

FÖRSLAG PÅ EXAMENSARBETEN I MATEMATISK STATISTIK

Maj, 2009



LUNDS UNIVERSITET

Matematikcentrum
Matematisk statistik

Innehåll

1	Algoritmer för mätning av ögonrörelser	1
2	Detection and interference rejection of non-circularly symmetric sources	1
3	Detection of liquid explosives	2
4	Double height or double sample in MCV-analyses and compression tests of clayey materials	2
5	Estimating the velocity of blood using ultrasound	3
6	Evaluation of different frequency bandwidths for power estimation of the Heart Rate Variability	3
7	Inference in SDEs	3
8	Joint Calibration of P and Q parameters	4
9	Kreditrisk	4
10	Modeling of wind fields	4
11	Optimering av pensioner	5
12	Prediktion av studieresultat	5
13	Premiesättning för försäkringar	5
14	Random networks and their applications	6
15	Simultaneous Calibration and Hedging of Options	6
16	Spatial modellering	7
17	Stochastic approximation applied to MLE	7
18	Time-frequency analysis of EEG from children	8
19	Värdering och hantering av finansiell risk vid stormskador	8

1 Algoritmer för mätning av ögonrörelser

Ögonrörelsemätning blir ett allt viktigare verktyg inte bara inom forskning, men också i flera tillämpade områden (t.ex. reklam, design, bilsäkerhet, HCI, human factors). Grundläggande analys av ögonrörelsedata innebär detektering av en mängd olika funktioner:

- Fixering: Ögat är i princip stilla.
- Sackad: Kroppens snabbaste rörelse.
- Smooth pursuit: En långsam målföljning.
- Nystagmus: En sågtandad ögonrörelse som är kopplad till balanssinnet.
- Mikrosackad: Ett sätt för ögat att hindra utmattning under långa fixeringar.

Dessa rörelser motsvarar olika kognitiva processer som man är intresserad att undersöka och dessutom finns olika typer av brus. Nuvarande (alltför enkla) algoritmer klarar inte alla typer av rörelser. Att hitta en smooth-pursuit-algoritm som kan kombineras med existerande eller nya fixerings- och sackad-algoritmer är av största vikt för att forskningen ska kunna utvecklas vidare. Bättre sackad-algoritmer behövs också; särskilt behöver vi bättre kunna skilja på olika faser i en sackad. I arbetet ingår att samla ögonrörelsedata på lämpliga animerade stimuli, m.h.a. videobaserade 1250 Hz högprecisions eye-trackers och 50–200 Hz mobila hjälmsystem, som finns i Humanistlaboratoriet. För hjälmsystemens data gäller också att kunna skilja på långsamma huvudrörelser och snabba ögonrörelser, som blandas i samma datafil. Projektet innebär modellering och detektering av olika typer av funktioner och processer. Kurser i stationära processer (och även icke-stationära) samt signalbehandling och gärna detektionsteori är värdefulla som bakgrund.

Handledare:

Maria Sandsten,

Kenneth Holmqvist, humlab@sol.lu.se,

Marcus Nyström, marcus.nystrom@eit.lth.se.

Kontaktperson: Maria Sandsten, sandsten@maths.lth.se, 046-222 49 53

2 Detection and interference rejection of non-circularly symmetric sources

Until recently, most array filtering problems have only considered second-order circular sources. As many forms of communication signals are non-circular, there is currently a significant interest in examining how to best detect desired non-circular sources, as well as how to best reject interference from undesired such sources. The suggested project will examine these issues in further detail.

Kontaktperson: Andreas Jakobsson, aj@maths.lth.se, 046-222 45 20

3 Detection of liquid explosives

On August 10, 2006, the British authorities announced that they had foiled an attempted terrorist attack aimed at blowing up airplanes using liquid explosives. The use of liquid explosives immediately caused worldwide restrictions on the allowed amount of liquids in carry-on luggage. Airport security can detect several forms of solid explosives, but are currently unable to detect liquid explosives with satisfactory reliability. However, recent research suggests that spectroscopic techniques can be used to uniquely identify liquid explosives. The proposed M.Sc. thesis project will form a part of an ongoing research project aimed at the classification of liquid explosives.

Kontaktperson: Andreas Jakobsson, aj@maths.lth.se, 046-222 45 20

4 Double height or double sample in MCV-analyses and compression tests of clayey materials

Will the proportions of the specimens in a sample affect the measurements of the shear strength and will it be possible to use these measurements for modelling the shear strength of clays of different volumes?

In Sweden and as an international accepted rule, double specimen height are used for testing the shear strength of clayey materials. Double specimen height means that the height is 2 times the diameter. The reason for this is to prevent boundary effects. Double specimen heights are used in unconfined compression tests and in triaxial test.

In most test situations there will be no limitations by this procedure. However, with experiments and measurements of MCV, (Moisture Condition Value), one are limited to use specimens of single height. It is of course important to investigate the dependence and relations between MCV-values and the two types of compression tests. Is it then possible to use these single height specimens instead of the traditional double height specimens for examining the shear strength?

One solution to this problem is to combine two single height specimens, one on top of the other, to form a double specimen. Will the contact surface between the two specimens affect the measurements, perhaps by initiating the shear-plane? The two specimens will of course have two different MCV-values, which shall we use, the minimum, maximum or perhaps the mean value?

Another solution is to use the single height specimens as they are and instead have twice as many specimens. In Denmark they are using single height specimens as a standard. Will we have boundary effects caused by the untraditional proportion or will we win inference by the double samples?

The purpose of this master thesis is to investigate the following questions:

- Are there any differences between the shear strength for single and double specimen height?
- If we are using two single height specimen instead of one double height, will this affect the compression tests?
- Can we scale the results from one specimen to another of different dimensions.
- How will we model and estimate the shear strength in practice?

Kontaktperson: Peter Gustafsson, pegus@maths.lth.se, 046-222 44 42

5 Estimating the velocity of blood using ultrasound

Medical ultrasound is a common tool to determine the blood velocity in a particular blood vessel. The velocity distribution is important for instance to detect turbulent blood flow, which can be used for diagnosing carotid artery stenosis. Typically, the blood spectral distribution is estimated using a windowed spectrogram, which suffers from several well-known limitations. In this project, we will examine and derive novel data-adaptive blood spectral estimation techniques allowing for substantially better ultrasonic imaging. This is a joint project with DTU.

Kontaktperson: Andreas Jakobsson, aj@maths.lth.se, 046-222 45 20

6 Evaluation of different frequency bandwidths for power estimation of the Heart Rate Variability

Methods to estimate Heart Rate Variability (HRV) power are highly interesting in medical diagnosis as well as physiological studies. The HRV high frequency (HF) power estimate is generally based on a frequency band ranging from 0.12–0.4 Hz, that is within the area that adults are normally breathing. However, recent research indicates that it seems reasonable to base the HF central frequency on the respiratory peak. It is well known, that in addition to the positive correlation between respiration rate and HF peak frequency, HF power decreases as respiration frequency increases.

Hence, in this study the respiratory peak is detected and the HF-band is defined as a frequency band around the respiratory frequency. The intention of the study is to evaluate the most optimal bandwidth of this frequency band and compare the result to the usual frequency band 0.12–0.4 Hz. It is also interesting to evaluate if it is possible to estimate the HF peak frequency directly from the HRV data to avoid measurement of the respiratory frequency. The data collection is made by Peter Jönsson at the Division of Occupational and Environmental Medicine, Dept. of Laboratory Medicine, Lund University.

Kontaktperson: Maria Sandsten, sandsten@maths.lth.se, 046-222 49 53

7 Inference in SDEs

An increasing number of financial models are formulated in using (systems of) Stochastic Differential Equations (SDE). Recent research has shown how the likelihood can be approximated for discretely observed, scalar models and even some discretely observed multivariate models.

Several extensions of the current knowledge is possible, including partially observed models (such as stochastic volatility), recursive (and possibly adaptive) estimation or by including jumps.

Kontaktperson: Erik Lindström, erikl@maths.lth.se, 046-222 45 78

8 Joint Calibration of P and Q parameters

Can we combine options and the underlying asset to derive efficient algorithms for estimating the P and Q dynamics simultaneously? A simple solution would be to invert the volatility from the options to estimate P parameters while the options are being used for the Q parameters.

This is probably suboptimal as it is possible to combine these estimation problems into a single problem through Kalman filters and other advanced statistical methods. The different estimation strategies should preferably be evaluated on simulated data for comparison.

Kontaktperson: Erik Lindström, erikl@maths.lth.se, 046-222 45 78

9 Kreditrisk

Bör försäkringsbolag utvärdera kreditrisken, det vill säga risken för uteblivna inbetalningar, från företag som de är finansiellt involverade i? Det hade varit intressant att göra modeller för sannolikheten att företag går i konkurs eller på annat sätt inte kan uppfylla sina betalningspåtaganden, vilket kan vara intressant för ett försäkringsbolag eftersom det i så fall borde påverka premiesättningen. Om ett företag har hög konkursrisk borde detta belastas med en högre premie eftersom det finns en stor sannolikhet att företaget inte kommer att betala premien kontraktstiden ut. Hur ska i så fall en modell som beskriver detta utformas och optimeras så att den så bra som möjligt följer marknadens och det aktuella företags utveckling. Här skulle det även vara intressant att koppla modelleringen av kreditrisk till BASEL II som är det kommande systemet för hur banker bör agera.

Kontaktperson: Nader Tajvidi, nader@maths.lth.se, 046-222 96 12

10 Modeling of wind fields

Wind power is becoming an increasingly important part of the Nordic power market. The difference between wind power and say hydro or nuclear power is that the actual supply is much harder to predict, making the price of electricity more volatile.

The purpose of the thesis is to model wind fields, i.e. spatio-temporal modeling of the evolution of wind over an area and possibly connect the predictions of wind with the short term variations of the electricity price.

This thesis is part of a joint EU-project with the Technical University of Denmark (DTU).

Kontaktperson: Erik Lindström, erikl@maths.lth.se, 046-222 45 78

11 Optimering av pensioner

Pension är uppskjutande av konsumtion idag till framtiden. Då denna framtid i allra högsta grad är okänd skulle det vara intressant att modellera detta som ett stokastiskt (slumpmässigt) dynamiskt system. Möjliga fokus kan vara på överlevnadsanalys, behandling av pensionen som ett derivat och prissättning av detta eller det ekonomiska lägets påverkan på pensionsmodellen. Även det ovan beskrivna copulas kanske kan användas för att göra en multivariat behandling av problemet.

Andra förslag är att inom pensionssystemet utveckla och utvärdera portföljoptimerande metoder så att maximal avkastning erhålls på den nivå av riskexponering som kunden har valt. Även här skulle metoder från extremvärdesanalysen kunna implementeras tillsammans med ekonomiska begrepp som ”value at risk”, det vill säga hur mycket pengar vi maximalt riskerar att förlora med en viss sannolikhet.

Kontaktperson: Nader Tajvidi, nader@maths.lth.se, 046-222 96 12

12 Prediktion av studieresultat

Hur det går för teknologerna vid LTH varierar från år till år av många olika anledningar. LTH:s ledning behöver kunna få bra prognoser över, bland annat, hur många examina som kommer att utfärdas, både på kort och lång sikt, och hur många helårsstudenter det kommer att finnas det närmsta året. Examensarbetet går ut på att, med hjälp av historiska data, bygga statistiska modeller som kan användas. De metoder som kan användas är, t.ex. regressionsanalys, markovkedjor och överlevnadsanalys.

Kontaktperson: Anna Lindgren, anna@maths.lth.se, 046-222 42 76

13 Premiesättning för försäkringar

Som en förbättring av den existerande modellen och på grund av den nya lagen om att försäkringspremier inte får baseras på kön är det intressant att finna nya faktorer som kan ge en mer rättvis premiesättning. Detta skulle vi kunna göra med multivariatanalys, det vill säga att vi i princip tar alla faktorer som vi tror kan påverka försäkringsanspråken och som vi har tillräcklig data om och undersöker vilka som har en signifikant betydelse för försäkringsanspråken. Ett intressant, nytt och spännande sätt att tillmötesgå multivariat risk är med Copulas, vilka är speciellt bra anpassade för att behandla finansiella riskfrågor där de inblandade slumpvariablerna inte är normalfördelade. Även här är en möjlig fortsättning att sätta premien så att den täcker kostnaden men samtidigt ligger på en nivå så att försäkringstagaren vill ta försäkringen.

Kontaktperson: Nader Tajvidi, nader@maths.lth.se, 046-222 96 12

14 Random networks and their applications

Random graphs became perhaps the most studied objects in probability theory over the last decade. Many "real-world networks" such as electrical power networks, social networks, railways, WWW, or even neuronal networks can be modelled with a help of random graphs. The most intriguing feature of random graphs is the phase transition, which is an abrupt change of global properties of a system due to a continuous change of parameters. Study of the phase transitions is central in the theory of random graphs, and information on the critical parameters is crucial for understanding of real world systems.

Here are some possible areas for a Master Thesis project:

- Phase transition in dynamical random graphs.
- Modelling and analysis of evolution of a neural network.
- Spread of epidemics in biological systems.

Kontaktperson: Tayana Turova, tatyana@maths.lth.se, 046-222 85 43

15 Simultaneous Calibration and Hedging of Options

The most common risk management technique for options is hedging, i.e. creating a portfolio of the underlying asset, possibly several options written on the same underlying asset and a bank account. The problem is then to combine the assets in the portfolio such that the evolution of the portfolio mimics the option in question.

This can be done for most option valuation models, but the allocation within the portfolio will depend on the parameters (P and Q) of the model. It is therefore crucial to use state of the art calibration methods as input to the calibration algorithm. Such algorithm can be based on non-linear Kalman filters.

The purpose of the master s thesis is to combine the non-linear filters with a suitable computational algorithm for calibration. Our preliminary results indicate that the calibration can be obtained from the filters with little extra cost.

Kontaktperson: Erik Lindström, erikl@maths.lth.se, 046-222 45 78

16 Spatial modellering

Under de senaste åren har ett nytt sätt att konstruera modeller för slumpmässiga spatiala (rums-
liga) fenomen utvecklats, baserat på kopplingar mellan Markov-fält och stokastiska partiella diffe-
rentialekvationer (SPDEer). Tillämpningarna är breda, med allt från analys av bottenstrukturer i
Östersjön eller sjukdomsutbredning, till vegetationsanalys i Afrika och global väder- och klimat-
modellering.

Metodernas styrka ligger i att man förvånansvärt enkelt kan specificera icke-triviala modeller,
med både oscillerande och icke-stationär kovariansstruktur, men ändå få praktiskt beräkningsbara
problem med hjälp av numeriska metoder för glesa matriser. Till hjälp för detta finns ett fritt pro-
grampaket, GMRFLib, som hanterar Markovfälten, och ett annat paket, INLA, som hanterar skatt-
ning av parametrar i de hierarkiska Bayesianska modeller som används. I exjobbet kan du behöva
implementera metoder i C/C++, Matlab och/eller R.

Förslag på projekt, där graden av teori och praktisk data-analys kan varieras:

- Utvidga de enkla modellerna med kovariater, det vill säga ytterligare information såsom to-
pografi eller flöden, som styr icke-stationariteten. Tillämpningar både på land och till havs.
- Integrera de nya statistiska metoderna med befintliga data-analyspaket från Naturgeografi,
och tillämpa på praktiska problem därifrån.
- Utveckla spatio-temporala varianter av modellerna, till exempel genom att utgå från enklare
befintliga modellerna med temporala trender, eller genom att konstruera dem direkt från
spatio-temporal SPDEer.
- Utveckla Markovfälts-modeller som imiterar havsvågor med givet spektrum.

Kontaktperson: Finn Lindgren, finn@maths.lth.se, 046-222 98 37

17 Stochastic approximation applied to MLE

A popular method for estimating parameters in partially observed models is to augment the latent
state vector with the parameters, often combined with random walk dynamics. The limitation of
this approach is that the random walk dynamics prevents convergence of the estimates. The average
of the filtered parameters is often different from the true parameters, making the estimates biased.

A technique would therefore be to decrease the variance of the random walk as additional
observations become available, with the purpose of decreasing the bias. Stochastic approximation is
a framework for balancing the decrease such that convergence is obtained while preserving the nice
properties of filter algorithm.

Kontaktperson: Erik Lindström, erikl@maths.lth.se, 046-222 45 78

18 Time-frequency analysis of EEG from children

How are small children affected by watching everyday communication situations? During the study, the EEG of 20–24 month old children was recorded, while they were watching a repetitive sequence of three main types of film clips showing different communication situations. The goal of the study is to investigate the difference in the measured EEG during the three types of clips. The data collection has been made at the Department of Linguistics and Scandinavian Studies, University of Oslo, Norway.

The recorded data is highly non-stationary and we are interested in estimating the time-dependent frequency behavior. Time-frequency analysis of non-stationary processes has been approached from different view-points and with different assumptions of the non-stationarity. One possible data model is the use a model based covariance function which is modeled from the grand average information from data. In this study, the idea is to use data modeled covariance functions to estimate data optimal time-frequency spectra, which are used to classify the measured signals from the three different types of film clips.

Kontaktperson: Maria Sandsten, sandsten@maths.lth.se, 046-222 49 53

19 Värdering och hantering av finansiell risk vid stormskador

På basis av de intressanta artiklarna om stormskador av Rootzen och Tajvidi, se nedan under litteratur, skulle man kunna studera dess omfattning och försäkringsmässiga konsekvenser, exempelvis genom:

Vinddata är ofta lokalt uppmätta, men verkan är spatial, dvs skadorna är spridda över stora ytor. Därtill kommer att vinddata har mycket stor variation. Också försäkringsanspråken är spatiala, men med andra gränser än vindpå verkan. Hur gör man en bättre modell som tar hänsyn till de lokala mätningarna, vinddata från vädertjänsten, skadornas utsträckning och försäkringstagarnas anspråk?

Det behöver inte enbart vara vindhastigheten som påverkar stormskadorna utan det kan vara en samling faktorer: vindriktning, regn, hagel, om det blir elavbrott, brand etc. Vidare kan det försäkrade objektets omgivning spela roll för skadorna. Kan vi hitta de faktorer som har en signifikant påverkan?

Finns det en uppgående trend i stormskador?

Hur mycket återförsäkring skall man köpa för sina försäkringar?

Alla fortsättningarna går egentligen ut på att behandla den information som vi har för att kunna skapa oss en bättre uppfattning om vad som kan hända i framtiden och därmed minska risken på förluster genom att sätta en bättre avvägd premie.

Möjliga fortsättningar på vårt arbete kan vara just premiesättning utifrån vad vi kommer fram till, kanske även kopplat till kundernas riskaversion och "willingness to pay".

Kontaktperson: Nader Tajvidi, nader@maths.lth.se, 046-222 96 12

Maj, 2009
Matematisk statistik
Matematikcentrum
Lunds universitet
Box 118, 221 00 Lund
<http://www.maths.lth.se/>