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

Gothenburg, August 15-19, 2005

Posters

UK Air Pollution: Modelling exceedances over a moving threshold

Eastoe, Emma *Lancaster University, UK*, e.eastoe@lancaster.ac.uk
Janet E. Heffernan *Lancaster University, UK*
Jonathan A. Tawn *Lancaster University, UK*

Air Pollution; Generalised Pareto Distribution; Moving threshold; Asymptotic independence; Cluster maxima:

Air pollution has a deleterious effect on human health, damages ecosystems, causes deterioration of many natural materials and reduces quality and quantity of crop yields. Discussions of such effects are well published although the PORG report [1] is of specific interest to a study of the sources and effects of air pollution in the UK. Within the last half century the need to impose limits on the production of air pollutants has become increasingly clear. Models of extreme levels of air pollutants can be used to create and test legislation designed to restrict anthropological emissions of air pollutants.

We consider the modelling of five air pollutants at twelve sites in the UK. These pollutants are Nitrogen Dioxide, Nitrogen Oxide, Ozone, Particulate Matter with a diameter of less than 10μm and Sulphur Dioxide. Our observations are the daily maxima. We treat the variables at each of the sites as independent. We also assume the sites are independent of each other. To demonstrate our methods we focus on a single site (Swansea) before giving a summary of all the sites. We look for similarities between models at different sites.

The data display short term dependence (clustering) and non-stationarity (seasonality). We follow the methods of Ferro and Segers [2] in our estimation of the extremal index (EI) to decluster the data. Since this requires that the data is stationary we first transform the data to the standard exponential scale using a local nonparametric method of transformation. The local nature of this transformation ensures that the transformed data are stationary. After threshold selection and declustering the data and threshold are transformed back to the original scale. This results in a threshold that changes in time, generally following the seasonal pattern of the data.

Estimation of the EI on the exponential scale shows that the data are asymptotically independent but that there is clustering at the finite threshold levels that we want to use. Following a result from Ledford and Tawn [3] we show that the GP distribution is an appropriate model for the cluster maxima which result from declustering at these levels. Accordingly, we fit GP distributions to the cluster maxima of each of the variables. To model seasonality and any linear trends we allow the model parameters to be functions of time. Seasonality can then be modelled by the inclusion of appropriate Fourier series to model the GP parameters, see, for example, Coles [4].

We assume that, given a site, the five pollutants are independent of each other. We also assume that the twelve sites are independent of each other. There are a number of reasons why it is reasonable that neither assumption should hold. Reactions between chemicals in the atmosphere may invalidate the first assumption and meteorological conditions common to several sites may invalidate the second. We hope to use the models fitted here as marginal models in future work, in which we model, separately for each site, the dependence between the variables. We then hope to fit a single global model to the full data set.

References

[1] Fowler, D., *et al* (1997) Fourth Report of the Photochemical Oxidants Review Group, Department of the Environment, Transport and the Regions.
[2] Ferro, C.A.T., Segers, J. (2003) *Inference for clusters of extreme values*, J. R. Statist. Soc. B. 65, 545–556.
[3] Ledford, A.W., Tawn, J.A. (2003) *Diagnostics for dependence within time series extremes*, J. R. Statist. Soc. B. 65, 521–543.
[4] Coles, S., (2001) An Introduction to Statistical Modelling of Extreme Values, Springer.

Testing the tail index using the right-spread function

Fernandez-Ponce, Jose Maria *Universidad de Sevilla, Spain*, ferpon@us.es
Rodriguez-Griñolo, Rosario *Universidad de Sevilla, Spain*
Pino-Mejias, Jose Luis *Universidad de Sevilla, Spain*
Gomez-Gomez, Teresa *Universidad de Sevilla, Spain*

Key words:Exponentiality; Right-spread Function; Strictly Algebraic Distributions;

Testing the tail index of a distribution function is an interesting problem which has widely been studied in the Statistical literature. Graphical methods were presented to characterize the nature of a distribution. In the algebraic case, that statistical tool provides an estimation procedure of the parameter characterizing the decrease of the survival function. This method is based on the relationship between the duration of exceeding an intensity threshold and the accumulation of the realizations of the random variables during this length of time. If the data proceed from an exponential distribution then the threshold will tend to infinity. This fact can not be appreciated by using graphical methods when the limit of the relative excess mean function is zero. Furthermore, the threshold should not be too high since the divergence of the bias. A high sample size is also needed to stable the tail index estimator for strictly algebraic distributions. These problems can be solved using the right-spread function. This function possesses a bounded domain and characterizes graphically the exponential distribution versus any continuous distribution function. In this paper, we study a new approach to testing exponentiality and an estimation for the tail index q in the algebraic case by using the right-spread function.

References

[1] Bahadur, R.R. (1966),*A note on quantiles in large samples,* Ann. Math. Stat., **37**, 577-580.
[2] Basu, D. (1975),*Statistical Information and likehood,* Sankhya, A, 37, 1-71.
[3] Borokov, A.A. (1988), *Estadística Matemática,* Mir Moscú.
[4] Chaouche, K; Hubert, P. and Lang, G. (2002), *Graphical characterisation of probability distribution tails,* Stochastic Environmental Research and Risk Assessment, 16, 342-357.
[5] Fernández-Ponce, J.M., Kochar, S.C., Muñoz-Perez, J. (1998), *Partial orderings of distributions based on right spread function,* J. Appl. Prob., **35**, 221-228.
[6] Karr, A.F.,(1993) *Probability*, Springer-Verlag, New York, Inc.
[7] Shaked, M. and Shanthikuman, J.G. (1998), *Two variability orders,* Prob. Eng. Infor. Sci., **12**, 1-23

A new model for time series dependence and the extremal index of higher-order Markov chains

Ferreira, Helena *University of Beira Interior, Covilhã, Portugal* ferreira@fenix2.ubi.pt

Extremal coefficient, dependence, extremes, extremal index, higher-order stationary Markov sequences:

We present a new model for time series dependence and derive one criterion for convergence of extremes.

The dependence structure of a stationary sequence is described by a sequence of extremal functions. Under a stability condition for the sequence of extremal functions, we obtain the asymptotic distribution of the sample maximum.

As a corollary, we derive a surprisingly simple method of computing the extremal index through a limit of a sequence of extremal coefficients.

The results may be used to determine the asymptotic distribution of extreme values from stationary time series based on copulas. We illustrate it with the study of the extremal behaviour of dth-order stationary Markov chains in discrete time with continous state space. For such sequences we present a way to compute the extremal index from the upper extreme value limit for its joint distribution of $d + 1$ consecutive variables.

The estimation of scale parameter based on fourth central moment in large deviation

Ghodsi, Maryam *Islamic Azad University, Iran*
Farnoosch, R. *Iran University of Science and Technology, Iran*
Jafarpour, H. *Islamic Azad University, Iran*

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

In this paper we consider actual values of standardized fourth central moment for scale distributions. Then we generate data from these distributions and compute estimations of scale parameter. We compare the estimators by actual values We discuss about their properties and finally we suggest the best estimators.

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] Stuart, A. and Ord, J. K. (1994), Kendalls Advanced Theory of Statistics, Volume 1: Distribution Theory, 6th ed., Arnold, London.

Analysis of flood in northern Moravia

Jarušková, Daniela *Czech Technical University, Czech Republic*, jarus@mat.fsv.cvut.cz
Hanek, Martin *Charles University, Czech Republic*

Flood in Northern Moravia; hydrological extremes; precipitation; case study:

In 1997 a big flood plagued the Northern Moravia (part of the Czech Republic). The water discharges of moravian rivers reached almost hundred times their average. According to public opinion the flood was caused by unusual weather conditions when severe storms occurred almost simultaneously in the basin drained by the rivers Opava and Opavice. In a case study we examine nine precipitation series and four water discharges series and try to find relationship between them. Using the theory of extremes we would like to answer the question whether such floods are likely to appear again.

References

[1] Coles S.G., Tawn J.A. (1994) *Statistical Methods for Multivariate Extremes: an Application to Structural Design*, Appl. Statist. 43, 1–48.
[2] Coles S.G., Tawn J.A. (1994) *A Bayesian Analysis of Extreme Rainfall Data*, Appl. Statist. 45, 463–478.
[3] Coles S.G., Hafernan J., Tawn J.A. (1999) *Dependence Measures for Extreme Value Analyses*, Extremes 2:4, 339–365.
[4] Engeland K., Hisdal H., Frigessi A. (2004) *Practical Extreme Value Modelling of Hydrological Floods and Droughts: A Case Study*, Extremes 7, 5–30.
[5] Embrechts P., Klüppelberg C., Mikosch, T. (1997) Modelling Extremal Events, Springer.

Ruin problem for integrated stationary Gaussian process

Kobelkov, Sergey *Moscow State University, Russia*, sergey@dodo.inm.ras.ru

Gaussian process, Rice method, ruin probability:

Exact asymptotic of ruin probability is found in the case when the profit rate has the form of Gaussian stationary process. The result is obtained by Rice's moment method.

Consider the ruin probability

$$P(u) = P\{\exists\, t \geq 0 : \int_0^t X_s\, ds - ct^\theta \geq u\},$$

where $\theta > 1/2$, X_t is a zero-mean real-valued stationary gaussian process with twice differentiable covariation function $R(t)$ such that $G = \int_0^\infty R(s)\, ds > 0$, $H = \int_0^\infty sR(s)\, ds < \infty$, and $u^{2-2/\theta} \int_{u^{1/\theta}}^\infty sR(s)\, ds \to 0, u \to \infty$. Then,

$$P(u) = \frac{\sqrt{R(0)}}{\sqrt{2\pi}} u^{-1+1/\theta} (2\theta-1)^{1/2-1/\theta} c^{-1/\theta} \exp\left\{-u^{2-1/\theta} \frac{(1+\tau_{\min}^\theta(2\theta-1)^{-1}))^2}{2G(u(2\theta-1)^{-1}/c)^{1/\theta}\tau_{\min} - 2Hu^{-1/\theta}}\right\}(1+o(1))$$

as $u \to \infty$, where $\tau_{\min} = \tau_{\min}(u)$ is a point of minimum of the function

$$v(\tau) = \frac{(1+\tau^\theta(2\theta-1)^{-1}))^2}{2G\left(u(2\theta-1)^{-1}/c\right)^{1/\theta}\tau - 2Hu^{-1/\theta}}.$$

In the case $\theta = 1$ we obtain

$$P(u) = \frac{\sqrt{R(0)}}{\sqrt{2\pi}} c^{-1} \exp\{-\frac{Hc^2}{G^2}\} \exp\{-uc/G\}(1+o(1)).$$

Comparing it with the result of Debicki, we obtain that the generalized Pickand's constant for the process $\eta(t) = \frac{c}{G\sqrt{2}} \int_0^t X_t\, dt$ equals to $\sqrt{R(0)}/(\sqrt{2\pi}Gc)$ if the covariation function $R(t)$ is twice differentiable.

The problem of the limit distribution of the time of ruin is also considered.

References

[1] Piterbarg, V. I. *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, American Mathematical Society, Ser. Translations of Mathematical Monographies, V. 148, Providence.

[2] Piterbarg, V. I. (1996) *Rice method for Gaussian random fields*, Fundamentalnaya i Prikladnaya Matematika, 2, 187–204.

[3] Hüsler, J., Piterbarg, V. (1999) *Extremes of a certain class of Gaussian processes*, Stochastic Processes and their Applications, 83, 257–271.

[4] Kobelkov, S. (2004) *Ruin problem for stationary Gaussian process*, Teorija veroyatnostej i ee primenenija, 49, 171–178.

[5] Debicki, K. (2002) *Ruin probability for Gaussian integrated process*, Stochastic Processes and their Applications, 98, 151–174.

[5] Norros, I., (1994) *A storage model with self-similar input*, Queueing Systems, 16, 387–396.

[6] Narayan, O. (1998) *Exact asymptotic queue length distribution for fractional Brownian traffic*, Advances in Performance Analysis, 1, 39–63.

Estimation of volatility and distribution function for the stochastic processes corresponding to stock price dynamics

Kritski, Oleg *Tomsk Polytechnic University, Russia*,
olegkol@yandex.ru, olegkol@mph.phtd.tpu.edu.ru

stochastic processes, volatility, evaluation of distribution function, risk, Dow Jones Industrial Index, RTS Index.

The statistical properties of the stochastic processes play a key role in the modeling of financial markets. For example, the knowledge of the stochastic nature of the price of a financial asset is crucial aspect for a sensible pricing of an option issued on it. And, of course, if the distribution function or conditional probability density of all orders are known, we can make a full characterization of stochastic process and forecast its stochastic behavior in any point of time.

The most common stochastic model of stock price dynamics assumes that stochastic processes to be used in the investigations are a diffusive process, and logarithms of its increments have Gaussian distribution. Such model, known as Brownian motion, provides the easiest way to estimate the behaviour of observed empirical data. But, in practice, there are the systematic deviations from the model predictions because empirical distributions have more leptokurtic distribution functions than Gaussian one. And we should estimate the probabilistic law corresponding to these data by applying some theoretical method.

Such estimations were made, for example, in [1-2] by using a fractal analysis or by plotting an empirical curve of probability density and doing subsequent comparison with Gaussian distribution respectively.

In this paper it is suggested the other method of evaluation of distribution function. Let we have some sample of daily stock prices that change in time (e.g. we have some time series). Computing a logarithm of daily price differences by formula

$$R_i = \frac{S_{i+1} - S_i}{S_i}, i = 1, 2, \ldots,$$

we can get raw data for an estimation.

After data processing the examination of χ^2– criterion of Pierson's goodness of fit [3-4] is made. The statistical estimation was chosen as follows:

$$\gamma_n = \sum \Big(\frac{(\nu_j - np_j(\Theta))^2}{np_j(\Theta)} \Big)$$

where n – total amount of points of the sample, $s = \log_2 n + 1$ – quantity of classification intervals, Θ –consistent estimate of evaluating parameters (in our case it is the vector consists of empirical estimation of average and dispersion), $p_j(\Theta) = F_{mod}(c_j, \Theta) - F_{mod}(c_{j-1}, \Theta)$—probability to fall into the j'–s classification interval $[c_{j-1}, c_j]$, $F_{mod}(x, \Theta)$–supposed theoretically known distribution function, ν_j– an amount of points of the sample that lie within $[c_{j-1}, c_j]$, $j = \overline{1, s}$.

The numerical computations of the data of companies entering Dow Jones Industrial Index and RTS Index (Russia) have shown that some companies satisfy to the Gaussian distribution (i.e. Intel Corp., RAO UES (EESR, http://www.rao-ees.elektra.ru/en/)) and some companies (in the most cases they are relatively small) don't. Made analysis allows to estimate an empirical volatility function, empirical average function and discover their approximate functional dependencies. It make us possible to forecast future prices of asset and minimize possible risk level.

References

[1] M.M. Dubovikov et al. (2004) *Dimension of the minimal cover and fractal analysis of time series*, Physica A, 339, 591–608.
[2] R.N. Mantegna, H.E. Stanley (2000). Applied statistics. An introduction to Econophysics: correlations and complexity in finance. Cambridge, Cambridge Univ. Press, 2000.
[3] Aivazian S.A., Mhitarian V.S. (2001). Applied statistics. Moscow, Unity–Press, V.1 (in russian).
[4] Aivazian S.A. et al. (1974). Classification of multidimensional data. Moscow, Statistics Press (in russian).

Regional analysis of extreme precipitation events in the Czech republic

Picek, Jan *Technical University in Liberec, Czech Republic*, jan.picek@vslib.cz
Kyselý, Jan *Institute of Atmospheric Physics AS CR, Prague, Czech Republic*
Huth, Radan *Institute of Atmospheric Physics AS CR, Prague, Czech Republic*

extreme precipitation event; regional analysis; L moments; tests for homogeneity of regions:

Extreme high precipitation amounts are among environmental events with the most disastrous consequences for the human society. Estimates of their return periods and design values are of great importance in hydrologic modelling, engineering practice for water resources and reservoirs design and management, planning for weather-related emergencies, etc. The L-moment based method of the regional frequency analysis of maximum annual 1- to 7-day precipitation totals is currently being utilized for the area of the Czech Republic. This contribution deals with the regional analysis.

Daily precipitation amounts over 1961-2000 measured at 78 stations are used as an input dataset. Candidate regions are formed by the cluster analysis of site characteristics (longitude, latitude, elevation, mean annual precipitation, mean ratio of summer half-year (May to October) to winter half-year (November to April) precipitation, and mean annual number of dry days), using the average-linkage clustering and Ward's method. Several tests for the homogeneity of regions are utilized, based on the 10-yr event, L-moment ratios, and the variation of L-moment statistics. In compliance with the results of the tests, the area of the Czech Republic has been divided into four homogeneous regions according to characteristics of extreme precipitation events. The last steps of the regional frequency analysis consists in the selection of the most appropriate distribution, and estimation of parameters and quantiles of the fitted distribution together with their uncertainty.

References

[1] Hosking, J.R.M. (1990) *L-moments: Analysis and estimation of distributions using linear combinations of order statistics*, J. Roy. Stat. Soc. 52B, 105–124.
[2] Hosking, J.R.M., Wallis, J.R., (1997) Regional Frequency Analysis. An Approach Based on L-moments. Cambridge University Press, New York.
[3] Lu, L.H., Stedinger, J.R., (1992) *Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test*, J. Hydrol. 138, 223–245.

Application of block bootstrap optimal choice in the extremal index estimation

Prata Gomes, Dora [1] *Universidade Nova de Lisboa, Portugal*, dsrp@fct.unl.pt
Neves, M. Manuela [2] *Universidade Técnica de Lisboa, Portugal*, manela@isa.utl.pt

extremal index; block bootstrap; estimation; blocking; simulation:

Let $\{X_n\}_{n\geq 1}$ be a stationary dependent sequence of random variables. Under suitable regularity conditions, the maximum, suitably normalized, converges to a non-degenerate distribution function $G_2(x)$, where $G_2(x) = G_1^{\theta}(x)$ and G_1 is the limit of the normalized maximum of the associated i.i.d. sequence, Leadbetter *et al.* (1983). The quantity θ $(0 < \theta \leq 1)$ is termed the *extremal index* and plays a key role in determining the intensity of cluster positions. The estimators for θ proposed in the literature depend crucially on the high level u_n.

Bootstrap methodology, providing answers to many complex problems, can help in obtaining better estimators or even in dealing with nuisance parameters. However Efron's (1979) classic bootstrap methodology is inadequate under dependence and there have been several attempts to extend i.i.d. case to the dependent case. A general approach consists of resampling blocks of data. The accuracy of block bootstrap estimation critically depends on the block size that must be supplied by the user. The orders of magnitude of the optimal block size are known in some inference problems.

Several ways of blocking have been recently proposed by several authors, such as Hall, Horowitz and Jing (1995), Lahiri, Furukawa and Lee (2003), Bühlmann and Künsch (1999) and Politis and White (2003).

Those procedures will be reviewed and applied in the extremal index estimation. A simulation study considering several dependent models will be carried out.

References

[1] Bühlmann, P. and Künsch, H. R. (1999) *Block length selection in the bootstrap for time series*, Computational Statistics and Data Analysis 31, 295-310.
[2] Efron, B.(1979). *Bootstrap methods: another look at the jakknife*, Ann. Statist 7, 1-26.
[3] Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995) *On blocking rules for the bootstrap with dependent data*, Biometrika 50, 561-574.
[4] Lahiri, S. N., Furukawa, K. and Lee, Y.-D. (2003) *A nonparametric plug-in rule for selecting the optimal block length for the block bootstrap methods*, Preprint, Department of Statistics, Iowa State University, Ames, IA.
[5] Leadbetter, M. R., Lindgren, G. e Rootzen, H. (1983) Extremes and related properties of random sequences and series, Springer Verlag, New York.
[6] Politis, D. N. and White, H. (2003) *Automatic block-length selection for the dependent bootstrap*, Preprint, Department of Mathematics, University of California at San Diego, LaJolla, CA.

[1] Support for this author was provided by Centro de Matemática e Aplicações of the Universidade Nova de Lisboa and by the project POCTI/MAT/58876/2004
[2] Support for this author was provided by the project POCTI/MAT/58876/2004

Statistics for asymptotic independence

Ramos, Alexandra *Universidade do Porto - FEP and CMUP, Portugal*, aramos@fep.up.pt
Anthony Ledford *Man Investments, U.K.*

Bivariate extreme values, Asymptotic independence; Ray independence; Inference; Joint tail dependence models:

A fundamental issue in applied multivariate extremes is how to model dependence within joint tail regions. In addressing this we developed a pseudo-polar framework that extends existing classical results to asymptotically independent random variables and additionally obtained parametric joint tail models suitable for applications with good performance for the important case of asymptotic independence.

In this presentation we show how these parametric joint tail models can be exploited to develop tests for asymptotic independence and symmetry. We also introduce some additional terminology for describing the characteristics of tail dependence structures such as convex and concave ray dependence and discuss inference for these. Our tests will be illustrated using bivariate simulated and environmental data.

Power-Gumbel distributions and its applications

Takahashi, Rinya *Kobe University, Japan*, r-taka@maritime.kobe-u.ac.jp

Gumbel distribution; Power Gumbel distribution; Power transformation:

Power-Gumbel distribution is defined and its several properties are discussed. Applications of the power-Gumbel distribution for some data sets are shown.

References

[1] Jeruchim, M. C. (1976) *On the estimation of error probability using generalized extreme-value theory* , IEEE Trans. Infor. Theory, 22, 108–111.
[2] Weinstein, S. B. (1973) *Theory and application of some classical and generalized asymptotic distributions of extreme values* , IEEE Trans. Infor. Theory, 19, 148–154.

Index