



ISSN 0803-8953

Tilfeldig gang

Nr. 1, årgang 32

Mars 2015

Utgitt av Norsk statistisk forening

Redaksjonelt

Fra lederen	1
Fra redaksjonen	2

Leserbidrag

Bannlysning, bruk eller misbruk av p-verdier innen anvendt forskning?	3
NHSTP i statistisk genomikk – andelen falske funn og reproduserbar forskning	4
Statistikere mot atomvåpen	8

Bokanmeldelse

R by Example	9
------------------------	---

Tankenøtter

Løsning på pusleri nummer 41	10
Pusleri nummer 42	11

Møter og konferanser

Kurs i Mixed Effects Models	12
30-årsjubileum for Statistisk forening i Bergen	13

TG for 25 år siden

Om uttrekning av meddommere enda en gang.	14
---	----

Meldinger

Nytt fra seksjonene 2 og 3 ved Matematisk institutt, Universitetet i Oslo	16
Nytt fra Norges teknisk-naturvitenskapelige universitet	17
Nytt fra Avdeling for biostatistikk, Det medisinske fa- kultet, Universitetet i Oslo	17
Nytt fra Universitetet i Bergen	17

Annonseringer

3 stipendiatstillinger i statistikk ved Mat.Inst. UiO	19
Ledige stillinger i Statistisk sentralbyrå	19
2 stipendiatstillinger i biostatistikk ved UiO	19
Det 18. norske statistikarmøtet	20
IAA colloquium in Oslo	22
NOBIM 2015	23

Scandinavian Journal of Statistics

Innholdsliste desember 2014	24
Innholdsliste mars 2015	26

Informasjon

Har du husket å betale kontingenten i år?	28
---	----

Fra lederen

Ragnar Bang Huseby | Oslo

Forberedelsene til årets norske statistikermøte er godt i gang. Bergensavdelingen har valgt å holde konferansen på Solstrand i Os utenfor Bergen. Det ser ut til å være et flott sted, og arrangements-komiteén reklamerer med adgang til stedets velværeavdeling. Vi regner med god deltagelse fra Bergensmiljøet, men håper også at mange fra andre deler av landet vil ta turen til Bergensområdet. Utdeling av Sverdrup-pris og foredrag fra prisvinnerne vil være blant høydepunktene. I tillegg vil vi gjerne ha foredrag fra andre bidragsytere. Dette ser jeg virkelig frem til.

Under konferansen vil foreningen holde sitt årsmøte. Vi håper dette blir en fin anledning for våre medlemmer til å få luftet sine meninger. Hvis noen har noe på hjertet, er det naturligvis også mulig å ta kontakt med meg før møtet.

Vi forventer at det kan bli diskusjon angående TG. Skal vi nøye oss med en digital versjon, eller vil vi ha tilbake papirutgaven? Det er sikkert flere enn de som kommer på årsmøtet, som har en mening om dette. Derfor har styret og redaksjonen laget en spørreundersøkelse. Jo flere som besvarer spørsmålene, desto bedre grunnlag vil vi ha for eventuelle beslutninger.



Fra redaksjonen

Turid Follestad | Trondheim

I TG for mars 2014 hadde vi ein introduksjon til eit notat av Jostein Lillestøl med ein presentasjon og diskusjon av ulike paradigmer i statistikken. Dette er framleis tilgjengelig på foreninga sine nettsider. Valg av tilnærming for ein statistisk analyse kan vere gjenstand for diskusjon, men like viktig som valg av metode er korleis resultata frå ein statistisk analyse blir tolka og presentert. Eit stadig tilbakevendande tema har vore bruk og misbruk av p-verdiar, kanskje særleg innan medisinsk forskning. Tidsskriftet «Basic and Applied Social Psychology» har tatt konsekvensen av desse utfordringane på sin måte - ut i det ekstreme vil nokon seie - ved å beslutte at dei ikkje lenger tar inn artiklar som innheld p-verdiar og konfidensintervall. Bayesianske metoder vil bli vurderte frå sak til sak. Dette har naturleg nok skapt diskusjon, og er eksempelvis kommentert av American Statistical Association og i Nature. Denne diskusjonen vil vi tru også er av interesse for TG sine lesarar, og redaksjonen har utfordra Stian Lydersen og Mette Langaas til å kommentere problemstillinga i dette nummeret.

ASA

Nature

Frå 2014 gjekk Norsk Statistisk Forening over til kun digital publisering av TG på foreninga sine nettsider. For nokre er dette ein naturleg konsekvens av den digitale utviklinga, mens andre saknar noko meir handfast å bla i, som ei motvekt til alt anna som kun er tilgjengeleg via ein skjerm. Som kommentert i spalta frå leiaren av foreninga, så vil publiseringsform for TG bli evaluert og diskutert på årsmøtet i foreninga i juni, og som grunnlag for evalueringa er det utforma eit uformelt spørreskjema som du svarer på ved å gå inn på følgjande nettside:

<https://skjemaker.app.uib.no/view.php?id=1093682>

Redaksjonen oppfordrar så mange som mogleg til å svare på undersøkinga, og dermed bidra til eit tilstrekkelig stort og representativt utvalg!

Til slutt: Har du irritert deg over at heisa «alltid» synest å vere på veg opp når du skal ned, eller omvendt? Om det kan hjelpe å rekne litt på det, så kan du supplere påskenøttene med denne utgåva av TG si «tankenøtt» med nettopp heiser som tema. Redaksjonen ønskjer alle ei god påske!

Redaksjon

Turid Follestad

Bo Lindqvist

Petter Arnesen

Bannlysning, bruk eller misbruk av p-verdier innen anvendt forskning?

Stian Lydersen | Trondheim

Redaktøren av «Basic and Applied Social Psychology» (BASP) har ettertrykkelig bannlyst p-verdier i tidsskriftet: "... the null hypothesis significance testing procedure (NHSTP) is invalid. ... From now on, BASP is banning the NHSTP. ... confidence intervals also are banned from BASP. ... Bayesian procedures are neither required nor banned from BASP". Her kastes altså all frekventistisk statistisk inferens ut av vinduet!

Ikke bare bannlyser han p-verdier, men også konfidensintervaller! Redaktøren vil neppe få mye støtte for dette. Imidlertid: Fremdeles finner jeg en overdreven vekt på p-verdier i publikasjoner innen medisin og atferdsfag. Jeg opplever dette i samarbeid med forskere, når jeg er statistisk fagfelle (statistical reviewer) for et medisinsk tidsskrift, og i enkelte lærebøker i anvendt statistikk. En kollega ved det medisinske fakultet fortalte at et tidsskrift hadde refusert hennes manuskript fordi p-verdien var 0.051. Om p-verdien hadde vært 0.049, ville manuskriptet blitt akseptert!

Debatten om hvilken vekt man skal legge på p-verdier og på konfidensintervall har pågått i mange tiår. Overdreven bruk av p-verdier er godt beskrevet i en nylig artikkel i Nature.

I 1978 møttes redaktørene for 14 medisinske tidsskrift i Vancouver og utarbeidet den første versjonen av retningslinjene for publikasjoner i deres tidsskrift. Disse inkluderer noen av verdens mest prestisjetunge tidsskrift som New England Journal of Medicine, the Lancet og British Medical Journal, og faktisk også de nordiske Tidsskrift for Den Norske Lægeforening, og Ugeskrift for Læger. Avsnittet om statistikk i retningslinjene er nå som følger:

"When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size and precision of estimates."

Disse retningslinjene blir i dag fulgt av over 500 medisinske tidsskrift. Her presiseres at konfidensintervall er generelt viktigere enn p-verdier. Tradisjonelt sett synes forskere innen psykologi og atferdsfag å ha vært enda mer fokusert på p-verdier enn forskere innen medisin. Men også her har det vært en endring i samme retning. "APA-manualen" (Publication

Trafimow, D. & Marks, M. 2015. Editorial. Basic and Applied Social Psychology, 37, 1-2

Editorial 2014. Number crunch. The correct use of statistics is not just good for science — it is essential. Nature, 506, (7487) 131-132 og

Nuzzo, R. 2014. Statistical Errors. Nature, 506, (7487) 150-152

International Committee of Medical Journal Editors 2014. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

Manual of the American Psychological Association) har en liknende status inne psykologi som “Vancouver-retningslinjene” har innen medisin. Gjeldende versjon (sjette utgave, 2010, side 34) sier: «The inclusion of confidence interval ... can be an extremely effective way of reporting results. ... they are, in general, the best reporting strategy.”

Den toneangivende epidemiologen Kenneth Rothman grunnla tidsskriftet “Epidemiology” i 1990. I dette tidsskriftet var p-verdier sterk uønsket i noen år, men ser ut til å være noe rehabilitert igjen. Ved å søke på «confidence interval p-value» i dette tidsskriftet finner man mange interessante innlegg om dette emnet. Blant annet diskuteres den såkalte konfidensintervallfunksjonen. Denne angir p-verdien som funksjon av nullhypotese-verdien for den aktuelle parameteren, se for eksempel link i marginen. Denne funksjonen viser også konfidensintervall som funksjon av konfidenskoeffisienten.

Jeg har vært statistisk fagfelle på 200 manuskript til et medisinsk tidsskrift, og finner ofte grunn til å kommentere feil eller svakheter knyttet til bruk av p-verdier . Dette kan være at det rapporteres bare om et resultat er signifikant ($p < 0.05$) uten å oppgi faktisk p-verdi, det kan være å ikke oppgi konfidensintervall, eller, aller verst, uten å oppgi estimat for den relevante (parameteren).

Jeg vil avslutte med å illustrere viktigheten av å rapportere konfidensintervall: Effekten av mammografi er noe omdiskutert, da forskjellige studier har vist forskjellige resultater. En studie publisert i British Medical Journal er basert på data fra 15 193 034 personår, i perioden 1986 til 2009 da mammografi ble gradvis innført i Norge. Den relative dødelighetsraten for brystkreft for de som var invitert til mammografi, var 0.72 (95% konfidensintervall 0.64 til 0.79). Forfatterne oppgir fornuftig nok ikke p-verdi, som er ganske irrelevant i denne sammenheng. P-verdien angir bare om effekten er statistisk signifikant, og er en funksjon av effektstørrelsen og utvalgsstørrelsen. Det som er relevant, er hvorvidt effekten er klinisk signifikant, altså stor nok til at man vil fortsette å gjennomføre et program som er ressurskrevende.

Kulldorff, M., Graubard, B., & Velie, E. 1999. The P-value and P-value function. *Epidemiology*, 10, (3) 345-347

Lydersen, S. 2015. Statistical review: frequently given comments. *Ann.Rheum.Dis.*, 74, (2) 323-325

Weedon-Fekjær, H., Romundstad, P.R., & Vatten, L.J. 2014. Modern mammography screening and breast cancer mortality: population study. *BMJ*, 348:g3701

NHSTP i statistisk genomikk – andelen falske funn og reproduserbar forskning

Mette Langaas | Perth

Tenk, nå har jeg fått lære en ny forkortelse: NHSTP, «null hypothesis significance testing procedure» . Kanskje det er det jeg skal skrive som klassifikasjon og nøkkelord på alle mine fremtidige arbeider. Ja, fordi det stemmer jo – i 80% av min forskningstid jobber jeg mot en NHSTP-tankegang. Jeg passer på at p-verdiene jeg regner ut holder type I-nivået sitt for alle

Trafimow, D. & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37, 1–2.

mulige signifikansnivå (ikke bare 0.05), og spesielt jobber jeg med å lage metoder som best mulig kontrollerer type I-nivået (FWER, family-wise error rate, og FDR, false discovery rate) når jeg tester mange nullhypoteser. Hvis NHSTP-tankegangen er bannlyst, så betyr det vel egentlig at alt jeg gjør også bør være bannlyst.

Så, hvorfor er det slikt fokus på NHSTP, og hvorfor er NHSTP noe vi ikke vil holde på med?

Stian Lydersen forteller i sitt diskusjonsinnlegg i dette nummer av TG om Vancouver-retningslinjene, der det advares mot å kun rapportere p-verdi ved analyse av medisinske data. Fokus bør være på informasjon om effektstørrelse og tilhørende usikkerhet, og så kan p-verdi være interessant tilleggsinformasjon. Jeg vet ikke om noen som bestrider denne strategien, og dette er det vi lærer bort til studenter i statistikk, uavhengig av studentens fagfelt.

Men, i noen tilfeller er det naturlig at NHSTP får mer. Mitt fagfelt er statistisk genomikk. Genetikk og DNA er i kjernen av faget, og så skiller genomikk seg fra genetikk ved at det ofte er mer enn ett gen eller én genetisk markør som studeres. Forskningen som gjøres er i de fleste tilfellene hypotese-genererende og multiplisitet er et viktig begrep. Da jeg startet å arbeide innen statistisk genomikk i 2000, var det mest populært å utvikle statistiske metoder for å analysere genuttryksdata fra microarray-forsøk (det er fremdeles populært), der det var i størrelsesorden ti tusen ulike cDNA-prober som ble studert, og man lurte på hvilke av disse som var påvirket av en gitt behandling. Videre kom genome-wide association studies (GWA-studier) med hundretusen genetiske markører som man lurte på om kanskje hadde en assosiasjon til sykdom eller fenotype. Nå er epigenetikk og sekvensdata i vinden, og med 3 milliarder basepar i det humane genom blir det trolig ikke mindre fokus på multiplisitet fremover.

Felles for alle disse typene av data er at forskningen er rettet mot å finne frem til lite sett av cDNA-prober eller genetiske markører som man tror er av interesse. Det vil si at dette er et første steg i en lengre analysevei. Forskerne ønsker en liste med navn på cDNA-prober eller genetiske markører som de kan studere i detalj og jobbe videre med i sin forskning. En slik liste kan gjerne lages ved å sortere på effektstørrelse, eller på en testobservator, eller på p-verdi. Rekkefølgen på en liste sortert på testobservator er i mange tilfeller den samme som en liste sortert på p-verdi. Noen forskningsgrupper setter deretter en cut-off på denne listen slik at de har en håndterbar liste (de klarer kanskje å jobbe videre med 10 eller 100 funn), og andre vil jobbe med å finne mønster i lengre lister. Dermed er fokus endret fra å være spesifikt på hver cDNA-probe eller hver genetiske markør, til å være settet av cDNA-prober eller genetiske markører som man ønsker å jobbe videre med. Det er

innen dette fagfeltet vanlig å knytte en estimert verdi for FWER eller FDR til en slik liste. For genuttrykksanalyser forventer man ofte at mange cDNA-prober har endret aktivitet på grunn av en behandling, og derfor har man lyst til å oppgi listens estimerte FDR – hvor mange falske positive funn man forventer å finne på denne listen. For GWA-studier forventer man ikke at mange genetiske markører er assosiert med en sykdom, og dermed ønsker man heller å rapportere estimert FWER – sannsynligheten for minst ett falsk positivt funn på listen. Det vil være svært vanskelig å få publisert en artikkel innen genomikk uten at et estimat av FWER eller FDR er oppgitt. I mange tidsskrifter er det også krav til at (kvalitetssikrede) rådata til studien som publiseres ligger tilgjengelig på offentlige repositorier (dette gjelder gjerne genuttrykksstudier), slik at andre forskere kan reprodusere funn og studere data, mens data på genetiske markører hos mennesker absolutt ikke kan gjøres offentlig tilgjengelig (en person kan identifiseres fra sin genetiske profil).

Etter man har funnet frem til listen av cDNA-prober eller genetiske markører man ønsker å jobbe videre med, blir effektstørrelse vurdert og gitt biologisk tolkning. Det er også populært å la en minste ønskelig effektstørrelse inngå i nullhypotesen, slik at funn som ikke er av biologisk relevans ikke dukker opp på listene.

Dette betyr at NHSTP ligger i kjernen i fagfeltet genomikk, og at p-verdier (eller testobservatorer) i hovedsak har sin rolle i estimeringen av FWER eller FDR. Det har så vidt jeg vet ikke vært noen bannlysning av bruk av FWER og FDR innen genomikk. Det har heller ikke vært noen bannlysning av NHSTP. Betyr det at fagfeltet genomikk skiller seg ut med sitt fokus på sett av funn og ikke enkeltfunn? Og, er det lettere for en forsker innen genomikk å liste opp alle nullhypoteser han/hun har studert enn det er for en forsker innen psykologi? Hva er det som gjør at NHSTP er OK innen genomikk og ikke innen psykologisk forskning? Min erfaring fra psykologisk forskning er meget begrenset, og jeg ser derfor ikke svaret på dette spørsmålet.

La oss til slutt bevege oss litt bort fra genomikk og over til nabofeltet medisin. Der var det stor ståhei da artikkelen «Why most published research findings are false» ble publisert i PLoS Medicine i 2005. Forfatter var J. P. A. Ioannidis, og dette er den mest nedlastede artikkelen fra PLoS Medicine. Ioannidis brukte Bayes' regel til å kombinere teststyrke og signifikansnivå med andelen sanne sammenhenger (pre-study odds) og andelen «utforskede analyser som ikke ville vært funn, men likevel endre opp presentert som funn». Det siste kalte han «bias». Han brukte disse størrelsene til å regne ut en positiv prediktiv verdi, som ble tolket som «probability that a research finding is true». Ved å sette inn ulike verdier for de fire størrelsene over var hans konklusjon at «most research findings are false for most

McCarthy, D. J. & Smyth, G. K. (2009): «Testing significance relative to a fold-change threshold is a TREAT». *Bioinformatics* 25(6), 765-71.

Ioannidis, J. P. A. (2005). «Why most published research findings are false», *PLoS Medicine*, 2 (8), e124.

research designs and for most fields». Han fant også at dette tallet varierte mye mellom fagfelt og type forskningsdesign (han fant at funn fra store randomiserte kliniske forsøk med liten bias hadde en sannsynlighet 85% for å være sanne med cut-off på p-verdi på 0.05). Ioannidis avsluttet sin artikkel med forslag til hvordan situasjonen kunne forbedres, og større studier og meta-analyser med lav bias var to hovedpunkter. Ioannidis så som et stort problem at det er vanskelig å vite hvor mange ikke-signifikante funn som er blitt studert før en forskningsgruppe har fått publisert sitt ene signifikante funn.

I 2013 kom Jager og Leek med artikkelen «An estimate of the science-wide false discovery rate (SWFDR) and applications to the top medical literature». Jager og Leek laget R-kode som automatisk søkte gjennom abstracts fra 77430 artikler publisert i fem medisinske tidsskrift i perioden 2000–2010, og tok vare på p-verdier. De fant 5322 rapporterte p-verdier og brukte disse til å estimere en SWFDR på 14%. Ja, de mente at «bare» 14% av funnene i disse medisinske tidsskriftene trolig var falske positive. Metoden bygget på hvordan man estimerer FDR innen genomikk, men med noen justeringer, blant annet for å modellere effekt av avrunding av p-verdier. Artikkelen var en diskusjonsartikkel i *Biostatistics*, og etter innspill fra bl.a. Cox, Benjamini og Hechtlinger, Gelman og O'Rourke og Ioannidis (spesielt var det kritikk av hvordan fordelingen til de sanne nullhypotestene var modellert) konkluderte Jager og Leek med at det var konkrete muligheter for å forbedre metoden deres og at et estimat på SWFDR trolig var noe større enn 14%, men ikke er større enn 50%. Jager og Leek publiserte sin R-code og datagrunnlag gjennom Github (<https://github.com>) slik at deltakerne i diskusjonen (og andre forskere) lett kunne se detaljene i hva de hadde antatt. Dette gjorde det svært lett å umiddelbart se den empiriske effekten av metodiske endringer.

Går det an å oppsummere erfaringene fra genomikk og disse to kontroversielle artiklene til Ioannidis og Jager og Leek med at «samvittighetsfull omgang med NHSTP er påkrevd»? Forskere vet hvilke effekter de ønsker å undersøke og hvilke hypoteser de ønsker å teste. Statistikere vet hvilke forutsetninger som ligger bak statistiske metoder som brukes til hypotesetesting, hvordan en p-verdi skal tolkes, og hvordan multiplisitet skal håndteres. Sammen må forskere og statistikere velge design og analysevei som gir de beste vilkår for reproduserbar forskning!

Jager, L. R & Leek, J. T (2013). «An estimate of the science-wide false discovery rate and application of the top medical literature», *Biostatistics* 15(1), 1–12. og

Jager, L. R & Leek, J. T (2013). «Rejoinder: An estimate of the science-wide false discovery rate and application of the top medical literature», *Biostatistics* 15(1), 39–44.

Cox, D. R. (2013). «Discussion: Comment on a paper by Jager and Leek». *Biostatistics* 15(1), 16–18.,

Benjamini, Y. & Hechtlinger, Y. (2013). «Discussion: An estimate of the science-wide false discovery rate and application of the top medical literature». *Biostatistics* 15(1), 13–16.,

Gelman, A. & O'Rourke, K. (2013). «Discussion: Difficulties in making inference about scientific truth from distributions of published p-values». *Biostatistics* 15(1), 18–22. og

Ioannidis, J. P. A. (2013). «Why “An estimate of the science-wide false discovery rate and application to the top medical literature” is false.» *Biostatistics* 15(1), 28–36.

Statistikere mot atomvåpen

Tore Schweder | Oslo

Det var på Engebret Café!

I Bernoulli News nummer 2, 2014, blir Shiryaev intervjuet om den første Bernoulli verdenskonferansen i Tashkent i 1986. Han sier at en hovedmann bak konferansen, Sagdu Khasanovich Shirazhinov, deltok på et møte ett eller annet sted i Europa, der ideen om en stor internasjonal konferanse kom opp. Det europeiske møtet som Shirazhinov deltok på må ha vært det ellefte europeiske statistikermøtet i Oslo 14 - 18 august 1978. Erik Mohn ledet organisasjonskommiteen. Jeg var også med, med ansvar for utstillinger. På en lunsj, eller kanskje en middag, på Engebret Café, satt jeg ved siden av en røsslig og hyggelig sovjet, Shirazhinov. I samtalen løp sa jeg at det ville være kjempefint om noen kunne ta initiativet til en verdenskonferanse i statistikk, kanskje i Uzbekistan der han var ifra. Det ble litt snakk om dette over bordet, men noe mere hørte ikke jeg om saken før det ble gjort kjent at den første verdenskonferansen skulle være 8-14 September 1986 i Tashkent. Konferansen ble en stor suksess.

Det er ikke bare fag på slike konferanser. Sammen med Geoffrey Grimmett gikk jeg en fin tur i fjellet sørøst for Tashkent. Vi pratet blant annet om den kalde krigen og faren for atomkrig. Legene hadde organisert seg internasjonalt mot atomvåpen, og spørsmålet ble reist om kanskje vi statistikere skulle samle oss i «Statistikere mot atomvåpen». Jeg husker ikke helt hvordan det gikk med den ideen før vi satt på det kjempestore Aeroflot-flyet på vei mot Moskva. Flyvertinnen hentet meg fram til første klasse der de fremste sovjetiske statistikerne (og partimedlemmene) satt. De var svært interessert i å få realisert en internasjonal organisasjon av statistikere mot atomvåpen. Grimmett var også med i samtalen. Avtalen ble at jeg skulle ta det videre initiativet. Det gjorde jeg ikke, dessverre. Gorbatsjov var ved makten og utover i 1987 tødde den kalde krigen opp. Om det var vel betenkt, eller om det var av latskap og en smule beskjedenhet, at jeg lot saken ligge, skal jeg ikke si, men angrer gjør jeg. Statistikerne fra Sovjet fikk annet å tenke på og noen henvendelse derfra om saken kom ikke. Grimmett var også taus. Men atomvåpnene ble ikke borte. Nå kjølner forholdene mellom Russland og NATO. Noen snakker om ny kald krig. Kanskje tiden er inne for å ta opp tanken om «Statistikere mot atomvåpen»?

R by Example

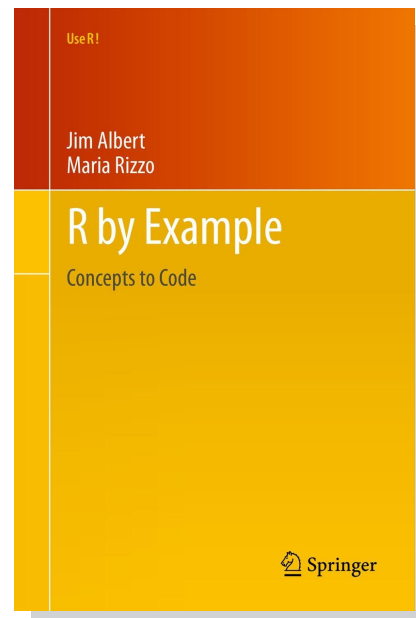
Karl Ove Hufthammer | Bergen

Ei god innføringsbok i R vil mange ha glede av, og ei *eksempelbasert* bok høyrest ut som ein veldig god idé, med stort potensial. Dessverre utnyttar ikkje denne boka potensialet. Boka skryt av «interesting applications with real data» og eksempel som «reflect good statistical practice». Eg synest ingen av delane stemmer.

Kvart kapittel tar for seg eit statistisk emne (for eksempel regresjon), som vert illustrert med ulike analyseeksempel og R-kode for desse, etterfølgt med eit lite sett oppgåver. Datasetta som vert presenterte er ofte uinteressante, og dei statistiske analysane for enkle/naive – og til dels direkte feil. Boka presenterer gammaldagse måtar å gjera ting på, der det no finst betre, enklare og meir fleksible alternativ. Presentasjon av grafikkmoglegheitane i R (ein styrke ved R) er svært svak.

Men alt i alt vil ein som nybegynnar få ei brukbar innføring i delar av R med denne boka. Om ein alt *har* kjøpt boka, bør ein klart ta seg tid til å arbeida seg gjennom ho.

Dette er ein svært nedkorta versjon av bokmeldinga, då det ikkje var plass til heile i *Tilfeldig gang*. Fullstending bokmelding, med mange eksempel på feil og manglar i boka, finn de på <http://huftis.org/artiklar/bokmelding-r-by-example/>.



Løsning på pusleri nummer 41

Jostein Lillestøl | Bergen

Oppgaven

Du skal simulere utfallet av et rettferdig terningkast, men til hjelp har du bare en sekssidet terning som ser skjev og rar ut. Greier du det likevel når du får lov til å trille den skjeve terningen flere ganger? Hint: $6 = 3!$

Løsning

Flere løsninger fins, og den enkleste å beskrive er funnet av Håvard Rue:

- 1 Kast terningen 6 ganger.
- 2 Hvis alle tallene 1...6 dukker opp, returner utfallet av første terningkast. Hvis ikke, gå tilbake til 1.

Løsningen bygger på at, gitt at alle kastene er ulike, så er alle 6! ordninger av utfallene like sannsynlige. Ulempen med denne løsningen er at vi risikerer å holde på ganske lenge. En «raskere» løsning er:

Navngi de seks sidene på terningen med bokstavene A, B, C, D, E, F. Resultatet av et antall terningkast danner da et «ord». Ulike ord kan da ordnes leksikografisk.

- 1 Start med å trille terningen tre ganger, og noter ordet. Er vi heldige er det tre ulike bokstaver i ordet, f.eks ECA. Dette er et av i alt seks ulike ord som som fås ved permutasjoner av de tre bokstavene, i leksikografisk rekkefølge er disse ACE, AEC, CAE, CEA, EAC, ECA. Disse seks ordene må være like sannsynlige. Hvis vi på forhånd har fastlagt at ordets plass i rekkefølgen bestemmer utfallet 1,2,3,4,5 eller 6. Siden vi fikk ECA, som kommer leksikografisk sist, har vi fått utfallet "6 øyne" simulert med sannsynlighet $1/6$. Dette gjelder uansett hvilke tre ulike bokstaver ordet består av.
- 2 Er vi mindre heldige får vi et ord med tre like bokstaver, f.eks AAA. Dette gir ingen informasjon, og vi starter på nytt med tre nye kast.
- 3 Er vi sånn passe heldige, får vi et ord med to like bokstaver, f.eks CAA. Dette er ett av i alt tre ulike ord, i leksikografisk rekkefølge AAC, ACA, CAA, som må være like sannsynlige. På forhånd har vi bestemt at ordets plass i rekkefølgen peker på en av de tre mengdene {1 øyne, 2



øyne}, {3 øyne, 4 øyne}, {5 øyne, 6 øyne}. Når vi fikk ordet CAA betyr det at utfallet er i den siste mengden, altså 5 eller 6 øyne. Alt vi nå trenger er en rettferdig trekning for hvilken av de to i den aktuelle mengden som skal velges. Dette kan skje ved to nye terningkast, med resultat f.eks FE. Dette er et av to mulige ord, i leksikografisk rekkefølge, EF, FE, som er like sannsynlige. At vi fikk den siste i rekkefølgen FE betyr det at vi velger den siste i vår mengde, altså 6 øyne. Dersom de to terningkastene skulle vise likt resultat, overser vi det, forsøker med to nye kast og følger samme rutine.

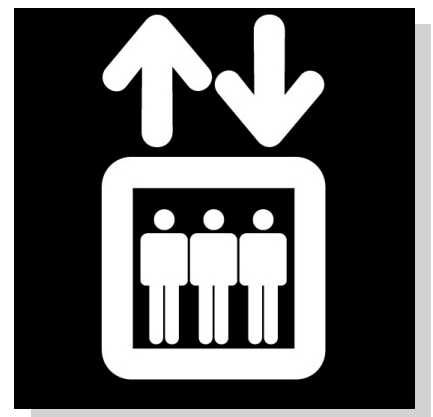
Det er forholdsvis liten sjanse for at vi trenger mange skjeve kast for å simulere et enkelt rettferdig kast. Det er ikke vanskelig å beregne forventet antall nødvendige kast dersom terningen i virkeligheten er rettferdig (selv om den ser rar ut). Den er $144/35=4.114$. Det er mulig å oppnå enda lavere forventning ved å kombinere informasjonen fra de to trinnene. En variant gir forventning 4.09, og det fins tabellariske løsninger med forventning ca. 3.60.

Pusleri nummer 42

Jostein Lillestøl | Bergen

Hansen arbeider i nest øverste etasje i en bygning med seks etasjer over o'te etasje. Den eneste heisen går kontinuerlig mellom etasjene $0,1,2,\dots,5,6,5,\dots,2,1,0,1,2,\dots$, og tar med folk der noen venter på heisen. Når han skal ta heisen ned og ut for å ta seg en røyk, så blir han ofte irritert over at heisen som oftest fortsetter opp. Han skjønner nok hvorfor, men spør seg hva sannsynligheten er, for at når heisen kommer, så går den ned. Han lurar også på hvordan situasjonen hadde vært med to parallelle heiser. Hva er da sannsynligheten for at den første heisen som kommer går ned? Han tror det må være det samme, men er ikke sikker.

Problemstillingen kan belyses ut fra ulike antakelser. Anta her heiser med samme fart og uavhengighet/ tilfeldig start, og at tiden det tar å tømme og fylle heisene er neglisjerbar i forhold til transporttiden. Det siste er selvsagt urealistisk, men tenk da på en såkalt paternosterheiser med kun en bås, der en hopper på og av i fart.



Send svar til Jostein.Lillestol@nhh.no.

møter og konferanser

Kurs i Mixed Effects Models

Anne Marie Fenstad | Bergen

17 statistikere knyttet til Helse Vest var med på Helse Vest-statistikermøte VI og kurset om blandede modeller (Mixed Effects Models). Kurset ga en enkel innføring i lineære og generaliserte lineære modeller. Det var først teori for hver av modelltypene, og så viste kursleder Roy Miodini Hansen hvordan vi kunne anvende disse modellene på reelle data. Det ble brukt Stata på kurset, men de kan også gjøres i R (f.eks. lme4 pakken) eller andre programpakker.

Det var et meget interessant kurs godt ledet av Roy, inspirerende og nyttig. Kurset gikk over to dager fra lunsj til lunsj og kvelden i mellom var så godt som alle deltagerne ute og spiste en tre retters middag. Denne inkluderte også en appetittvekker; rå fisk servert på rosa saltstein fra Himalaya. Er det noen som lurar på hvordan disse blir vasket før neste gjest kommer?



Glade deltagere «on a statistical stairway to heaven».

30-årsjubileum for Statistisk forening i Bergen

Anne Marie Fenstad | Bergen

Norsk statistisk forening avdeling Bergen (NSF Bergen) er NSF's lokalforening for Bergen og omland og ble opprettet 12. desember 1983. Da det kom noe brått på oss før jul i 2013 at vi hadde et jubileum dagen før Luciadagen valgte styret å utsette jubileumsmarkeringen til i år og feire 25 år i stedet.

Alle medlemmer av Norsk statistisk forening (NSF) avdeling Bergen ble invitert til middag (deilig tapas) og hyggelig selskap torsdag 5. mars. Det ble holdt et foredrag om: «Time series and panel data analysis by Hidden Markov models: basics and more recent advances». Foredragsholder var Jan Bulla, førsteamanuensis, Universitet i Bergen.

Det var et vellykket arrangement med mer enn 20 deltagere, fra hovedfag studenter til erfarne professorer. Dette er en klar rekord i deltagere for lokalforeningen.



Foto: Eva H. Dybvik



Foto: Eva H. Dybvik

Arrangementet var et samarbeid med Statistisk forening i Bergen og statistikkgruppa ved IGS.

Vi håper å se mange av TGs lesere på det 18. norske statistiker møte på Solstrand i sommer!

tg for 25 år siden

Om uttrekning av meddommere enda en gang....

(se TG mars 2014)

TIL DE TO DOMSMENN FRA OSLO.

Rolv Terje Lie, Forskningsstipendiat(NAVF)*

Døm ikke, for at dere ikke skal bli dømt! For dere skal selv dømmes etter den dom dere feller over andre, og det skal måles opp for dere i det målet dere selv bruker.

Matteus 7,1–2

Er det a priori gitt at det bare er høyrefolk som jukser? Man kan sitte igjen med det inntrykket etter å ha lest innlegget til Ørnulf Borgan og Bent Natvig i forrige nummer av Tilfeldig Gang (Desember 1989). De to forfatterne foregir å ville teste hypotesen om at uttrekningen av meddommere i Treholtsaken var lotterisk. Dette gjør de på bakgrunn av at to av de fire meddommerne viste seg å være medlemmer av Bygdøy Høyre. Det er lett å forstå at folk flest ville bli fokusert på disse to høyrefolkene. Når man imidlertid skal undersøke om alle personene i domsmannsutvalget hadde like stor sjanse for å komme med blant de fire, så må man vel se på muligheten for at politikere fra andre parti også kunne bli dommere i samme sak. Dersom forfatterne hadde spesielle grunner til bare å se på alternativet med Høyre, så burde de ha gjort rede for dette. I motsatt fall ligner dette på det å lage en ensidig test etter å ha observert at den aktuelle observasjon ligger i en av halene. Kan det ligge noe Bayesiansk i det å bare se på Høyre i dette tilfellet? I så fall er det også på sin plass med en klargjøring.

Den utledningen Borgan og Natvig gir av sannsynligheten for at én person fra Høyre kommer med i utvalget virker overbevisende utfra de tall som antakelig foreligger. Det virker allikevel noe grovt å sette to av de involverte sannsynlighetene ($P(D | C)$ og $P(E | D, C)$) til eksakt lik 0,5, og man må vel da være særlig tilbakeholden med konklusjoner etter å ha funnet en hale-sannsynlighet på 0,032, altså i området rundt 0,05. La oss, i et alternativ til Borgans og Natvigs fremgangsmåte, først akseptere at sjansen for at en H-mann kommer med er 0,077. Vi tar oss imidlertid den frihet å innføre et alternativt parti,

* Medisinsk fødselsregister, Universitetet i Bergen
Armauer Hansens hus, Haukeland sykehus
5021 Bergen

V-partiet som en illustrasjon. Vi gjetter på at sjansen for at en V-mann fra et typisk V-område kommer med er 0,06, altså litt forsiktig gjetning. Den begivenheten som egentlig er interessant i vårt tilfelle er da:

$A = \{\text{Blant de 4 dommerne er minst 2 fra samme parti}\}.$

A kan skrives som unionen av følgende disjunkte begivenheter:

$A = \{2 \text{ fra H, ingen fra V}\} \text{ eller } \{3 \text{ fra H, ingen fra V}\} \text{ eller } \{4 \text{ fra H}\} \text{ eller } \\ \{\text{ingen fra H, 2 fra V}\} \text{ eller } \{\text{ingen fra H, 3 fra V}\} \text{ eller } \{4 \text{ fra V}\} \text{ eller } \\ \{2 \text{ fra H, 1 fra V}\} \text{ eller } \{1 \text{ fra H, 2 fra V}\} \text{ eller } \{2 \text{ fra H, 2 fra V}\}.$

Sannsynligheten for begivenheten A kan nå beregnes som en sum av enkle multinomiske punktsannsynligheter:

$$\begin{aligned} P(A) &= 6(0.077)^2(1-0.137)^2 + 4(0.077)^3(1-0.137) + (0.077)^4 + \\ &\quad 6(0.06)^2(1-0.137)^2 + 4(0.06)^3(1-0.137) + (0.06)^4 + \\ &\quad 12(0.06)^2 \cdot 0.077(1-0.137) + 12(0.077)^2 \cdot 0.06(1-0.137) + 6(0.077)^2(0.06)^2 \\ &= 0,052 . \end{aligned}$$

Ut fra denne signifikanssannsynligheten (som overhodet ikke er ment å være realistisk), ville vi altså ikke velge å forkaste hypotesen om lotterisk utvalg.

Går det an å si noe eksplisitt om hva som er tenkt som alternativ hypotese i vårt tilfelle? Det ligger jo bak her en tanke om at politikere generelt har mulighet til å påvirke utfallet av uttrekningen, og dette kan altså gi et annet resultat enn om man regner som om bare høyrefolk har vilje og mulighet til å tukle med utfallet.

Nytt fra seksjonene 2 og 3 ved Matematisk institutt, Universitetet i Oslo

Nye ph.d-grader

Marcus Karl Viren Eriksson (20. november 2014)

Valuation of energy derivatives. A stochastic control approach.

(prøveforelesning: *Theory and application of LSS and CARMA processes*)

Bedømmelseskomité: Tiziano Vargiolu, University of Padova; Jeanette Woerner, Lehrstuhl LSIV; Tom Lindstrøm, Universitetet i Oslo. Veiledere: Fred Espen Benth, Universitetet i Oslo; Rudiger Kiesel, University of Duisburg; Jukka Mikael Lempa, HiOA.

Krzysztof Jaroslaw Paczka (16. januar 2015)

Stochastic calculus and optimal control under model uncertainty.

(prøveforelesning: *The Skorohod embedding problem and applications in finance*)

Bedømmelseskomite: Nizar Touzi, Ecole Polytechnique; Marie-Claire Quenez, Lehrstuhl Université Denis Diderot; Tom Lindstrøm, Universitetet i Oslo. Veiledere: Bernt Øksendal, Universitetet i Oslo; Giulia Di Nunno, Universitetet i Oslo.

Sara Ana Solanilla Blanco (27. februar 2015)

Stochastic modeling and pricing of energy and weather derivatives.

(prøveforelesning: *Cumulative prospect theory and its applications in mathematical finance, portfolio management and derivative pricing*)

Bedømmelseskomite: Svetlana Borovkova, University of Amsterdam; Michael Coulon, University of Sussex; Erik Bølviken, Universitetet i Oslo. Veiledere: Fred Espen Benth, Universitetet i Oslo; Giulia Di Nunno, Universitetet i Oslo.

Nye mastergrader

Kristina H. Vinvand

Preselection bias in high dimensional regression

Veiledere: Ingrid Glad og Linn Cecilie Bergersen

Simen Eide

Liquidity risk and valuation: valuing and assessing risk in an illiquid market

Veileder: Frank Proske

Peder E. Østbye

Cost information-value trade-off in covariate selection

Veileder: Nils Lid Hjort

Emil Huster

The stochastic transport equation with singular coefficient

Veileder: Frank Proske

Nytt fra Norges teknisk-naturvitenskapelige universitet

Ny mastergrad:

Marte Nevland Hansen

Analysis of Censored Data from Split-Plot Designs.

Veileder: John Tyssedal

Nytilsatte:

Jo Røislien

Førsteamanuensis II

Egil Ferkingstad

Forsker

Nytt fra Avdeling for biostatistikk, Det medisinske fakultet, Universitetet i Oslo

Ny ph.d-grad

Kristoffer Herland Hellton (9. februar 2015)

On high-dimensional principal component analysis in genomics: consistency and robustness.

Bedømmelseskomité: Professor James Stephen Marron, University of North Carolina; Associate Professor Rebecka Jörnsten, Chalmers University/University of Gothenburg; Førsteamanuensis Tron A. Moger, Universitetet i Oslo.

Hovedveileder: Magne Thoresen. Medveileder: Arnoldo Frigessi

Nytt fra Universitetet i Bergen

Nye mastergrader

Line Lægreid Sørensen

Prediksjon av skader i sjøforsikring

Veileder: Bård Støve

Anders Daasvand Sleire

Risk management in Energy markets: instruments and strategies

Veileder: Bård Støve

Linda Mon Xi

Portfolio optimization with the PCC-GARCH-CVaR model

Veileder: Kjersti Aas

Tove Aasland Torgersen

Generaliserte additive modeller med anvendelser på fødselsdata

Veileder: Hans A. Karlsen

Janne Mannseth

An analysis of numerical integrators for Hamiltonian Monte Carlo

Veileder: Tore Selland Kleppe

Nytilsatte

Jan Bulla

Professor

Janne Mannseth

PhD stipendiat

3 stipendiatstillinger i statistikk ved Mat.Inst. UiO

The statistics group at the Department of Mathematics has three PhD vacancies:

One PhD in biostatistics, working with statistical methodology for the analysis of effects of pharmaceuticals on human neurotoxicity and neurodevelopment.

Vacancy 1

One PhD in statistics, working with the development of statistical methodology in the area of structured personalised solutions and large, complex data sets.

Vacancy 2

One PhD in statistics, working with the development of graph based representations of genomes and statistical methodology for analyses on such graphs.

Vacancy 3

For all three positions, the deadline is April 13th 2015.
The statistics group offers an active and attractive work environment for the right candidates – please distribute to potential applicants!

Ledige stillinger i Statistisk sentralbyrå

Det er lyst ut to stillinger ved Seksjon for metoder i Statistisk sentralbyrå, begge har søknadsfrist 12. april.

Den ene stillingen er som rådgiver, seniorrådgiver eller forsker (II, I, o).

Stilling 1

– Den andre stillinger er som førstekonsulent, rådgiver eller seniorrådgiver.

Stilling 2

2 stipendiatstillinger i biostatistikk ved UiO

The positions are each for three years.
The research themes of the two phd positions are:

High-dimensional data analysis (application deadline April 10th.). Scientist in charge: Magne Thoresen.

Vacancy 1

Stochastic processes, event history analysis and causal inference (application deadline April 30th.). Scientist in charge: Kjetil Røysland.

Vacancy 2

SOMMAR STATISTIKK SOLSTRAND

Kor er du 15.–18. juni 2015? Kanskje på Noregs beste hotell*, med fagleg og sosialt fellesskap på det 18. norske statistikarmøtet? Der samlast me på Solstrand Hotel og Bad.

Bli gjerne med på forkurset, «Introduksjon til kausale grafer (DAGs)», som går måndagen og første del av tysdagen. Hein Stigum frå Folkehelseinstituttet/UiO gjev ein introduksjon til kausale grafar – med mykje vekt på eksempel og lite vekt på formalisme.

Sjølve statistikarmøtet startar med lunsj om tysdagen. I år har me fire inviterte føredragshaldarar:

- Erlend Aune (Intelcom)
- Thiago Martins (Yahoo)
- Ragnar Nymoen (UiO)
- Dag Tjøstheim (UiB)



* Kåra av TripAdvisor i 2015: <http://www.tripadvisor.com/TravelersChoice-Hotels-cTop-g190455>

Og om **du** ønskjer å bidra med eit føredrag, vert me veldig glade! Send ein e-post med namn, institusjon, tittel på føredraget og samandrag/abstrakt av føredraget til programkomitéen på **nsf.solstrand15@gmail.com** innan fredag **27. mars 2015**. Alle statistiske emne er av interesse!

I tillegg til føredrag, får me sjølvsagt også plass til utdeling av Sverdrup-prisen, med Svedrup-føredraga, og til årsmøtet til Norsk statistisk forening (NSF). Elles kan me frista med at inkludert i opphaldet får du pausemat, stor lunsjbuffé, trerettars middag kvar kveld (drikke kjem i tillegg), overnatting, frukost, gratis parkering og fri tilgang til hotellets velværeavdeling.



Meld deg på på nettsida til konferansen

<http://norskstatistiskforening.no/moeter/solstrand-2015>

Har du betalt NSF-kontingenten i år? Som medlem av NSF får du det rabatt på konferanseavgifta og på kursavgifta for forkurset.

Er du masterstudent i statistikk **og** medlem av NSF? Då spanderer me konferanseavgifta på kroner 2000. (Talet på plassar med masterstudentpris er avgrensa. Førstemann til mølla ...)

Arrangementskomitéen

Anne Marie Fenstad

Miriam Gjerdevik

Karl Ove Hufthammer

Øystein Ariansen Haaland

Bård Støve

Programkomitéen

Tore S. Kleppe (leiar)

Jo Eidsvik

Steffen Grønneberg

Genaro Succarat

Sigrunn H. Sørbye



Call for People

Den Norske Aktuarforening invites you to attend the Colloquium of the International Actuarial Association to be held in Oslo, Norway, from the 7th to the 10th of June 2015. The colloquium is a joint collaboration of the three IAA sections Pension Benefits and Social Security (PBSS), Life Insurance (LIFE) and The International Association of Consulting Actuaries (IACA).

You are cordially invited to attend
the Oslo Colloquium on 7-10 June 2015
REGISTRATION IS NOW OPEN!

Keynote Speakers

During the colloquium we will have the participation of contributed papers and also keynote speakers with international experience, who will share with us their ideas during their schedule presentations:

- Chief Executive Officer Paolo De Martin, SCOR Global Life, Italy
- Professor An Chen, University of Ulm, Germany
- Managing director Idar Kreutzer, Finance Norway, Norway
- Professor Antoon Pelsers, University of Maastricht, The Netherlands
- Professor Mogens Steffensen, University of Copenhagen, Denmark

Registration

Our early-bird registration period ends on **April 1st**, therefore please do not delay in registering for this important actuarial event and save. The [registration details](#) can be accessed online.

Social Activities

The social activities will include a Welcome Reception at Oslo City Hall, visit to Vigeland Park and the Munch Museum, and the Gala Dinner at Gamle Logen. Read more about the [social activities](#) online.

For any general inquiries relating to the Oslo Colloquium, please contact oslo2015@ccnorway.no



CALL FOR PEOPLE

NOBIM 2015

National Conference on Image Processing and Pattern Recognition

April 13-14, 2015
Hafjell Hotel & Apartments
www.hafjellhotell.no
Hafjell, Norway
www.nobim.no

Invited speakers to be announced

Organizing Committee

Per Christian Moen
HS NEWS SYSTEMS AS
Jan Tore Korneliussen
KUBICAM AS
Jon Yngve Hardeberg
GJØVIK UNIVERSITY COLLEGE
Siri Ø. Larsen
NORWEGIAN COMPUTING CENTER
Robert Jenssen
UNIVERSITY OF TROMSØ

The NOBIM biannual conference takes place in **Hafjell 13-14 April 2015**. It provides a meeting place for researchers and industry within *image processing and pattern recognition*, aiming at exposing relevant Norwegian R&D.

Accepted work will be presented as talks (no conference proceedings). Contributions need not be publishable, and may also have been *recently* published elsewhere. Submissions, in the form of Abstracts of at most 500 words, may be emailed to: conference@nobim.no.

Topics

Theoretical or applied image processing
Theoretical or applied pattern recognition

Important Dates

Submission of Abstract: March 31, 2015
Registration: April 4, 2015

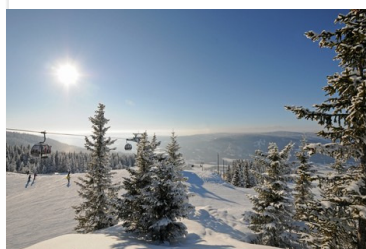
Registration

By email to: conference@nobim.no.

Prices in NOK (member/non-member)

Ordinary: 2500/2650, Ph.D. Stud: 1000/1150,
Master Stud: free (lunch not included).

NOTE: Master Stud. giving talk receives up to 2000 in reimbursement of traveling expenses.



Photos: www.hafjell.no

SCANDINAVIAN JOURNAL OF STATISTICS

Volume 41 • Number 4 • December 2014

Contents:

Statistical Inference for High-Dimensional Global Minimum Variance Portfolios Konstantin Glombek	845
Asymptotic Behavior of Conditional Least Squares Estimators for Unstable Integer-valued Autoregressive Models of Order 2 Mátyás Barczy, Márton Ispány & Gyula Pap	866
Extending Integrated Nested Laplace Approximation to a Class of Near-Gaussian Latent Models Thiago G. Martins & Håvard Rue	893
Strong Consistency of Reduced K-means Clustering Yoshikazu Terada	913
Likelihood Ratio Tests for Dependent Data with Applications to Longitudinal and Functional Data Analysis Ana-Maria Staicu, Yingxing Li, Ciprian M. Crainiceanu & David Ruppert	932
Non-Parametric Estimation of the Conditional Distribution of the Interjumping Times for Piecewise-Deterministic Markov Processes Romain Azaïs, François Dufour & Anne Gégout-Petit	950
Parameter Estimation for Hidden Markov Models with Intractable Likelihoods Thomas A. Dean, Sumeetpal S. Singh, Ajay Jasra & Gareth W. Peters	970
Non-parametric Estimation of Extreme Risk Measures from Conditional Heavy-tailed Distributions Jonathan El Methni, Laurent Gardes & Stéphane Girard	988
Computing Critical Values of Exact Tests by Incorporating Monte Carlo Simulations Combined with Statistical Tables Albert Vexler, Young Min Kim, Jihnehee Yu, Nicole A. Lazar & Alan D. Hutson	1013
Semiparametric Regression Analysis of Longitudinal Skewed Data Huazhen Lin, Ling Zhou & Xiaohua Zhou	1031
The Impact of Measurement Error on Principal Component Analysis Kristoffer Herland Hellton & Magne Thoresen	1051
Selection of Latent Variables for Multiple Mixed-outcome Models Ling Zhou, Huazhen Lin, Xinyuan Song & Yi Li	1064
MMCTest—A Safe Algorithm for Implementing Multiple Monte Carlo Tests Axel Gandy & Georg Hahn	1083
Sample Path Asymmetries in Non-Gaussian Random Processes Anastassia Baxevasi, Krzysztof Podgórski & Jörg Wegener	1102
A New Model for Multivariate Markov Chains João Nicolau	1124

Parameter Change Test for Poisson Autoregressive Models Jiwon Kang & Sangyeol Lee	1136
Local Influence Analysis in AB–BA Crossover Designs Chengcheng Hao, Dietrich Von Rosen and Tatjana Von Rosen	1153
On Efficient Estimators of the Proportion of True Null Hypotheses in a Multiple Testing Setup Van Hanh Nguyen and Catherine Matias	1167
Acknowledgement of Priority Weixin Yao and Longhai Li	1195

SCANDINAVIAN JOURNAL OF STATISTICS

Volume 42 • Number 1 • March 2015

Contents:

Semiparametric Time Series Models with Log-concave Innovations: Maximum Likelihood Estimation and its Consistency Yining Chen	1
Classification Error of the Thresholded Independence Rule Britta Anker Bak, Jens Ledet Jensen and Morten Fenger-Grøn	32
A Note on Estimation in Hilbertian Linear Models Siegfried Hörmann and Łukasz Kidziński	43
Towards Uniformly Efficient Trend Estimation Under Weak/Strong Correlation and Non-stationary Volatility Ke-li xu and Jui-Chung Yang	63
Marginal and Conditional Distribution Estimation from Doublesampled Semi-competing Risks Data Menggang Yu and Constantin T. Yiannoutsos	87
Instrument Assisted Regression for Errors in Variables Models with Binary Response Kun Xu, Yanyuan Ma and Liqun Wang	104
Regression Analysis of Current Status Data Under the Additive Hazards Model with Auxiliary Covariates Yanqin Feng, Ling Ma and Jianguo Sun	118
Inference of Seasonal Long-memory Time Series with Measurement Error Henghsiu Tsai, Heiko Rachinger and Edward M.H. Lin	137
Enriching Surveys with Supplementary Data and its Application to Studying Wage Regression Denis Heng Yan Leung, Ken Yamada and Biao Zhang	155
Deterministic Evolution of Strength in Multiple Comparisons Models: Who is the Greatest Golfer? Rose D. Baker and Ian G. McHale	180
A Non-parametric ANOVA-type Test for Regression Curves Based on Characteristic Functions Juan Carlos Pardo-Fernández, María Dolores Jiménez-Gamero and Anouar El Ghouh	197
Guided Censored Regression Majda Talamakrouni, Anouar El Ghouh and Ingrid Van Keilegom	214
Moderate Deviation Principles for Empirical Covariance in the Neighbourhood of the Unit Root Yu Miao, Yanling Wang and Guangyu Yang	234

A Central Limit Theorem in Non-parametric Regression with Truncated, Censored and Dependent Data Han-Ying Liang, Jacobo De Uña-Álvarez and María Del Carmen Iglesias-Pérez	256
Sparse Principal Component Analysis in Hilbert Space Xin Qi and Ruiyan Luo	270
Parametric Estimation of Menarcheal Age Distribution Based on Recall Data Sedigheh Mirzaei Salehabadi, Debasis Sengupta and Rituparna Das	290
Efficient Estimation of the Partly Linear Additive Hazards Model with Current Status Data Xuewen Lu and Peter X.-K. Song	306

Årskontingent Norsk Statistisk Forening 2015

Kontonummer: 0530 27 73299 Kasserer: E-mail: clara-cecilie.gunther@nr.no Tlf: 22 85 25 63

Medlemskontingenten i NSF er kr. 200,- pr. år for ordinære medlemmer (mastergradsstudenter får gratis medlemskap). Dersom du er usikker på når du betalte sist kan du kontakte kassereren per epost.

Hva gjør vi med pengene?

De største utgiftspostene for NSF er overføringer til lokallagene i Oslo, Bergen og Trondheim, utdeling av Sverdrupprisen, NSF's medlemskap i the European Mathematical Society (EMS) og the International Statistical Institute (ISI), og utgifter i forbindelse med diverse arrangementer (bla. støtte til studenter ved deltakelse på det norske statistikermøtet).

Medlemsregister

For at medlemsregisteret skal være best mulig oppdatert er det viktig at du oppgir evt. ny post/e-postadresse ved å sende meg en e-post. Vi har opprettet en egen epostliste for NSF's medlemmer, nsf-medlemmer@math.uio.no. Dersom du ikke mottar e-post fra denne listen, send meg din oppdaterte e-postadresse.

Betalingsfrist for innbetaling av kontingent er 1. mai 2015. **Vennligst oppgi hvem kontingenten gjelder for ved innbetaling, dette er spesielt viktig dersom din arbeidsgiver betaler kontingenten.**

PÅ FORHÅND TAKK!

Vennlig hilsen
Clara-Cecilie Günther
Kasserer NSF

GIRO F60-1 / Bl. 70.500,00 8 mill. 04.99	Betalingsinformasjon		GIRO		Betalingsfrist	01.05.2015
	Kontingent NSF 2015: kr. 200		Underskrift ved girering			
Betalt av		Betalt til		NSF v/ Clara-Cecilie Günther Norsk Regnesentral Postboks 114, Blindern 0314 Oslo		
Belast konto		Kvittering tilbake				
Kundeidentifikasjon (KID)		Kroner	Øre	Til konto	Blankettnummer	
H				< > 0530 27 73299	< >	