ISSN 0803-8953

TILFELDIG GANG
Nr 1, årgang 15
mars 1998

Utgitt av Norsk Statistisk Forening

INNHOLD

	Side
Fra redaktøren: Leiv Solheim	1
Fra formannen: Odd Aalen	2
Faglig stoff	
En vurdering av lærebøkene i statistikk i videregående skole	3
Statistikkoppgaver i videregående skole - gode idéer ønskes	9
Betinget sannsynlighet gitt en nullhendelse	10
Statistikeren med forventet gevinst som innsats	17
Yrkesveiledning for videregående skole	18
Rapport fra ISI 51. sesjon i Istanbul, 18.-26. august	19
Rapport fra 1. International Conference on Mathematical Methods in Reliability	21
Konferanser og kurs	
Short cource in biostatistics - Ås, Norge , 3. – 5. juni 1998	22
17th Nordic Conference on Mathematical Statistics	25
21. Nordiske Statistikermøte	29
Konferanseoversikt	37
Nytt fra universiteter og høyskoler	
Melding fra lokalforeningen i Trondheim	38
Nytt fra gruppene for statistikk - NTNU	38
Bokanmeldelser	
Donald A. Mackenzie: Statistics in Britain 1865-1930. The Social Construction of Scientific Knowledge.	40
Olav Kallenberg: Foundations of Modern Probability, Springer, New York 1997	41
Meldinger	
Medlemskap i Norsk Statistisk Forening	42
Bidrag til Tilfeldig Gang	42

Fra redaktøren: Leiv Solheim

Noe lettsindig sa jeg meg villig til å ta på meg ansvaret for å redigere TG på årsmøtet til Norsk Statistisk Forening på årsmøtet på Selbu 9. juni i 1997.

Litt seint har jeg kanskje oppdaget alvoret i dette siden mye av stoffet til dette nummeret kom inn i desember 97. Jeg lot dessverre stoffet ligge litt i dvale over på det nye året og etter hvert som jeg skulle forsøke å sette det hele sammen kom nye meldinger flytende, mens de daglige oppgaver tok mye av tiden. Nåvel - her er da TG endelig selv om det ikke ble nr 4 i 97, men nr 1 i 98. Jeg hadde håpet at det skulle blitt påskestoff, men det ser dessverre ut til at det ble heller stoff til å finne tilbake til hverdagen etter ski, sol og fritid i påskefjellet eller på andre kanter.

En travel konferansetid står for tur i Norden. De første dagene i juni innbyr danskene Nordens statistikere til konferanse i matematisk statistikk i Elsinore. To måneder seinere er Norges tur til å være vertskap for de samme lands statistikkprodusenter og brukere på Lillehammer i begynnelsen av august. Samtidig viser disse to konferansene både bredden i faget statistikk, dvs metodene og tabellene, men samtidig også avstanden mellom disse to sidene ved begrepet statistikk. For over 20 år siden holdt tidligere direktør i Statistisk sentralbyrå, Petter Jakob Bjerve , sin presidenttale ved ISI-møtet i 1977 der temaet hans nettopp var kløften mellom teoretisk statistikk på den ene siden og statistikkbyråene eller produksjonen av offisiell statistikk på den annen side. Utviklingen har ikke ført til noen tilnærming kanskje heller snarere omvendt.

Siden jeg selv arbeider som (metode)statistiker i Statistisk sentralbyrå ser jeg dette ofte daglig. Det burde derfor være en utfordring for Norsk statistisk forening å arbeide for både å utnytte bedre den bredde og tyngde faget burde ha i samfunnsdebatten og skape mer samarbeid mellom de statistiske metoder og teori på den ene siden og produksjon av statistikk(offisiell statistikk) på den andre siden.

Informasjonssamfunnet er over oss og i øyeblikket er det Internett som bombarderer oss nærmest i senk med informasjon. Norsk statistisk forening har opprettet en egen hjemmeside

http://www.math.uio.no/nst/

Det er meningen å lage forskjellige temasider og det er allerede laget en side for yrkesveiledning

http://www.math.uio.no/nst/yrkesveiledning/

se forøvrig en egen notis om dette.

I neste nummer av TG vil vi komme tilbake med mer informasjon om innholdet i vår hjemmeside, men TG vil etter hvert bli lagt ut på hjemmesiden

Ellers er det en interessant artikkel om statistikkundervisningen i videregående skole som jeg håper vil pirre statistikere til å sette igang en diskusjon om fagets plass og rolle i skolen og på universiteter og høyskoler.

I det redaktøren gjør seg klar til å reise på fjellet lover han på tro å ære at neste nummer skal komme ut til medlemmene innen sommerferien!

Leiv Solheim

Fra formannen:

Hvis man i sine tunge stunder lurer på om det er noe vits i å drive med statistikk, kan man for eksempel titte litt på forsidene til Dagbladet. Man vil da se at det til stadighet er oppslag relatert til helse, og at grunnlaget nesten alltid er statistiske data. En dag var det oppslag om «Dramatisk økning av brystkreft blant unge», med referanse til data fra Kreftregisteret. (Oppslaget var klemt inne mellom tøandre oppslag med noe mindre typer: «Vebjørns ville fest» og «Teigen vil knuse Grand Prix».) En annen dag kan man lese med krigstyper: «Livsfarlig sykehus-slurv: 250 000 pasientprøver er feil». Lignende, mer og mindre vederheftige oppslag om medisin og helse basert på talldata finner man svært hyppig. Mens de typiske helseoppslag tradisjonelt har vært enkeltmenneskers skjebner (bare se på ukebladene) har Dagbladet funnet ut at skal man si noe av generell gyldighet, så er det ikke nok med enkeltskjebnene, man må beskrive *mange* skjebner og summere opp deres erfaring, og hva er da bedre egnet enn tallene.

I disse oppslagene ligger det altså egentlig en anerkjennelse av vårt fag som vi kan ta til hjertet. Denne anerkjennelsen finner man selvfølgelig ikke bare på det medisinske fagfeltet, den faktiske respekten for statistikken er stigende på mange områder. Skal man komme med almengyldige utsagn, så er innsamling av tildels store datamengder og analyse og presentasjon av disse en av hovedmetodene. Trass i sine svakheter er likevel den kvantitative beskrivelsen av virkeligheten en av de desiderte hovedkilder til kunnskap om verden omkring oss.

Dagbladet har skjønt dette. Vårt problem er å sørge for at andre også begriper det. Hva for eksempel med ungdommen? I forrige TG diskuterte jeg den tildels begredelige presentasjon av statistikk i lærebøkene i videregående skole. En videre problemstilling her er hvordan vi kan få ungdom i videregående skole til å tenke på det å være *statistiker* som et attraktivt yrke. Vi vet alle at yrkesbetegnelsen vår ikke får ungpikehjertene til å banke eller svigermors øyne til å skinne. Dette ville nok kreve en revolusjon vi ikke kan håpe på. Men hvordan få presentert faget på en attraktiv måte slik at det kan sees på som et spennende alternativ ved siden av informatikk og teknisk fysikk, for eksempel?

Det er nå nedsatt et utvalg under Norsk statistisk forening som skal arbeide med yrkesveiledning overfor videregående skole. Utvalget ledes av Vigdis By Kampenes, og hennes samarbeidspartnere er Petter Prøsch og Therese Wettre. En av idéene er å utarbeide materiale for presentasjon på Internett. De tar sikkert villig imot innspill fra andre også. En spennende og attraktiv presentasjon på Internett av vårt yrke kunne bety mye og jeg ser på dette som av de viktigste konkrete oppgaver for vår forening i tiden fremover.

Odd Aalen

En vurdering av lærebøkene i statistikk i videregående skole

Ivar Heuch(Universitetet i Bergen), Jostein Lillestøl(Norges Handelshøyskole) og Henrik Dahl(Høgskolen i Agder)

Reform 94 innførte nye planer for undervisningen i videregåennde skole.Som en del av matematikkfaget i andre og spesielt tredje klasse kom det inn ganske mange emner i sannsynlighetsregning og statistikk.

Planene ble tatt i bruk gradvis på forskjellige nivåer, slik at den nye undervisningen på siste klassetrinn begynte i skoleåret 1996-97. Lærebøkene forelå i revidert form bare kort tid før undervisningsstart høsten 1996.

Det ble tidlig klart at fremstillingen i noen av bøkene var mangelfull, men det var ikke selvsagt hvordan statistikerne som gruppe burde reagere. Som medlemmer av undervisningsutvalget i Norsk Statistisk Forening gikk vi gjennom de aktuelle lærebøkene i løpet av det første skoleåret og utarbeidet sett med kommentarer rettet mot forfattere og forleggere. Tankegangen var at en kunne oppnå mer ved å kommunisere direkte med de ansvarlige for opplegget enn ved å gå ut med å pen kritikk. Hvert enkelt forfatterteam fikk bare oversendt bemerkningene som dreide seg om eget læreverk. Mange av problemene som ble tatt opp, hang sammen med hbox{mindre} misforståelser eller uheldig fremstilling av stoffet. Det forekom imidlertid også virkelig alvorlige feil som ble påpekt. Vi vil her omtale de generelle sidene ved vår kritikk. Det gjenstår å se hvilke reaksjoner de mer dyptgående kommentarene vil føre til fra lærebokforfatternes side. Saken vil fortsatt ha stor betydning for statistikkfaget i tiden fremover.

Lærebøkene,[2], [3], [4] og [5], som ble gjennomgått, gjelder for videregående kurs II i studieretning for allmenne og økonomisk/administrative fag, 3ts MX. Det finnes tilsvarende lærebøker i 3ts MY, men de følger hovedsakelig samme opplegg, slik at våre synspunkter stort sett også har gyldighet for dem.

Generelle problemer i statistikkundervisningen

Det er lett å se generelle problemer som oppstå når det plutselig blir innført et nytt emne som statistikk på det aktuelle nivået. Lærebokforfatterne har utvilsomt hatt en vanskelig oppgave, med liten tid til rådighet ved utarbeidelsen av manuskriptene. Ved vurderingen av bøkene må man også ta hensyn til at Reform 94 medførte andre vesentlige forandringer i opplegget for faget matematikk.

Med de nye fagplanene i matematikk skal mange emner behandles på kort tid. Dette blir også omtalt som et stort problem av lærerne i faget. Det skaper spesielle vanskeligheter i statistikkdelen, der det må innføres mange nye begrep som setter store krav til forståelsen. Fra annen elementærundervisning vet vi hvor vanskelig det er å gi en generell innføring i selve tankegangen bak de statistiske metodene. Læreplanene [1] er ganske ambisiøse i så måte, og de omtaler utføring av hypotesetesting og konstruksjon av konfidensintervaller som aktuelle hovedmomenter. Fremstillingen skal riktignok konsentreres om binomiske situasjoner og Gauss-modeller med kjent standardavvik, men de underliggende idéene må likevel forklares. I mange tilfeller får en inntrykk av at lærebokforfatterne går svært grundig til verks ved forklaring av enkelte emner, mens andre blir mer stemoderlig behandlet. Et typisk tilfelle er binomisk testing, som får langt fyldigere omtale enn tankegangen bak konfidensintervallene.

Den praktiske undervisningen vil ofte bli konsentrert om regneteknikkene som skal anvendes i bestemte situasjoner. Behandlingen av statistiske metoder krever mer omtale i vanlig tekst enn andre, rent matematiske teknikker. Denne omtalen faller ikke alltid heldig ut i de foreliggende lærebøkene. En kan forestille seg at mange elever vil lete gjennom teksten etter lett anvendelige oppskrifter, kanskje etter først å ha strevet med mer filosofiske forklaringer av grunnleggende idéer. Av og til får en inntrykk av at fremstillingen blir rotete fordi forfatterne selv ikke har klare forestillinger om behandlingen av konkrete problemstillinger.

Forfatterne prøver å sette emnene inn i en naturlig ramme, men nettopp fordi det er så mange temaer som skal tas opp, er det ofte vanskelig å se de logiske sammenhengene. Den fundamentale forbindelsen mellom sannsylighetsregning og statistikk blir uklar, selv om det gis en ganske bred innføring i sannsynlighetsregning. For elever som konsentrerer seg om konkrete regneteknikker, kan den logiske oppbygningen spille mindre rolle, men man skulle likevel håpe at varierte eksempler og oppgaver kunne gi større forståelse. Noen ganger virker det som om forfatterne har hentet aktuelt stoff fra større læreverk uten å få med seg kontinuiteten i fremstillingen.

Problemene med å få lærebøkene ferdig i tide, gjorde det sikkert vanskelig å finne realistiske eksempler som passer med teknikkene som skal innlæres. Noen ganske få omfattende eksempler blir gitt med omtale av reelle problemstillinger der en virkelig trenger statistiske metoder. Det er likevel ikke lett å tilpasse den praktiske beskrivelsen til det beskjedne utvalget av metoder som elevene får. Dette utvalget er fastlagt i læreplanen og kan ikke være forfatternes ansvar. Derimot stilles de overfor en stor utfordring når det skal vises at teknikkene kan være praktisk nyttige.

Et nytt trekk ved matematikkundervisningen er den utvidede bruken av kalkulator i mange sammenhenger. Dette kan få store konsekvenser for undervisningsmetodene i matematikk generelt, og kanskje spesielt i statistikk. Sammenlignet med tradisjonelle innføringer i statistikk, inneholder de aktuelle lærebøkene mye stoff som virker svært teknisk. Der mange statistikere i praksis vil bruke ferdig programvare f. eks. på personlig datamaskin, oppfordres skoleelevene til å taste inn tilsvarende programmer på kalkulator. Dette fører noen ganger til tilsynelatende avbrudd i den ordinære presentasjonen av statistiske metoder, gjerne med ulike sett detaljerte anvisninger for forskjellige kalkulatormerker. I ekstreme tilfeller kunne man nesten tro forfatterne prøvde seg på markedsføring av bestemte kalkulatorer! Prinsipielt bør en likevel se positivt på denne utviklingen i matematikkundervisningen, selv om lærebøkene hele tiden må gjøre det klart hvilket nivå fremstillingen befinner seg på. F. eks. kan simuleringsteknikker by på interessante illustrasjoner.

Det er naturlig å peke ut fire grunnleggende områder der forfatterne generelt har store problemer:

Innføring av sannsynligheter

Ifølge læreplanen skal elevene «kjenne den mengdeteoretiske formaliseringen av sannsynlighets-begrepet og kunne bruke de grunnleggende regnereglene for sannsynligheter». Den første delen av denne formuleringen har muligens gjort litt for sterkt inntrykk på lærebokforfatterne. En finner forskjellige versjoner av ``Kolomogorovs aksiomer'', som i stor grad fortoner seg som en ekstra teoretisk overbygning over sannsynlighetsregningen. Sammenhengen med de vanlige regnereglene som skal benyttes senere, blir ikke klarlagt. De aller fleste regneksemplene dreier seg om endelige utfallsrom, der det ville være naturlig å innføre sannsynligheter for sammensatte begivenheter som en sum av sannsynlighetene for utfallene som inngår. Denne enkle tankegangen blir imidlertid underslå tt i flere tilfeller, enda den tydeligvis skal utnyttes i konkrete oppgaver. I noen bøker kan det virke som om formalismen hovedsakelig blir et hinder som elevene må trenge gjennom før de kan læ re de aktuelle regnereglene.

I et ekstremt eksempel gir en lærebok en formalisering av sannsynlighetsbegrepet som avslutning

på kapittelet om sannsynligheter, der «sannsynlighetsfunksjonen» P defineres som «en funksjon fra S» (utfallsrommet) «inn i intervallet [0,1]». På forhånd har en innført det «generelle funksjons-begrepet»' med tilhørende «definisjonsmengde» og «verdimengde». Hele veien er det imidlertid også snakk om verdier P(A) som skal være definert for sammensatte begivenheter A, selv om det plutselig faller helt utenfor den formelle beskrivelsen. Det ligger naturligvis vanskelige begrep som sigma-algebrær bak disse tankene, som hører hjemme på et helt annet nivå enn skolen. Det er tydelig at kapitlene om sannsynlighetsregning må gjennomarbeides mye grundigere. Av og til virker det som om forfatterne er ikke klar over at enkle regneregler kan utledes direkte av ``aksiomene" som er stilt opp. F. eks. mangler noen bøker forklaring av regelen om «gunstige» på «mulige» utfall. Tanke-gangen må være at elevene skal kjenne regelen fra andre klasse, og at det derfor ikke er nødvendig å forklare den logiske sammenhengen med den generelle teorien som nå blir presentert.

En annen pussighet gjelder begrepene «umulige» og «sikre begivenheter». Et par lærebøker angir som definisjon at det dreier seg om begivenheter A med P(A)=0 og P(A)=1. Det henger ikke helt sammen når det litt senere blir presisert at P(X=x)=0 for hver bestemt verdi x dersom X er en stokastisk variabel med kontinuerlig fordeling.

Innføring av normalfordeling

Flere av bøkene bruker mye plass på å forklare hva en normalfordeling er. Vanlig utgangspunkt er empiriske fordelinger som beskriver observasjoner foretatt med stor nøyaktighet, der det blir innført finere og finere klasseinndeling. Resonnementet minner sterkt om det man finner i mange innføringsbøker på høyskolenivå. Det er ikke lett å forklare overgangen til kontinuerlige sannsynlighetsfordelinger på denne måten, men noen av lærebøkene har kommet spesielt uheldig ut. I stor grad fester de seg ved tekniske detaljer i selve beskrivelsen av den empiriske fordelingen som bare skal tjene som motivering. Deretter blir mye av tankegangen fra den deskriptive behandlingen av datasettet hengende ved selve normalfordelingen, slik at det knapt er forståelig at den har noe med sannsynlighetsregning å gjøre. Dette fører til underlige utsagn om at datasett må være «store» for at man skal kunne forutsette at variablene er normalfordelt.

Noe så elementært som utregning av sannsynligheter i en normalfordeling blir også forkludret i enkelte bøker. Akkurat her kan man virkelig lure på hvordan elevene er ment å få klare forestillinger om de aktuelle regneteknikkene. På høyskolenivå har vi vært vant til å oversette alle aktuelle begivenheter til utsagn om standardnormalfordelte variable, men denne tankegangen kommer sjelden til syne i skolebøkene. Dessuten ser det ut til å være til stede en motvilje mot å oppgi tetthetsfunksjonen for normalfordelingen eksplisitt, enda elevene bør være i stand til å behandle matematiske uttrykk av denne typen. Dette kan igjen lett føre til forestillinger om at alle entoppete, symmetriske fordelinger er normalfordelinger, selv om noen lærebøker gir advarsler her.

Grunnleggende problemstillinger i statistisk metodelære

Et vesentlig ankepunkt mot fremstillingen i de fleste bøkene er at den ofte ser ut til å underslå selve inferenstankegangen. Det gjøres rett og slett ikke klart hva som er ukjente parametre knyttet til selve spørsmålsstillingen, og hva som oppfattes som realisasjoner av stokastiske variable. En kan innvende at dette er vanskelige idéer, men nå r en f. eks. forklarer hypotesetesting, kommer en egentlig ikke utenom. For en statistiker virker det ofte som om forfatterne konsentrerer seg mer om det rent tekniske ved metodene enn prinsippene som ligger bak. På dette feltet kan det trolig ha en viss betydning at forfatterne ikke selv er statistikere, men har mer generell matematisk bakgrunn.

Lærebøkene legger mye vekt på utsagn som skal formuleres om «populasjoner» på grunnlag av «stikkprøver». Selve uttrykket «ukjent parameter» brukes så å si aldri. Her ser det ut som om forfatterne har hentet inspirasjon fra andre læreverk som omfatter mer generelle situasjoner

enn det som inngår i skolen. Det gis mange eksempler der det foretas utvalg av et visst antall elementer som skal skal belyse forholdene i en svært begrenset populasjon. Problemet er imidlertid at utvalg uten tilbakelegning ikke er inkludert i læreplanen i statistikk. Modellene som skal behandles, dreier seg derimot om serier med uavhengige observasjoner, og de vil ofte tilsvare problemstillinger der en ønsker å uttale seg om en bestemt underliggende prosess. Likevel har de motiverende forklaringene i flere bøker en sterk slagside mot utvalg fra endelige populasjoner.

I en lærebok ser begrepet «populasjon» i statistisk forstand ut til å være helt misforstått. Definisjonen er faktisk at «Samlingen av alle verdiene i et datamateriale kaller vi den totale populasjonen»! I en oppgave er det angitt ti målte kjøretider for en bestemt veistrekning, og elevene oppfordres til å finne μ og σ for alle ti verdiene, og tilsvarende $\overline{X}$ og s for de tre siste. En slik illustrasjon skal iallfall ikke bidra til noen bedre forståelse av de grunnleggende idéene.

Konfidensintervaller

Dette området har falt vanskeligst under behandlingen av de statistiske metodene. Det er også her man finner de mest graverende feilene i noen av bøkene. Det må likevel understrekes at andre bøker gir korrekte fremstillinger, selv om også de kan være vanskelige å følge for elevene.

I en lærebok er begrepet misforstått i den grunnleggende forklaringen. Forfatterne definerer et konfidensintervall som et intervall en bestemt stokastisk variabel må falle innenfor med en gitt sannsynlighet. Som statistikere kan vi kanskje anse dette som et slags prediksjonsintervall, men i en elementær fremstilling blir distinksjonen mellom ukjent parameter og observerbar stokastisk variabel straks rotet til. Enda verre blir det når samme lærebok litt senere går gradvis over til den vanlige tankegangen med intervaller for parametre som forventningen mu i normalfordeling eller suksesssannsynligheten p i binomisk fordeling. En student som blander sammen begrepene på denne måten i eksamener i elementærkurs i statistikk ved høyskoler, vil få vesentlig fratrekk.

Denne forvirringen gir seg utslag i flere eksempler og oppgaver. I andre eksempler blir det først konstruert et konfidensintervall etter en korrekt fremgangsmåte, men så ødelegges inntrykket ved feilaktig fortolkning av det endelige tallsvaret. Her dukker det opp utsagn om sannsynligheten for at den ekte parameterverdien ligger mellom to bestemte verdier. Et tilsvarende, kanskje mer alvorlig problem, oppstår ved fortolkningen av konklusjonene fra hypotesetester. Signifikansnivå er og verdier av teststyrke blir delvis betraktet som sannsynligheter for at bestemte hypoteser om parametrene skal væ re korrekte. En statistiker skulle av og til tro at forfatterne var bayesianere, men bare ved fortolkningen av resultatene!

Noe som volder mye besvær, er å begrunne hvorfor man oppnår konfidensintervaller med gitt konfidenskoeffisient ved bestemte metoder. Spesielt når man skal ta pedagogiske hensyn, er dette en stor problem. Bare én lærebok prøver seg på en fullstendig utledning, der man omformer den aktuelle begivenheten så man får et utsagn som gir det ønskede intervallet. Andre bøker forsøker på snarveier ved å trekke inn variasjon i en normalfordeling med et visst antall standardavvik i forhold til forventningen. Overgangen fra intervaller som f. eks. variabelen $overline X$ må ligge i, til et nytt intervall som skal inneholde parameteren mu, skjer svært raskt og forklares vanligvis bare med ord. En må være varsom med å kritisere lærebøkene på dette punktet, selv om fremstillingen i noen bøker utvilsomt er vanskelig å følge.

Totalvurdering

Utvalget har gitt relativt fyldige kommentarer til alle fire læreverkene.Det er likevel vesentlige forskjeller bøkene imellom når en tar i betraktning hvor alvorlige feil som forekommer. For noen bøkers vedkommende er det mest snakk om mindre justeringer og forslag som kan gjøre fremstillingen ryddigere. I andre tilfeller finnes det grove feil som en neppe skulle trodd kunne forekomme i godkjente læ reverk. At bøkene ellers er nokså ulike i fremstillingsform, kan kanskje ses på som en fordel når en skole skal foreta et valg. Markedskreftene ser ut til å følge sine egne regler nå r det fastsettes hvilke læreverk som skal bli populære. Popularitet og kvalitet i fremstillingen av statistikk ser ikke ut til å være sterkt korrelerte variable, selv om bildet kan endre seg på lengre sikt. I denne sammenhengen spiller selvsagt de generelle matematikk-kapitlene en større rolle. Mange forfattere står bak læreverkene, og en vurdering av avsnitt f. eks. om integrasjon kan falle ut helt forskjellig fra vurderingen i statistikk.

I det videre arbeidet vil det være mer konstruktivt å hjelpe forfatterne til å forbedre bøkene enn gå ut med kritikk. Likevel må det være berettiget å undre seg over organer høyere i systemet, og spørre hvorfor de ikke reagerte på elementære feil som forelå i manuskriptene. Kritikken må ramme forlagene, som ikke alltid har brukt konsulenter med nødvendig fagkunnskap. Alle bøkene inneholder en passus av denne typen: «Læreboka er godkjent av Nasjonalt Læremiddelsenter» ... «til bruk i videregående skole». va som egentlig ligger bak denne godkjennelsen, er noe uklart. Det er mulig at det bare dreier seg om en rent formell kontroll, f. eks. av språket i bøkene og av likestillingsaspekter. Hvis det overhodet dreier seg om en **faglig godkjennelse**, må senteret sies å utøve meget dårlig skjønn.

Aalen [6] har på pekt at eksempler og oppgaver i lærebøkene ofte virker kjedelige. Dette er et vanskeligere problem å løse på kort sikt. Forfatterne blir her stilt overfor en stor utfordring, som er nokså ulik problemene en møter i andre deler av matematikkfaget. På dette området vil det utvilsomt være en fordel om fagstatistikere i større grad blir trukket inn i arbeidet.

Reaksjoner på utvalgets kommentarer

Utvalget har iallfall i noen tilfeller fått en reaksjon som viser at arbeidet har vært satt pris på. I et tilfelle fikk to utvalgsmedlemmer besøk av lærebokforfatter og forlagskonsulent for å diskutere noen av problemene. En del av våre kommentarer dreier seg om forhold som ikke er lette å rette på umiddelbart. Det tar naturligvis tid å utarbeide nye utgaver av bøkene.For et læreverk er det senere produsert et tilleggshefte som gir en revidert behandling av spesielle avsnitt i statistikken. Heftet blir delt ut i bokhandler når elevene kjøper den aktuelle boken. Viktige ankepunkter er tatt hensyn til, selv om heftet av lett forstå elige grunner følger mye av opplegget fra den opprinnelige fremstillingen. Også i tilleggsheftet ser det imidlertid ut til å forekomme iallfall ett utsagn der det feilaktig oppgis en sannsynlighet for at en alternativ hypotese skal være korrekt, gitt en bestemt observasjon.

Hvordan har undervisningen klart seg?

Statistikk har hittil vært et forholdsvis ukjent område for mange matematikklærere i den videregående skolen. Undervisningen har sikkert budt på mange problemer, noe som bl.a. har gitt seg utslag i stor interesse for korte etterutdanningskurs i statistikk. En del lærere har trolig vært klar over manglene ved fremstillingen i noen lærebøker, men de har tilpasset seg situasjonen ved å falle tilbake på svært standardiserte oppgavetyper. At noe slikt skulle være nødvendig, er naturligvis høyst beklagelig, sett fra vår side. Vi har også hørt om lærere som faktisk har kopiert opp kapitler fra andre lærebøker for å avhjelpe manglene ved det verket som allerede var valgt for matematikkfaget.

Det har likevel vært lite åpen diskusjon i lærerkretser omkring problemene. I stor grad skyldes det sikkert at svært få lektorer i skolen har noen bredere bakgrunn i statistikk. Det går ubekreftede rykter om innlegg i saken som har vært sendt inn til Lærerforbundets organ Skolefokus. En rask undersøkelse av alle numrene av dette tidsskriftet som kom ut i løpet av de siste 18 månedene, viste ingen omtale av statistikkfaget på det aktuelle nivået. Et debattinnlegg [7] med den velkling-ende tittelen «Statistisk kortslutning» var viet problemer med manglende justering av statistiske resultater for aktuelle utenforliggende variable. Det dreide seg imidlertid om sammenhenger mellom syn på lærerlønninger og politiske sympatier, der det ble understreket en burde justere for lønn, ut-danning og bosted blant de forespurte! Det spørs om et generelt organ av denne typen er det riktige sted å ta opp en faglig diskusjon om undervisningen, men problemet er at det ikke finnes så mange alternativer. Magasinet Tangenten, som har undertittelen «Tidsskrift for matematikkundervisningen», hadde faktisk for kort tid siden et «temahefte om statistikk og sannsynlighetsregning». Det var naturligvis et utmerket tiltak, men med ett unntak [8] hadde ikke emnene som ble behandlet noen direkte tilknytning til undervisningen i videregående skole.

Hva kan statistikerne gjøre videre?

I tiden fremover vil det antagelig komme større revisjoner av de aktuelle læreverkene. Helst burde statistikerne engasjere seg mye sterkere i prosessen med utarbeidelse av lærebøker. Som understreket av Aalen [6] er det et vesentlig problem at de aktuelle avsnittene er skrevet av matematikere og ikke statistikere. Selv om vårt fag i denne sammenhengen blir ansett som en del av matematikken, kreves det en fremstillingsform som er temmelig ulik den som benyttes ellers. Det beste ville være om folk som både hadde statistisk bakgrunn og erfaring fra skoleundervisning kunne delta aktivt.

På lengre sikt kan det være aktuelt å ta opp en debatt om læreplanene. Mange vil mene at de foreliggende planene legger opp til et heseblesende løp, som egentlig ikke gavner statistikken som fag. Det kan sikkert være uenighet om hva som ville være et optimalt opplegg, og spørsmålene er mange:
Bør antallet temær reduseres noe, med mulighet for større fordypning?
Burde f. eks. et av emnene konfidensintervaller eller hypotesetesting utgå til fordel for en fyldigere behandling av det andre? Bør man fylle ut hull i den logiske oppbygningen? F. eks. behandles ikke reglene om forventning og varians til summer av uavhengige stokastiske variable generelt, men likevel utnyttes setningene $E(\overline{X}) = \mu$ og $SD(\overline{X}) = \sigma/\sqrt{n}$. Bør undervisningen i statistikk i det hele tatt organiseres som en del av matematikkfaget? Burde det heller legges vekt på andre sider ved vårt fag, f. eks. ved å knytte behandlingen i sterkere grad til informasjonsteknologi?

Akkurat nå er det trolig best å prøve å konsolidere statistikkens stilling i skolen. Andre fagfelt har kommet inn før i skolepensum, for så i stor grad å forsvinne ved senere reformer. En del matematikere beklager f. eks. at vektorregning er kommet dårlig ut ved siste revisjon. Vi må unngå at statistikk skal lide en lignende skjebne, spesielt hvis det fremtrer som en kjedelig og livsfjern disiplin. Samtidig må man huske at mange elever i videregå ende skole vil velge bort matematikk på det aktuelle nivå et. En porsjon statistikk inkludert i det videregående kurset i matematikk i tredje skoleår er derfor ingen garanti for at f. eks. begynnerstudenter ved universitet har noen bakgrunn i faget. Igjen er det viktig å gjøre statistikkundervisningen tiltrekkende ved realistiske eksempler og oppgaver.

Referanser

[1] Læreplan for videregående opplæring. Matematikk. Studieretningsfag i studieretning for allmenne og økonomisk/administrative fag. Kirke-, utdannings- og forskningsdepartementet, Oslo 1994.

[2] Erstad, G., Bjørnsgå rd, I., Heir, O.: Matematikk 3MX. Aschehoug, Oslo 1996.

[3] Oldervoll, T., Orskaug, O., Våje, A.: Sinus 3MX Grunnbok. Cappelen, Oslo 1996.

[4] Sandvold, K. E., Øgrim, S., Gravem, B., Jasper, P., Nordseth, T.: Matematikk 3MX Grunnbok. Gyldendal, Oslo1996.

[5] Voje Johansen, N., Ommundsen, J.ts B., Onstad, T., Solvang, R.: Aktiv matematikk. Matematikk for videregå ende skole. Videregående kurs II. 3 MX.} NKS-Forlaget, Oslo 1996.

[6] Aalen, O.: Statistikk - en vits? Tilfeldig Gang årgang 14, nr 2. 1997.

[7] Brotnov, J.ts H.: Statistisk kortslutning. Skolefokus, 1997, (10) 39.

[8] Topphol, A.ts K.: Monte Carlo-simulering. Terningar, ikkje berre til Ludo.Tangenten, (1997) (2), side 8-21.

Statistikkoppgaver i videregående skole - gode idéer ønskes

NSF's undervisningsutvalg(Henrik Dahl, Ivar Heuch og Jostein Lillestøl) har som kjent vurdert kapitlene om sannsynlighetsregning og statistikk i den videregående skoles matematikkbøker. Vi fant et klart behov for mer relevante og spennende eksempler og oppgaver, slik at elevene kunne få en mer positiv holdning til faget vårt. Spørsmålet er om vi statistikere kunne bidra positivt til forbedringer ved å skaffe fram og tilrettelegge materiale, gjerne av supplerende natur, som case, forslag til prosjektoppgaver e.l.

Vi ønsker å mobilisere NSF medlemmer, og ber foreløpig om gode idéer, som kan sendes til:

Jostein Lillestøl
e-mail: jostein.lillestøl@nhh.no
Adr: NNH
Helleveien 30
5035 Bergen - Sandviken

Betinget sannsynlighet gitt en nullhendelse?

Gunnar Taraldsen

17. november 1997

Sammendrag

Vi drøfter definisjonen og beregningen av betingede sannsynligheter motivert av et innledende Borel-paradoks. Spesielt utledes en formel (*) som generelt gir den betingede sannsynligheten gitt verdien til en variabel som har kontinuerlig fordeling. Vi mener denne formelen burde være mer kjent enn det den er. Som et eksempel beregnes den betingede forventningsverdien til variabelen $\phi(X_1)$ gitt observatoren X_{maks} tilsvarende et tilfeldig utvalg $X_1, \ldots, X_n$ fra en uniform fordeling. Her er ϕ en generell funksjon. Vi gir i tillegg to andre mulige beregningsmetoder. Den ene metoden er gitt ved å skifte variable slik at den elementære metoden gitt ved en simultantetthet kan benyttes. Den andre metoden gir den betingede sannsynligheten som en grenseverdi tilsvarende hendelser som krymper ned til nullhendelsen. Formelen (*) gir et bevis for gyldigheten av denne metoden. Leserne oppfordres til å presentere alternative beregningsmåter. Spesielt er beregningsmetoder for betingede forventningsverdier gitt en tilstrekkelig observator av stor praktisk interesse.

1 Et Borel-paradoks

La (X, Y) være koordinatene til et tilfeldig valgt punkt i enhetskvadratet $(0,1)^2$. Da gjelder

$$P(X < 1/2, Y < 1/2 \,|\, X - Y = 0) = 1/2 \quad \text{og} \quad P(X < 1/2, Y < 1/2 \,|\, X/Y = 1) = 1/4.$$

Dette ser paradoksalt ut fordi hendelsen $(X - Y = 0)$ er lik hendelsen $(X/Y = 1)$. Muligheten for et slikt Borel-paradokset skyldes at hendelsen $(X = Y)$ det betinges med hensyn på er en nullhendelse, dvs $P(X = Y) = 0$. Paradokset krever en nærmere forklaring. I første omgang kan vi beregne de to betingede sannsynlighetene hver for seg ved passende variabelskifte. Denne metoden begrunner vi i siste hovedavsnitt i dette notatet.

La f være tettheten til fordelingen til det tilfeldige punktet. Da er f lik 1 i enhetskvadratet og lik 0 utenfor. Innfør en ny variabel $U = X - Y$. Absoluttverdien til Jacobi-determinanten ved skifte fra koordinatene (x, y) til koordinatene (x, u) er lik 1. Dette gir oss simultantettheten $f_{X,U}$ til X, U og

$$P(X < 1/2, Y < 1/2 \,|\, X - Y = 0) = \int_{x=0}^{1/2} f_{X|U}(x \,|\, 0)dx = \int_{x=0}^{1/2} \frac{f_{X,U}(x,0)}{f_U(0)}dx = \int_{x=0}^{1/2} 1dx = 1/2.$$

Den andre betingede sannsynligheten beregnes enklest ved overgang til polarkoordinater. Arealelementet $dxdy$ erstattes da av $rdrd\theta$. Vi finner

$$P(X < 1/2, Y < 1/2 \,|\, X/Y = 1) = \int_{r=0}^{\sqrt{2}/2} f_{R|\Theta}(r \,|\, 45°)dr = \int_{x=0}^{\sqrt{2}/2} rdr = 1/2(\sqrt{2}/2)^2 = 1/4.$$

Alternativt kan beregningen gjøres ved å innføre variabelen $V = X/Y$, og ved skifte fra (x, y) koordinater til (x, v) koordinater. Beregningen ved polarkoordinater krever strengt tatt litt mer argumentasjon. Grunnen til at beregningen er korrekt er gitt ved at nivåkurvene $x/y =$ konstant er lik nivåkurvene $\theta =$ konstant. Disse nivåkurvene er ikke annet enn rette linjer gjennom origo. Nivåkurvene $x - y =$ konstant er rette linjer med stigningstall 1. Denne forskjellen er årsaken til det tilsynelatende paradokset. For å begrunne dette må vi være mer presise.

2 Definisjon og eksistens av betinget forventning

Når $P(B) = 0$, så kan ikke $P(A \mid B)$ defineres med utgangspunkt i $P(A \cap B) = P(A \mid B)P(B)$. Borel-paradokset vi startet med gir at med konvensjonelle definisjoner av betinget sannsynlighet så forblir $P(B \mid A)$ udefinert i tilfellet $P(A) = 0$. Vi vil allikevel gi mening til beregningene vi startet med. Løsningen er gitt ved å innføre den betingede forventningen før den betingede fordelingen. Dette krever litt mer teori som vi skal se på.

Når A er en hendelse, så er indikatorvariabelen $[A]$ definert ved

$$[A](\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c \end{cases}$$

Når X er en tilfeldig variabel, så er produktet $X\,[A]$ også en tilfeldig variabel og vi kan definere notasjonen $\mathrm{E}(X; A)$ ved $\mathrm{E}(X; A) = \mathrm{E}(X\,[A])$. Her må vi anta at forventningsverdien $\mathrm{E}\,X$ er endelig. Leseren kan selv overbevise seg om at funksjonen Q definert ved $Q(A) = \mathrm{E}(X; A)$ er en ikke-normalisert fordeling med fortegn, dvs Q er et fortegnsmål.

La P være en fordeling. Dersom Q er et fortegnsmål slik at

$$P(A) = 0 \text{ gir } Q(A) = 0,$$

så sies Q å være absolutt kontinuerlig med hensyn på P. Radon-Nikodym teoremet (Doob, 1994) sier at i så fall finnes det en tilfeldig variabel X slik at

$$Q(B) = \mathrm{E}(X; B).$$

Her spiller variabelen X rollen som en tetthet for Q i forhold til P. La X være gitt. Det er da lett å vise at forventningsverdien $\mathrm{E}(X; B)$ definerer et fortegnsmål som er absolutt kontinuerlig med hensyn på P. Radon-Nikodym teoremet sier at *alle* absolutt kontinuerlige fortegnsmål fremkommer på denne måten. Vi bemerker at det finnes flere variable enn X som gjør nytten. Enhver Y som oppfyller $P(X = Y) = 1$ er like god.

La X være en tilfeldig variabel og la T være en observator. Et typisk eksempel er gitt ved at T er en vektor av tilfeldige variable, dvs $T = (Y_1, Y_2, \ldots, Y_n)$, men det finnes mange andre muligheter. En annen viktig mulighet er gitt ved $T(\omega) = \omega$ hvor utfallsrommet Ω_T til T er lik utfallsrommet Ω, men hvor Ω_T er utstyrt med en redusert samling av hendelser (en del-σ-algebra). Vi vil definere den betingede forventningen

$$\mathrm{E}(X \mid T = t).$$

Her kan vi ha $P(T = t) = 0$, men vi kan til og med ha tilfeller hvor $(T = t)$ ikke er en hendelse! Først bemerker vi at indikatorvariabelen $[T \in A]$ er gitt ved

$$[T \in A]\,(\omega) = [\{\omega \mid T(\omega) \in A\}](\omega) = \begin{cases} 1 & T(\omega) \in A \\ 0 & T(\omega) \in A^c \end{cases}$$

Vi overlater til leseren å overbevise seg om at vi også har $[T \in A] = [A](T)$ hvor $[A](T)(\omega) = [A](T(\omega))$. Vi kan nå definere hva som menes med en betinget forventning i det generelle tilfellet.

En betinget forventning $\mathrm{E}(X \mid T)$ *er en tilfeldig variabel som er en funksjon av* T*, dvs* $\mathrm{E}(X \mid T) = \phi(T)$*, og som oppfyller*

$$\mathrm{E}(X; T \in A) = \mathrm{E}(\mathrm{E}(X \mid T); T \in A)$$

for alle hendelser A *i utfallsrommet til* T*. Videre defineres notasjonen* $\mathrm{E}(X \mid T = t) = \phi(t)$.

Denne definisjonen (Rao, 1973, s.96) gir samme resultat som den mer konvensjonelle definisjonen (Lehmann, 1986, s.44) (Doob, 1953, s.18). Konvensjonelt defineres først en betinget forventning

med hensyn på en redusert familie av hendelser i utfallsrommet Ω. Vi har valgt definisjonen over fordi den leder raskere til resultater. Det er kanskje også noen som foretrekker vårt valg fordi dette valget skyver σ-algebraene så langt inn i skyggen som mulig.

Radon-Nikodym teoremet gir et kort bevis av at den betingede forventningen finnes: Definer et fortegnsmål Q for utfallsrommet til T ved

$$Q(A) = \mathrm{E}(X; T \in A).$$

Det følger at Q er absolutt kontinuerlig med hensyn på fordelingen P_T til T fordi $P_T(A) = P(T \in A) = 0$ gir $Q(A) = \mathrm{E}(X; T \in A) = 0$. Dermed gir Radon-Nikodym teoremet at det finnes en tetthet ϕ slik at

$$Q(A) = \mathrm{E}_T(\phi; A) = \mathrm{E}(\phi(T); T \in A).$$

Den siste likheten er en konsekvens av det generelle variabelskifteteoremet

$$\mathrm{E}_T\, \psi = \mathrm{E}\, \psi(T)$$

brukt på tilfellet $\psi(t) = \phi(t)\ [A](t)$. Vi har bevist at

$$\mathrm{E}(X; T \in A) = Q(A) = \mathrm{E}(\phi(T); T \in A),$$

som gir

$$E(X \mid T) = \phi(T),\ E(X \mid T = t) = \phi(t).$$

Dermed har vi bevist at en betinget forventning $\mathrm{E}(X \mid T = t)$ finnes. En betinget forventning er ikke entydig fordi enhver $\tilde{\phi}$ som oppfyller $P_T(\tilde{\phi} = \phi) = 1$ også gir en betinget forventning. Fordelingen til $\mathrm{E}(X \mid T)$ er imidlertid entydig.

3 Beregning av betingede fordelinger

Et problem så langt er at det ikke uten videre er klart hvordan en betinget forventning kan beregnes. Situasjonen ligner på den en møter i teorien for differensialligninger. Én ting er å vise at en løsning finnes, en helt annen ting er å finne løsningen. Vi begynner med å gi en metode som synes å være lite kjent, dvs vi kjenner ingen referanse til denne metoden.

Utgangspunktet for metoden er den generelle sammenhengen

$$\mathrm{E}(X; T \le t) = \int_{-\infty}^{t} \mathrm{E}(X \mid T = s) P_T(ds),$$

som følger ved inspeksjon av ligningene i foregående hovedavsnitt med $A = (-\infty, t]$. Anta at T er en tilfeldig variabel med en tetthet f. Derivasjon av likheten

$$\mathrm{E}(X; T \le t) = \int_{-\infty}^{t} \mathrm{E}(X \mid T = s) f(s)\, ds$$

gir formelen

$$\mathrm{E}(X \mid T = t)\, f(t) = \frac{d}{dt} \mathrm{E}(X; T \le t),$$

som kan benyttes til beregning av $\mathrm{E}(X \mid T = t)$. I det overstående er X en generell tilfeldig variabel så vi kan erstatte X med indikatorvariabelen $[X \in A]$. Da finner vi formelen

$$(*) \qquad P(X \in A \mid T = t)\, f(t) = \frac{d}{dt} P(X \in A, T \le t)$$

Utledningen kan enkelt generaliseres til observatorer T som er flerdimensjonale og resultatet er likt, men derivasjonen erstattes av derivasjon mhp hver av komponentene til t. Et viktig poeng er at metoden ikke forutsetter at det finnes en simultantetthet for X, T. Et annet poeng er at utledningen viser at funksjonen vi skal derivere automatisk er deriverbar. Dette betyr desverre ikke at den deriverte eksisterer for alle t, men snarere for alle t i en mengde D som oppfyller $P(T \in D) = 1$. Spesielt eksisterer den deriverte for alle t dersom $\mathrm{E}(X \mid T = t)f(t)$ er kontinuerlig som funksjon av t. Formelen (*) kan omskrives til den intuitivt tiltalende formelen:

$$P(X \in A \mid T = t)\, f(t) = f(t \mid X \in A)\, P(X \in A).$$

Vi kan gå tilbake til det helt generelle tilfellet, dvs utfallsrommet Ω_T til T er ikke nødvendigvis lik mengden av reelle tall. Utgangspunktet for metoden gir også

$$E(X; T = t) = E(X \mid T = t)P(T = t).$$

Her må vi anta at $(T = t)$ er en hendelse. Med $X = [A]$ og $B = (T = t)$ finnes den velkjente

$$P(A \cap B) = P(A \mid B)P(B).$$

Vi ser på noen eksempler på bruk av beregningsformelen (*) for tilfellet hvor $P(T = t) = 0$. La $X_1, \ldots, X_n$ være uavhengige og uniformt fordelt på intervallet $(0, \theta)$. Vi vil beregne

$$\mathrm{E}(2X_1 \mid X_{\text{maks}})$$

som en illustrasjon av metoden over. Metoden en møter i de fleste lærebøker er ikke umiddelbart anvendelig fordi det ikke finnes noen simultantetthet for X_1, X_{maks}, eller X, X_{maks}. La $0 < t < \theta$. Derivasjon av

$$P(X_{\text{maks}} \leq t) = (t/\theta)^n$$

gir tettheten

$$f(t) = nt^{n-1}/\theta^n \qquad 0 < t < \theta$$

til X_{maks}. Integrasjon mhp simultantettheten til $X_1, \ldots, X_n$ gir

$$\mathrm{E}(2X_1; X_{\text{maks}} \leq t) = (t/\theta)^{n-1} \int_0^t 2s/\theta \; ds = t^{n+1}/\theta^n.$$

Derivasjon og divisjon med f gir $\mathrm{E}(2X_1 \mid X_{\text{maks}} = t) = \frac{n+1}{n} t$ og

$$\mathrm{E}(2X_1 \mid X_{\text{maks}}) = \frac{n+1}{n} X_{\text{maks}}.$$

La en statistisk modell være gitt ved en familie $\{P^\theta\}$, $\theta \in \Theta$, av fordelinger for et utfallsrom Ω. En observator T (for Ω) er en funksjon $T : \Omega \to \Omega_T$ slik at $(T \in A) = \{\omega \mid T(\omega) \in A\}$ er en hendelse (i Ω) for enhver hendelse A i utfallsrommet Ω_T til T. Dersom $\Omega_T = \mathbb{R}$ med den konvensjonelle hendelsesfamilien, så er T en tilfeldig variabel. Her er det viktig å merke seg at parameteren θ ikke inngår på noen måte i definisjonen av en observator. Fordelingen til T vil imidlertid avhenge av parameteren θ ved at

$$P_T^\theta(A) = P^\theta(T \in A).$$

Et eksempel på en statistisk modell er gitt ved å la utfallsrommet være $\Omega = \mathbb{R}^n$, ved å la parameterrommet være $\Theta = \mathbb{R}$, og ved å la fordelingen P^θ være gitt ved en tetthet som er lik $1/\theta^n$ i området $(0, \theta)^n$ og lik 0 ellers. Ved å definere $X_j(\omega_1, \ldots, \omega_n) = \omega_j$ følger det at vi er tilbake i samme situasjon som i vårt regneeksempel.

Når vi har gitt en statistisk modell så vil en betinget forventning generelt være på formen

$$\mathrm{E}^{\theta}(X \mid T)(\omega) = \psi(\theta, T(\omega)),$$

dvs en betinget forventning er en funksjon $\phi : \Theta \times \Omega \to \mathbb{R}$. Dermed er en betinget forventning strengt tatt ikke en tilfeldig variabel pga avhengigheten av parameteren. I vårt regneeksempel fant vi imidlertid at den betingede forventningen

$$\mathrm{E}^{\theta}(2X_1 \mid X_{\text{maks}}) = \frac{n+1}{n} X_{\text{maks}},$$

dvs i dette tilfellet er den betingede forventningen en tilfeldig variabel. Dette skyldes at observatoren X_{maks} er tilstrekkelig.

Vanligvis defineres observatoren T til å være tilstrekkelig (for parameteren θ i forhold til observatoren X) dersom den betingede fordelingen til X gitt T ikke avhenger av parameteren. Her kan det bemerkes at pga komplikasjoner i forbindelse med eksistensen av betingede fordelinger så synes denne definisjonen å være lite hensiktsmessig. *Vi definerer tilstrekkelighet ved å kreve at* $\mathrm{E}^{\theta}(\phi(X) \mid T)$ *er en tilfeldig variabel for enhver* ϕ *som er en tilfeldig variabel for utfallsrommet til* X. Vi krever altså at det finnes en funksjon $\psi : \Omega_T \to \mathbb{R}$ slik at

$$\mathrm{E}^{\theta}(\phi(X) \mid T) = \psi(T).$$

En grunn til dette valget er at det finnes betingede forventninger som ikke er gitt ved integrasjon med hensyn på en betinget fordeling (Doob, 1953, s.27) (Lehmann, 1986, s.48), dvs det er ikke alltid gitt at en betinget fordeling finnes. En annen grunn til dette valget er gitt ved at en estimator betinget med hensyn på en tilstrekkelig observator gir en estimator.

I vårt eksempel finnes imidlertid en betinget fordeling. Vi kan bevise at X_{maks} er tilstrekkelig for θ i forhold til X_1 ved å beregne fordelingsfunksjonen til X_1 gitt X_{maks}. I beregningen over erstattes $2X_1$ av indikatorvariabelen $[X_1 \leq x]$. For $0 < x < t$ finner vi

$$P(X_1 \leq x, T \leq t) = (t/\theta)^{n-1} \int_0^x 1/\theta \, ds = \frac{t^{n-1}x}{\theta^n}$$

$$P(X_1 \leq x \mid T = t) = \frac{(n-1)t^{n-2}x}{\theta^n} \cdot \frac{\theta^n}{nt^{n-1}} = \frac{n-1}{n}\frac{x}{t},$$

I tillegg har vi

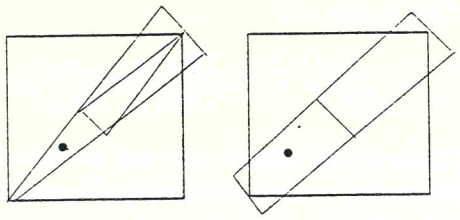
$$P(X_1 \leq t \mid T = t) = 1,$$

som er intuitivt meget rimelig, men dette kan også beregnes slik som over. Konklusjonen er at fordelingen til X_1 gitt $X_{\text{maks}} = t$ er en blanding av diskret og uniform fordi fordelingsfunksjonen har et sprang for $x = t$. Spesielt er sannsynligheten for $X_1 = t$ lik $1/n$. Dette resultatet kan også finnes ved andre argumenter (Lehmann, 1983, s.81). Fordelingsfunksjonen gir

$$\mathrm{E}^{\theta}(\phi(X_1) \mid X_{\text{maks}}) = \frac{1}{n}\phi(X_{\text{maks}}) + \frac{n-1}{n}\frac{1}{X_{\text{maks}}}\int_0^{X_{\text{maks}}} \phi(x)dx,$$

som beviser tilstrekkeligheten. Med $\phi(x_1) = 2x_1$ gjenvinnes det tidligere resultatet. Andre valg av ϕ gir forventningsrette estimatorer basert på X_{maks} for f eks variansen eller standardavviket til den uniforme fordelingen.

I det foregående er det verdt å merke seg at X_{maks} ikke er en funksjon av X_1 slik at faktoriseringsteoremet (Lehmann, 1986, s.55) ikke er direkte tilgjengelig. Løsningen ved faktoriseringsteoremet er gitt ved at dette kan brukes til å vise at X_{maks} er tilstrekkelig for $X_1, \ldots, X_n$ og derved for X_1.



Figur 1: Relativt areal er hhv 1/4 og 1/2.

4 To vanlige metoder

Intuitivt er det rimelig å gjette (Larsen and Marx, 1986, s.149)(Feller, 1971, s.157) på at

$$P(X \in A \,|\, T = t) = \lim_{h \to 0+} P(X \in A \,|\, t \leq T \leq t + h).$$

Et bevis av at denne formelen gjelder helt generelt finnes indirekte i (Doob, 1994, s.158-160), og en lignende formel bevises ved martingalteori i (Doob, 1953, s.343-348). Vi nøyer oss med å bevise at denne formelen gjelder under to forutsetninger. For det første antar vi at T har en tetthet f som er kontinuerlig og ulik 0 i punktet t. For det andre antar vi at $P(X \in A \,|\, T = t)$ er kontinuerlig i punktet t. Følgende beregning kan begrunnes ved L'Hopitals regel, kjerneregelen for derivasjon og ved beregningsformelen (*) fra forrige hovedavsnitt:

$$\begin{aligned}
\lim_{h \to 0+} P(X \in A \,|\, t \leq T \leq t + h) &= \lim_{h \to 0+} \frac{\frac{d}{dh} P(X \in A \cap (t \leq T \leq t + h))}{\frac{d}{dh} P(t \leq T \leq t + h)} \\
= \lim_{h \to 0+} \frac{\frac{d}{dh} P(X \in A \cap (T \leq t + h))}{\frac{d}{dh} P(T \leq t + h)} &= \lim_{h \to 0+} \frac{\frac{d}{dt} P(X \in A \cap (T \leq t + h))}{f(t + h)} \\
&= \frac{\frac{d}{dt} P(X \in A \cap (T \leq t))}{f(t)} = P(X \in A \,|\, T = t).
\end{aligned}$$

Forutsetningene i beregningen over er spesielt oppfylt for de betingede sannsynlighetene i Borel-paradokset vi innledet med. Leseren oppfordres herved til å overbevise seg om at formelen vi utledet nettopp leder til en geometrisk beregning av de nevnte betingede sannsynlighetene ved hjelp av areal mellom nivåkurver. Svarene kan finnes uten integrasjon fra figur 1.

En tredje metode er gitt ved variabelskifte (Lehmann, 1986, s.20) (Rao, 1973, s.99). Ideen er ganske enkelt å skifte koordinater slik at en finner en simultantetthet hvor variabelen det betinges med hensyn på er en av koordinatene. Deretter kan den elementære metoden benyttes. Denne metoden ble benyttet i det første hovedavsnittet i dette notatet. La $X = (X_1, \ldots, X_n)$ være en tilfeldig vektor hvor fordelingen er gitt av en tetthet f. Anta at vi ønsker å beregne $\mathrm{E}(\phi(X) \,|\, T = t)$ hvor $T = \psi(X) = (\psi_1(X), \ldots, \psi_m(X))$ med $m \leq n$. Anta videre at det finnes en stykkevis glatt en-en transformasjon $\chi : \mathbb{R}^n \to \mathbb{R}^n$ slik at $t = (\psi_1(\chi(t,s)), \ldots, \psi_m(\chi(t,s)))$. Da kan vi skifte fra $x_1, \ldots, x_n$ koordinater til $t_1, \ldots, t_m, s_{m+1}, \ldots, s_n$ koordinater ved at $x = \chi(t,s)$. La

$$\left| \frac{\partial(x_1, \ldots, x_n)}{\partial(t_1, \ldots, s_n)} \right|$$

være absoluttverdien til Jacobi-determinanten til variabelskiftet. Det følger at

$$\mathrm{E}(\phi(X) \,|\, T = t) = \int \phi(x) \frac{f(x)}{g(t)} \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(t_1, \ldots, s_n)} \right| ds$$

hvor g er tettheten til T. Den er gitt ved

$$g(t) = \int f(x) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(t_1, \ldots, s_n)} \right| ds$$

I begge formlene har vi forkortet notasjonen noe ved at koordinatene x skal oppfattes som en funksjon av koordinatene (t, s) ved $x = \chi(t, s)$. Betingede sannsynligheter finnes ved

$$P(X \in A \,|\, T = t) = \int [x \in A] \frac{f(x)}{g(t)} \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(t_1, \ldots, s_n)} \right| ds$$

Metoden gir gode eksempler på bruk av metodene som benyttes i kurs i flerdimensjonal analyse, men blir fort litt arbeidskrevende.

Den kritiske leser vil kunne innvende at $P(X \in A \,|\, T = t)$ kun er definert som en ekvivalensklasse av funksjoner ϕ av t hvor ψ er ekvivalent med ϕ dersom $P(\phi(T) = \psi(T)) = 1$. Slik sett gir det desverre ikke mening å snakke om funksjonsverdien $P(X \in A \,|\, T = t)$ for en t som oppfyller $P(T = t) = 0$, dvs konklusjonen er at $P(X \in A \,|\, T = t)$ forblir udefinert. Et svar til denne innvendingen er gitt ved at i tilfellene hvor ekvivalensklassen inneholder en kontinuerlig funksjon, så er denne entydig og formelen (*) gir funksjonsverdien til denne kontinuerlige funksjonen. Et bedre svar er gitt ved en drøfting av hvorvidt $Q(A) = P(X \in A \,|\, T = t)$ definerer en fordeling Q for gitt t, og om denne i så fall er unik. Spørsmålet om eksistens av en slik betinget fordeling er drøftet i (Doob, 1953), (Feller, 1971), (Lehmann, 1986), (Rao, 1973).

På tampen benytter vi anledningen til å utfordre leserne til å formidle alternative metoder for beregning av betingede forventningsverdier. Metodene bør kunne brukes på eksemplene vi har gitt. Dersom noen av leserne vet om en algoritme for simulering fra en betinget fordeling, så vil vi gjerne høre om dette. Vi tenker her selvsagt på tilfellet hvor en ikke kan ta utgangspunkt i en gitt betinget tetthet $f(x \,|\, t)$ fordi formen til $f(x \,|\, t)$ er ukjent. Problemet er å finne en algoritme som er effektiv i tilfeller hvor $P(T = t)$ er liten. En algoritme som virker i spesielle tilfeller er gitt i (Engen and Lillegård, 1997). Har noen av leserne andre forslag?

References

Doob, J. L. (1990 (1953)). *Stochastic Processes.* Wiley Classics Library. Wiley.

Doob, J. L. (1994). *Measure Theory.* Springer.

Engen, S. and Lillegård, M. (1997). Stochastic simulations conditioned on sufficient statistics. *Biometrika*, 84(1):235–240.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, vol. II.* Wiley series in probability and mathematical statistics. Wiley, second edition.

Larsen, R. and Marx, M. (1986). *An introduction to mathematical statistics and its applications.* Prentice-Hall.

Lehmann, E. L. (1983). *Theory of point estimation.* Wiley.

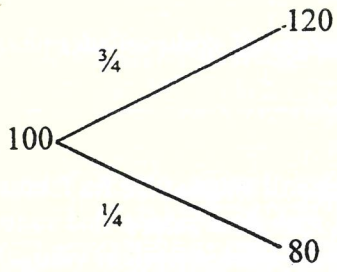
Lehmann, E. L. (1986). *Testing statistical hypotheses.* John Wiley & Sons Inc., second edition.

Rao, C. (1973). *Linear statistical inference and its applications.* Wiley.

Gunnar Taraldsen
Institutt for matematiske fag, seksjon Lade
Norges teknisk-naturvitenskapelige universitet
7034 Trondheim
gunnar@matstat.unit.no
http://www.matstat.unit.no/~gunnar

Statistikeren med forventet gevinst som innsats

Du er trolig kjent med at stokastiske modeller spiller en sentral rolle innen teoretisk og empirisk finans, og har muligens registrert at nobelprisen i økonomi i år gikk til to amerikanere Merton og Scholes, for arbeid som la grunnlaget for prising av opsjoner. Et marked er basert på forventninger og statistikere er gode til å regne med forventninger. Uten økonomisk kunnskap risikerer imidlertid vi «å gå fem på»! Betrakt følgende situasjon: En aksje står ved t = 0 i 100 og vil ved t = 1 gå opp til 120 med sannsynlighet ¾ eller ned til 80 med sannsynlighet ¼ som illustrert i figuren



La oss si du ved t = 0 tilbys en opsjon om å kjøpe aksjen ved t = 1 for 104. Hvor mye er du villig til å betale for denne opsjonen? Du tenker kanskje: Hvis det er oppgang kjøper jeg en aksje verd 120 for 104 og tjener 16. Hvis der er nedgang benytter jeg ikke opsjonen. Forventet gevinst er dermed $16 \cdot ¾ + 0 \cdot ¼ = 12$. Altså er jeg villig til å betale inntil 12. Opsjonsmakeren vil trolig bli henrykt og spørre etter telefonnummeret til flere statistikere. Opsjonen er faktisk ikke verd mer enn 8 som fremkommer ved $16 \cdot ½ + 0 \cdot ½ = 8$. La oss se det fra opsjonsmakerens side, som selger opsjonen for 8. Han kan da kjøpe 0,4 enheter av aksjer til 40 og må låne 32 for å finansiere dette kjøpet. Hvis aksjen går opp må han kjøpe ytterligere 0,6 aksje til markedspris $0,6 \cdot 120 = 72$ for å kunne levere til opsjonskjøper. Han mottar imidlertid 104 som betaling for 1 aksje og kan i tillegg tilbakebetale lånet. Går aksjen ned har opsjonsmakeren tapt $0,4 \cdot 20 = 8$ på denne, men har til gjengjeld fått 8 for opsjonen. Ved denne strategien går han altså ut i null uansett. En person som er villig til å betale mer enn 8 for opsjoner er derfor i stand til å generere inntekt uten risiko for opsjonsmakeren. Vi har sett bort fra lånerente, tidspreferanser og desslike, men dette rokker ikke ved det prinsipielle.

Vi ser at prisen ikke er avhengig av de gitte sannsynlighetene overhodet. Imidlertid fremkommer prisen som en forventning mht. de fiktive sannsynlighetene 0,5 og 0,5.

Satt på spissen kan vi si at store talls lov er irrelevant i et marked. Generelt vil markedsprisen på visse typer opsjoner (Europeiske) definert på prisprosesser under visse vilkår kunne finnes ved å bestemme disse fiktive sannsynlighetene og beregne forventninger med disse. Spiller så «virkelige» sannsynligheter ikke noen rolle overhodet? Jo, slike vil opplagt ha betydning for markedsprisen av aksjen til enhver tid, som er basert på forventninger om fremtid. Den teori for prising av opsjoner som anvendes i praksis er i hovedsak basert på modeller i kontinuerlig tid, der stokastiske integraler og martingal teori spiller en sentral rolle. Dette er et felt i rivende utvikling, og som årets nobelpristakere i økonomi åpnet opp.

Jostein Lillestøl

Yrkesveiledning for videregående skole

Kjære statistiker

Utvalget som skal organisere arbeidet med å lage yrkesveiledningen ønsker **din** hjelp. Våre langsiktige mål er følgende:

- Internett-sider med informasjon om utdannelsesmuligheter, yrkesmuligheter samt interessante anvendelser av vårt fag. Generell/overordnet informasjon skal ligge her, med pekere til de ulike undervisningssteder og bedrifter for mer detaljert informasjon.
- Brosjyrer til elevene og informasjonsbrev til studieveileder på skolen
- Stands på relevante arrangementer

Kortsiktig mål har vært at Russ'98 har hatt tilgang til sidene våre fra februar 1998. Første versjon vil være forholdsvis enkel mhp. layout og innhold, men disse sidene skal være «levende«, dvs. de skal kontinuerlig vedlikeholdes og forbedres. Det er her **dine** innspill er viktige! Vi ønsker at **du** kritisk går igjennom sidene og vurderer hvorvidt innholdet er korrekt, av riktig omfang, dekkende og representativt for ulike studiesteder og yrkesgrupper. **Dessuten trenger vi dine bidrag til «interessante anvendelsesområder»**- her er begrensningene kun at de må treffe målgruppen! Nå har vi en yppelig anledning til å profilere oss, for disse sidene vil være tilgjengelig for alle. Vi vet at det finnes mye spennende, men kan ikke hoste opp alt dette selv, så send inn tips eller helst ferdig tekst.

- Alle innspill sendes til Petter Prøsch : e-mail: ppr@math.uio.no. (Tlf. jobb: 63814987)
- Adressen til sidene er: http://www.math.uio./no/nst/yrkesveiledning/

Hilsen

Utvalget for yrkesorientering

RAPPORT FRA ISI 51. SESJON I ISTANBUL , 18. - 26. AUGUST 1997

fra Jan F. Bjørnstad, Statistisk sentralbyrå

Den 51. konferansen av The International Statistical Institute ble arrangert i Istanbul i perioden 18.-26. august. Det var ca. 1500 deltakere fra 93 nasjoner.

ISI-konferansene samler deltakere både fra statistiske sentralbyråer og universitetssektoren, noe som muliggjør tverrkontakter mellom offisiell sektor og akademiske miljøer. Den store spennvidden i deltakere medfører også at programmet for konferansen var meget variert med ialt 168 sesjoner. Proceedings fra konferansen finnes på biblioteket i Statistisk sentralbyrå.

Et utvalg av de forskjellige inviterte sesjonene vil gi et visst inntrykk av bredden i programmet: Statistics and the evaluation of legal evidence, Applications of longitudinal data analysis in sample surveys and official statistics, Mathematical statisticians in official statistics, Spatial-temporal modelling of meteorological fields with application tøclimate change and the greenhouse effect, Latent variable models, Challenges in meta-analysis, Teaching and training in statistics with sampling and sample surveys, Imputation and missing values, Generalized estimation in survey sampling and associated software, Algortihms for complex statistical problems, Bayesian approaches tømodel selection and prediction, Statistics: An inclusive interpretation.

I tillegg var det mange sesjoner med påmeldte bidrag, bl.a.: Environmental statistics, Statistics in medicine I, II, Sports statistics, Small area statistics, Sampling and sample surveys I-V, Survival analysis, Time series I-III, Image analysis and image reconstruction, Spatial Statistics.

Sammendrag fra noen få sesjoner

1. Mathematical Statisticians in Official Statistics

I denne sesjonen ble rollen til matematiske statistikere, vanligvis kalt metodestatistikere i Statistisk sentralbyrå, i offisielle sentralbyråer tatt opp.

Geoff Hole, Statistics Canada (SC), i samarbeid med Geoff Lee, Australian Bureau of Statistics (ABS) og Tim Jones, Office for National Statistics (ONS) i Storbrittania, presenterte et innlegg om rollen og organiseringen til metodestatistikere i de tre statistiske byråene. Metodestatistikerne i ABS og SC er organisert sentralt. I ABS, med ialt 3300 ansatte, er alle 100 metodestatistikere i Methodology Division, med 80 i Canberra og 20 fordelt på 6 distriktskontorer. I SC er det 240 metodestatistikere, av ialt 4500 ansatte, sentralt i Ottawa. ABS og SC rekrutterer mellom 10 og 20 metodestatistikere i året. I ONS har man en desentralisert organsisasjon og skiller ikke mellom fagstatistikere og metodestatistikere. Kun noen få personer har konsentrert seg om statistisk metodearbeid i ONS.

Metodeavdelingene i ABS og SC har tøhovedroller. Den ene er konsultasjonsvirksomhet for fagstatistikere,sosiologer og økonomer angående utvalgsplanlegging og estimering, editering og imputering, sesongjustering og tidsrekkeanalyse. Den andre rollen er å sørge for kvalitet og integritet i praktisk talt alle aspekter av statistikkproduksjonen gjennom utvikling og anvendelse av statistiske metoder til datainnsamling og analyse.

RistøLehtonen, Statistisk sentralbyrå i Finland beskrev hvordan metodestatistikerne (ialt 10) i byrået har et organisert samarbeid med universititet i Jyväskylä. Et professorat i statistikk er opprettet ved universitetet fra august 1997, 30% finansiert av sentralbyrået og 70% av universitetet. I samarbeid med universitetet har byrået et forskningsprosjekt i utvalgsmetoder med bl. a. følgende delprosjekter:

generell regresjonsestimering, justering for frafall, «small area» estimering, multilevel modeller, panelundersøkelser. Flere doktorgradsstudenter og post.doc. stipendiater er i dette programmet.

2. Teaching and Training in Statistics with Sampling and Sample Surveys

Sesjonen hadde tre innlegg hvorav tødreide seg om utvalgsteori som basis for introduksjonskurs i generell statistikk. Det siste innlegget var mer rettet mot bruk av data i undervisningen.

Richard L. Scheaffer fra University of Florida, USA holdt innlegget «Discovery of sampling concepts through activities». Han påpekte at selvom sampling er et fundamentalt begrep i anvendt statistikk så er det en tendens til at det blir forbigått i begynnerkurs i statistikk. Det ble gitt eksempler på hvordan begreper som målinger, variabilitet og skjevhet kan illustreres og forklares med egne studentaktiviteter. Det ble understreket hvor viktig det er at studentene forstår begrepene skjevhet, variasjon, tilfeldig utvalg, og «repeatability» før man kan «gjøre noe som helst». Idéene til Scheaffer kan uten tvil være til nytte i utarbeidingen av studieopplegget for den planlagte høyskoleutdanningen i anvendt statistikk i Kongsvinger.

Anne-Marie Dussaix fra Frankrike beskrev erfaringer fra undervisning av kurs i samplingmetoder for større komplekse utvalgsundersøkelser, for studenter med varierende bakgrunn. Hun diskuterte hvordan samplingkurs for brukere bør atskille seg fra kurs i en statistikkutdanning, og de pedagogiske utfordringene i slike kurs.

3. Imputation and Missing Values

Hovedtemær var frafallskjevhet, variansestimering for imputerte data og imputeringsmetoder.

Roderick J. Little fra University of Michigan,USA presenterte innlegget «Imputation of multivariate missing data on household assets» som dreide seg om Bayesiansk analyse for multivariate data hvor det vanligvis er en høy grad av frafall som medfører stor skjevhet i nettoutvalget. I denne undersøkelsen hadde man prøvd å redusere frafallet ved omformulering av svar om inntekt og formue til intervaller der hvor eksakte tall var upålitelige eller utilgjengelige. Flere imputeringsmetoder ble sammenlignet både empirisk og ved hjelp av simulering. Det viste seg at «intervallsvarene» ga en betydelig forbedring i informasjonen fra data.

Jun Shaøfra University of Wisconsin, USA holdt innlegget «Variance estimation for imputed survey data with non-negligible sampling fractions». Tema var variansestimering for Horvitz-Thompson type estimatorer under ignorerbar responsmekansime. Det ble hovedsakelig brukt Taylor linearisering, men også diverse resamplingmetoder kan brukes i anvendelse av metoden foreslått av Shao.

4. Generalized Estimation in Survey Sampling and Associated Software

Tre innlegg presenterte forskjellige generelle statistikksystemer for estimering/variansestimering. Det mest interessante innlegget var «Computing sampling errors from clustered, unequally weighted data using replications: WesVarPC» ble holdt av J. Michæl Brick i samarbeid med D. Morganstein Her ble programpakken WesVarPC, utviklet ved Westat, USA, presentert. Pakken dekker forskjellige estimeringsmetoder som etterstratifisering, raking og generaliserte regresjonsestimatorer. Hovedvekten var lagt på å vise hvordan resamplingteknikker som jackknife og bootstrap for variansestimering i kompliserte utvalgsplaner kan implementeres i programpakken, i situasjoner hvor Taylor linearisering ikke fungerer. Programpakken kan importere SAS-filer, er offentlig tilgjengelig og kan brukes i Windows 3.1 eller Windows 95. En ny versjon vil være tilgjengelig kun for Windows 95.

Rapport fra First International Conference on Mathematical Methods in Reliability (MMR 97)

Fra Bo Linquist,Institutt for matematiske fag, NTNU

Denne konferansen «The First International Conference on Mathematical Methods in Reliability» ble arrangert i Bucuresti i tiden 15.-19. september 1997. De fleste av de ca. 80 deltakerne var fra Øst-Europa, spesielt Romania (naturligvis) og det tidligere Sovjetunionen, men det var også spesielt mange deltakere fra Frankrike. Det teknologiske universitetet i Compiegne, Frankrike, var forøvrig medarrangør og har et formalisert samarbeid med det tekniske universitetet i Bucuresti. Såkalt «honorary chairman» for konferansen var statistikknestoren professor Marius Iosifescu, som bl.a. har gjesteforelest i Oslo og Trondheim. Undertegnede var den eneste skandinaviske deltakeren.

Som navnet tilsier, var dette ikke noen ren statistisk konferanse. Med «Mathemtical Methods» i tittelen ønsket man en pålitelighetskonferanse som også omfattet eksperter i operasjonsanalyse og matematisk modellering, samt representanter for brukerne.

Med et forholdsvis lite antall deltakere, ble dette en intim og hyggelig konferanse, noe ikke minst det meget gjestfrie rumenske vertskapet medvirket til. En middag på en typisk rumensk restaurant var således en opplevelse. Det samme var en ad hoc arrangert utflukt til et sted langt ute på landet der den berømte Lenin-statuen lå gjemt bort på et jorde bak et kloster.

Blant de inviterte foredragene kan nevnes «Statistical models of systems in reliability analysis» ved V.S. Korolyuk fra Ukraina, «Models building in accelerated experiments» ved M.S. Nikulin, Bordeaux, samt «Start-up demonstration testing» ved N. Balakrishnan (McMaster University, Canada). Balakrishnan var forøvrig den eneste deltaker fra «overseas».

Konferansen må betraktes som meget vellykket, og det ble bestemt at en ny konferanse skal finne sted i Frankrike, antageligvis i år 2000. Det tas da sikte på å få til en bredere deltakelse, bl.a. ved å henvende seg i sterkere grad til «survival analysis» miljøene.

SHORT COURSE IN BIOSTATISTICS

MEASUREMENT ERROR MODELING – LINEAR AND NON-LINEAR REGRESSION MODELS

AAS, NORWAY, JUNE 3 – 5, 1998

A course on **Measurement error modeling – linear and non-linear regression models** will be given at Aas, Norway, June 3 – 5, 1998.

John Buonnacorsi will teach the course. He is a Professor of Statistics, University of Massachusetts at Amherst. He has been working with models for measurement error for a long time; for references to some of his work, see for instance the recent book on this topic by Carroll, Ruppert and Stefansky (1995).

Course venue:
The venue of the course will be the Agricultural University of Norway at Aas, which is located 30 km south of Oslo. Aas is easily accessible from Oslo by train or by car.

COURSE DESCRIPTION:
This course will introduce measurement error models, investigate the consequences of measurement error, and then discuss approaches to correcting for measurement error based on additional data/information which allows for estimation of measurement error parameters. Problems treated include misclassification of categorical variables in analyzing proportions or contingency tables, and measurement error in the response and/or predictor variables in linear and nonlinear regression models. Both theory/methods and applications will be covered.

1. Measurement error models.
   What is measurement error? Misclassification of categorical variables.
   Measurement error models for quantitative variables
   (including additive and non-additive measurement error).

2. Types of additional data used to correct for measurement error.
   Replication, validation data (internal and external).

3. Misclassification in the analysis of categorical data.
   Estimation of a single proportion; analysis of contingency tables.

4. Measurement error in regression.

   A. Linear regression.
      Simple and multiple linear regression.

   B. Nonlinear regression models.
      General nonlinear models, generalized linear models,
      binary regression models.

The exercise work includes theoretical as well as practical exercises. Hands-on exercises with computing, using SAS, SAS/IML and SPLUS where possible, will also be included.

There will be a course exam, organized as a take-home exam, after the course. This may include some theoretical, as well as practical problems.

Course programme:

Lectures:	08.45 – 11.00
Exercise work:	11.15 – 12.15
Lunch:	12.15 – 13.00
Lectures:	13.00 – 15.00
Exercise work:	15.00 – 16.00
Concluding discussion:	16.00 – 16.30

Accomodation:
Accommodation in single rooms at student dormitories will be available at a price of NOK 350 including breakfast. The student dormitories are located approximately 500 meters from the course venue.

Course fee:
The course is sponsored by Biometric Society, Nordic Region. The course fee for members of Biometric Society will be NOK 2.500. For non-members the price is NOK 3.000. The course fee includes lunches during the course, hand-outs etc. Accomodation is not included.

Organizers:
The course is organized by Are Aastveit , Trygve Almøy, and Petter Laake

Registration and further information:
Final date for registration is **April 1, 1998**. The number of participants will be limited to about 30.

For registration and further information please contact:
Are Aastveit, Department of Mathematics, Agricultural University of Norway
P.O. Box 5035, 1432 Aas-NLH, NORWAY
Tel: +47 6494 8872 Fax: +47 64948879 E-mail: are.aastveit@imf.nlh.no

SHORT COURSE IN BIOSTATISTICS

Aas, Norway, June 3-5, 1998

Registration Form

Family name:

First name:

Affiliation

Mailing address:

Telephone:

Fax:

E-mail:

Member of Biometric Society: Yes ( ) No ( )

Accommodation:

Student dormitory: ( )

No accomodation: ( )

Date: Signature:

Kindly remit the amount to Union Bak of Norway, Oslo, 1654.25.21590, Course in Biostatistics 1998, P.O. Box 5035, 1432 Aas, NORWAY,
SWIFTCODE: UBNONOKK

Mail this registration form within April 1 to:

Are Aastveit
Department of Mathematics
Agricultural University of Norway
P.O. Box 5035
1432 Aas-NLH, NORWAY

17th Nordic Conference on Mathemathical Statistics

June 8 -12, 1998

Elsinore, Denmark

Final announcement and call for papers

Special Invited lectures

- Søren Asmussen, Department of Mathematical Statistics, Lund University, Sweden:
  Markovian distributions and point processes: Applied probability analysis and statistical inference

- Ib Skovgaard, Royal Veterinary and Agricultural University, Copenhagen:
  Likelihood asymptotics

Invited Lectures

- Elja Arjas, Section of Biometry, University of Helsinki:
  Bayasian inference

- Knut Conradsen, Institute of Mathematical Modelling, Technical University of Denmark:
  Image Analysis

- Ulla Holst, Department of Statistics, Lund University:
  Analysis of environmental data

- Per Mykland, Department of Statistics, University of Chicago:
  Likelihood and martingales

- Antti Penttinen, Department of Statistics, University of Jyväskylä:
  Spatial statistics in biologi

Contributed paper sessions

Nine special-topic sessions will be organized.

Contributed papers should be submitted directly to the organizers of the sessions, There is no special form for submission of papers. It is recommended that persons who wish to present a paper contact the session organizer first.

Deadline for submissions is May 1st.

1. *Statistical Methods for Diffusion Processes*: Bo Martin Bibby (bibby@dina.kvl.dk)

2. *Biopharmaceutical Statistics*: Eva Seim (Eva.Seim@nycomed.telemax.no)

3. *Incomplete Oberservations*: Tobias Rydén (tobias@sibeleius.maths.lth.se)

4. *Statistics in Genetics*: Søren Andersen (SA@ds-data.dk)

5. *Computational Statistics*: Eva B. Vedel Jensen (eva @mi.aau.dk)

6. *Inverse Problems*: Jens Ledet Jensen (jlj@mi.aau.dk)

7. *Industrial Statistics*: Jacques de Maré (demare@math.chalmers.se)

8. *Spatial Statistics and Image Analysis*: Martin Bøgsted Hansen (mbh@math.auc.dk)

9. *Longitudinal Data*: Jari Haukka (Jari. Haukka@ktl.fi)

Venue
The «LO-school», Elsinore. The School/conference centre is an independent institution with a strong connection to the Danish trade union movement. It is situated in the North-Western part of Elsinore in an old park with a view to Kattegat. The distance to Kronborg Castle and the old part of Elsinore is about 3 km.

Accommodation
Primarily at the school itself, and at the neighbouring «Højstrupgård». There are also camping facilities and a youth hostel nearby.

Social events
Wednesday afternoon there will be arranged a tour to Roskilde, (www.roskildekom.dk) including visit to the viking ship museum (www.natmus.dk/nmf/nb/8/english/art1.htm, www.natmus.dk/nmf/nb/8/english/art2.htm).

Organizer
Danish Society for Theoretical Statistics

Organizing Committee (nc-0@dsts.dk)

- Bendic Carstensen (chair) bxs@svs.dk
- Peter Dalgaard p.dalgaard@biostat.ku.dk
- Helle Holst hh@unidhp.uni-c.dk
- Kim Knudsen kimk@novo.dk

Programme Committee (nc-p@dsts.dk)

- Peter Dalgaard — p.dalgaard@biostat.ku.dk
- Dorte Kronborg (chair) — Kronborg@cbs.dk
- Steffen Lauritsen (representing SJS) — Steffen@math.auc.dk
- Mats Rudemo — rudemo@math.chalmers.se
- Michael Sørensen — michael@math.ku.dk

Web site

Current information updates are available at: www/dsts.dk/nordic98

Registration Form

17th Nordic Conference on Mathematical Statistics, June 8 - 12, 1998

First names:	
Last name:	
Mailing address: (if c/o organization, please state)	
Telephone: Fax: E-mail	
Arrival date: Departure data:	

Fees (in danish Kroner(DDK) and per person)		No.	Amount
Registration (including conference dinner) (normal/student)			
Before April 15, 1998	2500/2000		
After April 15, 1998	3000/2500		
Lunch tickets (Monday - Friday)	750		
Dinner tickets (Monday - Wednesday)	400		
Tour Wednesday June 10:			
Roskilde & Viking Museum	300		
Accommodation:			
LO-skolen (4 nights)	2700		
Højstrupgård (per night)	820		
Total			

NOTE: «Students» does not include PH.D.students

Payment All payments must be in Danish kroner (DKK), made payable to International Conference services A/S. Please remember to state your name, address and NCMS17 on all payments. (TICK)

	The amount is enclosed by banker's draft or cheque drawn on a Danish bank
	The amount has been transferred to A/C No 4001 4001 154656 Den Danske Bank, Holmens Kanal2, DK-1090 Copenhagen K, Denmark, SWIFT Address: DA BA DK KK (not Danish delegates)
	The amount has been transferred to Girobank 618 3522

Data: Signature:

Please return this form together with your payment to:
International Conference Services, Postbox 41, Strandvejen 171, DK-2900 Hellerup, Denmark

21. Nordiske statistikermøte - Utfordringer ved år 2000

Invitasjon
Norsk statistisk forening i samarbeid med Statistisk sentralbyrå inviterer med dette til det 21. Nordiske statistikermøtet i Lillehammer 5. - 8. august 1998.

Lillehammer
Lillehammer er en liten by med mange muligheter. Både byen og omegnen har flere attraksjoner, som teater- og konserthall på Maihaugen, Lillehammer kunstmuseum og Kulturhuset Banken.

De olympiske vinterlekene i 1994 var en stor begivenhet for Lillehammer. I Olympiaparken, nær Lillehammer sentrum, ligger Håkons hall med Norges olympiske museum, Lysgårdsbakkene hoppanlegg, freestylebakkene og Birkebeineren skistadion.

Lillehammer har flere fjellområder i umiddelbar nærhet av byen, og alle naturinteresserte kan oppleve naturen på nært hold. Det er hyppige togavganger bl.a. til Oslo, og kjøretiden til Oslo med bil er litt over to timer.

Quality Lillehammer Hotel
Quality Lillehammer Hotel er et konferansehotell med topp moderne AV-utstyr, grupperom og utstillingslokaler. Hotellet ligger i en privat park. Det har 250 værelser - 500 senger - og hyggelige salonger og solterasser.

Hotellet har også treningsstudio og svømmehall med badstuer og solarium. Hotellet ligger nær Olympiaparken og Maihaugen, og med kort avstand til byen.

Organisering
Statistikermøtet organiseres av Statistisk sentralbyrå med en egen organisasjonskomite i samarbeid med en nordisk komite som består av representanter fra de nordiske statistiske foreningene.

Tema og program
Det faglige programmet er organisert rundt temaet Utfordringer ved år 2000. Programmet består av tre plenumssesjoner.

Plenum A - torsdag kl 10.00-12.00
Kan historien forstås fra økonomiske tidsserier
Organisator: Professor Olav Bjerkholt,Universitetet i Oslo

Plenum B - fredag kl 09.00 - 11.30
Statistics and the role of statistics in the society.
Professor Steven Fienberg, Carnegie Mellom University, Pittsburg, innleder til en paneldebatt mellom sjefsstatistikerne i Norden.

Plenum C - fredag kl 16.30 - 17.30
Et mørkt kapittel i statistikkens historie?
Rådgiver Espen Søbye, Statistisk sentralbyrå

Språk
I plenumssesjonen fredag formiddag vil det bli benyttet engelsk, ellers vil det i hovedsak bli benyttet skandinaviske språk - dansk, norsk og svensk - ved møtet.

Parallelle sesjoner
Torsdag kl 13.30 - 16.30

1. Markedsføring	Lars Berglöf	Lars.berglof@scb.se
2. Regionalstatistikk	Paul Inge Severeide	pis@ssb.no
3. Bruk av administrative registre	Ib Thomsen	ith@ssb.no
4. Revisjon og dataeditering	Leopold Granquist	leopold.granquist@scb.se
5. Statistikk for offentlig sektor	Tuula Hausmann	tuula.hausmann@stat.fi

Fredag kl 13.00 -16.00

6. Datainnsamlingsmetoder	Gustav Haraldsen	gha@ssb.no
7. Nordiske levekårsundersøkelser	Ari Tyrkkö	ari.tyrkko@stat.fi
8. Arbeidsmarkedsstatistikk	Kirsetn Wismer	kwi.@dst.dk
9. Konjunkturstatistikk	Björn Rúnar Guðmundsson	
10. Nasjonalregnskap	Erling Joar Fløttum	ejf@ssb.no

Sosialt program

Onsdag 5. august
Bli kjent!
Deltakere og ledsagere inviteres til grillparty i hotellet park fra kl. 18.00 onsdag 5. august.

Fredag 7. august
Festmiddag
Deltakere og ledsagere inviteres til festmiddag fredag kveld.

Lørdag 8. august
Opplevelsestur til Sjoa/Heidal - Rafting, gårdsbesøk, barneaktiviteter
Heldagstur, Lillehammer - Sjoa - Lillehammer, varer ca. 7,5 - 8 timer, fra kl.09.00 til ca. kl. 17.00.

Deltagerne hentes med turbusser fra Lillehammer Hotel kl. 09.00. Turen går gjennom vakre Gudbrandsdalen med ankomst i Nedre Heidal. Deltagerne fordeles i grupper til aktiviteter de på forhånd har valgt, og aktivitetene settes i gang så snart fordelingen er klar. På bussene serveres kaffe og te til matpakken som er brakt med fra hotellet, og autoriserte Lillehammer-guider forteller om stedene vi passerer.

Etter arrangementet bringes deltagerene tilbake til hotellet eller til Oslo.
Vi håper at alle har anledning til å delta på turen.

Rafting

Rafting er vått, vilt og vakkert og bør prøves! Rafting er en populær aktivitet i Heidal, og deltagerne får gjennom raftingturen i Sjoa en unik mulighet til å oppleve den norske naturen på sitt beste. Dyktige guider tar hånd om gruppene etter ankomsten til Sjoa, deler ut utstyr, forteller om rafting og går gjennom opplegget og sikkerhetsrutinene. Utstyret består av våtdrakt (wet-suit), hjelm og redningsvest (safety jacket). Deltagerne deles i båtlag. Hver båt har plass for 7-8 personer og blir ledet av en guide. Guidene som følger turen er meget erfarne og kjenner Sjoa og legger stor vekt på sikkerheten. Etter turen skifter deltagerne til tørre klær - alt blir vått i løpet av turen - og samles rundt bålet i en lavvo.

Rafting I
Dette er en 11 km lang raftingtur med rolig start og frisk avslutning. Rafting I passer best for de som er komfortable med vannet som et element, og som tør å ta en utfordring! Turen går gjennom rolige partier, storslåtte juv og delvis gjennom spennende og utfordrende stryk.

Nedre aldersgrense for denne turen er 16 år.

Rafting II
En rolig raftingtur på ca. 4 km. Denne turen starter noe høyere oppe i elva hvor elven renner rolig. Turen avsluttes før de friskeste styrkene.

Denne turen anbefales for voksne som liker en roligere opplevelse, og for barn/ungdom 12 - 16 år.

Gårdsbesøk
For deltakere som ønsker å se på en norsk fjellgård, kan vi anbefale gårdsbesøk på en av Heidals vakre gårder. Deltakerne hentes og bringes til gården hvor de får en omvisning. Etter besøket samles deltagerne i basen rundt bålet.

Aktiviteter for barn
Barn er hjertelig velkomne til å delta i aktivitetene som er nevnt ovenfor. I tillegg kan vi tilby kano/kajakkpadling på en dam ved basen, "mini-baseball", fotballkamp eller "5-kamp".

Aktivitetene foregår ved basen og ledes av aktivitører.

Utstyr
Det deltagerne trenger å ta med seg/ha på seg er:

Rafting	Joggesko og ullsokker, stillongs og langermet ullundertrøye, superundertøy e.l. Alt dette blir vått under våtdrakten. I basen finnes det garderober hvor deltagerne kan skifte til tørre klær.
Andre aktiviteter	Fritidsklær som f.eks. joggedress eller annet fritidstøy som tåler en støyt.

Påmelding på fellesturen
Vennligst kryss av på påmeldingsskjemaet om du vil/ikke vil delta på heldagsturen til Sjoa.

Påmelding på de ønskede aktiviteter skjer i Lillehammer. Der gis det mer detaljert informasjon om hver enkelt aktivitet.

Lørdag 8. august
Møteavslutning - Møtet avsluttes med lunsch på Dale-Gudbrands gård

Ledsagerprogram

Torsdag 6. august

OL-sightseeing

Varighet ca. 3 timer

Gruppen hentes av buss med guide ved Quality Lillehammer Hotel. I løpet av sightseeingen besøkes Olympiaparken som består av Håkons hall, Lysgårsbakkene hoppanlegg, freestyleanlegg og Birkebeineren skistadion. I hoppanlegget går gruppen opp i hopptårnet i den største bakken og nyter panoramautsikten over byen.

Videre besøkes Håkons hall i Olympiaparken, Hafjell alpinanlegg i Øyer og Lillehammer olympiske bob- og akebane ved Hunderfossen. I løpet av turen får man også se blant annet kjempetrollet i Hunderfossen familiepark, Maihaugen museum og Kulturhuset Banken.

De Sandvigske samlinger på Maihaugen

Friluftsmuséet som ligger ca 5 minutters gangavstand fra Quality Lillehammer Hotel, viser hus og gårdstun fra Gudbrandsdalen og Lillehammer. Deltakerne får sin forhåndsbestilte billett ved registrering og kan besøke muséet når de selv ønsker. Det er mulighet for å delta i guided omvisning eller gjøre seg kjent med muséet på egen hånd.

Maihaugens nye attraksjon er en utstilling om Norges historie, Langsomt ble landet vårt eget, som viser viktige perioder i landets historie fra istid, gjennom vikingtid, middelalder og krigsårene 1940-1945. Utstillingen ender opp i vår egen tid.

Fredag 7. august

Hunderfossen familiepark
Heldagstur

Deltakerne hentes på hotellet med buss som kjører direkte til Hunderfossen familiepark, 18 km nord for Lillehammer. Familieparken er et eldorado både for små og store. I grotten under verdens største troll vises det scener fra norske folkeeventyr. Familieparkens store uteområde inneholder opplevelsessenter for iskrem, opplevelsessenter for energi, opplevelsessenteret for olje og gass, vokskabinett, bilbaner, bondegård, badeanlegg og mye mer. I Supervideografen vises en 20 minutters film fra Norge.

Aulestad og Segalstad gård
Varighet ca. 5 timer

Gruppen hentes på hotellet med buss som drar direkte til Aulestad i Follebu. Her blir det omvisning i Karoline og Bjørnstjerne Bjørnsons hjem fra 1875. Mange av gjenstandene på Aulestad forteller om forfatterens liv i sentrum av Norges litterære, kulturelle og politiske liv. Bjørnson fikk Nobels litteraturpris i 1903.

Etter omvisningen blir det servert en enkel lunsj i tradisjonsrike omgivelser på gården Segalstad i jordbruksbygda Gausdal. Vertinnen har satt brøddeigen før vi kommer, og brødet stekes i gammeldags ovn mens hun viser rundt på gården. En autorisert guide følger gruppen på turen og forteller om tradisjoner og steder som passeres på veien.

Påmelding og praktisk informasjon

Påmelding skjer til TS Forum AS som bidrar i den praktiske gjennomføring av 21. Nordiske statistikermøte i Lillehammer. TS Forum AS vil ha en kontaktperson, Sirpa Kokslien, til stede på møtet.

Påmelding - sosial program

Opplevelsestur lørdag 8. august 1998
Av hensyn til planlegging av lørdagens tur til Sjoa, ber vi deltagere og ledsagere allerede ved påmeldingen om å krysse av for deltagelsen på turen. Hele turen inklusive bespisning dekkes av konferanseavgiften. Registrering av deltagelsen på de alternative aktivitetene på turen vil skje i sekretariatet under konferansen.

Ledsagerprogram
Ledsagere bes vennligst krysse av deltagelsen på ønsket ledsagerprogam.

Busstransport Oslo/Lillehammer
Det settes opp busser fra Oslo. Bussen til Lillehammer onsdag 5. august går fra Oslo lufthavn Fornebu kl. 13.30 og fra Oslo sentrum ved Rådhuset kl. 14.00.

Overnatting
Quality Lillehammer Hotel er konferansehotellet der de fleste deltakere og ledsagere vil bo under møtet. Kryss av om du ønsker hotellrom utover møteperioden.

Arrangøren er behjelpelig med å skaffe overnatting i leiligheter (spesielt for barnefamilier) i et nærliggende motellanlegg hvis dette er ønskelig. Vennligst kontakt TS Forum AS om det er ønskelig med alternativ overnatting.

Hjemreise
Møte avsluttes med lunsj kl. 15.00 på Dale-Gudbrands gård. Bussene kjører deretter tilbake til Lillehammer med ankomst ca. kl. 17.00. Det er mulig å overnatte på Lillehammer til natt til søndag eller ta buss til Oslo og eventuelt overnatte på hotell der. Husk å krysse av på registreringsskjemaet.

Konferanseavgifter/avbestilling

Deltageravgift ved påmelding før 15. mai 1998	NOK 2250
Deltageravgift ved påmelding etter 15. mai 1998	NOK 3000
Ledsageravgift voksen	NOK 1400
Ledsageravgift barn (under 16 år)	NOK 500

Frist for avbestilling er 1. juli 1998. Ved avbestilling etter 1. juli refunderes innbetalt deltageravgift minus NOK 500. Ledsageravgiftene og avgifter for ledsagerprogram refunderes i sin helhet. Ved sykdom refunderes innbetalt deltageravgift mot fremvist legeerklæring.

Påmelding sendes til
21.Nordiske statistikermøte

v/TS Forum AS
Postboks 14
2601 Lillehammer, Norway

Tlf. 61 28 73 20
Faks. 6128 73 30
E-mail: lillarra@online.no

Påmeldingsfristen er 15. mai 1998. Innbetaling av konferanseavgifter skal sendes innen påmeldingsfristen. Hotellutgifter betales direkte til hotellet ved avreise.

Innbetalingen via bank sendes til:
21. Nordisk statistikermøte
co. TS Forum AS
P.O.Box 14, N-2600 Lillehammer, Norway
Konto no. 2000.08.97699. Adresse: The Union Bank og Norway
SWIFT CODE: UBNONOKKLHM

Bekreftelse
Deltakerne får tilsendt bekreftelse på at melding om dato for ankomst og avreise datoer og innbetaling er mottatt.

Registrering
Kurssekretariatet er på hotellet onsdag 5. august 1998 fra kl. 12.00. Deltagerne henvender seg til sekretariatet for registrering. Ved registrering fås praktisk informasjon og eventuelle billetter til ledsagerprogrammet. Sekretariatet vil være åpent hver dag under arrangementet.På baksida eller på siste omslagsside bør begge komiteene stå:

Nordisk arrangementskomité:

Leiv Solheim	Statistisk sentralbyrå, Oslo
Otto Andersen	Danmarks Statistik, København
Antero Pohjola	Statistikcentralen, Helsinki
Helga Einarsdottir	Hagstofa Islands, Reykjavik
Lars Bergløf	Statistiska centralbyrån, Stockholm
Lennart Norberg	Statistiska centralbyrån, Örebro
Bjørg Moen	Statistisk sentralbyrå, Oslo

Lokal organisasjonkomité:

Pirjo Grüner	Statistisk sentralbyrå, Oslo
Eirik Gaaseide	Statistisk sentralbyrå, Oslo
Jorunn Lajord	Statistisk sentralbyrå, Oslo
Bjørg Moen	Statistisk sentralbyrå, Oslo
Leiv Solheim	Statistisk sentralbyrå, Oslo
Kjell Strand	Statistisk sentralbyrå, Kongsvinger
Otto Vaagen	Statistisk sentralbyrå, Oslo

REGISTRERINGSKJEMA

21. Nordiske statistikermøte, Lillehammer 5. - 8. august 1998

Vennligst benytt blokkbokstaver

Deltager:

Fornavn: _____ Etternavn:_____

Arbeidssted: _____

Adresse, arbeid: _____

Postnummer: _____ Sted:_____ Land:_____

Telefon, arbeid: _____ Telefaks: _____

E-mail adresse: _____

Jeg ønsker å delta i følgende sesjoner:_____

Ledsager(e):

Fornavn: _____ Etternavn: _____ Alder på barn: _____

Fornavn: _____ Etternavn: _____ Alder på barn: _____

Fornavn: _____ Etternavn: _____ Alder på barn: _____

Avgifter:

		Antall	Beløp
☐ Deltageravgift ved påmelding før15. 05. 98	NOK 2250	_____	_____
☐ Deltageravgift ved påmelding etter 15.05.98	NOK 3000	_____	_____
☐ Ledsageravgift voksen	NOK 1400	_____	_____
☐ Ledsageravgift barn	NOK 500	_____	_____
Ledsagerturer;			
☐ OL-sightseeing 6.8.98	NOK 160	_____	_____
☐ Maihaugen friluftsmuseum 6.8.98 voksne/barn	NOK 60/30	_____	_____
☐ Hunderfossen familiepark 7.8.98	NOK 180	_____	_____
☐ Aulestad og Segalstad gård 7.8.98	NOK 370\320	_____	_____

Opplevelsestur til Sjoa lørdag 8.august

Jeg deltar på turen, antall personer _____(pris inkl. i avgift)

☐ Jeg deltar ikke på turen

Totalt **NOK** ___

Overnatting/bespisning:

Hotellregningen betales på hotellet ved avreise. Prisene inkluderer frokost og lunsj i konferanseperioden. Konferanseavgiften dekker grillparty på onsdag, festmiddag på fredag og bespisning under det sosiale programmet på lørdag. På torsdag dekker deltagerne selv eventuell middag.

Quality Lillehammer Hotel 5. -8. august. 98

☐ Enkeltrom, NOK 2800

☐ Dobbeltrom, deltager med ledsager, NOK 3950

☐ Dobbeltrom, 2 deltagere, NOK 4050

Barn:

☐ 0-5 år, gratis i foreldrenes værelse

☐ 6-15 år, NOK 450, ekstra seng i foreldrenes værelse

Jeg ønsker å dele rommet med:_____

VEND!

Quality Lillehammer Hotel, opphold før og etter konferansen

☐ Enkeltrom, inkl. frokost, NOK 645.- pr. døgn
☐ Dobbeltrom, inkl. frokost, NOK 860.- pr. døgn

Barn: ☐ 0-5 år, gratis i foreldrenes værelse
☐ 6-15 år, kr. 50.- pr. barn i foreldrenes værelse

Ankomst dato: _____ Avreisedato: _____

Jeg ønsker å dele rommet med: _____

Birkebeineren Motel og Leiligheter, Lillehammer, opphold fra 8. - 9. august 98

☐ Enkeltrom, NOK 505 inkl. frokost
☐ Dobbeltrom, NOK 710 inkl.frokost

Barn: ☐ 0-4 år, gratis i foreldrenes værelse
☐ 5-15 år, 50% rabatt i foreldrenes værelse

Jeg ønsker å dele rommet med: _____

Comford Hotel Børsparken, Oslo, opphold fra 8. - 9. august 98

☐ Enkeltrom, NOK 745 inkl. frokost
☐ Dobbeltrom, NOK 945 inkl.frokost
☐ Trippelrom/familierom, NOK 1130 inkl.frokost

Barn: Ekstra seng NOK 185 døgn

Jeg ønsker å dele rommet med: _____

Busstransport Oslo/Lillehammer

Det settes opp busser fra Oslo til Lillehammer onsdag 5. august. Avgang fra Oslo/Fornebu er kl. 13.30 og fra Oslo/ sentrum ved Rådhuset kl. 14.00. Det settes også opp buss fra Lillehammer til Oslo sentrum lørdag 8. august.

Onsdag 5. august

Jeg/vi ønsker å være med i fellestransport fra Oslo Fornebu flyplass til Lillehammer — Antall personer

Avgang fra Oslo/Fornebu kl. 13.30 _____
Avgang fra Oslo/Rådhuset kl 14.00 _____

Lørdag 8. august

Jeg/vi ønsker å være med i fellestransport fra Lillehammer til Oslo/Rådhuset — Antall personer

Avgang fra Lillehammer ca. kl. 18.00 _____

Betaling

Innbetalingen av **avgifter** må sendes i norske kroner (NOK) *innen 15. mai 1998.*
Hotellutgifter betales direkte til hotellet ved avreise.

Via Bank: 21. Nordiske statistikermøte
co. TS Forum AS
P.O. Box 14, N-2601 LILLEHAMMER, Norway
Konto no. 2000.08.97699. Adresse: The Union Bank of Norway.
SWIFT Code: UBNONOKKLHM

Vær vennlig å føre navnet på innbetalingen.

Vennligst returner skjemaet til : TS Forum AS , P.O. Box 14, N-2601 Lillehammer, Norway.
Tel. +47 61 28 73 20 Fax: +47 61 28 73 30
Kontaktperson: Sirpa Kokslien

Dato:_____ Signature:_____

KONFERANSEOVERSIKT

TID	STED	TITTEL	KONTAKT
1998			
8.-12. juni	Elsinore Danmark	17th Nordic Conference on Mathematical Statistics http://www.dsts.dk/nordisk.konf/	Dansk Selskap for Teoretisk Statistik
17-19. juni	Trondheim	European Safety and Reliability Conference (ESREL 98)	Stian Lydersen, Institutt for matematiske fag, NTNU, 7034 TRONDHEIM Fax: 73 59 35 24, Tlf: 73 59 70 53 E-mail: Stian.Lydersen@stat.ntnu.no
21.-26. juni	Singapore	The Fifth International Conference on Teaching Statistics http://www.nie.ac.sg:8000/ wwwmath/icots.html	Secretary,The Organising Committee ICOTS-5, c/o Department of Economics and Statistics, National University of Singapore 10 Kent Ridge Cresent, Singapore 119260, Fax: (65) - 775 2646 E-mail: ecsicots@nus.sg
5.-8.august	Lillehammer	21. Nordiske Statistikermøte	lso@ssb.no
9-13. aug.	Dallas, USA	1998 Joint Statistical Meetings	ASA, 1429 Duke St. Alexandria, VA 22314-3402, USA Tlf: 703 68 41221 Fax: 703 68 42037 E-mail: meetings@asa.mhs.compuserve.com
24.-28. aug.	Bistol, UK	COMPSTAT 98 http://www.stats.bris.ac.uk./ COMPSTAT	Prof. Roger Payne, Statistics Department, IACR-Rothhamsted, Harpenden, Herts AL5 2JQ, UK Fax: +44 1582 760981
6.-11.sept.	Glasgow, Skottland	RSS 98. International Conference of the Royal Statistical Society http://www.stams.strath.ac.uk/ rss98/	Prof. Eric Renshaw Dept. of Statistics and Modelling Science, University of Strathclyde, 6 Richmond Street,Glasgow,G1 1XH E-mail: rss98@stams.strth.ac.uk
6.-7. nov.	Linkøping Sverige	Konferens om statistikkutbildning	Studierektor Anders Nordgaard Matematiska Institutionen Liköpings Universitet 581 83 LINKÖPING Fax: 013-100746, Tlf: 013-281974 E-mail: annor@mai.liu.se

NYTT FRA UNIVERSITETER OG HØGSKOLER

Melding fra lokalforeningen i Trondheim

Årsrsmøte i Statistisk Forening i Trondheim ble avholdt 25. november 1997, paa Gløshaugen. Styret fra forrige periode ble gjenvalgt; Håvard Rue (leder), Ola Diserud, Alfhild Lien Eide og Oddgeir Samset. Revisorene fra forrige periode, John Tyssedal og Stian Lydersen, tar også en ny periode. Foreningen melder om godt oppmøte på foredragkveldene. Mer informasjon finnes på lokalforeningens hjemmeside http://www.math.ntnu.no/~hrue/sf

Nytt fra gruppene for statistikk - NTNU

Ferdige cand.scient.

Michiel Tarald Engeland: Veileder: Steinar Engen.	Simultan testing av uavhengige eksperimenter
Rune Rasmussen: Veileder: Magne Aarset	Modellering av etterspørsel etter flyreiser og beregning av prognoser basert på statiske og dynamiske tilstandsmodeller
Tom Rune Ervik: Veileder: Helga Vivås	Statistikk i videregående skole - Et kompromiss mellom omfang av pensum og god forståelse av statistikk.
Geir Klingenberg Hansen: Veileder: Tor Stålhane.	Dependability og distributed computer controlsystems.
Inge Brandsæter: Veileder: Steinar Engen	Inhomogene diffusjonsprosesser som stokastiske analogier til deterministiske vekstmodeller
Ragnar Furunes: Veileder: Øyvind Bakke.	Aldersstrukturerte populasjoner
Kristin Ødegård: Veileder: Magne Aarset.	Klassifisering av flypassasjerer ved hjelp av clusteranalyse

Forskningsrapporter

Rapportene er tilgjengelige i postscript-format via nettadressen
http://www.math.ntnu.no/techreports/lists/statreports.html

Statistics no.9/1997:	Block updating in constrained Markov chain Monte Carlo sampling by Merrilee Hurn, Håvard Rue and Nuala A. Sheehan
Statistics no.10/1997:	Predicting Soccer Matches in a League by Håvard Rue and Øyvind Salvesen

Statistics no.11/1997: Stochastic simulation of porosity and acoustic impedance conditioned to seismic data and well data
by Alfhild L. Eide, Bjørn Ursin and Henning Omre

Statistics no.12/1997: Stochastic simulation of porosity and acoustic impedance conditioned to seismic data and well data from the Troll field
by Alfhild L. Eide

Nye doktorgradsstudenter

Seksjon Gløshaugen:

Bjørn Axel Gran, finansiert av Institutt for Energiteknikk, Halden.

Oddvar Husby, finansiert av NFR.

Odd Kolbjørnsen, finansiert av NFR.

Helge Langseth, finansiert av NFR.

Midlertidige ansettelser
Seksjon Lade:

Gunnar Taraldsen: Ansatt av ALS (Avdeling for lærerutdanning og skoleutvikling for å undervise S 101 som etterutdanning for lærere (fra hele landet), sommerundervisning og to helgesamlinger.

Helge Blaker: Timelærer og ansvarlig for undervisning i S 301 (Asymptotisk teori) høsten 1997.

Donald A. Mackenzie: Statistics in Britain 1865-1930. The Social Construction of Scientific Knowledge.

Edinburgh University Press 1981. ISBN 0 854224 369 3.

Statistikkfaget er temmelig nytt. Det første universitetsprofessoratet i faget vårt kom så sent som i 1911. Midler donert av Francis Galton ble brukt til å opprette denne stillingen for Karl Pearson. Boka dreier seg om en sosiologisk studie av engelske pionerer i statistikk fra Galton til Fisher. Forfatteren argumenterer for at et ledemotiv for engelsk statistikk i nevnte tidsrom var «eugenikk», dvs. hvordan menneskslektens genetiske sammensetning kunne (og burde) forbedres gjennom ulike tiltak.

Francis Galton (1822–1911) var nevø av Charles Darwin og dette er muligens hovedgrunnen til at han tok opp spørsmålet om hvordan man ved kvantitative metoder kunne «måle» hvordan evolusjonen fungerte i praksis. I forbindelse med dette arbeidet ble begreper som korrelasjon og regresjon innført.

Karl Pearson (1857–1936) tok opp ideene til Francis Galton rundt 1890 og ga Galtons begreper den matematiske form som vi kjenner idag. Han bygget opp sin «Biometric school», som utgjorde en sterkt sammensveiset gruppe av forskere rundt tidsskriftet Biomtrika. Han videreførte Galtons studier av evolusjon og kunne f.eks. hevde at korrelasjonskoeffesienten mellom fysiske egenskaper som kroppshøyde i etterfølgende generasjoner var 0,5. Det ble etter hvert et sterkt motsetningsforhold mellom biomtrikerne på den ene siden og tilhengerne av Mendels arvelover representert ved biologen William Bateson på den andre. Også tidligere medarbeidere av Karl Pearson som G. Udny Yule (1871–1951) fikk føle bannstrålen fra «The Biomtric school» når han foreslo alternative mål for assosiasjon i en 2 x 2 – tabell til de Pearson brukte (tetrachoric correlation).

Boka antyder at W.S.Gosset (1876–1937) (Student) var den som skaffet R.A.Fisher jobb på Rothamstead forsøksstasjon i 1919. Det er også verd å merke seg at Gosset ikke var «eugeniker». Hans motiv var rett og slett kvalitetsforbedring for arbeidsgiveren Guinness bryggeri. Dette satte ham i stand til å frigjøre seg fra mange av biometrikernes forutinntattheter, f.eks. at det bare var meningsfullt å analysere store datasett. I 1908 kom det banebrytende arbeidet om student-fordelingen. Resultatene som Gosset gjettet seg frem til ut fra simuleringer ble senere bevist analytisk av Fisher.

R.A.Fisher (1890–1962) var som Galton og Pearson «eugeniker», men bortsett fra dette kunne det virke som om han og Pearson var uening om det meste. Det må ha ergret Karl Pearson at Fisher var i stand til å lage en syntese av evolusjon og Mendels arvelover (Nydarwinisme), slik at han tok Batesons parti i striden mellom disse. Ellers omskapte Fisher hele den statistiske metodelæren langt på vei slik den oppfattes idag. Men Fisher ga aldri opp sin tro på «eugenikk». Han skal visstok ha vært den eneste som praktiserte læren i praksis: Han fikk ikke mindre enn 8 barn! Han foreslo også rent praktiske tiltak for å oppmuntre de best tilpassede til å få mange barn: Barnetrygd proporsjonal med farens lønn. Det viste seg imidlertid at dette forslaget var dødfødt rent politisk.

Forfatteren diskuterer om statstisk metodelære ville blitt slik den ble uten «eugenikk» og konkluderer:

> «Without eugenics, statistical theory would not have developed in the way it did in Britain – and indeed might not have developed at all, at least till much later.»

Henrik Dahl

Olav Kallenberg: Foundations of Modern Probability Springer, New York 1997. ISBN 0-387-94957-7.

Da jeg deltok i Joint Statistical Meetings i Anaheim i august 1997, så jeg denne boken i hyllene til Springer Verlags utstillingsseksjon. Jeg hadde et visst forhold til forfatteren for 20 år siden. Jeg åpnet boken og fant at forordet begynte nokså ambisiøst. «Some thirty years ago it was still possible, as L`oeve so ably demonstrated, to write a single book in probability containing practically everything worth knowing in the subject. The subsequent development has been explosive. Thus it seems essential that we have at leat a general overview of the whole area, and we should do what we can to keep the subject together. The present volume is an ernest attemp in that direction.»

Selv om hans mål var å skrive en bok som inneholdt alt, var det klart at mye, som for eksempel «large deviations, Gibbs and Palm measures, interacting particle systems og Malliavin calculus», måtte utelates av plasshensyn. De er mulige emner for et eventuelt volum 2.

Kallenberg er akkurat som jeg opprinenlig utdannet ved Linjen for Teknisk Fysikk. Jeg husker fra våre samtaler den gangen for lenge siden hvordan vi var introdusert til målteoretisk sannsynlighets-teori - for oss begge var det L`oeves bok som åpnet den nye verden og som senere ble vår referensebok.

Jeg ble derfor interessert i å se hva han hadde skrevet. Jeg kjøpte boken og har lest i kapitlene om brownske bevegelser, stokastiske integral og differensiallikninger frem til det siste som er semimartingaler og generell stokastisk integrasjon. Disse kapitlene er spredt i boken, fra begynnelsen og til det siste. Jeg er imponert over å se hvor mange bevis jeg ikke har sett andre steder. Fremstillingen er meget økonomisk, noe som er nødvendig for å få så mye inn på bokens 500 sider.

Stoffet henger meget godt sammen. Dette er ikke stoff som er sakset litt fra en bok og litt fra en annen. Kallenberg gir en meget sammenhengende fremstilling. Dette gjør det lettere å få en oversikt over det hele. Det er ikke bare en skog full av trær - her får vi også et overblikk over helheten. Samtidig er bevisene klare og utførlige på de rette stedene.

Den økonomiske fremstillingen til Kallenberg er fin når en ønsker oversikt over den logiske strukturen. Hvorvidt det er den beste fremstillingen når en skal forelese etter boken, eventuelt ha dette som sin første innføring er en annen sak. Forfatteren sier at boken primært er tenkt som en generell referansebok selv om den også kan brukes til doktorgradskurs og seminar. Jeg tror at den er best egnet til det første. Den inneholder for få eksempler, og oppgavene er for teoretiske til at boken kan brukes alene i undervisningen. Da må den suppleres med Karatzas & Shreves, Revuz & Yor eller Øksendal.

For oss som har sett en del av dette stoffet i anvendelser tidligere er det en meget passende og velskrevet bok som anbefales på det varmeste, og jeg ser frem til volum 2.

Trygve Nilsen

Medlemskap i Norsk statistisk Forening

Kontingenten er kr 200 per år for regulære medlemmer og kr 50 for studenter. Korrespondanse om medlemskap og innbetaling skal skje til foreningens kasserer:

Norsk Statistisk Forening v/Trygve Almøy
Inst. for matematiske fag
NHL
Postboks 5035
1482 Ås

Innbetaling gjøres på foreningens postgirokonto 0802 51 51055. Medlemmer som er usikre på når de betalte kontingent sist kan nederst til høyre på adresselappen for Tilfeldig Gang se hvilket år siste betaling er registrert for. Vær imidlertid oppmerksom på at adresselappene for bladet kan ligge noe etter ajourføringen av medlemsregisteret, slik at nylig foretatte innbetalinger ikke vil være med her.

Normalt vil det i årets første nummer av Tilfeldig Gang ligge ved en giroblankett for innbetaling av kontingent. Den ligger derfor med dette nummer og kan benyttes av alle som ikke allerede har betalt inn kontingent for 1998.

Nye medlemmer melder seg til kassereren og får så tilsendt Tilfeldig Gang fra første påfølgende nummer av. Tidligere nummer av bladet kan dessverre ikke skaffes. ved utsendelsen av Tilfeldig Gang benyttes kopier av medlemsregisteret, slik at adresseforandringer ikke behøver meldes separat til bladet.

Annen korrespondanse med foreningen skal foregå med formannen:

Odd O. Aalen
Seksjon for medisinsk statistikk
Universitetet i Oslo
Postboks 1122, Blindern
0317 Oslo

Bidrag til Tilfeldig Gang

Neste nummer av Tilfeldig Gang vil komme ut i juni 1998. Alle slags bidrag. 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 redaktøren i hende senest 1. juni 1998.

Adressen til redaksjonen er

Postadresse:	Leiv Solheim Statistisk sentralbyrå Postboks 8131, Dep 0033 OSLO
Telefon:	22 86 46 87
Fax:	22 86 49 98
e-mail:	lso@ssb.no

SWENSEN, ANDERS RYGH
STATISTISK SENTRALBYRÅ
P.B. 8131 DEP.
0033 OSLO 1 97

B-post

Retur:

Tilfeldig Gang v/Trygve Almøy
Inst. for matematiske fag
NHL
Postboks 5035
1482 ÅS