The international workshop on the extreme value theory and its applications

Gu Lecture Hall
Shanghai Center for Mathematical Sciences

1. Program

	Monday	Tuesday	Wendesday	Thursday	Friday
Chair	**Bojan Basrak**	**John Einmahl**	**Yi Shen**		
9:00-9:45	Gennady Samorodnitsky	Johan Segers	Bojan Basrak	Richard Davis (I)	Thomas Mikosch (IV)
9:45-10:15	**Tea Break**	**Tea Break**	**Tea Break**	**Tea Break**	**Tea Break**
10:15-11:00	Yi Shen	Zhengjun Zhang	Yanlin Tang	Richard Davis (II)	Thomas Mikosch (V)
11:00-11:45	Jingping Yang	Jianqing Fan	Zhao Chen	Richard Davis (III)	Thomas Mikosch (VI)
12:00-14:00	Lunch	Lunch	Lunch	Lunch	Lunch
Chair	**Johan Segers**	**Yanxi Hou**			
14:00-14:45	John Einmahl	Juanjuan Cai		Thomas Mikosch (I)	Richard Davis (IV)
14:45-15:30	Xuan Leng	Zhonghao Fu		Thomas Mikosch (II)	Richard Davis (V)
15:30-16:00	**Tea Break**	**Tea Break**		**Tea Break**	**Tea Break**
16:00-16:45	Zuoxiang Peng	Chengxiu Ling		Thomas Mikosch (III)	Richard Davis (VI)
18:00-20:00	Conference Dinner				

2. Titles and Abstracts

Risk forecasting in the context of time series

Gennady Samorodnistsky

We propose an approach for forecasting risk contained in future observations in a time series. We take into account both the shape parameter and the extremal index of the data. This significantly improves the quality of risk forecasting over methods that are designed for i.i.d. observations and over the return level approach. We prove functional joint asymptotic normality of the common estimators of the shape parameter and and extremal index estimators, based on which statistical properties of the proposed forecasting procedure can be analyzed. (jointly with Xiaoyang Lu)

On discrete-time self-similar processes with stationary increments

Yi Shen

Abstract: In this talk we study the self-similar processes with stationary increments in a discrete-time setting. Different from the continuous-time case, it is shown that the scaling function of such a process may not take the form of a power function. More precisely, its scaling function can belong to one of three types, among which one type is degenerate, one type has a continuous-time counterpart, while the other type is new and unique for the discrete-time setting. We then focus on this last type of processes, construct two classes of examples, and prove a special spectral representation result for the processes of this type. We also derive basic properties of discrete-time self-similar processes with stationary increments of different types. This is a joint work with Zhenyuan Zhang.

Asymptotic Additivity of Value-at-Risk under Tail Dependence

Jingping Yang

We consider the asymptotic additivity of Value-at-Risk for multiple risks whose copula's stable tail dependence function (STDF) exists, where the asymptotic additivity includes asymptotic subadditivity and asymptotic superadditivity. For the purpose, a marginal region is defined by the marginal distributions of the multiple risks, and a stochastic order named tail concave order is presented for comparing the tail parts of the individual risks. We prove that the asymptotic subadditivity of VaR holds when the individual risks are smaller than regularly varying random variables with index -1 under the tail concave order. We also provide the sufficient conditions for VaR being asymptotically superadditive. We give the asymptotic diversification ratios for regularly varying and log regularly varying margins with specified copula structures. Some numerical analyses are provided for highlighting our results. We present a new method for discussing the asymptotic additivity of VaR for multiple risks by focusing on its marginal distributions and its copula separately. It is a joint work with Wenhao Zhu and Lujun Li.

Extreme value statistics in semi-supervised settings

John Einmahl

Heavy tailed phenomena are naturally analyzed by extreme value statistics. A crucial step in such an analysis is the estimation of the extreme value index, which describes the tail heaviness of the underlying probability distribution. We consider the situation where we have next to the n observations of interest another n + m observations of one or more other variables, like, e.g., financial losses due to earthquakes and the related amounts of energy released, for a longer period than that of the losses. This is called the semi-supervised model with n labeled data and m unlabeled data. For such data , we estimate the extreme value index of the variable of interest with an adapted version of the Hill estimator. For this adaptation the tail dependence between the variable of interest and the other variable(s) plays an important role. We establish the asymptotic normality of this new estimator. It shows greatly improved behavior relative to the Hill estimator, in particular the asymptotic variance is substantially reduced, whereas we can keep the asymptotic bias the same. A simulation study confirms the substantially improved performance of our adapted estimator. We also present an application to the aforementioned earthquake losses. Finally, we consider in a more general setting the case where the extreme value index is not necessarily positive. This is joint work with Hanan Ahmed and partially with Chen Zhou.

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

Xuan Leng

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

Asymptotics on extremes of bivariate Gaussian triangular arrays

Zuoxiang Peng

In this talk, we summarize our recent work on asymptotics of extremes of bivariate Gaussian triangular arrays. The seminal work of Hüsler and Reiss [3] showed that the limit distribution of maxima of bivariate Gaussian triangular arrays is a max-stable distribution provided that the Hüsler-Reiss condition holds. Our attention is on the asymptotic expansions and uniform convergence rates of the distributions of extremes. Hashorva et al. [1] established the higher-order expansions of the distribution of maxima under the refined Hüsler-Reiss condition, and Liao and Peng [5] derived its uniform convergence rates. Liao and Peng [6] also showed the higher-order expansions of the joint distribution of maxima and minima. For the asymptotics of maxima of independent and non-identically distributed bivariate Gaussian triangular arrays, Liao and Peng [7] established the higher-order expansions of the distribution of linear normalized maxima as the coefficient satisfies some regular conditions, and the asymptotics of maxima and minima were derived by Lu and Peng [9]. The

limit distribution in copula form was studied by Liao et al. [4] and distributional higher-order expansions were considered by Wang et al. [10]. For more results related to H-R models, see Weng and Liao [11] for extremes under power normalization, Zhou and Peng [12] for the powered-extremes of bivariate Gaussian triangular arrays, Hashorva et al. [2] for weakly dependent case with extremal index and Liao et al. [8] for the settings of bivariate elliptical triangular arrays.

One- versus multi-component regular variation

Johan Segers

One-component regular variation refers to the weak convergence of a properly rescaled random vector conditionally on the event that a single given variable exceeds a high threshold. Although the weak limit depends on the variable concerned by the conditioning event, the various limits are connected through an identity that resembles the time-change formula for regularly varying stationary time series. The formula is most easily understood through a single multi-component regular variation property concerning some (but not necessarily all) variables simultaneously.

The theory is illustrated for max-linear models, in particular recursive max-linear models on acyclic graphs, and for Markov trees. In the latter case, the one-component limiting distributions take the form of a collection of coupled multiplicative random walks generated by independent increments indexed on the edges of the tree. Changing the conditioning variable then amounts to changing the directions of certain edges and transforming their increment distributions in a specific way.

Reference:
Segers, J. (2019). "One- versus multi-component regular variation and extremes of Markov trees", https://arxiv.org/abs/1902.02226.

Simulated Distribution Based Learning for Non-regular and Regular Statistical Inferences

Zhengjun Zhang

Statistical research involves drawing inference about unknown quantities (e.g., parameters) in the presence of randomness in which distribution assumptions of random variables (e.g., error terms in regression analysis) play a central role. However, a fundamental issue of preserving the distribution assumptions has been more or less ignored by many inference methods and applications. As a result, the further inference of studied problems and related decisions based on the estimated parameter values may be inferior. This paper proposes a continuous distribution preserving estimation approach for various kinds of non-regular and regular statistical studies. The paper establishes a fundamental theorem which guarantees the transformed order statistics (to a given marginal) from the assumed distribution of a random variable (or an error term) to be arbitrarily close to the order statistics of a simulated sequence of the same marginal distribution. Different from the Kolmogorov-Smirnov test which is based on absolute errors between the empirical distribution and the assumed distribution, the statistics proposed in the paper are based on relative errors of the transformed order statistics to the simulated ones. Upon using the constructed statistic (or the pivotal quantity in estimation) as a measure of the relative distance between two ordered samples, we estimate parameters such that the distance is minimized. Unlike many existing methods, e.g., maximum likelihood estimation, which rely on some regularity conditions and/or the explicit form of probability density function, the new method only assumes a mild condition that the cumulative distribution function can be approximated to a satisfied precision. The paper illustrates simulation examples to show its superior performance. Under the linear regression settings, the proposed estimation performs exceptionally well regarding preserving the error terms (i.e., the residuals) to be normally distributed which is a fundamental assumption in the linear regression theory and applications.

Maximum Spurious Correlation in High Dimension: Guarding against false discoveries

Jianqing Fan

Many data-mining and statistical machine learning algorithms have been developed to select a subset of covariates to associate with a response variable. Spurious discoveries can easily arise in high-dimensional data analysis due to enormous possibilities of such selections. How can we know statistically our discoveries better than those by chance? In this paper, we define a measure of goodness of spurious fit, which shows how good a response variable can be fitted by an optimally selected subset of covariates under the null model, and propose a simple and effective LAMM algorithm to compute it. It coincides with the maximum spurious correlation for linear models and can be regarded as a generalized maximum spurious correlation. We derive the asymptotic distribution of such goodness of spurious fit for generalized linear models and L_1-regression. Such an asymptotic distribution depends on the sample size, ambient dimension, the number of variables used in the fit, and the covariance information. It can be consistently estimated by multiplier bootstrapping and used as a benchmark to guard against spurious discoveries. It can also be applied to model selection, which considers only candidate models with goodness of fits better than those by spurious fits. The theory and method are convincingly illustrated by simulated examples and an application to the binary outcomes from German Neuroblastoma Trials. (Joint work with Wenxinz Zhou)

A nonparametric estimator of the extremal index

Juanjuan Cai

Clustering of extremes usually has a large societal impact. The extremal index, a number in the unit interval, is a key parameter in modelling the clustering of extremes. We build a connection between the extremal index and the stable tail dependence function, which enables us to compute the value of extremal indices for some time series models. We also construct a nonparametric estimator of the extremal index and an estimation procedure to verify D(d)(un) condition, a local dependence condition often assumed when studying extremal index. We prove that the estimators are asymptotically normal. The simulation study compares our estimator to two existing methods, which shows that our method has good finite sample properties. We also apply our method to estimate the expected durations of heatwaves in the Netherlands and in Greece.

Estimation and Testing Distributional Changes via Characteristic Function

Zhonghao Fu

We estimate and test multiple structural breaks in joint distribution with unknown break dates via a characteristic function approach. By minimizing the sum of squared generalized residuals, we can consistently estimate the break dates and obtain the convergence rate for the estimated break fractions. We propose a sup-F type test to check the existence of structural breaks, and suggest an information criterion and a sequential testing procedure to determine the number of break dates. We further construct a class of derivative tests to gauge possible sources of structural breaks. Simulation studies show that our method performs well in determining the appropriate number of changes and consistently estimating the unknown break dates. Also, the proposed tests have reasonable size and

excellent power. In an application to exchange rate returns, we show that our tests are able to detect structural breaks in higher order moments, which are neglected by tests for structural changes in mean.

Extremes of stationary random fields on a lattice

Chengxiu Ling

Extremal behavior of stationary Gaussian sequences/random fields is widely investigated since it models common cluster phenomena and brings a bridge between discrete and continuous extremes. We establish extensively limit theorems of stationary random fields under certain mixing and dependence conditions, which are further illustrated by typical examples of order statistics of Gaussian random fields and skew-Gaussian random fields. The positivity of the cluster index involved and its link with the expected cluster size are discussed.

On subcritical branching processes with regularly varying immigration

Bojan Basrak

Provided that a stationary multivariate regularly varying time series satisfies certain weak dependence conditions, its extremal behavior can be elegantly characterized using the theory of point processes and the notion of the tail process. However, except in a few simple cases, establishing such conditions and determining exact distribution of the tail process in the multivariate setting remains a technically challenging task. We discuss a class of models where one can follow a somewhat different route to asymptotic analysis. The presentation is motivated by the study of conditional least squares estimator of the mean number of progeny in the branching process with heavy tailed immigration. We also exhibit the rate of convergence and precise asymptotic distribution of the estimator.

The talk is based on the joint work with M.Barczy, P.Kevei, G.Pap and H. Planinic.

Conditional marginal test in high dimensional quantile regression

Yanlin Tang

Analysis of tail quantiles of the response distribution is sometimes more important than the mean in biomarker studies. Inference in quantile regression is complicated when there exist a large number of candidate markers together with some pre-specified controlled covariates. In this paper, we develop a new and simple testing procedure to detect the effects of biomarkers in high dimensional quantile regression in the presence of protected covariates. The test is based on the maximum-score-type statistic obtained from conditional marginal regression. We establish the asymptotic properties of the proposed test statistic under both null and alternative hypotheses, and further propose an alternative multiplier bootstrap method with theoretical justifications. We demonstrate through numerical studies that the proposed method provides adequate controls of the family-wise error rate with competitive power, and it can also be used as a stopping rule in the forward regression. The proposed method is applied to a motivating genome-wide association study to detect single nucleotide polymorphisms associated with low glomerular filtration rates in Type 1 diabetes patients.

Ultrahigh Dimensional Precision Matrix Estimation via Refitted Cross Validation

Zhao Chen

This paper develops a new estimation procedure for ultrahigh dimensional sparse precision matrix, the inverse of covariance matrix. Regularization methods have been proposed for sparse precision matrix estimation, but they may not perform well with ultrahigh dimensional data due to spurious correlation. We propose a refitted cross validation (RCV) method for sparse precision matrix estimation based on its Cholesky decomposition. The proposed RCV procedure can be easily implemented with existing software for ultrahigh dimensional linear regression. We establish the consistency of the proposed RCV estimate and show that the rate of convergence of the RCV estimate without assuming banded structure is the same as those assuming the banded structure in Bickel and Levina (2008b). Monte Carlo studies were conducted to access the finite sample performance of the RCV estimate. Our numerical comparison shows that the RCV estimate can outperform existing ones in various scenarios. We further apply the RCV estimate for an empirical analysis of asset allocation.