

Book of Abstracts

**Dansk Selskab
for Teoretisk
Statistik**

Danish Society for Theoretical Statistics

1971 – 2021

**DSTS 50 years anniversary meeting
17th – 19th of November 2021**

Venue: Scandic CPH Strandpark, Amager Strandvej 401, 2770 Kastrup

Compiled: November 12, 2021

Conference Sponsors

Gold Sponsor



Silver Sponsors

Vestas®



Bronze Sponsor



Supported by

CARLSBERGFONDET

Conference program

Wednesday (17 November)

- 12:00 - 13:00 **Lunch** SPONSORED BY **LUNDBECK**
 13:00 - 13:15 WELCOMING SESSION
 13:15 - 14:05 **Niels Keiding** (KU)
 DSTS' HISTORIE
 14:05 - 14:20 *Break*
 14:20 - 15:10 **Asger Hobolth** (AU)
 MIXED MEMBERSHIP MODELS FOR MUTATIONAL SIGNATURES IN CANCER GENOMICS
 15:15 - 16:05 **Ann-Sophie Buchardt** (KU), **Sneha Das** (DTU) og **Nikolaj Thams** (KU)
 YOUNG STATISTICIANS DENMARK PRESENTS: YOUNG STATISTICAL RESEARCH IN DENMARK
 16:05 - 16:20 *Break*
 16:20 - 17:10 **Claudia Strauch** (AU)
 LEARNING TO REFLECT - DATA-DRIVEN STRATEGIES FOR STOCHASTIC CONTROL
 17:15 - 18:05 **Philip Hougaard** (Lundbeck/SDU)
 ANALYSIS OF RECURRENT EVENTS DATA
 19:00 **Dinner** SPONSORED BY **VESTAS**

Thursday (18 November)

- 09:00 - 09:10 OPENING SESSION
 09:10 - 10:00 **Jonas Peters** (KU)
 STATISTICAL TESTING UNDER DISTRIBUTIONAL SHIFTS AND ITS APPLICATION TO CAUSALITY
 10:00 - 10:15 *Break*
 10:15 - 11:05 **Birgitte Rønn** (Zealand Pharma), **Mette Krog Josiassen** (Lundbeck) og
Helle Lynggaard (Novo Nordisk)
 ESTIMANDS: THE ROSETTA STONE OF ESTIMATION OR OPENING PANDORA'S BOX?
 11:10 - 12:00 **Therese Graversen** (ITU)
 A STATISTICAL SLEUTH IN THE UK CRIMINAL JUSTICE SYSTEM
 12:00 - 13:00 **Lunch** SPONSORED BY **SAS INSTITUTE** *Lunch-presentation by Lars Kirdan (SAS Institute)*
 STATISTIKKENS SKABELSESBERETNING OG DENS MULIGE ENDELIGT (12:15 - 12:50)
 13:00 - 13:50 **Per Kragh Andersen** (KU)
 THE JOY OF PSEUDO-VALUES
 13:55 - 14:45 **Line Clemmensen** (DTU)
 ANALYSIS OF SPECTRAL DATA
 14:45 - 15:00 *Break*
 15:00 - 15:50 **Jacob von Bornemann Hjelmberg** (SDU)
 BIostatISTICS IN TWIN STUDIES
 15:55 - 16:45 **Sven Jesper Knudsen** (Vestas)
 THE THRILLING JOURNEY AND BENEFITS OF ANALYTICS AT VESTAS
 16:45 - 17:00 *Break*
 17:00 - 17:50 **Claus Thorn Ekstrøm** (KU)
 PREDICT(DSTS, N.AHEAD=50)
 19:00 **Anniversary dinner** SPONSORED BY **NOVO NORDISK**

Friday (19 November)

- 09:00 - 09:50 **Torben Martinussen** (KU)
 ASSUMPTION LEAN INFERENCE FOR SURVIVAL DATA
 09:55 - 10:45 **Jesper Møller** (AAU)
 25 YEARS MORE WITH POINT PATTERNS IN EUCLIDEAN SPACE AND BEYOND
 10:45 - 11:00 *Break*
 11:00 - 11:50 **Kajsa Kvist**, **Randi Grøn** og **Henrik Ravn** (Novo Nordisk)
 JOINT INITIATIVE FOR CAUSAL INFERENCE – FROM NOVO NORDISK PERSPECTIVE
 11:55 - 12:45 **Steffen Lauritzen** (KU)
 LOCALLY ASSOCIATED GRAPHICAL MODELS AND MIXED CONVEX EXPONENTIAL FAMILIES
 12:45 - 13:00 CLOSING SESSION
 13:00 - 14:00 **Lunch** SUPPORTED BY **CARLSBERGFONDET**

Contents

Wednesday	5
DSTS' Historie	5
<i>Niels Keiding</i>	
Mixed Membership Models for Mutational Signatures in Cancer Genomics	6
<i>Asger Hobolth</i>	
Young Statisticians Denmark presents: Young Statistical Research in Denmark	6
<i>Ann-Sophie Buchardt, Sneha Das and Nikolaj Thams</i>	
Learning to Reflect - Data-Driven Strategies for Stochastic Control . .	6
<i>Claudia Strauch</i>	
Analysis of Recurrent Events Data	7
<i>Philip Hougaard</i>	
Thursday	8
Statistical Testing under Distributional Shifts and its Application to Causality	8
<i>Jonas Peters</i>	
Estimands: The Rosetta Stone of Estimation or Opening Pandora's box? <i>Birgitte Rønn, Mette Krog Josiassen and Helle Lynggaard</i>	8
A Statistical Sleuth in the UK Criminal Justice System	9
<i>Therese Graversen</i>	
Statistikens Skabelsesberetning og Dens Mulige Endeligt	9
<i>Lars Kirدان</i>	
The Joy of Pseudo-Values	9
<i>Per Kragh Andersen</i>	
Analysis of Spectral Data	10
<i>Line Clemmensen</i>	
Biostatistics in Twin Studies	10
<i>Jacob von Bornemann Hjelmberg</i>	
The Thrilling Journey and Benefits of Analytics at Vestas	11
<i>Sven Jesper Knudsen</i>	
<code>predict(DSTS, n.ahead=50)</code>	11
<i>Claus Thorn Ekstrøm</i>	

<i>Contents</i>	4
Friday	12
Assumption Lean Inference for Survival Data	12
<i>Torben Martinussen</i>	
25 Years More with Point Patterns in Euclidean Space and Beyond . .	12
<i>Jesper Møller</i>	
Joint Initiative for Causal Inference - from Novo Nordisk Perspective .	13
<i>Kajsa Kvist, Randi Grøn og Henrik Ravn</i>	
Locally Associated Graphical Models and Mixed Convex Exponential Families	13
<i>Steffen Lauritzen</i>	

Wednesday

DSTS' Historie

Niels Keiding

Section of Biostatistics, University of Copenhagen

Abstract: Statistikfaget havde haft svært ved at etablere sig som disciplin i Danmark, men i foråret 1971 forelå den ene efter den anden præmis for at det var på tide at samle kræfterne: Siden 1949 havde professor A. Hald (København) drevet Det Statistiske Seminar med (over 150) aftenforedrag og heldagsarrangementer af inden- og udenlandske eksperter. I 1958 oprettedes Eksamen i Statistik ved Københavns Universitet. Siden 1960 eksisterede et effektivt samarbejde mellem Københavns og Aarhus Universiteter om hovedfags-statistikundervisningen. I 1965 initierede Aarhus-gruppen Nordisk Konference om Matematisk Statistik, som skulle gå på omgang mellem de 4 nordiske lande; det var Danmarks tur næste gang (1973) Der foregik forhandlinger om et nordisk statistisk tidsskrift, og der manglede en dansk organisation til at indgå i dette projekt. Professor G. Rasch ville fylde 70 år 21 september 1971, og mange ville gerne fejre denne respekterede men også kontroversielle pionerskikkelse i moderne dansk statistik. Et fagligt selskab ville give en fin klangbund for dette.

Min fremstilling starter med, hvordan den stiftende bestyrelse håndterede disse ganske varierede udfordringer. Vi skal se, hvordan Selskabet lige fra starten havde et imponerende program af foredrag dækkende ganske bredt i teoretisk og anvendt statistik. Et omfattende 'græsrods'-initiativ i 1979 førte til detaljerede rapporter om spørgsmålet, om Selskabet var for elitært; et udvalgsarbejde foreslog at erstatte 'teoretisk' med 'anvendt og teoretisk' i Selskabets navn. Sagen kulminerede på generalforsamlingen 26/2-1980, som vedtog en modererende kurs. Den centrale aktivitet har altid været møder med foredrag; den mest langtidsholdbare version de såkaldte todages-møder, normalt tirsdag eftermiddag til onsdag middag. Der har været en lang række samarbejds-møder med nærliggende selskaber. De nordiske initiativer (Nordisk Konference om Matematisk Statistik; Scandinavian Journal of Statistics) har vist sig levedygtige. Selskabet vedtog på generalforsamlingen 26/2-2019 at optage netværket Young Statisticians Denmark som ungdomsafdeling.

Mixed Membership Models for Mutational Signatures in Cancer Genomics

Asger Hobolth

Department of Mathematics, Aarhus University

Abstract: Somatic mutations in cancer can be described as a mixture of different mutational signatures. The signatures can often be attributed to exposures such as UV-light or tobacco smoking, and they can be decoded using a mixed membership model called non-negative matrix factorization. I will describe how non-negative matrix factorization works and why it is useful for cancer treatment and precision medicine. I will then proceed by explaining some fundamental methodological problems with applications of non-negative matrix factorization, and discuss novel solutions developed in my group.

The talk is based on joint work with Ragnhild Laursen (Department of Mathematics, Aarhus University), Marta Pelizzola (University of Veterinary Medicine Vienna) and Lasse Maretty (Department of Molecular Medicine, Aarhus University).

Young Statisticians Denmark presents: Young Statistical Research in Denmark

Ann-Sophie Buchardt, Sneha Das and Nikolaj Thams

Section of Biostatistics, University of Copenhagen

DTU Compute,

Copenhagen Causality Lab, University of Copenhagen

Abstract: What kind of questions and topics occupy the minds of young statistical researchers in Denmark? This presentation will give you a short tour around a few of our statistical research departments, as three young researchers from Danish universities give a brief introduction to their research. In this way you will get a glimpse of what kind of projects new statistical researchers spend their time on.

Learning to Reflect - Data-Driven Strategies for Stochastic Control

Claudia Strauch

Department of Mathematics, Aarhus University

Abstract: One of the fundamental assumptions in stochastic control of continuous-time processes is that the dynamics of the underlying process are known. This is, however, usually obviously not fulfilled in practice. On the other hand, a rich theory for nonparametric estimation of the characteristics of continuous-time processes has been developed over the last decades. In this talk,

we discuss how to bring together these two areas for developing purely data-driven strategies for stochastic control, which we explore for ergodic singular control problems associated to continuous diffusions and Lévy processes. Applications can be found in many areas of life such as natural resource management, engineering or in the financial area.

Analysis of Recurrent Events Data

Philip Hougaard

Lundbeck and University of Southern Denmark

Abstract: Recurrent events data refer to events that over time can occur several times for each individual. The frame for this presentation is events that can occur at most a few times during a clinical trial, such as hospitalizations or heart failures. A simple approach to analyze such data is by means of a Poisson process model, but it should be clear that this approach ignores the dependence between events, whether created by random subject differences or direct dependence between the events. In clinical trials, a classical and common solution is to consider only the time to the first event and thus avoiding the need to consider potential dependence. This ignores information that is important to the patients and therefore we should search for more complete analysis techniques. There are indeed many suggestions on how to address the dependence in recurrent events data. Common to many of these suggestions is that they are very simple to fit. This simplicity has implied that the area is crowded with techniques, which unfortunately give markedly different estimates for the treatment effect. I will try to make what could be called a 360° review of some of these techniques, following the principle laid out in the classical quote of George Box “All models are wrong, but some models are useful”. By being useful, I particularly focus on whether a treatment difference can be assessed in a fair and relevant way. This is in line with the recent development in the pharmaceutical industry covering estimands, which popularly can be described as an approach to define a treatment effect, covering the real world, where treatment may be stopped due to adverse events, additional treatment may be applied or the patient may die. At the same time, the approach aims at defining the treatment effect directly in the value of variables, instead of relying on assumptions specified in statistical models. This estimand framework is a step forward but may be too formalistic for more complex data types, such as recurrent events in the presence of mortality.

Thursday

Statistical Testing under Distributional Shifts and its Application to Causality

Jonas Peters

Department of Mathematical Sciences, University of Copenhagen

Abstract: We are interested in testing a hypothesis H_0 for a target distribution P , when observing data from a different distribution Q and assume that P is related to Q through a known shift. We propose a general testing procedure and prove theoretical guarantees, even when the shift needs to be estimated from data. Testing under distributional shifts has several applications: we argue that it may prove useful in conditional independence testing, reinforcement learning, covariate shift and causal inference.

Estimands: The Rosetta Stone of Estimation or Opening Pandora's box?

Birgitte Rønn, Mette Krog Josiassen and Helle Lynggaard

*Zealand Pharma
Lundbeck
Novo Nordisk*

Abstract: Estimands have been around for many years, but not until 2010 where the FDA* commissioned report from the National Research Council on the prevention and treatment of missing data in clinical trials was published, we began discussing this explicitly within the pharma industry. The introduction, (or actually the requirement ☺) that we use estimands was done in an effort to bring clarity and transparency into what treatment effect is being estimated. At the same time, the introduction has given rise to philosophical discussions regarding what it is we have been estimating in our clinical trials, what we should be estimating in our trials and not least whether what we are actually estimating is the relevant treatment effect. In the presentation we will briefly introduce the estimand concept and framework and illustrate it with a simple example. In addition, some of the interesting challenges we face as statisticians getting to terms with this concept will be discussed.

Reference: ICH E9(R1): Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

A Statistical Sleuth in the UK Criminal Justice System

Therese Graversen

IT University of Copenhagen

Abstract: I will talk about my experiences as an expert witness in various criminal cases involving complex DNA evidence. One example of a challenge is that to a statistician it is natural to carefully tailor the analysis to the data at hand, but it is not straightforward to adopt this approach in the legal setting where analysis methods are largely approved for pre-specified scenarios and, typically, for use by non-statistician experts.

Statistikens Skabelsesberetning og Dens Mulige Endeligt

Lars Kirdan

SAS Institute

Abstract:

- Hvorfor fik vi statistikere?: 1.000 års vestlig mentalitetshistorie på 10 minutter
- Hvorfor blev statistikerne institutionaliseret?: Baggrunden for etableringen af statslige statistikkontorer
- Data and the Big Why
- Hvorfor går det så galt?: Nøgleordet for fremtiden er 'Relevans'

The Joy of Pseudo-Values

Per Kragh Andersen

Section of Biostatistics, University of Copenhagen

Abstract: Survival analysis is characterized by the need to deal with incomplete observation of the outcome variable, most frequently caused by right-censoring, and several – now standard – inference procedures have been developed to deal with this. Examples include the Kaplan-Meier estimator for the survival function and partial likelihood for estimating regression coefficients in the proportional hazards (Cox) model. During the last decades, methods based on pseudo-values have been studied. Here, the idea is to apply a transformation of the incompletely observed survival data and, thereby, to create a more simple data set for which 'standard' techniques (i.e., for complete data) may be applied, e.g., methods using generalized estimating equations.

An advantage of this approach is that it applies quite generally to (marginal) parameters for which no or few other regression methods are directly available

(including average time spent in a state of a multi-state model). Another advantage is that it allows the use of a number of graphical techniques, otherwise unavailable in survival analysis. Disadvantages include that the method is not fully efficient and that it, in its simplest form, assumes covariate-independent censoring (though generalizations to deal with this have been developed).

We will review the development in the field since the idea was put forward in a 2003 *Biometrika* paper. Focus will be on graphical methods but the theoretical properties of the approach will also be touched upon.

Analysis of Spectral Data

Line Clemmensen

DTU Compute

Abstract: My talk will discuss multivariate statistics and machine learning methods used to analyse spectral data with applications to bio production data, astrophysics, and multi-spectral images.

Biostatistics in Twin Studies

Jacob von Bornemann Hjelmberg

Department of Public Health, University of Southern Denmark

Abstract: Studies with twins have been taking place at the Danish Twin Research Center since the 1950's. Thanks to great many people remarkable results on familial influences have been revealed from this unique resource of representative twin data. The statistical biometric modelling takes form in landmark papers by R.A. Fisher published back in 1918 providing a basis for further biostatistical developments.

Having cancer studies with twins in mind we will in this talk focus on such developments, in particular the role of survival analysis in biometrics which has improved knowledge on causing factors. The within-pair dependence becomes an essential part of any modelling, and we will consider modern approaches to this end under various scenarios.

Further, the matched twin pair design has proven useful in inferring association by reducing certain confounding elements, however, causal interpretation is of biostatistical consideration. We will highlight statistical problems, areas of development and encourage to collaboration as the data is quite extraordinary providing profound insight.

The Thrilling Journey and Benefits of Analytics at Vestas

Sven Jesper Knudsen

Vestas

Abstract: Once upon a time, a Vestas engineer started collecting all the data from our wind turbines for no particular reason and not knowing he was way ahead of his time. But data has been a part of the Vestas DNA ever since. Vestas is an excellent data & analytics use case, from a profitability perspective, how data science has evolved, and how we progress with data. Our journey continues fast as our customer's sustainable energy targets are aggressive, requiring bold and innovative solutions, where data and analytics play an exciting role.

predict(DSTS, n.ahead=50)

Claus Thorn Ekstrøm

Section of Biostatistics, University of Copenhagen

Abstract: We need to talk. DSTS is turning 50, and it is time to reflect on the deeper questions in life. Where did we come from and where are we going? Are we having a midlife crisis?

I'd like to take the temperature on our relationship. With each other, with the society, with the field of statistics, and with the larger public. Where would we like to go from here and what is needed to ensure that DSTS will thrive in the coming years?

Friday

Assumption Lean Inference for Survival Data

Torben Martinussen

Section of Biostatistics, University of Copenhagen

Abstract: Much statistical inference is based on parametric or semiparametric models. Robust standard error estimators exist that are reliable even if the model is misspecified. However, in that case, it is unclear what estimand is targeted and robust inference may be of little use. The approach taken by van der Laan and co-workers with their scientific roadmap is to define the estimand directly as a function of the underlying probability measure. This is preferable if it is obvious how to pick the estimand which may not always be the case, however. We take a different approach choosing an estimand that corresponds to well known parameters under certain models. Estimation is developed based on the efficient influence function.

25 Years More with Point Patterns in Euclidean Space and Beyond

Jesper Møller

Department of Mathematical Sciences, Aalborg University

Abstract: Since DSTS' 25 years anniversary in 1996 there have been many exciting developments in statistical methodology related to analyzing point patterns in Euclidean and more general metric spaces, including the sphere and linear networks (such as road networks and dendrite networks). This talk focuses on some highlights, omitting technical details and presenting some application examples. In particular we consider 1) Spatial inhomogeneity, 2) Spatial point processes specified by a covariance function, and 3) Perfect simulation and doubly intractable distributions. We conclude with briefly mentioning other topics.

Joint Initiative for Causal Inference - from Novo Nordisk Perspective

Kajsa Kvist, Randi Grøn og Henrik Ravn

Novo Nordisk

Abstract: Novo Nordisk has a priority of increased and sustainable use of data for evidence generation in the continuum of data originating from randomized controlled trials and observational databases. Several academic research collaborations have been initiated with a focus on enabling causal inference without the protection of randomization and on disease understanding including the underlying mechanistic of an intervention.

The presentation will dive into a project within the Joint Initiative for Causal Inference. The project targets research questions originating from several long-term cardiovascular outcomes trials (CVOTs). Treatment effects beyond the intention-to-treat analysis can be challenging as randomisation may no longer protect against selection bias. Issues like adherence to randomized treatment, withdrawals, and initiation of other treatments after randomization, which may even be unequally distributed in the randomization arms, will essentially make the CVOT be a mixture of randomised and observational data. This calls for causal inference methods which can analyse a randomised trial like a randomised trial.

Locally Associated Graphical Models and Mixed Convex Exponential Families

Steffen Lauritzen

Department of Mathematical Sciences, University of Copenhagen

Abstract: The notion of multivariate total positivity has proved to be useful in finance and psychology but may be too restrictive in other applications. In this paper we propose a concept of local association, where highly connected components in a graphical model are positively associated and study its properties. Our main motivation comes from gene expression data, where graphical models have become a popular exploratory tool. The models are instances of what we term mixed convex exponential families and we show that a mixed dual likelihood estimator has simple exact properties for such families as well as asymptotic properties similar to the maximum likelihood estimator. We further relax the positivity assumption by penalizing negative partial correlations in what we term the positive graphical lasso. Finally, we develop a GOLAZO algorithm based on block-coordinate descent that applies to a number of optimization procedures that arise in the context of graphical models, including the estimation problems described above. We derive results on existence of the optimum for such problems.

This is joint work with Piotr Zwiernik, Toronto. See also [arXiv:2008.04688](https://arxiv.org/abs/2008.04688)