ISSN 0803-8953

TILFELDIG GANG
Nr. 2, årgang 18
Mai 2001

Utgitt av Norsk Statistisk Forening

INNHOLD

	Side
Redaksjon.	1
Fra redaktøren. *Odd Aalen*	2
Fra leder i foreningen. *Geir Egil Eide*	3
Statistisk arbeidsliv	
Statistikk i økonomi og administrasjon: Del II. *Jostein Lillestøl*	4
Drømmejobben, statistisk poliklinikk og n=3. *Kathrine Frey Frøslie*	8
Litt personlig om tidsrekker. *Dag Tjøstheim*	11
Faglig stoff	
Bioinformatikk. *Mette Langaas*	14
Puslerispalten	
Pusleri nr. 7. *Jostein Lillestøl*	19
Mer fullstendig løsning av Pusleri nr. 5. *Jostein Lillestøl*	20
Løsninger Pusleri nr. 6. *Jostein Lillestøl*	22
Milepæl	
Pris til Bergensstatistikere.	23
Bo Lindqvist – 50 år. *Jon Tyssedal, Henning Omre og Stian Lydersen*	24
Leiv Solheim – 50 år. *Ib Thomsen*	25
Minnetale over Olav Reiersøl holdt i Det Norske Videnskaps-akademi. *Tore Schweder*	27
Minnetale ved Olav Reiersøls bisettelse. *Bent Natvig*	32
Møter, konferanser og annet	
Nytt om nettverk: NOREVENT. *Ellen Amundsen*	35
Det 11. norske statistikermøte.	36
NSF's etterutdanningsseminar 2001.	37
Medlemskap i Norsk statistisk forening.	38

Redaksjon:

Odd Aalen (redaktør)
Grete Usterud Fenstad (redaksjonssekretær)
Ingrid Glad
Nils Lid Hjort
Tron Anders Moger

Redaksjonens adresse:
Odd Aalen
Seksjon for medisinsk statistikk
Universitetet i Oslo
P.b. 1122 Blindern, 0317 Oslo

Tlf: 22 85 12 36
Fax: 22 85 13 13
e-mail: odd.aalen@basalmed.uio.no

Trykking og utsending av TG utføres av
Statistisk sentralbyrå

Bidrag til Tilfeldig Gang:

Alle slags bidrag til bladet er velkomne, enten det gjelder nye temaer av interesse for norske statistikere, eller man har kommentarer til innlegg som allerede har stått i bladet. Stoff til neste nummer må være redaksjonen i hende innen 1.9.2001.

Det greieste er at stoffet sendes inn som Word-fil til redaktørens e-mail-adresse, se ovenfor. Alternativt kan vi ta en ren tekst-fil, men vi vil helst ikke ha TEX-format med mindre dette er nødvendig på grunn av matematikken.

Vi har etablert en nettversjon av TG, se følgende internettadresse:
www.math.uio.no/nsf/web/index.htm . De som leverer bidrag til TG og som ønsker å reservere seg mot at bidraget deres legges ut på nettet, må gi beskjed om dette ved innsendelse av stoff.

Fra redaktøren:

Nok et nummer av TG er stablet på bena. Vi er takknemlige til alle som skriver for oss, og oppfordrer også folk til å sende inn innlegg uoppfordret.

Vi har nå en godt fungerende nettversjon: *www.math.uio.no/nsf/web/index.htm*. Muligheten for å legge ut bilder er naturligvis mye bedre på nettsidene enn i papirutgaven, og bilder er derfor velkomne selv om de ikke nødvendigvis vil bli trykt i papirversjonen.

Vi har denne gang to omtaler av femtiårsdager, for Bo Lindqvist og Leiv Solheim. Redaktøren benytter anledningen til å minne om at miljøene må være oppmerksomme når kolleger passerer slike milepæler og tipse TG-redaksjonen om det. Redaksjonen kan ikke selv ha oversikt over fødselsdager og andre anledninger.

En av fagets nestorer, professor Olav Reiersøl, er gått bort og Tore Schweder og Bent Natvig skriver omtaler av denne originale personligheten i norsk statistikk.

En viktig begivenhet er prisen til det medisinsk-statistiske miljøet i Bergen, se omtale i bladet. Vi gratulerer på det varmeste. Det er meget krevende å oppnå slik forskningsmessig anerkjennelse i et medisinsk fagmiljø.

Mette Langaas skriver en interessant artikkel om bioinformatikk. De avanserte biokjemiske og genetiske analyser som i dag er blitt mulige, resulterer i enorme mengder data som må håndteres. Bioinformatikken er vokst frem som et svar på dette. Mange av dens redskaper er veletablerte statistiske metoder, men faget oppfattes ikke som statistisk i sin natur. Mette etterlyser engasjement fra statistikere innen bioinformatikken. Et slikt engasjement vil sikkert være personlig meget utbytterikt og spennende. Det er også fagpolitisk viktig at et betydelig vekstfelt for statistisk metode og tankegang ikke blir stående utenfor vårt faglige virksomhetsområde.

Jostein Lillestøl forsetter sin utmerkede gjennomgang av statistikkfagets relasjon til den økonomisk-administrativt utdannelse. Med den posisjon denne utdannelsen har i dag er dette utvilsomt et viktig tema.

Vi fortsetter vår serie med personlig pregede betraktninger fra det statistiske arbeidsliv. Dag Tjøstheim skriver engasjerende om sin mangeårge virksomhet i analyse av tidsrekker; dette kan også leses som en fin populærvitenskapelig innføring på feltet. Kathrine Frey Frøslie representerer den unge garde av statistikere, og gir en spennende fremstilling av hvordan det er å jobbe innen det medisinske fagfeltet.

På vegne av redaksjonen ønsker jeg en riktig god sommer!

Odd Aalen

Fra leder i foreningen:

Kjære medlemmer.

En styreperiode går mot sin slutt og et nytt styre skal snart starte sitt arbeid. Et lite tilbakeblikk over perioden kan da kanskje være på sin plass. Hva kan vi egentlig forvente av et styre for Norsk statistisk forening, og hva har vi gjort? Styremedlemmene er alle frivillig valgte og ulønnete og har måttet passe sine vanlige jobber ved siden av, så selv om det ikke har stått på viljen til å prioritere foreningsarbeidet er det klart at dette kommer på toppen av en rekke andre oppgaver av faglig og profesjonell art. Dette gjør at selv om en har ideelle målsetninger om å ha en høyere aktivitet kan det ofte skorte på tiden en får til dette. Med en slik organisasjon blir det derfor viktig at mange er med å dra lasset. Det synes jeg det har vært: I tillegg til styret har vi bl.a. hatt et aktivt undervisningsutvalg, revisorer, arrangements- og programkomiteer for nordisk og norsk møte, lokalforeningenes styrer, TG-redaksjonen, samt bidragsytere til TG og NSF-forum m.fl. Dette har gjort at foreningsarbeidet har gått jevnt og trutt som det skulle, dog uten at vi i perioden har satt i gang store nye prosjekter. I retrospekt er kanskje de viktigste aktivitetene i perioden at vi har konsolidert arbeidet med foreningens hjemmesider og TG på nett, og at undervisningsutvalget har kommet godt i inngrep med fagplanene for den videregående skolen. Begge deler tror jeg er gode investeringer i fagets fremtid.

Et område der vi som forening ikke har vært særlige aktive er kontakt med andre nasjonale statistiske foreninger. Som leder får jeg tilsendt medlemsbladene fra den svenske og danske statistiske forening, men har ellers hatt liten kontakt med disse, så også med International Statistical Institute (ISI) som er en forening som søker å samle statistikere i alle land. - Når det gjelder de skandinavisk-talende land føler jeg at det kunne være nyttig med en bredere kontaktflate utenom de nordiske møtene. En mulig måte som en kunne diskutere internt og foreningene imellom, ville være en ordning med utvidete utsendingslister for de respektive medlemsbladene. F.eks. kunne en tenke seg at medlemskap i en nasjonal forening også innebar at en fikk tilsendt medlemsbladene til andre skandinavisk-talende land. På den måten ville annonsering av nasjonale kurs og konferanser, stillinger og faglige diskusjoner lett bli spredt over landegrensene innenfor vårt språkområde. Dette er en tanke som jeg sender videre til det nye styret. - Når det gjelder en utvidet internasjonal foreningsmessig kontakt er det kommet et initiativ fra ISI mht en bedre koordinering av dette. ISI vil som en start arrangere et eget møte i forbindelse med `53rd International Session of the ISI` i Korea i august i år. På den foreløpige sakslisten står bl.a. samarbeidet mellom ISI og de nasjonale foreninger, styrking av de eksisterende nasjonalforeninger f.eks. ved å etablere samarbeidsavtaler, støtte til statistikere i utviklingsland, utvikling av egne fagetiske regler for statistikkfaget, hvordan faget kan bidra til offisiell statistikk, godkjenningsordninger for tittelen `statistiker` og hvordan en skal kunne lage internasjonale standarder for å kvalitetssikre og oppmuntre statistisk arbeid og utdanning, og oppmuntre til videre profesjonell utvikling av faget. Alt i alt temaer av meget generell interesse for vårt fag. Her er det fritt fram for norske statistikere med interesse og initiativ til å aktivisere seg (se http://www.cbs.nl/isi/directory.htm).

Vennlig hilsen

Geir Egil Eide

Statistikk i økonomi og administrasjon: II

Jostein Lillestøl
Norges Handelshøyskole

I del I av Statistikk i økonomi og administrasjon i forrige nr. av TG forsøkte jeg å gi en oversikt over aktuelle områder innen økonomisk virksomhet, der en med fordel kan dra nytte av statistisk tenkning og statistiske metoder. I del II vil jeg drøfte hvordan det økonomiske fagmiljø oppfatter behovet, og hvordan vi statistikere har brukt og misbrukt den mulighet vi har hatt til å formidle vårt produkt, og videre hvordan det er å fungere som statistiker i et økonomisk-administrativt fagmiljø. I mitt tilfelle er det ved Norges Handelshøyskole, der vi i hovedsak utdanner siviløkonomer i et 4-årig studium, men også har studier frem til doktorgrad og mer profesjonsorienterte studier som Høyere revisorstudium.

De mange behov og problemstillinger av statistisk natur innen et foretak gjør det naturlig å hevde statistikkfagets selvstendige rolle i utdanninger for virke i næringslivet. Det er dessverre slik at næringslivets ledere i liten grad ser sine problemstillinger med statistiske briller, ellers ser potensialet i og verdien av statistisk kunnskap. Det samme gjelder i stor grad våre kolleger i økonomisk-administrative fag. Statistikk blir som regel sett på som et redskapsfag som har legitimitet bare i den utstrekning metodene blir brukt senere i studiet. Dette medfører ofte at faget kommer i skuddlinjen dersom studentene har problemer med det. Økonomene mener at det som er nødvendig, er relativt enkelt og at de selv kan ivareta det, fordi de er bedre til å gi emnene en relevant kontekst. Mange elementer i vårt fag fins i ulike økonomiske fag, og studentene kan like gjerne lære det der.

Jeg vil kort skissere den kontekst en statistiker ved et økonomisk-administrativt lærested vil befinne seg i. Det er viktig for å forstå utfordringene og velge aktuelle alliansepartnere.

De økonomisk-administrative fagene i et siviløkonomstudium vil i hovedsak være tre hovedblokker: bedriftsøkonomi, samfunnsøkonomi og organisasjons- og ledelsesfag. Bedriftsøkonomi, som mange betrakter som hovedkomponenten i studiet, omfatter flere delfag som betraktes som spesialiteter og der faglærerne i liten grad har ansvar for fagets bredde. Blant disse er regnskap, økonomisk styring og finans. Markedsanalyse og evt. produksjon ("operations management") tas også ofte med her. Samfunnsøkonomi omfatter generell mikroteori og makroteori, og også spesielle emner som internasjonal økonomi. Organisasjons- og ledelsesfag omfatter i hovedsak organisasjonsteori og strategi. Informasjonsbehandling kan også plasseres i denne båsen. Grensene mellom de tre hovedblokkene kan være vanskelig å trekke.

Den faglige bakgrunnen for de ansatte som dekker disse fagene på høyskolenivå er i hovedsak: For bedriftsøkonomi siviløkonomer med doktorgrad (hvis de er å få tak i), for samfunnsøkonomi sosialøkonomer eller tilsvarende, og for administrasjonsfag en blanding av siviløkonomer og cand.polit'er, med innslag av psykologer mm. De som dekker de ulike økonomiske fagene kan derfor ha ulik metodebakgrunn, og det er også utviklet egne, sterke metodetradisjoner innen hvert hovedområde som omfatter

til dels avanserte statistiske metoder som den realfaglig skolerte statistiker ikke nødvendigvis har godt kjennskap til. Eksempler på dette er samfunnsøkonomi, der økonomene regner økonometri som sitt genuine metodefag. Her starter en gjerne med den standard lineære regresjonsmodellen, men går raskt videre til de mange utfordrende problemstillinger som oppstår når standardforutsetningene brytes, og som er typiske for makroøkonomiske data og modeller: heteroscedastisitet, stokastisk regressor, tidsavhengighet, simultanitet, identifiserbarhet osv.

I markedsøkonomi som fag vil markedsundersøkelser være et sentralt tema. Her kan mange statistiske metoder finne praktisk anvendelse. Dette er også reflektert i mange gode amerikanske lærebøker som brukes i økonomistudier. Vi nevner estimering og testing i ett- og toutvalgssituasjoner (binomisk, normal, t) og i noen tilfeller også rangmetoder (Wilcoxon). Regresjonsanalyse og variansanalyse er med, og her har en gjerne med både en faktor og to faktor uten og med samspill. Forskjellen fra den realfaglige statistikers kontekst er at faktorene ikke er fast design-faktorer, men gruppetilhørigheten er bestemt ut fra utvalget. Enkelte bøker har også med den rangbaserte varianten (Kruskal-Wallis). I tillegg til de statistiske metodene vi er velkjent med, omtales gjerne metoder for datareduksjon og enkelte skaleringsmetoder, dvs. faktoranalyse, conjoint analyse mm. Disse bøkene tar opp hele prosessen fra planlegging av undersøkelser, gjennomføring av undersøkelser, tilrettelegging av data for analyse, metodevalg og gjennomføring av analyse. De ulike statistiske metodene blir derfor presentert i en kontekst. Studentene er kanskje bedre motivert for dem da, enn når en har et eget metodekurs i statistikk, der en gjennomgår f.eks. variansanalyse prinsipielt, selv om en med eksempler viser at det kan brukes både her og der.

Innen bedriftsøkonomi vil en i kurser i finans typisk omtale regresjon i forbindelse med kapitalmarkedsmodeller (CAPM), kanskje i en form der generalitet ikke sperrer for forståelse, men som derfor har umiddelbar overføringsverdi til andre områder. Regresjon kan også være omtalt i forbindelse med økonomisk styring og "operations management" (OM), men gjerne i en "myk" variant som ikke skiller mellom teori og empiri. I OM tar en også opp prognoseproblemet for tidsrekker, for eksempel salgstall. Metoder som tar omsyn til trend og sesong er som regel nok, dvs. en nøyer seg med enkel dekomponering og /eller eksponensiell glatting etter Holt-Winters metode. En legger da stor vekt på at prognosene skal kunne oppdateres for hver ny periode ettersom nye data kommer inn.

Enkelte fagmiljøer og studieprogrammer i økonomi benytter begrepet "Managerial Economics". Et kurs med tilhørende lærebok i Managerial Economics vil typisk inneholde en rekke temaer av statistisk karakter, bl.a. regresjon, og noen vil med en viss rett kunne hevde at en her langt på vei lærer statistikk i kontekst, og at et behovet for eget metodekurs i statistikk undervist av statistikere er overflødig. Et blikk på enkelte slike lærebøker kan gi en statistiker følelsen at "de stjeler vårt fag".

I de senere år er det påfallende at cand. polit'er har fått en sterkere stilling ved mange læresteder, tilsynelatende også innen de bedriftsøkonomiske fag. Dette skyldes neppe endringer i næringslivets behov, men mer at genuine bedriftsøkonomer er vanskelig å rekruttere til utdanningssystemet, mens cand. polit kanskje har et sysselsettingsproblem. I praksis har dette medført at studiene flere steder har blitt mindre kvantitativt orienterte og at metodeorienteringer har en tendens til å gå i retning av samfunnsvitenskapelig metode, istedenfor "realfaglig" modellorientert metode. Administrasjonsfagene føler seg, dels pga. stabens bakgrunn, nærmere knyttet til samfunnsvitenskapelig metode enn

til (norske) statistikeres metodesyn. De synes mere om en mykere variant, der en ikke blir plaget utidig med modeller og skiller mellom teori og empiri og flisespikkeri som parameter, estimator og estimat. De føler også at Stevens tanker om hvert målenivå sin metode gjør verden enklere for dem. Har du måledata, så er det t-test uavhengig av observasjonenes fordeling!

Det er etter hvert blitt en større utfordring å definere og rettferdiggjøre sin rolle som statistiker i studiet. Man forsøker å definere en kjerne av stoff de andre kan og bør bygge på, men opplever mange skuffelser. Mange holder seg ikke orientert om hva studentene har hatt før, gjentar det samme, bruker "dummere" metoder enn det studentene allerede har lært, eller simpelthen unngår kvantitative utfordringer.
Regresjon er kanskje den statistiske metode de fleste kan og vil kunne gjøre bruk av. Det er selvsagt uhensiktsmessig at alle gir en innføring i dette i hver sin kontekst, med fare for dobbeltdekning og ulik tilnærming. Det er imidlertid lite å slå i bordet med dersom dette blir et hovedargument for å ha en egen metodeseksjon.

Et problem er at metodekurs i statistikk oftest gis (og trengs?) første studieår, mens studentene som regel først møter anvendelser senere i studiet, kanskje først siste studieår. Ved NHH er det obligatoriske metodekurs i statistikk første studieår sammenlignbart med mange 2-4 vekttalls kurs ved våre universiteter og andre norske høyskoler. Det omfatter sannsynlighetsregning, stokastiske variable, fordelinger, forventning, varians og normaltilnærming. Videre gis en innføring i statistisk inferens med de generelle ideene knyttet til estimering og hypotesetesting. Dette anvendes i ett-utvalgssituasjoner (binomisk, normal) og i enkel regresjon. Stoffet kommer typisk i nevnte rekkefølge. Noen emner er tatt med som ikke nødvendigvis er med i tilsvarende pensa i økonomiutdanninger ved andre norske høyskoler, bl.a. utvalg fra endelige populasjoner, Poissonmodeller og styrkebetraktninger. Kurset følges opp senere i studiet av et obligatorisk kurs med hovedsakelig statistisk innhold i tredje studieår, som vi skal komme tilbake til nedenfor.

Ovennevnte kursinnhold er trolig modent for revisjon, noe som ofte kommer til uttrykk ved konferanser der undervisningen for "business"-studenter er tema (bl.a. ICOTS). Noen legger vekt på at regresjon bør komme først, fordi det er mest sentralt for økonomiske anvendelser, og at relevante problemer kan stilles med en gang. Det bør ikke komme etter omveien via sannsynlighetsregning og generell inferensteori. Andre mener at tidsrekker er den mest typiske datatypen i økonomi, og at det er et paradoks at dette først kommer i forlengelsen av det statistiske metodekurset. Andre igjen hevder med styrke at det relevante perspektiv for undervisningen må være "å forstå variasjon" og lære metoder for å innrette seg bedre ved variasjon, i mange tilfeller å redusere uønsket variasjon. I dag fins det gode amerikanske lærebøker i "business statistics" som ivaretar hvert av de ulike synspunktene ovenfor. En bok som snur opp ned på vante forestillinger om kursinnhold og rekkefølge har innholdslisten: Grafiske metoder, Prosesstyring og styringsdiagrammer, Regresjon, Tidsrekker, Forsøksplanlegging, Analyse av forskjeller. Det er først i dette siste kapitlet at vårt kjære begrepsapparat om hypotesetesting nevnes, og forfatteren marginaliserer dermed bevisst dette, vel vitende om at tiden som regel blir knapp.

Som nevnt har NHH et oppfølgende kurs i siviløkonomstudiets annen avdeling i tredje studieår. Høyskolen har opptak til annen avdeling fra de to-årige studiene ved de statlige høyskolene, og kurset starter derfor med en kort repetisjon og viderefører

det generelle teorigrunnlag på en del punkter, bl.a. kontinuerlige fordelinger. De statistiske emnene er to-utvalgsproblemet (binomisk, normal t), variansanalyse med en og to faktorer (med og uten samspill), multippel regresjon (der avvik fra standardforutsetningene vies en viss oppmerksomhet) og analyse av kategoridata (kontingenstabeller og logistisk regresjon). I denne sammenheng blir tokning av årsaksforhold viet en viss oppmerksomhet. Kurset omfatter videre faktoranalyse, en kort introduksjon til tidsrekkeanalyse (ARMA) og litt om økonometriske metodeproblemer. Metodevalg er tema for en egen dobbeltforelesning. Det samme er "Å forstå variasjon". der nytten av styringsdiagrammer ("control charts") blir vist. I dette kurset tar en sikte på integrasjon med de økonomiske fag som tas samme studieår: bedriftsøkonomi (finans), markedsundersøkelser, "operations management" (hos oss kalt vare- og tjenesteproduksjon).

Obligatoriske metodefag vil typisk omfatte forelesninger og øvelser for mange studenter, ved min institusjon er hvert kull på mer enn 400 studenter. Den letteste vei er tradisjonelle forelesninger, regneøvelser og der en står til rette bare ved en eksamen ved semesterslutt. Vårt fag innbyr til mer prosjektorientert undervisning, med tilbakemeldinger underveis. Med lav bemanning, er det vanskelig å være særlig innovativ. Vår undervisning har hittil være teori illustrert med pseudoeksempler og pseudoproblemer. Det er dels fordi vi selv ikke kjenner godt til de virkelige problemstillinger, og selv om vi gjorde det, vil studentene ikke ha tilstrekkelig bakgrunn for å diskutere problemer i en kontekst før etter et par års økonomistudier. Dette kan muligens bli bedre med de endringer som nå er under planlegging ved NHH, hvor studentene i større grad skal gjennomføre prosjekter med evaluering underveis.

I kurset Anvendt metode går planen ut på å ha en obligatorisk del som omfatter ovennevnte emner til og med regresjonsanalyse. Videre skal studentene i grupper på 5 velge et prosjekt blant 5-10 mulige, alt etter hva en tenker å satse på som fordypning i tredje avdeling (siste studieår). Hvert prosjekt starter med en introduksjonsforelesning og det tilbys også valgfrie emner som passer til ulike profiler og prosjekter . Hver student må velge 2 emner blant 6 valgemner som hvert foreleses over 3 timer. En student som er interessert i markedsanalyse vil derfor kunne velge prosjekt og spesialemner iht. dette, f.eks. ved å ta med faktoranalyse og utelate tidsrekker. Dette opplegget er ambisiøst, og krever deltakelse fra mange, og det er med spenning og litt frykt jeg ser frem til endringen.

Når det gjelder forskning, er de nevnte tre områdene bedriftsøkonomi, samfunnsøkonomi og organisasjon/ledelse stort sett brukere av eksisterende statistisk teori, men det foregår stadig generell teoriutvikling som er inspirert av disse områdene. I mange tilfeller er det kommet nye ideer som ikke stammer fra oss matematisk statistikere. I det siste tiåret er stokastiske modeller i matematisk finans kommet sterkt i fokus, med til dels avanserte modeller, f.eks. prisutvikling basert på diffusjon/hopp prosesser. Disse er bl.a. grunnlag for prising av opsjoner, der en kan bruke martingalteori. Parallelt med dette har fagfeltet empirisk finans utviklet seg, der ulike ikke-lineære modeller har fått en sentral rolle. Det gjelder i første rekke empirisk analyse av volatilitet, som er sentralt for prising av opsjoner. Koblingen med matematisk finans gir her mange spennende utfordringer for den matematisk skolerte statistiker. Ved NHH har vi nå tre professorer med statistikk hovedfag, men to av dem er ansatt i bedriftsøkonomiske stillinger. "If you can't beat them, join them!" .

Arbeidslivsrefleksjoner:

Drømmejobben, statistisk poliklinikk og n=3.

Kathrine Frey Frøslie
Senter for epidemiologi og sykehusstatistikk
Rikshospitalet

Den første jobben

Etter atskillige ST001-grupper og orakeltimer (og enda flere røde og svarte kuler opp og ned av urner), var det både befriende og skummelt å plutselig sitte med virkelige data mellom hendene på Folkehelsa. Legene og psykologene rundt meg var elskverdigheten selv, men det plaget meg litt at jeg ikke kunne diskutere logistisk regresjon, konfundering og interaksjon med den selvfølgeligheten de fremviste. Ivrig etter å komme på banen i diskusjonen, børstet jeg støv av ST202-permen min. Det første jeg støtte på var Nils' Oxford-notater ("Hillary term 1993") om minimering av Kullback-Leibler-avstanden for å vise at ML-estimatet gjør så godt det kan. Avstanden mellom mine forelesningsnotater og det praktiske behovet har også ved senere anledninger vist seg å være signifikant.

Fersk

Statistikk skiller seg ikke fra andre fag når det gjelder behovet for opplæring av nyutdannede i sin første jobb. Jeg mener allikevel at studiet burde vært lagt opp på en bedre måte når det gjelder å motivere og forberede til praktisk problemløsning. Det er tross alt dette de fleste av oss arbeider med etter studiet. En nyutdannet må selvsagt være lydhør overfor sine seniorer på jobben, men det er et stort gap mellom dette og det å føle seg forlegen fordi andre yrkesgrupper tilsynelatende kan ditt eget fag bedre enn deg selv. Jeg håper og tror at den pågående debatten om og omleggingen av studiet kan bedre på dette. Når alt kommer til stykket, kan jeg ikke se at det skulle være noen motsetning i det å gi studenter en grunnleggende magefølelse for anvendelser og det å gi dem et godt teoretisk grunnlag for å forstå metodene som brukes.

Drømmejobben

Ett års engasjement på Folkehelsa med en relativt bratt læringskurve ble etterfulgt av selveste drømmejobben. Stedet jeg arbeider på ligger sentralt til, er splitter nytt, var eksepsjonelt dyrt å bygge, har helikopterlandingsplass, basseng, trimrom, utkikkstårn, en av Norges største kunstsamlinger og helt sikkert flere toaletter enn hytta til Røkke. Sånn har vi det i staten, nærmere bestemt på Rikshospitalets Senter for epidemiologi og sykehusstatistikk. Senteret er et resultat av Rikshospitalets satsing på egen forskning og har vært operativt siden høsten 1996. Vår funksjon på senteret er å tilby hjelp og veiledning til alle som driver med forskningsprosjekter på Rikshospitalet. "Vi" er meg selv, Thore Egeland (statistiker), Akhtar Hussain (epidemiolog), og Stein Vaaler (lege). På sett og vis har jeg fortsatt med det jeg likte best på Blindern: Undervisning og orakeltjeneste. Klientellet og problemstillingene er imidlertid fullstendig forandret. Hundrevis av ST001-studenter med broket bakgrunn må sies å være en homogen gruppe i forhold til dem som nå kommer og spør om statistikk. Det som kanskje varierte mest hos ST001-studentene – deres motivasjon for statistikk – er kanskje det som er det mest homogent hos dem jeg møter i dag.

De som kommer til oss med statistiske spørsmål er leger, sykepleiere, kontoransatte, fysioterapeuter – mildt sagt en svært sammensatt gruppe. Felles for dem er at de er engasjerte og entusiastiske fagpersoner og forteller mer enn gjerne om sine spesialiteter. Jeg har hørt ivrige psykologer legge ut om irritabel tarm, hudlegene Mørk, Mørk og Mørk (ja, noen av dem er i slekt) har fortalt om behandlingsreiser til syden, og barneleger har dosert om effekten av tran under svangerskapet. Jeg har vært engasjert i prosjekter om uforklarlig fosterdød, om optimal lagringstemperatur for kaninhornhinner, og om klinisk vurdering av barn med epilepsi – spennvidden i problemstillingene er enorm. Det er også de ulike klientenes behov for statistisk veiledning. Mens noen har nok med å få forklart forskjellen på gjennomsnitt og median, har andre behov for å kjøre post-hoc-tester for interaksjonsledd i varianskomponentmodeller.

Timebestilling eller statistisk poliklinikk

De fleste henvender seg til oss per telefon. Ofte med innledningen " Hei, jeg har et lite statistisk spørsmål. Det tar bare et par minutter." Dette stemmer sjelden. Uklarheter omkring problemstillinger, studiedesign og medisinsk tolkning gjør at "et lite statistisk spørsmål" oftest er vanskelig å besvare over telefon. Alle som ønsker bistand fra oss må derfor bestille time ved å fylle ut et skjema med bakgrunnsinformasjon om sin problemstilling. Hver uke fordeler vi nye prosjekter oss i mellom og tar kontakt med dem som ønsker hjelp. Vanligvis vil det være 1 til 2 ukers ventetid på en veiledningstime. Den utålmodige med enklere statistiske problemer og et akutt behov for hjelp kan alternativt møte opp på vår *statistiske poliklinikk* på tirsdager.

Vi veileder i alt fra valg av studiedesign og diskusjon av datainnsamling, beregning av utvalgsstørrelse og protokollskriving, til valg av og forutsetninger for analysemetoder, analyse og fremstilling ved SPSS, etc. Vi gjør imidlertid ikke det *praktiske* arbeidet, som innleggelse av rådata i databaser og selve analysen av materialet. Det er den enkelte lege eller sykepleier som står ansvarlig for den rapporten/artikkelen som skrives, og de må derfor både utføre analysen og skrive om resultatene selv. Vi gir derimot så mange veiledningstimer som er nødvendig for den enkelte.

I noen utvalgte prosjekter – enten fordi analysen krever det, eller fordi vi har faglige interesser på området, går vi tyngre inn i prosjektdeltakelsen og gjør analysearbeidet og skriver statistikkdelen i rapporten/artikkelen. I slike tilfeller er vi også medforfattere når arbeidet publiseres.

Thore på kontoret.

De fleste ønsker en jobb i et miljø med flere statistikere. Jeg har vært så heldig at jeg ikke bare jobber sammen med, men jeg deler også kontor med en særdeles erfaren og hyggelig statistiker. Thore Egeland sitter ca 2 meter unna og svarer tålmodig på alle mine funderinger. Til gjengjeld tar jeg telefonen når han ikke er til stede, og alfabetiserer alle hyllemeterne med statistikkbøker. (Forfatteren Sesamgata var ny for meg. Tittelen på Sesamgatas rosa bok var like mystisk: "Bernt og Elling får en kattunge." Sesamgata-boken ble til slutt arkivert ved siden av "Næhei! Sa Albert Åberg".)

Assisterende underordnet reservelege

Den mest fascinerende fenomenet her på Rikshospitalet er i mine øyne den utpregede rangordningen mellom og innad i(!) de ulike yrkesgruppene. Tittelen i overskriften er litt overdrevet, men det *heter* faktisk "underordnet lege"! For en stakkars statistiker er det lett å gå i surr i overleger, avdelingssykepleiere, oversykepleiere og avdelingsoverleger. Jeg

fornærmet faktisk en av disse såpass for en stund siden at vedkommende måtte presisere at psykosomatisk ble skrevet med p. Jeg hadde svært lyst til å skrive sykopsomatisk. Men om ikke tittelen forteller hvem som er høyest på rangstigen, taler som oftest kroppsspråket sitt utvetydige språk. Så har vi da noe å forholde oss til. (Et spørsmål til TG-leserne: Hvilket kroppsspråk har egentlig en statistiker?)

Utfordringer

En viktig forutsetning for å trives på jobb er å møte utfordringer. Ethvert nytt prosjekt som kommer til oss bidrar i så måte til dette. Statistikere, sykepleiere, psykologer og leger snakker ikke alltid samme språk. Hva er det egentlig de mener? Hvorfor ønsker man i utgangspunktet å sette i gang denne studien? Å finne ut hva som er problemstillingen kan være en utfordring i seg selv. En statistiker vil ofte problematisere andre aspekter enn for eksempel leger. Når problemstillingen så er fastslått, kommer vi til studiedesign og utvalg. Jo, et randomisert utvalg er flott, men hva hvis det ikke finnes noe sentralt pasientregister? Skal vi bare plukke ut pasienter fra Rikshospitalet? Eller inkluderer vi pasienter fra andre sykehus? Hvem skal i så fall gjøre det? Og er det ikke slik at kun de sykeste havner på sykehus? Skal vi kanskje gå til pasientorganisasjonen i stedet? Hvem er registrert der? Hvor mange må vi rekruttere i studien? Et av mine første oppdrag var å beregne utvalgsstørrelse for å sammenligne to medisiner på en gitt type pasienter. Med de effektene legen opererte med, fant jeg ut at han måtte ha minst 76 pasienter i hver gruppe. Dette knegget han godt over, for hvis han var "heldig", fikk han 10-15 nye pasienter per år. Fokus for studien ble endret.

Om spørsmålene omkring studiedesign og utvalg kan være innfløkte, så har vi i disse tilfellene en viss kontroll over hva som er mulig å gjøre etter at dataene er samlet inn. Verre er det med de studiene der vi kontaktes etter at dataene er samlet inn. Vi kan for eksempel bli presentert for svært skjeve utvalg, et frafall på 70%, eller et design så innfløkt at vi ikke har gode metoder til å analysere det. Vel – arbeidet er gjort og dataene er samlet inn. Vi må gjøre det beste ut av det. En stipendiat kom inn og fortalte om en teori han og veilederen ville undersøke. Det dreide seg om blodsukkernivå hos en gitt pasientgruppe. For å gjennomføre målingene, måtte pasientene være på sykehuset i tre døgn, og det ble tatt prøver av dem hver fjerde time. Pasientgruppen var i utgangspunktet liten, og de som ble undersøkt kunne ikke gå på medisiner fra før. Han fikk 4 frivillige – én trakk seg i løpet av studien. Men resultatene var tydelige. Tittelen "pilotstudie", deskriptiv statistikk og tydelige grafer ble løsningen.

Personlig synes jeg det er disse aspektene som gjør faget interessant. Som bonus og tilvalgsstoff kommer etiske spørsmål, alle de rare sykdommene, snedige forsøk og de hvitkledte folkene rundt oss. Mye mer kunne sies, både om undervisning, samarbeid mellom de ulike medisinsk-statistiske miljøene og om spennende prosjekter.
Enn så lenge: Ha en riktig fin vår og sommer!

Litt personlig om tidsrekker

Dag Tjøstheim
Matematisk institutt, Universitetet i Bergen

Av redaktøren er jeg bedt om å skrive, gjerne litt personlig, om tidsrekker.

Mitt første møte med virkelige tidsrekker skjedde på NORSAR, Kjeller. Dit kom jeg i 1974 som nybakt Ph.D. i Applied Mathematics fra Princeton University. Der hadde jeg blant annet strevd med grupper av uniformt begrensete shift operatorer i Hilbert rommet, så overgangen til seismiske tidsrekker ved jordskjelvsobservatoriet på Kjeller ble brå. Tidsrekkene i dette tilfelle bestod av seismografregistreringer (vanligvis 10 målinger pr sekund), og oppgaven var å gjøre mønstergjenkjenning på seismiske signaler med formål å kunne skille mellom jordskjelv og underjordiske kjernefysiske eksperimenter.

Denne typen gjenkjenningsproblemer for tidsrekker påtreffes på flere andre fagområder, mest kjent er kanskje talegjenkjenningsproblemet. Jeg tilpasset vanlige autoregressive modeller til de seismiske signalene og brukte de autoregressive parametrene i en Fisher-type diskriminant. Det fungerte noenlunde bra og resulterte i noen publikasjoner i seismologiske tidsskrifter. Det var litt frustrerende for en som hadde mest lyst å arbeide med matematisk og statistisk teori, å måtte jobbe på heltid med computeranalyser av virkelige seismiske signaler, men det var særdeles lærerikt. Det arbeides annerledes innen anvendt forskning enn i grunnforskning, og jeg ville ikke ha vært foruten de impulsene jeg mottok på Kjeller. Egentlig burde enhver med planer om en universitetskarriere sørge for å få seg et par år ute i felten før en lar seg suge inn i universitetsmiljøet. Selv var jeg der i 2-3 år før jeg fikk jobb som dosent ved NHH og siden som professor ved Universitetet i Bergen. Også mitt opphold ved NHH betød svært mye for min faglige utvikling. Jeg kom i kontakt med Jostein Lillestøl og gjennom ham med økonomiske problemstillinger og økonometri, som har vært en hovedinteresse siden.

Da jeg begynte mitt yrkesliv som tidsrekkeperson, var det fremdeles ARMA modeller som var det sentrale, også innen forskning. Dette er nå totalt forandret. ARMA modeller brukes fremdeles mye i anvendelser, men i forskning er det neppe noe særlig mer å utrette. Interessen har forlengst skiftet til ikke-lineær og ikke-parametrisk modellering, og en er opptatt av ikke-stasjonære fenomener og av modeller som beskriver utvikling av varians over tid i motsetning til bare forventning. I det følgende skal jeg gi en kort beskrivelse av noen generelle forskningstemaer.

Ikke-linearitet

Ikke-lineære modeller kom for alvor på banen i 80-årene. Det er nå skrevet flere bøker om disse. Howell Tongs bok er vel den mest kjente. Han innførte terskelmodellene, som impliserer at en lineær modell erstattes av en stykkevis lineær modell. Liknende modeller hadde lenge vært kjent innen ikke-lineære differensiallikninger. Når overgangen mellom de forskjellige lineære regimene er glatt, fås STAR (smooth transition autoregressive processes) klassen, som er fleksibel, og som

blant andre Timo Teräsvirta har brukt på makroøkonomiske serier. I en tidligere fase var også bilineære og eksponensielt autoregressive modeller gjenstand for en viss interesse, men spesielt de bilineære er for tiden forsvunnet ut over sidelinjen. Det er fremdeles mange utfordrende problemer på området. Ett av dem er å avgjøre når en gitt ikke-lineær modell er stasjonær. Til dette kan en ofte bruke et Liapunov-funksjonskriterium, men problemet har vist seg vanskelig, og det har vært liten framgang å spore. Det er ikke en gang kjent for hvilke parameterverdier en annen ordens terskelmodell er stasjonær. Det er nylig utviklet en hybrid mellom ikke-lineære tidsrekke modeller og kaos. Selv er jeg negativ til dette som forskningsområde, og det er klart at dersom en har stokastisk støy, kan det være vanskelig å snakke om kaos i vanlig forstand.

Ikke-parametrikk

Ikke-parametrisk estimering av tetthet og betinget forventning går tilbake til Murray Rosenblatts arbeider midt på femtitallet. En viktig milepæl innen tids-rekkesammenheng er Peter Robinsons arbeid i 1983. Han påviste der at ikke-parametrisk kjernefunksjonsestimering kan gjøres matematisk rigorøs for tidsrekker, og de siste årene har ikke-parametriske teknikker vært svært mye brukt. Hvis en har relativt få datapunkter, kan ikke-parametrisk estimering være nyttig til å identifisere en parametrisk modell. Har en overflod av data (tenk for eksempel på sanntidsdata på børsen), kan de ikke-parametriske estimatene danne avslutningen for analysen. En viktig begrensning på ikke-parametrisk estimering generelt -- også innen tidsrekker -- er den såkalte "curse of dimensionality". De funksjoner som en estimerer, kan ikke ha et for høyt dimensjonalt definisjonsområde, da en ville måtte ha astronomisk mange datapunkter for å få til et godt estimat. Derfor sier man vanligvis stopp når dimensjonen kommer opp i 3-4. Det finnes imidlertid forskjellige måter å unngå dette problemet om en er villig til å gjøre noen ekstra forutsetninger. Den mest vanlige er å bruke såkalte additive modeller, der en generell ikke-lineær regresjonsfunksjon av flere variable byttes ut med en sum av funksjoner som hver er en funksjon kun av en enkel regressor. Tradisjonelt har backfitting vært brukt til å estimere additive modeller, men nå er marginal integrasjon kommet inn som et alternativ. Selv om den siste ikke er effisient, er asymptotisk teori forholdsvis enkel å utvikle, og metoden er ganske robust. Det bør nevnes at det de siste par årene er blitt utviklet en asymptotisk teori også for backfitting.

Betinget varians og volatilitet

En viktig hensikt med tidsrekkemodellering har vært å lage prognoser. Den beste prediktor ved bruk av minste kvadraters kriteriet er som kjent den betingete forventning, og svært mye av tidsrekkemetodikk har vært rettet mot en beskrivelse av denne, som altså reduserer seg til en lineær prediktor for Gaussiske ARMA modeller. For finansdata har dette vært av relativt liten nytte, da man ofte etter å ha tatt log differenser sitter igjen med noe som likner hvit støy, og som det derfor ikke kan sies så mye nyttig om via den betingete forventning. For denne type data har en funnet det mer meningsfyllt å konstrure prognoser av den betingete varians. Etter introduksjonen av ARCH modellene av Rob Engle midt på åtti-tallet har denne utviklingen virkelig skutt fart, og det er sikkert skrevet over tusen artikler om modellering av betinget varians og prognoser av volatilitet (tidsvariarende standardavvik). Det er klart et dette er viktig, da volatilitet uttrykker økonomisk risiko. Denne typen analyser er også benyttet i opsjonsprising, for eksempel ved Black and Scholes

formel, da volatilitet inngår som en avgjørende parameter, som må estimeres, i formelen. Avansert statistisk verktøy er tatt i bruk på området. Avhengig av angrepspunktet er det tilknytning til MCMC, faktoranalyse, lang hukommelse modeller og så videre.

Ikke-stasjonaritet

En tidsrekke er aldri helt stasjonær, men ofte er tilnærmingen god nok til at en stajonær modell kan brukes. Men ikke alltid; makroøkonomiske tidsrekker er vanligvis sterkt ikke-stasjonære. Den vanligste teknikken i slike tilfeller er å prøve å finne lineære kombinasjoner av komponentene i en multivariat tidsrekke som gir stasjonaritet for de transformerte rekkene. Dette kalles for kointegrasjon, og teknikken ble foreslått av Clive Granger for ca. 15 år siden. Nå er ikke alle relasjoner mellom økonomiske variable nødvendigvis lineære, og et temmelig åpent felt er å konstruere en teori for ikke-lineær kointegrasjon. Dette er vanskelig for det krever en estimeringsteori der begreper som mixing ikke kan benyttes. En mulig vei kan gå via null rekurrente Markov kjeder. Ved å bruke en Markov split kjede teknikk basert på rekurrens teori kan for eksempel en asymptotisk teori for ikke-parametrisk estimering i en nullrekurrent situasjon utvikles. Mange av de lineære kointegrasjonsmodeller kan formuleres som spesialtilfeller av nullrekurrente kjeder. En alternativ angrepsmåte er å bruke begreper som lokal stasjonaritet, forfektet blant andre av Rainer Dahlhaus.

Avsluttende kommentarer

Jeg var nylig i Bremen, og var der med på evalueringen av et "Schwerpunkt" program initiert av det tyske forskningsråd. Jeg fikk her min tidsrekkehorisont betydelig utvidet. Det foregår for tiden en utstrakt forsking i Tyskland -- og helt sikkert andre steder -- på biomedisinske tidsrekker. Disse synes alltid å være ikke-stasjonære og kommer ofte som et panel av tidsrekker. Nye teknikker brukes og kreves, men mye mangler på at de er skikkelig forstått. Jeg har også vært heldig det siste året å ha en bistilling på Havforskningsinstituttet. Her er det et eldorado for en tidsrekkeperson med et vell av utfordrende problemer innen ressursestimering og biologiske modeller, som blant annet Tore Schweder har pekt på i et tidligere innlegg i TG. Nok et område som jeg har vært engasjert i nylig er elektrisitetsmarkeder, der en har mye tidsrekkedata, og der en ønsker å ta i bruk finansielle modeller for prising av kraftkontrakter. Men det er store forskjeller fra tradisjonelle finansmarkeder.

Det er nok å ta seg til!

BIOINFORMATIKK – ET INTERESSANT FORSKNINGSFELT FOR STATISTIKERE?

METTE LANGAAS, NORSK REGNESENTRAL
Mette.Langaas@nr.no

HVA ER BIOINFORMATIKK?

Bioinformatikk er et tverrfaglig forskningsfelt der matematikk, statistikk og informatikk anvendes til å analysere data som er produsert av eksperimentelt arbeid innen biokjemi, cellebiologi og genetikk[1].

I en lederartikkel i tidsskriftet *Bioinformatics* foreslår Professor Russ Altman at pensum for en utdanning innen bioinformatikk bør deles inn i fem områder; biologi, informatikk, statistikk, etikk og "bioinformatiske kjernefag". Innen biologi bør det legges vekt på molekylærbiologi, cellebiologi og genetikk. Den statistiske delen av utdanningen bør inneholde kurs i sannsynlighetsteori, statistisk forsøksplanlegging, statistisk analyse og stokastiske prosesser. Spesielt blir optimering, dynamisk programmering, søkealgoritmer, klusteranalyse, klassifikasjon, nevrale nettverk, genetiske algoritmer og Bayesiansk inferens nevnt som metoder som ofte benyttes innen bioinformatikk. Innen informatikk er programmering, datastrukturer, algoritmer og håndtering av databaser grunnleggende. Bioinformatiske kjernefag består av blant annet av design og implementasjon av biologiske databaser, annotasjon av biologiske sekvenser, metoder som støtter laboratorie-arbeid (derunder DNA arrays), modellering og prediksjon av proteinstruktur, metoder for sammenligning av biologiske sekvenser (sequence alignment), Hidden Markov Models og fylogeniske trær.

I Norge finnes hovedfagsutdanning innen bioinformatikk ved Gruppe for bioinformatikk ved Institutt for informatikk ved Universitetet i Bergen og Gruppe for bioinformatikk ved Institutt for informatikk ved Universitetet i Oslo. Et nytt hovedfagsstudium innen bioinformatikk er planlagt opprettet fra 2002 ved Norges Landbrukshøgskole i Ås (naturvitenskapelige fag). Ved Institutt for datateknikk og informasjonsvitenskap (faggruppe for kunnskapssystemer) ved Norges Teknisk-Naturvitenskapelige Universitet i Trondheim arbeider man med bioinformatiske problemstillinger. Ved Norsk Regnesentral arbeider vi med statistiske aspekter av bioinformatikk, og planlegger en økt satsning på feltet fremover. Vi samarbeider med "Mikromatriseprosjektet på Radiumhospitalet" og ønsker å opprette ytterlig samarbeid med akademiske og kommersielle aktører innen bioinformatikk i Norge og i utlandet.

GENOM OG GEN

Kroppen vår består av mer enn 10^{12} celler. Inne i hver celle finner vi organeller som er ansvarlige for energiproduksjon, proteinsyntese og lagring og videreføring av genetisk informasjon til nye celler. Den totale genetiske informasjonen kalles *genomet*. Alle cellene i kroppen vår inneholder den samme genetiske informasjonen, men kun deler av denne informasjonen er aktiv i en gitt celle. Deler av informasjonen som er aktiv i en hjernecelle er inaktiv (sovende) i en levercelle og vice versa. Hvilken del av den genetiske informasjonen som er aktiv i en celle bestemmer hvordan cellen fungerer.

Det humane genomet er fordelt på 46 kromosomer, der hvert kromosom inneholder et molekyl deoksyribonukleinsyre (DNA). Byggesteinene i DNA-molekylet er nukleotide-enheter (sukkeret deoksyribose, en fosfatgruppe og en base), der fire mulige baser inngår[2]. DNA er laget som en dobbel helix, der to kjeder av nukleotider er festet sammen med hydrogenbindinger mellom basene[3]. Man har anslått at det humane genom består av 3 milliarder basepar. Et gen er et segment av et DNA molekyl, lokalisert i en spesifikk posisjon på et spesifikt kromosom. De aller fleste genene inneholder oppskriften på et protein (enzymer er proteiner som er viktige i produksjonen av ulike molekyler). Et viktig mellomprodukt i syntesen av proteiner heter mRNA (messenger RNA). Hvis vi ønsker å se om

---

[1] Denne definisjonen er i bruk ved Chalmers Universitet. Mange ulike definisjoner av bioinformatikk finnes; en meget bred definisjon er "bioinformatikk er å bruke informasjonsteknologi til å forstå biologi". Et relatert begrep er "computational biology".

[2] Basene heter G=guanine, C=cytosine, T=thymine og A=adenine

[3] Bindingene dannes mellom basene G og C og mellom T og A. Dette kalles basepar.

et gen er aktivt i en celle (*genekspresjon*) kan vi kvantifisere mengden av proteinet som genet koder for i denne cellen. Det er mer komplekst å måle proteinmengden enn mRNA mengden i en celle, og målinger av mRNA mengde er i dag ofte brukt når man ønsker å måle genekspresjon.

DET HUMANE GENOM PROSJEKT

Kartlegging av genomet vil si å identifisere posisjonen og DNA-sekvensen til hvert gen. Den komplette DNA-sekvensen for mange genomer[4] er kjent, mens sekvensiering av genomet til mange planter og dyr pågår. "Det humane genom prosjekt" (HGP) ble startet i 1990 under ledelse av The National Center for Human Genome Research i USA og har som mål å kartlegge hele det humane genomet. Et konkurrerende privat selskap, Celera Genomics, er blitt dannet med det samme målet. De to selskapene har benyttet to ulike teknikker for å sekvensiere genomet. I februar 2001 publiserte HGP sine resultater i Nature og Celera Genomics sine resultater i Science. Antall gener ble der anslått til å være rundt 30 000 (et mye mindre tall enn tidligere antatt). Kart over plasseringen av de ulike genene på de ulike kromosomene, og informasjon om egenskapene til det ulike genene er tilgjengelige fra HGPs WWW-sider[5]. Informasjonen legges ut kontinuerlig og kan benyttes fritt.

ULIKE FORSKNINGSOMRÅDER INNENFOR BIOINFORMATIKK

Ved at genomet til mange organismer nå er ferdig kartlagt, har vi beveget oss fra den pre-genome æra til den post-genome æra. Den viktigste utfordringen i den post-genome æra er å utforske funksjonen til genomet, dvs. rollen til hvert gen. Dette kalles *funksjonell genomikk*, og vil være hovedtemaet for resten av denne presentasjonen.

Det finnes imidlertid mange andre forskningsområder innen bioinformatikk, og mye av det eksperimentelle arbeidet og metodegrunnlaget har felles komponenter innen de ulike områdene. Noen av de mest profilerte områdene er som følger. Innen *proteomikk* ser man på hvilke proteiner som er tilstede i en celle[6] og hvordan de ulike proteinene samvirker. I *strukturell genomikk* ønsker man å bestemme den tredimensjonale strukturen til proteiner som er kodet i genomet. Hvordan ulike genetiske forskjeller påvirker en pasients respons på medisiner er hovedspørsmålet innen *pharmacogenetikk*. I *komparativ genomikk* ser man på funksjonen til menneskets gener og andre områder av menneskets DNA ved å studere deres paralleller i andre organismer (f.eks. mus).

FUNKSJONELL GENOMIKK: GENEKSPRESJON OG DATA FRA DNA MIKROARRAYS

Hvordan en celle fungerer kan generelt ikke beskrives ved aktiviteten til ett gen, den oppstår som et resultat av samspill mellom mange gener. Måling av aktiviteten til flere tusen gener samtidig kan danne grunnlaget for økt forståelse av ulike biologiske prosesser. Før ble ett og ett gen studert separat. Nye teknikker, som DNA mikroarrays, gjør det mulig å studere rollen til tusenvis av gener samtidig og dermed kunne avdekke komplekse samspill mellom gener.

Det finnes to hovedformater DNA mikroarrays; oligonukleotide- og cDNA mikroarray. Det mest brukte DNA mikroarray formatet i den akademiske forskningen er cDNA mikroarray for analyse av genekspresjon (se figur 1). I et cDNA mikroarray eksperiment inngår en glassplate, der små dråper av gener (kalt prober) er avsatt ved hjelp av en mikroarray-robot[7], og to ulike celleprøver som er bearbeidet og merket[8]. Antall gener som avsettes er i størrelsesorden 2000-10000, men i fremtiden vil glassplaten kunne inneholde hele det humane genomet. Genekspresjonen til de to celleprøvene anslås ved at celleprøvene reagerer (hybridiserer) med genene på glassplaten. Ved hjelp av en laserscanner avbildes glassplaten ved to ulike bølgelengder (jmf. merkingen). Resultatet er to gråtonebilder. Ved hjelp av bildeanalyse gjøres de to bildene om til et datasett med en intensitet for hvert gen for hver

---

[4] Blant annet gjær, bakterien E.coli, ormen C.elegans, bananflue.

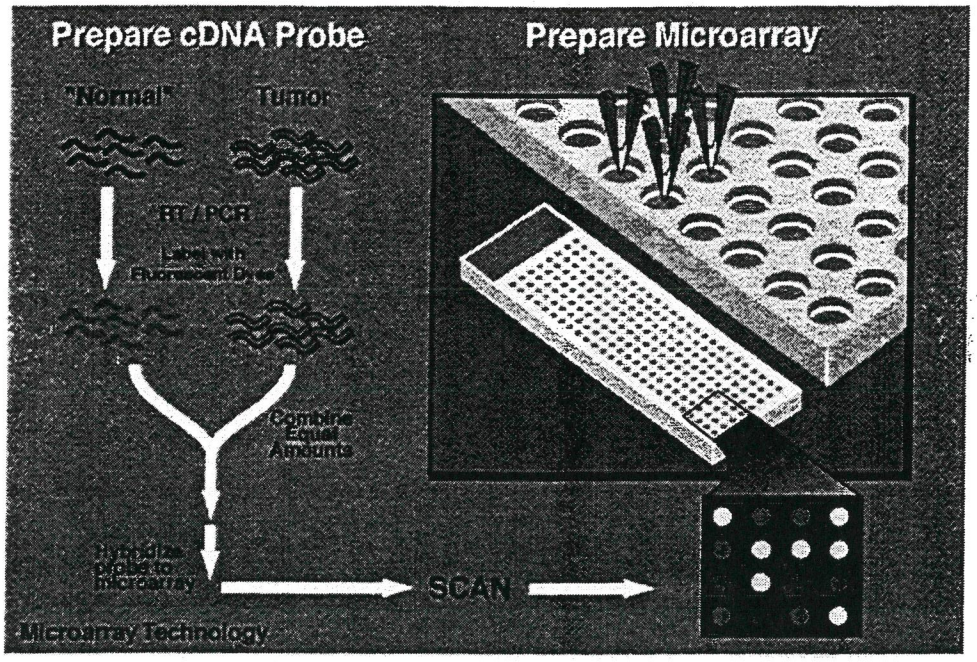
[5] http://www.ornl.gov/hgmis/

[6] Man har anslått at mennesket kan produsere i størrelsesorden 80000 ulike proteiner.

[7] I Norge finnes det to mikroarray-roboter; en ved NTNU i Trondheim og en ved Det Norske Radiumhospital. Minst to nye roboter er planlagt bygget.

[8] Vi tenker oss her at prøvene er merket med to typer fluorescent fargestoff (rødt og grønt). De to fargestoffene emitterer lys når de blir belyst med lys av to ulike bølgelengder. Farging med fluorescent fargestoff er den mest populære merketeknikken.

Figur 1: Et cDNA mikroarray eksperiment. Figuren er tatt fra HGPs sider http://www.ornl.gov/hgmis/



celleprøve. Intensitetene er anslag for mengden av mRNA assosiert med de ulike genene i de to celleprøvene (dvs. hvor mye et gen er "slått-på", aktivt, eller "slått-av", inaktivt").

Grunnen til at to celleprøver inngår i et cDNA mikroarray eksperiment er at det er vanskelig å kontrollere nøyaktig mengden av probematerial som avsettes i hvert område på mikroarrayen (probeområdene er kalt spots). For hver spot studeres derfor vanligvis ratioen mellom intensitetene for de to celleprøvene. Ønsker man å undersøke mer enn to celleprøver er det mest brukte forsøksdesignet et såkalt referansedesign, der en av celleprøvene defineres som en referanse. Hvis man har n celleprøver, inngår en av celleprøvene i alle cDNA mikroarray eksperimentene; dvs. man bruker (n - 1) mikroarray eksperimenter til å undersøke n celleprøver. F.eks. i et forsøk der man vil undersøke effekten av ulike doser av radioaktiv stråling på en kreftsvulst, kan referansen være en celleprøve som ikke har vært utsatt for stråling. Et cDNA mikroarray datasett inneholder typisk 2000 til 10000 gener (variable) og 2 til 100 celleprøver (observasjoner).

STATISTISKE METODER FOR Å ANALYSERE GENEKSPRESJONS-DATA

Teknikker som DNA mikroarrays produserer store datamengder. I tabell 1 har vi identifisert noen problemstillinger der statistisk kompetanse er viktig for å utnytte funksjonell genomikk data maksimalt. Dette kommenteres kort nedenfor.

La oss tenke oss at vi ønsker å undersøke hvilken effekt ulike doser av radioaktiv stråling har på ulike gener i kreftsvulster. F.eks. kan vi ønske å finne grupper av gener som har reagerer likt på strålingen. Det første vi må bestemme er hvordan cDNA mikroarray eksperimentene skal utføres. Det vanligste valget er å bruke et referansedesign, men ved hjelp av statistisk forsøksplanlegging kan vi finne andre design som kan være mer effektive[9]. Etter at eksperimentene er utført brukes bildeanalyse til å bestemme intensiteten til genene fra gråtonebildene. I bildeanalysen inngår segmentering, identifikasjon av spots og lokal bakgrunn, og man beregner intensiteter korrigert for at celleprøvene har reagert utenfor probeområdene på glassplaten. Det er mange kilder tilfeldige og systematiske feil i

[9] Logikken ved å benytte en referanse-celleprøve er å kunne gjøre indirekte sammenligninger, men introduksjon av en referanse er ineffektivt fordi halvparten av dataene samles inn fra denne referansen. Et referansedesign vil være et ufullstendig blokk-design, da referanse-celleprøven merkes likt i alle eksperimentene. Men, det er praktiske biologiske fordeler ved å bruke et referansedesign og forsøksplanlegging krever tett samarbeid mellom statistiker og biolog.

et cDNA mikroarray eksperiment. Benyttes de to mest populære merkestoffene (grønn og rød fluorescent farge) inkorporeres disse ulikt i celleprøvene og kan føre til at man for den ene celleprøven systematisk observerer en høyere intensitet enn for den andre. Korreksjon av denne type feil kalles i DNA mikroarrays for "normalisering". Observasjon av intensiteten til noen gener kan være manglende i noen eksperimenter. Det kan f.eks. skyldes at en eller annen form for forurensning (f.eks. et hårstrå) har kommet på glassplaten. Metoder for imputering av manglende data kan være nyttig.

Den mest populære statistiske metoden brukt på DNA mikroarray data er klusteranalyse. Her ønsker man å finne grupper av gener og/eller grupper av celleprøver som har lignende egenskaper. Hierarkisk klustring har vært mye benyttet i publiserte analyser av DNA mikroarray data, og det har også blitt utviklet nye statistiske metoder for to-veis klustring (gener og celleprøver samtidig). Klustring av celleprøver har i tillegg vært benyttet for å finne mulige undergrupper av sykdommer, f.eks. for noen kreftpasienter vil en behandling virke dårlig og en forklaring kan være at det finnes undergrupper av sykdommen som responderer ulikt på en gitt behandling. Har man celleprøver fra pasienter med forskjellige sykdomsdiagnoser, kan man bruke DNA mikroarray data til å finne gener som har ulik ekspresjon for de ulike sykdomsdiagnosene (diskriminantanalyse). Man kan også bruke disse genene videre til å lage en regel for å diagnostisere nye pasienter (klassifikasjon). En relatert oppgave er å finne et sett av gener som er uttrykt forskjellig i to eller flere celleprøver. Her har man i litteraturen sett på ulike måter å utføre multiple tester på. Data fra slike eksperimenter kan også analyseres med grunnlag i en variansanalysemodell (ANOVA), der hovedeffekten av gen, merking (rød og grønn), glassplate, stråledose og blant annet samspill mellom glassplate og gen (spot) inngår.

Det ligger store utfordringer i å utnytte data fra DNA mikroarray-eksperimenter effektiv. Disse datasettene er såkalte "stor p liten n" datasett, dvs. antall gener er mye større enn antall celleprøver. En liste av artikler som tar opp problemstillingene skissert her, kan finnes hos Rockefeller University [10] og Aas (2001) gir en bred oversikt over metoder som er blitt brukt til analyse av DNA mikroarrays. Det er ikke mange *metodiske* statistiske artikler som er publisert innen analyse av DNA mikroarrays. Grunnleggende spørsmål om kvaliteten til og reproduserbarheten av DNA mikroarray eksperimenter er i liten grad behandlet statistisk.

PROBLEM	STATISTISK METODE
Planlegge design av eksperiment	Statistisk forsøksplanlegging
Beregne genekspresjon fra gråtonebilde	Bildeanalyse
Håndtere systematiske feil (normalisering)	Generell statistisk analyse, variansanalyse, glatting
Håndtere manglende data	Analyse av manglende data
Finne gener eller celleprøver med lignende egenskaper	Klusteranalyse
Diskriminere mellom to eller flere grupper av celleprøver	Diskriminantanalyse
Klassifisere en ny celleprøve til en av mange klasser (eller beregne klassesannsynligheter)	Klassifikasjon
Finne gener som er forskjellig uttrykt i ulike celleprøver	Multippel testing, egenskapsutvelgelse

Tabell 1: Problemstillinger og tilhørende statistiske metoder for å analysere genekspresjonsdata.

HVORDAN KAN STATISTIKERE BIDRA?

Med noen få unntak ligger Norge langt etter genforskningen i utlandet. Det finnes i Norge i dag meget begrenset med forskningsmidler til arbeid innen bioinformatikk og genforskning generelt. Et samlet forsknings-Norge har laget en nasjonal plan for funksjonell genomforskning, forkortet til FUGE. FUGE-planen[11] ble presentert i januar 2001. Forskningsmiljøene ber om minst 300 millioner kroner årlig i de neste fem til ti årene.

[10] http://linkage.rockefeller.edu/wli/microarray/

[11] Les mer om FUGE-planen hos Norges Forskningsråd http://www.forskningsradet.no/fag/andre/fuge/

For å oppnå de beste resultatene innen funksjonell genomikk og bioinformatikk er det nødvendig med et tett samarbeid mellom forskere med statistisk (og bioinformatisk) bakgrunn og forskere innen genetikk, medisin og biologi. Behovet for statistisk tenkning er tilstede i alle faser av funksjonell genomforskning; fra planlegging av eksperimenter til analyser av eksperimentelle resultater. Det er et stort behov for statistiske analyser basert på klassisk metodikk og for utvikling av nye statistiske metoder for å imøtekomme spesielle datatyper og store datamengder.

BIOINFORMATIKK – ET INTERESSANT FORSKNINGSFELT FOR STATISTIKERE?

For å oppsummere; bioinformatikk er et fagfelt som er bygget på viktige biologiske og medisinske problemstillinger. Fagfeltet er nytt og svært aktivt. Hyppige teknologiske nyvinninger introduserer stadig nye utfordringer innen statistisk dataanalyse (datakvalitet og datamengde). Innen fagområdene genetikk, molekylærbiologi, biokjemi og cellebiologi er det et stort behov for statistisk tenkning og statistisk ekspertise. Kort sagt, det er store og mange statistiske utfordringer innen bioinformatikk, og bioinformatikk er uten tvil et meget interessant forskningsfelt for statistikere.

REFERANSER OG KILDER TIL MER INFORMASJON:

FRENGEN OG PRYDZ (2000):"Det humane genom prosjekt og kreft." Tilgjengelig fra Den Norske Kreftforenings WWW-sider på http://www.kreft.no/share/osas/cache/artid200244.html

THE HUMAN GENOME PROJECT (1996): "To know ourselves". Tilgjengelig fra http://www.ornl.gov/hgmis/publicat/tko/index.html/

AAS (2001): "Microarray Data Mining; A Survey", NR notat SAMBA/02/01, tilgjengelig fra http://www.nr.no/samba/bioinformatics.html

OVERSIKT OVER ARTIKLER OM ANALYSE AV DATA FRA MIKROARRAYS VED ROCKEFELLER UNIVERSITY, http://linkage.rockefeller.edu/wli/microarray/

GRUPPE FOR BIOINFORMATIKK VED INSTITUTT FOR INFORMATIKK VED UNIVERSITETET I OSLO, http://www.ifi.uio.no/bioinf/

MIKROMATRISEPROSJEKTET PÅ RADIUMHOSPITALET http://www.med.uio.no/microarray/

EN SAMLING BIOINFORMATIKK-LENKER FRA FORSKNINGSGRUPPEN FOR BIOINFORMATIKK VED UNIVERSITETET I BERGEN http://www.ii.uib.no/forskningsgrupper/bio/lenker/index-eng.shtml

GENOMIC CLASSIFIER RESEARCH VED NTNU http://www.idi.ntnu.no/grupper/KS-grp/GCR/

TIDSSKRIFT FRA BIOTEKNOLOGINEMNDA, WWW-sider http://www.bion.no

STATISTICAL METHODS IN BIOINFORMATICS, WWW-sider hos Norsk Regnesentral http://www.nr.no/samba/bioinformatics.html

Puslerispalten

Ved Jostein Lillestøl

Pusleri nr 7

Et spill består av gjentatte terningkast med gevinst sum øyne, men dersom du får en ener, er spillet slutt og du har tapt alt.

Hva er den optimale strategi for å maksimere forventet gevinst og hva blir den forventede gevinst når denne følges?

Hva blir sannsynlighetsfordelingen til gevinsten?

De som finner det siste spørsmålet for vanskelig eller for tidkrevende, kan skriv en makro som simulerer trekninger fra fordelingen. Hvem klarer å skrive makroen (f.eks. i Minitab) med færrest antall programlinjer?

Problemstillingen kan sikkert belyses på ulike måter, og *også denne gangen ønsker vi gode løsninger fra lesere for presentasjon i neste TG*, fra den enkle og intuitive til den matematisk sofistikerte under ulike antakelser. Send løsninger innen 4 uker fra mottatt TG til jostein.lillestol@nhh.no.

Løsninger Pusleri nr. 5

Løsningen i forrige nummer var dessverre ufullstendig, og en mer fullstendig løsning presenteres her. Vi minner om oppgaven:

Her er tre myntkastproblemer:

1. Hvordan kan du generere 50:50 sjanse med myntkast med en skjev mynt?

2. Hva er sannsynligheten for at KK forekommer i n rettferdige myntkast?

3. Hva er sannsynligheten for at KMK forekommer i n rettferdige myntkast?

Oppgave 1 (løsning):

Den enkleste løsningen på oppgaven er trolig følgende: Kast den skjeve mynten to ganger. Hvis utfallene er like, dvs. KK eller MM, se bort fra resultatet. Hvis utfallene er ulike, dvs. KM eller MK bruk utfallet i det første kastet. Hvis en ønsker en følge av uavhengige utfall med P(K)=P(M)=1/2, kan en praktisk generere to følger med den skjeve mynten, stille de opp under hverandre og sammenligne de to utfallene i samme posisjon. Hvis de matcher, stryk begge, hvis de er forskjellige, bruk utfallet i den første sekvensen. Eksempelvis:

M M K M K K K M K K M K M M K M.....
M K M K K K M K K M K M M K K M.....

Her blir den rettferdige sekvensen sammensatt av de som er understreket i første linje.

Et annet forslag er: Generer en følge med den skjeve mynten, f.eks. slik at vi får

M,M,K,M,K,K,K,M,K,K,M,K,M,M,K,M,..... (1)

Grunnet symmetri kan vi snu annenhver mynt, slik at (1) blir til

M,K,K,K,K,M,K,K,K,M,M,M,M,K,K,K,..... (2)

som kan betraktes som trekninger fra en rettferdig mynt.

Det er en hake ved dette resonnement, nemlig at P(K)=P(M)=1/2 bare gjelder i stasjonærsituasjonen, som vanskelig kan realiseres uten en rettferdighetsmekanisme, f.eks. ved å neglisjere et jamt eller odde antall kast i starten med lik sannsynlighet.

Men selv i stasjonærsituasjonen blir dette ikke uavhengige realisasjoner. Vi får nemlig P(KK)=P(MM)=p(1-p) og P(KM)=P(MK)=(p^2+(1-p)^2)/2, mens uavhengighet skulle gi 1/4. Dette forslaget er imidlertid et godt utgangspunkt for å diskutere sentrale begreper, som uavhengighet, Markovkjede, stasjonæritet etc.

Takk til Øystein Arild for innspill.

Oppgave 2 (løsning):

La g_n være sannsynligheten for at KK forekommer i n kast. Ved å betinge mhp. utfallet av første kast får vi

$$g_n = \frac{1}{2}g_{n-1} + \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}g_{n-2}\right) = \frac{1}{2}g_{n-1} + \frac{1}{4}g_{n-2} + \frac{1}{4} \qquad g_0 = g_1 = 0$$

Denne inhomogene differensligning kan løses på vanlig måte med løsning av form

$$g_n = 1 + A_1 r_1^n + A_2 r_2^2$$

der r er røttene i ligningen $r^2 - \frac{1}{2}r - \frac{1}{4} = 0$.

Dette gir

$$g_n = 1 - \frac{1}{2}\left(1 + \frac{3}{\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{4}\right)^n - \frac{1}{2}\left(1 - \frac{3}{\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{4}\right)^n$$

Det fins en lettere vei:

Sett $f_{n+2} = 2^n - 2g_n$. Da blir

$$f_{n+2} = f_{n+1} + f_n \qquad f_2 = 1 \quad f_3 = 2$$

som er den velkjente ligning som gir Fibonacci-tallene, som er tabellert bl.a. i Råde & Westergren: Mathematics Handbook.

Oppgave 3 (løsning): Har noen en vakker løsning, så kom med den. Min er det ikke! Fra Grete U. Fenstad er mottatt: Som en mulig oppfølging av Pusleri nr.5 har den læreboka* vi bruker i ST114 Modellering ved stokastiske prosesser et par oppgaver som sammenligner forventet antall kast før henholdsvis KKM og KMK oppnås. Det er kanskje litt overraskende at KKM gjennomsnittlig oppnås raskere enn KMK, uansett hva P(K)=p er. Ved en rettferdig mynt er svarene henholdsvis 8 og 10. Men resultatet kan forklares.

* Lærebokforfatterne er Howard M. Taylor og Samuel Karlin, og oppgavene det refereres til er Problems 3.5 og 4.1 (s.114 og s.130) i Kapittel III.

Løsninger Pusleri nr. 6

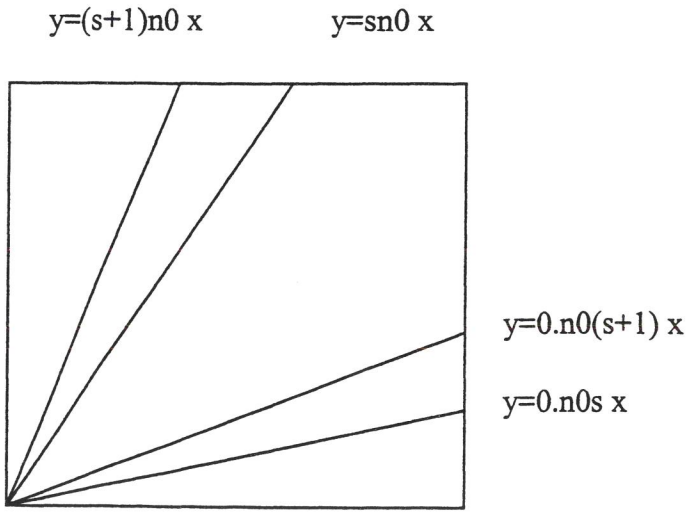
Vi minner om oppgaven:

Hva er sannsynligheten for at forholdet mellom to uavhengige R[0, 1] variable har et ett-tall som første gjeldende siffer?
Hva blir sannsynlighetsfordelingen til første gjeldende siffer?
De som er innesnødd i påsken kan løse den tilsvarende oppgaven med produkt, sum, differens, ...

Løsning:
Sannsynligheten for at første gjeldende siffer i kvotienten er 1 er lik 1/3!
Mer generelt er sannsynlighetsfordelingen til første gjeldende siffer:

$$p(s) = \frac{5}{9}\left(\frac{1}{10} + \frac{1}{s(s+1)}\right) \qquad \text{for s=1,2,...,9.}$$

Det er to hovedtilfeller: kvotient mindre enn 1 eller større enn 1. Problemet er redusert til å beregne og addere arealet av et uendelig antall trekanter av hvert slag. I figuren nedenfor er illustrert en trekant av hvert slag, der symbolikken n0 betyr n nuller. Den første trekanten med kvotient mindre enn 1 (for n=0) har areal lik 0.1 x 0.5 uansett s. Hver ny trekant av dette slaget har areal 1/10 av det forrige. Den geometriske rekken som fremkommer har sum 10/9, slik at det samlede arealet av disse trekantene fås ved å multiplisere med denne faktoren. Den første trekanten med kvotient større enn 1 (for n=0) har areal (1/s-1/(s+1)) x 0.5=1/s(s+1) x 0.5. Igjen har hver ny trekant areal 1/10 av den forrige, og vi får den samme geometriske rekken som ovenfor og samme faktor 10/9 å multiplisere med.



Statistikere fikk pris for beste vitenskapelige artikkel ved Det medisinske fakultet, UiB

Professorene Rolv Skjærven og Rolv T. Lie fikk i september 2000 pris av et medisinske fakultet ved Universitetet i Bergen. De fikk prisen for en artikkel i New England Journal of Medicine (Skjærven R, Wilcox AJ, Lie RT. A population-based study of survival and childbearing among female subjects with birth defects and the risk of recurrence in their children. N Engl J Med 1999 340:1057-62).

I denne epidemiologiske artikkelen følger forfatterne hele kohorten av jenter født i Norge mellom 1967 og 1982 fram til 1998. Spesielt fokus settes på jentene som ble født med misdannelser, og arbeidet beskriver deres overlevelse, deres sjanser for selv å få barn og risikoen for at deres egne barn skal bli født med misdannelser. Detaljer gies for hele 24 forskjellige kategorier av misdannelser.

Jenter med misdannelser hadde noe reduserte sjanser for å overleve til voksen alder. De hadde også en vel 30% redusert sjanse for selv å få barn. Ikke uventet ble det funnet en 1,6 ganger øket risiko for at disse barna ble født med en misdannelse.

Et hovedfunn var at hele denne overrisikoen kunne tilskrives en øket risiko for samme type misdannelse som moren selv var født med. Risikoen for andre misdannelser var ikke øket.

Et annet interessant funn var at mindre enn 1% av misdannelsene i andre generasjon i studien kunne tilskrives at mødre var bærere av en misdannelse. Dette kaster et underlig lys over tiltak som for eksempel sterilisering som ble aktivt brukt i enkelte land midt i forrige århundre for å redusere forekomsten av misdannelser i befolkningen.

Prisen ble delt ut av dekanus Rolf Reed. Dette var første gang fakultetet delte ut en slik pris, men det er meningen at dette skal bli en tradisjon som skal inspirere fakultetets forskere og synliggjøre god forskning.

Bo Lindqvist – 50 år

En svært sentral aktør i statistikkmiljøet i Trondheim fyller et halvt hundre år 7. mai. Bo er Skiensgutt utdannet fra Universitetet i Oslo der han ble cand. real i 1975 og senere dr. philos i 1982. Tittelen på avhandlingen var: ''On the memory of a Markov Chain. – A decision theoretical approach ''.

Bo lærte sine grunnleggende statistikkferdigheter i Oslo, men han var en av dem som skjønte at det fantes både muligheter og interessante utfordringer utenfor hovedstadsmiljøet. I 1979 tok han imot tilbudet om stilling som 1. amanuensis ved Institutt for matematisk statistikk, NTH og har fra 1988 vært professor i statistikk ved Institutt for matematiske fag, NTH. Alt fra starten av fattet han interesse for Sikkerhet og pålitelighet, et da oppblomstrende fagområde ved NTH og i SINTEF, og har i hele tiden etterpå hatt en fremtredende rolle innenfor dette fagmiljøet både som forsker og som rådgiver. De senere årene har han også interessert seg for problemstillinger innenfor medisin.

Studieretningen for Industriell matematikk ble startet opp av de matematiske miljøer ved NTH i 1982, og med den det som vel kan kalles et lite statistikk eventyr i teknologihovedstaden. Bo ble straks en av de mest populære veilederne på studieretningen, og i årenes løp har rundt regnet et halvt hundre hovedfags- og doktorgradsstudenter vandret fornøyde til og fra Bo sitt kontor.

Bo's arbeid ved instituttet har alltid vært sterkt preget av kvalitet. Forelesningene får de beste skussmål. I møter, komiteer og utval er det alltid en styrke å ha med Bo. En spesiell ære er blitt han til del i 2002. Da er han sammen med Kjell Doksum valgt til chairman for den tredje internasjonale konferansen om "Mathematical methods in Reliability" som blir arrangert i Trondheim i tidsrommet 17-20 juni.

Selv om Bo forlengst har funnet seg til rette i Trøndelag har vi ikke unngått å legge merke til at han vet hvorfra han har sine røtter. Det var en stor begivenhet da fotballklubben Odd rykket opp i eliteserien. For Bo betydde det at han kunne forflytte seg fra en kummerlig tilskuertilværelse i løpebanen på Dalgård til mer komfortable sittetribuner på Lerkendal for å iaktta sine fotballhelter.

Hjertelig til lykke med de runde år og de beste ønsker for fremtiden fra kolleger i Trondheim.

John Tyssedal, Henning Omre og Stian Lydersen.

Leiv Solheim – 50 år 24. april 2001

Min første kontakt med Leiv var å lese noen arbeider han hadde levert inn i forbindelse med at han søkte en stilling ved en Høgskole. Det ble beundring ved første blikk! Det som gjorde mest inntrykk på meg var hans evne til å formidle vanskelige statistiske begreper og metoder på norsk uten overdreven bruk av formler, og belyse dem ved praktiske eksempler. Dette var i "Head-huntingens" barndom, og jeg kom fort til at her var et "hode" som det var verdt å "jagte" på. Dessverre lyktes det meg ikke å få Leiv til Statistisk Sentralbyrå i første omgang, men noen tid senere tok han til min store glede kontakt for å høre om det var mulig å få jobb i SSB. Det var det, og 1. november 1987 startet han som forsker i Statistisk Sentralbyrå. Han ble etterhvert så begeistret for jobben at han senere kritiserte meg for ikke å ha mast tilstrekkelig på ham tidligere.

SSB var ikke Leivs første arbeidsplass. Etter endt cand.real. eksamen ved Universitetet i Oslo 1979 var han amanuensis ved Distriktshøgskolen i Telemark og senere konsulent ved NAVF's Utredningsinstitutt. Deretter gikk turen til Universitetet i Tromsø, et opphold som har satt betydelig preg på Leivs utvikling, både faglig og menneskelig. Den nordnorske kultur og natur gjorde et sterkt inntrykk og han ble for første gang introdusert til Likelihoodbegrepet. Likevel valgte han å reise sørover igjen etter 5 år som stipendiat og amanuensis i Tromsø. Først til Krigsskolen som Høgskolelektor og deretter som forsker ved Veterinærhøgskolen.

Det var altså ingen "novise" som startet i SSB høsten 1987, noe som fort ble klart for hans nye medarbeidere. Leiv kastet seg over en lang rekke metodeproblemer innen mange felter. Han viste en uvanlig vilje og evne til å sette seg inn i praktiske problemer og løse dem i samarbeid med sine kolleger på fagsiden. Som hans sjef fikk jeg mange klare signaler om hvor stor pris mange medarbeidere satte på hans råd og veiledning. Etterhvert konsentrerte han seg om statistiske metoder for produksjon av økonomisk statistikk, et meget vanskeligt og lite behandlet emne. I løpet av et par år ble han norsk mester på dette feltet, og kunne ha blitt europamester, -noe som fortsatt ikke er for seint!

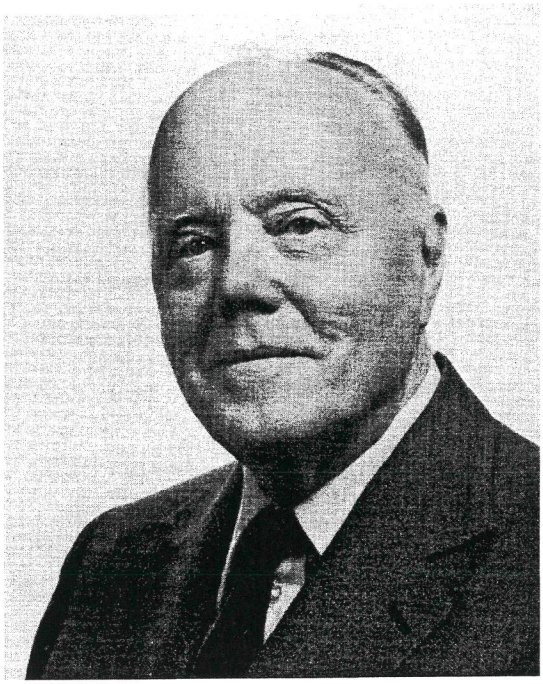
I 1994 ble Leiv utnevnt til Seksjonsleder for Seksjonen for Standarder og Statistiske Metoder. Etterhvert innså han at det var som forsker han ønsket å arbeide, og klarte å overtale sin venn og lærer fra Tromsø, Jan Bjørnstad til å overta stillingen som Seksjonsleder.

Jeg har lært Leiv å kjenne som et menneske med uvanlig stor spennvidde. Han er født odelsgutt, men overlot "gård og grunn" til sin bror. Han fant imidlertid tre mindre håndfaste, men like forpliktende holdepunkter i livet sitt: Familien, Ibsen og Likelihoodbegrepet. Dette har gjort ham til noe som jeg vil kalle en jordnær himmelstormer! Hans krav til empirisk forskning er like strenge som Ibsens moralske krav. Dette har til tider brakt ham i konflikt med en del samfunnsforskere, som ofte er

fornøyd med å ha forklart 20% av variasjonen i en variabel, eller som formulerer hypotesen etter å ha sett i data for deretter begeistret konkludere med at funnet er statistisk signifikant. Imidlertid er det ett trekk ved Ibsen som ikke synes å ha inspirert Leiv, nemlig Ibsens behov for å reise rundt i Europa for å hente inspirasjon. Leiv liker seg i Norge og føler lite behov for å reise lenger sør enn til Tønsberg. Mitt ønske for ham i de kommende år er at han litt oftere besøker oss her i Europas "statistiske sentralbyrå" og tar opp kampen om det europamesterskap som han kunne ha tatt for noen år siden.

Velkommen til Luxembourg!

Luxembourg 3.5.2001
Ib Thomsen



Olav Reiersøl

MINNETALE OVER PROFESSOR FIL. DR. OLAV REIERSØL holdt i Det Norske Videnskaps-akademi, den matematisk-naturvitenskapelige klasses møte den 5. april 2001 av Professor Tore Schweder

Olav Reiersøl døde 14. februar 2001, 92 år gammel. Med Olav Reiersøl er den siste av våre pionerer i matematisk statistikk og økonometri borte.

Reiersøl var født 28. juni 1908 i Krødsherad av Olava Kristine og Ole J. Reiersøl. Faren var folkeskolelærer. På hans todelte folkeskole fikk Olav god anledning til å forfølge sine matematiske interesser. Før han var 16 år gammel hadde han på egen hånd arbeidet seg gjennom pensumet i matematikk til realartium. I 1928 tok han artium på reallinjen ved Oslo katedralskole. Reiersøl studerte matematikk, fysikk, kjemi og botanikk ved Universitetet i Oslo og ble cand. real. med matematikk hovedfag i 1935.

Fra 1936 arbeidet Reiersøl som assistent for Ragnar Frisch ved Sosialøkonomisk institutt. I 1938 ble han "lærer i matematisk statistikk med sannsynlighetsregning og forsikringsmatematikk" samme sted. I 1941 kom Reiersøl med i produksjon og omdeling av den illegale avisen Avantgarden. Utover i 1943 arbeidet han med sin doktoravhandling, og mindre illegalt. Allikevel ble han bedt om å rømme til Sverige i oktober dette året da to av hans illegale kontakter ble tatt. Han fikk pålegg om ikke å ta på mye klær, og rømte tynnkledd. Dette førte til ørebetennelse og nedsatt hørsel resten av livet. I Stockholm fikk Reiersøl arbeidsplass ved Harald Cramérs Institutt for forsikringsmatematikk og matematisk statistikk. Han disputerte her i 1945 med avhandlingen Confluence analysis by means of instrumental sets of variables.

Reiersøl tok opp sitt arbeid ved Sosialøkonomisk institutt i 1945. Våren 1946 besøkte han Cambridge hvor han fikk arbeidsrom ved R.A. Fishers genetiske institutt. Høsten 1946 fikk Reiersøl stipendium fra Rockefeller Foundation og dro til Columbia University i New York. Etter opphold ved University of North Carolina, Cowles Commission i Chicago, Berkeley og Purdue University, kom han hjem til stilling som universitetsstipendiat i 1949. I 1950 ble Reiersøl dosent i matematikk ved Universitetet i Oslo, professor fra 1958. I 1961 tiltrådte han et nyopprettet professorat i matematisk statistikk. I 1964/65 hadde Reiersøl forskningstermin og var visiting professor ved Cornell University. Han gikk av med pensjon i 1975.

Reiersøl var medlem av Det Norske Videnskaps-Akademi fra 1951. Han var medlem av International Statistical Institute og Fellow både av Econometric Society og Institute of Mathematical Statistics.

Ved siden av matematikken og statistikken interesserte Reiersøl seg for esperanto, faktisk helt fra ungdommen. Han har skrevet en rekke artikler på esperanto, mange om matematiske og statistiske emner. Den siste kom ut i 1988. Disse arbeidene kan ikke omtales her. Reiersøl underviste i esperanto på Universitetet, noe direktøren likte dårlig. Han ønsket ikke at et kurs Reiersøl skulle holde i esperanto skulle føres opp i studiekatalogen. Kurset ble ført opp - fordi Reiersøl pekte på sin *jus docendi*.

Politisk sto Olav Reiersøl til venstre, og han var med på å starte Sosialistisk Folkeparti. Sitt frisinn og sin nøkternhet viste han ved mange anledninger. I 1961 lagde han sammen med Jan Erlvang, på eget initiativ, en prognose for studenttallet ved Universitetet i Oslo. Universitetsadministrasjonen la ingen vekt på prognosen, men da studenttallet vokste slik prognosen forutsa, ble det debatt. Reiersøl fikk ros, men med vanlig nøkternhet viftet han rosen bort ved å si at suksessen mye skyldtes flaks.

Sannsynlighetsregning og statistikk fikk sine fundamenter som fullverdige vitenskapelige disipliner i mellomkrigstiden, særlig gjennom arbeidene til R.A. Firsher, Jerzy Neyman, Harald Cramér, Abraham Wald og A.N. Kolmogorov. Dette ledet til en revolusjon innen forskning og undervisning innen disse områdene. I Norge ble det Erling Sverdrup og Olav Reiersøl som med stor dyktighet bygde opp statistikk med sannsynlighetsregning som universitetsfag etter krigen. Reiersøls originalitet og ytterste grundighet som universitetslærer er legendarisk. Han motsatte seg å undervise på lavere grad fordi han mente seg ansatt for å gi hovedfagsundervisning og drive med forskning. En gang ledet han et forskningsseminar som gikk over tre semestre. I sitt strev med å få grunnlaget i orden førte han oss som fulgte seminaret knapt nok lenger enn til bokas innledningskapittel.

I det vitenskapelige arbeidet Reiersøl mest med modeller med latent struktur. At strukturen er latent betyr at de variablene som inngår i strukturliknigene ikke kan observeres. Disse latente variable kan betraktes som genotypevariable. Det er bare gjennom observasjoner av fenotypevariable en kan få kunnskap om de strukturelle parametrene. Reiersøl brukte selv i et sent arbeid disse termene lånt fra genetikken, men de har ikke slått igjennom. Reiersøls viktigste bidrag avklarer om modellenes latente struktur er identifiserbar. Dette er tilfelle når det er en-entydighet mellom modellens struktur (vektoren av koeffisienter i strukturlikningene) og de observerte

variablenes sannsynlighetsfordelning. Uten identifiserbarhet kan en ikke skille strukturene fra hverandre, selv med uendelig mye data, og fullstendig statistisk analyse er fåfengt. Modellen kan, imidlertid, være delvis identifiserbar.

I konfluensanalysen til Frisch henger de latente variablene lineært sammen, og til hver latent variabel hører det en observasjon som er den latente variabelen pluss støy. Frisch hadde vist at i modeller med to latente variable som er normalfordelte på den linjen som stukturlikningen bestemmer, er denne linjen ikke identifiserbar når støyen er normalfordelt. Linjen ligger, imidlertid, mellom de to regresjonslinjene en får ved den lineære regresjonene av den ene observerte variabelen på den andre. Denne avgrensningen hadde Koopmans generalisert til flere dimensjoner under den forutsetning at den inverse til kovariansmatrisen til de observerte variablene er positiv, eventuelt etter skifte av fortegn for noen av variablene. Alt i 1941 publiserte Reiersøl en forløper for sin doktoravhandling der han viste at dette også var en nødvendig betingelse for slik områdeidentifserbarhet basert på kovariansmatrisen. Dette følges opp i doktoravhandlingen med en rekke resultater som gir avgrensing av strukturvektoren basert på kovariansmatrisen til de observerte variable. Her utnyttet Reiersøl matriseregning og lineær algebra. Slik matematikk hadde han ikke lært gjennom sitt hovedfagsstudium, men etter å ha sett hvor god nytte Frisch gjorde av matriseregningen, leste han seg opp på området.

Reiersøl er mest berømt for sin Econometrica-artikkel fra 1950. Her gir han en fullstendig analyse av identifiserbarheten i den aller enkleste modellen innen konfluensanalyse. Det er to latente variable som henger lineært sammen. Stigningskoeffisienten for denne strukturlinjen er den interessante parameteren. I det første tilfellet Reiersøl ser på, er observasjonsstøyen bivariat normalfordelt. Når også de latente variablene er normalfordelte, riktignok degenerert til strukturlinjen, er stigningskoeffisienten ikke identifiserbar. Dette var kjent. Ved en elegant og elementær håndtering av den bivariate karakteristiske funksjonen (Fourier-transformen) til observasjonsvektoren, viser Reiersøl at når stigningskoeffisienten ikke er identifiserbar, må de latente variablene være normalfordelt på strukturlinja. Et annet elementært bevis for at stigningskoeffisienten er identifiserbar viss og bare viss de latente variablene *ikke* er normalfordelte, presenterte Reiersøl sammen med Koopmans samme år. Her presenterer de for første gang identifiserbarhet som et spørsmål om en-entydighet. Tidligere var ikke problemet stilt i sin prinsipielle enkelhet.

Reiersøls resultat fra 1950 skapte oppsikt. I alminnelighet faller tingene pent på plass når observasjonene er normalfordelte. I konfluensanalysen er det motsatt. Det er når de latente variablene er normalfordelt at interesseparameteren ikke er identifiserbar. Og bare da. Ved sin artikkel fra 1941 og sin doktoravhandling fikk Reiersøl tidlig anledning til å besøke verdens fremste statistiker eller følge deres forelesninger (Harald Cramér, R.A. Fisher, Abraham Wald, P.L. Hsu, Tjalling Koopmans, Jerzy Neyman). Det hjalp nok også å komme fra miljøet til Frisch, der også Trygve Haavelmo hadde hatt sine læreår. Med sitt overraskende resultat om identifiserbarhet, fikk Reiersøl befestet sitt ry. Det merket jeg selv da jeg kom til Berkeley i 1972. Den gamle kjempen Jerzy Neyman var full av lovord.

I doktoravhandlingen innfører Reiersøl begrepet instrumentvariabel. Dette er hjelpevariable som er korrelert med de latente variablene, men ikke med

observasjonsstøyen i de observasjonene som svarer direkte til de latente variablene. Gode instrumentvariable kan gjøre strukturparametrene idenfiserbare og bidra til deres estimering. Dette begrepet har slått an, og lærebøker i økonometri drøfter vanligvis instrumentvariable og deres bruk.

Psykologenes modeller for faktoranalyse ligner modellene i konfluensanalysen. De har latent struktur, og identifkasjonsproblemet omtales her ofte som rotasjonsproblemet. Reiersøl klargjorde forbindelsen mellom faktoranalysen og konfluensanalysen, og han overførte sine resultater om identifisering dit. Reiersøl bidro også innen genetikk. Hos R.A. Fisher hadde Reiersøl sett arbeider om såkalte genetiske algebraer. Denne tråden tok han opp i 1962, med en elegant og kompakt representasjon av Mendelsk nedarving i koblete loci. For en populasjon med n individer representerte Reiersøl den simultane genfrekvensen i de k koblete loci (som kunne sitte på samme kromosom) som koeffisientene i et formelt polynom. Dette polynomet transformeres ved en differensialoperator, og det transformerte polynom representerer genotypefrekvensen i neste generasjon. Denne algebraiske tilnærmingen gir store forenklinger i studiet av genfordelingen i populasjonen, og hvordan denne fordelingen endrer seg over generasjonene.

Olav Reiersøl var matematisk statistiker, med vekt på matematisk. I hans publikasjoner er det lite innslag av empiriske studier. Det eneste unntaket er noen sider til slutt i doktoravhandlingen, hvor han bruker data for omsetting av egg og produksjon av gris til å illustrere sine metoder. I intervjuet som Willassen (2000) gjorde med Reiersøl, gjør han narr av de svenske avisene som etter diputasen skrev at avhandlingen dreide seg om griseavl.

Da Reiersøl ble bedømt for professoratet han fikk i 1958, ble det nevnt at han ikke framsto som spesielt original i sine problemstillinger. Det ble sagt at han kanskje hadde alt for stor respekt og trofasthet mot sine foregangsmenn. Med stor flid og systematisk grundighet, og med matematisk innsikt og fyndighet, videreutviklet han problemstillinger trukket opp av andre, og viktige spørsmål maktet han å avklare fullstendig. Hans originalitet som forsker ligger mer i hans evne til å komme til bunns i vanskelige problemer og til å se det prinsipielle i dem, enn i å legge ut på eller stimulere til nye forskningstokt i sitt fag.

Reiersøls mer personlige originalitet og uavhengighet, er uomtvistelig. Dette kommer blant annet til syne i hans hang til å konstruere terminologi. I doktoravhandlingen finner en nyskapninger som dito-momenter, forklart som disturbance+total i en note, struto-momenter og toto-momenter. Han betegner også punkter som hystrale av hypothetical+structural.

Vi vil huske Olav Reiersøl som en betydelig forsker innen matematisk statistikk og økonometri, og som en åpenhjertig, kunnskapsrik og imøtekommende veileder og fagfelle. Vi hedrer hans minne.

Utvalgte publikasjoner. Fullstendig bibliografi finnes i Willassen, Y. 2000. The ET interview: Professor Olav Reiersøl. *Econometric Theory*, 16: 113-125.

- Reiersøl, O. 1941. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, **9**: 1-24.
- Reiersøl, O. 1945. Confluence analysis by means of instrumental sets of variables. *Arkiv för Matematik, Astronomi och Fysik*, **32 A**, nr. 4: 1-119.
- Koopmans, T.C. and Reiersøl, O. 1950. The identification of structural characteristics. *The Annals of Mathematical Statistics*, **21**: 165-181.
- Reiersøl, O. 1950. Identification of a linear relation between variables which are subject to error. *Econometrica*, **18**: 375-389.
- Reiersøl, O. 1950. On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrica*, **15**: 121-149.
- Reiersøl, O. 1962. Genetic algebras studied recursively and by means of differential operators. *Mathematica Scandinavica*, **10**: 25-44.
- Reiersøl, O. 1962. Path coefficients, moments of moments and expectations of expectations in population genetics. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, Matematisk-Naturvidenskapelige Klasse, Ny Serie*, nr. 6: 1-23.

Minnetale ved Olav Reiersøls bisettelse 23. februar 2001

Ved Universitetet i Oslo flagges det i dag på halv stang til ære for Olav Reiersøl.

Olav Reiersøl ble dosent i matematikk ved Matematisk institutt, Universitetet i Oslo i 1950 og professor fra 1958. Han var professor i matematisk statistikk fra 1961 frem til 1975. Han hadde fast kontor her frem til 1987. I totalt 37 år var han således knyttet til instituttet.

Det er derfor naturlig at en representant for Matematisk institutt og for fagområdet matematisk statistikk sier noe ved denne bisettelsen. Jeg hadde selv Reiersøl som min veileder.

Ved Olav Reiersøls bortgang er hans kollega ved instituttet frem til midten av sekstiårene, senere professor ved daværende NTH, Arnljot Høyland, blitt fagets nestor. Arnljot er på vei hjem fra et 14 dagers opphold på Kanariøyene, og er derfor forhindret fra å delta noe han veldig gjerne skulle vært sammen med kona Liv, selv hovedfagsstudent i Olav Reiersøls tid.

I anledning av Olav Reiersøls 80 års dag i 1988 arrangerte instituttet en lunch for ham og kona Margrete i restauranten i Munchmuseet. De av oss som deltok der husker det som en svært hyggelig anledning der instituttet fikk takket Reiersøl for hans innsats for matematikkfaget.

Olav Reiersøl var ikke glemt av dagens generasjon statistikere. I septembernummeret i fjor av Tilfeldig Gang - TG, utgitt av Norsk Statistisk Forening, ble det trykket et intervju med ham, gjort av professor Yngve Willassen ved Økonomisk institutt, Universitetet i Oslo, i hans hjem på Nøtterøy i anledning av 90 års dagen i 1998. Dette intervjuet ble først trykket internasjonalt i Econometric Theory.

Willassen trekker her frem at boken til Ragnar Frisch om konfluensanalyse synes å være utgangspunktet for Reiersøls arbeider i økonometri, og fastslår at dette i moderne faglig språkbruk dreier seg om modeller med såkalte latente variable. Slike modeller er meget aktuelle i moderne forskning i fagområdet matematisk statistikk. Det var f. eks. tema for den oppgitte prøveforelesning da Xeni Dimakos helt nylig disputerte. Noen eksempler på anvendelser er psykometri, som Reiersøl selv ga viktige bidrag til, og

jordskjelvforskning og artsdiversitet i Skagerak som jeg selv for tiden arbeider med modeller for.

Den beste kilde til Reiersøls liv er hans selvbiografiske notater, som ligger i mappen hans ved instituttet. Jeg skal gjengi to utdrag herfra:

"Frisch hadde flere regneassistenter. Frisch arbeidet delvis med konkrete økonomiske data, dels med konstruerte data for å prøve ut statistiske metoder. Da jeg begynte som assistent hos Frisch var cand. oecon. Trygve Haavelmo leder for regnearbeidet. Senere sluttet Haavelmo ved instituttet, og jeg overtok som leder for regnearbeidet."

I våre dagers terminologi var opplagt miljøet rundt Frisch et "Senter for Fremragende Forskning".

"I 1951 diskuterte Sverdrup og jeg på ny om vi skulle foreslå for Fakultetet å innføre statistikk som realfag, men vi kom til at det ville være håpløst så lenge ingen av oss var medlem av Fakultetsrådet. På den tid var alle professorene, men ikke dosentene, medlemmer av Fakultetsrådet. Vi ble enig om at den første av oss som ble professor skulle foreslå innføring av statistikk som realfag."

På sett og vis er det kanskje tilfeldigheter som gjorde at Erling Sverdrup av mange blir betraktet som statistikkfagets far i Norge.
Egentlig var det to fedre.

Reiersøls selvbiografiske notater inneholder en kronologisk liste over all undervisning han ga ved Matematisk institutt. Denne listen viser at han supplerte Sverdrup på en utmerket måte ved bl.a. å gi forelesninger i det matematiske grunnlag for faget, i teori for stokastiske prosesser og da særlig for såkalte Markov prosesser. Denne siste teorien er igjen helt sentral i den moderne utviklingen av faget. Et stikkord her er Markovkjede Monte Carlo simuleringer, som igjen er tett koblet til modeller med latente variable allerede omtalt. Reiersøl ga også forelesninger i mer anvendte områder som populasjonsgenetikk, Shannons informasjonsteori og køteori.

Også sett med internasjonale øyne er det en imponerende bredde i statistikkfaget i Norge i dag. Mye av dette er frukter av impulser Reiersøl brakte hjem fra sine mange utenlandsopphold i topp internasjonale miljøer.

Jeg husker Olav Reiersøl som en vennlig mann med et hyggelig smil, men som pga. et ikke alltid velfunksjonerende høreapparat kunne være litt problematisk å kommunisere med.

Et brev til Jan Erlvang 28. september 1964 fra Reiersøl som var på studieopphold ved Cornell University, avsluttes med: "Det ville være fint om dere kunne frankere med jubileumsfrimerker. En av mine kolleger er nemlig frimerkesamler."

Erlvang skriver tilbake 30. mars året etter: " Når du kommer hjem, er det store forandringer på jordene utenfor Administrasjonsbygget. Lerken som tidligere slo sine triller der, er nå jaget vekk av reisverket til matematikkbygget som skal være under tak til jul." Svarbrevet fra Reiersøl 3. april avsluttes med: "Det gleder meg å høre om matematikkbygget."

I mappen til Reiersøl ligger også en liste over hans publikasjoner sirlig sortert. Listen omfatter 14 arbeider i internasjonale tidskrifter. Kvantiteten er kanskje ikke så imponerende, men spesielt i matematiske fag er det kvaliteten som teller. Her kommer Reiersøl med sine originale arbeider svært godt ut.

Matematisk forskning er sjelden lett å formidle til slekt, venner og samfunn. Reiersøl skrev dessuten alle sine arbeider alene. Matematikere er derfor avhengig av omsorg og omtanke kanskje mer enn mange andre. Vi vet at hans kone Margrete ga ham dette gjennom mer enn 30 år og helt til det siste. Vi er usedvanlig takknemlig for dette i dag.

Vi lyser fred over Olav Reiersøls minne.

I ærbødighet

Bent Natvig

Nytt om nettverk: NOREVENT

Ellen Amundsen
Statens institutt for folkehelse

Å treffe på en statistiker i Norge er en sjelden begivenhet. Mange arbeider alene i sitt miljø, andre har anledning til å klumpe seg noe mer sammen i miljøer hvor kravet til bruk av statistiske metoder er blitt innarbeidet som et nødvendig ledd i produksjonsprosessen. Som kunnskapsoaser ligger så universitetene og enkelte andre miljøer hvor statistikeren skal forske og deretter lære studenter, seg selv og oss andre om nye temaer eller metoder. Selv i større miljøer kan det faglige fokus være så ulikt at det blir nødvendig å søke andre miljøer for å finne god faglig respons.

I alle miljøer blir tiden til faglig utvikling knapp. Behovet for et nettverk som er opptatt av de samme faglige problemstillinger som deg selv, er velkjent. Noen greier å bygge slike. Et nettverk – en forskningsgruppe - knyttet til temaet *forløps- og overlevelsesanalyse* ble startet i august 2000 av Odd Aalen ved Seksjon for medisinsk statistikk, UiO. Gruppen ble etter noe diskusjon døpt til NOREVENT. Navnet spiller delvis på Nord eller Norge og også på den engelske betegnelsen Event History Analysis.

Norevent's viktigste målsetting er å bidra med hjelp og inspirasjon til forskere ved å diskutere mulige problemer i deres prosjekter i utviklingsfasen. Et hovedformål er å styrke rekrutteringen av yngre forskere inn på fagfeltet. Slik rekruttering forutsetter blant annet et godt fagmiljø. I skrivende stund har 18 personer i Oslo-regionen vist sin interesse, deltatt på møter og/eller beskrevet sine prosjekter på gruppens nettside: http://www.med.uio.no/imb/stat/norevent/ Anvendelsene som er representert er særlig innenfor kreftforskning, AIDS-forskning og landbruksforskning.

Norevent har en base ved Seksjon for medisinsk statistikk ved Universitetet i Oslo, men har tilknyttet miljøer ved blant annet Matematisk institutt og Institutt for informatikk, UiO, og ved Kreftregisteret. Odd Aalen [o.o.aalen@basalmed.uio.no] er for tiden koordinator for Norevent. I en sekretærfunksjon sitter stipendiat Tron Anders Moger [t.a.moger@basalmed.uio.no] Kontakter utover i Norge vil bli knyttet etterhvert og det søkes blant annet reisemidler som muliggjør deltagelse i arbeidsgruppen fra forskere i Trondheim, Bergen og andre deler av landet. Det er allerede et bredt internasjonalt kontaktnett. Det er kontakter i de nordiske land, i England (London, Cambridge), i USA (Los Angeles, Atlanta). *Aktivitetene* til Norevent er å avholde jevnlige møter om pågående forskningsprosjekter, årlige dagskonferanser ("Lysebumøtet") for gruppen, med invitert gjest, og å arrangere internasjonal "Workshop".

Gruppen har behov for midler til møter, reiser og stipendiater. Det tas sikte på å utvide gruppens kontaktflate nasjonalt og internasjonalt. Til dette kreves midler som gjør det mulig å invitere forskere til oss. Norevent har for tiden ingen økonomisk støtte, men har søknader inne i Jahrefondet, Den Norske Kreftforening og NORFA.

Vil du vite mer? Sjekk http://www.med.uio.no/imb/stat/norevent/ Eller skriv til Odd eller Tron.

Vil du lage noe lignende for ditt faglige fokus? Bruk TG som starthjelp! Har du noe lignende på gang? Fortell om det i TG!

VI MINNER OM PÅMELDING TIL DET 11. NORSKE STATISTIKERMØTE I ULVIK OG PREKONFERANSEKURS (etterutdanningsseminar) I BERGEN

OPPDATERT INFORMASJON /PÅMELDINGSSKJEMA: http://www.uib.no/smis/bsf/
(Merk tidspunkter for bussavgang fra Flesland flyplass)

PÅMELDINGSSKJEMA (ny frist 15. mai, 2001)
Det 11. norske Statistikermøte

Sted: Brakanes hotell, Ulvik
Tidspunkt: 20-22 juni 2001

Navn: _____
Adresse: _____
_____

Telefon/Faks: _____
Email: _____

HOTELLOPPHOLD
Hver enkelt deltager betaler direkte til Rica Brakanes hotell.

Kryss av det som passer:
[ ] Ønsker enkeltrom (kr 1.195,- pr døgn, fullpensjon)
[ ] Ønsker dobbeltrom (kr 995,- pr døgn, fullpensjon)
Navn på den du ønsker å dele rom med:

[ ] Lunsj ankomstdag [1]
[ ] Lunsj avreisedag [1]
[ ] Ønsker transport t/r Flesland/Bergen med buss (kr 300,-)
Avreise fra Flesland kl. 09.45
Ankomst Flesland ca. kl. 16:00 (beregn litt ekstra tid, ca ½ time?)

[1] *Dersom lunsj ønskes både ankomst og avreisedag; 165 kroner ekstra*

KONFERANSEAVGIFT
Konferanseavgift (600 kr) og evt busstransport (300 kr) betales til:

Kontonr: 52530502783
Statistisk Forening i Bergen
v/ Terje Schaathun
Øvre Fantoftåsen 39
5072 Bergen
Merk innbetalingen med navn og "11. norske Statistikermøte ".

Påmeldingsfrist gjelder også for innbetaling av konferanseavgift + evt transport.

Påmeldingsskjemaet sendes til:
Statistisk Forening i Bergen
v/ Terje Schaathun
Øvre Fantoftåsen 39
5072 Bergen
e-mail: terjesc@online.no

Hovedfagstudenter kan søke hovedforening (NSF) om stipend til deltagelse (se annonse i mars nummer av TG).

PÅMELDINGSSKJEMA (frist 15.mai, 2001)

Max antall deltagere: 30

NSF's ETTERUTDANNINGSSEMINAR 2001

Sted: PC-laboratoriet, Armauer Hansens Hus (kjelleretasje) ved Haukeland sykehus, Bergen.
Tidspunkt: 18-19 juni 2001

Navn: _____
Adresse: _____
_____

Telefon/Faks: _____
Email: _____

Konferanseavgift:

Medlem i NSF	kr. 2500,-
Ikke medlem	kr. 2600,-

Innbetaling av kursavgift skjer til konto:

Kontonr: 52530502783
Statistisk Forening i Bergen
v/ Terje Schaathun
Øvre Fantoftåsen 39
5072 Bergen

Merk innbetalingen med navn og "NSF etterutdanningsseminar 2001 ".

Påmeldingsfrist gjelder også innbetaling av kursavgift.

Påmeldingsskjemaet sendes til:
Statistisk Forening i Bergen
v/ Terje Schaathun
Øvre Fantoftsåen 39
5072 Bergen
e-mail: terjesc@online.no

Hovedfagstudenter kan søke hovedforening (NSF) om stipend til deltagelse (se annonse i mars nummer av TG).

Medlemsskap i Norsk Statistisk Forening

Hjemmeside for Norsk Statistisk Forening: http://www.math.uio.no/nsf/

Kontingent:
200 kr/år
50 kr/år for studenter (til og med hovedfag).
Kontonummer: 0530 2773299

Nummeret som er angitt på adresselappen på TG angir siste år kontingenten ble betalt.
Ved innbetaling må man opplyse om endret adresse (f.eks endret postnummer).

Korrespondanse om medlemsskap og innbetaling av kontingent skal skje til foreningens kasserer:

Norsk Statistisk forening v/Ågot Irgens
Kalfarlien 6A
5018 Bergen
Emailadresse: agot.irgens@haukeland.no

Nye medlemmer melder seg til kasserer og får så automatisk tilsendt Tilfeldig Gang fra første påfølgende nummer av.

Mailadresser:
Vi ønsker mailadresser til alle medlemmer i NSF:
Oppgi e-mailadressen din på merknadsfeltet på postgiroen, eller send en e-mail til agot.irgens@haukeland.no om mailadresse og evt adresseendring.

Annen korrespondanse med foreningen skal foregå med leder i foreningen:

Geir Egil Eide
Kompetansesenter for klinisk forskning
Haukeland sykehus
Armauer Hansens hus
5021 Bergen
Emailadresse: geir.egil.eide@haukeland.no

B-post

Retur:
Norsk Statistisk Forening
v/Ågot Irgens
Kalfarlien 6A
5018 BERGEN
Email: agot.irgens@haukeland.no