



ISSN 0803-8953

Tilfeldig gang

Nr. 2, årgang 40

Desember 2023

Utgitt av Norsk statistisk forening

Redaksjonelt

Leder	1
Fra redaksjonen	2

Tankenøtter

Pusleri nr 59	4
Pusleri nr 58 (løsning)	5

Møter og konferanser

Trondheim Statistikkosymposium 2023	8
Referat fra årsmøte i NSF, Trondheims-avdelingen	9
Invitasjon til det 21. statistikermøtet	10

Meldinger

Nytt fra NTNU	11
Nytt fra Universitetet i Bergen	13

Leder

Jacob Skauvold | Norsk Regnesentral

No er det nesten jul og slutt på året. La meg derfor først som sist nytta spalteplassen til å minna om at til neste år, i 2024, er det norsk statistikarmøte. Får du kalender til jul kan du innvia han med å merka tid og stad for møtet: 17. til 20. juni, Tønsberg.

Å ynskja at nokon skal «leva i interessante tider» vert av mange rekna meir som ei forbanning enn som ei velsigning. For å sikra seg kunne ein då tenkja seg å gjera det motsette, altså å senda ynskje om eit uinteressant nytt år. Det lèt ikkje heilt bra, så eg vel ei konvensjonell lukkeynsking. Men først, over til noko litt anna.

Byrjinga av eit nytt år er som kjent tida for sjølvutvikling og for å verta betre til ting. Eventuelt er det tida for å lesa bøker om dette. Personlege føremål ordnar de best sjølve. Her kjem likevel nokre stadfestingar (eng. affirmations) til fri bruk for deg som vil gjere din indre statistikar til ein betre versjon av seg sjølv:

- 1 Når eg lagar figurar skal eg merka alle aksane tydeleg. Eg skal velja skriftstorleik etter visningstilhøva og eg skal alltid hugsa å ta med einingar.
- 2 Eg skal ikkje venta at vanlege folk tolkar p-verdiar og konfidensintervall på den same, korrekte måten som eg og statistikarkollegaane mine gjer. Eg skal hugsa at slik er ikkje verda, og eg skal tilpassa kommunikasjonen min til dette.
- 3 (for studentar) Når eg får nye resultat ut av datamaskina, skal eg gå gjennom dei og sjekka at dei heng på greip før eg sender dei til veiledarane mine.

Og med det vil eg seia til alle som har tatt eksamen, eller laga eller retta eksamen, eller (som meg) ikkje har hatt noko med eksamen å gjera i det heile tatt, til dei som har skrive bok, til dei som har skrive artikkel, til dei som har skrive søknad, til dei som ikkje har skrive noko som helst, og sjølvsagt til dei som har tenkt å komme i gang på nyåret, til alle saman, Ha ei god og fredfylt jul og eit godt og produktivt nytt år.

Fra redaksjonen

Jarle Tufto, Gunnar Taraldsen | Trondheim

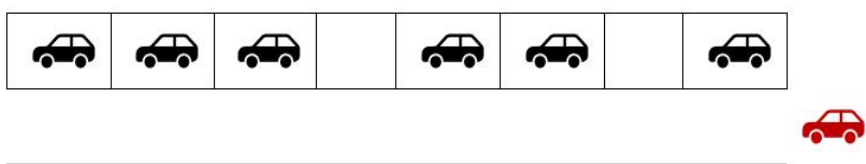
Redaksjonen ønsker å takke alle som har bidratt med stoff til denne utgaven. Vi tar sikte på ett nytt TG-nummer i juni 2024.



tankenøtter

Pusleri nr. 59

Lukas sin arbeidsplass har åtte langsgående parkeringslommer som vanligvis fylles opp ved arbeidshagens begynnelse. Lukas er «lukeredd», dvs. misliker å lukeparkere. Derfor håper han å finne en åpning med minst to parkeringslommer på rad eller en endeplass. Erfaringsmessig kommer seks av hans kolleger tidligere enn han, slik at det bare er to ledige plasser når han kommer. I figuren er vist en dag Lukas måtte lukeparkere.



Lukeproblemet er selvsagt ryddet av veien med enighet om at alle fyller opp plassene forfra ved ankomst. Men akk, for Lukas ser det ut som kollegene gir blaffen i dette.

Anta i utgangspunktet at kollegene velger sin parkeringslomme tilfeldig blant de som er ledige, men tenk også gjennom alternativer til dette: At kolleger muligens kan være omtenkssomme, «lukeredde» eller infame.

Hva er sannsynligheten for at Lukas må lukeparkere, dersom vi antar at de seks før ankomne har valgt sine plasser tilfeldig? Hva kan sies generelt dersom det er m parkeringslommer og n før ankomne?

jostein.lillestol@nhh.no

Pusleri nr. 58 (løsning)

Oppgaven var: I et TV-program, som TV2's Kompani Lauritzen, konkurrerer 14 kjendiser som rekrutter, 7 menn og 7 kvinner. Etter noen uker er 5 rekrutter dimittert, og 9 gjenstår, 5 menn og 4 kvinner. Anta at disse blir kommandert til å stille opp på linje i tilfeldig rekkefølge med ansiktet vendt samme vei. Så kommer ordren: «Menn: Høyre om! Kvinner: Venstre om!». Resultatet blir at noen står ansikt til ansikt med en av motsatt kjønn, mens noen ser inn i en rygg til en av samme kjønn eller ser ingen. Hvert par som står ansikt til ansikt skal fortsette som oppdragsteam, mens de øvrige tas ut til en drittjobb (rengjøring av latrine).

Hva er sannsynlighetsfordelingen til antall team?

Hva er forventet antall rekrutter tildelt drittjobben?

Hva kan sies mer generelt?

Løsning:

Betrakt generelt situasjonen med n_1 menn og n_2 kvinner tilfeldig oppstilt. Vi vil finne sannsynlighetsfordelingen til antall oppdragspar X etter regelen «gunstige på mulige» = g/m . Vi søker spesielt $(n_1, n_2) = (5, 4)$. Her er mønstret for en mulig oppstilling av de 9 rekruttene, der vi står bak rekken og observerer:

MKMMKKMMK

En mann og en kvinne vil stå ansikt til ansikt ved hver MK, dvs. M etterfulgt av K regnet fra venstre mot høyre. I den gitte oppstillingen skjer det tre steder. Dette er bare ett av 126 mulige mønstre, og bestemmelse av antall mønstre med et gitt antall slike par x krever tilsynelatende en komplisert opptelling. Betrakt det enklere tilfellet $(n_1, n_2) = (3, 2)$, der vi bare har følgende $m = 10$ mulige mønstre:

KKMMM	MKKMM	KMKMM	KMMKM	KMMM
MMKKM	MMMCK	MKMKM	MKMMK	MMKMK

Her forekommer ikke MK i det første mønstret, forekommer en gang i de seks neste og to ganger i de tre siste mønstrene. Altså har vi antall gunstige (g) og antall mulige (m) som følger

Antall team x	0	1	2	Mulige m
Gunstige g	1	6	3	10

Tilsvarende får vi (med en programsnutt til hjelp i opptellingen):

	$(n_1, n_2) = (5, 4)$					
Antall team x	0	1	2	3	4	Mulige m
Gunstige g	1	20	60	40	5	126

Resultatene er forenlig med den generelle formelen

$$P(X = x) = \frac{\binom{n_1}{x} \cdot \binom{n_2}{x}}{\binom{n_1 + n_2}{n_1}}, \quad x = 0, 1, \dots, \min(n_1, n_2)$$

Vi ser at telleren er det samme som antall måter vi kan plukke ut x menn fra de n_1 mennene og x kvinner fra de n_2 kvinnene til deltakelse i et eller annet team. Har vi med dette funnet en snarvei til løsningen? Nei! Utvalg av personer må ikke forveksles med vårt utvalg av plasser!

Uttrykket gjenkjennes som en hypergeometrisk fordeling med parametrene $(n_1 + n_2, n_1, n_2)$, der n_1 og n_2 kan bytte plass. Likevel er det ikke åpenbart hvordan dette kan begrunnes med enkel kombinatorikk (som kommer om litt). Forventningen i denne fordelingen blir

$$E(X) = \frac{n_1 n_2}{n_1 + n_2}$$

Denne kan utledes enkelt uten kjennskap til den eksakte fordeling (se sluttkommentar).

Antall rekrutter på drittjobben er $Y = n_1 + n_2 - 2X$ med forventning

$$E(Y) = n_1 + n_2 - 2E(X) = \frac{n_1^2 + n_2^2}{n_1 + n_2}$$

For situasjonen $(n_1, n_2) = (5, 4)$ blir $E(X) = 20/9$ og $E(Y) = 41/9$. Altså er forventet antall rekrutter til drittjobben marginalt større enn forventet antall rekrutter til oppdragsteam som er $40/9$.

Generell begrunnelse: Oppstilling av de $n = n_1 + n_2$ rekruttene i tilfeldig rekkefølge innebærer at alle $n!$ rekkefølger er like sannsynlige. Dette medfører at alle mønstre med n_1 menn og n_2 kvinner er like sannsynlige. Antall slike er antall måter å velge de n_1 mennenes plass blant de $n_1 + n_2$ plassene i rekkefølgen, dvs. $m = \binom{n_1 + n_2}{n_1}$. For å finne antall oppstillinger som medfører akkurat x

oppdragsteam kan vi gå fram slik (etter innspill av Per Manne): Still opp de n_1 mennene på rekke, og velg ut de x som hver skal inngå i et av de x oppdragsteamene. Dette kan gjøres på $\binom{n_1}{x}$ måter. For ethvert slikt valg skal vi plassere inn de n_2 kvinnene slik at det er en kvinne rett til høyre for hver av de x mennene som skal være med i oppdragsteamet. Så la oss gjøre det, og vi har igjen $n_2 - x$ kvinner som må plasseres uten å skape noen nye oppdragsteam. Det er $x+1$ steder hvor disse $n_2 - x$ kvinnene kan plasseres, enten til venstre for alle mennene, eller rett til høyre for en av de x kvinnene som vi allerede har plassert. På hvor mange måter kan dette gjøres?

Her kan en bruke følgende kombinatoriske resultat: Fordel n like objekter til r navngitte bokser (der enkelte tomme bokser tillates). Dette kan gjøres på $\binom{n+r-1}{r-1}$ måter. I vår situasjon tilsvarer det å sette $n = n_2 - x$ og $r = x + 1$, som innsatt gir $\binom{n_2}{x}$ måter.

Til sammen gir dette antall gunstige oppstillinger for x oppdragsteam lik $g = \binom{n_1}{x} \cdot \binom{n_2}{x}$. Den søkte sannsynlighet er da g/m som bekrefter påstanden ovenfor.

Kommentarer:

1. «Stars and bars» brukes ofte som navn på den kombinatoriske formel vi har benyttet ovenfor, etter måten den kan forklares: Dersom de n objektene («Stars») stilles opp på rekke kan plasseringen i navngitte bokser markeres med vertikale streker («Bars») som skilletegn. Med r bokser trengs $r-1$ skilletegn. Eksempelvis med $n=4$ og $r=3$ betyr $**|*|*$ to objekter i den første boksen og et objekt i de to neste boksene, mens $**|**|$ betyr to objekter i de to første boksene og ingen i den tredje boksen. Det er i alt $n+r-1$ symboler, og plasseringsalternativene fremkommer ved å velge de $r-1$ strekene blant disse (eller ekvivalent de n stjernene). «Stars and bars»-problemer kan dukke opp i mange sammenhenger, eksempelvis: På hvor mange måter kan vi skrive et heltall n som en sum av akkurat r positive heltall? Hadde det her stått ikke-negative heltall ville dette vært ekvivalent med

problemstillingen ovenfor med svar $\binom{n+r-1}{r-1}$. Situasjonen med kun positive heltall er imidlertid ekvivalent med å først legge et objekt i hver boks og så plassere de $n-r$ gjenværende i de r boksene. Vi kan derfor erstatte n med $n-r$ i formelen, og får isteden antallet $\binom{n-1}{r-1}$.

2. Tilbake til forventningen til X uten å kjenne den eksakte fordeling: X kan skrives som sum av $n_1 + n_2 - 1$ indikatorvariable, en for hver posisjoner å plassere M etterfulgt av K . For hver slik har vi sannsynligheten

$$P(MK) = \frac{n_1}{n_1 + n_2} \cdot \frac{n_2}{n_1 + n_2 - 1}$$

Forventningen til X som sum av indikatorer er summen av $n_1 + n_2 - 1$ forventninger som hver er lik $P(MK)$, slik at $E(X) = \frac{n_1 n_2}{n_1 + n_2}$. Merk at den hypergeometriske variable X her er representert ved indikatorer som ikke har samme parvise korrelasjoner, men avhenger av om indikatorene er «naboer» eller ikke.

3. Et alternativ til rekruttene tilfeldig oppstilt på linje vil være å ha dem tilfeldig oppstilt i sirkel. Dette kan analyseres på tilsvarende måte, og de to oppstillingene kan danne grunnlaget for en diskusjon av begrepet rettferdighet.

Mer litteratur:

Lillestøl, J and Manne, P (2023) Matched Dispatching in Randomized Settings, NHH Department of Business and Management Science, Discussion Paper 2023/23, <https://openaccess.nhh.no/nhh-xmlui/handle/11250/3106820>

møter og konferanser

Trondheim Statistikkosymposium 2023

Andrea Riebler | NTNU

Det femtende symposiet i statistikk ved Institutt for Matematiske Fag (IMF) ved NTNU ble høsten 2023 holdt på Gløshaugen, NTNU. 45 deltakere var samlet for å holde eller høre på spennende foredrag og/eller nyte godt samvær med kollegaer.

I år ble symposiet organisert som en 1-dags opplegg med fire foredrag. Dagen startet med kafee og te. Så ga John Paige fra IMF en oversikt over forskning som blir gjennomført på statistikkgruppa ved IMF. Han viste opp de forskjellige forskningsfeltene som gruppa har fokus på, felles interesseområder men også fagfelt som kanskje trenger mer støtte i framtida, som for eksempel industriell statistikk. Soledad Gonzalo Cogno, fra Kavli Institute for Systems Neuroscience, Trondheim, presenterte en foredrag med tittelen «Minute-scale periodic sequences in medial entorhinal cortex». Forskningen hun presenterte bidrar å forstå nettverksmekanismene og dynamikken som ligger til grunn for hjernens funksjon. Spesielt i fokus var spørsmålet om lignende koordineringsmekanismer eksisterer på atferdsmessige tidsskalaer.

Etter lunsj var det Pierluigi Salvo Rossi fra Institutt for elektroniske systemer, NTNU, sin tur. Han fortalte om «data fusion in wireless sensor networks». Målet med foredraget var å dekke design og analyse av fusjonstilnærminger for fremtidig «Internet-of-Things» (IoT)-oppsett. Pierluigi presenterte et statistisk signalbehandlingsperspektiv med fokus på den distribuerte versjonen av den binære hypotesetesten som støtter flere energieffektive IoT-praktiske applikasjoner angående robust deteksjon av et fenomen av interesse (f.eks. miljøfare, olje/gasslekkasje, skogbrann). Det siste foredraget ble holdt av Alex Lenkoski fra Norsk Regnesentral i Oslo. Han presenterte et samarbeidsprosjekt mellom Norsk Regnesentral (NR) og San Francisco-baserte startup Wheelhouse. Wheelhouse lar eiere av kortsiktige ferieleieeiendommer (Airbnb, vrbo, etc.) dynamisk prise enhetene sine og automatisk tilpasse seg endrede etterspørselsmiljøer for å maksimere inntektene. I bakgrunnen sitter et skreddersydd estimeringsrammeverk som bruker ideer av overlevingstids analyse of maskinlæring for å overvåke bestillingsatferd og gi pris anbefalinger.

Slides for alle foredragene er tilgjengelige på <https://wiki.math.ntnu.no/selbusymposium/2023/start>.

Da siste foredraget var over, takket deltakere for enda et vel overstått symposium og fortsatte videre til en sosial kveld med Bowling og en hyggelig middag i sentrum av Trondheim.

Referat fra årsmøte i NSF, Trondheims-avdelingen

Ingeborg Hem Sørmoen | NTNU

Tid: Den 23. november 2023. Sted: Lunsjrommet på IMF (1329, 13. etasje, Sentralbygg 2).

Underholdning: Clara-Cecilie Günther med foredraget “The yesterday, today and tomorrow of insurance pricing”. Under medlemsmøtet høsten 2022 stilte Trondheim med nydelig vær. I år var det derimot et voldsomt snøvær og veldig glatte veier, så pizzaen ble forsinket. Heldigvis er Clara-Cecilie sporty og delte foredraget i to deler, så vi fikk bakgrunn om forsikringsprising før matpausen, og så fylte hun på med modellering og statistikk etter pausen.

Tusen takk til dere ca. 20 som møtte opp, og en enda større takk til Clara-Cecilie som tok turen fra Oslo, vi sees neste år!

Referat fra årsmøtet:

- 1 Valg av møteleder og referent. Møteleder: John. Referent: Ingeborg. Godkjent av forsamlingen.
- 2 Årsberetning for de to foregående år. (Går gjennom vedlagt årsberetning). Ingen kommentarer til årsberetningen.
- 3 Revidert regnskap for de to foregående år. (Går gjennom vedlagt revidert regnskap). Ingen kommentarer til regnskapet. Merk: NSF gir alle lokallforeninger 5000 kr per år, men kun til lokallforeninger som har mindre enn 50.000 kr i oppsparte midler. Trondheims-avdelingen har mindre enn dette.
- 4 Medlemskontingent (lokalt tillegg til sentral medlemskontingent). Fortsetter uten ekstra medlemskontingent i Trondheim.
- 5 Øvrige saker fremlagt av styret. Ingen øvrige saker fra styret.
- 6 Innkomne forslag. Disse må være styret i hende minst 10 dager før årsmøtet. Ingen innkomne forslag.
- 7 Valg av styre og to revisorer for de neste to år. Leder: Ingeborg Hem Sørmoen (tidligere sekretær). Kasserer: Geir-Arne Fuglstad (fortsetter). Sekretær: Ole Bernhard Forberg (nytt medlem). Nytt styre godkjent av forsamlingen. Revisorer: Jo Eidsvik og Turid Follestad sier seg villige til å fortsette som revisorer. Godkjent av forsamlingen. Takk for god innsats i styret, John, og velkommen til Ole Bernhard!





Det 21. norske statistikermøtet

17. – 20. juni 2024

Invitasjon til det 21. norske statistikermøtet

Det 21. norske statistikermøtet arrangeres 17. - 20. juni 2024 på Quality Hotel i Tønsberg.

Påmeldingen vil åpne i januar og fristen for innsending av bidrag er satt 15. mars.

Det norske statistikermøtet arrangeres annethvert år av en av NSF's lokalforeninger, denne gangen er det Oslo-avdelingen som er arrangør. Målgruppen for møtet er statistikere både fra de ulike universitetsmiljøene, forskningsinstituttene, og ute i "industrien". Tematisk spanner møtet seg fra teoretisk statistikk og sannsynlighetslære, til big data-problematikk, maskinlæring, og rene anvendte og praktiske statistikkproblemer.

Som tidligere år, vil det de første dagene holdes et forkurs (17. - 18. juni).

Program fra tidligere møter finnes på NSF's nettsider, <https://sites.google.com/site/statistiskforening/>.

Nytt fra NTNU

Nyansatte

- Marcus Gehrmann, PhD student
- Daniel R Kjellevoid, PhD student
- Audun Rugstad, PhD student

Nye Mastergrader

- Nora Røhnebæk Aasen: Saddlepoint approximated methods for computing p-values in score tests. Veileder: Thea Bjørnland.
- André Egeli: Probabilistic Forecasting of Solar Radiation in Norway Using Spatial Varying Coefficients for NWP Forecasts. Veileder Ingelin Steinsland.
- Kristoffer Flaglien: Penalizing R2 Through Priors in Linear Regression and Poisson Regression. Veileder Geir-Arne Fuglstad.
- Selma Lerkørød Flo: Physics-Informed Bayesian Calibration Accounting for Model Discrepancy in a Linearized Homogeneous Ordinary Differential Equation. Veileder Ingelin Steinsland.
- Petter Giørtz: Fast Spatial Multi-level Models for Small Area Estimation. Veileder Geir-Arne Fuglstad.
- Eirik Jorstad: Music Generation Using Latent Diffusion Models. Veileder Erlend Aune.
- Pim Keer: Hypothesis testing in contingency tables. Veileder Øyvind Bakke.
- Sivert Laukli: Bayesian Analysis of Carbon Dioxide Production in Land-based Atlantic Salmon using Latent Gaussian State space Modeling. Veileder Ingelin Steinsland.
- Simon Liabø: Modeling Passenger Count Data Based on Automatic Counting. Veileder Ingelin Steinsland.
- Marte Bøe Ludvigsen: Suicide Crisis Syndrome in a Norwegian Acute Psychiatric Unit: Exploring Risk Factors using Statistical Learning and Inference. Veileder: Mette Langaas, Linde Melby and Terje Torgersen.
- Ask Moe Løite: Binary classification of Credit Card Users with Logistic Regression, Gradient Boosted Decision Trees and Deep Learning. Veileder John Tyssedal.

- Martine Middelthon: Delousing Effect and Welfare Implications of Non-Medicinal Delousing methods. Veileder John Tyssedal.
- Ingrid Langevei Mæland: A study of regression models for count data, with applications to salmon lice data. Veileder: Thea Bjørnland.
- Frederick Nilsen: Modeling non-Gaussian traits using Gaussian quantitative genetic models. Veileder Stefanie Muff.
- Johannes Padel: Predicting dry bulk vessel destination using historical AIS data. Veileder Jo Eidsvik.
- Sander Ruud: Deep attentive learning for intraday stock movement and volatility prediction. Veileder Arild Næss.
- Aleksander Johnsen Solberg: Musical Source Separation with Diffusion Models. Veileder Erlend Aune.
- Erling Fause Steen: Statistical analysis and modelling of decision-making among individuals with autistic and psychotic traits in the bunny task. Veileder Håkon Tjelmeland.
- Sunniva Bjelland Strømseng: Predicting Credit Card Activity for Passive Customers. Veileder John Tyssedal.
- Thomas Torkildsen. Stochastic Volatility Models. Veileder Jarle Tufto.
- Jim Totland: Geostatistical Modeling under Positional Uncertainty. Veileder Geir-Arne Fuglstad.
- Martin Gudahl Tufte: Vector Quantized Time Series Generation using Diffusion Models. Veileder Erlend Aune.
- Endre Børge Urheim: Bayesian statistical modelling and analysis of a DAS data set. Veileder Håkon Tjelmeland.

Nye PhD-grader

- Susan Anyosa «Statistical approaches for sampling and monitoring in Earth science applications» Supervisor: Jo Eidsvik, NTNU. Committee: Mimi Zhang, Trinity College, Ireland, Aojie Hong, University of Stavanger, Bob O'Hara, Department of Mathematical Sciences, NTNU
- Silius Mortensønn Vandeskog «Efficient high-dimensional modelling of temperature and extreme precipitation» Supervisor: Sara Martino. Co-supervisors: Ingelin Steinsland, Oddbjørn Bruland Committee: Philippe Naveau, CNRS, France, Emma Simpson, University College London, UK, Geir-Arne Fuglstad, Department of Mathematical Sciences, NTNU

- Kwaku Peprah Adjei «A statistical framework for modelling biodiversity data» Supervisor: Bob O’Hara, NTNU. Committee: Res Altwegg, University of Cape Town, South Africa, Ireland, Janine Illian, University of Glasgow, UK, Sara Martino, Department of Mathematical Sciences, NTNU

Nytt fra Universitetet i Bergen

Nye Mastergrader

- Thuvaragi Saravani: Testing independence of rainfall data. Veileder: Hans Karlsen
- Brage Grindheim: Modelling av overdispersjon i populasjonsdata. Veileder: Hans Skaug
- Marlin Firveld Andersen: Fallacy of a correction factor in automated computer vision-based lice counting: Application of mixed-effects models. Veileder: Jan Bulla
- Erling Bratli: Using Mixed-Effects Models to Analyze Speech Comprehension and Psychological Variables in Patients With Chronic Tinnitus. Veileder: Jan Bulla

Nye PhD-grader

- Anders Sleire: Essays on Portfolio Risk Management and Weather Derivatives. Veiledere: Bård Støve og Jan Bulla.
- Timothée Bacri: Efficient optimization for Hidden Markov Models using Template Model Builder. Veiledere: Bård Støve, Jan Bulla, Geir Berentsen (NHH) og Hans Karlsen.
- Francesca Azzolini: The heritability curve and its applications to human non-Mendelian traits. Veileder: Hans Julius Skaug (UiB) og Geir Berentsen (NHH). Komite: Jarle Tufto, Stephanie Muff (NTNU), Jan Bulla (Uib).