ISSN 0803-8953

TILFELDIG GANG
Nr. 4, årgang 9
Desember 1992

Utgitt av Norsk Statistisk Forening

INNHOLD

	Side
Fra redaktøren *Jan F. Bjørnstad*	1
Fra formannen i Norsk Statistisk Forening. *Trygve Nilsen*	2
Utdanning	
Nasjonalt doktorgradskurs om telleprosesser. *Ørnulf Borgan*	3
Antall cand.scient./dr.scient & siv.ing./dr.ing. i statistikk, 1984-92.	5
Kvalitet og statistikk - Utfordringer for elementærundervisningen. *Jostein Lillestøl*	6
Presentasjoner	
Norsk Regnesental. *Anne Marie Fenstad*	8
Faglige innlegg	
Noen statistiske problemstillinger innen helsetjenesteforskning. *Solfrid Lilleeng*	13
Sammenligning av forskjellige prediksjonsmetoder når bare noen få komponenter er relevante; et simuleringsforsøk. *Trygve Almøy*	15
Nevrale nett og statistikk. *Knut Liestøl*	20
Avisklipp:	
Vannhode mulig atomskade. *Aftenposten kveldsnummer 24.11.92*	28
Diverse fra *Natur & Miljø 20.11.92*	29
Non-cents: Laws of probability could end need for change. *The Milwaukee Journal, May 92*	31
Flerfaglighet i et amerikansk statistikkmiljø. *Ivar Heuch*	32
Konferanser	
Rapport fra hovedkonferansen 1992 i ASA/IMS. *Jan F. Bjørnstad*	36
Det norske statistikermøte 1993	37
Årsmøtet i American Statistical Association/Institute of Math. Statistics, 1993	38
XXVèmes Journees de statistique, 1993	39
15. Nordiske konferanse i matematisk statistikk, 1994	40
14th International Symposium on forecasting, 1994	40
3rd World Congress, 1994	41
Meldinger	
Nytt fra universiteter og høyskoler	42
Annonse: Stipendiatstilling ved AVH, Universitetet i Trondheim	44
Annonse: Statistiker. Statens Institutt for Folkehelse	45
Medlemskap i Norsk Statistisk Forening	46
Bidrag til TG	46

E.T.

Fra redaktøren

Dette "julenummeret" av TG inneholder faglige og andre innlegg av stor variasjon, forhåpentlig noe for enhver smak. Blant annet har vi en presentasjon av det første nasjonale doktorgradskurset i statistikk. Kurset har et spennende pedagogisk opplegg, med to atskilte ukers intensiv opplæring. Det skal ledes av Ørnulf Borgan og omhandler statistisk analyse for telleprosesser. Her bør alle dr.gradsstudenter og de av dere som planlegger å ta en doktorgrad, også utenfor Oslo, gripe anledningen. Uansett om temaet er sentralt for doktorgraden er det et relevant og interessant område av statistikken som her presenteres. Forhåpentlig vil vi i fremtiden få regelmessige nasjonale doktorgradskurs i forskjellige områder av faget, gitt av de "statistikkinstituttene" som har spisskompetansen. I denne forbindelsen er det viktig at dr.gradsstudentene støtter opp om tilbudet, og at deres institutter legger forholdene til rette økonomisk og praktisk.

Av andre innlegg i dette nummeret kan nevnes en presentasjon av Norsk Regnesentral, det tyngste statistikkmiljøet utenfor universitetene, ved Anne Marie Fenstad. Knut Liestøl gir en gjennomarbeidet og delvis kritisk introduksjon til nevrale nett, ett av de nye moteområdene i statistikk. Også i dette nummeret har vi et innlegg fra møtet på Røros, denne gangen fra Trygve Almøy om sammenlikning av prediksjonsmetoder. Det er interessant å se hvordan Almøy planlegger og analyserer simuleringseksperimentet som et faktorielt forsøk.

En liste over antall cand.scient./siv.ing. i statistikk for perioden 1984-92 viser at det har vært en betydelig økning de seneste årene. Vi får håpe at denne tendensen holder seg, og at det også vil medføre en økning i antall dr.scient./dr.ing.kandidater i fremtiden.

Vi har førstegangsannonseringer for det norske statistikermøtet 1993 som skal foregå i perioden 14. - 16. juni på Ustaoset høyfjellshotell, og for det nordiske møtet i 1994 som skal holdes i Lund i Sverige. Legg merke til at det er flyttet til august istedenfor juni.

Til slutt vil jeg på vegne av Norsk Statistisk Forening gratulere Emil Spjøtvoll med valget til det prestisjefylte vervet som rektor ved NTH. Selv om det kan være belastende for våre små statistikkmiljøer, er det positivt at statistikere markerer seg ved å ta lederverv slik Spjøtvoll og Bent Natvig (dekanus ved Mat.nat.-fakultetet i Oslo) har gjort. Det er ikke utenkelig at vår faglige ballast og erfaring med samarbeid over faggrenser kan være et godt grunnlag for lederverv av forskjellige slag.

Med et begrunnet håp om utgivelse før jul ønskes leserne en God jul og Godt nyttår!

Jan F. Bjørnstad

Fra formannen i Norsk Statistisk Forening

I selskapelig lag får vi statistikere stundom servert morsomheten: "Lies, damned lies and statistics". Det er ikke bare hyggelig å ha en profesjon med det omdømmet. Det er skrevet en del underholdende bøker om emnet. De to mest kjente er kanskje: Darrel Huff: *How to Lie with Statistics* og Robert Hooke: *How to Tell Liars from the Statisticians.*

Det finnes enkelte vitenskapsmenn som lever etter regelen: "Bestem deg først for konklusjonen og skaff deg deretter data som støtter opp under den." Stundom kan vedkommende være vel ivrig. En professor i kvinneforskning presterte følgende i Bergens Tidende 7. november 1986: "I 1974 satt det kvinner i 1.8 % av professoratene og 6.6 % av dosenturene. I alt var det altså for 12 år siden kvinner i 8.4 % av toppstillingene. I dag er dosenturer omgjort til professorater og kvinner utgjør 5 % av toppstillingene. Det har altså foregått en reduksjon av kvinners andel av toppstillingene." Her er det nokså klart for den oppmerksomme leser at noe er feil i resonnementet. Det førte da også til litt ablegøyer i avisen i ukene som fulgte. (Professoren fattet poenget da noen hjelpsomme sjeler ba henne om å utføre det tilsvarende regnestykket for menn.)

Nå er ikke alt misbruk like lette å oppdage som dette. I mange tilfeller holdes kortene, dvs. utvalgsmetode og data, tett inntil brystet. Det eneste som offentliggjøres, er bastante konklusjoner. For noen år siden var det en vel høy strykeprosent til eksamen for begynnerstudenter ved Universitetet i Bergen. En pedagog tilbød seg å utføre en undersøkelse, og de nødvendige midlene ble bevilget. Spørreskjema ble sendt til studenter og lærere. Av hensyn til personvern skulle det ikke oppgis data som ville tillate identifisering person. Dette førte da selvsagt i sin tur til at pedagogen ikke hadde mulighet til å sammenlikne forelesernes holdninger og meninger med studentenes svar og eksamensresultater. Etter et kort blaff ved avslutningen av prosjektet der publikum ble servert noen konklusjoner, døde det hele hen og universitetet var noen millioner kroner fattigere.

Det er ikke fagstatistikere som har utført disse arbeidene. Det første eksempelet er ikke av det slaget som vi betrakter som statistikk, det er elementær brøkregning. Men her faller ikke vår mening av hva som er statistikk sammen med den gjengse oppfatningen. Det andre eksempelet kommer fra sentrale deler av den praktiske statistikk.

Spørsmålet er om vi som statistikere skal være aktive i slike sammenhenger og peke på mangler og feil. En ting er hva vi ville gjøre dersom vi ble bedt om råd, en annen ting er om vi skal blande oss i saker der vi ikke er bedt om å uttale oss. Jeg har følelsen av at vår bakgrunn fra matematiske institutter har oppdradd oss til å være tilbakeholdne og ikke påpeke andres feil offentlig, dvs. ikke delta i polemikk om andres arbeider. Jeg er ikke sikker på at denne holdningen er riktig. Vi er samfunnets eksperter i faget og har med det et ansvar for at det ikke sjuskes vekk betydelige midler ved dårlig arbeid.

Trygve Nilsen

NASJONALT DOKTORGRADSKURS OM STATISTISKE MODELLER OG METODER FOR TELLEPROSESSER

Innhold og formål: De siste 10-15 årene har det vært en stor skandinavisk og internasjonal forskningsaktivitet om bruk av teorien for telleprosesser, martingaler og stokastiske integraler for å studere og videreutvikle statistiske metoder i levetids- og forløpsanalysen. Formålet med kurset er å gi deltagerne en oversikt over dette forskningsområdet, og å sette dem i stand til på egen hånd å følge med i faglitteraturen.

Deltagelse: Kurset er sentralt for dr.grads-studenter som benytter teknikker og metoder fra levetids- og forløpsanalysen i sitt forskningsarbeid. Dette gjelder uansett om forskningsarbeidet er relatert til problemstillinger innen biostatistikk eller teknologi (pålitelighetsanalyse). Kurset er også velegnet som et "breddekurs" under dr.scient-graden for studenter som ikke får direkte bruk for dette stoffet i sitt forskningsarbeid. Cand.scient-studenter som tar sikte på et dr.studium vil også kunne ha utbytte av kurset.

Forkunnskaper: Det er en fordel å ha kunnskaper om stokastiske prosesser og om levetids- og forløpsanalyse. Men kurset kan med stort utbytte også følges av studenter uten en slik bakgrunn. Da dette er et dr.grads-kurs i statistikk, forventes det imidlertid at kursdeltagerene har kunnskaper i statistikk på "cand.scient-nivå", og at deltagerne har evne til å arbeide selvstendig og er innstilt på å gjøre en innsats.

Tid, sted og lærere: Kurset avholdes ved Matematisk institutt, Universitetet i Oslo og går over to perioder, 15.-19. mars og 3.-7. mai 1993. Kursleder er førsteamanuensis Ørnulf Borgan, Matematisk institutt, Universitetet i Oslo. Undervisningen første periode vil bli gitt av kursleder, mens lektor Per Kragh Andersen, Statistisk Forskningsenhed, Københavns Universitet og Ørnulf Borgan vil dele på undervisningen i andre periode.

Kursplan: For hver av de to kurs-periodene vil dagene være delt i to med forelesninger før lunsj (kl. 9-12) og oppgavearbeid, diskusjon, o.l. etter lunsj (kl. 13-16). Kursplanen for første periode er:

Mandag 15/3: Presentasjon av noen motiverende eksempler og kort innføring i teorien for telleprosesser, martingaler, stokastiske integraler.

Tirsdag 16/3: Produkt-integral, likelihood og partiell likelihood. Spesifikasjon av telleprosess modeller. Sensurering og trunkering.

Onsdag 17/3: Ikke-parametrisk estimering av integrert intensitet ved Nelson-Aalen estimatoren.

Torsdag 18/3: Mer om produkt-integral. Den funksjonelle delta-metoden. Ikke-parametrisk estimering av overlevelsesfunksjonen ved Kaplan-Meier estimatoren.

Fredag 19/3: Ikke-parametrisk testing for ett og flere utvalg.

Kursplanen for andre periode er:

Mandag 3/5: Parametriske modeller og ML-estimering. Noen grafiske metoder for modellkontroll.

Tirsdag 4/5: Coxs regresjonsmodell for telleprosesser.

Onsdag 5/5: Coxs regresjonsmodell, fortsatt.

Torsdag 6/5: Modellkontroll for Coxs regresjonsmodell.

Fredag 7/5: Aalens lineære regresjonsmodell.

Selvstudium, oppgaver og eksamen: Det forventes at studentene ved selvstudium setter seg inn i nærmere oppgitt litteratur før første kursperiode (se nedenfor) og at de arbeider med stoffet også mellom de to kursperiodene. Eksamen vil bestå av to deler. Første del er et lite oppgavesett som vil bli tilsendt deltagerne i midten av april med innleveringsfrist ved starten av andre kursperiode. Andre del består av et "mini-prosjekt" som vil bli gitt ved avslutnigen av andre kursperiode med innleveringsfrist ca. 1. juni. Ved bedømmelsen vil første del telle 1/3 og andre del 2/3. Bedømmelsen vil bli gitt som bestått/ikke bestått for dr.scient-studenter og ved tallkarakter for (eventuelle) cand.scient-studenter.

Vekttall: Kurset er normert til 3 vekttall.

Litteratur: Kurset vil bli basert på monografien *Statistical Models based on Counting Processes* av P.K. Andersen, Ø. Borgan, R.D. Gill og N. Keiding som vil utkomme på Springer-Verlag rundt nyttår. En enklere introduksjon til emnet er gitt i artikkelen "Counting process models for life history data: A review" av P.K. Andersen og Ø. Borgan (*Scand. J. Statist.* **12**, 97-158;1985). Boka til T.R. Flemming og D.R. Harrington *Counting processes and survival analysis*, Wiley, 1991, behandler mye av det samme stoffet.

Foreløpig påmelding: For å få nærmere kartlagt interessen for kurset, bes de som tenker seg å delta melde fra om dette innen 15. januar 1993 til Ørnulf Borgan (postadresse: Matematisk institutt, P.b. 1053 Blindern, 0316 Oslo; elektronisk post: borgan@math.uio telefon (02)855862/855887). De som melder sin interesse vil bli nærmere orientert om hva de forventes å sette seg inn i på egen hånd før kursstart.

UTEKSAMINERTE CAND.SCIENT./SIV.ING. OG DR.SCIENT./DR.ING I STATISTIKK, 1984-1992.

Tabellen presenterer produksjonen av cand.scient.(UiO,UiB,UiTø,AVH) og siv.ing.(NTH) i statistikk for årene 1984-1992, inkludert noen cand.real.i årene 84-85. Tallene i parentes for UiO gjelder forsikringsmatematikk-delen av totalen. UNIT,AVH startet med cand.scient.-utdanningen i 1987. Tallene for 1992 er ufullstendige. Ellers tas forbehold om eksakt korrekthet da enkelte av årstallene er tatt fra opplysninger i tidligere TG.

Cand.scient./siv.ing.

	UiO	UiB	UiTø	UNIT,AVH	Ialt	UNIT,NTH
1984	1(1)	-	-	-	1	4
1985	6(0)	1	1	-	8	4
1986	7(4)	-	1	-	8	-
1987	5(3)	1	-	-	6	6
1988	9(1)	3	-	-	12	5
1989	4(2)	2	-	-	6	3
1990	7(3)	5	-	3	15	12
1991	14(3)	1	1	-	16	4
1992	9(2)	2	-	2	13	16
Ialt	62(19)	15	3	5	85	54

Dr.scient.(dr.ing. ved NTH)

1984	1	-	-	-	1	-
1985	-	-	-	-	-	-
1986	-	-	-	-	-	-
1987	-	-	-	-	-	-
1988	1	-	-	-	1	2
1989	2	-	-	-	2	-
1990	1	2	-	-	3	1
1991	-	-	-	-	-	-
1992	-	-	-	-	-	-
Ialt	5	2	-	-	7	3

Kvalitet og statistikk

- Utfordringer for elementærundervisningen

I de senere år er det kommet en rekke nye bøker over temaet kvalitet og statistikk. Disse er betydelig annerledes enn de tallrike eldre bøker i statistisk kvalitetskontroll, der stikkprøvemetoder for produktkontroll (aksept/forkast) var viet stor plass. I tråd med det moderne syn på kvalitetsstyring, der forebygging og forbedring prioriteres, blir hovedvekten nå lagt på prosesskontroll, dvs. at Shewhart og Demings idéer omsider vinner frem. Vi finner også innslag av moderne forsøksplanlegging.

Flere av disse bøkene kan faktisk brukes i innføringskurser i statistikk, der en sammen med sentrale statistiske begreper, blir innført i statistisk tenkemåte, og en systematikk for praktisk kvalitetsforbedringsarbeid. Stoffet er av interesse ikke bare i industriell og økonomisk virksomhet.

I denne litteraturen finnes også tilløp til ideologiske diskusjoner. Man stiller spørsmål ved verdier av tradisjonell inferensteori, og prioriterer prosessorientert tenkemåte, der prediksjon og søken etter årsaksforhold står mer sentralt. I denne sammenheng diskuteres ofte "enumerative studies" vs, "analytic studies", et tema som ble tatt opp av Deming allerede i 1950 årene.

En enumerativ studie tar sikte på å si noe om en veldefinert populasjon ut fra en stikkprøve, mens en analytisk studie skal hjelpe oss til å forstå årsaker (f.eks. til variasjon). I en analytisk studie tillater vi at populasjonen er udefinert eller kan undergå endringer over tid. Observasjon over tid vil være nødvendig for å forstå årsakene i alle fall, og antakelser om uavhengighet over tid er ofte en tvangstrøye. ("There is no such thing as a stable prosess"(Box)). For å konkretisere forskjellen mellom enumerative og analytiske studier:

Å beskrive resultatet av en dags produksjon for å se hva som skjedde denne dagen er i statistisk forstand en enumerativ studie. Hvis vi tar en stikkprøve fra dagsproduksjonen kan denne ha to ulike formål: Dersom hensikten er å bestemme hva vi skal gjøre med resten av produksjonen den dagen er studien fortsatt enumerativ, mens dersom hensikten er å lære hvordan vi skal gjøre det bedre i morgen trengs en analytisk studie.

Vi må da søke kunnskap om hvordan systemet bør endres for å forbedre ytelsen. Statistisk analyse kan ikke bidra til dette alene, men statistikere må bedre lære seg å se sine egne metoder innen en slik ramme. Dette er en klar utfordring, spesielt når det fra mange hold hevdes at de fleste problemstillinger av praktisk interesse krever analytiske studier.

Hypotesetesting av at to behandlinger er like har liten appell hos praktikere. De vet som regel på forhånd at det er forskjell. Spørsmålet er om forskjellen er stor nok til å forsvare kostnadene og risiko ved endringer.

"Tests of hypothesis and tests of significance belong to the philosophy of some other world, not this one" (Deming).

Litteratur:

Beuregard, Mikulak & Olson: A practical Guide to Statistical Quality Improvement, New York, Van Nostrand Reinhold 1992

Deming: Out of the Crisis, Cambridge Mass. , MIT-CAES 1986

Gitlow, Gitlow, Oppenheim & Oppenheim: Tools and Methods for the Improvement of Quality, Homewood IL, Irwin 1989.

Joiner: The key role of statisicians in the transformation of North American industry. The american Statisician, August, 1985.

Flynn: All answers are approximate... and temporary: Statistical thinking and total Quality Management. ASQC Quality Congress Transactions, San Francisco 1990.

Deming: On the statisticians contribution to quality. Invited paper 46th session of the International Statistical Institute, Tokyo 198

Jostein Lillestøl

Norsk Regnesentral

9. desember 1992

Norsk Regnesentral (NR) ble grunnlagt i 1952 av NTNF. I 1985 ble NR reorganisert til en uavhengig stiftelse. Det var NR som i sin tid utviklet det første objektorienterte programmeringsspråket i verden, nemlig SIMULA. Dette gjorde NR kjent internasjonalt.

NR er et oppdragsbasert forskningsinstitutt. Vi arbeider med anvendt forskning innen informasjonsteknologi (IT) og statistisk modellering. NR er et nasjonalt kompetansesenter med internasjonal anerkjennelse innen bildeanalyse, forvaltningsinformatikk, miljø- og ressursevaluering. NR har organisert sin virksomhet i seks forskningsgrupper hvor de tre som spesielt angår statistikk vil bli presentert nedenfor.

Det er tilsammen 43 forskere som utfører oppdrag innen statistisk modellering ved NR. Dette er det største statistikkmiljøet i Norge. NR holder til i samme bygning som Institutt for informatikk ved UiO og USIT (Universitetets sentrale EDB-tjeneste) i Gaustadbekkdalen. Den geografiske beliggenheten er en viktig faktor i det nære samarbeidet vi har med Universitetet i Oslo.

Bildebehandling og mønstergjenkjenning (BILD)

Enkeltbilder og sekvenser av bilder som film og video er essensielle informasjonsformidlere i moderne samfunn. Til nå har den alt vesentlige delen av bildene som benyttes vært analoge. Utnyttelsen av digitale bilder har vært betydelig hemmet av de store datamengdene selv små enkeltbilder representerer.

Imidlertid har den raske økningen av datamaskinenes ytelse og lagringskapasitet, kombinert med et samtidig stort fall i prisene, endret randbetingelsene radikalt. Kostnadene på maskinvaren er nå for svært mange anvendelser av digital bildebehandling og mønstergjenkjenning ikke lenger noen skranke for utviklingen av kostnadseffektive og konkurransedyktige produkter.

Den store utfordringen er nå å utvikle robuste algoritmer som fyller de krav et sterkt ekspanderende marked setter. Dette har vært og er BILDs mål for noen sentrale anvendelser

av høyteknologi innen norsk industri og forvaltning der det er behov for bildebehandling og mønstergjenkjenning. Slike høye mål er realistiske: BILD ble i 1989 utpekt av NTNF som Norges "Center of Excellence" innen dette området og tildelt programledelsen av et stort forskningsprogram. Dette ble i 1991 forlenget til 1994 og har nå høyteknologibedriftene Hafslund Nycomed AS, Simrad Subsea AS, SysScan Mapping AS, Tomra Systems AS og Vingmed Sound AS med i styringskomitéen.

Medisinsk bildebehandling

Arbeidet innen medisinsk bildeanalyse/mønstergjenkjenning har i 1992 vært konsentrert om tolkning av magnet-resonans-bilder av hjernen, ultralyd-undersøkelser av strømningsmønstre i hjertet, og gjenkjenning av mønstre i EEG (elektro-encephalografi).

Dokumentbehandling

Innenfor dette feltet har hovedvekten av aktivitetene vært lagt på automatisk og interaktiv datafangst fra kart og skjemaer. I kart ligger ofte den viktigste informasjonen i de geometriske strukturene og linjene som representerer høydekurver, kystlinjer, veier o.l. I skjemaer vil teksten og plasseringen av denne være det viktigste. Datamaskinen har imidlertid ingen forståelse av hva det digitale dokumentet inneholder.

I løpet av 1992 har vi jobbet spesielt med metoder for linjefølging slik at geometrisk informasjon kan trekkes ut fra kart. Vi har også vært opptatt av gjenkjenning av symboler som er aktuelt både for kart og skjemaer.

I tillegg til å drive metodeutvikling, har vi også hatt oppdrag der vi har vurdert eksisterende systemer innenfor bilde-/dokumentbehandling.

Fjernmåling

Avrenning fra høstpløyd mark er et stort forurensingsproblem, og Norge har forpliktet seg til å halvere slike utslipp til Nordsjøen innen 1995. I samarbeid med Norsk Institutt for Jord- og Skogkartlegging og Jordforsk har NR utført et prosjekt (1990-92) for Statens Forurensningstilsyn der vi har utprøvd forskjellige metoder for overvåkning av jordarbeiding i erosjonsutsatte områder ved hjelp av optiske satellittbilder. Prosjektet ga gode resultater, og i 1992 har NR i samarbeid med nevnte institutter laget et operativt system for slik overvåkning.

I samarbeid med Simrad Subsea AS har NR begynt å utvikle en pakke for klassifikasjon av havbunn basert på sonardata fra Simrad Subsea AS. Programvaren og metodene som NR utvikler skal brukes til å avgjøre om havbunnen består av stein, sand, mudder osv.

Industriell inspeksjon

Tomra Systems AS har godt etablerte produkter for gjenkjenning i sann tid av flasker og kasser med flasker. Sanntidskravet har gjort at robuste algoritmer basert på statistiske metoder tidligere ikke har vært prøvd. Potensialet i statistiske metoder for denne anvendelsen ble klart demonstrert i et forprosjekt, og metodene ble videreutviklet i 1992. Ved hjelp av avanserte matematisk-statistiske metoder, har det lykkes det NR å skille svært like flasketyper fra hverandre. Disse algoritmene vil bli del av en helt ny generasjon kasseautomater hvor prototypen utvikles i 1993.

Statistisk analyse av naturressursdata (SAND)

Utvinning av naturressurser krever store investeringer og det er ofte stor usikkerhet forbundet med dem. Statistisk metodikk er spesielt velegnet til å redusere og kvantifisere denne usikkerheten. Denne delen av statistikk kalles geostatistikk. I modellering av naturressurser er den romlige variasjonen viktig. For å modellere romlig variasjon er man avhengig av simulering på datamaskin.

Gruppen har valgt å rette seg mot petroleumsanvendelser. Dette er et betalingsdyktig marked med mange interessante faglige problemstillinger. Samtidig er det et marked som krever høyt faglig nivå samt god gjennomføringsevne.

Reservoardata

I dette innsatsområdet analyseres data fra reservoarer. Da det er svært dyrt å hente inn data, er det viktig at dataene analyseres best mulig. SCANDPOWER og NR har inngått en avtale om utviklingen av programmet DECISION. DECISION brukes til forsøksplanlegging i forbindelse med reservoar simulering.

Anslag på tilstedeværende ressurs

Før man beslutter utbygging av et felt er det viktig å ha et anslag på tilstedeværende mengde hydrokarboner i reservoaret med usikkerhet og estimater på hvordan ressursen fordeler seg i forskjellige deler av reservoaret. Gruppen har arbeidet med denne problemstillingen i flere år og har i tillegg til betydelig metodeutvikling også gjennomført studier av ca. 10 felt.

Reservoarkarakterisering

Under produksjonen fra et felt er det viktig å kunne karakterisere reservoaret best mulig. Det er en forutsetning for å kunne øke produksjonen. Det er viktig å kunne integrere all informasjonen konsistent og å kunne beskrive usikkerheten best mulig.

Produksjonsparametre

PROFIT PROGRAMMET er en sammenslutning av 13 oljeselskaper og OD. Det har gitt gruppen et prosjekt for å studere hvordan absolutt permeabilitet avhenger av skalaen og hvordan man kan finne blokkeffektive verdier på en grovere skala. Flere nye estimatorer er utviklet.

Statistisk dataanalyse (STAT)

Dagens beslutningstakere møter en virkelighet uttrykt i tall, og de trenger gode hjelpemidler for å forholde seg til denne informasjonen. En datamaskin gir oss tilgang til store mengder opplysninger ved et tastetrykk - men vår forståelse av dataene behøver ikke være på et tilsvarende nivå. Det produktet som gruppen tilbyr er nettopp metoder for å trekke forståelse og beslutningsgrunnlag ut fra slike datamaterialer. Vår kompetanse er i hovedsak matematisk statistikk og dataanalyse, men også nærliggende fagfelt som grafisk representasjon og geografiske informasjonssystemer er viktige.

STAT utfører forskningsoppdrag for forvaltning, industri og næringsliv. Vi søker å konsentrere aktiviteten rundt visse forskningstemaer og innsatsområder. I lengre tid har de viktigste temaene vært miljø- og ressurskartlegging, geografiske informasjonssystemer (GIS), trafikk-analyser samt økonomisk-administrativ modellering. I året som gikk har gruppen rettet sin forskning mot industrielle anvendelser.

86 736 vågehval

I flere år har vi deltatt i arbeidet med analyse av bestandsdata for vågehval. En foreløpig kulminasjon ble nådd da det endelige bestandsanslaget, med tilhørende konfidensintervall, (60 700, 117 400), kunne fremlegges for internasjonale hvalforskere i sommer. Tallene ble godkjent som det ofisielle anslaget til Den internasjonale hvalfangstkommisjonen.

Beregningene av vågehvalbestanden ble basert på en nyutviklet, EDB-intensiv metode, simulert sannsynlighetsmaksimering. Denne metoden gjør det mulig å utnytte effektivt de sannsynlighetsmodellene som er så komplekse at de bare kan studeres gjennom simulering.

Verktøy i industriell matematikk

I samarbeid med Senter for Industriforskning og Universitetet i Oslo har NR innledet et fire-årig prosjekt innen industriell matematikk, med STAT som ansvarlig for statistikkdelen. Prosjektet, som er finansiert av NTNF som et såkalt strategisk teknologiprosjekt, skal utvikle basis-programverktøy slik at terskelen for bruk av avanserte matematiske metoder skal være lavere, så lav at en vid utbredelse av metodene blir mulig.

I 1992 har vi startet dette arbeidet på følgende områder; romlig statistikk med anvendelser innen ressurs- og miljøkartlegging, moderne metoder for analyse av samvariasjon og årsak-virkningsrelasjoner mellom flere variable som har vide anvendelser innen mange av NRs innsatsområder, samt ikke-lineære dynamiske modeller med anvendelser fra petroleumsreservoarer til temperatur- og nedbørsprognoser.

Andre prosjekter

Gruppen har arbeidet med forskjellige prosjekter innenfor sjøpattedyr- og fiskeriforvaltning. Et omfattende datamateriale fra Havforskningsinstituttets tokter er analysert med henblikk på å øke forståelsen av fiskebestandens dynamikk og på å lage bedre prediksjonsmodeller for fiskebestanden.

Innen GIS-området fortsetter arbeidet med utvikling og vedlikehold av den geologiske feltdatabasen SAFARI med et konsortium av OD og de norske oljeselskapene som oppdragsgivere.

Denne beskrivelsen har tatt utgangspunkt i årsrapporten for NR. Meningen har vært å gjenspeile en del av hva statistikere på NR driver med. Som det går frem er statistikkmiljøet her stort og variert, noe som gjør at mulighetene når det gjelder arbeidsoppgaver er mange og utfordrende.

Anne Marie Fenstad

NOEN STATISTISKE PROBLEMSTILLINGER INNEN HELSETJENESTEFORSKNING

ved Solfrid Lilleeng, SINTEF-NIS.

Helsetjenesteforskning er et felt hvor problemstillingene ofte har både samfunnsfaglige og medisinske innfallsvinkler. Det dukker ofte opp problemstillinger hvor de tradisjonelle statistiske metodene må suppleres med mere avanserte metoder både på grunn av problemstillingens natur og datagrunnlaget man tar utgangspunkt i. Hensikten med dette innlegget er å fortelle om noen slike problemstillinger som har dukket opp i min jobb på SINTEF-NIS (Norsk institutt for sykehusforskning).

Pasientklassifiseringssystemer

Det finnes mange pasientklassifiseringssystem. Ett av dem er DRG (Diagnose Relaterte Grupper) og er utviklet for innlagte pasienter ved sykehusene, i motsetning til APG (Ambulatory Patient Groups) som er et tilsvarende system for polikliniske pasienter (pasienter som får behandling som ikke krever innleggelse ved sykehuset).

DRG er et amerikansk pasientklassifiseringssystem som ble utviklet ved Yale-universitetet i USA i slutten av 60-årene. Systemet ble utviklet som et informasjon - og kvalitetssikringssystem for amerikanske sykehus, men ble senere tatt i bruk til finansiering. Systemet klassifiserer pasientene i 470 grupper. Til hver DRG er det knyttet en kostnadsvekt som gjenspeiler det relative ressursforbruk ved behandling av "typiske" pasienter fra de ulike DRG-ene. Noen pasienter har av ulike årsaker særlig lange opphold ved sykehuset og kan betraktes som "atypiske" for den aktuelle DRG-en. Noen årsaker kan være ventetid før utskrivning eller overføring til andre omsorgsenheter. Disse pasientene vil være ferdigbehandlet og dermed mindre ressurskrevende enn de øvrige pasientene i DRG-en den siste tiden av oppholdet. Problemet består i å bestemme overgangen fra typisk til atypisk pasient. Grenseverdien kalles her et øvre trimpunkt hvor pasientene med liggetid over trimpunktet karakteriseres som outliere. Ressursbruk knyttet til liggetiden over det øvre trimpunktet varierer på grunnlag av andre faktorer (f.eks. alders-og sykehjemsplasser) enn ressursbruk til vanlig pasientbehamdling og bør synliggjøres bedre enn det hittil har vært gjort.

For å kunne identifisere outlierene; bestemme det øvre trimpunktet for hver DRG, er det nødvendig å foreta en statistisk modellering av liggetidene i de forskjellige DRG-ene. Variasjonen i liggetid for de typiske pasientene vil i hovedsak avhenge av pasientindividuelle trekk.

Behov for et nedre trimpunkt; identifisering av outliere i venstre hale, er tilstede i DRG-er hvor kort liggetid er urimelig i forhold til tradisjonell behandling. Årsaker kan være at behandlingen er blitt forbedret med moderne teknologi eller at pasienten er på sykehuset for etterkontroll for en alvorlig sykdom og blir registrert med de gamle medisinske

opplysningene.

Målet med dette vil blant annet være å forbedre DRG-systemets beskrivelse av pasientsammensetningen slik at det blir et bedre verktøy for planlegging og styring av ressursene ved norske sykehus. Liggetidsfordelingene varierer fra DRG til DRG, noe de tradisjonelle metodene ikke tar hensyn til. For å ta hensyn til denne variasjonen er det nødvendig å utvikle bedre metoder for identifisering av outliere med hensyn på liggetid for pasienter i de forskjellige DRG'ene.

Modellering av "typiske" pasienter ved Gamma, Weibull eller Lognormale fordelinger og anvendelse av likelihood-baserte prediksjonsmetoder (Bjørnstad, 1990), hvor identifisering av outliere her betraktes som et prediksjonsproblem, vil være aktuelt ved identifisering av outliere i disse typer modeller. Datagrunnlaget er sykehusinnleggelser i Norge; 600 000 pr. år,

Variasjonsforskning - analyser av variasjon i operasjonsaktivitet

Et eksempel på en operasjon som ofte er fokusert i variasjonsforskning er hysterektomi eller fjerning av livmor. Hysterektomi er et inngrep hvor det er knyttet endel faglig uenighet til indikasjonene for operasjonen. Litteraturen viser at det er internasjonalt sett store variasjoner i forekomsten av dette inngrepet. Stor variasjon mellom land, også land som ellers er relativt like med hensyn på helstjenestetilbud og sykelighet i befolkningen, har ført til stor diskusjon og mange forsøk på å finne forklaringer på dette.

Innen variasjonsforskning på operasjonsaktivitet ønsker man ofte å avdekke hvilke regioner som ligger høyt eller lavt i forhold til landsgjennomsnittet. Dette kan avdekkes ved modellering, f.eks. ved en multiplikativ Poisson modell. Antall tilfeller av inngrepet; X_{ij} følger en Poisson fordeling slik at forventet antall tilfeller i region i og aldersgruppe j er $E(X_{ij})=p_{ij}\ N_{ij}$. Sannsynligheten for å få utført hysterektomi for kvinner i aldersgruppe j, bosatt i region i, skrives på formen $p_{ij}=r_i\ a_j$ hvor r_i er den relative forekomsten i region i og a_j er sannsynligheten for hysterektomi for kvinner i aldersgruppe j. Parametrene estimeres ved sannsynlighetsmaksimeringsprinsippet og de relative forekomstene forteller hvor aktiviteten er høy eller lav. Dette er et eksempel på enkel modellering av operasjonsdata for å analysere regionale variasjoner med hensyn på operasjonsaktivitet. Datagrunnlaget her er basert på registerdata. Når man ønsker å se på faktorer som ikke inngår i databasen, for eksempel opplysninger vedrørende organisatoriske forhold ved operasjonsavdelingen, antall kirurger/gynekologer m.m., må dette samles inn via en spørreundersøkelse ved et utvalg operasjonsavdelinger.

Behandling av eventuelt høyt frafall i spørreundersøkelsen og utvalgsdata i kombinasjon med registerdata, er nye metodemessige utfordringer man må forholde seg til ved videre analyser av denne problemstillingen.

Trygve Almøy, Boks 5035 ,1432 Ås-NLH.

Sammenligning av forskjellige prediksjonsmetoder når bare noen få komponenter er relevante; et simuleringsforsøk.

(En kortversjon av innlegg på 14. Nordiske Konferanse i Matematisk Statistikk, Røros, 15.-19. juni 1992).

1: Generelt.

Anta vi har en stokastisk vektor $\mathbf{z}$ slik at: $\mathbf{z} = \begin{pmatrix}\mathbf{x}\\ y\end{pmatrix} \sim N_{p+1}(\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix}, \Sigma)$, der $\Sigma = \begin{pmatrix}\Sigma_{xx} & \sigma_{xy}\\ \sigma_{xy}' & \sigma_{yy}\end{pmatrix}$. Da er det velkjent at $y|\mathbf{x} \sim N(\mu_y + \beta'(\mathbf{x} - \mu_x), \sigma^2)$, $\beta = \Sigma_{xx}^{-1}\sigma_{xy}$, og $\sigma^2 = \sigma_{yy} - \sigma_{xy}'\Sigma_{xx}^{-1}\sigma_{xy}$. Ved n uavhengige observasjoner av $\mathbf{z}$, som gir $(\mathbf{x}_1', y_1)'$, $(\mathbf{x}_2', y_2)'$, . . . $(\mathbf{x}_n', y_n)'$, kan vi dermed anta modellen:

$$\underset{nx1}{\mathbf{y}} = \underset{nxp}{\mathbf{X}}\,\underset{px1}{\beta} + \underset{nx1}{\mathbf{e}}, \qquad (1)$$

der $\mathbf{y}$ og kolonnene i $\mathbf{X}$ er sentrerte om sine gjennomsnitt ($\bar{y}$ og $\bar{\mathbf{x}}$), og $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 I)$. Datamatrisen $[\mathbf{X},\mathbf{y}]$ (n x (p+1)) utgjør nå vårt såkalte kalibreringssett, hvorfra vi estimerer β med $\hat{\beta}$, noe som kan gjøres ved hjelp av mange forskjellige metoder. Dermed kan vi konstruere en lineære prediktor $\hat{y}$ for y, gitt en framtidig verdi $\mathbf{x}_0$ for $\mathbf{x}$ slik:

$$\hat{y} = \bar{y} + \hat{\beta}'(\mathbf{x}_0 - \bar{\mathbf{x}}), \qquad (2)$$

Prediksjonsevne evalueres etter forventet kvadrat tap, eller $v = E(\hat{y} - y)^2$. Dersom vi tar forventning kun over y, har vi bare informasjon om prediksjonsevne for en bestemt verdi $\mathbf{x}_0$. Dersom vi antar at den nye verdien $\mathbf{x}_0$ kommer fra samme fordeling som $\mathbf{x}_i$ ene i kalibreringssettet, er det naturlig å ta forventning simultant både over y og $\mathbf{x}_0$, og for enkelthets skyld også over $\bar{y}$ og $\bar{\mathbf{x}}$. Dermed har vi:

$$v = \frac{n+1}{n}(\sigma^2 + (\hat{\beta} - \beta)'\Sigma_{xx}(\hat{\beta} - \beta)). \qquad (3)$$

Gitt et bestemt kalibreringssett (f. eks. ved simulering av n uavhengige observasjoner av $\mathbf{z}$), er det en smal sak å finne v ved hjelp av (3), men da gjenstår problemet at v ikke er et generelt mål for prediksjon, men gjelder kun for dette bestemte kalibreringssettet. For å bøte på dette, tar vi forventningen til v over kalibreringssettet, eller:

$$g = Ev = \frac{n+1}{n}(\sigma^2 + E(\hat{\beta} - \beta)'\Sigma_{xx}(\hat{\beta} - \beta)). \qquad (4)$$

Fordi mange prediksjonsmetoder er svært kompliserte, er g ukjent eller bare kjent asymptotisk. Unntaket er dersom $\hat{\beta}$ er den vanlige minste kvadraters metode estimatoren ($\hat{\beta}_{mkm} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$), da gjelder:

$$g_{mkm} = \frac{n+1}{n}\,\frac{n-2}{n-p-2}\,\sigma^2, \qquad p < n-2. \qquad (5)$$

Resultatet i (5) er en følge av at $\hat{\beta}_{mkm}$ er forventningsrett, og dermed vil $E(\hat{\beta} - \beta)'\Sigma_{xx}(\hat{\beta} - \beta) = \text{trase}(\Sigma_{xx} E(\mathbf{X'X})^{-1})$, der $\mathbf{X'X}$ er Wishard fordelt $(n-1, \Sigma_{xx})$. Dermed vil $E(\mathbf{X'X})^{-1} = \Sigma_{xx}^{-1}/(n - p - 2)$. Dette er en mye brukt prediksjonsmetode, og i mange tilfeller fungerer den bra, men dersom forholdet p/n går mot 1 kan den gi særdeles dårlig prediksjon.

2: Modell-restriksjoner.

La oss anta følgende:

$$\Sigma_{xx} = \sum_{i=1}^{p} \lambda_i \mathbf{d}_i \mathbf{d}_i'$$

$$\sigma_{xy} = \sum_{i=1}^{m} \alpha_i \mathbf{d}_i \qquad m < p. \qquad (6)$$

Her er λ_i egenverdier med tilhørende egenvektorer $\mathbf{d}_i$ til Σ_{xx}. Dette er hva Helland og Næs (1992) kaller en modell med m relevante komponenter, og de viser samtidig hvordan mange prediksjons metoder kan tilpasses denne modellen. Antagelsen (6) medfører at $\mathbf{d}_i'\sigma_{xy} = 0$, og dermed $\mathbf{d}_i'\beta = 0$, for $i > m$. Merk at egenverdiene ikke nødvendigvis er ordnet etter størrelse.

3: De forskjellige prediksjonsmetoder.

De fleste metodene vi studerer her, kan tilpasses modellen i (6). Metodene er:

- Prinsipalkomponent regresjon etter største egenverdi (PCR1),
- Prinsipalkomponent regresjon etter største korrelasjon med y (PCR2),
- "Restricted" prinsipalkomponent regresjon (RPCR),
- "Partial least squares" (PLS),
- "Modifisert maximum likelihood" (MML).

RPCR er foreslått av Næs og Helland (1992), og MML er utviklet av Helland (1992).

Anta at vi reduserer dimensjonen av $\mathbf{X}$ ved $\mathbf{U} = \mathbf{XR}$, der $\mathbf{R}$ er en p x q matrise med rang q. Deretter finner vi $\hat{\beta}_U$ ved $\hat{\beta}_U = (\mathbf{U'U})^{-1}\mathbf{U'y}$, eller transformert tilbake til de opprinnelige data ved:

$$\hat{\beta} = \mathbf{R}(\mathbf{R'X'XR})^{-1}\mathbf{R'X'y}. \qquad (7)$$

Det er altså valget av $\mathbf{R}$ (og dermed også av q) som skiller metodene. For de nevnte metoder er den i-te kolonnen i $\mathbf{R}$ [$\mathbf{S}_{xx} = \mathbf{X'X}/(n-1)$, $\mathbf{s}_{xy} = \mathbf{X'y}/(n\ 1)$]:

PCR1: $\mathbf{r}_i = \hat{\mathbf{d}}_i$ (egenvektor til $\mathbf{S}_{xx}$ ordnet etter størrelse av tilhørende egenverdi, $\hat{\lambda}_i$).

PCR2: $\mathbf{r}_i = \hat{\mathbf{d}}_i$ (egenvektor til $\mathbf{S}_{xx}$ ordnet etter størrelse av $\hat{\mathbf{d}}_i'\mathbf{s}_{xy}/\sqrt{\hat{\lambda}_i}$).

PLS: $\mathbf{r}_i = (\mathbf{S}_{xx})^{i-1}\mathbf{s}_{xy}$.

RPCR: $\mathbf{r}_i = \mathbf{s}_{xy}$, $i = 1$, $\mathbf{r}_i = (\mathbf{S}_{xx} - \rho_i \mathbf{I})\mathbf{s}_{xy}$, $1 < i \leq q$, og der ρ_i er i-te egenverdi til $(\mathbf{I} - \mathbf{P})\mathbf{S}_{xx}(\mathbf{I} - \mathbf{P})$, og $\mathbf{P} = \mathbf{s}_{xy}\mathbf{s}_{xy}'/\mathbf{s}_{xy}'\mathbf{s}_{xy}$.

MML: Vi vil ikke beskrive metoden her, dette er en metode som finner maximum likelihood estimator for β under modell antagelsen (6).

Alle metodene bortsett fra PCR1 er konsistente under (6), og for disse har vi funnet g gitt ved (4) asymptotisk;

$$g = \frac{n+1+q}{n}\sigma^2 + \sum_{j=q+1}^{r}\lambda_j^2 f(\lambda_j) + o(n^{-1}), \qquad (8)$$

der $f(\lambda_j)$ er forskjellig for hver metode, se Helland og Almøy (1992).

4: Forsøksdesign.

Jeg har utført et større simuleringsforsøk, der v er beregnet ved hjelp av et kalibreringssett hvor radene er trukket fra fordeling $N_{p+1}(\mathbf{0}, \Sigma)$, Deretter er g estimert som gjennomsnittet av 100 slike verdier ($\bar{v}$). Skulle en simulere over hele parameterrommet til Σ, ville dette være uoverkommelig p.g.a. dets høye dimensjon, enten vi antar full modell eller restriksjonene i (6). Dimensjonen kan imidlertid reduseres ved en rotasjon av Σ_{xx}, slik at denne blir diagonal, og deretter skalering av **x** slik at den største relevante egenverdien blir 1. Vi kan selvsagt også skalere y slik at var(y) = 1. Det eneste vi antar er gitt er at m = 2 (to relevante komponenter). Vi har Σ_{xx} = diagonal(1, $\lambda_2, \lambda_3, \ldots, \lambda_p$), $\sigma_{xy} = (\alpha_1, \alpha_2, 0, \ldots .0)'$, $\sigma_{yy} = 1$. De relevante egenverdiene er ordnet slik at $1 > \lambda_2$, og de irrelevante egenverdiene er tilpasset en geometrisk rekke slik at $\lambda_j = \lambda_2 \delta \pi^{j-3}$, $\delta > 0$, $0 < \pi < 1$, $j \geq 3$. Dessuten er σ_{xy} gitt i polar-koordinater slik at $(\alpha_1)^2 = \Re^2\cos^2(\theta)$, $(\alpha_2)^2 = \Re^2\sin^2(\theta)\lambda_2$, $\Re^2$er multippel korrelasjonskoeffisient. Selv om vi også anser n og p som parametere, har vi redusert problemet til en representasjon av et 7-dimensjonalt rom bestående av elementene (n, p, π, δ, λ_2, $\Re^2$, θ). Dette har jeg valgt å legge opp som et faktorielt forsøk med 2 nivåer, og dermed står vi med 2^7 = 128 punkter i parameterrommet. De utvalgte nivåer er slik:

Nivå	n	p	π	δ	λ_2	$\Re^2$	θ
Lav	15	5	0.5	0.1	0.4	0.2	30°
Høy	60	10	0.9	2.0	0.9	0.9	75°

For å få et estimat for variasjonen i $\bar{v}$, er det gjort 3 gjenntak over alle parameterverdier. Fordi en lett kan drukne i for mange resultater, er metodene vurdert ut fra det antall komponenter (q) som gir lavest gjennomsnittsverdi av $\bar{v}$over disse 3 gjenntakene. Signifikant forskjell er beregnet etter Tuckey's test. Alle resultatene er skalert med verdien av σ^2. For fler detaljer, resultater og begrunnelser, se Almøy, 1992.

5: Konklusjon

Det er ikke mulig på grunnlag verken av asymptotiske beregninger eller denne simuleringsundersøkelsen å påstå at en metode er uniformt bedre enn en annen. Av de 128 parameter-punktene i denne undersøkelsen kom PCR2 og MML (for lav verdi av n) absolutt dårligst ut. PLS var bedre enn PCR1 i 70 tilfeller, derav 19 signifikant bedre, mens PCR1 var bedre enn PLS i 51 tilfeller herav 14 signifikant bedre. Det ser ut til at ved lav verdi av δ (de irrelevante egenverdiene er små), er PCR1 noe bedre enn PLS, mens det for høy verdi av δ er omvendt. Dette stemmer bra overens med resultatene i Helland og Almøy. Det

som kanskje er et større poeng, er å se på likheten mellom de 3 metodene PCR1, PLS og RPCR. Dersom δ er lav, gjør alle disse det bra, og forskjellen mellom dem er svært liten, slik at i dette tilfelle er det mye å tjene på å bruke en av disse (sammenlignet med vanlig multippel regresjon), mens det ikke er vesentlig hvilken av dem en bruker. Større blir imidlertid forskjellen når det er vanskelig å utføre god prediksjon (få observasjoner, mange variable og/eller store irrelevante egenverdier). I slike tilfeller kan det se ut som om PLS er noe bedre enn de andre.

I tabellen nedenfor presenteres bl.a. gjennomsnittet av minste prediksjonsfeil over alle parameterpunkter, og gjennomsnittlig antall komponenter dette er oppnådd med. Vi ser at PLS er den metoden som kommer best ut, og som finner minimum ved færrest antall komponenter. Ser vi derimot på hovedeffektene, ser vi at ved store n, er det svært liten forskjell, det samme gjelder for lave verdier av δ og p, men også for lav verdi av $\Re^2$. Verdiene av λ_2 og θ ser ikke ut til å spille noen vesentlig rolle for noen av metodene, derfor velger vi ta effektene av disse ut av tabellen, og å bruke gjennomsnittet ved de fire forskjellige verdiene av disse. Vi står derfor tilbake med 32 punkter som for alle metoder kan ordnes på samme måte, etter hvor god prediksjon en oppnår i punktet. Et ordnet punkt vil da se slik ut; (n, δ, $\Re^2$, p, π), der det "beste" punktet har lave verdier for alle parametere bortsett fra n. Nedenfor er det to plot av prediksjonsfeilen for PCR1, PLS og RPCR m.h.p. denne ordningen, det ene med de 16 beste (høy n), det andre med de 16 dårligste. Det mest iøynefallende er det metodene har felles, og ikke forskjellen dem imellom.

Litteratur:

-Almøy, T. 1992: A simulation study on comparison of prediction methods when only a few components are relevant", upublisert.

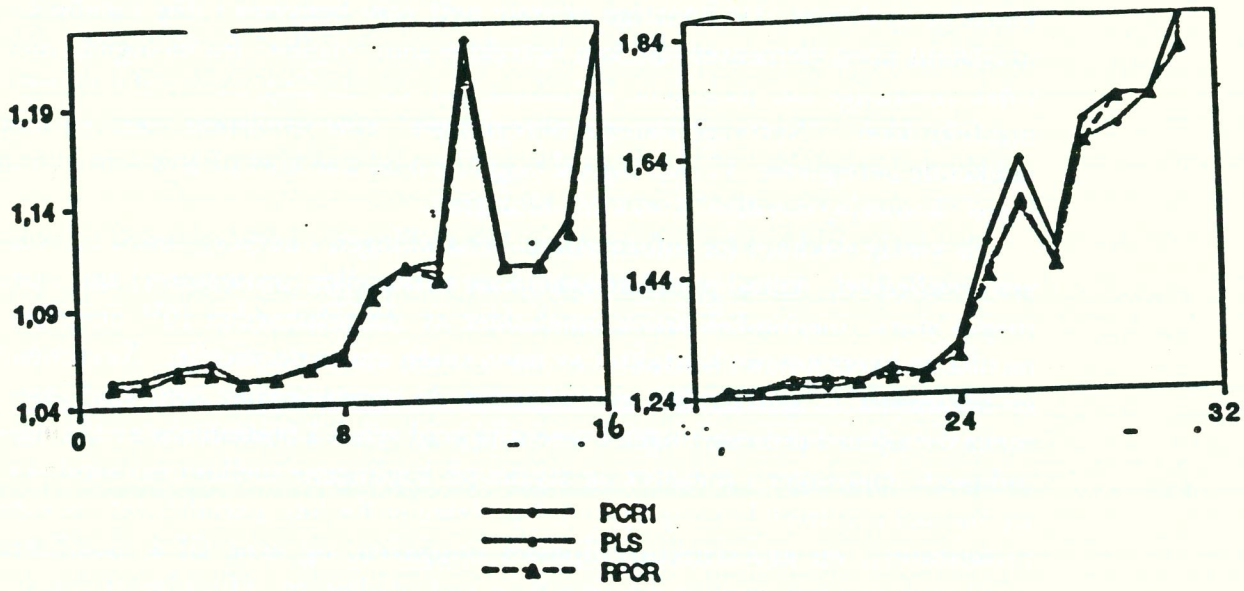
-Helland, I. S. og Almøy, T. 1992: "Comparison of prediction methods when only a few components are relevant", manuskript sendt JASA.

-Næs, T og Helland, I. S. 1992. "Relevant components in regression", antatt i Scand.J.Statist.

-Helland, I. S. 1992 "Maximum likelihood regression on relevant components", J.R. Statist. Soc. B, 54.

	PCR1		PCR2		PLS		RPCR		MML	
	p	k	p	k	p	k	p	k	p	k
Totalt gj.snitt	1.346	3.61	1.471	3.22	1.318	2.51	1.332	3.34	1.651	2.05
n = 15	1.597	3.26	1.831	2.54	1.540	2.17	1.569	2.83	2.201	1.73
n = 60	1.096	3.95	1.112	3.90	1.096	2.85	1.094	3.85	1.100	2.37
p = 5	1.266	3.22	1.315	3.05	1.255	2.58	1.256	2.99	1.322	2.03
p = 10	1.426	3.99	1.627	3.39	1.381	2.44	1.407	3.69	1.980	2.07
π = 0.5	1.266	3.21	1.412	2.63	1.257	2.27	1.256	2.96	1.595	1.56
π = 0.9	1.426	4.01	1.531	3.81	1.379	2.76	1.407	3.72	1.707	2.55
∂ = 0.1	1.161	2.02	1.287	1.73	1.165	1.76	1.160	1.91	1.532	1.24
∂ = 2.0	1.531	5.20	1.656	4.71	1.471	3.27	1.504	4.78	1.770	2.86
$\Re^2$ = 0.4	1.253	3.08	1.401	2.57	1.243	1.84	1.240	2.65	1.813	1.98
$\Re^2$ = 0.9	1.439	4.13	1.541	3.87	1.393	3.18	1.424	4.04	1.489	2.13

Tabellen viser foruten totalt gjennomsnitt, hovedeffekter for de fem viktigste faktorer, p er prediksjonsfeil og k er antallet komponenter.



NEVRALE NETT OG STATISTIKK

Knut Liestøl
Institutt for informatikk, UiO

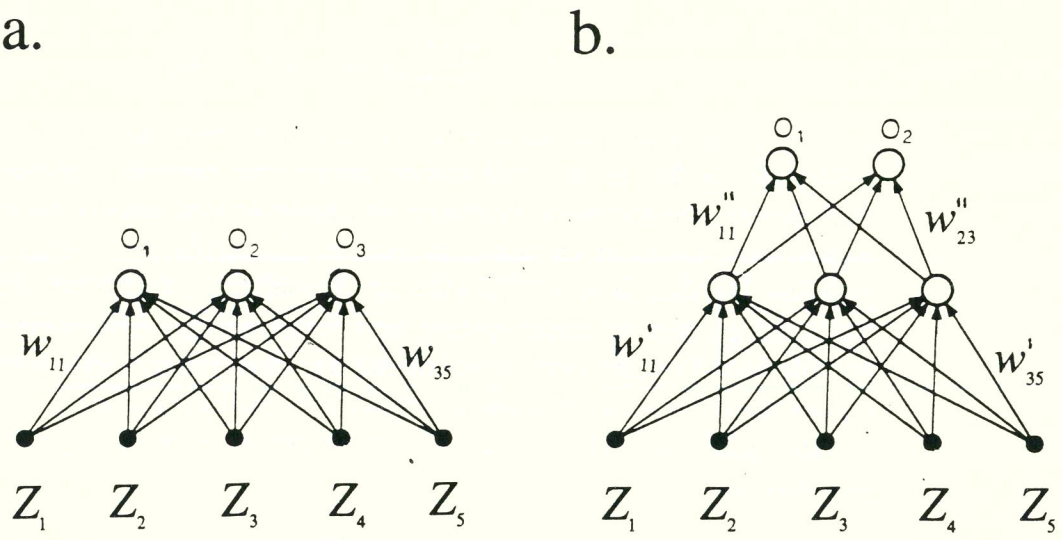
Forskningsaktiviteten rundt såkalte syntetiske nevrale nett og "neural computing" har vokst hurtig de siste årene. Fagfeltet har imidlertid en historie tilbake til slutten av 50-tallet. Hovedintensjonen til "pionerene" var at en gjennom studier av naturlige nerve-nettverk skulle forsøke å avsløre hva som gjør disse systemene spesielt effektive til visse oppgaver, spesielt signalbehandling og mønstergjenkjenning, for så å overføre disse prinsippene til kunstige systemer. De kunstige nevrale nett som benyttes i dag inneholder utvilsomt noen elementer som kan betraktes som "stjålet" fra biologien, men i det vesentlige bør metodene betraktes som en samling numeriske beregningsteknikker. "Nettverksbaserte beregninger" ville utvilsomt vært en mer dekkende betegnelse. Vi skal likevel begynne med å se kort på hva som kunne være "kopieringsverdige" trekk fra biologien.

Et viktig stikkord for informasjons-behandlingen i nervesystemer er *massiv parallellitet*. Antall parallelt-arbeidene nerveceller (prosessorer) kan være meget stort: menneskehjernen inneholder av størrelsesorden 10^{11} nevroner, og ofte er hvert nevron kontaktet av noen tusen andre nerveceller. Nevronene er organisert i "grovstrukturer" tilpasset de oppgavetyper som skal løses, mens detaljene i prosesseringen i vesentlig grad synes å bestemmes av læringsinduserte endringer i mønster og styrke på koplingene mellom nervecellene. Et biologi-inspirert kunstig system bør følgelig ha høy parallellitet (mange prosessorer), en grovstruktur tilpasset oppgaven, og evne til å modifisere koplingsmønsteret mellom prosessorene ved opplæring. Oppblomstringen i forskningen omkring nevrale nett skyldes i høy grad at en har sett muligheter for effektive hardware-implementasjoner av massivt parallelle nett, og at en for visse enkelt strukturerte nett har funnet anvendelige læringsregler basert på modifikasjon av koplingsstyrken mellom prosessorer.

Vi skal så se på to typer av kunstige nevrale nett. I statistikk-terminologi er den første typen beregnet på regresjon eller styrt klassifikasjon, og den andre en form for ikke-styrt klassifikasjon.

Vi skal først betrakte *lagdelte feed-forward nettverk*, også kalt perseptroner. Slike nett representerer funksjoner, typisk regresjonsfunksjoner: forklaringsvariable gis som input, mens nettet som output gir predikerte verdier for de avhengige variable. Oppførselen til disse nettene kan beskrives ved hjelp av vanlig regresjons-terminologi, men jeg velger her å bruke deler av den terminologien som vanligvis anvendes for nevrale nett. Nettverkene beskrives da som fysiske systemer av *noder* og *forbindelser*.

Nettverket i Fig. 1a består av et lag av inn-noder og et lag av ut-noder,



Figur 1: Ett-lags (a) og to-lags (b) feedforward nevrale nett.

med alle-til-alle forbindelser mellom lagene. Forklaringsvariablene (Z_j) kodes inn i inn-nodene. Under overføringen til ut-nodene multipliseres verdiene med parametre betegnet *vekter* (w_{kj} for forbindelsen fra inn-node j til ut-node k). Disse produktene blir så summert i ut-nodene og en konstant ("bias") addert slik at vi får et signal $v_k = \sum_j w_{kj} Z_j + w_{k0}$ i node k. Ut-verdien framkommer ved at v_k blir transformert av en *aktiverings-funksjon*. Hvis ut-nodene bruker en felles aktiverings-funksjon g vil dermed ut-verdien i node k bli

$$O_k^1(\underline{Z};\underline{w}) = g(\sum_j w_{kj} Z_j + w_{k0}). \tag{1}$$

Denne typen nett kalles vanligvis, men ikke helt konsekvent, for ett-lags nett (det "passive" inn-laget telles ikke).

Fig. 1b viser et to-lags nett. Mellom inn- og ut-noder er det innskutt såkalte *skjulte* noder. Hvis vekten mellom inn-node j og skjult node l kalles w'_{lj} og de skjulte nodene bruker en felles aktiverings-funksjon h, vil ut-verdien i skjult node l bli $h(v_l) = h(\sum_j w'_{lj} Z_j + w'_{l0})$. Under overføring til ut-nodene blir disse verdiene multiplisert med vekter w''_{kl}. Med en felles aktiverings-funksjon g i ut-nodene vil nettet implementere funksjonen

$$O_k^2(\underline{Z};\underline{w}) = g\{\sum_l w''_{kl} h(\sum_j w'_{lj} Z_j + w'_{l0}) + w''_{k0}\}. \tag{2}$$

For å lære opp nettet, dvs. estimere parameterene $\underline{w}$, defineres en kost-funksjon E. Mest brukt er

$$E^1 = \sum_i \sum_k \{O_k(\underline{z}_i; \underline{w}) - Y_{ki}\}^2 \tag{3}$$

hvor $\underline{z}_i$ er vektoren av forklaringsvariable for observasjon nr. i i vårt datamateriale og Y_{ki} er den tilhørende observerte respons i utnode k. Kostfunksjonen minimeres så, vanligvis ved hjelp av den såkalte *back-propagation* algoritmen. Dette er en algoritme som er gjenoppfunnet mange ganger, men populær ble den først i 1986 etter at Rumelhart (se litteraturliste sist i artikkelen) ga den en enkel nevral nett formulering. Algoritmen kan brukes for alle feed-forward nett hvor nodene har deriverbare aktivitets-funksjoner, og dessuten for en stor klasse av rekurrente (tilbakekoplede) nett. Det er en iterativ gradient-metode, dvs. oppdateringen av en vekt w_{kj} i en iterasjon er av formen

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}}, \quad \eta > 0. \tag{4}$$

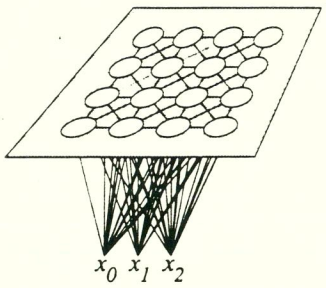
Algoritmen er beskrevet i flere lærebøker (f. eks. Hertz, Krogh og Palmer) og kjennskap til den er ikke nødvendig for å forstå grunnprinsippene bak nevrale nett. Siden algoritmen er så mye omtalt, gjengis likevel hovedtrekkene i et appendix. Nå skal vi imidlertid se hva disse nettene "egentlig" gjør.

Hvis vi velger ett-lagsnettet fra Fig. 1a, nøyer oss med en enkelt utnode, lar aktiveringsfunksjonen g være identitetsfunksjonen og velger kostfunksjonen E^1, skulle det være enkelt å se at vi utfører vanlig multippel lineær regresjon. Velger vi i stedet en logistisk aktiveringsfunksjon g og lar kost-funksjonen være

$$E^2 = -\sum_i \sum_k \{Y_{ki} \log O_k(\underline{z}_i; \underline{w}) + (1 - Y_{ki}) \log(1 - O_k(\underline{z}_i; \underline{w}))\}, \tag{5}$$

der responsen Y_{ki} er binær, så utfører vi maximum likelihood-estimering i en logistisk regresjons-modell. Avstanden til vanlig statistikk er m.a.o. ikke stor (jeg har imidlertid ennå til gode å se en lærebok i nevrale nett som klart og enkelt påpeker dette).

Med to-lags nett får vi ulike former for ikke-lineær regresjon. Hvilken modell vi da benytter avhenger naturligvis av valg av aktiveringsfunksjoner og antall skjulte noder, samt av eventuelle modifikasjoner av strukturen til nettet. Vi kan imidlertid merke oss en viktig forskjell fra de regresjonsmodeller statistikere normalt bruker. I sistnevnte spesifiserer en typisk at enkelte variable også skal inngå som et kvadratisk ledd, at det skal være interaksjon mellom to spesifiserte variable etc. I et nevralt nett som det i Fig. 1b inngår *et antall ulineære ledd*, men det er ikke foretatt noe utvalg av hvilke variable som skal inngå. Det "gode" med dette ligger i at nettene skulle ha mulighet til å finne viktige (ulineære) kombinasjoner av variable. Men det er også klare ulemper: parametrene er vanskelige å tolke,



Figur 2: Skisse av strukturen til et selvorganiserende nett med tre inn-noder.

og det blir lett mange parametre (vekter). Disse forhold har umiddelbare konsekvenser for hva som er hensiktsmessige bruksområder for modellene. Inferens vedrørende forklarings-variablene blir vanskeligere, både på grunn av tolknings-problemene, og fordi jeg vil tro de asymptotiske resultatene blir mindre troverdige (en del slike resultater kan "overføres" fra ikke-lineær regresjon, se White, 1989). Det som derfor er mer aktuelle anvendelsesområder er eksplorerende analyse og prediksjon/klassifikasjon. I begge disse tilfellene vil det være hensiktsmessig med en fleksibel klasse av modeller. Eksplorerende analyse er videre aktuelt fordi det er lett å programmere de tilpassende modellene opp mot grafiske verktøy, og dermed har en f. eks. mulighet til å bruke 3-D plot til å se på interaksjons-effekter fra forklaringsvariable. Kryss-validering vil være en nærliggende metode for å vurdere om modellene er adekvate.

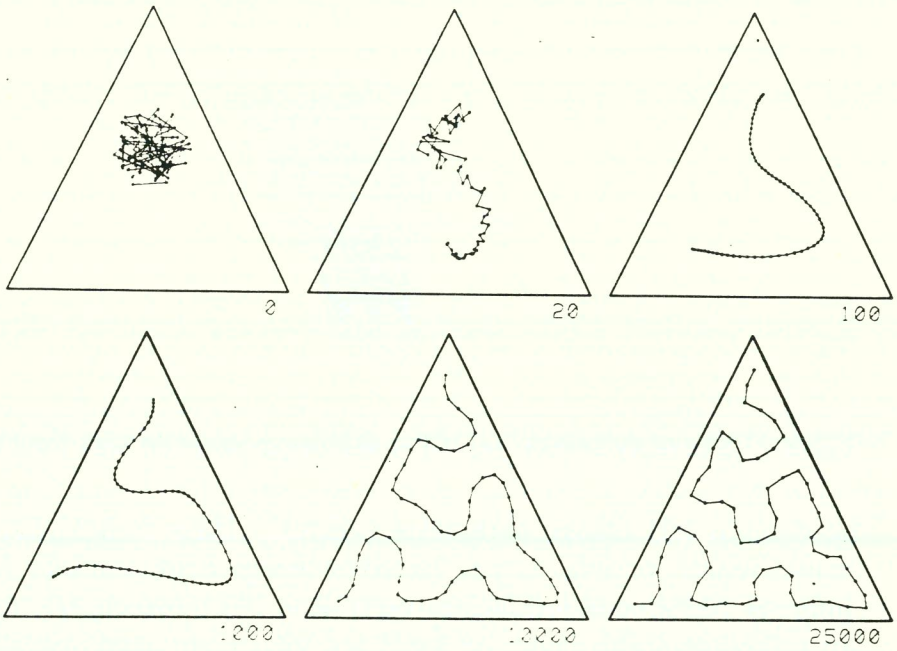
Feedforward-nett kan følgelig betraktes som systemer for prediksjon/ikke-lineær regresjon. Det finnes imidlertid en rekke andre typer kunstige nevrale nett. Noen av disse er laget for det en løselig kunne kalle ikke-styrt læring. F. eks. kan en ved hjelp av rimelig enkle nett finne prinsipal-komponenter eller foreta "clustring". Beslektet, men litt mer originale, er nettene som foretar (direkte oversatt fra engelsk) selv-organisert egenskaps-avbildning. Disse nettverkene har typisk en samling inn-noder med alle-til-alle forbindelser til et lag av ut-noder (Fig. 2). Nodene i ut-laget er normalt organisert som en rekke eller et plan, og hver node har tverrforbindelser til et område av naboer i utlaget. Påvirkningen fra naboer, som styres av faste parametre (vekter), vil typisk være av meksikaner-hatt type: positiv til nære naboer, negativ til fjernere og 0 til fjerne noder. Hvis vi kaller verdiene i inn-nodene for $\{x_i\}$ vil verdien i ut-node nr. m være

$$O_m = \sum w_{mi} x_i + input\ fra\ naboer$$

der w_{mi} er modifiserbare vekter mellom inn og ut-laget. Disse vektene modifiseres vanligvis etter en såkalt "Hebb-regel" (oppkalt etter psykologen Donald Hebb som foreslo at en tilsvarende mekanisme virket i naturlige nervenettverk)

$$\Delta w_{ij} = \eta x_i O_j + normaliseringsledd.$$

Endringen avhenger følgelig av produktet av "aktiviteten" i de to noder som



Figur 3: Illustrasjon av tidsutvikling i et selvorganiserende nett. Tallene under trekantene viser antall iterasjoner, se ellers teksten.

forbindelsen knytter sammen.

Når slike nett trenes på et utvalg av input-data skjer to ting:

- hver ut-node vil reagere på input fra et sammenhengende delområde av input-rommet.
- nabo-noder i ut-laget reagerer på nabo-områder i input-rommet (avbildningen er "topografisk").

Metodene er typisk utviklet for høydimensjonale input-data. For illustrasjonsformål bruker vi imidlertid et lavdimensjonalt leke-problem. I Fig. 3 er det illustrert hvordan et ut-lag som består av en rekke av noder reagerer på to-dimensjonalt input (dvs. input-koordinatene kodes i to noder). Input-verdiene blir trukket uniformt fra en trekant. Hvert punkt i det indre av trekantene i Fig. 3 viser hvor i input-rommet en bestemt ut-node gir kraftigst respons. Punktene for noder som er naboer i rekka av utnoder er forbundet med en linje. Initielt er det tilfeldige vekter, og følgelig intet system i ut-nodenes respons. Etter hvert som flere input-data presenteres og vektene som følge av dette modifiseres, "spesialiseres" ut-nodene til å respondere på input i ulike deler av inputrommet på en slik måte at nabo-noder i ut-laget reagerer på input fra nabo-områder i input-rommet.

Kohonen har utnyttet selvorganisering i sitt system for automatisk talegjenkjenning (for finsk, som visstnok er av de enklere språk!). Det selvorganiserende nettet blir fortløpende presentert for Fourier-komponenter beregnet fra det samplede lydsignal. Etter opplæring vil spesielle grupper av utnoder reagere på bestemte lyder. Et ekspertsystem er så utviklet for å analysere respons-sekvensen i utnodene og dekode hva som blir sagt.

Relativt få systemer er i dag realisert i hardware, men det arbeides langs flere utviklingsakser. Prinsipielt interessant er analog VLSI. Dette er integrerte kretser av samme type som finnes i mikroprosessorer, men hvor transistorene ikke fungerer som av/på-brytere, men som analoge regulatorer (det gjør transistorer ved lave spenningsnivåer). Analoge VLSI-systemer kan utføre kompliserte operasjoner i små kretser, men fungerer bare moderat presist og blir lett ustabile. Komponentene i naturlige nervenettverk fungerer imidlertid heller ikke presist, og derfor mener mange at en ved å ta prosesserings-prinsipper fra nervesystemet kan få ideer til effektive analoge VLSI-kretser. Carver Mead, som var sentral i arbeidet med å utvikle de første mikroprosessorene, har utviklet flere slike analoge kretser, blant annet modellert etter retina og hørsels-systemet (se ref. nedenfor).

Grovt sett kan en si at nevrale nett i dag oppnår resultater omtrent tilsvarende de en får med mer etablerte teknikker (rapportene pleier å si "litt bedre", men de er jo stort sett skrevet av entusiaster). Dette burde ikke overraske, siden nettene utfører operasjoner som minner sterkt om de statistiske teknikkene. Siden en vet mindre om egenskapene til de estimåter en får ut av de nevrale nettene, kan en spørre hva hensikten med nevrale nett egentlig er. Svaret ligger dels i mulighetene til massivt parallelle og følgelig hurtige implementasjoner (som jo er viktig f. eks. i mange mønstergjenkjennings-oppgaver), og dels i en tro på at bedre strukturerte nett etterhvert vil utvikles, og at disse vil få vesentlig bedre egenskaper enn dagens "primitive" nett. Tiden vil vise om det er en korrekt tro.

En god del bøker som omhandler nevrale nett har kommet ut de siste 2-3 årene, noen dessverre preget av at de er skrevet i betydelig hast for komme tidlig i den nye bølge av interesse for fagområdet. Den mest etterrettelige boka jeg har sett til nå er

- Hertz, Krogh og Palmer: Introduction to the theory of neural computation. Addison-Wesley, 1991.

Boka er imidlertid skrevet av fysikere og er preget av det. Det anbefales derfor å supplere med et av de få statistiker-bidrag til nevral-nett litteraturen:

- White, H.: Learning in artificial neural networks: a statistical perspective. Neural Computation, **1**, 425-464, 1989.

Hvis en er interessert i mer utførlig kunnskap om selvorganisering kan en prøve seg på:

- Kohonen, T.: Self-organization and assosiative memory. Springer-Verlag. 2nd ed., 1988.

Noen av kapitlene er lettleste, andre er det ikke.
Hvis en vil utvide sin horisont med litt kunnskap om hardware, spesielt analog VLSI, er det enkelt å komme med en anbefaling (det er ikke så mange alternativer):

- Mead: Analog VLSI and neural systems, Addison-Wesley, 1989.

Alle deler er ikke lettleste for folk som oss, men stoffet er avgjort interessant. Den nye bølge av interesse for nevrale nett ble i høy grad "sparket" igang av

- Rumelhart and McClelland: Parallel distributed processing. 2 bind. MIT press, 1986.

Etter min oppfatning har disse bøkene nå mer "historisk" interesse. Og er en først historisk interessert, bør en snarere se på

- Anderson and Rosenfeld: Neurocomputing. Foundations of research, 1988, MIT Press

der en finner en rekke av de (historisk sett) viktigste originalartiklene. Boken har vel 700 store sider. Lykke til!

Backpropagation: Som nevnt er backpropagation en gradientmetode, dvs. $\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}}$. Siden $E = \sum_i E_i$, hvor E_i er kostnaden for individ nr. i, er det vanlig å bruke en modifikasjon der oppdateringen av vektene foretas etter hvert individ, dvs.

$$\Delta w_{kj} = -\eta \frac{\partial E_i}{\partial w_{kj}}. \tag{6}$$

Vi skal først se hvordan denne varianten virker for ett-lagsnettet i Fig 1a. Vi merker oss at

$$\frac{\partial E_i}{\partial w_{kj}} = \frac{\partial E_i}{\partial v_k}\frac{\partial v_k}{\partial w_{kj}} = \frac{\partial E_i}{\partial v_k} z_{ji} \stackrel{def}{=} \delta_k z_{ji}, \tag{7}$$

der z_{ji} representerer forklarings-variablene for observasjon i. Videre har vi

$$\delta_k = \frac{\partial E_i}{\partial O_k^1}\frac{\partial O_k^1}{\partial v_k} = \frac{\partial E_i}{\partial O_k^1} g'(v_k). \tag{8}$$

Hvis vi videre velger kost-funksjonen E^1, slik at $\frac{\partial E_i^1}{\partial O_k^1}$ blir $2(O_k^1(\underline{z}_i;\underline{w}) - Y_{ki})$, og en logistisk aktiverings-funksjon $g(x) = (1+e^{-x})^{-1}$, slik at $g'(x) = g(x)(1-g(x))$, får vi

$$\delta_k = 2(O_k^1(\underline{z}_i;\underline{w}) - Y_{ki}) O_k^1(\underline{z}_i;\underline{w})(1 - O_k^1(\underline{z}_i;\underline{w})). \tag{9}$$

Kombinerer vi E^2 med samme aktiverings-funksjon, får vi tilsvarende

$$\delta_k = O_k^1(\underline{z}_i;\underline{w}) - Y_{ki}. \tag{10}$$

Siden O_k^1 er output-verdien fra node k er disse δ'ene lette å beregne, og dermed ser vi av (6) og (7) at hele algoritmen er enkel å implementere.

I flerlagsnett beregnes δ'ene på samme måte i ut-laget, og vektene w''_{kl} oppdateres ved

$$\Delta w''_{kl} = -\eta \delta_k H_l,$$

hvor H_l er output fra skjult node nr. l. For å oppdatere de øvrige parametre foretas en "bakover-propagering" av δ'ene. Ved å bruke kjerneregelen etc. kan en lett vise at for nettverket i Fig. 1b så gjelder

$$\frac{\partial E_i}{\partial w'_{lj}} = \{h'(v_l) \sum_k w''_{kl}\delta_k\} z_{ji},$$

og fra (6) får vi da oppdateringsregelen. Algoritmen er følgelig enkel å implementere, og en bør også merke seg at den er enkelt parallelliserbar: beregningene for alle vekter mellom to lag kan foretas samtidig.

En ulempe ved back-propagation er den alt annet enn imponerende konvergens-hastigheten. En rekke modifikasjoner er derfor lansert, og fornuftig bruk av disse er ofte avgjørende for å få konvergens i "endelig" tid. Uansett advares imidlertid personer med høyt blodtrykk og/eller generelt lav frustrasjons-terskel mot å kjøre backpropagation på sin lille hjemme-PC. Denne langsomme konvergens-hastigheten er naturlig nok en viktig drivkraft for utviklingen av hardware som effektivt utnytter mulighetene til parallellprosessering.

Som alle gradientbaserte metoder kan backpropagation kjøre seg fast i lokale minima. Mytologien, som er velutviklet omkring nevrale nett, sier at dette i praksis er et moderat problem.

Fra Aftenposten kveldsnummer 24.11.92 .

Vannhode mulig atomskade

Forskere har funnet en mulig sammenheng mellom radioaktivt Tsjernobyl-nedfall og antall nyfødte med vannhode i Norge.

BÅRD IDÅS
ARILD RYGNESTAD

I de delene av Norge som hadde mest radioaktivt nedfall, steg antall nyfødte med vannhode fra to til seks tilfeller i løpet av tre år. I områdene med mindre nedfall sank tvert om antallet tilfeller - fra 43 til 23.

Undersøkelsen har imidlertid ikke funnet noen sammenheng mellom andre typer fosterskader og Tsjernobyl-katastrofen i mai 1986. Direktør Ole Harbitz ved Statens institutt for strålehygiene sier undersøkelsen er grundig, men at det ikke er grunn til å slå alarm med bakgrunn i forskningsresultatene.

Vannhode, medisinsk betegnet hydrocephalus, er en sykelig økning av veske i hjernens hulheter og mellom hjernehinnene. Hos små barn fører dette til at hodets omfang vokser. Vannhode kan føre til alvorlige funksjonshemninger.

160 000 undersøkt

160 000 nyfødte i Norge er undersøkt i løpet av seks år av Medisinsk fødselsregister ved Universitet i Bergen og Statens institutt for strålehygiene. Forskerne sammenlignet antallet fosterskader på nyfødte i Norge i de tre årene før katastrofen, med de tre årene etter.

– Sammenhengen mellom radioaktivt nedfall og hydrocephalus er ikke påvist i andre studier vedrørende Tsjernobyl. Totalt har det vært en reduksjon i antallet vannhoder etter Tsjernobyl-ulykken, men det er påfallende at antallet tilfeller har økt i områdene med mye radioaktivt nedfall, sier forsker Rolv Terje Lie ved Medisinsk fødselsregister.

Stråling fra mat

I Norge ble det målt høyest doser med radioaktivitet i Trøndelag og øvre del av Gudbrandsdalen. I denne undersøkelsen har forskerne vurdert stråling fra både matvarer og radioaktivt nedfall.

Undersøkelsen ble nylig offentliggjort i det prestisjetunge forskningsmagasinet American Journal of Epidemiology. Forsker Rolv Terje Lie presenterte konklusjonene fra undersøkelsen på et seminar i regi av Norges allmennvitenskapelige forskningsråd i Oslo mandag.

– Vi er svært forsiktige med å trekke bastante konklusjoner av tallmaterialet. Problemer med en komplett melding av alle sykdomstilfeller og feilklassifiseringer av doser med radioaktivt nedfall gjør dette vanskelig, understreker Lie.

15 typer skader

Forskerne kartla forekomster av 15 forskjellige misdannelser på nyfødte i Norge. I annen europeisk medisinsk litteratur er det spekulert i om blant annet Downs syndrom, ryggmargsbrokk og lav fødselsvekt kan forklares med stråling fra Tsjernobyl.

– Vår undersøkelse gir ikke grunnlag for å tro at det finnes sammenheng mellom disse fosterskadene og Tsjernobyl-strålingen. Det synes vi er svært betryggende, sier forsker Rolv Terje Lie.

Bellona-leder Frederic Hauge er ikke fullt så betrygget av undersøkelsen.

– Resultatet av undersøkelsen er ikke spesielt sjokkerende, den viser at selv lavradioaktiv stråling kan få helsemessige konsekvenser. Spørsmålet er hvor mange barn som i årenes løp er født med vannhode som følge av atomprøvesprengningene på Novaja Semlja, sier Hauge.

Lite materiale

Direktør Ole Harbitz i Statens institutt for strålehygiene sier at det er vanskelig å ta stilling til om at stråling som følge av reaktorkatastrofen i Tsjernobyl er årsaken til økningen i antall nyfødte barn med vannhode.

– Arbeidet forskerne har gjort er meget grundig. Før vi kan ta stilling til om denne lidelsen er forårsaket av radioaktivt nedfall fra Tsjernobyl, må vi ha undersøkelser fra flere land. Først etter en tid kan man ta stilling til om det er noen sammenheng, slik den norske undersøkelsen antyder, sier Ole Harbitz. Han sier materialet er for lite til å kunne slå fast at det er en klar sammenheng mellom radioaktivt nedfall fra Tsjernobyl og økningen i antall nyfødte barn med vannhode.

Han understreker at selv om det er meget alvorlig for dem som blir rammet av denne typen fødselsskader, så er det ingen grunn til å slå alarm.

Nr. 21 – 20/11 – 1992

ISSN: 0802-5924

NATUR & MILJØ BULLETIN

Lakserømning kunne vært hindret side 3

Advarer mot kraftlinjer side 4

Venter nye atomuhell side 4

Bommet på hval side 5

Svensk nei til miljømerke side 7

– Atomtilsynet bagatelliserer

– Det er sjokkerende at Statens Atomtilsyn bagatelliserer uhellet ved Barsebäck atomkraftverk ved å overse hvor dårlige de svenske sikkerhetsanalysene viser seg å være, sier Bent Natvig, professor i pålitelighetsanalyse.

ANDERS GJESVIK

– Hver gang atomtilsynets leder Knut Gussgard uttaler seg om uhellet, graver han seg ned i tekniske detaljer, i stedet for å fokusere på svikten i hele sikkerhetsvurderingen. Han beroliger folk ved å pakke bort den usikkerhet som uhellet avslører. Et atomtilsyn bør i utgangspunktet være kritisk. Gussgard er ikke tillitsvekkende, og det er alvorlig, sier Natvig.

Natvig er professor i pålitelighetsanalyse ved Universitetet i Oslo. Hans fagfelt er vurdering av sikkerheten ved teknologiske systemer som atomkraftverk.

– Gussgard snakker som om dette er et hull som man kan tette igjen, så er alt bra. Det er ingen grunn til å bagatellisere dette uhell. Svikten i den svenske sikkerhetsanalysen er så graverende at den tyder på at hele nivået i svensk atomsikkerhetsanalyse er langt lavere enn man hittil har trodd, sier Natvig.

Sjefen for avdelingen for system- og pålitelighetsanalyse ved det svenske atomtilsynet (SKI), Lennart Carlsson, mener at uhellet ved Barsebäck har avslørt at risikoen for en kjernenedsmeltning er ti ganger større enn man har antatt i sikkerhetsanalysene.

Forsker Steinar Backe ved Statens institutt for strålehygiene sier at han er skuffet over at svenskene ikke har bedre sikkerhetsvurderinger.

– Uhellet avdekker at sikkerheten ikke er god nok, selv om et tilfelle ikke er nok til å endre prognosene for et alvorlig uhell ved et svensk atomkraftverk. Men vi må være villig til å la oss bevege av slike hendelser. Troen på sikkerheten ved vestlige kraftverk blir svekket, sier Backe.

Vant til kjeft

Gussgard tar Natvigs kritikk med fatning.

– Jeg får regelmessig kjeft, det ene øyeblikk for at jeg bagatelliserer, det neste øyeblikk fordi jeg overdriver farene. Jeg er ikke overrasket over at noe så banalt som svikten ved Barsebäck skjer, men det er alvorlig at svenske kolleger ikke gir tilstrekkelig rom for at det utenkte kan skje. Likevel er ikke min tillit til svenske sikkerhetsvurderinger svekket, sier Gussgard.

I dag skal det danske atomtilsynet ha et møte med sine svenske kolleger for å få nærmere opplysninger om uhellet ved Barsebäck. Gussgard skal ikke delta i dette møtet, eller i andre møter med SKI.

– Barsebäck ligger langt fra Norge. SKI gir oss god nok orientering. Vi har ikke behov for møter, sier Gussgard.

Atomkraftverket ved Ringhals har samme teknologi som førte til uhellet ved Barsebäck. Ringhals ligger bare 200 kilometer fra norskegrensa. Gussgard satt selv i et interdepartementalt utvalg som i vår la fram en rapport der det går fram at en total svikt i sikkerhets- og kjølesystemet ved Ringhals kan gi en forurensning over Østlandet som er ti ganger større enn den som ble forårsaket av Tsjernobyl.

– Vi har ikke tid til å delta i møter bare for å markere vår bekymring, sier Gussgard.

Kristensen enig

Olje- og og energiminister Finn Kristensen anser også norske initiativ som unødvendig. I Stortingets spørretime sa han at svenske myndigheter gjør det som skal til for at kjernekraftverkene fortsatt skal være blant de sikreste i verden, og at denne saken ikke gir grunnlag for noen norsk reaksjon nå.

Magnus Nilsson, som er utreder i atomkraftspørsmål i Svenska Naturskyddsföreningen, mener Kristensens innstilling er naiv.

– Sverige er ikke passiv i forhold til problemer ved litauiske atomkraftverk. Hvis noe hender her, får Norge det radioaktive nedfallet. Vi er uten tvil i en sensitiv politisk situasjon nå, der myndighetene er svært følsomme for utenlandske reaksjoner på svensk atomkraft, sier Nilssen.

Maner til varsomhet ved kraftledninger

Statens institutt for strålehygiene (SIS) har frårådet bygging av et nytt boligfelt i Lørenskog, fordi de mener det vil ligge for nært høyspentledning. SIS anbefaler 60 meters avstand, mens Statnett bruker 19 meter som minste avstand mellom kraftledning og bolig.

ANDERS GJESVIK

En svensk undersøkelse tyder på en klar sammenheng mellom magnetfelt fra høyspentledninger og blodkreft hos barn. Undersøkelsen viser at barn kan utvikle blodkreft (leukemi) ved svært svake magnetfelt. Undersøkelsen konkluderer med at allerede ved 60 meters avstand fra en vanlig kraftledning er magnetfeltet sterkt nok til at sjansen for at barn utvikler blodkreft mer enn tredobles. Lederen for det svenske elektrisitetstilsynet anslår at gjennomsnittlig 3,5 barn i Sverige årlig vil utvikle blodkreft som en følge av kraftlinjene.

Føre var

Den svenske undersøkelsen er oppsiktsvekkende klar. Tidligere undersøkelser viser samme tendens. Det er ikke påvist noen effekt for planter og dyr. I Norge er kreftregisteret i ferd med å gjennomføre en liknende undersøkelse, men resultatene vil ikke foreligge før om et år.

Sosialdepartementet har opprettet et embetsmannsutvalg som skal vurdere problemet. I mellomtiden anbefaler SIS en føre-var-strategi der man respekterer den svenske undersøkelsen.

– Vi anbefaler at man ikke bygger hus tett inntil kraftledninger, eller omvendt. Vi gå ut ifra at den svenske undersøkelsen er riktig, inntil vi har fått ny kunnskap som tilsier at den ikke er det, sier direktøren ved SIS, Ole Harbitz.

Maks 60 m

Derfor anbefaler SIS i samsvar med den svenske undersøkelsen en minsteavstand på 60 meter fra kraftlinjen til nærmeste bygning der mennesker skal oppholde seg. Det er ikke et pålegg, men et råd sett i lys av den store usikkerheten rundt problemet. Statnett som er ansvarlig for ledningsnettet opererer med en avstand på 19 meter. Forskrifter for elektriske anlegg krever bare 16 meters avstand.

SIS fraråder

Lørenskog kommune har henvendt seg til SIS i forbindelse med et nytt byggefelt som er planlagt med en avstand til kraftlinjen på 19 meter. SIS fraråder dette byggeprosjektet.

– Vi har ikke lovhjemmel til å hindre prosjektet, til det er usikkerheten for stor. Men den samme usikkerheten gjør at vi fraråder det, sier forsker Georg Thommesen ved SIS.

Lørenskog kommune har hittil praktisert 20 meters sikkerhetssone til ledningsnettet.

– Jeg etterlyser klarere retningslinjer fra myndighetene, men ser ikke noen grunn til at vi skal avvike praksisen denne gangen. Men det blir opp til politikerne å ta avgjørelsen. Vi venter en solid diskusjon i bygningsrådet, sier avdelingsingeniør Thorvald Sylte i Lørenskog kommune.

Takk for støtten!

Trondheim kommune
MILJØAVDELINGEN

Sørlaminering a·s
LAMINERTE TREKONSTRUKSJONER
4532 ØYSLEBØ · TLF 043 88 403
TELEFAX 043 88 479

Arendal Smelteverk
Vet du at vi lager silisiumkarbid som brukes til bl.a. slipepapir og slipeskiver?

HALLINGDAL RENOVASJON
3570 ÅL

RICOH
ore
MAGNE ORE A·S
Postboks 6217, Etterstad
0602 Oslo. Tlf. 02- 68 50 80

BORREGAARD
TILSLUTTET ORKLA A S

Venter nye svenske atomuhell

Svenske atomeksperter regner med at aldringen vil føre til flere uhell ved atomkraftverkene.

ANDERS GJESVIK

– Uhellet ved Barsebäck var sikkert bare den første av flere overraskende feil som kommer til å oppstå som en følge av alderskramper ved våre atomkraftverk, sier professor Jan-Olov Liljenzin til Dagens Nyheter.

Liljenzin bør vite hva han snakker om, han er leder for det svenske atomtilsynets (SKI) rådgivende nemd og Stråleverninstituttets forskningsnemd. Liljenzin sier at feilen som ble oppdaget ved Barsebäck var en ubehagelig overraskelse.

– Den skyldes at reaktorene er i ferd med å bli gamle, og at isoleringen eldes. Den burde kanskje ha blitt byttet tidligere, sier Liljenzin.

Overbelastning

Lijenzin mener at det ikke er mulig å legge ned de fem reaktorene som er berørt av uhellet.

– Da klarer vi ikke å dekke elektrisitetsbehovet når kulda kommer. Med fem reaktorer ute av drift blir de andre reaktorene overanstrengte, og det blir en stor risiko for havarier ved dem, sier Lijenzin.

Sjefen for avdelingen for system- og pålitelighetsanalyse ved SKI, Lennart Carlsson, sier at uhellet ved Barsebäck avslørte at risikoen for en kjernenedsmeltning var ti ganger større enn antatt.

– Vi aksepterer at svensk atomkraft kan oppleve et reaktorhavari med kjernenedsmeltning hvert hundre tusende reaktorår. Men feilen som ble oppdaget i nødkjølesystemet gjør det sannsynlig med et havari hvert ti tusende år. Det er uakseptabelt, og derfor stoppet vi reaktorene etter en måned, sier Carlsson.

From the Milwaukee Journal, May 1992

Non-cents: Laws of probability could end need for change

Chicago, Ill. —AP— Michael Rossides has a simple goal: to get rid of that change weighing down pockets and cluttering up purses.

And, he says, his scheme could help the economy.

"The change thing is the cutest aspect of it, but it's not the whole enchilada by any means," Rossides said.

His system, tested Thursday and Friday at Northwestern University in the north Chicago suburb of Evanston, uses the law of probability to round purchase amounts to the nearest dollar.

"I think it's rather ingenious," said John Deighton, an associate professor of marketing at the University of Chicago.

"It certainly simplifies the life of a businessperson and as long as there's no perceived cost to the consumer it's going to be adopted with relish," Deighton said.

Rossides' basic concept works like this:

A customer plunks down a jug of milk at the cash register and agrees to gamble on having the $1.89 price rounded down to $1 or up to $2.

Rossides' system weighs the odds so that over many transactions, the customer would end up paying an average $1.89 for the jug of milk but would not be inconvenienced by change.

That's where a random number generator comes in. With 89 cents the amount to be rounded, the amount is rounded up if the computerized generator produces a number from 1 to 89; from 90 to 100 the amount is

Please see **Money** *page* **22**



AP photo

A PENNY SAVED is a penny gone for inventor Michael Rossides (right) as he demonstrates his pay-by-gambling system to eliminate small change and round purchases to flat dollar amounts.

Money

From page **1**

rounded down.

Rossides, 29, says his system would cut out small transactions, reducing the cost of individual goods and using resources more efficiently.

The real question is whether people will accept it.

Rossides was delighted when more than 60% of the customers at a Northwestern business school coffee shop tried it Thursday.

Leo Hermacinski, a graduate student at Northwestern's Kellogg School of Management, gambled and won. He paid $1 for a cup of coffee and a muffin that normally would have cost $1.30.

Rossides is seeking financial backing and wants to test his patented system in convenience stores.

But a coffee shop manager said the system might not fare as well there.

"Virtually all of the clientele at Kellogg are educated in statistics, so the theories are readily grasped," said Craig Witt, also a graduate student. "If it were just to be applied cold to average convenience store customers, I don't know how it would be received."

Flerfaglighet i et amerikansk statistikkmiljø

Inntrykk fra et opphold ved University of Washington

Ivar Heuch

Høsten 1991 og våren 1992 hadde jeg forskningstermin og oppholdt meg ved Department of Statistics, University of Washington, i Seattle. Oppholdet gav mange nyttige impulser. Når man skal vurdere det utbyttet man får ved et gjesteopphold av denne typen, vil man naturligvis først og fremst tenke på de spesifikke faglige kontaktene. Men det betyr også mye at man får et omfattende innblikk i mer generelle spørsmål som angår virksomheten ved en universitetsenhet innenfor vårt fagfelt.

For en tilfeldig gjest fra Norge er det f. eks. svært lett å bli involvert i diskusjoner om grunnleggende amerikanske undervisningsspørsmål. Ofte blir man også trukket inn i diskusjoner om andre praktiske forhold som angår faget statistikk. Et aktuelt tema i Seattle var sammenhengen – eller muligens mangelen på sammenheng – mellom ferdighetene som statistikerne trenes opp i ved universitetet, og kravene som møter dem senere under yrkesutøvelsen. Diskusjonen på dette området var egentlig ikke så forskjellig fra den tilsvarende debatten som foregår i Norge. Likevel var det helt klart at spesialiserte statistikk-institutter i USA har langt flere ressurser til disposisjon enn vi er vant til her hjemme, slik at man i Seattle kunne by på adskillig bedre muligheter for spesialisering på ulike felt på "graduate" nivå.

Ved tidligere forskningsopphold i USA har det slått meg at man som gjest gjerne kommer i kontakt med et litt spesielt utvalg av personer som arbeider ved vertsinstituttet. I tillegg til ansatte med helt bestemte faglige interesser dreier det seg ofte om doktorstudenter og yngre folk som er i ferd med å etablere seg faglig. Dessuten oppstår det lett uformelle diskusjoner med andre ikke-amerikanere ved instituttet. En besøkende fra Norge må derfor være forsiktig med å karakterisere forholdene generelt på grunnlag av egne snevre inntrykk, spesielt fordi det tar lang tid før man kan danne seg et bilde av mer bakenforliggende interne forhold. Likevel slo det meg allerede fra begynnelsen av mitt opphold i Seattle at statistikerne der hadde spesielt bred kontakt med andre fagmiljøer ved universitetet. Ønsket om samarbeid førte nok ikke bestandig til oppsiktsvekkende konkrete resultater, men det var iallfall en klar vilje til stede til åpning utad. Jeg tror det kan være mulig å lære noe av denne innstillingen også for statistikere ved norske universiteter, selv om våre institusjoner er vesentlig annerledes organisert.

Det kan være mange årsaker til den mer fleksible holdningen som man ofte observerer i USA når det gjelder skiller mellom ulike fag. I Seattle kan noe av forklaringen ligge i historiske forhold. Tradisjonelt eksisterte det på den ene siden et sterkt institutt for biostatistikk ved University of Washington, med tilknytning til medisinske fagområder. I tillegg omfattet matematikk-instituttet bl. a. folk med interesser i sannsynlighetsteori og til dels også mer teoretisk statistikk. Da

Department of Statistics ble opprettet i 1979, kom utspringet fra disse miljøene, men ved nyrekrutteringer har det senere utviklet seg et sterkt eget særpreg. Ikke desto mindre virket det på meg som om de ansatte ved statistikk-instituttet stadig var seg sitt opphav bevisst, uten at det førte til noen mindreverdighetsfølelse overfor moderinstituttene.

Undervisningen var i enkelte avanserte kurs felles for studenter i statistikk og biostatistikk. Diskusjonen om hva man generelt skulle kreve av studenter på "graduate" nivå, foregikk også i stor grad i fellesskap. På mange måter stod statistikk-instituttet i en særstilling i forhold til biostatistikk, selv om de to instituttene hadde vidt forskjellig formell tilknytning under den sentrale organisering av universitetet. Department of Biostatistics var underlagt School of Public Health and Community Medicine, mens det noe mindre Department of Statistics tilhørte det ellers nærmest altomfattende College of Art and Sciences. Likevel ville mye av virksomheten ved instituttet for biostatistikk ut fra vanlig norsk målestokk bli karakterisert som hørende inn under det generelle faget statistikk. Det var f. eks. stor aktivitet på et område som overlevelsesanalyse. Samtidig var store deler av staben i biostatistikk også engasjert i arbeid mer direkte rettet mot anvendelser i epidemiologi.

I noen grad virket det som om de ansatte ved instituttet for statistikk prøvde å unngå dublering av undervisning og forskningsaktivitet som foregikk ved biostatistikk. Formelt sett hadde mange ansatte også "joint appointments" ved flere institutter, med særlig mange felles ansatte i statistikk og biostatistikk. Andre fagfelt som falt inn under denne ordningen med kombinert engasjement med statistikk, var genetikk, fiskeribiologi, skogbruk, datafag, matematikk, psykologi, økonomi og sosiologi. Det så likevel ikke ut som om disse fellestilknytningene fungerte like bra i alle tilfeller. Undervisningskatalogen for hvert institutt opererte med en liste over alle som hadde formell tilhørighet ved enheten, men man forstod fort at disse listene ikke gav noe reelt uttrykk for hvem som preget miljøene. Ved de generelle instituttseminarene arrangert av henholdsvis statistikk og biostatistikk var det f. eks. ytterst sparsomt med deltagere som hadde kontorplass ved det andre statistikk-instituttet. Faktisk kunne det virke som om gjesten fra Norge var den som hyppigst frekventerte begge seminarseriene! De to instituttene var geografisk adskilt i den forstand at de lå ca. 10–15 minutters gange fra hverandre innenfor samme universitetsområde, noe som egentlig ikke burde virke avskrekkende overfor vanlig kontakt.

Mens jeg oppholdt meg i Seattle, ble det foretatt et eget utspill fra statistikk-instituttets side for å oppmuntre til samarbeid med geofysikere, spesielt meteorologer og hydrologer. Sentral støtte med midler til en egen seminarserie gjorde det mulig å invitere foredragsholdere utenfra på dette grenseområdet. På denne måten ble de generelle institutt-seminarene sterkt preget av en bestemt type anvendelser i den aktuelle undervisningsperioden. Grunnlaget for samarbeid med geofysikerne var tidligere lagt på et individuelt nivå. Det vesentlig nye var oppfordringen til alle som arbeidet ved instituttet for statistikk om å følge seminarene hentet fra det aktuelle grenseområdet mot geofysikk. Tankegangen var at man skulle prøve å trekke flere personer inn på dette feltet. I praksis var det blandede reaksjoner på initia-

tivet. Det virket i stor grad som om studentene fulgte oppfordringen, muligens fordi seminardeltagelsen var en del av det offisielle studieopplegget. Ansatte uten spesiell tilknytning til geofysiske anvendelser glimret derimot ofte med sitt fravær. I rettferdighetens navn bør det likevel understrekes at det var klare skiller mellom seminarer holdt av geofysikere på anvendelsessiden og statistikere på metodesiden. Man kunne dessverre få inntrykk av at hver foredragsholder først og fremst snakket for folk som tilhørte samme leir. Brobygging mellom ulike fagfelt krever tydeligvis stor innsats, og den uformelle kontakten som ble etablert med seminarserien kan naturligvis ha vært et av de første trinnene i en lengre utvikling.

Dette seminaropplegget viser også hvor vanskelig det er å gjennomføre eksperimenter for utvikling av generelt samarbeid mellom statistikere og andre faggrupper. For nordmenn med begrensede muligheter til å gjennomføre slike eksperimenter hjemme blir det derfor enda viktigere å trekke korrekte konklusjoner av det vi kan observere ved gjesteopphold i land som USA. Mine observasjoner i Seattle viser at det kan være vanskelig å få i stand et godt *generelt* samarbeidsklima mellom et statistikk-institutt og en bestemt annen gruppering som gjør mye bruk av statistiske metoder. Derimot vil det i et åpent miljø lett kunne etableres verdifull kontakt mellom den *enkelte* statistiker og andre grupperinger på anvendelsessiden. Det var stor interesse blant de ansatte ved statistikk-instituttet for anvendte problemstillinger, men det virket for det meste som om forbindelsene berodde på personlige kontakter med andre fagfolk. Jeg var selv bl. a. interessert i forbindelseslinjer til matematisk og statistisk genetikk, noe som absolutt var representert innenfor faget statistikk i Seattle. Flere doktorstudenter arbeidet på dette området, men interessen blant de ansatte i statistikk var i det vesentlige begrenset til én person, som riktignok var meget aktiv. Innenfor dette feltet fungerte imidlertid kontaktene ellers med andre miljøer ved universitetet helt utmerket, både i retning av biologer og medisinere. Jeg kunne selv med stort utbytte dra nytte av disse kontaktene, til tross for at det ikke var noen gruppe med ansatte som virkelig samarbeidet om denne typen anvendelser i selve statistikermiljøet.

Jeg fikk også inntrykk av at mange andre ansatte ved instituttet for statistikk hadde sine egne spesielle interesser med tanke på anvendelser. Samme person kunne gjerne være opptatt av svært forskjellige problemområder på samme tid. Et noe tilfeldig eksempel på en litt underlig interessekombinasjon var en person som både gav seminarer om analyser av variasjon over tid av fruktbarhet i Irans befolkning, og estimeringsmetoder for størrelsen av hvalpopulasjoner. For noen personer var nok valg av interessefelt påvirket av mulighetene for økonomisk støtte for prosjektarbeid, men i mange tilfeller virket det som om orienteringen mot en bestemt type anvendelser representerte en genuin interesse som hadde holdt stand i årevis.

De ulike tilknytningene til anvendt statistikk så ikke ut til å lede til noen fragmentering av selve statistikk-instituttet. På et tidlig tidspunkt under mitt opphold i Seattle forhørte jeg meg hos forskjellige statistikere der om hva de syntes var det spesielle fellespreget ved akkurat deres institutt. Svaret var i stor grad at instituttet la stor vekt på anvendelser, uten at man dermed skulle miste den mer generelle metodologiske forankring. Jeg fikk også høre at en orientering i retning

av dataintensive metoder var et viktig fellestrekk. Akkurat det stemte godt med mitt eget inntrykk i løpet av oppholdet, og det kunne til tider virke som om motebegrep som "Gibbs sampling" og mer generelt "Markov chain Monte Carlo methods" på død og liv skulle presses inn i så mange ulike anvendelser som mulig. Alle institusjoner med stor aktivitet har vel sine spesielle "motebølger", og den generelle entusiasmen for disse metodene viste iallfall hvor effektiv spredningen var innad i instituttet av nye idéer .

Det er ikke helt enkelt å oversette erfaringer av denne typen fra et amerikansk universitetsmiljø i statistikk til norske forhold. Mitt inntrykk fra Seattle er i alle fall at det er fullt mulig for to relativt store statistikk-institutter å leve i fredelig sameksistens innenfor samme universitet. Jeg hørte aldri noen ymte om at instituttene for statistikk og biostatistikk egentlig burde vært slått sammen. Når det er naturlig med separate statistikk-enheter på forskjellige steder i universitetsstrukturen, burde tverrforbindelsene statistikerne imellom kunne ivaretas på andre måter, også med mer formelle bånd. Det var f. eks. meget vanlig i Seattle med aktive veiledningskomiteer for doktorgradsstudenter satt sammen på tvers av instituttgrensene.

Det positive som ligger i slike reelle og formelle kontakter mellom forskjellige fagmiljøer, er noe som lett burde kunne overføres til Norge. "Joint appointments" er sikkert vanskeligere å få i stand i Norge enn i USA, men skulle likevel kunne gjennomføres med den fleksibilitet som etter hvert er kommet i vårt eget undervisningssystem. I USA ser slik arbeidsdeling ut å virke godt når det faktisk foreligger et interessefelleskap. Men en bør være klar over at en formell tilknytning av en statistiker til et annet miljø ikke i seg selv garanterer noe virkelig samarbeidsforhold. Større og mer formaliserte samarbeidsprosjekter kan være vellykkede så sant de dreier seg om veldefinerte arbeidsgrupper av statistikere og andre fagfolk. Problemene oppstår når man prøver å skape generell entusiasme for bestemte typer anvendelser i et større statistikkmiljø. Det viktigste inntrykket fra Seattle er likevel at det kan virke fruktbart på det generelle statistikkmiljøet at hver enkelt ansatt får gode muligheter til faglige kontakter utad.

Rapport fra hovedkonferansen 1992 i American Statistical Association/ Institute of Mathematical Statistics, Boston , 9.-13. august

Hvert år arrangerer American Statistical Association, vanligvis sammen med Institute of Mathematical Statistics samt Biometric Society (Eastern og Western North American Regions), en hovedkonferanse som samler ca. 3000 deltagere. I 1992 var denne konferansen lagt til Boston. Hoveddelen av deltagerne holdt til i 2 gedigne nabohoteller, Sheraton og Marriott. Det var også i disse hotellene stort sett alle foredrag og aktiviteter var lagt. I de fleste store byer i USA finner man to og flere nabohoteller som samlet egner seg for slike store konferanser.

I forkant av konferansen ble det arrangert flere etterutdanningskurs (varighet 1-2 dager), bl.a. i statistisk numerikk og telleprosesser & overlevelsesanalyse. I tillegg til slike faglige kurs arrangeres hvert år et kurs i presentasjon av foredrag. De fleste av oss burde kanskje vært innom et slikt kurs !

På en slik omfattende konferanse er det selvfølgelig en meget stor bredde i foredragsemnene. Utenom de inviterte hovedbidragene er det til enhver tid ca. 25 parallellsesjoner. Her er det mulig å finne ut hva som foregår på forskningsfronten i de fleste områdene av statistikken, eksempelvis nevrale nett, wavelets, miljøstatistikk, utvalgsundersøkelser, ikkeparametrisk regresjon, Bayes analyse med Gibbs sampling, osv. Selv gikk jeg på foredrag som varierte fra glimrende forelesninger av Willem van Zwet om asymptotisk teori & resampling (Wald memorial lectures) og foredrag om kaos teori av Mark Berliner til foredrag om konsultasjonstjeneste , etterutdanning i statistikk for matematikklærere i skolen og statistikkundervisning i skolen.

Til disse årlige konferansene er det alltid en bred utstilling av nye bøker og programvare. De viktigste internasjonale forlagene som gir ut statistikkbøker er representert. Bestiller man bøker på konferansen får man vanligvis en rabatt på 10-15%. Den kanskje mest interessante delen av utstillingen gjelder det som foregår på programvaresiden hvor absolutt siste nytt utstilles og demonstreres, eksempelvis om S-plus, Minitab, SAS, Systat, SPSS, BMDP, NAG og Mathematica. Selv fikk jeg tak i et meget avansert, (og viktigst: lett å bruke) matematisk tekstbehandlingssystem for Macintosh, Mathwriter, på denne måten. For all programvare får man gratis demodisketter.

For tidligere USA studenter som undertegnede er disse møtene en fin måte å holde en viss kontakt med amerikanske kolleger, selv om det selvfølgelig er begrenset hvor mye faglig samarbeid som kan utføres i løpet av denne korte tiden møtet varer. Disse konferansene gir også muligheter for faglig kontakt i forskjellige miljøer. Selv har jeg i tillegg til kontakt med diverse universiteter også fått kontakt med kolleger fra Census Bureau, Bureau of Labor Statistics, Internal Revenue Service og IBM .

Jan F. Bjørnstad

Det norske statistikermøte 1993

Neste års store begivenhet for norske statistikere nærmer seg, og her kommer en **foreløpig** oversikt over arrangementet.

Statistikermøtet vil bli holdt på Ustaoset høyfjellshotel i tidsrommet 14. - 16. juni 1993.

Følgende organisering av møtedagene er foreslått:

Mandag 14. juni:	kl. 13.00-15.00	Registrering
		Lunsj
	kl. 15.00-17.00	Foredrag (plenum)
	kl. 18.00-19.00	Middag
	kl. 19.00-21.00	Årsmøte
Tirsdag 15. juni:	kl. 9.00-10.30	Parallellsesjon
	kl. 10.30-10.50	Kaffe
	kl. 10.50-12.30	Foredrag (plenum)
	kl. 12.30-16.00	Lunsj
	kl. 16.00-18.30	Foredrag (plenum) evt. parallellsesjon
Onsdag 16. juni:	kl. 9.00-10.30	Parallellsesjon
	kl. 10.30-10.50	Kaffe
	kl. 10.50-12.30	Parallellsesjon
	kl. 12.30-15.00	Lunsj (og pakking)
	kl. 15.00-17.00	Foredrag (plenum)

Ustaoset ligger på Hardangervidden og har gode togforbindelser med Oslo og Bergen. Ustaoset hotel ligger i gangavstand fra togstasjonen. Pris for kost og losji er ca. 2000 kr. (2 døgn + en ekstra lunsj). Studenter kan overnatte på Ustakvil til ca. halv pris. Dette er et pensjonat som drives av hotellet og har beliggenhet like bak hotellet.

Det er viktig at statistikkstudentene kommer med i det faglige miljøet. Hovedfagsstudenter ved universitetene og studenter på NTH (statistikk-fagretning) ønskes velkommen.

Som det fremkommer av programmet planlegger programkomiteen i år å holde parallellsesjoner, den ene i anvendt statistikk og den andre i teoretisk statistikk/sannsynlighetsteori. Da dette krever mange foredrag, oppfordres potensielle foredragsholdere å ta kontakt med et av medlemmene i programkomiteen snarest mulig. Komiteen består av:

Ørnulf Borgan	UiO	tlf. 02-855862
Liv Høyland	NTH	tlf. 07-593533
Jostein Paulsen	UiB	tlf. 05-212886

For å kunne bestille riktig antall plasser på Ustaoset hotel ber vi om at så mange som mulig fyller ut og returnerer svarslippen under til:

Marit Olsbø Holen
Ole Viks gate 9
5006 Bergen

Foreløpig (ikke bindende) påmelding

__Ønsker overnatting på hotellet.
__Ønsker overnatting på pensjonatet.

NAVN:_____
ADRESSE:_____
POSTNR:_____

American Statistical Association

1429 Duke Street, Alexandria, Virginia 22314-3402
(703) 684-1221 • Facsimile: (703) 684-2037

Dates: August 8–12, 1993
Location: San Francisco, California
Sponsors: ASA, ENAR, WNAR, IMS

Key Dates:

- •February 1 Abstracts for contributed papers due (ASA, ENAR, WNAR)
- •July 15 Last date for hotel reservations to arrive in San Francisco
- •July 15 Last date for preregistration forms to arrive at ASA office

Preregistration Fees for Meetings:

- •Members (ASA, ENAR, WNAR, IMS) – $125
- •Nonmembers – $185
- •Full–time Students – $ 42
- •Retired – $ 48

Co–Headquarters Hotels (all meetings activities and ASA Continuing Education courses held here):

- •San Francisco Hilton & Towers ($109 single; $126 double)
- •Parc Fifty Five ($109 single; $120 double)

Auxiliary Hotels

- •Hotel Nikko San Francisco ($115 single; $135 double)
- •St. Francis ($115 single; $135 double)
- •The Handlery Union Square ($88 single; $98 double)
- •Aston Pickwick ($80 single/double)
- •Shannon Court ($90 single/double)
- •Hotel Mark Twain ($85 single/double)
- •Hotel Union Square ($89 single/double)
- •Hotel Metropolis ($79 single/double)

Additional information will be available in forthcoming issues of Amstat News.

Further Questions? Contact:

Meetings Department
American Statistical Association
1429 Duke Street
Alexandria, VA 22314–3402
703/684–1221; FAX: 703/684–2037

ANNOUNCING
XXVèmes JOURNEES DE STATISTIQUE - VANNES (FRANCE) 24-28 May 1993

This congress, organised by the Association pour la Statistique et ses Utilisations (A.S.U.) brings together more than 500 statisticians from Universities and from Industry each year.
It offers sessions in Mathematical Statistics and Applied Statistics. The main themes in Vannes will be as follows :

Medical and Pharmaceutical Research
Agronomical Research
Quality and Reliability
Econometrics, Management and Marketing
Statistical Packages
Demography
History of Statistics
Teaching of Statistics
Classification and Discrimination
Troncated or Censored Data (life data)
Sample Survey Design
Image and Signal Processing
Bayesian Statistics
Neural Networks
Experimental Designs
Wavelets

The papers are to be presented in the form of a lecture or through posters. Non-French speakers may use English.
Vannes is an old town, full of character, on the south coast of Brittany, which celebrates being two thousand years old in 1993. It can be reached by train (TGV) or plane (via Lorient) and the Organising Committee will give you a warm welcome.

Deadlines

11 January 1993 :	Papers to be submitted
23 March 1993 :	Notification of acceptance of papers
1 April 1993 :	Increase in enrolment fees
1 April 1993 :	Despatch of definitive programme for the Congress

Enrolment Fees

Before 1/4/93 :	1500 FF
After 1/4/93 :	2000 FF
Reception Buffet at The Ile aux Moines (24/05/93) (Optional) :	130 FF
Set price for 5 lunches (Optional) :	300 FF

Additional information and enrolment forms may be obtained from the Organising Committee :

XXVèmes Journées de Statistique - Institut Universitaire de Technologie - 8 rue Montaigne -56014 VANNES - **TEL:** (33)97 46 06 00 - **FAX:** (33)97 46 04 40 -
MAIL: VANASU93@CERBERE.CICB.FR
Domiciliation Bancaire : ASU XXV JOURNEES DE STATISTIQUE, 16006 36011 92452004510 68
Hotel reservations (to be made as soon as possible) will be dealt with by the Tourist Office, 1 rue Thiers - 56000 VANNES - **FAX:** (33)97 47 29 49

15. NORDISKE KONFERANSE I MATEMATISK STATISTIKK

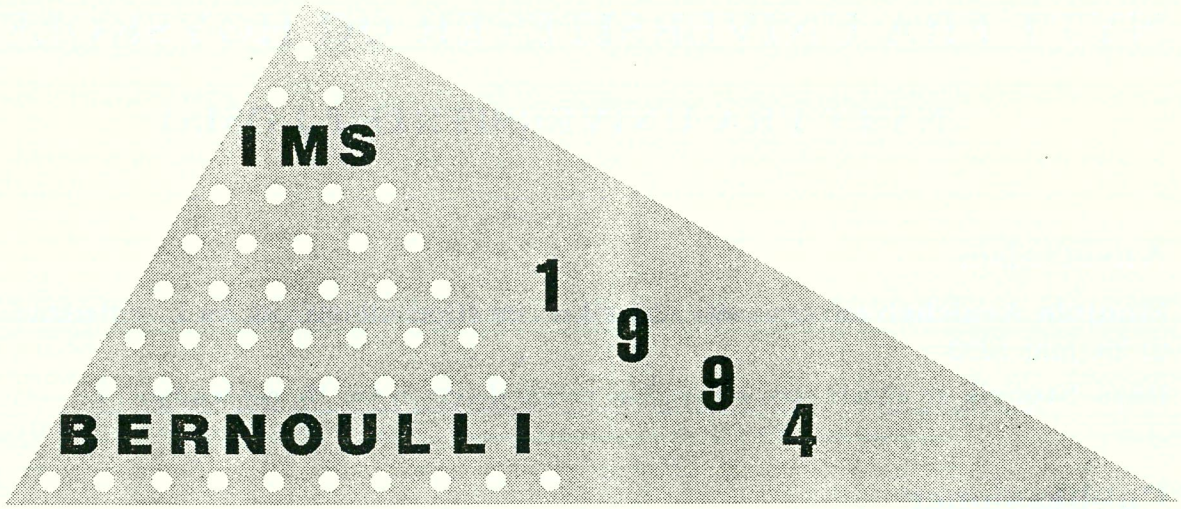
Konferansen arrangeres i Lund, **15. - 19. august. 1994.**

Organisasjonskomite:
Jan Lanke
Georg Lindgren
Lena Zetterqvist

Programkomite:
Georg Lindgren, Lund
Olle Nerman, Gøteborg
Bo Ranneby, Umeå
Daniel Thorburn, Stockholm

14TH INTERNATIONAL SYMPOSIUM ON FORECASTING

Arrangeres **12. -15. juni 1994** ved Handelshögskolan i Stockholm. For informasjon kontakt: Sune Karlsson, Handelshögskolan i Stockhom, Box 6501, S-113 83 Stockholm, e-mail: ISF94@HHS.SE eller ISF94@SEHHS.bitnet.



3rd World Congress of the Bernoulli Society for Mathematical Statistics and Probability and 55th Annual Meeting of the Institute of Mathematical Statistics

JUNE 20-25, 1994

The University of North Carolina at Chapel Hill

The William and Ida Friday Continuing Education Center

Program

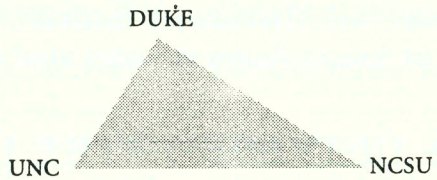
The program will include invited and contributed papers in a broad spectrum of areas of mathematical statistics and probability, reflecting the emphases of the two Societies. Speakers for the special named lectures will be announced in later information.

Planning Committees

The joint IMS-Bernoulli Society Program Committee is chaired by Professor Morris Eaton, University of Minnesota. The Conference Organizing Committee is chaired by Professor Ross Leadbetter, University of North Carolina at Chapel Hill.

Chapel Hill

Chapel Hill maintains its inviting small town flavor with shops, museums, historical landmarks, and sporting events within walking or trolley-ride distance. Chapel Hill is located in North Carolina's Triangle, which is home to Duke University, NCSU, UNC-CH, the state capital, and the Research Triangle Park. Late June in North Carolina is typically sunny, with daytime temperatures of 70–85°F (20–30°C) and nighttime temperatures of 55–65°F (13–18°C).



Triangle Area Universities

Travel and Accommodations

Chapel Hill and The Friday Center are in central North Carolina, approximately 15 miles from the Raleigh Durham International Airport (RDU) and near interstate highways 40 and 85. Accommodations will include both hotel and dormitories within Chapel Hill. Transportation arrangements are in the planning stages.

Social Events

A welcoming reception, banquet, and tours of local attractions are planned for participants.

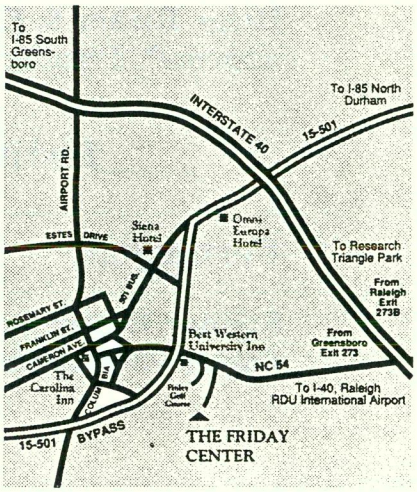
FRIDAY CENTER

Further Information

Subsequent announcements, including program information and registration forms, will be mailed in the spring and fall of 1993. It is planned that final program and registration information will be included in the January-February issue of the IMS Bulletin.

Address inquiries to:

IMS/Bernoulli Society Meeting
Division of Continuing Education
CB# 1020 The Friday Center
The University of North Carolina
at Chapel Hill
Chapel Hill, NC 27599-1020



NYTT FRA UNIVERSITETER OG HØYSKOLER.

NYTT FRA UNIVERSITETET I OSLO

Ansettelser

Manjula Arulchelvam er ansatt som vikar i en stipendiatstilling fra 1. september 1992 til 30. juni 1993

Bård Storvik er ansatt som stipendiat ved avdelingen fra 15. september i år.

Publikasjoner

Svein Eggen, Bent Natvig & Jørund Gåsemyr : ' "*Variation in torus palatinus prevalence in Norway*", Statistical Research Report No. 8, 1992

Svein Eggen & Bent Natvig : "*Concurrence of torus mandibularis and torus palatinus*", Statistical Research Report No. 9, 1992

Forskerkurs

Våren 1992 vil det bli arrangert et nasjonalt doktorgradskurs om statistiske modeller og metoder for telleprosesser ved Matematisk institutt. Kurset vil bli ledet av Ørnulf Borgan. Nærmere omtale av kurset finnes et annet sted i dette nummer av TG.

NYTT FRA UNIVERSITETET I TRONDHEIM

Institutt for matematiske fag - NTH

Rektor (NTH).

Emil Spjøtvoll er valgt til rektor ved NTH for perioden 1.1.93 - 31.12.95.

Publikasjoner.

Helge Blaker & Emil Spjøtvoll: "Preference functions and acceptability in statistical inference", Preprint Statistics No. 1/1992.

Institutt for matematikk og statistikk - AVH

Permisjon:

Steinar Engen har forskningsfri i 1993 og vil oppholde seg ved University of Oregon, Eugene i 3 måneder, fra juni til september.

Ansettelse:

Helga Vivås er ansatt som førsteamanuensis - vikar for 1993.

Publikasjoner.

Jan F. Bjørnstad: "On the generalization of the likelihood function and the likelihood principle", Preprint Statistics No. 2/1992.

Jan F. Bjørnstad & Frode Skjold: "Interval estimation in the presence of nonresponse", Preprint Statistics No. 3/1992.

NYTT FRA UNIVERSITETET I TROMSØ

Ansettelse:

Cand.scient. **Marita Melhus** er ansatt som stipendiat fra 1.1.93.

ANNONSE

UNIVERSITETET I TRONDHEIM
Den allmennvitenskapelige høgskolen

STIPENDIATSTILLING I MATEMATIKK/STATISTIKKK/INFORMATIKK FOR KVINNELIGE SØKERE (LP 17.515 KODE 1017)

Ved Universitetet i Trondheim, Den allmennvitenskapelige høgskolen, Det matematisk-naturvitenskapelige fakultet, er det ledig en stipendiatstilling øremerket for kvinnelige søkere. Dette er en utdanningsstilling som i hovedsak skal gi lovende kvinnelige forskerrekrutter anledning til faglig utvikling.

Stillingen vil bli lagt til fagområdet matematikk/statistikk eller informatikk.

Nærmere opplysninger om stillingen kan fås ved henvendelse til Institutt for matematikk og statistikk, professor Steinar Engen, tlf. 07 59 17 47, Institutt for informatikk, tlf. 07 59 17 24 eller professor Torbjørn Skramstad, tlf. 07 59 18 43/02 47 78 89.

Det kreves fullført cand.scient.-eksamen med hovedfag i informatikk, matematikk eller statistikk, eller tilsvarende fagutdanning. Søkeren må normalt ha bestått hovedfaget med godt resultat eller kunne dokumentere tilsvarende nivå gjennom senere forskningsarbeider og prosjektbeskrivelsen.

Den som blir tilsatt må rette seg etter de vilkår som til enhver tid gjelder for offentlige tjenestemenn.

Lønn etter ltr. 16 i statens lønnsregulativ.

Tilsetting skjer i samsvar med retningslinjer for tilsetting som stipendiat ved universiteter og høgskoler, og søkere vil måtte forplikte seg til å delta i organisert doktorgradsstudium innenfor tilsettingsperioden. Det vil videre bli inngått kontrakt om tilsettingstid og arbeidsplikt for stipendiaten.

Søknaden må inneholde opplysninger om utdanning, eksamener og tidligere praksis. Bekreftede kopier av vitnemål og attester må legges ved. Sammen med søknaden skal det følge en beskrivelse av forskningsarbeidet som er planlagt for doktorgraden og som tenkes gjennomført i stipendperioden. I prosjektbekskrivelsen skal det

- redegjøres for problemstillinger og metode
- refereres til relevant teori på området, slik at søkeren dokumenterer god kjennskap til aktuell forskning på feltet

Prosjektbeskrivelsen skal kunne danne grunnlag for å vurdere opptak på den organiserte forskeropplæringen ved fakultetet.

Publikasjoner og andre vitenskapelige arbeid som søkeren ønsker det skal bli tatt hensyn til ved vurderingen, må følge søknaden, helst i 3 eksemplarer.

Det vil bli tatt hensyn til fellesarbeider. Dersom det i fellesarbeider er vanskelig å identifisere den enkelte søkers innsats, må søkeren vedlegge en kort redegjørelse for sin medvirkning.

Søknaden sendes til Universitetet i Trondheim, Den allmennviteskapelige høgskolen, Personalkontoret, 7055 Dragvoll, innen 06.01.93. Merk søknaden med journalnummer 1633.

ANNONSE

FOLKEHELSA
Statens Institutt for Folkehelse

STATISTIKER

Folkehelsa arbeider for å forebygge sykdom og helseskade, fremme folkets helse og bedre helsetjenestene. Virksomheten baseres på at vi driver forskning og omsetter forskningsresultater til praktisk bruk, at vi forvalter nasjonale informasjonssystemer om helsetjenesten og at vi formidler og produserer vaksiner og produkter som forebygger sykdom. Arbeidet er rettet mot: mennesker og samfunn; kjemiske, fysikalske og biologiske faktorer i omgivelsene; infeksjonskontroll og helsetjenestene.

Ved Avdeling for samfunnsmedisin er ledig fast stilling som forsker (kode 1108/09) ved Seksjon for epidemiologi.

Stillingen ønskes besatt av erfaren og dyktig statistiker, gjerne med bakgrunn i biostatistikk. Ved seksjonen er det omtrent 25 medarbeidere med utdanning fra medisin, statistikk og samfunnsfag. Den som ansettes skal drive selvstendig forskning, men også delta i prosjekter som søker å forstå årsaker til sykdom, særlig knyttet til infeksjonsepidemiologi og reproduksjons- og barneepidemiologi. Veiledning av stipendiater vil inngå i arbeidet. Selvstendig kontakt og eventuelt forskningssamarbeid med statistikkmiljøer utenfor Folkehelsa er en fordel.

Kvinner oppfordres spesielt til å søke stillingen.

Det er røykfritt miljø ved instituttet. Vi har barnehage med mulighet for plassering av barn fra 2-års alder.

Stillingen lønnes i h.t. Statens lønnsregulativ, ltr. 15 - 21 (pt. kr. 196.541,- - 242.509,- pr. år). Spesielt godt kvalifiserte søkere kan eventuelt gis høyere avlønning (inntil ltr. 25, p.t. kr. 282.560,-). Det trekkes innskudd til Statens Pensjonskasse.

Nærmere opplysninger om arbeidsoppgaver og -forhold ved seksjonsleder Per Magnus, tlf. (02) 35 60 20.

Søknad merket "9241" bilagt attester/vitnemål og eventuelle publikasjoner, bes innsendt Statens Institutt for Folkehelse, Geitmyrsveien 75, 0462 Oslo 4, innen 8. januar 1993. På eget ark gis oversikt over utdanning og praksis i kronologisk rekkefølge.

Medlemskap i Norsk Statistisk Forening

Kontingenten i foreningen er på kr. 90 pr. år for regulære medlemmer og kr. 30 for studenter. Korrespondanse om medlemskap og innbetaling av kontingent skal skje til foreningens kasserer:

Norsk Statistisk Forening v/Ole Lauvskar
Vital Forsikring
5020 Bergen

Innbetalinger skal gjøres på foreningens postgirokonto 0802 5 15 10 55. Medlemmer som er usikre på når de betalte kontingent sist, vil nederst til høyre på adresseseddelen for Tilfeldig Gang finne angitt hvilket år siste betaling er registrert for. Vær imidlertid oppmerksom på at adressesedlene for bladet ligger noe etter ajourføringen av medlemsregisteret, slik at nylig foretatte innbetalinger ikke vil være tatt med her.

Nye medlemmer melder seg til kassereren under adressen oppgitt ovenfor. Slike medlemmer får automatisk tilsendt Tilfeldig Gang, med forbehold om at det tar noe tid før tilføyelser og forandringer i medlemsregisteret gjenspeiler seg i adressesedlene til bladet. Adresseforandringer meldes også til kassereren. Ved utsendelsen av Tilfeldig Gang benyttes det kopier av det vanlige medlemsregisteret, så adresseforandringer behøver ikke meldes separat til bladet.

Annen korrespondanse med foreningen skal foregå med formannen under denne adressen:

Norsk Statistisk Forening v/Trygve Nilsen
Matematisk institutt
Universitetet i Bergen
Allegt. 55
5007 Bergen

Bidrag til Tilfeldig Gang.

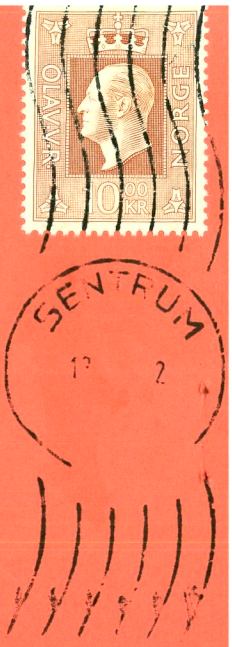
Neste nummer av Tilfeldig Gang vil forhåpentligvis komme ut i mars 1993. Alle slags bidrag til bladet er velkommen, enten det gjelder nye temaer av interesse for norske statistikere, eller man har kommentarer til innlegg som allerede har stått i bladet. Stoff til neste nummer må være redaktøren i hende senest 1. mars. Det kan gjerne sendes som ASCII-filer, eventuelt tilpasset LaTeX, på diskett eller med elektronisk post. Redaktørens adresse er:

Tilfeldig Gang v/Jan Bjørnstad
Institutt for matematikk og statistikk
Universitetet i Trondheim
Den allmenvitenskapelige høgskolen
7055 Dragvoll
Telefonnummer 07591751.

Elektronisk post: Jan.Bjoernstad@avh.unit.no
For LaTeX-dokumenter: Inger.Seehuus@avh.unit.no

TORGERSEN, ERIK
MAT. INST. AVD. C
P.B. 1053, BLINDERN
0316 OSLO 3 93

B



Returadresse:

Tilfeldig Gang v/ Ole Lauvskar
Vital Forsikring A.S
5020 Bergen