

Reserveret postvæsenet

MEDDELELSER

DANSK SELSKAB FOR TEORETISK STATISTIK

17. årg. nr. 1

januar 1992

ATS: Statistisk kollokvium.
Afholdes (hvis ikke andet er angivet) i lokale H2.28 på Aarhus
Universitet, Ny Munkegade, 8000 Aarhus C.
Arrangeres af: Afdeling for Teoretisk Statistik, Matematisk Institut,
Aarhus Universitet

AMD: Seminar.
Afholdes (hvis ikke andet er angivet) i lok. E3-109 på Aalborg
Universitetscenter, Frederik Bajers Vej 7, 9220 Aalborg Ø.
Arrangeres af: Afdeling for matematik og datalogi, Institut for
elektroniske systemer, Aalborg Universitetscenter.

Januar:

21. januar Henrik Stryhn, (Landbohøjskolen): Change Point Models in
Image Analysis.
DSTS, kl. 19.30.
27.-29. januar Symposium i anvendt statistik. Arrangeres af UNI*C og
afholdes på Århus Universitet.

Februar:

25. februar Generalforsamling i DSTS.

MØDE I SELSKABET TIRSDAG DEN 21. JANUAR KL. 19.30 I AUDITORIUM X PÅ H.C. ØRSTED INSTITUTET

CHANGE POINT MODELS IN IMAGE ANALYSIS

Henrik Stryhn, Den kongelige Veterinær- og Landbohøjskole

Abstract: The classical problem of estimating a change point in a sequence of random variables may be generalized to lattice variables and a boundary curve dividing the observations into contiguous regions. This leads to a statistical approach to image segmentation, by explicit modeling of the boundary curves and the distributions of the observations within regions. So far attention has been restricted to univariate boundaries, of which the starshaped curves constitute our main example. Results are shown from analyses of two-region images using maximum likelihood estimation of spline curves and homogeneous or regression-type region distributions.

In the second part of the talk asymptotic results are presented for a simple class of boundary estimators. The asymptotic distributions are given in terms of the locations of maxima of Brownian motions with triangular, parabolic or parabolic-triangular drift. In several examples good agreement was obtained between the empirical and asymptotic distributions of the boundary estimators.

Der afholdes **GENERALFORSAMLING** i DSTS tirsdag d. 25.
februar 1992. Nærmere herom i næste nummer af **MEDDELELSER**.

Institutsammenlægning er et aktuelt emne for tiden. Månedens kommentar handler om, hvorfor Institut for Matematisk Statistik ved Københavns Universitet mener, at den påtænkte sammenlægning, der vedrører deres institut, vil være en dårlig idé.

Desuden indeholder dette nummer en anmeldelse af programsystemet CoCo.

Godt nytår!

Helle Holst

Næste nummer af MEDDELELSER udkommer i begyndelsen af februar. Bidrag til dette nummer skal være redaktøren i hænde senest onsdag den 22. januar 1992. Bidrag bedes sendt til:

Meddelelser, v/Helle Holst
IMSOR, BYGNING 321
DTH
2800 LYNGBY

eller med e-mail til: hh@imsor.dth.dk

Trykt på IMSOR

Ansvarshavende Helle Holst

MÅNEDENS KOMMENTAR

INSTITUTSTRUKTUREN VED MATEMATISK CENTRALINSTITUT

Hans Brøns
Institut for Matematisk Statistik
Københavns Universitet

Med udgangspunkt i ministerielle krav gennemføres for tiden en omfattende centralisering på de højere læreanstalter. Som et led heri ønsker det naturvidenskabelige fakultetsråd på Københavns Universitet, at sammenlægge Institut for Matematisk Statistik og Forsikringsmatematisk Laboratorium med Matematisk Institut.

De involverede institutter har ved flere lejligheder afvist et sådant forslag som fagligt uhensigtsmæssigt. Alligevel er det blevet fremsat igen. I den anledning har medarbejderne ved Institut for Matematisk Statistik udarbejdet følgende begrundelse for vort ønske om at bevare et selvstændigt statistisk institut, idet vi henviser til internationale erfaringer og til de principper, der hidtil har været bærende i organiseringen af dansk statistik. Da spørgsmålet kunne tænkes at have interesse for en bredere kreds har vi bedt MEDDELELSER om at bringe udtalelsen som MÅNEDENS KOMMENTAR.

Hans Brøns
Institutbestyrer

Fakultetsrådet har i længere tid diskuteret institutsammenlægninger. Der har bl. a. været fremsat forslag om, at institutterne inden for Matematisk Centralinstitut skulle lægges sammen til et institut. Dette forslag er tidligere blevet afvist af centralinstitutet som administrativt, undervisningsmæssigt og forskningsmæssigt uhensigtsmæssigt. Da forslaget alligevel er bragt op igen, skal medarbejderne ved Institut for Matematisk Statistik hermed skriftligt begrunde afvisningen med henblik på en forestående faglig diskussion i fakultetsrådet.

Organisation

Matematisk Centralinstitut består for tiden af tre institutter: Matematisk Institut, Forsikringsmatematisk Laboratorium og Institut for Matematisk Statistik. Institutterne uddanner hovedfagsstuderende inden for henholdsvis matematik, statistik og forsikringsmatematik. De to sidste studier er inden for de sidste 15 år overflyttet fra det samfundsvidenskabelige hovedområde. Derudover varetages undervisning ved det i 1986 oprettede matematik-økonomi studium. I forbindelse med dette studium skal der i de kommende år besættes en række stillinger, hvis placering endnu ikke er afklaret.

Det naturvidenskabelige hovedområde omfatter fagligt og undervisningsmæssigt meget forskelligartede fag. Strukturen ved fakultetet bør afspejle fagenes indbyrdes placering og give de enkelte fagområder optimale udfoldelsesmuligheder. Den bør ikke være dikteret af overordnede organers ønsker om administrative lettelser.

Matematisk Centralinstituts nuværende struktur giver en effektiv administration, fordi sagerne bliver afgjort på det nærmest fagkyndige niveau. Samtidig følger institutgrænserne nøje den undervisningsmæssige opdeling. De enkelte institutter er gennem valgreglerne til centralinstitutrådet sikret repræsentation, når de fælles afgørelser træffes, idet der i Matematisk Centralinstituts forretningsorden er indført mindretalsbeskyttelse ved valg til institutråd og bestyrelse. Dette er muligt, da centralinstitutterne ikke er omfattet af standardforretningsordenen for Københavns Universitet. Det er imidlertid efter styrelsesloven umuligt at foretage en permanent opdeling i repræsentationsområder eller lignende ved et institut. Repræsentation af de små enheder under et institut er således helt baseret på velvilje fra de store enheder. Et sådant forhold er hverken holdbart eller acceptabelt. Det aktuelle størrelsesforhold mellem institutterne på Matematisk Centralinstitut er sådan, at Matematisk Institut reelt ville komme til at dominere det nye institut

fuldstændigt. Der er altså ikke tale om en institutsammenlægning i almindelig forstand. Her skal tilføjes, at de aktuelle bemanningstal ikke er resultatet af nogen bevidst planlægning, men først og fremmest afspejler den vanskelige rekrutteringssituation, som anvendt matematiske universitetsinstitutter står i, når man skal konkurrere med attraktive arbejdspladser i det private erhvervsliv.

Da det naturvidenskabelige fakultet i 1990 udskilte Datalogisk Institut som et selvstændigt centralinstitut, fjernede man en mulighed for at opbygge et samlet matematisk institut med en rimelig balance mellem ren matematik og de anvendt matematiske fag, således som man har det på de andre danske universiteter. I Århus, Ålborg, og Odense har man institutter, der omfatter både matematik, datalogi og statistik. Det samme gælder på Landbohøjskolen, mens man på DTH og Handelshøjskolen har selvstændige statistiske institutter. Også på det samfundsvidenskabelige hovedområde på Københavns Universitet har man et selvstændigt statistisk institut.

Forholdet mellem ren og anvendt matematik.

Næsten alle fag befinder sig i et spændingsfelt mellem teoretiske områder, som lægger hovedvægten på metodeudvikling, og mere praktiske anvendelser, der udnytter de teoretiske resultater. Det er et generelt problem, at dette teoretiseringshierarki fører til inkommensurable kvalitetsnormer, og det gælder i udpræget grad inden for matematik på grund af matematikkens ekstreme faglige placering. Problemet har i længere tid været genstand for en dybtgående diskussion i de matematiske miljøer og har været behandlet i adskillige tidsskriftartikler. Forholdet mellem ren og anvendt matematik var et af punkterne på det Landsmøde om Matematikken i Danmark, som Dansk Matematisk Forening afholdt i 1981. og i landsmøderapporten [1] hedder det:

"Disse spørgsmål (de inkommensurable kvalitetsnormer), har stor betydning i det daglige arbejde i forskningsmiljøerne, hvor den konstruktive kollegiale kritik forudsætter gensidig indsigt. Det er en almindelig (også international) erfaring, at rene matematikere kan have vanskeligt ved at tilpasse sig de mere nuancerede kvalitetsbegreber, der gælder i andre fag. Dette er særligt udpræget i forhold til nærtstående discipliner som statistik, datalogi, operationsanalyse, matematisk fysik, matematisk biologi og økonomisk teori....

De netop nævnte forhold motiverer, at de anvendt-matematiske discipliner har behov for at kunne udvikle sig på egne præmisser inden for selvstændige forskningsmiljøer, hvor den snævre rent-matematiske kvalitetsnorm ikke er altdominerende. Samtidig er det utvivlsomt nyttigt og frugtbart for alle parter, at organisatoriske og geografiske opdelinger ikke vanskeliggør samarbejde og gensidig inspiration og information om problemer og metoder."

Bringes de anvendt matematiske miljøer i den situation, at det bliver et flertal af forskere inden for ren matematik, der skal vurdere forskningens og undervisningens kvalitet, vil det svække de pågældende områders muligheder for at få de nødvendige bevillinger og stillinger, fordi deres indsats vil blive vurderet i forhold til et andet fags kvalitetsnormer. I en nedskæringsperiode vil dette være dræbende for de små fag.

Statistiske institutters rolle.

Problemerne om den rene og den anvendte matematik er som nævnt ikke nye. I det internationale statistiske miljø har de været diskuteret i de sidste 50 år. I slutningen af 30'erne og begyndelsen af 1940'erne formulerede en gruppe af tidens betydeligste amerikanske statistikere med Harold Hotelling, Jerzy Neyman, W.E. Deming, Milton Friedman og Paul Hoel i spidsen deres begrundelse for statistikkens selvstændige placering. Synspunkterne blev fremlagt i to skelsættende forelæsninger af Hotelling: "The Teaching of Statistics" [6] fra 1940 og "The Place of Statistics in the University" [7] fra 1946. Gruppen går med stor styrke ind for oprettelse af selvstændige institutter, idet de påpeger det statistiske instituts rolle som bro mellem matematik og de discipliner på universitetet, der anvender statistik som redskab. Samme synspunkter blev gjort gældende af det udvalg, der på foranledning af Københavns Universitets rektor i 1950 afgav betænkning om "Statistikkens stilling som selvstændig videnskab og som hjælpevidenskab ved de højere læreanstalter i København" [2].

Internationalt set er hovedparten af de steder, der spiller en væsentlig rolle inden for statistik, organiseret i selvstændige statistiske institutter. Dette gælder specielt i USA, der i løbet af en relativt kort årrække har udviklet et stærkt statistisk miljø. Ben-David [3] fremhæver, at den amerikanske tradition med selvstændige institutter har været medvirkende til den hurtige udvikling, ikke blot af statistik, men også af andre nye områder.

Da statistik har rødder i de praktiske og filosofiske problemer i forbindelse med induktive slutninger, har statistik kunnet spille en vigtig rolle i mange anvendelser. Statistik er ikke opstået som en matematisk disciplin, men er opstået fra dataindsamling og fra praktiske problemer først i astronomi og demografi (og endnu tidligere i jura og religion), senere fra problemer i medicin og socialvidenskab. Statistik har i dag en meget bred berøringsflade. Der er institutter eller afdelinger for anvendt statistik mange steder på universiteter og højere læreanstalter og på forskellige forskningsinstitutioner. Skal de teoretiske statistiske institutter kunne opfylde deres rolle som formidlere mellem den teoretiske statistik og anvendelser i forskningsmiljøer på universitetet og andre steder, kræver det en institutionel identitet.

Ressourcesituationen på Institut for Matematisk Statistik har hidtil medført, at instituttet ikke internt på fakultetet har haft mulighed for at

etablere et undervisningsmæssigt samarbejde med andre fag i den udstrækning, som vi føler at der behov for. En sammenlægning med matematik kan komme til at medføre at statistik bliver isoleret som en matematisk disciplin. Dette er i modstrid med den internationale udvikling inden for faget og i modstrid med de ønsker, der er på instituttet.

Kandidatuddannelser og undervisning.

Både aktuar- og statistikuddannelsen har som det fremgår deres oprindelse på det samfundsvidenskabelige fakultet. Dansk statistik og forsikringsmatematik er imidlertid stærkt matematisk baserede. De danske kandidater har en meget omfattende og moderne matematisk uddannelse, som gør det muligt at give dem en effektiv undervisning ved at bygge på deres gode matematikkundskaber. De er imidlertid ikke kandidater i matematik, og vore kandidater vil ikke være tjent med at blive vurderet som matematikere f. eks. når det gælder uddeling af stipendier og besættelse af videnskabelige stillinger.

De tre institutter udfører langt størsteparten af deres undervisning inden for rammerne af Matematisk Centralinstitut og underviser til en vis grad de samme studenter, men i forskellige fag. Det gælder for alle tre institutter, at lærerne har en meget stor undervisningsbyrde, og en sammenlægning af institutterne vil på grund af fagenes forskellighed ikke føre til en ændret udnyttelse af lærerstaben. Dette afspejler sig også i det faktum, at der ikke er nogen vekselvirkning når det gælder ansættelser i stipendie- eller andre stillinger. I statistik vil udefra kommende kandidater typisk være uddannet som statistikere i Århus eller Ålborg, være kandidater fra DtH eller cand. polit'er.

Det er et vigtigt træk ved de danske matematisk-statistiske miljøer, at man underviser og forsker både i statistik og i sandsynlighedsregning. Dette har givet en meget frugtbar vekselvirkning mellem de to discipliner. Her kan f. eks. henvises til de seneste års banebrydende arbejder inden for statistisk analyse af tælleprocesser, hvor Skandinavien har en meget fremtrædende position (se f.eks. David-rapporten [4]).

Institut for Matematisk Statistik står som nævnt i den situation, at det er svært at tiltrække og fastholde kvalificerede kandidater, da der findes attraktive stillinger mange andre steder. Vore muligheder for at rekruttere nye kandidater afhænger i høj grad af, at de kan være sikre på at blive vurderet som statistikere og ikke blive skubbet til side til fordel for hovedfagsmatematikere ved stillingsbesættelser. Samtidig må vi have muligheder for at slå stillinger op, når der er kvalificerede kandidater. Undervisningsbyrden ved instituttet er meget stor, da der i de sidste 10 år har været en betydelig afgang af lærere til nyoprettede topstillinger ved andre højere læreanstalter, uden at der er blevet foretaget de nødvendige nyansættelser. En institutsammenlægning vil medføre en sådan forringelse af arbejdsforholdene, at det vil blive svært at fastholde den

nuværende stab og bevare det faglige statistiske miljø.

Litteraturliste.

- [1] RAPPORT FRA LANDSMØDET OM MATEMATIKKEN I DANMARK 1981: *Forskning og Uddannelse i Matematikkens Anvendelser i andre Fagområder* p. 352-4. Dansk Matematisk Forening.
- [2] BETÆNKNING AF 22. JUNI 1950: *Statistikkens Stilling som Selvstændig Videnskab og som Hjælpevidenskab ved de Højere Lærestalter i København*. Københavns Universitet.
- [3] BEN-DAVID, J. (1971), *The Scientist's Role in Society: A Comparative Study*. Prentice-Hall, New York.
- [4] NATIONAL RESEARCH COUNCIL (1984), *Renewing U.S. Mathematics: Critical Resource for the Future. ("The David Report")*. National Academy Press, Washington D.C.
- [5] GARFUNKEL, A.S and YOUNG, G.S. (1990). *Mathematics outside of Mathematics Departments*. Not. Amer. Mat. Soc. 37 408-11.
- [6] HOTELLING, H. (1988a). *The Teaching of Statistics*. Statist. Sci. 3 63 -71.
- [7] HOTELLING, H. (1988b). *The place of Statistics in the University (with discussion)*. Statist. Sci. 3 72-83.
- [8] SHAFER, G. (1990). *The Unity and Diversity of Probability (with discussion)*. Statist. Sci. 5 435-443.



**Central Forskningsenhed for almen praksis
og
Statistisk Forskningsenhed
søger
statistikere
(cand.scient./stat./act., ingeniør)**

Projektet "Diabetesomsorg i almen praksis" er et opfølgingsstudie med eksperimentelt design. I projektet deltager 500 alment praktiserende læger og 1400 nyopdagede diabetikere. Tidshorisont: 1996. Dataindsamlingen foregår løbende, men projektet er nu nået analysefasen.

Der påregnes anvendt mange forskellige rutinemæssige statistiske og epidemiologiske analyseteknikker. Projektet indebærer imidlertid også validitetsundersøgelser, registerstudier (i samarbejde med Danmarks Statistik) og egtl. metodeudvikling - f.eks. 1) håndtering af "missing data" og 2) analyse af forløbsdata. Ansættelsen indebærer tæt samarbejde med projektlederen (Niels Olivarius) og med Statistisk Forskningsenhed (Niels Keiding).

Ansættelsen, som sker på fuld tid, påregnes at vare mindst 3 år. Tiltrædelse 1. marts 1992 eller snarest herefter. Løn og pension iht. overenskomst imellem AC og Staten.

Nærmere oplysning hos

prof. Niels Keiding, tlf. 31 35 79 00 og
læge Niels Olivarius, tlf. 35 37 31 00.

Skriftlig ansøgning senest 27. januar 1992 kl. 12 til

Statistisk Forskningsenhed
Panum Institut
Blegdamsvej 3
2200 København N

Software annoncering: CoCo

Version 1.0 af CoCo, et avanceret program for analyse af kontingenstabeller, er nu tilgængelig.

CoCo er specielt effektiv på *grafiske modeller*, og nogle af kommandoerne i CoCo er designet til at håndtere grafiske modeller. Grafiske modeller er for kontingenstabeller log-lineære vekselvirkningsmodeller, der kan repræsenteres ved en simpel ikke-orienteret graf, hvor antallet af knuder i grafen er lig tabellens dimension, en knude for hver faktor. Disse modeller kan gives en fortolkning i form af betingede uafhængigheder, og fortolkningen kan læses direkte af grafen i form af en Markov egenskab. Klassen af grafiske modeller er en delmængde af hierarkiske modeller, men indeholder alle dekomposable modeller.

CoCo er lavet med henblik på estimation og test i store kontingenstabeller. Ved hjælp af graf-teoretiske resultater (Rose, Tarjan & Lueker 1976, Tarjan & Yannakakis 1984 og Tarjan 1985) er de hierarkiske log-lineære vekselvirkningsmodeller dekomponeret. IPS-algorithm er ikke anvendt på den fulde tabel, men kun på atomer, der ikke kan dekomponeres yderligere. Hvis en model testes mod en anden, og de to modeller har fælles dekompositioner, så kan testet deles i test på mindre tabeller (Goodman 1971). Collapsibility (Asmussen & Edwards 1983) af test er anvendt.

Estimation og test mellem modeller med et ubegrænset antal faktorer (80 faktorer) kan da udføres på en PC, hvis det største ikke dekomposable atom i graferne for modellerne ikke har mere end 14 binære knuder (24 på arbejdsstationer).

Kommandoer for interaktiv modelsøgning er implementeret. En semi-automatisk modelsøgning mellem grafiske modeller er mulig ved backward elimination og forward selection. Den globale modelsøgningsprocedure fra Edwards & Havránek (1985) er også inkluderet i programmet i både den grafiske og den hierarkiske version. Et justeret antal frihedsgrader kan beregnes. Videre kan inkomplette tabeller og tabeller med ufuldstændige observationer håndteres. Eksakte test mellem to vilkårlige dekomposable modeller, hvor den ene er en delmodel af den anden, kan beregnes. CoCo kan loads ind i New S (S-plus) og XLISP-STAT. Tabeller og modeller kan sendes til CoCo-objektet i New S eller XLISP-STAT, og funktioner for returnering af beregnede stikprøvestørrelser og tabelværdier fra CoCo-objektet for f.eks. højopløsningsgrafik er lavet.

For tiden køre CoCo under Unix og DOS. CoCo kan hentes via anonym ftp over internet fra iesd.auc.dk (130.225.48.4). Eller få tidsendt CoCo på diskette (3.5" eller 5.25" High density). Send en formateret diskette for at sikre, at disken, som jeg skriver, kan læses af din PC'er. En manual på engelsk til CoCo findes (enten trykket og indbundet fra Afdelingen for Matematik og Datalogi eller i TeX eller postscript kode fra ftp adressen).

KEY WORDS: statistical analysis of discrete data, contingency table, decomposable, graphical and hierarchical models, log-linear model, interactions, independence graph, conditional independence, closed form expression for maximum likelihood estimates, IPS algorithm, large sparse tables, adjusted DF., zero partial association, exact conditional test by Monte Carlo approximation, exact conditional test between any two nested decomposable models, incomplete tables = Structural Zeros, conditional quasi-independence, incomplete data = missing values, EM-algorithm, tabulation, description, plotting, listing and exporting of observed counts and fitted values: probabilities, expected counts and residual (absolute, standardized, adjusted, Freeman-Tukey, etc.), partitioning and factorization of test, collapsibility, model editing commands, model selection strategy, coherence, incremental search procedures: backward and forward, restricted incremental search, global search procedures, restricted global search, XLISP-STAT, New S.

Jens Henrik Badsberg
Afdelingen for Matematik og Datalogi
Institut for Elektroniske Systemer
Aalborg Universitetscenter
Fredrik Bajers Vej 7
9220 Aalborg

Fax: 98 15 81 29
Telefon: 98 15 85 22 lokal 5074
Direkte: 98 15 42 11 tryk 5074
Email: coco@iesd.auc.dk

NYT & NAVNE

Cand. stat. Per Tanghøj er pr. 1. januar 1992 udnævnt til afdelingsleder for Biostatistisk afdeling i H. Lundbeck A/S.

Det meddeles hermed, at Selskabets formand, Philip Hougaard, har fået en velskabt datter.

KALENDER

- SFE:** Seminar i anvendt statistik.
Afholdes (hvis ikke andet er angivet) i lok. 21.1.25a på Panum Institutet, Blegdamsvej 3, 2200 København N.
Arrangeres af: Statistisk Forskningsenhed, Københavns Universitet, Panum Institutet.
- KUIMS:** Seminar i matematisk statistik og sandsynlighedsregning.
Afholdes (hvis ikke andet er angivet) i aud. 8 på H.C. Ørsted Institutet, Universitetsparken 5, 2100 København Ø.
Arrangeres af: Institut for Matematisk Statistik, Københavns Universitet, H.C. Ørsted Institutet.
- DSTS:** Aftenmøde.
Afholdes (hvis ikke andet er angivet) i aud. X på H.C. Ørsted Institutet, Universitetsparken 5, 2100 København Ø.
Arrangeres af: Dansk Selskab for Teoretisk Statistik.
- KUFL:** Forsikringsmatematisk kollokvium.
Afholdes (hvis ikke andet er angivet) i aud. 8, H.C. Ørsted Institutet, Universitetsparken 5, 2100 København Ø.
Arrangeres af: Forsikringsmatematisk Laboratorium, Københavns Universitet, H.C. Ørsted Institutet.