, MI 48109-1107
UNITED STATES

Dr. Kirstin Strokorb
School of Mathematics
Cardiff University
Senghennydd Road
Cardiff CF24 4AG
UNITED KINGDOM

Max Thannheimer
Institut für Stochastik und
Anwendungen
Universität Stuttgart
Allmandring 5b
70569 Stuttgart
GERMANY

Victor Verma
Department of Statistics
University of Michigan
1085 South University
Ann Arbor, MI 48109-1107
UNITED STATES

Stanislav Volgushev
Dept. of Statistics
University of Toronto
Toronto ON M5S 3G3
CANADA

Dr. Jennifer Wadsworth
Dept. of Mathematics and Statistics
Lancaster University
Fylde College
Lancaster LA1 4YF
UNITED KINGDOM

Dr. Phyllis Wan
Econometrisch Instituut
Erarsmus Universiteit Rotterdam
Burgemeester Oudlaan 50
3062 PA Rotterdam
NETHERLANDS

Dr. Tiandong Wang
Shanghai Center for Mathematical
Sciences
Room 533, Jiangwan Campus
Fudan University
No. 2005 Songhu Road
Shanghai Shi 200438
CHINA

Prof. Dr. Olivier Wintenberger
Sorbonne Université
Laboratoire d'Informatique de Paris 6
4 Place Jussieu
P.O. Box 169
75252 Paris
FRANCE

Prof. Dr. Chen Zhou
Econometrisch Instituut
Erarsmus Universiteit Rotterdam
Burgemeester Oudlaan 50
3062 PA Rotterdam
NETHERLANDS