



ISSN 0803-8953

**TILFELDIG
GANG**

Nr. 2, årgang 25
April 2008

Utgitt av Norsk Statistisk Forening

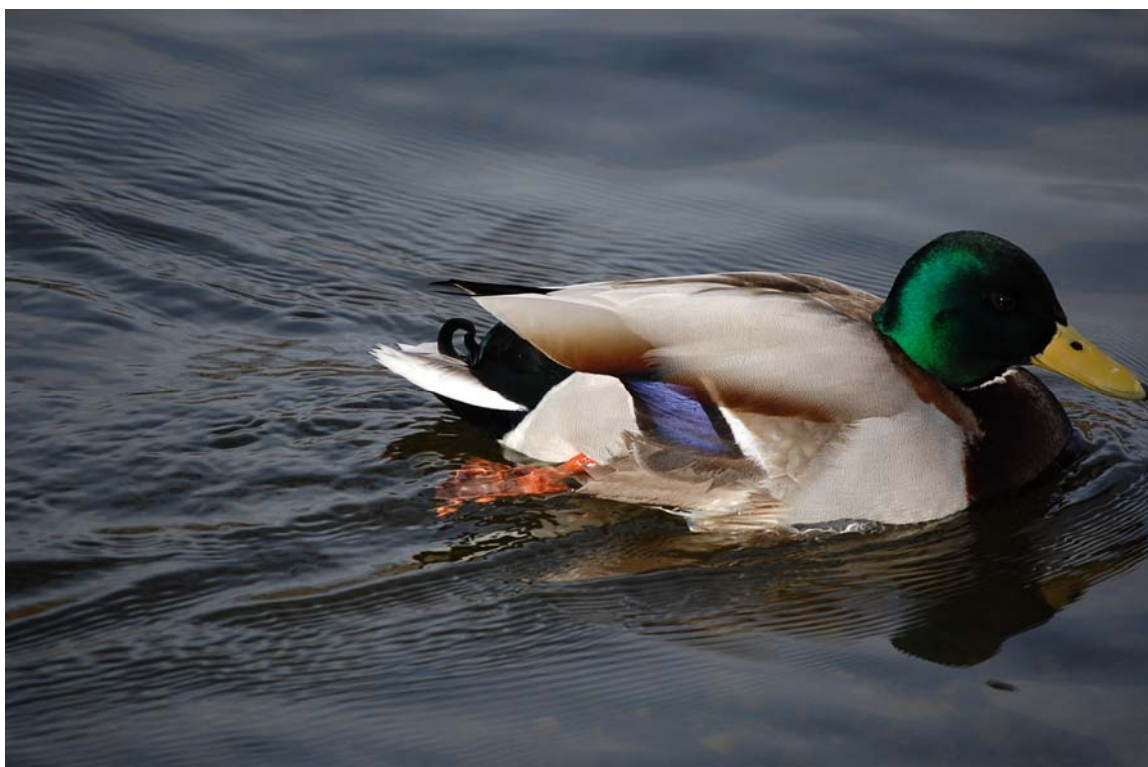
INNHold

	Side
Fra lederen, <i>Magne Aarset</i>	3
Fra redaksjonen, <i>Mette Langaas</i>	5
Brev	
Reisebrev fra Bristol, <i>Geir Storvik</i>	7
Vismannen Confuseus om statistikk, <i>Øystein Evandt</i>	9
Jubileum	
Erik Bølviken 60 år, <i>Tormod Næs og Bent Natvig</i>	10
Stipendiatpresentasjoner	
Presentasjon av doktoravhandling, <i>Egil Ferkingstad</i>	12
Selvpresentasjon, <i>Kristian Hindberg</i>	14
Møter og konferanser	
Abelfinalen 2008 og Walkers algoritme, <i>Øyvind Bakke</i>	17
NORDSTAT 2008	19
Puslerispalten	
Pusleri nr. 24, <i>Jostein Lillestøl</i>	20
Løsning til pusleri nr. 23, <i>Jostein Lillestøl</i>	21
Korrelasjonen mellom gjennomsnitt og median, <i>Nils Lid Hjort</i>	22
Utlysninger	
Stipendiatstilling i biostatistikk, UiO	26
Meldinger	
Nytt fra Norges teknisk-naturvitenskapelige universitet	27
Nytt fra Universitetet i Bergen	28
Nytt fra Universitetet i Oslo, Avd. C	29
Nytt fra Universitetet i Oslo, Avd. for biostatistikk	30
Nytt fra Universitetet i Tromsø	31
Årsberetning, Statistisk forening i Bergen	32
Referat fra årsmøte, Statistisk forening i Bergen	33
Lokalforeningene i Norsk statistisk forening	34

Nå er våren kommet (tror jeg)!

Det er ikke til å komme fra at man blir litt lettere til sinns når solen skinner, fuglene synger og det er vår. Ikke bare blir det lysere så vel morgen som kveld, men mange som meg ser faktisk lysere på sin umiddelbare framtid. Det er like før jeg gjør som Bjørnson og velger meg april. Dette er tiden for å planlegge avslutningen på årets kurs (eksamensforberedelser der jeg heldigvis bare skal lage og ikke besvare eksamensoppgaver), de siste forelesninger på så lang tid jeg greier (og er interessert i) å planlegge framover og ikke minst; deltakelse på konferanser jeg faktisk har tid til å delta på.

Og som mange av dere ganske sikkert vil være enige i, så er konferanser viktige. Jeg har riktignok nettopp hatt bokettersyn i mitt lille aksjeselskap og oppdaget at det er ikke alle som er enige i at dette er viktig (eller nødvendig for den saks skyld). Men til og med dette er det mulig å fortrenge nå når solen titter fram. Og her kommer et av poengene mine denne gang. Ikke bare la oss være snill mot hverandre og småfuglene; la oss være snill mot studentene også. Som jeg har nevnt tidligere har jeg alltid hatt stor sans for at foreningen støtter studenter som ønsker å reise på den norske eller nordiske konferansen i statistikk. Dette vil vi selvfølgelig gjøre også i år (og neste år også, såfremt ikke noe som ligger utenfor Normalfordelingens haler skulle inntreffe). Så vennligst spre det glade budskap!



Det ligger imidlertid (som vanlig) noen mørke skyer bak mine optimistiske funderingene. Jeg har tidligere veiledet en hovedfagsstudent som like før innlevering døde av en overdose narkotika. Og i helgen som var tok også en av studentene mine dette semesteret sitt eget liv. Forferdelig trist! Jeg synes det er viktig å minne oss såkalt etablerte som omgås ungdom under utdanning daglig på at livet til tider er tøft (også i dag). Jeg tenker ofte på at jeg bør være like tålmodig og forståelsesfull som mine mentorer var mot meg da jeg studerte. Det er mange jeg burde ha takket, men Arnljot (Høyland) i Trondheim og Nils (Lid Hjort) i Oslo var eksemplariske. Jeg mener selvfølgelig ikke å si at vi som foreleser har mange liv på samvittigheten, men det er klart at vi er en viktig del av mange unge menneskers skjebne. Og selv så skremmende som det er å innse, så er vi til tider også rollemodeller.

En annen mørk sky (ikke så veldig mørk, riktignok) er at når sommeren kommer, så øker gjerne reisevirksomheten vår. I seg selv oftest en positiv opplevelse, men samtidig med at all transport forurensar så er reising også noe av det farligste vi gjør. Statistikk kan som kjent presenteres på så mange måter, men det faller faktisk ned ett rute- eller charterfly sånn cirka annen hver uke (på verdensbasis). På de verste årene var antallet faktisk oppe i nesten ett hver uke (for eksempel 48 i 1995 ifølge Flight International). Men alle var jo ikke norske, da. Og så er det også et tankekors med den positive korrelasjonen mellom miljøengasjement og så mye den enkelte av oss forurensar.

Men la oss løfte hodet igjen og, om nødvendig, bruke vårt felles fag til å spre både glede og kunnskap rundt oss. Ifølge mediene er det igjen barna og ungdommen vi må satse på. (Selv om jeg syntes det var lettere å akseptere når det ble påstått for 30 år siden.) Så la oss filosofere som Harald Heide Steen jr. i en av sine mange uforglemmelige stunder:

*Nå er det snart sommer, og vi har ingen vinter hatt.
Men hva skal vi med en vinterhatt, når det er snart sommer?*

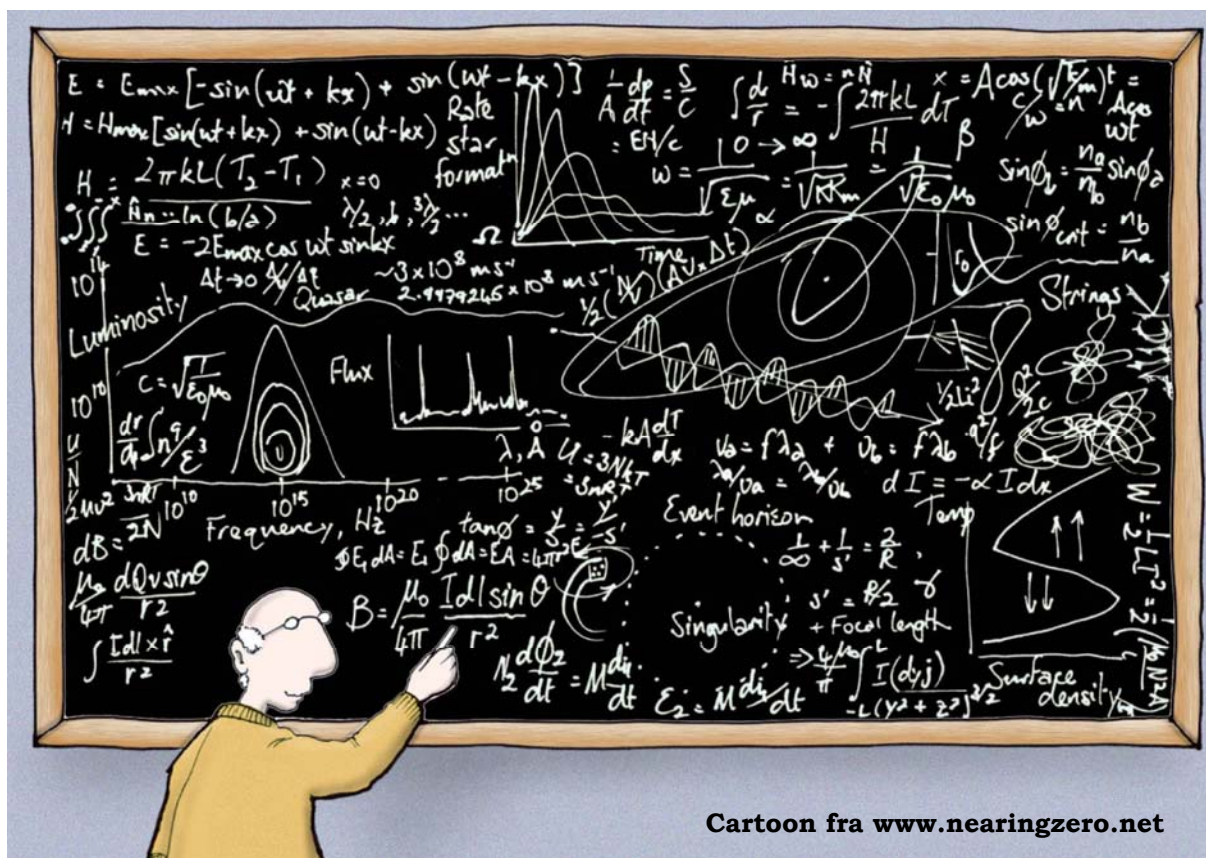
Jeg ønsker alle en riktig god sommer, gode konferanser og "Have a safe flight"!

Magne

Fra redaksjonen

Velkommen som leser av TG nummer 2, 25.årgang!

Vårsemesteret 2008 nærmer seg slutten, og studenter og forelesere starter forberedelsene til avsluttende eksamen. Studentene klør seg i hodet og tenker over hva de egentlig har lært, og foreleserne forsøker å oppsummere teori godt hjulpet av gode eksempler fra den virkelige verden. Og, alt skjer på krittavle, med foreleser lett hostende av krittstøvet, kanskje med innslag av lysark på gammel overheadprojektor eller Powerpoint-filer. Slik er det og slik har det vært i mange år – man endrer ikke uten videre på en oppskrift som fungerer. Og, heldigvis (?): "Nei takk til søvndyssende Powerpoint-forelesninger uten bruk av tavle", er dommen fra studentene, ihvertfall fra mine studenter.



Og det er vel også noen økonomiske føringer. Jeg hadde gledet meg til dette semesteret – kanskje jeg nå skulle snuse litt på nye tekniske hjelpemidler? Men ikke på universitetet, nei. I grunnskolen fikk jeg prøve trykksensitiv tavle – og elevene på 3. trinn behersket dette til fingerspissene. Her var det mulig å lagre alt jeg hadde skrevet på tavla, og jeg kunne ha lagt ut notatene slik at elevene som ikke var tilstede (enten fysisk eller mentalt) ikke gikk glipp av hovedpunktene i undervisningen.

Ved NTNU har et studentresponsystem (klikkere, mentometerknapper) vært forsøkt i undervisningen (førsteamanuensis Dan Jacobsen, <http://www.svt.ntnu.no/psy/Dan.Jacobsen/>), med svært positive tilbakemeldinger fra studentene. Foreleser har stilt et spørsmål med fem mulige svaralternativer – av ulik grad av riktighet. Studentene har (i auditoriet) diskutert spørsmålet i grupper à to–fire

studenter og kommet frem til hvilket svaralternativ de synes er mest korrekt. Dette alternativet har de så valgt ved hjelp av hver sin klikker. Prosentvis fordeling av de valgte svaralternativene blir så vist momentant, og foreleser går gjennom de ulike svaralternativene. Studentene har vært veldig takknemlige for at de endelig har blitt sett, fått tilbakemelding på egen oppfatning av det som er forelest, blitt engasjert, fått delta i auditoriumsundervisningen, og blitt kjent med andre studenter. Dette har virkelig falt i smak hos studentene. Har du tanker om, eller erfaringer med, bruk av studentresponssystemer eller andre aktiviserende tiltak i forelesningene, så vil vi gjerne høre fra deg i TG! TG bør være en viktig arena for pedagogiske tanker rettet mot undervisning i statistikk.

Dette nummer av TG står tuftet på våre faste spalter, med friske innspill fra vårt statistiske miljø. Du kan lese stipendiatpresentasjoner, reisebrev, jubileumshyllest, invitasjon til konferanse (Nordstat 2008), referat fra Abelkonkurransen med et statistisk problem, kloke sitater fra vismannen Confuseus, Jostein Lillestøls puslerier og løsninger, utlysning av PhD-stipend, nytt fra universitetene og kontaktinformasjon til NSF's lokalforeninger. Dette nummer av TG kan du også lese fra <http://www.math.uio.no/nsf/> (velg "Tilfeldig gang" i menyen).

Vi hører gjerne fra deg om alle dine hjertesaker. Frist for innsending av stoff til TG 3/2008 er 15.september 2008.

Vi ønsker deg en vakker vår og en varm sommer!



Fotograf: Hege Løkken Hansen

Redaksjonen v/Mette Langaas

Reisebrev fra Bristol

Geir Storvik

Høsten 2007 var det igjen muligheter for et “sabbatsår” i det store utland. Med kone i full jobb og barn i skole-/barnehagealder var det langt fra opplagt at dette skulle bli en realitet, men etter en del diskusjoner ble vi enige om å satse. Selv om min kone Ulla kanskje heller ville til et mer eksotisk sted, endte vi opp i Bristol, England. England ble valgt pga. språket mens Bristol fristet med sitt faglig sterke miljø innen Bayesianske og Monte Carlo metoder samt statistikk generelt.

Den mest kjente statisiker i Bristol er Peter Green. Ved å sitte på hans nabokontor med ikke alt for lydtette vegger har jeg merket hvor mye tid han bruker på å snakke med andre, kolleger og studenter, og fungerer som en slags faglig nestor i miljøet. Han har også vært drivkraften bak et større forskningsprogram, SuSTaIn (*Statistics underpinning Science, Technology and Industry*), som både finansierer phd- og post doc studenter samt har tilført miljøet mange nye og unge fast ansatte. Miljøet er stort sett delt opp i to hovedgrupper, en innen Bayesianske metoder og Monte Carlo beregninger og en innen ikke-parametriske metoder. Utover det har de også folk innen sannsynlighetsteori og mer generell statistikk. For min del har jeg mest snakket med Christophe Andrieu som stort sett jobber med MCMC og particle filters.

Mye inspirasjon får jeg også gjennom seminarseriene her (som holder veldig høy kvalitet) samt deltagelse i et par lesegrupper, “Bayesian cake club” tar opp ulike artikler innen Bayesianske metoder. Her hopper vi litt fra emne til emne, og lærer dermed mye om temaer vi kanskje ellers ikke ville ha sett så mye på. “Variational methods” har sett på artikler med fokus på algoritmer for å tilnærme komplekse a posteriori fordelinger med enklere fordelinger, typisk produkt av uavhengige komponenter. Etter å ha brukt endel tid på dette, er jeg langt fra overbevist om at dette fungerer så godt i praksis (svært lite konkrete eksempler i dette feltet ennå). Nylig har vi også startet en “ $p \gg n$ ” lesegruppe.

Jeg deler ellers kontor med en annen skandinavisk gjest, Søren Højsgaard fra det jorbruksvitenskapelige fakultet ved Århus universitet. Som meg bruker han også mye tid på å analysere reelle datasett og vi bruker mye tid på å legge ut om våre frustrasjoner på at data aldri er så enkle som en får inntrykk av når de bliver presentert i metodiske artikler og metodene dermed ikke fungerer. Men samtidig genererer dette jo også ideer for forbedringer av metodene og forhåpentligvis kommer det noen artikler ut av det.

Foruten den faglige inspirasjonen som ligger i å komme til et nytt miljø (og som min romkamerat sier “ved å være i et miljø der de fleste er mer intelligente enn meg”) er ellers en viktig gevinst med et slikt opphold å kunne sitte i fred og ro over lenger tid for å drive med forskning. Det er svært behagelig å kunne slette masse mail om møteinnkallinger og administrative saker som en ellers bruker mye tid på.

Sosialt ligner miljøet i Bristol på mange andre universitetsmiljøer. Folk sitter mye bak sine lukkede dører og mange ser man kanskje kun når det er seminarer. Lunsj tar mange på kontoret, men heldigvis klarer Søren og jeg stort sett å dra med oss noen til “Howthorns” for en sandwich og en kopp kaffe. Etter seminarene (som de alltid arrangerer med to foredrag fredag ettermiddag) samles som regel endel av de fast ansatte (og av og til noen PhD studenter) til en liten pub-tur. Vi har ellers god kontakt med Søren og Christophe og deres familier utenfor jobb.

Den største utfordringen med oppholdet har vært tilpasning for våre to jenter Emilie på 9 og Louise på 4. Begge skulle starte på ny skole. Hvilken fikk vi først vite en uke før avreise, men vi var ekstremt heldige i og med at de kom inn på vårt førstevalg, en skole der også Christophe’s datter skulle begynne i parallell-klasse med Louise (de starter med reception class som 4-åringer der). For å venne dem litt til det engelske språk, dro vi over

en måned før skolen startet, men i praksis var de nærmest blanke da skolen startet. For Louise gikk dette svært greit. Vi hadde lært henne å si “do you want to play with me”, som hun med friskt mot prøvde, og selv etter et par klare “NO” fikk hun et positivt “yes” til slutt. Etter dette klarte hun fint å kommunisere på en blanding av norsk og noen få engelske ord og nå er hun nærmest flytende i engelsk og den eneste i familien med klar Bristol aksent. For Emilie har det vært adskillig hardere, men med en mor som har deltatt i skolearbeidet et par timer hver dag samt en veldig fin gjeng av klassekamerater har dette også godt seg til. For Ulla har det nok også tatt litt tid å vende seg til et nytt miljø, men med antydninger om at det hadde vært fint å være her et år til, har nok hun også begynnt å trives!

Ellers bruker vi fridager til å utforske England. Etter en start der vi først og fremst dro på turer i nærområdet (innkludert flere turer til Bath) har vi nå våget oss ut på litt lengre turer, bl.a. til nasjonalparken Dartmoor. Med nok en skoleferie igjen før avreise til Norge, håper vi å se enda litt mer av England.

Vismannen Confuseus om statistikk

Vismannen Confuceus, som ikke må forveksles med den kinesiske filosofen og vismannen Confucius (551 - 479 f.Kr.), har uttalt mye vist, og nesevist, om statistikere og statistikkfaget.

Noen eksempler, i ikke-tilfeldig rekkefølge, følger:

- Statisticians are always tossing coins but do not own many.
- Frequentists think that it is the thought that counts whereas the Bayesians count the thoughts.
- If the statistics of Bayesians is all about belief, why are they so sure of everything?
- Biostatisticians are always talking about power but they do not have any.
- Statisticians are more rigorous in interpreting data but physicians are more imaginative.
- A statistician can tell you how large your trial will be but you will not like the answer.
- Even if you have a significant relation with a statistician you may not find it relevant.
- Studying many subjects gives insight, but statistics brings significance.
- Do not worry about uniformly most powerful tests; they do not worry about you.
- Parallel trials meet at infinity but that is more patients than you can study.
- Do not give up hope, the statistician may be wrong.
- There is only one way to truth, but ANOVA often has many.
- How can deviation be standard?
- How to win friends and influence people? You took the wrong step by becoming a statistician.

Nå er det ikke bare Confuceus som har avstedkommet mer eller mindre vise, og nesevise, utsagn om statistikere og statistikk. Her er noen eksempler:

- En innadvendt statistiker ser på skoene sine når han/hun snakker med noen, en utadvendt statistiker ser på sine egne sko når vedkommende snakker med noen.
- Two men are flying in an air balloon. It is heavily clouded and they get completely lost. When they finally see an opening in the clouds, they head down to figure out where they are. On the ground, they see a man. From the balloon they shout down to the man on the ground: "Where are we?" The man looks up and replies: "You're in a balloon." The two men look at each other and say: "Must be a statistician. The answer is absolutely correct, but completely irrelevant".
- To call in the statistician after the experiment is done may be no more than asking him to perform a post mortem examination: he may be able to say what the experiment died of.

Sir Ronald Fisher (1890- 1962)

- In God we trust; all others have to bring their data.
W. Edwards Deming (1900-1993)

Hvis du kan tenke deg en hyggestund med flere av Confuseus sine dypsindige betraktninger om statistikkfaget, kan du klikke på lenken <http://www.senns.demon.co.uk/Confuseus.htm>. Husk imidlertid på at visdomsordet "alt med måte" gjelder også for kloke ord, liksom for god mat og andre nytelser. Det kan lønne seg å studere Confuseus sine skrifter i flere bolker.

Legg merke til at på slutten av dokumentet som fremkommer når du klikker på lenken nevnt ovenfor, er det en lenke til hjemmesiden til Stephen Senn ved University of Glasgow. Denne hjemmesiden er vel verdt å avlegge et besøk. Senn er en av de internasjonalt ledende innen medisinsk statistikk, og kanskje den ledende når det gjelder kryssover forsøk. Dessuten er han en god humorist.

Øystein Evandt

Erik Bølviken 60 år

Det er ikke til å tro, men det er sant! Professor Erik Bølviken ved Matematisk institutt, avd. C, UiO fylte 60 år 15. april.

Etter endt utdanning ved UiO jobbet Erik ved FFI før han dro til Universitetet i Tromsø i 1977. Dette var i en tid da dette universitetet var meget ungt, men hadde en statistikkgruppe som kunne måle seg med de beste i landet. For statistikkfaget generelt og for Erik spesielt var dette en viktig oppbyggingsperiode.

Det var også i denne perioden Eriks personlig syn på statistikkfaget begynte å modnes. Fra et utgangspunkt som klassisk orientert statistiker med betydelige teoretiske arbeider innen blant annet multiple sammenligninger, utviklet Erik en personlig stil og et syn der den matematiske delen av faget dempes ned til fordel for beregninger og effektiv bruk av datamaskin. Relevansbegrepet og den nære kobling til den konkrete og gode anvendelse av faget fikk etter hvert en dominerende plass i hans forskning. Teoretiske utlegninger uten et klart siktemål og en klar forankring i en god anvendelse er i dag fjernt for Erik.

I 1984 flyttet han til Oslo der han siden har jobbet ved UiO som førsteamanuensis, dosent og professor. Når det gjelder anvendelser vil vi spesielt nevne oljeleting, biologi, finans og aktuarfag. Erik ble tidlig engasjert i arbeid knyttet til oljeleting i Nordsjøen. Mye av hans bidrag innen dette feltet skjedde mens han hadde en bistilling ved Norsk Regnesentral. Etter dette fulgte en periode med biologiske anvendelser knyttet til økologi i samarbeid med blant annet Biologisk institutt, UiO. Dette samarbeidet varte fram til 1999 da Erik tok på seg et nytt og stort ansvar som hovedansvarlig for aktuarfaget ved UiO. Hans idé var å fornye aktuarfaget i samsvar med sin personlige profil og faglige holdning. Etter dette viktige valget har Erik lagt hele sin faglige energi og tyngde bak arbeidet med blant annet å forankre aktuarfag innenfor en videre finansfaglig ramme. Dette var et viktig valg og skritt i Eriks karriere som i ettertid har vist seg å være en stor suksess både for han selv personlig og ikke minst for avd. C og aktuarfagets utvikling i Norge.

Erik er en meget populær og god lærer og veileder, noe som manifesterer seg gjennom det store antall studenter som til enhver tid følger hans kurs og hans hovedfagsveiledning. Som et ledd i hans store innsats for aktuarfaget må også



nevnes hans kommende bokverk: "Actuarial and Financial Risk. How you make modelling and Monte Carlo work". Erik har meget gode kontakter i finansindustrien, blant annet gjennom sitt engasjement i Gabler & Partners, noe som også sikrer direkte anvendelser for hans arbeid. Han er dessuten en mye brukt ekspert i rettssaker.

Privat er Erik en meget kunnskapsrik og inspirerende samtalepartner. Han har mange interesser, der vi spesielt vil fremheve historie og utenrikspolitikk. Erik er også en meget habil bridgespiller med blant annet et norgesmesterskap og solide plasseringer i internasjonale konkurranser å se tilbake på. Dette er en interesse han fortsatt dyrker. Erik har også i perioder vært en ivrig tennisspiller, og han er glad i reiser og friluftsliv.

Tormod Næs og Bent Natvig

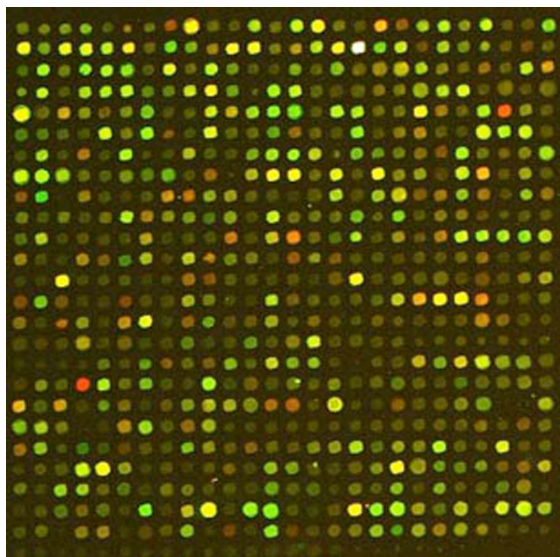
Presentasjon av doktoravhandling

Egil Ferkingstad, Statistics for Innovation (sfi)², Norsk Regnesentral. egil@nr.no

Den 13. februar forsvarte jeg doktoravhandlingen min, og jeg gir en kort presentasjon av den her. Tittelen er "Statistical methods for biological discovery: causal structures and multiple decisions". Det dreier seg altså om statistiske metoder innenfor biologi; for det meste (tre av fire artikler) genomikk. Metodologisk har avhandlingen to hovedtema: multippel hypotesetesting og kausal modellering. "Discovery" i tittelen peker på at en forkastet nullhypotese ofte kan ses på som en "oppdagelse", og det man ønsker å oppdage er gjerne årsakssammenhenger.

Mangfoldig testing

To av artiklene i avhandlingen dreier seg om multippel testing. Dette er et spesielt aktuelt problem i genomikk, men også i andre anvendelser som f.eks. hjerneskaning (brain-imaging) og astrofysikk. I biologi er multippel testing særlig viktig for mikromatrise-studier av genekspresjon. Selv om alle celler inneholder det samme genetiske materialet, er det stor forskjell i hvilke gener som er "skrudd av og på" i de ulike celletypene, dvs. hvilke gener som gir opphav til proteiner, og hvor stor mengde protein som blir produsert. Dette blir kalt genekspresjon, og med mikromatriser er det mulig å måle ekspresjonen av titusener av gener samtidig. (For ordens skyld: normalt måler man mellomproduktet mRNA, ikke protein direkte.)



Positiv, men falsk

Med denne teknologien kan man f.eks. sammenligne genekspresjon i kreftvev og normalt vev for hver av 30 000 gener. Dette kan si noe om hvilke gener som er involvert i kreftutvikling. Det statistiske problemet her er at vi for hvert gen har en hypotesetest der nullhypotesen er "ingen forskjell i ekspresjon mellom kreftvev og normalt vev". Om vi nå gjør 30 000 tester og forkaster alle nullhypoteser med $p < 0.05$, så vil det gi ca 1 500 signifikante resultater selv om alle nullhypotesene i virkeligheten er sanne.

En eller annen form for korreksjon for multippel testing er derfor nødvendig, og det finnes mange metoder for dette; de mest kjente er ulike varianter av "false discovery rates". En viktig input for de fleste multippel-testing-metodene (og en interessant størrelse i seg selv) er et estimat for hvor stor andel av nullhypotesene som faktisk er sanne. Artikkel 1 i avhandlingen,

"Estimating the proportion of true null hypotheses, with application to DNA microarray data", presenterer en ny metode for å estimere denne andelen. En simuleringsstudie viser at denne er mer presis enn tidligere foreslåtte metoder. Artikkel 3, "Unsupervised empirical Bayesian multiple testing with external covariates", viser hvordan man kan bruke tilgjengelig tilleggsinformasjon (kovariater) knyttet til hver test for å oppdage flere signifikante funn uten å øke antallet falske positive. For denne artikkelen vant jeg en pris: David P. Byar Young Investigator Award 2007 fra American Statistical Association, Biometrics Section.

Årsak og virkning

Den andre hoveddelen av avhandlingen handler om metoder for å ta i bruk informasjon om årsakssammenhenger i statistisk analyse. Tradisjonelt har statistikere vært svært forsiktige med å snakke om kausalitet, bortsett fra å påpeke at "kausalitet og korrelasjon ikke er det samme", men dette har endret seg noe i den senere tid.

På gjengrodde stier

Artikkel 2, "Dynamic path analysis — a new approach to analyzing time-dependent covariates", beskriver "dynamisk stianalyse", som er en metode for kausal modellering for overlevelsedata. I artikkelen analyseres et kjent datasett på pasienter med skrumplever, der protrombinnivå i blodet er en intern tidsavhengig kovariat: Behandling påvirker protrombinnivå, som igjen påvirker overlevelse. Samtidig kan behandling påvirke overlevelse direkte (eller gjennom variabler som vi ikke har målinger på). I det første tilfellet har vi en indirekte effekt (gjennom protrombin) på overlevelse; i det andre tilfellet en direkte effekt. Gjennom dynamisk stianalyse kan man skille fra hverandre direkte og indirekte effekter, og dermed få økt innsikt i mekanismen som genererer dataene.



Artikkel 4, "Indirect genomic effects on survival from gene expression data", viser hvordan dynamisk stianalyse kan anvendes på studiet av effekter av gener på overlevelse for kreftpasienter. En ny metode kombinerer data på pathways og genekspressjon for å studere om gener virker direkte på overlevelse, eller indirekte via andre gener, noe som kan være nyttig for utvikling av nye medisiner og personalisert behandling.

Livet videre

Siden 22. oktober har jeg vært ansatt som forsker ved Statistics for Innovation (sfi)², Norsk Regnesentral.

Selvpresentasjon Kristian Hindberg

Planlagt disputas: November 2010
 Institutt for matematikk og statistikk, UiT
 Veileder: Professor Fred Godtliebsen

Bakgrunnen min

Som en del andre statistikere så har jeg ikke en «standard» statistikkbakgrunn. Hovedfagsoppgaven min kalt «Flerstråledeteksjon i syntetisk aperture sonar-data» fra 2003 er avlagt ved Institutt for fysikk ved UiT med professor Alfred Hanssen som veileder. Jeg fikk deretter direkte jobb ved FFI og der jobbet jeg på et prosjekt innen forsvar mot ballistiske missiler. Dette viste seg etter hvert å ikke være drømmejobben, selv om det å kunne kalle seg «rakettforsker» var artig nok. Litt ut av ingenting så søkte jeg på og fikk et stipend fra Norsk romsenter for å gå på International Space University i Strasbourg i Frankrike.



Dette er en skole som tilbyr et ettårig masterprogram som berører alle tenkelige aspekter ved «space». Så nå kan jeg litt om romlovgivning, romfinans, biologi i rommet, banemekanikk osv. Som en del av skoleåret var jeg på et tre måneder langt «internship» ved Canada Centre for Remote Sensing i Ottawa sommeren 2006. Der fikk jeg ekstremt god hands-on-erfaring med å jobbe med og drive forskning med «synthetic aperture radar»-bilder, et tema jeg hadde gode teorikunnskaper innen og stor interesse for fra før av. Tiden i utlandet og Oslo var svært lærerik, men etter hvert ble savnet til hjembyen alt for stort for denne Tromsø-fantasten. Da det ble klart at jeg ble tilbudt en stipendiatstilling ved UiT i statistikk med professor Fred Godtliebsen som veileder, så var valget enkelt og jeg har ikke angret et sekund. Med min litt utradisjonelle bakgrunn har jeg måttet bruke en del tid på tette en del hull i «barnelærdomsstatistikken», men etter hvert har jeg kommet riktig så godt i gang synes jeg.

Vi her i Tromsø

Her i Tromsø er vi en relativt stor gruppe som samarbeider. Gruppen er sentrert rundt Godtliebsen og innbefatter folk fra NTNU, Nasjonalt Senter for Telemedisin (NST), Universitetssykehuset i Nord-Norge, Havforskningsinstituttet, Norsk Polarinstitut og Nofima (tidligere Fiskeriforskning), pluss utenlandske samarbeidspartnere fra USA, Tyskland, Finland, Taiwan og India.

Mange av disse aktørene er tilknyttet denne delvis uformelle gruppen via eVITA-prosjektet «The Power of Scale-Space Methods and Gaussian Markov Random Fields Applied in Climatology and Medicine» som er ledet av Godtliebsen. I prosjektet jobbes det med problemstillinger hvor vi kan koble Gaussiske Markov-felt (GMRF) opp til flerskalateknikker. For GMRF-delen samarbeider vi med professor Håvard Rue ved NTNU. Mer spesifikt søker vi å benytte Rues eminente C-bibliotek for superrask behandling av GMRF. I den forbindelse besøkte jeg Rue en måned ved NTNU i oktober-august 2007. I tillegg er flere i gruppen tilknyttet SFI-senteret Tromsø Telemedicine Laboratory (TTL) med NST som vertsinstitusjon (se TG fra august 2006).

(noen av) Mine prosjekter

Med utgangspunkt i Rue og Martinos «Integrated Nested Laplace Approximation» (INLA) har vi utviklet en «spekter-SiZer». Her brukes INLA til å approksimere likelihooden til et standard periodogram for jevnt samplede data over alle interessante skalaer, hvor vi antar en «random walk 2» som priormodell. Skalaen bestemmes av glattingsparameteren i RW2'en. Resultatet av algoritmen er et SiZer-lignende plott hvor statistisk signifikante frekvenstopper kan leses av simultant på tvers av alle skalaene. Metoden skal videreutvikles til også å håndtere ujevnt samplede data ved hjelp av Lomb-Scargle-periodogrammet. Verktøyet er ment brukt for å sannsynliggjøre at detekterte frekvenstopper er reelle. En kan si at jo flere skalaer en topp detekteres for, jo mer sannsynlig er det at frekvenstoppen er en reelle «feature» i dataene/prosessen.

Fremover er tanken at jeg skal jobbe med problemstillinger innen blodsukkernivå og blodkoaguleringssevne. I samarbeid med NST får vi nå tilgang til begge typer målinger fra forsøkspersoner. Blodsukkerdata er ikke noe nytt og er jobbet mye med, men vi håper å kunne kombinere flerskalateknikker og GMRF på en god måte for å kunne forstå blodsukkerets variasjoner bedre og for eksempel kunne skille blodsukkeravvik som skyldes infeksjoner og avvik som skyldes andre faktorer som mat, medisiner og trening. Så vidt vi vet er det ikke gjort noe tilsvarende studier som det vi ser for oss i forbindelse med blodkoaguleringssevnen. Normalt må pasienter regelmessig til legen for å kontrollere at blodkoaguleringssevnen ligger innenfor bestemte grenser. Fra daglig, men irregulært, samplede data søker vi å modellere blodtykkelsen. Det vil særlig være av interesse å se hvor store variasjoner pasientens blodtykkelse viser og om vi kan identifisere faktorer som bidrar til variasjonene. Gjennom NST er teknologien tilgjengelig til å gjennomføre en større undersøkelse med mange pasienter der målingene daglig blir overført via spesielle mobiltelefonheter.

Et annet spennende prosjekt som vi håper å komme i gang med snarlig handler om overvåking av smittsomme sykdommer og utbrudd av slike. Norge er i en særstilling i verden i den forstand at alle legekontorer og sykehus er koblet sammen og at det skjer en stadig toveis kommunikasjon. Idéen bygger på at ved alle legekonsultasjoner så markerer legen av for pasientens symptomer, og dette lagres så i den elektroniske journalen til pasienten. Per i dag benyttes ikke denne gigantiske datamengden (effektivt) til noen monitorering av befolkningens helsetilstand. Tanken er å benytte disse symptomdataene sammen med en demografisk modell til å tidlig oppdage at noe er «galt» med

helsetilstanden til befolkningen. Mulighetene er enorme, og et enkelt tenkt eksempel kan være forurensning av en drikkevannskilde. En slik forurensning kan føre til flere tilfeller av omgangssyke blant tilknyttede innbyggere. Det vil kunne ta tid før en slik oppblomstring merkes for den enkelte lege. Dette kan føre til at legen ikke finner det nødvendig å undersøke og teste pasienter så nøye som det trengs for å oppdage at omgangssyken skyldes noe annet enn en «vanlig» omgangssyke. Med vårt tenkte system vil en slik oppblomstring av omgangssyketilfeller i et område oppdages tidlig og systemet kan da sende ut et varsel til legene i området om å være observant på at det er unormalt mange med omgangssyke nå. Legen kan dermed gi en tettere oppfølging av pasientene med symptomene og deteksjonen av den bakenforliggende årsaken kan skje raskere. I et slikt system er det naturlig å ta i betraktning befolkningens bevegelsesmønster (jobbpendling f.eks.) og geografiske begrensninger. Her er tilgangen til pasientenes bostedspostnummer nøkkelen. Etter hvert ser vi for oss å kombinere dette med modeller for spredning av smittsomme sykdommer.

Rekrutteringsarbeid som pliktarbeid

Pliktarbeidet mitt ved UiT har så lagt bestått av ulike former for studentrekruttering. Som følge av å være den unge og kule(!) her på statistikkavdelingen er det jeg som har representert oss ved diverse skolebesøk både her på UiT og ute ved skoler. Dette er noe jeg tror er veldig nyttig og som passer meg veldig godt. Godtliebsen har i tillegg blitt tildelt penger fra eVITA til å gjennomføre et rekrutteringsprosjekt rettet mot ungdomsskoletrinnet. Håpet vårt er at vi kan lage et statistikkopplegg som skal hindre at ungdomsskoleelever mister interessen for matematikk/statistikk, en interesse vi føler ofte er sterkt tilstede ved inngangen til ungdomsskoletrinnet, men som for ofte blir borte der. Vi skal sammen med Grønnåsen skole i Tromsø teste ut et opplegg jeg har laget der vi lar elevene prøve å bruke Matlab til å løse oppgaver og utføre simuleringer. Valget av Matlab kontra gratisalternativer er nøye diskutert. Til slutt kom vi frem til at Matlabs ekstreme brukervennlighet og gode grafiske løsninger var avgjørende. Vi håper å teste dette ut før sommeren kommer, men som så mye annet, så tar ting tid. I tillegg har jeg vært noe involvert i UiTs storsatsing kalt Kaskjer.no. Dette er en internettrekrutteringsportal rettet mot tredjeklassinger ved videregående skoler i hele landet, hvor målet naturlig nok er å få flest mulig til å søke seg til UiT.

Utenlandsopphold i USA

Til høsten drar jeg til USA for et ti måneder langt utenlandsopphold. Jeg skal tilbringe høsten ved University of North Carolina sammen med Godtliebsen (som er der på sabbatsår) og professor Steve Marron. Om det går som vi tror så blir jeg deretter å tilbringe våren ved University of Washington i Seattle. Der vil jeg og en annen stipendiat fra Tromsø være gjester hos professor Donald Percival og planen er at vi der skal jobbe med flerskalateknikker og wavelets med anvendelse innen hyperspektrale bilder.

Til sommeren har jeg tenkt meg til Nordstat i Vilnius, så kanskje vi sees der! Ellers så storkoser jeg meg i mitt kjære Tromsø! Så er det bare å komme nordover i 2018!!!! :-D

Abelfinalen 2008 og Walkers algoritme

Det er ikke så ofte det er oppgaver med relevans for stokastiske simuleringer i finalen i Abelkonkurransen, men i år var det en.

Abelkonkurransen er en landsomfattende matematikkonkurranse for elever i videregående skole. 3700 elever deltok skoleåret 2007–2008, og 6. mars kom 22 av dem til finalen ved NTNU i Trondheim.

Oppgaven gikk essensielt ut på å bevise Walkers algoritme (Walker, 1977) for å generere observasjoner fra en generell diskret fordeling med verdier $1, 2, \dots, n$.

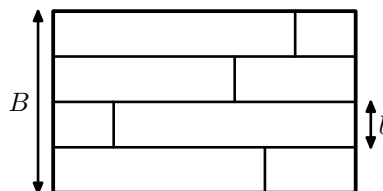
Metoden går ut på at et heltall x , $1 \leq x \leq n$, trekkes tilfeldig – hvert tall med sannsynlighet $1/n$. På forhånd (før første generering) er det beregnet heltall $y_1, y_2, \dots, y_n$, der $1 \leq y_i \leq n$, og sannsynligheter $q_1, q_2, \dots, q_n$, og x aksepteres som en observasjon med sannsynlighet q_x (som kan være lik én) og erstattes med sitt «alias» y_x med sannsynlighet $1 - q_x$. At det alltid fins slike y_i -er og q_i -er var det abelfinalistene måtte vise.

Algoritmen likner altså på en forkastningsmetode, men har den store fordel at en forkastning ikke er bortkastet, men gir en annen verdi. Walkers metode brukes i R-funksjonen `sample` (R Development Core Team, 2008).

Walker viste ikke at algoritmen alltid gir den ønskede fordelingen, men sjekket fra tilfelle til tilfelle. Beviset kom først i 1979 (Kronmal og Peterson, 1979).

Og her er oppgaven fra Abelfinalen – det skulle være en plankeoppgave i dobbel forstand, men det var faktisk den eneste oppgaven som ingen fikk full score på:

Vi skal legge planker på et golv som har bredde B i den retningen som er på tvers av plankene. Vi har n planker av bredde b , og B/b er et helt tall, og $nb \leq B$. Det er nok planker til å dekke golvet, men plankene kan ha forskjellige lengder. Vis at vi kan sage plankene på en slik måte at hver plankelengde på golvet høyst får én skjøt.



For å se oppgaven som en illustrasjon av Walkers metode, kan vi la $b = 1$,

$B = n$ og la golvet ha lengde $1/n$ fra venstre til høyre på figuren. Plankenes lengder svarer til punktsannsynlighetene i fordelingen, og det er nøyaktig nok plank til å dekke golvet. Vi kan tolke plankelengde til venstre for skjøten som q_i (eventuelt er $q_i = 1$ hvis lengden ikke har noen skjøt). Hvis vi nummererer de n plankene fra 1 til n før vi begynner å sage, svarer nummereringen av plankebitene til høyre for skjøtene til y_i -ene. Selve genereringen svarer til å trekke en plankelengde tilfeldig og så velge et punkt tilfeldig på plankelengden, og notere seg plankens nummer.

Løsningen av oppgaven kan finnes på `abelkonkurransen.no`.

Kronmal, R. A. og Peterson, A. V. 1979. On the alias method for generating random variables from a discrete distribution. *The American Statistician* (33) 214–8.

R Development Core Team. 2008. R: A language and environment for statistical computing. <http://www.R-project.org>

Walker, A. J. 1977. An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software* (3) 253–6.

Øyvind Bakke



**22ND NORDIC CONFERENCE
ON MATHEMATICAL STATISTICS**
16-19 JUNE 2008 VILNIUS, LITHUANIA



Nordstat 2008 blir arrangert i perioden mandag 16.juni- torsdag 19.juni i Vilnius, Litauen.

Konferansen har fire plenumsforedrag;

- Assisterende forskningssjef Kjersti Aas (Norsk Regnesentral):
"Pair-copula constructions of multiple dependence"
- Professor Hans Künsch (ETH Zürich):
"Biases and uncertainty in climate predictions"
- Professor Aad van der Vaar (Vrije Universiteit):
"Some frequentist results about posterior distributions on infinite-dimensional parameter spaces"
- Professor Willem van Zwet (Amsterdam University):
"A centenary of Student's"

Inviterte og påmeldte foredrag finnes innen følgende seksjoner:

Bioinformatics, Computational Statistics, Econometrics, Graphical Models, High Dimensional and Multivariate Methods, Industrial and Business Statistics, Nonparametric Methods, Observer Agreement, Reliability, Event History and Survival Analysis, Spatial and Spatio-temporal Modelling, Statistics in Finance, Survey Sampling and Social Statistics.

Programmet for konferansen er klart – se konferansens www-sider fra:

<http://www.nordstat2008.com>

Påmelding innen 1.mai koster 220 €, og senere 250 €.

Velkommen!



Pusleri nr. 24

Du bor i en by med kvadratisk gatestruktur, og vil besøke Per som bor m kvartaler øst og n kvartaler nord for deg. Alle gangfelt er lysregulert med vekslende grønt og rødt lys, og møter du et rødt lys i gangretningen, som er øst evt. nord, skifter du gangretning uten stopp til nord evt. øst, krysser gaten, og fortsetter i samme retning etter samme prinsipp, så lenge en slik mulighet fortsatt finnes. Anta at trafikklysene er usynkroniserte, og at rødt eller grønt i gangretningen er like sannsynlig. Hva er forventet antall nødvendige stopp for rødt lys inntil du er framme hos Per? Angi spesielt tilfellet $m=n$.

Løsning til: jostein.lillestol@nhh.no

Pusleri nr. 23 - løsning

Oppgaven var: For n uavhengige standardnormalfordelte variable, hva blir (evt. tilnærmet) korrelasjonen mellom gjennomsnittet $\bar{X}_n$ og medianen M_n ?

Hva kan sies om dette dersom vi fjerner antakelsen om normalitet?

Denne gang har vi fått en løsning fra Nils Lid Hjort, som har brakt fram mange flere fasetter ved problemet enn jeg kunne drømme om. Hans bidrag kommer i sin helhet etter mine kommentarer. Som dere vil se har vi begge funnet den samme enkle "bakpå konvolutten" løsning i det normale tilfellet, og jeg begrenser meg til noen korte refleksjoner rundt bruk av Basu's teorem i denne sammenheng. Les gjerne Nils sin løsning først!

Løsning (Lillestøl):

For uavhengige normalfordelte variable vet vi at $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ er uavhengig av $\bar{X}$, og dermed også ordningsobservatoren $(X_{(1)} - \bar{X}, X_{(2)} - \bar{X}, \dots, X_{(n)} - \bar{X})$ uavhengig av $\bar{X}$. Følgelig $\text{cov}(X_{(i)} - \bar{X}, \bar{X}) = \text{cov}(X_{(i)}, \bar{X}) - \text{cov}(\bar{X}, \bar{X}) = 0$, slik at $\text{cov}(X_{(i)}, \bar{X}) = \text{var}(\bar{X})$ for alle i .

Vi har derfor spesielt for medianen $\rho(M, \bar{X}) = \sigma(\bar{X}) / \sigma(M)$ (både for n jamn og odde).

For uavhengige standardnormalfordelte er* $\text{var}(M) \approx \pi / 2n$, slik at $\rho(M, \bar{X}) \approx \sqrt{2/\pi} = 0.798$.

Uavhengigheten kan blant annet fås fra Basu's teorem, fordi $\bar{X}$ er suffisient for forventningen μ og $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ er ansillær (fordeling som ikke avhenger av μ). Jeg tenkte så at det samme må gjelde for enhver lokasjonsfamilie som kan karakteriseres slik, men fant ingen slik familie, og med god grunn skal vi tro Nils. Imidlertid har vi for den dobbel-eksponensielle fordeling med lokasjon μ at M er suffisient og $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ ansillær, som ifølge Basu gir uavhengighet. På samme vis som ovenfor får vi da $\rho(M, \bar{X}) = \sigma(M) / \sigma(\bar{X})$, med asymptisk resultat $\rho(M, \bar{X}) \approx 1/\sqrt{2} = 0.707$, som Nils har bekreftet.

* Se for eksempel Sverdrup: Lov og tilfeldighet Bind I, s.149

Korrelasjonen mellom gjennomsnitt og median

Nils Lid Hjort, april 2008

Jostein Lillestøl spør i TG 1/2008 om vi kan finne korrelasjonen mellom empirisk gjennomsnitt og median, (i) når data stammer fra normalfordelingen, og (ii) utenfor normalitetsantagelsen. Dette uskyldig utseende problem er ganske riktig noe kinkig – og det er en grunn til at Jostein deler problemet i (i) og (ii), siden det man kan kalle «en relativt enkel løsning» er mulig kun for normalsituasjonen. For det generelle tilfellet finnes det riktig nok eksplisitte formler, men pene er de ikke, og numeriske beregninger må til, med integraler og summer. Bildet er penere når man lar sin sample-størrelse n bevege mot uendelig, som vi skal se. Vi trenger et minimum av notasjon: våre observasjoner $X_1, \dots, X_n$ er i.i.d. fra en tetthet $f(x)$, og har gjennomsnitt $\bar{X}$ og median M_n . Vi skal utlede en generell formel for grensekorrelasjonen mellom disse to størrelser, altså uttrykt ved f , men vi tar normalsituasjonen først.

For normalfordelte data: Her gjelder det at vektoren av $X_i - \bar{X}$ er stokastisk uavhengig av $\bar{X}$. Dette kommer av at kovarians null medfører uavhengighet for multinormalt fordelte variable. Men da er også $M_n - \bar{X}$ uavhengig av $\bar{X}$, hvorav følger at

$$\text{cov}(M_n, \bar{X}) = \text{Var } \bar{X} = \sigma^2/n. \quad (1)$$

Skriver vi variansen til M_n som τ_n^2/n får vi

$$\rho_n = \text{corr}(M_n, \bar{X}) = \frac{\sigma^2/n}{(\sigma/\sqrt{n})(\tau_n/\sqrt{n})} = \frac{\sigma}{\tau_n}. \quad (2)$$

(Om vi ikke visste det fra før, vet vi derfor nå at $\bar{X}$ er mer presis enn M_n .) Men det er kjent at M_n er asymptotisk normalfordelt med standardavvik $(\frac{1}{2}/f(\mu))/\sqrt{n}$, der μ er populasjonsmedianen. For normalfordelingen betyr dette varians $(\pi/2)\sigma^2/n$, altså at τ_n over vil konvergere mot $\sqrt{\pi/2}\sigma$. Korrelasjonen er derfor $\sqrt{2/\pi} = 0.7979$ for stor n .

Mer nøyaktig analyse må bruke den eksplisitte tettheten til M_n , som for tilfellet n oddetall, si $n = 2m + 1$, generelt er lik

$$\frac{(2m+1)!}{m!m!} F(x)^m \{1 - F(x)\}^m f(x),$$

der F er den kumulative. Dens eksakte varians blir

$$\frac{\tau_n^2}{n} = \frac{(2m+1)!}{m!m!} \int_0^1 \{F^{-1}(u) - F^{-1}(\frac{1}{2})\}^2 u^m (1-u)^m du.$$

Tar man seg bryderiet med å beregne dette integralet numerisk, oppdager man at approksimasjonen $\tau_n \doteq \sigma\sqrt{\pi/2}$ er meget god, selv for temmelig lave n .

Når kan knepet over benyttes? Vi kom rimelig kjapt frem til svaret $\sqrt{2/\pi}$ over, ved å bruke det hendige faktum at vektoren med $X_i - \bar{X}$ er uavhengig av $\bar{X}$. Det er dette

som medfører (1) og (2). Men grunnlaget for (1) og (2) holder *meget sjelden*, viser det seg. La for enkelhets skyld forventningen være null og la $M(t)$ være den momentgenererende funksjon for X_i . Uavhengigheten medfører da

$$\mathbb{E} \exp\{t(X_1 - \bar{X}) + t\bar{X}\} = \mathbb{E} \exp\{t(X_1 - \bar{X})\} \mathbb{E} \exp(t\bar{X})$$

for alle t , som gir

$$M(t) = M((1 - 1/n)t)M(-t/n)^{n-1}M(t/n)^n.$$

Sier vi for enkelhets skyld også at f er symmetrisk (en antagelse man kan kvitte seg med etterpå, med ytterligere analyse), får vi spesielt at $M(t) = M(t/2)^4$. Poenget er at dette etter noe analyse viser seg å holde utelukkende for tilfellet $M(t) = \exp(\frac{1}{2}\sigma^2 t^2)$, som karakteriserer normalfordelingen. – For enhver ikke-normal fordeling må vi altså ty til andre knep og verktøy enn (1) og (2).

Det generelle tilfellet. En helt presis formel, om vi for et konkret tilfelle virkelig trenger et eksakt tallsvar, må anvende noe slikt som

$$\text{cov}(\bar{X}, M_n) = n^{-1} \sum_{i=1}^n \text{cov}(X_{(i)}, X_{(m+1)}), \quad (3)$$

sammen med den eksakte simultantetthet for $(X_{(i)}, X_{(m+1)})$, der altså $X_{(i)}$ er den i -te største observasjon. Dette kan gjennomføres, men blir ikke særlig opplysende.

Alt blir så meget bedre om vi går til grensen. Sentralgrenseteoremet gir $\sqrt{n}(\bar{X}_n - \xi) \rightarrow_d A \sim N(0, \sigma^2)$ direkte, der ξ er forventningen, mens man finner det ovennevnte resultatet $\sqrt{n}(M_n - \mu) \rightarrow_d B \sim N(0, \tau^2)$ i diverse lærebøker, med $\tau = \frac{1}{2}/f(\mu)$. Det er derimot ikke så lett å finne det *simultane* resultatet vi nå trenger, av typen

$$\begin{pmatrix} \sqrt{n}(\bar{X}_n - \xi) \\ \sqrt{n}(M_n - \mu) \end{pmatrix} \rightarrow_d \begin{pmatrix} A \\ B \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}\right).$$

Vi kjenner A alene og B alene, men hvordan samvarierer de? Jostein har altså bedt oss om å finne ρ – der vi kjenner svaret for normalfordelingen, men ikke for andre f .

Det beste grepet her blir å se også gjennomsnittet som en operasjon utført på ordningsobservatoren, idet $\bar{X}_n = (1/n) \sum_{i=1}^n X_{(i)}$, og så anvende teori for ordningsobservatorprosessen. Man har

$$\sqrt{n}\{F_n^{-1}(t) - F^{-1}(t)\} \rightarrow_d q(t)W^0(t) \quad \text{over } (0, 1),$$

der F_n^{-1} er den empiriske kvantilprosessen, med $q(t) = (F^{-1})'(t) = 1/f(F^{-1}(t))$, og $W^0(\cdot)$ er en såkalt Brownsk bro over $[0, 1]$: en Gaußisk prosess med forventning null og kovarians $s(1-t)$ for $s \leq t$. Det følger at B kan ses som $q(\frac{1}{2})W^0(\frac{1}{2})$, mens

$$A = \int_0^1 q(t)W^0(t) dt = \int W^0(F(x)) dx.$$

Dette henger sammen med at $\bar{X}_n$ er nær nok $\int_0^1 F_n^{-1}(t) dt$ mens $\int_0^1 F^{-1}(t) dt$ er identisk med $\xi = \int xf(x) dx$. Svaret for B krever den milde antagelse at tettheten er positiv i medianen μ .

Nå har vi simultan kontroll over A og B , og kan beregne kovariansen

$$E AB = q(\tfrac{1}{2}) \left[\int_{-\infty}^{\mu} F(x)(1 - \tfrac{1}{2}) dx + \int_{\mu}^{\infty} \tfrac{1}{2} \{1 - F(x)\} dx \right]. \quad (4)$$

Svaret på Josteins spørsmål, for en helt generell fordeling, blir dermed

$$\rho = \frac{1}{\sigma} \left\{ - \int_{-\infty}^{\mu} x f(x) dx + \int_{\mu}^{\infty} x f(x) dx \right\},$$

der jeg har anvendt delvis integrasjon for å forenkle svaret. Når fordelingen er symmetrisk, la oss si rundt null, får man

$$\rho = 2 \int_0^{\infty} x f(x) dx / \left\{ 2 \int_0^{\infty} x^2 f(x) dx \right\}^{1/2}. \quad (5)$$

For normaltettheten blir svaret $\sqrt{2/\pi}$, som funnet tidligere.

Hvilke verdier kan korrelasjonen anta? Hvordan ser da denne ρ ut, for ulike alternative tettheter? Vi finner $1/\sqrt{2} = 0.7071$ for den doble eksponentialfordeling, $\sqrt{3}/2 = 0.8660$ for den uniforme fordeling, osv. Hvor lav kan den bli? Man kan tvinge (5) ned mot null (tilnærmet uavhengighet mellom gjennomsnitt og median, for store n), men essensielt bare på den noe uinteressante måten at man velger tetthet med stor σ , f.eks. en t -fordeling med antall frihetsgrader 2.01. Og hvor høy kan den bli? Bruker man Cauchy–Schwarz på $x f(x)^{1/2} \cdot f(x)^{1/2}$ finner man for det første at $\rho \leq 1$, som man vet fra før, men også at ρ når høyest når $x f(x)^{1/2}$ er omtrent proporsjonal med $f(x)^{1/2}$ overalt der det teller, som kan oversettes til at den symmetriske $f(x)$ bør ha mesteparten av sin masse sentrert nær ± 1 – men den må på den annen side også ha en positiv verdi i null (ellers har ikke medianen en grensefordeling). Man kan derfor eksperimentere litt med skarpe tri-miksturer av typen

$$f = \tfrac{1}{2}(1-p) N(-1, \sigma^2) + p N(0, \sigma_0^2) + \tfrac{1}{2}(1-p) N(1, \sigma^2),$$

med lave standardavvik σ og σ_0 og lav p . Da vil ρ komme opp mot sin teoretiske maksimalverdi 1. Her er imidlertid konvergensens $\rho_n \rightarrow \rho$ langsom.

Andre spesialtilfeller. En formel for den eksakte korrelasjon trenger altså både variansen til M_n og kovariansene i (3). Kun for noen ganske få tettheter kan man finne rimelig eksplisitte formler, og med mitt presumptivt øvede øye ser jeg bare to gode kandidater.

For den *eksponensielle modellen* (der $\bar{X}$ og M_n har to forskjellige siktemål) leder formel (4) til grensekorrelasjon $\log 2 = 0.6931$, men eksakte formler lar seg utlede med bakgrunn i representasjonen

$$X_{(i)} = E_1/n + E_2/(n-1) + \cdots + E_i/(n-i+1),$$

der E_i -ene er i.i.d. og standard eksponensielle; poenget er at mellomrommene $X_{(i)} - X_{(i-1)}$ er uavhengige akkurat for denne fordelingen. Dette leder med litt plunder til $\text{corr}(M_n, \bar{X}) = c_n/\tau_n$, der

$$\begin{aligned}\tau_n^2 &= n \text{Var } M_n = (2m+1) \left\{ \frac{1}{(2m+1)^2} + \frac{1}{(2m)^2} + \cdots + \frac{1}{(m+1)^2} \right\}, \\ c_n &= n \text{cov}(M_n, \bar{X}) = \sum_{i=1}^m \left\{ \frac{1}{(2m+1)^2} + \cdots + \frac{1}{(2m+2-i)^2} \right\} \\ &\quad + (m+1) \left\{ \frac{1}{(2m+1)^2} + \cdots + \frac{1}{(m+1)^2} \right\}.\end{aligned}$$

Det er nå en utmerket gymnastikkøvelse å vise at $\tau_n \rightarrow 1$ (ganske raskt) mens $c_n \rightarrow \log 2$ (en anelse langsommere).

For den *uniforme modellen* vet vi igjen grensefasitsvaret $\sqrt{3}/2$ (fra (5)), men kan supplere dette med eksakte formler for endelig n , via resultatene

$$\mathbb{E} X_{(i)} = \frac{i}{n+1} \quad \text{og} \quad \text{cov}(X_{(i)}, X_{(j)}) = \frac{1}{n+2} \frac{i}{n+1} \left(1 - \frac{j}{n+1}\right) \text{ for } i \leq j$$

(disse kan utledes direkte, eller via det hjelperesultat at $(X_{(i)}, X_{(j)} - X_{(i)}, 1 - X_{(j)})$ følger en Dirichlet-fordeling med parametre $(i, j-i, n-j+1)$). Dette leder til $\rho_n = c_n/(\sigma\tau_n)$, der $\sigma = \sqrt{1/12}$ er standardavviket til X_i , og der

$$\begin{aligned}\tau_n^2 &= n \text{Var } M_n = \frac{n}{n+2} \frac{1}{4}, \\ c_n &= \sum_{i=1}^n \text{cov}(X_{(i)}, X_{(m+1)}) = \frac{1}{n+2} \left\{ \sum_{i=1}^m p_i \frac{1}{2} + \frac{1}{2} \frac{1}{2} + \sum_{i=m+2}^n (1-p_i) \frac{1}{2} \right\} = \frac{1}{4} \frac{m+1}{n+2}.\end{aligned}$$

Sluttresultatet blir

$$\rho_n = \frac{\sqrt{3}}{2} \frac{n+1}{\sqrt{n(n+2)}}.$$

Her er $p_i = i/(n+1)$, og $n = 2m+1$ er altså odde.

PS: Jeg skrev først til Jostein i en mail at jeg hadde kommet frem til svaret $\sqrt{2/\pi}$ på den proverbiale «baksiden av en konvolutt» – men da brukte jeg det generelle resonnementet, og formelen (4), ikke dette knepet forbundet med (1) og (2) som altså gir regning som får plass på en enda mindre konvolutt, for normalsituasjonen. Det kom jeg først på etterpå ...

Stipendiat (statistikk / biostatistikk)

STILLING SOM STIPENDIAT I STATISTIKK/BIOSTATISTIKK

er ledig ved Matematisk institutt, Universitetet i Oslo med tilsetting fra 1. august 2008 og for en periode på 4 år med 25% undervisningsplikt. I henhold til gjeldende retningslinjer vil eventuell tid som forskningsstipendiat eller tilsetting i annen rekrutteringsstilling helt eller delvis bli fratrukket tilsettingstiden.

Stillingen hører inn under Det matematisk-naturvitenskapelige fakultet og er knyttet til forskningsgruppen i statistikk og biostatistikk ved Matematisk institutt. Forskningsgruppen i statistikk og biostatistikk arbeider med grunnleggende metodespørsmål i statistikkfaget og med problemer som springer ut av anvendelser, særlig i biologi, medisin og teknologiske og økonomiske risikovurderinger. Den som tilsettes må dokumentere faglige kvalifikasjoner innen statistikk/biostatistikk. Søkerne vil bli vurdert i forhold til gruppens forskningsvirksomhet. Se <http://www.math.uio.no/research/groups/StatisticsBiostatistics/index.shtml?intro> for en nærmere beskrivelse.

Aktuelle søkere må ha fullført utdanning på mastergrads-nivå i statistikk/biostatistikk eller et annet matematisk fag, eventuelt dokumentere at de har til hensikt å fullføre slik utdanning i løpet av vårsemesteret 2008. Stipendiatstillinger er utdanningsstillinger der hensikten er at stipendiaten skal ta en Ph.d.-grad. Søknad om opptak til Ph.d.-studiet skal være godkjent innen 3 måneder etter tilsetting. En orientering om Ph.D.-studiet ved Det matematisk-naturvitenskapelige fakultet finnes på

<http://www.uio.no/forskning/doktorgradsutdanning/phdgraden.html>

Søknaden må inneholde fullstendige opplysninger om søkerens utdanning, eksamenskarakterer (vedlegges), faglig virksomhet og relevant praksis. Eventuelle publikasjoner og andre arbeider skal legges ved søknaden, og søkere kan ikke regne med at det vil bli tatt hensyn til arbeider som ikke er vedlagt. Det vises for øvrig til gjeldende retningslinjer og utfyllende bestemmelser for tilsetting i stipendiatstillinger ved Universitetet i Oslo:

<http://www.uio.no/admhb/reglhb/personal/tilsettingvitenskapelig/ansettelsesvilkar/tilsetting-stipendiat.xml>

Universitetet i Oslo ønsker flere kvinner i vitenskapelige stillinger. Kvinner oppfordres til å søke. Universitetet i Oslo har en overtakelsesavtale for alle tilsatte, med formål å sikre rettighetene til forskningsresultater m.m.

Nærmere opplysninger om stillingen kan rettes til:
professor Ørnulf Borgan, tlf. 22855862, e-post: borgan@math.uio.no.

Ltr.: LR 20, Ltr. 43-48 (avh. av tjenesteansiennitet), SKO 1017.

Søknadsfrist: 1. juni 2008 REF. NR.: 2008/5006

Søknad, merket med stillingens referansenummer 2008/ 5006 sendes sammen med CV, vitnemål, attester og liste over evt. publiserte og upubliserte arbeider til:
Det matematisk-naturvitenskapelige fakultet, v/personalkonsulent Grete Andresen, Postboks 1032, Blindern, 0316 Oslo. Søknad og CV kan sendes pr E-post til:
grete.andresen@matnat.uio.no.

Nytt fra NTNU

Nye PhD

Sivilingeniør **Eirik Mo** forsvarte 31. januar 2008 sin avhandling *Nonlinear Stochastic Dynamics and Chaos by Numerical Path Integration*.

Tittel på prøveforelesning *Numerical Methods for Stochastic Differential Equations*.

Komité: Professor Igor Rychlik, Chalmers Institute of Technology, Sweden; Professor S. John Hogan, University of Bristol, UK; Professor Harald E. Krogstad, NTNU.

Arbeidet er utført ved Institutt for matematiske fag, NTNU, med Professor Arvid Næss som hovedveileder.

Sivilingeniør **Hugo Lewi Hammer** forsvarte 25. april 2008 sin avhandling *Topics in stochastic simulation, with an application to seismic inversion*.

Tittel på prøveforelesning *Sequential Monte Carlo for static and sequential problems - and when is it preferable to MCMC*.

Komité: Professor Nicholas Chopin, Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE), France; Professor Erik Bølviken, Department of Mathematics, University of Oslo; Professor Henning Omre, NTNU.

Arbeidet er utført ved Institutt for matematiske fag, NTNU, med førsteamanuensis Håkon Tjelmeland som hovedveileder.

Diverse

Studieretningen 'Industriell Matematikk', som utdanner de fleste av NTNU sine master-studenter i statistikk blir ranket som nummer 2 av alle NTNU studiene i en reportasje gjort av Teknisk Ukeblad. (<http://www.tu.no/karriere/article142291.ece>) Score veies fra tre kriterier som er 'Faglig nivå på studiet', 'Faglig nivå på personalet', og 'Pedagogisk nivå på personalet'. 'Industriell Matematikk' er kun slått av 'Materialteknologi', med 'Teknisk Kybernetikk' på tredje plass.

Nytt fra UIB

Uteksaminerte kandidater

Trond Ove Gangsøy: Chain Ladder Metoden og Macks modell.

Veileder: Trygve Nilsen

Diverse

Jostein Paulsen har permisjon i vår semesteret og foreleser for studentene på finansprogrammet ved University of Chicago.

Nytt fra Avdeling C, Mat-nat fakultet, UiO

Nye PhD-kandidater

Daniel Berg forsvarte 14. mars 2008 sin avhandling:

Statistical analysis of credit risk – topics in default and dependence modelling
ved Det matematisk-naturvitenskapelige fakultet

Prøveforelesning: *Using and selecting among copulae: frequentist and Bayesian perspectives*

Veiledere: **Fred Espen Benth**, **Xeni Kristine Dimakos** og **Kjersti Aas**

Ida Scheel forsvarte 2. april 2008 sin avhandling:

Influence of incomplete knowledge in inference: cases in the life sciences
ved Det matematisk-naturvitenskapelige fakultet

Prøveforelesning: *Computations and applications of Bayesian networks in the life sciences*

Veiledere: **Arnoldo Frigessi**, **Ingrid K. Glad** og **Nils Lid Hjort**

Ny Master:

Hongchuan Xia: *The Swedish ALM pension system under the solvency II regime*

Veileder: **Frank N. Proske**

Tekniske rapporter:

Pål N. Henriksen: *Lognormal moment matching and pricing of basket options.*

Fred Espen Benth, **Pål N. Henriksen**: *Pricing of basket options using univariate normal inverse Gaussian approximations.*

Ansettelses:

Ida Scheel er tilsatt som postdoktor i statistikk.

Permisjoner:

Geir Storvik og **Anders Rygh Swensen** har forskningstermin høsten 2007 og våren 2008.

Nytt fra Avdeling for biostatistikk, Med. fakultet, UiO

Avhandlinger

Johan Fosen forsvarte 7 desember 2007 sin avhandling *Internal time-dependent covariates in event history analysis* for graden ph.d.

Tittel på prøveforelesningen var *EM-algoritmen og anvendelse i forløpsanalyse*.

Komit : Lektor Torben Martinussen, Den Kgl. Veterin r- og Landboh jskole, Danmark, Forsker Per Henrik Zahl, Nasjonalt folkehelseinstitutt, Professor Eva Skovlund, Universitetet i Oslo.

Veiledere har v rt Odd Aalen, Avd. for biostatistikk og  rnulf Borgan, Matematisk Institutt, Universitetet i Oslo.

Harald Fekj r forsvarte 3. januar 2008 sin avhandling *Modelling breast cancer incidence, progression and screening test sensitivity using screening data* for graden ph.d.

Tittel p  pr veforelesningen var *Favourable and adverse effects of mammographic screening in Norway*.

Komit : Professor J. Dik. F. Habbema, Departement of Public Health, Erasmus MC, Nederland, Professor Giske Ursin, Universitetet i Oslo, Forsker Randi Selmer, Nasjonalt folkehelseinstitutt.

Veiledere har v rt Steinar Tretli, Kreftregisteret, Lars Vatten og Bo Lindqvist, NTNU, Odd Aalen, Avd. for biostatistikk.

Egil Ferkingstad forsvarte 13. februar sin avhandling *Statistical methods for biological discovery: causal structures and multiple decisions* for graden ph.d.

Tittel p  pr veforelesningen var *A critical review of competing risk models*.

Komit : Professor Hans van Houwelingen, Leiden University Medical Center, Nederland, Professor Peter M ller, The University of Texas/ MD Anderson Cancer Center, US Professor Knut Liest l, Universitetet i Oslo

Veiledere har v rt Arnolando Frigessi og Odd Aalen, Avd. for biostatistikk.

Nyansettelser

Birgitte Freiesleben de Blasio er ansatt i en midlertidig f rsteamanuensis stilling.

Diverse

Professor Arnolando Frigessi er tatt opp som medlem i Det Norske Videnskaps-Akademi.

Nytt fra Universitetet i Tromsø

Ny PhD

Christina Alsvik Pedersen forsvarte 3. mars sin avhandling: *Optical Properties of Snow and Sea Ice - Field Measurements, Parametrization Schemes and Validation*.

Tittel på prøveforelesning: *Evidence, mechanisms and impacts of accelerated Arctic warming*

Komit : Richard Essery (University of Edinburgh), Alf Harbitz (Havforskningsinstituttet) og Sigrunn Holbek S rbye (UiT ).

Veiledere: Jan-Gunnar Winther (Norsk Polarinstitutt) og Fred Godtliebsen (UiT ).

STATISTISK FORENING I BERGEN

Opprettet 15/12 1983

Styret 2007

Styreleder	Grethe Albrektsen	Seksjon for epidemiologi og medisinsk statistikk, ISF, UiB	Tlf. 55 58 61 58
Styremedlem	Ågot Irgens	Yrkesmedisinsk avdeling, Haukeland Universitetssykehus	Tlf. 55 97 38 82
Kasserer	Terje Schaathun	Vital forsikring ASA	Tlf. 934 07 037
Sekretær	Astrid Lunde	Nasjonalt Folkehelseinstitutt, avd. Bergen	Tlf. 55 58 61 84

Årsberetning, 2007

Årsmøtet for 2006 ble holdt 20. februar 2007 og møtet godkjente årsberetning og regnskap for 2006. Grethe Albrektsen ble gjenvalgt som formann, og fullførte dermed styreperioden 2006-2007 (tok gjenvalg kun for ett år i 2006). Astrid Lunde ble valgt som ny sekretær, Ågot Irgens ble valgt som nytt styremedlem. Kasserer stod ikke på valg. Bård Støve ble valgt som ny revisor. Styret for 2007 har dermed bestått av følgende personer:

Styreleder:	Grethe Albrektsen	Valgt for 2006-2007
Sekretær:	Astrid Lunde	Valgt for 2007-2008
Kasserer:	Terje Schaathun	Valgt for 2006-2007
Styremedlem:	Ågot Irgens	Valgt for 2007-2008

Det har vært avholdt to medlemsmøter i 2007, begge i vårsemesteret. Tidspunkt og tema for foredragene var:

20.02.07 Tittel: 'Patient allocations according to circumstances and preferences: Modelling based on the Norwegian patient list system'

v/ professor Jostein Lillestøl, Norges Handelshøyskole

20.03.07 Tittel: 'Comparing and combining effort and catch estimates from aerial-access designs, with applications to a large-scale angler survey in the Delaware River'

v/ forsker Jon Helge Vølstad, Versar Inc. i Colombia, USA

STATISTISK FORENING I BERGEN

Opprettet 15/12 1983

Styret 2007

Styreleder	Grethe Albrechtsen	Seksjon for epidemiologi og medisinsk statistikk, ISF, UiB	Tlf. 55 58 61 58
Styremedlem	Ågot Irgens	Yrkesmedisinsk avdeling, Haukeland Universitetssykehus	Tlf. 55 97 38 82
Kasserer	Terje Schaathun	Vital forsikring ASA	Tlf. 934 07 037
Sekretær	Astrid Lunde	Nasjonalt Folkehelseinstitutt, avd. Bergen	Tlf. 55 58 61 84

Referat fra årsmøte 2007

Møtet ble avholdt torsdag 27.mars 2008, kl. 19:00 på institutt for samfunnsmedisinske fag, Kalfarveien 31, i forbindelse med første møte i lokalforeningen 2008. Det var 12 tilstede på møte, men ikke alle var medlemmer. Referat i.h.t. sakslisten :

- Sakslisten ble godkjent
- Grethe Albrechtsen meldte seg og ble valgt som møteleder og referent
- Årsberetning ble gjennomgått og godkjent; det har vært avholdt 2 møter i 2007
- Regnskap ble gjennomgått og godkjent etter kommentar om trykkfeil i årstall for renteinntekter; saldo per 31.12.07 var 22 996,64 kroner
- VALG: Styremedlem (Ågot Irgens), sekretær (Astrid Lunde) og revisor (Bård Støve) stod ikke på valg, Terje Schaathun stilte til gjenvalg som kasserer, men vi manglet kandidat til styrevervet; ingen enkeltpersoner har fått direkte forespørsel dette året.

Det ble foreslått å beholde et 'sovende' styret uten styreleder, og samtidig opprette en valgkomite som kunne lete aktivt etter mulige kandidater. Dette synes å være en bedre løsning enn å legge ned foreningen. Til neste år er det trolig Bergen som må arrangere nasjonalt årsmøte og vi trenger et styre til den tid.

Valgkomité: Ågot Irgens, Terje Schaathun og Hans Julius Skaug meldte seg frivillig som kandidater til valgkomite. I tillegg utnevnte møtet Bård Støve som medlem i valgkomite; da har vi representater med kontakter i ulike miljøer. Valgkomite fikk i oppgave å finne en kandidat som kunne bli presentert og valgt på neste møte i lokalforeningen, fortrinnsvis iløpet av perioden 10.-20. juni.

Vi fikk en litt flytende overgang til saker under 'Eventuelt', men alle kandidater (i mangel av motkandidater) må ansees som valgt da møtet ble avsluttet.

- Eventuelt: Mulige tema for framtidige møter ble diskutert, bl.a. kom følgende forslag opp
 - Statistiske modeller for klimaendringer (Stordal, med invitert diskutant, jmf. møte i Oslo)
 - Google-søk: statistiske prinsipper for effektive søkemotorer (Google har kontor i Norge)
 - presentasjon av studentoppgaver

REFERENT : Grethe Albrechtsen

Lokalforeningene i Norsk Statistisk Forening

Styret i Avdeling Oslo (2007-)

<http://www.math.uio.no/div/nsf/oslo/>



Leder: Ingunn Fride Tvette , Norsk Regnesentral (bildet)
 Sekretær: Øyvind Langsrud, RBUP Øst og Sør og NKVTS
 Kasserer: Tor Cato Gjerdingen, Storebrand
 Web-ansvarlig: Ståle Nygård, Avd. C, Matematisk
 Institutt, UiO
 Styremedlem: Hege Marie Bøvelstad, Avd. C, Matematiske
 Institutt, UiO
 Styremedlem: Anders Løland, Norsk Regnesentral

Styret i Avdeling Bergen (2008-2009)



Styreleder: Ingen valgt, se referat fra årsmøtet
 27.03.2008.
 Sekretær: Astrid Lunde,
 Nasjonalt Folkehelseinstitutt, avd Bergen
 Kasserer: Terje Schaathun, Vital forsikring ASA
 Styremedlem: Ågot Irgens,
 Yrkesmedisinsk avdeling, Haukeland
 Universitetssykehus

Styret i Avdeling Trondheim (2007-2009)

<http://www.math.ntnu.no/tsf/>



Leder: Stian Lydersen, Det medisinske fakultet,
 NTNU (bildet)
 Sekretær: Rupali Akerkar, Institutt for matematiske
 fag, NTNU.
 Kasserer: Jo Eidsvik, Institutt for matematiske fag,
 NTNU
 Styremedlem: Ola Eirik Mo, SONOWAND AS.

RETUR:

*Norsk Statistisk Forening
v/Hilde Grude Borgos
Schlumberger Stavanger Research
Postboks 8013
4068 Stavanger*

B

