

Title and abstracts:

Plenary Lecturer 1: Anna Kiriliouk – University of Namur (Belgium)

Title: Climate extreme event attribution using causal counterfactual theory

Abstract: The aim of extreme event attribution is to assess by how much human activities have increased the risk of occurrence of an extreme climate event (e.g., a heatwave, a flood,...). Such an assessment is obtained by calculating the Fraction of Attributable Risk (FAR), i.e., the relative ratio of two probabilities of occurrence of an extreme event, one calculated in the factual world, the other in the counterfactual world. The latter is defined as a world without anthropogenic forcing, i.e., a world that “might have been” if no humans inhabited the earth. Because randomized control trials cannot be performed on the Earth’s climate, causal statements must be based on output from General Circulation Models (GCMs), which simulate climate variables on a grid of spatial locations. These models allow to conduct sensitivity experiments in which the suspected causal link or driver is switched on or off, thereby assessing the role of this causal mechanism or driver on the probability of occurrence of the extreme event under study. Current attribution studies are commonly based on univariate methods, ignoring spatial tail dependence. We discuss two different estimators of the FAR, one based on the multivariate generalized Pareto distribution (the asymptotically justified model for multivariate threshold exceedances), the other based on a completely positive decomposition of the so-called tail pairwise dependence matrix. Applied to heavy winter precipitation in central Europe, the proposed estimators lead to a higher FAR with narrower confidence intervals when compared to a univariate attribution analysis.

Plenary Lecturer 2: Mirjam Moerbeek - Utrecht University (Netherlands)

Title: Optimal design of cluster randomized trials

Abstract: With cluster randomized trials complete groups such as schools, households or family practices are randomized to treatment conditions. Cluster randomized trials are less efficient than individual randomized trials, hence it is important they are designed in an optimal way.

This presentation gives a short introduction to the cluster randomized trial and the statistical model to analyze data from such a trial. It then shows how to calculate the optimal number of clusters and cluster size such that highest power for the test on treatment effect is achieved. Software for calculating the optimal design is demonstrated. After that, an extension is made to multiperiod designs where clusters and the subjects therein are measured multiple times in a longitudinal trial. Examples are the cross-over trial and the stepped-wedge design. The optimal design of such trials under attrition of subjects and/or clusters over time is discussed and it shown how to repair for the loss of efficiency due to attrition.

Plenary Lecturer 3: Piet Daas - Statistics Netherlands (Netherlands)

Title: Big Data and Official Statistics: Challenges and Applications at Statistics Netherlands

Abstract: The use and application of Big Data in official statistics has made considerable progress at Statistics Netherlands. The major contributors are the increased attention for Big Data in the methodological research program, in the creation of experimental statistics and in its use for regular statistics production. To stimulate this the Center for Big Data Statistics was setup in 2016. The most important research challenges identified are: 1. Concept: What (derived) concept is measured in Big Data? 2. Population: What part of the target population is included in Big Data? 3. Methods: What new methods (or new ways of thinking) are needed? 4. Infra: What infrastructural requirements are needed? The infrastructural (IT) challenge is ignored here. The fact that there is a steady increase in the application of Big Data at the office indicates the need (and progress made) in the study of the research challenges identified. The statistics that make use of Big Data and are either in production or for which an implementation process has started at Statistics Netherlands are: 1. Using scanner data and scraped prices for the Consumer Price Index 2. Using road sensor data for Traffic Intensity statistics 3. Using website texts for Online Platform Economy statistics 4. Using social media for the Social Tension indicator 5. Using texts of online job advertisement for Vacancy statistics 6. Using solar panel output and weather data for Solar Energy production The presentation will discuss the research challenges and how this has affected the use of Big Data for official statistics at the office.

Plenary Lecturer 4: Holger Dette - Ruhr-Universitaet Bochum (Germany)

Title: Equivalence of regression curves

Abstract: This paper investigates the problem whether the difference between two parametric models m_1, m_2 describing the relation between a response variable and several covariates in two different groups is practically irrelevant, such that inference can be performed on the basis of the pooled sample. Statistical methodology is developed to test the hypotheses $H_0: d(m_1, m_2) \geq \epsilon$ versus $H_1: d(m_1, m_2) < \epsilon$ to demonstrate equivalence between the two regression curves m_1, m_2 for a pre-specified threshold ϵ , where $d(\cdot)$ denotes a distance measuring the distance. Our approach is based on the asymptotic properties of a suitable estimator $d(\widehat{m}_1, \widehat{m}_2)$ of this distance. In order to improve the approximation of the nominal level for small sample sizes a bootstrap test is developed, which addresses the specific form of the interval hypotheses. In particular, data has to be generated under the null hypothesis, which implicitly defines a manifold for the parameter vector. The results are illustrated by means of a simulation study and a data example. It is demonstrated that the new methods substantially improve currently available approaches with respect to power and approximation of the nominal level. The results have been applied in cooperation with the EMA and FDA for comparing dissolution profiles.