Van Dantzig seminar April 12, 2024

Programme

10.00-10.45 (NU 4B-47):	Sara van de Geer
10.45-11.00 Coffee break	
11.00-11.45 (NU 4B47):	Michael Sørensen
11.45-12.45 Lunch (not included)	
12:45-13.30 (NU 4B47):	Ernst Wit

At 15.45 Mathisca de Gunst will give her valedictory address in the VU main building. Title: *Over ivoren torens en een glazen plafond*. There is a reception afterwards.

Titles and abstracts of talks are given below.

Rates of convergence for tensor denoising

Sara van de Geer, ETH Zürich

We consider least squares estimation of a regression function on $[0,1]^d$ with a constraint on its derivatives. Let μ be Lebesgue measure on $[0,1]^d$ and $\mathcal{F} \subset L_2(\mu)$ be a class of functions f with $\|Df\|_1 := \int |Df| d\mu \leq 1$, where $D := \prod_{j=1}^d \partial^m/(\partial x_j)^m$. One calls $\|Df\|_1$ the m-th order Vitali total variation. We generalize the entropy result of Blei et al. [2007] for the case $m = 1$ to $m \in \mathbb{N}$, and apply this generalization to the tensor denoising regression problem.

References

R. Blei, F. Gao, and W. Li. Metric entropy of high dimensional distributions. *Proceedings of the American Mathematical Society*, 135(12):4009–4018, 2007.

From Toegepaste Statistiek to Causality and back

Ernst C. Wit, Università della Svizzera italiana

Mathisca de Gunst taught me statistics 30 years ago in her class Toegepaste

Statistiek. In our project, Marcus Westra and I were exceptionally pleased to prove that shoe size was one of the drivers of academic performance among bachelor students in mathematics. Reproducibility nor causality were particularly bothering us, but Mathisca put several big question marks in the report. We learned the hard way that statistics is more than significance.

Causality is the holy grail of science, but humankind has struggled to operationalize it for millennia. In recent decades, a number of more successful ways of dealing with causality in practice, such as propensity score matching, the PC algorithm, and invariant causal prediction, have been introduced. However, approaches that use a graphical model formulation tend to struggle with computational complexity, whenever the system gets large. Finding the causal structure typically becomes a combinatorial-hard problem.

In this presentation, I will go back to our linear regression setting from the Toegepaste Statistiek class. By replacing combinatorial optimization with continuous causal regularization, we are not only able to make our method applicable to large systems, but, more importantly, our approach allows a precise formulation of the trade-off between in-sample and out-of-sample prediction error.

Models of time series of angular data: diffusion processes on the torus

Michael Sørensen, University of Copenhagen

Two types of diffusion processes on the multivariate torus with related statistical methodology are presented. The aim is to model time series of angular data. The diffusion processes are ergodic and time-reversible and can be constructed for any pre-specified stationary distribution on the torus. Applications to the evolution of proteins and to ants' movement are briefly presented. First we consider Langevin diffusions of the torus and compare four approximations to the likelihood function. We also present a class of diffusion models with explicit transition probability densities, which enables exact likelihood inference. Asymptotic likelihood theory is presented, and it is shown how exact diffusion bridge simulation can easily be done for these models. A class of circular jump processes with similar properties is proposed too.