

Contents

Program	2
Abstract (Talks)	3
Susanne Ditlevsen APPROXIMATE LIKELIHOOD INFERENCE IN PARTIALLY OBSERVED HYPOELLIPTIC DIFFUSIONS	4
Kang Li. NEURAL STATE-SPACE MODELS DESCRIBING PARALLEL AND SERIAL VISUAL PROCESSING.	5
Frederik Riis Mikkelsen. DEGREES OF FREEDOM IN MODEL SELECTION	6
Brice Ozenne ASSESSING TREATMENT EFFECTS ON REGISTRY DATA IN PRESENCE OF COMPETING RISKS	7
Vladimir Pastukhov ESTIMATING A MONOTONE PROBABILITY MASS FUNCTION WITH KNOWN FLAT REGIONS	8
Umberto Picchini A LIKELIHOOD-FREE VERSION OF THE STOCHASTIC APPROXIMATION EM ALGORITHM (SAEM) FOR INTRACTABLE MODELS	9
Christian Bressen Pipper SEMIPARAMETRIC MULTI-PARAMETER REGRESSION SURVIVAL MODELLING	10
Liis Starkopf MEDIATION ANALYSIS WITH STOCHASTIC MEDIATOR PROCESSES	11
Stas Volkov BORDER AGGREGATION MODEL	12

LUND COPENHAGEN MEETING 2017

PROGRAMME

Venue: Auditorium MH:R
Mathematics Building (Matematikcentrum), Sölvegatan 18

Tuesday, May 23

- 10:00 – 10:05 Dragi Anevski: WELCOME AND PRACTICAL DETAILS
- 10:05 – 10:45 Susanne Ditlevsen(Mathematical Sciences, Copenhagen): APPROXIMATE LIKELIHOOD INFERENCE IN PARTIALLY OBSERVED HYPOELLIPTIC DIFFUSIONS
- 10:45 – 11:00 Fika
- 11:00 – 11:40 Liis Starkopf (Biostatistics, Copenhagen): MEDIATION ANALYSIS WITH STOCHASTIC MEDIATOR PROCESSES
- 11:40 – 12:20 Umberto Picchini (Mathematical Sciences, Lund): A LIKELIHOOD-FREE VERSION OF THE STOCHASTIC APPROXIMATION EM ALGORITHM (SAEM) FOR INTRACTABLE MODELS
- 12:20 – 13:00 Frederik Riis Mikkelsen (Mathematical Sciences, Copenhagen): DEGREES OF FREEDOM IN MODEL SELECTION
- 13:00 – 14:00 LUNCH AT THE MATHEMATICS CENTRE
- 14:00 – 14:40 Christian Bressen Pipper (Biostatistics, Copenhagen): SEMIPARAMETRIC MULTI-PARAMETER REGRESSION SURVIVAL MODELLING
- 14:40 – 15:20 Vladimir Pastukhov (Mathematical Sciences, Lund): ESTIMATING A MONOTONE PROBABILITY MASS FUNCTION WITH KNOWN FLAT REGIONS
- 15:20 – 16:00 Kang Li (Mathematical Sciences, Copenhagen): NEURAL STATE-SPACE MODELS DESCRIBING PARALLEL AND SERIAL VISUAL PROCESSING
- 16:00 – 16:30 Fika
- 16:30 – 17:10 Brice Ozenne (Biostatistics, Copenhagen): ASSESSING TREATMENT EFFECTS ON REGISTRY DATA IN PRESENCE OF COMPETING RISKS
- 17:10 – 17:50 Stas Volkov (Mathematical Sciences, Lund): BORDER AGGREGATION MODEL
- 19:00 – DINNER AT RESTAURANT SMAK, MALMÖ KONSTHALL
-

ABSTRACTS

TALKS

Approximate likelihood inference in partially observed hypoelliptic diffusions

SUSANNE DITLEVSEN

Department of Mathematical Sciences, University of Copenhagen

Hypoelliptic diffusion processes appear naturally in a variety of applications, but most parameter estimation procedures are ill conditioned when only partial observations are available. Hypoellipticity means that the diffusion matrix of the stochastic differential equation (SDE) defining the multidimensional diffusion process is not of full rank, but its solutions admit a smooth density. A prominent example is the large class of stochastic damping Hamiltonian systems, also called Langevin equations, describing the motion of a particle subject to potential, dissipative and random forces. These models have many applications, such as molecular dynamics, stochastic volatility models, paleoclimate research, neural mass models, random mechanics or classical physics. However, many applications need to allow for a more flexible formulation of the drift. For example, it can be convenient to model parts of a large deterministic system exhibiting multiple time scales by a low dimensional stochastic model, leading to a hypoelliptic structure on the reduced model. An important field of application is neuronal models of membrane potential evolution, where the noise only acts on the input, or on the ion channel dynamics, leading to hypoelliptic SDEs. Ergodicity of these models have been studied, based on the hypoellipticity of the system. But even if the model is ergodic, the degenerate noise structure complicates the statistical analysis and many standard tools break down. I will discuss ways to estimate parameters in hypoelliptic diffusion models.

Joint work with Adeline Samson.

Neural state-space models describing parallel and serial visual processing

KANG LI

Department of Mathematical Sciences, University of Copenhagen

Serial and parallel processing in visual search have been long debated in psychology but the processing mechanism remains an open issue. Serial processing allows only one object at a time to be processed, whereas parallel processing assumes that various objects are processed simultaneously. Here we present novel neural models for the two types of processing mechanisms based on analysis of simultaneously recorded spike trains. We construct state-space models combining mathematical description for neuronal attention and point process models for spike trains. The same model can explain both serial and parallel processing by adopting different parameter regimes. We present statistical methods to distinguish between serial and parallel processing based on both maximum likelihood estimates and decoding analysis of the attention. Simulation studies show the models can be successfully inferred. Results on experimental data show that both processing mechanisms are in play for the simultaneously recorded neurons, but neurons tend to follow parallel processing in the beginning after the onset of the stimulus pair, whereas they tend to serial processing later on.

Degrees of Freedom in Model Selection

FREDERIK RIIS MIKKELSEN

Department of Mathematical Sciences, University of Copenhagen

Degrees of freedom in a regression setup are generally useful to quantify bias-variance tradeoffs across models. According to Stein's Lemma one can obtain unbiased estimates of the degrees of freedom using the trace operator. This, however, relies on the estimator in question being differentiable, but most estimators with data-adaptive model selection are discontinuous. An extension to Stein's Lemma, which applies to most model selection procedures, is presented. Though this extended lemma presents a closed form expression for the degrees of freedom, it is generally very difficult to evaluate in practice. But model selection via the LASSO is an examples in which the computations simplifies considerably and yields directly applicable results.

Assessing treatment effects on registry data in presence of competing risks

BRICE OZENNE

Section of Biostatistics, Department of Public Health, University of Copenhagen

Registry data offer the opportunity to estimate the incidence of a disease at a national scale and to compare treatment effects. Estimation techniques, however, have to be corrected for the confounders that will arise due to the absence of randomization and have to handle the occurrence of competing events (e.g. death). Here we investigate the use of cause-specific Cox models to estimate the cumulative incidence of a disease over time and the use of the G-formula to compare this incidence between treatment modalities. We show how analytic expressions for the confidence interval of the cumulative incidence, and their difference, can be derived using the influence formula. Using a resampling technique, we also obtained confidence bands for the cumulative incidence function. These confidence bands can be used to compare the risk of two treatment modalities over time. We illustrate our work on data from the Danish national registries comparing the effect of preventing treatments on the risk of recurrent stroke. The proposed estimators for the cumulative incidence and their confidence interval are implemented in the `riskRegression` package that is available on Github.

Joint work with Christian Torp-Pedersen and Thomas Alexander Gerds.

Estimating a monotone probability mass function with known flat regions

VLADIMIR PASTUKHOV

Centre for Mathematical Sciences, Lund University

We propose a nonparametric ML estimator of a discrete monotone probability mass function under the further assumption that it has (known) flat regions. We analyse the estimators asymptotic properties, in particular we derive the limiting distribution, and prove that it performs better than the Grenander estimator and the monotone rearrangement estimator in both l^2 metric and Hellinger distance.

Joint work with Dragi Anevski.

A likelihood-free version of the stochastic approximation EM algorithm (SAEM) for intractable models

UMBERTO PICCHINI

Centre for Mathematical Sciences, Lund University

We present an approximate maximum likelihood methodology for the parameters of incomplete-data models. A likelihood-free version of the stochastic approximation expectation-maximization (SAEM) algorithm is constructed to maximize the likelihood function of model parameters. While SAEM is best suited for models having a tractable "complete likelihood" function, its application to moderately complex models is difficult, and results impossible for models having so-called intractable likelihoods. The latter are typically treated using approximate Bayesian computation (ABC) algorithms or synthetic likelihoods, where information from the data is carried by a set of summary statistics. While ABC is considered the state-of-art methodology for intractable likelihoods, its algorithms are often difficult to tune. On the other hand, synthetic likelihoods (SL) is a more recent methodology which is less general than ABC, it requires stronger assumptions but also less tuning. By exploiting the Gaussian assumption set by SL on data summaries, we can construct a likelihood-free version of SAEM. Our method is completely plug-and-play and available for both static and dynamic models, the ability to simulate realizations from the model being the only requirement.

Semiparametric multi-parameter regression survival modelling

CHRISTIAN BRESSEN PIPPER

Section of Biostatistics, Department of Public Health, University of Copenhagen,

We consider a log-linear model for survival data, where both the location and dispersion parameters depend on covariates and the baseline hazard function is completely unspecified. It is argued that this model provides the flexibility needed to capture many interesting features of survival data at a relatively low cost in model complexity.

Estimation procedures are developed based on identifying the counting process martingales. Moreover, asymptotic properties of the resulting estimators are derived using empirical process theory. Finally, a resampling procedure is suggested to estimate the limit distributions of the estimators. The finite sample properties of the estimators are investigated by simulation

Mediation analysis with stochastic mediator processes

LIIS STARKOPF

Section of Biostatistics, Department of Public Health, University of Copenhagen

Keywords: causal inference, mediation analysis, direct and indirect effects, time-dependent mediator:

In epidemiology and many other scientific disciplines, mediation analysis is an important tool for understanding causal mechanisms. Specifically, mediation analysis allows to disentangle the indirect effect of an exposure on outcome through a given intermediate variable, the mediator. Developments in causal inference have greatly extended the theoretical framework providing definitions and conditions for identification of direct and indirect effects as well as a number of distinct estimation procedures. The majority of the existing literature on causal mediation analysis focuses on applications with a mediator measured at one or a few specific time points thereby ignoring the fact that in any biological system the mediator is likely to be a continuous time process evolving from treatment initiation to outcome measurement. This talk will discuss the problems and challenges arising in mediation analysis when the stochastic process nature of the mediator is ignored.

Joint work with Thomas Alexander Gerds and Theis Lange.

Border aggregation model

STAS VOLKOV

Centre for Mathematical Sciences, Lund University

Consider a graph with a subset of vertices called *the border*. A particle released from the origin performs a random walk on the graph until the moment when it comes to the immediate neighbourhood of the border, at which point it joins it, thus increasing the border set by one point. Then a new particle is released from the origin, and the process repeats until the origin becomes a part of the border itself. We are interested in the total number ξ of particles to be released by this final moment. Incidentally, this model covers OK Corral model as well as the erosion model, and we obtain distributions and bounds for ξ in cases where the graph is star graph, regular tree, and a d -dimensional lattice.

Joint work with Debleena Thacker (Lund).