Fens, Forests and Formulas

John Einmahl Workshop

September 6–7, 2023

ABSTRACT BOOK

Table of Contents

Day 1 — Morning Presentations

September 6, 10:45 – 12:30

Opening 10:45-11:00

Chair: Yi He

Opening Talk
Bertrand Melenberg
Head of Department of Econometrics and Operations Research
Tilburg University

Session 1: Our Journey With John and the Joint Works 11:00-12:30

Chair: Yi He

Almost 40 years
Laurens de Haan
Erasmus University Rotterdam
I am not going to present a piece of new research. Instead, I want to retrace some steps John Einmahl and I made together in the development of extreme value theory. It is not a review of John's work because he made his mark in other subjects as well. Our joint research was generally in cooperation with PhD students.

On our work with John
Estate Khmaladze
Victoria University of Wellington
I will describe atmosphere and mathematical content of two projects which we worked on in late 1990 and mid-2010. One was the two-sample problem in d-dimensional space and another was on local empirical processes at the boundaries of convex bodies. It is now strange to realise they were about 10 years apart.

John's co-authorship and other contributions; from high-quantiles to heteroskedasticity
Ana Ferreira
Universidade de Lisboa
John's intrinsic enthusiasm on Mathematical Statistics has undoubtedly been reflected with important contributions in Extreme Value Theory. For instance, John's earlier developments on empirical processes for understanding high-quantile and tail probability estimation, and most recently its extensions considering a scedasis function to achieve most recent advancements under non-stationary conditions. We shall overview my experience including main results obtained so far.

Day 1 — Afternoon Presentations

September 6, 14:00 – 17:00

Session 2: Advances in Extreme Value Statistics 14:00-15:30

Chair: Cees de Valk

Heteroscedastic extremes

Chen Zhou
Erasmus University Rotterdam

One of John Einmahl's seminal contributions to extreme value statistics in the last 10 years is to relax the condition that observations are identically distributed. This talk briefly reviews the history of heteroscedastic extremes and presents recent developments in multivariate heteroscedastic extremes. The first topic is to estimate the tail dependence function with the presence of common marginal trends. The second topic is to estimate a time varying tail dependence function, allowing for both marginal and joint heteroscedasticity.

Estimation of tail parameters with missing largest observations

Jan Beirlant
Katholieke Universiteit Leuven

We consider the case where an unknown number m of the highest data is missing assuming an underlying Pareto-type distribution. We provide solutions for estimating the extreme value index, the number of missing data and extreme quantiles. We derive an asymptotic result of the parameter estimators and an adaptive selection method for the number of top data used in the estimation is proposed for the case where all missing data are beyond the observed data. An estimator of the number of missing extremes spread over the largest observed data is also proposed. To this purpose, we use a likelihood solution based on exponential representations of spacings between the largest observations. We also establish an effective and fast optimization procedure using regularization, and finally comment on simulation experiments. We illustrate the methodology in a practical case from the diamond mining industry, where large-carat diamonds are expected to be missing from the dataset.
This is a joint work with Martin Bladt.

Conditional tail moment and reinsurance premium estimation under random right censoring

Armelle Guillou
Université de Strasbourg

We propose an estimator of the conditional tail moment (CTM) when the data are subject to random censorship. The variable of main interest and the censoring variable both follow a Pareto-type distribution. We establish the asymptotic properties of our estimator and discuss bias-reduction. Then, the CTM is used to estimate, in case of censorship, the premium principle for excess-of-loss reinsurance. The finite sample properties of the proposed estimators are investigated with a simulation study and we illustrate their practical applicability on a dataset of motor third party liability insurance.
This is a joint work with Yuri Goegebeur and Jing Qin.

Session 3: Asymptotic Statistics for Large Data Set 16:00-17:00

Chair: Umut Can

Asymptotic analysis of PCA for extremes

Holger Drees
Universität Hamburg

Drees and Sabourin (2021) examined the PCA projection of the angular part of a multivariate regular varying random vector. In particular, they derived uniform bounds for the risk of the PCA approximation, that is the expected squared norm of the difference between the angular part and its PCA approximation. In this talk we analyze the asymptotic behavior of the PCA projection and the resulting excess risk in more detail.

Semiparametrically point-optimal rank-based panel unit root tests

Bas Werker
Tilburg University

We study optimality of unit-root tests in a panel setting with both large time-series and cross-sectional dimension. While, with finite cross-sectional dimension, the limit experiments are Locally Asymptotically Brownian Functional (LABF), we find the classical Locally Asymptotically Normal (LAN) structure with large cross-sectional dimension. This leads to an upper-bound for the power of tests using sequential asymptotics. We propose a hybrid rank-based test, based on cross-sectional ranks per time unit, that is point optimal.

Day 2 — Morning Presentations

September 7, 9:30 – 12:00

Session 4: Common Sense and Stylized Facts 09:30-10:30

Chair: Pavel Cizek

On common sense in statistics with applications to extreme value theory

Jan Magnus
Vrije Universiteit Amsterdam

This presentation is about common sense in statistics.

Arbitrage and statistics

Casper de Vries
Erasmus University Rotterdam

The tail risk of equity returns is heavy tail distributed. Arbitrage by portfolio investors who are concerned with tail risk suggests that tail indices of different stocks should be equal as stocks with heavier tails would be excluded from their portfolios. If tail indices are equal but scales do differ, however, investors do benefit from composing diversified portfolios. To investigate these hypotheses, we devise tests for tail and scale (in-)equality based on the R(1,1) measure but correct for bias that stems form first and second order parameters. We investigate the statistical properties of these tests and present an application.
This is a joint work with Laurens de Haan and Chen Zhou.

Session 5: Empirical Process Theory — 11:00-12:00

Chair: Martien van Zuijlen

Bivariate quantile transform empirical process.

Philippe Berthet
Institut de Mathématiques de Toulouse

Given two samples of distributions P and Q on $\mathbb{R}^2$ we construct an empirical bivariate quantile transform that is easy to compute and converges to a transport map from P to Q. We deduce CLT's for new non parametric statistics of geometrical nature.

In [1] we define a universal generator G through what we call the Kendall geometry, made of two families of time-mass curves — the z_1 and z_2 curves. The generator $G(P)$ of P is such that if $z = (z_1, z_2)$ is uniform on the unit square then the intersection of the z_1 and z_2 curves of $G(P)$ is a random variable with distribution P. This point $G(P)(z)$ is the bivariate quantile of P with bivariate rank z. The coupling $[G(P)(z), G(Q)(z)]$ has the property to optimally transport the conditional laws of P and Q on their corresponding z1 curves with respect to a large class of costs, including Wasserstein ones. It moreover coincides with the global optimal transport if P, Q share the same copula. The quantile transform $\tau_{P,Q} = G(Q) \circ (G(P)^{-1})$ is then a transport map having a closed form expression and a probabilistic meaning, which allows to apply empirical processes theory, but not directly, as seen in [2].

The empirical counterparts built from the natural estimators $G(P, n)$ and $G(Q, m)$ can be managed in practice with n, m of order several millions. The specific new difficulty is that $\tau_{n,m} = G(Q, m) \circ (G(P, n)^{-1})$ involves non independent global and local empirical processes. We establish non standard weak convergences thanks to a non asymptotic joint Brownian coupling of all these empirical processes. The local processes along z_1-curves induce a Brownian white noise component, however the limiting processes and covariances are explicit. The uniform approximation of the empirical generator geometry implies explicit, standard CLT's for statistical quantities such as transport costs and a new, bivariate Kendall tau. We also derive the weak convergence of new contours, clusterings, mode localizations, trimmed supports and local depth fields that we derive from $G(P, n)$, and comparison or goodness of fit tests that we derive from $\tau_{n,m}$.

[1] Berthet P. and Fort J. C. (2022) Kendall quantile ordering on $\mathbb{R}^2$, probabilistic transport maps and their empirical counterpart.

[2] Berthet P. and Fort J. C. (2023) Convergence of bivariate quantile, rank and transport empirical processes.

The empirical copula process in non-rectangular sets

Johan Segers
Université catholique de Louvain

The copula of a random vector with unknown marginals can be estimated non-parametrically, based on iid observations, by the empirical copula, akin to the empirical distribution. However, the asymptotic analysis of the empirical copula is made considerably more involved than that of the empirical distribution by the use of pseudo-observations, involving the marginal empirical distribution functions. In particular, it is still unknown whether the empirical copula evaluated at a non-rectangular set is asymptotically normally distributed. In this work, sufficient conditions under which this is the case are identified. The analysis exploits tools from empirical processes indexed by random functions and from the concept of differentiation of sets in measure.

This is a joint work with Axel Bücher, Michaël Lalancette, and Stanislav Volgushev

Day 2 — Afternoon Presentations

September 7, 14:30 – 17:00

Session 6: Copula-based Dependence 14:30-15:30

Chair: Xuan Leng

X-Vines: modelling extremal dependence by vine constructions for exponent measures

Anna Kiriliouk
Université de Namur

Regular vines are a way to organize the variables in a random vector along a sequence of trees. The first tree corresponds to a Markov random field whereas the other trees capture higher-order effects. Pair copula constructions based on vines have become greatly popular in dependence modelling because they allow arbitrary bivariate copulas to be combined into flexible high-dimensional distributions. Both for simulation and inference, computations are typically performed by recursive algorithms. In this project, we explore the opportunities and limitations of vine decompositions for the density of the exponent measure of a multivariate max-stable distribution. The homogeneity property that such densities satisfy leads to some simplifications in comparison to the copula case. The decomposition sheds new light on existing parametric models and facilitates the construction of new ones.
This is a joint work with Jeongjin Lee and Johan Segers.

Copula-based divergence measures for dependence between random vectors

Irène Gijbels
Katholieke Universiteit Leuven

In this talk we discuss copula-based dependence quantification between multiple groups of random variables of possibly different sizes via the family of Phi-divergences. An axiomatic framework for this purpose is provided, and we illustrate the divergence measures by means of examples. For statistical inference we focus on the absolutely continuous setting assuming copula densities exist. We consider parametric and semi-parametric frameworks, discuss estimation procedures, and establish asymptotic properties of the proposed estimators. Simulations indicate finite-sample performances, and practical use is discussed.
This talk is based on a joint work with Steven De Keyser.

Session 7: Advances in Modeling and Estimation 16:00-17:00

Chair: Otilia Boldea

Extending Extreme Value Inference to General Heterogeneous Data

Yi He
University of Amsterdam

We fully upgrade the extreme value statistics to independent data with possibly very different distributions. This is a joint work with John.

Instrumental variable estimation of dynamic treatment effects on a duration outcome

Ingrid Van Keilegom
Katholieke Universiteit Leuven

This paper considers identification and estimation of the causal effect of the time Z until a subject is treated on a duration T. The time-to-treatment is not randomly assigned, T is randomly right censored by a random variable C, and the time-to-treatment Z is right censored by $\min(T, C)$. The endogeneity issue is treated using an instrumental variable explaining Z and independent of the error term of the model. We study identification in a fully nonparametric framework. We show that our specification generates an integral equation, of which the regression function of interest is a solution. We provide identification conditions that rely on this identification equation. We assume that the regression function follows a parametric model for estimation purposes. We propose an estimation procedure and give conditions under which the estimator is asymptotically normal. The estimators exhibit good finite sample properties in simulations. Our methodology is applied to evaluate the effect of the timing of a therapy for burnout.