



ISSN 0803-8953

**TILFELDIG
GANG**

Nr. 3, årgang 23
Desember 2006

Utgitt av Norsk Statistisk Forening

INNHold

	Side
Fra redaktøren, <i>Ingunn Fride Tvete</i>	2
Lederen har ordet, <i>Sven Ove Samuelsen</i>	3
NR presenterer seg selv	
SAND, Oljereservoarmodellering, <i>Ragnar Hauge</i>	4
SAMBA, Bioinformatikk, <i>Marit Holden</i>	5
SAMBA, Marin biologi, <i>Magne Aldrin</i>	7
SAMBA, Bank, finans og forsikring, <i>Kjersti Aas</i>	10
Referat fra møter	
Temakveld, Statistisk forening, Oslo.....	12
Årsmøte i Statistisk forening, Trondheim.....	13
Jubileer	
André Teigland 40 år, <i>Ingunn Fride Tvete</i>	14
Puslerispalten	
Pusleri nr. 20, <i>Jostein Lillestøl</i>	16
Løsning til pusleri nr. 19, <i>Jostein Lillestøl</i>	17
Stipendiater presenterer seg selv	
Fra UMB, <i>Kristian Hovde Liland</i>	18
Fra UIO, <i>Daniel Berg</i>	19
Juks med forskning siden antikken, Øystein Evandt	21
CAS, Sven Ove Samuelsen	24
SFI, Arnaldo Frigressi	25
Kvinner bak tallene, Ingunn Fride Tvete	27
Meldinger og annonser	
Nytt fra UIO, Matematisk institutt, Avdeling for statistikk.....	30
Nytt fra NTNU, Institutt for matematiske fag, Seksjon for statistikk	30
Nytt fra UMB, Institutt for matematiske realfag og teknologi.....	31
Nytt fra UIB, Matematisk institutt, Statistikk.....	33
Nyhet fra Vest, <i>Geir Egil Eide</i>	33
Møter og kurs.....	34
Stillingsutlysninger.....	42
Medlemskap i Norsk Statistisk Forening.....	44
Årskontingent Norsk Statistisk Forening	45

Fra redaktøren



Du holder nå i hånden årets tredje og siste utgave av Tilfeldig Gang. Denne gangen er det NRs tur til å ta ansvar for utgaven, og undertegnede har blitt redaktør. Det har ikke gått så mange måneder siden forrige utgave av Tilfeldig Gang (august 2006).

Denne utgaven var relativt fyldig, med spennende beretninger om SFier og referater fra flere av sommerens konferanser. Så mye nytt og spennende er det kanskje ikke å berette om siden den gang. Vi her på NR så dermed en gyllen mulighet til å formidle hva vi driver på med. Ulike fagområder, og noe av det man gjør innenfor disse, blir presentert av Ragnar Hauge, Marit Holden, Magne Aldrin og Kjersti Aas.

Statistisk forening, avdeling Oslo, hadde høstens temakveld 18 oktober, der vi hadde interessante foredrag av Bjørn Helge Mevik og Øyvind Langsrud. Referat fra dette møtet, samt referat fra årsmøtet i Statistisk forening, Trondheim, er gitt i dette nummeret.

Det er nok delte meninger om hva et "nyhetsbrev" som Tilfeldig Gang skal inneholde. Jeg kunne for eksempel tenke meg en fast spalte der noen stipendiater presenterer seg selv og det de driver på med. Siden dette er første gang jeg er redaktør vil jeg benytte sjansen til å innføre nettopp dette. Jeg har derfor bedt noen stipendiater om å presentere seg selv og arbeidet sitt. Kristian Hovde Liland og Daniel Berg stilte sporty opp. Mens Kristian er helt i startfasen av sin PhD, har Daniel kommet langt i sitt arbeide. Vi får altså presentert to stadier av en stipendiats tilværelse.

Året 2006 nærmer seg slutten. Ser vi tilbake på året som har gått finner vi at forskning har vært et hyppig tema i media, på godt og vondt. Spesielt har historiene om juks og uredelighet i forskningen skapt store overskrifter i landets aviser. Dette har vel vært sjokkerende lesning for de fleste av oss, men som mange påpeker; dette er ikke noe nytt i historien. Øystein Evandt forteller mer om dette her i TG.

Som noen av TGs faste lesere kanskje er kjent med, har en av gruppene ved Senter for grunnforskning, ledet av Odd Aalen og Ørnulf Borgan, jobbet med forløpsanalyse. Sven Ove Samuelsen forteller nærmere om Senter for grunnforskning.

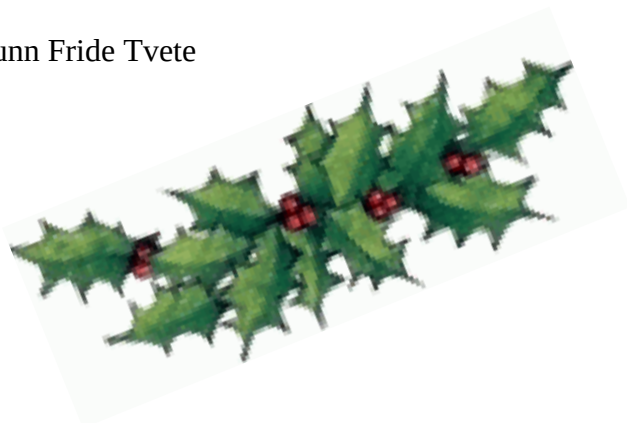
I "Kvinner bak tallene" ser vi litt på kvinner og statistikk før og nå.

Faste spalter som Puslerispalten, Jubileer, Meldinger og annonser er selvfølgelig på plass også i denne utgaven av TG.

Leder i Statistisk Forening, Sven Ove Samuelsen, har hjulpet til med innsamling av data, i tillegg til å være skribent og rådgiver. Takk for god hjelp! En stor takk også til alle de andre bidragsyterne! Dere bidrar til at Tilfeldig Gang består.

Jeg vil med dette ønske alle lesere og deres familier en riktig god og fredelig jul, og et godt nytt år!

Ingunn Fride Tvete



Lederen har ordet

Denne gangen skriver jeg i utlendighet og dette blir nok mer et reisebrev enn en leder i vanlig forstand. Jeg har nemlig forskningsfri dette året og høsten tilbringes ved London School of Economics and Political Science (LSE) hvor jeg besøker Anders Skrondal. Anders ble ansatt på LSE som professor i sosialstatistikk for knapt 2 år siden.



LSE ligger omtrent så sentralt i London som det er mulig, ca. 500 meter øst for Covent Garden. I deler av tiden har jeg bodd i en leilighet på Campus, så tilgjengeligheten til sentrum har altså vært upåklagelig. Dvs. kanskje den nesten har vært litt for god, London kan jo være ganske ståkete og jeg føler meg av og til noe ør. Prisen for å leie så sentralt er også verdt å nevne, 50£ om dagen som med en kurs opp mot 13 kroner blir ca. 19000 i måneden. Så det koster!

Som mange vet er det ytterligere en nordmann på Department of Statistics på LSE, nemlig Ragnar Norberg som er professor i forsikringsmatematikk her. Ragnar har jeg imidlertid ikke truffet i høst siden han også har sabbatical og tilbringer den på sitt landsted i Provence. Noe fredeligere skulle jeg tro. Men dermed har jeg fått låne hans meget romslige, om enn ikke spesielt godt lydisolerte eller luftavkjølte, kontor.

Det er selvfølgelig mulig å lure på om dette er noen god måte å lede en statistisk forening på, men jeg føler ikke jeg har forsømt så mye i høst. For det første spiller det liten rolle om man skriver eller leser en email i Oslo eller London (i dag er det jo blitt et problem å komme seg bort, selv norsk TV kommer jo rett inn på PC-en). Friåret mitt var også allerede avtalt da jeg ble foreslått som leder. Dessuten har det egentlig vært ganske få saker i høst.

Den eneste større sak er neste års norske møte og forkurs på Sommerøya utenfor Tromsø, men som dere kan lese mer om i dette nummeret synes planleggingen av møtet å være i god gjenge. Så jeg håper å treffe mange av dere på Sommerøya i juni, både på forkurset som vil bli gitt av Anders Skrondal og hans bokmedforfatter Sophia Rabe-Hesketh, og på selve møtet.

Jeg snakket en gang med tidligere leder av foreningen, Geir Egil Eide, om å skrive til TG. Vel, til julenummeret er det greit, sa Geir, da kan man bare ønske god jul. Det synes jeg er tradisjon som det er hyggelig å følge. Derfor ønskes alle foreningens medlemmer en riktig god jul!

Sven Ove Samuelsen



SAND-gruppa på NR

Av Ragnar Hauge

På Norsk Regnesentral er det to grupper som driv med statistikk: SAMBA (Statistisk Analyse, Mønstergjenkjenning og BildeAnalyse), og SAND (Statistisk Analyse av NaturressursData). I praksis er "naturressurs" i all hovudsak olje, og vi 14 på SAND driv for tida berre med oljerelaterte prosjekt, medan SAMBA dekkjer alt anna.



Hovudtyngda av det vi driv med er metodar som vert brukt til reservoarmodellering. Dette går ut på å laga ein best mulig modell for reservoaret for å predikera flyt av olje, gass og vatn. Denne flyten vert approksimert av Darcy-likninga, som igjen er avhengig av porøsitet og permeabilitet i mediet der flyten skjer. Ettersom samanhengen mellom reservoarmodell og flyt er svært kompleks vert det i stor grad nytta Monte Carlo teknikkar for prediksjon. Fokus for metodane våre er derfor å kunne sampla reservoarmodellane korrekt, gitt data.

Dei tilgjengelege datakjeldene er brønndata og seismiske data. Brønndata vert rekna som presise, men eksisterer berre i ein svært liten del av volumet. Seismiske data er overalt, men er mindre presise. Oppløysing er dårleg, og både dei høge og låge frekvensane manglar. Seismikken gir dessutan berre informasjon om dei såkalla akustiske parametrane, og krev at ein modellerer samanhengen mellom desse og det ein er interessert i. Dessutan vert seismiske data målt ved gangtid til refleksjonar, og omrekninga frå denne tida til faktisk djup er og usikker.

Det tynne datagrunnlaget er ein av grunnane til at vi i stor grad nyttar Bayesianske modellar. Den andre viktige grunnen er at geologane ofte har ein ide om korleis reservoaret ser ut, og kan gi a priori informasjon basert på dette, og på reservoaranalogiar (geologiske formasjonar som ligg i dagen, og som ein meiner har same geologi som reservoaret).

Den stokastiske delen av reservoarmodellering foregår i tre steg:

1. Flatemodellering.
2. Bergartsmodellering.
3. Petrofysikkmodellering (porøsitet og permeabilitet).

Flatemodelleringa har som mål å finna flater som avgrensar den geologiske formasjonen som utgjør reservoaret. (Ei flate som ein finn nær sagt over heile verda er "Bunn Kritt" – som representerar tida då Kritt-periode byrja.) Slike flater kan typisk tolkast direkte frå seismikk, men då med uvisse. Vi modellerer desse som gaussiske felt, og nyttar kriging for å få samsvar i brønnobservasjonane. Kriging er ein eksakt interpolasjonsteknikk, som gir korrekt betinging i gaussiske felt. Flatemodellen vert brukt både til å gi geometrien til vidare modellering, og til volumutrekning.

Bergartsmodellering er essensielt i dei tilfella der reservoaret er samansett av ulike bergartar der nokon er oljeførande og andre ikkje. Flyten i reservoaret vil då vera veldig styrt av geometrien til desse bergartane. Dette er ein diskret modell, som vert laga på eit grid der topp og botn er definert av flatemodellen. SAND-gruppa er leiande på bruk av merka

punktprosesser til slik modellering, men jobbar og med andre metodar, som trunkerte gaussiske felt og markovfelt.

Det neste steget er å modellera petrofysikk for kvar bergart. Til dette nyttar vi essensielt sett gaussiske felt igjen, men gjerne kombinert med eit sett transformasjonar som gjer at den endelege fordelinga kan vera langt frå normal. Utfordringa her er å få krigingprosessen til å gå raskt nok. Kriging krev i utgangspunktet inversjon av kovariansmatrisa for alle datapunkt, men sidan det kan vera hundre tusen av dei er dette ikkje praktisk gjennomførbart. Her har vi jobba med teknikkar for å kunna splitta opp problemet og løysa det som mange småproblem utan at det i særleg grad påverkar resultatet.

Etter at petrofysikken er modellert, vert resultatet oftast oppskalert til eit grovare grid. Dette er ein deterministisk prosess, som vert gjort sidan flytsimuleringa er svært tidkrevande på fine grid. Til slutt kjem så målet, flytsimuleringa. Ved å gjenta denne løkka kan ein få ut alternative flytmønster, og ut frå dei planlegga utvikling og drift av reservoaret.

Utover reservoarmodellering driv vi og litt med metodar for oljeleiting, men det er eit område med mykje hysj-hysj og paranoia, så meir er det ikkje å seia om det. Vi driv og litt med andre spesialiserte modellar for delar av verksemda i eit oljeselskap, som estimering av boretid for nye brønner, og enkel avgjerdsstøtte på høgare nivå.

Mange av modellane vi har utarbeidd er tilgjengelege som kommersiell software frå eit profesjonelt softwarefirma (Roxar) som NR samarbeidar med. Dei er i jevnleg bruk blant store oljeselskap verda over.

Bioinformatikk – et tverrfaglig forskningsfelt

Litt om bioinformatikkaktiviteten ved Norsk Regnesentral

Av Marit Holden

Det genetiske materialet i menneskets celler er stort sett kartlagt. Den store utfordringen i årene framover blir å forstå rollen til hvert gen, dvs. hvilken funksjon det har og hvordan de ulike genene virker sammen. Dette gir ny forståelse av biologiske prosesser. Slik kunnskap kan bl.a. brukes til å utvikle nye og bedre medisiner, til å stille bedre og mer presise diagnoser og til å skreddersy behandling til det enkelte individet.



Nye teknikker, for eksempel cDNA mikromatriser, gjør det mulig å studere rollen til tusenvis av gener samtidig. Slike teknikker produserer enorme mengder data. For å få mest mulig ut av disse dataene er det nødvendig å bruke og å utvikle avanserte statistiske metoder. Det finnes mange ulike typer genomdata, for eksempel mikromatrisedata (genekspresjonsdata), CGH-data, SNP-data og proteomikkdata. NR har først og fremst jobbet med cDNA mikromatrisedata og også noe med SNP-data, så bare disse vil bli omtalt under.

Mikromatriseforsøk er svært ressurskrevende. Blant annet derfor er et godt gjennomtenkt forsøksdesign før mikromatriseforsøkene utføres svært viktig. For å oppnå gode resultater er det nødvendig med nært samarbeid mellom forskere i statistikk og forskere i genetikk, biokjemi, medisin og/eller biologi. De siste årene har statistikere ved NR, ofte sammen med statistikere ved universitetene, samarbeidet med biologer, biofysikere og medisinerer ved de norske universitetssykehusene. I slike samarbeidsprosjekter identifiserer sykehusene de biologisk mest interessante problemstillingene og produserer data som statistikerne så analyserer ved hjelp av ulike statistiske metoder. For at resultatet av den statistiske analysen skal bli best mulig, legger vi stor vekt på at vi sammen planlegger hvilke mikromatriseforsøk som bør gjøres.

Det som måles ved hjelp av mikromatriseteknologien er genuttrykk, dvs. hvor aktive de ulike genene i en celle er. Ulike celletyper har ulikt genuttrykk. Det er disse forskjellene i genuttrykk man er interessert i å måle. Man kan for eksempel være interesserte i å finne ut: Hvilke gener er mer/mindre uttrykt i kreftceller sammenlignet med normale celler? Hvilke gener er forskjellig uttrykt hos pasienter som responderer godt og dårlig på behandling? Hvilke gener er knyttet til overlevelse? NR har jobbet med alle disse problemstillingene.

Dataene fra mikromatriseforsøk er svært støyfylte. Denne støyen er komplisert og kommer fra de mange forskjellige stegene i prosessen fra man har en celleprøve til man har data for genuttrykk. De fleste statistiske analysemetoder er basert på å analysere intensitetsratioer mellom to biologiske prøver, der intensitetsmålingene for de to prøvene oppnås fra et enkelt mikromatriseeksperiment. NR har brukt slike ratiobaserte metoder for flere ulike biologiske problemstillinger de siste årene. Intensitetsratioene er et mål på forholdet mellom transkriptnivået i to ulike celleprøver. Det er vanskelig å oppnå absolutte transkriptnivåer fra mikromatrisedata pga. eksperimentell variasjon som kommer fra produksjons- og hybridiseringsprosessen (dvs. prosessen der biologisk materiale binder seg til mikromatrisen). Slik informasjon er på den annen side svært nyttig siden den gjør det mulig å sammenligne resultater fra ulike mikromatrisestudier direkte. NR har vært med på å utvikle en metode for å finne slike absolutte transkriptnivåer.

Et annet område innen bioinformatikk der NR er aktiv, er analyse av SNP-data. DNA-sekvensen med bokstavene A, T, C, G for mennesker er kjent, og den inneholder ca. 3 milliarder bokstaver fordelt på 23 kromosomer. Det som gjør oss genetisk ulike er at denne sekvensen varierer noe fra menneske til menneske. Den vanligste formen for DNA-variasjon er SNP-er. En SNP er en posisjon i DNA-sekvensen hvor noen individer har f.eks. en C mens andre har f.eks. en A. Ulike bokstaver i en posisjon kan medføre at ulike proteiner blir produsert, evt. at enkelte proteiner ikke blir produsert i det hele tatt. Dette kan igjen føre til utvikling av ulike sykdommer. En aktuell problemstilling er da å finne ut om noen (grupper av) SNP-er er assosiert med sykdom.

Det fins i dag mange metoder og mye programvare for å finne lister med gener som er differentielt uttrykte eller lister med SNP-er som er assosiert med sykdom. Mange av disse metodene er basert på hypotesetesting. Siden vi har data for titusenvis av gener/SNP-er på en gang, så betyr det at vi ønsker å teste mange hypoteser på en gang. For hvert gen tester vi om genet er forskjellig uttrykt i to vev, og for hver SNP tester vi om SNP-en er assosiert med sykdom. Det fins forskjellige metoder for å korrigere for at mange hypoteser testes samtidig. Det vanligste for SNP- og mikromatrisedata er å kontrollere FDR (false discovery rate), men det fins også metoder der både FDR og FNR (false non-discovery rate) kontrolleres. Uansett fremgangsmåte for korrigere for at det gjøres mange tester, så er det et problem at testene får lav styrke så lenge man tester ett og ett gen / en og en SNP av gangen.

Utvikling og bruk av metoder der man tester grupper av gener/SNP-er, istedenfor å teste dem enkeltvis, er derfor viktig for å kunne trekke mest mulig biologisk interessant informasjon ut av dataene. Da vil det også være lettere å kunne ta hensyn til og utnytte informasjon om at genene/SNP-ene virker sammen i kompliserte genetiske nettverk, dvs. det at det er mye korrelasjon mellom genene/SNP-ene. I tillegg til dette vil man kunne få mye mer ut av dataene dersom man bruker dem sammen med andre typer genomdata og annen type informasjon knyttet til det aktuelle biologiske materialet. Utvikling av modeller og metoder som klarer dette vil være svært viktig i tiden fremover.

FUGE er et program i Norges forskningsråd for funksjonell genomforskning, dvs. forskning for å finne ut hvilken funksjon de enkelte genene og proteinene har og hvordan de virker inn på hverandre. Sammen med blant annet UiO, Rikshospitalet-Radiumhospitalet og Norges landbrukshøgskole (NLH) deltar NR i flere FUGE-plattformer/-prosjekter og andre Forskningsrådsprosjekter innen bioinformatikk. Vi er med i følgende FUGE-plattformer: Bioinformatikkplattformen (the Norwegian Bioinformatics Platform); Mikromatriseplattformen (the Norwegian Microarray Consortium); og CIGENE (the Norwegian Centre for Integrative Genetics). I tillegg har NR flere prosjekter innen statistiske deler av genforskningen for flere sykehus, og vi veileder UiO-ansatte og studenter i bl.a. forsøksplanlegging og statistiske analyser. Vi har også holdt en del kurs og seminarer.

Miljø, marin og helse ved Norsk Regnesentral

Av Magne Aldrin

Miljø, marin og helse er ett av forskings- og anvendelsesområdene i SAMBA-avdelinga ved Norsk Regnesentral (NR), med et omfang på i overkant av fire årsverk. Oppdragene er i hovedsak finansiert av offentlige myndigheter eller Norges forskningsråd, men noe er også finansiert av næringslivet. I det følgende gir jeg noen smakebiter på hva vi arbeider med. Mer informasjon finnes på http://www.nr.no/pages/samba/area_emr.



Miljø

Innen miljøfeltet har vi i noen år arbeidet med å modellere forurensning i vassdrag sammen med Norsk institutt for vannforskning. Vi har tatt utgangspunkt i en veletablert deterministisk hydrogeokjemisk modell og koblet denne med en empirisk stokastisk modell for å kunne estimere parametre, predikere og beregne usikkerhet usikkerhet innafor et statistisk rammeverk.

Et annet miljøarbeid som har pågått i flere år har vært modellering av trafikkrelatert luftforurensning som funksjon av trafikkmengde, meteorologiske forhold og eventuelle tiltak for å begrense luftforurensninga. Et tiltak det er stor interesse for er salting med MgCl med tanke på å begrense mengden av svevestøv. Vinteren 2004/2005 blei det gjennomført et eksperiment i Strømsåstunnellen i Drammen, hvor det blei saltet med MgCl 47 ganger i løpet av forsøksperioden. Dataene består av timevise registreringer av svevestøvkonsentrasjon, trafikkvolum, meteorologiske variable som temperatur, luftfuktighet og lignende, samt tidspunkt for salting. Svevestøvmengde måles som konsentrasjonen av PM_{2.5} (partikler

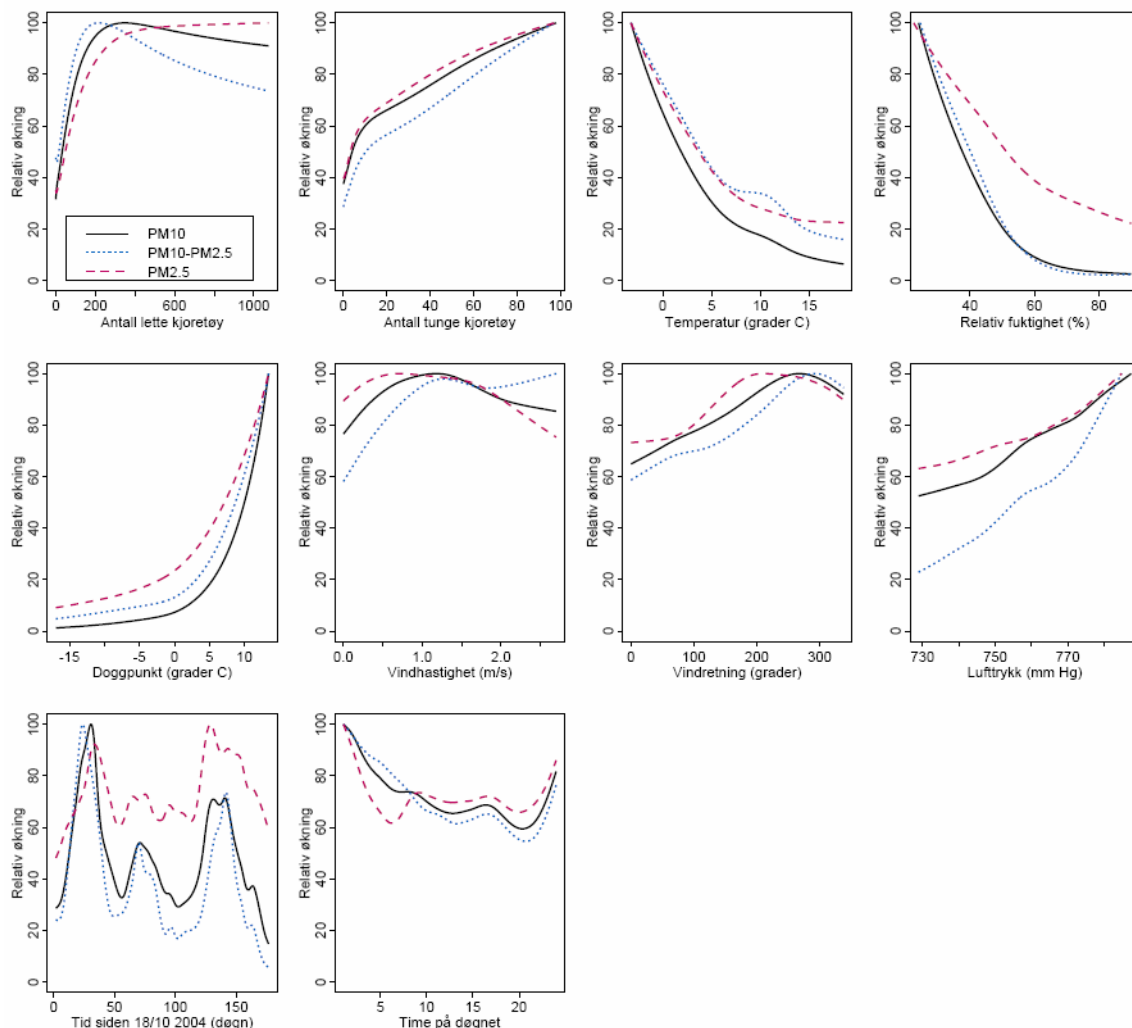
mindre enn 2.5 nanometer), PM_{10} (partikler mindre enn 10 nanometer) og grovfraksjonen $PM_{10}-PM_{2.5}$. Dette modelleres ved hjelp av en generalisert lineær modell

$$\log(y_t) = s_1(x_{1t}) + \dots + s_p(x_{pt}) + \varepsilon_t,$$

hvor y_t er en av svevestøvskomponentene i time t , x_{it} -ene er forklaringsvariable, s_i -ene er glatte funksjoner og ε_t er et restledd. Ved tilbaketransformering til original skala kan modellen skrives på multiplikativ form

$$y_t = S_1(x_{1t}) \cdots S_p(x_{pt}) \cdot e_t,$$

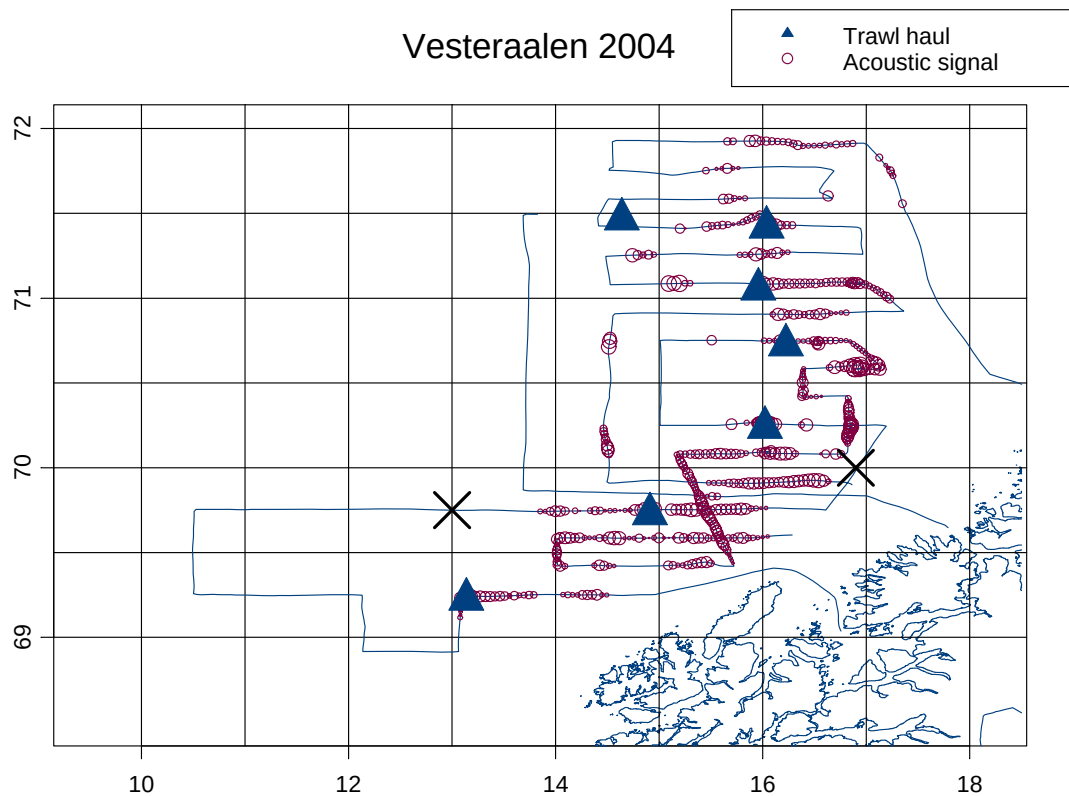
hvor $S_i(\cdot) = \exp(s_i(\cdot))$. Figuren under viser de estimerte S_i -funksjonene for disse dataene, så langt uten at effekt av salting er inkludert i modellen. Ved å utvide modellen med effekt av salting finner vi at salting med $MgCl$ reduserer konsentrasjonen av grovfraksjonen $PM_{10}-PM_{2.5}$ med i middel 60% umiddelbart etter tiltaket, med en avtakende effekt 5-10 dager framover. Effekten på små partikler ($PM_{2.5}$) er derimot langt mindre.



Marin

Prosjekter i tilknytning til oppdrettsnæringa utgjør en voksende andel av arbeidet innen marin. Det siste året har vi arbeidet med spredning av smitte mellom oppdrettsanlegg (i samarbeid med Veterinærinstituttet og Universitetet i Oslo), prediksjon av mengde slakteklare fisk og produksjon av fiskefor. Den største aktiviteten er likevel bestandsestimering av villfisk og vågehval, og dette arbeidet foregår i tett samarbeid med Havforskningsinstituttet.

Figuren under viser et eksempel på data fra tokt. Dataene er samlet inn av forskningfartøyet Johan Hjorth utafor Troms i tre uker desember 2004. Båten har gått langs transektorer og målt intensitet av fisk ved hjelp av ekkolodd. Størrelsen på sirklene i figuren indikerer signalstyrken. Flesteparten av målingene har verdien 0, mens de positive målingene har en svært skeiv fordeling. I tillegg er det tatt trålprøver på 7 steder, markert med trekkanter. For hvert trålhal måles det lengde og alder på omtrent 100 fisk. Lengdemålingene brukes til å beregne antall fisk basert på det akustiske signalet, hvor det tas hensyn til at større fisk gir kraftigere signal. Aldersmålingene brukes til å fordele totalt antall fisk inn i årsklasser. Her er det mange kompliserende detaljer som gjør det utfordrende å lage en tilfredsstillende romlig modell.



Helse

Bioinformatikk (se egen artikkel i dette nummer av TG) utgjør hovedtyngden av aktivitetene innen helse. Andre aktiviteter i seinere tid har vært prediksjon av oppmøte av blodgivere i samarbeid med Blodbanken i Oslo, og i øyeblikket er vi i gang med et flerårig prosjekt angående modellering av sammenheng mellom forbruk av antibiotika og mengden av resistente bakterier, i samarbeid med forskere ved tre sjukehus i Osloregionen.

STATISTISK ANALYSE FOR FINANS, FORSIKRING OG RÅVAREMARKEDER

Av Kjersti Aas



De siste 6 årene har NR hatt et forsknings- og anvendelsesområde der vi har jobbet med statistiske problemstillinger innenfor finans, forsikring og råvaremarkeder. Omfanget har vært drøyt 5 årsverk årlig, og mesteparten av dette (rundt 85%) er anvendt oppdragsforskning for banker, forsikringsselskaper og råvareselskaper. Resten er grunnforskning finansiert av Norges Forskningsråd og Finansmarkedsfondet. Vi har også de siste årene holdt en del kurs for næringslivet.

Oppdragsforskningen kan munne ut i litt ulike resultater, enten er leveransen en analyse dokumentert i en rapport eller en skreddersydd programvare. Utvikling av programvare er det vi gjør klart mest av. DnB NOR og Hydro er noen av bedriftene som daglig benytter programvare NR har utviklet. For førstnevnte har vi utviklet en modell for å måle total økonomisk kapital for hele konsernet, mens et 13 årig samarbeid med Hydro har resultert i et titalls programmer for å simulere og predikere ulike typer råvarepriser, samt prise komplekse derivater.

Når det gjelder grunnforskning har vi de siste årene hovedsakelig konsentrert oss om to ulike temaer; skjeve og tunghalede fordelinger og avhengighetsstrukturer. Det har lenge vært kjent at fordelingen til avkastningene til finansielle tidsserier som aksjekurser, renter, valutakurser, råvarepriser har tunge haler og i tillegg er skjev. Dette betyr at ekstreme begivenheter forekommer langt oftere enn det normalfordelingen skulle tilsi. I tillegg skjer det mer ekstreme endringer i den ene retningen, enn i den andre.

NR har jobbet med et spesialtilfelle av den generaliserte hyperbolske fordelingen som vi har kalt GH skjev t-fordelingen. Denne fordelingen har fire parametere, som gjør den mer fleksibel enn normalfordelingen, og dermed i langt større grad i stand til å representere tunghalede og skjeve data. Simulering fra GH skjev-t fordelingen er rett-fram og det er også forholdsvis enkelt å estimere parametrene (benytter EM-algoritmen).

Modellering av avhengighetsstrukturer er et annet viktig tema innen finans. Det har etter hvert vist seg at ingen kjent parametrisk fordeling representerer multivariate finansielle data perfekt. Både i den akademiske verdenen og etter hvert også i praksis, har man derfor en stund jobbet med copulas. En copula er en multivariat fordeling med uniforme $[0,1]$ marginalfordelinger. Ved å kombinere ulike typer copulas med vidt forskjellige marginalfordelinger har man en veldig stor fleksibilitet i den multivariate modelleringen. NR har i det siste jobbet med en enda mer fleksibel multivariat struktur, som vi har kalt en "par-copula dekomposisjon". Denne strukturen, som er basert på et hittil lite kjent arbeid av Tim Bedford og Roger M. Cooke, er en dekomponering av en multivariat sannsynlighetstetthet i kun bivariate copulas og marginal-fordelinger.

NR blir fra 2007 vertsinstitusjon for et senter for forskningsdrevet innovasjon, kalt "Statistics for Innovation". Senteret, som er et samarbeid med UiO, NTNU og 12 bedriftspartnere, skal produsere statistisk metodikk som skal gi innovasjon og økonomisk vekst hos partnerne spesielt og i Norge generelt. Finans er et av fire hovedsatsings-områder i dette senteret, med DnB NOR, Gjensidige og Hydro som partnere. "Statistics for Innovation" får til sammen 80

MNOK fra Norges Forskningsråd de neste 8 årene. Vi foreventer derfor et ytterligere løft i NRs satsing på finans, forsikring og råvarer i tiden framover!

For mer informasjon om NRs aktiviteter innen dette feltet, se http://www.nr.no/pages/samba/area_bff



Temakveld i Statistisk Forening, Oslo

Høstens temakveld i Statistisk forening, avdeling Oslo, gikk av stabelen den 18. oktober. Sted for møtet var som vanlig møterommet Alfa & Omega på Norsk Regnesentral, men kun Omega-delen denne gang; det skiftes nemlig tak på bygget der NR holder til, og kraftig nedbør i høst har ført til en temmelig stor lekkasje som har gitt deler av gulvet på Alfa-delen en slags trekkspillfasong. De nesten 20 deltagerne ble derfor sittende betydelig tettere enn vanlig, men desto mer sosialt. Tema denne gang var "Multivariabel analyse med programvare".

Men først var det forfriskninger, det vil si pizza og brus. Bestiller man pizza til et møte er det jo essensielt at pizzaen kommer når den skal. Det ble derfor avkrevd et dyrt og hellig løfte av Spice om å levere pizzaene til kl 17.00. Kl 17.15 var det fortsatt ingen pizza, kun mange sultne statistikere. Med dårlig batteri på gammel mobiltelefon, en stakkars forfjamset ektemann som ringer og får brølt i øret "ER PIZZAEN HER NÅ?!" følte det litt mer hektisk enn behagelig. Men pizzaen kom da til slutt. Etter at alle var mette og fornøyde kunne vi endelig komme i gang med selve temakvelden.

Først ute var Bjørn-Helge Mevik fra UMB. Han har, sammen med Ron Wehrens, vært med på å utvikle den såkalte «pls»-pakken i R. Partial Least Squares Regression (PLSR) og Principal Component Regression (PCR) er regresjonsmetoder som er godt kjent og mye brukt i en del statistikkmiljøer og andre tilgrensende miljøer (som kjemometrimiljøet). Sentralt i vurderingen av modellene står visuell inspeksjon, og programpakken inneholder et bredt utvalg av funksjoner for å plote ulike aspekter ved PLSR- og PCR-modeller. Bjørn-Helge guidet oss gjennom en rekke funksjoner i pls, illustrert med eksempler.

Neste mann ut var Øyvind Langsrud fra Matforsk. I vitenskapelige og industrielle eksperimenter kan man gjerne ha flere responser. Man trenger da et alternativ til ordinær variansanalyse (ANOVA). Signifikanstester som baserer seg på multivariabel ANOVA (såkalt MANOVA) er ikke tilfredsstillende i tilfeller der responsene er sterkt korrelerte. Er antall responser større enn antallet observasjoner kan slike tester uansett ikke brukes. Øyvind presenterte programmet 50-50 MANOVA, som inneholder en ny metode for å håndtere slike problemer. Et viktig element i denne metoden er Principal Component analyse. Programmet inneholder tester basert på klassiske testobservatorer og deres fordelinger, og en modifisert variant av den klassiske MANOVA p-verdien blir beregnet. I tillegg til å rapportere signifikans beregnes et mål på forklart varians for alle termene i modellen. Ved bruk av rotasjonstesting beregner programmet også justerte enkeltrespons-p-verdier ("familywise error rate" eller "false discovery rate"). Øyvind ga oss en kortfattet innføring i teorien for rotasjonstesting som er beslektet med permutasjonstesting. Forskjellen er at rotasjonstesting bygger direkte på den multivariable normalfordelingen. Alle de klassiske testene (t-test, F-test, Hotelling T^2 test ...) kan sees på som spesialtilfeller av rotasjonstesting. Dette ble illustrert med et datasett.

Det hele ble avsluttet, noe på overtid som vanlig, med applaus fra lytterne og gave til foredragsholderne. Er leserne interessert i å lære mer om dette anbefales det at man ser nærmere på referansene på neste side eller tar kontakt med foredragsholderne.

Referanser:

PLS:

R-pakken «pls» implementerer PLSR og PCR, og er fritt tilgjengelig fra CRAN (<http://cran.r-project.org/>) under GNUs General Public License (GPL).

50-50 MANOVA:

Programmene er tilgjengelig på <http://www.matforsk.no/ola/program.htm>



Over til venstre: nesten 20 sultne statistikere har endelig fått pizza.

Over til høyre: to fornøyde foredragsholdere, Øyvind Langsrud (til venstre) og Bjørn Helge Mevik (til høyre).

Ingunn Fride Tvette

Statistisk Forening Trondheim

Årsmøte i Trondheim statistiske forening 2006 ble avholdt tirsdag 14 November.

17 personer møtte.

Styret bestående av Turid Follestad (leder),

Ola Diserud, Jo Eidsvik og Hugo Hammer ble gjenvalgt.

Under møtet ble nye vedtekter godkjent. De nye vedtektene gjelder fra 2007 og ligger tett opptil dem man har laget for Osloavdelingen, blant annet med valg og årsmøte kun annethvert år, det vil si 2007, 2009, etc.

En viktig del av møtet var pizzaspising og sosialt samvær.

Møtet ble avsluttet med foredrag av gjest Ting-Li Su fra Lancaster University med tittel *Application of spatial statistics to space-time disease surveillance data*.

Jo Eidsvik.

André Teigland 40 år

André Teigland fylte 40 år den første oktober 2006. André startet sine studier ved Universitetet i Oslo, Matematisk institutt, i 1985, og ble Cand. Scient i 1992 med oppgaven ”Blandingsprosesser i dynamisk lineære modeller. Noen bidrag til statistisk metodikk”.

Like etter avsluttet hovedfag, i august 1992, begynte han å jobbe ved Norsk Regnesentral (NR), i det som den gang het STAT gruppen (STAT stod for Statistisk analyse). André orienterte seg raskt mot lederoppgaver, og hadde tidlig et vikariat som forskningssjef i STAT, fra desember 1994 til ut februar 1995.

Allerede i mai samme år ble han permanent forskningssjef for det som den gang het BILD (BILD stod for Bildeanalyse). Ansatte på BILD minnes tiden som faglig interessant og meget hyggelig på det sosiale planet. Senere ble han sjef for både STAT- og BILD gruppene, som i 1998 ble sammenslått til det som nå heter SAMBA (SAMBA står for Statistisk analyse, bildegjenkjenning og mønstergjenkjenning).

André har siden den tid ledet SAMBA med stødig hånd. Gruppen har vokst raskt i antall ansatte, og med et bredt spekter av faglig interessante oppdrag. Dette er noe André har jobbet hardt og målbevisst for. Som Kjersti Aas skrev i NRs interne månedsblad ” André har alltid hatt store ambisjoner på SAMBAs og egne vegne og har aldri lagt skjul på det. Jeg tror det er en hovedårsak til at vi på NR i dag har det klart største anvendte statistikkmiljøet i Norge og sannsynligvis også i Europa”.

I tillegg til å være forskningssjef for SAMBA har André også vært leder for Norsk Statistisk Forening. Dette var i årene 2001 -2003.

Et viktig og spennende kapittel står for døren de neste årene. Som TGs trofaste lesere vet, ble NR sammen med UIO i vår tildelt det viktige Senter for forskningsdrevet innovasjon, ”Statistics for innovation”. André skal være innovasjonsleder ved dette senteret. Her vil det være bruk for administrative og diplomatiske egenskaper. Å samordne NR-ansatte, Professorer og folk fra næringslivet, og få alle til å jobbe sammen i henhold til de forpliktelser man har inngått i et slikt program, kan være litt av en utfordring. Den er André absolutt kvalifisert for.

Ellers er André en meget sporty person. Han er en habil fotballspiller, med fartstid i Fossum i andredivisjon på 80- og 90-tallet. Han er også, som mange andre på NR, meget glad i å gå på ski. I måneder med snø går tilnærmet 70 % av lunsjtiden i SAMBA til å snakke om hvor man gikk på ski sist helg (mandag), føre/ lysløypeforhold/ værmelding/ utstyr/ smørning (tirsdag-torsdag) og hvor man skal gå på ski i helga (fredag). André er en god bidragsyter i disse samtalene. I tillegg er han ikke redd for å vurdere alternativer til bilen som framkomstmiddel. I måneder uten snø sykler han gjerne til jobb. Leserne synes kanskje ikke dette høres så veldig imponerende ut, men mannen bor på Bærums Verk, veien til NR er veldig langt i fra flat, og avstanden er på nesten 20 km!

Vi på SAMBA synes vi har vært veldig heldige med sjefen vår. Han er engasjerende, flink til å delegere oppgaver og ikke minst er han meget god til å lytte når ansatte har noe på hjertet. Vi gleder oss til mange år med André som sjef, og gratulerer med jubileet!



En ung og litt eldre utgave av jubilanten.
Bilde til venstre: André som nyutdannet statistiker i 1992.
Bilde til høyre: André som forskningssjef i 2002.

Ingunn Fride Tvette

Pusleri nr. 19 - løsning

Oppgaven var:

La $X_1, X_2, \dots, X_n$ være uavhengige variable med samme forventning og samme endelige varians. Variansen til summen S_n vokser da lineært med n mot uendelig.

Spørsmål: Er det slik at variansen til kvadratroten av summen vokser med kvadratroten av n mot uendelig? Hvis ikke, hva skjer? Hva med variansen til kubikkroten av summen?

Løsning:

For at oppgaven skal gi mening, må vi selvsagt forutsette ikke-negative variable. Selv om $\text{var} S_n$ er proporsjonal med n , vil $\text{var} S_n^{1/2}$ ikke være proporsjonal med $n^{1/2}$, men faktisk holde seg begrenset, mens $\text{var} S_n^{1/4}$ vil gå mot null med n !

For begrunnelse kan like gjerne gå rett på $\text{var} S_n^r$ for $r \in [0, 1]$. For ikke-negative variable gjelder generelt at (bevis nedenfor):

$$\text{var} S^r \leq \frac{\text{var} S}{(ES)^{2(1-r)}}$$

Anvender vi dette på S_n får vi

$$\text{var} S_n^r \leq \frac{n\sigma^2}{(n\mu)^{2(1-r)}} = n^{-1+2r} \frac{\sigma^2}{\mu^{2(1-r)}}$$

For $r=1/2$ avhenger skranken ikke av n , mens for $r \in [0, 1/2]$ vil den gå mot null.

Bevis for ulikheten: For $t \geq 0$ gjelder at $|t^r - 1| \leq |t - 1|$ og følgelig $(t^r - 1)^2 \leq (t - 1)^2$. For $a \geq 0$ og $b > 0$ gjelder derfor $((a/b)^r - 1)^2 \leq (a/b - 1)^2$ dvs. $(a^r - b^r)^2 \leq (a - b)^2 / b^{2(1-r)}$. Vi får derfor

$$\text{var} S^r = E(S^r - ES^r) \leq E(S^r - (ES)^r)^2 \leq \frac{E(S - ES)^2}{(ES)^{2(1-r)}}$$

der den første ulikheten følger av at forventning minimerer forventet kvadratisk tap.

Pusleri nr. 20

Dreidel er et gammelt spill med jødisk opprinnelse, der en bruker en firesidig terning (tetraeder) med sider markert med N, G, H og S. Spillet tillater et vilkårlig antall spillere p , og starter med at alle legger en pengeenhet i potten. Så kastes dreidelen av spillerne etter tur. Resultatet for spiller ved de ulike utfall er N="Nun"=ingen gevinst, G="Gimel"= Hele potten, H="Hay"=Halve potten, S="Shin"=Legg en pengeenhet i potten. Hver gang potten tømmes bidrar alle spillere med en pengeenhet, og spillet fortsetter.

Hva er forventet utbetaling for spilleren i n 'te spilleomgang?

Hva er forventet utbetaling til første spiller som tømmer potten?

*Hva er forventet nettogevinst for hver spiller for en spillerunde, dvs. inntil potten tømmes?

Spillet foregår typisk ved at hver spiller har et gitt beløp, for eksempel 20 pengeenheter, og varer inntil en spiller sitter igjen med alt. Dette gir opphav til ytterligere og mer utfordrende sannsynlighetsproblemer.

Forslag til løsning av problemene, og gjerne forslag og løsning på ytterligere problemer sendes til jostein.lillestol@nhh.no.



Kristian Hovde Liland, en tilfeldig statistiker



Fram til nå

Ingen som kjenner meg vil påstå at jeg har villet bli statistiker hele mitt liv. Helt opp mot ungdomsskolen sto bussjåfør høyest på ønskelista. Til gjengjeld er det vel ingen som ville ha protestert om jeg hadde sagt at statistikk var min store drøm. Jeg har fått merkelappene teknisk naturvitenskapelig og ordensmenneske i en utdanningsundersøkelse og kan vel ikke nekte for at disse betegnelse passer svært godt. Selv etter fullført videregående skole hadde jeg det ikke helt klart for meg hva jeg skulle bli. Noen år med klassisk gitarspilling gjorde at jeg valgte å ta et år på Toneheim Folkehøgskole. Dette, i kombinasjon med at jeg traff en ung pike fra Kristiansand, ga nok mersmak til at jeg prøvespilte ved Musikkonservatoriet i Kristiansand, ble tatt opp og holdt på i to år med faglærerutdanning musikk.

Etter å ha funnet ut at musikkskolelærer ikke egentlig var det jeg ville bli, var det naturlige valget å gå over på bachelor i informatikk og matematiske metoder ved Høgskolen i Agder. Her møtte jeg på statistikken skikkelig for første gang gjennom Henrik Dahl. Dette var et svært positivt møte, spesielt da jeg fikk sjansen til å være timelærer i statistikk før endt bachelor. En kammerat av meg ville til Ås for å ta master i anvendt matematikk og statistikk. Dette syntes jeg virket som en god idé, så jeg kastet meg med. Kompetansen og miljøet på Ås er unikt, så da jeg mot slutten av masteren øynet muligheten til å få en stilling som PhD-stipendiat var jeg ikke i tvil om at det var dette jeg ville satse på.

Framover

Mastergraden min ble veiledet av Ulf Indahl og omhandlet klassifikasjonsproblemer med hovedvekt på utvidelse av Ulfs Powered Partial Least Squares til diskriminanttilfellet (PPLS-DA). Samarbeidet fungerte godt, og for tiden viderefører og omarbeider vi blant annet metoder og ideer som kom fram i prosessen rundt masteroppgaven. I veiledergruppa mi sitter i tillegg Are Aastveit og leder Trygve Almøy. Jeg burde altså være i trygge hender i de kommende årene.

Framover håper jeg å fortsette samarbeidet med Ulf i en periode og innlede samarbeid med Bjørn-Helge Mevik med problemstillinger blant annet knyttet mot hans prosjekt mot EWOS Inovation i Dirdal. Her er baselinjekorreksjon og skiftproblematikk i NMR- og spektroskopispektre naturlige områder å fokusere på. Siden PPLS-DA ser ut til å ha sin største styrke på nettopp spektroskopidata, vil altså doktorgraden min ta for seg hele forløpet fra forbehandling til endelig analyse av slike data.

På privaten jobber jeg i perioder som konsulent for Norges Geotekniske Institutt mot Petromaks. Jeg har stor glede av å kunne teste mine kunnskaper på et praktisk prosjekt der jeg er nødt til å kunne presentere stoffet på en konkret og overbevisende måte for folk som ikke har statistisk bakgrunn. Det føler jeg også er nyttig i forbindelse med artikkelsskriving, da jeg blir bevisstgjort på at ikke alle andre tenker i samme baner som meg og kjenner min side av problemstillingen.

Statistisk analyse av kredittrisiko

Doktoravhandling - planlagt innlevering november 2007

Daniel Berg

Norsk Regnesentral / Universitetet i Oslo



Avhandlingen er en anvendt studie av ulike aspekter ved kredittrisiko. Den planlegges satt sammen av to deler. I den første delen fokuserer jeg på modeller for prediksjon av konkurs. Denne delen er basert på et stort datasett som inneholder regnskapene til alle små- og mellomstore selskaper registrert i Brønnøysund registeret i perioden 1995-2001, samt konkursdata for tilsvarende periode. Delen vil bestå av to artikler:

1. Bankruptcy prediction by generalised additive models

I denne første artikkelen sammenligner jeg flere populære modeller for prediksjon av konkurs, basert på regnskapsvariable. Generaliserte additive modeller blir introdusert som en overlegen modell for konkursprediksjon i forhold til diskriminantanalyse, generaliserte lineære modeller og nevrale nettverk. Viktige praktiske tema blir diskutert: tids horisont for konkurs, modellens forringelse over tid og antall år man går tilbake i tid ved konstruksjonen av treningsdata.

2. Bankruptcy prediction: spatial effects in industry classification

Medforfattere: Rada Dakovic, Claudia Czado.

I denne artikkelen ser vi på industrieffekter i konkursprediksjon. Vi ser på industri klassifiseringen av selskaper som en romlig modell, et tre, og undersøker om det en slik romlig modell finner signifikante industri effekter i dataene, etter at de allerede er justert for slike effekter i henhold til tradisjonelle metoder.

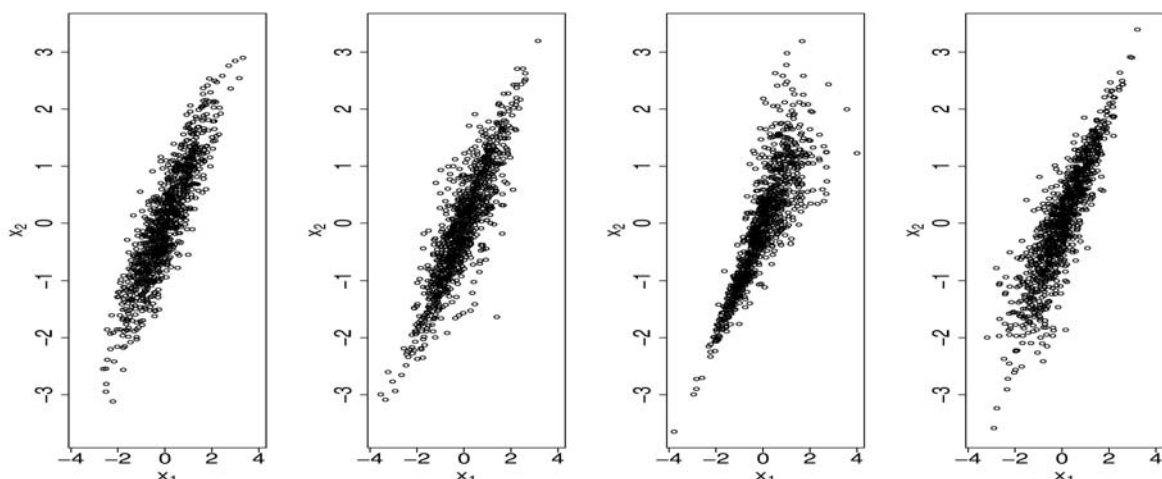
Del to omhandler avhengighetsmodellering, spesielt såkalte copulas.

En copula er en multivariat fordeling hvor marginene er standard uniformt fordelt på $[0,1]$.

Enhver multivariat fordeling består av en copula og marginalfordelinger. Et sentralt og illustrerende resultat er som følger. Den simultane tettheten, f , til en kontinuerlig fordelingsfunksjon F med strengt økende marginalfordelinger $F_1, \dots, F_d$ kan skrives som

$$f(x_1, \dots, x_d) = c[F_1(x_1), \dots, F_d(x_d)] \cdot f_1(x_1) \cdots f_d(x_d),$$

for en d -dimensjonal copula tetthet c . Figuren under viser 1000 simulerte observasjoner fra fire ulike copulas, alle med standard normalfordelte marginalfordelinger. Denne figuren illustrerer litt av fleksibiliteten man har med copulas.



Del to vil bestå av fire artikler:

3. A copula goodness-of-fit test based on the probability integral transform

Medforfatter: Henrik Bakken.

I denne artikkelen blir en ny goodness-of-fit test for copulas presentert. Testen er veldig fleksibel mhp vekting av ulike områder av copulan og umeriske tester viser at den har god styrke for haletyngde og skjevhet. Testen blir anvendt på porteføljer bestående av amerikanske aksjer og resultater viser at den gaussiske copulan blir forkastet mens student-t copulan ser ut til å gi en mye bedre tilpasning.

4. Copula goodness-of-fit tests: a comparative study

Medforfatter: Henrik Bakken.

Her blir flere ulike goodness-of-fit tester for copulas sammenlignet. Vi oppsummerer styrker, svakheter, testprosedyrer og undersøker deres statistiske styrke for ulike null- og alternative hypoteser.

5. Construction of higher dimensional dependency structures

Her tar jeg for meg konstruksjoner av høyere dimensjonale avhengighetsstrukturer der de ulike metodene som er foreslått blir oppsummert. En ny simuleringsalgoritme for en av strukturene blir foreslått. De ulike konstruksjonene blir testet med goodness-of-fit tester og til slutt anvendt på en finansportefølje for å undersøke deres effekt på sentrale risikomål som f.eks. value-at-risk.

6. The probability integral transform and its use in copula goodness-of-fit testing

Til slutt presenteres en anvendt studie av en metode (PIT) som er mye brukt i goodness-of-fit tester for copulas. Metoden kan tolkes som det inverse av simulering. Anta at vi har et datasett $\mathbf{X}$ med en viss avhengighetsstruktur. Vi konstruerer så, under en antagelse om en parametrisk copula, et iid $U(0,1)$ datasett. Vi presenterer teorien bak denne transformasjonen, en del sentrale eksempler og noen numeriske tester.

I tillegg til de to hoveddelene vil avhandlingen inkludere en manual til en R-pakke jeg har bygd underveis. Denne inneholder et knippe metoder for simulering og estimering av copulas, goodness-of-fit testing av copulas og simulering og estimering av høyere dimensjonale avhengighetsstrukturer.

Arbeidet er utført ved Norsk Regnesentral og Matematisk Institutt, Universitetet i Oslo, med professor Fred Espen Benth (UiO) som hovedveileder, sjefsforsker Xeni Kristine Dimakos (NR) som medveileder og assisterende forskningsdirektør Kjersti Aas (NR) som viktig bidragsyter. Arbeidet er en del av prosjektet StAR, finansiert av Norsk forskningsråd.

Verdensberømte vitenskapsfolk har jukset med forskning siden antikken

Av Øystein Evandt



Verden ble i oktober 2005 presentert for en oppsiktsvekkende artikkel i det prestisjetunge medisinske tidsskriftet *The Lancet*, der det konkluderes med at en del vanlig brukte smertestilende medisiner reduserer faren for kreft i munnhulen, men øker faren for hjerte- og karsykdommer. Artikkelen, og ”forskningen” den var basert på, bygget imidlertid på fabrikkerte data og var derfor et bedrageri.

Da bedrageriet ble avslørt i januar 2006 uttalte sjefsredaktør Richard Horton i *The Lancet* at ”Dette er den største svindelen fra en forsker verden noen gang har sett”. Denne svindelen har imidlertid mange konkurrenter til å fortjene en slik beskrivelse. Det gjelder også publikasjoner skrevet av personer som tilhører de fremste vitenskapsfolkene gjennom historien.

Fusk i forskning er grundig behandlet i boken ”Betrayers of the Truth – Fraud and Deceit in Science” (Sannhetens forredere – svindel og bedrag i vitenskap) av William Broad og Nicholas Wade, utgitt i 1985 på Oxford University Press. Noen eksempler på fusk utført av verdensberømte vitenskapsfolk beskrives nedenfor, bl.a. ved sitater fra denne boken.

- Claudius Ptolemeius (85-165 e. Kr), kjent som antikkens største astronom, gjorde ikke de fleste av observasjonene sine om natten ved kysten av Egypt, men på dagtid i det store biblioteket i Alexandria. Han tok utgangspunkt i arbeidet til en gresk astronom og lot som om det var sitt eget.

En rekke av de stjernene Ptolemeius angivelig observerte, er ikke engang synlige fra kysten av Egypt!

- Galileo Galilei (1564-1642) kalles ofte for grunnleggeren av moderne vitenskap fordi han gjorde seg til talsmann for det syn at eksperimenter, ikke arbeidene til Aristoteles (384-322 f. Kr), skulle danne grunnlaget for å dømme om hva som er sant og hva som er usant. Men kollegene til den italienske 1600-talls fysikeren hadde store vanskeligheter med å reproducere resultatene hans, og tvilte sterkt på at han hadde gjennomført alle de eksperimentene han refererte til.
- Isac Newton (1642-1727), guttegeniet som formulerte lovene for gravitasjon, baserte sitt hovedverk på usømmelig fusk for å gi inntrykk av at arbeidet hans ga bedre grunnlag for å forutsi hendelser enn det i virkeligheten gjorde.

I det anerkjente tidsskriftet *Science*, i 1973, beskrives Newton som ”en mester i å manipulere data slik at de passet inn i hans beregninger”.

- John Dalton (1766-1844), den store 1800-talls kjemikeren som oppdaget lovene for kjemiske bindinger og påviste eksistensen av ulike typer atomer, publiserte elegante resultater som ingen kjemikere i dag er i stand til å reproducere.
- Gregor Mendel (1822-1884), den østerrikske munken som grunnla den genetiske vitenskap, publiserte arbeid om erter der de rapporterte statistiske resultatene stemmer for godt til at de kan være sanne.

En av teoriene til Mendel tilsa at, under visse forutsetninger, vil erterplanter som har hhv røde og hvite blomster forekomme i forholdet 3 til 1, i den forstand at hver gang en slik plante blir til, er sannsynligheten 75 % for at den får røde blomster og 25 % for at den får hvite blomster. Det innebærer imidlertid slett ikke at når, for eksempel, 100 planter blir til, så vil nødvendigvis akkurat 75 av disse ha røde blomster og akkurat 25 ha hvite blomster. Mendel rapporterte imidlertid mange datasett der prosentandelene av avkom med ulike egenskaper, hos ulike typer levende organismer, *nesten* sammenfalt med de teoretiske sannsynlighetene, noe som er helt usannsynlig, med mindre data er manipulert.

- Den amerikanske fysikeren Robert Milikan (1868-1953) fikk Nobelprisen [i 1923] for å være den første som målte størrelsen av den elektriske ladningen på et elektron. Men Milikan feiltolket bevisst arbeidet sitt slik at de eksperimentelle resultatene virket mer overbevisende enn de i virkeligheten var.

Milikan sløyfet de målingene som passet dårligst med teorien hans, og baserte sine beregninger på de målingene som passet best med teorien.

I boken til Broad og Wade beskrives også mange eksempler på juks i forskning i nyere tid, innenfor ulike felt. Et av eksemplene, innen medisin fra begynnelsen av 1980 årene, er en svindel som kan sammenlignes med artikkelen i The Lancet fra oktober 2005. En ung amerikansk forsker ved Cornell University i Itacha i staten New York, Marc Spector, og en anerkjent professor i biokjemi samme sted, Efraim Racker, utviklet en enkel og elegant teori om årsaker til kreft. Teorien, som ble kalt Racker-Spector teorien, hadde støtte i eksperimenter som Spector utførte i 1980. Blant medisinere med kjennskap til teorien var det en utbredt oppfatning at den ville komme til å revolusjonere behandling av kreft, og mange av dem antok at Racker og Spector ville få Nobelprisen for arbeidet. En kollega av Spector fattet imidlertid mistanke om at resultatene fra eksperimentene var fabrikkert. Han prøvde å reproducere eksperimentene, men klarte det ikke. Han var tilbakeholdende med å melde fra om sin mistanke, av frykt for konsekvensene dersom han tok feil. Han ble imidlertid etter hvert sikker i sin sak, og meldte fra. Jukset ble bevist, selv om Spector aldri innrømmet det. Hans karriere tok imidlertid en brå slutt. Professor Racker var uskyldig i det som skjedde. Han hadde hatt full tillit til Spector. Det hadde aldri falt Racker inn at Spector kunne ha jukset med eksperimentene.

Broad og Wade, som har utført grundige og omfattende studier av juks i forskning, herunder hvordan juks kan skjules, er relativt lite optimistiske når det gjelder mengden av juks som ikke blir oppdaget, eller som blir oppdaget etter lang tid. Men hva er det som får vitenskapsfolk, som i prinsippet har viet sitt liv til å søke sannhet, til å jukse?

Det er vel grunn til å tro at en viss uredelighet i vitenskapelig virksomhet forekommer, slik det utvilsomt forekommer i for eksempel politikk, juridisk praksis, religion og forretningsvirksomhet. Fusk i forskning kan ta mange former. Når det gjelder data, kan det bl.a. fuskes ved å fabrikke et datasett, eller ved å forandre litt på noen målinger og/eller sløyfe noen målinger for å få et datasett til å passe bedre med egen teori. Andre vanlige

former for juks er bl.a. å unnlate å referere til arbeid som andre har gjort, men som ligger til grunn for eget arbeid, og å stjele andres arbeid før det er publisert. I blant stjeles til og med andres arbeid *etter* at det er publisert. Det kan ta lang tid før dette oppdages dersom det dreier seg om perifere forskningsområder og det stjalne arbeidet blir ”publisert” om igjen i lite kjente tidsskrift.

Publisering har ikke bare den funksjon at forskningsresultater blir gjort kjent. Det er også et middel for forskere til å gjøre karriere, få bevilgninger og å bli med i interessante og meritterende fremtidige forskningsprosjekter. Vi vet at det er stor forskjell på mennesker når det gjelder ærlighet. Noen forskere lar seg helt sikkert lettere friste av ”snarveier” i karrieren enn andre, hvis de mener at sjansen for å bli oppdaget er liten. Det er nødvendig å være først ute med nye resultater. Å være nummer to gir ingen uttelling.



Gruppe om forløpsanalyse ved Senter for grunnforskning

Av Sven Ove Samuelsen

I perioden august 2005 – juni 2006 var en av gruppene ved Senter for grunnforskning eller Centre of advanced study (CAS) (<http://www.cas.uio.no/index.php>) ledet av Odd Aalen og Ørnulf Borgan. Tittel for gruppens prosjekt var *Statistical Analysis of Complex Event History Data*.

CAS holder til Det norske vitenskapsakademi's ærverdige lokaler i Drammensveien i Oslo. Hvert år inviteres tre forskergrupper til å arbeide ett år ved Senteret, vanligvis én gruppe innen hvert av feltene: humaniora / teologi, samfunnsvitenskap / jus, naturvitenskap / matematikk. Hver gruppe skal ha internasjonal sammensetning og skal ledes av en eller to anerkjente norske forskere. Inviterte utenlandske forskere får romslig støtte til reise og opphold, og kan om nødvendig kjøpes fri fra undervisning og administrasjon ved sine respektive universiteter eller forskningsinstitusjoner. Inviterte norske forskere får etter avtale mellom senteret og universitetene innvilget ekstra forskningsfri for perioden de er invitert til senteret. Dessuten stiller senteret til rådighet leiligheter sentralt i Oslo, og det legges til rette for at deltagerne kan ha med familie.

Sentrene finansieres av regjeringen, med bidrag fra Forskningsrådet, Vitenskapsakademiet samt norske universiteter. Sentrene skal være "devoted to ensuring that outstanding, innovative researchers get the resources they need to make further scientific advances. In other words, CAS is intended to serve as a catalyst to enhance the quality of the Norwegian research system, to raise the standards of Norwegian basic and interdisciplinary research to the highest international level."

Gruppen til Ørnulf og Odd hadde deltagere fra Danmark, Holland, Tyskland, England samt fra Stavanger, Bergen, Trondheim og Oslo. En oversikt over deltagerne finnes på <http://www.cas.uio.no/Groups/0506Statistical/index.html>. Selv var jeg så heldig å få delta i gruppen hele høsten 2005. Siden jeg bor i Oslo var omstillingen mindre enn for tilreisende, men det var uansett en stimulerende miljøforandring og nyttig å bli bedre kjent med mange kolleger. Lunsj og diverse utflukter var dessuten felles med de andre gruppene og forhindret at vi låste oss helt fast i vår egen faglige lille verden. Gruppen fra samfunnsvitenskap hadde som tema miljø-økonomi og deres perspektiv sammenfalt rimelig godt med vårt. Temaet for gruppen fra humaniora derimot var "narrativ teori" med fokus på forfatterne Joseph Conrad og Kafka samt på fortellinger om Holocaust og hadde derfor, ihvertfall ikke i utgangspunktet, så mye med statistikk å gjøre.

Faglig sett er jeg også fornøyd, selv om man alltid håper å få gjort mer. Men i hvert fall påbegynte jeg og ble nesten ferdig med å skrive en artikkel i løpet av halvåret. Et ytterligere prosjekt som det hadde gått litt trått med fikk også den nødvendige siste dytten. Dessuten gikk det ganske mye tid til en såkalt "fast-track" revisjon for en artikkel til et medisinsk tidsskrift. Vel, i det minste for meg kvalifiserer dette som et meget produktivt halvår.

Et prosjekt ved Senter for grunnforskning passer kanskje ikke for alle typer fag, men det passet utmerket for vår gruppe om forløpsanalyse. Jeg vil tro at det tilsvarende vil fungere meget godt for andre statistiske temaer, og jeg vil derfor anbefale andre statistikere å søke om å få ha et prosjekt ved senteret.

Statoil once said, we have started a run, and we are running, but we shall not stop too early. Our project portfolio describes our current thoughts and plans and will adapt along the next months and years, in terms of focus, research teams, funding. Nevertheless, we are very proud of our collection of 19 projects that we wish to start in 2007. There are strong projects in each of the four sectors, marine, finance, health, petroleum. There are several important methodological projects, motivated by and feeding to innovation ideas. We have created the terrain for research team that include senior and junior staff from UiO, NTNU and NR, current students and staff from our non-academic partners. Some projects are very preliminary and explorative, more working groups around central themes; other projects are already precise and concrete. Some projects are focused on a specific important aspect of a general application, other projects are grand challenges, which will be ongoing for many years and with large research teams. This agrees with the original NFR application and with our Research Charter, which we have written and approved, that describes the characteristics of the research we wish to house at (sfi)2.

Without going into details, here is a table with the names of the 19 projects.

P1	FindOil	Statoil, NR, UiO, NTNU
P2	TotalRisk	DnBNOR, NR, UiO
P3	StatInsure	Gjensidige, NR, UiO
P4	Commodity	Hydro, NR, UiO, NTNU
P5	SampleFish	HI, NR, UiO
P6	ByCatch	HI, NR, UiO
P7	VariousMarine	HI, NR, UiO
P8	MissingStripes	Biomolex, NR, UiO
P9	NetAlgebra	Pubgene, RR, UiO, NR
P10	GenomeAtlas	DNR-RR, UiO, NTNU, NR
P11	MetaAnalysis	Smerud, NR, UiO
P12	CompareSeq	Sencel, NR
P13	MaleContracept	Spermatech, NR

More methodological projects include:

P14	ToolHSSS	NR, UiO, NTNU
P15	Multicompare	NTNU, NR, UiO
P16	EventAnalysis	UiO
P17	TailDepend	NR, UiO, NTNU
P18	LongPredict	NR, UiO, NTNU
P19	GreenHouse	ALL

(Soon we will have material describing these projects on the web; for the time being, guessing is allowed!)

It is statistical methodology that represents the meeting point for projects originating in the four sectors, and the aim is to strengthen very much in the future synergies, building bridges that will enhance research plans and results and lead to a good outcome of the conspicuous investments given to (sfi)2.

I will try to keep the statistics community in Norway regularly informed about the activities at (sfi)2. It is very important to underline that the centre is open, all research results will be published. At all our seminars and workshop everybody is warmly welcomed. We also wish to establish new contacts and links, hence do not hesitate to contact me in the future.

Best regards, Arnaldo Frigessi

Kvinner bak tallene

Av Ingunn Fride Tvette

Stipendiater har også ferie. 5 uker i året, som alle andre arbeidstakere. I hvert fall i teorien. Selv hadde jeg 8 dager på hytta i sommer. I 7-8 grader og tilnærmet kontinuerlig pøsregn. Den eneste som smilte fra øre til øre var min fiskegale ektemann, for nå var det endelig laks i elva. (Ja, det er sant som Aftenposten skriver, menn fisker og damer plukker bær. Vi kvinner har nemlig foretatt en rask vurdering av sannsynlighet og mulig fortjeneste for fisk i forhold til bær, og kommet fram til at forventet gevinst er mye større ved bærplukking). Vi, som nøyde oss med toppen et par timer med fiskestang per dag i ferien, måtte prøve å aktivisere oss så godt vi kunne på andre måter.

Når man er på ferie og det regner så tett at man ikke engang skimter Hurtigruta der ute på havet, og Hjortefot og Senjas årbøker 1984-1990 for lengst er lest om igjen, er de gamle ukebladene gode å ty til.

Mine besteforeldre hadde visse oppfatninger av hva som utgjorde et dannet og opplyst hjem. Dette omfattet abonnement på ukebladet URD. For de som aldri har hørt om URD kan jeg referere fra Universitetsbibliotekets hjemmesider. URD var et blad som "...var preget av en kristelig grunnholdning med tro på fremskritt og folkeopplysning, brakte mer stoff om kunst og kultur enn det som var vanlig i dameblader".

For meg hører sommerferier, regnvær og gamle ukeblader naturlig sammen. Halvveis ut i min andre Pepita kjekspakke, blant artikler om "den genuine grønlandske anorakk ... er som kjent fremstilt av et så merkelig materiale som tygget fuglehud" (som kjent?!), "Mennesket Fartein Valen" og annonser fra for eksempel Oslo Plissefabrikk, finner jeg en lang artikkel med tittelen "Kvinner bak tallene". En artikkel om kvinner og statistikk! I 1958! Journalist Liv Florelius har intervjuet underdirektør Signy Arctander i Statistisk Sentralbyrå, og oppdaget at "statistikk er langt fra kjedelig", og "anbefaler de som tror Statistisk sentralbyrå er en gørr [sic] papirmølle å ta en tur inn på underdirektørens kontor". Hun er også villig til "å gå så langt som til å påstå at statistikk kan være morsom lesning".

Leserne i 1958 (og her i 2006) får vite at mens "Det faste Personalet utgjorde i 1879 foruden Direktøren og de to Sekretærer, 2 Fuldmægtige samt 10 Kopister og Ekstraskrivere, derunder medregnet 2 damer..." var andelen kvinner en helt annen i 1958. Nemlig "Av 289 funksjonærer ved årsskiftet [januar 1958] var 194 kvinner, 6 av dem sitter i høyere stillinger som underdirektør, byråsjef, hovedbibliotekar og førstesekretærer. Blant 39 sekretærer er 4 kvinner, og de teller ikke mindre enn 42 av 44 fullmektiger i særklasse og førstefullmektiger. De fleste av disse er arbeidsledere. Alle spesialstillinger i Statistisk Sentralbyrå er besatt med kvinner". Journalist Florelius kommer også med en fremtidsvisjon i sin artikkel: "Det er ikke verst, selv om neste generasjon sikkert vil mene at vår tid var utrolig bakstreversk og fordomsfull nettopp på dette punkt. Men til den tid er det nok større tilgang på kvinnelige økonomer også!"

Der satt jeg, en onsdag formiddag i juli, med Pepita smuler og regnvær rundt meg, temmelig overrasket. Og litt imponert. Kanskje i likhet med mange av min generasjon ville jeg ikke gjettet at kvinneandelen i Statistisk Sentralbyrå var såpass høy i 1958.

I følge Likestillingssenteret, forgjengeren til Likestillings- og diskrimineringsombudet, var Norge på 1960-tallet blant landene i Vesten med færrest kvinnelige yrkesaktive. Og blant de kvinner som var yrkesaktive var ikke realfagsrelaterte yrker det dominerende valg. Derfor synes jeg dette er ekstra artig lesing. I årsmeldingen til SSB for 2005 kan vi lese at ved inngangen til 2006 var andelen kvinner blant lavere nivåes saksbehandlere 83 %, høyere nivåes saksbehandlere 53 %, høyere fagstillinger/forskere 37 % og i ledelsen 31 %. Ikke så dårlig det heller, selv om det sikkert kunne vært bedre. Her på NRs statistikkavdeling (SAMBA) utgjør kvinner også en betydelig andel, i underkant av 40 %. Ved Avdeling for Statistikk, UIO, har stipendiatandelen for kvinner det siste året vært på hele 50 %! Vi hører stadig vekk at det er få kvinner som velger realfag, men statistikk synes kanskje å være mindre hardt rammet enn en del andre realfag? Det har blitt pekt på, for eksempel i Aftenposten 19 April 2006, at jenter "velger vekk de harde fagene til fordel for de «myke», de som har med mennesker og ikke tall å gjøre". Selv om statistikk så absolutt kan sies å omfatte tall oppfattes det kanskje av mange jenter som nokså "mykt", i den forstand at det kan anvendes på så mange spennende fagfelt, og dermed er mer motiverende? Ikke vet jeg. Jeg er i hvert fall glad for at såpass mange jenter har valgt statistikkfaget som levevei, og at vi gjennom studietid og stipendiattid har hatt et meget godt miljø, sosialt så vel som faglig. Gutter så vel som jenter. Kanskje vi kan håpe at dameblader i 2007 igjen kan ha reportasjer à la "Kvinner bak tallene"?

<http://www.ssb.no>

<http://forbruker.no/jobbogstudier/studier/article1283073.ece>

<http://www.likestilling.no/undertema.html?id=48>

<http://www.nb.no/nbvev/eksternvev/html/nyhetsformidlere.html>

Liv Florelius, Kvinner bak tallene, URD, 15ende mars 1958

Under og på neste side: de tre første bildene fra artikkelen i Urd (med henvisning ovenfor).



Bilde over: deltakerne i det første internasjonale statistikkmøtet i Norge, i Kristiania, den 8ende september 1899. Legg merke til antrekket. Det må vel kunne sies å være litt mer formelt enn ved nåtidens statistikermøter.



Bildene over: kvinner med tall anno 1958. Til venstre: Fullmektig Elsa Bertheau overvåker hullkortavdelingen. Til høyre: Punchedame Tove Gulbrandsen overfører statistiske opplysninger fra skjema til hullkort.



Bilde over: kvinne med laks (i regnvær) anno 2006 (undertegnede).

Nytt fra UIO

Ansettelseser

Pål Henriksen ble ansatt som fakultetsstipendiat i forsikring 1. november 2006

Uteksaminerte Masterstudenter i Modellering og dataanalyse juli - oktober 2006

Raabe Dag: Tilskuerforsikring i norsk toppfotball: Tariffering og metode
Veileder: Erik Bølviken

Haugen Kristoffer: Mathematical Models for the Measure of Risk and Applications
Veileder: Bernt Øksendal

Lunde John Alexander: Norwegian interest rate derivatives
Veileder: Erik Bølviken

Mohammed Hayat: Estimation of the binomial parameters N and p when both are unknown
Veiledere: Nils Lid Hjort/ Ingrid Glad

Statistical Research Report

Oktober: S. Nygård, Ø. Borgan, O. Chr. Lingjærde, H. L. Størvold:
Partial least squares Cox regression on genomic data.

Nytt fra NTNU

Ny PhD, oktober 2006:

Hans Christian Karlsen
Statistics of wave induced nonlinear loads and responses
Veileder: Arvid Næss

Nye Master, Sommer 2006:

Arne Andresen
Romlig interpolering av nedbør
Veileder: John Tyssedal

Sanjay Manikarnika
A General Face Recondition System
Veileder: John Tyssedal

Bjarne Sæther

Rekursiv blokkoppdatering av Isingmodellen

Veileder: Håkon Tjelmeland

Tekniske Rapporter 2006:

Statistical hypothesis testing of association between two lists of genes for a given gene class,
Clara-Cecilie Günther, Mette Langaas and Stian Lydersen

Approximate Bayesian Inference in Spatial
Generalized Linear Mixed Models,
Jo Eidsvik, Sara Martino and Håvard Rue

The Value of Information in Spatial Decision Making,
Debarun Bhattacharjya, Jo Eidsvik and Tapan Mukerji

Nytt fra UMB**Nye Dr. Scient:**

Heidi Cecilie Henriksen Disputas 14. oktober 2005

Avhandlingens tittel:

Empirical models in sulphite pulp production

Emne for forelesning:

Standardisering av respons og validering av multivariable prediksjonsmetoder

Veiledere:

Professor Are H. Aastveit IKBM, UMB

Professor Tormod Næs Matforsk

Bedømmelseskommité:

Paul Geladi, Head of Research NIR Infrared Center of Excellence, The unit
of Biomass Technology and Chemistry

Kari Thyholt, Mills DA

Førsteamanuensis, Trygve Almøy, IKBM, UMB

Kjetil Jørgensen Disputas 28. oktober 2005

Avhandlingens tittel:

Combining designed experiments and spectroscopic measurements for improved process
modeling and interpretation

Emne for forelesning:

The use of experimental design for optimization of industrial processes

Veiledere:

Are Aastveit UMB,

Tormod Næs MATFORSK

Bedømmelseskommité:

Professor Tom Fearn, University College London, UK

Ass. Professor Lars Nørgaard, Dept. of Food Science, KVL, Denmark

Førsteamanuensis Ulf Indahl, IKBM, UMB

Guri Feten Disputas 24 mars 2006

Dissertation title:

Statistical analysis of microarray data

Lecture title:

Class prediction based on microarray expression data

Supervisor:

Associate Professor Trygve Almøy

Evaluation committee:

Professor Mats Rudemo, Chalmers Univ. of Technology, Sweden

Post doc Dirk Repsilber, Potsdam University, Germany

Associate Professor Torfinn Torp, IKBM, UMB

Henrik Rene' Cederkvist Disputas 31 mars 2006

Avhandlingens tittel:

Optimization, Prediction and Restricted Variable Selection in Linear Models.

Emne for forelesning:

Bootstrapmetodikk - med vekt på ferske bidrag.

Veiledere:

Professor Are Halvor Aastveit, IKBM

Professor Tormod Næs, Matforsk

Bedømmelseskomite:

1. Professor Per M. Bruun Brockhoff, Technical University of Denmark

2. Førsteamanuensis John Sølve Tyssedal, NTNU

3. Førsteamanuensis Trygve Almøy, UMB

Ingrid Måge Disputas 25. august 2006

Avhandlingens tittel:

Modelling and optimisation of industrial processes with raw material variation

Emne for forelesning:

Forsøksplanlegging ved industriprosesser

Veiledere:

Professor Are Aastveit (UMB)

Professor Tormod Næs (MATFORSK)

Bedømmelseskomite:

Professor Inge Helland (UiO)

Professor Rasmus Bro (KVL)

Førsteamanuensis Ulf Indahl (UMB)

Nye master

Kristian Hovde Liland Juni 2006

Avhandlingens tittel:

PLS_DA: En vri på Partial Least Square Discriminant Analysis

Veiledere:

Førsteamanuensis Ulf Indahl, UMB

Professor Tormod Næs, MATFORSK

Sensor

Professor Tormod Næs, MATFORSK

Trond Lien Juni 2006

Normalisering av microarray data

Veiledere:

Førsteamanuensis Trygve Almøy, UMB

Førsteamanuensis Lars Snipen, UMB

Sensor

Første amanuensis Anja Braathen, UiO

Nytt fra UIB

To master i statistikk i høstsemestert 2006:

Lars Aga Reisæter:

Hedging av rentederivat

Veileder:

Jostein Paulsen

Jorunn Slagstad:

Blanda modeller i R

Veileder:

Hans Julius Skaug

Den 13. desember vil **Astrid Lunde** disputere for PhD graden med avhandlingen:

Mixed-effects models with application to insurance and genetic epidemiology.

Veileder: Jostein Paulsen

Komite: Magne Thorsen, Universitetet i Oslo

Per Mykland, University of Chicago

Trygve Nilsen, Universitetet i Bergen

Astrid Lunde er tilknyttet Seksjon for epidemiologi og medisinsk statistikk ved UiB.

Nyhet fra Vest:

Statistikernettverk etableres i Helse Vest.

Statistisk metode er viktig for medisinsk og helsefaglig forskning og utvikling, og i Helse Vest finnes det statistikere spredt på mange avdelinger. Kompetansesenter for klinisk forskning, Haukeland universitetssykehus har nå tatt initiativet til å etablere et nettverk for denne faggruppen i regionen og vil arrangere Det første Helse Vest-statistikermøte i Bergen den 12. januar 2007. Interesserte statistikere i Helse Vest som ikke allerede er blitt kontaktet kan kontakte biostatistiker Geir Egil Eide for nærmere informasjon (Geir.Egil.Eide@Haukeland.No).

Vennlig hilsen Geir.

Nasjonal møteserie i eVitenskap 2007-2010
Dr. Holms Hotel, Geilo, 28 jan. – 2 feb., 2007



Vinterskole: Monte Carlo metoder

Forelesere:

- Morten Hjorth-Jensen (UiO)
- Geir Evensen (Norsk Hydro)
- Laurent Bertino (Nansensenteret)
- Fred Espen Benth (UiO)
- Håkon Tjelmeland (NTNU)

Faglig ansvarlig: Einar Rønquist (NTNU)

Passer for:

doktorgrads- og hovedfagsstudenter

Det er gode muligheter for å gå/stå på ski!

Påmeldingsfrist: 15. desember 2006

<http://www.sintef.no/eVITA>

Forskermøte

- Presentasjon av eVITA programmet
- Norsk infrastruktur for tungregning
- Nye prosjekter i eVITA
 - Discrete optimization methods in maritime and road-based transportation
 - The power of scale-space methods and Gaussian Markov random fields applied in climatology and medicine
 - Coral distribution and conditions for growth in Norwegian waters
 - Parallel algorithms for combinatorial scientific computing
 - ...
- Inviterte foredragsholdere
 - Peter Deuffhard (ZIB, Berlin)
 - Risto Nieminen (Helsinki)
 - Arnoldo Frigessi (UiO/NR)
 - Kenneth Ruud (UiT)

Course number: O-22856

Analysis of Repeated Categorical Response Data

20. - 23. August 2007

Target group: Ph.D. students in biostatistics, medical researchers with good knowledge of statistics, and others with an interest in statistical analysis. The course will be given in English.

Teaching goal: The aim of this course is to provide theoretical and practical training analysis of repeated categorical data. The course will acquaint students with the modern modeling techniques. Examples will be emphasized, and in the computer lab students will analyze data using software. Instruction will be in SAS, but students are also welcome to use R.

Requirements: It is required that the students are familiar with likelihood ideas of inference and regression analysis, and some previous background in logistic regression would be helpful. The course will use the SAS statistical language, but some guidance will also be given about R functions for performing the analyses.

Contents:

1. Review of logistic regression
 - Interpretation of model parameters
 - Maximum likelihood model fitting
 - Extensions for multinomial responses
2. Models for Matched Pairs
 - Comparing dependent proportions: McNemar test et al.
 - Conditional vs marginal models for binary matched pairs
 - Comparing margins of square contingency tables
3. Marginal Models
 - Marginal logit models for repeated binary response
 - Maximum likelihood (ML) and weighted least squares (WLS)
 - Generalized estimating equations (GEE) approach
 - Cumulative logit modeling of repeated ordinal responses
4. Conditional Models: GLMs with Random Effects
 - Conditional logistic regression of clustered binary data
 - Generalized linear mixed models (GLMMs): ML fitting and inference
 - Logistic GLMMs for clustered binary data
5. Additional Topics about Mixture Models
 - GLMMs for ordinal data
 - Multivariate GLMMs for multi-center clinical trials
 - Beta-binomial model for clustered binary data: Teratology studies

Location: Domus Medica, Universitetet i Oslo

Course committee: Morten Fagerland, Center for Clinical Research, Ullevål University Hospital, Stian Lydersen, Unit for Applied Clinical Research, University of Science and Technology, Petter Laake, Department of Biostatistics, University of Oslo, e-mail: petter.laake@medisin.uio.no, Øivind Skare, Department of Biostatistics, University of Oslo.

Instructor: Professor Alan Agresti, Department of Statistics University of Florida, USA

Registration: Den norske lægeforening, Kontor for legers videre- og etterutdanning, Postboks 1152 Sentrum, 0107 Oslo.

Registration e-mail: koordinatorkontoret.oslo@legeforeningen.no. In the registration please make reference to course number O-22856

Deadline: 20. June 2007

Number of participants: 30

Course hours: 28

Credits: 4

Course fee: NOK 2800

Course literature: Alan Agresti: Categorical data analysis, 2nd edition, Wiley. The textbook may be ordered from <http://www.amazon.co.uk/>

Other information: The course will be given in English. There will be a take home exam after the course.

Programme:

Monday 20. August 2007

08.30-09.00: Registration
09.00-11.00: Review of logistic regression
11.00-12.00: PC-lab
12.00-13.00: Lunch
13.00-15.30: Models for matched pairs
15.30-16.30: PC-lab

Tuesday 21. August 2007:

09.00-11.00: Marginal models
11.00-12.00: PC-lab
12.00-13.00: Lunch
13.00-15.30: Marginal models
15.30-16.30: PC-lab

Wednesday 22. August 2007:

09.00-11.00: Conditional Models: GLMs with Random Effects
11.00-12.00: PC-lab
12.00-13.00: Lunch
13.00-15.30: Conditional Models: GLMs with Random Effects
15.30-16.30: PC-lab

Thursday 23. August 2007:

09.00-11.00: Additional Topics about Mixture Models
11.00-12.00: PC-lab
12.00-13.00: Lunch
13.00-15.30: Additional Topics about Mixture Models
15.30-16.30: PC-lab

PREKONFERANSEKURS

GENERALIZED LINEAR LATENT AND MIXED MODELS

SOMMARØYA (Tromsø)

18.-19. juni 2007

Norsk statistisk forening ønsker velkommen til prekonferansekurs før statistikarmøtet på Sommarøya (Tromsø)

Tidspunkt

Kurset startar med lunsj måndag 18. juni klokka 13.00 og blir avslutta tysdag 19. juni klokka 12.00. Statistikarmøtet startar på same stad etter at kurset er ferdig.

Lokalisering

Sommarøy kurs og konferansesenter, sjå omtale under presentasjonen av statistikarmøtet.

Foredragshaldarar

Anders Skrondal <<mailto:a.skrondal@lse.ac.uk>> ,
London School of Economics & Norwegian Institute of Public Health.
<http://www.gllamm.org/anders.html>

Sophia Rabe-Hesketh <<mailto:sophiarh@berkeley.edu>> ,
University of California, Berkeley & Institute of Education, University of
London

Tema for kurset

Generalized linear mixed (or multilevel) models (GLMMs) are useful for longitudinal data, cluster-randomized trials, surveys with cluster-sampling, genetic studies, meta analysis, etc. The random effects in GLMMs are latent variables representing between-cluster variability and inducing within-cluster dependence. Latent variables also are used often to represent true values of variables measured with error. Measurement models specifying the relationship between measured and latent variables can form part of regression models, giving structural equation models (SEMs). SEMs also can be used to model dependence between processes. Taking a unified view of these models is beneficial because developments for one model-type are applicable to other model-types and the same software can be used for seemingly different models. The course will be structured in three parts: GLMMs, measurement models, and SEMs. Ideas will be developed by starting from simple versions of the models and motivating extensions through examples. Methods of estimation and prediction will be surveyed. Real applications will be considered from different disciplines.

Litt reklame

Kurset er tidligere halde ved to "Joint Statistical Meetings" (Toronto og Seattle) der det var eitt av dei mest populære og med best evaluering. Elles er kurset gitt m.a. i Firenze, Pacific Grove, Washington DC, Munster, Montreal, Research Triangle Park og Boston.

Litteratur

Skrondal, A. and Rabe-Hasketh, S. (2004): Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Chapman and Hall/CRC Press. ISBN: 1-58488-000-7.

Målgrupper

Kurset vil vere av interesse for alle som arbeider med statistikk, epidemiologi, biologi, økologi, økonomi m.fl.

Påmelding

Påmelding og hotellbestilling skjer via nettsidene for statistikarmøtet:
<http://uit.no/matstat/statistikermote>. Påmeldingsfristen er 1. juni, og kursavgifta er 1000 kroner.
Hotellovernatting for 1 dag med heilpensjon vil komme på omtrent 1400 kroner.

Transport

Det vil bli sett opp (gratis) buss frå Tromsø lufthamn til Sommarøya klokka 11.45 måndag 18/6. Den vil då korrespondere med flyruter frå Oslo og Trondheim med ankomst mellom 11.00 og 11.15. Bussturen tar omtrent 45 minutt.
tysdag 19/6 blir det sett opp buss frå Sommarøya klokka 12.30 slik at den korresponderer med fly til Oslo og Trondheim med avgang 14.00.

Informasjon

For ytterligare informasjon sjå nettsida <http://uit.no/matstat/forkurs> der det etter kvart vil bli lagt ut meir materiale om kurset.

Velkommen!

Det 14. norske statistikarmøtet

SOMMARØYA (Tromsø)

19.-21 juni 2007

Norsk statistisk forening avdeling Trondheim ønskjer velkommen til:

Det 14. norske statistikarmøtet, 19.-21. juni 2007
på Sommarøy kurs og konferansesenter

Tidspunkt

Møtet startar med lunsj tirsdag 19. juni klokka 13.00 og blir avslutta torsdag 21. juni klokka 12.30.

Lokalisering

Sommarøy kurs og konferansesenter, www.sommaroy.no, ligg på Sommarøya, som er lokalisert på vestsida av Kvaløya omtrent 50 km vest for Tromsø, altså omtrent ytst mot havgapet. Det er frå gammalt av eit fiskeribasert samfunn der det i dag òg blir drive ein del med turisme. Øya har brusamband med Kvaløya og vidare til Tromsø. På sommarstid er det og mogeleg å ta ferje frå Brensholmen (5 km frå Sommarøya) til Senja og følgje ytre veg vidare derfrå til Vesterålen og Lofoten utan å dra innom fastlandet.

Program

Sjå eigen presentasjon av programkomite og program nedanfor.

Påmelding

Påmelding og hotellbestilling skjer via konferansenettsidene <http://uit.no/matstat/statistikermote> Påmeldingsfristen er 1. mai, og konferanseavgifta er sett til 1000 kroner. Det er mogeleg å melde seg på fram til 1. juni, då med ei konferanseavgift på 1250 kroner.

For studentar er konferanseavgifta sett til 500 for påmelding før 1. mai, og 750 mellom 1. mai og 1. juni. Det vil i tillegg bli gitt støtte til dekning av (delar av) utgiftene til reise og opphald for studentar som vil delta, meir informasjon om dette vil kome på nettsidene.

Hotellovernatting for 2 dagar med heilpensjon vil komme på omtrent 2800 kroner. I tillegg kan dei som ønskjer det bestille lunsj siste dagen, dette er ikkje inkludert i pensjonsprisen.

Transport

Det vil bli sett opp (gratis) buss frå Tromsø lufthamn til Sommarøya klokka 11.45 tirsdag 19/6. Den vil då korrespondere med flyruter frå Oslo og Trondheim med ankomst mellom 11.00 og 11.15. Bussturen tar omtrent 45 minutt.

Torsdag 21/6 blir det sett opp buss frå Sommarøya klokka 12.30 slik at den korresponderer med fly til Oslo og Trondheim med avgang 14.00.

Det går òg rutebussar frå flyplassen til Sommarøya, sjå www.tromsbuss.no (rute 1) for nærmare informasjon.

Sosiale aktivitetar

På ettermiddagen onsdag 20. vil det bli eit tilbod om anten tur med båt eller eit opplegg med «naturbaserte spenningsaktivitetar og moro» frå <http://www.naturogutfordring.no>. Påmelding og betaling for den enkelte deltakar skjer via konferansenettsidene. På kvelden den 20. er det òg konferansemiddag.

Forkurs

Det vil bli arrangert eit kurs på Somarøy kurs og konferansesenter frå måndag 18/6 til statistikarmøtet startar tisdag den 19. Temaet for kurset er «Generalized Linear Latent and Mixed Models». Kurshaldarar er Anders Skrondal og Sophia Rabe-Hesketh. Sjøå elles eigen presentasjon.

Arrangementskomité

Georg Elvebakk (Tromsø), leiar
Sigrunn Sørbye (Tromsø), økonomi
Elinor Ytterstad (Tromsø), påmelding, transport, websider
Turid Follestad (Trondheim)

Vi ønskjer velkommen til det første norske statistikarmøtet nord for polarsirkelen!

DET 14. NORSKE STATISTIKERMØTET: INFORMASJON FRA PROGRAMKOMITEKOMITÉEN

Det 14. norske statistikermøtet arrangeres på Sommarøy utenfor Tromsø, 19.-21.juni 2006. Informasjon fra arrangementskomitéen (bla. om påmelding og kurs med Skrondal og Rabe-Hesketh i forkant av møtet), finner du et annet sted i dette TG.

Programkomitéen er ansvarlig for det vitenskapelige programmet for møtet, og har følgende medlemmer:

- Mette Langaas, Institutt for matematiske fag, NTNU, Mette.Langaas@math.ntnu.no (leder).
- Fred Godtliebsen, Institutt for matematikk og statistikk, Universitetet i Tromsø.
- Ivar Heuch, Matematisk institutt, Universitetet i Bergen.
- Magne Aldrin, Norsk Regnesentral, Oslo.

Vi ønsker at det vitenskapelige programmet skal inneholde følgende elementer:

1. Inviterte foredrag innen aktuelle statistiske temaer:
 - Her jobber vi med ulike tema, og vil komme med program i neste TG. Har du forslag til tema og foredragsholder så ta kontakt.
2. Foredrag fra hovedfagsstudenter og stipendiater:
 - Her ønsker vi at veiledere tar kontakt med oss og forteller om spennende hovedoppgaver og PhD arbeid som er underveis eller avsluttet.
3. Påmeldte foredrag:
 - Tittel sendes innen 15.mars og kort abstract innen 1.mai til programkomitéen.
4. Postere/plakater og "læremiddel" under temaet "fargerik statistikk og visualisert undervisning":
 - Har du funnet på noe spennende i statistikkforelesningene dine, så ønsker vi at du deler det med andre statistikere - ved å vise ideen din på stand eller som en plakat.
 - Forslag: todimensjonale fordelinger i plastelina, spill Monty Hall med skoer, to-utvalgs t-test brukt på seigmennsstrekking, Poisson-sokk, Florence Nightingale - statistikk i det godes tjeneste.

Innspill og påmelding av foredrag sendes Programkomitéen v/Mette Langaas,
email: Mette.Langaas@math.ntnu.no.

Vi ser frem til å se deg på det 14.norske statistikermøtet, og vil arbeide for at konferansen får et spennende vitenskapelig program!

Med vennlig hilsen
Programkomitéen

Norsk Regnesentral (NR) er nasjonalt ledende innen anvendt statistisk oppdragsforskning. Våre 40 statistikere arbeider med en rekke spennende anvendelser innen petroleum, finans, forsikring, råvaremarkeder, marine ressurser og bioinformatikk.



Fra og med 2007 er NR vertsinstitusjon for **Statistics for Innovation**, et senter for forskningsdrevet innovasjon (SFI) utnevnt av Norges forskningsråd. Statistisk kompetanse blir stadig mer etterspurt i finansinstitusjoner, forsikringsselskaper og ellers i næringslivet. Med SFien og økende oppdragsmengde, trenger vi å styrke oss ytterligere og søker derfor

2 STATISTIKERE

Vi søker etter forskere:

- Med solide teoretiske kunnskaper innen statistisk analyse og modellering (minst hovedfag/master i statistikk eller industriell matematikk med hovedvekt på statistikk)
- Erfaring fra og/eller interesse for å jobbe med anvendelser innen finans, forsikring og råvaremarkeder
- Programmeringskompetanse (C/C++, S-plus)
- Erfaring fra og/eller interesse for anvendelser innen finans, forsikring og råvaremarkeder
- Entusiasme og stå-på vilje!

NR tilbyr spennende og varierte prosjekter og hyggelige kollegaer i et unikt fagmiljø. Innen finans og forsikringsområdet jobber vi med Norges ledende miljøer. Mer informasjon om dette området finnes på http://www.nr.no/pages/samba/area_bff.

I tillegg kan vi tilby konkurransedyktig lønn, kollektiv pensjons- og forsikringsordning, bedriftshelsetjeneste og mulighet for barnehageplass. NR holder til ved Forskningsparken på Blindern i Oslo.

For nærmere informasjon kontakt forskningssjef André Teigland, 2285 2675, eller assisterende forskningssjef og ansvarlig for NRs aktiviteter innen finans, forsikring og råvaremarkeder Kjersti Aas, 2285 2694.

Søknad med cv, vitnemål og attester sendes HR-kontoret innen 31.12.2006, helst pr e-post til human.resources@nr.no.



Norsk Regnesentral

Gaustadalléen 23

Postboks 114 Blindern

NO-0314 Oslo

(+47) 22 85 25 00

nr@nr.no

www.nr.no

Statistics for Innovation is one of the 14 Norwegian Centres for Research-based Innovation, with the aim to enhance the capability of the Norwegian economy to innovate by focusing on long-term research, forging alliances between research-intensive enterprises and prominent research groups. **Statistics for Innovation** will develop core statistical methodologies, strategically necessary to achieve innovation goals in four key sectors: petroleum, finance, marine and health.

The centre will operate from 1.1.2007 for eight years, based at the Norwegian Computing Center (NR) in Oslo and in partnership with the University of Oslo (UiO), the Norwegian University of Science and Technology (NTNU) and 11 partners: Statoil, DnBNOR, Gjensidige, Hydro, Institute of Marine Research, Sencel, Biomolex, Pubgene, Rikshospitalet-Radiumhospitalet, Smerud Medical research and Spermatech.

Statistics for Innovation is funded by the Research Council of Norway and by the partners, with an annual budget of about 4 mill Euro. About 40 researchers will actively join the research projects of the centre. **Statistics for Innovation** participates to the UiO and NTNU training programmes and will fund many PhD grants and postdocs. The centre considers applications of outstanding candidates ongoing.



1-2 Postdoctoral fellowships (Ref. SFI-01)

1-3 Ph.D research fellowships (Ref. SFI-02)

available at the Centre for Research-based Innovation (SFI) Statistics for Innovation,
Norwegian Computing Center and University of Oslo, Norway

1-2 Postdoctoral fellowships (up to two years):

Applicants should have a doctoral degree in statistics, actuarial sciences, mathematics, computer science or equivalent. It is acceptable if the degree will be obtained within 3 months from the deadline of the present call. The ideal candidate has a strong CV, with methodological and/or applied expertise, an outstanding potential and genuine interest to develop new statistical methodology aimed to innovation. Fellows will be key scientific personnel in the centre, engaging in methodological and collaborative research. Salary is approx. 48 000 – 51 000 Euro annually according to qualifications.

1-3 Ph.D fellowships (three year appointments):

Applicants should have a master of science degree in statistics, actuarial sciences, mathematics, computer science, or equivalent. It is acceptable if the degree will be obtained within 3 months from the deadline of the present call. The ideal candidate has some experience in methodological and /or applied statistics, and an outstanding potential and genuine interest to develop statistical methodology aimed to innovation. The research project will be defined together with the candidate and the supervising team, including staff from the university and the other partners. Salary is approx. 38 000 – 42 000 Euro annually according to qualifications.

General: The successful candidates will join a very active scientific group in Oslo conducting research in statistical methods directly relevant to innovation in the four sectors and at our partners. Active research areas include biostatistics, statistical genomics, geostatistics, time series, simulation based inference, Bayesian methods, rare events, complex models, huge datasets, data integration, event history analysis, among others. The posts will give ambitious researchers an outstanding opportunity to develop their scientific careers in a vibrant interdisciplinary centre. The Norwegian Computing Center and the University of Oslo seek to increase the number of female scientists. Women are therefore particularly encouraged to apply.

For more information visit our website <http://www.nr.no/sfi> or contact Professor Arnoldo Frigessi (frigessi@medisin.uio.no).

Applicants should submit for all positions: A letter of application, indicating the position they apply for, a brief summary of research interests, curriculum vitae with full list of publications with internet links to recent papers and the names of at least two references. Applicants for the Ph.Ds should also send copies of grades.

Applications should be sent to: human.resources@nr.no, preferably as a single PDF file.

Application deadline: January 10, 2007



Medlemskap i Norsk Statistisk Forening

Kontingenten er kr. 200 per år for regulære medlemmer og kr. 50 for studenter (til og med hovedfag). Korrespondanse om medlemskap og innbetaling av kontingent skal skje til foreningens kasserer:

Norsk Statistisk Forening v/ Hilde Grude Borgos
Schlumberger Stavanger Research
P.O.Box 8013, N-4068 Stavanger, Norway
Tel: (+47) 51946466, Fax: (+47) 51946001
Email: HBorgos@stavanger.oilfield.slb.com

Innbetalinger gjøres på foreningens konto 0530 2773299. Medlemmer som er usikre på når de betalte kontingent sist, kan nederst på adresselappen for Tilfeldig Gang se hvilket år siste betaling er registrert for. Vær oppmerksom på at adresselappene for bladet kan ligge noe etter ajourføringen av medlemsregistret, slik at nylig foretatte innbetalinger ikke vil være med her.

Nye medlemmer melder seg til kassereren og får så automatisk tilsendt Tilfeldig Gang fra første påfølgende nummer. Ved innmelding skal både epost- og vanlig adresse oppgis.

Annen korrespondanse med foreningen skal foregå med formannen:

Sven Ove Samuelsen
Matematisk institutt, UiO
Postboks 1053 Blindern
0316 Oslo
Emailadresse: nsf-styret@math.uio.no

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 Sven Ove Samuelsen (osamuels@math.uio.no) i hende innen 01.03.2007.

Artikkelstoff vil normalt bli lagt ut i nettversjonen av TG. De som ønsker å reservere seg mot dette, må gi beskjed ved innsendelse av stoff.

Årskontingent Norsk Statistisk Forening 2006

Medlemskontingent i NSF er kr. 200,- pr. år for ordinære medlemmer og kr. 50,- for studenter (til og med hovedfag), og betales inn på kontonummer 0530.27.73299. Mange har allerede betalt medlemskontingent for 2006, men fortsatt er det en stor andel som mangler. Vennligst sjekk adresselappen på baksiden av bladet hvilket år vi sist har registrert innbetaling fra deg. De som ikke har betalt for tidligere år gjør dette sammen med kontingenten for 2006. Dersom du mener betalingsinformasjonen ikke stemmer, gi beskjed til hborgos@slb.com.

NB: Dersom din arbeidsgiver betaler kontingenten, sørg for at navnet ditt kommer med på innbetalingen.

Hva gjør vi med pengene?

De største utgiftspostene for NSF er overføringer til lokallagene i Oslo, Bergen og Trondheim, NSF's medlemskap i the European Mathematical Society (EMS) og the International Statistical Institute (ISI), utgivelse av TG, og økonomiske bidrag i forbindelse med diverse arrangementer. For i fjor gjelder dette det norske statistikermøte og the European Meeting of Statisticians i juli, hvor vi blant annet sponset deltakelse for studenter. I år har vi bidratt økonomisk til jubileumsseminaret for 50 år med hovedfag i matematisk statistikk, som ble arrangert av UiO og NSF på Voksenåsen 22. mai. Det siste året har vi også brukt penger på å fornye hjemmesiden vår, se <http://www.math.uio.no/nsf/>.

Medlemsregister

For at medlemsregisteret skal være best mulig oppdatert er det viktig at du oppgir evt. ny post/epost-adresse enten ved å sende meg en epost eller oppgi dette på innbetalingsgiroen. Vi har opprettet en egen mailingliste for NSF's medlemmer, nsf-medlemmer@math.uio.no. Dersom du ikke mottar epost på denne lista, send meg din oppdaterte epostadresse.

PÅ FORHÅND TAKK!

Vennlig hilsen
Hilde Grude Borgos
Kasserer NSF
Epost: hborgos@slb.com. Tlf: 51 94 64 66

Betalingsinformasjon		GIRO		Betalings- frist
Kontingent NSF 2006: kr. 200 (kr. 50 for stud.)		Underskrift ved girering		
Betalt av		Betalt til		
		NSF v/ Hilde Grude Borgos Schlumberger Stavanger Research Postboks 8013 4068 Stavanger		
Belast konto			Kvittering tilbake	
Kundeidentifikasjon (KID)	Kroner	Øre	Til konto	Blankettnummer
		< >	0530 27 73299	< >