4th Conference on Extreme Value Analysis: Probabilistic and Statistical Models and their Applications

Gothenburg, August 15-19, 2005

Scientific Programme of the Meeting and Abstracts

Monday, 15 August

08:55	Opening
09:00-10:30	Chair: Claudia Klüppelberg
09:00-09:30	Laurens de Haan *Failure set estimation in multi- and infinite-dimensional space*
09:30-10:00	Armelle Guillou *Return level bounds for extreme values*
10:00-10:30	Jery Stedinger *GEV flood quantile estimators with Bayesian GLS shape-parameter regression*

10:30-11:00 Refreshments

11:00-12:30	Chair: Ross Maller
11:00-11:30	Lars Holst *On single and double records*
11:30-12:00	Ross Leadbetter *Comments on extremes in naval architecture and the environment - some pitfalls, promises, and progress*
12:00-12:30	Allyson Abrams *Empirical/asymptotic p-values for Monte Carlo-based hypothesis testing: an application to cluster detection using the scan statistic*

14:30-16:00	The Aula, NHV: Extreme Value Statistics Chair: Philippe Soulier	Vinga: Large Deviations and Percolation Chair: Gennady Samorodnitsky
14:30-15:00	Ivette Gomes *Second order reduced bias tail index estimators*	Frank Redig *Maximal clusters in non-critical percolation and related models*
15:00-15:30	Christian Robert *Inference for the limiting cluster size distribution of extreme values*	Jeffrey Collamore *Large deviation estimates for certain heavy-tailed dependent sequences arising in risk management*
15:30-16:00	Johan Segers *A closer look at the Hill estimator: Edgeworth expansions and confidence intervals*	Mikhail Kozlov *On large deviations for branching processes in random environment*

16:00-16:30 Coffee

16:30-17:30	The Aula, NHV: Finance Chair: Casper de Vries	Vinga: Extreme Value Statistics Chair: Ishay Weissman
16:30-17:00	Jose Olmo *Contagion versus flight to quality in financial markets*	Amelie Fils-Villetard *Least-squares estimation of a convex function*
17:00-17:30	Juan Cajigas *Dynamic conditional correlation models with asymmetric multivariate Laplace innovations*	Maria Isabel Fraga Alves *Reduced bias semi-parametric quantile estimators with a linear-type property*

17:30 Reception with cheese and wine and poster session

Tuesday, 16 August

09:00-10:30	The Aula, NHV: Fatigue Chair: Jacques de Maré	Vinga: Extreme Value Statistics Chair: Ivette Gomes
09:00-09:30	Clive Anderson *Extreme Value Theory in metal fatigue*	Luisa Canto E Castro *A class of distribution functions with unbiased estimators for the extreme value index*
09:30-10:00	Pär Johanneson *Extrapolation of fatigue loads*	Fabrizio Laurini *Smoothing sample extremes: the mixed model approach*
10:00-10:30	Anastassia Baxevani *Fatigue life prediction for a vessel sailing the North Atlantic route*	Jana Jurečková *Extreme R-estimator in regression model*

10:30-11:00 Refreshments

11:00-12:30	The Aula, NHV: Fatigue Chair: Lars Holst	Vinga: Climate Chair: Richard Davis
11:00-11:30	Aidan Kerrigan *SKF experience with Extreme Value Analysis on rolling bearing steels and the relation to fatigue properties*	Ana Cebrian *Modelling heat waves using a non homogeneous Poisson process*
11:30-12:00	Allan Gut *Shock models*	Hang Choi *Rethink on the inference of annual maximum wind speed distribution*
12:00-12:30		Viatchelsav Kharin *Changes in temperature and precipitation extremes as simulated in the IPCC multi-model ensemble of global coupled model simulations*

14:30-16:00	The Aula, NHV: Lévy Processes Chair: Robert Adler	Vinga: Software Chair: Johan Segers
14:30-15:00	Vicky Fasen *Extremal behavior of supOU processes*	Alec Stephenson *A survey of software for the analysis of extreme values*
15:00-15:30	Ilya Pavlyukevich *Kramers' type law for Lévy flights*	Alec Stephenson - continued
15:30-16:00	Chang Dorea *Anomalous diffusion index for Lévy motions*	Eric Gilleland *The Extremes toolkit: weather and climate applications of Extreme Value Statistics*

16:00-16:30 Coffee

16:30-18:00	The Aula, NHV: Extreme Value Statistics Chair: Jana Jurečková	Vinga: Software Chair: Johan Segers
16:30-17:00	Natalia Markovich *Smoothing of variable bandwidth kernel estimate of heavy-tailed density function*	Sofia Åberg *WAFO - a MATLAB toolbox for analysis of random waves and loads*
17:00-17:30	Deyuan Li *Tail approximations to the density function in EVT*	Myriam Garrido *The EXTREMES software*
17:30-18:00	Hedieh Jafarpour *A modified measures of kurtosis for heavy tail distributions*	

18:00 Reception with cheese and wine and poster session

19:30 Copulas a discussion
Chair: Klüppelberg. Discussants: Mikosch, de Haan, Drees, Segers, ...

Wednesday, 17 August

08:30-10:00	Chair: Jery Stedinger
08:30-09:00	Robert Adler *Random fields over manifolds*
09:00-09:30	Philippe Naveau *Modeling spatial dependence for extremes in climate studies*
09:30-10:00	Chris Ferro *Blue skies research? Extremes in atmospheric science*

10:00-10:30 Refreshments

10:30-12:00	Chair: Holger Drees
10:30-11:00	Filip Lindskog *On Kesten's counterexample to the Cramér-Wold device for regular variation*
11:00-11:30	Jürg Hüsler *On testing the extreme value conditions*
11:30-12:00	Niels Richard Hansen *Local maximal stack scores with general loop penalty function*

12:00-20:00 Excursion with lunch

Thursday, 18 August

09:00-10:30	The Aula, NHV: Telecommunications Chair: Sid Resnick	Vinga: Hydrology Chair: Philippe Naveau
09:00-09:30	Gennady Samorodnitsky *Poisson Cluster Process as a model for teletraffic arrivals and its extremes*	Benjamin Renard *Bayesian analysis of extremes in hydrology: a powerful tool for knowledge integration and uncertainties assessment*
09:30-10:00	Francois Roueff *Estimation of the long memory parameter using an Infinite Source Poisson model applied to transmission rate measurements*	Pál Rakonczai *Extreme Value Analysis: focusing on the fit and the conditions, with hydrological applications*
10:00-10:30	Krzysztof Dębicki *On a storage process for fluid networks with multiple Lévy inputs*	Maria Isabel Ortego *Log-scaling rainfall data: effects on GPD Bayesian goodness-of-fit*

10:30-11:00 Refreshments

11:00-12:30	The Aula, NHV: Finance Chair: Vicky Fasen	Vinga: Extreme Value Statistics Chair: Armelle Guillou
11:00-11:30	Casper de Vries *Weak & strong financial fragility*	Ion Grama *Adaptive estimation of the excess d.f.*
11:30-12:00	Björn Vandewelle *On univariate extreme value statistics and the estimation of reinsurance premiums*	Cláudia Neves *Statistical inference for heavy and super-heavy-tailed distributions*
12:00-12:30	Gabriel Kuhn *Dimension reduction with heavy tails*	Jan Picek *Testing the tail index in autoregressive models*

14:30-16:00	The Aula, NHV: Extreme Value Theory Chair: Ross Leadbetter	Vinga: Extreme Value Statistics Chair: Niels Richard Hansen
14:30-15:00	Christer Borell *Inequalities for Gaussian measures and Brownian exit times*	Holger Drees *Validation of the Ledford & Tawn Model*
15:00-15:30	Cecile Mercadier *Numerical bounds for the distribution of the maximum of a one- and two-parameter process*	Rene Michel *Estimation of the angular density in multivariate Generalized Pareto Models*
15:30-16:00	Janusz Kawczak *On extreme quantiles in a nilpotent markov chain with application to PRNG testing*	Bojan Basrak *Extreme values, copulas and genetic mapping*

16:00-16:30	Coffee

16:30-18:00	Chair: Francois Roueff
16:30-17:00	Sandra Dias *Large quantile estimation for distributions in the domain of attraction of a max-semistable law*
17:00-17:30	Fernanda Otilia Figueiredo *Comparison of semi-parametric reduced bias' quantile estimators*
17:30-18:00	Salaheddine El Adlouni *Estimation of non-stationary GEV model parameters*

19:00	Conference dinner at Sjömagasinet

Friday, 19 August

09:00-10:30	The Aula, NHV: Spatial problems Chair: Jonathan Tawn	Vinga: Stochastic processes Chair: Henrik Hult
09:00-09:30	Daniel Cooley *A spatial Bayesian hierarchical model to compute a precipitation return levels map*	Pavle Mladenović *On maxima of complete and incomplete samples from stationary sequences*
09:30-10:00	Gabriel Huerta *Time-varying models for extreme values*	Sinisa Stamatovic *Cox limit theorem for high level a-upcrossings by* χ*-process*
10:00-10:30	Marta Nogaj *Analysis of climatic extreme events under non-stationary conditions*	

10:30-11:00 Refreshments

11:00-12:30	The Aula, NHV: Multivariate problems Chair: Filip Lindskog	Vinga: Time series Chair: Clive Anderson
11:00-11:30	Zhengjun Zhang *Asymptotically (in)dependent multivariate maxima of moving maxima processes*	Alexander Lindner *Extremal behaviour of moving average processes with light-tailed innovations*
11:30-12:00	Ishay Weissman *Two dependence measures for multivariate extreme value distributions*	Péter Elek *Extremal cluster characteristics of a regime switching model, with hydrological applications*
12:00-12:30	Ana Ferreira *A simple representation of max-stable processes*	

14:00-16:00	Chair: Laurens de Haan
14:00-14:30	Jonathan Tawn *Practical issues in applications of multivariate extreme values*
14:30-15:00	Ross Maller *Some results on extremal and maximal processes associated with a Lévy process*
15:00-15:30	Henrik Hult *Extreme behaviour for stochastic integrals driven by regularly varying Lévy processes*
15:30-16:00	Sid Resnick *Data network models of burstiness*

16:00-16:30 Coffee

Empirical/Asymptotic p-values for Monte Carlo-based hypothesis testing: an application to cluster detection using the scan statistic

Abrams, Allyson (speaker) *Harvard Medical School and Harvard Pilgrim Health Care, USA*, allyson_abrams@hphc.org
Kulldorff, Martin *Harvard Medical School and Harvard Pilgrim Health Care, USA*
Kleinman, Ken *Harvard Medical School and Harvard Pilgrim Health Care, USA*

Scan statistics; Monte Carlo p-values; empirical; asymptotic:

SaTScan is a freely available software that uses the scan statistic to detect clusters in space, time or space-time. SaTScan uses Monte Carlo hypothesis testing in order to produce a p-value assessing the null hypothesis of no clusters being present. Monte Carlo hypothesis testing can be a powerful tool when asymptotic theoretical distributions are inconvenient or impossible to discover; the main drawback to this approach is that precision for small p-values can only be obtained through greatly increasing the number of Monte Carlo replications, which is both computer-intensive and time consuming.

We ran SaTScan on a sample map using 100,000,000 Monte Carlo replicates in order to generate the 'true' log-likelihood ratio needed to obtain certain p-values. We also ran SaTScan 1000 times on the same map, each time generating 999 Monte Carlo replicates. In each of these 999 replicates the maximum log-likelihood ratio, among all distinct circles, is the statistic reported. The ordinary Monte Carlo p-value is the rank of the observed maximum log-likelihood ratio among the 999 Monte Carlo log-likelihood ratios, divided by 1000.

We found the mamximum likelihood estimates of the parameters of various distributions, assuming the 999 replicates came from that distribution, for each of the 1000 SaTScan runs. The empirical/asymptotic p-value under a given distribution is the area to the right of the observed log-likelihood assuming the estimated parameters for that distribution. For each distribution, we generated: (1) empirical/asymptotic p-values based on the 'true' log-likelihood value and (2) the log-likelihoods that would have been required to generate a specified set of p-values.

Intuitively, an extreme value distribution should be the best fit since the Monte Carlo replicates generate maximum log-likelihood ratios, and in fact the empirical/asymptotic p-values from the Gumbel distribution appear unbiased. In contrast, other tested distributions, including the Gamma, Normal, and Lognormal, all resulted in biased p-values. Interestingly, the ordinary Monte Carlo p-values reported from SaTScan based on 999 Monte Carlo replicates had greater variance than the Gumbel-based p-values.

Empirical/asymptotic p-values can be preferable to true Monte Carlo p-values even when both can be generated from the same set of Monte Carlo replicates. Empirical/asymptotic p-values can also accurately generate p-values smaller than is possible with Monte Carlo p-values with a given number of replicates. We suggest 'Empirical/Asymptotic' p-values as a hybrid method to obtain small p-values with a relatively small number of Monte Carlo replicates and view this as an important and interesting application of extreme value theory.

Random fields over manifolds

Adler, Robert *Technion, Israel*, robert@ieadler.technion.ac.il

Random fields; extrema; manifolds; Lipschitz-Killing curvatures:

I shall start by briefly discussing some statistical problems related to controlling false discovery rates in detecting activity in the brain, which has motivated recent deep results describing the geometry of random fields f on abstract manifolds M, which I shall describe and relate back to the original problem.

The new results centre on the mean Lipschitz-Killing curvatures of the random 'excursion sets'

$$\{t \in M : f(t) \geq u\}, \quad u \in \Re,$$

for which there are now very elegant and explicit formulae for both Gaussian and non-Gaussian random fields over Whitney stratified manifolds.

Furthermore, it has recently been shown, at least in the setting of smooth, centered, constant variance Gaussian processes, that these explicit formula provide an excellent (asymptotic in u) approximation to the extremal probabilities

$$P\left\{\sup_{t \in M} f(t) \geq u\right\}.$$

The error in this approximation has also been shown to be super-exponentially small with an identifiable rate.

The talk will be based on the joint papers [1] and [2], and a full treatment of the subject is currently being put together in [3].

References

[1] Taylor J.E. and Adler R.J. (2002) *Euler characteristics for Gaussian fields on manifolds*, Annals of Probability, 30, 533-563.
[2] Taylor J.E., Takemura A. and Adler R.J. (2005) *Validity of the expected Euler characteristic heuristic,* Annals of Probability, in print.
[3] Adler R.J. and Taylor J.E. (2005) *Random Fields and Geometry*, Birkhaüser, Boston. Most chapters available at `ie.technion.ac.il/Adler.phtml`.

Estimation of non-stationary GEV model parameters

El Adlouni, Salaheddine (speaker) *Institut National de la Recherche Scientifique INRS-ETE, Canada*,
salaheddine_el-adlouni@ete.inrs.ca
Ouarda, Taha B.M.J. *Institut National de la Recherche Scientifique INRS-ETE, Canada*;
Zhang, Xuebin *Environment Canada, Canada*;
Roy, René *OURANOS Consortium on Climate Change, Canada*;
Bobée, Bernard *Institut National de la Recherche Scientifique INRS-ETE, Canada.*

Generalized Extreme Value distribution ; Hydrology ; Maximum likelihood ; Non-stationarity ; Generalized Maximum Likelihood ; Bayesian estimation ; Covariables.

In frequency analysis, data must generally be independent and identically distributed (i.i.d) which implies that they must meet the statistical criteria of independence, stationarity and homogeneity. In reality, the probability distribution of extreme events can change with time, indicating the existence of non-stationarity. The criterion of stationarity can then be jeopardized. The objective of the present study is to develop efficient estimation methods for the use the GEV distribution for quantile estimation in the presence of non-stationarity. Parameter estimation in the non-stationary GEV model is generally done with the Maximum Likelihood Estimation method. In this work, we suggest two other estimation methods: the Generalized Maximum Likelihood Estimation (GML) and the Bayesian approach with a non-informative prior distribution. A simulation study is carried out to compare the performances of these three estimation methods in the case of the stationary GEV model (GEV0), the non-stationary case with a linear dependence of the parameters on covariates (GEV1), and the non-stationary case with a quadratic dependence on covariates (GEV2). The non-stationary GEV model is also applied to a case study from the State of California to illustrate its potential.

Extreme value theory in metal fatigue

Anderson, Clive *University of Sheffield, UK*, c.w.anderson@shef.ac.uk

Fatigue in metals is the deterioration in their load-bearing capability leading to ultimate failure, caused by repeated application of stress. The understanding of fatigue and of ways to avoid it is crucial to safety and reliability in many of the systems fundamental to modern living: in motors and engines, for example, and in aircraft, railways, ships, cars, and in fact in any machine or structure subject to varying stress.

Randomness is intrinsic to fatigue. Loads, environmental conditions and material quality are all variable, and the fundamental mechanisms of fatigue, the initiation and propagation of cracks, are governed by the internal microstructure of the metal, which is naturally described in stochastic terms. *Extreme values* of random quantities enter the study of fatigue in at least two fundamental ways. In the load process it is found that the sequence of local maxima and minima of stress are what determine fatigue life; and at the microstructural level it is the largest cracks, often initiated where internal stresses are highest at inclusions of nonmetallic material, that lead to failure. Extreme value models and statistical procedures are therefore central aids to scientific understanding and prediction of fatigue properties.

The talk will review results and problems in the area, concentrating on aspects of the load process and associated stochastic models; the analysis of test results and prediction of lifetimes and the fatigue limit; estimation of the distribution of large 3-dimensional inclusions from planar and other measurements; and simple stochastic models for the genesis of large inclusions in the light of the physico-chemical processes acting in metal-making.

References

[1] Anderson, C. W., de Maré, J. & Rootzén, H. (2005) *Methods for estimating the sizes of large inclusions in clean steels*, Acta Materialia 53, 2295–2304.
[2] Murakami, Y. (2002) Metal Fatigue: Effects of Small Defects and Nonmetallic Inclusions, Elsevier.

Extreme values, copulas and genetic mapping

Basrak, Bojan *University of Zagreb, Croatia*, bbasrak@math.hr

Genetic mapping, linkage analysis, extreme values, Gaussian processes, copulas.

We describe the method of linkage analysis in human genetics. Linkage analysis is a genetic mapping technique that compares genetic similarity between two individuals to similarity of their physical or psychological traits - *phenotype*. Linkage allows us to find regions of chromosomes that are likely to be associated with a specific trait or phenotype. In this talk we consider the quantitative traits, such as a person's height or cholesterol level.

In human genetics measurements for linkage analysis come from related individuals. For instance, our sample may consist of n pairs of nonidentical twins, for each of whom we have one measurement of the quantitative trait. In addition to that, geneticists also measure degree or relatedness between two twins on many different places along their genome using the concept of *identity by descent - IBD status*. It is an important problem to decide if there is any region on our genome where higher IBD status translates into more similar phenotypic values.

One commonly used approach assumes that the two trait values have bivariate normal distribution conditionally on the IBD status. Under this assumption, the correlation coefficient between the two variables becomes an obvious measure of their similarity. However, many traits are not normally distributed. Further difficulties are caused by the fact that genes sitting close to each other on our genome are usually inherited together. That makes statistical inference more difficult and one has to deal with the issue of multiple testing.

We shall discuss how some of these problems can be resolved using extreme value theory for Gaussian processes and copulas.

References

[1] Basrak, B., Klaassen, C., Beekman, M., Martin, N.G. and Boomsma, D. (2004). *Copulas in QTL mapping*. Behavior Genetics 34, 161–171.
[2] Dupuis, J. and Siegmund, D. (1999). *Statistical methods for mapping quantitative trait loci from a dense set of markers.* Genetics. 151(1). 373–86.
[3] Lander, E.S. and Botstein D. (1989). *Mapping mendelian factors underlying quantitative traits using RFLP linkage maps.* Genetics., 121(1):185–99.

Fatigue life prediction for a vessel sailing the North Atlantic route

Baxevani, Anastassia (speaker) *Lund University, Sweden*, baxevani@maths.lth.se
Rychlik, Igor *Lund University, Sweden*

significant wave height; damage; fatigue; Gaussian random fields:

In this talk, we compute the distribution of the fatigue life of a vessel sailing along the North Atlantic route, (NAr). For simplicity, the load the vessel experiences is assumed to be proportional to the encountered significant wave height, H_s, process. Hence it is important to have an accurate spatio-temporal model for the variability of H_s along the NAr. The field consisting of the H_s values is assumed to be log-normally distributed. The parameters in the spatial model are estimated using the data collected by the TOPEX-Poseidon satellite, while the temporal parammeters are estimated using the velocities the wave systems are moving with as well as the temporal correlation of the H_s time series.

The methodology may be extended for more realistic responses. The proposed method improves the already existing ones, by making use of the information contained in the variance of the fatigue damage accumulated during the voyages. The method is illustrated through numerical examples.

References

[1] Baxevani, A., Rychlik, I. (2005) *Fatigue life prediction for a vessel sailing the North Atlantic route.*, submitted to Probabislistic Mechanical Engineering.
[2] Baxevani, A., Rychlik, I. (2004) *A new method for modelling the space variability of significant wave height.* , submitted to Extremes.

Inequalities for Gaussian measures and Brownian exit times

Borell, Christer *Chalmers University of Technology, Sweden*, borell@math.chalmers.se

Gaussian measure; Brownian motion; exit time; maximum of Brownian motion:

The purpose of this talk is to exhibit some new results for Gaussian measures and Brownian exit times.

If A, B are subsets of $\mathbf{R}^n$ and $\alpha \in \mathbf{R}$, $A + B = \{x + y;\ x \in A,\ y \in B\}$, and $\alpha A = \{\alpha x;\ x \in A\}$. Moreover, γ_n stands for the standard Gaussian measure in $\mathbf{R}^n$, $\Phi(a) = \gamma_1(]-\infty, a[)$, $-\infty \leq a \leq \infty$, and $\Psi(a) = 2\Phi(a) - 1$, $0 \leq a \leq \infty$.

Now suppose $\alpha, \beta > 0$. Then

$$\Phi^{-1}(\gamma_n(\alpha A + \beta B)) \geq \alpha\Phi^{-1}(\gamma_n(A)) + \beta\Phi^{-1}(\gamma_n(B))$$

for all A, $B \in \mathcal{B}(\mathbf{R}^n)$ with positive γ_n-measure if and only if $\alpha + \beta \geq 1$, and $\mid \alpha - \beta \mid \leq 1$. The special case $\alpha + \beta = 1$ is treated in my paper [1] .

Next let $W = (W(t))_{t \geq 0}$ be Brownian motion in $\mathbf{R}^n$ and if C is a domain in $\mathbf{R}^n$, denote by $T_C = \inf\{t > 0;\ W(t) \notin C\}$ the Brownian exit time from C. Then, if $x \in C$, $y \in D$, and $t > 0$,

$$\Phi^{-1}(P_{\alpha x + \beta y}\,[T_{\alpha C + \beta D} > t]) \geq \alpha\Phi^{-1}(P_x\,[T_C > t]) + \beta\Phi^{-1}(P_y\,[T_D > t])$$

for all domains C and D in $\mathbf{R}^n$ and all $\alpha, \beta > 0$ such that $\alpha + \beta \geq 1$, and $\mid \alpha - \beta \mid \leq 1$. In the special case $\alpha = \beta = 1$ a stronger result holds, viz.

$$\Psi^{-1}(P_{x+y}\,[T_{C+D} > t]) \geq \Psi^{-1}(P_x\,[T_C > t]) + \Psi^{-1}(P_y([T_D > t])$$

where equality occurs if C and D are parallel affine half-spaces in $\mathbf{R}^n$. Recall that if H is an open affine half-space in $\mathbf{R}^n$, the Bachelier formula for the maximum of real-valued Brownian motion yields

$$P_x\,[T_H > t] = \Psi(\frac{d(x, H^c)}{\sqrt{t}}),\ t > 0,\ x \in H$$

where $d(x, H^c) = \min_{y \notin H} \mid x - y \mid$.

References

[1] Borell, Ch. (2003) *The Ehrhard inequality*, C. R. Acad. Paris, Ser. I 337, 663–666.

Dynamic conditional correlation models with asymmetric multivariate Laplace innovations

Cajigas, Juan (speaker) *Cass Business School, London UK*, j.p.cajigas@city.ac.uk
Urga, Giovanni *Cass Business School, London UK*

In this paper we propose a multivariate (GARCH) asymmetric generalised dynamic conditional correlation (AGDCC) model where the vector of standardised residuals is assumed to follow an asymmetric multivariate Laplace (AML) distribution. This multivariate distribution is able to capture leptokurtosis and asymmetry which characterise returns from financial assets. It preserves, under general conditions, desirable properties such as finiteness of moments and stability under geometric summation. The empirical validity of this form is tested in the context of a Value-at-Risk (VaR) model. We illustrate the methodology by fitting a sample of 21 FTSE All-World stock indices and 12 bond return indices. We provide clear evidence that in our data set this distribution overwhelmingly outperforms the case in which we assume normality of innovations. :

References

[1] Barndorff-Nielsen, O. (1977) *Exponentially decreasing distributions for the logarithm of particle size*, Proc. Roy. Soc. Lond. A 353, 401-419.

[2] Bauwens, L., and Laurents, S. (2004) *A new class of multivariate skew densities, with application to GARCH models*, working paper.

[3] Bauwens, L., Laurents, S. and J. V. K. Rombouts (2003) *Multivariate GARCH models: A survey*, CORE discussion paper 2003-31. Catholic University of Louvain.

[4] Bertero, E. and Mayer, C. (1990) *Structure and performance: Global interdependence of stock markets around the crash of October 1987*, European Economic Review, 34, 1155-1180.

[5] Billio, M., Caporin, M., and M. Gobbo (2004) *Flexible dynamic conditional correlation multivariate GARCH for asset allocation*, GRETA Working Paper, n.04.03.

[6] Bollerslev, T. (2001) *Financial econometrics: past developments and future challenges*, Journal of Econometrics, 100, 41-51.

[7] Cappiello, L., Engle, R. and K. Sheppard (2004) *Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns*, ECB Working Paper No. 204.

[8] Christoffersen, P. and Pelletier, D. (2003) *Backtesting Value-at-Risk: A Duration-Based Approach*, CIRANO, working paper.

[9] Doornik, J. A. and Hansen, H. (1994) *An omnibus test for univariate and multivariate normality*, unpublished manuscript.

[10] Dufour, J-M. (2004) *Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics*, Working paper, Universite de Montreal

[11] Engle, R. (2001) *Financial econometrics – A new discipline with new methods*, Journal of Econometrics, 100, 53-56.

[12] Heyde, C. C. and Kou, S. G. (2004) *On the controversy over tailweight of distributions*, Operations Research Letters, 32, 399-408.

[13] Johnson, N. L. and Kotz, S. (1972) Distributions in Statistics: Continuous Multivariate Distributions, Wiley, NY

[14] Khindanova, I., Rachev, S., and E. Schwartz. (2001) *Stable modeling of Value at Risk*, Mathematical and Computer Modeling, 34, 1223-1259.

[15] Kotz, S., Kozubowski, T. J., and K. Podgorski (2003) *An asymmetric multivariate Laplace Distribution*, Working paper

[16] Kozubowski, T. J., and Podgorski, K. (2001) *Asymmetric Laplace laws and modeling financial data*, Mathematical and Computer Modelling, 34, 1003-1021.

[17] Mencia, F. J., and Sentana, E. (2004) *Estimation and testing of dynamic models with generalised hyperbolic innovations*, CEMFI working paper No. 0411

A class of distribution functions with unbiased estimators for the extreme value index

Canto E Castro, Luisa (speaker) *University of Lisbon, Portugal*, luisa.loura@fc.ul.pt
de Haan, Laurens *Erasmus University of Rotterdam, Holland*

second order condition; tail quantile process; extreme value index estimation:

Let $X_1, X_2, \ldots$ be i.i.d.random variables with distribution function F. Suppose F is in the domain of attraction of some extreme value distribution. Suppose further that the second order condition holds. Holger Drees (1998) proved under these conditions a very useful expansion for the tail quantile process. Now suppose that the aproach of $(U(tx) - U(t))/a(t)$ (with $U := (1/(1-F))^{\leftarrow}$ and a a scale normalization) to $(x^\gamma - 1)/\gamma$ is not of order $t^{-\rho}$ with some $\rho \leq 0$ as in Drees's Theorem, but that the difference is of lower order than $t^{-\rho}$ for *all* $\rho < 0$. Then a similar result holds but without bias term, i.e., for all $\epsilon > 0$

$$\sup_{0<s\leq 1} \min(1, s^{\gamma+1/2+\epsilon}) \left| \sqrt{k} \left(\frac{X_{n-[ks],n} - X_{n-k,n}}{c\left(\frac{n}{k}\right)^\gamma} - \frac{s^{-\gamma}-1}{\gamma} \right) - s^{-\gamma-1} W_n(s) + W_n(1) \right| = o_p(1)$$

$n \to \infty$, provided there is a $\delta > 0$ with $k = k(n) \to \infty$, $k(n) = o(n^{1-\delta})$. Note that $X_{i:n}$ is the i-th order statistic of $X_1, X_2, \ldots, X_n$, c is a positive constant and W_n is Brownian motion. As a result, under this condition, the known estimators of the extreme value index show no bias for all $k(n)$ with $k(n) = o(n^{1-\delta})$.

References

[1] Drees, H. (1998) *On smooth statistical tail functionals*, Scandinavian Journal of Statististics 25, 187–210.

Modelling heat waves using a non homogeneous Poisson process

Cebrian, Ana C. (speaker) *University of Zaragoza, Spain*, acebrian@unizar.es
Abaurrea, J., Asin, J., and Centelles, A. *University of Zaragoza, Spain*

Poisson process; non homogeneous process; heat wave

Extremely hot events, or heat waves, can have important consequences on agriculture, water resources, energy demand and even on human mortality. Although descriptive analysis of some particular extremely severe events have been done, not much is known about the heat wave general behaviour and how it can change in the near future. The aim of this work is the statistical modelling of heat wave occurrence in order to answer questions such as: 'Are heat waves becoming more frequent? or 'Is heat wave severity increasing over time?'. The answer to these questions becomes even more important due to the observed increase of the temperature level observed during the last years.

The first step for this analysis is to fix the heat wave definition since no generally accepted definition exists. We are interested in extreme heat waves, so we will only consider that a hot period is a heat wave if it yields to an unusual increase of human mortality. High maximum air temperature is the main factor affecting mortality considered in a heat wave, but we also study, empirically, the influence of other factors such as humidity or daily minimum temperature in order to decide the signals that should be taken into account in a heat wave definition.

Finally, we use an operational definition based on the 'Excess over threshold' approach, trying to improve and complete the definition by Kysely (2000). First, a hot spell is defined as a group of consecutive days with daily maximum air temperature over a threshold, and a heat wave is defined as a cluster of dependent hot spells, verifying certain conditions concerning length and temperature during the period. Values for the thresholds defining hot spells an the clustering thresholds are empirically determined.

According to results from Extreme Value theory, the occurrence of excesses over increasing thresholds converges to a Poisson process. Moreover, the use of a point process leads to a likelihood definition that enables a simple formulation of non-stationary processes and allows us to contrast the hypothesis of change in the occurrence rate of the heat waves. Thus, we consider as model a non-homogeneous Poisson process (NHPP), where points occur randomly at a time dependent rate $\lambda(t)$; for the intensity function of the process, we assume a parametric form depending on observed variables $z(t)$, such as time or temperature information. Some tools and adequate residuals are developed to check the validity of these time dependent models.

The model is applied to analyze some data series from about 1950, recorded for the summer period and located in the Ebro river basin, a region in the NE of Spain. The first results suggest that, during last years, there is some evidence of increase of the heat wave occurrence, but it is not linear in time. It is also observed that the evolution of the heat wave occurrence rate is very close to the one of the maximum temperature signal; this contributes to the better fit provided by the models including temperature information, which are also simpler.

The models including temperature information can be used not only to establish the existence of changes in the occurrence of heat waves during the considered period but also for predicting the future evolution of the heat wave occurrence, using as input the temperature projections of general circulation models, GCMs. A GCM is a model whose purpose is to numerically simulate changes in different climate signals, such as temperature, resulting from slow changes in some boundary conditions; it allows to predict long term changes, concerning mean and variability.

References

[1] Kysely, J. (2002). *Temporal fluctuations in heat waves at Prague, the Czech republic, from 1901-97, and their relationship to atmospheric circulation.* Int. J. Climatol. 22, 33-50.

Rethink on the inference of annual maximum wind speed distribution

Choi, Hang (speaker) *formerly The University of Tokyo, JAPAN*, number3choi@unitel.co.kr
Kanda, Jun *The University of Tokyo, JAPAN*

annual maximum wind speed; non-stationarity; Monte Carlo simulation; the law of large number:

The distribution of annual maximum wind speed plays an important role in the field of structural engineering and the classical theory of extreme value distribution has been utilized for the appropriate inference of the distribution. However, despite long history of utilization we are still discussing the appropriateness of the results and the applicability of the theory is frequently viewed in doubt because of the differences between the idealized theoretical framework and the complex characteristics of natural phenomena. A representative example is independent and identically distributed random variable (i.i.d. rv) assumption in the theory. Whether the assumption holds has not been considered in the application of the theory. If annual wind speed does not follow unique probability distribution aside from the independency of extremes, the application of classical theory may not give reasonable inference results because the class of extreme value distribution in non-i.i.d. case is much larger than that in i.i.d. case [1].

In this presentation the non-stationarity of wind speed fluctuation is discussed based on the statistical analysis results of past 42 years observation records at 155 meteorological observatories in Japan. The Monte Carlo simulation results considering non-identical parent distributions [2] are also compared with the representative historical annual maximum wind speed data from 1929 to 1999 compiled by Ishihara *et al.*[3] and newly extracted data from 2000∼2002 observatory records. From the comparison, the simulation results shows different type of quantile function compared with those defined in the classical theory and the simulated quantile functions coincide well with the empirical qunatile functions based on the historical data in normalized form. However, the estimated attraction coefficients reflect that the appropriateness of estimation strongly depends on the number of extremes according to the law of large number. On the other hand, if the annual wind speed distribution is assumed as the mean distribution of annual distributions, which is usually assumed in the practice, the variance of extremes is significantly underestimated but the mean of extremes is insensitive to the assumption.

References

[1] Falk, M., Hüsler, J. and Reiss, R.-D. (1994) Laws of Small Numbers: Extremes and Rare Events, Birkhäuser
[2] Choi, H. and Kanda, J. (2005) *An approach to the extreme value distribution of non-stationary process*, abstract of Cherry Bud Workshop 2005 Quantitative Risk Management: Theory and Practice, Keio University, Japan, 6-8
[3] Ishihara, T. *et al.* (2002), *A database of annual maximum wind speed and corrections for anemometers in Japan*, Wind Engineers 92, JAWE, 5-54 (in Japanese)

Large deviation estimates for certain heavy-tailed dependent sequences arising in risk management

Collamore, Jeffrey F. (speaker) *University of Copenhagen, Denmark*, collamore@math.ku.dk
Höing Andrea *ETH Zürich, Switzerland*

Keywords: Financial risk, subexponential distributions, Harris recurrent Markov chains, repetitive operational risk, GARCH processes.

We consider the large deviation behavior of the sums,

$$S_n = F(X_1) + \cdots + F(X_n),$$

where $\{X_i\}$ is a general Harris recurrent Markov chain, F is a random function, and the tail decay of $F(X_i)$ is subexponential. Motivated by certain problems in risk management, we are particularly interested in the following two cases:

(i) The process $\{X_i\}$ is "light-tailed," so that large exceedences arise as a result of large jumps of $F(\cdot)$. This parallels the finite-state space setting, and could be used as a model for "repetitive" operational risk losses.

(ii) $F(X_i) = X_i$, where $\{X_i\}$ is a sequence of random variables satisfying a stochastic recurrence equation,

$$X_i = A_i X_{i-1} + B_i, \tag{1}$$

where $\{(A_i, B_i)\} \subset \mathbf{R}_+^2$ is i.i.d. For example, $\{X_i\}$ could denote a GARCH(1,1) financial process.

In either case, we establish a "small-time" ruin estimate, namely,

$$\mathbf{P}\{S_n \geq u, \text{ some } n \leq \delta u\} \sim Cu\bar{F}(u) \text{ as } u \to \infty,$$

where F denotes the distribution of $\{S_n\}$ attained over its regenerative cycles. Furthermore, we show that for both (i) and (ii),

$$\bar{F}(u) \sim D\mathbf{P}_\pi \{F(X) > u\} \text{ as } u \to \infty,$$

where π denotes the stationary measure of $\{X_i\}$, and D is a constant which is actually different in the two separate cases—as is the method of proof. (In particular, for (ii) the large exceedence results from a "build-up" in the multiplicative terms, A_i.)

In the setting of (ii), we note that related results have recently been obtained in [2], but under a dominant tail assumption on the B_i-terms in (1). This leads to tail behavior which is different from either of the two cases considered here.

The results in this talk have recently been obtained in [1].

References

[1] Collamore, J. F. and Höing, A. (2005). Small-time ruin for a financial process modulated by a Harris recurrent Markov chain. Submitted.

[2] Mikosch, T. and D. Konstantinides. (2005). Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations. To appear in *Ann. Probab.*

A spatial Bayesian hierarchical model to compute a precipitation return levels map

Cooley, Daniel (speaker) *University of Colorado at Boulder, USA*, Daniel.Cooley@colorado.edu
Nychka, Douglas *National Center for Atmospheric Research, Boulder, CO, USA*
Naveau, Philippe *Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, France*

Bayesian hierarchical model, precipitation, spatial

Quantification of extreme values is important for planning purposes. To aid with the understanding of potential flooding along Colorado's Front Range, we are developing a map of precipitation return levels for the region.

We model daily precipitation above a high threshold at 56 weather stations throughout the region with the generalized Pareto distribution (GPD). Each station's GPD parameters are modeled within a Bayesian hierarchical structure which allows us to pool the data from all the stations. GPD parameter estimates also take into account geographical covariates such has elevation or mean annual precipitation. This strategy yields parameter and return-level estimates which have more spatial consistency. The Bayesian model allows us to estimate the spatial structure between the parameters at each station, and once the spatial structure is estimated, it is used to interpolate over the entire region. These region-wide parameter estimates can then be converted into the desired return levels.

Model inference is obtained using a straightforward MCMC algorithm, through which draws for the posterior distribution are obtained. These draws yield a natural method for obtaining uncertainty estimates for the precipitation return levels.

The flexibility of the Bayesian hierarchical structure allows us to test different models which can be compared. The model testing and comparison process provides meteorologists insight into how extreme precipitation behaves in Colorado.

References

[1] Casson, E. and Coles, S.G. (1999) *Spatial regression models for extremes*, Extremes 1, 449-468.
[2] Poncet, P., Naveau P., and Cooley, D. (2005) *Variograms for Spatial Max-stable Random Fields*, Manuscript.

On a storage process for fluid networks with multiple Lévy inputs

Dębicki, Krzysztof *University of Wroclaw, Poland*, debicki@math.uni.wroc.pl

Lévy process; Skorokhod problem; Stochastic network; Storage process :

We consider a fluid stochastic network with spectrally positive Lévy input. Under some conditions on the geometry of the network, we derive the Laplace transform of the joint steady-state workload and running busy periods distribution.

Moreover we present a new representation for the steady state distribution of the workload of the second queue in a two-node tandem network. It involves the difference of two suprema over two adjacent intervals. Additionally we obtain the exact distribution of the workload in the case of Brownian and Poisson input, as well as some insightful formulas representing the exact asymptotics for α-stable Lévy inputs.

The talk is based on works [1], [2] jointly written with Ton Dieker, Michel Mandjes, Miranda van Uitert and Tomasz Rolski.

References

[1] Dębicki, K., Mandjes, M., Uitert, M. (2005) *A Tandem Queue with Lévy Input: A New Representation of the Downstream Queue Length*, Mittag-Leffler Institute Report 10/2004/2005; submitted for publication.
[2] Dębicki, K., Dieker, A. B., Rolski, T. (2005) *On fluid networks for multiple Lévy inputs*, in preparation.

Large quantile estimation for distributions in the domain of attraction of a max-semistable law

Dias, Sandra (speaker) *University of Trás-os-Montes e Alto Douro, Portugal*, sdias@utad.pt
Canto e Castro, L. *University of Lisbon, Portugal*

max-semistable domains of attraction; geometrically growing sequences; ratios of differences of order statistics:

Let G be a distribution function with $G(0) = e^{-1}$, $G(1) = \exp(-r^{-1})$ and continuous at $x = 0$. As can been seen in Canto e Castro *et al.* (2000), G is max-semistable if and only if for every $x \in [0, 1]$

$$-\log(-\log G(a^m x + s_m)) = m \log r + y(x),\ m = 0, \pm 1, \pm 2, ...,$$

for some constant $a > 0$, some constant $r > 1$ and some function y defined in $[0, 1]$ with values in $[0, \log r]$, non decreasing, right continuous and continuous at $x = 1$. The sequence s_m is given by $s_m = (a^m - 1)/(a - 1),\ a \neq 1$ and $s_m = m,\ a = 1$. From this representation estimators of large quantiles follow easily as a function of the estimators of the parameters a and r and of the function y.

To apply the results we used simulated data according to the fact that the distribution function of the waiting time in non-homogeneous Poisson processes with periodic (or log-periodic) intensity function is in a max-semistable domain of attraction. We also present a real data study consisting in the analysis of the major earthquake inter-arrival times registered in the period between January 1st, 1973 and March 31st, 2005.

References

[1] Canto e Castro, L., de Hann, L., Temido, M.G. (2000) *Rarely observed sample maxima.*, Theory of Probab. Appl. 38, 779–782.

Anomalous diffusion index for Lévy motions

Dorea, Chang C.Y. (speaker) *University of Brasilia, Brazil*, cdorea@mat.unb.br
Medino, Ary V. *University of Brasilia, Brazil*

Anomalous diffusions; Lévy motions; Diffusion index:

In modelling complex systems as diffusion processes it is common to analyse its diffusive regime through the study of approximating sequences of random walks. For the partial sums $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ one considers the approximating sequence of processes $X^{(n)}(t) = a_n \left(S_{[k_n t]} - b_n\right)$. Where b_n are centering constants, a_n perform a scaling of the state variable and k_n perform the required time scaling. Then, under sufficient smoothness requirements we have the convergence to the desired diffusion, $X^{(n)}(t) \to X(t)$. A key assumption usually presumed is the finiteness of the second moment, and, hence the validity of the Central Limit Theorem. A class of distributions that possesses similar properties are the stable distributions that result as limits of properly stabilized sums of random variables. The asymptotic behavior of $\dfrac{S_n - b_n}{a_n}$ may well be non-Gaussian and $\dfrac{1}{n}E(S_n^2) \to \infty$. Such random walks have been referred by physicists as Lévy motions or Lévy flights. And the following limit has been used to classify different diffusive regimes, $D_X = \lim_{t\to\infty} \dfrac{E(X^2(t))}{2t}$ ($D_X = 0$ subdiffusion; $0 < D_X < 0$ normal diffusion; and $D_X = \infty$ superdiffusion). In this work we introduce an alternative notion to classify these regimes, the diffusion index γ_X, that constitutes a refinement of the diffusion constant D_X. For some γ_X^0 properly chosen let $\gamma_X = inf\{\gamma : 0 < \gamma \leq \gamma_X^0\ ,\ \limsup_{t\to\infty} \dfrac{E|X(t)|^{1\gamma}}{t} < \infty\}$. Relationship between γ_X, the infinitesimal diffusion coefficients and the diffusion constant D_X will be explored. Illustrative examples as well as estimates, based on extreme order statistics, for γ_X will also be discussed.

References

[1] Dorea, C.C.Y. and Medino, A.V. (2005) *Anomalous diffusion index for Lévy Motions*, Universidade de Brasilia, Brazil, preprint.
[2] Costa, I.V.L.; Morgado, R.; Lima, M. V. B. T. and Oliveira, F. A. (2003) *The Fluctuation-Dissipation Theorem fails for fast superdiffusion*, Europhysics Letters, vol. 63, 173-179.
[3] Metzler, R. and Klafter, J. (2000) *The random walk's guide to anomalous diffusion: a fractional dynamics approach*, Physics Reports, vol. 339, 1-77.

Research partially supported by CNPq, CAPES/PROCAD, FAPDF/PRONEX and FINATEC/UNB.

Validation of the Ledford & Tawn Model

Drees, Holger (speaker) *University of Hamburg, Germany*, drees@math.uni-hamburg.de
Müller, Peter *University of Hamburg, Germany*

Ledford & Tawn model; model validation; dependence structure:

The approach proposed by Leford and Tawn (1996, 1997) has proved to be useful for modelling the dependence structure of the joint extremes of bivariate random vectors, if one can take neither asymptotic dependence nor exact independence of the components for granted. For example, the large claims in different lines of business of a non-life insurer often exhibit a clear positive dependence which vanishes asymptotically when one considers the exceedances over increasing thresholds.

Let (X,Y) denote the bivariate random vector with marginal df's F_X and F_Y. The central assumption of the Ledford & Tawn model can be reformulated as

$$\frac{P\{1-F_X(X)<tx,\ 1-F_Y(Y)<ty\}}{P\{1-F_X(X)<x,\ 1-F_Y(Y)<y\}} \longrightarrow t^{1/\eta}$$

as $x, y \to 0$ (in a suitable way). We discuss a graphical tool to evaluate the accuracy of this "scaling law". In analogy to the well-known Hill pp-plot, to this end one checks whether differences of the logarithm of certain empirical probabilities lie approximately on a certain plane. In addition to this purely data-analytic tool, we derive pointwise asymptotic confidence intervals which enables us to check whether the observed deviations from the ideal plane can be explained by random effects or whether they indicate that the model assumptions are violated. These asymptotic results are based on approximations to certain empirical processes established by Draisma et al. (2004).

References

[1] Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004). *Bivariate tail estimation: dependence in asymptotic independence.* Bernoulli 10, 251–280.
[2] Ledford, A. and Tawn, J. (1996). *Statistics for near independence in multivariate extreme values.* Biometrika 83, 169–187.
[3] Ledford, A. and Tawn, J. (1997). *Modelling dependence within joint tail regions.* J. Roy. Statist. Soc. Ser. B 59, 475–499.

Extremal cluster characteristics of a regime switching model, with hydrological applications

Elek, Péter (speaker) *Eötvös Loránd University, Hungary*, elekpeti@cs.elte.hu
Vasas, Krisztina *Eötvös Loránd University, Hungary*, v_krisz@ludens.elte.hu
Zempléni, András *Eötvös Loránd University, Hungary*, zempleni@ludens.elte.hu

aggregate excess; cluster length; conditional heteroscedasticity; Markov chain; regime switching model

Studying extremal characteristics - other than the tail behaviour - of stochastic processes is often of very much importance. These characteristics include the duration of clusters beyond high thresholds or the distribution of aggregate excesses during extremal events. However, apart from the estimation of the extremal index, the general theory is of little help in determining these quantities. Empirically, there is a relationship between the cluster maxima and the aggregate excesses, which one may exploit by fitting a transformed generalised Pareto distribution, suggested by [1]. Another way to resolve the problem is to adopt a Markovian (see e.g. [2] and [3]) or a k-th order Markovian framework for the threshold exceedances, which then allows the simulation of extremal cluster functionals.

Alternatively, one may choose a time series model - possibly motivated by physical considerations - to describe the whole data generating process. After parameter estimation such a model allows the simulation of cluster functionals or even the exact calculation of their distribution. We examine two families of such models, suitable for the hydrological datasets we analysed. Our light-tailed conditionally heteroscedastic model ([4]) - although its theoretical extremal index is equal to one - has desirable subasymptotic properties so it provides realistic flood length and flood volume simulations. The other family is the regime switching family, more often used in the hydrological literature. We develop a baseline Markov-switching autoregressive model with state-dependent innovations and autoregressive coefficients, whose extremal index, distribution of cluster lengths and distribution of aggregate excesses can be calculated exactly. Extremal behaviour of some extensions of this model to allow other types of innovations or non-Markovian state transitions are also examined.

We apply and compare the above methods to Hungarian river discharge data.

References

[1] Anderson, C. W. and Dancy, G. P. (1992) *The severity of extreme events* , Research Report 409/92, University of Sheffield
[2] Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) *Markov chain models for threshold exceedances*, Biometrika 84, 249–268.
[3] Bortot, P. and Tawn, J. A. (1998) *Models for the extremes of Markov chains*, Biometrika 85, 851–867.
[4] Elek, P. and Márkus, L. (2005) *A light-tailed conditionally heteroscedastic model with applications to river flows*, in preparation.

Extremal behavior of supOU processes

Fasen, Vicky *Munich University of Technology, Germany*, fasen@ma.tum.de

cluster; extreme value theory; independently scattered random measure; long range dependence; subexponential distribution; supOU process; tail behavior:

We study the extremal behavior of superpositions of Ornstein-Uhlenbeck (supOU) processes

$$Y(t) = \int_{R_+ \times R} \exp(-r(t-s))\, d\Lambda(r,s) \quad \text{for } t \in R,$$

where Λ is an infinitely divisible independently scattered random measure. Under specific conditions this class of processes introduced by Bandorff-Nielsen and Shephard for modelling volatility processes exhibits long range dependence and models upward jumps. Depending on the tail behavior of the Lévy process $L(t) = \Lambda(R_+, [0,t])$ we show that they are heavy tailed and model clusters on high levels. We restrict our attention to subexponential Lévy processes. The extremal behavior is modelled as marked point processes at a properly chosen discrete-time skeleton by the jump times of the Lévy process L. We obtain also convergence of partial maxima.

References

[1] Fasen, V. (2005) *Extremes of regular varying mixed moving average processes*, submitted for publication.
[2] Fasen, V. (2005) *Extremes of subexponential Lévy driven moving average processes*, submitted for publication.

A simple representation of max-stable processes

Ferreira, Ana (speaker) *ISA, Technical University of Lisbon, Portugal*, anafh@isa.utl.pt
de Haan, Laurens *Erasmus University Rotterdam, The Netherlands*

Keywords: Max-stable process; Poisson point process

Let S be a compact subset of a Euclidean space and $C(S)$ the space of continuous functions f on S equipped with the supremum norm $|f|_\infty = \sup_{s\in S}|f(s)|$. Let ξ be a stochastic process on $C(S)$ with non-degenerate marginals, i.e. $\xi(s)$ is non-degenerate for all $s \in S$. The process ξ is max-stable if there exist continuous functions $a_n > 0$ and b_n, defined on S, such that if $\xi_1, \xi_2, \cdots, \xi_n$ are independent and identically distributed copies of ξ,

$$\left\{ \bigvee_{i=1}^{n} \frac{\xi_i(s) - b_n(s)}{a_n(s)} \right\}_{s\in S} \stackrel{d}{=} \{\xi(s)\}_{s\in S} .$$

The probabilistic structure of those processes is fully captured, through a transformation of the marginal distributions, by a corresponding standardized max-stable process. We shall give a simple representation of the standardized max-stable process.

References

[1] E. Giné, M. G. Hahn and P. Vatan (1990) *Max-infinitely divisible and max-stable sample continuous processes*, Probab. Th. Rel. Fields 73, 139–165.

Blue skies research? Extremes in atmospheric science

Ferro, Chris *University of Reading, UK*, c.a.t.ferro@reading.ac.uk

What do climate scientists mean by ‘extreme events’ and what questions do they want to answer? I shall review some relevant problems in atmospheric science and describe how extreme-value analysis might help to solve them.

Comparison of semi-parametric reduced bias' quantile estimators

Figueiredo, Fernanda Otilia (speaker) *Faculdade de Economia da Universidade do Porto and CEAUL, Portugal*, otilia@fep.up.pt
Gomes, Maria Ivette *CEAUL and DEIO (FCUL), Universidade de Lisboa, Portugal*

Heavy tails; High quantiles; Semi-parametric estimation; Bias reduction.

One of the most important issues in Statistics of Extremes is the statistical modelling of rare events, and consequently the estimation of high quantiles. The estimation of a high quantile, a value which is exceeded with a small probability p, is very important in many areas of research, for instance in Insurance, in Finance and in Statistical Quality Control. This issue has already been addressed by several authors; however, the widely used classical quantile estimator usually provides a high bias, and in order to reduce it, we propose some alternative estimators. We are here going to base quantile estimation either on adequate reduced bias' tail index estimators, like the ones used in Gomes and Figueiredo (2003), or on an adequate direct accommodation of bias of high quantiles, like in Matthys and Beirlant(2003), or on both techniques.

In this paper we assume to be working in a context of heavy-tailed models with a positive tail index γ. More specifically we shall work in Hall's class of models (Hall and Welsh, 1985), i.e., we shall assume that there exist $\gamma > 0$, $\rho < 0$, $C > 0$ and $\beta \neq 0$, such that

$$U(t) := F^{\leftarrow}(1 - 1/t) = C\, t^{\gamma} \left(1 + \gamma\, \beta\, t^{\rho}/\rho + o(t^{\rho})\right), \qquad \text{as } t \to \infty. \tag{2}$$

We shall base inference on the largest k top order statistics (o.s.), where k is an intermediate sequence of integers. We consider different approximations for the quantile function in (2), and we are led to different quantile estimators. Given a tail index estimator $\widehat{\gamma}(k)$, we easily derive the associated quantile estimator,

$$Q_{\widehat{\gamma}}^{(p)}(k) := X_{n-k:n} \left(k/(np)\right)^{\widehat{\gamma}(k)}. \tag{3}$$

The quantile estimation depends thus heavily on the tail index estimation. For heavy tails, the classical tail index estimator is the Hill estimator, $H(k)$, the average of the scaled log-spacings $U_i := i\left\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\right\}$, $1 \le i \le k$, and if we plug $H(k)$ in (3), we are led to the classical quantile estimator. Since the Hill estimator exhibits usually a strong bias for moderate k and sample paths with very short stability regions around the target value γ, researchers have recently considered the possibility of dealing with the bias term in an appropriate way, building new estimators, $\widehat{\gamma}_{R}(k)$ say, the so-called second order reduced bias' tail index estimators. Such a tail index estimator may thus be plugged in (3) in order to reduce the bias of the quantile estimator. Matthys and Beirlant (2003) try also to reduce the bias of the classical quantile estimators, going directly into the second order framework, and suggesting the consideration of the estimator

$$\overline{Q}_{\widehat{\gamma}}^{(p)}(k) := X_{n-k:n} \left(k/(np)\right)^{\widehat{\gamma}(k)} \exp\left(\widehat{\gamma}(k)\widehat{\beta}\,(n/k)^{\widehat{\rho}} \left(\left((k/(np))^{\widehat{\rho}} - 1\right)/\widehat{\rho}\right)\right). \tag{4}$$

It is known (Gomes and Figueiredo, 2003) that the use of a reduced bias' tail index estimator $\widehat{\gamma}_{R}$ in (3) provides better results than the use of the classical Hill estimator H. The obvious question, that we shall try to answer both theoretically and computationally, is the following: is it better to work with

1. the estimator in (3) and a reduced bias tail index estimator $\widehat{\gamma}_{R}$ of γ,
2. the estimator in (4) and a classical estimator of γ, like the Hill estimator H,
3. or the estimator in (4) and a reduced bias tail index estimator $\widehat{\gamma}_{R}$ of γ?

References

[1] Gomes, M.I., Figueiredo, F. (2003) *Bias reduction in risk modelling*, Notas e Comunicações, CEAUL 12/2003. To appear in Test.
[2] Hall, P., Welsh, A.H.(1985) *Adaptive estimates of parameters of regular variation*, Ann. Statist 13, 331-341.
[3] Matthys, G., Beirlant, J. (2003) *Estimating the extreme value index and high quantiles with exponential regression models*, Statistica Sinica 13, 853-880.

Least-squares estimation of a convex function

Fils-Villetard, Amelie (speaker) *Université Pierre et Marie Curie, France*, fils@ccr.jussieu.fr
Guillou, Armelle *Université Pierre et Marie Curie, France*
Segers, Johan *Tilburg University, the Netherlands*

Projection theory; Least-squares estimation; Constrained estimates; Pickands dependence function.

We consider the problem of estimating a convex function θ_0 when the function space Θ is a closed and convex subset of a real Hilbert space $\mathcal{H}$. Typically, $\mathcal{H}$ is an L^2-space on some real interval and elements of Θ are square-integrable functions subject to convexity constraint. The function θ_0 of interest could for instance be a density function, a regression function or the Pickands dependence function of an extreme-value copula.

Given an initial estimator $\hat{\theta}$ of the function θ_0, we define our new estimator as the orthogonal projection of $\hat{\theta}$ on the function space Θ, that is we define the least-squares estimator of θ_0 by

$$\hat{\theta}^{ls} = \arg\min_{\theta \in \Theta} \|\hat{\theta} - \theta\|.$$

Although the initial estimate $\hat{\theta}$ may not belong to Θ, by definition we always have $\hat{\theta}^{ls} \in \Theta$. Moreover, since projections are non-expansive, $\|\hat{\theta}^{ls} - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$. So by projecting the initial estimate, $\hat{\theta}$, onto Θ, we obtain a new estimate satisfying all the constraints imposed by Θ and which is at least as accurate.

Our method extends the theory already existing in the literature on least-squares estimation of convex density or regression functions (see e.g. Groeneboom, Jongbloed and Wellner 2001), and yields new estimators of, for instance, Pickands dependence functions of extreme value copulas.

From the knowledge of the rate of convergence of $\hat{\theta}$, we deduce the one of $\hat{\theta}^{ls}$. If the asymptotic behavior of $\hat{\theta}$ is given, it can be transferred to $\hat{\theta}^{ls}$ using the delta-method, since orthogonal projections on closed and convex subsets are one-sided Hadamard differentiable.

Geometric characterizations of the projection estimator $\hat{\theta}^{ls}$ lead to algorithms for actual computation of the estimator. Alternatively, approximations to $\hat{\theta}^{ls}$ can be computed by solving a quadratic program under contraints.

References

[1] Groeneboom, P., Jongbloed, G., Wellner, J. (2001) *Estimation of a convex function: characterizations and asymptotic theory*, Annals of Statistics 29, 1653–1698.

Reduced Bias Semi-parametric Quantile Estimators with a linear-type property

Fraga Alves, M.I. (speaker) *Universidade de Lisboa, CEAUL and DEIO (FCUL), Portugal*, isabel.alves@fc.ul.pt
Gomes, M. Ivette *Universidade de Lisboa, CEAUL and DEIO (FCUL), Portugal*
Santos, P. Araújo *Politécnico de Santarém, ESGS, Portugal*

Extreme Quantiles; Semi-parametric Estimation; Reduced Bias; Shifted Data:

In practical applications to fields such as finance, environment or insurance, involving statistical analysis of extreme values, we are often interested in the estimation of a *high quantile*, χ_p, a level which is exceeded with a small probability p. Here we deal with the estimation of χ_p, based on the k largest observations from a random sample of size n, from an underlying random variable X, with a heavy-tailed distribution. Fraga Alves and Araújo Santos (2004) proposed a simple modification of the Weissman-type estimator (Weissman, 1978) which enjoys a desirable property in the presence of linear transformations of the data, in accordance with the empirical counterpart of the theoretical property for the distribution quantiles, namely, $\chi_p(aX + b) = a\chi_p(X) + b$. In that paper the tail index γ is estimated through the use of classical semi-parametric estimators, which exhibit high bias for low thresholds. Here, instead of using classical semi-parametric tail index estimators, we follow the approach introduced by Gomes and Figueiredo (2002) and incorporate "asymptotically unbiased" tail index estimators, in order to improve the performance of the modified Weissman estimator. The exact performance of the new estimators will be compared with the classical semi-parametric quantile estimators, through Monte Carlo simulation techniques.

References

[1] Fraga Alves, M.I. and Araújo Santos, P. (2004) *Extreme quantiles estimation with shifted data from heavy tails*, Nota e Comunicações CEAUL 11/2004.
[2] Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2003) *A new class of semi-parametric estimators of the second order parameter*, Portugaliae Mathematica, vol.60, 2, 193–213.
[3] Gomes, M.I. and Figueiredo, F. (2002) *Bias reduction in risk modelling: semi-parametric quantile estimation.* To appear in Test.
[4] Weissman, I. (1978) *Estimation of parameters and large quantiles based on the k largest observations*, J. Amer. Statist. Assoc., 373, 812–815.

The EXTREMES software

Girard, Stéphane *SMS/LMC-IMAG, Université Grenoble 1, France*, Stephane.Girard@imag.fr
Garrido, Myriam (speaker) *ENAC / LSP, U. Toulouse 3, France*
Ecarnot, Jérôme *is2, INRIA Rhône-Alpes, France*

Distribution Tail, Extreme Quantiles, Goodness-of-fit Test, Bayesian Statistics:

The EXTREMES software gathers different tools dedicated to extreme values study and more precisely to extreme quantiles estimation and model selection for distribution tails. It is written in C++ with a graphical user interface developped with Matlab. This solution matches rapid execution and user-friendliness. Avaible functions can be grouped in three classes:

1) Usual statistical functions. These functions are not dedicated to extreme value study: sample simulation, ploting distribution related functions, parameter estimation, non parametric estimation of density, parametric estimation of quantiles, Anderson-Darling or Cramer-von Mises test.

2) Usual functions for extreme value analysis. These are well known functions for estimation and test in extreme value analysis context.
– Checking excess exponentiality. The goal is to check if the data distribution is in the Gumbel maximum domain of attraction and if the number of excesses is well choosen. Exponentiality of the excesses is graphically checked drawing a qq-plot. A test is also proposed.
– Estimation of Generalized Pareto Distribution parameters. Some usual methods are implemented: Hill (19), Dekkers *et. al.* (1989), Hosking & Wallis (1987), maximum likelihood and Zipf (Beirlant *et. al.*, 2002) estimates.
– Extreme quantiles estimation using POT method and the previous estimates.

3) New procedures
– The GPD test is a goodness-of-fit test for the distribution tail of usual global models belonging to all the maximum domains of attraction (Gumbel, Weibull and Fréchet). We compare the parametric estimate using the global model and the POT method estimate of an extreme quantile. For the POT estimation, different estimates exist (Hill, Dekkers ...) leading to different tests.
– The ET test is a particular case of the GPD test for which we suppose the data distribution is in the Gumbel maximum domain of attraction. To compute the POT estimate, we then use an exponential approximation of the distribution tail.
– Bayesian regularisation procedure is a method to improve the extremal fit of previous models using an expert opinion on distribution tail.
When one wants to know the data distribution both in central (most likely) and extremal ranges, an usual model can be looked for. Central fit is checked by usual tests like Anderson-Darling or Cramer-von Mises. Then the GPD (or ET) test allows to check extremal fit of these models. If no distribution is accepted both by central and extremal tests, the bayesian regularisation procedure can improve the extremal fit of a central adapted model.

acknowledgements: Financial support from EDF is gratefully acknowledged.

The Extremes toolkit: weather and climate applications of extreme value statistics

Gilleland, Eric (speaker) *National Center for Atmospheric Research (NCAR), U.S.A.*, ericg@ucar.edu
Katz, Richard W. *National Center for Atmospheric Research (NCAR), U.S.A.*

R software; graphical user interface (GUI); weather and climate applications; :

The *Extremes Toolkit* has been developed with funding from the Weather and Climate Impact Assessment Science Initiative undertaken at the National Center for Atmospheric Research (NCAR). Weather and climate impact assessment often involves estimation of the probability of occurrence and distribution of the severity of events that might be very rare if not unprecedented: for example, the health impacts from a heat wave, such as happened recently in Europe; or the economic damage from a hurricane. The extremes toolkit and the accompanying tutorial [3] are intended to provide atmospheric scientists, as well as other researchers investigating weather and climate extremes and their environmental and societal impact, with an easy-to-use interface to extreme value software. In addition, it is desired to accomplish this goal without needing to learn a new programming language and to reduce the learning curve for the application of the theory of extreme values.

The software is provided as a package on CRAN [4] called `extRemes`, and is essentially a graphical user interface (GUI) for Stuart Coles' S-Plus routines ([1] and [2]) as ported to R by Alec Stephenson as the `ismev` package.

References

[1] Coles, Stuart (2001). An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.
[2] Coles, Stuart, (2001). S-plus functions for extreme value modeling: An accompaniment to the book An introduction to statistical modeling of extreme values.`http://www.stats.bris.ac.uk/~masgc/ismev/uses.ps`
[3] Gilleland, Eric and Katz, Richard W. (2005). Tutorial to the Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics, `http://www.assessment.ucar.edu/toolkit`
[4] R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Second order reduced bias tail index estimators under a third order framework

Gomes, M. Ivette (speaker) *Universidade de Lisboa, Portugal*, ivette.gomes@fc.ul.pt
Martins, M. João and Neves, M. Manuela *Universidade Técnica de Lisboa, Portugal*

Statistics of Extremes; Semi-parametric estimation; Heavy tails; Maximum likelihood:

Heavy-tailed models are quite useful in most diversified fields, like telecommunication networks and finance. Power laws, such as the Pareto income distribution and the Zipf's law for city-size distribution, have been observed a few decades ago in some important phenomena in economics and biology and have seriously attracted scientists in recent years. We shall consider here *heavy-tailed* models F, i.e., we shall assume that the quantile function $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$, is of regular variation with index γ.

For intermediate k, we shall consider, as basic statistics, the scaled log-spacings $U_i := i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$, $1 \leq i \leq k < n$, where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic, $1 \leq i \leq n$, associated to a random sample $(X_1, X_2, \cdots, X_n)$. We shall further assume that there exist functins $A(t)$ and $B(t)$ such that $\lim_{t\to\infty}\ \{(\ln U(tx) - \ln U(t) - \gamma \ln x)\,/A(t) - (x^\rho - 1)\,/\rho\}\,/B(t) = \left(x^{\rho+\rho'} - 1\right)/(\rho+\rho')$, for every $x > 0$, where $|A(t)$ and $|B(t)|$ must then be of regular variation with index ρ and ρ', respectively, $(\rho,\ \rho' \leq 0)$, and that we are in a sub-class of Hall's class of models, where we may choose $A(t) = \alpha\ t^\rho =: \gamma\ \beta\ t^\rho,\ B(t) = \alpha'\ t^{\rho'} =: \beta'\ t^{\rho'},\ \beta,\ \beta' \neq 0,\ \rho,\ \rho' < 0$. The adequate accommodation of the bias of Hill's estimator, $H(k) := \sum_{i=1}^{k} U_i/k$, has been extensively addressed in recent years by Beirlant *et al.* (1999), Feuerverger and Hall (1999), Gomes and Martins (2002), among others. All these authors have been led to second order reduced bias tail index estimators, with asymptotic variances larger or equal to $(\gamma(1-\rho)/\rho)^2$, the minimal asymptotic variance in Drees' class of functionals (Drees, 1998). More recently, Gomes *et al.* (2004) and Caeiro *et al.* (2004), deal, in different ways, with a joint external estimation of both the "scale" and the "shape" parameters, β and ρ, respectively, in the A function, being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator. We have here proceeded to the same external estimation of the unknown parameter β in the ML tail index estimator derived in Gomes and Martins (2002), considering the estimator $ML_{\widehat{\beta},\ \widehat{\rho}}(k) := \sum_{i=1}^{k} U_i/k - \widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \sum_{i=1}^{k} (i/k)^{-\widehat{\rho}}\, U_i/k$.

Again, the estimation of the second order parameters in the bias, at a level k_1 of a larger order than that of the level k at which we compute the tail index estimators, enables us to keep the asymptotic variance of the new estimators equal to γ^2. To enhance the interesting performance of this type of estimators, we also consider the estimation of the shape second order parameter only, at the same level k used for the tail index estimation, as well as the estimation of all unknown parameters at the same level k. The asymptotic distributional properties of the proposed class of γ-estimators are derived under a third order framework and the estimators are compared not only asymptotically, but also for finite samples, through Monte Carlo techniques. A case-study in the field of finance will illustrate the performance of these new second order reduced bias' tail index estimators.

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. (2004) *Direct reduction of bias of the classical Hill estimator*, Notas e Comunicações CEAUL 18/04. Submitted.
[3] Drees, H. (1998) *A general class of estimators of the extreme value index*, J. Statist. Plann. and Infer. 98, 95-112.
[4] Feuerverger, A. and Hall, P. (1999) *Estimating a tail exponent by modelling departure from a Pareto distribution*, Ann. Statist. 27, 760-781.
[5] Gomes, M. I., de Haan, L. and Rodrigues, L. (2004) *Tail index estimation through accommodation of bias in the weighted log-excesses*, Notas e Comunicações C.E.A.U.L. 14/2004. Submitted.

Adaptive estimation of the excess d.f.

Grama, Ion *University of South Brittany, France*, ion.grama@univ-ubs.fr

Heavy tails; tail index; excess over the threshold; adaptive estimation; quantiles; value at risk:

The regular variation gives theoretical foundation for approximating a heavy tailed distribution function (d.f.) via estimating its index of regular variation γ, see [1]. However, practical applications concerned with extreme value analysis often suggest approximating the tail of a distribution function directly without passing through the step of estimating γ. In the present talk we develop an adaptive approach for estimating the excess of an arbitrary distribution function using approximations by Pareto models. The proposed approach is fully adaptive to the structure of the underlying d.f. which is in contrast to the case when γ is estimated.

Assume we are given the i.i.d. observations $X_1, ..., X_n,\ n \geq 1$ with the same strictly increasing d.f. $F(x),\ x \geq x_0 \geq 0$. The excess d.f. over a threshold $t > x_0$, is defined by $F_t(x) = 1 - \frac{1-F(xt)}{1-F(t)},\ x \geq 1$. Denote by $K(P, Q)$ the Kullback-Leibler entropy between probability measures P and Q for $P \ll Q$. For any $t \geq x_0$ an approximation of the excess d.f. F_t is defined by "projecting" F_t on the set of Pareto distributions. Let $\theta_t(F) = \arg\min_{\theta>0} K(P_\theta, F_t)$ be the minimum Kullback-Leibler entropy Pareto parameter, called for short *fitted Pareto parameter.* Let $\widehat{\theta}_{n,t},\ t \geq x_0$ be the family of estimators of the fitted parameters $\theta_t(F),\ t \geq x_0$ obtained by substituting empirical d.f. $\widehat{F}_n$ for F.

Let the sequence $\tau_n \geq x_0,\ n \geq 1$ be such that the excess d.f. F_{τ_n} is "nearly" Pareto, which is expressed by the following condition:

$$\int_1^\infty \log^2 \frac{p_{\theta_{\tau_n}(F)}(x)}{f_{F_{\tau_n}}(x)} \exp\left(\left|\log \frac{p_{\theta_{\tau_n}(F)}(x)}{f_{F_{\tau_n}}(x)}\right|\right) F_{\tau_n}(dx) \leq c_1 \frac{\log n}{n\left(1 - F(\tau_n)\right)},$$

for n sufficiently large. Here and below c_i denotes a constant not depending on n, F and τ_n. The left hand side represents a measure of the "squared bias" between F_{τ_n} and $P_{\theta_{\tau_n}(F)}$ while the right hand side is nearly the variance of the stochastic fluctuations corresponding to the expected number of observations in the interval $[\tau_n, \infty)$.

Based on the observations $X_1, ..., X_n$, we give an adaptive selection rule $\widehat{\tau}_n$ for the location τ_n such that, *with overwhelming probability* as $n \to \infty$, the adaptive estimator $\widehat{\theta}_n = \widehat{\theta}_{n,\widehat{\tau}_n}$, satisfies

$$K\left(P_{\widehat{\theta}_n}, F_{\tau_n}\right) \leq c_2 \frac{\log n}{n\left(1 - F(\tau_n)\right)}.$$

The latter bound says merely that the adaptive estimator $\widehat{\theta}_n$, constructed without any knowledge of the location τ_n, is "rate optimal", since it provides the same rate of convergence as the "oracle" estimator $\widehat{\theta}_{n,\tau_n}$, constructed under the additional information that τ_n is known.

The choice of the location of the excess in the procedure mentioned above is based on the consecutive testing for the hypothesis that the tail belongs to Pareto family against the alternative that it is a Pareto change-point model. The selected location of the excess corresponds to the first detected change point. We refer to [2] where a similar approach was proposed.

Note that the rate of convergence depends on the unknown d.f. F, which itself does not affect the optimality property of the proposed adaptive estimator. The obtained results are applied then to give explicit rates of convergence depending on parameters of some particular classes of d.f.'s.

We shall illustrate our theoretical results by simulations and applications to quantile estimation.

References

[1] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987) Regular variation, Cambridge University Press, Cambridge.

[2] Grama I. and Spokoiny V. (2002) *Tail index estimation by local exponential modelling*, Preprint WIAS.

Return level bounds for extreme values

Guillou, Armelle (speaker) *Université de Paris VI, France*, guillou@ccr.jussieu.fr
Diebolt, Jean *Université de Marne-la-Vallée, France*
Naveau, Philippe *Laboratoire des Sciences du Climat et de l'Environnement, France*
Ribereau, Pierre *Université de Paris VI, France*

Empirical processes; Generalized Extreme Value; Probability weighted moment.

For a wide range of applications (hydrology, insurance, etc), the return level is a fundamental quantity to built dykes, define insurance policies, study large wind speeds or other phenomena linked to the behavior of the upper tail of a distribution. More precisely, z_t is called the return level associated with a given return period t if the level z_t is expected to be exceeded on average once every t years. It is also called the t-year event in insurance language. To estimate this level in the iid setting, Extreme Value Theory (EVT) is classically used by assuming that exceedances above an high threshold approximately follow a Generalized Pareto distribution. This approximation is based on an asymptotic argument but the rate of convergence is slow. In this talk, we propose and study a simple estimator of different bounds for the return level. Still, there is a clear link with EVT because the estimator is based on the probability weighted moments that has been particularly used in hydrology for their good properties for small sizes. We derive the properties of our estimators and illustrate our results with a few simulation studies.

Shock models

Gut, Allan *Uppsala University, Sweden*, allan.gut@math.uu.se

Shock, stopped random walk, first passage time, convergence:

Shock models are systems that at random times are subject to shocks of random magnitudes. Traditionally, one distinguishes between two major types; cumulative shock models (systems that break down because of a cumulative effect) and extreme shock models (systems that break down because of one single large shock).

A variation of the theme is to consider mixed models, that is to assume that the system breaks down either because of a cumulative effect, or by a single, large shock, depending on which attains its critical level first.

A more realistic approach in the cumulative case would be to say that "minor shocks" have no effect in the long run. Instead of considering the total accumulation of shocks one might therefore prefer to consider only a final "window" of at each step, so that failure occurs as soon as the total shock load in the window exceeds some given level.

In the extreme case one would e.g. like to include the fact that objects subject to shocks may wear out and introduce some kind of "discount" of earlier shocks.

Most of this research is joint work with Jürg Hüsler, Universität Bern.

References

[1] Gut, A., Hüsler, J. (1999) *Extreme shock models*, Extremes 2, 295–307.
[2] Gut, A. (2001) *Mixed shock models*, Bernoulli 8, 407–421.
[3] Gut, A., Hüsler, J. (2003) *Realistic variation of shock models*, Statist. Probab. Lett. (to appear).

Failure set estimation in multi- and infinite-dimensional space

de Haan, Laurens (speaker) *Erasmus university Rotterdam, The Netherlands*, ldhaan@few.eur.nl
Lin, Tao *Xiamen university, China*
Ferreira, Ana *Technical university of Lisbon, Portugal*

multidimensional extremes; extremes in C[0,1]; rare event estimation

The study of extreme value distributions in various spaces is important and rewarding. So is the study of how to estimate the parameters of the extreme value distributions. But the real goal and purpose of the excercise is to apply these results for estimating the probability of sets that are *extreme* i.e. outside the range of the available observations. In one-dimensional space the solution of this problem has reached its final form. But in finite-dimensional space it is easy to get lost in a forest of not completely transparant conditions. I shall develop an approach that seems quite intuitive and natural. This approach can be extended to the space C[0,1].

Local maximal stack scores with general loop penalty function

Hansen, Niels Richard *University of Copenhagen, Denmark*, richard@math.ku.dk

Extreme value theory; maximal free energy score; local stacks; loop penalty; Poisson approximations; RNA; maximal structure scoring:

Strong limit laws for the maximal free energy score M_n of a random, single stranded RNA molecule of length n was show in [6], and the existence of a phase transition between logarithmic and linear growth of M_n was described. By a random RNA molecule we essentially mean a sequence of iid stochastic variables from a finite alphabet. In [6] it was conjectured that in the logarithmic phase a normalized score, $\lambda M_n - \log Kn$, asymptotically follows a Gumbel distribution. Under the additional, restrictive, assumption that the structure considered is a stack with a single hairpin loop, i.e. no internal loops, bulges or multi-branch loops are allowed, we prove their conjecture. In fact, we show that by counting the number of structures with a high score in a suitably declumped way, we can obtain a Poisson limit by using the results in [1]. This technique has been used successfully in a number of similar cases when considering the comparison of DNA or protein sequences [2,5], and it has also been used in [4] for analyzing maximal stack scores with no loop penalty. However, in the present case the application of [1] differs in several ways from those previous applications. Though our assumptions on the structure are restrictive in comparison with [6], we can treat general hairpin-loop penalties by applying recent results obtained in [3]. In [6] only a linear penalty function is considered, but we show that allowing for general penalty functions provides a refinement of the understanding of the logarithmic phase. For sufficiently fast decaying loop penalty functions the conjecture holds as stated. On the other hand, with no penalty on the loop size, it was shown in [4] that one should use a normalization like $\log K'n(n-1)$ instead of $\log Kn$. Finally we discuss some practical consequences of the results for RNA or DNA database searching with the aim of locating sequences that possess certain structural properties.

References

[1] Arratia, R. A., Goldstein, L., Gordon, L. (1989) *Two moments suffice for Poisson approximation: The Chen-Stein method*, Ann. Probab., 17, 9-25.
[2] Dembo, Amir, Karlin, Samuel, and Zeitouni, Ofer (1994) *Limit distribution of maximal non-aligned two-sequence segmental score*, Ann. Probab. 22(4), 2022–2039.
[3] Hansen, N. R. (2004) *The Maximum of a Random Walk Reflected at a General Barrier* Accepted for publication in Ann. Appl. Prob.
[4] Hansen, N. R. (2004) *Local Stacks in a Markov Chain*, Submitted.
[5] Siegmund, David and Yakir, Benjamin (2000) *Approximate p-values for local sequence alignments*, Ann. Statist. 28(3), 657–680.
[6] Xiong, M., Waterman, M. S. (1997) *A Phase Transition for the Minimum Free Energy of Secondary Structures of a Random RNA*, Adv. in Appl. Math. 18(2), 111–132.

On single and double records

Holst, Lars *Royal Institute of Technology, Sweden*, lholst@math.kth.se

Records; Poisson-approximation :

Consider independent continuous random variables $X_1, X_2, \ldots$ with distribution functions $F_k = F^{a_k}$ where $a_k > 0$ for $k = 1, 2, \ldots$ Let $I_k = I(X_k > X_1, \ldots, X_{k-1})$ indicate a record at trial k. Note that the indicators are independent.

We discuss the distributions of the number of records and double records (that is two successive records) in n trials for a_k's such that for some constant $\theta > 0$

$$P(I_k = 1) = a_k/(a_1 + \cdots + a_k) = \theta/(\theta + k - 1), \quad k = 1, 2, \ldots$$

In the *iid* case $a_k = 1$ we have $\theta = 1$. The distributions are asymptotically Poisson as $n \to \infty$.

Time-varying models for extreme values

Huerta, Gabriel (speaker) *University of New Mexico, USA*, ghuerta@stat.unm.edu
Sansó, Bruno *University of California-Santa Cruz, USA*

Spatio-temporal process; Extreme values; GEV distribution; Process convolutions; MCMC; ozone levels:

We propose a new approach for modeling extreme values that are measured in time and space. First we assume that the observations follow a Generalized Extreme Value (GEV) distribution for which the location, scale or shape parameters define the space-time structure. The temporal component is defined through a Dynamic Linear Model (DLM) or state space representation that allows to estimate the trend or seasonality of the data in time. The spatial element is imposed through the evolution matrix of the DLM where we adopt a process convolution form. We show how to produce temporal and spatial estimates of our model via customized Markov Chain Monte Carlo (MCMC) simulation. We illustrate our methodology with extreme values of ozone levels produced daily in the metropolitan area of Mexico City and with rainfall extremes measured at the Caribbean coast of Venezuela.

References

[1] Huerta, G. and Sansó, B. (2005) *Time-Varying Models for Extreme Values*, Technical report, University of New Mexico. Available from http://www.stat.unm.edu/∼ghuerta/listofpapers.html

Extreme behaviour for stochastic integrals driven by regularly varying Lévy processes

Hult, Henrik (speaker) *Cornell University, Ithaca, USA*, hh228@cornell.edu
Lindskog, Filip *Royal Institute of Technology, Sweden*

We study the extremal behavior of a stochastic integral driven by a multivariate Lévy process that is regularly varying with index $\alpha > 0$. For predictable integrands with a finite $(\alpha + \delta)$-moment, for some $\delta > 0$, we show that the extremal behavior of the stochastic integral is due to one big jump of the driving Lévy process and we determine its limit measure associated with regular variation on the space of càdlàg functions.

References

[1] de Haan, L., Lin, T. (2001) *On convergence toward an extreme value limit in C[0, 1].* Ann. Probab. 29, 467-483.
[2] Hult, H., Lindskog, F. (2005) *Extremal behavior for regularly varying stochastic processes.* Stochastic Process. Appl. 115, 249-274.

On testing the extreme value conditions

Hüsler, Jürg (speaker) *University of Bern , Switzerland*, huesler@imsv.unibe.ch
Li, Deyuan *University of Bern, Switzerland*

Extreme value conditions; Statistical tests; Power :

In applications of the extreme value statistical procedures, one assumes that the distribution of the underlying data belongs to the domain of attraction of an extreme value distribution. Thus the verification of this assumption is necessary for such applications. There exist test procedures of these assumptions which will be discussed in this talk. We will present extensions of the known tests and investigate the statistical properties of these tests. The results are based on an extensive simulation study.

A modified measures of kurtosis for heavy tail distributions

Jafarpour, Hedieh (speaker) *Islamic Azad University, Iran*
Farnoosch, R. *Iran University of Science and Technology, Iran*
Ghodsi, M. *Islamic Azad University, Iran*

Estimators, Large deviation, Scale parameter, Standardized fourth central moment:

The standardized fourth central moment coefficient of kurtosis measures peakedness and tail heaviness of distributions. In this paper three disadvantages of the kurtosis measure has been discussed. It can be misleading as a measure of departure from normality. This measure doesn't work well for some families, for example Ali's scale contaminated normal distributions. It is infinite for heavy tail distribution.

It is so sensitive to extreme values. We introduce a modified measure of kurtosis which it is robust against outlier values and measures peakedness and heaviness of distributions like as the usual kurtosis measure. This introduced measure is finite for all distributions and works well for Ali's scale contaminated normal distribution and like as the usual kurtosis sort the distributions based on Van Zwet's ordering.

References

[1] Ali, M. M. (1974) *Stochastic ordering and kurtosis measure*, JASA, Vol. 69, No. 346, pp. 543-545.
[2] Brys, Hubert, M and Stryyf, A (2004), *Robust measure of tail weight*, Computational statistics and Data Analysis.
[3] Finucan, H. M. (1964), *A note on kurtosis*, JRSS, Ser. B, 26, No. 1, pp.111-112.
[4] Groeneveld, R. A. (1998), *A class of Quantile Measure for kurtosis*, The American Statistician, Vol. 51, No. 4, pp. 325-329.
[5] Hannu, O.(1981), *On Location, Scale, Skewness and kurtosis of univariate Distributions*, Scand J Statist, Vol.8, pp.154-168.
[6] Kevin P. Balanda and H. L. Macgillivray (1988), *Kurtosis: a critical review.* JASA, Vol. 42, No. 2, pp. 111-119.
[7] Richard A. Groeneveld and Glen Meeden (1984), *Measuring skewness and kurtosis*, The Statistician, Vol. 33, pp. 391-399.
[8] Stuart, A. and Ord, J. K. (1994), Kendalls Advanced Theory of Statistics, Volume 1: Distribution Theory, 6th ed., Arnold, London.

Extrapolation of fatigue loads

Johannesson, Pär *Fraunhofer-Chalmers Centre, Sweden*, par.johannesson@fcc.chalmers.se

In fatigue life assessments both the material properties and the load characteristics are essential parameters. The life of a component can be experimentally found by performing fatigue tests. The load often is a sequence of constant amplitude cycles, however, it is well known that it often results in a systematic error when predicting the life in service. In order to get reliable predictions, the tests should be performed using load signals that are representative for the service loads. We focus on the problem of extrapolating a measured load signal to a longer time period, for example to a full design life. Two methods, both based on statistical extreme value theory, for extrapolating a load measurements are presented. The first is a simulation based method, while the second method approximates the tails of the so-called rainflow cycle distribution.

When performing tests with service loads, it is customary to use a measured load signal, and repeat this load block until failure. The first method is based on repeating load blocks, however the largest maxima and the lowest minima of each block are randomly regenerated based on statistical extreme value theory (Johannesson [1]). More precisely, we use the Peak Over Threshold (POT) technique for modelling the exceedances of maxima above a high threshold, and the minima below a low threshold. An example of three load blocks is shown in Figure 1.

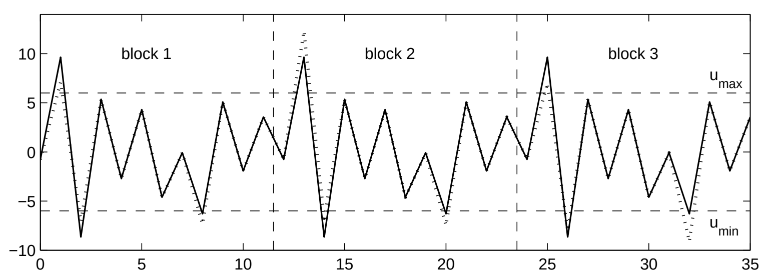

Figure 1: Three repetitions of a measured block (solid lines), compared to three extrapolated blocks (dotted lines). The vertical dashed lines represent the threshold levels for the POT extrapolation starts.

The damaging events in the load signals are found by counting cycles. The rainflow cycle count method is most often used, and it is equivalent to counting crossings of intervals of the load signal. The second method is to extrapolate in the rainflow cycle domain, using the method in Johannesson & Thomas [2], resulting in an asymptotic expression for the tails of the 2-dimensional rainflow cycle distribution.

The extrapolation methods have been verified on many types of load measurements from different applications, e.g. a service load on a train, and test track measurements on cars. The two methods give about the same results in terms of the so-called load spectrum, which is the cumulative number of cycles with amplitudes above a given value. However, the time domain method results in an extrapolated time signal where the order of the cycles is preserved, while the rainflow domain method gives an extrapolated rainflow cycle distribution.

References

[1] Johannesson, P. (2004) *Extrapolation of Load Histories and Spectra*, Proceedings of 15th European Conference on Fracture.
[2] Johannesson, P. and Thomas, J.-J. (2001) *Extrapolation of Rainflow Matrices*, Extremes 4, 241-262.

Extreme R-estimator in regression model

Jurečková, Jana *Charles University in Prague, Czech Republic*, jurecko@karlin.mff.cuni.cz

Key words: Extreme regression quantile; R-estimator; asymptotic distribution

The extreme (maximal) regression quantile in the linear model $Y_i = \beta_0 + \mathbf{x}_i'\beta + e_i,\ i = 1, \ldots, n$, is defined as a solution of the linear program $\sum_{i=1}^{n}(b_0 + \mathbf{x}_i'\mathbf{b}) =:$ min under the restrictions $\beta_0 + \mathbf{x}_i'\beta \geq Y_i,\ i = 1, \ldots, n,\ \beta_0 \in \mathbf{R}, \beta \in \mathbf{R}^p$. Jurečková and Picek (2005) showed that the extreme regression quantile can be equivalently written in a two step version, starting with a suitable R-estimator of the slope parameters and then ordering the residuals. We shall show that, for a class distributions of the e_i from the domain of attraction of the Gumbel distribution, the initial R-estimator is even a consistent estimator of the slopes components, and derive its asymptotic distribution. The resulting estimator of the intercept component converges to the Gumbel distribution, after a proper standardization.

References

[1] Portnoy, S., Jurečková, J. (2000) *On extreme regression quantiles*, Extremes 29, 467–483.
[2] Jurečková, J., Picek, J. (2005) *Two-step regression quantiles*, Sankhya (to appear).

On extreme quantiles in nilpotent markov chain with application to PRNG testing

Kawczak, Janusz (speaker) *University of North Carolina at Charlotte, U.S.A.*, jkawczak@math.uncc.edu
Molchanov, Stanislav *University of North Carolina at Charlotte, U.S.A.*

Markov Chain; Berry-Esseen; PRNG; Extreme Quantiles:

In [1] we study the spectrum of the covariance operator of the nilpotent Markov Chain. This is a special case of the general Markov Chain with the Doeblin condition and under the assumption that there exist a finite k such that for all $l \geq k$ $(P - \Pi)^l = 0$, where Π is an invariant distribution and P is the transition operator associated with the chain. The chains with this property arise naturally in testing Pseudo Random Number Generators (PRNG) when it is understood that a finite sequence of digits is produced to form a word. We specialize our approach to the study of weak convergence with the improved estimation of the remainder term for the Marsaglia [2] permutation type test statistics. This is used to estimate the extreme quantiles of the limiting distribution of the functionals of Markov Chain with nilpotent property.

A complete analysis of the spectrum of the covariance operator is presented for the $L^2(X, \mu)$ space. We give an explicit decomposition of $L^2(X, \mu)$ into the direct sum of the eigenspaces associated to the eigenvalues of the covariance operator. This decomposition allows for the development of efficient computational algorithms when establishing the limiting distribution of the functional Central Limit Theorem generated by a general Markov Chain.

We also present some results of Berry-Esseen type for general Markov chains with and without nilpotent property [3], [4].

References

[1] Kawczak, J., Molchanov, S. and Al Hakim, A. (2004) *On the Class of Nilpotent Markov Chains, I. The Spectrum of Covariance Operator*, Markov Processes and Related Fields. Vol. 10, Issue 4, 629–652.
[2] Marsaglia, G. (1985) *A Current View of Random Number Generators.*,Computer Science and Statistics. Elsevier Science Publisher B.V. North-Holland.
[3] Prokhorov Yu.V. and Statulevičius, V. (2000) Limit Theorem of Probability Theory. Springer-Verlag.
[4] Woodroofe, M. A. (1992) *A central limit theorem for functions of a Markov chain with applications to shifts.* Stochastic Processes and Their Applications. Vol. 41, 33–44.

SKF experience with Extreme Value Analysis on rolling bearing steels and the relation to fatigue properties

Kerrigan, Aidan aidan.kerrigan@skf.com

Changes in temperature and precipitation extremes as simulated in the IPCC multi-model ensemble of global coupled model simulations

Kharin, Viatchelsav (speaker) *Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada*, slava.kharin@ec.gc.ca
Zwiers, Francis *Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada*

Climate extremes; IPCC; global climate models:

Changes in extreme temperatures and precipitation are examined in the IPCC multi-model ensemble of global coupled climate model simulations from more than a dozen of climate modelling groups. The projected future changes are estimated for a number of climate change scenarios including the IPCC SRES A1B, B2 and A2 conditions. The extremes are described in terms of return values estimated from the Generalized Extreme Value distribution. fitted to samples of annual extremes.

This study is still ongoing as more and more model output becomes available for the analysis. The first preliminary results indicate that there is substantial inter-model variability in simulated climate extremes, in particular, precipitation extremes. Changes in temperature extremes on global scale are largely associated with changes in the location of the distribution of annual extremes and are comparable to changes in mean temperature. Changes in precipitation extremes generally exceed the corresponding changes in the annual mean precipitation. The results are broadly consistent with finding in previous studies (Kharin and Zwiers, 2005; Kharin et al., 2005).

References

[1] Kharin, V. V., and Zwiers, F. W. (2005) *Estimating extremes in transient climate change simulations*, J. Climate, 18, 1156–1173.
[2] Kharin, V. V., Zwiers, F. W., and Zhang, X. (2005) *Intercomparison of near surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations*, J. Climate, accepted.

On large deviations for branching processes in random environment

Kozlov, Mikhail *Moscow State University , Russia*, kozlov_mikhail @ mail.ru

large deviations; branching process; random environment:

Let us introduce the sequence $\eta_. := (\eta_0 \equiv \eta, \eta_1, \eta_2, \dots)$ of i.i.d. random elements on the space $(\mathcal{Y}, \mathcal{B})$ considered as the *random environment* and the matrix of r.v. ζ_{in} row independent i.i.d. conditionally on the environment $\eta_.$ with the following conditional generating function:

$$\mathbf{E}(s^{\zeta_{in}}|\eta_.) := f_{\eta_n}(s) \text{ a.s.}, \ f_y(s) := 1 - \left(\frac{1}{a(y)(1-s)} + \frac{b(y)}{2a(y)^2}\right)^{-1}, \quad |s| \le 1\,.$$

The recurrent relation $Z_{n+1} = \sum_{i=1}^{Z_n} \zeta_{in}$ determines a process which is known as the *branching process in random environment* (BPRE).

Introduce the notations:

$$X_i := \ln f'_{\eta_{i-1}}(1), \quad S_n := \sum_{i=1}^{n} X_i, \quad R(h) := \mathbf{E}e^{hX} = \int e^{hx}\, dF(x), \quad \mu := \mathbf{E}X,$$

$$h^+ := \sup\{h : R(h) < \infty\}, \quad m(h) := R'(h)/R(h) \quad ,$$

$$h(\theta) \text{ is a root of the equation } m(h) = \theta \text{ for } \mu < \theta < \theta^+\,,\ \theta^+ := \lim_{h \to h^+} m(h)\,,$$

$$\Lambda(\theta) := h(\theta)\theta - \ln R(h(\theta))\,.$$

Theorem 1. *Let us suppose that* $b(y)/(2a(y)^2) \equiv 1,\ y \in \mathcal{Y}$, *and that the r.v.* $X = \ln f'_\eta(1)$ *fulfills right hand Cramer condition. Let* $0 \le \mu \equiv \mathbf{E}X < \infty$. *Then uniformly on* θ *from any compact in* (μ, θ^+) *the following asymptotic holds*:

$$\mathbf{P}(\ln Z_n > \theta n) \sim I(\theta)\mathbf{P}(S_n > \theta n) \sim I(\theta)\Big(\sqrt{2\pi n}c(\theta)\Big)^{-1} \exp\{-\Lambda(\theta)n\},\ n \to \infty. \tag{5}$$

Theorem 2. *Let us assume the general conditions of theorem 1. Suppose that* $\mu < 0$ *and* $h(0) < h^+$. *Then the asymptotic (5) takes place uniformly on* θ *from any compact: (I) in* $(0, \theta^+)$ *if* $h(0) \le 1$ *and (II) from* (θ^*, θ^+) *if* $h(0) > 1$ *and there is the unique value* $1 < h^* < h^+$ *such that* $R(h^*) = R(1)$ *and* θ^* *is determined by the equation* $m(\theta^*) = h^*$.

Theorem 3. *Let us assume the general conditions of theorem 1. Suppose that* $\mu < 0$, $R'(1) < 0$, $\mu < \theta < \theta^*$. *Then*

$$\mathbf{P}(\ln Z_n > \theta n) \sim I^*(\theta) \exp\{-\Lambda^*(\theta)n\},\ n \to \infty\,,$$

$$I^*(\theta) > 0,\ \Lambda^*(\theta) = \theta h^* - \ln R(1) < \Lambda(\theta),\ \theta \in (0, \theta^*)\,.$$

Dimension reduction with heavy tails

Kuhn, Gabriel *Munich University of Technology, Germany*, gabriel@ma.tum.de

dimension reduction, heavy tails:

We extend the standard approach of factor analysis. The classical normal assumption is dropped and replaced by the more general class of elliptical distributions and copulas. A special focus lies on the possibility for factorization of heavy-tailed models where we also show how to decompose the dependence structure in the extremes. In a simulation study we observe that the factorization is quite stable independent of the underlying model – the only difference lies in the different interpretations, depending on the estimation method of the correlation structure. Analyzing a sample of data, we observed quite different factors, depending if we focus on the extremes or not.

References

[1] Kuhn, G. (2004) *Tails of Credit Default Portfolios.*, Submitted for publication.
[2] Lucas, A., Klaasen, P., Spreij, P. and Straetmans, S. (2003). *Tail behaviour of credit loss distribution for general latent factor models. Applied Mathematical Finance 10(4), 337-357.*

Smoothing sample extremes: the mixed model approach

Pauli, Francesco *University of Padova, Italy*, fpauli@stat.unipd.it
Laurini, Fabrizio (speaker) *University of Parma, Italy*, flaurini@stat.unipd.it

Spline; Smoothing; Mixed effects model:

Smoothing sample extremes has previously been dealt with by local likelihood models [1] and by spline smoothing. Furthermore, estimates of spline smoothers can be obtained by penalized likelihood either by classical [2] or Bayesian [3] inference.

An alternative approach to spline estimation involves representing the spline function as a linear combination of basis functions and estimating coefficients of these functions as random effects in a mixed model. This idea dates back to [4] but was recently thoroughly discussed in [5] and [6]. To our knowledge, however, this approach to spline estimation is confined to samples from distributions within the exponential family.

We implemented the mixed effect approach for smoothing sample extremes both using the Generalized Extreme Value distribution and the Poisson Point Process models. We tackled the problem by adopting a Bayesian approach to inference and used MCMC algorithms for exploration of the posterior distribution. We took the Bayesian approach since it is easier to implement in this context and, moreover, it does not share the interpretation issues which arise in the classic framework concerning the role of random coefficients of the spline in the model, and whether their randomness should be taken into account in doing inference on the model (see Nelder, Morton and Green in the discussion of [7]).

This approach allows to straightforward inclusion of splines in arbitrarily complex hierarchical models: we can naturally deal with multiple time series as well as correlated data. Moreover, the role of the smoothing coefficient is taken by a function of model parameters and so is easily estimated from data.

We applied this approach to model extremes of pollutant concentrations in a major Italian city (Milan) in order to establish whether a trend exists in recent years. Data include daily maxima of pollutant concentrations measured at different locations and, as such, fall quite naturally within a hierarchical model structure. By adopting a mixed model approach to analyze the data we model the temporal trend by introducing random effects for calendar years, while a smoothing spline is included to allow for the seasonal pattern, which is found to be definitely non linear.

References

[1] Davison, A. C., Ramesh, N. I. (2000) *Local likelihood smoothing of sample extremes*, J. R. Statist. Soc. B, 62, 191

[2] Chavez-Demoulin, V., Davison, A.C. (2005) *Generalized additive modelling of sample extremes*, Appl. Statist. 54, 1, 207–222.

[3] Pauli, F., Coles, S. G. (2001) *Penalized likelihood inference in extreme value analyses*, J. Appl. Statist., 28, 547

[4] Robinson, G.K. (1991) *That BLUP is a good thing: The estimation of random effects* (with discussion), Statistical Science, 6, 15–51.

[5] Lin, X., Zhang, D. (1999) *Inference in generalized additive mixed models by using smoothing splines*, J. R. Statist. Soc. B, 61, 381–400.

[6] Ruppert, D., Wand, M.P., Carroll, R.J. (2003) *Semiparametric regression*, Cambridge University Press.

[7] Verbyla, A.P., Cullis, B.R., Kenward, M.G., Welham, S.J. (1999) *The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion)*, J.R. Statist. Soc. C - Applied Statistics, 48, 269–300.

Comments on extremes in naval architecture and the environment - some pitfalls, promises, and progress

Guillou, Armelle *Université de Paris VI, France*
Leadbetter, Ross (speaker) *University of North Carolina, Chapel Hill, USA*, mrl@email.unc.edu
Rychlik, Igor *Lund University, Sweden*

Extreme values, Palm probabilities, structural safety, stability of ships:

The use of Extremal Theory understandably provides a natural immediate attack on description of phenomena involving high level behavior. However while it may be trite, it is nevertheless true to note that the high values are not necessarily "extreme" for application of the theory simply because their occurrence may have "extreme" consequences, and that moreover the theory requires appropriate coordination of high levels and long observation periods.

In this talk I plan to briefly describe three projects in which the allure of formal use of extreme value theory was shattered and other methods (albeit related) required. Two of these involve Naval Architecture, the first concerning structural integrity of ocean structures and vessels under "extreme" (hurricane) conditions. This study for the US Navy was "underway" at the time of EVA 1998 at which it was described in part, and was brought to a successful conclusion using models based on the machinery of Palm distributions at level crossings.

The second (current) project concerns vessel capsize risk, and the need for techniques beyond the simple fitting of Type I and II extreme value distributions for ship roll, extensively relied on for naval use. Specifically US Coastguard proposals for relating capsize risk to "encountered wave" geometry are being further developed using wave statistics routines of I Rychlik and colleagues as part of the Lund University extensive "WAFO" program system.

Finally a study of environmental (ozone) data indicates the need for modeling by "intermediate" rather than "extreme" levels for USEPA environmental regulation and assessment programs.

Tail approximations to the density function in EVT

Li, Deyuan (speaker) *University of Bern, Switzerland*, deyuan.li@stat.unibe.ch
Hüsler, Jürg *University of Bern, Switzerland*

tail approximation; density function; maximum; extreme value distribution; differentiable domain of sttraction:

Let $X_1, X_2, ..., X_n$ be independent identically distributed random variables with common distribution function F, which is in the max domain of attraction of an extreme value distribution, i.e. there exist sequences $a_n > 0$ and $b_n \in R$ such that the limit of $P(a_n^{-1}(\max_{1 \leq i \leq n} X_i - b_n) \leq x)$ exists. Assume the density function f (of F) exists. We obtain an uniformly weighted approximation to the tail density function f, and an uniformly weighted approximation to the tail density function of $P(a_n^{-1}(\max_{1 \leq i \leq n} X_i - b_n) \leq x)$ under some second order condition.

References

[1] Cheng, S. and Jiang, C. (2001). The Edgeworth expansion for distributions of extreme values. *Sci. China* **44**, 427-437.
[2] Drees, H. (1998). On smooth statistical tails functionals. *Scand. J. Statist.* **25**, 187-210.
[3] Drees, H., de Haan, L. and Li, D. (2003). On large deviation for extremes. *Statis. Probab. Letters* **64**, 51-62.
[4] Drees, H., de Haan, L. and Li, D. (2005). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statis. Plan. Infer.* to appear.
[5] de Haan, L. and Resnick, S. I. (1982). Local limit theorems for sample extremes. *Ann. Probab.* **10**, 396-413.
[6] de Haan, L. and Resnick, S. I. (1996). Second order regular variation and rates of convergence in extreme value theory. *Ann. Probab.* **24**, 97-124.
[7] Gabaix, X. and Laibson, D. (2003). Some Game Theory with Extreme Value Theory. *Working paper.*
[8] Pickands, J. (1986). The continuous and differentiable domains of attraction of the extreme value distributions. *Ann. Probab.* **14**, 996-1004.
[9] Smith, R. L. (1982). Uniform rates of convergence in extreme value theory. *Adv. Appl. Peob.* **14**, 600-622.

Extremal behaviour of moving average processes with light-tailed innovations

Lindner, Alexander (speaker) *Munich Technical University, Germany*, lindner@ma.tum.de
Klüppelberg, Claudia *Munich Technical Univerisity, Germany*, cklu@ma.tum.de

domain of attraction, light-tailed innovations, moving average process:

We consider infinite moving average processes of the form

$$X_n = \sum_{i=-\infty}^{\infty} c_i Z_{n-i}, \quad n \in \mathbf{Z},$$

where $(Z_i)_{i\in\mathbf{Z}}$ is a sequence of iid random variables with "light tails" and $(c_i)_{i\in\mathbf{Z}}$ is a sequence of positive, suitably decreasing constants. By light tails we mean that Z_0 has a bounded density f satisfying the asymptotic

$$\lim_{t\to\infty} \frac{f(t)}{\gamma(t)\exp(-\psi(t))} = 1,$$

where $\gamma(t)$ behaves roughly like a constant as $t \to \infty$ and ψ is a strictly convex C^2 function such that $\psi'(t) \to \infty$ as $t \to \infty$ and such that $1/\sqrt{\psi''(t)}$ is self-neglecting. (The latter is an asymptotic variation condition). It is shown that the iid sequence associated with X_0 is in the maximum domain of attraction of the Gumbel distribution. Under further regular variation conditions on ψ, it is shown that the stationary sequence $(X_n)_{n\in\mathbf{N}}$ has the same extremal behaviour as its associated iid sequence. This generalizes results of Rootzén [2,3] on moving average processes where Z_0 has a density behaving asymptotically like $\exp(-x^p)$, $p > 1$. The talk is based on [1].

References

[1] Klüppelberg, C., Lindner, A. (2005) *Extreme value theory for moving average processes with light-tailed innovations*, Bernoulli, to appear.
[2] Rootzén, H. (1986) *Extreme value theory for moving average processes*, Ann. Probab., 612-652.
[3] Rootzén, H. (1987) *A ratio limit theorem for the tails of weighted sums*, Ann. Probab., 728-747.

On Kesten's counterexample to the Cramér-Wold device for regular variation

Lindskog, Filip (speaker) *Royal Institute of Technology, Sweden*, lindskog@math.kth.se
Hult, Henrik *Cornell University, USA*

heavy-tailed distributions; linear combinations; multivariate regular variation:

In 2002 Basrak, Davis, and Mikosch showed that an analog of the Cramér-Wold device holds for regular variation of random vectors if the index of regular variation is not an integer. This characterization is of importance when studying stationary solutions to stochastic recurrence equations. In this paper we construct counterexamples showing that for integer-valued indices, regular variation of all linear combinations does not imply that the vector is regularly varying. The construction is based on unpublished notes by Harry Kesten.

References

[1] Basrak, B., Davis, R., Mikosch, T. (2002) *A characterization of multivariate regular variation*, Ann. Appl. Probab., 12, 908-920.
[2] Hult, H., Lindskog, F. (2005) *On Kesten's counterexample to the Cramér-Wold device for regular variation*, to appear in Bernoulli. Available at www.math.kth.se/~lindskog

Some results on extremal and maximal processes associated with a Lévy process

Maller, Ross (speaker) *Australian National University, Canberra, Australia Ross.Maller@anu.edu.au*
Erickson, K. Bruce *University of Washington, USA*

extremal process, Lévy process, two-sided maximum process, two-sided passage time, overshoot

We compare $Y_t = \sup\{|X_s - X_{s-}| : s \le t\}$, the largest jump in a Lévy process X_t up till time t, to the two-sided maximal value of the process, $M_t = \sup\{|X_s| : s \le t\}$. With $T(r) = \inf\{t > 0 : |X_t| > r\}$, $r > 0$, as the two-sided passage time out of the strip $[-r, r]$, we show that Y_t is negligible with respect to M_t for small times, i.e., $\lim_{t\downarrow 0} Y_t/M_t = 0$ a.s., iff the overshoot $X_{T(r)} - r$ is relatively stable in the sense that $\lim_{r\downarrow 0} |X_{T(r)}|/r = 1$ a.s. These are further equivalent to the a.s. convergence of the stochastic integral $\int_0 1\overline{\Pi}(M_t)dt$, where $\overline{\Pi}(\cdot)$ is the Lévy measure associated with X, and to the bounded variation (with nonzero drift) of X. Negligibility of Y_t with respect to M_t as $t \to \infty$ can similarly be characterised.

Smoothing of variable bandwidth kernel estimate of heavy-tailed density function

Markovich, Natalia *Institute of Control Sciences Russian Academy of Sciences, Russia*, markovic@ipu.rssi.ru

Heavy-tailed density; smoothing; discrepancy method;kernel estimate:

A practical version of the variable bandwidth kernel estimate

$$\widetilde{f}^A(x|h_1,h) = \frac{1}{nh}\sum_{i=1}^{n} \hat{f}_{h_1}(X_i)^{1/2} K\left((x-X_i)\hat{f}_{h_1}(X_i)^{1/2}/h\right), \tag{6}$$

constructed by the empirical sample $X_1,...,X_n$ is often used for the estimation of a heavy-tailed probability density function $f(x)$. Here, $\hat{f}_{h_1}(X_i)$ is a pilot estimate of $f(x)$, $K(\cdot)$ is a kernel function. This estimate may provide the mean squared error (MSE) of order 8/9 without the disadvantage of negativity if its bandwidth h is of order $n^{1/9}$ and fourth derivative of $f(x)$ is continuous. The same MSE can be reached for a non-variable kernel estimate $\hat{f}_h(x) = (nh)^{-1}\sum_{i=1}^{n} K((x-X_i)/h)$ with the fourth order kernels, that are non-positive. However, the value of $h \sim n^{-1/9}$ that is recommended by the theory, depends on the fourth derivative of $1/f(x)$ that is unknown.
Here, we introduce the discrepancy method that produces data-driven, asymptotically optimal estimators of h. This method is alternative to different versions of the cross-validation method (e.g., see [1], [2]). Namely, let h in (6) be a solution of the equation

$$\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| = n^{-4/9}\psi_1(n)\left(\psi_2(n)/2+1\right), \tag{7}$$

where $F_h^A(x) = \int_0^x \widetilde{f}^A(t|h_1,h)dt$ is the estimate of the distribution function (DF) $F(x)$, $F_n(t)$ is the empirical DF, $\Omega^* \subseteq [0,\infty)$ is some finite interval, the functions $\psi_1(n)$, $\psi_2(n)$ obey the conditions: $\psi_1(n)\to\infty$, $\psi_2(n)\psi_1(n)^{9/8}\to 0$, $(\psi_1(n)\psi_2(n))^{-2} = o\left(n^{1/9}\right)$ as $n\to\infty$. This implies, that (7) generates the spectrum of methods for the estimation of h in (6). The application of (7) requires the preliminary transform of the data to a finite interval. Hence, the method may be applied to heavy-tailed densities, too. The equation (7) can be represented as

$$\sup_{x\in\Omega^*} |F_n(x) - F_h^A(x)| = \frac{(\ln\ln n)^{\gamma}}{n^{4/9}}\left(1+\frac{1}{2\left(\ln\ln n\right)^{(5/4)\gamma}}\right) \approx 1.5\left(\frac{\ln\ln n}{n}\right)^{1/2}\frac{n^{1/18}}{(\ln\ln n)^{1/2-\gamma}}, \tag{8}$$

when $\psi_1(n) = (\ln\ln n)^{\gamma}$, $\gamma > 0$, $\psi_2(n) = \psi_1(n)^{-5/4}$. For sufficiently small $\gamma < 1/2$ the right-hand side of (8) is close to $c\left(\ln\ln n/n\right)^{1/2}$, $c > 1$. The discrepancy method was investigated in [3] in a slightly different formulation.
It is proved that the method (7) provides

$$\mathbb{P}\{\overline{\lim}_{n\to\infty} n^{8/9} var(\widetilde{f}^A(x)) \le c_1^*\} = 1,$$

where $var(\widetilde{f}^A(x))$ is the variance of the estimate $\widetilde{f}^A(x)$, and

$$\mathbb{P}\{\overline{\lim}_{n\to\infty} \psi_1(n)^{-2} n^{8/9} MSE(\widetilde{f}^A(x)) \le c_2^*\} = 1,$$

where c_1^*, c_2^* are the constants those are independent of n, if the fourth derivative $f(x)^{(4)}$ and $\varphi(x) = (d/dx)^4\left(1/f(x)\right)$ are continuous and the kernel function obeys some standard conditions. It implies, that the order of $MSE(\widetilde{f}^A(x))$ is close to the value $n^{-8/9}$, when $\psi_1(n)$ tends to infinity sufficiently slowly. For example, if $\psi_1(n) \sim (\ln\ln n)^{1/8}$ we have $MSE(\widetilde{f}^A(x)) \sim (\ln\ln n)^{1/4} n^{-8/9}$.

References

[1] Bowman, A.W. (1984) *An alternative method of cross-validation for the smoothing of density estimates*, Biometrika. 71, 2, 353–360.
[2] Hall, P. (1992) *On global properties of variable bandwidth density estimators*, Annals of Statistics. 20, 2, 762–778.
[3] Markovich, N.M. (2005) *The accuracy of retranformed kernel estimates of heavy-tailed densities*, Automation and Remote Control. 2, 55–72.

Numerical bounds for the distribution of the maximum of a one- and two-parameter process

Mercadier, Cecile *Toulouse* mercadie@cict.fr

Random fields; Distribution of the maximum; Abs. continuous sample paths; First time passage; WAFO Toolbox.:

We consider the class of real valued stochastic processes indexed on a compact subset of $\mathbb{R}$ or $\mathbb{R}^2$ with almost surely absolutely continuous sample paths. We obtain an implicit formula for the distribution of their maximum. The main result is the derivation of numerical bounds that turn to be very accurate for levels which are not large. We also present the first explicit upper bound for the survival function of the maximum in the two-dimensional framework.
Numerical comparisons are performed with known tools as Rice upper bound and expansions based on Euler characteristic. We deal numerically with the determination of the persistence exponent.

Estimation of the angular density in multivariate Generalized Pareto models

Michel, Rene *University of Würzburg, Germany*, michel@mathematik.uni-wuerzburg.de

Multivariate Generalized Pareto Distribution; Angular Density; Pickands Coordinates; Kernel Density Estimator:

Let $X = (X_1, \ldots, X_d) \in (-\infty, 0)^d$ be a random vector which is distributed according to a generalized Pareto distribution (GPD)

$$W(x) = 1 + \left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, \ldots, \frac{x_1}{\sum_{i=1}^{d} x_i}\right)$$

in a neighborhood of 0, following the definition of Falk et al. (2004). A transformation T based on Pickands coordinates is presented such that the random vector $T(X)$ asymptotically follows a distribution generated by the angular density belonging to the GPD W. The angular density is a useful tool in measuring the degree of dependence of $X_1, \ldots, X_d$. This has been done by Coles and Tawn (1991 and 1994), Coles et al. (1999) and Joe et al. (1992) for the case of an extreme value distribution. In the above references an estimation of the angular density with real data sets gave valuable information in different areas such as oceanography or meteorology.

With the help of the transformation T, estimation of the angular density in GPD models can be done in a nonparametric way via kernel density estimators. Under certain regularity conditions the asymptotic normality of these estimators can be shown. Using the simulation algorithm of Michel (2004) the results of these estimations are illustrated by a simulation study.

References

[1] Coles, S., Tawn, J. (1991) *Modelling Extreme Multivariate Events*, Journal of the Royal Statistical Society 53, 377–392.
[2] Coles, S., Tawn, J. (1994) *Statistical Methods for Multivariate Extremes*, Applied Statistics 43, 1–48.
[3] Coles, S., Heffernan, J., Tawn, J. (1999) *Dependence Measures for Extreme Value Analysis*, Extremes 1, 339–365.
[4] Falk, M., Hüsler, R., Reiss, R.-D. (2004) Laws of Small Numbers: Extremes and Rare Events - 2nd Edition, Birkhäuser.
[5] Joe, H., Smith, R., Weissman, I. (1992) *Bivariate Threshold Methods*, Journal of the Royal Statistical Society 54, 171–183.
[6] Michel, R., (2004) *Simulation of Generalized Pareto Distributions of Logistic Type and Applications*, Technical Report

On maxima of complete and incomplete samples from stationary sequences

Mladenović, Pavle (speaker) *University of Belgrade, Serbia and Montenegro*, paja@matf.bg.ac.yu
Piterbarg, Vladimir *Lomonosov Moscow State University, Russia*

Stationary sequences, Extreme values, Missing observations, Exponent of regular variation:

Let (X_n) be a strictly stationary random sequence with the marginal distribution function $F(x) = P\{X_1 \leq x\}$. Suppose that some of random variables $X_1, X_2, X_3, \dots$ can be observed, and let ε_k be the indicator of the event that random variable X_k is observed and $S_n = \varepsilon_1 + \cdots + \varepsilon_n$. We suppose that ε_k is independent of the sequence (X_n). Let us denote

$$\widetilde{M}_n = \max\{X_j : 1 \leq j \leq n, \varepsilon_j = 1\}, \quad M_n = \max\{X_1, \dots, X_n\}.$$

Suppose that d.f. F belongs to the maximum domain of attraction of some of extreme value distributions. The limiting distribution of the random vector $(\widetilde{M}_n, M_n)$ and "asymptotic independency" of $\widetilde{M}_n$ and M_n are obtained under a condition imposed on asymptotic behavior of the sum S_n and a condition of weak dependency of random variables from the sequence (X_n), which is more restrictive than Leadbetter's $D(u_n)$ condition.

Some results concerning estimation of the exponent of regular variation using a sample with missing observations will also be presented.

References

[1] Hsing, T. (1991) *On tail index estimation using dependent data*, Ann. Statist. 19, 1547-1569.
[2] Hüsler, J. and Piterbarg, V. (2004) *Limit theorem for maximum of the storage process with fractional Brownian motion as input*, Stoch. Proc. Appl. 114, 231-250.
[3] Mladenović, P. and Piterbarg, V. (2004) *On estimation of the exponent of regular variation using a sample with missing observations*, Submitted.
[4] Mladenović, P. and Piterbarg, V. (2004) *On asymptotic distribution of maxima of complete and incomplete samples from stationary sequences*, Preprint.

Modeling spatial dependence for extremes in climate studies

Naveau, Philippe (speaker) *Laboratoire des Sciences du Climat et de l'Environnement, France*,
philippe.naveau@lsce.cnrs-gif.fr
Cooley, Dan *University of Colorado, USA*
Poncet, Paul *Université Paris X, France*

Max-stable fields; Variograms; Extremal coefficient; Madograms:

For a wide range of scientific applications in climate research, the observations are scattered in space, either on a regular grid or at irregularly spaced locations. For example, climatological or pollutant data are recorded at different locations and measurements classically exhibit some degree of spatial dependence. While the mean behavior of most spatial processes such as daily temperatures or wind fields is well modeled and understood by the statistical and scientific communities, our understanding of how to measure the spatial dependence for extreme events is still incomplete from a statistical perspective.

Within the statistical community, there has been a growing interest in the analysis of spatial extremes in recent years. For example, de Haan and Pereira [2] proposed two specific stationary models for extreme values. These models depend on one parameter that varies as a function of the distance between two points. Davis and Mikosch [1] proposed space-time processes for heavy tail distributions by linearly filtering i.i.d. sequences of random fields at each location. Schlather and Tawn [4] simulated stationary max-stable random fields and studied the extremal coefficients for such fields. Bayesian or latent processes for modeling extremes in space has been also investigated by several authors. In geostatistics, a classical approach called "Gaussian anamorphosis" consists of transforming a spatial field into a Gaussian one (e.g. [5]), but it does not take advantage of the theoretical foundation provided by extreme value theory.

In comparison with all these past developments, our research can be seen as a further step in the direction taken by Schlather and Tawn [4]. We work with stationary max-stable fields [3] and focus on capturing the spatial structure with extremal coefficients. The novelty is that we propose different estimators of extremal coefficients that are more clearly linked to the field of geostatistics. Our estimators are based on the variogram which has been the cornerstone of the field of geostatistics. Hence, one of the main advantages of our research is conceptual. It provides a natural bridge between extreme value theory and geostatistics, two research fields that have been rarely interconnected in the past. We illustrate our results by studying a variety of climate fields such as precipitation and temperatures.

References

[1] Davis, R. and Mikosch, T. (2004). Extreme value theory for space-time processes with heavy-tailed distributions. *Lecture notes from the MaPhySto Workshop on "Nonlinear Time Series Modeling", Copenhagen, Denmark.*

[2] de Haan, L. and Pereira, T. (2005). Spatial extremes: the stationary case. *submitted.*

[3] de Haan, L. and Resnick, S. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie*, 4:317–337.

[4] Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90:139–156.

[5] Wackernagel, H. (2003). *Multivariate Geostatistics. An Introduction with Applications.* Springer, Heidelberg, third edition.

Statistical inference for heavy and super-heavy-tailed distributions

Neves, Cláudia (speaker) *University of Aveiro, Portugal*, claudia@mat.ua.pt
Fraga Alves, M. Isabel *University of Lisbon, Portugal*
de Haan, Laurens *Erasmus University of Rotterdam, The Netherlands*

estimation; max-domain of attraction; regular variation theory; test of hypothesis:

Heavy-tailed distributions constitute a fundamental tool in the study of rare events. These distributions have been extensively used to model phenomena for which extreme values occur with a relatively high probability. In fields related to computer science and telecommunications, environmental sciences, finance or economics, it is not uncommon to stumble upon examples of large long-tailed data sets. Nowadays however, even distributions with slowly varying tails are shown to be of great value in practice. Although there is not a unified agreement on terminology, in the literature *super-heavy-tailedness* has been attached to a degree of tail heaviness associated with slow variation. For example, distributions such as log-Pareto, log-Cauchy or log-Weibull, all possessing logarithmically decaying tails, lay in the class of super-heavy-tailed distributions.

In this talk, rather than just using slow variation to characterize the super-heavy-tailed nature of a distribution, we consider a more restrictive definition resulting from Generalized Regular Variation (cf. Geluk and de Haan, 1987; de Haan and Stadtmüller, 1996), with the ulterior advantage of providing a simple division between heavy and super-heavy tails. In a quite intuitive way, this framework enables us to develop statistical inference methodologies, namely estimation and hypothesis testing procedures, for a non-negative parameter measuring the tail heaviness of the underlying distribution.

References

[1] Geluk, J., de Haan, L. (1987) Regular Variation, Extensions and Tauberian Theorems, CWI Tract 40, Center of Mathematics and Computer Science, Amsterdam, Netherlands.
[2] de Haan, L. and Stadtmüller, U. (1996) *Generalized regular variation of second order*, Journal of the Australian Mathematical Society. 61 (A), 381–395.

Analysis of climatic extreme events under non-stationary conditions

Nogaj, Marta (speaker) *LSCE, FRANCE*, marta.nogaj@cea.fr
Dacunha-Castelle, Didier *University Paris XI, FRANCE*
Yiou, Pascal *LSCE, FRANCE*

Non-stationary extreme events:

One prevalent question rising in the scientific community concerns the change in occurrence and amplitude of abrupt and widespread climate events with major impacts in the past decades. It is conceivable that the anthropic forcingof climate change could increase the probability of extreme events, such as floods or heat waves. Using Extreme Value Theory, we analyze non-stationary time-series of temperature and precipitation. The probability of an extreme event under non-stationary conditions depends on the rate of change of the parameters of the distribution as well as on the rate of change of the frequency of their occurrence. In this study, we use the NCEP reanalysis data (1948-2004) of temperature and precipitation over the extended region of the North Atlantic. The data being highly dependent, a pre-processing by declustering (elimination of data aggregates) is needed. We then investigate the distribution of extremes over a given threshold, so that the resulting dates of exceedances follow a non-stationary Poisson process and the associated peaks are fitted by a Generalized Pareto Distribution with time-dependent scale parameters (sigma). These conditions are checked with likelihood tests. Within this framework, the concept of the return period is altered, since the return level is highly dependent on the extrapolated period of consideration. Moreover, the scale and Poisson intensity parameters are explained by different covariates, such as time, the North Atlantic Oscillation (NAO) or the Greenhouse gas content.

In summary, this research brings us a step further in the estimation of climate change impacts on abrupt climate events. Indeed, amplitude and frequency variations of extreme temperatures and heavy precipitation are evaluated at different locations.

References

[1] Naveau P. et al. (2005) *Statistical Analysis of Climate Extremes.* , “Comptes rendus Geosciences de l’Academie des Sciences”, in press.
[2] Coles S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag.
[3] Katz R. et al. (2002) *Extremes in hydrology. Advances in Water Resources.*, Vol. 25, 1287-1304.

Contagion versus flight to quality in financial markets

Olmo, Jose (speaker) *Universidad Carlos III de Madrid, Spain*, jose.olmo@uc3m.es
Gonzalo, Jesus *Universidad Carlos III de Madrid, Spain.*

Contagion; Copula functions; Causality in the Extremes; Flight to quality; Interdependence; Multivariate extreme value theory.

None doubts that financial markets are related (interdependent). What is not so clear is whether there exists contagion among them or not, its intensity, and its causal direction. The aim of this paper is to define properly the term contagion (different from interdependence) and to present a formal test for its existence, the magnitude of its intensity, and for its direction. Our definitions of interdependence and contagion lie on tail dependence measures. Interdependence is defined by positive quadrant dependence (PQD) as in Lehmann (1966),

$$P\{Z_1 > z_1, Z_2 > z_2\} \geq P\{Z_1 > z_1\}P\{Z_2 > z_2\}, \tag{9}$$

or equivalently if

$$P\{Z_1 \leq z_1, Z_2 \leq z_2\} \geq P\{Z_1 \leq z_1\}P\{Z_2 \leq z_2\}, \tag{10}$$

with Z_1, Z_2 representing two financial markets. In the same way, contagion in intensity is defined by nonincreasing or nondecreasing tail dependence. In particular contagion in the upper tails is characterized by increasing tail monotonicity of the function $P\{Z_1 > z_1, Z_2 > z_2\} - P\{Z_1 > z_1\}P\{Z_2 > z_2\}$.

One of the main goals in this paper is testing directional contagion between the variables. This is interpreted as causality in the extremes. Contagion in this context occurs when one variable is influencing the other, that is, a large value in one variable is raising the likelihood of a large value in the other variable. It is defined by these properties

$$P\{Z_2 > z_2 | Z_1 > z_1\} \geq P\{Z_1 > z_2 | Z_2 > z_1\}, \tag{11}$$

for the upper tails, with $z_2 \geq z_1$, and

$$P\{Z_2 \leq z_2 | Z_1 \leq z_1\} \geq P\{Z_1 \leq z_2 | Z_2 \leq z_1\}, \tag{12}$$

for the lower tails, where $z_2 \leq z_1$.

These definitions are made operational through their equivalence with some copula properties. In the latter directional contagion boils down to see if $\widetilde{C}_G(u_1, u_2; \Theta) > \widetilde{C}_G(u_2, u_1; \Theta)$ with $u_1,\ u_2 \in [0, 1]$, and $\widetilde{C}_G$ a copula function. In turn, the relation between the variables under directional contagion must be asymmetric. In this paper we define a NEW copula, a variant of the Gumbel type, that is sufficiently flexible to describe different patterns of dependence, as well as being able to model asymmetric effects of the analyzed variables (something not allowed with the standard copula models). This copula function takes this expression

$$\widetilde{C}_G(u_1, u_2; \Theta) = \exp^{-D(u_1, u_2; \gamma, \eta)[(-log\ u_1)^{\theta} + (-log\ u_2)^{\theta}]^{1/\theta}}, \tag{13}$$

with $\theta \geq 1$, and

$$D(u_1, u_2; \gamma, \eta) = \exp^{\gamma(1-u_1)(1-u_2)^{\eta}}, \quad \gamma \geq 0, \quad \eta > 0. \tag{14}$$

Finally, we estimate our copula model to test the intensity and the direction of the extreme causality between bonds and stocks markets (in particular, the flight to quality phenomenon) during crises periods. We find evidence of a substitution effect between Dow Jones Corporate Bonds Index with 2 years maturity and Dow Jones Stock Price Index when one of them is through distress periods. On the contrary, if both are going through crises periods a contagion effect is observed. The analysis of the corresponding 30 years maturity bonds with the stock market reflects independent effects of the shocks.

References

[1] Lehmann, E.L. (1966) *Some Concepts of Dependence*, Ann. Math. Statist. 37, 1137–1153.

Log-scaling rainfall data: effects on GPD Bayesian goodness-of-fit

Ortego, Maria Isabel (speaker) *Universitat Politècnica de Catalunya, Spain*, ma.isabel.ortego@upc.edu
Egozcue, J.J *Universitat Politècnica de Catalunya, Spain*

predictive; hazard; model checking:

Goodness of fit testing is a problem of general concern. In the case of hazard problems extremal distributions are required, e.g. Generalized Pareto distributions (GPD). Then, additional difficulties arise due to the systematic scarcity of data. For instance, estimation of long return periods may strongly depend on the goodness of fit to the extreme model. When studying dependence between extremal variables, estimation of the associated copulas may depend dramatically on the estimation of the marginal distributions. In order to deal with the unavoidable uncertainty of the results, Bayesian methods are useful and, consequently, a Bayesian assessment of goodness of fit is appropriated. Bayesian estimation of GPD models has been used successfully, e.g. [1],[2].

At the same time, the selection of proper scales to describe phenomena arises as an important issue. Lots of phenomena are better described by a relative scale (e.g. positive data where the null value is unattainable) and are thus suitably treated in a logarithmic scale. Logarithmic scale has been used successfully for ocean-wave-height [1], and seems to be also adequate for daily rainfall data [2]. A new question then arises: which scale, raw or logarithmic, does provide better fit of the GPD model to the data?

Validation of the results requires goodness of fit testing. In a Bayesian framework, two approaches are selected to check the model, e.g. [4], [5]: the predictive p-value and the Bayesian p-value model checking. Both approaches evaluate discrepancies between the model and observed data. These issues are illustrated using a set of 30 years daily rainfall data. Hazard assessment of the rainfall data set is carried out with a standard model. Time-occurrence of events is assumed to be Poisson distributed, and the magnitude of each event is modelled as a random variable which upper tail is described by a GPD. Independence between this magnitude and occurrence in time is assumed, as well as independence from event to event. A Bayesian joint estimation of parameters (Poisson rate, scale and shape of the GPD) using BGPE, [3], is carried out. Posterior and predictive distributions are available.

GPD models are assessed using (a) predictive p-values and (b) Bayesian model checking. In (a)-method, for each value of the GPD shape-scale parameters, (ξ, β), both Kolmogorov-Smirnov and multinomial χ^2 goodness of fit test to the data is carried out so obtaining an asymptotic p-value, $v(\xi, \beta)$. The predictive p-value, p_v, is

$$p_v = \int_{\xi,\beta} v(\xi,\beta) f(\xi,\beta|\mathbf{x})\ d\beta d\xi\ ,$$

where $\mathbf{x}$ represent the data and $f(\xi, \beta|\mathbf{x})$ is the posterior joint density of the GPD parameters.

For (b)-method, a suitable discrepancy function $D(\mathbf{x}; \xi, \beta)$ is chosen. To obtain predictive replicates of the data, $\mathbf{x}^{(i)}$, simulated values $(\xi, \beta)^{(i)}$ are drawn from the posterior, and then $\mathbf{x}^{(i)}$ is simulated from $\mathrm{GPD}((\xi, \beta)^{(i)})$. The Bayesian p-value is computed as

$$p_b \simeq \mathrm{freq}_{(i)}[D(\mathbf{x}^{(i)}; (\xi,\beta)^{(i)}) \geq D(\mathbf{x}; (\xi,\beta)^{(i)})]\ .$$

Results of goodness of fit are discussed with reference to data scaling and GPD domain of attraction.

References

[1] Egozcue, J.J., Pawlowsky-Glahn, V. and Ortego, M.I. (2005) *Wave-height hazard analysis in Eastern Coast of Spain- Bayesian approach using generalized Pareto distribution*, Advances in Geosciences 2, 25–30.
[2] Egozcue, J.J., Ramis, C. (2001) *Bayesian hazard analysis of heavy precipitation in Eastern Spain*, Int. J. Climatology 21, 1263–1279.
[3] Egozcue, J.J. and Tolosana-Delgado, R. (2002) Program BGPE: Bayesian Generalized Pareto Estimation. Ed. Diaz-Barrero, J.L., ISBN 84-69999125, Barcelona, Spain.
[4] Gelman, A., Carlin. J.B., Stern, H.S. and Rubin, D.B. (1995) Bayesian Data Analysis, Chapman&Hall/CRC.
[5] Gelman, A., Meng, X.L. and Stern, H. (1996) *Posterior predictive assessment of model fitness via realized discrepancies*, Statistica Sinica 6, 733–807.

Kramers' type law for Lévy flights

Pavlyukevich, Ilya (speaker) *Humboldt-University of Berlin, Germany*, pavljuke@mathematik.hu-berlin.de
Imkeller, Peter *Humboldt-University of Berlin, Germany*

Lévy flights; first exit time; α-stable process; Kramers' law; extreme events; heavy tails; paleoclimate :

Motivated by the Greenland ice-core data analysis performed in [1], we study the climate dynamics X^ε driven by the α-stable Lévy noise:

$$X_t^\varepsilon = x - \int_0^t U'(X_{s-}^\varepsilon)\, ds + \varepsilon L_t, \quad x \in \mathbf{R}, t \geq 0, \varepsilon > 0,$$

where the process L is composed of a standard Brownian motion and a symmetric α-stable Lévy process, $\alpha \in (0,2)$

First, we study the exit problem of X^ε from intervals which contain the unique asymptotically stable critical point of the deterministic dynamical system $\dot{Y}_t = -U'(Y_t)$, (see [2]). Using probabilistic estimates we show that in the small noise limit $\varepsilon \to 0$, the exit time of X^ε from an interval is an exponentially distributed random variable and determine its expected value. In particular, if $\sigma_x(\varepsilon) = \inf\{t \geq 0 : X_t^\varepsilon > a\}$, $a > 0$, and 0 is the stable point, then

$$\mathbf{E}_x \sigma(\varepsilon) \approx \frac{\alpha a^\alpha}{\varepsilon^\alpha}, \quad \varepsilon \to 0, \quad x < a.$$

Due to the heavy-tail nature of the α-stable component of L, the results differ strongly from the well known case in which the deterministic dynamical system undergoes purely Gaussian perturbations (Kramers' law, Freidlin–Wentzel theory), where the mean exit time turns out to be exponentially large in ε^{-2}

Second, we consider the case of a double-well potential U, with wells' minima corresponding to colder and warmer climate states. We study transitions of X^ε between the wells and questions of convergence of X^ε to a two-state process (metastability).

References

[1] Ditlevsen, P. (1999) *Observation of α-stable noise induced millenial climate changes from an ice-record*, Geoph. Res. Lett. 10 (26), 1441–1444.
[2] Imkeller, P., Pavlyukevich, I. (2004) *First exit times of solutions of non-linear stochastic differential equations driven by symmetric Lévy processes with α-stable component*, arXiv:math.PR/0409246.

Testing the tail index in autoregressive models

Picek, Jan (speaker) *Technical University in Liberec, Czech Republic*, jan.picek@vslib.cz
Jurečková, Jana *Charles University in Prague, Czech Republic*
Koul, Hira L. *Michigan State University, U.S.A.*

Heavy tailed distribution; Feigin-Resnick estimator; Pareto tail index; GR-estimator:

Testing the hypothesis on the tail index of a heavy tailed distribution is an alternative inference to the classical point estimation, surprisingly not yet much elaborated in the literature. The tests often work under weaker conditions than the point estimators, can be easily reconverted into the confidence sets, and have an intuitive interpretation.

In the contribution, we construct a class of tests on the tail index of the innovation distribution in a stationary linear autoregressive model. Such tests would find many applications in the environmental, financial and other time series.

The tests are nonparametric and are based on the series of residuals with respect to an appropriate estimator of the AR parameters; more precisely, they are based on the empirical process of maximal residuals of non-overlapping segments of such series. The simulation study illustrates a very good level performance of the tests, comparable to the tests for the sequence of *i.i.d.* observations, constructed in Jurečková and Picek (2001). The methodological tools are based on Jurečková and Picek (2001) and on Koul(2002).

References

[1] Jurečková, J., Picek, J. (2001) *A class of tests on the tail index*, Extremes 4:2, 165–183.
[2] Koul, H., L., (2002) Weighted empirical processes in dynamic nonlinear models. Lecture Notes in Statistics 166, Springer-Verlag.

Extreme-value analysis: focusing on the fit and the conditions, with hydrological applications

Rakonczai, Pál (speaker) *Eötvös Loránd University, Budapest, Hungary*, paulo@cs.elte.hu
Bozsó, Dávid *Eötvös Loránd University, Budapest, Hungary*, davidszg@cs.elte.hu
Zempléni, András *Eötvös Loránd University, Budapest, Hungary*, zempleni@ludens.elte.hu

Anderson-Darling test; extreme-value distributions; generalized Pareto distributions; mixing condition:

Extreme value theory has been one of the most quickly developing area of mathematical statistics in the last decades. The methods, based on the Fisher-Tippett theorem about the possible limits of normalized maxima of iid random variables (fitting GEV distributions) or the analogous model based on threshold excesses and the corresponding GPD distribution are now routinely applied in very different branches of areas from financial mathematics to environmetrics, from reliability to internet traffic.

As the conditions which ensure the convergence to a GEV distribution are rather mild for absolutely continuous, iid observations, it is common practice to assume the asymptotics to hold and estimate the parameters of the limiting distribution if one has a data set like block maxima of random variables. In most cases some graphical methods, like Q-Q plots, are used for model validation. These are rather subjective, formal tests based on the plot are rarely applied. Another reason for not using them can be that the typical goodness-of-fit procedures, like the Kolmogorov-Smirnov statistics are not strong enough.

In the GPD case, threshold selection is a crucial step, which in practice is usually based on graphical tools, like the mean excess plot, or the investigation of the behaviour of the parameter estimates.

So an easy, objective procedure for this purpose would definitely be interesting, especially in cases, if - usually due to the vast amount of data sets to be analysed - one has to automatise the procedure as much as possible. In the paper we summarize the available tests for checking if the distribution of the sample is indeed that of a GEV/GPD. We propose and investigate a modified Anderson-Darling test, adapted to the problems, where the emphasis is on the fit for one of the tails of the distribution.

Another important question is the independence of the observations. Even if the marginals belong to the GEV/GPD class, the dependencies may seriously disturb the calculations such as those for the return levels. If one is aware of this problem, and uses the extremal index for example, the question about the validity of the convergence results should be investigated. Unfortunately conditions like $D(u_n)$ from [1] are by far not easy to check, but we show partial results for this problem by investigating the time-dependencies of the two dimensional distributions for selected thresholds.

We present applications of the proposed approach to Hungarian hydrological data sets.

References

[1] Leadbetter, M. R. - Lindgren, G. and Rootzén, H. (1983). Extremes and Related Properties of Random Sequences and Processes, Springer.

Maximal clusters in non-critical percolation and related models

Redig, Frank *Leiden University, the Netherlands*,redig@math.leidenuniv.nl

Let M_n be the cardinality of the largest cluster in a box of size n in the standard percolation model or FKG variants of it (such as random cluster measures with $q > 2$). We prove a "discrete Gumbel law" with error bounds for M_n. The main input is an exponential law with error bound for the occurrence of rare patterns in Gibbsian random fields. This is joint work with Remco van der Hofstad (TU Eindhoven).

Bayesian analysis of extremes in hydrology: a powerful tool for knowledge integration and uncertainties assessment

Renard, Benjamin (speaker) *CEMAGREF Lyon, France*, renard@lyon.cemagref.fr
Garreta, V., Lang, M. *CEMAGREF Lyon, France*
Bois, P. *CNRS-LTHE, Grenoble, France*

Bayesian analysis; hydrological extremes; historical information; stationarity;

In hydrology, the probabilistic behavior of streamflow extremes is of crucial importance for floods and droughts mitigation. Extreme value theory is a powerful tool to assess properties of hydrological extremes. Among different methods used to estimate the models parameters, Bayesian inference has become more and more popular in recent years (Coles and Powell, 1996). The aim of this talk is to present two possible applications and to illustrate the improvement we obtained compared to classical methods.

The first application deals with the incorporation of several sources of knowledge thanks to the prior distribution. We hope to improve the estimations accuracy in this way. On a French River with 50 years of data, we will take into account information about the general properties of the catchment (surface, geographic localization, rainfall properties), historical data about very high past events, and statistical behavior of extreme rainfalls. This knowledge will be translated into a prior distribution, using the procedure suggested by Coles and Tawn (1996). Finally, the parameters posterior distribution will be compared to the maximum likelihood estimators distribution.

The second application deals with detecting and taking into account changes in hydrological time series. Impacts of climatic change have been observed worldwide on several hydro-meteorological variables. Nevertheless, there is no clear evidence of a consistent change in extreme hydrological regimes, although some significant trends have been reported at-site. Consequently, stationarity of extreme values time series is not ensured, and should therefore be considered as an uncertainty. The approach we used is adapted from the Bayesian framework proposed in the Gaussian case by Perreault et al.(2000a, b). It is applied here on a French river with 93 years of data. We will consider several models (stationary, step-change and linear trend models) for the distributions of peaks-over-threshold and occurrence process. Prior distributions are specified by using regional knowledge on quantiles. Posterior distributions are used to estimate parameters for each model. The posterior probability of models can then be computed and used to derive a realistic frequency analysis, which takes into account estimation and stationarity uncertainties. Results will be compared to classical likelihood-based methods. These approaches consists in choosing the more suitable model with a statistical test, and then carrying out the frequency analysis with this model, thus ignoring the fact that test decision may be false.

Finally, we discuss about potential improvements and methodological difficulties we face. As an example, generalization of the approaches presented here to the multivariate case is appealing, for taking into account spatial dependance (regional trend detection) or processes dependance (combined peak/volume/duration analysis). Unfortunately, such generalization is far from being easy, as it implies both numerical and theoretical complications.

References

[1] Coles, S. G., Powell, E.A. (1996) *Bayesian methods in extreme value modelling: A review and new developments*, International Statistical Review. 64, 119–136.

[1] Coles, S. G., Tawn, J.A. (1996) *A Bayesian analysis of extreme rainfall data*, Applied Statistics-Journal of the Royal Statistical Society Series C. 45, 463–478.

[1] Perreault, L., Bernier, J., Bobee, B., Parent, E. (2000) *Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited*, Journal of Hydrology 235, 221–241.

[1] Perreault, L., Bernier, J., Bobee, B., Parent, E. (2000) *Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting*, Journal of Hydrology 235, 242–263.

Data network models of burstiness

Resnick, Sid *Cornell University, USA*, sir1@cornell.edu

We review characteristics of data traffic which we term *stylized facts*: burstiness, long range dependence, heavy tails, bursty behavior determined by high bandwidth users, dependence determined by users without high transmission rates. We propose an infinite source Poisson input model which accounts for traffic in adjacent time slots. This model has the ability to account for many of the stylized facts.

Inference for the limiting cluster size distribution of extreme values

Robert, Christian Y. *CNAM and CREST, France*, christian.robert@cnam.fr

Exceedance point processes; Limiting cluster size distribution; Extremal index; Strictly stationary sequences.

Let (X_n) be a strictly stationary sequence of random variables (rvs) with distribution F. We assume that for each $\tau > 0$ there exists a sequence of levels $(u_n(\tau))$ such that $\lim_{n\to\infty} n(1 - F(u_n(\tau))) = \tau$. The point process of time normalized exceedances $N_n^{(\tau)}(\cdot)$ is defined by $N_n^{(\tau)}(B) = \sum_{i=1}^n I_{\{i/n \in B, X_i > u_n(\tau)\}}$ for any Borel set $B \subset E := (0,1]$. If (X_n) is a sequence of independent and identically distributed (iid) rvs, $N_n^{(\tau)}$ converges weakly to a homogeneous Poisson process with intensity τ (see e.g. Embrechts *et al.* (1998)). If the iid assumption is relaxed and a long range dependence condition is assumed, the limiting point process for $N_n^{(\tau)}$ is necessarily a homogeneous compound Poisson process with intensity $\theta\tau$ ($\theta > 0$) and limiting cluster size distribution π (Hsing *et al.* (1988)). θ is referred to as the extremal index and its reciprocal θ^{-1} is equal to the mean of π. It may be shown that $\theta \leq 1$ and that the compound Poisson limit for $N_n^{(\tau)}$ becomes Poisson when $\theta = 1$.

Although estimation of the extremal index has generated a huge literature, there is very few papers which investigate the inference for the limiting cluster size distribution (see Hsing (1991, 1993) or more recently Ferro (2003)). In this paper we introduce new estimators of the cluster size probabilities which are constructed from the compound probabilities of the point process through a recursive algorithm. Indeed it is very simple to estimate the compound probabilities and then to use a declustering algorithm to form estimates of the cluster size probabilities. More specifically, let us denote by $p^{(\tau)} = \left(p^{(\tau)}(m)\right)_{m\geq 0}$ the distribution of the weak limit of $N_n^{(\tau)}(E)$ as $n \to \infty$ when it exists. It can be shown that there exist differentiable functions such that

$$\pi(m) = f_m\left(p^{(\tau)}(0), p^{(\tau)}(1), ..., p^{(\tau)}(m)\right), \; m \geq 1. \tag{15}$$

The paper is organized as follows. First, we explain more precisely how we construct the estimators of $p^{(\tau)}$ and π, and how we derive three new estimators of the extremal index. We emphasize that we consider estimators which rely on estimated thresholds. Second, general theory for weak convergence of multi-level exceedance point processes is presented. Indeed the asymptotic joint distribution of the estimators of the cluster size probabilities is related to the asymptotic distribution of the two-level exceedance point process. Third, we discuss conditions for consistency and asymptotic normality of the estimators. Finally we inspect the finite sample behavior of the estimators on simulated data.

References

[1] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1998) Modelling extremal events. Springer, Applications of Mathematics.

[2] Ferro, C.A.T. (2003) Statistical methods for clusters of extreme values. Phd Thesis, Lancaster University.

[3] Hsing, T. (1991) *Estimating the parameters of rare events*, Stoch. Proc. Appli. 37, 117-139.

[4] Hsing, T. (1993) *On some estimates based on sample behavior near high level excursions*, Proba. Theory Relat. Fields 95, 331-356.

[5] Hsing, T., Hüsler, J. and Leadbetter M.R. (1988) *On the exceedance point process for a stationary sequence*, Proba. Theory Relat. Fields 78, 97-112.

Estimation of the long memory parameter using an Infinite Source Poisson model applied to transmission rate measurements

Roueff, Francois *Télécom Paris* roueff@tsi.enst.fr

Infinite Source Poisson Model, Heavy tails, long range dependence, Traffic modelling:

We present a long memory processes related to a Poisson point process, give its main properties, asymptotic behaviour and discuss some statistical issues with a view on Internet traffic analysis. The Infinite Source Poisson model is a generalisation of the M/G/∞ queue. Arrivals are driven by a homogeneous Poisson process, durations of active periods are independent and identically distributed (iid) and independent of the arrivals. Each active periods (say dowload sessions, or flows) is assumed to have a constant transmission rate and the available bandwidth to be unlimited. Theses rates are iid, independent of the arrivals but possibly depending on the durations. In a traffic modelling context, the obtained process $X(t)$ can serve for modelling the bandwith occupation, often called the *workload*. The stability of the model depends on the tail behavior of the duration distribution. Both in the stable and unstable cases, the tail behavior of the durations can be recovered from the dependence structure of $X(t)$. In particular, heavy-tailed durations will result in long range dependence (LRD) for $X(t)$ and the corresponding tail and Hurst indices α and H satisfy $H = (3 - \alpha)/2$ for all $\alpha \in (0, 2)$. In practical situations, the process $X(t)$ is observed through passive measurements, by counting packets going trough a point of the network, and then by evaluating the instantaneous workload. Such measurements are much simpler than collecting complete characterizations of the flows. However, from a queuing point of view, as mentionned above about the stability, the important parameter is the flow duration tail index α. The object of this work is to rely on the relationship between α and H for estimating α from measurements on $X(t)$.

Poisson Cluster Process as a model for teletraffic arrivals and its extremes

Samorodnitsky, Gennady (speaker) *Cornell University, USA*, gennady@orie.cornell.edu
González-Arévalo, Bárbara *University of Louisiana at Lafayette, USA*
Mikosch, Thomas *University of Copenhagen, Denmark*

Poisson cluster process; heavy tails; long memory; invariance principle:

We consider a Poisson cluster process N as a generating process for the arrivals of packets to a server. This process generalizes in a more realistic way the infinite source Poisson model which has been used for modeling teletraffic for a long time. At each Poisson point Γ_j, a flow of packets is initiated which is modeled as a partial iid sum process $\Gamma_j + \sum_{i=1}^{k} X_{ji}$, $k \leq K_j$, with a random limit K_j which is independent of (X_{ji}) and the underlying Poisson points (Γ_j). We study the covariance structure of the increment process of N. In particular, the covariance function of the increment process is not summable if the right tail $P(K_j > x)$ is regularly varying with index $\alpha \in (1,2)$, the distribution function of the X_{ji}'s being irrelevant. This means that the increment process exhibits long-range dependence. If $\mathrm{var}(K_j) < \infty$ long-range dependence is excluded. We study the asymptotic behavior of the process $(N(t))_{t \geq 0}$ and give conditions on the distribution of K_j and X_{ji} under which the random sums $\sum_{i=1}^{K_j} X_{ji}$ have a regularly tail. Using the form of the distribution funtion of the interarrival times of the process N under the Palm distribution , we also conduct an exploratory statistical analysis of simulated data and of Internet packet arrivals to a server. We illustrate how the theoretical results can be used to detect distributional characteristics of K_j, X_{ji}, and of the Poisson process.

A closer look at the Hill estimator: Edgeworth expansions and confidence intervals

Segers, Johan (speaker) *Tilburg University, the Netherlands*, jsegers@uvt.nl
Haeusler, Erich *University of Giessen, Germany*

Key Words: Confidence interval; Edgeworth expansion; Hill estimator; Sample fraction

We establish Edgeworth expansions for the distribution function of the centered and normalized Hill estimator for the positive extreme value index [4]. The expansions differ from classical Edgeworth expansions because of the presence of extra terms reflecting the bias of the Hill estimator.

The expansions are used to assess the accuracy of confidence intervals for the extreme value index based on the Hill estimator. For one-sided intervals, this was already done in [3] based on first-order Edgeworth expansions in [1] and [2]. For two-sided intervals, however, a second-order Edgeworth expansion for the Hill estimator is required in order to obtain the main correction term in the coverage probability expansion. The resulting coverage probability predictions are surprisingly accurate and motivate data-driven methods to determine the sample fraction for which the true coverage probability matches the nominal one.

References

[1] Cheng, S., de Haan, L. (2001) "Penultimate Approximation for Hill's estimator", *Scandinavian Journal of Statistics* **28**, 569–575.

[2] Cheng, S., Pan, J. (1998) "Asymptotic expansions of estimators for the tail index with applications", *Scandinavian Journal of Statistics* **25**, 717–728.

[3] Cheng, S., Peng, L. (2001) "Confidence intervals for the tail index", *Bernoulli* **7**, 751–760.

[4] Cuntz, A., Haeusler, E., Segers, J. (2003) "Edgeworth expansions for the distribution function of the Hill estimator", Tilburg University, CentER discussion paper 2003-08, http://center.uvt.nl/pub/dp/.

Cox limit theorem for high level a-upcrossings by χ-process

Stamatovic, Sinisa *University of Montenegro, Serbia and Montenegro*, sins@cg.ac.yu

Gaussian process; high level; Berman inequality:

This paper is a continuation and a refinement of researches which have been initiated in [2], [3]. Cox limit theorem for a-points of upcrossings of a high level by trajectories of the process $\chi(t)$ is established. This result is more general then previously, we include the case of higher dependence for remote points of the basic process.

References

[1] Piterbarg, V. (1996) Asymptotic Methods in the Theory of Gaussian Processes and Fields, AMS, Providence.
[2] Piterbarg, V., Stamatovic, S. (2003) *Limit theorem for high level* $a-$*upcrossings by* $\chi-$*process*, Probability Theory and Applications, 48, 734–741.
[3] Konstant, D., Piterbarg, V., Stamatovic, S. (2004) *Gnedenko-type limit theorems for cyclostationary* χ^2 *processes*, Lith. Math. J., 44, 157–167.

GEV flood quantile estimators with Bayesian GLS shape-parameter regression

Reis Jr., Dirceu S *Fundaçâo Cearense de Meteorologia e Recursos Hídricos, Fortaleza, Brazil*
Stedinger, Jery R. (speaker) *Cornell University, Ithaca, NY, USA*, jrs5@cornell.edu
Martins, Eduardo Savio *Fundaçâo Cearense de Meteorologia e Recursos Hídricos, Fortaleza, Brazil*

Floods; GEV distribution; GLS regression; Bayesian estimators

The GEV distribution is often used to estimate flood and precipitation quantiles, though estimation of the shape parameters with records of 50 years or less is problematic unless supplemental or prior shape information is provided (Coles and Dixon, 1999; Martins and Stedinger, 2000). Reis et al. (2003) developed a quasi-analytic Bayesian analysis of GLS-regression regional shape parameter estimators wherein the sampling variance of the estimators is much larger than the regional variation in the true values, and shape estimators are cross-correlated (Martins and Stedinger, 2002). The resulting regression model provides a mean and variance for the GEV shape parameter κ that defines site-specific prior distributions that are the basis for Generalized Maximum Likelihood estimators (GMLE) of flood quantiles, rather than the general and less informative geophysical prior developed by Martins and Stedinger (2000). For the Illinois River (USA), the best regional κ model employs a binary basin variable Z, log(Area), and log(Slope) as explanatory variables, and remarkably was as precise as a 114 year record! Use of ordinary or weighted least squares estimators yield very different results; other data sets yielded similar conclusions (Reis, 2005). Comparisons of maximum likelihood estimation (MLE), GMLE [geophysical prior], and GMLE [site specific prior] show that such precise site-specific regional κ-priors significantly reduce the uncertainty in extreme flood quantile estimators, even for sites with long record lengths.

[Oral presentation preferred.]

References

[1] Coles, S.G. and M.J. Dixon, "Likelihood-Based Inference for Extreme Value Models," Extremes, 2(1), 5-23, 1999.
[2] Martins, E.S. and J.R. Stedinger, "Generalized Maximum Likelihood GEV Quantile Estimators for Hydrologic Data," Water Resources Research, 28(11), 3001-3010, 2000.
[3] Martins, E.S., and J.R. Stedinger, "Cross-correlation among estimators of shape", Water Resources Research, 38(11), doi:10.1029/2002WR001589, 2002.
[4] Reis, D.S., Jr., Flood Frequency Analysis Employing Bayesian Regional Regression and Imperfect Historical Information, Ph.D. Thesis, Cornell University, Ithaca, NY, USA, 2005.
[5] Reis, D. S., Jr., J.R. Stedinger, and E.S. Martins, Bayesian GLS Regression with application to LP3 Regional Skew Estimation, Proceedings World Water and Environmental Resources Congress 2003, Paper No. 811, Editors P. Bizier and P. DeBarry, Philadelphia, PA, American Society of Civil Engineers, June 23-26, 2003.

A survey of software for the analysis of extreme values

Stephenson, Alec *National University of Singapore, Singapore*, stasag@nus.edu.sg

Extreme value theory; Software development; Software environments:

The last few years has seen a significant increase in publicly available software specifically targeted to the analysis of extreme values. This reflects the increase in the use of extreme value methodology by the general statistical community. Extreme value theory motivates a large number of statistical techniques, even among those that are regarded as standard elements of a practitioners toolkit. The software that is available for the analysis of extremes has evolved in essentially independent units, with most being extensions of larger software environments. An inevitable consequence is that these units are spread about the statistical landscape, and hence the techniques incorporated into each cannot be determined without considerable time and effort. We seek to simplify the process somewhat by reviewing the software that is currently available, and summarising the types of analyses that can be performed.

Practical issues in applications of multivariate extreme values

Tawn, Jonathan (speaker) *Lancaster University, UK*, j.tawn@lancaster.ac.uk
Caroline Keef *Lancaster University, UK*
Mark Latham *Lancaster University, UK*

Conditional Modelling; Missing Data; Multivariate Extremes; Spatial Extremes; River Flows; Surge Levels:

Statistical methods for the application of multivariate extreme value theory have developed rapidly over the last 20 years. A wide range of methods, including parametric and nonparametric approaches, have been proposed. Despite this development almost all published applications are bivariate, with one or two notable exceptions, and they tend not to address the practical data handling issues that arise with real applications. In this talk we will outline two major environmental studies that we are involved in where multivariate extreme value methods are being applied.

In the first study the spatial extremes of sea surges in the North Sea are modelled. We have numerical model hindcast hourly surge data on a grid of 259 sites for the period of 1955-2000. The numerical model was driven by a high-quality hindcast of the meteorology for this period and had been shown to reproduce the surge well at observational sites. The talk will focus on estimating the characteristics of the spatial field of extreme values; this is important for the insurance industry for determining offshore and coastal loss distributions. The large dimension of the problem leads to new problems related to the visualisation of the fitted dependence structure, testing for asymptotic dependence, parsimony, identifying the minimum number of sites to capture all asymptotic dependence, and the generation of super-storms. Given the finding of decadal scale trends in marginal distributions [1], in the longer term we are interested in assessing whether there are temporal trends in the dependence structure.

In the second study we have assess to daily river flow data at the entire UK network of sites. We are interested in estimating the risk of flooding over different gauges along a river and for gauges in different rivers given that flooding is observed at a given gauge. Unlike the surge data this application has irregularly located sites with no clear distance measure of dependence. However, the feature we will emphasis is the large amount of missing data over the entire network. Currently no methods exist for handling missing data in multivariate extreme values, consequently the default is to analyse only those data observed simultaneously on all variables. Such an approach is potentially highly inefficient relative to having access to the full data. In the talk we will introduce an approach for handling the missing data in multivariate extremes and illustrate its efficiency properties.

The multivariate extreme value method that we use as a starting point in each application is the conditional approach of [2]. That model gives a semi-parametric dependence model with a natural regression interpretation where the conditional mean and variance functions are fully parametric but the residual distribution is non-parametrically modelled. In particular if X is a Gumbel variables, $\mathbf{Y}$ is a d-dimensional variable with Gumbel marginal variables, and u is a high threshold, then the conditional model gives that for $X = x$ with $x > u$

$$\mathbf{Y} = \mathbf{a}x + x^{\mathbf{b}}\mathbf{Z}$$

where $\mathbf{0} < \mathbf{a} \leq \mathbf{1}$ and $-\mathbf{1} < \mathbf{b} < \mathbf{1}$ are d-dimensional parameters and Z is a d-dimensional residual variable which is independent of X and is modelled non-parametrically. Consequently dependence is parametrised conditionally through pairwise relationships with all higher order structure absorbed by the non-parametric component of the model. The variables X and Y_j are asymptotically dependent if $a_j = 1$ and $b_j = 0$ but are asymptotic independent otherwise. Estimation of $\mathbf{a}$ and $\mathbf{b}$ is therefore fundamental to estimation of the multivariate extreme value behaviour of $(X, \mathbf{Y})$.

References

[1] Butler, A., Heffernan, J. E., Tawn, J. A., Flather, R. A. and Horsburgh, K. (2005). *Recent trends in North Sea storm surge elevations*. Submitted to Coastal Shelf Research.
[2] Heffernan, J. E. and Tawn, J. A. (2004). *A conditional approach to modelling multivariate extreme values (with discussion)* J. Roy. Statist. Soc., B, 66, 497–547.

On univariate extreme value statistics and the estimation of reinsurance premiums

Vandewalle, Björn (speaker) *Katholieke Universiteit Leuven, Belgium*,
Bjorn.Vandewalle@wis.kuleuven.ac.be
Beirlant, Jan *Katholieke Universiteit Leuven, Belgium*

Excess-of-loss reinsurance rating; Wang's premium principle; heavy-tailed distributions:

In this talk, we consider univariate extreme value statistics in the background of certain (re)insurance applications, more specifically relating to the calculation of insurance premiums for excess-of-loss reinsurance policies in excess of a high retention level. Special attention is paid to heavy-tailed distributions and Wang's premium principle (Wang, 1996), as a generalization to the popular net premium principle.

Through these principles and following the line of reasoning as in Beirlant et al. (2001), estimators of the tail index and small exceedance probabilities allow for the correct estimation of reinsurance premiums. Next to the construction of estimators, we also consider the corresponding asymptotic results and illustrate the finite sample behavior through several simulations.

References

[1] Beirlant, J, Matthys, G. and Dierckx, G. (2001) *Heavy-tailed distributions and rating*, ASTIN Bulletin. 31, 41–62.
[2] Wang, S. (1996) *Premium calculation by transforming the layer premium density*, ASTIN Bulletin. 26, 71–92.

Weak & strong financial fragility

de Vries, Casper G. (speaker) *Erasmus University Rotterdam, The Netherlands*, cdevries@few.eur.nl
Geluk, Jaap *Erasmus University Rotterdam, The Netherlands*
de Haan, Laurens *Erasmus University Rotterdam, The Netherlands*

Market Linkages, Crisis Periods, Multivariate Extreme Value Analysis, Asymptotic Dependence :

The linkages in terms of returns and volatility between financial service institutions asset returns during periods of crisis is one of two types, depending on whether the returns are either asymptotically independent (weak fragility) or asymptotically dependent (strong fragility), regardless their correlation. If asymptotically independent, the dependency when present, eventually dies out completely at the more extreme quantiles. We study the joint loss behavior of correlated bank portfolios, due to e.g. loan syndication, under the weak assumption that the asset and liability return distribution are in the domain of attraction of an extreme value distribution. Thus we cover both discrete and continuous compounding. It is shown that due to the bank portfolio induced linearity, the type of extreme value distribution to which the marginal distributions are attracted determines whether the fragility is weak or strong and we provide an index for the dependency. This permits a characterization of systemic risk inherent to different bank network structures. The theory also suggests the functional form of the economically relevant copulas.

Two dependence measures for multivariate extreme value distributions

Weissman, Ishay *Technion - Israel Institute of Technology, Israel*, ieriw01@ie.technion.ac.il

Keywords: Fréchet margins; Pickands dependence function :

Let $G(x_1, x_2, \cdots, x_d) = G(\mathbf{x})$ be a multivariate extreme value distribution on R^d and let $\mathbf{X} = (X_1, X_2, \cdots, X_d)$ be a random vector G-distributed. We are interested in measuring the dependence among the components of $\mathbf{X}$ (beyond the conventional pair-wise dependence) in a way which is marginal-free. Thus, without loss of generality, we assume that all the marginal distributions G_j of G are unit-Fréchet, namely, $G_j(x) = \exp(-1/x)$ $(x > 0,\ j = 1, 2, \cdots, d)$. Let $\lambda = -\log G$ be the exponent function of G, then it is known that

$$\max \frac{1}{x_j} \leq \lambda(\mathbf{x}) \leq \sum \frac{1}{x_j} \qquad (\mathbf{x} \in \mathrm{R}^d_+),$$

where the upper bound corresponds to total independence and the lower bound to complete dependence. Define the (generalized) Pickands function on the unit-simplex $\Omega = \{\mathbf{v} = (v_1, \cdots, v_d) : v_j \geq 0, \sum v_j = 1\}$

$$A(v_1, v_2, \cdots, v_d) = \lambda(v_1^{-1}, v_2^{-1}, \cdots, v_d^{-1}) \quad (\mathbf{v} \in \Omega).$$

Then, A is convex and satisfies

$$\max v_j \leq A(\mathbf{v}) \leq 1$$

(see Beirlant et al, 2004). Again, the cases $A(\mathbf{v}) \equiv 1$ and $A(\mathbf{v}) \equiv \max v_j =: A_0(\mathbf{v})$ correspond, respectively, to total independence and complete dependence. Let

$$S_A = \int_\Omega (1 - A(\mathbf{v}))d\mathbf{v}$$

be the volume under the function $1 - A$. It is suggested to measure the dependence among the components of $\mathbf{X}$ by either

$$\tau_1 = \frac{d}{d-1}\left(1 - A\left(\frac{1}{d}, \frac{1}{d}, \cdots, \frac{1}{d}\right)\right)$$

or by

$$\tau_2 = \frac{S_A}{S_{A_0}}.$$

Both coefficients are equal to 0 or 1 in the two extreme cases of independence or complete dependence. In this paper we discuss some of their properties and raise some interesting open questions.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., de Wall, D. and Ferro, C. (2004) Statistic of extremes: Theory and applications, Wiley.

Asymptotically (in)dependent multivariate maxima of moving maxima processes

Janet E. Heffernan, Janet E. *Lancaster University, United Kingdom*
Tawn, Jonathan *Lancaster University, United Kingdom*
Zhang, Zhengjun (speaker) *Washington University in Saint Louis, USA*, zjz@math.wustl.edu

In this paper, we first extend Smith and Weissman's M4 processes to include non-Fréchet margins such as generalized extreme value (GEV) distributions. We then introduce a new model which contains independent random variables (cross sections and cross time), and M4 processes. The underlying moving random variables are either unit Fréchet or unit exponential in this new model. This new model has the ability to model negative dependence, near independence, positive dependence, asymptotic (in)dependence. Properties of each model will be studied carefully. In particular, asymptotic dependence indexes, coefficients of asymptotic dependence, and extremal indexes are derived for each case.

WAFO - a MATLAB toolbox for analysis of random waves and loads

Åberg, Sofia *Lund University, Sweden*, s_aberg@maths.lth.se

Gaussian processes; wave spectra; estimation; simulation; rainflow cycles; fatigue:

WAFO (Wave Analysis for Fatigue and Oceanography) is a toolbox of Matlab routines for statistical analysis and simulation of random waves and loads. Using WAFO you can, for example, calculate theoretical distributions of wave characteristics from observed or theoretical power spectra of the sea or find the density of rainflow cycles from parameters of random loads. The routines are based on algorithms for extreme values and crossing analysis. The toolbox, which is freely available at `http://www.maths.lth.se/matstat/wafo/`, is developed for a wide audience including ocean engineers, mechanical engineers as well as researchers who are interested in statistical analysis of random processes. The free availability makes it possible for anyone to reproduce results in scientific articles that were obtained using WAFO. In this talk an overview of the toolbox will be given and illustrative examples will show how it can be used. Furthermore examples from recent research will point out some of its possibilities and show that WAFO is still in progress.

References

[1] Brodtkorb, P.A., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E. (2000) *WAFO - a Matlab toolbox for analysis of random waves and loads*, Proceedings of the 10th International Offshore and Polar Engineering conference, Seattle, Vol III, 343–350.

[2] The WAFO-group: Brodtkorb, P.A., Frendahl, M., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E. (2000) *WAFO - Tutorial*, http://www.maths.lth.se/matstat/wafo/documentation/

Index