

7th Conference on

Extreme Value Analysis

Probabilistic and Statistical Models and their Applications

June 27th-July 1st, 2011

Lyon, France

Abstracts

(updated June 30)



Abstracts of Plenary Talks

New Phenomena/New Algorithms for Rare Events

Peter W. GLYNN – *Stanford University, USA*
Email: glynn@stanford.edu

Abstract

In this talk, we will survey several recent developments associated with rare events. In the light-tailed setting, it has long been known that sums of Markov-dependent bounded random variables exhibit large deviations that are qualitatively similar to those arising in the iid setting. In particular, in the bounded case, the summands, conditioned on a large deviation, exhibit conspiratorial behavior in which all the summands contribute in a roughly equal way to the large deviation of the sum. However, when the summands are unbounded, we will show, via a study of the single-server queue, that the large deviations behavior can be very different. In the unbounded context, a large deviation for the sum can arise as a consequence of unusual behavior involving only a small fraction of all the summands.

In the heavy-tailed setting, we will discuss the “sharp threshold” phenomenon that arises in settings in which heavy-tailed models can be approximated by diffusions. In particular, there is frequently a sharp threshold below which the tail behavior is described by diffusion-scale analysis, and above which heavy tails are the dominant influence on tail behavior; blending the two approaches yields tail approximations that are better than either one by itself. We will also discuss several heavy-tailed simulation algorithms, in the context of queues, that illustrate the promise and challenges of building effective rare-event methods in the presence of heavy tails.

This talk discusses joint work with Jose BLANCHET, Sean MEYN, and Mariana OLVERA-CRAVIOTO.

References

- [1] J. Blanchet, P. Glynn, and J.C. Liu. *Efficient rare event simulation for heavy-tailed multiserver queues*, Technical Report, 2011
- [2] J. Blanchet, P. Glynn, and S. Meyn, *Large deviations for the empirical mean of an M/M/1 queue*, submitted for publication, 2011.
- [3] M. Olvera-Cravioto, *On the distribution of the nearly unstable AR(1) process with heavy-tails*, Advances in Applied Probability, Vol. 42, No. 1, pp. 106-136, 2011.
- [4] M. Olvera-Cravioto, J. Blanchet, and P. Glynn, *On the transition from heavy traffic to heavy tails for the M/G/1 queue: The regularly varying case*, Annals of Applied Probability, Vol. 21, No. 2, pp. 645-668, 2011.
- [5] M. Olvera-Cravioto and P. Glynn, *Uniform approximations for the M/G/1 queue with subexponential processing times*, to appear in Queueing Systems, 2011.

Everything is getting worse... Or not?

Geurt JONGBLOED – *TU Delft, Netherlands*
Email: G.Jongbloed@tudelft.nl

Abstract

In reliability theory and survival analysis, the hazard rate h is a natural function to characterize the distribution of a nonnegative random variable X with probability density function f . If X describes the time a typical electrical component from a batch can be effectively used, monotonicity properties of h indicate whether the components in the batch become more / less reliable during a time period in which these are used (decreasing / increasing h on time interval).

Let X_1, X_2, \dots be i.i.d. random variables with hazard rate h . In this presentation, the problem of testing the null hypothesis that h is increasing on a certain time interval will be considered, based on X_1, \dots, X_n . An L_1 -type test statistic based on an L_2 -projection estimator will be introduced. The asymptotic distribution of this statistic will be given, as well as an outline of its derivation. Moreover, a practical bootstrap-based procedure will be described and simulation results will be presented, comparing the procedure with natural competitors introduced in the literature.

Random Convex Hulls and Extreme Value Statistics

Satya MAJUMDAR – *Université Paris-Sud, France*
Email: majumdar@lptms.u-psud.fr

Abstract

Convex hull of a set of points in two dimension roughly describes the shape of the set. In this talk, I will discuss the statistical properties of the convex hull of a set of N independent planar Brownian paths. We compute exactly the mean perimeter and the mean area of this convex hull, both for open and closed paths. We show that the area and perimeter grows extremely slowly (logarithmically) with increasing population size N . This slow growth is a consequence of extreme value statistics and has interesting implication in ecological context in estimating the home range of a herd of animals with population size N .

Anthropogenic Influence on Long Return Period Daily Temperature and Precipitation Extremes

Francis ZWIERS – *Pacific Climate Impacts Consortium, University of Victoria, Canada*
Email: fwzwiwers@uvic.ca

Abstract

There is now a well established approach to detecting and attributing the causes of observed changes in mean climatic conditions that has been applied progressively from global scales to regional scales to temperature, precipitation and other climate variables. While this research has provided a great deal of useful information about the causes of climate change observed during the past century or more, policy makers and others have also been demanding answers about whether there are attributable changes in the frequency and/or intensity of extreme weather and climate events. Climatologists have made some initial attempts to respond to these questions, but much remains to be done. This talk will describe a standard technique that is used in climate change detection and attribution research and will describe several parallel approaches that have been proposed to assess whether there is a detectable human influence in the far tails of the distribution of climate variables such as daily maximum air temperature or daily precipitation amount. We also describe initial applications of these approaches, and discuss limitations and further areas of improvements. These applications suggest that human influence on the climate system has affected the extremes of daily maximum and minimum temperatures, and extreme daily precipitation amounts, altering the waiting times for events of a fixed amplitude. For example, on a global scale, waiting times for extreme annual minimum daily minimum and daily maximum temperatures events that were expected to recur once every 20 years in the 1960s are now estimated to exceed 35 and 30 years respectively. In contrast, waiting times for circa 1960s 20-year extremes of annual maximum daily minimum and daily maximum temperatures are estimated to have decreased to less than 10 and 15 years respectively.

Abstracts of Invited Talks

New models for multivariate extremes: a construction principle and related properties

F. BALLANI – *Technische Universität Bergakademie Freiberg, Germany*
Email: ballani@math.tu-freiberg.de
Coauthor: Martin SCHLATHER

Abstract

We present a construction principle for the spectral density of a multivariate extreme value distribution. Although a few parametric models already exist for the bivariate case, the number of flexible parametric models in higher dimensions is relatively small.

We generalize the pairwise beta model introduced in the literature recently [1] in several steps. The main idea is to construct new (parametric) models by a suitable composition of lower-dimensional spectral densities. Furthermore, we discuss some related properties of this new class of models, for instance the possible range of the pairwise extremal coefficients, and apply it to a wind speed data set.

References

- [1] D. Cooley, R. A. Davis and P. Naveau, The pairwise beta distribution: A flexible parametric multivariate model for extremes, *J. Multivariate Anal.*, **101**(9), 2103–2117, 2010.

Rare Event Analysis of Heavy-tailed Systems with Multiple Jumps

Jose H. BLANCHET – *Columbia University, USA*
Email:jose.blanchet@columbia.edu

Abstract

The extreme behavior of stochastic systems is often caused by a single jump, specially when thinking about one dimensional systems. Nevertheless, in the context of multidimensional processes, it is also often the case that one is interested in large deviations behavior that is caused by more than one jump. For instance, consider a network of individual sub-systems whose output feeds as input to other sub-systems in the network. In this situation, a rare event such as the collapse of a specific individual sub-system might be most likely caused by the collapse of other sub-systems that feed into the specific individual subsystem in consideration. Large deviations still occur as a consequence of big jumps, but now these jumps occur in a specific order and thus there is "cooperative" behavior that leads to the occurrence of a rare event in question.

This talk focuses on a systematic approach to compute large deviations asymptotics in these types of multiple-jump situations. Our method involves two main elements: a) Reduction to an associated fluid system with jumps, and b) the application of the contraction principle for large deviations of light-tailed systems. We illustrate the approach in the analysis of multi-server queues and stochastic fluid networks.

A comparative simulation study of the annual maxima and the peaks-over-threshold methods

Sofia CAIRES – *Deltares - Hydraulic Engineering, Netherlands*
Email: Sofia.Caires@deltares.nl

Abstract

Two approaches are usually considered for the estimation of extreme values of environmental data: the Annual Maxima / Generalized Extreme Value distribution (AM/GEV) approach, and the Peaks-over-Threshold / Generalized Pareto Distribution (POT/GPD) approach. In general, and especially in view of the typically short (covering a few decades) time series available of environmental data, the POT/GPD approach, thanks to the

possibility of choosing a threshold that is exceeded on average by more than one peak per year, is expected to yield more accurate estimates of parameters and exceedance probabilities as it is based on more observations. Although the choice of the threshold in the POT/GPD approach represents an opportunity to increase the number of observations used to estimate the parameters and hence to reduce the errors of the estimates, it is often seen by some practitioners as a problem and source of subjectivity. Consequently, because of its comparative straightforwardness the AM/GEV approach often gets the better of the POT/GEV approach and the latter is not even considered.

In order to exhibit the relative merits of the AM/GEV and POT/GPD approaches in for the estimation of extreme values of environmental data, we have compared them in terms of their accuracy in estimating exceedance probabilities on the basis of time series with various lengths and, most importantly, with characteristics that mimic those of real time series. The aim was to provide a global assessment and comparison of the two approaches on the basis of small to large time series of simulated data possessing realistic features such as non-stationarity and serial dependence. It is hoped that our results will help practitioners to choose between the two methods, for instance on the basis of the length of the time series and of the expected underlying shape parameter of the data available to them.

Three types of simulation studies were carried out taking into account the characteristics of the time series of environmental data currently available. The first study focused on the finite-sample properties of the Maximum Likelihood and Probability Weighted Moments estimators of the GEV and the GPD models, based on independent and identically distributed observations. The second study focused on the finite-sample properties of the AM/GEV and the POT/GPD approaches applied to non-stationary and serially dependent observations. The third study, in which non-stationary and serially dependent observations were also simulated, assessed the relative merits of a formal procedure and of the ad hoc procedure used in the second study for the choice of the threshold in a Peaks-over-Threshold analysis.

Non-stationary frequency analysis of heavy rainfall events with climatic covariates in the french mediterranean region

Julie CARREAU – *Université Montpellier 2, France*

Email: julie.carreau@univ-montp2.fr

Co-authors: Yves TRAMBLAY, Luc NEPPEL and Kenza NAJIB

Abstract

Heavy rainfall events are prevalent in the French Mediterranean region during the fall. To characterize these events, we developed a non-stationary Peaks-over-Threshold (PoT) model with climatic covariates. An event is defined as rainfall occurring in the fall which exceeds 100 mm. The first step of the analysis consists in selecting potential covariates from correlation analysis. The covariates considered include a synoptic pattern classification, temperature, pressure and geopotential heights from NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis data and sea surface temperature (SST) from the NOAA (National Oceanic and Atmospheric Administration) database. The correlation analysis reveals that the occurrence of heavy rainfall events is linked to 1) the occurrence of a specific southern-circulation synoptic patterns and 2) the mean SST in the Western Mediterranean Sea. On the other hand, the magnitude of heavy rainfall events is correlated with air temperature which indicates higher rainfall amounts during warmer months. In the second step of the analysis, several non-stationary PoT models are compared. Each model has two parts: a Poisson model for the occurrence of heavy rainfall events and a Generalized Pareto distribution (GPD) for the amount. The rate parameter of the Poisson and the scale parameter of the GPD are allowed to vary log-linearly with the covariates. Based on the deviance test for nested models, some non-stationary PoT models are shown to perform better than the more conventional stationary model. Non-stationary PoT models, thanks to the climatic covariates which can be computed for a future period by global or regional climate models, can serve to evaluate the impact of climate change on heavy rainfall in this region.

Multivariate, multimodal extremes for novelty detection

David CLIFTON – *Institute of Biomedical Engineering, University of Oxford, UK*

Email: davidc@robots.ox.ac.uk

Abstract

Novelty detection, or one-class classification, classifies data with respect to a model of normality, such that data appearing significantly different with respect to that model are classified as being "novel". Such an approach is particularly suitable to the analysis of data from complex systems such as jet engines or human vital-signs. We describe extensions of EVT for use in describing the tails of multivariate, multimodal probability distributions, as are typically required for modelling complex engineering applications. We provide examples from jet engine monitoring, where our techniques have been used to monitor the engines of the Eurofighter, the Airbus A380, and the Boeing "Dreamliner", and examples from human vital-sign monitoring, where probabilistic patient monitoring systems have been installed in hospitals in the UK and USA.

Policy Impacts of Fat Tails

Roger COOKE – *Ressources for the Future - Delft University of Technology, Netherlands*

Email: Cooke@rff.org

Co-author: Carolyn KOUSKY

Abstract

Insurers, investors and policy officials have yet to realize the implications of fat tailed distributions. For example, the historical average may have infinite variance, or may not even exist, the probability of the current record being doubly exceeded by the next record may be very large, and aggregation may not produce normal distributions with vanishing variance. Policy implications of these facts are discussed.

A Model for Extremes on a Regular Spatial Lattice

Dan COOLEY – *Department of Statistics, Colorado State University, USA*

Email: cooleyd@stat.colostate.edu

Abstract

We propose a model which can be used to characterize extremes on a regular spatial lattice. Analogous to lattice models from spatial statistics, the proposed model creates an overall model composed from many smaller models. The spatial domain is covered by a number of small and overlapping subregions and data on these subregions is modeled with parametric multivariate extreme value models of low dimension. We show that these subregion models can be combined in such a way that the angular measure of the overall model meets the requirements of an extreme value distribution.

The model is designed specifically for threshold exceedance data and is used to model data, such as extreme precipitation which at any given time is likely to be extreme only in limited areas of the study region.

Modelling of extreme rainfall in space and time

Anthony DAVISON – *Ecole Polytechnique Fédérale de Lausanne, Switzerland*

Email: Anthony.Davison@epfl.ch

Co-author: Raphaël HUSER

Abstract

This talk will describe the space-time modelling of extreme hourly rainfall based on max-stable processes fitted to data at a number of spatial locations. The ingredients are large amounts of data, random set modelling and max-stable processes fitted using composite threshold likelihoods. Though essentially a conceptual model, the approach seems to work surprisingly well.

Risk Aggregation (A geometric calculation for rare event probabilities)

Paul EMBRECHTS – *ETH Zürich, Switzerland*

Email: paul.embrechts@math.ethz.ch

Abstract

Both in Solvency 2 as well as Basel III, the aggregation of risk across different lines of business, risk classes or stress scenarios is a crucial task with considerable impact on the resulting solvency numbers. In this talk I will review some of the key underlying mathematical issues. In particular, geometrically based algorithms (AEP, GAEP) for the calculation of exceedance probabilities for sums (AEP) and more general functionals (GAEP) of d dependent random variables will be produced. A careful mathematical analysis of the convergence properties of the algorithms will be discussed.

Spatial modeling of peak wind speed using max-stable processes

Petra FRIEDERICHS – *Meteorological Institute, University of Bonn, Germany*
Email: pfried@uni-bonn.de

Abstract

Gust warnings are one of the most frequent type of weather warnings issued by national weather services. It is thus of great importance to improve the quality of extreme wind warnings. Numerical weather forecast models provide information about the general conditions in which extreme peak wind speeds occur. However, the strength of such extremes is determined by small scale processes that are not resolved even by high resolution weather forecast models. For a reliable probabilistic analysis it is thus necessary to statistically model the extremal process in space and time.

The study aims at providing area-wide analyses of the occurrence probability of extreme peak wind speed given a relatively sparse observational network and numerical weather prediction model analyses of mean wind speed. This work is done within the framework of extreme value theory: the response surfaces of the generalized extreme value distribution parameters are modeled, and the concept of max-stable spatial processes is employed in order to account for the spatial dependencies that are not explained by the conditioning variables. The ability of max-stable processes to represent peak wind speed observations is assessed.

Copula structure analysis based on extreme dependence

Stephan HAUG – *Technische Universität München*
Email: haug@ma.tum.de
Co-authors: Claudia KLÜPPELBERG and Gabriel KUHN

Abstract

Classical correlation structure analysis is mainly based on linear modelling. We replace the linearity assumption by a copula model. We assume an elliptical copula for the dependence structure, an approach, which preserves a correlation-like structure, but allows for different marginal distributions and the non-existence of second moments. The estimation of the dependence is then possible via Kendall's tau (in a robust way), as introduced in [1], or via the tail dependence function (focussing on extreme dependence). In this talk we will introduce the latter concept.

References

- [1] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society Series B*, 71: 737–753, 2009.

Event sets for flood catastrophe models

Caroline KEEF – *JBA Consulting, United Kingdom*
Email: caroline.keef@jbaconsulting.co.uk
Co-authors: Jonathan A. TAWN and Rob LAMB

Abstract

Flooding is a natural phenomenon that can cause havoc for the people affected. In dealing with the aftermath of floods two main types of organisation are involved. The first is governmental bodies such as central government, councils and emergency services who manage flood defences and respond to flood events. The second are insurance and reinsurance companies who provide cover for the financial losses associated with floods.

Knowing the maximum damage due to flooding that can be expected to occur in a year is useful for both these organisations. In particular insurance and reinsurance companies need to know how much money they need in order to be able to pay the largest losses, which may be related to floods that occur at the same time, or close in time, over a large geographical area. To help solve this problem the insurance industry uses catastrophe models. A fundamental part of a flood catastrophe model is a flood event set. A flood event set consists of a set of simulated events that represent a large number of years of data.

In this talk we will describe how an event set can be generated from an extreme value analysis of river flow data. We will also discuss some of the issues faced when creating a flood event set starting with real flow data. In particular the complications caused by differing levels of dependence in different geographical regions.

Improved modelling of extreme pesticide residues for probabilistic dietary exposure modelling

Marc KENNEDY – *The Food and Environment Research Agency, York, United Kingdom*

Email: marc.kennedy@fera.gsi.gov.uk

Co-authors: Vicki ROELOFS and Clive ANDERSON

Abstract

Pesticides are widely used in crop treatment, and it is important that any residues from them in the food chain can reach the consumer in at most negligible amounts. In the development of new plant protection products, regulatory authorities are required to carry out risk assessments based on small samples of pesticide residue data submitted by the pesticide developer. Software for probabilistic modelling of dietary exposure to residues is designed to produce realistic or conservative estimates of the distribution of exposures within a population. Consumption patterns are derived from national surveys that are often very detailed. By contrast, algorithms for simulating residues to assign to individual consumed items rely on strong assumptions (such as assuming log-normal residue distributions) and are based on sparse data sets of composite samples rather than individual units. Individual residues are more variable than composite values. Furthermore, the small samples available for new pesticides provide very limited information about the true shape of the residue distribution, especially in the upper tail.

We describe a new Bayesian procedure which pools information from multiple datasets on extreme individual item residues to learn about the range of upper tail shapes seen in reality and also to take account of uncertainty. We have been working with Syngenta and UK, European and US regulators, together with software developers to explore options for using these and other results in future probabilistic dietary risk models.

Validation of the algorithms is a major focus. By using a selection of synthetic and real validation datasets we will show the relative accuracy of alternative methods.

Second order tail asymptotics for the sum of dependent, subexponential random variables

Dominik KORTSCHAK – *Université de Lausanne, Switzerland*

Email: Dominik.Kortschak@unil.ch

(Supported by the Swiss National Science Foundation Project 200021-124635/1)

Abstract

For the sum of independent, subexponential random variables, second order asymptotic results for the tail are well studied in the literature. In the case that a mean exists, a heuristic interpretation of the second order approximation is that the sum is large if one summand is large and the others summands behave in a normal way, i.e. $n\mathbb{P}(X_1 > u - (n-1)\mathbb{E}(X_1))$ is asymptotically a better approximation than the first order asymptotic $n\mathbb{P}(X_1 > u)$. In this talk we provide conditions under which similar results hold for dependent random variables.

Vine Regression for Predicting the effects of Climate Change using Geo-Economic data

Dorota KUROWICKA – *Delft University of Technology, Netherlands*
Email: D.Kurowicka@tudelft.nl
Co-author: Roger COOKE

Abstract

The Yale G-Econ data base relates geographic and economic data, and is invaluable for predicting the effects of climate change. However, the data is extremely complex and ill suited for standard regression modeling, as the empirical copulae are not remotely normal. A vine-copula representation is used to enable predictions of economic impacts of climate change.

Dependence Measures for joint fat tailed Distributions

Jolanta MISIEWICZ – *University of Warsaw, Poland*
Email: J.Misiewicz@mini.pw.edu.pl

Abstract

When the variance is infinite in one or more variables traditional methods of measuring dependence fail, and rank dependence may be unrelated to the physical units in which the variables are measured. New methods for characterizing dependence in these cases are developed.

An entropy based approach to detect changes in climate extremes

Philippe NAVEAU – *Laboratoire des Sciences du Climat et l'Environnement (LSCE) CNRS, France*
Email: naveau@lsce.ipsl.fr
Co-authors: Théo RIETSCH, Armelle GUILLOU and James MERLEAU

Abstract

The Kullback-Leibler information (Kullback, 1968) is defined as

$$I(f; g) = E_f \left\{ \log \left(\frac{f(\mathbf{Z})}{g(\mathbf{Z})} \right) \right\}, \quad (1)$$

and it measures the entropy distance (Robert, 2001) between the probability densities f and g for a random variable \mathbf{Z} . Kullback (1968) also refers to this quantity as the directed divergence to distinguish it from the divergence given by

$$J(f; g) = I(f; g) + I(g; f), \quad (2)$$

which is a symmetrical measure relative to f and g . This notion has been exhaustively used and studied in many research fields. Here we explore this concept within the framework of climatology and Extreme Value Theory (EVT). While the divergence (2) is expressed in function of densities, it is more convenient to work the tail distribution $1 - F$ when analyzing large excesses. In this talk, we propose an approximation to bypass the need of computing densities. This allows us to derive and study new estimators of the entropy. We apply this approach to the important problem of detecting changes in our warming climate.

References

- [Kullback, 1968] Kullback, S. (1968). *Information Theory and Statistics*, 2nd ed. New York: Dover.
- [Resnick, 2007] Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Operations Research and Financial Engineering, Springer, New York.
- [Robert, 2001] Robert, C. P. (2001). *The Bayesian Choice*. New York: Springer-Verlag.

Threshold modelling of spatially-dependent non-stationary extremes

Paul NORTHORP – *University College London, UK*
Email: paul@stats.ucl.ac.uk

Abstract

It is common for environmental extremes to exhibit non-stationary behaviour. We consider the use of quantile regression to set a covariate-dependent threshold for extreme value modelling. As an example we present the marginal modelling of wave heights, for which adjustment for spatial dependence is necessary.

A finite-dimensional construction of a max-stable process for spatial extremes

Brian J. REICH – *Department of Statistics, North Carolina State University, USA*
Email: brian_reich@ncsu.edu
Co-author: Benjamin A. SHABY

Abstract

From heat waves to hurricanes, often the environmental processes that are the most critical to understand probabilistically are extreme events. Such extremal processes manifestly exhibit spatial dependence. Max-stable processes are a class of asymptotically-justi

ed models that are capable of representing spatial dependence among extreme values. While these models satisfy modeling requirements, they are limited in their utility because their corresponding joint likelihoods are unknown for more than a trivial number of spatial locations, preventing, in particular, Bayesian analyses. In this paper we propose an approximation to the Gaussian extreme value process (GEVP) that, critically, is amenable to standard MCMC and inclusion in hierarchical models. We show that this model is max-stable and approximates the GEVP arbitrarily well. The proposed model also leads to a non-stationary extension, which we use to analyze the yearly maximum temperature in the southeast US for years 1983-2007.

Nonparametric Bayesian Inference on Bivariate Extremes

Johan SEGERS – *Université catholique de Louvain, Belgium*
Email: johan.segers@uclouvain.be
Co-authors: Simon GUILLOTTE and François PERRON

Abstract

The tail of a bivariate distribution function in the domain of attraction of a bivariate extreme-value distribution may be approximated by the one of its extreme-value attractor. The extreme-value attractor has margins that belong to a three-parameter family and a dependence structure which is characterised by a spectral measure, that is a probability measure on the unit interval with mean equal to one half. As an alternative to parametric modelling of the spectral measure, we propose an infinite-dimensional model which is at the same time manageable and still dense within the class of spectral measures. Inference is done in a Bayesian framework, using the censored-likelihood approach. In particular, we construct a prior distribution on the class of spectral measures and develop a trans-dimensional Markov chain Monte Carlo algorithm for numerical computations. The method provides a bivariate predictive density which can be used for predicting the extreme outcomes of the bivariate distribution. In a practical perspective, this is useful for computing rare event probabilities and extreme conditional quantiles. The methodology is validated by simulations and applied to a data-set of Danish fire insurance claims.

References

- [1] S. Guillotte, F. Perron and J. Segers, Nonparametric bayesian inference on bivariate extremes, *JRSS B*, to appear, 2011.

Extreme Value Theory and Single-Event Attribution in Climatology

Richard L. SMITH – *University of North Carolina, Chapel Hill, USA*

Email: rls@email.unc.edu

Co-author: Michael WEHNER

Abstract

Recently, climate scientists have become very interested in the problem of how to attribute a single extreme event, such as a heatwave or hurricane, to different forcing factors in the climate. At a crude level, the question is whether an event such as Hurricane Katrina of 2005 or the European heatwave of 2003 is in some sense “due to” human-induced climate change as opposed to natural variation in the earth’s climate. Following a much-cited paper by Stott, Stone and Allen (*Nature*, 2004), we focus on the June-July-August mean temperature for a region of western Europe corresponding to the European heatwave of 2003. A time series of observational data is calculated using the University of East Anglia land-only variance-adjusted temperature anomalies, and this is compared with two sets of climate model calculations of the same quantity: one from control runs, the other from twentieth century runs containing all the known climate forcings. By fitting an extreme value model to the observational data in combination with either the control or the twentieth century model runs, we are able to derive point and interval estimates of the probability of the observed event under either a control or forced climate scenario, and hence to derive point and interval estimates of the risk ratio or, equivalently, the fraction of attributable risk. The analysis is intended to point towards a general methodology for making calculations of this nature.

References

- [1] P. Stott, D. Stone and M. Allen, Human contribution to the European heatwave of 2003, *Nature*, 432, pp.610-614, 2004.

Decomposability for max–stable processes

Stilian A. STOEV – *University of Michigan Ann Arbor, USA*

Email: sstoev@umich.edu.

Coauthors: Yizao WANG and Parthanil ROY

Abstract

Consider the decomposition

$$\{X_t\}_{t \in T} \stackrel{d}{=} \{X_t^{(1)} \vee \cdots \vee X_t^{(n)}\}_{t \in T}, \quad (n \geq 2)$$

where $X^{(k)} = \{X_t^{(k)}\}_{t \in T}$, $k = 1, \dots, n$ are independent processes and $\stackrel{d}{=}$ denotes equality in law. We refer to the $X^{(k)}$ ’s as to (max–)components of X .

We characterize all possible max–stable components $X^{(k)}$ ’s of a max–stable process X in terms of its spectral representation. In particular, we characterize the independent stationary max–stable components of stationary max–stable processes. One simple consequence of our results is that all stationary components of the moving maxima processes are trivial. As a main application, we show that the standard Brown–Resnick process has a moving maxima representation. This complements a result of Kabluchko, Schlather and de Haan [2], who established mixed moving maxima representations for these processes.

We obtain parallel characterization results for infinite variance symmetric sum–stable processes by using the notion of association.

References

- [1] B. M. Brown and S. I. Resnick. Extreme values of independent stochastic processes, *J. Appl. Probability*, 14(4):732–739, 1977.
- [2] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions, *Ann. Probab.* 37(5):2042–2065, 2009.
- [3] Y. Wang, S. Stoev, and P. Roy (2011) Decomposability for Stable Processes, *Department of Statistics, the University of Michigan, Technical Report 520*. <http://arxiv.org/abs/1105.1765>

Abstracts of Contributed Talks

Local Asymptotic Normality in δ -Neighborhoods of Standard Generalized Pareto Processes

Stefan AULBACH – *University of Würzburg, Germany*

Email: stefan.aulbach@uni-wuerzburg.de

Co-author: Michael FALK

Abstract

De Haan and Pereira [1] provided models for spatial extremes in the case of stationarity, which depend on just one parameter $\beta > 0$ measuring tail dependence, and they proposed different estimators for this parameter. This framework was supplemented in [2] by establishing local asymptotic normality (LAN) of a corresponding point process of exceedances above a high multivariate threshold, yielding in particular asymptotic efficient estimators.

The estimators investigated in these papers are based on a finite set of points t_1, \dots, t_d , at which observations are taken. We generalize this approach in the context of functional extreme value theory (EVT). Contrary to multivariate EVT, functional EVT does not investigate finite dimensional random variables but stochastic processes. Actually it turns out that, in a proper setup, multivariate results can easily be deduced from the functional case.

The more general framework of functional EVT allows estimation over some spatial parameter space, i. e., the finite set of points t_1, \dots, t_d is replaced by $t \in [a, b]$. In particular, we derive efficient estimators of β based on those processes in a sample of iid processes in $C[0, 1]$ which exceed a given threshold function.

References

- [1] L. de Haan and T. Pereira, *Spatial extremes: models for the stationary case*, Ann. Statist. **34**, 146-168, 2006.
- [2] M. Falk, *Local asymptotic normality in a stationary model for spatial extremes*, J. Multivariate Anal. **102**, 48-60, 2011.

Quantile estimation for a bivariate sample

Marie-Francoise BARME-DELCROIX – *Lab. Painlevé, University Lille 1, France*

Email: marie-francoise.barme@math.univ-lille1.fr

Co-author: Margarida BRITO

Abstract

There is no natural ordering of a multidimensional space and the extension of the definition of univariate quantiles to multivariate distributions is, therefore, not straightforward. We use a definition based on the ordering of a multivariate sample according to an increasing family of curves, we have called isobars. For a given level u , an u -level isobar is defined as a level curve of the conditional distribution function of the radius given the angle (that is, as a conditional quantile), the sample points being defined by their polar coordinates. In this way, the maximum value of the sample is defined as the point which belongs to the upper level isobar and it is really a sample point. Moreover, the so-defined order statistics may be characterized by an unidimensional approach ([1], [2]). For the applications the distribution function is, in general, unknown and we have to estimate the isobars.

For an unknown distribution and for a given level, here we investigate the behaviour of an histogram-type estimator of this conditional quantile based on a partition of the unit circle. Under additional regularity and mild Lipschitz conditions on the underlying distribution, we establish the uniform asymptotic normality and the uniform almost sure consistency of this estimator.

References

- [1] M. F. Barne-Delcroix and M. Brito, Multivariate stability and strong limiting behaviour of intermediate statistics, *J. Multivariate Anal.* 79, 157–170, 2001.
- [2] M. F. Barne-Delcroix and U. Gather, Limit laws for multidimensional extremes, *Statistics and Probability Letters* 77, 1750–1755, 2007.

Prediction of Catastrophes in Space over Time

Anastassia BAXEVANI – *University of Gothenburg, chalmers University of Technology, Sweden*

Email: anastass@chalmers.se

Co-authors: Richard WILSON and Manolo SCOTTO

Abstract

An optimal alarm policy for detecting catastrophes in spatio-temporal settings plays a central role in the analysis of the occurrence and impact of future rare events. The key research objective is to develop optimal alarm systems based on excursions of an alarm field to predict whether another field will enter a catastrophic situation at a later time.

In this work an optimal alarm system in space over time is introduced and studied in detail. To this extend, we generalize the results obtained by [1] and ([2], [3]) for stationary stochastic processes evolving in continuous time to the present setting. Specifically, we say that a catastrophe is considered to commence at some time if the random field exceeds the location specific level in some spatial neighborhood of the location whatever the situation at the other locations of interest, i.e., catastrophes that start outside the region and work into it are ignored. In this framework we will consider the case that the locations of possible catastrophes are discrete, lie along a smooth bounded one dimensional curve through a two-dimensional space (think of locations along the coastline) and the case the locations lie in an bounded connected region in a two dimensional space.

This latter, is qualitatively more difficult than the setting covered previously as it involves conditioning on a fixed level upcrossing in time for the first time in a spatial region; i.e. it involves conditioning on the occurrence of a point at which the random field has a level upcrossing in time and a local maximum in space. We can do this relatively easy for stationary (both in space and time) Gaussian random fields but it is moderately straightforward for Gaussian random fields stationary in time but non-stationary in space. Obviously, the choice of the alarm would reflect information which gives a higher probability of a catastrophe and should be chosen according to some optimality criteria. The celebrated Rice formula will be used to obtain Palm distributions and the probabilities of interest will be calculated by obtaining Slepian model representations of the alarm and catastrophe random fields, both by conditioning on the event that a catastrophe commences at a specific time or that an alarm commences at a specific time.

A natural extension of the above setting is for the case the underlying field is no longer Gaussian but is substituted by a Laplace field. In this case, we need to develop Palm distributions and Slepian model representations for Laplace fields.

References

- [1] De Maré, J., (1980). Optimal prediction of catastrophes with applications to Gaussian processes. *Ann. Probab.*, **8**, pp. 841-850.
- [2] Lindgren, G., (1980). Model processes in nonlinear prediction with application to detection and alarm. *Ann. Probab.*, **8**, pp. 775-792.
- [3] Lindgren, G., (1985). Optimal Prediction of Level Crossings in Gaussian Processes and Sequences *Ann. Probab.*, **13**, Number 3, pp. 804-824.

Light-Tail Estimation: a Negative Moment EVI-estimator and an Application to Environmental Data

Frederico CAEIRO – *Universidade Nova de Lisboa, Portugal*

Email: fac@fct.unl.pt

Co-authors: M. Ivette GOMES and Lígia HENRIQUES-RODRIGUES

Abstract

Whenever interested in extreme large events, the most common models in *statistics of extremes* are semi-parametric or even non-parametric in nature, with the imposition of a few “regularity conditions” in the right-tail, $\bar{F}(x) := 1 - F(x)$, as $x \rightarrow +\infty$, of an unknown model F underlying the available data. The primordial parameter is the *extreme value index* (EVI). For large values, the EVI is the shape parameter γ in the distribution function (d.f.) $EV_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma})$, $1 + \gamma x > 0$, the (unified) *extreme value* (EV) distribution. The EVI needs to be estimated in a precise way, because such an estimation is one of the basis for the estimation of other parameters of extreme and large events, like a *high quantile* of probability $1 - p$, with p small, the *right endpoint* of the model F underlying the data, $x^F := \sup\{x : F(x) < 1\}$, whenever finite, and the *return period* of a high level, among others.

For the EVI-estimation, we first refer one estimator, valid for all $\gamma \in \mathbb{R}$, the *moment* (M) estimator (Dekkers *et al.*, 1989), with the functional form $\hat{\gamma}_{k,n}^M := M_{k,n}^{(1)} + \frac{1}{2} \left\{ 1 - \left(M_{k,n}^{(2)} / [M_{k,n}^{(1)}]^2 - 1 \right)^{-1} \right\}$, with $M_{k,n}^{(j)} := \sum_{i=1}^k \{ \ln X_{n-i+1:n} - \ln X_{n-k:n} \}^j / k$, $j \geq 1$. Note that the statistic $M_{k,n}^{(1)}$, is the well-known *Hill* estimator (Hill, 1975), often denoted $\hat{\gamma}_{k,n}^H \equiv M_{k,n}^{(1)} := \sum_{i=1}^k \{ \ln X_{n-i+1:n} - \ln X_{n-k:n} \} / k$, the average of the log-excesses and valid only for $\gamma \geq 0$. We can thus write the *moment* estimator as

$$\hat{\gamma}_{k,n}^M = \hat{\gamma}_{k,n}^H + \hat{\gamma}_{k,n}^{NM}, \quad \hat{\gamma}_{k,n}^{NM} := \frac{1}{2} \left\{ 1 - \left(M_{k,n}^{(2)} / [M_{k,n}^{(1)}]^2 - 1 \right)^{-1} \right\}, \quad (3)$$

with *NM* standing for *negative moment* estimator. Indeed, whereas the H-estimator, $\hat{\gamma}_{k,n}^H$, is consistent for the estimation of $\gamma_+ := \max(0, \gamma)$, the NM-estimator, in (3), is consistent for the estimation of $\gamma_- := \min(0, \gamma)$, provided that $k = k_n$ is an intermediate sequence, i.e., a sequence of integers such that $k = k_n \rightarrow \infty$ and $k_n = o(n)$, as $n \rightarrow \infty$. Under these same conditions, the M-estimator is consistent for the estimation of any real γ , which can be written as $\gamma = \gamma_+ + \gamma_-$. Most of the classical EVI-estimators have usually a high variance for small k and a high bias when k is large. This problem affects both the *moment* and the *Hill* estimator, and leads to a difficult choice of the “optimal” k , in the sense of the value k that minimizes the asymptotic mean squared error (MSE).

On the basis of a comment in Fraga Alves (1998), where it is noticed that when $\gamma < 0$, $\hat{\gamma}_{k,n}^{NM}$ and $\hat{\gamma}_{k,n}^M$ have the same asymptotic variance, Caeiro and Gomes (2010) were led to the introduction of a semi-parametric class of consistent estimators for $\gamma < 0$, which generalizes both the M and the NM-estimators. Such a class, denoted $NM(\theta)$, is given by

$$\hat{\gamma}_{k,n}^{NM(\theta)} := \hat{\gamma}_{k,n}^{NM} + \theta M_{k,n}^{(1)}, \quad \theta \in \mathbb{R}. \quad (4)$$

Apart from the usual integer parameter k , related with the number of top order statistics involved in the estimation, these estimators depend on an extra real parameter θ , which makes them highly flexible and possibly second-order unbiased for a large variety of models in $\mathcal{D}_{\mathcal{M}}(EV_{\gamma})_{\gamma < 0}$. In this paper, we are interested not only on the adaptive choice of the *tuning* parameters k and θ , but also on an application of these semi-parametric estimators to the analysis of sets of environmental and simulated data. Note that we get the *negative moment* estimator in (3) for $\theta = 0$ and the *moment estimator*, for $\theta = 1$. With the appropriate choice of θ , $\hat{\gamma}_{k,n}^{NM(\theta)}(k)$ enables us to have access to an estimator of a negative EVI with a smaller asymptotic bias and the same asymptotic variance as the M-estimator. We shall further provide an *Algorithm* for the adaptive choice of the *tuning* parameters under play in the semi-parametric estimation of the EVI through such an estimator. Finally, we apply the *Algorithm* to the analysis of a set of environmental data, the daily average wind speeds in knots (one nautical mile per hour), collected in Dublin airport, in the period 1961-1978, as well as to a set of simulated data. Due to the seasonality of wind data, we restrict ourselves to the Spring and Summer months, and analyze the whole data-set, merely as a curiosity.

References

- [1] Caeiro, F., Gomes, M.I. (2010): An asymptotically unbiased moment estimator of a negative extreme value index. *Discussiones Mathematica: Probability and Statistics* **30**:1, 5–19.
- [2] Dekkers, A.L.M., Einmahl, J.H.J., Haan, L. de, (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* **17**(4), 1833–1855.
- [3] Fraga Alves, M.I. (2001). Weiss-Hill estimator. *Test* **10**, 203–224.
- [4] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**(5), 1163-1174.

Bias Correction in Extreme Value Statistics with Extreme Value Index around Zero

Juan-Juan CAI – *Tilburg University, The Netherlands*

Email: j.cai@uvt.nl

Co-authors: Laurens de HAAN and Chen ZHOU

Abstract

The major difficulty in applying extreme value statistics is to produce accurate estimates of the extreme value index. Most of existing estimators of the extreme value index exhibit convenient statistical properties such as the asymptotic normality under a speed of convergence $k^{-1/2}$, with k denoting the number of high order statistics used in estimation. Although a higher choice of k is desired for better estimation accuracy, it may leads to a higher asymptotic bias. Thus, employing a bias correction procedure allows a higher choice of k . Bias correction for estimators of the extreme value index has been studied extensively in literature, however, either

for positive (see, e.g. [4]) or negative (see, e.g. [5]) extreme value index. In other words, prior knowledge on the sign of the extreme value index is required to perform existing bias correction procedures.

In meteorology and environmental science, empirical literature has documented that most random variables exhibit extreme value indices around zero (see, e.g. [2]). The probability weighted moment (PWM) estimator proposed by [3] is popular in such a situation (see, e.g. [1]). In this paper, we provide a bias correction procedure for the PWM estimator by estimating the asymptotic bias and subsequently subtracting the bias from the original estimator. We also produce a bias corrected quantile estimator connected to the PWM estimator, when the extreme value index is around zero. Lastly, we provide a bias corrected estimator of the endpoint of a distribution when the extreme value index is negative.

The main advantage of the bias corrected estimators explored in this paper is that one gets more flexibility in choosing the number of high order statistics used in estimation. Compared to the original estimators without bias correction, with choosing a higher choice of k in the bias corrected estimators, the mean squared errors are subsequently of a lower order. This advantage becomes apparent in simulations and an environmental application.

References

- [1] T. A. Buishand, Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal*, 36(4):345–365, 1991.
- [2] S. G. Coles and J. A. Tawn, Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B(Methodological)*, 58(2):377–392, 1996.
- [3] J. R. M. Hosking and J. R. Wallis, Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3):339–349, 1987.
- [4] M. I. Gomes and D. Pestana, A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association*, 102(477):280–292, 2007.
- [5] D. Li and L. Peng, Does bias reduction with external estimator of second order parameter work for endpoint? *Journal of Statistical Planning and Inference*, 139(6):1937–1952, 2009.

Nonparametric estimation for multivariate extremes

Miguel DE CARVALHO – *Ecole Polytechnique Fédérale de Lausanne, Switzerland*
Email: miguel.carvalho@epfl.ch
Co-author: Anthony DAVISON

Abstract

The Ramos–Ledford spectral model [2] has received considerable attention as an elegant tool for modelling multivariate extremes. A major advantage over existing models is its ability to unify in a single framework the cases of asymptotic dependence and asymptotic independence. Despite its attractive features, for purposes of estimation only a parametric asymmetric logistic model is available, thus restricting the range of its applications. In this paper we propose nonparametric estimation and inference procedures for the Ramos–Ledford spectral model by empirical likelihood techniques [1].

References

- [1] A. Owen, *Empirical Likelihood*, Boca Raton: Chapman and Hall, 2001.
- [2] A. Ramos and A. Ledford, A new class of models for bivariate joint tails, *J. R. Statist. Soc. B*, 71, 219–241, 2009.

Group theoretic dimension of stationary symmetric α -stable random fields

Arijit CHAKRABARTY – *Indian Statistical Institute, New Delhi, India*
Email: arijit@isid.ac.in
Co-author: Parthanil ROY

Abstract

The growth rate of the partial maximum of a stationary stable process was first studied in the works of [4, 5], where it was established, based on the seminal works of [1, 2], that the growth rate is connected to the ergodic theoretic properties of the flow that generates the process. The results were generalized to the case of stable random fields indexed by \mathbb{Z}^d in [3], where properties of the group of nonsingular transformations generating the stable process were studied as an attempt to understand the growth rate of the partial maximum process. This

work generalizes this connection between stable random fields and group theory to the continuous parameter case, that is, to the fields indexed by \mathbb{R}^d .

References

- [1] J. ROSIŃSKI (1995): On the structure of stationary stable processes. *The Annals of Probability* 23:1163–1187.
- [2] J. ROSIŃSKI (2000): Decomposition of stationary α -stable random fields. *Annals of Probability* 28:1797–1813.
- [3] P. ROY and G. SAMORODNITSKY (2008): Stationary symmetric α -stable discrete parameter random fields. *Journal of Theoretical Probability* 21:212–233.
- [4] G. SAMORODNITSKY (2004a): Extreme value theory, ergodic theory, and the boundary between short memory and long memory for stationary stable processes. *Annals of Probability* 32:1438–1468.
- [5] G. SAMORODNITSKY (2004b): Maxima of continuous time stationary stable processes. *Advances in Applied Probability* 36:805–823.

Alarm systems and catastrophes from a different point of view

Pasquale CIRILLO – *University of Bern, Switzerland*

Email: pasquale.cirillo@stat.unibe.ch

Co-author: Jürg HÜSLER

Abstract

Using a chain of urns, we build a Bayesian nonparametric alarm system to predict catastrophic events. In observing a given phenomenon, it is often important to predict if the variable or quantity of interest will or will not reach a predetermined critical level sometime in the near future. This type of analysis becomes fundamental when the studied phenomenon may have a strong impact on human life, as in the case of floods, tides, epidemics or large power outages. In these rare and extreme situations, the critical level frequently corresponds to some sort of catastrophe, and the possibility of giving an alarm in due time is invaluable.

An alarm system is hence a tool that predicts the occurrence of a catastrophic event at a specified time in the future on the basis of the available information. To be more exact, an alarm system is meant to activate when the probability of a catastrophe in a given time horizon overcomes a certain threshold of tolerance.

In the literature (e.g. [1]), an alarm system is said optimal when, for a set of available data, it possesses the highest probability of correct alarm. This means that when the alarm is given, the probability that the catastrophic event actually manifest itself is the highest among all possible systems.

A naive alarm system is represented by the predictor $\hat{X}_{t+h} = E[X_{t+h}|X_s, -\infty < s \leq t, h > 0]$, where an alarm is cast every time the predictor exceeds some risk level. Naturally this system is not optimal at all, because it does not show good performances in detecting exceedances, in correctly locating them in time and in reducing the number of false alarms.

Differently from other alarm systems in the literature, we propose a model that is constantly updated on the basis of the available information, according to the Bayesian paradigm. Thus one of our main assumptions is that catastrophic events over time may be assumed to be exchangeable in the sense of de Finetti, i.e. their joint distribution is immune to permutations.

Our idea is to use a special urn process that can only assume a finite number of values over time but that, thanks to Polya-like reinforcement mechanism, is able to learn from the past. Every value must be seen as a level of risk, on a scale that goes from 0, no risk, to L, catastrophe. We will assume time to be discrete, in order to have an intuitive description of the model. Nevertheless, we also show that it is possible to extend our alarm system to continuous time.

The use of a finite-valued process may seem reductive in many cases where the observed phenomenon assumes a continuous range of values. In reality, if $\{X_n\}$ is our process, it can be seen as the “simplified” (or discretized) version of another underlying process $\{Y_n\}$, characterized by a (much) larger space of states.

The construction we propose is a generalization of the model presented in [2], which belongs to the wider class of reinforced urn processes (RUP). RUP have been introduced in [3] as reinforced random walks on a state space of urns.

In the papers we discuss some interesting probabilistic properties of our model, showing how they can be used to predict catastrophic events. The construction we propose is rather flexible and it can be applied to very different problems.

In the last part of the work, we test our alarm system on a couple of interesting time series related to catastrophic events.

References

- [1] M. Antunes, M.A. Amaral-Turkman and F.K. Turkman, *A Bayesian approach to event prediction*, Journal of Time Series Analysis 24, 631-646, 2003.

- [2] P. Cirillo and J. Hüslér, *Extreme shock models: an alternative perspective*, Statistics and Probability Letters 81, 25-30, 2010.
- [3] P. Muliere, P. Secchi and S. Walker, *Urn schemes and reinforced random walks*, Stochastic Processes and their Applications 88, 59-78, 2000.

Extremes values for continuous diffusions and its skeleton: the bounded case and FARCH approximations.

Didier DACUNHA-CASTELLE – *Université Paris Sud, France*
 Email: Didier.Dacunha-Castelle@math.u-psud.fr
 Co-author: Thi Tu HUONG HOANG

Abstract

We address extreme theory for bounded ergodic diffusion with inaccessible boundary and non zero drift at the boundary.

We show that in this case the square volatility is linear at the boundary and it depends only of the shape coefficient of extremes if it exists a limit GEV theorem. We then give a detailed study of the copula of the skeleton imbedded Markov chain observes at equispaced times $k\Delta$. This kind of copula has very specific properties. The basic tool is a representation of the transition density given by [1]. This is a way to study different problems for the dependence on Δ of the extremes and on the link continuous /discrete time for this theory. The method can be applied to every ergodic stationary diffusion on \mathbb{R} with of course more complicated behaviors of extremes.

We apply these results to temperature data. These results are used also to get important improvement for models of simulation of temperatures. Here the main tool are FARCH (functional autoregressive process with functional variance coefficient) approximations of the previous skeleton, FARCH processes being studied as misspecified ones. These processes are also geometrically ergodic, we give a proof using the approximated renewal atom technic.

References

- [1] D. Dacunha-Castelle, D. Florens-Zmirou. Estimation of the coefficients of a diffusion for discrete observations. Stochastics 19, 1986.
- [2] R.A.Davis. Minimum and maximum of one dimensional diffusions. Stochastic processes and their applications 13, 1982.
- [3] Robinson & J.A. Tawn. Extremal analysis of processes sampled at different frequencies. J. Roy. Statist. Soc. Ser B 62, 2000.

A methodology for hyperbolic graph efficiency assessment via extremes

Abdelaati DAOUIA – *Toulouse School of Economics, France*
 Email: daouia@cict.fr

Abstract

We consider a sample of independent businesses in a sector of production activity where a vector of inputs X is used to produce multiple outputs Y . An important question is : what are the most efficient businesses that might be useful to emulate? Using ideas from extreme value theory in conjunction with the concept of hyperbolic graph efficiency measure, we have been able to come up with a reasonable answer. From a statistical viewpoint, the set of all possible producers may be viewed as the joint support of the population of businesses (X, Y) . The identification of the top sample observations lying near its optimal boundary is clearly a problem belonging to statistical modeling of extremes.

The hyperbolic distance function measures the potential input decrease and output increase along a hyperbolic path to achieve the efficient support boundary. Its free disposal hull estimator and two alternative extreme-value based estimators are discussed. Practical guidelines to effect the necessary computations of these estimators are described and generated data sets illustrate how they work out in practice. The Dekkers-Einmahl-de Haan [1] type of estimator appears to give frank improvement in fit to data and performs better than the free disposal hull estimator [2] in terms of both bias and mean-squared error even in presence of outliers. A case study is provided on frontier and efficiency analysis of the delivery activity of French postal services.

References

- [1] A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan, *A moment estimator for the index of an extreme-value distribution*, *The Annals of Statistics*, 17(4), 1833–1855, 1989.
- [2] D.C. Wheelock and P.W. Wilson, *Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations*, *Journal of Econometrics*, 145 (1-2), 209–225, 2008.

Exploratory Plots in Extreme Value Analysis

Bikramjit DAS – *ETH Zürich, Switzerland*

Email: bikram@math.ethz.ch

Co-author: Souvik GHOSH

Abstract

Exploratory plotting tools are widely used to diagnose the goodness-of-fit of data sets to a hypothesized distribution. Some of them have found extensive use in diverse areas of finance, telecommunication, environmental science, etc. in order to detect sub-exponential or heavy-tailed behavior in observed data. Here we look at two such methodologies: the Quantile-Quantile plots for heavy-tails and the Mean Excess plots. Under the assumption of heavy-tailed behavior of the underlying sample the convergence in probability of these plots to a fixed set in a suitable topology of closed sets of \mathbb{R}^2 has been studied in [1] and [2]. These results give theoretical justifications for using the plots to test the null hypothesis that the underlying distribution is heavy-tailed by checking if the observed plot is “close” to the limit under the null hypothesis. In practice though one set of observations would lead to only one plot of each kind. This is often not good enough to verify proximity of the plot to the fixed set of interest under the null hypothesis of heavy-tails. Here we provide weak limits for these two plots when the underlying distribution of the sample is heavy-tailed and as an application we are able to construct confidence bounds around the plots which enables us to check whether the underlying distribution is heavy-tailed or not.

References

- [1] B. Das and S.I. Resnick. QQ plots, random sets and data from a heavy tailed distribution. *Stochastic Models*, 24(1):103–132, 2008.
- [2] S. Ghosh and S.I. Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applications*, (120):1492–1517, 2010.

A new semi-parametric family of estimators for the second order parameter

El-hadji DEME – *Université Gaston Berger de Saint-Louis, Sénégal*

Email: eltinoyaw@yahoo.fr

Co-authors: Laurent GARDES and Stéphane GIRARD

Abstract

An important parameter in extreme value theory is the extreme value index γ . It controls the first order behavior of the distribution tail. In the literature, numerous estimators of this parameter have been proposed especially in the case of heavy tailed distributions (which is the situation considered here). The most known estimator was proposed by [2]. It depends on the k largest observations of the underlying sample. The bias of the tail index estimator is controlled by the second order parameter ρ . In order to reduce the bias of γ 's estimators or to select the best number k of observations to use, the knowledge of ρ is essential. Some estimators of ρ can be found in the literature, see for example [1, 2, 3]. We propose a semiparametric family of estimators for ρ that encompasses the three previously mentioned estimators. The asymptotic normality of these estimators is then proved in a unified way. New estimators of ρ are also introduced.

References

- [1] G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, **13**, 55–87, 2010.
- [2] M.I. Fraga Alves, M.I. Gomes, and L. de Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, **60**(2), 193–213, 2003.
- [3] Y. Goegebeur, J. Beirlant, and T. de Wet. Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, **140**, 2632–2652, 2010.
- [4] B.M. Hill. A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174, 1975.

Robust estimators of extreme value index and linear models

Jan DIENSTBIER – *Technical University of Liberec, Czech Republic*
Email: jan.dienstbier@tul.cz

Abstract

In the contribution we deal with estimation of extremal properties of heavy-tailed linear models. We use quantile sensitive linear regression estimators such as regression quantiles. The idea is to plug their residuals or intercepts into the robust tail index estimators proposed in [1]. Derived estimators of extreme value index are consistent and asymptotically normal which can be proven using the theory of smooth functionals of tail quantile function [1] and an improved version of the law of iterated logarithm for regression quantiles similar to the one derived in [3]. We show that described location and scale invariant estimators work for simulated datasets with a reasonable reliability. A practical adaptation of the method is also demonstrated on Condroz dataset of calcium levels in Belgium Condroz region.

References

- [1] J. Beirlant, J.L. Teugels, and P. Vynckier, *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, 1996.
- [2] H. Drees, On Smooth Statistical Tail Functionals, *Scandinavian journal of Statistics* 25, 1998, 187–210.
- [3] J. Jurečková, Regression rank scores testes against heavy-tailed alternatives, *Bernoulli* 5, 1999, 659-676.

Statistical Analysis of Extremal Dependence: Bootstrapping the Extremogram

Holger DREES – *University of Hamburg, Germany*
Email: holger.drees@uni-hamburg.de

Abstract

The complexity of the serial dependence between extreme values of a time series cannot be captured by a single number like the extremal index. Recently Davis and Mikosch [1] introduced the extremogram which is given by the limiting conditional probability that an observation lagged by a given number of time points belongs to some extreme set tB given the present observation belongs to another set tA as $t \rightarrow \infty$. They examined the asymptotics of an empirical version of the extremogram for fixed sets A and B , but the *joint behavior* of these estimators over an infinite family of sets is needed to get a full picture of the extremal serial dependence.

In the talk we will present the asymptotics of a process of empirical extremograms indexed by sets. As the limiting Gaussian process usually has a quite complex covariance structure, we indicate how to use a multiplier bootstrap to construct uniform confidence bands. The main technical tools are limit theorems for a general class of empirical cluster functionals established by Drees and Rootzén [2] and a recent result which proves consistency of bootstrap versions of these processes.

References

- [1] Richard A. Davis and Thomas Mikosch, *The extremogram: A correlogram for extreme events*, *Bernoulli* 15 (4), 977–1009, 2009.
- [2] H. Drees and H. Rootzén, *Limit Theorems for Empirical Processes of Cluster Functionals*, *Annals of Statistics* 38 (4), 2145–2186, 2010.

Evaluating extreme snow avalanches in long term forecasting

Nicolas ECKERT – *Cemagref UR ETNA, France*
Email: nicolas.eckert@cemagref.fr
Co-author: Ophelie GUIN

Abstract

Mountain communities are exposed to snow avalanche risk in winter. Hence, evaluating extreme avalanches is a crucial question for hazard zoning and the design of defense structures, i.e. long term forecasting. Different damageable quantities (traveled distance, impact pressure, flow depth, deposit volumes) have to be considered,

but multivariate extreme value theory is for now not used in this field. Even the use of univariate EVT remains difficult because the most critical variable, the traveled distance, strongly depends on topography [1]. The recently developed alternative is the combination of a mechanical model for flow propagation with a stochastic model describing the variability of the different inputs/outputs, leading to a multivariate Peak Over Threshold (POT) model where the correlation between the different magnitude variables is constrained by physical rules [2]. Crucial technical problems are model identifiability and finding a reasonable compromise between precision of the description of the flow and computation times. These points can be addressed with a depth-averaged set of equation describing the propagation of snow avalanches within a hierarchical Bayesian framework [1]. First, the joint posterior distribution of model unknowns is estimated using a sequential Metropolis-Hastings algorithm. Second, the point estimates are used to predict the joint distribution of different variables of interest for hazard mapping. With regards to a purely statistical approach, this has the advantage of introducing reliable physics into the modeling. On the other hand, consistency with EVT in terms of attraction domain and asymptotic dependence of the different variables is not granted and remains an open space for further research. The different steps of the method are illustrated and discussed with real case studies from the French Alps.

References

- [1] D. McClung and K. Lied, *Statistical and geometrical definition of snow-avalanche runout*, Cold Regions Science and Technology, 13:107-119, 1987.
- [2] M. Barbolini and C. J. Keylock, *A new method for avalanche hazard mapping using a combination of statistical and deterministic models*, Natural Hazards and Earth System Sciences, 2:239-245, 2002.
- [3] N. Eckert, M. Naaim and E. Parent, *Long-term avalanche hazard assessment with a Bayesian depth-averaged propagation model*, Journal of Glaciology, 56:563-586, 2010.

A Novel Approach to Brown-Resnick Processes and Mixed Moving Maxima Processes

Sebastian ENGELKE – *Georg-August-Universitaet Goettingen, Germany*

Email: sengelk@uni-goettingen.de

Co-authors: Martin SCHLATHER and Zakhar KABLUCHKO

Abstract

Brown and Resnick [1] introduced a max-stable process that is obtained as a limit of maxima of independent Ornstein-Uhlenbeck processes. This so-called Brown-Resnick process and its generalizations are in fact essentially the only limits which arise as maxima of independent Gaussian processes [5]. Another representation of the Brown-Resnick process is given by the maximum of Brownian motions $(B_t)_{t \in \mathbb{R}}$ drifting to negative infinity for $t \rightarrow \pm\infty$. Stoev [4] showed that substituting the Brownian motions in this construction for suitable Lévy processes still yields a stationary, max-stable process. This results, however, only holds for one-sided processes on the positive axis.

To obtain a two-sided stationary process we cannot simply reflect the Lévy process at the y -axis as in the Brownian case. In fact, we show that a certain transformation of the Lévy process is necessary to ensure stationarity on the whole real line. Moreover, we analyze properties of these processes and show that they have a mixed moving maxima representation under certain conditions. For the special case when the Lévy process is a Brownian motion, i.e. the original Brown-Resnick process, we explicitly derive the distribution of the processes used in its mixed moving maxima representation. It turns out that it equals the distribution of the Brownian motion conditioned to stay negative. This fact naturally entails a fast and efficient simulation method of the Brown-Resnick process [2].

Furthermore, we replace the Lévy-process by a d -dimensional Lévy-Mori field [3] and obtain a new class of max-stable, stationary random fields in d dimensions.

References

- [1] B. M. Brown and S. I. Resnick, Extreme values of independent stochastic processes, *J. Appl. Probability*, 14:139-156, 1977.
- [2] S. Engelke, Z. Kabluchko and M. Schlather, An equivalent representation of the Brown-Resnick process, *Stat. Probab. Lett.*, In Press, 2011.
- [3] T. Mori, Representation of linearly additive random fields, *Probab. Theory Related Fields*, 92(1):91-115, 1992.
- [4] S. A. Stoev, On the ergodicity and mixing of max-stable processes, *Stochastic Process. Appl.*, 118(9):1679-1705, 2008.
- [5] Z. Kabluchko, M. Schlather and L. de Haan, Stationary max-stable fields associated to negative definite functions, *Ann. Probab.*, 37:2042-2065, 2009.

Sojourn Times and the Fragility Index

Michael FALK – *University of Würzburg, Germany*

Email: falk@mathematik.uni-wuerzburg.de

Co-author: Martin HOFMANN

Abstract

We investigate the sojourn time above a high threshold of a continuous stochastic process $Y = (Y_t)_{t \in [0,1]}$ on $[0, 1]$ with continuous identical margins. It turns out that the limit, as the threshold increases, of the expected sojourn time given that it is positive, exists if the copula process corresponding to Y is in the functional domain of attraction of an extreme value process η as defined in [1]. This limit coincides with the limit of the fragility index ([2]) corresponding to $(Y_{i/n})_{1 \leq i \leq n}$ as n and the threshold increase. It turns out that this limit is the reciprocal of the mass of the (finite) spectral measure corresponding to η ([3]) or, equivalently, of the functional D -norm of the constant function one ([1]).

If the process is in a certain neighborhood of a generalized Pareto process, then we can replace the constant threshold by a general threshold function and we can compute the asymptotic sojourn time distribution. An extreme value process is a prominent example.

References

- [1] S. Aulbach, M. Falk and M. Hofmann, *On extreme value processes and the functional D -norm*. Technical report, University of Würzburg, 2011.
- [2] J. L. Geluk, L. de Haan and C. G. de Vries, *Weak and strong financial fragility*. Tinbergen Institute Discussion Paper, TI 2007-023/2, 2007
- [3] L. de Haan and A. Ferreira, *Extreme Value Theory. (An Introduction)*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

Poisson process approximation for point processes of exceedances using Stein's method

Anne FEIDT – *University of Zurich, Switzerland*

Email: anne.feidt@math.uzh.ch

Abstract

A fundamental result of extreme value theory is that a point process of exceedances can be shown to asymptotically behave like a Poisson process. Stein's method for point process approximation, developed by Barbour and Brown in [1], additionally provides bounds on the errors involved in approximating a point process by a Poisson process with the same mean measure. These bounds depend on the sample size n and thus give a more precise idea on the accuracy of the approximation for finite sample sizes.

With a generalization to higher dimensions in mind, we study the behaviour of the point process of occurrences of iid bivariate random variables in sets in which they can be considered extreme. We aim to use Stein's method to approximate by a Poisson process with an intensity measure that reflects the copula structure of the distribution of the random variables. We show this on the example of a point process of simultaneous exceedances over some thresholds of both components of iid bivariate geometric random pairs whose survival copula belongs to the Marshall-Olkin family of copulas.

More precisely, we first determine the accuracy in total variation of the approximation by a Poisson process having the same intensity as the original point process, defined on the lattice \mathbb{Z}_+^2 . For more flexibility, we would prefer a Poisson process with a continuous intensity defined on \mathbb{R}_+^2 . To approximate further by such a process, the total variation metric proves too strong and we need to use a weaker Wasserstein metric to obtain sharp error bounds as well as rely on the distributional parameters to vary with the sample size at an appropriate rate. The latter condition however implies that the continuous intensity of the approximating Poisson process also depends on n . By imposing conditions on the parameters we are able to show that the approximation by a further Poisson process with a continuous intensity no longer varying with n will not add errors of a bigger order to the final error bound.

References

- [1] A. D. Barbour and T. C. Brown, Stein's method and point process approximation, *Stochastic Processes and their Applications*, 43(1): 9-31, 1992.

The Generalized Pareto Process and its Domain of Attraction

Ana FERREIRA – *ISA, Universidade Técnica de Lisboa and CEAUL, Portugal*

Email: anafh@isa.utl.pt

Co-author: Laurens DE HAAN

Abstract

The Generalized Pareto distribution has played a fundamental role in Extreme Value Theory ([1], [2] and e.g. [3] for possible extension to multivariate theory). We present a generalization of this concept for extremes in function space. We call this generalization the ‘Generalized Pareto Process’, a concept to formalize in agreement with some desirable properties. We present ways to construct it and characterize its domain of attraction.

References

- [1] A.A. Balkema and L. de Haan: Residual life time at great age. *Ann. Probab.* 2, 792–804 (1974)
[2] J. Pickands III: Statistical inference using extreme order statistics. *Ann. Statist.* 3, 119–131 (1975)
[3] H. Rootzén and N. Tajvidi: Multivariate generalized Pareto distributions. *Bernoulli* 12, 917–930 (2006)

Investigating the interplay of excess kurtosis and tail events in financial series

Anna M. FIORI – *University of Milano-Bicocca, Italy*

Email: anna.fiori@unimib.it

Abstract

The well established *fat tail issue* of financial returns is routinely supported by high values of the conventional kurtosis statistic [2]. Based on fourth powers of the observations, the measure is however unable to discriminate between heavy tailed distributions as d.g.p. or simply the presence of a few extremes in the sample ([3]).

For this reason, we suggest alternative measures of right and left kurtosis which are less sensitive to outlying observations and are consistent with common risk perceptions of investors and risk managers. Based on a recent interpretation of kurtosis as inequality at either side of the median ([4]), the new measures $R(D)$ and $R(S)$ presume a partial ordering of distributions induced by nested *kurtosis curves* and are easily computed by *Gini indices* in both symmetric and asymmetric contexts ([1]).

Using the theory of L-statistics, we construct consistent estimators of the new measures in the i.i.d. case and prove their asymptotic normality under minimal assumptions on the underlying distribution. While the sampling variance of the conventional kurtosis coefficient is related to the population moment of order eight, the new measures can be appropriately estimated under the milder requirement that second moments are finite. In addition to an explicit formula for the asymptotic variance, we estimate the latter by nonparametric and bootstrap techniques. Extensive Monte Carlo simulations are designed to compare the traditional kurtosis measure with its right and left counterparts in finite samples.

For purposes of financial applications, the interplay of kurtosis and clustering in the volatility dynamics needs to be considered. Consequently, we first fit ARMA-GARCH models to several series of daily stock market returns by QML techniques. The kurtosis curves and related measures $R(D)$ and $R(S)$ are then estimated on standardized residuals. While the stylized facts about excess kurtosis in financial series may have been accepted too readily, the new approach provides a deeper insight into the interplay of peakedness, tail events and right/left excess kurtosis.

References

- [1] A. M. Fiori, Measuring kurtosis by right and left inequality orders, *Communications in Statistics - Theory and Methods*, 37(17), 2665-2680, 2008.
[2] A. M. Fiori and M. Zenga, Karl Pearson and the origin of kurtosis, *International Statistical Review*, 77(1), 40-50, 2009.
[3] T. H. Kim and H. White, On more robust estimation of skewness and kurtosis, *Finance Research Letters*, 1(1), 56-73, 2004.
[4] M. Zenga, Kurtosis, in: Kotz, S., Read, C.B., Balakrishnan, N. and Vidakovic, B. (Eds.), *Encyclopedia of Statistical Sciences*, 2nd ed., New York: Wiley, 2006.

On double exceedances of stationary Gaussian processes

Jasmin FURRER – *University of Bern, Switzerland*

Email: jasmin.furrer@stat.unibe.ch

Co-author: Jürg HÜSLER

Abstract

We investigate the clustering of double exceedances of stationary Gaussian processes $X(t)$, $t \in [0, T]$. Exceedances over identical levels have been analyzed thoroughly in [1]. Based on these results, we consider the case with two different thresholds. Applications of such a setup will be particularly interesting in climate sciences, e.g. earth quakes which are of different strength. Exceedances over different thresholds which are separated by at least a time lag ε are considered. We assume that the correlation function has the following properties:

$$\begin{aligned} r(t) &= 1 - |t|^\alpha + o(|t|^\alpha) \quad \text{as } t \rightarrow 0, \\ r(t) &< 1 \quad \forall t > 0, \end{aligned}$$

with $\alpha \in (0, 2)$. Additionally, we assume that in the interval $[\varepsilon, T]$ there exists only one point of maximal correlation t_m , being an interior point of the interval, and that r is twice continuously differentiable in a neighborhood of t_m . Furthermore, the relationship between the two thresholds u and v is the following: $v = c \cdot u$ with $c \in (r(t_m), 1)$. With these assumptions, the following probability of a cluster of ε -separated exceedances can be derived exactly where the levels to be exceeded by the extreme events tend to ∞ :

$$P(u, c; \varepsilon, T) = P(\exists(s, t) : s, t \in [0, T], |t - s| \geq \varepsilon, X(s) > v, X(t) > u).$$

The shape of the excursion sets depends on the behavior of the conditional mean $E(X(t) | X(0) = u, X(t_m) = v)$, a function defined by u, c and the correlation function r . We will discuss some examples to indicate the pattern of exceedances depending on the given correlation function.

References

- [1] Hüsler, J., Ladneva, A. and Piterbarg, V. I. (2010) On clusters of high extremes of Gaussian stationary processes with ε -separation. *Electron. J. Probab.* **15** (59), 1825-1862.

Asymptotic comparison at optimal levels of reduced-bias EVI-estimators

M. Ivette GOMES – *University of Lisbon, Portugal*

Email: ivette.gomes@fc.ul.pt

Co-author: Frederico CAEIRO

Abstract

We are interested in the asymptotic comparison, at optimal levels, of a set of semi-parametric reduced-bias extreme value index (EVI) estimators, valid for a wide class of heavy-tailed models, underlying the available data. For heavy-tailed models in the domain attraction for maxima of an extreme value distribution $EV_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma})$, $x > -1/\gamma$, $\gamma > 0$, with γ the EVI, the classical EVI-estimators are the Hill estimators (Hill, 1975), which are the average of the scaled log-spacings as well as of the log-excesses, given by $U_i := i \{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ and $V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}$, $1 \leq i \leq k < n$, respectively. But the Hill EVI-estimators have often a strong asymptotic bias for moderate up to large values of k . Consequently, the adequate accommodation of the bias of Hill's estimators has been extensively addressed in recent years by several authors. We mention the pioneering papers by Beirlant *et al.* (1999) and Feuerverger and Hall (1999), among others. In all these papers, authors are led to second-order reduced-bias EVI-estimators, with asymptotic variances larger than or equal to $(\gamma(1 - \rho)/\rho)^2$, the minimal asymptotic variance of an “asymptotically unbiased” estimator in Drees' class of functionals, where $\rho < 0$ is a “shape” second-order parameter ruling the rate of convergence of the normalized sequence of maximum values towards the limiting *extreme value* random variable (r.v.). Recently, Caiiro *et al.* (2005) and Gomes *et al.* (2008), among others, considered, in different ways, the problem of corrected-bias EVI-estimation, being able to reduce the bias without increasing the asymptotic variance, which was shown to be kept at the value γ^2 , the asymptotic variance of Hill's estimator, the maximum likelihood estimator of γ for an underlying Pareto distribution. Those estimators, called *minimum-variance reduced-bias* (MVRB) EVI-estimators, are all based on an adequate external estimation of a pair of second-order parameters, $(\beta, \rho) \in (\mathbb{R}, \mathbb{R}^-)$, done through estimators denoted $(\hat{\beta}, \hat{\rho})$, and outperform the classical estimators for all k . We

shall here consider a set of MVRB statistics, denoted generally $UH_{\hat{\beta}, \hat{\rho}}(k)$, with UH standing for *unbiased Hill*. We shall also consider, similarly to what has been done in Beirlant *et al.* (1999), among others, the statistics

$$UH_{\hat{\rho}}^*(k) := UH_{\hat{\beta}(k; \hat{\rho}), \hat{\rho}}(k), \quad (5)$$

with an asymptotic variance that is no longer γ^2 but $\gamma^2(1-\rho)^2/\rho^2 > \gamma^2$ for every ρ . As mentioned before, in the $UH_{\hat{\beta}, \hat{\rho}}(k)$ and $UH_{\hat{\rho}}^*(k)$ classes, $(\hat{\beta}, \hat{\rho})$ or $\hat{\rho}$, respectively, need to be adequate consistent estimators of the second-order parameters, if we want to keep the same properties of $UH_{\beta, \rho}(k)$ or $UH_{\rho}^*(k)$, the associated r.v.'s. A third possibility is to estimate both the scale and shape second-order parameters at the same level k , used for the estimation of γ , like in Feuerverger and Hall (1999), also among others. We could then work with statistics of the type,

$$UH^{**}(k) := UH_{\hat{\beta}(k; \hat{\rho}(k)), \hat{\rho}(k)}(k), \quad (6)$$

with an asymptotic variance surely larger than $\gamma^2(1-\rho)^2/\rho^2$, for every ρ . We shall here consider the estimator in Feuerverger and Hall (1999), denoted $FH^{**}(k)$, as an illustration of estimators of the type of the ones in (6). We shall compare asymptotically, at optimal levels, the above mentioned set of MVRB statistics, denoted generically $UH(k)$, the reduced-bias statistics $UH^*(k)$, in (5) (assuming thus that β and ρ are known or adequately estimated) and the $FH^{**}(k)$ -statistics, as illustrative of reduced-bias EVI-estimators associated with an “internal” estimation of second-order parameters, like the ones in (6). Again, as in the classical case, there is not any estimator that can always dominate the alternatives, but interesting clear-cut patterns are found. Consequently, and in practice, a suitable choice of a set of EVI-estimators will jointly enable us to better estimate the EVI, the primary parameter of extreme events.

References

- [1] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**:2, 177-200.
- [2] Caeiro, F. Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**:2, 111-136.
- [3] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals Statistics* **27**, 760-781.
- [4] Gomes, M.I., de Haan, L. and Henriques-Rodrigues, L. (2008). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statistical Society* **B70**:1, 31-52.
- [5] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals Statistics* **3**, 1163-1174.

Bias-reduced estimators for bivariate tail modelling

Armelle GUILLOU – *Strasbourg University and CNRS, France*

Email: armelle.guillou@math.unistra.fr

Co-authors: Jan BEIRLANT and Goedele DIERCKX

Abstract

Ledford and Tawn [3] introduced a flexible bivariate tail model based on the coefficient of tail dependence and on the dependence of the extreme values of the random variables. In this paper, we extend the concept by specifying the slowly varying part of the model as done by Hall [2] with the univariate case. Based on Beirlant et al. [1], we propose a bias-reduced estimator for the coefficient of tail dependence and for the estimation of small tail probabilities. We discuss the properties of these estimators via simulations and a real-life example. Furthermore, we discuss some theoretical asymptotic aspects of this approach.

References

- [1] Beirlant, J., Joossens, E. and Segers, J., 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distribution. *Journal of Statistical Planning and Inference* **139**, 2800-2815.
- [2] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B* **44**, 37-42.
- [3] Ledford, A.W. and Tawn, J.A., 1997. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B* **59**, 475-499.

Lifting wind storms

Laurens DE HAAN – *Erasmus University, Netherlands*

Email: ldhaan@ese.eur.nl

Abstract

In view of establishing a novel method for determining hydraulic boundary conditions Deltares is planning to use wind data not only from a single source but over a wide area as input. These (artificial) wind field data should be used to gain insight on how a severe storm behaves (over time and space), one that is so severe that it is not expected to show up in the data. The idea is that extreme value theory can help to perform this task.

The above is a simplification of the problem as stated in Deltares report 1202120-001-HYE-0002 but it may capture the essence of the problem.

A method is proposed to produce these extreme storms not by estimating the structure of the limiting max-stable process but by "lifting" less extreme storms.

Products in conditional extreme value model

Rajat S. HAZRA – *Indian Statistical Institute, Kolkata, India*

Email:rajatmaths@gmail.com

Co-author: Krishanu MAULIK

Abstract

Classical multivariate extreme value theory tries to capture the extremal dependence between the components under a multivariate domain of attraction condition and it requires each of the components to be in domain of attraction of a univariate extreme value distribution as well. Multivariate extreme value (MEV) model has a rich theory but has some limitations as it fails to capture the dependence structure in presence of asymptotic independence. A different approach to MEV was given by [1], where they examined MEV distributions by conditioning on one of the components to be extreme. Here we assume one of the components to be in Fréchet or Weibull domain of attraction and study the behavior of the product of the components under this conditional extreme value model.

References

- [1] Janet E. Heffernan and Jonathan A. Tawn. A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):497–546, 2004. ISSN 1369-7412. With discussions and reply by the authors.

D-Norm of Extreme Value Processes and Functional Domain of Attraction

Martin HOFMANN – *University of Würzburg, Germany*

Email: hofmann.martin@mathematik.uni-wuerzburg.de

Co-authors: Stefan AULBACH and Michael FALK

Abstract

We introduce the distribution function G of a continuous max-stable process on $[0, 1]$ (cf. [2]) and it is shown that G can be represented via a norm on functional space, called D -norm. This is in complete accordance with the multivariate case and leads to a functional domain of attraction approach for stochastic processes, which is more general than the usual one based on weak convergence, cf. [3], [4].

This framework leads to a natural definition of functional generalized Pareto distributions (GPD) W in $C[0, 1]$, which satisfy $W = 1 + \log(G)$ in their upper tails. Moreover, we characterize the functional domain of attraction condition for copula processes in terms of tail equivalence with a functional GPD and introduce δ -neighborhoods of a functional GPD. It is shown that these are characterized by a polynomial rate of convergence of functional extremes, which all is in great analogy with the well-known finite-dimensional case (cf. [1]).

References

- [1] M. Falk, J. Hüsler and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*. 3rd ed., Birkhäuser, Basel, 2010.
- [2] E. Giné, M. Hahn, and P. Vatan, *Max-infinitely divisible and max-stable sample continuous processes*. *Probab. Theory Related Fields* 87, 139-165, 1990.
- [3] L. de Haan and A. Ferreira, *Extreme Value Theory. (An Introduction)*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
- [4] L. de Haan and T. Lin, *On convergence toward an extreme value distribution in $C[0, 1]$* , *Ann. Probab.* **29**, 467-483, 2001

An Efficient Estimation Method for Risk Models

Mei Ling HUANG – *Dept. of Mathematics, Brock University, St. Catharines, Ontario, Canada L2S 3A1*
Email: mhuang@brocku.ca
Co-author: Percy BRILL

Abstract

The extreme value distributions are Pareto type distributions which have many applications in risk analysis, survival analysis and queueing networks ([1]). There is difficulty in the inference of the Pareto distribution which has infinite moments in the heavy tailed situation. This paper studies the truncated Pareto distribution to avoid this difficulty. We also propose an optimal weighted estimation method for the truncated Pareto distribution. For the new estimator, an exact efficiency function relative to the classical estimators, and its properties, are derived. L1-optimal weights and L2-optimal weights are also proposed. Monte Carlo simulation results confirm the theoretical conclusions. The new estimation method has been applied to analyze hurricane and forest fire loss data. These studies show that the new estimation method is more efficient relative to existing methods, and fits data well in risk analysis.

References

- [1] P. Embrechts and C. Kluppelberg and T. Mikosch, *Modelling Extremal Events, for Insurance and Finances*, Springer, New York, 2003.
- [2] I. B. Aban and M. M. Meerschaert and A. K. Panorska, Parameter Estimation for Truncated Pareto Distribution, *Journal of the American Statistical Association*, 2006, Vo. 101, No. 473, pp. 270-277.

On the shape of excursions of Gaussian processes

Jürg HÜSLER – *University of Bern, Switzerland*
Email: juerg.huesler@stat.unibe.ch
Co-author: Vladimir PITERBARG

Abstract

We investigate the shape of excursions above a high level u by a stationary Gaussian process $X(t)$. Typically the excursions of the process $X(t)$ above a large value u are rather short and tend to 0 as u tends to ∞ . We call such a case a cluster of excursions since the path may cross the level u many times. The probability of such excursion events can be accurately approximated for large u , as is well-known.

However, the length of excursions can be much larger if one conditions on two clusters of excursions which are separated in time by a positive time lag, not depending on u . Under certain assumptions on the correlation function of the process, the asymptotic behavior of the probability of such a pattern of clusters of exceedances is derived exactly where the level u tends to ∞ (see [1]). The shape of such excursions is quite interesting and does depend on the conditioned mean and covariances of the underlying process.

We focus on the shape of paths in-between the two clusters or after one such cluster. The paths vary slightly around a deterministic trend. In addition, the probability of such events can be determined asymptotically exact for $u \rightarrow \infty$.

References

- [1] J. Hüsler, A. Ladneva and V. Piterbarg, *On clusters of high extremes of Gaussian stationary processes with ε -separation*. Elect. J. of Probability **15**, 1825-1862, 2010.

Assessing the Index of Regular Variation for Random Difference Equations

Anja JANSSEN – *University of Hamburg, Germany*
Email: janssen@math.uni-goettingen.de
Co-author: Ulf-Rainer FIEBIG

Abstract

We analyze multivariate random difference equations (RDEs), which are of the form

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{N}, \quad (7)$$

for $\mathbf{Y}_t \in \mathbb{R}^d$, where \mathbf{A}_t are random $\mathbb{R}^{d \times d}$ -matrices and \mathbf{B}_t are random \mathbb{R}^d -vectors. We assume that $(\mathbf{A}_t, \mathbf{B}_t)_{t \in \mathbb{N}}$ are i.i.d. and independent of \mathbf{Y}_0 .

In a seminal paper [3], Kesten derived conditions for $(\mathbf{A}_1, \mathbf{B}_1)$ (mainly for the case of non-negative entries of \mathbf{A}_1 and \mathbf{B}_1) which guarantee the existence of a stationary solution $(\mathbf{Y}_t)_{t \in \mathbb{N}_0}$ to (7). Furthermore, he derived a property of the stationary distribution which could later be shown to imply multivariate regular variation, cf. [1]. Now, an important characteristic for the extremal behavior of the process $(\mathbf{Y}_t)_{t \in \mathbb{N}_0}$ is given by the index of regular variation $\kappa > 0$, which satisfies

$$\lim_{y \rightarrow \infty} \frac{P(\|\mathbf{Y}_t\| > xy)}{P(\|\mathbf{Y}_t\| > y)} = x^{-\kappa}$$

for all $x > 0$ and an arbitrary norm $\|\cdot\|$ on \mathbb{R}^d . However, for most cases and especially for higher-dimensional matrices, the formula for the derivation of κ which is stated in [3], namely that $\kappa > 0$ is the unique solution to

$$0 = \lim_{n \rightarrow \infty} \frac{1}{n} \ln E \|\mathbf{A}_n \cdot \dots \cdot \mathbf{A}_1\|_{\text{op}}^\kappa,$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm, is not analytically solvable. This instance is especially problematic since κ is not only of interest in itself but needed for many further derivations and simulations of the extremal behavior of a stationary solution to (7). We will make use of a characteristic of κ that is introduced quite implicitly in [3] and which does not involve a limit but a certain spectral measure ν_κ on $\mathbb{S}_+^{d-1} = \{\mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\| = 1\}$ which satisfies

$$\int_{\mathbb{S}_+^{d-1}} E \left[\|\mathbf{A}_1 \mathbf{x}\|^\kappa f \left(\frac{\mathbf{A}_1 \mathbf{x}}{\|\mathbf{A}_1 \mathbf{x}\|} \right) \right] \nu_\kappa(d\mathbf{x}) = \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x}) \nu_\kappa(d\mathbf{x})$$

for all continuous functions f on \mathbb{S}_+^{d-1} . We propose a numerical algorithm which allows us to approximate ν_κ and thereby evaluate κ .

The talk is based on [2].

References

- [1] J. Boman and F. Lindskog, Support theorems for the Radon transform and Cramér-Wold theorems, *J. Theoret. Probab.*, **22**(3) 683710: 2009
- [2] A. Janßen, *On Some Connections between Light Tails, Regular Variation and Extremes*, Ph.D. thesis, University of Göttingen: 2010
- [3] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, *Acta Math.*, **131** 207248: 1973

Asymptotic behaviour of a test of no breaks versus fixed number of breaks

Daniela JARUSKOVA – *Czech Technical University, Prague, Czech Republic*
Email: jarus@mat.fsv.cvut.cz

Abstract

A sequence of independent random variables X_1, \dots, X_n is observed. We test the null hypothesis H_0 claiming that all observations are identically distributed with the same mean value μ :

$$H_0 : X_i = \mu + e_i, \quad i = 1, \dots, n,$$

against the alternative that there exists at least one and at most d breaks in the mean (at unknown positions):

$$A_d : \text{there exist } 1 < n_1 < n_2 < \dots < n_d < n \text{ satisfying } n_1 \geq \epsilon n, n_2 - n_1 \geq \epsilon n, \dots, n - n_d \geq \epsilon n$$

$$X_i = \mu_1 + e_i, \quad i = 1, \dots, n_1, \quad X_i = \mu_2 + e_i, \quad i = n_1 + 1, \dots, n_2, \quad \dots, \quad X_i = \mu_{d+1} + e_i, \quad i = n_d + 1, \dots, n,$$

such that $\mu_{j_0} \neq \mu_{j_0+1}$ for some $1 \leq j_0 \leq d$. The $\{e_i\}$ are i.i.d. satisfying $E e_i = 0$, $E e_i^2 = 1$ and $E |e_i|^{2+\Delta} < \infty$.

Under H_0 the test statistic that is based on the standardized differences of least squares estimates of $\{\mu_j\}$ converges in distribution to a maximum of a χ^2 - process (see [1]):

$$\max_{\substack{0 < t_1 < t_2 < \dots < t_d < t_{d+1} = 1 \\ t_1 \geq \epsilon, t_2 - t_1 \geq \epsilon, \dots, 1 - t_d \geq \epsilon}} X_1^2(\mathbf{t}) + \dots X_d^2(\mathbf{t}),$$

where

$$X_i(\mathbf{t}) = \frac{1}{\sqrt{t_{i+1}t_i(t_{i+1} - t_i)}} \left(t_i W(t_{i+1}) - t_{i+1} W(t_i) \right).$$

The approximate critical values may be obtained from the approximation of the exceedance probability (see [2] and [3]):

$$P\left(\max_{\substack{0 < t_1 < t_2 < \dots < t_d < t_{d+1} = 1 \\ t_1 \geq \epsilon, t_2 - t_1 \geq \epsilon, \dots, 1 - t_d \geq \epsilon}} X_1^2(\mathbf{t}) + \dots X_d^2(\mathbf{t}) > u^2\right) \sim \frac{1}{\pi^{(d-1)/2}} I_d u^{3d-1} (1 - \Phi(u)) \quad \text{as } u \rightarrow \infty.$$

An explicit expression for the constant I_d may be applied to obtain values of I_d for d small.

References

- [1] J. Bai and P. Perron *Estimating and testing linear models with multiple structural changes*, *Econometrica* 66, 47–78, 1998.
- [2] V. I. Piterbarg *Asymptotic methods in the theory of Gaussian processes and fields*, American Mathematical Society, 1996.
- [3] D. Jarušková and V. I. Piterbarg *Log-likelihood ratio test for detecting transient change*, *Statistics and Probability Letters*, 2011, doi:10.1016.

Max-stable Processes for Threshold Exceedances in Spatial Extremes

Soyoung JEON – *University of North Carolina at Chapel Hill, U.S.A.*

Email: soyoung@email.unc.edu

Co-author: Richard L. SMITH

Abstract

The analysis of spatial extremes requires the joint modeling of a spatial process at a large number of stations. Multivariate extreme value theory can be used to model the joint extremal behavior of environmental data such as precipitation, snow depths or daily temperatures. Max-stable processes are the natural generalization of extremal dependence structures to infinite dimensions arising from the extension of multivariate extreme value theory. However, there have been few or no works on the threshold approach of max-stable processes.

Padoan, Ribatet and Sisson [1] proposed the maximum composite likelihood approach for fitting max-stable processes to avoid the complexity and unavailability of the multivariate density function. We propose the threshold version of max-stable process estimation and we apply the pairwise composite likelihood method to it. We assume a strict form of condition, so called the second-order regular variation condition, for the distribution satisfying the domain of attraction. It is well known that the condition was used to prove the asymptotic properties of estimators in univariate threshold approach (see Smith [2]) and the second-order condition was studied for bivariate extremes by de Haan and Ferreira [1]. To obtain the limit behavior, we also consider the increasing domain structure with stochastic sampling design based on the setting and conditions in Lahiri [2] and we then establish consistency and asymptotic normality of the estimator for dependence parameter in the threshold method of max-stable processes. The method is studied by simulation and illustrated by the application of temperature data in North Carolina, United States.

References

- [1] S. Padoan, M. Ribatet and S. Sisson, *Likelihood-based inference for max-stable processes*, *Journal of the American Statistical Association* 105(489), 263–277, 2010.
- [2] R. L. Smith, *Estimating tails of probability distributions*, *Annals of Statistics* 15, 1174–1207, 1987.
- [3] L. de Haan and A. Ferreira, *Extreme Value Theory, An Introduction*, New York: Springer, 2006.
- [4] S. N. Lahiri, *Central Limit Theorems for weighted sums under some stochastic and fixed spatial sampling designs*, *Sankhya: The Indian Journal of Statistics, Ser. A* 65, 356–388, 2003.

Conditional extremes incorporating covariates

Philip JONATHAN – *Shell Technology Centre Thornton, U.K.*

Email: philip.jonathan@shell.com

Co-authors: David RANDELL, Kevin EWANS and Jan FLYNN

Abstract

Reliable design and assessment of marine structures requires estimation of both marginal and joint characteristics of extreme ocean environments. The conditional extremes model of Heffernan and Tawn [1] motivates a particular parametric model for the distribution of variables given a large value of a conditioning variate, for suitably transformed variables. In this work, we extend the conditional extremes approach to incorporate covariate effects in both marginal and conditional extremes. The smoothness of marginal and conditional extremes model parameters with covariate is regulated by roughness-penalised maximum likelihood (see, e.g., [2]). The approach is applied to the estimation of extreme quantiles of ocean storm severity and conditional values of wave peak periods given storm wave direction, for a number of ocean locations, extending [3].

References

- [1] Heffernan, J. E. and J. A. Tawn, *A conditional approach for multivariate extreme values*, J. R. Statist. Soc. B, 2004, **66**, 497.
- [2] Ewans, K. C. and P. Jonathan, *The effect of directionality on Northern North Sea extreme wave design criteria*, J. Offshore Mechanics Arctic Engineering, 2008, **130**, 10.
- [3] Jonathan, P., J. Flynn and K. C. Ewans, *Joint modelling of wave spectral parameters for extreme sea states*, Ocean Engineering, 2010, **37**, 1070.

Extreme value statistics of wind speed data by the ACER method

Oleh KARPA – *CeSOS, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway*

Email: oleh.karpa@ntnu.no

Co-author: Arvid NAESS

Abstract

The paper concentrates on the application of a method that appears to be uniquely appropriate for predicting extreme wind speeds. Standard methods for estimating extreme wind speeds from observed data series are commonly based on assuming either that episodic extreme values are Gumbel distributed or by adopting a peaks over threshold (POT) approach, assuming that the exceedances above high thresholds follow a generalized Pareto distribution. A weakness of these approaches is that they depend on adopting asymptotic distributions. However, it is hard to fully verify the applicability or correctness of such procedures. An alternative method for predicting extreme wind speeds that avoids invoking the ultimate asymptotic distributions, but rather try to capture the sub-asymptotic behavior of extreme value data has been developed by Naess and Gaidai [2]. This approach has been used in this work and seems to be the appropriate way to deal with the recorded data time series of, for example, the hourly largest wind speeds observed at a given location.

Assume that the observed data X_1, \dots, X_N are allocated to the discrete times t_1, \dots, t_N . In general, these data may be strongly dependent. Our goal is to accurately determine the distribution function of the extreme value $M_N = \max\{X_j; j = 1, \dots, N\}$. Specifically, we want to estimate $P(\eta) = \text{Prob}(M_N \leq \eta) = \text{Prob}(X_1 \leq \eta, \dots, X_n \leq \eta)$ accurately for large values of η . The joint distribution function on the right hand side of this equation cannot in general be estimated directly from the data. However, this problem can be solved in practice by introducing a cascade of conditioning approximations $P_k(\eta)$ of $P(\eta)$, where $P_k(\eta) \rightarrow P(\eta)$ as k increases, and $P_k(\eta) \approx \exp(-\sum_{j=k}^N \alpha_{kj}(\eta))$, for $N \gg 1$ and $k = 1, 2, \dots$, where $\alpha_{kj}(\eta) = \text{Prob}(X_j > \eta | X_{j-1} \leq \eta, \dots, X_{j-k+1} \leq \eta)$.

For the empirical estimation of the quantities in the $P_k(\eta)$, average conditional exceedance rates (ACER) are introduced as follows, $\varepsilon_k(\eta) = \sum_{j=k}^N \alpha_{kj}(\eta) / (N - k + 1)$. Estimation of the ACER function $\varepsilon_k(\eta)$ proceeds by counting the total number of favorable incidents, that is, exceedances conditioned on the requisite number of preceding non-exceedances, for the total data time series.

For the prediction problem, it is argued in [1] that the ACER functions can be represented in the tail as $\varepsilon_k(\eta) \approx q_k \exp\{-a_k(\eta - b_k)^{c_k}\}$, $\eta \geq \eta_1$, where a_k, b_k, c_k and q_k are suitable constants, that in general will be dependent on k . The optimal values of the parameters are obtained by optimizing the fit on the log level by minimizing a mean square error function. Note that the values $c_k = q_k = 1$ correspond to the Gumbel asymptotic form.

Wind speed data, measured at Sula and Torsvåg Fyr weather stations in Norway, were analyzed to obtain numerical results. Hourly maximum wind gust speeds were recorded during 12 years 1998-2010 at the first station and 13 years 1997-2010 at the second.

For the analysis of the data with the ACER method, $\varepsilon_k(\eta)$ for $k = 1, \dots, 5$ were chosen for both stations. As expected, it was revealed that there is significant dependence between the data. However, this dependence is largely accounted for by $k = 2$; and for $k \geq 5$ full convergence has been achieved for all practical purposes.

Optimal curve fitting and estimation of 100-year return period values were achieved for both stations. In the table 100-year return period values are listed together with 95% confidence intervals.

k	Sula		Torsvåg	
	η_{100} , m/s	95% CI(η_{100}), m/s	η_{100} , m/s	95% CI(η_{100}), m/s
1	45.27	(42.68, 47.32)	49.07	(43.05, 52.79)
2	46.10	(43.58, 48.60)	51.11	(45.66, 56.06)
3	47.11	(44.27, 50.27)	51.19	(45.16, 57.28)
4	47.28	(44.35, 50.21)	50.86	(46.44, 55.99)
5	47.30	(44.35, 50.00)	51.13	(45.52, 55.96)

The new ACER method offers a unique approach to the prediction of extreme wind speeds. The converged empirical ACER functions provide an estimate of the exact extreme value distribution inherent in the data. It has been observed that the ACER method seems to give consistent and accurate results compared with e.g. the POT method. The ACER method appears to be quite robust with respect to the choice of parameter values. Even if the parameter values may deviate somewhat from the optimal values, the obtained predictions are still good. Under the premise that the estimated ACER function can be used to make extreme value predictions, another advantage of the proposed ACER method over the POT method is also the lack of sensitivity to outliers.

References

- [1] Naess, A., Gaidai, O., 2009. *Estimation of extreme values from sampled time series.*, Structural Safety 31, 325-334.
- [2] Naess, A., Gaidai, O., 2008. *Monte Carlo Methods for Estimating the Extreme Response of Dynamical Systems.* Journal of Engineering Mechanics. ASCE 134(8), 628-636.

Extremal dependence and the ACE algorithm

Keith KNIGHT – *University of Toronto, Canada*

Email: keith@utstat.toronto.edu

Abstract

The maximal correlation between two random variables X and Y was defined by Gebelein in [3] as the supremum of $E[\psi(X)\phi(Y)]$ over functions ψ and ϕ satisfying $E[\psi(X)] = E[\phi(Y)] = 0$ and $E[\psi^2(X)] = E[\phi^2(Y)] = 1$. Given a sample from a joint distribution, the maximal correlation (as well as the optimal functions ψ and ϕ) can be estimated via the ACE algorithm introduced in [1]; see also [2]. In this talk, we will define a version of maximal correlation for bivariate extremal dependence, which depends on the random variables only via their copula, and discuss its estimation using an appropriate modification of the ACE algorithm.

References

- [1] L. Breiman and J. Friedman (1985) Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association.* **80**, 580–598.
- [2] A. Buja (1990) Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics.* **18**, 1032–1069.
- [3] H. Gebelein (1941) Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitschrift für Angewandte Mathematik und Mechanik.* **21**, 364–379.

Multi–element generalized Polynomial Chaos for Extreme Quantile Estimation

Jordan KO – *Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris 7, France*

Email: jordan.ko@mac.com

Co-author: Josselin GARNIER

Abstract

The estimation of extreme quantile is a challenging numerical topic and has received a considerable amount of attention in many research disciplines. In the context of this paper, we wish to evaluate the α -quantile, Y_α , of the solution of a numerical function, $f(\boldsymbol{\xi})$, whose input, $\boldsymbol{\xi}$, is a multi–dimensional random vector. Monte Carlo (MC) method and its variants are the traditional approach for quantile approximation but large number of MC samples are needed to accurately determine extreme quantiles. Therefore, it may not be practical in cases where the numerical function is computational costly to evaluate. Importance sampling (IS) reduces the numerical cost by concentrating MC samples near design points, *i.e.* input values associated with the quantiles; however,

determining the new IS distribution can be costly. Within the context of structural reliability, First-Order Reliability Method (FORM) and Second-Order Reliability Method (SORM) have been developed to estimate quantiles but the methods are accurate only for moderate quantiles and the error is difficult to estimate [1].

In the current study, we investigate the quantile estimation by multi-element generalized Polynomial Chaos (gPC) metamodels. gPC is a generalization of the original Homogeneous Chaos introduced by Wiener [2] and it is an expansion of the function solution with multivariate orthogonal polynomials, *i.e.*

$$f_r(\boldsymbol{\xi}) = \sum_{m=0}^M f_m \phi_m(\boldsymbol{\xi}),$$

where r denotes that the function is a metamodel and $\phi_m(\boldsymbol{\xi})$ is an orthogonal polynomial whose weight function is similar to the *pdf* of $\boldsymbol{\xi}$. The original Wiener polynomial chaos used the Hermite polynomials in terms of the Gaussian random variables [2]. The coefficients of expansion f_m are determined through a collocation method where the Galerkin projection of $f_r(\boldsymbol{\xi})$ with respect to $\phi_m(\boldsymbol{\xi})$ is approximated with numerical quadrature. The gPC has recently been applied to many uncertainty quantification studies [3, 4]. From the spectral gPC metamodel, different statistical measures such as mean, variance and sensitivity indices can be readily computed [5]. Function evaluation on the gPC metamodel can be considered as essentially free; thus, large number of MC samples from the metamodel can be used to estimate the Y_α , for moderate α .

As the gPC metamodels are expansions about the means of the inputs, their accuracy may worsen away from these mean values where the extreme events may occur. By increasing the quadrature accuracy, we may eventually improve accuracy of the quantile but it can be very expensive. Thus, a multi-element approach is used by combining the global metamodel with supplementary local metamodels centered at the design points. The design point can be sought by solving a minimum constraint problem where the limit-state function is the feasibility surface, $f(\boldsymbol{\xi}) = Y_\alpha$, and the objective function is the Euclidean norm of $\boldsymbol{\xi}$. It is solved with the Lagrange multiplier algorithm where the gPC metamodel is used in the limit-state function.

Supplemental local refinement metamodels are constructed in a bounded domain centered on the design point. To improve the accuracy and to minimize the sampling cost, sparse-tensor and anisotropic quadratures are tested in addition to the full-tensor Gauss quadrature in the computation of local metamodel. The global and local metamodels are combined in the multi-element gPC (MEgPC) approach for extreme quantile estimation. After the MEgPC quantile estimation approach is validated with some test functions, it is applied to estimation extreme quantiles in the context of a flooding model, a General Circulation Model for precipitation prediction and a turbulence model. It is shown that MEgPC can be more accurate than MC or IS for extreme quantile estimations for input dimensions less than $N = 5$ to $N = 7$.

References

- [1] R. Rackwitz, *Reliability analysis - a review and some perspectives*, Structural Safety 23 (2001) 365395.
- [2] N. Wiener, *The homogeneous chaos*, American Journal of Mathematics 60 (1938) 897936.
- [3] R. Ghanem, P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, Berlin, 1991.
- [4] D. Xiu, D. Lucor, C.-H. Su, G. Karniadakis, *Stochastic modeling of flow-structure interactions using Generalized Polynomial Chaos*, Journal of Fluids Engineering 124 (2001) 5159.
- [5] J. Ko, D. Lucor, P. Sagaut, *Sensitivity of two-dimensional spatially developing mixing layers with respect to uncertain inflow conditions*, Physics of Fluids 20 (2008) 077102.

Estimation of High Conditional Quantiles

Deyuan LI – *Fudan University, China*

Email: deyuanli@fudan.edu.cn

Co-authors: Xuming HE and Huixia WANG

Abstract

Estimation of conditional quantiles at very high or low tails is of interest in numerous applications such as climate studies, finance and economics, medical cost studies, studies of infant birthweights. Quantile regression provides a convenient and direct tool to assess the impact of covariates at different tails of the response distribution. However, due to lack of information in the tail areas, estimations from quantile regression are often not precise at tails especially for heavy-tailed distributions. In this paper, we develop two new estimation methods for high conditional quantiles based on the extreme value theory. The new methods operate by first estimating the intermediate conditional quantiles using quantile regression, and then extrapolating these estimates to the high tails using different assumptions on the tail behavior. We rigorously establish the asymptotic properties of the proposed conditional quantile estimators and the associated tail index estimators. Compared to the conventional

quantile regression, the proposed extrapolation methods demonstrate higher accuracy in a simulation study, and they lead to more insightful estimation of high conditional quantiles of precipitation in a downscaling analysis of daily precipitation in Chicago urban area.

References

- [1] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.* **25**, 187-210.
- [2] de Haan, L and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer.
- [3] Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91, 74-89.
- [4] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.
- [5] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36, 1108-1126.

Asymptotic Approximations of Integrals of Gaussian Random Fields

Jingchen LIU – *Columbia University*

Email: jcliu@stat.columbia.edu

Abstract

In this talk, we present asymptotic approximations of $P(\int_T e^{f(t)} dt > u)$ as $u \rightarrow \infty$ for a smooth Gaussian random field, f , living on a d -dimensional compact set T . The results for the homogeneous cases are provided by [1]. Further results when f has a spatially varying mean are given in [2]. From a theoretical point of view, this analysis is a generalization of the study of the probability $P(\sup_T f(t) > u)$. From an application point view, the integral of exponent of Gaussian random field is an important random variable for many generic models in spatial point processes, portfolio risk analysis, asset pricing, material failure theory, and so forth.

The analysis technique consists of two steps: 1. evaluate the tail probability $P(\int_{\Xi} e^{f(t)} dt > u)$ over a small domain Ξ depending on u , where $mes(\Xi) \rightarrow 0$ as $b \rightarrow \infty$ and $mes(\cdot)$ is the Lebesgue measure; 2. with Ξ appropriately chosen, we show that $P(\int_T e^{f(t)} dt > b) = (1 + o(1))mes(T)mes^{-1}(\Xi)P(\int_{\Xi} e^{f(t)} dt > u)$.

References

- [1] Jingchen Liu. Tail Approximations of Integrals of Gaussian Random Fields. *Annals of Probability*. Accepted.
- [2] Jingchen Liu and Gongjun Xu. Some Asymptotic Results of Gaussian Random Fields with Varying Mean Functions and the Associated Processes. Preprint.

Asymptotic portfolio risk in multivariate regularly varying models

Georg MAINIK – *ETH Zürich, Switzerland*

Email: georg.mainik@math.ethz.ch

Abstract

This talk comprises a series of results on the extremal behaviour of portfolio losses in multivariate regularly varying models. Asymptotic distributions of extreme portfolio losses are characterized by a functional of the portfolio weights and the intrinsic parameters of the multivariate regular variation, given by the tail index and the spectral measure.

Existence, uniqueness, and location of the optimal portfolio are analysed and applied to risk minimization. The analysis of the optimization problem shows that the asymptotic portfolio risk is a convex function of the portfolio vector if the tail index is higher than or equal to 1. On the other hand, if the components of the multivariate regularly varying vector are non-negative and the tail index is lower than 1, the asymptotic portfolio risk is concave.

Estimation of the asymptotic portfolio risk factor in the i.i.d. setting is approached by a semiparametric method. Strong consistency and asymptotic normality are established as a functional law of large numbers and a functional central limit theorem. The estimated portfolio risk factor converges uniformly on compact portfolio sets.

To compare the diversification effect across different models, the notion of asymptotic portfolio loss ordering is introduced. The initial definition is stated in terms of uniform ordering of asymptotic portfolio losses for all portfolios with non-negative weights. Under the assumption of multivariate regular variation this ordering has equivalent criteria in terms of marginal distributions and spectral measures. These results are applied to copula

models and to the characterization of the worst and the best dependence structures for the diversification of multivariate regularly varying risks.

References

- [1] G. Mainik and L. Rüschendorf, *On optimal portfolio diversification with respect to extreme risks*, Finance and Stochastics, 2010.
- [2] G. Mainik, *Estimating asymptotic dependence functionals in multivariate regularly varying models*, preprint, 2011
- [3] G. Mainik and L. Rüschendorf, *Ordering of multivariate risk models with respect to extreme portfolio losses*, preprint, 2011

Estimating bivariate tails, a copula based approach

Véronique MAUME-DESCHAMPS – *Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 69366, France*
Email: veronique.maume@univ-lyon1.fr

Co-authors: Elena DI BERNARDINO and Clémentine PRIEUR.

Abstract

We are interested in the problem of fitting the joint distribution of bivariate observations exceeding high thresholds. To this end, we develop a two-dimensional extension of the POT method, mainly based on a version of the Pickands-Balkema-de Haan Theorem in dimension 2. This extension allows to consider a two-dimensional structure of dependence between both continuous random components X and Y . The modelling of this dependence is done via a copula C , which is supposed to be unknown. Asymptotic dependence as well as asymptotic independence are considered.

Starting from the work by Juri and Wüthrich ([4]) and Charpentier and Juri ([2]), we construct a two-dimensional tail estimator and study its asymptotic properties. A parameter that describes the nature of the tail dependence is also introduced and estimated. We also present real data examples which illustrate our theoretical results.

Modeling the dependence by using copulas leads to exploit the one dimensional generalized Pareto distribution of the excesses of the marginal laws. Using directly the multivariate generalized Pareto distribution developed by Falk and Reiss ([3]) and Rootzen and Tajvidi ([5]) could be another attempt to estimating the bivariate tail distribution. Nevertheless, estimation of the scaling parameters of the multivariate generalized Pareto distribution have to be addressed first. The model proposed by Ledford and Tawn provides a way to take into account the dependence structure in the tail (see [1] for a recent work). We propose an alternative model based on regularity conditions of the copula and on the explicit description and estimation of the dependence structure in the joint tail.

References

- [1] Beirlant J., Dierckx G. and Guillou A., *Reduced bias estimators for bivariate tail modelling* Insurance: Mathematics and Economics, in press (2011).
- [2] Charpentier A. and Juri A. *Limiting dependence structures for tail events, with applications to credit derivatives.*, J. Appl. Probab., **43**, 2, 563–586, (2006).
- [3] Falk, M. and Reiss, R.-D., *On the distribution of Pickands coordinates in bivariate EV and GP models*, J. Multivariate Anal., **93**, 2, 267–295, (2005).
- [4] Juri A. and Wüthrich M.V., *Tail Dependence From a Distributional Point of View*, *Extremes*, **3**, 213-246, (2003).
- [5] Rootzén H. , Tajvidi N., *Multivariate generalized Pareto distributions*, Bernoulli, **12**, 5, 917–930, (2006).

POT methods: a new insight into the estimation of extreme value distributions

Franck MAZAS – *Sogreah Maritime, ARTELIA Group, France*.

Email: Franck.Mazas@arteliagroup.com

Co-authors: Philippe GARAT and Luc HAMM

Abstract

For a decade, a broad consensus has emerged in favor of a general methodology for determining extreme values of environmental data. Roughly speaking, this methodology consists in the five following steps. First, an independent and identically distributed (i.i.d.) sample is extracted from the raw time series. For this purpose, the time series is homogenized in order to identify the different populations it may contain, then declustering is applied with a Peaks-Over-Threshold (POT) approach to extract the independent extreme data from these homogeneous sub-samples. Second, a statistically meaningful threshold is set, usually by theoretical properties of the Generalized Pareto Distribution (GPD). Third, a two-parameter GPD is fit to the threshold excesses using

the Maximum Likelihood Estimator (MLE). Fourth, quantiles (return values) are calculated for the return periods of interest (assuming the number of events per year is a Poisson process). Fifth, confidence intervals are computed.

This methodology, often referred to as the GPD-Poisson model, has sound theoretical justifications (see for instance [1]). The GPD is the asymptotic law for the threshold excesses. As for the MLE, it is particularly valued by statisticians for its asymptotic properties: consistency, efficiency, asymptotic normality, and hence is widely used among analysts.

However, we wish to discuss in this communication two questionable points of this methodology: one concerning the general form of the GPD-Poisson model itself, the other one concerning the estimation techniques.

First, one should keep in mind that the GPD is an asymptotic law, meaning that it is an approximation of the true law of excesses over a threshold valid only when this threshold is high enough. Consequently, if the asymptotic domain is not reached, other distributions may fit the data better. As presented in [2], we extend the univariate and stationary GPD-Poisson model by fitting two additional distributions (namely the Weibull and Gamma distributions), though others could of course be tested. Objective criteria are then used to select the best-fitting law.

Such a multi-distribution approach is useful for analysts and/or engineers who wish to cover a wide range of situations and is justified since the GPD is outperformed in several practical cases.

Second, MLE exhibits strange behavior when the threshold is allowed to vary between two consecutive data values [3]: the estimated parameters of the distribution (scale and shape) and consequently the estimated quantiles do not remain constant. Thus, a slight translation of the sample leads to a significant change in the estimation, although the sample size remains identical. This phenomenon is particularly noteworthy for the Weibull and Gamma distributions, but also exists for the GPD. For the former distributions, the likelihood tends to infinity in many cases when the threshold tends to the open upper bound of its interval of validity, i.e. the first data. For the latter, the maximum of the likelihood is reached at this open upper bound, but with a non-nil derivate. Still, it must be kept in mind that the asymptotic properties of the MLE are valid only when the maximum is reached on an interior point of an open set [4].

Furthermore, resampling techniques from real extreme data show that when MLE is applied to 2-parameter GPD, the empirical density (histogram) of estimated parameters exhibits 2 or 3 peaks, resulting in many peaks in the density of quantiles.

Actually, the widely spread 2-parameter GPD-Poisson model makes a confusion between the threshold, whose role should be limited to data selection/censorship, and the location parameter, whose role is to accurately set the origin of the distributions. Hence it is necessary to fit 3-parameter distributions and MLE must be rejected since it is not valid in many practical cases. We use the L-moments estimator [5] which allows a proper estimation of the three parameters for the studied distributions. This estimator is unbiased, although its variance is still significant.

In conclusion: i) a multi-distribution approach to POT methods is recommended since in practical applications it is unsure whether the asymptotic domain for the GPD is reached; ii) we strongly favor including a location parameter, clearly distinct from the threshold, in the distributions for extreme values, and iii) MLE should not be used without a close examination of its validity conditions. Our tests with real environmental data (extreme wave heights for the design of coastal structures) show that in this case, MLE should be rejected whereas the 3-parameter L-moments estimator behaves well.

References

- [1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001.
- [2] F. Mazas, L. Hamm, A multi-distribution approach of POT methods for determining extreme wave heights, *Coastal Engineering*, **58**, 385-394, 2011.
- [3] F. Mazas, P. Garat, L. Hamm, Questioning MLE: a new insight into the estimation of extreme wave distributions. Submitted to *Coastal Engineering*, 2011.
- [4] E.L. Lehmann, *Theory of point estimation*, Wiley & Sons, Australia, 1983.
- [5] J.R.M. Hosking, L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, *J. R. Statist. Soc.*, **52**(1), 105-124, 1990.

Limit theorems for coupon collector's problem and generalized Pareto distributions

Pavle MLADENVIĆ – *University of Belgrade, Serbia*

Email: paja@matf.bg.ac.rs

Co-author: Jelena JOCKOVIĆ

Abstract

Limit theorems for waiting times that arise in coupon collector's problem will be presented. New results that give connection with generalized Pareto distributions are also included.

References

- [1] J. Jocković and P. Mladenović, Coupon collector's problem and generalized Pareto distributions, *Journal of Statistical Planning and Inference*, **141**, 2348-2352, 2011.
- [2] P. Mladenović and J. Vukmirović, Rates of convergence in certain limit theorems for extreme values, *Journal of Mathematical Analysis and Applications*, **363**, 287-295, 2010.
- [3] P. Mladenović, Limit distributions for the problem of collecting pairs, *Bernoulli*, **14**(2), 419-439, 2008.

Estimation of parameters of extreme events for randomly censored data

M. Manuela NEVES – CEAUL, University of Lisbon and ISA, Technical University of Lisbon, Portugal

Email: manela@isa.utl.pt

Co-author: M. Ivette GOMES

Abstract

In *Statistics of Extremes* we deal essentially with the estimation of parameters of extreme or even rare events. The most common assumption on any set of univariate data, (X_1, X_2, \dots, X_n) , is to consider them as a *complete* sample of size n , with observations either independent and identically distributed or weakly dependent and stationary, from an unknown distribution function $F = F_X$. There is a large variety of parameters of extreme events, but in all applications of extreme value theory (EVT), the estimation of the *extreme value index* (EVI), denoted γ_X , is of primordial importance and the basis for the estimation of all other parameters of extreme events. Among the most relevant parameters of extreme events, and assuming that we are interested in large values, i.e., in the right tail of the underlying model F , we mention:

- the *probability of exceedance* of a high level $x \equiv x_H$, $p_x := \mathbb{P}(X > x) = 1 - F(x)$,
- the *return period* of a high level x , which is given by $r_x := 1/(1 - F(x))$, in an i.i.d. scheme,
- the *right endpoint* of an underlying model F , $x^* \equiv x^F := \sup\{x : F(x) < 1\}$, and
- a *high quantile* of probability $1 - p$, p small, situated in the border or even beyond the range of the available data, defined as $\chi_{1-p} := \inf\{x : F(x) \geq 1 - p\} =: F^{\leftarrow}(1 - p)$, $p < 1/n$.

But in many real situations, censored observations can occur. For example,

- in the analysis of lifetime data or reliability data,
- in the analysis of some physical phenomena such as wind speeds, earthquake intensities or floods, where extreme measurements are sometimes not available due to damage in the instruments.

We shall give here special attention to the estimation of the extreme value index γ_X , under *random censorship*, where apart from two recent papers by Einmahl *et al.* (2008) and Gomes and Neves (2010), there is only, as far as we know, a brief reference to the topic in Reiss and Thomas (1997, Section 6.1) and a paper by Beirlant *et al.* (2007). We first provide a few details on the EVI and max-domains of attraction. Next, we introduce a set of semi-parametric EVI estimators, valid for complete samples, providing some details on their asymptotic non-degenerate behaviour, and we illustrate the effect of random censorship on the EVI of the potential, non-available sample $\underline{\mathbf{X}} = (X_1, \dots, X_n)$. This is done in order to motivate the functional expression of the EVI estimators for randomly censored data. Indeed, if we are under a random censor scheme, any of the common EVI estimators needs to be slightly modified in order to be consistent. We shall also present the results of a small-scale Monte-Carlo simulation devised to obtain the behaviour of the EVI-estimators under study. Finally, we illustrate the behaviour of the same EVI-estimators for a few sets of survival data, available in Klein and Moeschberger (2005).

References

- [1] Beirlant, J., Guillou, A., Dierckx, G. and Fils-Viletard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* **10**:3, 151-174.
- [2] Einmahl, J.H.J., Fils-Viletard, A. and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli* **14**:1, 207-227.

- [3] Gomes, M.I. and Neves, M.M. (2010). A note on statistics of extremes for censoring schemes on a heavy right tail. In Luzar-Stiffler, V., Jarec, I. and Bekic, Z. (eds.), *Proceedings of ITI 2010*, SRCE Univ. Computing Centre Editions, 539-544.
- [4] Klein, J.P. and Moeschberger, M.L. (2005). *Datasets for Survival Analysis – Techniques for Censored and Truncated Data*, Second Edition, Springer.
- [5] Reiss, R.-D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other Fields*, Birkhäuser Verlag, Basel.

Dense classes of multivariate max stable distributions

John P. NOLAN – *American University, Washington, DC, USA*

Email: jpnolan@american.edu

Co-authors: Anne-Laure FOUGÈRES and Cécile MERCADIER

Abstract

Our goal is to describe flexible max-stable models in high dimensions. The first class is the max stable distributions with discrete angular measures. It is tractable in all dimensions, with simple exact methods to simulate and compute the distribution function. The second class is max stable models with piecewise polynomial density for the angular measure. Currently, these are only accessible in two dimensions. A third class of models is the family of generalized asymmetric logistic models. We detail the connection between this class of models and the closely related class of generalized stable mixtures. It is shown that all three classes are dense in the space of max stable models.

References

- [1] Fougères, A.-L., Nolan, J. P. and Rootzén, H., Models for dependent extremes using stable mixtures, *Scand. J. Stat.*, **36**, pp. 42–59, 2009.
- [2] J.P. Nolan, *Metrics for multivariate stable distributions*, To appear in Banach Center Publications, 2011.

Asymptotic dependence for light-tailed homothetic densities

Natalia NOLDE – *University of British Columbia, Canada*

Email: natalia@stat.ubc.ca

Co-author: Guus Balkema

Abstract

In many real life applications, dependence between coordinate extremes is of great importance for risk assessment. Hence, it is often of interest to know if the components of a given multivariate random vector exhibit asymptotic independence or asymptotic dependence, and in the latter case give its characterization. Since in the multivariate setting it is more common to have an explicit form of the density rather than the distribution function, it is convenient to have criteria for asymptotic (in)dependence in terms of the density. So we assume that the stochastic behaviour of the data may be described by a multivariate probability density, and in addition we restrict our attention here to homothetic densities with light tails. The study of the relation between the (asymptotic) shape of the level sets of densities and the asymptotic dependence properties of the underlying light-tailed distributions has been initiated in [1]. The latter paper provides a sufficient condition on the limiting shape of the level sets of a light-tailed density for asymptotic independence to hold. We extend this result to state necessary and sufficient conditions for asymptotic independence. We then look into a more delicate problem of characterizing the (asymptotic) shape of the level sets of the density when the components are asymptotically dependent. Unlike in the case of asymptotic independence, the shape of the level sets alone does not suffice to determine the presence of asymptotic dependence, and additional conditions need to be imposed.

References

- [1] A.A. Balkema and N. Nolde, Asymptotic independence for unimodal densities, *Adv. Appl. Prob.* **42**, 411-432, 2010.

Conditional sampling for mixed moving maxima on the real axis

Marco OESTING – *Georg-August-Universität Göttingen, Germany*

Email: oesting@math.uni-goettingen.de

Co-author: Martin SCHLATHER

Abstract

Max-stable processes are frequently used for modelling extreme events with spatial or temporal dependencies. In this context the question of conditional sampling and prediction turns out to be a challenging problem. Recently, an exact algorithm for the max-linear model

$$X_i = \max_j a_{i,j} Z_j, \quad Z_j \text{ i.i.d. Fréchet,}$$

has been introduced [5].

Here, we consider the class of stationary max-stable processes which allow for a mixed moving maxima representation, see [3], i.e.

$$Z(t) = \max_{(s,u) \in \Pi} u F_{s,u}(t-s), \quad t \in \mathbb{R}^d,$$

where Π is a Poisson point process on $\mathbb{R}^d \times (0, \infty)$ with intensity measure $ds \times u^{-2} du$, and $F_{s,u}$ are independent copies of a random measurable “shape function” $F : \mathbb{R}^d \rightarrow (0, \infty)$.

We develop an exact procedure for conditional sampling using of the Poisson point process structure. Since explicit calculations turn out to be quite sophisticated, we restrict ourselves to the case where $d = 1$ and use a finite number of shape functions satisfying some regularity conditions. For more general shape functions approximation techniques using MCMC methods are presented.

Our algorithm is applied to the Gaussian extreme value process where the shape function is the Gaussian density ([4]) and, secondly, to a Brown-Resnick process with a mixed moving maxima representation [1], [2]. Finally, we compare our computational results to other approaches including the algorithm of [5].

References

- [1] B. M. Brown and S. I. Resnick, Extreme Values of Independent Stochastic Processes, *J. Appl. Probab.*, 14(4): 732–739, 1977.
- [2] S. Engelke, Z. Kabluchko and M. Schlather, An equivalent representation of the Brown-Resnick process, *Statist. Probab. Lett.*, To appear, 2011.
- [3] M. Schlather, Models for Stationary Max-Stable Random Fields, *Extremes*, 5(1):33–44, 2002.
- [4] R. L. Smith, Max-stable processes and spatial extremes, Unpublished manuscript, 1990.
- [5] Y. Wang and S. A. Stoev, Conditional sampling for spectrally discrete max-stable random fields, *ArXiv e-prints*, 1005.0312, 2010.

New inferential tools for max-stable processes

Thomas OPITZ – *University Montpellier 2, France*

Email: Thomas.Opitz@math.univ-montp2.fr

Co-authors: Jean-Noël BACRO and Pierre RIBEREAU

Abstract

Modelling the extreme behaviour of random spatial phenomena is indispensable for the management of risks related to climate and environment. Phenomena of interest include precipitation, snowfall, storm tides/surges, wind and temperatures. Max-stable fields arise as limits of adequately normalised iid sequences of spatial processes and thus are obvious candidats for the modelisation of spatial dependence between extreme events. We recall some of their properties, the current principal inferential methods and some common models, see for example [1], [2] and [3]. We propose new inferential approaches to characterize the extremal dependence structure, based on the so-called spectral measure of the bivariate distributions for site pairs. In particular, a novel graphical tool and a threshold-based (composite) likelihood approach are presented.

References

- [1] S.A. Padoan, M. Ribatet and S.A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105:263-277, 2010.
- [2] Z. Kabluchko, M. Schlather and L. de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37:2042-2065, 2009.
- [3] R.L. Smith. Max-stable processes and spatial extremes. *Unpublished Manuscript*, 1990.

A diffusion model for the simulation of observed temperature series

Sylvie PAREY – *EDF R&D, France*

Email: sylvie.parey@edf.fr

Co-authors: Thi Tu HUONG HOANG and Didier DACUNHA-CASTELLE

Abstract

Stochastic simulation provides a tool to study the occurrence of hazards potentially penalizing for industrial activities. Extreme low or extreme high temperatures are one of those hazards regarding electricity production and consumption. A methodology is thus proposed to simulate observed daily minimum or maximum temperature series. In a first step, the observed temperature is analyzed in order to remove low frequency processes, like trends and seasonality. The non parametric LOESS technique is used to derive trends in mean and variance, and a modified cross-validation technique is proposed to select the smooth parameter. Then, seasonality of the mean and variance of the detrended reduced variable is computed as a trigonometric polynomial and removed. The remaining process is checked regarding possible remaining trends or seasonality in different moments, as well as in its extremes. Finally, a diffusion process is designed to simulate it in the best way, especially regarding extreme events. Examples will be given for temperature series in Europe or in the United States.

References

- [1] Dacunha-Castelle D. and Florens-Zmirou D.: Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19 :263-284,1986.
- [2] Hoang T.T.H: Modélisation de séries chronologiques non stationnaires, non linéaires: application à la définition des tendances sur la moyenne, la variabilité et les extrêmes de la température de l'air en Europe. *Thèse de Doctorat de l'Université Paris 11 Orsay*, 2010.
- [3] Hoang T.T.H., Parey S., and Dacunha-Castelle D.: Multidimensional trends : the example of temperature. *European Physical Journal-Special Topics*, 174 :113-124, 2009.
- [4] Klein Tank, A.M.G., J.B. Wijngaard, G.P. Können, R. Bhm, G. Demare, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pilsdtir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Miletus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio Lopez, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, and P. Petrovic. Daily datasets of 20th-century surface air temperature and precipitation series for the European climate assessment. *Int. J. of Clim.*, 22 :1441-1453, 2002

Modelling extreme values of processes observed at irregular time step

Nicolas RAILLARD – *University Brest, France*

Email: nicolas.raillard@univ-brest.fr

Co-authors: Pierre AILLIOT and Jian-Feng YAO

Abstract

When modelling the extremal properties of a given phenomenon is needed, two main approaches are under statistical interest: modelling the maxima over a large block of consecutive observations or modelling the excesses over a high threshold.

The first approach, denoted as the 'block maxima' method relies on results that describe the acceptable distributions for the maximum of a random variable. The main drawback of this approach is the waste of data induced by taking the maximum over a large block (e.g. a year). Hence another approach, denoted **POT** for Peaks Over Threshold, consists in taking into account any value that is high enough to lie in the tail of the distribution, i.e that exceeds a sufficiently high threshold.

A specific problem related to the **POT** approach is that consecutive data are dependent. Indeed, every data that exceeds a threshold has to be taken into account, but it is very likely that clusters of consecutive excess are being observed, even more if the time-lag between observations is smaller than the characteristic duration of an extremal event. This is a real problem since the theory only deals with independent observations and forgetting this dependence leads to under- or over-estimation of the extremes and despite its importance, this fact is often forgotten in applications. To deal with this dependence, two main approaches are now well-developed: declustering and markov-chains models. The first scheme is based on a probabilistic result stated by Leadbetter and well explained by Coles [1], Chapter 5: the extremes of a dependant stationary sequence show clusters of consecutive large values, and there exists a parameter, named the **extremal index** which summarize this dependence and is found to be, under some assumptions, the reciprocal of the mean cluster size. This result induces a filtering of the dependent observations to obtain nearly independent excesses: once the clusters are identified, the parameters can be estimated thanks to the maxima over the clusters and the clusters length.

As seen above, two specific problems arise: first, a definition of a cluster of consecutive exceedances is needed, and this scheme leads to a waste of data, since only maxima within each clusters are used to estimate the parameters. Hence another approach is to keep all excess but with the introduction of an appropriate dependence structure. This has been used with a **first-order Markov-chain** by Smith [4, 5] with tools to

compute several features of the fitted model. More recently, Ribatet [3] applied this approach to several flood characteristic along with diagnostic tools.

The present study takes place within the latter framework. The idea guiding this work was the modelling of extremes of the significant wave height along satellite tracks: indeed, those data exhibits a complex structure, both in space and time, in addition to the presence of missing values along a track. It may thus be quite difficult to apply the usual declustering since identifying clusters when there is missing values might be a hard task. The need for extension come from the irregular time sampling on the data: the complex repartition of the satellite induce a complex recurrence time of passage at a given location, hence there is no natural time-lag between consecutive points, and a markovian structure is quite inadequate. So extensions of the aforementioned method to this irregular sampling in time will be under study in this talk. The main idea is to approximate the data above a threshold by a **censored max-stable process**, whereas previous works used a bivariate extreme value distribution. Based on work of Padoan [2], an estimation procedure is provided thanks to a composite likelihood method. Performance of this new estimator will be assessed both on synthetic data and on significant wave height data that comes from buoy and from numerical models.

References

- [1] Stuart Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.
- [2] S. A. Padoan, M. Ribatet, and S. A. Sisson. *Likelihood-based inference for max-stable processes*. Journal of the American Statistical Association, 105(489):263-277, 2010.
- [3] M. Ribatet, T. B. M. J. Ouarda, E. Sauquet, and J. M. Gresillon. *Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics*. Water Resources Research, 45(3), March 2009.
- [4] Richard L. Smith. *The extremal index for a markov chain*. J. Appl. Probab., 29(1):37-45, 1992.
- [5] Richard L. Smith, Jonathan A. Tawn, and Stuart G. Coles. *Markov chain models for threshold exceedances*. Biometrika, 84(2):249-268, 1997.

Conditioning Exceedances Processes on Covariate Information

Rolf-Dieter REISS – *University Siegen, Germany*

Email: rolf.reiss@gmx.net

Co-author: Ulf CORMANN

Abstract

This talk concerns the statistical inference for the upper tail of the conditional distribution of a response variable Y given a covariate $X = x$ based on n random vectors within the parametric extreme value framework. Pioneering work in this field was done by Smith [2] and Smith and Shively [3]. We propose to base the inference on a conditional distribution of the point process of exceedances given the point process of covariates. It is of importance that the conditional distribution merely depends on the conditional distribution of the response variable given the covariates. In the special case of Poisson processes such a result may be found in a book by Reiss [1]. Our results are valid within the broader model where the response variables are conditionally independent given the covariates.

References

- [1] Reiss, R.-D., *A Course on Point Processes*. Springer, New York (1993).
- [2] Smith, R.L., *Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone*. Statistical Science 4, 367-393 (1989).
- [3] Smith, R.L. and Shively, T.S., *Point process approach to modeling trends in tropospheric ozone based on exceedances of high threshold*. Atmospheric Environment 29, 3489-3499 (1995).

On the comparison of methods for predicting extremes: a data-based methodological framework.

Benjamin RENARD – *Cemagref, UR HHLY Hydrology-Hydraulics, Lyon, France*

Email: benjamin.renard@cemagref.fr

Co-authors: Krzysztof KOCHANEK, Michel LANG, Federico GARAVAGLIA, Emmanuel PAQUET, Luc NEPPEL, Kenza NAJIB, Julie CARREAU, Patrick ARNAUD, Yoann AUBERT, Francois BORCHI, Jean-Michel SOUBEYROUX, Sylvie JOURDAIN, Jean-Michel VEYSSEIRE, Eric SAUQUET, Annick AUFRAY

Abstract

Extreme value analysis is one of the cornerstones of hazard quantification and risk assessment. Its basic objective is to estimate the distribution of some environmental variable X , e.g. annual maximum of the areal rainfall over some catchment, annual maximum flood, etc. This distribution can be used to estimate the exceedance probability of a given value of X (often expressed in terms of return period), or alternatively, to estimate the p -quantile of X . The estimation of quantiles is of primary importance since they are used to design civil engineering structures (e.g. dams, reservoirs, bridges) or to map hazard-prone areas where restrictions may be enforced (e.g. building restrictions in flood zones).

Extreme value analysis has been the subject of extensive research, yielding an abundance of approaches. In Hydrology, several families of methods exist, including (but not limited to):

- Standard application of extreme value theory (EVT), i.e. estimation of an extreme value distribution based on a sample of block maxima or peaks over a high threshold
- Climate/Weather-informed application of EVT. This family of methods uses additional meteorological (e.g., weather type, [1]) or climatic (e.g. Interdecadal Pacific Oscillation IPO, [2]) information.
- Regional approaches, conjointly using data from several sites to perform the inference, which may improve the precision of estimates.
- Model-based approaches, using a simulation model reproducing the main characteristics of the environmental variable (e.g. [1]).

In practice, users and practitioners of extreme value analyses may feel lost facing such an abundance of methods. Consequently, it is necessary to provide them with practical guidelines to choose and implement adequate methods, depending on the conditions of application (e.g. availability of long series, geographical area, type of hydrological regime, etc.).

This presentation describes a methodological framework to perform a data-based comparison of competing approaches for predicting extremes. This framework is based on the following principles:

- The objective is to assess the *predictive* performance of competing methods (as opposed to standard goodness-of-fit evaluations). This requires decomposing the available dataset into estimation / validation sub-samples.
- Reliably quantifying uncertainties is recognized as a primary objective, and the issue of scrutinizing uncertainty estimates is discussed. To this aim, we make use of predictive distributions for extremes, obtained by integrating out parameter uncertainty. Such predictive distributions are standard in a Bayesian context [4] but can also be derived in a frequentist context [5].
- Reliability indices are derived in order to compare the performances of competing methods on an objective basis.

In a second step, this framework is used to perform a thorough comparison between approaches currently used in France for extreme prediction. The comparison is based on an extensive dataset of long series of rainfall and runoff (about 40-50 years of daily data), available for hundreds of sites over France. Results demonstrate the ability of the comparison framework to distinguish between "good" and "bad" approaches, and yield valuable insights into the optimal ambit of each approach.

References

- [1] F. Garavaglia, M. Lang, E. Paquet, J. Gailhard, R. Garcon, and B. Renard, *Robustness and reliability of a rainfall compound distribution model based on weather pattern sub-sampling*, Hydrology and Earth System Sciences, 15, 519-532, 2011.
- [2] T. Micevski, S. W. Franks, and G. Kuczera, *Multidecadal variability in coastal eastern Australian flood data*, J. Hydrol., 327(1-2), 219-225, 2006.
- [3] P. Arnaud, and J. Lavabre, *Using a stochastic model for generating hourly hyetographs to study extreme rainfalls*, Hydrol. Sci. J., 44(3), 433-446, 1999.
- [4] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, 2 ed., 696 pp., Chapman & Hall, 2004.
- [5] I. R. Harris, *Predictive Fit for Natural Exponential-Families*, Biometrika, 76(4), 675-684, 1989.

Tail estimation methods for the number of false positives in high-throughput testing

Holger ROOTZÉN – *Chalmers, Sweden*

Email: hrootzen@chalmers.se

Co-author: Dmitrii Zholid

Abstract

This paper develops methods to handle false rejections in high-throughput screening experiments. The setting is very highly multiple testing problems where testing is done at extreme significance levels and with low degrees of freedom, and where the true null distribution may differ from the theoretical one. We show that the conditional distribution of the number of false positives, given that there is in all r positives, approximately has a binomial distribution, and develop efficient and accurate methods to estimate its success probability parameter. Furthermore we provide efficient and accurate methods for estimation of the true null distribution resulting from a preprocessing method, and techniques to compare it with the theoretical null distribution. Extreme Value Statistics provides the natural analysis tools, a simple polynomial model for the tail of the distribution of p-values. We provide asymptotics which motivate this model, exhibit properties of estimators of the parameters of the model, and point to model checking tools, both for independent data and for dependent data. The methods are tried out on two large scale genomic studies and on an fMRI brain scan experiment. A software implementation, SmartTail, may be downloaded from the web.

References

- [1] K. F. Kerr. Comments on the analysis of unbalanced microarray data. *Bioinformatics* **25**(16):20352041, 2009.
- [2] T. A. Knijnenburg, L. F. A. Wessels, J. T. M. Reinders, and I. Shmulevich. Fewer permutations, more accurate P-values. *Bioinformatics* **25** (12):161168, 2009.

Ergodic Properties of Stable Random Fields via Group Actions

Parthanil ROY – *Michigan State University, USA*

Email: parthanil.roy@gmail.com

Co-authors: Stilian A. STOEV and Yizao WANG

Abstract

Ergodic properties of stochastic processes play a very important role in the investigation of their extremal behaviors (e.g., estimation of ruin probability, large deviations, etc.). In this work, we establish characterization results for ergodicity of symmetric α -stable (S α S) stationary random fields. We first show that the result of Samorodnitsky [1] remains valid in the multiparameter setting, i.e., a stationary S α S ($0 < \alpha < 2$) random field is ergodic (or equivalently, weakly mixing) if and only if it is generated by a null group action. We also give a criterion for ergodicity of these random fields which is valid for all dimensions and new even in the one-dimensional case. The similarity of the spectral representations for sum- and max-stable random fields yields parallel characterization results in the max-stable setting.

References

- [1] G. Samorodnitsky. *Null flows, positive flows and the structure of stationary symmetric stable processes*, Annals of Probability, 33:1782–1803, 2005.

Estimating an endpoint using high order moments

Gilles STUPFLER – *Université de Strasbourg, France*

Email: stupfler@math.unistra.fr

Co-authors: Stéphane GIRARD and Armelle GUILLOU

Abstract

In 2008, Girard and Jacob [1] presented a new method to estimate the frontier of a multidimensional sample. This method relies on a kernel regression on high order moments of the data, and was further used by the same authors in [2] to develop local polynomial estimators of a frontier. Using this approach, we present a new way to estimate the endpoint of a unidimensional sample when the distribution function belongs to the Weibull domain of attraction. The estimator is based on computing high order empirical moments of the variable of interest. Contrary to most popular methods, this one is not threshold-based. It is assumed that the order of the moments goes to infinity, and provided a second-order assumption holds, we give conditions on its rate of divergence to get the asymptotic normality of the estimator. The good performance of the estimator is illustrated on some finite sample situations.

References

- [1] Girard, S., Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data, *Journal of Multivariate Analysis*, **99**(3):403–420.
- [2] Girard, S., Jacob, P. (2009). Frontier estimation with local polynomials and high power-transformed data, *Journal of Multivariate Analysis*, **100**(8):1691–1705.

Ionosphere Severe Storms and Occurrence Risk Estimation

Norbert SUARD – *CNES, Toulouse, France*

Email: Norbert.Suard@cnes.fr

Co-authors: Jean-Marc AZAÏS, Sébastien GADAT and Dominique DEBAILLEUX

Abstract

Even if a lot of analyses, often based on last solar cycle data have been done to preserve the integrity of navigation systems such as Satellite Based Augmentation System (SBAS) or Ground Based Augmentation System (GBAS) in case of ionosphere severe storm threat, the in place integrity preservation means have often a drawback in the continuity availability domain. In order to be able to assess the continuity (and availability) risk that can be caused by such ionosphere severe storm, CNES has on one hand, contracted a study led by TASF and supported by IMT (Institute of Mathematics in Toulouse) to assess the occurrence risk of severe/extreme ionosphere storms taking into account the solar cycle profile and based on innovative statistical methods and has on an other hand, established a classical risk by using a min, max, average approach along a solar cycle based on several solar cycle information. In addition to that and in order to support different other activities, CNES has already analysed different historical data sets of solar activity indices and the resulting earth geomagnetism/ionosphere activities over a period covering several solar cycles. This includes the SSN, 10.7cm flux, Dst, Kp and Ap indexes, going back to as far as 1932 for those available, and an assessment of the relationships between these different indexes. Looking into the future, several predictions of the new solar cycle activity are also available, for example from the Marshall Space Flight Centre. Among all the analyses indices, the Ap index (daily or 3 hours), measuring the earth geomagnetism/ionosphere activities and the monthly SSN measuring the solar activity have been used in this ionosphere storm risk occurrence assessment work. More in details, the innovative statistical methods were based on rare event theory, firstly by a global and simple approach over the 80 years, considering a stationary system and resulting in an average risk. This approach is de facto not satisfactory due the evident dependence between geomagnetic/ionosphere activity and solar one that is a cyclic system. So, secondly a different approach, the Cox Model (Proportional Hazard Model with Time-Dependent Covariates), better known in the epidemiology and medical treatment domain has been experienced. Basically, the Cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. It allows to estimate the hazard (or risk) of death, or other event of interest, for individuals, given their prognostic variables. Our interest was driven by the proportional hazards assumption. this assumption is that covariates multiply hazard. In other terms, the hazard function for an observation depends on the values of the covariates and the value of the baseline hazard. Given two observations with particular values for the covariates, the ratio of the estimated hazards over time will be constant - hence the name of the method: the proportional hazard model as in the Cox model [1]. This gives an example on non-stationary extreme model that is different from that of [2]. The validity of this assumption may often be questionable. Nevertheless, the covariate influence create a relative risk that is constant in regards to the reference and this risk can be expressed exponentially as a function of a linear covariates combination. In our approach one of these covariates is a solar activity index. The baseline hazard is estimated by a kernel-smoothing estimator and the coefficients of the covariate by maximum likelihood. After having determine the different parameters of our Cox model and having at disposal some predictions for the new solar cycle, it is then possible to extrapolate the ionosphere severe storm occurrence risk over the new solar cycle and to introduce it in a system continuity risk assessment. The classical method was just based for each solar cycle on a counting year per year of the number of days in regards to a daily Ap ionosphere activity ranking and then to determine the min, max, average values year per year having at our disposal seven complete solar cycles in the analysed dataset.

The paper will present a brief overview of relations between solar events, earth geomagnetism and ionosphere status the data sets that were used in our analyses the different approach to determine an ionosphere severe storm risk of occurrence a focus on our Cox model with its current description, the estimation process and questions that are still open for future work and the already achieved results and comparison between them.

References

- [1] O. Aalen, O. Borgan, H. Gjessing, *Survival and event history analysis*, Springer, 2008

A Bayesian analysis of extreme precipitation in Mediterranean France using non-stationary GEV models

Xun SUN – *Cemagref Lyon, France*
Email: xun.sun@cemagref.fr

Abstract

Extreme value theory becomes increasingly important for hydrologists in order to assess the hazard related to extreme rainfall and flooding. The annual maximum daily precipitation at one site is often modelled with a GEV distribution[1].

The objective of the present study is to describe the extreme rainfall in French Mediterranean region with different covariate models[2]. 92 precipitation gauges within this region are utilized. Based on different covariates (like time, weather type and climate indices) and regression models (e.g. linearity on location parameter, scale parameter or both), several non-stationary GEV models are constructed. The Bayesian approach and the Markov chain Monte Carlo (MCMC) methods are used to infer the posterior parameter distributions[1], to assess uncertainties and to examine the performance of different probability models. Moreover, the predictive distributions are computed to make predictions on future observations.

Due to the limited information content of at-site data, the identification of at-site non-stationary models becomes challenging. The construction of a general modelling framework for regional non-stationary models will thus be discussed.

References

- [1] Coles, S.; Pericchi, L. R. and Sisson, S. *A fully probabilistic approach to extreme rainfall modeling*, Journal of Hydrology, 273, 35 - 50, 2003
- [2] Katz, R. W., Parlange, M. B. and Naveau, P. *Statistics of extremes in hydrology*, Advances in Water Resources, 25, 1287 - 1304, 2002
- [3] Renard, B.; Garreta, V. and Lang, M. *An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima* Water Resources Research, Amer Geophysical Union, 42, W12422, 2006

Extremes of locally self-similar Gaussian processes

Kamil TABIŚ – *University of Wrocław, Poland*
Email: kamil.Tabis@math.uni.wroc.pl
Co-author: Krzysztof DĘBICKI

Abstract

Pickands' double-sum method allows us to obtain exact asymptotics for the supremum distribution of Gaussian processes ([1], [2]). The classical application of this method is based on the use of stationary or locally stationary structure of the analyzed process $X(\cdot)$, i.e. the assumption that

$$\mathbf{Var}(X(t) - X(s)) = \text{Const} \cdot |t - s|^\alpha (1 + o(1)) \quad \text{as } t, s \rightarrow t^*,$$

for some $\alpha \in (0, 2]$, where t^* is a point in which variance function of $X(\cdot)$ attains its maximum.

In the talk we focus on the exact asymptotics of supremum distribution of Gaussian processes such that

$$\mathbf{Var}(X(t) - X(s)) = \text{Const} \cdot \mathbf{Var}(Y(t) - Y(s)) (1 + o(1)) \quad \text{as } t, s \rightarrow t^*,$$

where $Y(\cdot)$ is a self-similar Gaussian process (with not necessarily stationary increments).

We will point out difficulties of the application of the double-sum method to our problem and we present a new approach, that allows us to determine the exact asymptotics of $\mathbf{P}\left(\sup_{t \in [0, T]} X(t) > u\right)$ as $u \rightarrow \infty$. In the obtained asymptotics there appear new analogous of Pickands' constants with interesting properties. The theory will be illustrated by some examples.

References

- [1] J. Pickands III, (1969) *Upcrossing probabilities for stationary Gaussian processes*, *Trans. Amer. Math. Soc.* **145**, 51–73.
- [2] V. I. Piterbarg, *Asymptotic methods in the theory of Gaussian processes and fields*, Translations of Mathematical Monographs 148, AMS, Providence, 1996.
- [3] K. Dębicki, K. Tabiś, (2011) *Extremes of the time-average of stationary Gaussian processes*, In revision.

Probabilistic wind gust forecasting using non-homogeneous Gaussian regression

Thordis L. THORARINSDOTTIR – *Heidelberg University, Germany*
Email: thordis@uni-heidelberg.de
Co-author: Matthew S. JOHNSON

Abstract

Skillful probabilistic forecasts for wind speed and wind gust are called for both in specialized weather related risk estimation and in every-day situations in aviation, agriculture, and energy production. Probabilistic forecasts are usually obtained with statistical post-processing of an ensemble and recently, an ensemble post-processing method for wind speed based on non-homogeneous Gaussian regression (NGR) has been proposed [1]. However, this method does not apply directly to wind gust as gust speed forecasts are not a standard output from numerical weather prediction models. We propose a solution to this, where we combine the NGR probabilistic forecasts for wind speed with a gust factor to obtain a probabilistic forecast for wind gust at a very low additional computational cost. We apply our framework to 48-hour ahead forecasts of wind speed over the North American Pacific Northwest obtained from the University of Washington mesoscale ensemble. The resulting density forecasts for gust speed are calibrated and sharp and offer substantial improvement in predictive performance over the raw ensemble multiplied with a gust factor or climatological reference forecasts.

References

- [1] T. L. Thorarinsdottir and T. Gneiting, Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression, *Journal of the Royal Statistical Society, Series A*, **173**, pp. 371-388, 1999.

A characterization of extreme precipitation in the Mediterranean region

Andrea TORETI – *Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University of Giessen, Giessen, Germany and Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland*
Email: andrea.toreti@giub.unibe.ch
Co-authors: E. XOPLAKI and J. LUTERBACHER

Abstract

The Mediterranean region has been considered a hot-spot of climate change [1] that is (and will be more) vulnerable and exposed, especially to climate extreme events, such as heat waves and flood events. Therefore, a characterization of extreme precipitation by using observations and global/regional climate models is essential. In this frame, a set of more than 400 daily precipitation series (mostly covering the period 1950-2006) has been collected and quality checked. Then, extended winter (October to March) precipitations have been analyzed by applying a procedure based on a declustered Peak Over Threshold approach and a recently developed minimum density power divergence estimator [2]. Results reveal remarkable spatial differences, in terms of the parameters of the extreme distributions. Besides observations, a new set of climate runs, carried out in the framework of the EU-FP6 'Climate Change and Impact Research: the Mediterranean Environment' CIRCE project by using innovative high resolution global/regional climate models, has been analyzed.

References

- [1] Giorgi, F. 2006. Climate change hot-spots, *Geophys. Res. Lett.*, 33, L08707, doi:10.1029/2006GL025734.
- [2] Juárez, S. F., and Schucany, W. R. 2004. Robust and efficient estimation for the Generalized Pareto Distribution, *Extremes*, 7, 237-251.

Multivariate tail probability representations

Jennifer L. WADSWORTH – *Lancaster University, UK*
Email: j.wadsworth@lancaster.ac.uk
Co-author: Jonathan A. TAWN

Abstract

Existing theory for multivariate extreme values focusses upon characterisations of the distributional tails when all components of a random vector grow at the same rate (for example as presented in [1], [2]). In this talk we consider the effect of allowing the components to grow at different rates, and characterise the link between these marginal growth rates and the multivariate tail probability decay rate. The approach leads to some theoretical results which, for asymptotically independent random vectors, mirror the rich variety of descriptions of max-stable dependence which are available through the exponent measure. In addition we describe a simple inferential approach to joint survivor probability estimation. The key feature of the methodology is that extreme set probabilities can be estimated by extrapolating upon rays emanating from the origin when the margins of the variables are exponential. This offers an appreciable improvement over existing techniques where extrapolation in exponential margins is upon lines parallel to the diagonal.

References

- [1] L. de Haan and J. de Ronde, *Sea and wind: Multivariate extremes at work*, *Extremes* 1(1) pp.7–45, 1998.
- [2] A. W. Ledford and J. A. Tawn, *Modelling dependence within joint tail regions*, *J. Roy. Stat. Soc. B* 59 pp.475–499, 1997.

Conditional sampling for spectrally discrete max-stable random fields

Yizao WANG – *University of Michigan, Ann Arbor, USA*
Email: yizwang@umich.edu
Co-author: Stilian A. STOEV

Abstract

Max-stable random fields play a central role in modeling extreme value phenomena. We will present an explicit formula for the conditional probability in general max-linear models, which include a large class of max-stable random fields. As a consequence, we obtain an algorithm for efficient and exact sampling from the conditional distributions. Our method provides a computational solution to the prediction problem for spectrally discrete max-stable random fields. This work offers new tools and a new perspective to many statistical inference problems for spatial extremes, arising in many applications.

References

- [1] Y. Wang and S. A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. To appear in *Adv. in Appl. Probab.*, 2011. <http://arxiv.org/abs/1005.0312>

Empirical likelihood based confidence regions for first order parameters of heavy-tailed distributions

Rym WORMS – *LAMA (CNRS UMR 8050), University Paris-Est-Créteil, France*
Email: rym.worms@u-pec.fr
Co-author: Julien WORMS

Abstract

Consider X_1, \dots, X_n some i.i.d. observations from a heavy tailed distribution F , i.e. such that the common distribution of the excesses Y_1, \dots, Y_{N_n} over a high threshold u_n can be approximated by a Generalized Pareto Distribution $G_{\gamma_0, \sigma_{0n}}$ with $\gamma_0 > 0$. Well-known estimators for the couple (γ_0, σ_{0n}) are the maximum likelihood (ML) estimators of Smith and the probability weighted moments estimators of Hosking & Wallis. However the performance of the confidence regions (for the couple) which are issued from the asymptotic normality of these estimators has never been investigated.

In this work an alternative method is considered, namely the empirical likelihood (EL) method. An empirical likelihood ratio (ELR below) is constructed, starting from the less-known estimating equations proposed by

Zhang in [3], which are close to the classical likelihood equations but possess good computational properties. By proceeding so, classical advantages of the EL methodology are expected, namely the avoidance of covariance estimation and the possibility of profiling the likelihood for estimating γ_0 alone.

The estimating equations of Zhang are based on the following function g (where the constant $r < 0$ is arbitrary)

$$g(y, \gamma, \sigma) := \left(\frac{\log(1 + \gamma y/\sigma) - \gamma}{(1 + \gamma y/\sigma)^{r/\gamma} - \frac{1}{1-r}} \right).$$

which, when Z is $G_{\gamma, \sigma}$ -distributed, satisfies $E[g(Z, \gamma, \sigma)] = 0$. Reminding that the excesses over the high threshold u_n are denoted by Y_1, \dots, Y_{N_n} (where N_n is therefore random and binomial distributed), the estimators $(\hat{\gamma}, \hat{\sigma})$ resolving in (γ, σ)

$$\frac{1}{N_n} \sum_{i=1}^{N_n} g(Y_i, \gamma, \sigma) = 0$$

are shown to be always well-defined and computable, and therefore to maximize the empirical likelihood ratio

$$ELR(\gamma, \sigma) := \sup \left\{ \prod_{i=1}^{N_n} (N_n p_i) \middle/ \forall i, p_i \geq 0, \sum_{i=1}^{N_n} p_i = 1 \text{ and } \sum_{i=1}^{N_n} p_i g(Y_i, \gamma, \sigma) = 0 \right\}$$

If $c_{k, \alpha}$ denotes the $(1 - \alpha)$ -quantile of the $\chi^2(k)$ distribution, then a consequence of our asymptotic results is that the Wilks-type region

$$\mathcal{R} = \{ (\gamma, \sigma) ; -2 \log ELR(\gamma, \sigma) \leq c_{2, \alpha} \}$$

defines a confidence region for (γ_0, σ_{0n}) with asymptotically correct level $1 - \alpha$, and

$$\mathcal{I} = \{ \gamma ; -2 \log ELR(\gamma, \hat{\sigma}_\gamma) \leq c_{1, \alpha} \}$$

defines a confidence interval for γ_0 with asymptotically correct level $1 - \alpha$, where $\hat{\sigma}_\gamma$ designs the maximizer, in σ , of $ELR(\gamma, \sigma)$. Proof of these statements are based on the methodology developed in [2], with difficulties arising from the fact that the distribution of the excesses Y_i depend on the sample size.

Simulations results will be presented, in order to show the performance, in terms of coverage probability: (i) of the confidence regions \mathcal{R} w.r.t. those derived from the gaussian approximation of the ML estimators ; (ii) of the confidence intervals \mathcal{I} for γ w.r.t. those proposed by Lu & Peng in [1] by the EL methodology, and connected to the Hill estimator.

References

- [1] J.C. Lu and L. Peng (2002), *Likelihood based confidence intervals for the tail index*, *Extremes* 5, pages 337-352.
- [2] Y. S. Qin and J. Lawless (1994), *Empirical likelihood and general estimating equations*, *Annals of Statistics* 22(1), pages 300-325.
- [3] J. Zhang (2007), *Likelihood moment estimation for the generalized Pareto distribution*, *Australian & New Zealand J. Stat* 49(1), pages 69-77.

Systematic risk under extremely adverse market conditions

Chen ZHOU – *De Nederlandsche Bank and Erasmus University Rotterdam*

Email: zhou@ese.eur.nl

Co-author: Maarten VAN OORDT

Abstract

Extreme losses are the major concern in risk management. The dependence between financial assets and the market portfolio changes under extremely adverse market conditions. We develop a measure of systematic tail risk, the *tail regression beta*, defined by an asset's sensitivity to large negative market shocks, and establish the estimation methodology.

Building on extreme value theory, the estimator of the tail regression beta consists of the asymptotic dependence measure and the marginal risk measures. Theoretically, it has a similar structure as the estimator of the regular beta from regression analysis. Simulations show that our estimation methodology yields an estimator that has a lower mean squared error than performing regressions in the tail.

Empirical results based on analyzing 46 industrial portfolios demonstrate that the regular portfolio sensitivity to the systematic risk is in general different from the sensitivity to systematic risk in severe market downturns. Furthermore, the tail regression beta is a useful instrument in both portfolio risk management and systemic risk management. We demonstrate its applications in analyzing Value-at-Risk (VaR) and Conditional Value-at-Risk (CoVaR).

Abstracts of Contributed Posters

Extreme events cutoff, long-range correlation and linearisation effect in multifractal analysis

Florian ANGELETTI – *Laboratoire de physique de l'ENS Lyon, CNRS UMR 5672, France*

Email: florian.angeletti@ens-lyon.fr

Co-authors: Marc MÉZARD, Patrice ABRY, and Eric BERTIN.

Abstract

The analysis of the linearization effect in multifractal analysis [2, 3], and hence of the estimation of moments for multifractal processes, is revisited borrowing concepts from the statistical physics of disordered systems, notably from the analysis of the so-called Random Energy Model [4]. Considering a standard multifractal process (compound Poisson motion [5]), chosen as a simple representative example, we show: i) the existence of a critical order q^* beyond which moments, though finite, cannot be estimated through empirical averages, irrespective of the sample size of the observation; ii) that multifractal exponents necessarily behave linearly in q , for $q > q^*$. Tayloring the analysis conducted for the Random Energy Model to that of Compound Poisson motion, we provide explicative and quantitative predictions for the values of q^* and for the slope controlling the linear behavior of the multifractal exponents. These quantities are shown to be related only to the definition of the multifractal process and not to depend on the sample size of the observation. Monte-Carlo simulations, conducted over a large number of large sample size realizations of compound Poisson motion, confirm and extend these analyses.

References

- [1] F. Angeletti, M. Mézard, E. Bertin, and Abry P. Linearization effect in multifractal analysis: Insights from random energy model analysis. *Accepted by Physica D*, 2011.
- [2] B.B. Mandelbrot. Intermittent turbulence in self-similar cascades; divergence of high moments and dimension of the carrier. *J. Fluid Mech.*, 62:331–358, 1974.
- [3] G. M. Molchan. Turbulent cascades: Limitations and a statistical test of the lognormal hypothesis. *Phys. Fluids*, 9(8):2387–2396, 1997.
- [4] B. Derrida. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613, 1981.
- [5] J. Barral and B. Mandelbrot. Multiplicative products of cylindrical pulses. *Probab. Theory Relat. Fields*, 124:409–430, 2002.
- [6] P. Abry, V. Pipiras, and H. Wendt. Extreme values, heavy tails and linearization effect: A contribution to empirical multifractal analysis. In *21st GRETSI Symposium on Signal and Image Processing*, Troyes, France, 2007.

Multivariate extreme values for assessing simultaneous over-exposure to food chemicals

Emilie CHAUTRU – *Télécom ParisTech, INRA & ANSES, France*

Email: emilie.chautru@gmail.com

Abstract

If eating is the privileged way of providing the necessary nutrients for the human organism, it also conveys toxic elements that, due to various environmental causes, contaminate the food. When consumed over certain tolerable doses, these can have a non-negligible impact on health. Similar phenomena also occur when diets are either too rich or too poor in nutrients. For public health organisms such as the ANSES (the French agency for food, environmental and occupational health safety), it is then of major interest to understand the mechanisms that lead to over- or under-exposure to various chemical elements and quantify the level of risk in a given population. In the present study, we focus our attention on the relationships between high - and relatively rare - exposures to various nutrients and contaminants, that are assessed with the use of the multivariate extreme value theory.

Our variables of interest, i.e. exposures to a group of nutrients or contaminants, are not directly available; they result from a combination of both the consumption data base INCA2 (the national individual survey

about eating habits in France) and the contamination data bases TDS2 (total dietary survey) and CIQUAL (information centre about food quality). Thus, extreme exposures can arise from over-consumption of products containing these elements, from ingestion of food with excessively high amounts of contaminants, or even from the association of moderately contaminated foodstuffs. To take all three situations into account, we estimate exposures with an incomplete generalized U-statistic as developed in P.Bertail and J.Tressou [1] and then study the dependence of the multivariate extremes with classical methods.

In reference to the previous work of J.Tressou and al. [2] on the probability of over-exposure to methylmercury, we focus on the case of heavy-tailed marginals. After a brief univariate analysis, we determine pairs of variables that can be studied simultaneously by assessing the coefficient of tail dependence and testing for asymptotic independence as in [3]. For seemingly dependent pairs we compare various estimates of the spectral measure, based on either parametrical or non-parametrical reductions to standard Fréchet.

From these results we draw conclusions about the probabilities of simultaneous over-exposure to methylmercury, dioxins and PCB, iron and sodium.

References

- [1] P. Bertail and J. Tressou, *Incomplete Generalized U-Statistics for Food Risk Assessment*, Biometrics, 2006.
- [2] J. Tressou, A. Crépet, P. Bertail, M.H. Feinberg, J.Ch. Leblanc, *Probability exposure assessment to food chemicals based on extreme value theory. Application to heavy metals from fish and sea products*, Food and Chemical Toxicology, 2004.
- [3] M. Falk and R. Michel, *Testing for tail independence in extreme value models*, Annals of the Institute of Statistical Mathematics, 2006.
- [4] J. Beirlant, Y. Goegebeur, J. Segers and J. Teugels, *Statistics of Extremes. Theory and Applications*, Wiley, 2004.
- [5] S.I. Resnick, *Heavy-tail phenomena: probabilistic and statistical modeling*, Springer Verlag, 2007.

Random field with random dimensions

Sandra DIAS – *CM-UTAD, Department of Mathematics, University of Trás-os-Montes e Alto Douro, Portugal*
Email: sdias@utad.pt

Co-authors: Luísa CANTO E CASTRO and Maria da Graça TEMIDO

Abstract

We study the limiting distribution of the maximum value of a stationary bivariate random field. In a first part, when the double dimensions have geometric growing pattern, a max-semistable distribution is obtained. Moreover, in a second part, considering the random field with geometric growing random dimensions, a mixture distribution is established for the maximum.

References

- [1] M. G. Temido and L. Canto e Castro, Max-semistable laws in extremes of stationary random sequences, *Theory Probab. Appl.*, 47(2), 365-374, 2003.
- [2] A. Freitas, J. Hüsler and M. G. Temido, Limit laws for maxima of a stationary random sequence with random sample size, *Test* (to appear), DOI 10.1007/s11749-011-0238-2.

Estimation of a new parameter discriminating between Weibull tail-distributions and heavy-tailed distributions

Jonathan EL METHNI – *LJK & INRIA Rhône-alpes, France*

Email: jonathan.el-methni@inrialpes.fr

Co-authors: Laurent GARDES, Stéphane GIRARD and Armelle GUILLOU

Abstract

The Gnedenko theorem is a general result in extreme value theory establishing the asymptotic distribution of extreme order statistics. The maximum of a sample of independent and identically distributed random variables after proper renormalization converges in distribution to one of the three possible maximum domain of attraction: Fréchet, Weibull and Gumbel. In a lot of applications (hydrology, finance, etc . . .), the Fréchet maximum domain of attraction and the Gumbel maximum domain of attraction are used. The Gumbel maximum domain of attraction encompasses a large variety of distributions. Here, we focus on a subfamily of distributions called Weibull tail-distributions which depends on the Weibull tail-coefficient. Numerous works are dedicated to the estimation of this coefficient (see for instance, [1, 2, 3]), and to the estimation of the tail index (see [5] for a

review). In order to explain why the same methodology can be used to estimate the Weibull tail-coefficient and the tail index, the authors proposed in [4] a family of distributions which encompasses the whole Fréchet maximum domain of attraction as well as Weibull tail-distributions. These distributions depend on 2 parameters $\tau \in [0, 1]$ and $\theta > 0$. The first one, τ allows us to represent a large panel of distribution tails ranging from Weibull-type tails ($\tau = 0$) to distributions belonging to the maximum domain of attraction of Fréchet ($\tau = 1$). The main goal of this communication is to propose an estimator for τ independent of θ . Under some assumptions we establish the asymptotic distribution of this estimator.

References

- [1] J. Beirlant, M. Broniatowski, J.L. Teugels and P. Vynckier. (1995), The mean residual life function at great age: Applications to tail estimation. *Journal of Statistical Planning and Inference*, **45**, 21–48.
- [2] J. Diebolt, L. Gardes, S. Girard and A. Guillou. (2009). Bias-reduced estimators of the Weibull tail-coefficient. *Test*, **17**, 311–331.
- [3] L. Gardes and S. Girard. (2006). Comparison of Weibull tail-coefficient estimators. *REVSTAT-Statistical Journal*, **4**, 163–188.
- [4] L. Gardes, S. Girard and A. Guillou. (2011). Weibull tail-distributions revisited: a new look at some tail estimators. *Journal of Statistical Planning and Inference*, **141**, 429–444.
- [5] L. de Haan and A. Ferreira. (2006), *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering, Springer.

DPOT Methodology: a duration based POT method with application to VaR

M. Isabel FRAGA ALVES – *CEAUL & DEIO University Lisbon, Portugal*

Email: isabel.alves@fc.ul.pt

Co-author: Paulo ARAÚJO SANTOS

Abstract

Threshold methods, based on fitting a stochastic model to the excesses over a threshold, were developed under the acronym POT (peaks over threshold). In order to eliminate the tendency to clustering of violations, a model based approach within the POT framework, that uses the durations between excesses as covariates, is presented. This is named by DPOT methodology. Based on this approach, models for forecasting one-day-ahead Value-at-Risk were applied to real data. Comparative studies provide evidence that they can perform better than state-of-the-art risk models and much better than the widely used RiskMetrics model.

Acknowledgements

Research partially supported by Fundação para a Ciência e Tecnologia (FCT/PROTEC and FCT/OE).

References

- [1] Araújo Santos, P. and Fraga Alves, M.I.(2011) . Forecasting Value-at-Risk with a Duration based POT method. Technical Report, <http://www.ceaul.fc.ul.pt/getfile.asp?where=notas&id=303>, CEAUL 08/11.
- [2] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, Heidelberg.
- [3] Smith, R. (1990). Models for Exceedances over High Thresholds. *J. R. Statist. Soc. B*, (52), no.3, 393-442.
- [4] Tsay, R.(2010). *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics.

Modeling monthly maximum 24 hour precipitation in Iceland

Oli P. GEIRSSON – *University of Iceland, Iceland*

Email: olipalli@gmail.com

Co-author: Birgir HRAFNKELSSON

Abstract

We develop a Bayesian hierarchical model for the prediction of monthly maximum 24 hour precipitation in Iceland. This model incorporates observed data, outputs from meteorological models and other covariates. The data are modeled with the generalized extreme value distribution. The hierarchical model is similar to that of [1]. Parametric inference is based on the Bayesian approach.

References

- [1] B. Hrafnkelsson, J. S. Morris and V. Baladandayuthapani , *Spatial modeling of annual minimum and maximum temperatures in Iceland*, to appear in *Meteorology and Atmospheric Physics*.

Spatial modeling of annual minimum and maximum temperatures

Birgir HRAFNKELSSON – *University of Iceland, Iceland*

Email: birgirhr@hi.is

Co-authors: Jeffrey S. MORRIS

Abstract

Bayesian hierarchical modeling is applied to the analysis of annual minimum and maximum temperatures. In both cases the generalized extreme value distribution is selected as the marginal distribution at each site due to its flexibility and theoretical basis. This distribution has three unknown parameters, a location parameter, a scale parameter and a shape parameter. The location and scale parameters are assumed to vary across sites while the shape parameter is assumed to be the same for all sites. Our model is very similar to the models presented in [1] and [2]. The location and the scale parameters are modeled as two independent Gaussian spatial processes which are governed by the Matérn correlation function. These spatial processes play a central role in spatial predictions at unobserved sites. The location parameter is also modeled with a common time effect which includes a linear trend and a linear site effect with altitude, distance to open sea, latitude and longitude as covariates. This model allows for spatial predictions for any set of unobserved sites, the scale of the grid can be as fine as possible as long as the covariates are observed at each of the unobserved sites.

Data on the annual minimum and maximum temperatures in Iceland from 1961 to 2009 at 72 sites are analyzed and used to predict the 2nd percentile of the minimum temperature for the 12-month period from July 2011 to June 2012 and the 98th percentile of maximum temperature for the year 2011 for a large set of unobserved sites across Iceland at which the covariates are available. The spatial predictions reveal that the 2nd percentile of minimum temperature for 2011 to 2012 reaches -35°C in the central part of Iceland and ranges from -12 to -22°C around the coast of Iceland with the lowest coastal temperatures in the Northern part and the Northeastern part. The 98th percentile of maximum temperature in 2011 ranges from 10 to 30°C where the warmest areas are found in the Eastern part, the North-Eastern part and the Southern part of the country. The estimated increase in minimum and maximum temperatures over the years 1961 to 2009 is 0.71 and 0.47°C per decade, respectively, while the average annual temperature increased 0.24°C per decade over the same period.

References

- [1] E. Casson and S. Coles (1999) Spatial regression models for extremes. *Extremes* 1:449–468
- [2] H. Sang and A. Gelfand (2009) Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* 16:407–426

Extreme value models based on homogeneous and non-homogeneous Poisson process in climate change simulations

Jan KYSELÝ – *Institute of Atmospheric Physics, Prague; Global Change Research Centre, Brno, Czech Republic*

Email: kysel@ufa.cas.cz

Co-authors: Jan PICEK and Romana BERANOVÁ

Abstract

The study compares statistical models for estimating high quantiles of daily temperatures based on the homogeneous and non-homogeneous Poisson process, and their applications to global climate model (GCM) simulations. Both types of the models make use of a non-stationary peaks-over-threshold method and the Generalized Pareto distribution (GPD) for modelling extremes, but they differ in how the dependence of the model parameters on time index is captured. The homogeneous Poisson process model assumes that the intensity of the process is constant and the threshold used to delimit extremes changes with time; the non-homogeneous Poisson process assumes that the intensity of the process depends on time while the threshold is kept constant [1]. The model for time-dependency of GPD parameters is selected according to the likelihood ratio test. Statistical arguments are provided to support the homogeneous Poisson process model, in which temporal dependence of the threshold is modelled in terms of regression quantiles [2]. Dependence of the results on the quantile chosen for the threshold (95-99%) is evaluated. The extreme value models are applied to estimate changes in high quantiles of daily temperatures (20-yr and 100-yr return values) in transient simulations of several GCMs for the 21st century.

References

- [1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp., 2001.
- [2] J. Kysely, J. Picek, R. Beranová, Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72, 55–68, 2010.

Location invariant Weiss-Hill estimator

Chengxiu LING – *University of Lausanne, Department of Actuarial Science, Switzerland*

Email: Chengxiu.Ling@unil.ch

Co-authors: Zuoxiang PENG and Saralees NADARAJAH

Abstract

In this paper, a novel location invariant Weiss-Hill estimator of extreme value index $\gamma \in \mathbf{R}$ is proposed. The new estimator is a combination of two estimators proposed by Weiss (1971) and Fraga Alves (2001a)[1], respectively. Its asymptotic behaviors such as weak and strong consistencies and asymptotic representations are obtained, including also a bias corrected location invariant Weiss-Hill estimator and the optimal choice of sample fraction asymptotic mean squared error (see also [2]). Simulation studies and comparison with other location invariant estimators given in [3],[4] are further considered.

References

- [1] Fraga Alves, M. I. (2001a). A location invariant Hill-type estimator. *Extremes* 4, 199–217.
- [2] Ling, C., Peng, Z. and Nadarajah, S. (2009). Selecting the optimal sample fraction in location invariant moment-type estimators. *Theory of Probability and Its Applications*, in press.
- [3] Ling, C., Peng, Z. and Nadarajah, S. (2007a). A location invariant Moment-type estimator I. *Theory of Probability and Mathematical Statistics* 76, 23–31.
- [4] Ling, C., Peng, Z. and Nadarajah, S. (2007b). A location invariant Moment-type estimator II. *Theory of Probability and Mathematical Statistics* 77, 177–189.

Conditional Analysis for Multivariate Extreme Financial Risks

Ye LIU – *Lancaster University, U.K.*

Email: y.liu10@lancaster.ac.uk

Co-author: Jonathan. A. TAWN

Abstract

This poster discusses some recent theoretical development in the area of multivariate extreme value theory, based on the model proposed in Heffernan and Tawn (2004). The first half of the poster focuses on some of the unsolved issues in the original paper, namely the self-consistency issue and the residual distribution. The second half of the poster explores the an application of Heffernan-Tawn model in risk management and proposes a factor-based hierarchical multivariate extreme value model.

Introduction

This section starts by reinforcing the importance of extreme risk management in the financial industry and moves onto a quick introduction to the background of multivariate extreme value theory. This is then followed by an overview of a conditional approach to analyze multivariate extreme value data, proposed in Heffernan and Tawn (2004), and the key concepts involved in the model.

Theory

This section covers some of the recent research results developed from the original Heffernan and Tawn model, with regard to the the following two issues, which were previously unsolved

- Self-consistency requires the joint tail density being identical regardless of the conditioning margin

$$\mathbb{P}(Y_j = y_j | Y_i = y_i) f_{Y_i}(y_i) = \mathbb{P}(Y_i = y_i | Y_j = y_j) f_{Y_j}(y_j)$$

- The residuals of the standardized extreme observations converge to a non-degenerate yet unidentified distribution

So this section summarizes the potential problems that these issues are associated with and proposes a feasible resolution.

Application

Motivated by some initial findings when applying the Heffernan and Tawn model to the real world financial data, this section introduces an intuitive multivariate extreme value model that accounts for the classes of assets in a typical investment portfolio. The main purpose of this section is to

- Provide some evidence to support a proposed factor-based hierarchical approach
- Describe the overall model formulation
- Illustrate its main application in cross-asset portfolio construction
- Discuss other potential applications in risk management

Conclusion

This section summarizes the main findings to date and suggests areas of future work.

References

- [1] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, Heidelberg.
- [2] Heffernan, J.E. and Tawn, J.A. (2004). A conditional approach for multivariate extreme values (with disucssion). *JRSS B* **66**:3, 497-546.
- [3] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169-187.
- [4] Ledford, A.W. and Tawn, J.A. (1997). Modelling depenence within join tail regions. *JRSS B* **59**:3, 475-499.

Paleoclimate Extremes in Proxy Data

Elizabeth MANNSHARDT-SHAMSELDIN – *Duke University, USA*

Email: elizabeth@stat.duke.edu

Co-authors: Peter CRAIGMILE and Martin TINGLEY

Abstract

There is much debate about the relationship between the changing climate and extreme climate events. Possible trends identified in temperature reconstructions based on proxy data lead to several questions about long-term climate behavior, and how to interpret this behavior given the patterns seen in proxy series. For example, using proxy data to address questions such as "Is there evidence that the extreme events of recent decades are more extreme than previous decades?" This leads to the question of whether the distributions of climate extremes are changing over time, which can provide important insights into how the climate system is changing. The methodology of extreme value theory has not been widely applied to this problem. This paper looks at what the statistics of extremes has to offer the field of paleoclimatology through modeling of the original proxy series, and seeks to address several of the emerging areas of research that intertwine extreme value analysis and paleoclimate reconstructions.

References

- [1] Briffa, K.R., Osborn, T. J. Schweingruber, F.H., Jones, P.D., SG Shiyatov, S.G., and Vaganov, E.A., 2002. "Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals", *The Holocene*, (6), Vol 12, pg 737.
- [2] IPCC: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007.
- [3] Tingley, M. P., and P Huybers, P., In Prep. "A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record", 2011.

Strong invariance principles for tail quantile processes with applications to extreme value index estimation

Ludovic MENNETEAU – *Université Montpellier 2, France*

Email: mennet@math.univ-montp2.fr

Co-author: Stéphane GIRARD

Abstract

Many estimators of the extreme value index are functions of the k_n largest observations of the sample and therefore can be seen as a functional of the k_n upper tail quantile process. Under classical second order assumptions, this quantile process can be approximated, via a quantile transformation, by a non linear functional of the tail uniform empirical process (see [1]). Here, we prove a strong invariance principle for this non linear functional. In some ways, this result improves the approximation result obtained by [1] (Theorem 2.1) since it is convenient to prove strong limit theorems. In particular, we obtain a functional law of the iterated logarithm for the quantile process. As an application, we establish a compact law of the iterated logarithm for the classical Hill estimator [2].

References

- [1] H. Drees. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, **25**, 187–210, 1998.
- [2] B.M. Hill. A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174, 1975.

Statistical inference for the right endpoint of a light-tailed distribution

Cláudia NEVES – *CEAUL and University of Aveiro, Portugal*
Email: claudia.neves@ua.pt

Abstract

In extreme value statistics we often encounter testing procedures for assessing the presence of the Gumbel domain, attached to the simple null hypothesis of the shape parameter being equal to zero. The problem of assessing for light-tailed distributions with finite or infinite right endpoint is seldom referred. We present two testing procedures which enable us to distinguish light-tailed distribution functions with finite right endpoint from those with infinite endpoint lying in the Gumbel domain. An estimator for the finite right endpoint is addressed with application to taxiway centerline deviation data.

A non copula based model for extreme value dependence. Theory and applications.

Antonio ORTIZ BARRANON – *University of Kent, UK*
Email: aao33@kent.ac.uk
Co-author: Stephen WALKER

Abstract

We present a new way of modeling dependent extreme value dependences. To date, the most common way of working with bivariate extremes is to separate and deal separately with the marginals and the dependence structure. See Coles et al in [1], and Jonathan Tawn et al in [2], for example. This approach proceeds by fitting the extreme marginals, by transforming the variables depending on the marginal distributions, and computing a couple of dependence measures with the transformed variables. Depending on the outcome, a type of copula would be chosen and then fitted.

In our model, both the marginal and the dependence structure are modeled simultaneously. This is done in such a way that the parameter modeling the tail dependence can be identified and estimated, i.e. the value of the tail dependence is a simple function of this parameter. Therefore, we have found a simple way to study whether two variables are tail dependent. We present simulation studies and a real data illustration involving financial data, which is very similar to the data used by [2], in order to compare the results of our model.

References

- [1] S.G. Coles, J. Heffernan, and J. Tawn, *Dependence Measures for Extreme Value Analyses*, *Extremes*, **3**, 1999, pp. 5-38.
- [2] J. Tawn, et al, *Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications*, Oxford Journals, *The Review of Financial Studies*, Vol. 17, No 2, 2004, pp. 581-610.

Conditional Modelling of Extreme Values: Drug Induced Liver Injury

Ioannis PAPANASTATHOPOULOS – *Lancaster University, United Kingdom*

Email: i.papastathopoulos@lancaster.ac.uk

Co-authors: Jonathan A. TAWN and Harry SOUTHWORTH

Abstract

Drug induced liver injury (DILI) is a major public health issue and of serious concern for pharmaceutical industry. Early detection of signs of a drugs potential for DILI is vital for pharmaceutical companies evaluation of new drugs. Combination of extreme values of liver specific variables indicate DILI (Hy's Law). We estimate the probability of severe DILI using [1] conditional dependence model which arises naturally in applications where a multidimensional random variable is extreme in at least one component. We extend the current model by including the assumption of stochastically ordered survival curves for different doses in a Phase 2 study.

References

- [1] J. E. Heffernan and J. A. Tawn, A Conditional Approach for Multivariate Extreme Values, *J. R. Statist. Soc. B*, **66**, 1–34

On the distribution of the maximum of the Gaussian fields

Viet Hung PHAM – *University Toulouse 3, France*

Email: viet-hung.pham@math.univ-toulouse.fr

Abstract

The problem of finding the distribution of the maximum of a random field has many applications in spatial statistics. In this poster, by using the Rice formulas developed by J.M. Azaïs and M. Wschebor [1] and extending the record method of C. Mercadier [5], we give some approximations of this distribution in the case of stationary Gaussian field for very general parameter sets.

In dimension 2, we provide an upper bound and a lower bound for the tail of the distribution at each level and prove that it is the exactly asymptotic result.

In dimension 3, from the result of the expectation of the absolute value of quadratic forms by W. Li and A. Wei [3], some similar results are obtained.

References

- [1] R.J. Adler and J. Taylor, *Randoms fields and Geometry*, Springer, New York, 2007.
[2] J.M. Azaïs and M. Wschebor, *Level sets and extrema of random processes and fields*, John Wiley and Sons, 2009.
[3] W. Li and A. Wei, *Gaussian integrals involving absolute value functions*, IMS lecture Notes-Monograph Series, 2009.
[4] M.L. Mehta, *Random matrices*, 3rd ed, Academic Press, San Diego, CA, 2004.
[5] C. Mercadier, *Numerical bounds for the distribution of the maximum of one-and two-dimensional processes*, Adv.APl. Probab. 38(1), 149-170, 2006.

Estimation of extreme value index and two-step regression quantiles

Jan PICEK – *Technical University, Liberec, Czech Republic*

Email: jan.picek@tul.cz

Co-author: Jan DIENSTBIER

Abstract

The contribution deals with estimators of extreme value index based on two-step regression quantiles in the linear regression model. Two-step regression quantiles can be seen as a possible generalization of the quantile idea and as an alternative to regression quantiles. We derive the approximation of the tail quantile function of errors. Following Drees (1998) we consider a class of smooth functionals of the tail quantile function as a tool for the construction of estimators in the linear regression context. Pickands, maximum likelihood and probability weighted moments estimators are illustrated on simulated data.

References

- [1] H. Drees, On Smooth Statistical Tail Functionals, *Scandinavian Journal of Statistics*, **25** 187–210, 1998.
- [2] J. Picek and J. Dienstbier, Extremes of two-step regression quantiles, *Nonparametric and Robustness in a Broader Perspective: A Festschrift in Honor of Professor Jana Jureckova*, Volume No. 7, IMS Collection, 204 - 214, 2010.

An empirical study of an adaptive resampling scheme for estimating the extremal index

Dora PRATA GOMES – *CMA and FCT/UNL, Portugal*

Email: dsrp@fct.unl.pt

Co-author: M. Manuela NEVES

Abstract

Extreme Value Theory deals with extreme or rare events, this is, events that have not been observed yet. It has been widely used in environmental applications, finance, hydrology, etc. There are several parameters that need to be estimated such as: tail index, γ , high quantiles, the return level and the *extremal index*, θ . This parameter plays a special role in the description of the dependence between exceedances over a high threshold.

There exist several interpretations of the *extremal index* from which several estimators have been derived. One of those interpretations, due to [5], considers θ as the reciprocal of the limiting mean cluster size. The identification of clusters of high level exceedances is then required.

The most basic form of cluster identification is to consider that a cluster occurs whenever there is an up-crossing of the high threshold u_n . This suggested the *up-crossing estimator*, $\hat{\theta}_n^{UC}$, ([6] and [1]), defined for a random sample (X_1, X_2, \dots, X_n) and a suitable threshold u_n as:

$$\hat{\theta}_n^{UC} := \frac{\sum_{i=1}^{n-1} I(X_i \leq u_n < X_{i+1})}{\sum_{i=1}^n I(X_i > u_n)}.$$

The estimation of the extremal index is then performed on the basis of the k order statistics in the sample or on the exceedances over a high level u_n . That estimator and other given in the literature, despite of having good asymptotic properties, present high variance for high levels and a high bias when the level decreases, showing a strong dependence on the high threshold u_n , for finite samples. Regarding the compromise between these two measures given by the mean squared error, MSE, a resampling scheme and an adaptive procedure is performed for estimating the optimal level u_n (that minimizes MSE) and then for obtaining an estimate of θ . We are here mainly interested in the use of the bootstrap methodology for dependent data to estimate adaptively the optimal sample fraction.

For dependent data several bootstrap procedures have appeared in the literature, like the block bootstrap (non-overlapping blocks, “moving blocks”, circular blocks), sieve bootstrap, local bootstrap or jackknife-after-bootstrap.

The block bootstrap tries to mimic the behavior of an estimator by resampling blocks of consecutive observations; the blocking is used to preserve the original structure within a block. The performance of a block resampling method critically depends on the particular block length employed in finding the bootstrap estimator. [3] proposed a subsampling method to define a data-based version of *MSE* function which is minimized and rescaled to produce an estimator of the optimal block size. [4] proposed a plug-in rule as an alternative approach for empirical choice of the optimal block size. The key idea of the method is based on the bootstrap estimation of the variance and the bias of the block bootstrap estimator. The proposed rule is based on the Jackknife-After-Bootstrap (JAB) that yields a nonparametric estimator of the variance of a block bootstrap estimator, for the moving block bootstrap case, which is considered in this work.

Once obtained an estimate for the optimal block length, an adaptive procedure is carried out and the optimal sample fraction is obtained and finally θ is estimated.

A simulation study as well as a real case study has been considered. The real data set consists of daily mean river levels from hydrometric station at Fraga, Portugal, during the years from 1946/47 to 1996/97. A subset of this data has already been studied to obtain the estimate of the extremal index, [2].

References

- [1] Gomes, M.I. (1990). Statistical inference in an extremal Markovian model. *Compstat*, 257-262.
- [2] Gomes, M.I. (1993). On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment* (Barnett and K. F. Turkman eds.), 225-241.
- [3] Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, **50**, 561-574.

- [4] Lahiri, S.N., Furukawa, K. and Lee, Y.-D. (2007). A nonparametric plug-in rule for selecting the optimal block length for the block bootstrap methods. *Statistical Methodology*, 4, 292-321.
- [5] Leadbetter, M. R. (1983). Extremes and local dependence in a stationary sequence. *Z. Wahrsch. Verw. Gebiete*, **65**, 291-306.
- [6] Leadbetter, M. R. and Nandagopalan, L. (1989). On exceedance point process for stationary sequences under mild oscillation restrictions. In *Extreme Value Theory: Proceedings, Oberwolfach 1987*, ed J. Hüsler and R. D. Reiss, Lecture Notes in Statistics **52**, 69-80. Springer-Verlag, Berlin.

Accommodating Measurement Scale Uncertainty in Extreme Value Analysis of Northern North Sea Storm Severity

David RANDELL – *Shell Technology Centre Thornton, U.K.*

Email: david.randell@shell.com

Co-authors: Daniel REEVE and Philip JONATHAN

Abstract

Modelling storm severity is critical to the design and reliable operation of marine structures. Extreme hindcast storm peak significant wave heights (H_S) for a neighbourhood of 50 locations in the northern North Sea are modelled, using the four parameter Poisson point process model of [1], which incorporates measurement scale variability via a Box-Cox transformation. The model allows estimation of predictive distributions for both measurement scale and point process parameters within a Bayesian framework. The effect of measurement scale on return values of significant wave height is quantified by comparison with a three parameter Poisson point process model ignoring measurement scale uncertainty. It is found that, within the neighbourhood of locations examined, that the Box-Cox parameter varies from approximately 0.4 to 2, suggesting that appropriate measurement scales range from $H_S^{\frac{2}{5}}$ to H_S^2 . Moreover, return values corresponding to a return period of 3000 years can vary by as much as 15 to 20% when measurement scale uncertainty is accommodated.

References

- [1] J. L. Wadsworth, J. A. Tawn and P. Jonathan, Accounting for Choice of Measurement Scale in Extreme Value Modelling, *Annals of Applied Statistics* **4**, pp1558-1578, 2010.

Estimation of the truncation point of a truncated distribution

Mathias RASCHKE – *Fa.M.Raschke Leipzig, Germany*

Email: mathiasraschke@t-online.de

Abstract

The upper tail of the distribution of a random variable X can be truncated, what can be modelled by a truncated exponential distribution if the original uncut distribution belongs to the Gumbel or the Frechet domain of attraction. The variable X has to be transformed in last case by $Y = \ln(X)$. The consideration of the truncation point is not necessary in the most situations, but it influences considerably the distribution functions of wild fire sizes and earthquake magnitudes, what is motivation for the development and improvement of inference methods. There exist already estimation methods being described or introduced by Hannon and Dahiya [1]. Independent on this, estimation methods for the truncation point are discussed and developed in seismology (s.[2]). They base on different statistical approaches. The established inference methods are compared here with a new estimator [3] being presented in the poster. The new estimator has an excellent performance in a important range of parameters and has interesting properties from the users point of view. Beside this I show by numerical researches that the classical method of block maxima works only well for the estimation of the truncation point in special cases although the truncated exponential distribution belongs to the Weibull domain of attraction with extreme failure index $\gamma = 1$ (s.[4]). This performance of the classical method of extreme value statistics explains the interpretation of "instability" by Pisarenko et al. [5]. Finally, the influence of the estimated truncation point to the inference of the scale parameter of the truncated exponential distribution is also researched.

References

- [1] P.M. Hannon and R.C. Dahiya, *Estimation of parameters for the truncated exponential distribution*, Communications in Statistics - Theory and Methods **28**: 2591-2612, 1999.

- [2] A. Kijko, *A modified form of the first Gumbel distribution: model for the occurrence of large earthquakes, part II: estimation of parameters*, Acta Geophysica 31: 27-39, 1983.
- [3] M. Raschke, *Inference for the truncated exponential distribution*, Stochastic Environmental Research and Risk Assessment, DOI 10.1007/s00477-011-0458-8, 2011.
- [4] M.R. Leadbetter, G. Lindgren and H. Rootzn, *Extremes and Related Properties of Random Sequences and Processes*, Springer, Berlin Heidelberg New York, 1983.
- [5] V.F. Pisarenko, A. Sornette, D. Sornette, and T.V. Rodkin, *New approach to the characterization of M_{max} and of the tail of the distribution of earthquake magnitudes*, Pure and Applied Geophysics 165: 847-888, 2008.

Flexible Extreme Value Mixture Modeling - Towards a Black Box?

Carl J. SCARROTT – *University of Canterbury, New Zealand*

Email: carl.scarrott@canterbury.ac.nz

Co-authors: A. MACDONALD and D. S. LEE

Abstract

A plethora of recent articles have proposed various extreme value mixture models for threshold estimation, some of which also tackle the issue of quantifying the corresponding uncertainty. These mixture models typically treat the threshold as a parameter, so it can be objectively estimated using standard inference tools, avoiding the traditional graphical diagnostics which require expert (subjective) judgment. Some of these mixture models are easy to automate for application to multiple datasets, or in forecasting situations, for which in the past various ad-hoc adaptations had to be made in the past to overcome the threshold estimation problem.

This paper will outline one particularly flexible mixture model (MacDonald *et al*, 2011) which splices together the usual extreme value model for the upper tail behavior, with the threshold as a parameter, and the “bulk” of the distribution below the threshold captured by a non-parametric kernel density estimator. This representation avoids the need to specify a-priori a particular parametric model for the bulk distribution, and only really requires the trivial assumption of a smooth density which is realistic in most applications. The model flexibility also overcomes sensitivity to the specification of the bulk distribution (and in particular it’s lower tail). The existing mixture models in the literature which use parametric models for the bulk distribution can suffer from sensitivity to the specification of the lower tail, such that you can end up in the perverse situation where the fit for the lower tail can have strong influence on the inferences for the upper tail which is undesirable.

Inference for all the parameters, including threshold and kernel density bandwidth, is carried out in a Bayesian paradigm, potentially allowing sources of expert information to be included which can supplement the inherently sparse of extremal sample information.

A known problem with kernel density estimators is that they suffer from edge effects if the (lower) tail does not decay away to zero at the boundary. Various adaptations have been proposed in the nonparametric density estimation literature to overcome this problem. This paper explores one such approach which is flexible enough to cope with distributions which at the boundary have either: 1) a pole; 2) a shoulder; or 3) tail decaying to zero. An alternative approach of replacing the upper and lower tails by extreme models will also be shown to overcome these edge effects. The proposed boundary corrected extreme value mixture appears to provide a good step forward towards a black box solution for threshold estimation and uncertainty quantification for the well behaved population distributions that are typically observed in applications.

A version of the mixture model with extreme models for both tails is also shown to overcome two known problems with non-parametric density estimators: 1) tendency to oversmooth densities with heavy tails; and 2) sensitivity to outliers. In a quid pro quo, this model has led to a sharing of ideas and solutions between the extreme value and non-parametric density estimation literature.

References

- [1] A. MacDonald, C.J. Scarrott, D.S Lee, B. Darlow, M. Reale and G. Russell, *A flexible extreme value mixture model*, *Computational Statistics and Data Analysis*, 55(6), 2137-2157, 2011.

Properties of the extremal correlation function and construction of spatial max-stable processes

Kirstin STROKORB – *Georg-August-Universitaet Goettingen, Germany*

Email: strokorb@math.uni-goettingen.de

Co-authors: Martin SCHLATHER and Zakhar KABLUCHKO

Abstract

An adequate characterisation of the spatial dependence of a stationary max-stable random field $(X_t)_{t \in T}$ is a non-trivial issue, specifically in view of the non-existence of covariances for certain marginal distributions. Being a bivariate positive definite characteristic that can be inferred from data the extremal correlation function

$$\chi(t) := \lim_{x \rightarrow \infty} \mathbb{P}(X_t \geq x | X_0 \geq x)$$

has been proposed as an analogue of the covariance function [1].

We investigate the set of all valid extremal correlation functions $\chi(t)$ on general spaces T , including \mathbb{Z}^d or \mathbb{R}^d , and relate it to other collections of positive definite functions with values in $[0, 1]$, namely set covariance functions, uncentered covariance functions of binary fields, and pointwise limits of extremal correlation functions.

In particular, we raise the following question: Does the set of extremal correlation functions coincide with the set of functions $f(t) = \lim_{x \rightarrow \infty} \mathbb{P}(X_t \geq x | X_0 \geq x)$ where X is a (not necessarily max-stable) process such that the respective limit exists? For processes on $T = \mathbb{Z}$ this is true, for example. Moreover, in this case any extremal correlation function χ with bounded support can be represented as a set covariance function [2]. This allows the construction of a max-stable process realising $\chi(t)$.

As a byproduct, we generalise a distribution model from [1] and obtain an explicit model for any valid collection of extremal coefficients

$$\theta(A) := \frac{-\log \mathbb{P}(\max_{t \in A} X_t \leq x)}{-\log \mathbb{P}(X_t \leq x)}.$$

Thus, we also establish an intriguing connection to the capacity functional of a binary field (compare [3]).

References

- [1] M. Schlather and J. A. Tawn, A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference, *Biometrika*, 90:139-156, 2003.
- [2] A. Ehlert and M. Schlather, A Constructive Proof for the Extremal Coefficient of a Dissipative Max-Stable Process on \mathbb{Z} being a Set Covariance, *Courant Research Centre PEG - Discussion Papers*, 31, 2010.
- [3] I. Molchanov, Convex geometry of max-stable distributions, *Extremes*, 11(3):235-259, 2008.

Stable estimations for extreme wind speeds

Hans VAN DE VYVER – *Royal Meteorological Intstitute of Belgium*

Email: hvijver@oma.be

Co-author: Andy DELCLOO

Abstract

The Generalized Pareto distribution (GPD)

$$G(y) = 1 - \left(1 + \gamma \frac{y}{\sigma}\right)^{-1/\gamma}, \quad (8)$$

is frequently applied for the statistical analysis of extreme wind speeds. Here, y is the excess over a sufficiently high threshold, γ the extreme-value index or the shape parameter, and σ is the scale parameter. A central topic in extreme-value theory is the adaptive estimation of γ . A characteristic feature of wind speed measurements is the absence of decimal units. Unfortunately, such a loss accuracy may seriously affect the parameter estimation of GPD. In this context, several authors have demonstrated a high sensitivity of γ against the threshold when analysing extreme wind speeds, see for example [2]. This undesirable effect introduces the difficulty to provide reliable quantile estimates. In [3] we aim to bring this problem to meteorologists and proposes the estimator of Beirlant *et al.* [1] (the Zipf-estimator) for γ . The method is based on regression in the so-called generalized quantile plots. We summarize our main conclusions of the comparative tests with the Zipf-estimator and classical estimators (probability weighted method, maximum likelihood, ...). Firstly, we have shown that the estimation of γ becomes stable as the threshold moves when using the Zipf-estimator. This could allow a more objective prior identification of the sign and range of γ . Taking into account that the observations were rounded would change the tail estimation rather little for the Zipf-estimator. In contrast, the rounding increased substantially the errors in the classical parameter estimation. Secondly, extensive Monte-Carlo simulations indicate that the Zipf-estimator is much more accurate and reliable than classical methods. These simulations were designed in such a way that they share the characteristics of Belgian extreme wind speeds as close as possible. Thirdly, convergence of γ provided by the Zipf-estimator is generally achieved for lower threshold values compared to

classical methods, so that more data can be included in the extreme-value analysis. Finally, the new methodology is applied to get improved prediction of extreme wind gusts in Belgium. A considerably shortening of the confidence intervals for the parameters and quantiles is found when the Zipf-estimator is used. Furthermore, the spatial differences of GPD's parameter are greatly reduced.

References

- [1] J. Beirlant, G. Dierckx and A. Guillou, Estimation of the extreme-value index and generalized quantile plots, *Bernoulli* **11** (2005) 949–970.
- [2] J. Galambos and N. Macri, Classical extreme value model and prediction of extreme winds, *J. Struct. Eng.* **125** (1999) 792–794
- [3] H. Van de Vyver and A. W. Delcloo, Stable estimations for extreme wind speeds. An application to Belgium, *Theor. Appl. Climatol.* In press, published online 27 Januari 2011. Doi: 10.1007/s00704-010-0365-9.

Rates of convergence of extremes for mixed exponential distributions

Zhichao WENG – *University of Lausanne, Switzerland*

Email: wengzhichao166@126.com

Co-authors: Zuoxiang PENG and Saralees NADARAJAH

Abstract

One interesting topic in extreme value theory is to consider the uniform convergence rates of distributions of extremes to its ultimate extreme value distribution. de Haan and Resnick [2], under the second Von-Mises conditions, considered the convergence rate in the uniform metric and the total variance metric. For the uniform convergence rates of the extremes of some special distributions, see Hall [3] for normal distribution and Hall and Wellner [4] for exponential distribution. Recently Peng et al. [5] extended Hall's work to the general error distribution, and Lin et al. [1] considered the uniform convergence rates of extremes of short tailed symmetric distribution.

Let F denote the finite mixed exponential distribution defined by

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots + p_r F_r(x),$$

where weights $0 < p_i < 1$, $\sum_{i \leq r} p_i = 1$ and F_i follows exponential distribution with parameter λ_i for $1 \leq i \leq r$. In this short note, we consider the limiting distributions of the extremes of finite mixed exponential distribution and the associated uniform converge rates. The practical values of the results are illustrated by a numerical study.

References

- [1] F. Lin, X. Zhang, Z. Peng and Y. Jiang, *On the convergence of STSD extremes*, Communications in Statistics-Theory and Methods, 40, 1795-1806, 2011.
- [2] L. de Haan and S. I. Resnick, *Second-order regular variation and rates of convergence in extreme-value theory*, Ann. Prob. 24,97-124, 1996.
- [3] P. Hall, *On the convergence of normal extremes*, J. Appl. Prob. 16, 433-439, 1979.
- [4] W. J. Hall and J. A. Wellner, *The rate of convergence in law of the maximum of an exponential sample*, Statist. Neerlandica, 33, 151-154, 1979.
- [5] Z. Peng, S. Nadarajah and F. Lin, *Convergence rate of extremes from general error distribution*, Journal of Applied Probability, 47, 668-679, 2010.

Index

- Angeletti, Florian, 47
Aulbach, Stefan, 12
- Ballani, F., 4
Barme-Delcroix, Marie-Francoise, 12
Baxevani, Anastassia, 13
Blanchet, Jose H. , 4
- Caeiro, Frederico , 13
Cai, Juan-Juan, 14
Caires, Sofia, 4
Carreau, Julie, 5
Chakrabarty, Arijit, 15
Chautru, Emilie, 47
Cirillo, Pasquale, 16
Clifton, David , 5
Cooke, Roger, 6
Cooley, Dan, 6
- Dacunha-Castelle, Didier, 17
Daouia, Abdelaati, 17
Das, Bikramjit, 18
Davison, Anthony, 6
de Carvalho, Miguel, 15
de Haan, Laurens, 24
Deme, El-hadji, 18
Dias, Sandra, 48
Dienstbier, Jan, 19
Drees, Holger, 19
- Eckert, Nicolas, 19
El Methni, Jonathan, 48
Embrechts, Paul, 6
Engelke, Sebastian, 20
- Falk, Michael, 21
Feidt, Anne, 21
Ferreira, Ana, 22
Fiori, Anna M., 22
Fraga Alves, M. Isabel, 49
Friederichs, Petra, 7
Furrer, Jasmin, 23
- Geirsson, Oli P., 49
Glynn, Peter W., 2
Gomes, M. Ivette, 23
Guillou, Armelle, 24
- Hüsler, Jürg, 26
Haug, Stephan, 7
Hazra, Rajat S., 25
Hofmann, Martin, 25
Hrafinkelsson, Birgir, 50
Huang, Mei Ling, 26
- Janßen, Anja, 26
Jaruskova, Daniela, 27
Jeon, Soyoung, 28
Jonathan, Philip, 28
- Jongbloed, Geurt, 2
- Karpa, Oleh, 29
Keef, Caroline, 7
Kennedy, Marc, 8
Knight, Keith, 30
Ko, Jordan, 30
Kortschak, Dominik , 8
Kurowicka , Dorota , 9
Kyselý, Jan, 50
- Li, Deyuan, 31
Ling, Chengxiu, 51
Liu, Jingchen, 32
Liu, Ye, 51
- Mainik, Georg, 32
Majumdar, Satya, 3
Mannshardt-Shamseldin, Elizabeth, 52
Maume-Deschamps, Véronique, 33
Mazas, Franck, 33
Menneteau, Ludovic, 52
Misiewicz, Jolanta, 9
Mladenović, Pavle, 34
- Naveau, Philippe, 9
Neves, Cláudia, 53
Neves, M. Manuela, 35
Nolan, John P., 36
Nolde, Natalia, 36
Northorp, Paul, 10
- Oesting, Marco, 36
Opitz, Thomas, 37
Ortiz Barranon, Antonio , 53
- Papastathopoulos, Ioannis, 54
Parey, Sylvie, 37
Pham, Viet Hung, 54
Picek, Jan, 54
Prata Gomes, Dora, 55
- Raillard, Nicolas, 38
Randell, David, 56
Raschke, Mathias, 56
Reich, Brian, 10
Reiss, Rolf-Dieter, 39
Renard, Benjamin, 39
Rootzén, Holger, 40
Roy, Parthanil, 41
- Scarrott, Carl J., 57
Segers, Johan, 10
Smith, Richard L., 11
Stoev, Stilian A., 11
Storkorb, Kirstin, 57
Stupfler, Gilles, 41
Suard, Norbert, 42
Sun, Xun, 43

Tabiś, Kamil, 43
Thorarinsdottir, Thordis L., 44
Toreti, Andrea, 44

Van de Vyver, Hans, 58

Wadsworth, Jennifer L., 45
Wang, Yizao, 45
Weng, Zhichao, 59
Worms, Rym, 45

Zhou, Chen, 46
Zwiers, Francis, 3