Extreme Value Statistics for General Heterogeneous Data through the Average Tail

Speaker: Yi He

Eastern Institute of Technology, Ningbo

Time&Venue:

SCMS Room 102, Apr. 10 (Friday), 10:30-11:30

Abstract:

In the general setting of independent data with possibly very different distributions extreme value estimators inevitably target the tail of the average distribution function. We consider all possible cases, that is, the extreme value index of the average distribution can be negative, zero, or positive, and we present novel asymptotic theory for the moment estimator. Our results require a different and much more challenging proof than those for the power-law case and are based on a uniform central limit theorem for the underlying weighted tail empirical process.We find that, due to the heterogeneity of the data, the asymptotic variance of the moment estimator can be much smaller than that in the i.i.d. case. We also unravel the improved performance of high quantile and endpoint estimators in this setup. In case of a heavy tail, we ameliorate the Hill estimator by taking an optimal combination of the Hill and the moment estimator. Simulations show the good finite-sample behavior of our limit results. Finally we present applications to the maximum lifespan of monozygotic twins, the ultimate 200m running world records, and to the tail heaviness of energies of earthquakes around the globe.

Biography:

Professor Yi He's research spans extreme value theory, high-dimensional statistics, and financial econometrics. In extreme value theory, he focuses on developing novel frameworks for heterogeneous large-scale data, proposing new statistical laws and their theoretical foundations, and advancing super efficient prediction of natural disasters and climate extreme events. In high-dimensional statistics, he extends random matrix theory to study non-sparse models and derives optimal statistical inference and forecasting methods through rigorous theoretical analysis. In financial econometrics, he concentrates on building data-driven statistical inference frameworks for financial time series, and providing methodological support and theoretical guarantee for real-time risk management.