# Undersøgelse af indlæring via forelæsninger og opgaveregning

Thomas Mejer Hansen, Peter Harremoës,
Bo Markussen, Jesper Lund Pedersen

4. November 2004

— Forprojekt på adjunktpædagogikum —

# Indhold


**1 Formål med forsøget** **4**

**2 Statistik 1B** **4**

2.1 Valg af emnet for undervisningsforsøget . . . . . . . . . . . . . 5

**3 Materiale til forsøget** **5**

3.1 Undervisningsmateriale . . . . . . . . . . . . . . . . . . . . . . . . . 5

3.1.1 Overvejelser angående forbedring af materialet . . . . . 6

3.2 Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.3 Spørgeskemaer . . . . . . . . . . . . . . . . . . . . . . . . . . 7

**4 Forsøget** **7**

4.1 Forsøgsprotokol . . . . . . . . . . . . . . . . . . . . . . . . . 7

4.2 Forelæsningerne . . . . . . . . . . . . . . . . . . . . . . . . . 8

4.3 Øvelserne . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

**5 Resultater og behandling af data** **9**

5.1 Behandling af testdata . . . . . . . . . . . . . . . . . . . . . 9

5.1.1 Beregning af indlæringsparametre og model kontrol . . 9

5.1.2 Signifikans test . . . . . . . . . . . . . . . . . . . . . . 11

5.2 De studerendes forberedelse og opfattelse af undervisningsform før forsøget. . . . . . . . . . . . . . . . . . . . . . . . . . 13

5.2.1 Forberedelsestid vs. testscore, grupperet i undervisningshold . . . . . . . . . . . . . . . . . . . . . . . . . . 16

5.2.2 Forberedelsestid vs. testscore, grupperet efter om de studerende regner uopfordret eller ej . . . . . . . . . . 17

**6 Konklusion** **18**

**Appendix A** **20**

**Spørgeskemaer** **20**

Spørgeskema 1 : Før forsøget . . . . . . . . . . . . . . . . . . . 20

Spørgeskema 2 : Efter forsøget . . . . . . . . . . . . . . . . . . 21

**Appendix B** **23**

**Forelæsning** **23**

**Opgaver om Markovkæde Monte Carlo** **27**

**Evaluering** **31**

**Uafhængighedstest i tovejstabel** **32**

Statistisk Model . . . . . . . . . . . . . . . . . . . . . . . . . . 32
MCMC metoden . . . . . . . . . . . . . . . . . . . . . . . . . . 33
Uafhængighedstest . . . . . . . . . . . . . . . . . . . . . . . . . 34
Implementation i MatLab . . . . . . . . . . . . . . . . . . . . . 36
Plots vedrørende uafhængigstestet . . . . . . . . . . . . . . . . . 38

# 1 Formål med forsøget

Forelæsning og assisteret opgaveregning er to vidt forskellige måder at videregive viden til studerende. Formålet med denne undersøgelse er at undersøge hvilke af de to formidlingsformer, der giver den bedste indlæring hos de studerende, når der er begrænset til til rådighed. I forsøget er det 60 minutter. Undersøgelsen er foretaget ved at inddele studerende på et statistikkursus i to hold: Første hold modtager deduktiv undervisning via forelæsning. Andet hold modtager induktiv undervisning gennem opgaveregning.

Emnet for undervisningen er valgt således at begge former for undervisning skulle kunne gennemføres inden for 60 minutter.

Et andet formål med undersøgelsen er at undersøge de studerendes mening om og opfattelse af kurset. Resultaterne af undersøgelse er mest til gavn for forelæseren, men flere af de studerendes kommentarer er relevante og bruges i rapporten.

# 2 Statistik 1B

*Statistik 1B* er andetårs-kurset i sandsynlighedsregning på statistikuddannelsen. Bo er forlæser på kurset og øvelseslærer er Ander Tolver Jensen. Kurset løber over to semestre med 2 timers forelæsninger og 1 times regneøvelser i første semester og 2 timers forelæsninger og $1 + \frac{1}{2}$ timers regneøvelser i andet semester. Den samlede belastning er 10 ECTS punkter, hvilket svarer til 1/6 belastning for en fuldtidsstuderende eller $7\frac{1}{2}$ time per uge. Dermed skal de studerende ifølge normeringen bruge omkring $4\frac{1}{2}$ time på at forberede og efterlæse hjemme i første semester. Kurset afsluttes med en 4 timers skriftlig eksamen med hjælpemidler. Lærebogen på kurset er *J.R. Norris: Markov Chains, Cambridge 1997.* På kursets hjemmeside

`http://www.stat.ku.dk/~markusb/Stat1B.htm`

findes følgende indholdsbeskrivelse:

**Indhold:** Markovkæder i diskret og kontinuert tid, stokastiske processer. De store tals lov. Svag konvergens, karakteristiske funktioner, den centrale grænseværdisætning.

**Kompetencer:** Kurset skal give de studerende fortrolighed med stokastiske processer og følger af sandsynlighedsfordelinger.

Så vidt vides er kurset kun obligatorisk for statistik studerende. Af ikke-statistik studerende på kurset er de to største grupper aktuar og mat-øk studerende. Disse studerende kan være på et senere tidspunkt i deres studium.

## 2.1 Valg af emnet for undervisningsforsøget

Som emne for undervisningsforsøget blev *Markov Chain Monte Carlo* (MCMC) metoden valgt af følgende grunde:

1. Den grundlæggende idé i og teori for MCMC udgør et afrundet hele, som kan læres på 60 minutter givet forudkendskab til teorien for Markovkæder.

2. MCMC udgør en vigtig praktisk anvendelse af Markovkæder, og er dermed et relevant emne for *Statistik 1B*. Endvidere er MCMC hverken pensum eller kernestof for kurset, hvorfor det overfor de studerende er forsvarligt at lave et randomiseret forsøg i dette emne.

# 3 Materiale til forsøget

For at kunne gennemføre undervisningsforsøget udarbejdede vi følgende materiale:

- Undervisningsmateriale for forelæsninger og opgaveregning (afsnit 1 og 2, notat 4 i appendix B)
- Test (afsnit 3, notat 4 i appendix B, her kaldet 'Evalueringsskema')
- Spørgeskemaer (appendix A)
- Opfølgende undervisningsmateriale (afsnit 4, notat 4 i appendix B)

Her er en beskrivelse af de 4 punkter.

## 3.1 Undervisningsmateriale

Til undervisningen udarbejdede vi to notater om Markovkæde Monte Carlo og Metropolis-Hastings algoritmen. Det ene notat, forelæsningnoterne, blev brugt på forelæsningsholdet. Det andet notat, opgavesættet, blev brugt på øvelsesholdet:

**Forelæsningsnoterne** er struktureret på klassisk vis indenfor matematik. Først udvikles teorien med definitioner og resultater fra Markovkædeteorien, der skal bruges til konstruktionen af Metropolis-Hastings algoritmen. Derefter præsenteres et simpelt eksempel.

Strukturen i **opgavesættet** er baseret på 10 opgaver. Udgangspunktet for opgavesættet er det simple eksempel som også er brugt i forelæsningsnoterne. Definitioner og resultater fra Markovkædeteorien og algoritmen bliver løbende introduceret i opgaven efterhånden som der er brug for dem i eksemplet.

For at sikre at forelæsningsholdet og øvelsesholdet havde mulighed for at lære det samme, er forelæsningsnoterne og opgavesættet gjort så identiske som muligt i følgende forstand: Der er de samme matematiske definitioner, resultater og fraser i begge sæt, og de har samme ordlyd. Yderligere er eksemplet det samme. Netop med eksemplet blev vi nødt til at gå på kompromis. Eksemplet er søgt og ikke motiverende for at bruge MCMC metoden, men er tilpas nemt til at øvelsesholdet kan regne opgaver med det.

Et egentligt motiverende eksempel er et statistisk problem for tovejstabler. I en normal forelæsning vil dette emne være en del af undervisningen, som så vil være presset i tid. Men dette eksempel er for svært for et øvelseshold at regne på. Ved induktiv undervisning egner dette eksempel egner sig bedre som projektopgave. Derfor valgte vi at bruge det simple eksempel i undervisningen i forsøget. Vi forventede at forelæsningerne ville være mindre pressede end normalt, og til øvelserne ville de studerende alligevel være under et tidspres for at nå gennem opgavesættet. Eksemplet fra tovejstabler er kommet i et selvstændig notat. Dette notat blev brugt til forelæsningerne ugen efter som opfølgning på undervisningen fra forsøget.

### 3.1.1 Overvejelser angående forbedring af materialet

Vi vil kort skitsere hvordan materialet eventuelt kunne forbedres således at konklusionerne ville blive mindre diskutable.

- Forsøget kunne konstrueres som en konkurrence mellem en svoren tilhænger af henholdsvis induktiv og deduktiv undervisning. De skulle blive enige om et emne og i fællesskab lave en database af spørgsmål inden for emnet. Databasen skulle være tilpas stor således at man ikke ville kunne give en "facitliste" til alle spørgsmålene. Idet både tilhængeren af induktiv og deduktiv undervisning begge kendte hele databasen ville de hver for sig kunne lave hvad de selv anså for den optimale undervisning.

- Forsøget skulle gentages med forskellige typer studerende og forskellige typer emner. Idet tidsfaktoren er helt central i vores konklusion, skulle forsøget også gentages med forskel i afsat tid.

- Forsøget kunne med fordel udvides til at omfatte andre former for undervisning såsom selvstudie, men det ville naturligvis kræve at man havde flere studerende til rådighed.

## 3.2 Test

Hvad de studerende havde lært ved de 60 minutters undervisning, blev undersøgt ved en efterfølgende test. Testen bestod af 8 spørgsmål, der skulle besvares uden hjælpemidler. Der var både åbne og lukkede spørgsmål om begreber og matematisk forståelse af Markovkæder og MCMC metoden. Det var tilstræbt, at hverken forelæsnings- eller øvelseshold blev favoriseret. Der var afsat så meget tid til testen, at tidspres ikke burde være nogen faktor.

## 3.3 Spørgeskemaer

I undersøgelsen uddelte vi to spørgeskemaer. Det første blev brugt til at starte forsøget med, og det andet til at afrunde forsøget.

Spørgeskema 1 indeholder spørgsmål om de studerendes tidsforbrug på henholdvis forelæsninger og øvelser, samt deres vurdering af om de lærer mest ved forelæsninger eller øvelser. Spørgeskema 2 indeholder spørgsmål om de studerendes mening om selve undervisningsforsøget og *Statistik 1B*.

# 4 Forsøget

Forsøget foregik mandag den 18. oktober fra klokken 9 til 12, hvor der normalt er to timers forelæsninger og en times øvelser. Der deltog ialt 18 studerende i forsøget, hvor 4 af dem først kom med fra klokken 10. De 4 nye blev af praktiske grunde placeret på forelæsningsholdet. Der var dermed 11 på forelæsningsholdet og 7 på øvelsesholdet.

## 4.1 Forsøgsprotokol

De studerende var ikke på forhånd informeret om, at de skulle deltage i et undervisningsforsøg. Dermed var ingen af de studerende forberedt til undervisningen om MCMC. For at de studerende kunne være anonyme i forsøget og vi samtidig kunne holde rede på hver enkelt studerendes besvarelser, modtog hver en kuvert, som besvarelserne skulle lægges i. På kuverterne var der skrevet enten et “F” for forelæsning eller et “Ø” for øvelser. Forsøget løb af stablen efter planen:

- Klokken 9:15, vi uddelte spørgeskema 1 som de havde omkring 10 minutter til at svare på.
- Klokken 9:30, kurverter med lige mange “F”’er og “Ø”’er blev uddelt tilfældigt. Vi opdelte de studerende i to hold, et forelæsningshold og

et øvelseshold. De studerende, der havde en kuvert med et "F", var på forelæsningsholdet, og skulle blive og have forelæsninger med Bo. De studerende, der havde en kuvert med et "Ø", var på øvelsesholdet, og skulle have øvelser med Peter.

- Klokken 9:45-11:00, der var undervisning i 60 minutter med forelæsninger ved Bo og øvelser ved Peter.
- Klokken 11:15, de studerende fik en test uden hjælpemidler, som de havde 25 minutter til at svare på.
- Klokken 11:45, spørgeskema 2 blev udleveret som afrunding på forsøget.

## 4.2 Forelæsningerne

Forelæsningerne blev afholdt på to gange 30 minutter med 15 minutters pause. De studerende fik udleveret forelæsningsnoterne. Forelæsningerne forsøgtes gennemgået ordret efter disse noter, på nær at beviset for vurderingen givet i ligning (3) ikke blev gennemgået. Dette bevis er med for fuldstændighedens skyld i noterne, som udgør en del af *Statistik 1B*, men er temmelig omfattende og teknisk. Idet beviset var umuligt at indpasse i øvelserne, blev det heller ikke gennemgået til forelæsningerne.

Efter pausen dukkede der yderligere 4 studerende op. To af disse mangler altid i første forelæsningstime, hvor de er instruktorer for regneøvelser. De må derfor antages at være dygtige studerende, hvorfor der er en lille selektion bias. De 4 ekstra studerende blev naturligt indlemmet på forelæsningsholdet. Idet det var vigtigere for forelæseren at levere god undervisning end at følge forsøgsprotokollen besluttedes at give en hurtig repetition af stoffet. De studerende, der har været til begge forelæsningstimer, har således fået repeteret ergodesætningen og definitionen af detaljeret balance en ekstra gang, mens de 4 ekstra studerende kun har fået en hurtig repetition af disse emner.

## 4.3 Øvelserne

De 7 som kom på øvelsesholdet fulgtes, med Peter op til Matematisk Instituts frokoststue som i denne anledning var blevet inddraget til undervisningslokale. De studerende fik udleveret opgavesættet og fik at vide hvordan det ville foregå. Der var 60 minutter til rådighed, som de skulle bruge til at løse flest muligt opgaver. De kunne arbejde sammen, som de havde lyst, og de kunne spørge Peter om stort og småt, men det ville ikke være en klassetime, så der ville ikke blive nogen fælles gennemgang af dele af stoffet. Deres besvarelser af opgaverne skulle ikke afleveres, så de behøvede ikke at gøre noget ud af skrive det pænt op.

De regnede koncentreret i de følgende 60 minutter. De samarbejdede stort set ikke om opgaverne - måske fordi dem, de plejer at arbejde sammen, med var til forelæsninger. I begyndelsen var de helt stille og spurgte stort set ikke. Stemningen var åbenbart lidt prøveagtig, så Peter fandt en undskyldning for at forlade lokalet i 5 minutter. Da han kom tilbage, var stemningen lidt lettere, og de var mere villige til at spørge. Efter en time sluttede øvelsestimen, og der blev holdt en pause indtil den sidste del af dagens program. De fleste nåede hovedparten af de spørgsmål, der ikke drejede sig om bevis for ergodesætningen. Ingen nåede alle spørgsmål. Der var almindelig enighed om, at øvelsestimen havde været ret tidspresset.

# 5 Resultater og behandling af data

## 5.1 Behandling af testdata

Tabel 1 indeholder resultatet af testen. Spørgsmål 1 i testen adskiller sig fra spørgsmål 2 til 8 idet det ikke vedrører indholdet af undervisningen under forsøget, men omhandler stof som skulle være kendt på forhånd. På trods af dette er spørgsmål 1 blevet behandlet sammen med de øvrige spørgsmål.

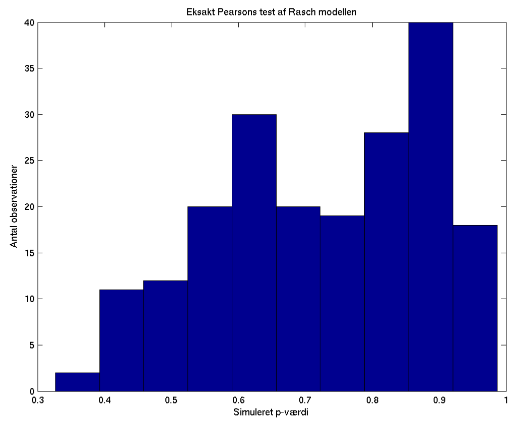
### 5.1.1 Beregning af indlæringsparametre og model kontrol

Vi ønsker at tildele hver enkelt studerende en parameter, der kvantificerer indlæringen under undervisningsforsøget. Hvis rækkerne og søjlerne i tabel 1 er uafhængige, så kan det antages at alle spørgsmålene måler den samme slags viden om MCMC hos de studerende. Idet denne viden kan antages at stamme fra forsøgsundervisningen, vil rækkesummerne dermed kvantificere indlæringen. Hvis alle testspørgsmålene enten er besvaret helt rigtige eller helt forkerte, så haves en tabel af 0'er og 1'er. Uafhængighed mellem rækker og søjler svarer i dette tilfælde til at data kan beskrives ved den såkaldte Rasch model. For at omforme delvis rigtige svar, som har fået tildelte en score mellem 0 og 1, omsættes disse værdier til enten 0 eller 1, idet 1 vælges med sandsynlighed givet ved scoren. For denne tilfældige tabel af 0'er og 1'er testes Rasch modellen. Vi har gjort dette for 100 forskellige realiseringer af disse tabeller, og et histogram over de tilhørende p-værdier[1] findes i figur 1.

[1]MCMC metoden bruges til at beregne p-værdierne af det såkaldte eksakte test baseret på Pearsons teststørrelse.

| Stud. nr. | Undervis. f/ø | sp.1 0-1 | sp. 2 0-1 | sp. 3 0-1 | sp. 4 0-1 | sp.5 0-1 | sp.6 0-1 | sp. 7 0-1 | sp.8 0-1 | i alt 0-8 | rang |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | ø | 0.6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 1 |
| 11 | ø | 0.1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1.1 | 2 |
| 2 | f | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2.0 | 3.5 |
| 12 | ø | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2.0 | 3.5 |
| 10 | ø | 1 | 0.2 | 0 | 0 | 1 | 0 | 0 | 0.5 | 2.7 | 5 |
| 9 | ø | 1 | 0.8 | 1 | 0 | 0 | 0.5 | 0 | 0 | 3.3 | 6 |
| 6 | ø | 0.6 | 0 | 1 | 0 | 1 | 0.8 | 0 | 0 | 3.4 | 7.5 |
| 15 | f | 1 | 0 | 0 | 1 | 1 | 0.4 | 0 | 0 | 3.4 | 7.5 |
| 1 | f | 0.6 | 1 | 1 | 0.7 | 0.5 | 0.5 | 0 | 0 | 4.3 | 9 |
| 13 | f | 0.4 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 4.4 | 10 |
| 16 | ø | 0.3 | 0.7 | 1 | 0.5 | 1 | 0 | 0 | 1 | 4.5 | 11 |
| 5 | f | 0.4 | 1 | 1 | 1 | 1 | 0.2 | 0 | 0 | 4.6 | 12 |
| 18 | f | 0.4 | 1 | 1 | 0 | 1 | 0.8 | 0 | 0.5 | 4.7 | 13 |
| 17 | f | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 5.0 | 14 |
| 4 | f | 1 | 0.2 | 1 | 1 | 1 | 0 | 1 | 0.5 | 5.7 | 15 |
| 7 | f | 1 | 0.3 | 1 | 1 | 1 | 0.5 | 1 | 0 | 5.8 | 16 |
| 3 | f | 1 | 1 | 0.8 | 1 | 1 | 1 | 0 | 0.5 | 6.3 | 17.5 |
| 14 | f | 1 | 1 | 1 | 1 | 1 | 0.8 | 0 | 0.5 | 6.3 | 17.5 |
| | snit fore. | 0.80 | 0.68 | 0.82 | 0.40 | 0.96 | 0.38 | 0.18 | 0.18 | 4.40 | |
| | snit øvl. | 0.66 | 0.24 | 0.43 | 0.07 | 0.71 | 0.19 | 0 | 0.21 | 2.51 | |

Tabel 1: Tabel for resultat af test af undervisning og opgaveregning, sorteret efter rang.

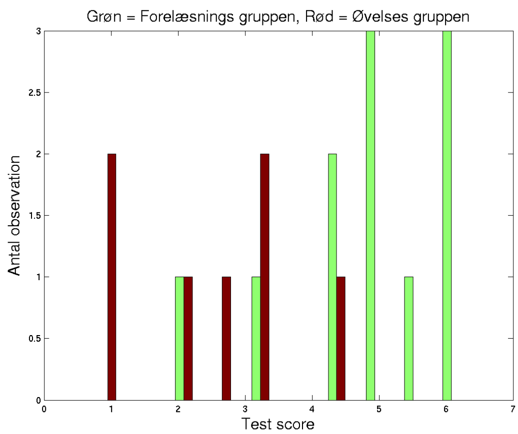



Figur 1: p-værdier for accept af Rasch modellen for randomiserede tabeller

Disse p-værdier udviser stor variation i området fra 0.3265 til 0.9865, men idet ingen af de 100 tilfældige tabeller fører til forkastelse af Rasch modellen accepteres hypotesen om uafhængighed mellem rækker og søjler.

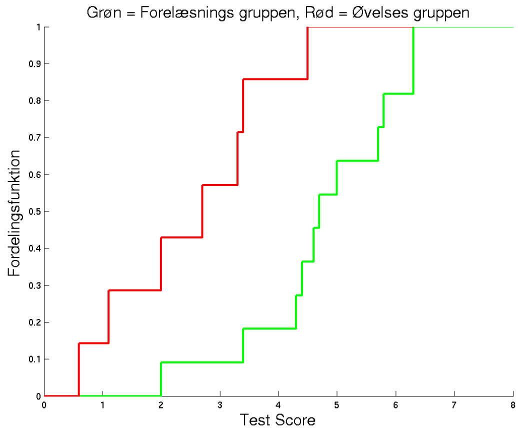
### 5.1.2 Signifikans test

Figure 2 viser et histogram og Figur 3 viser fordelingsfunktionen for testscoren for forelæsnings- og øvelsesholdet.




Figur 2: Histogram over testscore for de to hold

Tilsyneladende klarede forelæsningsholdet sig markant bedre end øvelsesholdet. Nedenfor undersøges om dette kan tilskrives en tilfældighed ved




Figur 3: Fordelingsfunktion for testscore for de to hold

lodtrækningen. Hvis forskellen ikke kan forklares ved en tilfældighed i lodtrækningen, så er konsekvensen at forskellen må skyldes forskellen mellem de to undervisningsforløb.

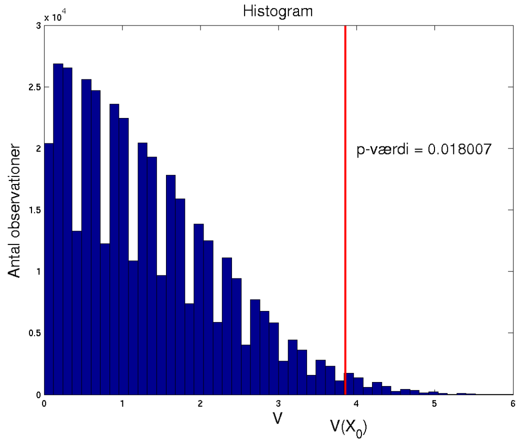
Vi rangordner de 18 studerende efter deres score og giver dem en placering fra 1 til 18, hvor 1 er dårligst. De to studerende på plads 3,4 henholdvis 7,8 henholdvis 17,18 har opnået den samme score, hvorfor de begge får placering 3.5 henholdvis 7.5 henholdvis 17.5. Herefter glemmes de studerende score, og kun deres placering huskes. Hvis placeringerne i forelæsnings- og øvelsesholdet skyldes tilfældigheden via lodtrækningen, så skal de studerende i øvelsesholdet i middel have opnået placering (1+18)/2. Vi bruger dermed teststørrelsen

$$V = \left| \frac{1}{7} \sum_{k=1}^{7} \text{placering af studerende } \# \, k \text{ ved øvelserne} - \frac{1+18}{2} \right|.$$

Dette giver $p$-værdien[2] $\mathbb{P}(V \geq 3.8571) = 0.0180$, se Figur 4. Det er altså højst usandsynlig, at forskellen kan forklares ved lodtrækningen alene. [3]

[2] Denne p-værdi beregnes ligeledes via MCMC

[3] Man kan argumentere for at de 4 senere tilkomne studerende på forelæsningsholdet skulle udelades af forsøget. Vi har gentaget signifikanstestet med de to bedste studerende udeladt fra forelæsningsholdet. Dette giver p-værdien $\mathbb{P}(V \geq 2.8571) = 0.0465$. Konklusionen er derfor robust overfor den potentielle selektion bias.




Figur 4: Signifikanstest for forskel mellem forsøgsholdene

## 5.2 De studerendes forberedelse og opfattelse af undervisningsform før forsøget.

Tabel 2 og Tabel 3 sammenfatter de studerendes besvarelse af spørgeskema 1.

**Forberedelsestid** Som forventet er der stor spredning på de studerendes forberedelsestid. Der er en svag tendens til at der bruges mindre tid til at forberede sig før en forelæsning i forhold til den tid derbruges til forberedelse til øvelser og forståelse af en forelæsning. Som nævnt tidligere er studiet normeret til at tage $7\frac{1}{2}$ time af de studerendes tid per uge, hvoraf de $4\frac{1}{2}$ time eller 270 minutter er afsat til forberedelse/efterbehandling hjemme. Der er kun 4 ud af de 18 studerende, som forbereder sig mere end normeringen.

**Undervisningsform** Alle på nær een studerende mener øvelser vil give den mest effektive udnyttelse af en times undervisning, og næsten alle mener at de får mere ud af øvelser end af forelæsninger.

**Stoffets sværhedsgrad** De studerende mener generelt at stoffet og forelæsning er alt for svært til udfordrende. Der er en tendens til at øvelser opfattes som en anelse lettere end stof til forelæsninger.

| Stud. nummer | For-. F min | Efter F. min | Forb. Øv. min | SUM min |
|---|---|---|---|---|
| 1 | 0 | 60 | 60 | 120 |
| 2 | 60 | 120 | 60 | 240 |
| 3 | 120 | 120 | 120 | 360 |
| 4 | 30 | 30 | 30 | 90 |
| 5 | 20 | 20 | 30 | 70 |
| 6 | 30 | 30 | 60 | 120 |
| 7 | 30 | 0 | 60 | 90 |
| 8 | 0 | 30 | 60 | 90 |
| 9 | 120 | 60 | 120 | 300 |
| 10 | 20 | 30 | 10 | 60 |
| 11 | 120 | 120 | 60-120 | 330 |
| 12 | 30 | 30 | 60 | 120 |
| 13 | 15 | 30 | 30 | 75 |
| 14 | 60 | 60 | 120 | 240 |
| 15 | 30 | 30 | 30 | 90 |
| 16 | 60 | 180-240 | 60 | 330 |
| 17 | 120-180 | 60-120 | 120-180 | 390 |
| 18 | 60 | 60 | 60 | 180 |

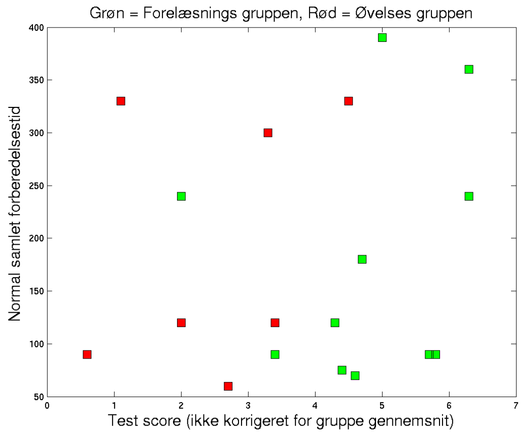
Tabel 2: Resultaterne af spørgeskema 1 vedrørende tidsforbug. Hvor et interval af forbrugt tid er angivet benyttes gennemsnittet til at beregne summen af forbrugt tid.

| Stud. nr. | Mest udb. For./Øv. | Selv opg. Ja/Nej | Mest eff. For./Øv. | Stoffet S/U/T/L | Forelæsninger S/U/T/L | Øverlser S/U/T/L |
|---|---|---|---|---|---|---|
| 1 | ø | n | ø | u | u | u |
| 2 | ø | j | ø | s | u | u |
| 3 | f/ø | n | ø | u | u | u |
| 4 | ø | n | ø | u | u | u |
| 5 | ø | j | ø | u | t | u |
| 6 | ø | n | ø | u | u | u |
| 7 | ø | n | ø | s | u | u |
| 8 | f/ø | n | ø | s/u | u | s |
| 9 | ø | j | ø | s | s | t |
| 10 | ø | n | ø | u | u | u |
| 11 | ø | n | ø | s | s | u |
| 12 | ø | n | ø | u | u | t |
| 13 | ø | j | ø | s | s | u |
| 14 | ø? | j | ø | u | u | u |
| 15 | ø | j | ø | u | u | u |
| 16 | ø | n | ø | u | s | u |
| 17 | ø | n | ø | u/s | u/s | u |
| 18 | f | j | selvstudie/f | u | u | t |

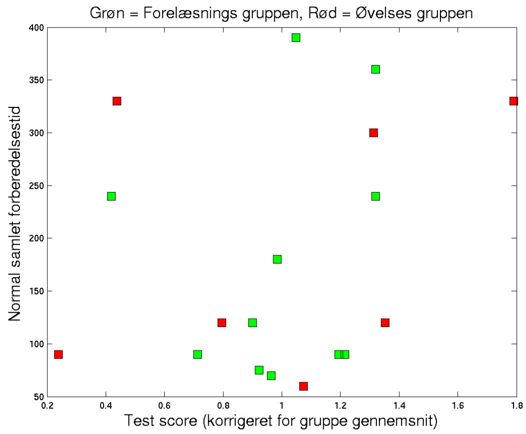
Tabel 3: Resultaterne af spørgeskema 1: S:Alt for svært, U:Udfordrende, T:Tilpas, L:Let.

### 5.2.1 Forberedelsestid vs. testscore, grupperet i undervisningshold

Figur 5-6 viser sammenhængen mellem forberedelsestid og testscore, for forelæsnings- og øvelsesholdet. I Figur 6 er testscoren korrigeret for hold gennemsnit.




Figur 5: Plot af forberedelsestid mod testscore



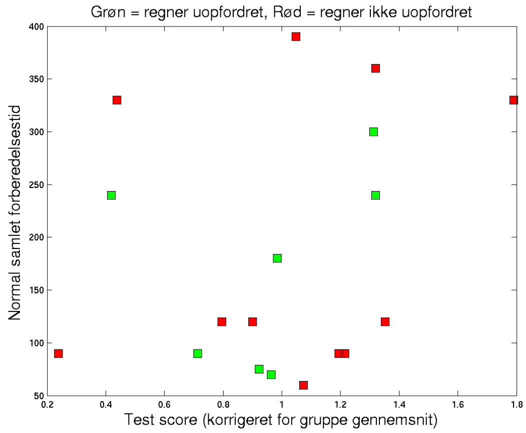

Figur 6: Plot af forberedelsestid mod testscore

Figur 5 viser at for dem, der forbereder sig lidt ($\leq 150$ minutter), klarer forelæsningsholdet sig markant bedre end øvelsesholdet. Ud af de 18 studerende er der to der både bruger lang tid på at forberede sig og klarer sig dårligt. Pædagogisk set er dette uheldigt. Ellers synes studerende med en stor arbejdsindsats at klare sig godt i testen. Der er også en gruppe af studerende som klarer sig godt på trods af lav forberedelsestid.

### 5.2.2 Forberedelsestid vs. testscore, grupperet efter om de studerende regner uopfordret eller ej

Figure 7-8 viser sammenhængen mellem forberedelsestid og testscore for de studerende der henholdsvis 'regner' og 'ikke regner' uopfordret. Umiddelbart forventende vi at studerende som på eget initiativ regner opgaver er engagerede og har en større forståelse af stoffet, og dermed ville klare sig bedre i testen. En smule overraskende ser det ikke ud til at der er nogen sammenhæng med om man uopfordret regner opgaver og klarer sig godt.


Figur 7: Plot af forberedelsestid mod testscore.




Figur 8: Plot af forberedelsestid mod testscore.

# 6 Konklusion

Kort er konklusionen at med en tidsbegrænset undervisning på 60 minutters er forelæsninger mere effektive end øvelser. Den klare tendens i vores data står i modsætning til vores egne forventninger. Konklusionen står ligeledes i modsætning til de studerendes opfattelse af, at en times øvelser er mere effektiv end en times forelæsning. Denne opfattelse skyldes formodelig at de studerende ikke har taget i betragtning at øvelser kræver mere tid end forelæsninger. Videre er øvelser formodenlig mere effektive hvis der har været forudgående forelæsninger.

**Tidspres for øvelsesholdet.** Da vi lavede undervisningsmaterialet, var det ret klart for os, at øvelsesholdet ville være under tidspres. Det var svært at lave et opgavesæt, som man ville kunne nå at regne inden for den afsatte tid og samtidig lære de begreber, som ville blive gennemgået til forelæsningen.

Vi forventede, at de enkelte studerendes almene dygtighed ville slå igennem og overskygge forskellene mellem undervisningsmetoderne. Det var imidlertid ikke tilfældet. øvelsesholdet klarede sig klart dårligere end forelæsningsholdet. Selv i spørgsmål, der var stort set identiske med, hvad der blev regnet på til øvelserne, klarede forelæsningsholdet sig bedst. Man kunne forestille sig, at dem på øvelsesholdet var blevet gode til visse af spørgsmålene, og dem på forelæsningsholdet var blevet gode til andre. Men der er intet i talmaterialet der tyder på dette (test af Rasch-modellen). Ligeledes er det ret usandsynligt (2%) at resultatet skyldtes, at de dygtige studerende tilfældigvis var kommet på forelæsningsholdet. Alt tyder på at tidsfaktoren har været det helt afgørende, som flere af de studerende også noterede det som svar på skema 2, eksempelvis:

> "*Idet jeg ikke fik regnet alle opgaverne, har det ikke været muligt at danne sig et samlet overblik over stoffet. Derfor er en det ikke rimeligt at sammenligne (forelæsninger og øvelser).*" (studerende fra øvelsesholdet)
>
> "*Jeg føler jeg har spildt min tid og er gået glip af en forelæsning.*" (studerende fra øvelsesholdet)
>
> "*Det har været spændende men man havde fået langt mere ud af det hvis man kunne have regnet eller set på opgaverne hjemme.*" (studerende fra øvelsesholdet)

Når de studerende føler, at de lærer mest ved øvelser, er det muligvis fordi det er her, de kommer til en endelig afklaring af stoffet. Denne undersøgelse tyder dog på, at de ville have langt mindre udbytte af øvelserne dersom de ikke først havde været til forelæsninger. Videre er det velkendt at man kun kan bevare koncentrationen i et begrænset stykke tid til forelæsninger, mens man ved regning af opgaver kan lade tingene fungere mere i ens eget tempo og derved vil kunne regne relativt koncentreret i ganske lang tid.

Det er væsentligt at skelne mellem rene regneøvelser og klasseundervisning. Når de studerende i det første spørgeskema svarede at de mente at man ville lære mere ved en times øvelser end ved en times forelæsninger, har de muligvis tænkt på klasseundervisning og ikke rene regneøvelser. Flere studerende angav at de ville have fået mere ud af normal klasseundervisning end øvelsen i forsøget:

> "*Ret ubrugeligt da jeg kun nåede til opgave 6. Man mangler fuldstændigt overblik og forståelse af stoffet. Det er svært at regne opgaver om noget man aldrig har set før. Det kunne være at det havde virket bedre med en introduktion til stoffet før man begyndte at regne. Det tager meget længere tid at lære ved at regne opgaver end at sidde til en forelæsning. Man lærer nok mere af at regne opgaver generelt, men ikke i dette tilfælde da der var alt for kort tid.*" (studerende fra øvelsesholdet)

**Lettere forelæsning end normalt?** Der er noget der tyder på at forelæsningen var en nemmere end de studerende er vant til. Dels var der god tid og forelæsningen fulgte ikke den sædvanlige lærebog som får en del negative kommentarer med på vejen af de studerende. En studerende bemærker:

> "*Ja, jeg syntes jeg har lært noget, forstod hvad der foregik, men synes også at stoffet virkede lettere end normalt. Muligvis fordi Norris (red.: lærebogen) er svært tilgængelig.*" (studerende fra forelæsningsholdet)

Det afsluttende spørgeskema og afrundende diskussion med de studerende viste at de er positive over for denne type eksperimenter og forsøg på at forbedre undervisningen. For eksempel:

> "*Jeg syntes det var udemærket. Specielt kunne jeg godt lide testen, hvor man fik testet hvad man havde lært på en uformel måde og man kan se hvad man ikke har forstået og derfor med fordel kan kigge mere på.*" (studerende fra forelæsningsholdet)

Selv synes vi også det har været yderst interessant at arbejde med undervisningsforsøget, som har givet os en bedre forståelse af forskellige undervisningsmetoder.

# Appendix A

## 1 Spørgeskemaer

### 1.1 Spørgeskema 1 : Før forsøget

# Spørgeskema vedrørende undervisningen i Statistik 1B

| |
|---|
| Hvor meget tid bruger du typisk på at forberede dig til en forelæsning? |
| Hvor meget tid bruger du typisk efter en forelæsning på at forstå teorien? |
| Hvor meget tid bruger du typisk på at forberede dig til øvelserne? |
| Hvad synes du, at du plejer at få mest ud af? ☐ forelæsning, ☐ øvelser. |
| Regner du nogle gange opgaver på eget initiativ? ☐ Ja, ☐ Nej |
| Hvis du har en time til rådighed, hvad tror du så vil være<br>den mest effektive indlæringsmetode? ☐ forelæsning, ☐ øvelse. |
| Synes du stoffet er: ☐ alt for svært, ☐ udfordrende, ☐ tilpas, ☐ for let. |
| Synes du forelæsningerne er: ☐ alt for svære, ☐ udfordrende, ☐ tilpas, ☐ for lette. |
| Synes du øvelserne er: ☐ alt for svære, ☐ udfordrende, ☐ tilpas, ☐ for lette. |

### 1.2 Spørgeskema 2 : Efter forsøget

## Afrundende spørgeskema

| |
|---|
| Hvad synes du om dagens undervisningsforløb? Hvorfor? |
| Har du lært noget nyt om Markov kæder?<br>Havde det været bedre med "almindelige" forelæsninger og øvelser? |
| Hvad har du lært om undervisning i sig selv? |
| Giv gerne forslag til forbedring/ændring af undervisningen på Statistik 1B: |

# Appendix B

De følgende sider udgør undervisningsmaterialet udleveret til de studerende efter gennemførelsen af forsøget. Materialet til forelæsningerne og øvelserne findes i henholdvis afsnit 1 og afsnit 2. Dette materiale blev udlevet til de respektive grupper i starten af forsøget.

Jesper Lund Pedersen, Peter Harremoës & Bo Markussen

# Markovkæde Monto Carlo (MCMC) med Metropolis-Hastings algoritmen

## Undervisningsforsøg

Dette notat udgør undervisningsmaterialet til det didaktiske forsøg gennemført på kurset Statistik 1B mandag d. 18. oktober 2004. Formålet med forsøget var at undersøge effekten af indlæring ved henholdvis klassiske deduktive forelæsninger og induktiv opgaveregning. Forelæsningerne findes i afsnit 1, opgavematerialet findes i afsnit 2 og testmaterialet findes i afsnit 3. Endelig indeholder afsnit 4 et ikke-trivielt eksempel der gennemgås ved forelæsningerne mandag d. 25. oktober.

## 1 Forelæsning

Lad $(X_n)_{n\geq 0}$ være Markov $(\lambda, P)$ på $I$. Ved forelæsningerne mandag d. 4. oktober blev følgende sætning diskuteret.

**Sætning (1.7.2, udvidet version).** *Hvis $p_{ij}^{(n)} \to \pi_j$ for $n \to \infty$ for alle $i, j \in I$ og $\sum_{j\in I} \pi_j = 1$, så er $\pi = (\pi_j)_{j\in I}$ en invariant fordeling for $(X_n)_{n\geq 0}$.*

*Bevis.* Ifølge Fatous lemma (målteori) gælder

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in I} p_{ik}^{(n)} \, p_{kj} \geq \sum_{k\in I} \lim_{n\to\infty} p_{ik}^{(n)} \, p_{kj} = \sum_{k\in I} \pi_k \, p_{kj}.$$

Dermed giver Tonelli's sætning (målteori)

$$1 = \sum_{j\in I} \pi_j \geq \sum_{j\in I} \sum_{k\in I} \pi_k \, p_{kj} = \sum_{k\in I} \pi_k = 1,$$

hvoraf vi ser at $\pi_j = \sum_{k\in I} \pi_k \, p_{kj}$. □

En Markovkæde siges at være periodisk med periode $d$ hvis der findes en klassedeling $I = I_1 \cup \cdots \cup I_d$ således at

$$\mathbb{P}\big(X_{n+1} \in I_{k+1} \bigm| X_n = i\big) = 1 \quad \text{for alle } i \in I_k \text{ og } k = 1, \ldots, d,$$

hvor $I_{d+1} = I_1$. En 2-periodisk Markovkæde er således skiftevis i "lige" og "ulige" tilstande. En Markovkæde der ikke er $d$-periodisk for noget $d \geq 2$, kaldes aperiodisk. Ved forelæsningerne mandag d. 25. oktober vil den omvendte af Sætning 1.7.2 blive vist.

**Sætning (1.8.3).** *Antag* $(X_n)_{n\geq 0}$ *er irreducibel, aperiodisk og positiv rekurrent med invariant fordeling* $\pi$. *Så gælder* $p_{ij}^{(n)} \to \pi_j$ *for* $n \to \infty$ *for alle* $i, j \in I$.

Vi kan nu vise:

**Sætning (Ergodesætningen for Markovkæder).** *Antag* $(X_n)_{n\geq 0}$ *er irreducibel og positiv rekurrent med invariant fordeling* $\pi$. *For alle begrænsede* $f\colon I \to \mathbb{R}$ *og alle* $\varepsilon > 0$ *gælder da*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \sum_{i\in I} \pi_i\, f(i)\right| > \varepsilon\right) \xrightarrow[N\to\infty]{} 0. \tag{1}$$

*Bevis.* Vi kan antage at $(X_n)_{n\geq 0}$ er aperiodisk. At slippe af med denne antagelse er en teknisk øvelse, som overlades til læseren. Ifølge Sætning 1.8.3 gælder dermed $p_{ij}^{(n)} \to \pi_j$ for $n \to \infty$. Vi regner

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right] &= \frac{1}{N}\sum_{i,j\in I}\sum_{n=0}^{N-1} \lambda_i\, p_{ij}^{(n)}\, f(j) = \sum_{i,j\in I} \lambda_i \left(\frac{1}{N}\sum_{n=0}^{N-1} p_{ij}^{(n)}\right) f(j) \\ &\xrightarrow[N\to\infty]{} \sum_{i,j\in I} \lambda_i\, \pi_j\, f(j) = \sum_{j\in I} \pi_j\, f(j).\end{aligned} \tag{2}$$

Idet, som vil blive vist nedenfor,

$$\sup_{n\geq 0} \left|\operatorname{Cov}\left[f(X_n), f(X_{n+m})\right]\right| \xrightarrow[m\to\infty]{} 0, \tag{3}$$

gælder

$$\operatorname{Var}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right] = \frac{1}{N^2}\sum_{n,m=0}^{N-1} \operatorname{Cov}\left[f(X_n), f(X_m)\right] \xrightarrow[N\to\infty]{} 0. \tag{4}$$

Resultatet (1) følger ved at indsætte (2) og (4) i Chebychevs ulighed. Altså

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \sum_{i\in I}\pi_i\, f(i)\right| > \varepsilon\right) \\ &\leq \mathbb{P}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right]\right| > \varepsilon - \left|\mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right] - \sum_{i\in I}\pi_i\, f(i)\right|\right) \\ &\leq \left(\varepsilon - \left|\mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right] - \sum_{i\in I}\pi_i\, f(i)\right|\right)^{-2} \operatorname{Var}\left[\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right] \xrightarrow[N\to\infty]{} 0.\end{aligned}$$

Vi mangler at vise (3). Der gælder

$$\begin{aligned}
\operatorname{Cov}\big[f(X_n), f(X_{n+m})\big] &= \mathbb{E}\big[f(X_n)\, f(X_{n+m})\big] - \mathbb{E}\big[f(X_n)\big]\,\mathbb{E}\big[f(X_{n+m})\big] \\
&= \sum_{i,j,k\in I} \lambda_i\, p_{ij}^{(n)}\, p_{jk}^{(m)}\, f(j)\, f(k) - \Big(\sum_{i,j\in I} \lambda_i\, p_{ij}^{(n)}\, f(j)\Big)\Big(\sum_{g,k\in I} \lambda_g\, p_{gk}^{(n+m)}\, f(k)\Big) \\
&= \sum_{g,i,j\in I} \lambda_g\, \lambda_i\, p_{ij}^{(n)} \sum_{k\in I} \big(p_{jk}^{(m)} - p_{gk}^{(n+m)}\big)\, f(j)\, f(k).
\end{aligned}$$

Lad $\delta > 0$ være givet. Idet der for alle endelige delmængder $I_0$ af $I$ gælder at $\sum_{g,i,j\in I_0} \lambda_g\, \lambda_i\, p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \sum_{g,i,j\in I_0} \lambda_g\, \lambda_i\, \pi_j$ kan vi vælge en endelig delmængde $I_0$ således at

$$\inf_{n\geq 0} \sum_{g,i,j\in I_0} \lambda_g\, \lambda_i\, p_{ij}^{(n)} > 1 - \delta.$$

Videre gælder

$$\sup_{g,j\in I_0} \sup_{\nu\geq 0} \sum_{k\in I} \big|p_{jk}^{(m)} - p_{gk}^{(\nu+m)}\big| \leq \sup_{g,j\in I_0} \sum_{k\in I} \big|p_{jk}^{(m)} - \pi_k\big| + \sup_{g,j\in I_0} \sup_{\nu\geq 0} \sum_{k\in I} \big|\pi_k - p_{gk}^{(\nu+m)}\big|$$

som ifølge majoriseret konvergens (målteori) konvergerer mod 0 for $m \to \infty$. Hvis vi uden tab af generalitet antager at $|f(i)| \leq 1$ for alle $i \in I$, så fås dermed

$$\begin{aligned}
\sup_{n\geq 0}\big|\operatorname{Cov}\big[f(X_n), f(X_{n+m})\big]\big| &\leq \sup_{n\geq 0} \sum_{g,i,j\in I} \lambda_g\, \lambda_i\, p_{ij}^{(n)} \sup_{\nu\geq 0} \sum_{k\in I} \big|p_{jk}^{(m)} - p_{gk}^{(\nu+m)}\big| \\
&\leq \sup_{n\geq 0} \sum_{g,i,j\in I_0} \lambda_g\, \lambda_i\, p_{ij}^{(n)} \sup_{\nu\geq 0} \sum_{k\in I} \big|p_{jk}^{(m)} - p_{gk}^{(\nu+m)}\big| + 2\Big(1 - \inf_{n\geq 0} \sum_{g,i,j\in I_0} \lambda_g\, \lambda_i\, p_{ij}^{(n)}\Big) \\
&\leq \sup_{g,j\in I_0} \sup_{\nu\geq 0} \sum_{k\in I} \big|p_{jk}^{(m)} - p_{gk}^{(\nu+m)}\big| + 2\delta \xrightarrow[m\to\infty]{} 2\delta \xrightarrow[\delta\to 0]{} 0.
\end{aligned}$$

Dette afslutter beviset. $\square$

**Konklusion:** Vi kan altså tilnærmelsesvis beregne $\sum_{i\in I} \pi_i\, f(i)$ ved at simulere $\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)$ for passende stort $N$ for en irreducibel Markovkæde $(X_n)_{n\geq 0}$ med invariant fordeling $\pi$.

**Spørgsmål:** Hvordan konstrueres en sådan Markovkæde?

**Definition.** *En stokastisk matrix $P$ og et mål $\lambda$ siges at være i "detaljeret balance" hvis der for alle $i, j \in I$ gælder*

$$\lambda_i\, p_{ij} = \lambda_j\, p_{ji}.$$

**Lemma.** *Hvis $P$ og $\lambda$ er i detaljeret balance, så er $\lambda$ invariant for $P$.*

*Bevis.* Der gælder

$$\big(\lambda P\big)_i = \sum_{j \in I} \lambda_j \, p_{ji} = \sum_{j \in I} \lambda_i \, p_{ij} = \lambda_i.$$

$\square$

**Svar:** Vi kan altså vælge overgangssandsynlighederne således at $P$ og $\pi$ er i detaljeret balance.

**Spørgsmål:** Hvordan kan vi konstruere en stokastisk matrix $P$, som er i detaljeret balance med en given fordeling $\pi$?

**Svar: Metropolis-Hastings algoritmen.** Antag vi har en stokastisk matrix $Q$ således at vi for alle $i \in I$ let kan simulere fra fordelingen $(q_{ij})_{j \in I}$. Vi definerer sandsynlighederne $\alpha_{ij}$ ved

$$\alpha_{ij} = \begin{cases} 1 & \text{hvis } 0 \le \pi_i \, q_{ij} \le \pi_j \, q_{ji}, \\ \dfrac{\pi_j \, q_{ji}}{\pi_i \, q_{ij}} & \text{hvis } \pi_i \, q_{ij} > \pi_j \, q_{ji} \ge 0. \end{cases}$$

Vi konstruerer nu en Markov kæde $(X_n)_{n \ge 0}$ udfra følgende algoritme:

(1) Antag $X_n = i$.

(2) Med sandsynlighed $q_{ij}$ bliver der forelået et spring fra $i$ til $j$.

(3) Uafhængigt af (2) bliver springet fra $i$ til $j$ accepteret med sandsynlighed $\alpha_{ij}$, dvs. vi sætter $X_{n+1} = j$. Ellers forbliver kæden i tilstand $i$, dvs. vi sætter $X_{n+1} = i$.

Dermed gælder for $i \neq j$,

$$\begin{aligned} p_{ij} &= \mathbb{P}\big(X_{n+1} = j \;\big|\; X_n = i\big) \\ &= \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver foreslået}, \text{springet fra } i \text{ til } j \text{ bliver accepteret}\big) \\ &= \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver foreslået}\big) \, \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver accepteret}\big) \\ &= q_{ij} \, \alpha_{ij} \end{aligned}$$

og dermed

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij} = 1 - \sum_{j \neq i} q_{ij} \, \alpha_{ij} = q_{ii} + \sum_{j \neq i} q_{ij} \, \big(1 - \alpha_{ij}\big).$$

Det indses at $P$ og $\pi$ er i detaljeret balance. Thi hvis $\pi_i\, q_{ij} = \pi_j\, q_{ji}$ så gælder

$$\pi_i\, p_{ij} = \pi_i\, q_{ij} = \pi_j\, q_{ji} = \pi_j\, p_{ji}$$

og hvis $\pi_i\, q_{ij} > \pi_j\, q_{ji}$ så gælder

$$\pi_i\, p_{ij} = \pi_i\, q_{ij}\, \alpha_{ij} = \pi_i\, q_{ij}\, \frac{\pi_j\, q_{ji}}{\pi_i\, q_{ij}} = \pi_j\, q_{ji} = \pi_j\, p_{ji}.$$

**Konklusion:** Antag overgangssandsynlighederne $Q$ og acceptsandsynlighederne $\alpha$ er givet på tilstrækkelig eksplicit form til at vi kan simulere fra disse fordelinger. Så kan vi også simulere fra overgangssandsynlighederne $P$, og $P$ og $\pi$ er i detaljeret balance.

**Eksempel.** *Antag $\pi$ er ligefordelingen på $I = \{1,2,3,4\}$ og overgangssandsynlighederne $Q$ er givet ved*

$$Q = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

*Idet $Q$ og fordelingen $\lambda = (\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ er i detaljeret balance kan vi altså ikke bruge $Q$ direkte til at lave MCMC for fordelingen $\pi \neq \lambda$. Vi finder acceptsandsynlighederne $\alpha$ og overgangssandsynlighederne $P$ givet ved*

$$\alpha = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 1/3 & 0 & 0 & 2/3 \end{pmatrix}.$$

*Det eftervises let at $P$ og $\pi$ er i detaljeret balance.*

# 2 Opgaver om Markovkæde Monte Carlo

Dette er en opgave om Markov kæde Monte Carlo. Der er afsat 60 minutter til opgaven.

## Indledning:

I denne opgave betragter vi følgende problem: Givet en sandsynlighedsfordeling $\pi$ på $I$ og en begrænset funktion $f : I \to \mathbb{R}$ ønsker vi ved simulation at tilnærme $\sum_{j \in I} \pi_j f(j)$.

Hvis $X_0, X_1, X_2, \ldots$ er uafhængige og alle med fordeling $\pi$, da gælder der ifølge Store Tals Lov at

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \sum_{j\in I}\pi_j f(j)\right| > \varepsilon\right) \to 0 \text{ for } N \to \infty.$$

Hvis vi kan simulere $\pi$ direkte kan vi bruge Store Tals Lov til at finde en tilnærmet værdi af $\sum_{j\in I}\pi_j f(j)$.

*Spørgsmål 1.* Lad $\pi$ være en ligefordeling på fire punkter med $I = \{1, 2, 3, 4\}$. Lad $U$ være en variabel med en ligefordeling på intervallet $[0, 1]$, hvordan kan man da konstruere en ny variabel $Y$, som har $\pi$ som fordeling?

I komplicerede situationer bruges en anden metode, kaldet Markov Chain Monte Carlo (MCMC). Ideen er at konstruere en Markov kæde $(X_n)_{n\geq 0}$ (eller en stokastisk matrix $P$) med invariant fordeling $\pi$. Hvis vi simulerer kæden for tilstrækkelig lang tid $n$, så vil fordelingen af $X_n$ være tæt på $\pi$. Målet med resten af opgaven er at studere MCMC metoden.

## Detaljeret balance:

Den givne fordeling $\pi$ fra spørgsmål 1, vi ønsker at simulere fra, er en ligefordeling på fire punkter. Antag at vi er i stand til at simulere fra en Markov kæde på de fire punkter med den stokastiske matrix

$$Q = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

*Spørgsmål 2.* Tegn overgangsgrafen for Markov kæden.

**Definition.** En stokastisk matrix $P$ og en fordeling $\lambda$ siges at være i detaljeret balance hvis der for alle $i, j \in I$ gælder, at $\lambda_i\, p_{ij} = \lambda_j\, p_{ji}$.

*Spørgsmål 3.* Vis at $Q$ og fordelingen $(\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ er i detaljeret balance.

*Spørgsmål 4.* Vis: Hvis $P$ og $\lambda$ er i detaljeret balance, så er $\lambda$ invariant for $P$. (Vink: regn på $(\lambda P)_i$).

Antag vi har en stokastisk matrix $Q$ givet således at vi for alle $i \in I$ let kan simulere fra fordelingen $(q_{ij})_{j\in I}$.

**Definition.** Vi definerer såkaldte acceptsandsynligheder $\alpha_{ij}$ ved

$$\alpha_{ij} = \begin{cases} 1 & \text{hvis } 0 \le \pi_i\, q_{ij} \le \pi_j\, q_{ji}, \\ \frac{\pi_j\, q_{ji}}{\pi_i\, q_{ij}} & \text{hvis } \pi_i\, q_{ij} > \pi_j\, q_{ji} \ge 0. \end{cases}$$

*Spørgsmål 5.* Eftervis at acceptsandsynlighederne ud fra $Q$ og $\pi$ er givet ved

$$\alpha = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 \end{pmatrix}.$$

**Metropolis-Hastings algoritmen:** Vi konstruerer nu en Markov kæde $(X_n)_{n\ge 0}$ ved algoritmen: Start med $X_0 = 1$.

(1) Antag $X_n = i$.

(2) Med sandsynlighed $q_{ij}$ bliver der foreslået et spring fra $i$ til $j$.

(3) Hvis $j$ er foreslået accepteres $j$ med sandsynlighed $\alpha_{ij}$ og vi sætter $X_{n+1} = j$. Ellers forbliver kæden i tilstand $i$, det vil sige vi sætter $X_{n+1} = i$.

Dermed gælder for $i \neq j$,

$$\begin{aligned} p_{ij} &= \mathbb{P}\big(X_{n+1} = j \;\big|\; X_n = i\big) \\ &= \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver foreslået}, \text{springet fra } i \text{ til } j \text{ bliver accepteret}\big) \\ &= \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver foreslået}\big)\, \mathbb{P}\big(\text{springet fra } i \text{ til } j \text{ bliver accepteret}\big) \\ &= q_{ij}\, \alpha_{ij} \end{aligned}$$

*Spørgsmål 6.* Vis at $p_{ii} = 1 - \sum_{j\neq i} p_{ij} = q_{ii} + \sum_{j\neq i} q_{ij}\,\big(1 - \alpha_{ij}\big)$.

*Spørgsmål 7.* Indse at den nye stokastiske matrix $P$ er givet ved

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 1/3 & 0 & 0 & 2/3 \end{pmatrix}$$

og at $P$ og $\pi$ er i detaljeret balance.

*Spørgsmål 8.* Vis at $P$ fra Metropolis-Hastings algoritmen og $\pi$ altid er i detaljeret balance.

Udfra den givne stokastiske matrix $Q$ (som antages valgt så vi kan simulere herfra) og fordelingen $\pi$ konstruerer Metropolis-Hastings algoritmen en ny stokastisk matrix $P$, som er i detaljeret balance med $\pi$. Yderligere kan vi simulere fra $P$ udfra $Q$ og $\alpha$.

## Ergodesætningen:

Lad $(X_n)_{n\geq 0}$ være Markov $(\lambda, P)$ på $I$. Ved en tidligere forelæsning blev følgende sætning gennemgået.

**Sætning (1.7.2, udvidet version).** *Hvis $p_{ij}^{(n)} \to \pi_j$ for $n \to \infty$ for alle $i, j \in I$ og $\sum_{j\in I} \pi_j = 1$, så er $\pi = (\pi_j)_{j\in I}$ en invariant fordeling for $(X_n)_{n\geq 0}$.*

I MCMC metoden får vi brug det omvendte resultat af Sætning 1.7.2. Resultatet vil blive gennemgået ved en kommende forelæsning. Begrebet aperiodisk, som optræder nedenfor, er en teknisk detalje, og vi får ikke brug for den præcise definition.

**Sætning (1.8.3).** *Antag $(X_n)_{n\geq 0}$ er irreducibel, aperiodisk og positiv rekurrent med invariant fordeling $\pi$. Så gælder $p_{ij}^{(n)} \to \pi_j$ for $n \to \infty$ for alle $i, j \in I$.*

Opgaven er nu at bevise Ergodesætningen.

**Sætning (Ergodesætningen for Markov kæder).** *Antag $(X_n)_{n\geq 0}$ er irreducibel og positiv rekurrent med invariant fordeling $\pi$. For alle begrænsede $f : I \to \mathbb{R}$ og alle $\varepsilon > 0$ gælder så*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \sum_{j\in I} \pi_j f(j)\right| > \varepsilon\right) \to 0 \text{ for } N \to \infty.$$

*Spørgsmål 9.* Overvej hvordan Ergodesætningen som teoretisk resultat kan benyttes til at approksimere $\pi_j$ og $\sum_{j\in I} \pi_j f(j)$.

*Spørgsmål 10.* Bevis Ergodesætningen under den ekstra betingelse at kæden er aperiodisk ved at bruge Chebyshevs ulighed. Som hjælp kan man besvare følgende underspørgsmål.

*Spørgsmål 10(i).* Vis at

$$\frac{1}{N}\mathbb{E}\Big[\sum_{n=0}^{N-1} f(X_n)\Big] \to \sum_{j\in I} \pi_j f(j) \text{ for } N \to \infty.$$

(Vink: brug Sætning 1.8.3 samt at hvis $a_n \to a$ så vil $\frac{1}{N}\sum_{n=0}^{N-1} a_n \to a$).

*Spørgsmål 10(ii).* Lad $\varepsilon$ være givet. Ifølge spørgsmål 10(i) findes $N_0$ således at

$$\Big|\frac{1}{N}\mathbb{E}\Big[\sum_{n=0}^{N-1} f(X_n)\Big] - \sum_{j\in I} \pi_j f(j)\Big| < \frac{\varepsilon}{2}$$

for $N > N_0$. Benyt Chebychevs ulighed: $\mathbb{P}(|Y - \mathbb{E}[Y]| > a) \leq a^{-2}\operatorname{Var}(Y)$ til at vise

$$\mathbb{P}\left(\Big|\frac{1}{N}\sum_{n=0}^{N-1} f(X_n) - \sum_{j\in I} \pi_j f(j)\Big| > \varepsilon\right) \leq \left(\frac{\varepsilon}{2}\right)^{-2} \operatorname{Var}\left(\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right)$$

for $N > N_0$.

*Spørgsmål 10(iii).* Vis at

$$\operatorname{Var}\left(\frac{1}{N}\sum_{n=0}^{N-1} f(X_n)\right) \to 0 \text{ for } N \to \infty$$

og benytte dette til at vise sætningen.

# 3 Evaluering

Dette er en evaluering af MCMC undervisningen. Der er **25** minutter til at svare på **8** spørgsmål. Du må gerne svare med matematik og/eller med ord.

1. Hvad er en Markov kæde?
2. Hvad er ideen bag Markov Chain Monte Carlo?
3. Hvad siger Ergodesætningen?
4. Kan man bruge Stor Tals Lov i stedet for Ergodesætningen?

5. Hvad vil det sige at $P$ og $\lambda$ er i detaljeret balance?

6. Lad $P$ være den stokastiske matrix fra Metropolis-Hastings algoritmen. Hvad skal der gælde om $P$ for at bruge Ergodesætningen?

7. Lad $Q$ og $\pi$ være i detaljeret balance. I Metropolis-Hastings algoritmen, hvad er da $\alpha$ og $P$?

8. På baggrund af den undervisning du har modtaget, mener du, at du ville være i stand til at skrive et computer program der udfører MCMC?

# 4 Uafhængighedstest i tovejstabel

I dette afsnit beskrives en ikke-triviel anvendelse af MCMC metoden til at gennemføre uafhængighedstest i tovejstabeller. Givet en tovejstabel

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1S} \\ X_{21} & X_{22} & \cdots & X_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ X_{R1} & X_{R2} & \cdots & X_{RS} \end{pmatrix}$$

ønsker vi altså at test nul hypotesen givet ved

$$H_0 : \text{rækkerne og søjlerne er uafhængige.}$$

## 4.1 Statistisk Model

Lad $R$ og $S$ være to tællelige mængder, som vi bruger til at indicere henholdvis rækker og søjler i en tovejstabel

$$X = \big(X_{rs}\big)_{r\in R, s\in S} \in \mathbb{N}_0^{R\times S}.$$

Række- og søjle summerne defineres ved

$$X_{r\cdot} = \sum_{s\in S} X_{rs}, \qquad\qquad X_{\cdot s} = \sum_{r\in R} X_{rs}.$$

Lad $M \in \mathbb{N}$ og en fordeling $\rho = (\rho_{rs})$ på $R \times S$ være givet. Antag at $X$ er polynomialfordelt med antalsparameter $M$ og sandsynlighedsparameter $\rho = (\rho_{rs})$. Altså for $x \in \mathbb{N}_0^{R\times S}$ med $\sum_{r\in R}\sum_{s\in S} x_{rs} = M$ gælder

$$\mathbb{P}\big(X = x\big) = \frac{M!}{\prod_{r\in R}\prod_{s\in S} x_{rs}!} \prod_{r\in R}\prod_{s\in S} \rho_{rs}^{x_{rs}}.$$

Række- og søjleindex er uafhængige netop hvis

$$H_0\colon \rho_{rs} = \Big(\sum_{s\in S}\rho_{rs}\Big)\Big(\sum_{r\in R}\rho_{rs}\Big) = \rho_{r\cdot}\,\rho_{\cdot s}. \tag{5}$$

I det følgende antages at nul hypotesen (5) er opfyldt. Givet række- og søjlesummer $x_{r\cdot}$ og $x_{\cdot s}$ betragtes fordelingen $\pi$ givet ved

$$\begin{aligned}\pi_x &= \mathbb{P}\big(X = x \bigm| X_{r\cdot} = x_{r\cdot}, X_{\cdot s} = x_{\cdot s}\big)\\ &= \frac{\frac{M!}{\prod_{r\in R}\prod_{s\in S} x_{rs}!}\prod_{r\in R}\prod_{s\in S}\big(\rho_{r\cdot}\,\rho_{\cdot s}\big)^{x_{rs}}}{\big(\frac{M!}{\prod_{r\in R} x_{r\cdot}!}\prod_{r\in R}\rho_{r\cdot}^{x_{r\cdot}}\big)\big(\frac{M!}{\prod_{s\in S} x_{\cdot s}!}\prod_{s\in S}\rho_{\cdot s}^{x_{\cdot s}}\big)}\\ &= \frac{\big(\prod_{r\in R} x_{r\cdot}!\big)\big(\prod_{s\in S} x_{\cdot s}!\big)}{M!}\Big(\prod_{r\in R}\prod_{s\in S} x_{rs}!\Big)^{-1}.\end{aligned} \tag{6}$$

Antag at vi for en given begrænset funktion $f\colon \mathbb{N}_0^{R\times S}\to\mathbb{R}$ ønsker at beregne middelværdierne

$$\mathbb{E}\big(f(X) \bigm| X_{r\cdot} = x_{r\cdot}, X_{\cdot s} = x_{\cdot s}\big) = \sum_{y\colon\, y_{r\cdot}=x_{r\cdot},\, y_{\cdot s}=x_{\cdot s}} \pi_y\, f(y). \tag{7}$$

Umiddelbart synes dette at være svært at gøre i praksis. Dette skyldes ikke blot at punktsandsynlighederne $\pi_y$ er ubehagelige små, men i særdeleshed også at mængden hvorover summen tages,

$$\big\{y\in\mathbb{N}_0^{R\times S}\colon\, y_{r\cdot} = x_{r\cdot}, y_{\cdot s} = x_{\cdot s}\big\},$$

er yderst vanskelig at ordne i en rækkefølge der kan gennemløbes i et computerprogram.

## 4.2 MCMC metoden

Nedenfor beskrives hvorledes vi kan simulere (7) via MCMC. Givet række- og søjlesummerne $x_{r\cdot}$ og $x_{\cdot s}$ indføres de endelige mængder

$$R_0 = \big\{r\in R\colon\, x_{r\cdot}\neq 0\big\}, \qquad S_0 = \big\{s\in S\colon\, x_{\cdot s}\neq 0\big\}.$$

Dermed kan $x$ på naturlig måde opfattes som et element i $\mathbb{N}_0^{R_0\times S_0}$. Vi indfører nu et lille trick idet vi opfatter $\mathbb{N}_0^{R_0\times S_0}$ som en delmængde af $\mathbb{Z}^{R_0\times S_0}$. Pointen med dette vil blive klar nedenfor, hvor vi konstruerer en matrice af overgangssandsynligheder på $\mathbb{Z}^{R_0\times S_0}$ således at

$$q_{yz} = q_{zy} \qquad \text{for alle } y,z\in\mathbb{Z}^{R_0\times S_0}.$$

Vi siger at to tovejstabeller $y, z \in \mathbb{Z}^{R_0 \times S_0}$ er naboer hvis der findes $r_1, r_2 \in R_0$ og $s_1, s_2 \in S_0$ således at

$$z_{rs} = \begin{cases} y_{rs} & \text{for } r \notin \{r_1, r_2\} \text{ eller } s \notin \{s_1, s_2\}, \\ y_{rs} + 1 & \text{for } (r, s) = (r_1, s_1) \text{ eller } (r, s) = (r_2, s_2), \\ y_{rs} - 1 & \text{for } (r, s) = (r_1, s_1) \text{ eller } (r, s) = (r_2, s_2). \end{cases}$$

Det indses let, at hvis $y$ og $z$ er naboer, så har de samme række- og søjlesummer. Matricen af overgangssandsynligheder $Q$ for MH-algoritmen vælges ved

$$q_{yz} = \begin{cases} \dfrac{2}{|R_0| \big(|R_0| - 1\big) |S_0| \big(|S_0| - 1\big)} & \text{hvis } y \text{ og } z \text{ er naboer,} \\ 0 & \text{ellers.} \end{cases}$$

Vi vælger altså to forskellige rækker i $R_0$ og to forskellige søjler i $S_0$ og foreslår at ændre tovejstabellen via additionsmasken

| | $s_1$ | $s_2$ |
|---|---|---|
| $r_1$ | $+1$ | $-1$ |
| $r_2$ | $-1$ | $+1$ |

De tilhørende acceptsandsynligheder er så givet ved

$$\begin{aligned} \alpha_{yz} &= \min\left\{1, \frac{\pi_z\, q_{zy}}{\pi_y\, q_{yz}}\right\} \\ &= \min\left\{1, \frac{\pi_z}{\pi_y}\right\} \\ &= \min\left\{1, \frac{y_{r_1,s_1}!\, y_{r_1,s_2}!\, y_{r_2,s_1}!\, y_{r_2,s_2}!}{z_{r_1,s_1}!\, z_{r_1,s_2}!\, z_{r_2,s_1}!\, z_{r_2,s_2}!}\right\} \\ &= \min\left\{1, \frac{y_{r_1,s_2}\, y_{r_2,s_1}}{(y_{r_1,s_1} + 1)\, (y_{r_2,s_2} + 1)}\right\}. \end{aligned}$$

Bemærk, at disse acceptsandsynligheder sikrer at kæder aldrig accepterer et spring til en tilstand der ikke ligger i $\mathbb{N}_0^{R_0 \times S_0}$. Hvis vi tror på at den konstruerede Markov kæde er irreducibel (hvilket rent faktisk er tilfældet), så kan vi dermed simulere (7).

## 4.3 Uafhængighedstest

Antag vi er givet en tovejstabel

$$X = \big(X_{rs}\big)_{r \in R, s \in S} \sim \text{polynomial}\big(M, \rho_{rs}\big).$$

Fra Statistik 0 kendes kvotientteststørrelsen $V$ for hypotesen om uafhængighed mellem rækkerne og søjlerne. Denne er givet ved

$$\begin{aligned} V(X) &= -2\log Q \\ &= 2\Big(\sum_{r,s} X_{rs}\,\log(X_{rs}) - \sum_{r} X_{r\cdot}\,\log(X_{r\cdot}) - \sum_{s} X_{\cdot s}\,\log(X_{\cdot s}) + M\,\log(M)\Big). \end{aligned}$$

På Statistik 0 bruges at for $M \to \infty$ er $V(X)$ asymptotisk $\chi^2$-fordelt med antal frihedsgrader givet ved

$$RS - 1 - \big(R-1\big) - \big(S-1\big) = \big(R-1\big)\big(S-1\big).$$

Antag vi har observeret $X = x$. Så kan $p$-værdien for accept af nulhypotesen approksimeres ved

$$p \approx \mathbb{P}\big(\chi^2_{\mathrm{df}=(R-1)(S-1)} \geq V(x)\big).$$

Denne approksimation er dog kun god når $M$ er stor, så hvad kan man gøre når dette ikke er tilfældet? Vi bemærker at den betingede fordeling af $X$ givet række- og søjle summerne $X_{r\cdot}$ og $X_{\cdot s}$ ikke afhænger af sandsynlighedsparameteren $\rho$ under nul-hypotesen. Det ligger derfor ligefor at foretage kvotienttestet i denne betingede fordeling. Dette test, kaldet det *eksakte test*, har altså $p$-værdi givet ved

$$\begin{aligned} p &= \mathbb{P}\big(V(X) \geq V(x) \mid X_{r\cdot} = x_{r\cdot}, X_{\cdot s} = x_{\cdot s}\big) \\ &= \mathbb{E}\big(1_{V(X)\geq V(x)} \mid X_{r\cdot} = x_{r\cdot}, X_{\cdot s} = x_{\cdot s}\big). \end{aligned}$$

Denne $p$-værdi kan altså simuleres via MCMC. I afsnit 4.4 ses MatLab-programkode, som gør dette. Den første lille stump af programmet generer en tovejstabel. Derefter findes $p$-værdien for det eksakte test via MCMC. Endelig udskrives et histogram og et plot af udfaldsstien. Vi har sammenlignet det asymptotiske test med det eksakte test for følgende to tovejstabeller.

| | | | |
|---|---|---|---|
| 2 | 1 | 0 | 0 |
| 0 | 3 | 2 | 3 |
| 0 | 0 | 4 | 1 |

| | | | |
|---|---|---|---|
| 8 | 5 | 5 | 4 |
| 4 | 10 | 4 | 4 |
| 1 | 23 | 4 | 4 |

Resultaterne er givet ved

| Tabel | Kvotientteststørrele | Asymptotisk test | “Eksakt” test |
|---|---|---|---|
| 1 | 16.1301 | 0.0130 | 0.0232 |
| 2 | 7.5230 | 0.2751 | 0.3236 |

MCMC $p$-værdierne er baseret på $N = 100000$ trin i Metropolis-Hastings Markov kæden. Bemærk, at dette antal formodenlig ikke er tilstrækkeligt til at retfærdiggøre at $p$-værdierne er angivet med 4 decimaler. Histogrammer og udfaldsstier for simulationerne ses i figurerne i afsnit 4.5 nedenfor. Bemærk, at man direkte kan se at tilstandsrummet er endeligt for Markov kæden svarende til Tabel 1.

## 4.4 Implementation i MatLab

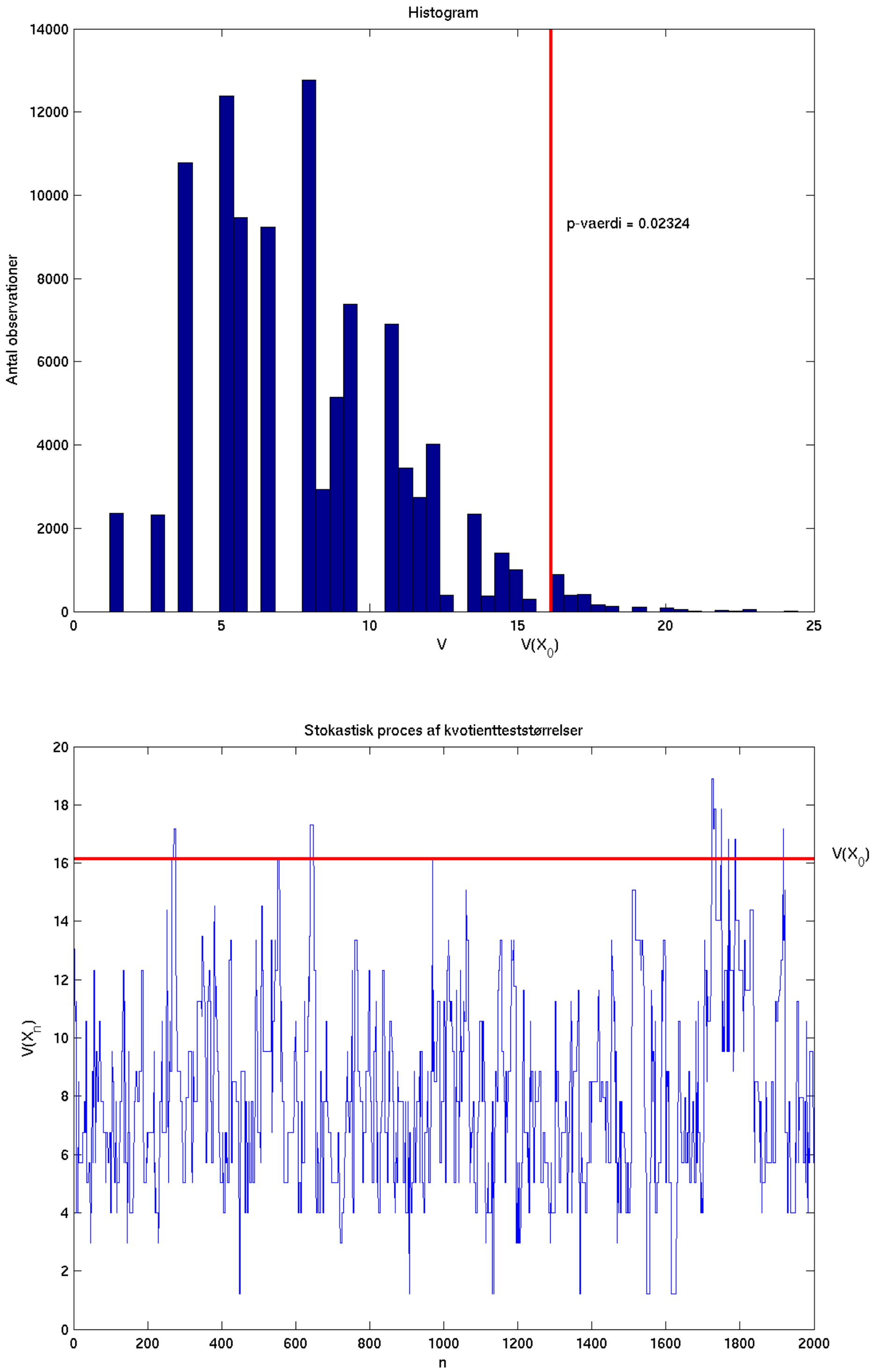
```
% Dette MatLab program gennemfører det "eksakte" uafhængighedstest for en
% given tovejstabel. Bemærk, at koden ikke er designet for at opnå maksimal
% beregningshastighed.

% antal iterationer
N = 100000;
K = 100;

% initial similering af tovejstabel:
% først simuleres en tovejstabel med uafhængige rækker og søjler
R = 10; S = 16; M = 100;
X0 = zeros(R,S);
for m = 1:M
    r = ceil(rand*R); s = ceil(rand*S);
    X0(r,s)=X0(r,s)+1;
end
% dertil lægges en tovejstabel, der evt. ødelægger uafhængigheden
X0 = X0 %+ reshape([2,0,0,0,2,0,0,0,2,0,2,0],R,S)

psim = zeros(1,K);
for k=1:K
    % initialisering af tovejstabellen
    X = X0;

    V = zeros(1,N);

    % beregning af (den vigtige del af) kvotientteststørrelsen
    V(1) = sum(sum(X.*log(X+(X==0))));

    for n = 1:N

        % vælger tilfældige række og søjle par
        r1=ceil(rand*R); r2=ceil(rand*R);
        while (r1==r2) r2=ceil(rand*R); end

        s1=ceil(rand*S); s2=ceil(rand*S);
        while (s1==s2) s2=ceil(rand*S); end

        %beregning af acceptsandsynlighederne
        alpha = min(1,X(r1,s2)*X(r2,s1)/((X(r1,s1)+1)*(X(r2,s2)+1)));

        % accepter springet med sandsynlighed alpha, ellers bliv stående i
```

```
        % nuværende tilstand (dvs. gør ikke noget)
        if alpha > rand
            X(r1,s1) = X(r1,s1)+1; X(r1,s2) = X(r1,s2)-1;
            X(r2,s1) = X(r2,s1)-1; X(r2,s2) = X(r2,s2)+1;
        end

        % beregning af kvotientteststørrelsen
        V(n+1) = sum(sum(X.*log(X+(X==0))));

    end

    % den endelige beregning af kvotientteststørrelsen incl. led fra
    % marginalerne
    V = 2*(V-sum(sum(X,1).*log(sum(X,1))) ...
            -sum(sum(X,2).*log(sum(X,2))) ...
            +sum(sum(X))*log(sum(sum(X))));

    % beregning af p-værdien for uafhængighedstestet
    pvalue = sum(V(2:N+1) >= V(1))/N;
    fprintf('Simulation nr. %d ud af %d : \n',k,K)
    fprintf('Kvotientteststørrelsen: %f \n',V(1))
    fprintf('p-værdi for kvotienttest af uafhængighedshypotesen: %f \n',pvalue)

    psim(k) = pvalue;
end

% Dette afslutter den egentlige MCMC del. Resten af koden laver nogle pæne
% plots.

% histogram for kvotientteststørrelse med 50 søjler og indtegning af den
% observerede teststørrelse
figure(1); clf
hist(V,50)
hold on; plot([V(1),V(1)],ylim,'-r','LineWidth',2); hold off
xlabel('V')
ylabel('Antal observationer')
title('Histogram')
f = ylim; g = xlim;
text(V(1)-1,-f(2)/15,'V(X_0)')
text(g(1)+(g(2)-g(1))*2/3,f(1)+(f(2)-f(1))*2/3,['p-vaerdi = ',num2str(pvalue)])
clear f; clear g

% plot af de 2000 først simulerede værdier af kvotientteststørrelsen
```

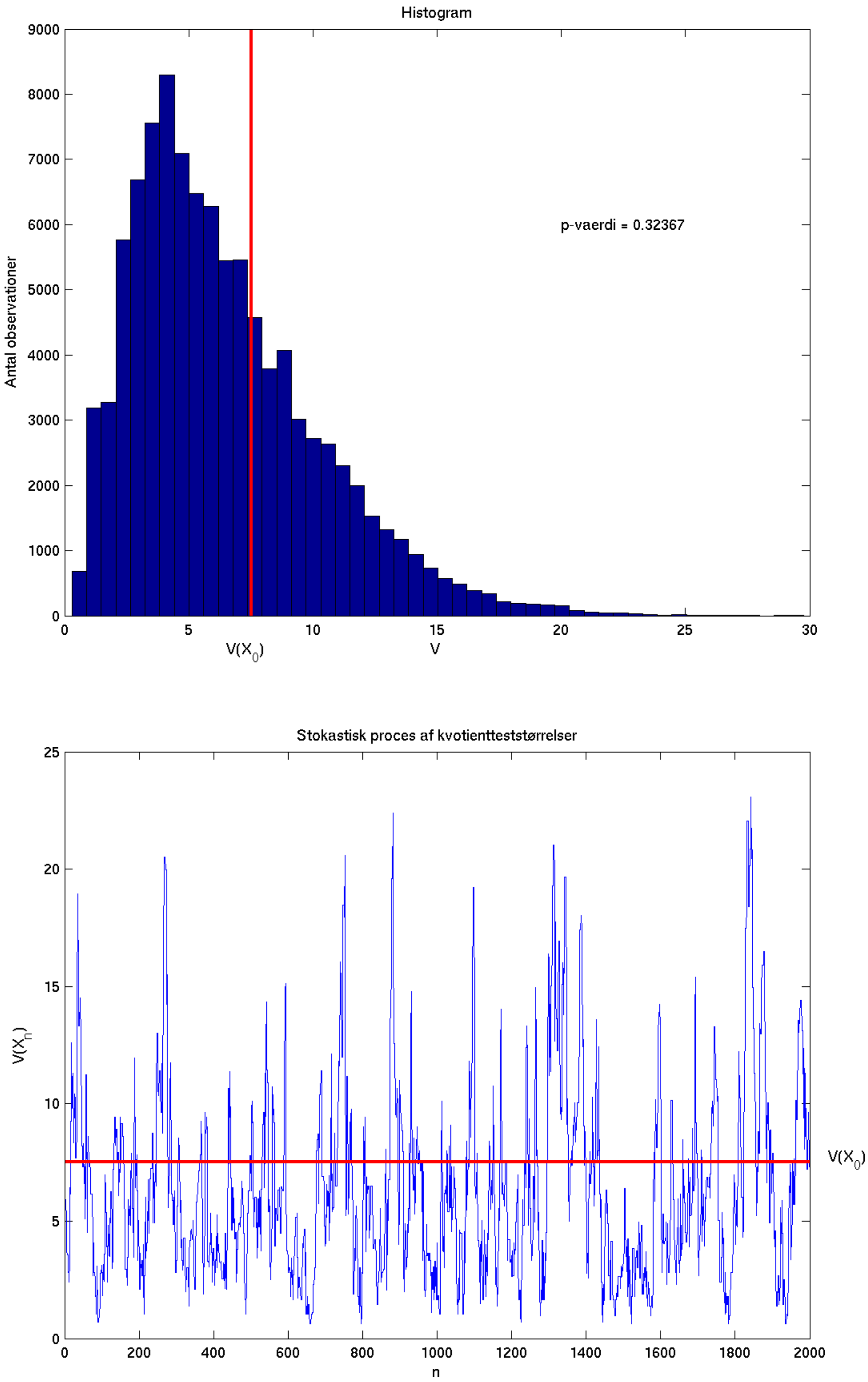
```
figure(2); clf
plot(0:2000,V(1:2001))
hold on; plot([0,2000],[V(1),V(1)],'-r','LineWidth',2); hold off
xlabel('n')
ylabel('V(X_n)')
title('Stokastisk proces af kvotientteststørrelser')
text(2050,V(1),'V(X_0)')

% histogram for de simulerede p-værdier
figure(3); clf
hist(psim)
xlabel('Simuleret p-værdi')
ylabel('Antal observationer')

% proces af estimaterede p-vaerdier
figure(4)
plot(cumsum(V(2:length(V)) >= V(1))./(1:length(V)-1))
title('Stokastisk proces af p-vaerdi estimater')
xlabel('n')
ylabel('p-value(n)')
```

## 4.5 Plots vedrørende uafhængigstestet

Ses på de næste sider.




Figur 9: Fordeling af kvotientteststørrelsen for tabel nr. 1




Figur 10: Fordeling af kvotientteststørrelsen for tabel nr. 2