TEKST NR 283 1994

Grænser for tilfældighed

(en kaotisk talgenerator)

Erwin Dan Nielsen
Niels Bo Johansen

TEKSTER fra
IMFUFA ROSKILDE UNIVERSITETSCENTER
INSTITUT FOR STUDIET AF MATEMATIK OG FYSIK SAMT DERES FUNKTIONER I UNDERVISNING, FORSKNING OG ANVENDELSER

IMFUFA, Roskilde Universitetscenter, Postbox 260, DK-4000 Roskilde

Grænser for tilfældighed (En kaotisk talgenerator)

Af: Erwin Dan Nielsen og Niels Bo Johansen
Vejleder: Peter Harremoës
IMFUFA tekst nr. 283/1994 53 sider ISSN 0106-6242

Abstract

Vi vil i denne fremstilling fremkomme med den ide, at kaotiske ligninger kan anvendes i forbindelse med generering af pseudotilfældige tal.

Der konstrueres en talgenerator, som er baseret på en kaotisk ligning, og talgeneratoren bliver udsat for en statistisk analyse.

Indhold

Indledning **3**

1 To kriterier for uforudsigelighed **5**

1.1 Ligefordeling 5

1.2 Stokastisk uafhængighed 7

2 Talgeneratorer **9**

2.1 Den klassiske talgenerator 9

2.2 Den kaotiske talgenerator 10

3 Statistisk analyse **17**

3.1 De anvendte statistiske tests 17

3.1.1 Korrelations-testen 20

3.1.2 Permutations-testen 22

3.1.3 χ^2-testen 23

3.1.4 Kolmogorov-Smirnov-testen 24

3.1.5 Seriel-testen 26

3.1.6 Afstand-testen 27

3.1.7 Partitions-testen 30

3.2 Valg af test-metode 33

3.3 Valg af parametre 36
3.3.1 Korrelations-testen 37
3.3.2 Permutations-testen 37
3.3.3 χ^2-testen 38
3.3.4 Kolmogorov-Smirnov-testen 38
3.3.5 Seriel-testen 38
3.3.6 Afstand-testen 39
3.3.7 Partitions-testen 39
3.4 Analyse af testresultaterne 40
3.4.1 Hypotese om stokastisk uafhængighed 40
3.4.2 Hypotese om ligefordeling 41
3.4.3 Analyse af testsandsynlighederne 42

4 God talgenerator = stor K-entropi + simpel transformation **43**

Konklusion **47**

A Beregningspræcision **49**

Litteraturliste **51**

Indledning

Pseudotilfældige tal bliver ofte anvendt i samkvem med computere. De nedenstående eksempler skal illustrere en række computer-anvendelser, hvori pseudotilfældige tal indgår:

Simulering: Til simulering af "uforudsigelige" eller "naturtro" hændelser. F.eks militære "kamp-situationer"[1] eller til statistiske tests.

Kryptologi: Både civil kommunikation ("dankort") og militær kommunikation.

Databaser: Placering (og søgning) af data i databaser er ofte baseret på talgeneratorer ("hashing").

Underholdning: Denne kategori dækker fra kommercielle "computerspil" til købmændenes "Lyn-Lotto".

Programmering: I forbindelse med udvikling af programmel — pseudotilfældige tal kan indgå som test-data.

Det er således klart, at talgeneratorer ofte bliver implementeret i computer-programmel.

[1] Der er konstrueret programmeringssprog til udvikling af "simulationsprogrammel". "SIMULA" er et eksempel herpå, og i standard klassen "SIMULATION" optræder 10 forskellige procedurer til generering af pseudotilfældige tal.

Talgeneratorer, som følger med til diverse programmer, er ofte af dårlig kvalitet. Disse talgeneratorer er som regel baseret på modulo-operatoren — i den primitive form kan talgeneratoren se således ud:

$$x_{i+1} = Ax_i \bmod m.$$

Vi kan højest få m forskellige værdier, og talgeneratoren er periodisk.

At den modulo-baserede talgenerator er periodisk giver anledning til ideen til dette projekt: *kaotiske ligninger* er ikke *periodiske* — måske kunne vi ud fra en kaotisk ligning konstruere en talgenerator[2], som er bedre end de *klassiske*.

Vor talgenerator er baseret på den simple iterative ligning (hvor A er valgt således, at periodicitet ikke forekommer)

$$x_{i+1} = Ax_i(1 - x_i).$$

Denne ligning er valgt, da det er den mest simple og kendteste kaotiske ligning.

For at vurdere, hvorvidt vor kaotiske talgenerator er *god eller dårlig*, vil vi gennem en række statistiske tests sammenligne denne med "en god klassisk talgenerator"[3].

Projekt-rapporten er opbygget på følgende måde: i kapitel 1 angives to statistiske kriterier for en *god* talgenerator. Med disse to kriterier i hånden forsøger vi i kapitel 2 at konstruere den kaotiske talgenerator. I dette kapitel præsenteres også en udvalgt klassisk talgenerator, som vor kaotiske skal sammenlignes med. Denne sammenligning finder sted i kapitel 3. I dette kapitel er en nøje beskrivelse af, hvorledes testene er udført, og hvilke resultater, der er opnået. I kapitel 4 ser vi på en anden side af problematikken vedrørende talgeneratorer — hvorledes en kaotisk talgenerator skal konstrueres, således, at denne bliver *hurtig*.

[2]Denne talgenerator kaldes fremover *den kaotiske talgenerator*. Hvis denne talgenerator anvendes i computer-programmel, så er de genererede sekvenser selvfølgelig periodiske, da en computer kun har et *endeligt* antal tilstande.

[3]Netop den anvendte talgenerator bliver anbefalet af Donald Ervin Knuth, side 27 vol. 2.

Kapitel 1

To kriterier for uforudsigelighed

Formålet med dette projekt er at konstruere en kaotisk talgenerator, der kan danne sekvenser, der kan betragtes som værende uforudsigelige. Derfor vil vi opstille to statistiske kriterier for uforudsigelighed, som vi vil referere til i resten af denne projekt-rapport.

På det intuitive plan har vi alle en idé om, at uforudsigelige sekvenser må eksistere — vi har nærmest en uformel definition: elementerne i sekvensen fremkommer uafhængigt af hinanden, og de er ligefordelte[1].

Ud fra dette intuitive plan skal vi i afsnit 1.1 behandle sekvenser, hvori elementerne er ligefordelte; i afsnit 1.2 skal vi behandle emnet stokastisk uafhængighed.

1.1 Ligefordeling

Lad os til en begyndelse betragte en sekvens af reelle tal

$$y_1, y_2, \ldots, y_n. \tag{1.1}$$

[1]Det er klart, at vi kunne være interesseret i andre fordelinger. Disse fordelinger kan imidlertid afledes af ligefordelingen, hvis den inverse til den pågældende fordelingsfunktion kendes.

Sekvensen kan betragtes som værende en observation af en n-dimensional stokastisk vektor

$$Y = (Y_1, Y_2, \ldots, Y_n) \tag{1.2}$$

(eller: elementerne i sekvensen kan betragtes som observationer af en stokastisk proces $\{Y_i, i \in \{1, 2, \ldots, n\}\}$). Vi er interesseret i, at de n stokastiske variable i 1.2 er *identisk fordelte* stokastiske variable, som er *ligefordelte* på intervallet $[0, 1]$.

Til hver af de n stokastiske variable i 1.2 har vi en tæthedsfunktion, som kan skrives således:

$$f(y) = \begin{cases} 1 & \text{hvis } y \in [0,1] \\ 0 & \text{ellers.} \end{cases}$$

Vi vil nu vise et eksempel[2] på, hvorledes vi kan afgøre, hvorvidt de stokastiske variable i 1.2 er ligefordelte.

Fordelingsfunktionen $F(y)$ vil se således ud:

$$F(y) = \begin{cases} 0 & \text{hvis } y \in]-\infty, 0[\\ y & \text{hvis } y \in [0,1] \\ 1 & \text{ellers.} \end{cases}$$

Denne funktion kan vi sammenligne med den *empiriske* fordelingsfunktion (eller den *kumulerede frekvensfunktion*) $F_n(y)$, som defineres på følgende måde:

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} I_i(y),$$

hvor

$$I_i(y) = \begin{cases} 1 & \text{hvis } y_i \leq y \\ 0 & \text{ellers.} \end{cases}$$

Det er klart, at $F_n(y)$ er en diskontinuert funktion, da sekvensen 1.1 består af et endeligt antal elementer.

[2]Eksemplet er *Kolmogorov-Smirnov-testen*, som er nærmere beskrevet i afsnit 3.1.4.

Hvis de stokastiske variable i 1.2 er ligefordelte (og uafhængige), så vil der være følgende relation mellem $F(y)$ og $F_n(y)$:

$$\lim_{n\to\infty} |F(y) - F_n(y)| = 0, \ \forall y \in \mathbf{R}. \tag{1.3}$$

Nu omhandler 1.3 imidlertid en grænseværdi for $n \to \infty$. I den statistiske analyse har vi givet en konkret værdi af n, så den statistiske analyse består i at afgøre, hvorvidt størrelsen

$$|F(y) - F_n(y)|, \ \forall y \in \mathbf{R} \tag{1.4}$$

afviger *signifikant* fra 0.

I kapitel 3 har vi en række andre eksempler på statistiske tests, som kan afgøre, hvorvidt 1.2 består af (uafhængige) identisk fordelte ligefordelte stokastiske variable.

1.2 Stokastisk uafhængighed

Det er imidlertid ikke nok, at elementerne i en given sekvens er ligefordelte. Elementerne skal også fremkomme *stokastisk uafhængigt* af hinanden.

I praksis bliver vi nødt til at tage til takke med et andet kriterium end *stokastisk uafhængighed*, nemlig med kriteriet om *ringe korrelation*[3].

Eksempelvis[4] kan korrelationen vurderes ud fra den såkaldte *korrelationskoefficient*, som vi skal definere herunder:

Hvis sekvenserne

$$y_1, y_2, \ldots, y_n$$

og

$$y_2, \ldots, y_n, y_{n+1}$$

[3]Vi vil dog alligevel anvende frasen *stokastisk uafhængighed*, hvor der egentligt menes *ringe korrelation*.

[4]Eksemplet er *Korrelations-testen*. Denne er beskrevet nærmere i afsnit 3.1.1.

bliver opfattet som værende observationer af de stokastiske vektorer $X = (Y_1, Y_2, \ldots, Y_n)$ og $Y = (Y_2, Y_3, \ldots, Y_{n+1})$, kan vi vurdere korrelationen mellem Y_k og Y_{k+1}. Korrelationskoefficienten $\mathrm{corr}(X, Y)$ for de stokastiske vektorer X og Y er givet ved følgende:

$$\mathrm{corr}(X, Y) = \frac{\mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}}.$$

Størrelsen $\mathrm{corr}(X, Y)$ antager værdier i intervallet $[-1, 1]$. Vi har særtilfældene:

- Hvis der er en (lineær) korrelation mellem X og Y (dvs, at $Y = \alpha X + \beta$), så er $\mathrm{corr}(X, Y)$ enten 1 eller -1.
- Hvis de stokastiske vektorer X og Y er stokastisk uafhængige, så er $\mathrm{corr}(X, Y) = 0$.

Som vi kan se, er der *ikke* ensbetydende mellem stokastisk uafhængighed og $\mathrm{corr}(X, Y) = 0$ — vi kan konstruere eksempler hvor $\mathrm{corr}(X, Y) = 0$ og hvor $Y = g(X)$ (i disse tilfælde er $g(x)$ en ikke-lineær funktion). Imidlertid vil korrelationskoefficienten give et fingerpeg — specielt hvis den er *signifikant* tæt på ± 1. I dette tilfælde vil der være en udpræget korrelation mellem de stokastiske variable.

Vi skal i kapitel 3 give flere eksempler på statistiske tests, som kan afsløre forskellige former for korrelation.

Kapitel 2

Talgeneratorer

Vi vil i dette kapitel beskrive *de udvalgte talgeneratorer* — en klassisk og vor kaotiske talgenerator. Den klassiske talgenerator beskrives kun kort, mens der er en detaljeret beskrivelse af konstruktionen af den kaotiske talgenerator.

Vi skal anvende den klassiske talgenerator til at vurdere vor kaotiske talgenerator — den kaotiske talgenerator og den udvalgte klassiske talgenerator udsættes for de samme tests. Den udvalgte klassiske talgenerator kan således betragtes som værende en *repræsentant*.

2.1 Den klassiske talgenerator

De klassiske talgeneratorer er en fællesbetegnelse for en række talgeneratorer, som ofte hører med til diverse program-pakker og andet programmel. De klassiske talgeneratorer er hurtige, da de er baseret på simple iterative ligninger — ofte indgår *modulo-operatoren*, som kan implementeres særdeles effektivt i computere.

Vi har valgt en klassisk talgenerator, som Knuth anbefaler til praktiske anvendelser[1]:

[1] Se Knuth side 27 vol. 2.

$$\begin{aligned} x_{i+1} &= (x_{i-55} + x_{i-24}) \bmod m, \\ y_{i+1} &= \frac{x_{i+1}}{m-1}. \end{aligned}$$

Elementerne x_i tilhører mængden $\{0, 1, \ldots, m-1\}$, og elementerne $y_i \in [0,1]$ udgør den genererede sekvens. Vi har, at y_i er rationale tal.

Vi har valgt $m = 10^9$ — ifølge Knuth er perioden af længden $2^c(2^{55}-1)$, hvor $0 \leq c < \log_2(m)$. Der er således tale om en meget lang periode. Der er i alt m mulige værdier for y_i.

2.2 Den kaotiske talgenerator

Den kaotiske talgenerator vil vi basere på følgende iterative ligning (eller transformation):

$$x_{i+1} = Tx_i = 4x_i(1 - x_i), \tag{2.1}$$

hvor $x_i \in [0,1]$. Denne iterative ligning er et skoleeksempel — eller i hvert fald eksemplet, som *kaos-tilhængere* altid benytter — på en *kaotisk ligning*[2].

Lad os til en begyndelse betragte sekvensen genereret af transformationen T:

$$x_0, x_1, \ldots .$$

Vi kan opfatte sekvensen som værende observationer af en stokastisk proces: $\{X_i, i \in \mathbf{N_0}\}$.

[2]Der er tale om den logistiske vækst ligning $x_{i+1} = Ax_i(1 - x_i)$. Der optræder kaotisk opførsel for flere værdier af A (se Schuster side 33). Vi har valgt værdien 4 for A, da dette kan udnyttes i implementeringen af talgeneratoren — multiplikation med 4 i det binære talsystem er specielt simpelt og hurtigt — samt, at K-entropien netop er højest for $A = 4$ (se senere). Vi bemærker, at der ikke optræder kaotisk opførsel for alle værdier af $x_0 \in [0,1]$ — f.eks $x_0 = 0$ og $x_0 = 1$, som jo er hhv begyndelsespunkt og endepunkt i det betragtede interval. Imidlertid er sandsynligheden for at vælge værdier af x_0, som ikke fører til kaotisk opførsel, lig nul.

Hvis de stokastiske variable $X_0, X_1, \ldots$ har samme fordelingsfunktion $F(x)$, har vi at gøre med en stationær proces (eller med andre ord: transformationen T er målbevarende: $P(T^{-1}A) = P(A)$). Vi kan danne en sekvens $y_0, y_1, \ldots$, hvori elementerne er ligefordelte. Idet vi benytter $F(x)$, kan vi danne den ønskede sekvens $y_0, y_1, \ldots$ på følgende måde[3]:

$$\begin{aligned} x_{i+1} &= 4x_i(1-x_i), \\ y_{i+1} &= F(x_{i+1}). \end{aligned}$$

Det er nu et spørgsmål om at finde $F(x)$ — dette kan selvfølgelig være et problem, men i dette tilfælde er $F(x)$ elementær[4]; $F(x)$ er arcussinus-fordelingsfunktionen:

$$F(x) = \frac{2}{\pi} \arcsin \sqrt{x}.$$

Vi kan kontrollere, at $F(x)$ faktisk er den rigtige fordelingsfunktion.

Transformationen T skal være målbevarende: $f(Tx) = f_T(x)$, og pga symmetrien $Tx = T(1-x)$ har vi:

$$f_T(x) = \frac{f(x)}{|T'(x)|} + \frac{f(1-x)}{|T'(1-x)|} = \frac{2f(x)}{|T'(x)|}.$$

De nedenstående udregninger viser, at hvis

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \tag{2.2}$$

[3] $g(y) = 1$ er tæthedsfunktionen for ligefordelingen i $[0,1]$. Det ses umiddelbart, at hvis $t(x)$ sættes lig $F(x)$ så gælder:

$$g(y) = \frac{f(x)}{|t'(x)|} = \frac{f(x)}{F'(x)} = 1.$$

[4] Schuster (side 58) nævner den pågældende tæthedsfunktion: $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$. Vi har, at denne er tæthedsfunktionen for arcussinus-fordelingen, som er et specielt tilfælde af betafordelingen: $f(x) = \mathrm{Be}(x, \frac{1}{2}, \frac{1}{2})$.

og $Tx = 4x(1-x)$, så gælder $f(Tx) = f_T(x)$[5]:

$$\begin{aligned} f(Tx) &= f_T(x) \Leftrightarrow \\ f(Tx) &= \frac{2f(x)}{|T'(x)|} \Leftrightarrow \\ \frac{1}{\pi\sqrt{4x(1-x)(1-(4x(1-x)))}} &= \frac{2}{\pi\sqrt{x(1-x)}}\frac{1}{|4-8x|} \Leftrightarrow \\ &\vdots \\ \frac{1}{\sqrt{1-4x+4x^2}} &= \frac{1}{|1-2x|}. \end{aligned}$$

Vi får da den kendte fordelingsfunktion

$$F(x) = \int_0^x \frac{1}{\pi\sqrt{t(1-t)}}dt = \frac{2}{\pi}\arcsin\sqrt{x}.$$

Talgeneratoren kan derfor se således ud:

$$\begin{aligned} x_{i+1} &= 4x_i(1-x_i), \\ y_{i+1} &= \frac{2}{\pi}\arcsin\sqrt{x_{i+1}}. \end{aligned}$$

Imidlertid viser det sig efter et par små-eksperimenter[6], at denne talgenerator danner sekvenser, hvori elementerne kan betragtes som værende ligefordelte — men til gengæld kan elementerne *ikke* betragtes som værende stokastisk uafhængige. Dette er også umiddelbart klart: hvis vi

[5] Vi viser kun, at pågældende tæthedsfunktion er en mulighed. Vi viser ikke, at denne er den *eneste* mulighed.

[6] En mere grundig analyse vil dog også vise, at talgeneratoren *ikke* genererer ligefordelte tal. Imidlertid bliver en hypotese om stokastisk uafhængighed prompte forkastet.

blot kender elementet y_0, kan vi uden videre levere samtlige[7] elementer i sekvensen $y_1, y_2, \ldots$. Vi skal i det følgende vise en metode til at "tilsløre" denne åbenbare determinisme.

Betragt tæthedsfunktionen 2.2 for arcussinus-fordelingen; denne er symmetrisk omkring $x = \frac{1}{2}$.

Vi ser, at

$$\int_0^{\frac{1}{2}} \frac{1}{\pi\sqrt{t(1-t)}} dt = \int_{\frac{1}{2}}^1 \frac{1}{\pi\sqrt{t(1-t)}} dt = \tfrac{1}{2}.$$

Opfatter vi sekvensen $x_0, x_1, \ldots$ genereret af transformationen T som værende observationer af den stationære stokastiske proces $\{X_i, i \in \mathbf{N_0}\}$, har vi sandsynlighederne

$$P(X_i \in [0, \tfrac{1}{2}]) = P(X_i \in]\tfrac{1}{2}, 1]) = \tfrac{1}{2}. \tag{2.3}$$

Vi kan transformere processen $\{X_i, i \in \mathbf{N_0}\}$ over i processen $\{Z_i, i \in \mathbf{N_0}\}$ ved følgende transformation:

$$\{Z_i, i \in \mathbf{N_0}\} = \begin{cases} 0 & \text{hvis } X_i \in [0, \frac{1}{2}] \\ 1 & \text{hvis } X_i \in]\frac{1}{2}, 1]. \end{cases} \tag{2.4}$$

Denne proces er en Bernoulli-proces[8] — dette kan erkendes af det følgende: at processen $\{Z_i, i \in \mathbf{N_0}\}$ er en Bernoulli-proces betyder, at $Z_{i+1}, Z_{i+2}, \ldots$ er stokastisk uafhængige af Z_i.

Da $Z_{i+1}, Z_{i+2}, \ldots$ er determineret af X_{i+1}, skal vi vise, at Z_i er uafhængig af X_{i+1}. Dette er det samme som at vise, at fordelingen af X_{i+1} er uafhængig af om $x_i \in [0, \frac{1}{2}]$ eller $x_i \in]\frac{1}{2}, 1]$.

---

[7] Vi har, at

$$y_{i+1} = F \circ T \circ F^{-1}(y_i) = \begin{cases} 2y_i & \text{hvis } y_i \in [0, \frac{1}{2}] \\ 2(1-y_i) & \text{hvis } y_i \in]\frac{1}{2}, 1]. \end{cases}$$

[8] En Bernoulli-proces $\{Z_i, i \in \mathbf{N_0}\}$ er stationær, og den kan befinde sig i h tilstande $0, 1, \ldots, h-1$. Sandsynligheden for at processen befinder sig i tilstand k til "tiden" i, er givet ved $P(Z_i = k) = p_k, k = 0, 1, \ldots, h-1$. Hvis Bernoulli-processen er symmetrisk, så gælder $p_k = \frac{1}{h}$. For Bernoulli-processer gælder, at Z_i er *stokastisk uafhængig* af Z_k, hvor $k = 0, 1, \ldots, i-1, i+1, \ldots$.

Betragt endnu en gang tæthedsfunktionen 2.2. Vi ser, at transformationen S, givet ved $S\colon x \mapsto 1-x$, bijektivt transformerer tæthedsfunktionens restriktion på $[0, \frac{1}{2}[$ til tæthedsfunktionens restriktion på $]\frac{1}{2}, 1]$, dvs, at

$$f_{S\big|_{[0,\frac{1}{2}[}} = f_{\big|_{]\frac{1}{2},1]}}. \tag{2.5}$$

Transformationen T transformerer 2.5 til

$$f_{T\circ S\big|_{[0,\frac{1}{2}[}} = f_{T\big|_{]\frac{1}{2},1]}}.$$

Da $T = T \circ S$ har vi

$$f_{T\big|_{[0,\frac{1}{2}[}} = f_{T\big|_{]\frac{1}{2},1]}}, \tag{2.6}$$

hvilket netop viser, at fordelingen af X_{i+1} er uafhængig af om $x_i \in [0, \frac{1}{2}]$ eller $x_i \in]\frac{1}{2}, 1]$.

Vi har nu opnået, at processen $\{Z_i, i \in \mathbf{N_0}\}$ er en *symmetrisk* (pga 2.3) *Bernoulli-proces* (pga 2.6). Sekvensen

$$z_1, \ldots, z_m, z_{m+1}, \ldots, z_{2m}, \ldots, z_{(n-1)m+1}, \ldots, z_{nm}, \tag{2.7}$$

som er genereret af 2.4, består således af symmetrisk-fordelte elementer, og elementerne fremkommer stokastisk uafhængigt af hinanden. En forudsætning for dette er naturligvis, at vi har valgt en *tilfældig* værdi af x_0 — vi skal først sætte processen $\{X_i, i \in \mathbf{N_1}\}$ igang.

Vi kan nu danne den ønskede sekvens $y_0, y_1, \ldots, y_{n-1}$ af rationale tal ud fra 2.7 på følgende måde:

$$y_k = \sum_{j=1}^{m} 2^{-j} z_{km+j},$$

for $k = 0, 1, \ldots, n-1$.

Elementerne i sekvensen $y_0, y_1, \ldots, y_{n-1}$ er ligefordelte og stokastisk uafhængige i intervallet $[0, \sum_{j=1}^{m} 2^{-j}]$. Der er netop 2^m mulige værdier for y_k. Da $\sum_{j=1}^{m} 2^{-j} \to 1$ for $m \to \infty$, og antallet af mulige værdier er 2^m, bør m være "stor". Vi har til den kaotiske talgenerator valgt $m = 20$; vi får derved over 10^6 mulige værdier af y_k.

Transformationen 2.4 transformerer $\{X_i, i \in \mathbf{N_0}\}$ til en *symmetrisk* Bernoulli-proces $\{Z_i, i \in \mathbf{N_0}\}$. Vi kunne have valgt andre klasseinddelinger af $[0,1]$ end $[0, \frac{1}{2}]$ og $]\frac{1}{2}, 1]$, og stadig opnå en symmetrisk Bernoulli-proces. F.eks kunne vi vælge klasseinddelingen $[0, \frac{1}{4}] \cup]\frac{1}{2}, \frac{3}{4}]$ og $]\frac{1}{4}, \frac{1}{2}] \cup]\frac{3}{4}, 1]$. Men til gengæld skal parameteren A i

$$Tx_i = Ax_i(1 - x_i)$$

være lig 4, hvis der skal eksistere en transformation af $\{X_i, i \in \mathbf{N_0}\}$ til en symmetrisk Bernoulli-proces. I figur 2.1 ses K-entropien[9] K af $\{X_i, i \in \mathbf{N_0}\}$ afbildet som funktion af parameteren A (hvor $A \in [3.85; 4]$). K-entropien er størst for $A = 4$; her er den $\log_e(2)$.

Det er beviseligt, at vi kan transformere en stationær proces $\{S_i, i \in \mathbf{N_0}\}$ til en vilkårlig Bernoulli-proces $\{B_i, i \in \mathbf{N_0}\}$, hvis blot K-entropien af $\{S_i, i \in \mathbf{N_0}\}$ er større end entropien af Bernoulli-processen. I visse tilfælde eksisterer der også en Bernoulli-proces, hvis K-entropien af $\{S_i, i \in \mathbf{N_0}\}$ er lig entropien af Bernoulli-processen. Det sidste er

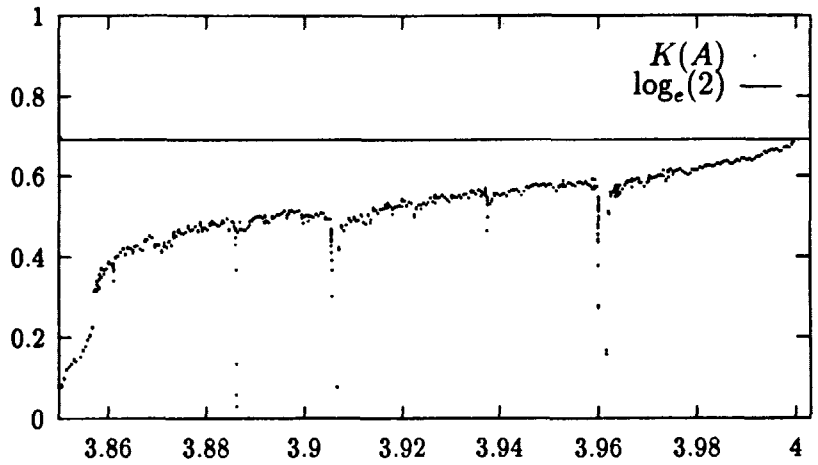
[9] K-entropien af $\{X_i, i \in \mathbf{N_0}\}$ defineres således: Hvis $A_0, A_1 \ldots, A_{h-1}$ er en klasseinddeling (her: af intervallet $[0,1]$), hvis $P(X_i \in A_k) < f, k = 0, 1, \ldots h-1$, og hvis $P(A_{i_0} \ldots A_{i_{n-1}})$ er sandsynligheden for, at $X_i \in A_{i_0}, X_{i+1} \in A_{i_1}, \ldots, X_{i+n-1} \in A_{i_{n-1}}$, så er K-entropien af processen $\{X_i, i \in \mathbf{N_0}\}$ givet ved

$$K(\{X_i, i \in \mathbf{N_0}\}) = -\lim_{f \to 0} \lim_{n \to \infty} \frac{1}{n} \sum_{i_0 \ldots i_n} P(A_{i_0 \ldots i_n}) \log_e(P(A_{i_0 \ldots i_n})).$$

Det er ikke hensigtsmæssigt at anvende denne definition, hvis vi skal udregne K-entropien af en stationær proces vha computer-kraft (stiller helt urimelige krav til *lagerkapacitet* og hastighed). I stedet kan vi med fordel anvende den til transformationen T hørende Liapunov-eksponent; størrelsen af denne er i det en-dimensionale tilfælde lig størrelsen af K-entropien. Liapunov-eksponenten λ er givet ved

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log_e(|T'(T^i x_0)|)$$

for næsten alle x_0.



Figur 2.1: K-entropi som funktion af parameteren A

tilfældet med vor proces $\{X_i, i \in \mathbf{N_0}\}$; entropien af den symmetriske Bernoulli-proces med *to* tilstande er netop

$$-\left(\tfrac{1}{2}\log_e(\tfrac{1}{2}) + \tfrac{1}{2}\log_e(\tfrac{1}{2})\right) = \log_e(2).$$

Det er derfor klart, at valget af A er begrænset til $A = 4$, da K-entropien af $\{X_i, i \in \mathbf{N_0}\}$ ellers bliver *mindre end* $\log_e(2)$, og derved kan der *ikke* eksistere en symmetrisk Bernoulli-proces med to (eller flere) tilstande.

Det kan synes underligt at foretage en række statistiske tests på talgeneratoren, da vi *ved*, at den danner stokastisk uafhængige og ligefordelte tal. Imidlertid skal den kaotiske talgenerator anvendes i praksis, dvs i forbindelse med computer-programmel. En computer kan kun regne med endelig præcision; der regnes med et antal betydende cifre. I appendiks A forsøger vi at overbevise læseren om betydningen af dette problem. Det fremgår af dette appendiks, at elementerne i sekvensen genereret af 2.1 skal repræsenteres med omkring 11 betydende cifre, hvis talgeneratoren skal fungere ordentligt. En konsekvens heraf, er, at vi i alle testene i afsnit 3.4 benytter 19 betydende cifres præcision (den maksimale præcision ved det anvendte programmeringsværktøj).

Kapitel 3

Statistisk analyse

Som beskrevet i kapitel 1 er vi interesseret i talgeneratorer, der producerer sekvenser, hvori elementerne er ligefordelte og fremkommer stokastisk uafhængigt af hinanden.

I afsnit 1.1 så vi *Kolmogorov-Smirnov-testen* for ligefordeling, og i afsnit 1.2 så vi *Korrelations-testen* for stokastisk uafhængighed. Det er ikke nok, blot at udføre disse to tests på de genererede sekvenser. Der skal anvendes en række forskellige tests, da de enkelte tests kontrollerer helt bestemte egenskaber ved de genererede sekvenser.

Afsnit 3.1 omhandler en række tests, der med rimelighed kan anvendes i forbindelse med alle talgeneratorer, som forventes at generere ligefordelte og stokastisk uafhængige sekvenser.

I afsnit 3.2 gives en beskrivelse af, hvorledes den samlede test af talgeneratorer finder sted (kombinationen af de tests, som er beskrevet i afsnit 3.1); i afsnit 3.3 foretages en række valg for hver enkelt test, og i afsnit 3.4 analyseres resultaterne af den samlede test.

3.1 De anvendte statistiske tests

Vi skal vurdere en talgenerators anvendelighed ved udførsel af en række statistiske tests. Testene udføres for at afgøre, om elementerne i den genererede sekvens kan betragtes som *ligefordelte* og fremkommet *stokastisk uafhængigt* af hinanden. I en test opstilles en af hypoteserne:

- Elementerne i sekvensen er ligefordelte i intervallet $[0, 1]$.
- Elementerne i sekvensen er stokastisk uafhængige af hinanden.
- Elementerne i sekvensen er ligefordelte i intervallet $[0, 1]$, og elementerne fremkommer stokastisk uafhængigt af hinanden.

Under hypotesen vil vi dernæst, i den pågældende test, undersøge om der optræder en bestemt systematik (i generel forstand) i den genererede sekvens. Hvis der ikke opdages systematik, kan vi ikke forkaste den pågældende hypotese.

I de tests vi udfører, betragter vi sekvensen

$$y_1, y_2, \ldots, y_n \tag{3.1}$$

af reelle tal eller heltals-sekvensen

$$x_1, x_2, \ldots, x_n, \tag{3.2}$$

hvor tallene i sekvensen 3.2 er frembragt vha en funktion, der først ganger et positivt heltal med et reelt tal, og dernæst nedrunder til nærmeste heltalsværdi. Funktionen skrives på følgende måde:

$$\text{int}(d \times y_i), \tag{3.3}$$

hvor d er et positivt heltal. At omforme sekvensen 3.1 til sekvensen 3.2 ved brug af funktionen 3.3 vil vi illustrere på følgende måde:

$$x_i = \text{int}(d \times y_i),$$

hvor $i = 1, 2, \ldots, n$. Derved kommer elementerne i 3.2 til at tilhøre mængden $\{0, 1, \ldots, d-1\}$.

De fleste af testene går ud på, at vi skal vælge en måde at danne *kategorier* på. Vi skal danne kategorierne ud fra elementerne i sekvenserne 3.1 eller 3.2 eller delsekvenser af disse. Eksempelvis kan vi i en test[1]

[1] χ^2-testen, som bliver beskrevet i afsnit 3.1.3.

danne kategorier ud fra elementerne i 3.2. Da elementerne kan antage værdierne $0, 1, \ldots, d-1$, har vi d forskellige kategorier. I en anden test[2] kan vi danne kategorier ud fra antallet af forskellige elementer i delsekvenser af 3.2. Har delsekvenserne længden t, har vi $\min(d, t)$ kategorier.

Når vi i en test har valgt en måde at kategorisere på, kan vi finde det *observerede* antal forekomster i hver kategori. I det første eksempel finder vi det observerede antal ved at tælle antallet af elementer i sekvensen 3.2, som kan antage værdierne $0, 1, \ldots, d-1$. I det andet finder vi det observerede antal ved at tælle antallet af delsekvenser, med længde n, af 3.2 med hhv $1, 2, \ldots, \min(d, t)$ forskellige elementer.

Under en hypotese i en given test kan vi, vha de størrelser der optræder, når vi danner kategorier, beregne det *forventede* antal forekomster i hver kategori. I det første eksempel beregner vi vha d og længden n af sekvensen 3.2, samt i det andet vha d, t og n, det forventede antal forekomster i hver kategori.

Når vi udfører en test, har vi altså et observeret antal forekomster og et forventet antal forekomster. Det, som vi nu ønsker at nå frem til, er et mål for, hvor meget det observerede antal forekomster afviger fra det forventede antal forekomster. For at kunne angive et sådant mål, har vi brug for en *teststørrelse*. Denne kan defineres på flere måder. Et krav til den er blot, at vi skal kende dens fordeling.

En ofte benyttet teststørrelse er en teststørrelse, der under visse forudsætninger vil være approksimativt χ^2-fordelt. Pointen med at benytte en sådan approksimation er, at vi får en nemmere måde at udregne en testsandsynlighed på, end hvis vi skulle bruge (finde) den eksakte fordeling. Vi anvender følgende teststørrelse[3]:

$$q(p_1, p_2, \ldots, p_m) = \sum_{i=1}^{k} \frac{(\text{obs}_i - \text{for}_i)^2}{\text{for}_i}, \tag{3.4}$$

hvor obs_i og for_i angiver hhv det observerede og det forventede antal forekomster i den i'te kategori. k er det samlede antal kategorier. $p_1, p_2, \ldots, p_m$ er parametre, der bestemmer: antallet af kategorier, det

[2] Partitions-testen, som bliver beskrevet i afsnit 3.1.7.

[3] Pearson's X^2.

samlede antal observationer, samt det forventede antal forekomster hørende til de enkelte kategorier. Teststørrelsen 3.4, er approksimativt χ^2-fordelt med $k-1$ frihedsgrader, når det forventede antal observationer i enhver af kategorierne er større end 5.

Testsandsynligheden ε er sandsynligheden for at få en værdi Q, der er større end den observerede værdi $q(p_1, p_2, \ldots, p_m)$ i χ^2-fordelingen med $k-1$ frihedsgrader. Denne findes således

$$\varepsilon = P(Q > q(p_1, p_2, \ldots, p_m)). \tag{3.5}$$

Vi har nu skitseret, hvorledes de fleste af testene er opbygget, dvs de tests, hvor teststørrelserne er approksimativt χ^2-fordelte.

To af testene afviger fra det ovenfor beskrevet, idet vi benytter teststørrelser, som følger andre fordelinger. Det vil dog fremgå af de følgende afsnit, hvordan teststørrelserne beregnes, og hvorledes testsandsynlighederne findes ud fra de respektive fordelinger eller approksimative fordelinger. I de følgende afsnit, afsnittene 3.1.1–3.1.7, vil vi give en nærmere beskrivelse af samtlige anvendte statistiske tests[4].

3.1.1 Korrelations-testen

Vi vil i denne test vurdere hypotesen om, at elementerne i sekvensen

$$y_0, y_1, \ldots, y_{n-1} \tag{3.6}$$

fremkommer stokastisk uafhængigt af hinanden.

Vi har allerede i afsnit 1.2 omtalt et kvantitativt mål for vekselvirkningen mellem elementerne i sekvensen 3.6, idet vi indførte korrelationskoefficienten $\mathrm{corr}(X, Y)$. Vi kiggede på vekselvirkningen mellem y_i og dens efterfølger y_{i+1}.

I denne test vil vi vurdere vekselvirkningen mellem elementerne y_i og $y_{(f+i) \bmod n}$, hvor $i = 0, 1, \ldots, n-1$ og hvor $f \in \{1, 2, \ldots, n-1\}$. f kaldes *forskydningsparameteren*.

[4] De anvendte tests er alle hentet fra Knuth, kapitel 3 vol. 2.

Den *empiriske korrelationskoefficient* $C(f,n)$ mellem elementerne y_i og $y_{(f+i) \bmod n}$ kan skrives på følgende måde:

$$\begin{aligned} C(f,n) &= \frac{\sum_{i=0}^{n}(y_i-\overline{y})\left(y_{(f+i) \bmod n}-\overline{y}\right)}{\sqrt{\sum_{i=0}^{n}(y_i-\overline{y})^2}\sqrt{\sum_{i=0}^{n}\left(y_{(f+i) \bmod n}-\overline{y}\right)^2}} \\ &= \frac{n\sum_{i=0}^{n} y_i \times y_{(f+i) \bmod n} - \left(\sum_{i=0}^{n} y_i\right)^2}{n\sum_{i=0}^{n} y_i^2 - \left(\sum_{i=0}^{n} y_i\right)^2}, \end{aligned}$$

hvor $\overline{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$.

Vi vil benytte

$$q(f,n) = C(f,n)\sqrt{\frac{n-2}{1-C(f,n)^2}}$$

som teststørrelse. Teststørrelsen $q(f,n)$ er approksimativt *t-fordelt* med $n-2$ frihedsgrader — når $n \to \infty$ er $q(f,n)$ med tilnærmelse (0,1)-normalfordelt.

Testsandsynligheden ε bliver udregnet på følgende måde:

$$\varepsilon = 2P(\mathcal{Q} > |q(f,n)|),$$

da t-fordelingen er symmetrisk omkring 0. Store positive værdier af $q(f,n)$ er lige så signifikante som tilsvarende store negative værdier.

3.1.2 Permutations-testen

Vi vil i denne test vurdere følgende hypotese:

Elementerne i sekvensen

$$y_0, y_1, \cdots, y_{n-1}, \tag{3.7}$$

hvor $n = st$, er fremkommet stokastisk uafhængigt af hinanden.

Vi har en inddeling af 3.7 i s delsekvenser hver med t elementer. Vi kan referere til elementerne i den j'te delsekvens

$$y_{jt}, y_{jt+1}, \cdots, y_{jt+t-1}, \tag{3.8}$$

hvor $0 \leq j < s$. Vi har at

$$P(Y_i = Y_j) = 0, \tag{3.9}$$

for alle $i \neq j$ og $i, j \leq s - 1$. Delsekvenserne kan inddeles i kategorier, idet vi udnytter, at der findes en og kun en permutation π: $M \mapsto M$ der opfylder

$$y_{\pi(jt \bmod j)} < y_{\pi((jt+1) \bmod j)} < \cdots < y_{\pi((jt+t-1) \bmod j)}.$$

Vi har antaget, at lighed mellem elementerne ikke forekommer. Denne antagelse skyldes 3.9. For hver af de $t!$ mulige permutationer har vi en kategori.

Vi lader nu Π betegne mængden af de $t!$ mulige permutationer. Vi kan nu foretage følgende observationer:

$$\text{obs}_\pi = (\text{Antal forekomster af permutationen } \pi),$$

hvor $\pi \in \Pi$. Hvis alle elementerne i 3.7 er fremkommet stokastisk uafhængigt af hinanden, må den forventede sandsynlighed, for, at en vilkårlig valgt delsekvens tilhører en bestemt kategori, være $\frac{1}{t!}$. Da der er s delsekvenser, må det forventede antal i hver kategori være

$$\text{for}_\pi = (\text{Forventede antal forekomster af permutationen } \pi) = \frac{s}{t!}, \quad (3.10)$$

hvor $\pi \in \Pi$. For at finde ud af, hvorvidt det forventede og det observerede antal stemmer overens, vil vi benytte følgende teststørrelse

$$q(s,t) = \sum_{\pi \in \Pi} \frac{(\text{obs}_\pi - \text{for}_\pi)^2}{\text{for}_\pi},$$

som med tilnærmelse vil være χ^2-fordelt, når $s > 5t!$ (mindst 5 elementer i hver kategori). Antallet af frihedsgrader vil være $t! - 1$.

Testsandsynligheden ε udregnes som

$$\varepsilon = P(\mathcal{Q} > q(s,t)).$$

3.1.3 χ^2-testen

I denne test vurderes hypotesen:

Elementerne i den reelle sekvens

$$y_1, y_2, \ldots, y_n \quad (3.11)$$

er ligefordelte i intervallet $[0,1]$.

Vi vil antage, at elementerne i sekvensen er stokastisk uafhængige.

I stedet for at betragte den reelle sekvens 3.11, vil vi reducere problemet til at betragte en heltallige sekvens af formen:

$$x_i = \text{int}(d \times y_i), \quad (3.12)$$

hvor $i = 1, 2, \ldots, n$.

Under hypotesen om ligefordeling, vil elementerne i den heltallige sekvens 3.12 være symmetrisk fordelte.

Elementerne x_i vil tilhøre mængden $\{0, 1, \ldots, d-1\}$ — da der således er d mulige værdier for x_i, har vi umiddelbart d kategorier. De observerede værdier kan opdeles på følgende vis:

$$\begin{aligned} \text{obs}_0 &= \text{(Antallet af gange, hvor } x_i = 0) \\ \text{obs}_1 &= \text{(Antallet af gange, hvor } x_i = 1) \\ &\vdots \\ \text{obs}_{d-1} &= \text{(Antallet af gange, hvor } x_i = d-1). \end{aligned}$$

hvor $i = 1, 2, \ldots, n$.

Da elementerne i sekvensen 3.12 forventes at være symmetrisk fordelte, vil vi forvente $\frac{n}{d}$ elementer i hver kategori, dvs, at $\text{for}_i = \frac{n}{d}$, hvor $i = 0, 1, \ldots, d-1$.

Som teststørrelse vil vi benytte

$$q(d, n) = \sum_{i=0}^{d-1} \frac{(\text{obs}_i - \text{for}_i)^2}{\text{for}_i},$$

der med tilnærmelse vil være χ^2-fordelt med $d-1$ frihedsgrader (under forudsætning, at der er mindst 5 elementer i hver kategori).

Testsandsynligheden ε udregnes som

$$\varepsilon = P(Q > q(d, n)).$$

3.1.4 Kolmogorov-Smirnov-testen

I denne test vil vi vurdere hypotesen:

Elementerne i sekvensen

$$y_1, y_2, \ldots, y_n \tag{3.13}$$

er ligefordelte i intervallet $[0,1]$. Elementerne i sekvensen 3.13 antages at være fremkommet stokastisk uafhængigt af hinanden.

Vi har allerede i afsnit 1.1 omtalt *Kolmogorov-Smirnov-testen*. Vi definerede her den *empiriske fordelingsfunktion*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x),$$

hvor

$$I_i(x) = \begin{cases} 1 & \text{hvis } y_i \leq x \\ 0 & \text{ellers,} \end{cases}$$

og vi ville vurdere, hvorvidt størrelsen

$$|F(x) - F_n(x)|, \ \forall x \in \mathbf{R} \tag{3.14}$$

afveg signifikant fra 0.

$F(x)$ er igen den i afsnit 1.1 viste fordelingsfunktion for ligefordelingen i intervallet $[0,1]$. I stedet for at vurdere størrelsen 3.14, vil vi vurdere, hvorvidt følgende størrelser:

$$\begin{aligned} K_n^+ &= \sqrt{n} \max_{1\leq j\leq n} \left(\frac{j}{n} - y_{(j)}\right), \\ K_n^- &= \sqrt{n} \max_{1\leq j\leq n} \left(y_{(j)} - \frac{j-1}{n}\right) \end{aligned}$$

afviger signifikant fra 0.

Sekvensen $y_{(i)}$, $i = 1,2,\ldots,n$ angiver den *ordnede* sekvens af 3.13 — der gælder således, at $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$.

K_n^+ og K_n^- er begge *Kolmogorov-Smirnov-fordelte*. Kolmogorov-Smirnov-fordelingen er kendt for enhver størrelse af n (se Knuth side 48, vol. 2), men når $n \to \infty$ får vi følgende simple fordelingsfunktion:

$$G(x) = 1 - e^{-2x^2}, \ \ x \geq 0. \tag{3.15}$$

Når n er *stor*, kan vi benytte fordelingsfunktionen 3.15, da den derved begået approksimationsfejl er lille. I dette tilfælde skriver vi teststørrelserne K_n^+ og K_n^- som hhv K_∞^+ og K_∞^-.

Testsandsynlighederne ε^+ og ε^- udregnes som hhv

$$\varepsilon^+ = P(Q > K_\infty^+)$$

og

$$\varepsilon^- = P(Q > K_\infty^-).$$

3.1.5 Seriel-testen

Vi vil i denne test vurdere følgende hypotese:

Betragt talparrene

$$(y_{2j}, y_{2j+1}), \tag{3.16}$$

hvor $j = 0, 1, \ldots, n-1$. Disse talpar, som er udplukket fra sekvensen

$$y_0, y_2, \ldots, y_{2n-1}, \tag{3.17}$$

er ligefordelte og fremkommer stokastisk uafhængigt af hinanden.

Grunden til, at vi ser på talparrene (y_{2j}, y_{2j+1}) og ikke, som vi kunne have forventet på talparrene (y_j, y_{j+1}), er, at de sidstnævnte talpar netop *ikke* er stokastisk uafhængige — når vi har udplukket talparret (y_j, y_{j+1}), så kendes allerede den første komponent i det efterfølgende talpar (y_{j+1}, y_{j+2}).

I stedet for at betragte den reelle sekvens 3.17 reducerer vi problemet til at betragte en heltallig sekvens

$$x_i = \text{int}(d \times y_i), \tag{3.18}$$

hvor $i = 0, 1, \ldots, 2n-1$.

Elementerne x_i vil derefter tilhøre mængden $\{0, 1, \ldots, d-1\}$.

Talparrene (x_{2j}, x_{2j+1}) fra 3.18 kan derefter inddeles i d^2 kategorier, idet vi i alt kan danne d^2 forskellige talpar (s,r), når

$$s,r \in \{0,1,\ldots,d-1\}.$$

Vi har således, at

$$\text{obs}_{(s,r)} = (\text{Antal forekomster af } (x_{2j}, x_{2j+1}) = (s,r)),$$

hvor $j = 0,1,\ldots,n-1$.

Hvis elementerne i sekvensen 3.17 er ligefordelte, og talparrene 3.16 fremkommer stokastisk uafhængigt af hinanden, vil vi forvente $\frac{n}{d^2}$ elementer i hver kategori, dvs, at $\text{for}_{(s,r)} = \frac{n}{d^2}$.

Som teststørrelse benyttes følgende:

$$q(d,n) = \sum_{s,r\in\{0,1,\ldots,d-1\}} \frac{(\text{obs}_{(s,r)} - \text{for}_{(s,r)})^2}{\text{for}_{(s,r)}}.$$

$q(d,n)$ vil med tilnærmelse være χ^2-fordelt, hvis $n > 5d^2$ (mindst 5 elementer i hver kategori), idet antallet af frihedsgrader er $d^2 - 1$.

3.1.6 Afstand-testen

Vi vil i denne test vurdere følgende hypotese:

Elementerne i sekvensen

$$y_1, y_2, \ldots, y_n \tag{3.19}$$

er ligefordelte i intervallet $[0,1]$, og elementerne er fremkommet stokastisk uafhængigt af hinanden.

Under hypotesen om ligefordeling, er sandsynligheden for, at Y_i ligger i delintervallet $[\alpha,\beta] \subset [0,1]$ følgende størrelse:

$$p = P(Y_i \in [\alpha, \beta]) = \beta - \alpha.$$

Og følgelig bliver sandsynligheden for, at Y_i *ikke* ligger i $[\alpha, \beta]$

$$P(Y_i \in [0,1] \setminus [\alpha, \beta]) = 1 - p.$$

Lad sekvensen

$$y_j, y_{j+1}, \ldots, y_{j+r-1}, y_{j+r}$$

være en delsekvens af 3.19.

Under hypotesen om ligefordeling og stokastisk uafhængighed, er sandsynligheden for, at $Y_j, Y_{j+1}, \ldots, Y_{j+r-1}$ tilhører intervallet $[0,1] \setminus [\alpha, \beta]$, mens Y_{j+r} tilhører $[\alpha, \beta]$, givet ved udtrykket:

$$P(Y_j, Y_{j+1}, \ldots, Y_{j+r-1} \in [0,1] \setminus [\alpha, \beta]), Y_{j+r} \in [\alpha, \beta]) = p(1-p)^r. \tag{3.20}$$

Vi vil nu opdele sekvensen 3.19 i en række delsekvenser:

$$\begin{gathered} y_1, y_2, \ldots, y_{r-1}, y_r \\ y_{r+1}, y_{r+2}, \ldots, y_{r+s-1}, y_{r+s} \\ \vdots \\ y_{r+s+\cdots+1}, y_{r+s+\cdots+2}, \ldots, y_{r+s+\cdots+t-1}, y_{r+s+\cdots+t}, \end{gathered}$$

således, at elementerne $y_r, y_{r+s}, \ldots, y_{r+s+\cdots+t}$ tilhører intervallet $[\alpha, \beta]$, og de resterende elementer tilhører $[0,1] \setminus [\alpha, \beta]$.

Størrelserne $r, s, \ldots, t$ er afstandene mellem forekomster af elementer, som tilhører intervallet $[\alpha, \beta]$.

De observerede afstande kan inddeles efter deres størrelse i $k+1$ kategorier:

$$\begin{aligned}
\text{obs}_0 &= (\text{Antallet af afstande} = 0)\\
\text{obs}_1 &= (\text{Antallet af afstande} = 1)\\
\text{obs}_2 &= (\text{Antallet af afstande} = 2)\\
&\vdots\\
\text{obs}_{k-1} &= (\text{Antallet af afstande} = k-1)\\
\text{obs}_k &= (\text{Antallet af afstande} \geq k).
\end{aligned}$$

Størrelserne $\text{obs}_0, \text{obs}_1, \ldots, \text{obs}_k$ skal sammenlignes med de tilsvarende *forventede* størrelser. Lad m være det samlede antal af afstande. Vi får da:

$$\begin{aligned}
\text{for}_0 &= (\text{Det forventede antal afstande} = 0) = mp\\
\text{for}_1 &= (\text{Det forventede antal afstande} = 1) = mp(1-p)\\
\text{for}_2 &= (\text{Det forventede antal afstande} = 2) = mp(1-p)^2\\
&\vdots\\
\text{for}_{k-1} &= (\text{Det forventede antal afstande} = k-1) = mp(1-p)^{k-1}\\
\text{for}_k &= (\text{Det forventede antal afstande} \geq k) = m(1-p)^k.
\end{aligned}$$

De forventede størrelser $\text{for}_0, \text{for}_1, \ldots, \text{for}_{k-1}$ fremkommer umiddelbart af ligning 3.20.

Udtrykket for for_k kræver en nærmere forklaring: antallet af forventede afstande, som er mindre end k er jo

$$\begin{aligned}
&\text{for}_0 + \text{for}_1 + \cdots + \text{for}_{k-1} =\\
&m\sum_{i=0}^{k-1} p(1-p)^i = m\left(1-(1-p)^k\right).
\end{aligned}$$

Antallet af forventede afstande større end k bliver da

$$m - m\left(1 - (1-p)^k\right) = m(1-p)^k.$$

I testen benyttes følgende teststørrelse:

$$q(\alpha, \beta, k, m) = \sum_{i=0}^{k} \frac{(\text{obs}_i - \text{for}_i)^2}{\text{for}_i}.$$

$q(\alpha, \beta, k, m)$ vil med tilnærmelse vil være χ^2-fordelt, hvis $mp(1-p)^{k-1} > 5$ og $m(1-p)^k > 5$, således, at det forventede antal i hver kategori er større end 5. Antallet af frihedsgrader vil være k.

Testsandsynligheden ϵ udregnes som

$$\epsilon = P(\mathcal{Q} > q(\alpha, \beta, k, m)).$$

3.1.7 Partitions-testen

Vi vil i denne test vurdere følgende hypotese:

Elementerne i sekvensen

$$y_0, y_1, \ldots, y_{n-1}, \tag{3.21}$$

hvor $n = st$, er ligefordelte i intervallet $[0, 1]$, og elementerne er fremkommet stokastisk uafhængigt af hinanden. Vi vil vurdere hypotesen ved i testen at betragte heltal. Derfor omformer vi 3.21 til en heltallig sekvens

$$x_i = \text{int}(d \times y_i), \tag{3.22}$$

hvor $i = 0, 1, \ldots, n-1$, således, at elementerne tilhører mængden $\{0, 1, \ldots, d-1\}$.

Vi har en inddeling af sekvensen 3.22 i s delsekvenser hver med t elementer. Vi kan referere til elementerne i den j'te delsekvens på følgende måde:

$$x_{jt}, x_{jt+1}, \ldots, x_{jt+t-1}, \tag{3.23}$$

hvor $0 \leq j < s$. Delsekvenserne kan nu inddeles i kategorier, efter antallet af forskellige elementer. I delsekvensen 3.23 er der t elementer. Hver plads udfyldes med et element fra mængden $\{0, 1, \ldots, d-1\}$. Der er r forskellige elementer på de t pladser og $1 \leq r \leq \min(d, t)$. Fremover vil vi antage, at $d \geq t$.

Vi skal nu løse følgende kombinatoriske problem:

> Givet r forskellige elementer udtaget fra en mængde bestående af d forskellige elementer; på hvor mange forskellige måder kan vi placere disse på t forskellige pladser.

Løsningen hertil er først at udtage r forskellige elementer fra mængden af de d elementer. Dette kan gøres på $d(d-1)\cdots(d-r+1)$ måder. Dernæst findes antallet af måder, hvorpå vi kan inddele en mængde bestående af t elementer i r delmængder. Dette kan gøres på $\left\{ \begin{matrix} t \\ r \end{matrix} \right\}$[5] forskellige måder. Kombineres de to resultater, bliver løsningen til problemet $d(d-1)\cdots(d-r+1)\left\{ \begin{matrix} t \\ r \end{matrix} \right\}$ måder. I det vi minder om at samtlige måder at placere d forskellige elementer på hver af de t pladser er d^t, bliver sandsynligheden for, at en delsekvens befinder sig i en kategori med r forskellige elementer:

$$p_r = \frac{d(d-1)\cdots(d-r+1)}{d^t} \left\{ \begin{matrix} t \\ r \end{matrix} \right\}. \tag{3.24}$$

Vi kan nu inddele observationerne i henhold til de t forskellige kategorier på følgende måde:

[5] $\{\ \}$ er notationen for stirling tal, som er defineret ved rekursionsformlen: $\left\{ \begin{matrix} t \\ r \end{matrix} \right\} = m \left\{ \begin{matrix} t-1 \\ r \end{matrix} \right\} + \left\{ \begin{matrix} t-1 \\ r-1 \end{matrix} \right\}$. Se Knuth, side 73 vol. 1.

$$
\begin{aligned}
\text{obs}_1 &= \text{(Antal delsekvenser med } t \text{ ens elementer)} \\
\text{obs}_2 &= \text{(Antal delsekvenser med 2 forskellige elementer)} \\
&\vdots \\
\text{obs}_{t-1} &= \text{(Antal delsekvenser med } t-1 \text{ forskellige elementer)} \\
\text{obs}_t &= \text{(Antal delsekvenser med } t \text{ forskellige elementer).}
\end{aligned}
$$

Da der er s forskellige delsekvenser, har vi, ved brug af 3.24, følgende *forventede* størrelser:

$$
\begin{aligned}
\text{for}_1 &= \text{(Antal delsekvenser med } t \text{ ens elementer)} = sp_1 \\
\text{for}_2 &= \text{(Antal delsekvenser med 2 forskellige elementer)} = sp_2 \\
&\vdots \\
\text{for}_{t-1} &= \text{(Antal delsekvenser med } t-1 \text{ forskellige elementer)} = sp_{t-1} \\
\text{for}_t &= \text{(Antal delsekvenser med } t \text{ forskellige elementer)} = sp_t,
\end{aligned}
$$

hvis hypotesen om ligefordeling og stokastisk uafhængighed har sin gyldighed. Som teststørrelse benyttes

$$q(s,d,t) = \sum_{i=1}^{t} \frac{(\text{obs}_i - \text{for}_i)^2}{\text{for}_i},$$

der med tilnærmelse vil være χ^2-fordelt, hvis $sp_r > 5$, for $1 \leq r \leq t$, således, at det forventede antal for hver kategori er større end 5. Antallet af frihedsgrader vil være $t-1$.

Testsandsynligheden ε udregnes som

$$\varepsilon = P(Q > q(s,d,t)).$$

3.2 Valg af test-metode

Vi skal, som tidligere beskrevet, vurdere en talgenerators anvendelighed vha en række statistiske tests. De tests, som vi vil udføre, er beskrevet i afsnit 3.1. Vi skal vurdere både den kaotiske og den udvalgte klassiske talgenerator.

Inden vi går igang med at udføre testene, skal vi først vælge en test-metode. En test-metode er en måde at udføre testene på. Valg af test-metode afhænger af, hvilket formål testene har.

Hvis vi f.eks har mistanke om, at en given talgenerator er dårlig, kan vi benytte os af en test-metode kaldet *sekventiel test*. En måde at konstruere en sekventiel test på vil være først at opdele intervallet $[0,1]$ i 3 regioner. Afhængigt af hvilken region en testsandsynlighed tilhører, vil det føre til, at vi hhv: forkaster en given hypotese, ikke forkaster en given hypotese eller, at vi udvider sekvensens længde, således at nye observationer tilføjes de tidligere observationer. Derved kan vi beregne en ny teststørrelse, hvis beregning er baseret på observationer bestående af både de tidligere og de ny observationer. Herefter kan vi finde en testsandsynlighed, som pånyt evalueres. Vi kan nu udnytte, at, hvis hypotesen er falsk (det er den, hvis den givne talgenerator er dårlig), gælder der om sekvensen af testsandsynligheder ε_i, at $\lim_{n\to\infty} \varepsilon_n = 0$ (n er sekvensens længde). I de tests, hvor elementerne i sekvenserne 3.1 eller 3.2 eller delsekvenser af disse kategoriseres, gælder der, at $\lim_{k\to\infty} \lim_{n\to\infty} \varepsilon_{(k,n)} = 0$ (k er antallet af kategorier). Fordelen ved at benytte sekventiel test er, at vi kan starte med relativt korte sekvenser (og hvis der i en test foretages kategorisering desuden et lille antal kategorier), og relativt hurtigt få konstateret, om den pågældende talgenerator nu også er dårlig.

Hvis en talgenerator ikke viser sig at være dårlig, ved benyttelse af den ovenfor beskrevne test-metode, vil vi endvidere gerne undersøge, hvor god den er. Hvis vi i praksis ikke kan forkaste hypotesen i en test, når blot n og k er relativt store, vil vi betegne den pågældende talgenerator som værende god. Hvor stor vi kan gøre n og k, uden at det vil føre til, at hypotesen må forkastes, giver os et mål for, hvor god den pågældende talgenerator er.

Knuth (side 27 vol. 2) giver en begrundelse for, at den udvalgte klassiske talgenerator må opfattes som værende god. Vi har således ikke

en begrundet mistanke om, at testsandsynlighederne i de enkelte tests vil konvergere mod 0 for relativt korte sekvenser, før vi går igang med at teste denne. Ud fra denne betragtning, sammen med ønsket om at kunne sammenligne de to talgeneratorer, vil vi ikke benytte os af en sekventiel test.

Den måde vi har valgt at udføre testene på, er ved at beregne teststørrelserne (og testsandsynlighederne) ud fra delsekvenser af sekvensen:

$$y_1, \ldots, y_{n_1}, y_{n_1+1}, \ldots, y_{n_1+n_2}, \ldots, y_{\Sigma_{i=1}^{t-1} n_i+1}, \ldots, y_{\Sigma_{i=1}^{t-1} n_i+n_t}.$$

Delsekvenserne er successive blokke af denne:

$$\begin{array}{c} \underbrace{y_1, y_2, \ldots, y_{n_1}}_{\varepsilon_1} \\ \underbrace{y_{n_1+1}, y_{n_1+2}, \ldots, y_{n_1+n_2}}_{\varepsilon_2} \\ \vdots \\ \underbrace{y_{\Sigma_{i=1}^{t-1} n_i+1}, y_{\Sigma_{i=1}^{t-1} n_i+2}, \ldots, y_{\Sigma_{i=1}^{t-1} n_i+n_t}}_{\varepsilon_t}. \end{array}$$

y_i'erne er de genererede tal, n_i'erne er delsekvensernes længde og ε_i'erne er testsandsynlighederne, dvs resultatet af hver af testene. t angiver antallet af tests.

Fra afsnit 3.1 ved vi, at hypoteserne i testene er forskellige. De er forskellige på den måde, at vi med nogle af testene kan afgøre, om elementerne i en sekvens er stokastisk uafhængige[6], andre om de er ligefordelte, samt nogle om de er ligefordelte og stokastisk uafhængige. Hypotesen i en test indrettes selvfølgelig efter, hvad der kan afgøres med den pågældende test. I de tests, med hvilke vi kan afgøre, om elementerne i sekvensen er stokastisk uafhængige og ligefordelte, har vi dobbelt-hypotesen: stokastisk uafhængighed og ligefordeling. Disse tests strider imidlertid mod almindelig praksis for hypotese-test, hvor kun én hypotese vurderes ad gangen. Derfor er testene med dobbelthypotesen tilrettelagt således, at vi har reduceret dobbelt-hypotesen til

[6]Dvs ringe korrelation, se afsnit 1.2.

en enkelt hypotese med en antagelse. Dette kræver selvfølgelig, at der er foretaget tests, hvori vi har vurderet om antagelsen holder. F.eks, hvis vi har reduceret dobbelt-hypotesen til en hypotese om ligefordeling med antagelse om stokastisk uafhængighed, da skal der først foretages tests med stokastisk uafhængighed som hypotese. Spørgsmålet er nu, hvilken rækkefølge testene skal udføres i.

I χ^2-testen og Kolmogorov-Smirnov-testen kræves det, at vi antager, at elementerne i en sekvens er stokastisk uafhængige. Det skyldes, at hvis elementerne i en sekvens er stokastisk afhængige og ligefordelte, vil det f.eks i χ^2-testen kunne lade sig gøre, ud fra en teststørrelse, at finde en testsandsynlighed i χ^2-fordelingen, der ville føre til, at hypotesen om ligefordeling ikke kunne forkastes. Dette har dog ikke nogen mening, da teststørrelsen ikke er χ^2-fordelt. Analoge resultater vil, under de givne omstændigheder, gælde for teststørrelserne i Kolmogorov-Smirnov-fordelingen.

Det er derfor oplagt først at udføre tests med stokastisk uafhængighed som hypotese, hvor vi ikke behøver at antage, at elementerne er ligefordelte. Til det formål benyttes Korrelations-testen og Permutations-testen, da der i disse tests ikke kræves nogen antagelse om, at elementerne i en sekvens skal være ligefordelte. Dernæst kan vi udføre tests, hvor vi har ligefordeling som hypotese og stokastisk uafhængighed som antagelse. Til det formål benyttes χ^2-testen, Kolmogorov-Smirnov-testen, Seriel-testen, Afstand-testen og Partitions-testen. Det bemærkes, at vi har reduceret dobbelt-hypotesen i Seriel-testen, Afstand-testen og Partitions-testen til en hypotese om ligefordeling med antagelse om stokastisk uafhængighed[7].

Hvis hypotesen i hver test er sand, vil testsandsynlighederne $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t$ være ligefordelte. Om testsandsynlighederne er ligefordelte kan afgøres vha Kolmogorov-Smirnov-testen, idet vi bruger $K = \max(K_t^+, K_t^-)$ som teststørrelse.

[7]Selvom vi har reduceret dobbelt-hypotesen til en hypotese om ligefordeling med antagelse om stokastisk uafhængighed, vil Seriel-testen, Afstand-testen og Partitions-testen, hvis antagelsen ikke holder, stadig kunne afgøre om elementerne i en sekvens er stokastisk uafhængige.

3.3 Valg af parametre

I dette afsnit angives hvilke parametre vi har valgt til de enkelte tests, som er beskrevet i afsnit 3.1. For hvert valg af parametre, er der udført to tests: en for den kaotiske talgenerator, og en for den klassiske talgenerator.

I testene indgår to grupper af parametre. Den første gruppe udgør parametrene α og β i Afstand-testen og f i Korrelations-testen. Valget af disse parametre er beskrevet i afsnittene 3.3.1 og 3.3.6.

Den anden gruppe af parametre er mere interessant: det er først og fremmest parameteren n, som angiver længden af den betragtede sekvens:

$$y_1, y_2, \ldots, y_n.$$

Vi er interesseret i store værdier af n, da sandsynligheden for ikke at forkaste en falsk hypotese derved bliver lille (se afsnit 3.2).

I Permutations-testen, χ^2-testen, Seriel-testen, Afstand-testen, og Partitions-testen indgår endnu en størrelse k, som er antallet af kategorier. Vi er interesseret i store værdier af denne størrelse, da igen sandsynligheden for ikke at forkaste en falsk hypotese bliver lille.

For Korrelations-testen og Kolmogorov-Smirnov-testen er vi således interesseret i, at $n \to \infty$ — og i resten af testene skal også $k \to \infty$.

Imidlertid må vi vælge værdier af n og k, således, at testene kan udføres indenfor en rimelig tidshorisont. Valget af n svinger mellem 10000 og 1680000 — men ofte vælges $n = 100000$.

k afhænger af valget af n (da $k < n$) — i alle tilfælde er det forventede antal forekomster i hver kategori større end 5 (af hensyn til χ^2-approksimationen).

3.3.1 Korrelations-testen

I Korrelations-testen er følgende parametre: n, der er antallet af elementer i sekvensen 3.6, samt forskydningsparameteren f. Vi har valgt længden n af sekvensen 3.6 "stor", nemlig $n = 100000$.

Følgende kombinationer af f og n anvendes:

	f	n
1	1	100000
2	16667	100000
3	33334	100000
4	50000	100000

$C(f,n)$ er en funktion af f, og der vil gælde, at

$$C(f,n) = C(n-f,n).$$

Dette er af praktisk betydning, da vi herved kan reducere antallet af tests. I stedet for at betragte værdier af f i mængden $\{1,2,\ldots,n-1\}$, kan vi nøjes med at betragte værdier af f i mængden $\{1,2,\ldots,n/2\}$ for n lige eller i mængden $\{1,2,\ldots,(n-1)/2\}$ for n ulige. I vort tilfælde betragter vi således mængden $\{1,2,\ldots,50000\}$ — vi har valgt "yderpunkterne" 1 og 50000, samt "inderpunkterne" 16667 og 33334.

3.3.2 Permutations-testen

I Permutations-testen er følgende parametre: s er antallet af delsekvenser 3.8, og t er antallet af elementer i disse delsekvenser.

Antallet af elementer n i sekvensen 3.7 er givet ved $n = st$. n skal være "stor". Vi vil også have et stort antal kategorier, som er givet ved $t!$. Vi har valgt $t = 7$ (således, at der er 5040 kategorier) og $s = 100800$ (det forventede antal er 20). Vi har da $n = 705600$.

3.3.3 χ^2-testen

I χ^2-testen er følgende parametre: n, som er antallet af elementer i sekvensen 3.11, samt d, der angiver antallet af kategorier.

Vi har valgt en "stor" værdi af d, nemlig $d = 5000$. n skal også være "stor". Da det forventede antal forekomster i hver kategori pga χ^2-approksimationen skal være større end 5, har vi valgt $n = 100000$ (dvs, at det forventede antal er 20).

3.3.4 Kolmogorov-Smirnov-testen

I Kolmogorov-Smirnov-testen er parameteren n, som er antallet af elementer i sekvensen 3.13. Som sædvanlig skal n være "stor". Normalt plejer vi at vælge $n = 100000$ eller større, men af tekniske årsager[8] måtte vi her vælge $n = 10000$. Alligevel er n stor nok til, at approksimation 3.15 kan anvendes — approksimationsfejlen bliver meget lille.

Med teststørrelserne K_∞^+ og K_∞^- er Kolmogorov-Smirnov-testen udført på uafhængige sekvenser.

3.3.5 Seriel-testen

I Seriel-testen er parametrene n og d. $2n$ er antallet af elementer i sekvensen 3.17; d^2 er antallet af mulige talpar 3.16.

Vi har valgt d "stor", nemlig $d = 70$, således, at der er 4900 kategorier.

Tilsvarende har vi valgt n "stor", nemlig $n = 98000$, således, at det forventede antal forekomster i hver kategori er 20.

[8] Elementerne i sekvensen 3.13 skal gemmes og sorteres — dette er ikke nødvendig i de andre tests.

3.3.6 Afstand-testen

I Afstand-testen er parametrene α og β, der udgør hhv begyndelses- og endepunkt i det betragtede interval $[\alpha, \beta]$; k som afgør antallet af kategorier (der er $k+1$ kategorier), samt m, der er antallet af afstande.

Vi har valgt følgende værdier:

	α	β	k	m
1	0	0.1	40	10000
2	0.7	1	20	30000
3	0.2	0.7	13	50000
4	0.3	1	8	100000
5	0	0.9	4	100000

Det ses, at vi har valgt interval-længderne 0.1, 0.3, 0.5, 0.7, og 0.9; intervallerne er *ad hoc* "spredt" ud på intervallet [0,1]. Størrelserne k og m er bestemt ud fra valget af α og β. Helst vil vi have k "stor", samt m "stor", da den betragtede sekvens 3.19 skal være lang — det estimerede antal elementer n i sekvensen 3.19 kan udregnes således $n \approx \frac{m}{\beta - \alpha} \cdot n$ forventes i alle testene at være større end eller lig med 100000. I alle tilfælde er m valgt, således, at det forventede antal forekomster i hver kategori større end 5 (pga χ^2-approksimationen).

3.3.7 Partitions-testen

I Partitions-testen er følgende parametre: d, som er det mulige antal forskellige elementer i delsekvensen 3.23; t, som angiver antallet af elementer i delsekvensen 3.23, samt s, som er antallet af delsekvenser. Vi har, at antallet af elementer n i sekvensen 3.21 er givet ved $n = st$. Som sædvanlig vil vi have n "stor", og vi har valgt $s = 120000$ og $t = 14$, således, at $n = 1680000$. d skal være større end eller lig med t, og vi har valgt $d = 17$.

t afgør antallet af kategorier — og t skal derfor være "stor". Imidlertid er der en række kategorier, hvor det forventede antal forekomster er *meget små*. Vi bliver nødt til at sammenlægge disse kategorier af hensyn til χ^2-approksimationen. Efter sammenlægningen af de første 5 kategorier (se side 32) har vi reelt 10 kategorier.

3.4 Analyse af testresultaterne

I afsnit 3.2 er beskrevet, hvorledes testene udføres. Testene udføres for både den kaotiske talgenerator og for den klassiske talgenerator.

Valget af parametre er beskrevet i afsnit 3.3.

3.4.1 Hypotese om stokastisk uafhængighed

Som det første skridt i analysen, vil vi vurdere, hvorvidt hypotesen om stokastisk uafhængighed kan forkastes eller ej. Til dette formål har vi Korrelations-testen og Permutations-testen.

Vi har udført Korrelations-testen og Permutations-testen på hhv den kaotiske og klassiske talgenerator med de parametre, som er beskrevet i afsnit 3.3 — resultaterne af disse tests er placeret i de nedenstående skemaer:

Korrelations-testen			
Parametre		Kaotisk	Klassisk
f	n	ϵ	ϵ
1	100000	0.0885	0.8149
16667	100000	0.3234	0.7838
33334	100000	0.5522	0.9512
50000	100000	0.1383	0.2816

Permutations-testen			
Parametre		Kaotisk	Klassisk
t	s	ϵ	ϵ
5040	100800	0.6479	0.8122

Som det ses, er de opnåede testsandsynligheder for hhv den kaotiske og klassiske talgenerator alle større end 0.05. Da 0.05 er et ofte anvendt signifikans-niveau, vil vi ikke forkaste hypotesen om stokastisk uafhængighed.

3.4.2 Hypotese om ligefordeling

Vi vil i dette afsnit antage, at hypotesen om stokastisk uafhængighed er sand — vi testede jo for stokastisk uafhængighed i afsnit 3.4.1. Vi vil nu afgøre, hvorvidt hypotesen om ligefordeling kan forkastes. Vi har χ^2-testen, Kolmogorov-Smirnov-testen, Seriel-testen, Afstand-testen og Partitions-testen til dette formål (Seriel-testen, Afstand-testen og Partitions-testen vil også kunne afsløre korrelation).

Disse tests er udført på den kaotiske og klassiske talgenerator, og resultaterne heraf er placeret i de nedenstående skemaer:

χ^2-testen			
Parametre		Kaotisk	Klassisk
d	n	ε	ε
5000	100000	0.9710	0.4350

Kolmogorov-Smirnov-testen			
$K_\infty^\pm$	Parameter	Kaotisk	Klassisk
	n	ε	ε
K_∞^+	10000	0.3899	0.2018
K_∞^-	10000	0.7137	0.1522

Seriel-testen			
Parametre		Kaotisk	Klassisk
d	n	ε	ε
70	98000	0.4288	0.1243

Afstand-testen					
Parametre				Kaotisk	Klassisk
α	β	k	m	ε	ε
0	0.1	40	10000	0.0961	0.5968
0.7	1	20	30000	0.1064	0.3396
0.2	0.7	13	50000	0.9530	0.6998
0.3	1	8	100000	0.4716	0.9159
0	0.9	4	100000	0.5549	0.1631

Partitions-testen				
Parametre			Kaotisk	Klassisk
d	t	s	ε	ε
17	14	120000	0.7978	0.0779

Med et signifikans-niveau på 0.05 kan vi ikke på dette grundlag forkaste hypotesen om ligefordeling.

3.4.3 Analyse af testsandsynlighederne

Vi har i de ovenstående 2×15 tests opnået to sekvenser af testsandsynligheder — og testsandsynlighederne er alle over et signifikans-niveau på 0.05. Imidlertid er testsandsynlighederne ligefordelte, hvis de pågældende hypoteser er sande. Derfor udføres en Kolmogorov-Smirnov-test på hver af de to sekvenser af testsandsynligheder.

For sekvensen bestående af de 15 testsandsynligheder for den kaotiske talgenerator, får vi

$$K_{15}^{+} = 0.4972$$

og

$$K_{15}^{-} = 0.3428.$$

Idet vi bruger $K_{15} = \max(K_{15}^{+}, K_{15}^{-})$ som teststørrelse, får vi ved tabelopslag[9] testsandsynligheden $\varepsilon > 0.5$. På dette grundlag kan vi ikke forkaste hypotesen om ligefordeling og stokastisk uafhængighed — men må acceptere den kaotiske talgenerator som værende udmærket til generering af pseudotilfældige tal.

For den klassiske talgenerator får vi

$$K_{15}^{+} = 0.5094$$

og

$$K_{15}^{-} = 0.4537.$$

Ved anvendelse af $K_{15} = \max(K_{15}^{+}, K_{15}^{-})$ som teststørrelse, får vi $\varepsilon > 0.5$. På dette grundlag kan vi ikke heller ikke forkaste hypotesen om ligefordeling og stokastisk uafhængighed — og vi vil ligeledes acceptere den klassiske talgenerator som værende udmærket til generering af pseudotilfældige tal.

[9] Knuth, side 48 vol. 2.

Kapitel 4

God talgenerator = stor K-entropi + simpel transformation

Vi har i dette projekt undersøgt muligheden for at danne pseudotilfældige tal vha en kaotisk talgenerator.

Ud over, at en talgenerator skal opfylde, at de genererede elementer er ligefordelte og fremkommer stokastisk uafhængigt af hinanden, skal talgeneratoren også være *hurtig*.

De klassiske talgeneratorer udmærker sig ved, at disse kan implementeres meget effektivt i lav-niveau programmeringssprog — de bliver meget hurtige. Den kaotiske talgenerator, som er beskrevet i denne projektrapport, vil i et lav-niveau programmeringssprog være langsom i forhold til klassiske talgeneratorer. Vi har implementeret den kaotiske og den udvalgte klassiske talgenerator i et høj-niveau programmeringssprog — i dette tilfælde er den klassiske talgenerator ca. 17 gange hurtigere end den kaotiske.

Muligvis kunne vi reducere dette gab, hvis vi vælger en *anden* transformation end

$$Tx_i = 4x_i \times (1 - x_i) \tag{4.1}$$

som udgangspunkt for en kaotisk talgenerator.

I afsnit 2.2 opfattede vi sekvensen $x_0, x_1, \ldots$, der er genereret af transformationen T, som værende observationer af en stationær stokastisk proces $\{X_i, i \in \mathbf{N_0}\}$. Det fremgår af afsnittet, at K-entropien af denne proces er lig $\log_e(2)$. Processen

$$\{Z_i, i \in \mathbf{N_0}\} = \begin{cases} 0 & \text{hvis } X_i \in [0, \frac{1}{2}] \\ 1 & \text{hvis } X_i \in]\frac{1}{2}, 1], \end{cases}$$

er en Bernoulli-proces, og vi kunne generere den ønskede sekvens $y_0, y_1, \ldots, y_{n-1}$ ud fra sekvensen

$$z_1, \ldots, z_m, z_{m+1}, \ldots, z_{2m}, \ldots, z_{(n-1)m+1}, \ldots, z_{nm}, \tag{4.2}$$

på følgende måde:

$$y_k = \sum_{j=1}^{m} 2^{-j} z_{km+j},$$

for $k = 0, 1, \ldots, n-1$. Det ses, at vi fortolker delsekvensen $z_{km+1}, z_{km+2}, \ldots, z_{km+m}$ i 4.2 som en repræsentation af y_k i det binære talsystem: $y_k = 0.z_{km+1} z_{km+2} \cdots z_{km+m}$.

Vi vil nu generalisere dette: lad os opfatte sekvensen $x_0, x_1, \ldots$ genereret af transformationen G som værende observationer af den stationære proces $\{X_i, i \in \mathbf{N_0}\}$. Hvis K-entropien af $\{X_i, i \in \mathbf{N_0}\}$ er større end (og i visse tilfælde lig med) $\log_e(h)$, så *eksisterer* klasseinddelinger $A_0, A_1, \ldots, A_{h-1}$ af intervallet $[0, 1]$ (eller hvilken mængde vi end betragter), således, at

$$\{Z_i, i \in \mathbf{N_0}\} = \begin{cases} 0 & \text{hvis } X_i \in A_0 \\ 1 & \text{hvis } X_i \in A_1 \\ \vdots & \vdots \\ h-1 & \text{hvis } X_i \in A_{h-1} \end{cases} \tag{4.3}$$

er en *symmetrisk* Bernoulli-proces. Entropien af denne proces er netop lig $\log_e(h)$[1]. Pointen er nu at repræsentere den ønskede sekvens i h-talsystemet.

Har vi sekvensen

$$z_1, \ldots, z_m, z_{m+1}, \ldots, z_{2m}, \ldots, z_{(n-1)m+1}, \ldots, z_{nm}, \tag{4.4}$$

som er observationer af 4.3, kan y_k, hvor $k = 0, 1, \ldots, n-1$, dannes på følgende måde:

$$y_k = \sum_{j=1}^{m} h^{-j} z_{km+j}. \tag{4.5}$$

Elementerne y_k vil tilhøre intervallet $[0, (h-1)\sum_{j=1}^{m} h^{-j}]$. Det er klart, at vi er interesseret i at kunne generere et *stort* antal *forskellige* elementer ved hjælp af formel 4.5. Antallet af mulige genererede tal i dette interval er givet ved h^m.

Helst vil vi have h "stor", men m "lille". Grunden er, at vi skal have m observationer af processen $\{X_i, i \in \mathbf{N_0}\}$ for hvert element i den ønskede

[1]Vi kan også anvende andre Bernoulli-processer, hvor entropien er mindre end $\log_e(h)$. I disse tilfælde skal vi dog "snyde". F.eks kunne vi betragte Bernoulli-processer $\{Z_i, i \in \mathbf{N_0}\}$ med tilstandene $0, 1, \ldots, h-1, h, h+1, \ldots, s$, hvor

$$P(Z_i = 0) = P(Z_i = 1) = \cdots = P(Z_i = h-1)$$

og

$$\sum_{k=0}^{s} P(Z_i = k) \log_e(P(Z_i = k)) < K(\{X_i, i \in \mathbf{N_0}\}),$$

samt

$$\sum_{k=h}^{s} P(Z_i = k) > 0.$$

Ved at betragte sandsynlighederne

$$\begin{aligned} P(Z_i = 0 | Z_i \in \{0, 1, \ldots, h-1\}) &= \\ P(Z_i = 1 | Z_i \in \{0, 1, \ldots, h-1\}) &= \cdots = \\ P(Z_i = h-1 | Z_i \in \{0, 1, \ldots, h-1\}) & \end{aligned}$$

er vi tilbage til "det symmetriske tilfælde", idet vi ignorerer forekomster af tilstandene $h, h+1, \ldots, s$.

sekvens $y_0, y_1, \ldots, y_{n-1}$. Ideelt set behøvede vi kun én observation af processen $\{X_i, i \in \mathbf{N_0}\}$, hvis K-entropien af $\{X_i, i \in \mathbf{N_0}\}$ var "stor". I dette tilfælde har vi, at $y_i = h^{-1} z_i$.

Hvis vi skulle danne en *hurtig* kaotisk talgenerator, er der to ting vi skal være opmærksomme på:

- Hvilket talsystem, vi kan tillade os at repræsentere den ønskede sekvens $y_0, y_1, \ldots, y_{n-1}$ i (afhængig af hvor stor K-entropien er).
- Hvor kompliceret transformationen G er (dvs antallet af multiplikationer, subtraktioner osv i G).

Som det fremgår af side 15, skal der for hvert tal, som vor kaotiske talgenerator leverer, udføres 20 iterationer af 4.1. For hver iteration skal der udføres en multiplikation (vi ser bort fra multiplikationen med 4 i 4.1, da denne er triviel), samt en subtraktion. Det højeste antal forskellige elementer, som den kaotiske talgenerator kan danne, er 2^{20}.

Hvis nu K-entropien var større end (evt. lig med) $\log_e(4)$, så kunne vi bruge den symmetriske Bernoulli-proces med *fire* tilstande som model, og kunne nøjes med 10 iterationer af 4.1 for at få $4^{10} = 2^{20}$ mulige elementer. Dvs, at vi i dette tilfælde har 10 multiplikationer og 10 subtraktioner — da multiplikation er meget tidskrævende, bliver talgeneratoren en del hurtigere. Dette (fiktive) eksempel skulle vise, at jo større K-entropien er, og jo simplere (få multiplikationer, subtraktioner osv) transformationen er, jo hurtigere bliver talgeneratoren.

En udvidelse af dette projekt kunne være at finde (eller konstruere) en transformation, således, at vi har en stor K-entropi at arbejde med. Transformationen skal selvfølgelig være simpel — men hvor simpel den *skal være*, afhænger af K-entropien; hvis K-entropien er stor, kan vi tillade at benytte en knap så simpel transformation, og dog bevare effektiviteten. Vi vil dog ikke forvente, at kaotiske talgeneratorer kan blive hurtigere end de klassiske.

Konklusion

Vi har i dette projekt fremkommet med ideen, at kaotiske ligninger kan anvendes i forbindelse med generering af pseudotilfældige tal. Som eksempel på en kaotisk ligning er følgende benyttet:

$$x_{i+1} = 4x_i(1 - x_i).$$

Den kaotiske talgenerator, som er baseret på ovenstående ligning, er blevet sammenlignet med den klassiske talgenerator, som er blevet anbefalet af Knuth (se afsnit 2.1).

I kapitel 2.2 har vi fremført teoretiske argumenter for kaotiske ligningers anvendelighed i forbindelse med generering af pseudotilfældige tal.

Tilsvarende har vi i kapitel 3 fremført statistiske argumenter. Vi har lavet en statistisk sammenligning af den kaotiske talgenerator og den klassiske talgenerator; i afsnit 3.4 så vi, at den kaotiske talgenerator klarede sig lige så godt som den udvalgte klassiske talgenerator. Sekvensen af testsandsynligheder, som blev dannet ud fra de enkelte tests, var tilsyneladende ligefordelte. Vi vil derfor acceptere hypotesen: den kaotiske talgenerator danner ligefordelte sekvenser, og elementerne i disse sekvenser fremkommer stokastisk uafhængigt af hinanden.

Det væsentlige punkt for dette projekt er følgende: talgeneratoren, der er baseret på den kaotiske ligning må betragtes som værende udmærket til generering af pseudotilfældige tal. Den er hverken dårligere eller bedre end den udvalgte klassiske talgenerator. Imidlertid er vor kaotiske talgenerator al for langsom i forhold til den udvalgte klassiske talgenerator, så talgeneratoren i sin nuværende form er ikke aktuel for praktiske anvendelser.

Appendiks A

Beregningspræcision

Effektiviteten af den kaotiske talgenerator er afhængig af, hvor mange betydende cifre der regnes med. Som regel bliver elementerne $x_k, x_{k+1}, \ldots, x_{n-1}$ fra sekvensen $x_0, x_1, \ldots, x_k, \ldots, x_{n-1}$, der er genereret af den kaotiske ligning

$$x_{i+1} = 4x_i(1 - x_i),$$

0, hvis beregningspræcisionen er for lille. I det nedenstående skema er resultaterne af fire testudførelser. Disse testudførelser skal illustrere, hvor stor betydning beregningspræcisionen har på de genererede sekvenser. I alle fire testudførelser er der brugt χ^2-testen med parametrene $d = 5000$ og $n = 100000$, og samme "startværdi" x_0 er benyttet. ε er de opnåede testsandsynligheder.

Antal betydende cifre	ε
7	0.0000
11	0.8718
15	0.6198
19	0.4330

Testsandsynlighederne viser, at de genererede sekvenser i de fire tilfælde er *meget forskellige* — specielt viser de, at vi *ikke* kan nøjes med at anvende 7 cifres præcision.

Til testudførelserne, som er beskrevet i afsnit 3.4, har vi anvendt en beregningspræcision på 19 cifre — og da testene "godkendte" talgeneratoren, er der således ikke til praktisk brug behov for en større beregningspræcision.

Litteraturliste

Borch, Tommy
Statistik
Fag
1981

Borland
Turbo Pascal Reference Guide
(version 5.0)
Borland
1989

Brown, James R.
Ergodic theory and topological dynamics
Academic press, inc.
1976

Christiansen, Peder Voetmann
Tilfældighedens nødvendighed ifølge Peirce og fysikken
IMFUFA, Roskilde Universitetscenter
1985

Hansen, Peter & Henriksen, Per
Turbo Pascal 5.5
Teknisk Forlag
1989

Jørsboe, Ole Groth
Sandsynlighedsregning
Matematisk Institut, DTH
1990

Knuth, Donald Ervin
Seminumerical algoritms
The art of computer programming, vol. 1, Second Edition
Stanford University
1973

Knuth, Donald Ervin
Seminumerical algoritms
The art of computer programming, vol. 2, Second Edition
Stanford University
1981

M. Kubcek, M. Marek
Computational Methods in Bifurcation Theory and Dissipative Structures
Springer-Verlag New York Inc.
1983

Lamperti, John
Stochastic Processes
Springer-Verlag
1977

Larsen, Jørgen
Basisstatistik 1. Diskrete modeller, anden udgave
IMFUFA, Roskilde Universitetscenter
1989

Larsen, Jørgen
Basisstatistik 2. Kontinuerte modeller, anden udgave
IMFUFA, Roskilde Universitetscenter
1989

Larsen, Jørgen
Grundbegreber i sandsynlighedsregningen, anden udgave
IMFUFA, Roskilde Universitetscenter
1989

Larsen, Jørgen
Introduktion til ISP, anden udgave
IMFUFA, Roskilde Universitetscenter
1989

Ripley, Brian D.
Stochastic Simulation
John Wiley & Sons, Inc.
1987

Schuster, Heinz Georg
Deterministic chaos
Physik-verlag
1984

Liste over tidligere udkomne tekster tilsendes gerne. Henvendelse herom kan ske til IMFUFA's sekretariat tlf. 46 75 77 11 lokal 2263

217/92 "Two papers on APPLICATIONS AND MODELLING IN THE MATHEMATICS CURRICULUM"
by: Mogens Niss

218/92 "A Three-Square Theorem"
by: Lars Kadison

219/92 "RUPNOK - stationær strømning i elastiske rør"
af: Anja Boisen, Karen Birkelund, Mette Olufsen
Vejleder: Jesper Larsen

220/92 "Automatisk diagnosticering i digitale kredsløb"
af: Bjørn Christensen, Ole Møller Nielsen
Vejleder: Stig Andur Pedersen

221/92 "A BUNDLE VALUED RADON TRANSFORM, WITH APPLICATIONS TO INVARIANT WAVE EQUATIONS"
by: Thomas P. Branson, Gestur Olafsson and Henrik Schlichtkrull

222/92 On the Representations of some Infinite Dimensional Groups and Algebras Related to Quantum Physics
by: Johnny T. Ottesen

223/92 *THE FUNCTIONAL DETERMINANT*
by: Thomas P. Branson

224/92 *UNIVERSAL AC CONDUCTIVITY OF NON-METALLIC SOLIDS AT LOW TEMPERATURES*
by: Jeppe C. Dyre

225/92 *"HATMODELLEN" Impedansspektroskopi i ultrarent en-krystallinsk silicium*
af: Anja Boisen, Anders Gorm Larsen, Jesper Varmer, Johannes K. Nielsen, Kit R. Hansen, Peter Bøggild og Thomas Hougaard
Vejleder: Petr Viscor

226/92 *"METHODS AND MODELS FOR ESTIMATING THE GLOBAL CIRCULATION OF SELECTED EMISSIONS FROM ENERGY CONVERSION"*
by: Bent Sørensen

227/92 "Computersimulering og fysik"
af: Per M.Hansen, Steffen Holm, Peter Maibom, Mads K. Dall Petersen, Pernille Postgaard, Thomas B.Schrøder, Ivar P. Zeck
Vejleder: Peder Voetmann Christiansen

228/92 "Teknologi og historie"
Fire artikler af:
Mogens Niss, Jens Høyrup, Ib Thiersen, Hans Hedal

229/92 "Masser af information uden betydning"
En diskussion af informationsteorien i Tor Nørretranders' "Mærk Verden" og en skitse til et alternativ basseret på andenordens kybernetik og semiotik.
af: Søren Brier

230/92 "Vinklens tredeling - et klassisk problem"
et matematisk projekt af
Karen Birkelund, Bjørn Christensen
Vejleder: Johnny Ottesen

231A/92 "Elektrondiffusion i silicium - en matematisk model"
af: Jesper Voetmann, Karen Birkelund, Mette Olufsen, Ole Møller Nielsen
Vejledere: Johnny Ottesen, H.B.Hansen

231B/92 "Elektrondiffusion i silicium - en matematisk model" Kildetekster
af: Jesper Voetmann, Karen Birkelund, Mette Olufsen, Ole Møller Nielsen
Vejledere: Johnny Ottesen, H.B.Hansen

232/92 "Undersøgelse om den simultane opdagelse af energiens bevarelse og især deles om de af Mayer, Colding, Joule og Helmholtz udførte arbejder"
af: L.Arleth, G.I.Dybkjær, M.T.Østergård
Vejleder: Dorthe Posselt

233/92 "The effect of age-dependent host mortality on the dynamics of an endemic disease and
Instability in an SIR-model with age-dependent susceptibility
by: Viggo Andreasen

234/92 "THE FUNCTIONAL DETERMINANT OF A FOUR-DIMENSIONAL BOUNDARY VALUE PROBLEM"
by: Thomas P. Branson and Peter B. Gilkey

235/92 OVERFLADESTRUKTUR OG POREUDVIKLING AF KOKS
- Modul 3 fysik projekt -
af: Thomas Jessen

236a/93 INTRODUKTION TIL KVANTE HALL EFFEKTEN

af: Anja Boisen, Peter Bøggild

Vejleder: Peder Voetmann Christiansen
Erland Brun Hansen

236b/93 STRØMSSAMMENBRUD AF KVANTE HALL EFFEKTEN

af: Anja Boisen, Peter Bøggild

Vejleder: Peder Voetmann Christiansen
Erland Brun Hansen

237/93 The Wedderburn principal theorem and Shukla cohomology

af: Lars Kadison

238/93 SEMIOTIK OG SYSTEMEGENSKABER (2)

Vektorbånd og tensorer

af: Peder Voetmann Christiansen

239/93 Valgsystemer - Modelbygning og analyse

Matematik 2. modul

af: Charlotte Gjerrild, Jane Hansen, Maria Hermannsson, Allan Jørgensen, Ragna Clauson-Kaas, Poul Lützen

Vejleder: Mogens Niss

240/93 Patologiske eksempler.
Om sære matematiske fisks betydning for den matematiske udvikling

af: Claus Dræby, Jørn Skov Hansen, Runa Ulsøe Johansen, Peter Meibom, Johannes Kristoffer Nielsen

Vejleder: Mogens Niss

241/93 FOTOVOLTAISK STATUSNOTAT 1

af: Bent Sørensen

242/93 Brovedligeholdelse - bevar mig vel

Analyse af Vejdirektoratets model for optimering af broreparationer

af: Linda Kyndlev, Kare Fundal, Kamma Tulinius, Ivar Zeck

Vejleder: Jesper Larsen

243/93 TANKEEKSPERIMENTER I FYSIKKEN

Et 1.modul fysikprojekt

af: Karen Birkelund, Stine Sofia Korremann

Vejleder: Dorthe Posselt

244/93 RADONTRANSFORMATIONEN oq dens anvendelse i CT-scanning

Projektrapport

af: Trine Andreasen, Tine Guldager Christiansen, Nina Skov Hansen og Christine Iversen

Vejledere: Gestur Olafsson og Jesper Larsen

245a+b /93 Time-Of-Flight målinger på krystallinske halvledere
Specialerapport

af: Linda Szkotak Jensen og Lise Odgaard Gade

Vejledere: Petr Viscor og Niels Boye Olsen

246/93 HVERDAGSVIDEN OG MATEMATIK - LÆREPROCESSER I SKOLEN

af: Lena Lindenskov, Statens Humanistiske Forskningsråd, RUC, IMFUFA

247/93 UNIVERSAL LOW TEMPERATURE AC CONDUCTIVITY OF MACROSCOPICALLY DISORDERED NON-METALS

by: Jeppe C. Dyre

248/93 DIRAC OPERATORS AND MANIFOLDS WITH BOUNDARY

by: B. Booss-Bavnbek, K.P.Wojciechowski

249/93 Perspectives on Teichmüller and the Jahresbericht Addendum to Schappacher, Scholz, et al.

by: B. Booss-Bavnbek

With comments by W.Abikoff, L.Ahlfors, J.Cerf, P.J.Davis, W.Fuchs, F.P.Gardiner, J.Jost, J.-P.Kahane, R.Lohan, L.Lorch, J.Radkau and T.Söderqvist

250/93 EULER OG BOLZANO - MATEMATISK ANALYSE SET I ET VIDENSKABSTEORETISK PERSPEKTIV

Projektrapport af: Anja Juul, Lone Michelsen, Tomas Højgård Jensen

Vejleder: Stig Andur Pedersen

251/93 Genotypic Proportions in Hybrid Zones

by: Freddy Bugge Christiansen, Viggo Andreasen and Ebbe Thue Poulsen

252/93 MODELLERING AF TILFÆLDIGE FÆNOMENER

Projektrapport af: Birthe Friis, Lisbeth Helmgaard Kristina Charlotte Jakobsen, Marina Mosbæk Johannessen, Lotte Ludvigsen, Mette Hass Nielsen

253/93 Kuglepakning

Teori og model

af: Lise Arleth, Kåre Fundal, Nils Kruse

Vejleder: Mogens Niss

254/93 Regressionsanalyse

Materiale til et statistikkursus

af: Jørgen Larsen

255/93 TID & BETINGET UAFHÆNGIGHED

af: Peter Harremoës

256/93 Determination of the Frequency Dependent Bulk Modulus of Liquids Using a Piezoelectric Spherical Shell (Preprint)

by: T. Christensen and N.B.Olsen

257/93 Modellering af dispersion i piezoelektriske keramikker

af: Pernille Postgaard, Jannik Rasmussen, Christina Specht, Mikko Østergård

Vejleder: Tage Christensen

258/93 Supplerende kursusmateriale til "Lineære strukturer fra algebra og analyse"

af: Mogens Brun Heefelt

259/93 STUDIES OF AC HOPPING CONDUCTION AT LOW TEMPERATURES

by: Jeppe C. Dyre

260/93 PARTITIONED MANIFOLDS AND INVARIANTS IN DIMENSIONS 2, 3, AND 4

by: B. Booss-Bavnbek, K.P.Wojciechowski

261/93 OPGAVESAMLING
Bredde-kursus i Fysik
Eksamensopgaver fra 1976-93

262/93 Separability and the Jones Polynomial
by: Lars Kadison

263/93 Supplerende kursusmateriale til "Lineære strukturer fra algebra og analyse" II
af: Mogens Brun Heefelt

264/93 FOTOVOLTAISK STATUSNOTAT 2
af: Bent Sørensen

265/94 SPHERICAL FUNCTIONS ON ORDERED SYMMETRIC SPACES
To Sigurður Helgason on his sixtyfifth birthday
by: Jacques Faraut, Joachim Hilgert and Gestur Olafsson

266/94 Kommensurabilitets-oscillationer i laterale supergitre
Fysikspeciale af: Anja Boisen, Peter Bøggild, Karen Birkelund
Vejledere: Rafael Taboryski, Poul Erik Lindelof, Peder Voetmann Christiansen

267/94 Kom til kort med matematik på Eksperimentarium - Et forslag til en opstilling
af: Charlotte Gjerrild, Jane Hansen
Vejleder: Bernhelm Booss-Bavnbek

268/94 Life is like a sewer ...
Et projekt om modellering af aorta via en model for strømning i kloakrør
af: Anders Marcussen, Anne C. Nilsson, Lone Michelsen, Per M. Hansen
Vejleder: Jesper Larsen

269/94 Dimensionsanalyse en introduktion
metaprojekt, fysik
af: Tine Guldager Christiansen, Ken Andersen, Nikolaj Hermann, Jannik Rasmussen
Vejleder: Jens Højgaard Jensen

270/94 THE IMAGE OF THE ENVELOPING ALGEBRA AND IRREDUCIBILITY OF INDUCED REPRESENTATIONS OF EXPONENTIAL LIE GROUPS
by: Jacob Jacobsen

271/94 Matematikken i Fysikken.
Opdaget eller opfundet
NAT-BAS-projekt
vejleder: Jens Højgaard Jensen

272/94 Tradition og fornyelse
Det praktiske elevarbejde i gymnasiets fysikundervisning, 1907-1988
af: Kristian Hoppe og Jeppe Guldager
Vejledning: Karin Beyer og Nils Hybel

273/94 Model for kort- og mellemdistanceløb
Verifikation af model
af: Lise Fabricius Christensen, Helle Pilemann, Bettina Sørensen
Vejleder: Mette Olufsen

274/94 MODEL 10 - en matematisk model af intravenøse anæstetikas farmakokinetik
3. modul matematik, forår 1994
af: Trine Andreasen, Bjørn Christensen, Christine Green, Anja Skjoldborg Hansen. Lisbeth Helmgaard
Vejledere: Viggo Andreasen & Jesper Larsen

275/94 Perspectives on Teichmüller and the Jahresbericht
2nd Edition
by: Bernhelm Booss-Bavnbek

276/94 Dispersionsmodellering
Projektrapport 1. modul
af: Gitte Andersen, Rehannah Borup, Lisbeth Friis, Per Gregersen, Kristina Vejrø
Vejleder: Bernhelm Booss-Bavnbek

277/94 PROJEKTARBEJDSPÆDAGOGIK - Om tre tolkninger af problemorienteret projektarbejde
af: Claus Flensted Behrens, Frederik Voetmann Christiansen, Jørn Skov Hansen, Thomas Thingstrup
Vejleder: Jens Højgaard Jensen

278/94 The Models Underlying the Anaesthesia Simulator Sophus
by: Mette Olufsen(Math-Tech), Finn Nielsen (RISØ National Laboratory), Per Føge Jensen (Herlev University Hospital), Stig Andur Pedersen (Roskilde University)

279/94 Description of a method of measuring the shear modulus of supercooled liquids and a comparison of their thermal and mechanical response functions.
af: Tage Christensen

280/94 A Course in Projective Geometry
by Lars Kadison and Matthias T. Kromann

281/94 Modellering af Det Cardiovaskulære System med Neural Pulskontrol
Projektrapport udarbejdet af:
Stefan Frello, Runa Ulsøe Johansen, Michael Poul Curt Hansen, Klaus Dahl Jensen
Vejleder: Viggo Andreasen