Entropie

Bas Janssens

Herfst 2008

Inhoudsopgave

1 Het Project **5**

1.1 Praktische zaken 5

1.2 Verslag en presentatie 5

1.2.1 Een aantal voor de hand liggende opmerkingen over het verslag 6

1.2.2 Een aantal voor de hand liggende opmerkingen over het praatje 6

1.3 Geen Meden en Perzen 7

2 Entropie **7**

2.1 Definitie 8

2.2 Karakterisering 10

2.3 Opdracht voor allen: een stelling van Shannon 12

3 Voor allen één opdracht **14**

3.1 Vul zelf in 14

3.2 Large deviations 15

3.3 Maak je eigen datacompressor 16

3.4 Een coderingsstelling voor Markovketens 16

3.5 Ruis 17

3.6 Statistische mechanica 17

3.7 Chaos en Kolmogorov-Sinai-entropie 18

3.8 Quantumentropie 19

1 Het Project

De bedoeling van het vak 'Project' is drieledig. Ten eerste leer je met anderen samenwerken om problemen op te lossen. Ten tweede leer je onderweg wat informatietheorie. En ten derde oefen je je presentatievaardigheden.

1.1 Praktische zaken

We pakken het als volgt aan. De eerste twee weken (09–09 en 16–09) zal ik gebruiken om kort wat te vertellen over het begrip 'entropie', de hoeksteen van de informatietheorie.

Vorm groepjes van vier. De tweede week (16–09) hoor ik graag wat de groepjes zijn.

Hoofdstuk 2 bevat een opgave die ieder groepje moet maken. Van ieder groepje verwacht ik één verslag, op of voor 07–10. (Iedere dag te laat is een punt eraf.)

Hoofdstuk 3 bevat een aantal keuzeopdrachten, waarvan ieder groepje er een maakt. Stel gezamenlijk een top 3 op van meest begeerde opdrachten, en laat me deze op of voor 30–09 weten. Door van tevoren met elkaar te overleggen valt het vast wel zó te regelen dat elke groep een leuke opdracht krijgt. (Alleen al omdat ze allemaal leuk zijn.) Van de keuzeopdracht ontvang ik graag een verslag op of voor 16–12. (Iedere dag te laat is een punt eraf.)

Voorts organiseert ieder groepje twee praatjes. Een op 21–10 en een op 16–12. Ik wil graag dat ieder praatje gehouden wordt door twee mensen, zodat iedereen een keer 'voor de klas' staat. Op die data wil ik graag dat iedereen aanwezig is, ook degenen die geen praatje houden. Leg in het eerste praatje uit wat het probleem is dat je op hoopt te lossen. Leg in het tweede praatje nogmaals het probleem uit (het eerste praatje is iedereen dan al weer vergeten), en breng verslag uit van je bevindingen.

Let op! Het laatste praatje valt dus samen met de datum waarop het verslag dient ingeleverd! Houd hier rekening mee met je planning.

1.2 Verslag en presentatie

Maak van tevoren een planning en verdeel de taken. Jullie mogen zelf bepalen hoe, onder voorwaarde dat iedereen meedenkt over de wiskundige inhoud van het verslag, en dat iedereen mee schrijft aan het verslag. Je zou er bijvoorbeeld aan kunnen denken in je groepje van vier één persoon verantwoordelijk te maken voor de vorm van het verslag, een voor de inhoud van het verslag, een voor de praatjes en een voor de planning . (Waarbij verantwoordelijk zijn voor iets natuurlijk niet hetzelfde is als iets helemaal alleen opknappen.) Het verslag is uiteindelijk de verantwoordelijkheid van de hele groep.

Integraal onderdeel van het verslag is een 'logboek', waarin jullie aangeven wie hoelang waaraan gewerkt heeft, en hoe de taakverdeling lag. Als onderdeel van het verslag, is het logboek de verantwoordelijkheid van de hele groep.

1.2.1 Een aantal voor de hand liggende opmerkingen over het verslag

Schat je doelgroep in en sluit aan bij de voorkennis van je lezers. Je schrijft dit verslag voor je medestudenten; probeer het voor hen interessant en begrijpelijk te maken.

Zorg voor een overzichtelijke structuur, met kop, lijf en staart. In de inleiding leg je uit wat je gaat doen, en waarom. Geef eventueel kort aan wat waar gebeurt in je verslag. In het slot vat je samen wat je gedaan hebt, trek je conclusies en kijk je kritisch terug op het voorafgaande. Beantwoord eventuele vragen uit de inleiding.

Zorg voor een logische indeling in hoofdstukken.

Correct Nederlands oogt professioneel en leidt niet af van de inhoud. Let dus op je taalgebruik, in het bijzonder op werkwoordstijden. Vooral de derde persoon enkelvoud tegenwoordige tijd wordt nogal eens verkeerd geschreven ('dat betekend'). Een elektronische spellingchecker is geen substituut voor gezond verstand, maar kan veel tikfouten voorkomen.

Soms zeggen plaatjes meer dan duizend woorden. Schuw ze niet als je denkt dat ze verhelderend kunnen werken. Zorg dat figuren een genummerd bijschrift hebben, (Bijvoorbeeld "Fig.7: Grafiek in de vorm van een smurf.") en besteed aandacht aan een verzorgde opmaak.

Formules vormen een integraal onderdeel van de grammaticale structuur van je verslag. Let bijvoorbeeld op interpunctie in "De oppervlakte van een cirkel wordt gegeven door de formule

$$A = \pi r^2 \,,$$

waarbij A de oppervlakte en r de straal is."

En tenslotte het belangrijkste: vermeld *altijd* je bronnen.

Lees al vóór je aan het verslag begint een keer de 'Schrijfwijzer' [Sc] door, met nuttige tips over hoe je een verslag opzet.

1.2.2 Een aantal voor de hand liggende opmerkingen over het praatje

De gouden regel is: maak heldere keuzes over wat je wel en niet kunt vertellen. Wat is het ene centrale punt dat je wilt overbrengen? Bouw daar de rest omheen.

Als je bijvoorbeeld het volledige bewijs van een stelling wilt geven, denk dan goed na of dit in korte tijd wel overkomt bij je publiek. Soms is het duidelijker om iets te vertellen over de context van je stelling (Waarom is het belangrijk?),

of om een schets te geven van je bewijs. Stel jezelf de vraag: 'Als ik in het publiek zat, wat zou er bij mij blijven hangen?'

Reken bij je praatje alleen op de direct parate kennis van je publiek. Een lezer kan zijn 'sluimerende kennis' wakker roepen door nog eens terug te lezen en rustig na te denken. Een luisteraar moet echter direct begrijpen wat er gebeurt, informatie komt maar een keer langs.

Besteed voldoende tijd aan het uitleggen van definities, en geef aansprekende voorbeelden. Onderschat niet hoeveel tijd het kost voordat een nieuw begrip is 'ingedaald' in de hersenpanne(n) van je toehoorders. Een verstandige kreet luidt 'nooit meer dan 3 nieuwe begrippen per college'.

Houd oogcontact met het publiek. Kijk steeds één persoon enkele seconden lang aan, en dan weer een ander. Pauzeren met oogcontact wekt de indruk van zekerheid.

Articuleer duidelijk, spreek voldoende luid, en gebruik intonatie. Houd een rustig tempo aan. Je voelt wellicht de neiging om te snel te spreken, geef hier niet aan toe.

Als je met een beamer of projector werkt, geeft je slide kort de kernpunten aan van je betoog. Gebruik steekwoorden, geen volzinnen. Een volzin op je slide leidt de aandacht af van wat je op dat moment zegt. Maximaal 7 regels met 7 woorden per slide is een goede richtlijn. Behandel niet meer dan een onderwerp per slide. Gebruik letters die voldoende groot zijn om op afstand leesbaar te zijn. Liever niet uitsluitend hoofdletters: een lezer herkent het woordbeeld sneller in kleine letters, omdat dit bekend voorkomt.

Realiseer je dat een schoolbord langzamer is dan powerpoint, wat zo zijn voor- en zijn nadelen heeft. Let op je bordgebruik, bij een groot schoolbord is het vaak handig om eerst de linker, en dan de rechterhelft te gebruiken.

De opmerking over kop en staart van het verslag geldt mutatis mutandis voor een praatje.

Zie voor deze en meer tips de tekst 'Presenteren van wiskunde' [Pr].

1.3 Geen Meden en Perzen

Het doel van presentatie en verslag is om zo helder mogelijk over te brengen wat je te zeggen hebt. Hiertoe zijn alle vreedzame middelen geoorloofd. Negeer in het bijzonder de voorgaande richtlijnen als je daar een goede reden toe hebt.

2 Entropie

We bekijken het begrip 'entropie'. Er volgt een opdracht voor alle groepjes.

2.1 Definitie

Sommige kansmaten geven meer onzekerheid dan andere. Kijk bijvoorbeeld eens naar de volgende drie kansmaten. (Horizontaal staat x, verticaal de kansdichtheid $p(x)$.)

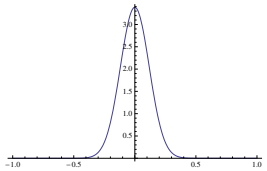

Fig.1: Kansdichtheid (a)

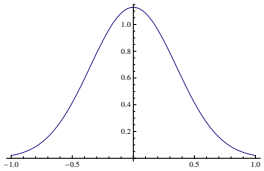

Fig.2: Kansdichtheid (b)

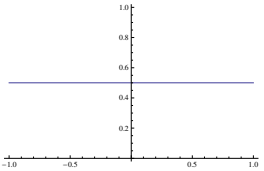

Fig.3: Kansdichtheid (c)

Het is intuïtief duidelijk dat de kansmaat (c) het meeste onzekerheid laat, kansmaat (a) het minste, en dat (b) daar tussenin zit. Voortaan beperken we ons tot kansmaten op eindige kansruimten $(\Omega, \Sigma, \mathbb{P})$, of kortweg $(\Omega, \mathbb{P})$ als $\Sigma = \mathscr{P}(\Omega)$. Aan iedere kansmaat willen we een getal toekennen, dat zijn 'onzekerheid' uitdrukt.

We definiëren de *entropie* van $\mathbb{P}$ als

$$H(\mathbb{P}) := -\sum_{\omega\in\Omega} \mathbb{P}(\omega)\log_2(\mathbb{P}(\omega))\,. \tag{1}$$

Dit ziet er wellicht merkwaardig uit, maar komend half jaar hoop ik jullie ervan te overtuigen dat entropie echt de juiste maat voor het begrip 'onzekerheid' is.

Laten we eerst eens controleren dat de puntmaat $\mathbb{P}(\omega) = \delta(\omega, \omega_0)$, met intuïtief gezien de kleinste 'onzekerheid', ook de minste entropie bezit, en dat de vlakke maat $\mathbb{P}(\omega) = 1/\#\Omega$, die de meeste 'onzekerheid' laat, maximale entropie heeft.

Onderzoek van de functie $\eta(x) := -x\log_2(x)$ levert het volgende plaatje op:

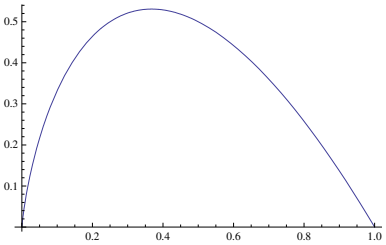

Fig.4: $\eta(x) = -x\log_2(x)$

Merk op dat $\lim_{x\downarrow 0} -x\log_2(x) = 0$. Mocht ergens per ongeluk $0\log_2(0)$ staan, (zoals soms in formule 1), lees daar dan 0. Uit $\eta(x) \geq 0$ blijkt direct dat $H(\mathbb{P}) \geq 0$. Onze entropie is dus positief of 0. De enige manier waarop $H(\mathbb{P})$ gelijk aan nul kan zijn, is door $\eta(\mathbb{P}(\omega))$ nul te kiezen voor alle ω. Dat kan alleen als de kansmaat geconcentreerd is op één punt ω_0. (Waarom?)

We zitten in een kansruimte $(\Omega, \mathbb{P})$ met $\#\Omega = N$. Om de maximale entropie te vinden maximaliseren we $\sum_{i=1}^{N} -p_i \log_2(p_i)$ onder de randvoorwaarde $\sum_{i=1}^{N} p_i = 1$. Deze eis geeft een Lagrange-multiplier. We stellen dus de variatie van

$$S(\{p_i\}, \lambda) = \sum_{i=1}^{N} -p_i \log_2(p_i) + \lambda(\sum_{i=1}^{N} p_i - 1)$$

naar p_i gelijk aan nul, hetgeen $-\ln(2)\lambda = 1 + \log_2(p_i)$ oplevert. Kennelijk zijn alle kansen gelijk, dus $p_i = 1/N$. Dit geeft entropie $\sum_{i=1}^{N} -\frac{1}{N} \log_2(\frac{1}{N}) = \log_2(N)$. (Ik ben slordig geweest met de 'grenzen', maar dat blijkt achteraf gerechtvaardigd. Waarom?)

Kort samengevat: *de entropie is minimaal 0 voor puntmaten, en maximaal* $\log_2(N)$ *voor de vlakke kansmaat. Alle andere maten zitten hier tussenin.* Dat komt wel overeen met je (of in ieder geval mijn) intuïtieve idee van onzekerheid. Merk op dat de entropie (onzekerheid) van de vlakke kansmaat stijgt met het aantal punten. Ook dat lijkt me ook niet onredelijk.

Als $\#\Omega = 2$, wordt een kansverdeling gegeven door één parameter p. Immers, als $\mathbb{P}(\omega_1) = p$, dan $\mathbb{P}(\omega_2) = 1 - p$. Als $\#\Omega = 3$ zijn er 2 parameters p en q. Het volgende plaatje laat $H(p, 1-p)$ en $H(p, q, 1-p-q)$ zien.

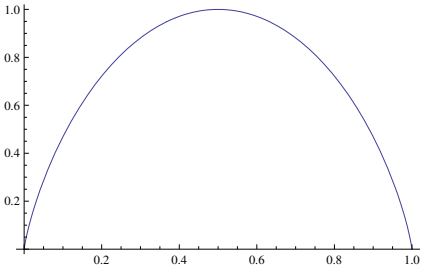

Fig.5: De entropie van de kansverdeling $(p, 1-p)$. Horizontaal staat p, verticaal $H(p, 1-p)$

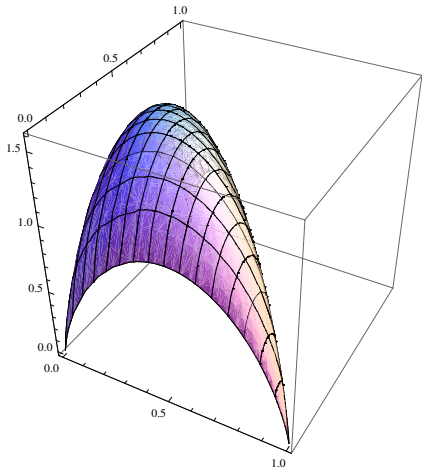

Fig.6: De entropie van de kansverdeling $(p, q, 1-p-q)$. Horizontaal staat (p, q), verticaal $H(p, q, 1-p-q)$

Links zie je dat de entropie maximaal is voor de eerlijke verdeling $(\frac{1}{2}, \frac{1}{2})$, en minimaal voor $(0, 1)$ en $(1, 0)$. Rechts zie je dat de entropie maximaal is voor $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, en minimaal voor $(1, 0, 0)$, $(0, 1, 0)$ en $(0, 0, 1)$. Merk op dat voor $p = 0$ het linker plaatje terugkomt!

2.2 Karakterisering

We pakken het iets formeler aan. Stel we hebben een andere manier F om aan iedere kansmaat $\mathbb{P}$ een onzekerheid $F(\mathbb{P})$ toe te kennen. De volgende stelling noemt een viertal eigenschappen op van het intuïtieve idee 'onzekerheid'. De bewering is, dat iedere 'onzekerheid' F die hieraan voldoet, op een factor na gelijk is aan onze entropie H.

Stelling 1 *Een afbeelding F van de verzameling kansmaten op eindige kansruimten naar $\mathbb{R}$ heet een '*onzekerheidsfunctie*' als deze voldoet aan de volgende 4 eisen:*

- *F hangt niet af van de naamkaartjes die aan je kansruimte hangen. Ofwel: als er een bijectie $b : \Omega \to \Omega'$ bestaat zó dat $\mathbb{P}'(b(\omega)) = \mathbb{P}(\omega)$ voor alle ω in Ω, dan $F(\mathbb{P}) = F(\mathbb{P}')$.*
- *Continuïteit*
- *De onzekerheid wordt groter als je meer punten hebt. Preciezer: definieer $f(N)$ als $F(\mathbb{P}_N)$, waarbij $\mathbb{P}_N$ de kansmaat op N punten is, die aan ieder punt kans $\frac{1}{N}$ toekent. Dan is $N \mapsto f(N)$ een monotoon stijgende functie van N.*
- *Stel, je deelt je kansruimte op in K stukken. Dan is de totale onzekerheid het gewogen gemiddelde van de onzekerheden van ieder stuk, plus de onzekerheid die van de opdeling zelf komt. Preciezer: verdeel Ω in K disjuncte stukken, $\Omega = \bigcup_{i=1}^{K} \Omega_i$. Op ieder stuk Ω_i leeft een voorwaardelijke kansmaat $\mathbb{P}_i$, gedefinieerd door $\mathbb{P}_i(\omega) = \mathbb{P}(\omega)/\mathbb{P}(\Omega_i)$. Ook leeft er natuurlijk een kansmaat op de opdeling $\overline{\Omega} = \{\Omega_i | i = 1 \dots k\}$ zelf, namelijk $\overline{\mathbb{P}}(\Omega_i) = \mathbb{P}(\Omega_i)$. We eisen dan dat*

$$F(\mathbb{P}) = F(\overline{\mathbb{P}}) + \sum_{i=1}^{K} \mathbb{P}(\Omega_i) F(\mathbb{P}_i)\,.$$

  Dit heet ook wel de compositiewet.

Iedere onzekerheidsfunctie F is een positief veelvoud van de entropie H.

Bewijs (uit [Sh]) Formuleer zelf een zinnige versie van de compositiewet als er een i is met $\mathbb{P}(\Omega_i) = 0$. Eerst bewijzen we dat $f(N) = k \log_2(N)$.

Voor alle positieve $N, r \in \mathbb{N}$ bestaat er wel een $s \in \mathbb{N}$ zó dat $2^s \leq N^r < 2^{s+1}$. Dan ook $s \leq r \log_2(N) < s + 1$, zodat

$$\left| \log_2(N) - \frac{s}{r} \right| < \frac{1}{r}\,. \tag{2}$$

Verdeel een verzameling met xy punten in x verzamelingen van y punten. Met de compositiewet kun je dan inzien dat $f(xy) = f(x) + f(y)$ voor alle x en y.

(Hoe?) In het bijzonder is $f(2^s) = sf(2)$ en $f(N^r) = rf(N)$. Omdat $N \mapsto f(N)$ monotoon is, volgt uit $2^s \leq N^r < 2^{s+1}$ dat $f(2^s) \leq f(N^r) < f(2^{s+1})$. Derhalve geldt $sf(2) \leq rf(N) < (s+1)f(2)$, zodat in het bijzonder

$$\left| \frac{f(N)}{f(2)} - \frac{s}{r} \right| < \frac{1}{r}\,. \tag{3}$$

Tezamen met vergelijking 2 levert dit $|\frac{f(N)}{f(2)} - \log_2(N)| < \frac{2}{r}$. Dit geldt voor alle r, dus moet wel gelden dat $f(N) = f(2)\log_2(N)$. Met $k = f(2)$ hebben we dus inderdaad $f(N) = k\log_2(N)$.

Vervolgens bekijken we kansverdelingen met slechts rationale kansen, dat wil zeggen met $\mathbb{P}(\omega) \in \mathbb{Q}$. Breng alle kansen op gelijke noemer N. We hebben dan $\mathbb{P}(\omega) = n_\omega/N$. Bekijk een verzameling van N punten, met daarop de vlakke kansmaat die ieder punt gelijke kans toedicht. De onzekerheid is $f(N) = k\log_2(N)$. Verdeel nu de punten in $\#\Omega$ groepjes, voor iedere $\omega \in \Omega$ kies je een groepje met n_ω punten. De kansmaat die op de n_ω punten leeft is weer vlak, met onzekerheid $k\log_2(n_\omega)$. De kansmaat die op de verdeling zelf leeft is natuurlijk weer $\mathbb{P}$, omdat de kans om bij de n_ω punten in 'groepje ω' te belanden gelijk is aan n_ω/N. Dit geeft onzekerheid $F(\mathbb{P})$. De compositiewet leert ons dat

$$k\log_2(N) = F(\mathbb{P}) + \sum_{\omega\in\Omega} \frac{n_\omega}{N} k\log_2(n_\omega)\,,$$

ofwel $F(\mathbb{P}) = k\sum_{\omega\in\Omega} -\frac{n_\omega}{N}\log_2(\frac{n_\omega}{N})$.

We weten nu dat $F(\mathbb{P}) = kH(\mathbb{P})$ als $\mathbb{P}$ louter rationale waarden aanneemt. Maar F is continu, en je kunt iedere kansmaat willekeurig dicht benaderen met kansmaten van het rationale slag. Het moet dus wel zo zijn dat $F = kH$ op *alle* kansmaten. □

Als je vindt dat 'onzekerheid' de vier bovenstaande eigenschappen heeft, dan is entropie voor jou de enige juiste maat van onzekerheid!

Entropie kun je ook zien als een hoeveelheid 'informatie'. Dat gaat als volgt. Je doet een 'trekking' uit een kansruimte $(\Omega, \mathbb{P})$, en je neemt de uitkomst ω waar. Hoeveel informatie levert dit op?

Stel de kansmaat is geconcentreerd op ω_0. Dan wist je van tevoren al dat je $\omega = \omega_0$ zou vinden. De trekking levert in het geheel geen informatie op: $H = 0$.

Hoe meer onzekerheid de kansmaat bevat, hoe meer informatie je krijgt bij het zien van de uitkomst. Bij een vlakke kansmaat krijg je het meeste informatie.

Stel de kansmaat is vlak. Als $\#\Omega = 2$ komt het geven van de uitkomst neer op het geven van een 0 of een 1, waarvan je eerder niets wist. Dit is één *bit* aan informatie. Als $\#\Omega = 2^n$, hoeveel informatie krijg je dan? Een trekking uit Ω is equivalent met een n-tal nullen of enen, waarvan je eerder niets wist. Ofwel n bits aan informatie. Dit klopt mooi met $H(\mathbb{P}_N) = \log_2(N)$ als je $N = 2^n$ invult.

2.3 Opdracht voor allen: een stelling van Shannon

Het begrip 'entropie' treedt duidelijk naar voren als 'hoeveelheid informatie' wanneer we proberen informatie zo kort mogelijk te coderen. Dat gaan we zo dadelijk doen. Ook volgen in dit stukje de opdrachten die voor iedereen bestemd zijn.

Laat nu $\Omega = \{a, b\}$ een verzameling met twee elementen zijn. Een element $\omega \in \Omega$ heet een *letter*, en Ω heet het *alfabet*. Een element $\vec{\omega} \in \Omega^N$ heet een *woord* van lengte N. Het woord $\vec{\omega} = (b, a, b, a, b, a)$ van lengte 6 wordt ook wel genoteerd als $\vec{\omega} = bababa$.

Stel dat $\mathbb{P}$ een kansverdeling op Ω is. De kansverdeling $\mathbb{P}$ op de letters geeft een kansverdeling (die ik weer $\mathbb{P}$ noem) op de woorden, door de letters van een woord onafhankelijk te kiezen. Als $\vec{\omega} = \omega_1\omega_2 \ldots \omega_N$, dan $\mathbb{P}(\vec{\omega}) = \mathbb{P}(\omega_1) \cdot \mathbb{P}(\omega_2) \cdot \ldots \cdot \mathbb{P}(\omega_N)$.

$\mathcal{R}$ is de verzameling van alle eindige rijtjes nullen en enen,

$$\mathcal{R} = \{0, 1\} \cup \{0, 1\}^2 \cup \{0, 1\}^3 \ldots .$$

L is de afbeelding $\mathcal{R} \to \mathbb{N}$ die aan elk rijtje zijn lengte toekent. Als een rijtje lengte l heeft, zeggen we dat hij l *bits* aan geheugenruimte inneemt.

Een *code* op N letters is een injectieve afbeelding $c : \Omega^N \to \mathcal{R}$. De verzameling codes op N letters geven we aan met $\mathcal{C}_N$.

Een code is iets dat woorden vervangt door rijtjes nullen en enen, zó dat je uit het rijtje het oorspronkelijke woord weer kunt terugvinden. (Dat is de injectiviteit.) We zijn geïnteresseerd in codes die gemiddeld zo kort mogelijke rijtjes opleveren. Dat is bijvoorbeeld handig als je data zo snel mogelijk via een kabel met kleine bandbreedte wilt versturen, of als je gegevens zo efficiënt mogelijk op een harde schijf wilt opslaan.

De verwachtingswaarde van de lengte van een gecodeerd woord is $\mathbb{E}(L(c(\vec{\omega}))) = \sum_{\vec{\omega} \in \Omega^N} \mathbb{P}(\vec{\omega}) L(c(\vec{\omega}))$. Het gemiddelde aantal bits per letter dat je nodig hebt bij code c is dus $\mathbb{E}(\frac{1}{N} L(c(\vec{\omega})))$. We vragen ons af wat, voor grote N, het kleinste aantal bits per letter is dat een code kan opleveren.

De volgende stelling van Shannon (die wel bekend staat als 'Shannon's noiseless channel coding theorem') geeft het antwoord.

Stelling 2 *Als woorden van twee letters gecodeerd worden in rijtjes nullen en enen, dan levert de best mogelijke code asymptotisch H bits per letter op.*

$$\lim_{N \to \infty} \inf_{\mathcal{C}_N} \mathbb{E}(\tfrac{1}{N} L(c(\vec{\omega}))) = H \,,$$

waarbij H de entropie van $\mathbb{P}$ is: $H = -\mathbb{P}(a) \log_2(\mathbb{P}(a)) - \mathbb{P}(b) \log_2(\mathbb{P}(b))$.

Opgave 1 Bewijs de stelling.

Elk juist bewijs is een goed antwoord, maar je zou bijvoorbeeld de volgende lijn kunnen volgen.

De zwakke wet van de grote aantallen zegt dat, als N groot is, de frequentie $f_a(\vec{\omega}) := \frac{1}{N}\sum_{i=1}^{N}\delta(\omega_i, a)$ waarmee de letter a voorkomt in het woord $\vec{\omega}$, ongeveer gelijk zal zijn aan $\mathbb{P}(\omega)$. Evengoed zal $f_b(\vec{\omega}) \sim \mathbb{P}(b)$.

Er zijn 2^N mogelijke woorden in Ω^N. Verdeel deze woorden in *typische* en *atypische* woorden. Een woord $\vec{\omega}$ heet *ε-typisch* als a en b ongeveer met frequentie $\mathbb{P}(a)$ en $\mathbb{P}(b)$ voorkomen, waarbij ε de tolerantie in 'ongeveer' aangeeft. Noteer de verzameling ε-typische woorden als $T_{\varepsilon,N}$:

$$T_{\varepsilon,N} := \{\, \vec{\omega} \in \Omega^N \mid |f_a(\vec{\omega}) - \mathbb{P}(a)| < \varepsilon \text{ en } |f_b(\vec{\omega}) - \mathbb{P}(b)| < \varepsilon \,\}\,.$$

Voor vaste ε en grote N zullen bijna alle woorden ε-typisch zijn. Maar er zijn niet zoveel ε-typische woorden: ongeveer $\binom{N}{\mathbb{P}(a)N}$, (waarom?), wat véél minder is dan 2^N. Je kunt dit gebruiken om een slimme code te construeren: ken eerst rijtjes toe aan alle ε-typische woorden. dat kan met rijtjes korter dan $\log_2(\#T_{\varepsilon,N})$ (waarom?). Nummer dan de atypische woorden. De verwachte lengte $\mathbb{E}(L(c(\vec{\omega})))$ wordt vooral bepaald door de typische woorden, en zal dus $\sim \log_2(\#T_{\varepsilon,N})$ zijn.

- Bewijs Chebyshev's ongelijkheid. Bewijs daarmee de zwakke wet van de grote aantallen. Bewijs daarmee dat $\mathbb{P}(T_{\varepsilon,N}) \geq 1 - \frac{1}{N\varepsilon^2}$.

- Het hart van het bewijs. Laat zien dat $\#T_{\varepsilon,N} = \sum_{\mathcal{S}} \binom{N}{f_a N}$, waar de som loopt over de toegestane frequenties: $\mathcal{S} := \{f_a | f_a \in (\mathbb{P}(a) - \varepsilon, \mathbb{P}(a) + \varepsilon) \cap \frac{1}{N}\mathbb{N}\}$. Bewijs (een zwakke vorm van) de formule van Stirling,

  $$N\log_2(N) - \log_2(e)(N+1) \leq \log_2(N!) \leq (N+1)\log_2(N+1) - \log_2(e)N\,.$$

  Bewijs de volgende uitspraak. Er bestaat een keuze $N \mapsto \varepsilon(N)$ zó dat ten eerste $\lim_{N\to\infty}\varepsilon(N) = 0$, ten tweede $\lim_{N\to\infty}\mathbb{P}(T_{\varepsilon,N}) = 1$, en ten derde: $\exists\,\alpha > 0$ zó dat de ongelijkheid $NH - \alpha N\varepsilon \leq \log_2(\#T_{\varepsilon,N}) \leq NH + \alpha N\varepsilon$ geldt voor $\varepsilon = \varepsilon(N)$ en N groot genoeg.

- Construeer voor iedere N een 'slimme' code c_N. Laat zien dat voor alle $\tau > 0$, er een N_0 bestaat zó dat $\mathbb{E}(\frac{1}{N}L(c_N(\vec{\omega}))) \leq H + \tau$ voor alle $N \geq N_0$. Compressie kan dus in H bits per letter.

- Nu nog laten zien dat het niet beter kan. Bewijs dat voor een optimale code, $\mathbb{E}(L(c(\vec{\omega}))) \geq N(H-\tau)\mathbb{P}([\vec{\omega} \in T_{\varepsilon,N}, L(\vec{\omega}) \geq N(H-\tau)])$. Deze kans gaat naar 1 voor N groot en $\tau \geq 0$. (Hint: laat zien dat $\mathbb{P}([\vec{\omega} \in T_{\varepsilon,N}, L(\vec{\omega}) < N(H-\tau)]) \to 0$, door te tonen dat er een $\beta > 0$ bestaat zó dat $\mathbb{P}(\vec{\omega}) \leq 2^{-NH}2^{\beta\varepsilon N}$ voor $\vec{\omega} \in T_{\varepsilon,N}$ en N groot.)

Opgave 2 Wat verandert er in de stelling van Shannon als je niet in rijtjes nullen en enen codeert, maar in $\{0,1,2\}$? Of, algemener, in Ω' met $\#\Omega' < \infty$? Deze verzameling Ω' heet het *doelalfabet*.

Opgave 3 Wat verandert er in de stelling van Shannon als je niet werkt met een alfabet van 2 letters, $\Omega = \{a, b\}$, maar met 3 letters $\Omega = \{a, b, c\}$ of, algemener, in Ω met $\#\Omega < \infty$?

De naamgevingen 'letter' en 'woord' vallen nu letterlijk te nemen. Stel je bijvoorbeeld voor dat Ω het alfabet is, verenigd met de punt en de spatie. Dit is eindig, $\#\Omega = 29$. Een woord van 5 letters (b.v. '*woord*') is dan een element van Ω^5. Als een boek N letters, spaties en punten bevat, is het een woord van lengte N.

Opgave 4 'De ontdekking van de hemel' van H. Mulisch telt 905 bladzijden. Boze tongen beweren dat het ook wel in de helft van dat aantal had gekund. Spreken deze boze tongen de waarheid?

Verslag Formuleer en bewijs een versie van de bovenstaande stelling die geldt voor een eindig alfabet Ω, dat gecodeerd wordt als een rijtje symbolen in een eindig doelalfabet Ω'. Als je een bewijs hebt voor het algemene geval is een apart bewijs voor het speciale geval (opdracht 1) natuurlijk overbodig. Beantwoord wel de vraag in opgave 4. Maak dit een hoofdstuk voor het uiteindelijke verslag. Dit wil zeggen dat inleiding en conclusie overbodig zijn, maar dat je er wel een lopend verhaal van maakt, dat begrijpelijk is voor een lezer van buitenaf.

3 Voor allen één opdracht

Kies met je groepje een van de volgende opdrachten om de rest van het semester aan te werken. Alle opdrachten vereisen wat basiskennis van kansrekening. De opdrachten 'ruis' en 'een coderingsstelling voor Markovketens' blijven misschien het dichtst bij het voorafgaande.

3.1 Vul zelf in

Het is vaak leuker om je eigen vragen te beantwoorden dan die van iemand anders. Heb je zelf een idee? Stel het voor als opdracht!

Er is een drietal caveats:

- Ik moet genoeg van het onderwerp weten om jullie te kunnen begeleiden.
- De vraag moet een beetje 'op niveau' zijn; niet te eenvoudig, maar ook niet zó moeilijk dat er niets meer over te zeggen valt.
- Het moet iets van doen hebben met 'informatietheorie' of 'entropie'. Laat dit laatste je vooral niet afschrikken, je zou verbaasd zijn hoeveel wiskunde daar zijdelings mee verbonden is!

Verder kunnen de andere opdrachten, in overleg, altijd worden aangepast. Ook als je halverwege de rit een interessant zijpad ontdekt.

3.2 Large deviations

Stel je werpt een dobbelsteen N maal, en je turft de frequenties f_1 t/m f_6 van iedere uitkomst. Voor grote N verwacht je natuurlijk dat $(f_1, f_2, f_3, f_4, f_5, f_6)$ dicht bij $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ komt te liggen. Je zou dit zelfs kunnen zien als de betekenis van de uitspraak: "De dobbelsteen wordt beschreven door de kansmaat $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$. "

Inderdaad heb je zoiets als de zwakke wet van de grote aantallen, die dit wat preciezer maakt. Wat betekent dichtbij? Laten we zeggen dat de afstand tussen de kansmaten $\mathbb{P}$ en $\tilde{\mathbb{P}}$ gelijk is aan

$$d(\mathbb{P}, \tilde{\mathbb{P}}) := \mathbf{max}_{i=1,\ldots,6} |p_i - \tilde{p}_i| \,.$$

De zwakke wet van de grote aantallen zegt dan dat, voor alle $\epsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}([d(\mathbb{P}_f, \mathbb{P}) \geq \epsilon]) = 0 \,,$$

waar $\mathbb{P}_f$ de kansmaat is met als dichtheid $(f_1, f_2, f_3, f_4, f_5, f_6)$.

Kort gezegd: de kans dat de gevonden frequentie nog steeds verder dan ϵ verwijderd is van wat hij zou moeten zijn, wordt klein als N groot wordt.

Ok, maar hoe snel wordt dit klein? Dat is voor toepassingen nogal van belang; als de frequentie zich de eerste duizend worpen niets van de kansmaat aan zou trekken, zou de hele kanstheorie niet meer zijn dan een leuke hobby voor wiskundigen.

Gelukkig blijken de frequenties erg snel te convergeren. Exponentieel snel zelfs:

$$\mathbb{P}([d(\mathbb{P}_f, \mathbb{P}) \geq \epsilon]) \simeq e^{-nH(\epsilon)}$$

met $H(\epsilon) > 0$. Dit soort resultaten staat bekend onder de naam 'Large Deviation-results'. De coëfficiënt H blijkt een soort 'relatieve entropie' te zijn.

Als de wet van de grote aantallen zegt waarom kansrekening überhaupt betekenis heeft, dan zeggen large deviation results waarom kansrekening nut heeft. In zekere zin wordt het nut van kansrekening voor de dobbelaar dus bepaald door een vorm van entropie!

Zoek uit wat relatieve entropie is. (Deze heet ook wel 'Kullback-Leibler divergentie'. Kijk uit! Conditionele entropie is weer iets anders, en Shannon [Sh] geeft niet het juiste antwoord.) Probeer er een beetje gevoel voor te krijgen, bijvoorbeeld door wat eigenschappen te bewijzen, of misschien een Shannon-achtige stelling.

Zeg dan eens iets zinnigs over een 'Large Deviation' resultaat. (Het eerste hoofdstuk van [El] is misschien van enig nut.)

3.3 Maak je eigen datacompressor

Implementeer je favoriete datacompressiemethoden, en zeg daar iets intelligents over. (Vergelijk ze bijvoorbeeld op een zinnige manier met elkaar, met commerciële programma's, en met de coderingsstelling van Shannon.)

Misschien heb je geen lievelingscompressiemethode, en wil je jezelf uit deze weinig benijdenswaardige toestand verheffen. Je zou dan eens kunnen kijken naar mijn favorieten:

- De Huffman-code is wereldberoemd in kleine kring. Zie bijvoorbeeld [MK]. Op de website [Au] staat er zelfs een animatie van!
- Het LZ77-algoritme. Het originele artikel van Ziv en Lempel uit 1977, [ZL], is kort, maar wel zo bondig. (GIF en ZIP werken ongeveer volgens dit principe.)
- Voor bijvoorbeeld meetdata en plaatjes: wavelets en de 'lifting scheme'. Zie bijvoorbeeld [SS]. Bedenk dan wel even hoe je hier een compressiemethode van maakt. (Kijk voor de lol eens op [Wi] hoe JPEG2000 werkt!)

Doe deze opdracht alleen als je een beetje vertrouwen in je eigen 'programming-skills' hebt. De kans is groot dat ze beter zijn dan de mijne, en dat ik daarin derhalve niet veel hulp kan bieden.

3.4 Een coderingsstelling voor Markovketens

De coderingsstelling van Shannon zegt het volgende.

Je hebt r 'letters' $a(1)$ t/m $a(r)$. Stel dat een informatiebron een reeks letters uitspuugt, waarbij op iedere positie de kans dat er $a(i)$ staat gelijk is aan p_i, *onafhankelijk van de letters op overige posities.*

Dan is de efficiëntie van de best mogelijke datacompressie gelijk aan de entropie $H = -\sum_{i=1}^{r} p_i \log_2(p_i)$ van deze verdeling.

Deze stelling is altijd waar, maar meestal nutteloos. (Vertel dat s.v.p. niet aan de mensen die deze opdracht niet doen.) Als je een reeks data hebt, zijn er bijna altijd wel correlaties tussen de opeenvolgende tekens.

Neem bijvoorbeeld een boek. Het feit dat de 'E' vaker voorkomt dan de 'X' kun je gebruiken om woorden efficiënter te coderen. Dat is de portee van bovenstaande stelling. Maar het is zeker niet zo dat opeenvolgende letters onafhankelijk zijn! Als ergens een 'Q' staat, is de kans aanzienlijk dat de volgende letter een 'U' is. (In het Nederlands zelfs 100%.) De correlaties kun je gebruiken om nog beter te coderen dan de stelling van Shannon zou voorspellen! Maar wat is nu het beste dat je kunt bereiken?

Een aardig model voor een boek is wat dat betreft een Markovketen; de kans op de volgende letter hangt af van de vorige k letters, maar niet van de letters daarvoor. We nemen een stationaire kansverdeling.

In plaats van de entropie H_1 van de kansverdeling op één enkele letter, kijken we nu naar de gemiddelde entropie $\frac{1}{n}H_n$ van de eerste n letters.

Je zou kunnen uitzoeken waarom de 'groei van de entropie' $H := \lim_{n\to\infty} \frac{1}{n}H_n$ bestaat, en vooral ook waarom deze echt de efficiëntie van de optimale codering geeft. Misschien kun je ook een expliciete formule voor H in termen van de overgangskansen geven.

De grootheid H heet 'entropy rate' in de informatietheorie, en 'Kolmogorov-Sinai-entropie' in de wondere wereld van dynamische systemen.

Het originele artikel [Sh] uit 1948 is nog steeds een erg heldere uiteenzetting. (Om te bewijzen dat de limiet bestaat zou je daarnaast bijvoorbeeld eens in hoofdstuk 2 van [Bi] kunnen bladeren.)

3.5 Ruis

Als je informatie door een kanaal stuurt, is er eigenlijk altijd wel ruis. Stel je een draadje voor waar je een 1 instopt, en waar met kans $1-\varepsilon$ ook een 1 uitkomt, maar met kans ε een 0. Is dit draadje daarmee van nul en generlei waarde voor het versturen van informatie?

Neen. Je kunt die ruis terugdringen door bijvoorbeeld redundant informatie te versturen: in plaats van één 1 stuur je drie enen het kanaal in, en aan de uitgang bepaalt de meerderheid van de bits wat de uitkomst wordt. De kans dat minstens twee maal een 1 in een 0 verandert is $\mathcal{O}(\varepsilon^2)$ i.p.v. $\mathcal{O}(\varepsilon)$. Je kunt dus ruis terugdringen ten koste van de snelheid waarmee een signaal verzonden wordt. (Drie enen duurt langer.)

Zo op het eerste gezicht zou je verwachten, dat bij het verder en verder terugdringen van de ruis, de snelheid van je signaal verder en verder afneemt.

Wonderlijk genoeg blijkt dit niet het geval! Er blijkt een kritieke 'capaciteit' C te bestaan, zó dat de ruis willekeurig ver kan worden teruggedrongen terwijl toch de snelheid niet onder C komt.

Zoek eens uit hoe dit zit. Het originele artikel [Sh] van Shannon is misschien geen slechte plek om te beginnen. (Shannon werkt met een Markov-model voor de informatiebron, zie opdracht 3.4, zodat de enen en nullen niet per se onafhankelijk zijn. Je zou jezelf, in ieder geval in het begin, kunnen beperken tot bronnen waarbij dit wel het geval is.)

3.6 Statistische mechanica

Laten we onze maat voor de hoeveelheid informatie van een kansverdeling 'informatie-entropie' noemen. In de thermodynamica bestond al veel langer een 'fysische entropie', en er is natuurlijk een relatie tussen deze twee begrippen.

Sterker nog, eind jaren '50 van de vorige eeuw heeft Edwin Jaynes gepoogd de statistische fysica op het concept van 'informatie-entropie' te bouwen. Het idee is ruwweg als volgt.

Een systeem kan in toestand ω_1 t/m ω_n verkeren, waar toestand ω_i energie E_i heeft.

Stel je weet helemaal niets. Wat is dan de kansmaat die je moet gebruiken om het systeem te beschrijven? Dat is intuïtief duidelijk: $p_i = \frac{1}{n}$ voor alle i. Dit is de kansmaat met maximale entropie.

Stel je weet de gemiddelde energie $\langle E \rangle$ van het systeem, maar verder niets. Wat doe je dan? Het antwoord van Jaynes is: zoek onder de kansmaten die gemiddelde energie $\langle E \rangle$ opleveren weer degene met maximale informatie-entropie.

Dat blijkt de Bolzmann-verdeling te zijn, en je kunt dit als fundering gebruiken om de hele statistische fysica op te bouwen.

Het resultaat is elegant, maar niet onomstreden. Laat eens zien wat de mogelijkheden van deze manier van kijken zijn: probeer bijvoorbeeld eens om, op deze manier, de machinerie van de statistische fysica op te bouwen, (temperatuur, toestandssom, etc), tot en met een interessante toepassing.

Wat vind jij van deze aanpak?

Als je geïnteresseerd bent in de 'filosofische achtergrond' van de rekenpartij, zou je eventueel ook eens naar het originele stuk [Ja] van Jaynes kunnen kijken.

Als je deze opdracht kiest, is het handig als je al eens wat statistische fysica hebt gezien. Een mooie tekst die bovenstaande lijn volgt, (en wonderlijk genoeg doorspekt is met opbeurende Prediker-citaten,) is [Ve].

3.7 Chaos en Kolmogorov-Sinai-entropie

Een eenduidige definitie van het begrip chaos in dynamische systemen is nog niet zo makkelijk te geven. Een manier om 'chaos' wiskundig te vangen, is door te zeggen dat een chaotisch systeem positieve 'Kolmogorov-Sinai-entropie' heeft.

Zeg hier iets zinnigs over. Zoek bijvoorbeeld uit wat de definitie is, en waarom de Kolmogorov-Sinai-entropie hetzelfde is voor dynamische systemen die isomorf zijn. (En wat het betekent dat twee dynamische systemen isomorf zijn.) Voordat je hieraan begint, is het raadzaam om vertrouwd te geraken met de ergodenstelling. (Het bewijs nazoeken voert wat ver, maar zorg dat je de uitspraak begrijpt. Bewijs eventueel een wat 'lichtere' versie van de stelling, waarbij de aannamen sterker zijn of het resultaat zwakker is.)

Probeer eventueel deze entropie eens uit te rekenen in een eenvoudig geval. De stelling van Kolmogorov en Sinai helpt je daarbij.

Ik raad je aan om deze opdracht alleen te doen als je vertrouwd bent met maten, Σ-algebras en Lebesgue-integralen. Een goede referentie is [Bi].

3.8 Quantumentropie

Quantummechanica valt niet op te vatten als een 'klassieke' kanstheorie. (Althans, de bezwaren hiertegen zijn overweldigend.) Hiervoor heb je een *quantumkanstheorie* nodig. Deze werkt met C^*-algebras in plaats van Σ-algebras, Hermitische operatoren in plaats van stochasten, en met quantumtoestanden in plaats van kansmaten.

Zoek uit wat een quantumkansruimte is, en overtuig jezelf ervan dat de 'klassieke' kansruimte, zoals je die kent uit het vak 'kansrekening', hiervan een speciaal geval is. Ga na dat de quantummechanica zoals je die kent, een ander speciaal geval is. Maak jezelf vertrouwd met het begrip 'volledig positieve operatie', en vraag jezelf af of dit de juiste generalisatie is van het begrip 'operatie' zoals dat in de kansrekening bestaat.

Beperk jezelf bij dit alles tot eindigdimensionale C^*-algebras. (Voor een quantumsysteem betekent dit dat je een eindigdimensionale Hilbertruimte hebt, denk aan een spin-$\frac{1}{2}$-systeem.) Kijk bijvoorbeeld in [Ma].

Net zoals je een kansmaat een entropie kunt geven, kun je aan een quantumtoestand een quantumentropie toekennen. Deze quantumentropie (ook wel *von Neumann-entropie* naar de uitvinder) valt, net als de Shannon-entropie, te karakteriseren aan de hand van een klein aantal redelijke eisen, à la stelling 1.

Een belangrijke eigenschap van Shannon-entropie (of eigenlijk van relatieve entropie) is monotoniciteit onder operaties van kansruimten. Hier is ook een quantumversie van, die zegt dat de von Neumann-entropie (of eigenlijk relatieve entropie) monotoon is onder volledig positieve operaties.

Zie bijvoorbeeld [OP]. Ook [NC] valt erg aan te raden, diep doch licht verteerbaar. Het bevat een nogal spectaculaire toepassing van von Neumann-entropie: het bewijs dat quantumcryptografie 100% veilig is!

(Voor een aanloop die wat meer via de thermodynamica loopt zou je eens in [vN] kunnen bladeren. Merk op dat von Neumann twee volledig positieve operaties kent. Welke?)

Ik raad je aan deze opdracht alleen te doen als je je een beetje op je gemak voelt met quantumfysica. Het is niet makkelijk in te schatten, maar ik denk dat dit een van de moeilijker opdrachten is. Wie neemt de handschoen op?!

Referenties

[Au] `www.cs.auckland.ac.nz/software/AlgAnim/huffman.html`

[Bi] P. Billingsley, *Ergodic Theory and Information*, Wiley, (1965).

[El] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag (1985).

[Ja] E.T. Jaynes, *Information Theory and Statistical Mechanics*, The Physical Review, vol. 106, No. 4, 620–630, (1957).

[Ma] J.D.M. Maassen, *Quantum Probability, Quantum Information Theory and Quantum Computing*, KUN (2004).
`www.math.ru.nl/~maassen/lectures/qpqiqc.pdf`

[MK] D. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press (2003).
`www.inference.phy.cam.ac.uk/itprnn/book.html`

[vN] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Berlijn, Springer, (1932).
Engelse vertaling van R. T. Beyer, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, (1955).

[OP] M. Ohya, D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, (1993).

[NC] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, (2000).

[Pr] J. Daemen, M. Ruijgrok, *Presenteren van wiskunde*,
`www.math.uu.nl/people/ruijgrok/overdragen/Overdragen.pdf`

[Sc] *De Schrijfwijzer*, `www.math.uu.nl/Onderwijs/Schrijfwijzer/SchrijvenInDeWiskunde.pdf`

[Sh] C.E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal, vol. 27, p. 379–423 en 623–656, (1948) .

[SS] W. Swelders en P. Schröder, *Building Your Own Wavelets at Home.*

[Te] G. Tel, *Cryptografie, beveiliging van de digitale maatschappij*, Addison-Wesley, (2002). `www.cs.uu.nl/docs/vakken/cry/index.htm`

[Ve] G. Vertogen, *Een inleiding tot de statistische mechanica*, KUN (2003).

[Wi] `en.wikipedia.org/wiki/JPEG_2000`

[ZL] J. Ziv, A. Lempel, *A Universal Algorithm for Sequential Data Compression*, IEEE Transactions on Information Theory, vol. IT-23, No.3, p. 337–343 (1977).