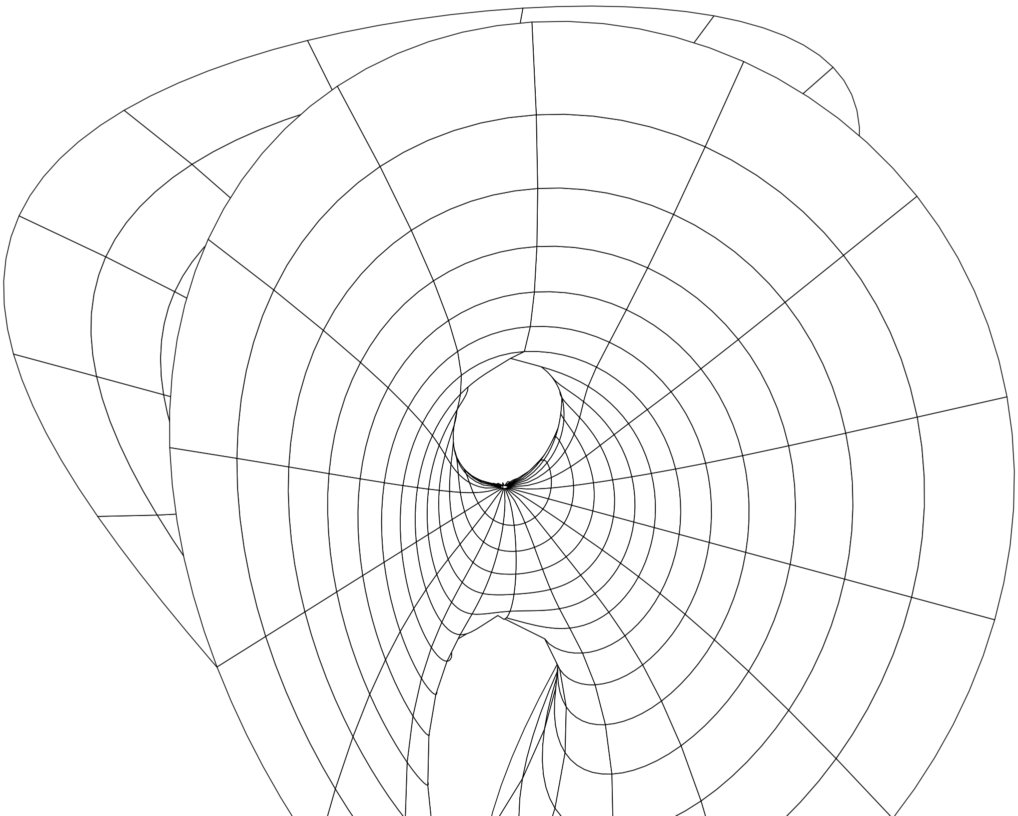
FAMØS

Fagblad for Aktuar, Matematik, -Økonomi og Statistik
12. årgang, nr. 4, maj 1999



Chen-Gackstatter-fladen. Den først opdagne minimalflade med et hul.

FAMØS 12.4; maj 1999.
Fagblad for Aktuar-, Matematik-, Økonomi- og Statistikstuderende ved Københavns Universitet.

Redaktionsgruppe:

Henrik Christian Grove (ansvh.)
Rasmus Borup Hansen
Anders Bo Nielsen

Deadline for næste nummer:
Torsdag den 24. september 99

Indlæg modtages gerne og kan sendes til `famos@math.ku.dk` (meget gerne skrevet i LaTeX), eller afleveres på Matematisk Afdelings sekretariat i E 103.

FAMØS er et internt fagblad.

Fagbladet FAMØS
c/o Institut for matematiske fag
Matematisk Afdeling
Universitetsparken 5
2100 København Ø

World Wide Web adresse:
`http://www.math.ku.dk/famos/`

Tryk: HCØ Tryk

Oplag: 700 stk.

ISSN 1395-2145

Indhold

Leder 3
Nye bidrag fra slipseforskningen . . 4
Markov chain Monte Carlo 9
Forskerdag 1999 20
Choquets sætning – til hverdag og fest 21
Opgaveløsninger 30
Nyt fra naturfagsbogladen 32

Leder

Nu er vejret endelig ved at blive godt, og som så ofte før betyder det at det er tid til at stikke næsen i bøgerne, og få læst pensum inden eksamen. Når denne så om en måned til halvanden forhåbentlig er veloverstået, er der tid til at komme ud og nyde det gode vejr.

Den 23. marts deltog mellem 1.000 og 2.000 studerende på naturvidenskab i demonstrationen på Frue Plads mod udsultningen af naturvidenskab. Arbejdet sluttede absolut ikke dermed, og det sidste resultat var paneldebatten den 6. maj i audiotorium 1, hvor dekanen, en repræsentant fra både undervisning- og forskningsministeriet, Christine Antorini fra SF, Brian Mikkelsen fra det konservative folkeparti og tre repræsentanter fra det private erhvervsliv, redegjorde for deres syn på problemerne og besvarede spørgsmål.

Da institutlederen for Forsikringsmatematisk Laboratorium (FML) Ragnar Norberg, forlader instituttet næste år, har han foreslået at FML lægges ind som en afdeling under Institut for Matematiske Fag.

Ragnar har endvidere foreslået at de tre studienævn (matematik, mat-øk og stat-act) lægges sammen. Der er både gode og dårlige sider ved dette. På plussiden er der udsigten til et forbedret samarbejde, på minus-siden er at det vil bringe diskussionerne længere væk fra de faglige miljøer. Fra studentersiden mener vi at det vil være bedst at bevare de små studienævn og så sikre samarbejdet gennem et koordinationsudvalg.

Man kan finde mange mærkelige ting på internet, blandt de mere brugbare, kan man nu også finde en komplet udgave af Euklids elementer. Kig på
http://aleph0.clarku.edu/~djoyce/java/elements/elements.html

Nye bidrag fra slipseforskningen

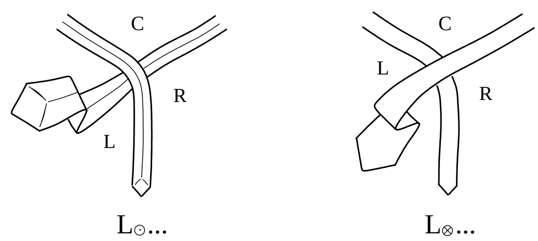
Anders B. Nielsen

Indledning

Det følgende er uddrag fra to spændende artikler skrevet af Thomas M. Fink og Yong Mao (http://www.tcm.phy.cam.ac.uk/∼tmf20/). Artiklerne indeholder oplysninger som er uundværlige, hvis du deltager i formelle sociale arrangementer i ny og næ. Artiklerne er en introduktion til disciplinen „Anvendt slipseknudeteori“ indenfor anvendt knudeteori. Enhver med en smule forstand på slipseknuder ved at det kræver en universitetsuddannelse i matematk for at kunne binde sit slips ordentligt. Hvis du undrer dig over, hvorfor du bliver taget mere seriøst når du bærer jakke og slips kommer svaret her: Det er ikke fordi folk går op i ligegyldige overflader, men snarere fordi de respekterer din store matematiske viden. Lad os starte dette indlæg med at få klargjort hvad vi overhovedet taler om.

Definition af slips

En slipseknude tilhører mængden af „slip“ knuder. Det vil sige knuder som er dannet ved at lægge slipset omkring en hals og hvor den tykke (aktive) ende manipuleres på en sådan måde omkring den tynde (passive) ende at den sidstnævnte kan glide frit igennem den dannede slipseknude. Bemærk at slipseknuder er en ægte delmængde af „slip“knuder.



Figur 1: Områder C, R og L. Retningsindikatorer $\otimes$ og $\odot$.

En slipseknudes første trin består i at lægge den aktive (tykke) ende til venstre eller højre. Enden lægges enten over eller under den passive ende (den tynde). Herved dannes en trekantsformet basis som deler mulighedsrummet op i tre: højre, midten og venstre. Disse benævnes (R, C, L). Se figur 1.

Alle knuder hvis første trin består i at lægge den aktive ende til højre er spejlinger af dem hvis første trin består i at lægge den aktive ende til venstre. Derfor vil vi ignorere de første. Alle de knuder vi viser kan naturligvis udføres spejlvendt. Det sker naturligvis ved at erstatte venstre med højre og vice versa i vejledningerne.

En slipseknude dannes ved at lægge den aktive ende rundt om trekantsbasis. For at gengive hvordan knuderne konstrueres er det praktisk at benytte 6 operatorer.

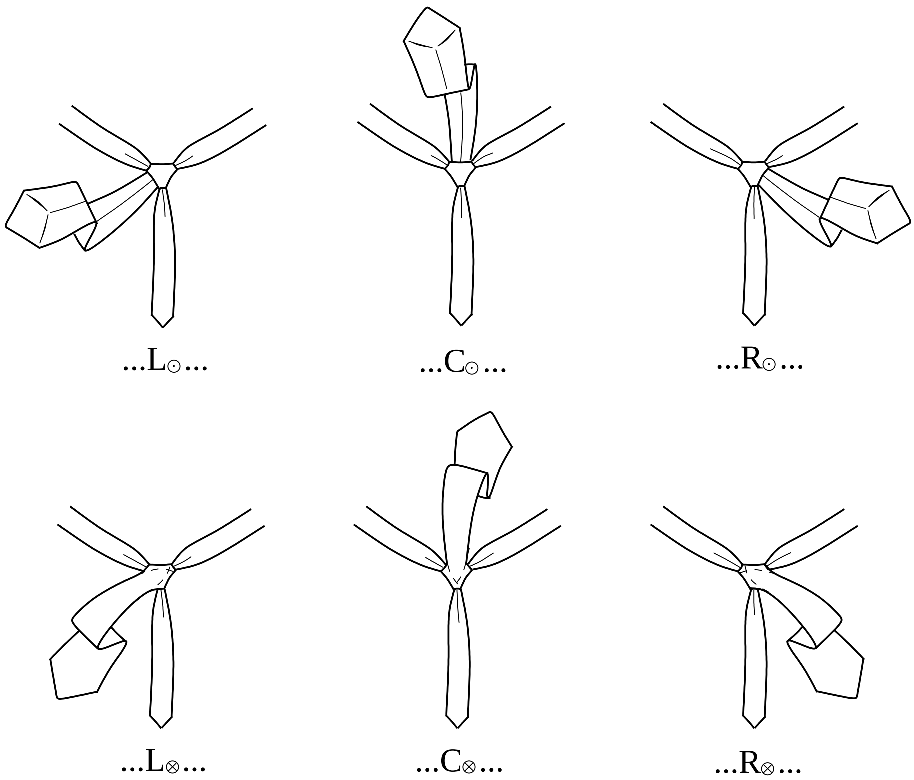
Operatorer

Operatorerne defineres udfra de 6 tilstande:

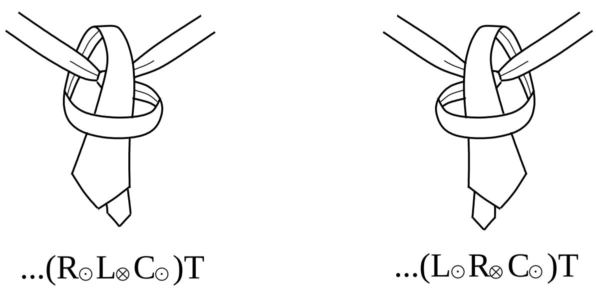
$$\{L\otimes, C\otimes, R\otimes, L\odot, C\odot, R\odot\}$$

Tilstandene er gengivet i figur 2 herunder. R, L og C angiver fra hvilken del af mulighedsrummet den aktive ende kommer fra. $\odot$ og $\otimes$ angiver den retning den aktive ende bevæger sig i. $\odot$ betyder at retningen er udfra skjorten og $\otimes$ betyder at retningen er ind imod skjorten. De to forskellige retningsindikatorer er vist i figur 1.

I alt bliver der dermed 6 forskellige operationer. Se figur 2.



Figur 2: De 6 operationer.



Figur 3: Ender.

Man kan ikke lade $\odot$ eller $\otimes$ efterfølge hinanden. $R\odot$ er den inverse til $R\otimes$. Det giver derfor ikke mening at lade R efterfølge R eller tilsvarende for C og L. En slipseknude består derfor af vekslinger mellem $\odot$ og $\otimes$. En knude afsluttes ved operationen T. Operationen ses i figuren herunder. De 2 slutoperationer i figur 3 er de eneste der bruges.

Vi definerer nu formelt en slipseknude ved en sekvens af operationer fra sættet

$$\{R\odot, R\otimes, C\odot, C\otimes, L\odot, L\otimes\},$$

hvor der startes med enten $L\odot$ eller $L\otimes$ og hvor der afsluttes med en af delfølgerne $R \odot L \otimes C \odot T$ eller $L \odot R \otimes C \odot T$. Ingen følge må have to efterfølgende operatorer med ens retninger eller kildemulighedsområde. Som tidligere nævnt vil to sådanne operationer ophæve hinanden.

Slipseknuden vil bestå af et endeligt antal knuder eftersom et slips har endelig længde.

Succeskriterier for en god knude.

Størrelse Slipseknuderne kategoriseres i henhold til det antal operationer slipseknuden består af. Forfatterne vælger af praktiske grunde at sætte den øvre begrænsning på antallet af operationer til 9. Det er ikke realistisk, at nogen vil bruge en slipseknude med mere end 9 operationer.

Form Formen af en knude karakteriseres ved det relative antal af venstre, højre og midteroperationer. Eftersom vi i næste afsnit indfører s der karakteriserer antallet af venstreoperationer i forhold til antallet af højreoperationer indfører vi γ, antallet af midteroperationer. Det bruges til at klasssificere knuder af samme størrelse, h. En knude med et lavt γ/h index indikerer en smal knude hvorimod et højt γ/h index indikerer en bred knude. Vi begrænser vores eftersøgning til slipseknuder med γ/h index i intervallet $[1/4, 1/2]$

Symmetri For at få en æstetisk tiltalende slipseknude er det ønskeligt at lave slipseknuder der er symmetriske. Symmetriindexet $s = \sum_{i=1}^{h} x_i$ for en knude defineres til at være antallet af højre operationer minus antallet af venstre operationer. Eftersøgningen af slipseknuder begrænses til de knuder der minimerer symmetriindexet s. Det vil sige 0 hvis antallet af venstre og højreoperationer, $h-\gamma$ er lige og -1 eller 1 hvis summen af antallet af venstre og højreoperationer er ulige.

Balance Centeroperationsantallet og symmetriindexet s angiver hvilke operationer en knude indeholder. Balanceindexet $b = \frac{1}{2}\Sigma_{i=2}^{h-1}|\omega_i - \omega_{i-1}|$ beskriver derimod hvorledes operationerne er fordelt. Jo mindre indexet, b, er desto mere jævnt er operationerne fordelt. Dette er den anden æstetiske betingelse. Definitionen af b er således: Kald den $i$te operation σ_i. Omvindingsretningen $\omega_i(\sigma_i, \sigma_{i+1})$ er lig 1 hvis skiftet fra σ_i til σ_{i+1} er med uret(det vil sige hvis (σ_i, σ_{i+1}) er lig $(L, C), (C, R)$ eller (R, L)) og -1 hvis skiftet fra σ_i til σ_{i+1} er mod urets retning.

Løsninger til problemet

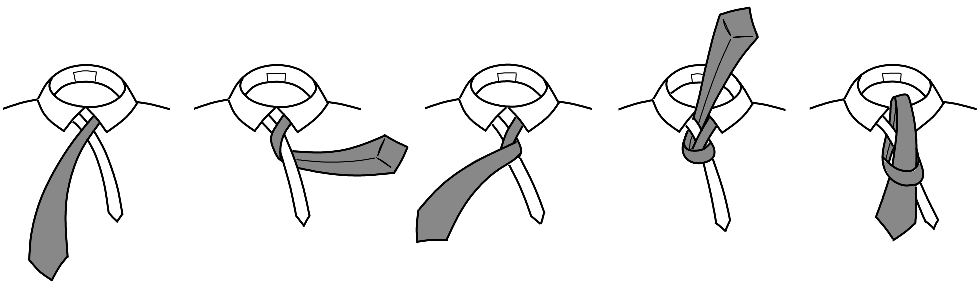
I dette afsnit præsenteres slipseknuder der er gode ifølge vores opstillede kriterier. Alle de 10 løsninger er præsenteret i tabellen nedenfor. Af de 10 løsninger til slipseproblemet er kun 5 allerede kendte.

h	γ	γ/h	$K(h,\gamma)$	s	b	Navn	Følge	KS	SKT
3	1	0.33	1	0	0	Communist knot	$L_\odot R_\otimes C_\odot T$	y	0_1
4	1	0.25	1	-1	1	Four-in-hand	$L_\otimes R_\odot L_\otimes C_\odot T$	n	3_1
5	2	0.40	2	-1	0	Pratt Knot	$L_\odot C_\otimes R_\odot L_\otimes C_\odot T$	n	0_1
6	2	0.33	4	0	0	Half-Windsor	$L_\otimes R_\odot C_\otimes L_\odot R_\otimes C_\odot T$	y	0_1
7	2	0.29	6	-1	1		$L_\odot R_\otimes L_\odot C_\otimes R_\odot L_\otimes C_\odot T$	n	0_1
7	3	0.43	4	0	1		$L_\odot C_\otimes R_\odot C_\otimes L_\odot R_\otimes C_\odot T$	y	3_1
8	2	0.25	8	0	2		$L_\otimes R_\odot L_\otimes C_\odot R_\otimes L_\odot R_\otimes C_\odot T$	y	7_4
8	3	0.38	12	-1	0	Windsor	$L_\otimes C_\odot R_\otimes L_\odot C_\otimes R_\odot L_\otimes C_\odot T$	n	3_1
9	3	0.33	24	0	0		$L_\odot R_\otimes C_\odot L_\otimes R_\odot C_\otimes L_\odot R_\otimes C_\odot T$	y	4_1
9	4	0.44	8	-1	2		$L_\odot C_\otimes R_\odot C_\otimes L_\odot C_\otimes R_\odot L_\otimes C_\odot T$	n	5_2

Først skal det nævnes at det ikke er alle sociale lejligheder der lige velegnet til brugen af slips og dermed slipseknuder. Eksempelvis er jakkesæt og slips ikke velegnet til visse udendørs sociale arrangementer såsom finaleopgør mellem Brøndby og FCK. Heller ikke til møder i politiske foreninger til venstre for socialdemokratiet. Dog er der en undtagelse for denne regel. Denne nævnes herunder.

Kommunistisk ungdomsforenings slips Man bør undgå at bruge den første slipseknude i tabellen, hvis man er inkarneret liberalistisk ideolog. Denne slipseknude bruges hyppigt af medlemmer af den kommunistiske ungdomsorganisation i Kina.

Four-in-hand Four-in-hand er den anden slipseknude i tabellen. Det er den simpleste af de 4-5 kendte slipseknuder. Den stammer fra sent i det 19. århundrede i England.



Figur 4: Four-in-hand knuden.

Pratt Pratt knuden er den tredje i tabellen. Pratt knuden blev præsenteret på forsiden af New York Times i 1989. Pratt knuden blev officielt registreret hos "Neckwear Association of America"i 1989. Knuden var den første nye slipseknude i 50 år. Med Fink og Maos artikel er der blevet taget et tigerspring (kvantespring) for slipsefanatikere. Med deres artikel er resten af de praktiske slipseknuder blevet fundet på en systematisk og hurtig måde. Uden denne artikel kunne vi have vente mange årtier før vi havde fundet de 5 nye slipseknuder.

Windsor Windsor knuden er nummer 8 i tabellen. Hertugen af Windsor er krediteret for at have fundet på denne knude omkring år 1936. Windsorknuden er relativt kompliceret. Det er den mest komplicerede af de 4-5 almindeligt brugte slipseknuder.

Half-Windsor Half-Windsor knuden er nummer 4 i tabellen. Som navnet antyder er knuden et afkom af Windsor knuden. På trods af navnet er knudens størrelse trefjerdedele af Windsor knudens størrelse snarere end halv størrelse.

De 5 nye slipseknuder Tilbage er der 5 ukendte slipseknuder i tabellen. Disse 5 knuder er dermed et af Fink og Maos bidrag til denne verden. Da 3 af de 5 nye slipseknuder er mindre komplicerede end den gammelkendte Windsor knude er der ingen undskyldninger for ikke progressivt at indføre de nye spændende knuder. Dette vil uden tvivl pifte det danske indavlede selskabsliv gevaldigt op.

Opfordring og konklusion

Bidraget fra Fink og Mao har potentiale til gøre denne verden til et mere spændende sted at være. Initiativet er nu op til jeg kære læsere. Herfra kan jeg kun opfordre til at bidrage det danske selskabsliv nye eksotiske indtryk: brug de nye knuder aktivt. Til slut vil jeg opfordre til at oprette en dansk forening, der svarer til det amerikanske Neckwear Association of America.

Markov chain Monte Carlo

Søren Feodor Nielsen
Afdelingen for Teoretisk Statistik

Introduktion

Markov chain Monte Carlo – på dansk burde det hedde Markovkæde Monte Carlo, men alle bruger tilsyneladende det engelske udtryk eller forkortelsen MCMC – er et meget aktivt forskningsområde i statistik i dette årti. I virkeligheden er det, som vi skal se, bare en metode til at udregne visse summer (eller mere generelt integraler), men anvendelserne indenfor statistik (og fysik!) er vigtige og utallige.

Teorien bag er i sin simpleste form ikke særlig svær. Jeg vil vove den påstand, at man i løbet af et kursus som STATISTIK 0 (eller MAT 2SS?) lærer nok til at man kan forstå *alt* hvad der vil foregå i denne artikel! Bevares, nogle af argumenterne er nok sværere end hvad man udsættes for i indledende kurser i statistik og sandsynlighedsregning, men ingen af dem kræver begreber, som ikke er gennemgået på STATISTIK 0.

I næste afsnit vil jeg give et eksempel på et statistisk problem, som kan løses vha. MCMC. Dernæst skal vi gennem en introduktion til Markovkæder på endelige tilstandsrum – bare rolig, ordene bliver defineret i afsnittet „Markovkæder på endelige tilstandsrum"; det er her alt det tekniske står. Der er også nogle få øvelser. Lad være med at regne dem! De indeholder bare forskellige fakta, som jeg ikke gider bevise, og som ikke er sværere end du kan vise dem selv, hvis du ikke tror mig. Endelig benytter vi de fundne resultater til at løse det problem, vi begyndte med.

God fornøjelse.

Motivation

Som motiverende eksempel skal vi se på et datamateriale om alkoholpromillens betydning for dødelige trafikulykker. Materialet skyldes Anders Siboni og Jørgen B. Dalgaard (*Ugeskrift for læger*, 1986; 2790-92) og er gengivet i tabel 1. Vi vil nøjes med at se på om der er en sammenhæng mellem ulykkestypen (hvad bilisten er kørt ind i) og alkoholpromillen. Normalt vil vi gøre dette vha. Pearsons χ^2-teststørrelse

$$K = \sum_{i,j} \frac{(X_{ij} - E_{ij})^2}{E_{ij}}$$

hvor X_{ij} er antallet i celle (i, j) i tabellen og $E_{ij} = \frac{X_{i\cdot}X_{\cdot j}}{X_{\cdot\cdot}}$ er de forventede under hypotesen om uafhængighed mellem ulykkestype og alkoholpromille. Hvis K er stor, svarer det til, at der er stor forskel på de observerede og de forventede under hypotesen, og vi vil så afvise hypotesen.

	Uheldstype: Bilist mod			
Promille	Cyklist/knallert	Fodgænger	Bilist	Alene
0.10-0.79	5	2	7	4
0.80-1.19	3	8	8	3
1.20-1.59	8	5	8	11
1.60-3.19	12	8	22	34

Tabel 1: Blodalkoholkoncentration i ulykkesøjeblikket hos 148 bilister aktivt involveret i dødelige trafikulykker, opdelt efter promille og ulykkens art (STAT 0, opg 4.14)

For vores datamateriale er teststørrelsen $16,4767$. Pearsons χ^2-teststørrelse er approksimativt $\chi^2_{(4-1)(4-1)}$-fordelt, hvis $E_{ij} \geq 5$, men det er ikke tilfældet her. For eksempel er det forventede antal ulykker af typen „bilist mod fodgænger" ved lav alkoholpromille 2,8. Så vi har et problem.

Det, vi er interesserede i, er testsandsynligheden, dvs. $P\{K > 16,4767\}$ beregnet under hypotesen. Når vi nu ikke kan bruge χ^2-approksimationen, så er vi nødt til at beregne testsandsynligheden. Det første problem, vi støder ind i her, er, at fordelingen af $X = (X_{ij})_{i=1,\ldots,4,j=1,\ldots,4}$ – og dermed fordelingen af K – afhænger af de marginale fordelinger af rækkesummer og søjlesummer, som jo er ukendte parametre.

Løsningen på dette problem er kun at sammenligne vores observerede tabel med visse 4×4-tabeller i stedet for med alle 4×4-tabeller. Lad S være mængden af 4×4-tabeller med samme række- og søjle-summer som tabellen i vores datamateriale, og lad $p = (p_i)_{i=1,\ldots,4}$, hhv. $q = (q_j)_{j=1,\ldots,4}$ være de ukendte sandsynlighedsparametre i fordelingen af rækkesummer, hhv. søjlesummer. Så er den betingede fordeling af en tabel betinget med at den har samme række- og søjle-summer som den observerede givet ved

$$
\begin{aligned}
P\{X = x | X \in S\} &= \frac{P\{X = x, X \in S\}}{P\{X \in S\}} = \frac{P\{X = x\}}{P\{X \in S\}} \\
&= \frac{\dfrac{x_{..}}{\prod_{ij} x_{ij}} \cdot \prod_{i=1}^{4} p_i^{x_{i\cdot}} \prod_{j=1}^{4} q_j^{x_{\cdot j}}}{\sum_{x' \in S} \dfrac{x'_{..}}{\prod_{ij} x'_{ij}} \cdot \prod_{i=1}^{4} p_i^{x'_{i\cdot}} \prod_{j=1}^{4} q_j^{x'_{\cdot j}}} \\
&= \frac{\dfrac{1}{\prod_{ij} x_{ij}}}{\sum_{x' \in S} \dfrac{1}{\prod_{ij} x'_{ij}}} \qquad (1)
\end{aligned}
$$

for $x \in S$. Det vigtige her er ikke så meget udtrykket (1), men det at det ikke afhænger af de ukendte parametre, p og q. Det betyder nemlig, at fordelingen af X

givet række- og søjle-summer er en kendt fordeling. Dermed er også fordelingen af K *givet* $X \in S$ en kendt fordeling. Vi kan altså *i princippet* godt beregne

$$P\{K \geq 16,4767|X \in S\} = \sum_{x \in S : K(x) \geq 16,4767} P\{X = x|X \in S\} \tag{2}$$

hvor summen altså går over alle tabeller, x, med række- og søjlesummer som den observerede tabel, men med en Pearson-teststørrelse, $K(x)$, større end den, vi har observeret.

Dette kaldes et *betinget test*. Fidusen ved at betinge er, at vi får noget, som kun afhænger af det, vi egentlig er interesserede i at teste. I vores tilfælde er vi interesseret i at teste om der er uafhængighed, mens vi er ganske lige glade med de marginale fordelinger, p og q. Ved at betinge får vi et test, som ikke afhænger af de irrelevante parametre, p og q, men kun den eventuelle afhængighed; hvis teststørrelsen bliver for stor, kan det kun skyldes afhængighed!

Når der ovenfor står, at vi "i princippet" kan beregne (2), så skyldes det, at det ikke er let at finde alle de tabeller, som har samme række- og søjle-summer som vores datamateriale, og dermed bliver det svært at beregne summen direkte.

Lad nu $K_1, \ldots, K_n$ være uafhængige, identisk fordelte stokastiske variable med fordeling givet ved

$$P\{K_i = k|X \in S\} = \sum_{x \in S : K(x) = k} P\{X = x|X \in S\} \tag{3}$$

og sæt $Y_i = 1_{\{K_i \geq 16,4767\}}$. Så er $Y_1, \ldots, Y_n$ uafhængige, identisk fordelte stokastiske variable med middelværdi (2). Derfor vil for et vilkårligt $\varepsilon > 0$

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - E(Y_1)\right| \geq \varepsilon\right\} \to 0 \tag{4}$$

når $n \to \infty$, ifølge Store tals lov.

Vi kan altså beregne testsandsynligheden (2), hvis vi har mange uafhængige teststørrelser, $K_1, \ldots, K_n$, med den rette fordeling ved at tage et gennemsnit. Denne idé kaldes *Monte Carlo integration*. For at få fat på sådanne teststørrelser, kan vi simulere, dvs. få en computer til at generere dem tilfældigt.

Det er igen ikke helt simpelt at se, hvordan vi kan generere teststørrelser med den rette fordeling. Vi skal i virkeligheden heller ikke gøre præcis dette men i stedet noget, der er meget tæt på. Men først skal vi vide noget om Markovkæder på endelige tilstandsrum.

Markovkæder på endelige tilstandsrum

Definition af Markovkæder

Inden vi formelt definerer Markovkæder, skal vi indføre en smule notation. I dette afsnit vil vi lade S være en vilkårlig *endelig* mængde. $\mu = (\mu_i)_{i \in S}$ er en fordeling på

S dvs. en (række-)vektor med ikke-negative elementer, som opfylder at $\sum_{i \in S} \mu_i = 1$. Endelig skal $\mathcal{P} = [p_{ij}]_{i,j \in S}$ være en ($S \times S$-)matrix, således at hver række i matricen er en fordeling på S; dvs. $p_{ij} \geq 0$ og $\sum_{j \in S} p_{ij} = 1$.

Øvelse: Vis at $\mu\mathcal{P}$ (et almindeligt matrixprodukt) er en fordeling på S.

Definition 1. *Vi siger, at en følge $(X_n)_{n \in \mathbf{N}_0}$ af stokastiske variable er en Markovkæde med tilstandsrum S, begyndelsesfordeling μ og overgangssandsynligheder $\mathcal{P}$, hvis*

(i) $P\{X_0 = i\} = \mu_i$ *for alle* $i \in S$

(ii) $P\{X_n = j | X_{n-1} = i\} = p_{ij}$ *for alle* $n \in \mathbf{N}$ *og alle* $i, j \in S$.

En Markovkæde er altså en følge af stokastiske variable med værdier i tilstandsrummet S således at fordelingen af den første (X_0) er begyndelsesfordelingen, mens fordelingen af de øvrige er givet ved betingede fordelinger givet den forrige. Vi siger, at p_{ij} er *sandsynligheden for at gå fra i til j*.

Øvelse: Vis at fordelingen af X_n givet X_0 er $(p_{ij}^{(n)})_{j \in S}$, hvor $(p_{ij}^{(n)})_{j \in S}$ er den $i$te række i matricen $\mathcal{P}^n$ ($\mathcal{P}$ ganget med sig selv n gange). (*Vink: Benyt induktion.*) Vi siger, at $p_{ij}^{(n)}$ er *sandsynligheden for at gå fra i til j i n skridt*. Vis også at $\mu\mathcal{P}^n$ netop er fordelingen af X_n.

Hvordan ville vi lave os en Markovkæde, hvis vi skulle? Jo, det er simpelt nok. Vi starter med at simulere X_0 (på en computer), således at fordelingen af X_0 netop er μ. Dernæst, når vi har værdien af X_0 – lad os sige at $X_0 = i$ – simuleres X_1 så den har fordelingen $(p_{ij})_{j \in S}$. Når vi så har værdien af X_1, kan værdien af X_2 simuleres på nøjagtig samme måde. Altså

Algoritme til simulering af Markovkæde

X_0 `trækkes fra fordelingen` μ
`n<-1`
`begin`
  `givet` $X_{n-1} = i$ `trækkes` X_n `fra fordelingen` $(p_{ij})_{j \in S}$
  `n<-n+1`
`end`

Det er nyttigt at tænke på Markovkæder, som noget vi kan simulere vha. denne algoritme, og jeg vil ofte henvise til denne tankegang i de følgende beviser. For eksempel er følgende øvelse triviel, hvis man tænker „algoritmisk“.

Øvelse: Vis at $(X_{nm})_{n \in \mathbf{N}_0}$ er en Markovkæde (med overgangssandsynligheder $\mathcal{P}^m$) for et vilkårligt $m \in \mathbf{N}$, hvis $(X_n)_{n \in \mathbf{N}_0}$ er en Markovkæde.

Vi skal gøre os nogle yderligere antagelser. Først og fremmest skal vi antage, at der er positiv sandsynlighed for at komme fra en vilkårlig tilstand i til en vilkårlig anden tilstand j; ikke nødvendigvis i et skridt, men i et eller andet antal skridt. Med andre ord antager vi, at for alle $i, j \in S$ findes et $n \in \mathbf{N}$ således at $p_{ij}^{(n)} > 0$; bemærk at n må afhænge af i og j. Når denne antagelse er opfyldt, siger vi, at Markovkæden er *irreducibel.*

Vi vil også antage, at der findes en tilstand $i_0 \in S$ således at der er positiv sandsynlighed for at blive i denne tilstand, dvs. at sandsynligheden for at gå fra i_0 til i_0 er positiv, $p_{i_0 i_0} > 0$. I modsætning til antagelsen af irreducibilitet, er denne antagelse ikke strengt nødvendig – mindre kan gøre det – men den er opfyldt i de tilfælde, jeg er interesseret i, og den simplificerer de følgende argumenter meget.

Vi skal ud fra disse to betingelser vise to resultater om Markovkæder. Først viser vi at „n'et i irreducibiliteten" ikke behøver at afhænge af i og j.

Lemma 2. *Der findes $N \in \mathbf{N}$ således at for alle $n \geq N$ og alle $i, j \in S$ er $p_{ij}^{(n)} > 0$.*

Bevis: For alle $i, j \in S$ findes der $n_1, n_2 \in \mathbf{N}$ (som kan afhænge af i og j), således at

$$p_{ii_0}^{(n_1)} > 0 \qquad \text{og} \qquad p_{i_0 j}^{(n_2)} > 0$$

Derfor er $p_{ij}^{(n_1+n_2)} \geq p_{ii_0}^{(n_1)} p_{i_0 j}^{(n_2)} > 0$; sandsynligheden for at gå fra i til j i $n_1 + n_2$ skridt er ikke mindre ned sandsynligheden for først at gå fra i til i_0 i n_1 skridt og derefter at gå fra i_0 til j i n_2 skridt. På samme måde er $p_{ij}^{(n)} \geq p_{ii_0}^{(n_1)} p_{i_0 i_0}^{n-n_1-n_2} p_{i_0 j}^{(n_2)} > 0$ for alle $n > n_1 + n_2$; her kræver vi jo, at vi skal blive stående i i_0 i $n - n_1 - n_2$ skridt. Dermed har vi vist, at for alle $i, j \in S$ findes et N (f.eks. $n_1 + n_2$) så $p_{ij}^{(n)} > 0$ for $n \geq N$. Vi mangler at vise, at dette N kan vælges, så det ikke afhænger af $i, j \in S$. Men det er nemt; der er jo kun endeligt mange $i$er og $j$er og dermed kun endeligt mange Ner, så vi kan bare vælge det største. □

Det næste resultat er lidt mere dybt. Lad os for hvert $i \in S$ definere τ_i som det første tidspunkt n, hvor Markovkæden besøger tilstand i. F.eks. hvis $X_0 \neq i$ men $X_1 = i$, så er $\tau_i = 1$. τ_i antager altså værdier i $\mathbf{N}_0 \cup \{\infty\}$, hvor $\tau_i = \infty$ betyder, at Markovkæden aldrig besøger tilstand i. Vi kunne frygte, at dette kunne ske med positiv sandsynlighed, altså at $P\{\tau_i = \infty\} > 0$. Det næste lemma siger, at det ikke er tilfældet:

Lemma 3. *For alle $i \in S$ er $P\{\tau_i = \infty\} = 0$.*

Bevis: Først bemærker vi, at det er nok at indse at $P\{\tau_i > t\} \to 0$ når $t \to \infty$. Dernæst erindrer vi, at vi kan vælge n så $p_{ji}^{(n)} > 0$ for alle $j \in S$. Igen fordi tilstandsrummet er endeligt, kan vi sikre, at der findes $\varepsilon > 0$ således at $p_{ji}^{(n)} > \varepsilon$ for alle $j \in S$ (vælg $\varepsilon = \min_{j \in S} p_{ji}^{(n)}$). Sandsynligheden for at vi til tid n er i tilstand i er dermed ikke mindre end ε uanset hvilken tilstand vi starter i. Vi indser, at sandsynligheden for at $X_n = i$ givet $X_0 \neq i$ er mindst ε:

$$\begin{aligned} P\{X_n = i | X_0 \neq i\} &= \frac{P\{X_n = i, X_0 \neq i\}}{P\{X_0 \neq i\}} = \frac{\sum_{j \neq i} \mu_j p_{ji}^{(n)}}{\sum_{j \neq i} \mu_j} \\ &\geq \varepsilon \frac{\sum_{j \neq i} \mu_j}{\sum_{j \neq i} \mu_j} = \varepsilon \end{aligned}$$

På samme måde (tænk på hvordan vi simulerer en Markovkæde!) er sandsynligheden for at $X_{mn} = i$ givet $X_{(m-1)n} \neq i$ mindst ε for alle $m \in N$. Nu er (tænk på hvordan vi ville simulere en Markovkæde!)

$$\begin{aligned}P\{\tau_i > mn\} &= P\{X_0, X_1, \ldots, X_{nm} \neq i\} \leq P\{X_0, X_n, \ldots, X_{mn} \neq i\} \\ &= P\{X_0 \neq i\}P\{X_n \neq i | X_0 \neq i\} \ldots P\{X_{mn} \neq i | X_{(m-1)n} \neq i\} \\ &\leq (1-\varepsilon)^m \to 0 \quad \text{når } m \to \infty\end{aligned}$$

hvilket viser det ønskede. □

Konvergens af overgangssandsynligheder

Vi får behov for endnu en antagelse, nemlig at der findes en fordeling π på S, således at hvis X_0 har fordelingen π, så har X_1 også fordelingen π. Vi kalder en sådan fordeling for en *stationær begyndelsesfordeling*. I virkeligheden er dette slet ikke en antagelse, fordi de to antagelser, vi allerede har gjort, faktisk er nok til at sikre eksistensen af en stationær begyndelsesfordeling, men at vise dette er mere end vi får plads til i denne artikel. Desuden vil vi i de tilfælde, vi er interesserede i, vide at den stationære begyndelsesfordeling eksisterer.

Øvelse: Vis at π er en stationær begyndelsesfordeling netop hvis $\sum_{i\in S} \pi_i p_{ij} = \pi_j$ for alle $j \in S$. Vis også at hvis X_0 har fordelingen π, så er fordelingen af X_n også π. (*Vink: Induktion.*)

Med denne antagelse kan vi nu vise hvad der faktisk er hovedresultatet for de Markovkæder, vi betragter her.

Sætning 4. *For alle $i, j \in S$ vil $p_{ij}^{(n)} \to \pi_j$ når $n \to \infty$*

Inden vi giver beviset for sætningen, skynder vi os at give et korollar:

Korollar 5. *For alle begyndelsesfordelingen μ vil $P\{X_n = j\} \to \pi_j$ når $n \to \infty$ for alle $j \in S$.*

Bevis (for korollar): Dette følger direkte af sætningen, fordi $P\{X_n = j\} = \sum_{i\in S} \mu_i p_{ij}^{(n)}$ og summen er endelig. □

Inden vi viser sætningen, skal vi lige bemærke hvad den egentlig siger. Den siger, at fordelingen af X_n givet $X_0 = i$, når n er stor, ligger tæt på den stationære begyndelsesfordeling. Med andre ord vil Markovkæden uanset den faktisk begyndelsesfordeling (se korollaret) glemme sin begyndelsesfordeling og stabilisere sig (i fordeling) omkring den stationære begyndelsesfordeling. Nu til beviset:

Bevis (for sætningen): Lad $(X_n)_{n\in\mathbf{N}_0}$ være en Markovkæde med $P\{X_0 = i\} = 1$ og overgangssandsynligheder $\mathcal{P}$. Lad $(X'_n)_{n\in\mathbf{N}_0}$ være en anden Markovkæde med overgangssandsynligheder $\mathcal{P}$ men med begyndelsesfordeling π, uafhængig af den første Markovkæde. Ud fra disse to Markovkæder kan vi danne en tredje Markovkæde $(Y_n)_{n\in\mathbf{N}_0}$ ved at sætte $Y_n = (X_n, X'_n)$. At dette er en Markovkæde på $S \times S$ er klart; tænk på hvordan vi ville simulere! Den er også irreducibel,

$$P\{Y_n = (j, j')|Y_0 = (i, i')\} = P\{X_n = j|X_0 = i\}P\{X'_n = j'|X_0 = i'\} > 0$$

bare $n \geq N$ fra lemma 2. Lad os nu se på

$$\begin{aligned}
|p_{ij}^{(n)} - \mu_j| &= |P\{X_n = j\} - P\{X'_n = j\}| \\
&= |P\{X_n = j, X_n = X'_n\} + P\{X_n = j, X_n \neq X'_n\} \\
&\quad - P\{X'_n = j, X_n = X'_n\} - P\{X'_n = j, X_n \neq X'_n\}| \\
&= |P\{X_n = j, X_n \neq X'_n\} - P\{X'_n = j, X_n \neq X'_n\}| \leq 2P\{X_n \neq X'_n\}
\end{aligned}$$

Men den sidste sandsynlighed er jo mindre end sandsynligheden for at Y-kæden endnu ikke har besøgt (f.eks.) tilstanden (j, j) til tid n, og denne går jo mod 0 når $n \to \infty$ ifølge lemma 3. □

Store tals lov

Vi er nu klar til at vise Store tals (svage) lov for de Markovkæder, vi er interesserede i. Vi starter med at minde om Chebychevs ulighed:

Lemma 6. *Hvis $E(Z^2)$ eksisterer, er $P\{|Z| > \varepsilon\} \leq \frac{1}{\varepsilon^2} E(Z^2)$ for alle $\varepsilon > 0$.*

Beviset findes f.eks. i Stat 0-noterne. Nu til Store tals lov

Sætning 7. *For enhver funktion $h\colon S \to \mathbf{R}$ og alle $\varepsilon > 0$ vil*

$$P\left\{\left|\frac{1}{n}\sum_{k=0}^{n-1} h(X_k) - Eh(X_0)\right| > \varepsilon\right\} \to 0$$

når $(X_n)_{n \in \mathbf{N}_0}$ er en Markovkæde med $P\{X_0 = i\} = \pi_i$ for $i \in S$.

Bevis: Det følger af Chebychevs ulighed, at det er nok at vise at variansen på gennemsnittet $\frac{1}{n}\sum_{k=0}^{n-1} h(X_k)$ går mod 0, fordi $Eh(X_k) = Eh(X_0)$ for alle $k \in \mathbf{N}$ pga den stationære begyndelsesfordeling. Denne varians er

$$Var\left(\frac{1}{n}\sum_{k=0}^{n-1} h(X_k)\right) = \frac{1}{n^2}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} Cov(h(X_k), h(X_l)) \tag{5}$$

Lad os se på $Cov(h(X_k), h(X_l))$ for $k \leq l$. Den er

$$\begin{aligned}
Cov(h(X_k), h(X_l)) &= \sum_{i \in S}\sum_{j \in S} [h(i) - Eh(X_0)][h(j) - Eh(X_0)]\pi_i p_{ij}^{(l-k+1)} \\
&= Cov(h(X_0), h(X_{l-k+1}))
\end{aligned}$$

fordi fordelingen af X_k er π, og fordelingen af X_l givet X_k netop er $(l-k+1)$-trinsovergangssandsynlighederne $\mathcal{P}^{l-k+1}$. Det betyder at vi kan simplificere (5). Man indser (forholdsvis) let at

$$\begin{aligned}
Var\left(\frac{1}{n}\sum_{k=0}^{n-1} h(X_k)\right) &= \frac{1}{n^2}[n\,Var(h(X_0)) + 2\sum_{k=1}^{n-1}(n-k)Cov(h(X_0), h(X_k))] \\
&= \frac{1}{n}Var(h(X_0)) + \frac{2}{n}\sum_{k=1}^{n-1}\frac{(n-k)}{n}Cov(h(X_0), h(X_k))
\end{aligned}$$

Her går det første led tydeligvis mod 0. I det følgende gennemsnit er $\frac{(n-k)}{n} < 1$ og

$$\begin{aligned} Cov(h(X_0), h(X_k)) &= \sum_{i\in S}\sum_{j\in S}[h(i) - Eh(X_0)][h(j) - Eh(X_0)]\pi_i p_{ij}^{(k)} \\ &\to \sum_{i\in S}\sum_{j\in S}[h(i) - Eh(X_0)][h(j) - Eh(X_0)]\pi_i \pi_j = 0 \end{aligned}$$

når $k \to \infty$ ifølge sætning 4. Dermed går ledene i gennemsnittet mod 0 og det sikrer at gennemsnittet går mod 0 når $n \to \infty$ (et ε-argument lyder „for vilkårligt $\varepsilon > 0$ vælges K så den numeriske led af led med $k > K$ er mindre end ε. Så er grænsen af gennemsnittet mindre end ε"). Dette viser det ønskede. □

Bemærkning. Konklusionen i Store tals lov for Markovkæder holder også selv om vi ikke bruger den stationære begyndelsesfordeling men bare har $P\{X_0 = i\} = \mu_i$ for $i \in S$. Det burde virke rimeligt nok: Fordelingen af X_n når n er stor er jo tæt på den stationære begyndelsesfordeling, så derfor burde gennemsnittet, hvor de første endeligt mange led ikke betyder noget for grænsen, ikke ligge langt fra det vi får ved brug af sætningen. Denne mere generelle udgave af sætning 3 er meget vigtig i andre anvendelser end den, vi er interesserede i. Man kan uden det store besvær modificere beviset for sætning 3 en lille smule og derved bevise det mere generelle resultat.

Store tals lov sikrer, at vi kan bruge simulerede Markovkæder til at udregne summer med. For at beregne $\sum_{i\in S} h(i)\pi_i$, skal vi bare simulere en Markovkæde med stationær begyndelsesfordeling π og benytte at

$$\sum_{i\in S} h(i)\pi_i = Eh(X_0) \approx \frac{1}{n}\sum_{k=0}^{n-1} h(X_k)$$

for n stor. Det er dette, der er idéen i Markov chain Monte Carlo!

Metropolis-Hastings algoritmen

Det eneste, vi endnu mangler, er at hitte på en passende Markovkæde, dvs. en Markovkæde med en given stationær begyndelsesfordeling π. Her kommer Metropolis-Hastings algoritmen os til hjælp.

Lad $\mathcal{Q} = [q_{ij}]_{i,j\in S}$ være en vilkårlig overgangsmatrix på S; den behøver ikke have noget med π at gøre. Vi skal kalde denne for *forslagsfordelingen.* Definer så en ($S \times S$-)matrix $\mathcal{A} = [\alpha_{ij}]_{i,j\in S}$ ved

$$\alpha_{ij} = \begin{cases} 1 & \text{hvis } 0 \le \pi_i q_{ij} \le \pi_j q_{ji} \\ \frac{\pi_j q_{ji}}{\pi_i q_{ij}} & \text{hvis } \pi_j q_{ji} \le \pi_i q_{ij} \end{cases} \tag{6}$$

for $i, j \in S$. α_{ij}erne ligger i $[0;1]$, og vi kalder dem *acceptsandsynligheder.* Metropolis-Hastings algoritmen forløber således:

Metropolis-Hastings algoritmen

```
X_0 simuleres fra π
n<-1
begin
   givet X_{n-1} = i simuleres X fra fordelingen (q_ij)_{j∈S}
   U simuleres fra en ligefordeling på [0;1]
   if U<α_{i,X} then X_n<-X
   else X_n<-X_{n-1}
   n<-n+1
end
```

Altså, vi bruger forslagsfordelingen $\mathcal{Q}$ til at generere et *forslag* j (baseret på den nuværende værdi i) og *accepterer* forslaget med sandsynlighed $\alpha_{ij} = P\{X_n = X|X_{n-1} = i, X = j\} = P\{U \leq \alpha_{ij}\}$. Ellers beholder vi den gamle værdi.

Metropolis-Hastings algoritmen giver os tydeligvis en Markovkæde på S. Overgangssandsynlighederne for $i \neq j$ er givet ved

$$p_{ij} = P\{X = j|X_0 = i\}P\{X_1 = X|X_0 = i, X = j\} = q_{ij}\alpha_{ij}$$

(vi skal først foreslå et spring til j fra i og derefter acceptere det), mens sandsynligheden for at blive stående er 1 minus sandsynligheden for at springe væk, dvs.

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij}$$

Vi skal vise, at π er en stationær begyndelsesfordeling for Metropolis-Hastings algoritmen.

Lemma 8. *π er en stationær begyndelsesfordeling for Metropolis-Hastings algoritmen med forslagsfordeling $\mathcal{Q}$ og acceptsandsynligheder givet ved (6).*

Bevis. Vi skal vise, at $\sum_{i \in S} \pi_i p_{ij} = \pi_j$. Først observerer vi at for $i \neq j$ er

$$\pi_i p_{ij} = \pi_i \alpha_{ij} q_{ij} = \begin{cases} \pi_i q_{ij} & \text{hvis } \pi_i q_{ij} \leq \pi_j q_{ji} \\ \pi_j q_{ji} & \text{hvis } \pi_i q_{ij} \leq \pi_j q_{ji} \end{cases} = \min(\pi_i q_{ij}, \pi_j q_{ji})$$

hvorfor $\pi_i p_{ij} = \pi_j p_{ji}$ for alle $i, j \in S$ (tilfældet $i = j$ er naturligvis trivielt). Dermed er

$$\sum_{i \in S} \pi_i p_{ij} = \pi_j \sum_{i \in S} p_{ji} = \pi_j$$

som viser det ønskede. □

I konkrete tilfælde er det let nok at vise, at der findes en tilstand $i_0 \in S$ med $p_{i_0 i_0} > 0$; man skal bare finde en tilstand, hvor der er positiv sandsynlighed for at man ikke accepterer at springe (dvs. vi skal lede efter en $\alpha_{i_0,j} < 1$!). Vi kan endda være på den sikre side og vælge $\mathcal{Q}$ så $q_{ii} > 0$. At vise irreducibilitet er ofte noget sværere; den antagelse er ikke automatisk opfyldt.

Tilbage til eksemplet

Lad os nu vende tilbage til testproblemet i afsnittet „Motivation“. Med det vi har lært i sidste afsnit, ser vi, at vi kan finde testsandsynligheden (2) ved Markov chain Monte Carlo integration. Vi skal altså bare lave en Markovkæde med (1) som stationær begyndelsesfordeling, og det kan vi gøre vha. Metropolis-Hastings algoritmen.

Helt konkret vil vi gøre følgende. Ud fra en given tabel $x \in S$ vil vi tilfældigt vælge to rækker, nummer r_1 og r_2, og to søjler, nummer s_1 og s_2. Vores forslag x' er så den tabel, vi får, hvis vi lægger 1 til $x_{r_1s_1}$ og $x_{r_2s_2}$ og trækker 1 fra $x_{r_1s_2}$ og $x_{r_2s_1}$ (se figur 1). Det er klart, at vi får en ny tabel med de samme række- og søjlesummer som den forrige. Det eneste problem, vi kan få, er, at vi kunne risikere at $x'_{r_1s_2}$ eller $x'_{r_2s_1}$ kunne blive negative (vi trækker jo 1 fra). Sker dette, lader vi bare være med at acceptere forslaget og bliver stående (for et sådant forslag er $\pi_{x'} = 0$).

Acceptsandsynlighederne kan let findes. Fordi vi trækker rækker og søjler tilfældigt er $q_{x,x'} = q_{x',x}$, så

$$\alpha_{x,x'} = \min\left(1, \frac{\pi_{x'}}{\pi_x}\right) = \min\left(1, \frac{1/\prod_{ij} x'_{ij}}{1/\prod_{ij} x_{ij}}\right) = \min\left(1, \frac{x_{r_1s_2}x_{r_2s_1}}{(1+x_{r_1s_1})(1+x_{r_2s_2})}\right)$$

når $x, x' \in S$.

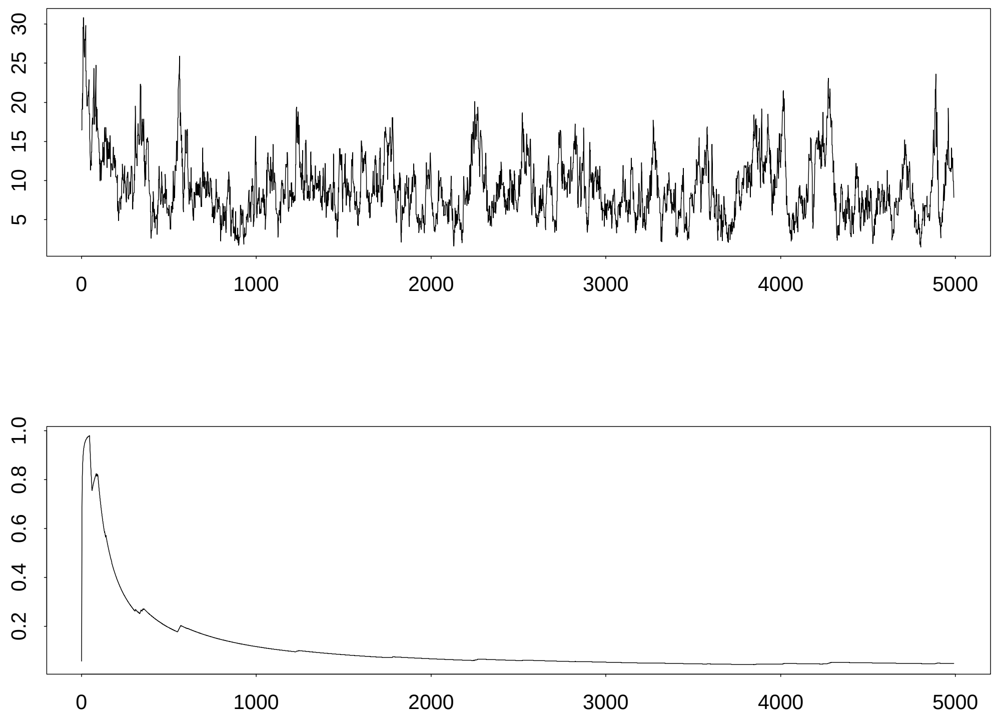
x'	$\dots$	s_2	$\dots$	s_1	$\dots$	$\sum_j$
$\vdots$		$\vdots$		$\vdots$		$\vdots$
r_1	$\dots$	$x_{r_1s_2}-1$	$\dots$	$x_{r_1s_1}+1$	$\dots$	$x_{r_1.}$
$\vdots$		$\vdots$		$\vdots$		$\vdots$
r_2	$\dots$	$x_{r_2s_2}+1$	$\dots$	$x_{r_2s_1}-1$	$\dots$	$x_{r_2.}$
$\vdots$		$\vdots$		$\vdots$		$\vdots$
$\sum_i$	$\dots$	$x_{.s_2}$	$\dots$	$x_{.s_1}$	$\dots$	$x_{..}$

Figur 1: x' genereres fra x.

Det er klart, at der findes en tabel x_0 således at $p_{x_0x_0} < 1$; vi kan jo bare vælge den tabel som har mindst sandsynlighed (dvs. vælge x_0 så $\pi_{x_0} = \min_{x\in S} \pi_x$); det er muligt fordi S er endelig. Det er meget sværere at vise irreducibiliteten, så det springer vi over.

Endelig skal vi –for at bruge sætning 3 direkte– sikre at fordelingen af X_0 netop er π. Men det er ganske let; under hypotesen er fordelingen af den observerede tabel jo netop π, så vi starter bare kæden med den observerede tabel som udgangspunkt.

Alt, der er tilbage nu, er at sætte computeren til at arbejde. Jeg har valgt at benytte 5000 iterationer. I figur 2 ser vi øverst, hvordan fordelingen af de simulerede teststørrelser ser stabil ud; der er måske et lille problem i begyndelsen, men ellers er det godt nok. Nederst i figur 2 ses vores testsandsynlighed beregnet ved Markov chain Monte Carlo integration som funktion af antal iterationer. Efter 5000 iterationer har den tydeligvis stabiliseret sig.



Figur 2: Simulerede teststørrelser (øverst) og testsandsynligheden beregnet ved MCMC (nederst); i begge tilfælde er førsteaksen antal iterationer.

Vi får dermed en testsandsynlighed på $0,047$. Der er altså noget, som tyder på, at hypotesen skal forkastes. Dette støttes også af at de simulerede teststørrelser ser lidt „uldne" ud i begyndelse, altså der hvor tabellerne ikke er særligt meget forskellige fra den observerede.

Her forlader vi så vores datamateriale. En videre analyse bør prøve at afklare hvorfor der ikke er uafhængighed mellem de to inddelingskriterier. Man kunne f.eks. se på om det er risikoen for ene-uheld, der er væsentlig forskellig fra risikoen for andre typer uheld, eller om risikoen for uheld, som involverer bløde trafikanter, adskiller sig væsentligt fra hinanden. Begge dele leder til nye Pearson teststørrelser, som vi kan finde testsandsynlighederne for ved MCMC.

Bemærkninger

Hvis man har fået lyst til at vide mere om Markovkæder, kan man følge ATSs kursus i Markovkæder i diskret tid til efteråret. Her gennemgås teorien for Markovkæder på tællelige tilstandsrum. Kurset er en del af Statistik 1B, hvor man også ser på Markovkæder i kontinuert tid (hvor man indicerer Xerne ved $t \in [0;\infty[$ fremfor $n \in \mathbf{N}$ (diskret tid)). Endelig gennemgås Markovkæder på generelle tilstandsrum på

Statistik 2B.

Hvis man vil læse mere om Markov chain Monte Carlo er Gilks, Richardson & Spiegelhalter *Markov chain Monte Carlo in practice* (Chapman & Hall, 1996) et godt udgangspunkt; den forudsætter dog mere end Statistik 0. For at få et kig på hvad der er „hot“ i MCMC lige nu, kan man besøge MCMC preprint serveren på `www.stats.bris.ac.uk/MCMC/`

Forskerdag 1999

Henrik Chr. Grove

Mandag den 26. april holdt matematisk afdeling sin årlige forskerdag. Normalt ser vi kun noget til de ansatte i forbindelse med undervisning, men det er kun knap halvdelen af de ansattes tid, der går med undervisning, resten[1] går med at forske. I år fik vi mulighed for at høre lidt om hvad Jesper Lützen, Henrik Schlichtkrull, Anders Thorup, Asmus Schmidt og Niels Grønbæk forsker i.

Jesper Lützen lagde ud med at beklage at han ikke havde noget igangværende matematisk forskning at berette om, men leverede derefter en glimrende foredrag, om hvad han selv kaldte „Lidt stillestående matematisk-fysisk-historisk forskning“, under overskriften „Mekanistiske billeder på geometrisk form“.

Henrik Schlichtkrull fortalte om sin forskning indenfor „Harmonisk analyse“.

I Anders Thorups foredrag om „Antalsgeometri“, fik vi at vide hvordan man regner ud at der i planen er 8 cirkler der tangerer 3 givne, og hvorfor der ikke er 7776 keglesnit der tangerer 5 givne (og hvorfor man kunne tro det), men kun ca. 3000.

Asmus Schmidt gav en grundig introduktion til algebraisk talteori, og de emner herunder der forskes i i øjeblikket.

Niels Grønbæk sluttede foredragsrækken af med at fortælle om „Banachalgebraer — strukturer i samspil“. Niels lagde ud med at binde sit eget emne sammen med Henrik Schlichtkrulls, ved at fortælle at Banachalgebraer var opfundet pga. de repræsentationer Henrik havde talt om i sit foredrag. Herefter gav han et godt indblik i emnet, og sluttede af med at præsentere en imponerende liste med (en del af) de forskellige grene af matematikken, han var kommet i berøring med ud fra sit arbejde med Banachalgebraerne.

Det var rart at se at selv førsteårsstuderende turde komme, jeg håber også det har givet dem endnu større lyst til at fortsætte med matematikken. Det er min egen erfaring at oplevelsen ændrer sig meget, efterhånden som man kommer længere hen i sit studieforløb. Fra den første forskerdag, der blev holdt da jeg var på andet år, hvor jeg absolut ikke forstod hvad Jesper Michael Møller talte om, men derimod fik

[1] Bortset fra lidt administration

en idé om, hvordan en forsker i matematik arbejder, til årets forskerdag, hvor jeg kunne forstå en god del af foredragene.

Citater fra forskerdagen

„Det var før Gödel det her.“ – *Jesper Lützen om en fysikers opfattelse af konsistens*

„Selvom vi lige har forkastet at det skulle være en gruppe, kommer gruppeteorien ind alligevel“ – *Henrik Schlichtkrull*

„Det kan være ligegyldigt hvad det er, det er noget man tror på!“ – *Asmus Schmidt om den genereliserede Riemann hypotese*

Choquets sætning – til hverdag og fest

Troels Roussau Johansen

Dette lille indlæg er delvist inspireret af et studenterkollokvium, der blev afholdt for noget tid siden, og for nogle vil det følgende derfor virke velkendt. Der vil dog forhåbentligt være noget nyt at hente for de fleste, og læseren opfordres hermed til trygt at læse videre. Jeg vil dog af hensyn til læserens almene velbefindende straks nævne, at jeg forudsætter et vist kendskab til funktionalanalyse (for eksempel svarende til Matematik 3AN og hvad dette implicit medfører af anden baggrundsviden).

Introduktion til konveksitet

Et lidt forsømt begreb på de to første studieår er begrebet *konveksitet*, men de fleste studerende, der har fulgt kurset Matematik 3AN i en tidligere udgave end den nuværende vil være stødt på det i forbindelse med studiet af de såkaldte *lokalkonvekse topologiske vektorrum*. Det er egentlig lidt trist, at man først på dette forholdsvis sene tidspunkt i sit studium lærer om konveksitet[1], for i sig selv er konveksitet ikke specielt kompliceret:

Definition 1. En delmængde K af et reelt vektorrum V siges at være **konveks**, hvis der for alle $x, y \in K$ og alle $t \in [0,1]$ gælder, at $tx + (1-t)y \in K$. Den mindste konvekse mængde indeholdende en given – ikke nødvendigvis konveks – mængde M, kaldes for **det konvekse hylster** af M og betegnes $\mathrm{co}(M)$.

Dette svarer åbenbart til at sige, at hvis to punkter x og y tilhører mængden K, da skal linjestykket, der forbinder de to punkter, forløbe *helt* i mængden. Et godt eksempel på en konveks delmængde af $\mathbb{R}^3$ kunne være den afsluttede enhedskugle

[1] Hvis man er matematik-økonomi studerende, sker det dog allerede på Matematik 2OK.

$S^2 = \{x \in \mathbb{R}^3 \mid \|x\| \leq 1\}$, men der er naturligvis mange andre eksempler at tage af. Man indser desuden let, at

$$\mathrm{co}(K) = \left\{ \sum_{i=1}^{n} t_i x_i \,\middle|\, x_1, \ldots, x_n \in K, t_i \geq 0, \sum_{i=1}^{n} t_i = 1, n \in \mathbb{N} \right\}.$$

Definition 2. Lad K være en konveks delmængde af et reelt vektorrum V, og lad $x \in K$ være skrevet på formen $x = (1-t)y + tz$ for et eller andet par af punkter $y, z \in K$, $y \neq z$ og et eller andet $t \in]0,1[$. Hvis man heraf kan konkludere, at $y = x = z$, da siges x at være et **ekstremt punkt**. Mængden af ekstreme punkter for en mængde K betegnes $\mathrm{ex}\, K$.

Sætning 1 (Krein-Milman)**.** *Lad C være en ikke-tom kompakt konveks delmængde af et lokalkonvekst topologisk vektorrum X. Da gælder der følgende to ting:*

1. *Der findes et ekstremt punkt i C, dvs.* $\mathrm{ex}\, C \neq \emptyset$*,*
2. *C er præcis afslutningen af det konvekse hylster af sine ekstreme punkter, dvs.* $C = \overline{\mathrm{co}}(\mathrm{ex}\, C)$.

Bevis. Jeg vil ikke give mig til at bevise denne klassiske sætning, for dels hører det naturligt til på MATEMATIK 3AN, og dels vil jeg efter at have bevist Choquets sætning vise, hvordan dette giver os Krein-Milmans sætning. Læsere, der gerne vil præsenteres for et direkte bevis, kan kigge i W.Rudin: "*Functional Analysis*", side 75-76. □

Krein-Milmans sætning er åbenbart et topologisk udsagn vedrørende de ekstreme punkter for en kompakt konveks mængde. Choquets sætning, der følger nedenfor, er et målteoretisk udsagn om de ekstreme punkter, og der viser sig at være en pæn sammenhæng mellem de to resultater.

Iøvrigt er det ikke kun mængder, der kan være konvekse; også reelle funktioner kan være det. Hvis K er en konveks mængde, da siges en funktion $f : K \to \mathbb{R}$ at være *konveks*, hvis $f((1-t)y + tz) \leq (1-t)f(y) + tf(z)$ for alle $t \in]0,1[$ og $y, z \in K$. Hvis der i denne ligning gælder '=' siges f at være *affin*.

Repræsentation af punkter vha. et integral

Man kan vise, at hvis K er en kompakt konveks delmængde af et endelig-dimensionalt vektorrum V, da kan et vilkårligt punkt $x \in K$ skrives som en endelig, konveks kombination af ekstreme punkter for K. Dette resultat kaldes Minkowskis sætning, og betydningen er, at der er "nok" ekstreme punkter til at give en fuldstændig beskrivelse af punkterne i K. Vi kan oversætte dette til et udsagn om integraler på følgende måde: Givet $y \in K$ lader vi ϵ_y betegne Borelmålet defineret ved $\epsilon_y(B) = 1$ hvis Borelmængden B indeholder punktet y, og $\epsilon_y(B) = 0$ ellers. Skriver vi ved hjælp af Minkowskis sætning y som $y = \sum_{i=1}^n t_i x_i$, opnås ved at lade $\mu := \sum_{i=1}^n t_i \epsilon_{x_i}$, at μ bliver et regulært, positivt Borelmål på K, så $\mu(K) = 1$. Desuden vil der for alle $f \in V'$ gælde, at $f(y) = \sum_{i=1}^n t_i f(x_i) = \int_K f\, d\mu$. Vi siger, at μ *repræsenterer* punktet y, samt $\mu(\mathrm{ex}\, K) = \sum_{i=1}^n t_i \epsilon_{x_i}(\mathrm{ex}\, K) = \sum_{i=1}^n t_i = 1$, og Minkowskis sætning kan således omformuleres til følgende udsagn: *Til hvert $y \in K$ svarer der et diskret mål μ koncentreret på $\operatorname{ex} K$, der repræsenterer y.*

Nu er de interessante vektorrum i analysen sjældent endelig-dimensionale, så Minkowskis sætning skal erstattes af noget mere slagkraftigt. Det "at repræsentere" et punkt kan vi dog umiddelbart give mening til med følgende

Definition 3. Lad K være en ikke-tom kompakt delmængde af et lokalkonvekst topologisk vektorrum E og lad μ være et sandsynlighedsmål på K. Så siges et punkt $y \in K$ at være **repræsenteret** af μ, hvis $f(y) = \int_K f\,d\mu$ for enhver kontinuert lineær funktional f på E.

Grunden til at forlange, at E skal være lokalkonvekst er den simple, at E^* da ifølge Hahn-Banachs sætning skiller punkter, så at der findes *højst ét* mål, der repræsenterer et givent punkt.

Hvis man er blot det mindste optimistisk anlagt, kunne man nu håbe på, at denne formodede generalisering af Minkowskis sætning altid ville gælde. Det viser sig at være korrekt, hvis man accepterer en antagelse om, at mængden K i betragtning er konveks, kompakt og metrisérbar, men inden jeg begynder at vise dette, vil jeg understrege, at denne måde at repræsentere punkter på allerede er kendt i form af Riesz' repræsentationssætning. Denne siger følgende: Lad Y være et kompakt Hausdorffrum og lad X være mængden af kontinuerte lineære funktionaler Λ på $(C(Y), \|\cdot\|_u)$ som opfylder, at $\Lambda(1) = 1 = \|\Lambda\|$. Så er X en svag*-kompakt konveks delmængde af det lokalkonvekse topologiske vektorrum $(C(Y)^*, w*)$ og Riesz' Repræsentationssætning udsiger, at der til hvert $L \in X$ svarer et entydigt bestemt sandsynlighedsmål μ på Y således, at $\Lambda f = \int_Y f\,d\mu$ for alle $f \in C(Y)$. Nu kan man imidlertid vise (se for eksempel Dunford & Scwartz: "*Linear operators*, part I, side 442), at Y er homeomorf med $\operatorname{ex} X$, så vi kan opfatte μ som et sandsynlighedsmål på Boreldelmængderne af X. For at indse, at dette netop giver et repræsentationsudsagn af den ønskede form behøver man nu blot at erindre, at rummet E^* af w^*-kontinuerte lineære funktionaler på $C(Y)^*$ præcis er funktionalerne af formen $\Lambda \mapsto \Lambda f$, hvor $f \in C(Y)$.

Choquets sætning

I dette afsnit vil jeg bevise den formodning, der allerede blev omtalt i det sidste afsnit. Af hensyn til læserens generelle ve og vel vil jeg dog holde mig til det metrisérbare tilfælde.

Sætning 2 (Choquets sætning)**.** *Lad X være et reelt, lokalkonvekst topologisk vektorrum, lad K være en metrisérbar, kompakt konveks delmængde af X og lad E betegne mængden af ekstreme punkter for K. Da gælder der følgende:*

1. *E er en Borel-delmængde af K,*
2. *For alle $x \in K$ findes der et Borel-sandsynlighedsmål μ_x på E så*

$$f(x) = \int_E f\,d\mu_x$$

for alle $f \in E'$, dvs. for alle kontinuerte lineære funktionaler f på E.

Bevis. Lad $A = K \times K \setminus \Delta := \{(x, y) \in K \times K \; x \neq y\}$. Så er A en åben delmængde af det kompakte metriske rum $K \times K$ og kan derfor skrives som foreningsmængden $A = \bigcup_{n=1}^{\infty} A_n$ af en stigende følge $(A_n)_{n \in \mathbb{N}}$ af kompakte delmængder af $K \times K$. Definér $I :]0,1[\times A \to K$ ved $I(t, y, z) = (1-t)y + tz$. Da er $\mathrm{Im}(I) = K \setminus E$, og da I er kontinuert, er $\mathrm{Im}(I)$ den tællelige forening af de kompakte mængder $I([\frac{1}{n}, 1 - \frac{1}{n}] \times A_n)$. Herved bliver $K \setminus E$ en F_σ-delmængde af K, og dermed er E en G_δ-mængde, specielt en Borelmængde. Dette viser den første påstand.

Lad nu Y betegne vektorrummet af alle kontinuerte, affine funktioner på K, dvs. alle de kontinuerte (reelle) funktioner g på K, der for alle $y, z \in K$ og alle $t \in [0,1]$ opfylder, at $g((1-t)t + tz) = (1-t)g(y) + tg(z)$. Bemærk i denne forbindelse, at restriktionen til K af en kontinuert lineær funktional på X er et element i Y. Da Y er et underrum af $C(K)$, som er et separabelt normeret vektorrum (i uniform norm), bliver Y selv et separabelt, normeret vektorrum. Lad således $\{g_i\}$ være en tællelig, tæt delmængde i enhedskuglen $B_1(0)$ af Y og definér g' ved $g' = \sum_{i=1}^{\infty} 2^{-i} g_i^2$. Så er g' kontinuert på K og opfylder desuden, at $g'((1-t)y + tz) < (1-t)g(y) + tg(z)$ for alle $y, z \in K$, $y \neq z$ og $0 < t < 1$. Dette indses på følgende måde: Ifølge Weierstrass' M-test vil rækken, der definerer g', være uniformt konvergent, og sumfunktionen g' er dermed kontinuert. Hvis nu $y, z \in K$, $y \neq z$, findes der en kontinuert lineær funktional ϕ på X som skiller y og z, dvs. $\phi(y) \neq \phi(z)$ [2]. Altså vil der for mindst ét $i \in \mathbb{N}$ gælde, at $g_i(y) \neq g_i(z)$, og for et sådant i vil der for alle $t \in]0,1[$ gælde, at $g_i^2((1-t)y + tz) < (1-t)g_i^2(y) + tg_i^2(z)$. (En lille øvelse ligger gemt her...) Hvis på den anden side i opfylder, at $g_i(y) = g_i(z)$, da er $g_i^2((1-t)y+tz) = (g_i((1-t)y+tz))^2 = g_i^2(y) = (1-t)g_i^2(y) + tg_i^2(z)$, hvorved

$$
\begin{aligned}
g'((1-t)y + tz) &= \sum_{i=1}^{\infty} 2^{-i} g_i^2((1-t)y + tz) < \sum_{i=1}^{\infty} 2^{-i}((1-t)g_i^2(y) + tg_i^2(z)) \\
&= (1-t)g'(y) + tg'(z).
\end{aligned}
$$

Lad dernæst $Y_1 := Y \oplus \mathbb{R}g'$, dvs. $Y_1 \ni h = g + rg'$ for ét $r \in \mathbb{R}$. Givet $x \in K$ defineres ρ_x på $C(K)$ ved $\rho_x(h) = \inf c(x)$, hvor der tages infimum over alle kontinuerte, konkave[3] funktioner c på K, der desuden opfylder, at $h(y) \leq c(y)$ for alle $y \in K$. Fordi summen af to konkave funktioner igen er konkav (se selv efter), og et positivt multiplum af en konkav funktion ligeledes er en konkav funktion, følger det direkte, at ρ_x er en subadditiv lineær funktional på $C(K)$. Bemærk også, at hvis c er en kontinuert konkav funktion på K, da er $\rho_x(c) = c(x)$. Definér nu en lineær funktional ψ_x på Y_1 ved $\psi_x(g + rg') = g(x) + r\rho_x(g')$. Da 1-funktionen I er affin, vil $I \in Y_1$, hvorved $\psi_x(I) = 1$. Det er desuden ikke så svært at se, at $\psi_x \leq \rho_x$ på Y_1, og ifølge Hahn-Banachs sætning findes der således en lineær funktional ϕ_x på $C(K)$, som dels udvider ψ_x og dels opfylder vurderingen $\phi_x \leq \rho_x$ på $C(K)$. Bemærk nu, at hvis $h \in C(K)$ og $h \leq 0$, da er $\rho_x(h) \leq 0$ (0-funktionen er konkav og $0 \geq h$), og dermed $\phi_x(h) \leq \rho_x(h) \leq 0$. Det følger, at ϕ_x er en positiv lineær funktional. Ifølge Riesz' Repræsentationssætning (i den udgave, der oftest kaldes for Riesz–Markoffs sætning) findes der således et ENTYDIGT bestemt endeligt Borelmål ν_x på K, for

[2]Denne funktional eksisterer fordi X var antaget at være et *lokalkonvekst* topologisk vektorrum.

[3]Dette betyder bare, at $c((1-t)y + tz) \geq (1-t)c(y) + tc(z)$ for alle $y, z \in K$ og $t \in [0,1]$.

hvilket der gælder, at

$$\phi_x(h) = \int h \, d\nu_x$$

for alle $h \in C(K)$. Lader vi atter I betegne den identiske funktion på K, får vi specielt, at $\nu_x(K) = \int I \, d\nu_x = \phi_x(I) = \psi_x(I) = 1$, så ν_x er faktisk et sandsynlighedsmål. Hvis nu f er en kontinuert lineær funktional på X, da er

$$\int f \, d\nu_x = \phi_x(f) = \psi_x(f) = fx,$$

idet restriktionen af f til K er en kontinuert affin funktion og derfor et element i Y_1.

Vi vil dernæst vise, at $\nu_x(\complement E) = 0$, dvs. at ν_x er båret af E. For at gøre dette lader vi $(c_n)_{n\in\mathbb{N}}$ være en følge af kontinuerte, konkave funktioner på K for hvilke $c_n \geq g'$ for alle $n \in \mathbb{N}$, og $\rho_x(g') = \lim_n c_n(x)$. Sæt $c = \liminf_n c_n$. Så er c en Borelmålelig funktion, dermed ν_x-målelig, og $c(y) \geq g'(y)$ for alle $y \in K$. Det betyder, at $\int (c - g') \, d\nu_x \geq 0$. Men

$$\begin{aligned}
\int (c - g') \, d\nu_x &= \int (\liminf_n c_n - g') \, d\nu_x \leq \liminf_n \left(\int (c_n - g') \, d\nu_x \right) \\
&= \liminf_n \phi_x(c_n - g') = \liminf_n \phi_x(c_n) - \phi_x(g') \\
&= \liminf_n \phi_x(c_n) - \rho_x(g') \leq \liminf_n \rho_x(c_n) - \rho_x(g') \\
&= \liminf_n c_n(x) - \rho_x(g') = \lim_n c_n(x) - \rho_x(g') \\
&= 0.
\end{aligned}$$

Dette viser, at ν_x er båret af mængden, hvor c og g' stemmer overens. Lad os vise, at $c(w) \neq g'(w)$, når $w \notin E$. Vi lader derfor $w = (1-t)y + yz$ for $y, z \in K$, $y \neq z$ og $t \in]0, 1[$. Så er

$$\begin{aligned}
c(w) &= \liminf_n c_n(w) = \liminf_n c_n((1-t)y + tz) \geq \liminf_n ((1-t)c_n(y) + tc_n(z)) \\
&\leq (1-t)g'(y) + tg'(z) > g'((1-t)y + tz) = g'(w),
\end{aligned}$$

Lad nu $\mu_x = \nu_x\big|_E$. Så er μ_x et Borelsandsynlighedsmål på E og

$$\int_E f \, d\mu_x = \int_K f \, d\nu_x = f(x)$$

for alle kontinuerte lineære funktionaler f på X. Dette viser punkt 2 i sætningen. $\square$

Vi kan som et umiddelbart korrolar hertil få Riesz' repræsentationssætning i et specialtilfælde: Lad X være et kompakt metrisk rum og sæt $K = \{\Lambda \in C(X)^* \mid \Lambda \geq 0, \Lambda(1) = 1\}$. Så er K svag*-kompakt, konveks og metrisérbar, så vi kan anvende Choquets sætning. Til hvert $\Lambda_0 \in K$ svarer derfor et sandsynlighedsmål μ på $\operatorname{ex} K$ således, at der for alle $f \in C(X)$ gælder, at

$$\Lambda_0(f) = \int_{\operatorname{ex} K} \Lambda(f) \, d\mu.$$

Afbildningen $\phi : X \to \operatorname{ex} K$ defineret ved $\phi(x) = \epsilon_x$ er en homeomorfi, thi da $f \in C(X)$ er kontinuert (!) er ϕ kontinuert fra X ind i $(K, w*)$. Da $C(X)$ skiller punkter i X er ϕ injektiv, og da X er kompakt bliver ϕ derfor en homeomorfi. Det betyder, at hvis vi definerer ψ ved $\psi(B) = \mu(\phi(B))$, hvor B er en Boreldelmængde af X, så kan ovenstående formel udtrykkes ved

$$\Lambda_0(f) = \int_X \phi(x)(f)\, d\mu(\phi(x)) = \int_X f(x)\, d\psi(x),$$

og det er jo netop Riesz' Repræsentationssætning.

Korollar 1 (Krein-Milman). *Lad K være en kompakt konveks delmængde af et lokalkonvekst topologisk vektorrum E og lad K_0 være afslutningen af mængden af alle endelige konvekse kombinationer af ekstreme punkter fra K. Da er $K_0 = K$.*

Her vil vi ved en *konveks kombination* af punkterne $x_1, \ldots, x_n$ forstå et udtryk af formen $\sum_{i=1}^n t_i x_i$, hvor $t_i \geq 0$ og $\sum_{i=1}^n t_i$.

Overvej, hvorfor dette præcist er det resultat, der tidligere blev kaldt Krein-Milmans sætning. Det vil her være en hjælp at prøve at indse, at hvis $A \subseteq K$, da er $\overline{\text{co}}(A) = K$.

Bevis. Det er let at se, at mængden af alle konvekse kombinationeraf punkter i $\operatorname{ex} K$ er konveks, og at denne mængdes afslutning K_0 ligeledes er konveks. Siden K_0 er en afsluttet delmængde af K, er K_0 kompakt. Antag nu, at K_0 er en **ægte** delmængde af K. Vi kan da uden indskrænkning antage (hvorfor det?), at $0 \in K \setminus K_0$. Siden K_0 er kompakt, findes der ifølge Separationssætningen[4] en kontinuert affin funktion h på K med egenskaben, at $\sup\{h(x) \mid x \in \operatorname{ex} K\} < h(0)$. Ifølge Choquets sætning findes der et sandsynlighedsmål μ på $\operatorname{ex} K$ som repræsenterer 0 og således, at $\int_K f(x)\, d\mu = f(0)$ for enhver kontinuert affin funktion f på K. Men siden $\mu(\operatorname{ex} K) = 1$, betyder dette, at

$$h(0) = \int_K h(x)\, d\mu(x) \leq \mu(\operatorname{ex} K) \sup\{h(x) \mid x \in \operatorname{ex} K\} < h(0),$$

hvilket er en modstrid. Altså må $K = K_0$. □

Bernsteins sætning om fuldstændigt monotone funktioner

Kan alt denne fine, abstrakte så "bruges" til noget, i betydningen, at det giver os værdifuld information i konkrete situationer? Det kan man heldigvis svare 'ja' til, og i dette afsnit skal vi se, hvad Choquets sætning siger i forbindelse med såkaldte fuldstændigt monotone funktioner. Det, man i det følgende bevis skal bemærke, er, at Choquets sætning IKKE benyttes og at det er overraskende kompliceret at bringe os i stilling til at benytte terminologien allerede udviklet.

[4]. . . der siger, at hvis C er en afsluttet konveks delmængde i et lokalkonvekst topologisk vektorrum X, og hvis $x \notin C$, da findes der en kontinuert lineær funktional Λ på X, så $\sup\{\Lambda y \mid y \in C\} < \Lambda x$. Hvorfor bliver det aktuelle h affin?

Definition 4. En funktion $f :]0,\infty[\to \mathbb{R}$ kaldes **fuldstændigt monoton**, hvis f er uendeligt ofte differentiabel og hvis $(-1)^n f^{(n)} \geq 0$ for alle $n \in N_0$.

Et oplagt eksempel på sådan en funktion fås ved at lade $f(x) = x^{-\alpha}$ for $\alpha \geq 0$ og $x > 0$. At f faktisk ER fuldstændigt monoton overlades trygt som en øvelse til læseren at indse.

Vi stiler efter at vise følgende klassiske sætning af Bernstein (der viste den uden brug af Choquet-teori, men det er en anden historie...)

Sætning 3 (Bernstein)**.** *Hvis f er fuldstændigt monoton og begrænset på $]0,\infty[$, da findes der et entydigt bestemt Borelmål μ på $[0,\infty]$, så der for alle $x > 0$ gælder, at*

$$f(x) = \int_0^\infty e^{-\alpha x}\, d\mu(\alpha).$$

Det er faktisk ikke nogen væsentlig indskrænkning kun at se på begrænsede funktioner, og Bernsteins sætning gælder også i det mere generelle tilfælde, hvor f blot antages at være fuldstændigt monoton.

Bevis. Lad CM betegne den konvekse kegle af alle fuldstændigt monotone funktioner f så $f(0+) := \lim_{x\to 0^+} f(x) < \infty$ (da f er ikke-voksende, vil denne grænseværdi altid eksistere, selvom den kunne blive ∞). At CM er en kegle betyder blot, at $\forall f, g \in CM\, \forall a, b \geq 0 : af + bg \in CM$, og at dette er tilfældet er oplagt, thi med $h(x) = af(x) + bg(x)$ er $(-1)^n h^{(n)}(x) = a(-1)^n f^{(n)}(x) + b(-1)^n g^{(n)}(x)$. Det ses på samme måde, at CM er konveks. Lad dernæst K betegne mængden af de funktioner f fra CM med $f(0^+) \leq 1$. Denne mængde er ligeledes konveks, for med $f, g \in K$ og $t \in [0,1]$ er $\lim_{x\to 0^+}(tf(x) + (1-t)g(x)) = t\lim_{x\to 0^+} f(x) + (1-t)\lim_{x\to 0^+} g(x) \leq t + (1-t) = 1$. Endvidere bemærkes det, at hvis $f \in CM$ og $f \neq 0$, da vil $\frac{1}{f(0^+)} f \in K$, så vi kan nøjes med at vise Bernsteins sætning for funktioner i K – og det vil vi derfor gøre. (Hvorfor er funktionerne i K begrænsede?)

Siden funktionerne i K alle er uendeligt ofte differentiable, kan K naturligt opfattes som en delmængde af $\mathcal{E} = C^\infty(]0,\infty[)$, hvor dette rum er udstyret med initialtopologien mht. den tællelige familie $(p_{m,n})_{m,n}$ af seminormer

$$p_{m,n}(f) = \sup\left\{ |f^{(k)}(x)| \,\middle|\, \tfrac{1}{m} \leq x \leq m; 0 \leq k \leq n \right\}.$$

Det er da velkendt fra Matematik 3AN, at $\mathcal{E}$ herved får Heine-Borel egenskaben (dvs. at afsluttede, begrænsede delmængder er kompakte). Lad os derfor forsøge at vise, at K er begrænset – idet jeg overlader det til læseren at udføre den øvelse i leg med seminormer, det er at vise, at K er afsluttet.

At K er begrænset kommer åbenbart ud på at vise, at

$$\forall m, n \in \mathbb{N} : \sup\{|f^{(n)}(x)| \mid f \in K\} < \infty,$$

hvortil det er nok at vise, at $\sup\{|f^{(n)}(x)| \mid \frac{1}{m} \leq x \leq m, f \in K\} < \infty$ for alle $n \geq 0$ og $m \geq 1$. Lad til dette formål $K_n := \{(-1)^n f^{(n)} \mid f \in K\}$, hvor $n \in N_0$. Så kan vi klare opgaven ved at vise, at der for alle $a > 0$ og alle $n \in \mathbb{N}_0$ gælder, at funktionerne i K_n er opad begrænset på $[a,\infty[$ ved konstanten $a^{-n} 2^{\frac{n(n+1)}{2}}$. Dette klares ved induktion efter n og for $n = 0$ er det klart pr. konstruktion af K, idet jo $K_0 = K$. Antag derfor påstanden er korrekt for K_n og betragt mængden K_{n+1}. Siden funktionerne i K_{n+1} er ikke-voksende, er det nok at vise den ønskede vurdering i punktet $x = a$. Benytter vi Middelværdisætningen[5] på $f^{(n)}$, findes der et $c \in]\frac{a}{2}, a[$ så $\frac{a}{2} f^{(n+1)}(c) = f^{(n)}(a) - f^{(n)}(\frac{a}{2})$. Sammenholdes dette med induktionsantagelsen får vi, at $\left(\frac{a}{2}\right)^{-n} 2^{\frac{n(n+1)}{2}} \geq (-1)^n f^{(n)}(\frac{a}{2}) \geq (-1)^{n+1} \left(\frac{a}{2}\right) f^{(n+1)}(c) \geq \left(\frac{a}{2}\right) (-1)^{n+1} f^{(n+1)}(a)$, og det viser netop den ønskede vurdering.

Næste skridt er at identificere de ekstreme punkter for K, og det viser sig, at disse netop er funktionerne $x \mapsto e^{-\alpha x}$, hvor $0 \leq \alpha \leq \infty$. For at indse dette lader vi først $f \in \operatorname{ex} K$. For et fast $x_0 > 0$ lader vi $u(x) = f(x + x_0) - f(x)f(x_0)$. Så vil $f \pm u \in K$, og lad os straks indse hvorfor: med $b := f(x_0) \in [0,1]$ vil $(f+u)(0^+) = (1-b)f(0^+) + b \leq 1$ og $(f-u)(0^+) = f(0^+) - b(1 - f(0^+)) \leq f(0^+) \leq 1$, samt $(-1)^n (f+u)^{(n)}(x) = (1-b)(-1)^n f^{(n)}(x) + (-1)^n f^{(n)}(x + x_0) \geq 0$ såvel som $(-1)^n (f-u)^{(n)}(x) = ((-1)^n f^{(n)}(x) - (-1)^n f^{(n)}(x + x_0)) + b(-1)^n f^{(n)}(x) \geq 0$, da $(-1)^n f^{(n)}$ er ikke-voksende. Men da $f \in \operatorname{ex} K$, må $u = 0$, og det betyder, at $f(x + x_0) = f(x)f(x_0)$ når $x, x_0 > 0$. Siden f er kontinuert på $]0, \infty[$, betyder dette enten, at $f \equiv 0$, eller at $f(x) = e^{-\alpha x}$ for et eller andet α. Siden $0 \leq -f'(x) = \alpha e^{-\alpha x}$, må $\alpha \geq 0$, og vi har dermed vist, at alle ekstreme punkter i K har formen $x \mapsto e^{-\alpha x}$ for et $\alpha \geq 0$. Faktisk er alle funktionerne $x \mapsto e^{-\alpha x}; \alpha \geq 0$ ekstreme punkter, og for at vise dette lader vi for $r > 0$ $T_r : K \to K$ være defineret ved $(T_r f)(x) = f(rx)$. Siden T_r oplagt er lineær, bijektiv og bevarer konvekse kombinationer, er $T_r(\operatorname{ex} K) = \operatorname{ex} K$. Da K er kompakt, gælder der ifølge Krein-Milmans sætning, at $K = \overline{\operatorname{co}}(\operatorname{ex} K)$, og derfor er der mindst ét ikke-konstant ekstremt punkt (dvs. en funktion). Ifølge det lige viste er dette ekstreme punkt af formen $x \mapsto e^{-\alpha x}$ for et eller andet $\alpha \geq 0$, og derfor er $x \mapsto e^{-r\alpha x}$ også ekstrem (hvorfor det?). Da dette gælder for alle $r > 0$, er *alle* eksponentialfunktionerne ekstreme, ganske som de konstante funktioner 0 og 1 er det.

Vi kan nu færdiggøre beviset for Bernsteins sætning: Idet det er let at vise, at afbildningen $T : \alpha \to e^{-\alpha \cdot}$ (dvs. $T(\alpha)(x) = e^{-\alpha x}$) fra $[0, \infty]$ ind i K er kontinuert, og da $[0,1]$ er kompakt, vil afbildningens billede, $\operatorname{ex} K$, ligeledes være kompakt. Ifølge Krein-Milmans sætning findes der til hvert $f \in K$ et regulært Borelsandsynligheds-mål m på $\operatorname{ex} K$ som repræsenterer f i betydningen, at der for enhver kontinuert lineær funktional Λ på E gælder, at

$$\Lambda f = \int_{\operatorname{ex} K} \Lambda \, dm.$$

Hvis $x > 0$, er evalueringsfunktionalen $\Lambda_x : f \mapsto f(x)$ kontinuert på E, så

$$f(x) = \int_{\operatorname{ex} K} \Lambda_x \, dm$$

for alle $x > 0$. Definér nu μ op en vilkårlig Boreldelmængde af $[0, \infty]$ ved $\mu(B) = m(T^{-1})$. Siden $\Lambda_x(T\alpha) = e^{-\alpha x}$ har vi, at

$$f(x) = \int_{\operatorname{ex} K} \Lambda_x \, dm = \int_{T^{-1}(\operatorname{ex} K)} \Lambda_x \circ T \, d(m \circ T^{-1}) = \int_0^\infty e^{-\alpha x} \, d\mu(\alpha)$$

[5] Hvem havde troet, den skulle dukke op her?

for alle $x > 0$.

Vi mangler nu blot at vise, at dette mål μ er entydigt bestemt. Antag derfor, der er et andet mål ν på $[0, \infty]$ så $f(x) = \int_0^\infty e^{-\alpha x}\, d\nu(\alpha)$ for $x > 0$. For hvert $x \geq 0$ er $\alpha \mapsto e^{-\alpha x}$ kontinuert på $[0, \infty]$. Lad A betegne delalgebraen af $C([0, \infty])$ frembragt af disse funktioner. Så består A af alle endelige linearkombinationer af disse og som lineære funktionaler på $C([0, \infty])$ vil μ og ν stemme overens på de ikke-konstante funktioner i A. Men benyttes Lebesgues Majorantsætning på funktionerne $\alpha \mapsto e^{-\frac{\alpha}{n}}$ indses, at μ og ν også bliver sammenfaldende på de konstante funktioner i A. Da nu A skiller punkter i $[0, \infty]$ følger det af Stone-Weierstrass' sætning, at A er tæt i $C([0, \infty])$, og dermed er $\mu = \nu$ på hele $C([0, \infty])$. Dette viser den ønskede entydighed. $\square$

Opgave: Kan du reducere ovenstående bevis i længde, hvis du må benytte Choquets sætning??!

Hvad med alt det andet?

Et spørgsmål, som Choquets sætning *ikke* besvarer, er spørgsmålet om entydighed af det repræsenterende mål μ_x. Det viser sig, at man ved at kræve en geometrisk betingelse opfyldt for K KAN opnå dette (kig efter i kapitel 9 i Phelps bog). Et andet problem er kravet om, at K skal være metrisérbart – det er jo ikke sikkert, at K er det i lige netop den situation, man har valgt at undersøge. Det viser sig, at $\operatorname{ex} K$ da ikke længere behøver være en Borelmængde, og så falder vores hidtidige definition vedrørende integralrepræsentation jo lidt til jorden. Det viser sig, at man kan generalisere Choquets sætning, hvis man fremfor at betragte Borelmængder istedet ser på såkaldte *Bairemængder* – det vil dog føre for vidt at komme på dette, og jeg vil i stedet atter henvise den interesserede læser til Phelps bog (dvs. *Lectures on Choquet's Theorem*). Hvis man finder denne emnekreds interessant, ville det iøvrigt være en idé at kigge nærmere på Choquets egne bøger…

Litteraturliste

[1] Baggett, Lawrence W.: "*Functional Analysis — A Primer*", Marcel Dekker, Inc., New York, 1992

[2] Choquet, Gustave: "*Lectures on Analysis*", bind I–III, W.A.Benjamin, Inc., New York 1969

[3] Phelps, Robert R.: "*Lectures on Choquet's Theorem*", D.van Nostrand Company, Inc., Princeton, New Jersey 1965

[4] Phelps, Robert R.: "*Integral Representations for Elements of Convex Sets*", bragt i R.G.Bartle (Ed.): "*Studies in Functional Analysis*", The Mathematical Association of America, 1980

[5] Rudin, Walter: "*Functional Analysis*", 2$^{\text{nd}}$ Edition, McGraw-Hill, Inc., 1991

Opgaveløsninger

Rasmus Borup Hansen

Opgave 1

Lad os forestille os, at hver af de 27 kuber i osten er farvet enten sort eller hvid, således at to kuber, der støder op til hinanden har forskellige farver (ligesom et skakbræt i tre dimensioner). Vi ser nu, at en kube i et af hjørnerne ikke kan have den samme farve som centerkuben, og da musen spiser skiftevis sorte eller hvide kuber, og da der er et ulige antal kuber i alt, kan det ikke lade sig gøre, at musen slutter med at spise den midterste kube.

Opgave 2

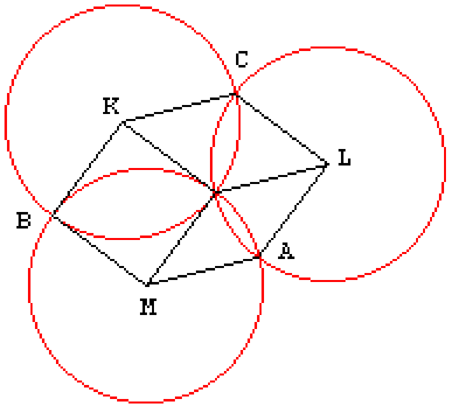
De tre letteste udgør $\frac{5}{13}$, dvs. 65% af hele fangsten. De mellemste fisk udgør dermed 40%. I mellemgruppen må der følgelig være mindst 4 fisk (da de tre tungeste kun vejer 35% tilsammen). På den anden side kan der heller ikke være mere end 4 fisk i mellemgruppen: Hver mellemfisk vejer jo mindst så meget som gennemsnittet af de tre letteste fisk, altså mindst $\frac{1}{3} \cdot 25\% = 8\frac{1}{3}\%$, og 5 mellemfisk ville dermed tilsammen veje over 40%. Altså er der præcis 4 fisk i mellemgruppen, og dermed i alt $3 + 4 + 3 = 10$ fisk.

Opgave 3

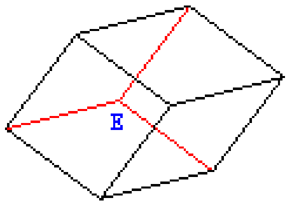
Betragtes mindst 2000 forskellige tal bestående af lutter 1-taller, vil to af disse give samme rest ved division med 1999 (da der jo højst kan forekomme 1999 forskellige rester ved division med 1999), dvs., deres differens er delelig med 1999. Denne differens har formen $11 \cdots 1 \cdot 10^n$, og da 1999 ikke har fælles faktorer med 10^n, går 1999 op i det foranstillede tal. Dette viser, at der findes et tal med de ønskede egenskaber.

Opgave 4

Lad os betegne centrene i de tre cirkler k, l og m med henholdsvis K, L og M. Da ved vi, at alle de linjestykker, der er indtegner på figuren nedenfor, er lige lange:



Tilføjes tre linjestykker af samme længde som de ni linjestykker, så vi får figuren



ser vi, at punktet E må ligge lige langt fra A, B og C og derfor må være centrum i cirklen, der går gennem disse punkter. Men alle linjestykkerne på figuren ovenfor er lige lange, så denne cirkel må have samme radius som de tre andre cirkler.

Opgave 5

Lad r_1 og r_2 betegne de to radii i de to små cirkler. Da er $r = r_1 + r_2$. Da (for eksempel) den øverste halvdel af korden er mellemproportional med diametrene i de to små cirkler, er $2r_1 \cdot 2r_2 = \left(\frac{t}{2}\right)^2$. Alt i alt er det søgte areal

$$\pi r^2 - \pi r_1^2 - \pi r_2^2 = \pi((r_1 + r_2)^2 - r_1^2 - r_2^2) = \pi(2r_1 r_2) = \pi \frac{t^2}{8}.$$

Nyt fra naturfagsbogladen

J.Nyhus: Rapportskrivning med WORD 97	kr. 69,-
P.Forchhammer: Bestå statistikken	kr. 155,-
P.S.Mortensen: Stikprøver- udvælgelse og estimation	kr. 175,-
J.Mathews: Numerical Methods using MATLAB, 3.edt.	kr. 435,-
M.Gardner: The last Recreations	kr. 230,-
R.Banks: Towing Icebergs, Falling Dominoes, and other Adventures	kr. 375,-
D.Benson: The Moments of proof, Mathematical Epiphanies	kr. 395,-
B.S.Everitt: The Cambridge Dictionary of Statistics	kr. 335,-
H.Kragh: Videnskabens væsen	kr. 198,-
P.Naur: Antifilosofisk leksikon	kr. 140,-
Hvad er metafysik - idag? 13 bud på et svar	kr. 225,-
R.Penrose: The large, the small and the hum an Mind	kr. 175,-
Hurtigt i gang med RedHat LINUX 5.2, m/2 CD-Rom	kr. 69,-
J.Buyens: Avancerede WEB Tricks	kr. 299,-
An electronic Companion to STATISTICS, bog m/ CD-Rom	kr. 335,-