



ISSN 0803-8953

Tilfeldig gang

Nr. 2, årgang 39

November 2022

Utgitt av Norsk statistisk forening

Redaksjonelt

Leder	1
Fra redaksjonen	1

Artikler

Empiriske Metoder ved NHH	2
Hva påvirker strømprisen?	4
Pappkasser, klosser og mennesker som spredningsplott.	6
The Rousseeuw Prize for Statistics 2022	11
Geir Storvik 60 år	14

Tankenøtter

Pusleri nr 57	16
Pusleri nr 56 (løsning)	18

Møter og konferanser

Emeritusfeiring – Henning Omre	23
Seminar til ære for Bo Lindqvist	24
Trondheim statistikk symposium, 2022	26

Meldinger

Nytt fra Oslo senter for biostatistikk og epidemiologi	28
Nytt fra Tromsø	29
Nytt fra NTNU	29
Nytt fra UiB	32
Nytt fra Stavanger	32
Nyheter fra UiO	32

Leder

Ingrid Hobek Haff | Universitetet i Oslo

En av pliktene jeg som leder av NSF har, er å skrive et kort innslag til hvert nummer av TG. Det kan ikke sies å være noen overveldende byrde, og redaksjonen gir meg en påminnelse god tid i forveien. Like fullt kommer tidsfristen som julekvelden på kjerringa hver gang. Jeg ser meg derfor nødt til å tolke «kort innslag» veldig bokstavlig denne gangen.

Det er kanskje flere som har begynt å lure på når neste norske statistikermøte skal finne sted. Nå er det jo også tre og et halvt år siden sist. I 2021 skulle det normalt vært et norsk møte, men så måtte jo Nordstat i Tromsø flyttes fra 2020 til 2021 på grunn av pandemien. Neste norske møte skulle så vært i juni 2023, men Nordstat i Gøteborg, som egentlig skulle vært arrangert i år, har blitt flyttet til juni 2023. Dermed blir det ikke noe norsk møte da heller. Når vil så det neste norske møte bli arrangert? Det er et veldig godt spørsmål, som jeg ikke kan gi et presist svar på i skrivende stund. Jeg kan imidlertid love at en avgjørelse om tidspunktet vil bli tatt i nær framtid.

I mellomtida kan vi glede oss til Nordstat neste år. Jeg har i alle fall tenkt meg dit, og håper å treffe mange av dere der også. Da gjenstår det bare å ønske dere alle en riktig god innspurt til slutten av året!

Fra redaksjonen

Jarle Tufto, Gunnar Taraldsen | Trondheim

Redaksjonen ønsker å takke alle som har bidratt med stoff til denne utgaven.

Empiriske Metoder ved NHH

Håkon Otneim og Geir Drage Berentsen | Norges Handelshøyskole,
hakon.otneim@nhh.no, geir.berentsen@nhh.no

Norges Handelshøyskole (NHH) tilbyr ett enkelt bachelorprogram i økonomi og administrasjon. Høsten 2022 var det 500 plasser ved dette programmet. Bachelorprogrammet består av en rekke obligatoriske kurs, blant annet statistikkursene MET2 – Statistikk for Økonomer (som typisk ligger i 2. semester) og MET4 – Empiriske Metoder (som studentene vanligvis tar i 4. eller 5. semester). MET2 er et klassisk innføringskurs i grunnleggende statistikk, mens MET4 i stor grad er et kurs i anvendt økonometri som danner grunnlaget for empiriske masterkurs i økonometri og det vi kaller «business analytics».

MET4 består av fem moduler: Grunnleggende statistikk, hypotesetesting, regresjonsanalyse, tidsrekkeanalyse, samt en sekkepost som vi har kalt «avansert regresjonsanalyse og maskinlæring» der vi går gjennom paneldatamodeller, logistisk regresjon, samt enkel klassifisering ved hjelp av kNN. Ved årsskiftet 2019/2020 bestemte vi oss for å eksperimentere litt med undervisningsformen. Vi kan godt innrømme at en viktig motivasjon for dette var å gjøre kurset mer effektivt å undervise – vi må tross alt levere MET4 både i høst- og vårsemesteret – og en stor del av dette arbeidet innebar å digitalisere den tradisjonelle pensumgjennomgangen i plenumsforelesninger, og heller bruke forelesningstiden på oppgaveregning og gruppediskusjoner.

Som pilotprosjekt valgte vi tidsrekkeanalyse, som var siste del av kurset. Vi delte pensum opp i 7 mindre tema, og spilte inn korte videoer for hver av dem. Til hver video skrev vi så opp noen korte kommentarer, noen kontrollspørsmål, og noen empiriske oppgaver som skulle løses i R. Timingene for dette opplegget var meget god, for tidsrekkeforelesningene våren 2020 var berammet til andre halvdel av mars. Det skulle være unødvendig å peke på fordelene med å ha klart et godt digitalt opplegg på den tiden, og vi fikk følgelig en mindre sjokkartet overgang til «koronatiden» enn mange andre.

Når det er sagt så var det ikke lenge til vi måtte snu bunken. Kurset skulle gå igjen allerede høsten 2020, så vi bestemte oss ganske enkelt for å gå «all in» og digitalisere resten av kurset med en gang. Vi måtte henge i, men i løpet av vår, sommer og høst 2020 klarte vi å spille inn over 100 små videoer som dekket hele kurset. Videoene kunne være en kort innføring til et spesifikt tema, enten ved at vi snakker og peker over lysbilder, tegner og regner på en iPad, eller demonstrerer relevante kommandoer i R. Vi la disse videoene ut på en nettside (skrevet ved hjelp av RMarkdown), der vi kunne legge ut kommentarer, spørsmål, oppgaver, datalabber og tidligere eksamensoppgaver i tillegg. Denne nettsiden er åpent tilgjengelig: hotneim.github.io/met4.

Gjennom koronatiden har vi eksperimentert med ulike måter å

fylle «skoletiden» på. Det har ikke vært enkelt, siden det har vært så mye frem og tilbake mellom stengt, delvis stengt, og mer eller mindre åpen campus. Vi har fortsatt noen få oversiktsforelesninger i plenum for å knytte sammen ulike deler av pensum på en måte som ikke lar seg gjøre i små videoer. Vi har sesjoner der studentene jobber sammen i grupper med relativt enkle oppgaver, og så har vi oppgaveseminarer i plenum for problemene som er litt vanskeligere. Vi har også hatt rene kontakttimer/orakeltimer der vi er tilgjengelig for spørsmål. Det var likevel veldig vanskelig å engasjere studentene gjennom aktiviteter som både skulle gjennomføres fysisk og digitalt – gjerne begge deler samtidig.

På tross av det ser vi nå konturene av et opplegg som fungerer veldig godt, både for oss som underviser og for studentene. Undervisningsformen minner om det som kalles «flipped classroom», der ideen er at studentene bruker egen tid på å gå gjennom teori ved hjelp av videosnutter, og at de i større grad aktiviseres i fellesskap med medstudenter og kursansvarlige. Vi skriver litt mer detaljert om opplegget vårt i en artikkel som vi presenterte ved ØA-konferansen i Bodø i april 2022. Artikkelen kan lastes ned fra [denne lenken](#).

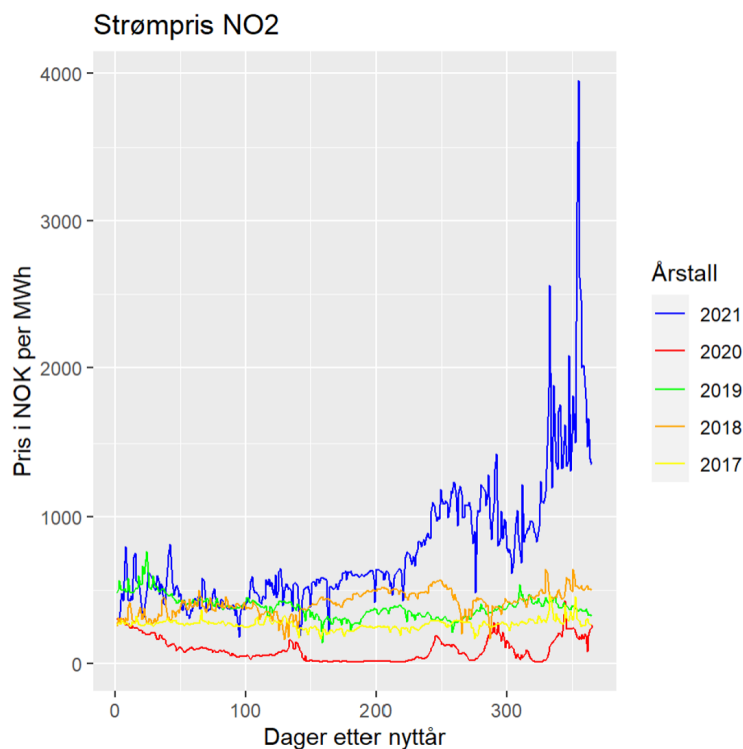
Vi har tro på dette. Vi vet også at undervisere med andre institusjoner lager lignende opplegg. Kan det være på sin plass å diskutere erfaringer fra koronatiden og muligheter for fremtiden i et forum organisert av NSF? Vi er i alle fall interessert i tilbakemeldinger på det vi har gjort, og er åpne for diskusjon og samarbeid.

Hva påvirker strømprisen?

Benjamin Sigbjørnsen | Universitetet i Stavanger

Det har det siste året vært mye diskusjon rundt strømprisen i Norge. Prisen har vært mye høyere enn tidligere, og privatpersoner og bedrifter har derfor fått betydelig høyere strømrregninger. Prisene har vært så høye at staten har sett seg nødt til å subsidiere strømrregningene. Det har også vært stor diskusjon rundt årsaken til de høye strømprisene, og blant folk flest virker det som om ACER og strømkabler har fått mye av skylden. Funn i bacheloroppgaven min tyder på at årsaken er mer sammensatt.

Jeg har i bacheloroppgaven min, sammen med veilederen min Alex Bentley Nielsen, gjennomført en uavhengig statistisk analyse av dagsgjennomsnittet til spotprisen i NO2. NO2 er Sør-Norge sitt prisområde for strøm. Oppgaven ble skrevet mellom januar og mai i 2022. Målet var å finne modeller som kunne forklare størst mulig del av variasjonen i strømprisen, predikere framtidig strømpris og som kunne si noe om hva som påvirker strømprisen og hvordan. Dette ble gjort ved å bruke metoder fra statistisk læring. Før jeg begynte å skrive så tok jeg kontakt med Lyse Produksjon, og fikk fra dem tips til hvilke prediktorer som kunne være aktuelle. Det ble hentet metrologiske data og data fra strømmarkedet fra 2017 til noen måneder ut i 2022 til bruk som prediktorer. Forskjellige delmengder av årstallene fra 2017 til 2021 ble brukt som treningsdata, mens data fra 2022 ble brukt som testdata. Flere forskjellige typer modeller ble tilpasset, lineære og ikke-lineære. Den beste modellen ble funnet å være en lineær modell, med logaritmen til spotprisen som responsvariabel og data fra 2021. Denne modellen klarte best å predikere testdataene fra 2022.



Et anslag for prosentvis økning i strømprisen ved elektrifisering av

norsk sokkel, ved bruk av den beste modellen, er 11.7 prosent. Dette er anslått å ville gi 10% høyere forbruk i NO₂. Hvis det i gjennomsnitt overføres 50 prosent mer strøm ut av NO₂, enn det som ble gjort i 2021, så vil prisen bli cirka 8.25% høyere ifølge modellen. Denne økningen vil være tilsvarende som om North Sea Link kjører strøm ut av NO₂ for fullt, hele året. Analyser gjort for nesten 10 år siden viser omtrent det samme, her anslås det 10 prosent økning i strømprisen (Lie, 2014) ved åpning av North Sea Link. Ved å plote de forskjellige variablene og se på forskjellene fra år til år er det et par variabler som peker seg ut som interessante for året 2021. Summen av overføring ut av NO₂ i 2021 er ikke spesielt mye høyere enn for tidligere år, men forbruket i NO₂ er en god del høyere enn tidligere år. Nedbørsmengden i 2021 ved en valgt værstasjon i NO₂ er mye mindre enn tidligere år, og plott av magasinnivået viser at magasinene ikke blir fylt opp på samme måte som tidligere.

Disse resultatene gir ikke et entydig svar på hvorfor strømprisen er så høy, men det virker som om utenlandskablene bare forklarer en begrenset del av prisøkningen. North Sea Link vil, ifølge modellen, bare øke prisene med 8 prosent. Mye har gått «galt» samtidig, og derfor vil løsningen trolig være sammensatt for at prisen skal gå ned. Strømmarkedet er hele tiden i endring, og det har spesielt skjedd mange endringer i 2022 på grunn av Russlands invasjon av Ukraina. Endrede betingelser for strømmarkedet vil kunne gjøre at modellene bør oppdateres med nye data og at slutningene derfor blir annerledes. Andre ting som vil kunne påvirke nøyaktigheten til modellen er hvis prediktorene beveger seg utenfor verdier som er sett tidligere. Magasinfyllingen har i 2022 vært spesielt lav, og modellen som har data fra 2021 kan ikke nødvendigvis forklare hvordan dette vil påvirke strømprisen.

Lenke til R-kode: <https://github.com/BenSig00/Statistical-analysis-of-the-power-market-in-NO2>

Lenke til oppgaven: <https://hdl.handle.net/11250/3004299>
Kilde:

Lie, Ø. (2014, 20. august). Norges kraftoverskudd forsvinner til England [Teknisk Ukeblad]. <https://www.tu.no/artikler/norges-kraftoverskudd-forsviner-til-england/230246>

Pappkasser, klosser og mennesker som spredningsplott.

Jo Røislien | Universitetet i Stavanger

Hvordan lager man nyttige videofilmer om grunnleggende statistikk til bruk i statistikkundervisning?

Statistikk har på relativt kort tid gått fra å være et fag for de spesielt interesserte til et fag de fleste som tar høyere utdanning må gjennom. Om de studerer ingeniørfag eller samfunnsøkonomi, markedsføring eller helsefag, må de vite noe om hva man kan – og ikke kan – gjøre med tall, de må kunne lese vitenskapelige artikler innen eget fagfelt, og vurdere kvaliteten på forskningsresultater.

At flere får innblikk i kvantitativ forskningsmetode og statistikk er positivt. Samtidig har mange av disse nye studentene ikke den nødvendige matematiske kompetansen – eller interessen – for å forstå all den matematikken som gjennomsyrrer statistikkfaget. Det gjør undervisning av faget mer utfordrende enn det var tidligere.

Statistikk og levende bilder

Våren 2021 ble jeg kontaktet av MatRIC – Centre for Research, Innovation and Coordination of Mathematics Teaching – ved Universitetet i Agder [1]. Senterets fokus er ikke primært matematikkstudenter, men studenter i fag som bruker matematikk og der matematikk er en naturlig del av utdanningen, som lærere, sykepleiere og økonomer. MatRIC kontaktet meg fordi de ønsket å få laget et sett med videofilmer til bruk i grunnkurs i statistikk. Og siden jeg er glad i både statistikk og i filmmediet måtte jeg selvsagt svare ja.

Video er et formidlingsmedium med potensielt enorm tiltrekningskraft: 80% av all internettrafikk er video [2]. Samtidig er hjernens tålmodighet relativt lav når den ser på video, noe som setter krav til kreativiteten, virkemidlene som brukes, og formidlingsfarten.

Nøyaktig hvordan et sett med videofilmer til bruk i grunnutdanning i statistikk bør utformes er på ingen måte åpenbart. Statistikk står med en fot godt plantet i matematikkens abstrakte univers, og den andre i den virkelige verden der talldataene samles inn fra. Hvilken fot man legger mest tyngde på vil avhenge av hva man selv er mest interessert i, og hvilke studenter man underviser faget til.

Skal man lykkes med å nå brukerne av statistikkfaget – de studentene som kanskje aller helst ville vært et helt annet sted enn på statistikkforelesning – med videofilmer om statistikk blir det ekstra viktig å ta formatet video på alvor. Hva er det video er bra på, og hva er det ikke er så bra på. Video er et audiovisuelt medium. Skal man vise bilder som ikke beveger seg kunne man like gjerne tatt et fotografi. Skal man bare snakke rett inn i kamera uten å vise noe kunne man laget en podkast.

Mentale modeller

Altfor ofte når vi har noe å formidle kaster vi oss for fort ut i selve formidlingen. Vi snakker i vei, uten først å ha tenkt tilstrekkelig gjennom hva det er vi faktisk vil si, og hvordan vi bør si det.

Forskning innen utdanning, helsekommunikasjon, kognitiv antropologi og psykologi viser at folk tolker ny informasjon i lys av hva de allerede vet og tror på – deres mentale modeller. Mentale modeller som tilnærming har blitt benyttet flittig i risiko-kommunikasjon [3,4,5], og er en trinnvis prosedyre inspirert av kognitiv psykologi [6, 7].

Mentale modeller-tilnærmingen har også med stort hell blitt benyttet i science communication – formidling av vitenskap til folk utenfor fagmiljøet [8]. Det er en fire-trinns prosess. Først, finn ut hva det er viktig for folk å vite om. Dernest, finn ut hva mottageren allerede vet, og hvor de henter informasjonen fra. Tre, lag kommunikasjonen. Til sist: Test om den virket.

1. Hva vil du si?

Statistikk er omfattende fag. Ikke nødvendigvis fordi den totale mengden stoff er så stor, men fordi faget spenner så bredt. Fra vitenskapsteori og filosofi, via teoretisk og anvendt matematikk, til praktisk datainnsamling, databehandling og tolkning av konkrete resultater.

Høsten 2021 samlet jeg derfor sammen en liten gruppe med erfarne statistikkforelesere. Kathrine Frey Frøslie fra NMBU, Mette Langaas fra NTNU, Njål Foldnes fra BI, og meg selv, fra Det helsefaglige fakultetet på Universitetet i Stavanger. Hele høsten hadde vi jevnlig diskusjonsmøter. Hva er det viktig å få formidlet. Hva det er vi sliter med å få studentene til å forstå. Hva det er statistikk virkelig er? Noen av de vanskeligste statistiske konseptene for studentene å forstå er gjerne de mer filosofiske. Det som forankrer statistisk metode i overordnede ideer og teoretisk matematikk. Men er dette de samme tingene som det er viktig for brukerne av statistikk å ha en inngående forståelse av?

2. Hva trenger mottageren å høre?

En ting er hva du som ekspert mener det er viktig å få sagt. Noe annet er hva mottageren trenger å høre. Kanskje vet mottageren allerede det du vil si. Eller kanskje ligger det du har mest lyst til å formidle så langt forbi der mottageren befinner seg at de ikke har sjans til å få det med seg. Kanskje er de på et sted i livet der de trenger å høre noe annet enn det du opprinnelig hadde ønsket å si. Kanskje ønsker du å få mottageren til å forstå, mens det eneste de bryr seg om er eksamen. Vi snakker ulikt til folk i starten og slutten av et utdanningsløp. Til teoretikere og praktikere. Til folk som vil bli matematikere, og til dem som sitter på en statistikkforelesning fordi de må.

Høsten 2021 samlet jeg derfor også sammen en gruppe studentassistenter ved MatRIC-senteret til en workshop for å høre deres erfaringer med å veilede yngre studenter i oppgaveløsning i statistikk. Mye av det de framhevet som utfordrende var det vi forelesergruppen også hadde notert. Det var en større overraskelse at det studentene slet mest med ikke nødvendigvis var det teoretisk mest avanserte. Faktisk syntes studentene syntes det var lettere å regne på sannsynlighetsfordelinger enn å forstå hvilke fordelinger

som hørte hjemme hvor. Den store utfordringen lå i å skulle koble sannsynlighetsteori til statistikk og tall fra den virkelige verden. De erfarne studentene innrømmet åpenhertig at de sleit med å forstå hva det hele egentlig gikk ut på.

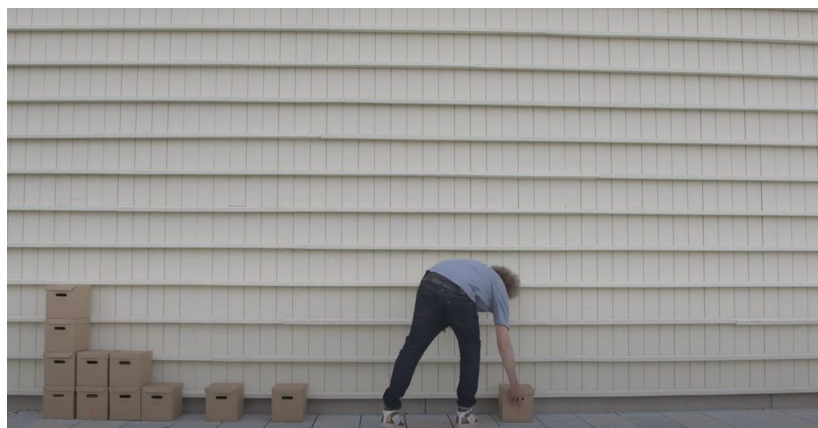
Hva er det statistikk er? Det hadde ingen av de erfarne studentassistentene noe svar på.

3. Lag kommunikasjonen.

En videofilm er bevegelige bilder på en liten, todimensjonal flate. Det vil aldri fullt ut kunne erstatte en god foreleser i full 3D. En live-konsert og en musikkvideo er to forskjellige ting.

Skal man lage videofilmer til bruk i grunnkurs i statistikk må man bruke filmmediet til det det er bra på: Det audiovisuelle. Det å kunne se med øynene. Å visualisere statistikken. Demonstrere med fysiske hjelpemidler og props. Forskningsstudier viser at konkrete er lettere å huske enn abstrakter [9]. Konkrete og bildebruk øker forståelsen og evnen til å huske informasjonen [10]. Å bruke konkrete, visuelle ord hjelper. Legg til et faktisk bilde, og forståelsen økes ytterligere. Særlig i starten av det å skulle lære seg noe nytt. Studentene som representerer framtidige statistikkbrukere er ikke statistikkeksperter, og dermed trenger de også konkrete å holde fast. Noe som abstraksjon kan bygges videre på. Heller enn å gi dem dypdykket inn i abstraksjon og teori kan videofilmer hjelpe dem til å forstå hvor statistikk kommer fra. «Abstraction is the luxury of the expert.» [11]

Manus til filmene ble skrevet vårsemesteret 2022. Levende mennesker lager alltid liv, og vi ønsket derfor å få med statistikk i noen av filmene. Filmene ble derfor filmet på sensommeren 2022 når studentene var tilbake på campus. Klipping har foregått høsten 2022, og filmene er planlagt å være klare til starten av vårsemesteret 2023.



Figur 1: Pappkasser som enkeltobservasjoner (av tid mellom to siste tekstmeldinger på telefonen til 25 statister), og stablet til et (veldig skjevfordelt) histogram.

Til sammen er det planlagt 15 filmer, hver på rundt tre minutters lengde, med visualisering og enkel forklaring av grunnleggende emner i univariat og bivariat statistikk. Det blir stablet av pappkasser for å demonstrere histogrammer og søylediagrammer (Figur 1), balansering av klosser for å visualisere gjennomsnittet som det geometriske tyngdepunktet (Figur 2), og «The human scatterplot» der enkeltpunktene i diagrammet er enkeltpersoner (Figur 3). Alt for å

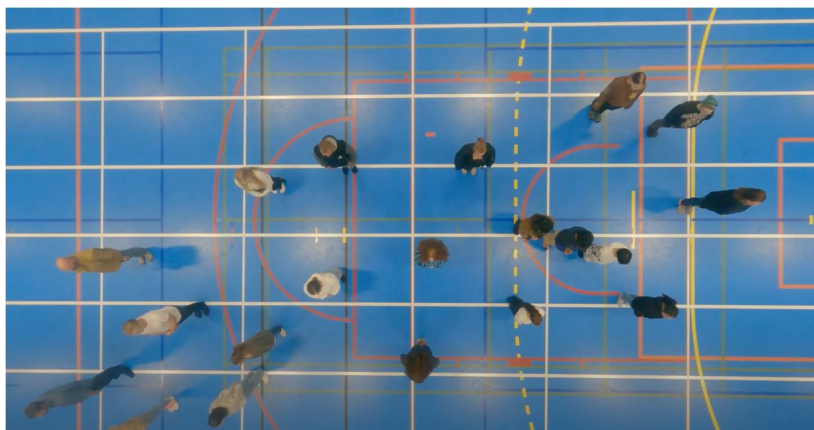
bidra til å gjøre det abstrakte statistikkfaget mer konkret, og hjelpe ikke-statistikere inn i det statistiske universet.



Figur 2: Treklosser stablet langs en tallinje, som balanserer på gjennomsnittet.

4. Virket det?

Det er vanskelig å evaluere om ett konkret undervisningstiltak har effekt. Særlig et tiltak som skal være et supplement til eksisterende undervisning. Det er lettere å måle effekten av en ny pille mot hodepine. Så et konkret suksesskriterium er vanskelig å sette. Noe av håpet med filmene går også forbi den rene læringen, men også å bidra til mer positive holdninger til statistikkfaget. At det å lære statistikk ikke kun er noe man gjør i forelesningssaler og øvingsrom og ved skrivebordet, men også på bussen, ved et cafebord, eller i sofaen hjemme i stua.



Figur 3: «The human scatterplot»: Enkeltmennesker og -observasjoner danner et tydelig system i sammenhengen mellom høyde og skostørrelse.

Statistikk til alle

At MatRIC putter penger på å få laget et sett med filmer til bruk i grunnkurs i statistikk er et flott tilgakt. Vårt svar på oppgaven er å bruke spraybokser, klistrelapper, statister og droner, for å gi formidling og læring av statistikk et visuelt dytt, og henvender seg mer til den praktiske enn den teoretiske delen av faget. Filmene kommer til å ligge åpent tilgjengelig på MatRIC sine hjemmesider og på YouTube, til fri benyttelse for alle som vil.

Så får vi se om noen vil. Kanskje funker denne måten å gjøre det på. Kanskje ikke. Men det som er sikkert er at med den stadig viktigere plassen talldata og statistisk analyse får i samfunnet, koblet

med den endrede studentmassen som nå sitter på statistikkurs på landets høyskoler og universiteter, så kan vi ikke lene oss ensidig på å introdusere folk for faget gjennom matematiske ligninger, og eksemplifisere med svarte og røde baller i en urne. Landets framtidige ingeniører og økonomer, markedsførere og helsearbeidere, trenger noe annet. Fortjener noe annet. Sann at de også kan se skjønnheten i det statistikkfaget vi er så glade i.

Referanser

- [1] Hjemmeside til MatRIC – Centre for Research, Innovation and Coordination of Mathematics Teaching: <https://www.uia.no/senter-og-nettverk/matric>
- [2] CISCO. 2018. «VNI Complete Forecast Highlights ». CISCO. Accessed 08.12.2020. http://cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2022_Forecast_Highlights.pdf.
- [3] de Bruin, W. B., & Bostrom, A. (2013). Assessing what to address in science communication. *Proceedings of the National Academy of Sciences*, 110(Supplement 3), 14062-14068.
- [4] Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk communication: A mental models approach*: Cambridge University Press.
- [5] Boase N, White M, Gaze W, Redshaw C. Evaluating the Mental Models Approach to Developing a Risk Communication: A Scoping Review of the Evidence. *Risk Anal.* 2017 Nov;37(11):2132-2149. doi: 10.1111/risa.12789. Epub 2017 Mar 17. PMID: 28314068.
- [6] Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: an interdisciplinary synthesis of theory and methods. *Ecology and Society*, 16(1).
- [7] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124-1131.
- [8] Røislien J, O'Hara JK, Smeets I, Brønnick K, Berg SH, Shortt MT, Lungu DA, Thune H, Wiig S Creating Effective, Evidence-Based Video Communication of Public Health Science (COVCOM Study): Protocol for a Sequential Mixed Methods Effect Study *JMIR Res Protoc* 2022;11(3):e34275 doi: 10.2196/34275
- [9] Mark Sadoski, Ernest T. Goetz and Maximo Rodriguez, "Engaging Texts: Effects of Concreteness on Comprehensibility, Interest, and Recall in Four Text Types." *Journal of Educational Psychology* 92, 2000, pp. 85-95
- [10] Suzanne Hidi, "Interest and Its Contribution as a Mental Resource for Learning," *Review of Educational Research*, Winter 1990, Vol. 60, No. 4, pp. 549-571
- [11] Chip Heath, Dan Heath. *Made to Stick: Why Some Ideas Survive and Others Die*. Random House Publishing Group, 2007.

The Rousseeuw Prize for Statistics 2022

Stian Lydersen | NTNU

Den 11 og 12 oktober 2022 var jeg i Leuven i Belgia, på den første utdelingen av The Rousseeuw Prize for Statistics. Prisen er på 1 million USD, og er planlagt delt ut annethvert år. Årets pris ble tildelt James Robins, Miguel Hernán, Thomas Richardson, Andrea Rotnitzky, og Eric Tchetgen Tchetgen for deres arbeider innen kausal inferens med anvendelser innen medisin og folkehelse. Hernán og Robins er bl.a. kjent for boka «Causal Inference: What If». Denne boka har vært tilgjengelig på nett i flere år, i stadig oppdaterte og reviderte utgaver. På amazon.com har det i flere år stått at den vil være tilgjengelig neste år. Den kan uansett lastes ned i sin helhet, per 9 november 2022 med 334 sider, fra nettstedet <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>.



Den 11 oktober ble det arrangert et seminar ved Universiteit Gent, der prisvinnerne og andre forskere innen feltet gav presentasjoner, etterfulgt av en paneldebatt. Til slutt fikk vi fingermat og drikke, med en enestående mulighet å snakke med prisvinnerne og andre tilstedeværende. Dagen etter, 12 oktober, var arrangementet flyttet til Leuven, til den ærverdige «Promotiezaal» ved Katholieke Universiteit Leuven (KU Leuven) i sentrum av byen. Om formiddagen var det flere presentasjoner holdt av prisvinnerne. Etter lunsj var det en høytidelig seremoni der kong Philippe av Belgia delte ut prisen.

Professor Peter Rousseeuw var til stede. Han fortalte at han lenge hadde ønsket å etablere en pris innen statistikk, gjerne med en prestisje som kan sammenliknes med nobelprisene innen andre vitenskaper, samt den norske Abelprisen i matematikk. Rousseeuw hadde tidlig i sin karriere arbeidet med aksjefond og pengefond i New York. Han har donert penger til et fond som skal kunne brukes til en pris på 1 million USD annethvert år. Fondet administreres av The King Baudouin Foundation i Belgia, og denne stiftelsen oppnevner komiteen som foreslår hvem som skal få prisen.



Fra venstre: Thomas Richardson (University of Washington), Eric Tchetgen Tchetgen (University of Pennsylvania), James Robins (Harvard University), Peter Rousseeuw, Andrea Rotnitzky (Universidad Torcuato di Tella in Argentina) og Miguel Hernán (Harvard University).

Foto: Skjermdump fra

<https://rousseeuwprize.org/ceremony>

Det var verdt å reise til Belgia for å høre presentasjonene fra

prisvinnerne. I tillegg fikk jeg en halv formiddag som turist i Gent, som viste seg å være en idyllisk by uten for mange turister.

Det fundamentale problemet i kausal inferens er at vi ikke observerer kontrafaktiske utfall. Og vi kan aldri være helt sikre på at det ikke er noen residualkonfundering. Den ideelle studien for å konkludere om kausalitet, er en randomisert kontrollert studie. Men ofte er det ikke praktisk mulig å gjennomføre randomiserte kontrollerte studier. Beslutninger må likevel tas, basert på observasjonelle data.

Bruk av en rettet asyklisk graf (directed acyclic graph – DAG) er egnet til å strukturere antatte retninger på årsak-virkning som ligger under en analyse. Dette burde være en del av alle introduksjonskurs i statistikk, ifølge prisvinnerne. Thomas Richardson sin presentasjon fokuserte mye på generaliseringer av DAGs.

Når man skal gjøre en analyse basert på observasjonelle data er følgende anbefalt ifølge: Start med å beskrive en randomisert kontrollert studie, en målstudie som ville være ideelt egnet til å studere det aktuelle forskningsspørsmålet, dersom det var mulig å gjennomføre den. Dette kalles «target trial emulation». Skriv en plan (protokoll) for en slik studie. Så analyserer man data fra den observasjonelle studien på en slik måte at det ligger så nær som mulig den tenkte randomiserte kontrollerte studien. Hernán viste blant annet til nyere studier av Covid-vaksiner, der forskningsspørsmålet gjaldt effekt av tredje eller fjerde dose, effekt hos barn, eller effekt hos gravide. Det var uaktuelt å gjennomføre store randomiserte kontrollerte studier for disse forskningsspørsmålene. så studiene ble basert på observasjonelle data bl.a. fra Israel. Da man skulle studere effekt av tredje eller fjerde dose, ble dette gjort med matching av personer der den ene var vaksinert og den andre ikke var vaksinert, men ellers så like som mulig. For eksempel en vaksinert 65 år gammel mann av arabisk herkomst ble matchet med en uvaksinert 65 år gammel mann av arabisk herkomst fra samme nabolag i Israel. Matching på samme nabolag viste seg å være en viktig del av matchingen. På denne måten kan man sammenlikne en gruppe som er vaksinert med en gruppe som er uvaksinert, og ellers er så lik som mulig mht antatt viktige bakgrunnsvariable. Hvis jeg har talt rett, har Hernán medforfattet 23 publikasjoner knyttet til effekt av Covid-vaksiner i 2020 til 2022, hvorav halvparten i prestisjetidsskriftene New England Journal of Medicine, The Lancet, The BMJ, og Annals of Internal Medicine.

James Robins beskrev metoder for kausal analyse eksemplifisert med studier som ble gjort på effekt av HAART (highly active antiretroviral therapy) hos HIV-positive. Der ble det typisk gjennomført gjentatte behandlinger og endringer i behandlingsregimer underveis for de enkelte pasientene, basert på tilstand etter forrige behandling. Analysene som han beskrev, var basert på den såkalte g-formelen (Robins generaliserte metoder, 1986). Disse metodene baserer seg på antakelser om kontrafaktisk konsistens, omskiftbarhet (exchangeability), og positivitet (at alle behandlingsforløpene har en sannsynlighet ekte mellom 0 og 1).

Hvem får neste Rousseeuw Prize for Statistics? På hjemmesiden <https://rousseeuwprize.org/about> er følgende områder av statistikk



Foto: Stian Lydersen

En av de mange idylliske kanalene i Gent.



Foto: Stian Lydersen

Kveldsstemning i Leuven sentrum.

listet opp som en retningslinje for å ivareta en bredde:

- I. General Statistical Methodology
- II. Computational Statistics and Data Science
- III. Biostatistics and Envirometrics
- IV. Statistics in the Physical Sciences and Industry
- V. Statistics in Economics and Humanities

Ved avslutningen av seremonien ble vi oppfordret til å komme med forslag til kandidater til neste pristildeling, når det blir åpnet for dette i 2023. Årets pris var innen område III, så prisen neste gang vil tildeles innen et av de andre fire områdene. Kanskje noen av leserne av TG har forslag når det åpnes for nominering våren 2023?

Geir Storvik 60 år

Ingrid Glad, Thore Egeland, Sven Ove Samuelsen

Vår kjære kollega Geir Olve Storvik, professor i statistikk på Matematisk institutt, UiO, fylte 60 år 19 juli. Geir vokste opp i Grimstad og begynte sine studier ved Høgskolen i Agder. Han fortsatte sitt statistikkstudium på Universitetet i Oslo og tok hovedfagseksamen der i 1986. Etter hovedfag begynte Geir på Norsk Regnesentral (NR) og var heltidstilknyttet inntil 1993, men han har fortsatt en bistilling der. I deler av denne perioden var Geir doktorgradsstipendiat. Han tok Dr. Scient-eksamen i 1993 på en avhandling om medisinsk billedbehandling.

Samme år ble Geir ansatt som førsteamanuensis i statistikk på MI, UiO, og han er i dag professor samme sted. Han har pleiet sine internasjonale kontakter blant annet gjennom forskningsopphold i Stanford, Bristol og Odense. Han har vært knyttet til flere sentre på Institutt for biovitenskap, UiO, og bl.a. arbeidet med estimering av fiskeribestander også knyttet opp mot Havforskningsinstituttet. Videre har Geir vært sterkt involvert i SFI-ene «Statistics for innovation» (2007-2014) og «BigInsight» (2015-2024), der han i sistnevnte jobbet mye med data fra maritim sektor.

Med utgangspunkt i doktoravhandlingen om klassifikasjon og billedbehandling, har Geir arbeidet mye med simuleringsteknikker slik om 'adaptiv importance sampling'. Videre har han skrevet om numerisk implementering i situasjoner der det beregningsmessige byr på utfordringer. Geir har et omfattende faglig nettverk og han har arbeidet med anvendt statistikk innenfor mange områder. Hans solide matematiske bakgrunn kombinert med en evne og vilje til praktisk og effektiv implementering har gjort ham til en verdifull samarbeidspartner i også i andre sammenhenger slik som bestandsestimering og rettsmedisinske problemstillinger knyttet til DNA-basert identifisering. De senere årene har han også skrevet flere arbeider innenfor Bayesiansk maskinlæring sammen med yngre kollegaer.

Geir var en selvfølgelig og sentral fagperson i arbeidet med covid-modellering og R-talls-estimering. Han var rett mann på rett sted og kombinerte solid faglig innsikt med praktisk implementering, tidvis under sterkt tidspress. I juni i år var Geir med og presenterte noe av dette arbeidet i London, der artikkelen hans var valgt ut til å diskuteres i Royal Statistical Society.

For tiden er Geir en stødig seksjonsleder for Seksjon 2 for statistikk og data science. Han er varm og forståelsesfull som leder, og går alltid foran som et godt eksempel når det gjelder å yte sitt beste. Han påtar seg mange oppgaver, er et ja-menneske, og har for eksempel stilt opp frivillig som foreleser for studenter i Etiopia, på vinterskoler og UiOs etterutdanningstilbud. Han har også vært helt sentral under oppbyggingen av et nytt masterprogram i Data Science, som er et samarbeid mellom Matematisk Institutt og Institutt for Informatikk.

Av fritidssysler er det særlig badminton, som har opptatt Geir.



Dette engasjementet må sies å overskride begrepet fritidssysler i vanlig forstand: Geir har en rekke gode NM-plasseringer gjennom årene og deltar fortsatt på elitenivå i Norge. Dessuten vinner han stadig senior-verdensmesterskap, nå sist høsten 2022. Han sykler fra Bærum til Blindern i all slags vær, og kommer alltid så tidlig på kontoret at han er fast kaffekoker om morgenen.

I nyere tid har kontakten med Danmark blitt forsterket gjennom kjøp av hytte i nærheten av Frederikshavn.

Det er en stor glede å gratulere Geir med 60-årsdagen! Alt tyder på at han står på videre. Det er vi svært takknemlige for.

tankenøtter

Pusleri nr. 57

I en sjakkklubb står kampen om klubbmesterskapet mellom A og B. For å redusere risikoen for mange kjedsomme remis-partier er det foreslått at mesterskapet avgjøres ved at mester blir den som først vinner to partier på rad. Totalt antall partier er derfor ikke kjent på forhånd, men det er enighet om at opptil rundt 10 partier kan aksepteres. Som medlem av klubben blir du bedt om å vurdere forslaget ut fra en sannsynlighetsteoretisk synsvinkel. Ta utgangspunkt i følgende antakelser:

For hvert parti er det 50% sjanse for remis, 30% sjanse for at A vinner og 20% sjanse for at B vinner (dvs. uavhengig av hvem som spiller svart og hvit).

Hva er sannsynligheten for at A blir mester?

Hva er forventet antall partier som må spilles for å kåre mesteren?

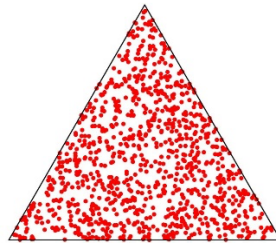
Vil det være et bedre forslag at mester blir den første som har vunnet minst en gang med hvit og minst en gang med svart, ikke nødvendigvis etter hverandre?

Jostein.lillestol@nhh.no

Pusleri nr. 56 (løsning)

Oppgaven var:

Figuren viser simulerte punkter uniformt fordelt innenfor en likesidet trekant med sidelengde 1.



Finn en enkel geometrisk simuleringsmetode (dvs. med passer og linjal).

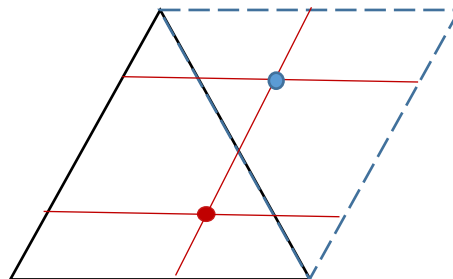
Finn en analytisk simuleringsmetode (dvs. med formel og bruk av regnekraft), helst en som lar seg generalisere til vilkårlig trekant og til rommet (tetraeder).

Tilgjengelig: Uavhengige Uniform[0,1]-variable U .

Dersom du raskt finner en analytisk løsning med tre U 'er og tre programlinjer i R for å simulere, beregne og plote resultatet, kan du jo prøve å finne en alternativ løsning som krever bare to U 'er og to programlinjer i R.

Løsning: Her vises fire metoder: to geometriske, som fritar helt fra tall- og formelregning, og to analytiske metoder som krever en viss bakgrunn i fordelingsteori og som lar seg generalisere til n -dimensjonale simplekser.

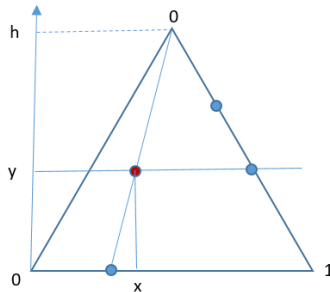
Metode 1: En enkel geometrisk løsning er illustrert i Figur 1, der den likesidige trekanten er supplert med en tilsvarende trekant (stiplet), slik at vi får en rombe med sidekanter av lengde 1. Trekk et tilfeldig tall mellom 0 og 1 og avsett det på den horisontale sidekanten, og trekk linje parallell med venstre sidekant. Trekk et nytt tilfeldig tall og avsett langs venstre side og trekk horisontal linje. Skjæringspunktet vil være et tilfeldig innen romben. Punkter som faller utenfor trekanten kan enten fjernes eller flyttes, enten ved å erstatte det siste tilfeldige tall r med $1-r$ som vist i figuren, eller brette over den stiplede trekant langs den felles sidekant.



Figur 1

Metode 2: Et annet geometrisk resonnement, illustrert i Figur 2, går som følger: Legg den likesidige trekanten med sidelengde 1 og dermed høyde $h = \sqrt{3}/2$ inn i (x,y) -koordinatsystem. Betrakt et tilfeldig punkt (X,Y) . Den marginale tetthet til Y må være proporsjonal med avstanden $h-y$ fra toppunktet, en proporsjonalitet som like gjerne kan avsettes langs sidekanten fra toppunktet (0) til

høyre hjørne (1), dvs. langs et linjestykke av lengde 1. Simuler to Uniform[0, 1] variable (de to blå punkter), og velg deres maksimum (laveste punkt), som nettopp har sannsynlighetstetthet proporsjonal med avstanden. Dette gir Y og den horisontale linjen i figuren. X kan da realiseres ved å avsette en Uniform[0, 1] variabel langs x-aksen (blått punkt), og så trekke linjen til toppunktet. Skjæringspunktet av de to linjene (rødt punkt) er da den realisererte (X,Y).



Figur 2

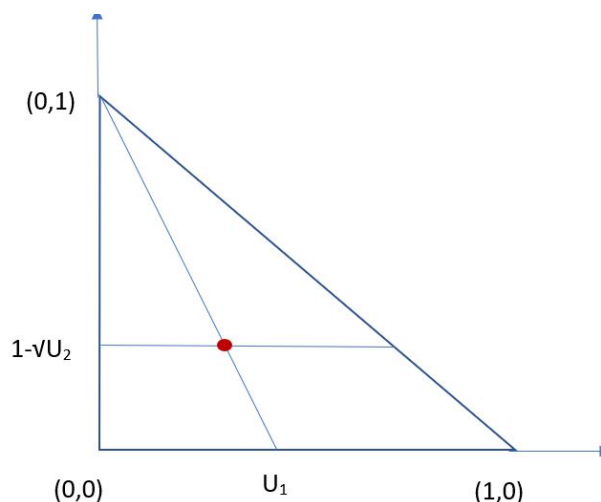
Merknader. Trekanten behøver ikke være likesidet. Resonnementet går like greit for enhver trekant med to sider av lengde 1, f.eks. den rettvinklede trekanten i Figur 3 nedenfor. Hva dersom bare en side (basis) har lengde 1? Da lager vi bare en formlik trekant med tilstøtende side av lengde 1, der vi avsetter de to trekningene som punkt og velger det laveste. Med velvalgte hjelpelinjer overføres dette punktet til den opprinnelige trekanten. Dette gir oss y og dermed den horisontale linjen. Den kryssende linjen for bestemmelsen av x gis da som ovenfor.

Hovedideen ovenfor lar seg generalisere til tetraeder i rommet med sidelengder 1, der areal av horisontale snitt er proporsjonal med kvadratet av avstanden til toppunktet. Simuler tre uavhengige Uniform[0,1] og avsett langs sidekant fra toppunktet. Legg snittflate gjennom det laveste punkt. Simuler så et tilfeldig punkt i snitt-trekanten som ovenfor. Merk at ideen med lavest av tre punkter kan brukes også ved pyramider og kjegler, der et tilfeldig punkt i snittflaten selvfølgelig må simuleres på annet vis.

Merk også at $U \sim \text{Uniform}[0,1]$ medfører at $\sqrt{U}$ har tetthet proporsjonal med u . Med kalkulator kan vi derfor bruke dette istedenfor å gå veien om maksimum av to U 'er. Tilsvarende kan vi bruke $\sqrt[3]{U}$ istedenfor maksimum av tre for tetraederet.

Metode 3: Vi ønsker en eksplisitt formel for simulering. Dersom vi har en formel for en fritt valgt trekant, har vi i realiteten løst problemet for en vilkårlig trekant, fordi en slik alltid fås ved en affin transformasjon, dvs. transformasjon som bevarer linjer som linjer og bevarer parallellitet. Det er hensiktsmessig å ta utgangspunkt i trekanten i Figur 3 med sidekanter langs aksene, og der U_1 og $\sqrt{U_2}$ avsettes. Geometrien i figuren innebærer at $X/U_1 = \sqrt{U_2}/1$, som gir simuleringsformelen

$$(X, Y) = (\sqrt{U_2} \cdot U_1, 1 - \sqrt{U_2})$$



Figur 3

En alternativ formel som er symmetrisk i X og Y blir

$$(X, Y) = (\sqrt{U_2} \cdot U_1, \sqrt{U_2} \cdot (1 - U_1))$$

Dette ser vi ved å betrakte summen $S = X + Y$, med verdi mellom 0 og 1 og fordeling med tetthet proporsjonal med s , og som simuleres med $\sqrt{U_2}$. Betinget gitt $S = s$, er X fordelt Uniform $[0, s]$, dvs. X/s er Uniform $[0, 1]$. Denne simuleres med U_1 , som gir ubetinget $X = \sqrt{U_2} \cdot U_1$ og samtidig $Y = S - X = \sqrt{U_2} \cdot (1 - U_1)$. Den tilsvarende formel for en likesidet trekant med samme basis (se Figur 2) får vi med den lineære transformasjonen $X' = X + Y/2$, $Y' = (\sqrt{3}/2)Y$. En figur produsert med to kodelinjer i R:

```
u <- matrix(runif(2000), ncol = 2)
plot(0.5*sqrt(u[,2])*(1+u[,1]), sqrt(3/4)*sqrt(u[,2])*(1-u[,1]), pch=20, col="red")
```

Uten den geometriske innsikt faller det nok naturlig å ta utgangspunkt i den simultane sannsynlighetstetthet, og bruke tilgjengelig fordelingsteori. Med sikte på generalisering trenger vi et egnet koordinatsystem, og har valgt mellom euklidske og barysentriske. Barysentriske koordinater har fordelen at koordinatene er invariant ved affine transformasjoner, dvs. forskyvning, speiling, rotasjon, ekspansjon og kompresjon. Tar vi utgangspunkt i trekanten i Figur 3 vil den og dens indre være (x_1, x_2) der $x_1 + x_2 \leq 1$. I barysentriske koordinater uttrykkes dette som (x_0, x_1, x_2) , der $x_0 + x_1 + x_2 = 1$. De tre hjørnene i trekanten blir da $(1, 0, 0)$, $(0, 1, 0)$ og $(0, 0, 1)$ som er henholdsvis venstrehjørne (euklidisk origo), høyrehjørne og topp. Med vårt valg av trekant vil vi gjenfinne de euklidske koordinatene som de to siste barysentriske.

På grunn av symmetrien kan vi velge to av elementene i (X_0, X_1, X_2) som grunnlag for simuleringen. Vi ser at den barysentriske løsningen

$$(X_0, X_1, X_2) = (1 - \sqrt{U_2}, \sqrt{U_2} \cdot U_1, \sqrt{U_2} \cdot (1 - U_1))$$

og de to euklidske løsningene ovenfor svarer til ulike valg av to elementer fra den barysentriske.

Dersom trekanten isteden har hjørner med barysentriske koordinater henholdsvis b_0, b_1, b_2 er trekanten og dens indre simulert ved $(X'_0, X'_1, X'_2) = X_0 \cdot b_0 + X_1 \cdot b_1 + X_2 \cdot b_2$. Dersom trekanten er likesidet med samme basis som ovenfor (som i Figur 2), er $b_0 = (1, 0, 0)$, $b_1 = (0, 1, 0)$ og $b_2 = (1 - 1/2 - \sqrt{3}/2, 1/2, \sqrt{3}/2)$, som gir samme euklidske koordinater som ovenfor for denne trekanten.

Metode 4: Uniform fordeling over trekanten vil i euklidske koordinater være $f(x_1, x_2) = 2$ for $x_1, x_2 \geq 0, x_1 + x_2 \leq 1$. Alternativt uttrykt i barysentriske koordinater $f(x_0, x_1, x_2) = 2$ for $x_0, x_1, x_2 \geq 0, x_0 + x_1 + x_2 = 1$. Hvordan vi så går fram vil avhenge av våre kunnskaper i fordelingsteori. Noen vil innse at fordelingen er $\text{Dir}(1,1,1)$, dvs. en Dirichlet-fordeling med parametre $(1,1,1)$, og som utvidelsen av Beta-fordelingen til flere dimensjoner har den Beta-marginaler og nedskalerte Beta-betingede fordelinger. Denne fordeling oppstår blant annet ved andeler for uavhengige Gammafordelte variable med samme skalaparameter, herunder spesielt eksponensialfordelingen med forventning 1, som nettopp gir $\text{Dir}(1,1,1)$. En eksponensialfordelt variabel får vi ved minus logaritmen til en $\text{Uniform}[0,1]$ variabel. For (X_0, X_1, X_2) har vi derfor:

$$X_i = -\log(U_i) / \log(U_0 \cdot U_1 \cdot U_2).$$

Dette gir derfor de to euklidske koordinatene (X_1, X_2) uttrykt ved tre uavhengige $\text{Uniform}[0,1]$ -variable (U_1, U_2, U_3) . Resultatet generaliseres direkte til et n -dimensjonalt simpleks, der $n=2$ svarer til trekant og $n=3$ til tetraeder.

Metode 3 (forts.): De to formlene basert på geometrisk resonnement kan generaliseres til n -dimensjonalt simpleks ved hjelp av den såkalte invers-metoden anvendt rekursivt på betingede fordelinger. En mulighet går som følger:

Betrakt $(X_1, X_2, \dots, X_n)$ i euklidske koordinater og la $S_k = X_1 + X_2 + \dots + X_k$, for $k=1, 2, \dots, n$ (som innebærer at $S_1 = X_1$). Uniform modell på n -simplekset innebærer at S_k er fordelt Beta($k, n+1-k$) for $k=1, 2, \dots, n$. Vi merker oss spesielt $k=n$, der fordelingen er Beta($n, 1$) og derfor har kumulativ fordeling $F_n(s) = s^n$. Modellen innebærer videre at S_k gitt S_{k+1} har kumulativ fordeling $F_k(s | S_{k+1}) = (s / S_{k+1})^k$ for $0 \leq s \leq S_{k+1}$, gjeldende for $k=1, 2, \dots, n-1$. Med n monotont voksende kumulative fordelinger og n uavhengige $U_1, U_2, \dots, U_n$ $\text{Uniform}[0,1]$ anvender vi invers-metoden, og får

$$F_n(S_n) = (S_n)^n = U_n \quad \& \quad F_k(S_k | S_{k+1}) = (S_k / S_{k+1})^k = U_k, \text{ for } k=1, 2, \dots, n-1$$

Dette gir (ubetinget)

$$S_n = (U_n)^{1/n} \quad \& \quad S_k = (U_k)^{1/k} S_{k+1} \text{ for } k=n-1, n-2, \dots, 2, 1.$$

Koordinatene er da gitt ved $X_1 = S_1$ og $X_k = S_k - S_{k-1}$ for $k=2, 3, \dots, n$.

For tetraederet får vi med $n=3$ simuleringsformelen:

$$(X_1, X_2, X_3) = (\sqrt[3]{U_3} \cdot \sqrt{U_2} \cdot U_1, \sqrt[3]{U_3} \cdot \sqrt{U_2} \cdot (1-U_1), \sqrt[3]{U_3} \cdot (1-\sqrt{U_2}))$$

Her har selvfølgelig hver komponent samme fordeling, og komponentene kan permuteres. Den korresponderende barysentriske løsningen er pga. summasjon til 1 gitt ved

$$(X_0, X_1, X_2, X_3) = (1 - \sqrt[3]{U_3}, \sqrt[3]{U_3} \cdot \sqrt{U_2} \cdot U_1, \sqrt[3]{U_3} \cdot \sqrt{U_2} \cdot (1-U_1), \sqrt[3]{U_3} \cdot (1-\sqrt{U_2}))$$

og nye løsningsvarianter får vi ved å velge ut tre av disse komponentene for simuleringen og også permutere disse. Denne metode, utledet på en annen måte, er benyttet av Eivind Stensholt ved NHH i sine studier av valgordninger.

Problemstillinger knyttet til simplekser og barysentriske koordinater finner vi på mange områder, ofte som representasjon av andeler for kategoriske data. Det gjelder i første rekke samfunnsfag og

adferdsfag, eksempelvis i sosiologi og demografi. Det er nyttig både for teoretisk analyse og grafisk illustrasjon, som oftest med trekant for tre kategorier. Det gjelder både rent empirisk, f.eks. utvikling over tid, og generering av scenarier ved simulering. Innen naturvitenskap finner vi simplekser anvendt i kjemi og geologi, blant annet som bestanddeler ved prøvetaking. I operasjonsanalyse og i industriell sammenheng finner vi simplekser ved optimalisering av blandingsforhold og innen IT forekommer det i algoritmer, blant annet ved konstruksjon av grafikk.

møter og konferanser

Emeritusfeiring – Henning Omre

Jo Eidsvik

Torsdag 25 august ble Professor Henning Omre feiret ved NTNU i Trondheim. Henning har en lang og imponerende karriere bak seg innen statistikk: Han startet sitt studium ved NTH, med Arnljot Høyland som veileder på sin diplomoppgave. Han arbeidet deretter ved NR i mange år, og bidro i oppbyggingen av SAND gruppa der. Årene ved NR ble delt opp med en doktorgrad fra Stanford University, California. Henning startet i sin stilling som professor ved NTH (senere NTNU) i 1992, og han har dermed jobbet der i 30 år.



Henning Omre, Kvilhaugen Gård, 2022.

Seminaret inneholdt 10 foredrag inspirert av Henning's bidrag i statistikk og sannsynlighetsregning. Det var noen internasjonale gjester med fagfelt nokså likt Henning, samt gjester fra Etiopia der Henning har ledet et NORAD prosjekt for å utvikle studieprogram i matematikk og statistikk, og seminaret var selvsagt preget av hans store aktivitet med PhD studenter der Henning har veiledet 21 stykker!

Henning's forskningsinteresser er i hovedsak innen romlig statistikk og beregningsorientert statistikk, med anvendelser. I seminaret ble det mimret tilbake til tidlige resultat innen Bayesiansk

Kriging, utviklingen av Bayesiansk inversjon av seismiske data, og om nyanserte modeller og metoder for å assimilere data i rom-tid prosesser.

Dagen ble avsluttet med en middag på Kvilhaugen Gård. Trondheim viste seg fra sin beste side med klar himmel og en vakker solnedgang bak fjellene på den andre siden av fjorden.

Seminar til ære for Bo Lindqvist

Jan Terje Kvaløy og Mette Langaas

Bo Lindqvist rykket ved nyttår 2022 opp til emeritus, og torsdag 15. september 2022 feiret vi Bo og hans karriere. Feiringen var et seminar, som ble holdt i Hovedbygningen på NTNU, og en festmiddag med taler og mimring på restauranten «To rom og kjøkken».



Instituttleder Einar Rønquist ønsker velkommen til «Seminar til ære for Bo Lindqvist».

Bo har gjennom en lang karriere gjort en imponerende innsats på mange områder. Han har vært en aktiv forsker, og har veiledet 100 master og doktorgradsstudenter. Han har hatt mange redaksjonelle og fagadministrative verv, inkludert redaktør i *Scandinavian Journal of Statistics*, og undervist emner på alle nivå, <https://folk.ntnu.no/bo/>.

Bo begynte sin karriere ved UiO, hvor han i 1975 fullførte sitt hovedfag med Erik Torgersen som veileder. I 1978 avla han sin Dr.philos-grad på samme sted. Han ble i 1979 ansatt ved NTH, som

senere ble til NTNU, og har jobbet der siden – totalt over 40 år. De siste 15 årene har han også hatt en professor-II stilling ved UiO.



Rundebordsdiskusjon ledet av Bo Lindqvist, med Ingrid Glad, Arne Bang Huseby og John Tyssedal rundt bordet.

Bo har brede forskningsinteresser og har publisert innen mange områder av matematisk statistikk og beslektede fagområder. Noen stikkord er inferensteori, stokastiske modellering, pålitelighets- og forløpsanalyse. Seminaret reflekterte godt spennvidden i Bo sin karriere, ved foredrag av tidligere studenter, venner og kollegaer. I tillegg inkluderte seminaret en interessant rundebordsdiskusjon om temaet «Kunstig intelligens – eller rett og slett statistikk?», ledet av Bo. Mer informasjon om seminaret, med navn på foredragsholder og titler, finnes her: <https://wiki.math.ntnu.no/selbusymposium/bo70>



Seminardeltakerne på «Seminar til ære for Bo Lindqvist»

Trondheim statistikk symposium, 2022

Kenneth Aase, Mina Spremic og Karina Lilleborge | NTNU

En regnfull oktoberdag i kjent trøndersk stil samlet statistikere, og andre interesserte, seg til den årlige begivenheten Trondheim Statistics Symposium - TSS (ikke til forveksling med Trondheim Studentsangforening eller toksisk sjokksyndrom). Årets utgave ble avholdt ved Bårdshaug Herregård i Orkanger den 28. til 29. oktober. De fleste reiste med buss fra Hovedbygningen på Gløshaugen. Bussturen var fylt med skravling, latter og glede hele veien til Orkanger, tilsvarende 45 minutter kjøretid. Alle og enhver sjekket inn på sine respektive hotellrom og ventet i spenning på kakeutvalget som skulle serveres kort tid etterpå. Ventetiden var kort, og i et konferanserom ventet kaffe, vafler, navnelapper og brownies. Deretter startet foredragene.



Forsamlingen følger nøye med på foredragene.

Professor Thorsten Dickhaus fra University of Bremen var først ut med et foredrag om multipl testing av partiell konjunksjon null hypoteser med anvendelser på replikabilitetsanalyse av høydimensjonale studier. For en gitt γ er målet å teste om oppdagelsene gjort i γ av $s \geq 2$ uavhengige studier stemmer.

Etter en kort pause var det Professor Emeritus Geir Egil Eide fra Universitet i Bergen sin tur, og han fortalte om hans perspektiv på *attributable risk* i epidemiologi. Hvor har vi vært, hvor er vi nå (kausalitetsanalyser) og hvor fremtiden kan lede, samt interessante paralleler til spillteori og lambda-bygget ble presentert.

Da klokka nærmet seg seks var det tid for å forberede seg til middag, som skulle serveres klokka sju. Timen ble fylt med mingling, brettspill eller generell avslapning. Maten var helt akseptabel. Etter middagen, fortsatte de spennende samtale over drinker i jaktbaren.

Lørdagen startet med frokostbuffet (for de aller fleste). Klokka 9 var det klart for dagens første foredragsholder, Valeria Vitelli fra Det medisinske Fakultet, Universitetet i Oslo. Hun snakket om nylige utviklinger innen bayesianske mallowmodeller, med anvendelser innen filmrangeringer og transkriptomikk.

Ingrid Måge fra Nofima presenterte forskningen og fokuset av SFI-en DigiFood og målet om å, blant annet, skape verdi ut av dataene som samles under matproduksjonen. Dette omfatter måling



Thorsten Dickhaus



Geir Egil Eide (t.v.)



Valeria Vitelli



Ingrid Måge

av kvaliteten på den produserte matvaren, variabiliteten i produktene og dens forståelse slik at kvaliteten på det ferdige produktet som ankommer kundene har minst mulig variabilitet.

Symposiets siste taler, Magne Aldrin fra Norsk Regnesentral, snakket om bayesianske modeller for klimasensitivitet. Han presenterte et modellrammeverk der fundamentet for modellen var elementære fysiske lover slik som energibalanse, mens prediksjonene er årlig hemisfærisk overflatetemperatur og global havvarme som en funksjon av historisk strålingspådriv.

Da siste foredraget var over, takket deltakere for enda et vel overstått symposium og fortsatte videre til det avsluttende måltidet - nemlig lunsjen. Deretter gikk bussturen tilbake til Trondheim by, og alle lengtet til neste års TSS.



Magne Aldrin

Nytt fra Oslo senter for biostatistikk og epidemiologi

Avhandlinger

- Magnus Nygård Osnes forsvarte 3. Mai 2022 sin avhandling “Genome epidemiology of *Neisseria gonorrhoeae* and SARS-CoV-2 - strain dynamics from the level of individuals to global dissemination” for graden PhD. Bedømmelseskomite: Professor Edward J. Feil, University of Bath, Associate Professor Nicole Wheeler, University of Birmingham, Professor Fredrik Müller, University of Oslo. Hovedveileder: Seniorforsker Vegard Eldholm, Folkehelseinstituttet. Medveileder: Birgitte Freiesleben de Blasio, Folkehelseinstituttet og Avd. for biostatistikk, UiO.
- Leiv Rønneberg forsvarte 20. juni 2022 sin avhandling “Bayesian modelling of high-throughput drug combination screens: Accounting for uncertainty when searching for drug synergy” for graden PhD. Bedømmelseskomite: Professor Juho Rousu, Department of Computer Science, University of Aalto, Professor Andrea Riebler, Department of Mathematical Sciences, Norwegian University of Science and Technology, Professor Marit Inngjerdingen, University of Oslo. Hovedveileder: Manuela Zucknick, Avdeling for biostatistikk, UiO.
- Simon Lergenmuller forsvarte 22. Juni 2022 sin avhandling “Indoor tanning, sunscreen use, sunburns, and risk of cutaneous squamous cell carcinoma: Understanding for better prevention” for graden PhD. Bedømmelseskomite: Professor Margaret Karagas, Center for Molecular Epidemiology, Geisel School of Medicine at Dartmouth, Assistant Professor Sven Sandin, Karolinska Institutet, Professor Tone K. Omsland, University of Oslo. Hovedveileder: Marit Veierød, Avdeling for biostatistikk, UiO.
- Fei-Chih (Phillip) Liu forsvarte 23. september 2022 sin avhandling “Ultraviolet exposure and night shift work in relation to risk of skin, breast and prostate cancer in offshore petroleum workers” for graden PhD. Bedømmelseskomite: Vice Dean of Research, Professor Pål Richard Romundstad, NTNU, Professor Eva Schernhammer, Medical University of Vienna, Department Chief Physician, Associate Professor II Ingrid Sivesind Mehlum, University of Oslo. Hovedveileder: Jo Stenehjem, Kreftregisteret. Medveileder: Marit Veierød, Avdeling for biostatistikk, UiO.
- Chi (Andrea) Zhang forsvarte 13. Oktober 2022 sin avhandling “Representation and Utilization of Hospital Electronic Health Records Data” for graden PhD. Bedømmelseskomite: Associate Professor Cristina Soguero Ruiz, Universidad Rey Juan Carlos,

Spain, Professor Hugo Lewi Hammer, Oslomet, Professor II
Åslaug Helland, University of Oslo Hovedveileder: Magne
Thoresen, Avdeling for biostatistikk, UiO

Nyansettelser

- PhD stipendiater: Nils Leitzinger, Elisabeth Griesbauer, Gergely Dániel Lukáts, Amirhossein Kazemi.
- Postdoktorer: Charlotte Castel, Anna Plaksienko, Fatih Kizilaslan, Pål Ryalen, Erlend Ignacio Fleck Fossen, Jessica Calland.
- Forsker: Euloge Clovis Kenne Pagui.

Nytt fra Tromsø

Nye Mastergrader

- Petter Støtvig: Probabilistic wind power and electricity load forecasting with echo state networks. Veileder: Filippo M. Bianchi
- Preben Drangsholt: Personalized optimization of blood glucose regulation - a Diabetes Mellitus case study. Veiledere: Fred Godtliebsen, Miguel A. T. Hernandez og Phuong D. Ngo

Nye PhD-grader

- Pedro G. Nicolau: Boreal rodents fluctuating in space and time - Tying the observation process to the modeling of seasonal population dynamics. Disputas 3. juni 2022. Bedømmelseskomite: Sarah Cubaynes, Center d'Ecologie Fonctionnelle et Evolutive, Montpellier, France, Aline Magdalena Lee, Centre for Biodiversity Dynamics, NTNU og Martin Rypdal, UiT. Veiledere: Sigrunn H. Sørbye og Nigel Yoccoz

Nytt fra NTNU

Nyansatte

- Kenneth Aase, PhD student
- Guillermina Senn, PhD student
- Øyvind S Auestad, PhD student
- Karina Lilleborge, PhD student

Nye Mastergrader

- Erik Holst Aasland: «Shapley Values for dependent features using Divisive Clustering» Veileder Kjersti Aas.
- Sebastian Øiungen Ankill: «Statistical methods for analysis of gene expression count data applied to a Crohn's disease dataset». Hovedveileder Mette Langaas, biveileder Atle van Beelen Granlund, (Institutt for klinisk og molekylærmedisin, Det medisinske fakultet, NTNU).
- Eivind Hagemann Brataas: "Statistical Machine Learning on Covid-19 Time Series using Econometrics" Veileder Gunnar Taraldsen.
- Gunnar Grotmol: «A Horizon Metadata Transformer for Multi-horizon Forecasting» Veileder Erlend Aune.
- Hilde Hefte Haug: «Predicting coupled, economical behaviour using Random Forests for longitudinal data» Veileder John Tyssedal.
- Aurora Christine Hofman: «A shared Parameter Model Accounting for Dropout Not at Random -A Case Study on Blood Pressure in the HUNT Study» Veileder Ingelin Steinsland.
- Ole-Magnus Høiback «Estimation in credibility models using Bayesian hierarchical models» Veileder Håkon Tjelmeland, biveileder Anne-Randi Syversveen (If P&C Insurance).
- Vegard Kolaas: «Penalised complexity priors in hierarchical models» Veileder Geir-Arne Fuglstad.
- Piotr Ludvig Kopczynski «Video understanding with the Youtube-8M dataset» Veileder Thiago Martins
- Karina Lilleborge: «Multivariate Spatial Modeling using SPDEs with Application to Ocean Sampling» Veileder Geir-Arne Fuglstad.
- Andreas Matre: «Relative Variable Importance Approaches for Linear Models with Random Intercepts» Veileder Stefanie Muff
- Helene Minge Olsen: «Bayesian Calibration for Modelling the Cardiovascular System Using the Two Element Windkessel Model» Veileder Ingelin Steinsland.
- Lene Tillerli Omdal: «Statistical Analysis of the Association between MicroRNAs in Breast Milk, Perinatal Probiotic Supplement and the Development of Atopic Dermatitis» Hovedveileder Turid Follestad (Klinisk forskningsenhet Midt-Norge, Det medisinske fakultet), biveiledere Melanie Rae Simpson (Klinisk forskningsenhet Midt-Norge, Det medisinske fakultet) og Mette Langaas.

- William Scott Grundeland Olsen: «Spatial Non-Stationary Bayesian Analysis of Deterioration of Pavement With Uncertainty -A Case Study of the European Route E16 Between Bergen and Voss in Western Norway» Veileder Ingelin Steinsland.
- Arne Rustad: «TabGAN: A Framework for Utilizing Tabular GAN for Data Synthesizing and Generation of Counterfactual Explanations» Veileder Kjersti Aas.
- Hennie Schau-Hansen: «Improving Direct Response modelling through Hyperparameter Optimization of Extreme Gradient Boosting and Random Forests» Veileder John Tyssedal.
- Mads Adrian Simonsen: «Approximate Filtering Approaches for Switching Linear Dynamical Systems» Veileder Jo Eidsvik.
- Casper Stenersen: «Forex Trading Volatility Prediction using Machine Learning» Veileder Erlend Aune.
- Elsie Backen Tandberg: «Modeling of decisions made by test persons in a dice task» Veileder Håkon Tjelmeland
- Sara Wøllo: «Correcting for under-reporting of violence against women in Italy using INLA» Veileder Sara Martino.
- Magnus Wølneberg: «Exploring and modeling several missing not at random processes with shared parameter models in a longitudinal study» Veileder Ingelin Steinsland.

Nye PhD-grader

- Bert van der Veen «Statistical advances in multispecies models for community ecology» Supervisor: Professor Robin Brian O'Hara, NTNU. Co-supervisors has been Senior Adviser Knut Anders Hovstad, Norwegian Biodiversity Information Centre and Dr. Francis K.C. Hui, The Australian National University. Committee: Cajo ter Braak, Wageningen University & Research, The Netherlands, Mollie Elizabeth Brooks, DTU, Technical University of Denmark, Stefanie Muff, Department of Mathematical Sciences, NTNU
- Pål Vegard Johnsen «Explainability and validity of statistical methods for genome-wide association studies: Extending Shapley-based explanation methods and adapting saddlepoint approximations» Supervisor: Professor Mette Langaas. Co-supervisor Research Scientist Signe Riemer-Sørensen at SINTEF Digital Committee: Hien Duy Nguyen, University of Queensland, Australia, Valeria Vitelli, University of Oslo, Bo Henry Lindqvist, Department of Mathematical Sciences, NTNU

Nytt fra UiB

Nyansatte

- Samuel Modee, stipendiat
- Jessica Renz, stipendiat

Nye Mastergrader

- Eirik Grasdal Økland: Spatial Prediction with Long Short-Term Memory Models. Veileder: Hans Karlsen
- Odin Fossli: Risk Predictions Using Generalized Additive Models Combined With Bivariate Copulas. Veileder: Antonello Maruotti
- Eirik Lund Rikstad: Stochastic Automatic Gradient Tree Boosting Veiledere: Hans Skaug, Berent Lunde
- Øystein Helle Nordberg: Deep learning-based cross-sensor super resolution of satellite images. Veileder: Hans Karlsen
- Christoffer Brokstad: A Comparison of the local Gaussian correlation and the local dependence function . Veileder: Hans Skaug
- Philip B. Holm: Exploring Neural Networks and Risk Factors for Non-life Insurance Pricing. Veiledere: Kristian Gundersen, Bård Støve

Nytt fra Stavanger

Nyansatte

- Arild Buland er ansatt som Professor II i statistikk ved institutt for matematikk og fysikk.

Nye Mastergrader

- Jimmy Huy Tran, tittel på avhandling: Relative Survival Methods – Theory, Applications and Extensions to Monitoring. Veileder: Jan Terje Kvaløy.

Nyheter fra UiO

Nyansatte

Ingen siden sist.

Nye Mastergrader

- Erik Lien Bolager: Formalising Flows – Introducing Affine Extended Transformations and Flow-Structures. Veileder: Geir Storvik.
- Fredrik Johannessen: Finding Money Launderers Using Heterogeneous Graph Neural Networks. Veiledere: Martin Jullum og Ingrid Hobæk Haff.
- Lars Einar Skaaret-Lund: Improving latent binary Bayesian neural networks using the local reparametrization trick and normalizing flows. Veilder: Geir Storvik.
- Fabio Rodrigues Pereira: Trading Financial Markets Using Reinforcement Learning: Application and Analysis. Veileder: Salvador Ortiz Latorre.
- Anna Kejvalova: Total least squares estimation of maritime battery capacity. Veiledere: Ingrid Glad, Morten Stakkeland og Erik Vanem.
- Thomas Bødkergaard Balser: Linear and Non-Linear Illiquidity Models in Finance and Insurance. Veileder: Frank Proske.
- Christian Grindheim: Methods for battery state of health estimation – Comparative study of machine learning methods. Veiledere: Ingrid Glad, Morten Stakkeland og Erik Vanem.
- Martin Strøm Olsen: Twisting targets in Sequential Monte Carlo. Veileder: Geir Storvik.
- Xiaofei Li: Optimizing bivariate reinsurance with dependent risks. Veileder: Arne Bang Huseby.
- Marte Kittilsviken Tendrup: Two-phase Validation Correction – A simulation study using dfbeta residuals for robust variance estimation. Veileder: Sven O. Samuelsen.
- Simon Kudsk Skoffer : Markovianity and Time Changed Lévy Processes. Veileder: Giulia Di Nunno.
- Yoshimi Kujime: Priority boosting and Lasso-based block boosting: two novel approaches for multi-omics analysis. Veileder: Riccardo De Bin.
- Erik Martinez Jensen: Reserving under stochastic mortality. Veiledere: David Ruiz Banos og Frank Proske.
- Thomas Løland: Modeling and Forecasting Norwegian mortality for pension and life insurance purposes. Veileder: David Ruiz Banos.
- Mathias Alexander von Ottenbreit: Automatic Piecewise Linear Regression. Veileder: Riccardo De Bin.

- Vibeke Gwendoline Fængsrud: Optimal liquidation of market moving portfolios. Veileder: Tom Lindstrøm.
- Anette Fredriksen: Transfer Effects in Scandinavian Languages
– A case study on dependency parsing. Veiledere: Vinit Ravishankar og Erik Velldal.
- Shuijing Liao: The application of penalized logistic regression for fraud detection Studying measures of prediction performance for class imbalanced and high-dimensional data. Veileder: Ingrid Hobæk Haff.
- Marius Johan Franco Dioli: Fair Imputation – Reducing bias under missing data. Veileder: Christos Dimitrakakis.
- Erlend Høigaard: Partial internal models for a Norwegian life insurance company. Veileder: Arne Bang Huseby.

Nye PhD-grader

- Mihaela-Alexandra Puica: Advances in Wind Power Modeling: Merging Research and Market Experience. Veiledere: Fred Espen Benth og Salvador Ortiz Latorre. Bedømmelseskomité: Førsteamanuensis Svetlana Borovkova, Vrije Universiteit Amsterdam, Professor Andreas Wagner, Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM, Professor Tom Louis Lindstrøm, Universitetet i Oslo.
- Andreas Nakkerud: Approaches for Real-Time Traffic Management. Veiledere: Carlo Mannino og Ingrid Kristine Glad. Bedømmelseskomité: Professor Alessandro Agnetis, Università degli Studi di Siena, Professor Irina Gribkovskaia, Høgskolen i Molde, Førsteamanuensis Ingrid Hobæk Haff, University of Oslo.
- Alexander Lobbe: Machine Learning for the Stochastic Filtering Problem. Veiledere: Salvador Ortiz-Latorre og Dan O. Crisan. Bedømmelseskomité: Førsteamanuensis Samuel N. Cohen, University of Oxford, Professor Josef Teichmann, ETH Zurich, Førsteamanuensis Håkon Hoel, University of Oslo.