Wiskunde in je vingers

Wiskunde in je vingers

Een oppepcursus voor liefhebbers en andere freaks

Joost Hulshof en Ronald Meester

VU University Press, Amsterdam

VU University Press
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

www.vuuniversitypress.com
info@vuuitgeverij.nl

Fotografie omslag: Peter Valckx, Amsterdam
Vormgeving omslag: Titus Schulz, Arnhem
Eindredactie: Yda Smets, Den Haag

ISBN 978 90 8659 715 4
NUR 918

De auteursinkomsten uit dit boek komen ten goede aan Het Orkest van de Achttiende Eeuw en de Stichting Communication Middle East (COME):

`http://www.orchestra18c.com` `http://www.stichting-come.nl`

VOORWOORD

In dit boek zetten we een aantal basisbegrippen uit de wiskunde in een wat ander daglicht dan gebruikelijk. We zullen daarbij een poging doen om doorlopende leerlijnen te schetsen die beginnen bij rekenen op basisschoolniveau en eindigen bij onderwerpen die in het voortgezet en hoger onderwijs aan bod komen. Van breukrekenen via letterrekenen naar differentiaalrekening, van tellen via faculteiten naar combinatoriek en kansrekening, van oppervlakte als lengte keer breedte naar de integraalrekening en moderne maattheorie, van de wortel uit 2 naar de getallen e en π, van sinterklaaslootjes naar de wet van de grote aantallen, van kleine naar extreem grote getallen, van eindig naar oneindig en meer, met dwarsverbanden tussen de diverse deelgebieden van de wiskunde, en dat alles gelardeerd met wat persoonlijk beschouwingen over wat wel en wat niet waar is. Met korte en lange zinnen.

Ons boek is losjes gebaseerd op twee nascholingscursussen die we op de Vrije Universiteit aan een groep enthousiaste docenten hebben gegeven. Vakinhoudelijke nascholing is niet alleen belangrijk, het is ook gewoon leuk. En leuk is wat we hopen dat dit boek zal zijn voor iedereen met belangstelling voor wiskunde. Weggezakte wiskunde van de middelbare school is de voorkennis die we van de lezer verwachten, en voor de meeste stof is niet meer nodig dan wat vertrouwdheid met getallen en getalrepresentaties. Veel onderwerpen betreffen gewone schoolstof in het voortgezet onderwijs, en sluiten zo hopelijk aan bij de belevingswereld van de docent, van zijn of haar leerlingen en van de lezer die misschien niet meer precies weet hoe het allemaal ook al weer zat met die wiskunde van school. Wij hopen dat de invalshoeken die we daarbij kiezen inspirerend zijn voor al deze doelgroepen. We denken dat docent en leerling prima samen met deze stof aan de slag kunnen gedurende het hele schoolcurriculum.

Waar gaan we het over hebben? Sommige onderwerpen in dit boek hebben we nadrukkelijk gekozen, andere werden ons door de cursisten aangedragen. Centraal staat e, het getal van Euler. Waarom is e zo bijzonder en waarom zou je eigenlijk alleen met dit grondtal willen werken als je aan het machtsverheffen slaat? En waarom speelt e zo'n interessante rol in de kansrekening?

We laten de lezer het antwoord ontdekken, rechtstreeks vanuit een rente-op-rentecontext. Deze maken we zo expliciet mogelijk, op basis van een getalbegrip dat we opbouwen vanuit het welbekende pizzamodel voor breuken in de rekenles op de basisschool, en met als uitgangspunt de happende staartdeling waarmee we die breuken als tienvingerigen decimaal representeren, met zich herhalende ontwikkelingen als $0,042042042042\ldots = 0,\underline{042}$.

Het 'eerste' getal waarbij zo'n repeterende ontwikkeling misgaat is niet e maar de wortel uit 2. Dat rijmt goed met de introductie van een systematisch aanpak voor worteltrekken die we aan de introductie van e vooraf laten gaan. Niet met de cijferende aanpak die ooit op school door de bovenmeester aan de betere leerlingen werd uitgelegd, maar met de methode van Newton. Newton geeft ons een algoritme dat duizenden jaren terug al werd gebruikt, lang voor zijn differentiaalrekening.

Die differentiaalrekening introduceren we door middel van het begrip raaklijn aan de grafiek bij simpele verbanden gegeven door wat we sinds Leibniz meestal functies noemen. We gaan in detail in op de vraag waarom Newtons methode zo goed is, bijvoorbeeld bij het benaderen van $\sqrt{2}$, meestal het eerste getal waarvan je leert dat het geen breuk is. En tja, dat is ook meteen het enige getal waarvan we het niet-rationaal zijn in dit boek kort bespreken.

In dit boek dus geen getaltheorie. Tegenover (of naast) deze keuze staat een tweede keuze, en die komt er in het kort op neer dat we het limietbegrip zo veel mogelijk willen vermijden. Dus aanvankelijk ook geen analyse. Veel te moeilijk. We kiezen ervoor de vanzelfsprekendheid die we zien in de representatie van de reële getallen door middel van doorlopende decimale ontwikkelingen niet in twijfel te trekken (nou ja, even dan), maar juist door te trekken via een lang doorlopende leerlijn rechtstreeks het 'hoger' onderwijs in.

Dat betekent dat we ons in de differentiaalrekening van Newton eerst zo veel mogelijk beperken tot formules met de spreekwoordelijke x die je wil weten en de a, die je naar men soms zegt niet kunt weten, en voor de rest alleen maar plussen, minnen, keer en gedeeld door. Kortom, de rationale expressies uit de algebra, met de algebraïsche operaties die daarbij horen. De limieten komen later wel.

De didactische insteek hier is dat puntjes die we accepteren in de decimale ontwikkeling

$$\frac{42}{999} = 0{,}\underline{042} = 0{,}042042\ldots$$

niet echt anders zijn dan de puntjes in expressies als

$$1 + x + x^2 + x^3 + x^4 + \cdots ,$$

en die puntjes zijn weer niet anders dan de puntjes in

$$1 + x + \frac{x^2}{2} + \frac{x^3}{3 \times 2} + \frac{x^4}{4 \times 3 \times 2} + \cdots ,$$

of in 2,718281828...

Als je maar weet wat er staat, op wat voor manier dan ook, modulo details waarover vanaf het allereerste begin niet moeilijk gedaan hoeft te worden, als je maar precies bent. Eulers e krijgt wat dat betreft eerst de behandeling die het getal verdient, en daarna een toepassing die je niet verwacht: hoe groot is de kans dat niemand zichzelf trekt bij het jaarlijks sinterklaaslootjesgebeuren?

Kansrekening is een vak apart. Enerzijds leidt het tot fundamentele en bijna filosofische vragen over zaken die soms maar moeilijk vanuit de wiskunde te begrijpen zijn. Anderzijds is kansrekening buitengewoon concreet bij bijvoorbeeld wetten over grote aantallen. In dit boek gaan we op beide aspecten in. Het eerste omdat we daar bij de eerste cursus bewust voor gekozen hebben, het tweede omdat de cursisten vragen stelden die tot nadenken stemden. De vraag bijvoorbeeld waarom die normale verdeling altijd en overal opduikt.

Wiskunde begint met het stellen van vragen. Iedereen kan vragen stellen: de gek, de student, de docent en de wetenschapper. Alle vier, en ook de vakdidacticus en misschien zelfs de onderwijskundige, kunnen veel opsteken van de antwoorden op goede vragen, en van wat je nodig hebt om die vragen te beantwoorden. Dat laatste liefst met minimaal geschut, waarmee zo mogelijk van hogere wiskunde weer lagere wiskunde gemaakt wordt. Het gaat natuurlijk ook niet om het verschil tussen hoog en laag, maar om goede wiskunde, wat dat ook moge zijn.

De vraag hoe groot $n!$ eigenlijk is en waarom heeft ons geïnspireerd. Combinatoriek en kansrekening zijn ondenkbaar zonder de faculteiten waarmee we producten als $4 \times 3 \times 2 \times 1 = 4!$ benoemen. Dobbelen en pokeren gaan niet zonder, en omdat ook God vanuit ons perspectief soms lijkt te dobbelen, met atomen in plaats van dobbelstenen, is deze vraag essentieel in de bètawetenschappen. De formule van Stirling die iets zegt over de grootte van $n!$ is niet prominent aanwezig in de wiskundecurricula, maar bij ons wel. Hij verenigt ook onze wiskunde in dit boek met dat andere getal π. We maken er een beetje een ontdekkingstocht van hoe $n!$ via de integraalrekening en e tot π leidt, en gebruiken het resultaat voor een bewijs van de centrale limietstelling. Met die prachtige formule voor de normale verdeling die vroeger op het Duitse tienmarkbiljet stond, maar nu onder een knop op de rekenmachine verstopt is.

Juist ook de vraag over $n!$ leidde tot de stap naar differentiaalrekening voor formules met puntjes, rechtstreeks vanuit wat voor x, x^2 en al die monomen apart zo makkelijk is met het gereedschap der gereedschappen: de staartdeling. We maken er een beetje een sport van om calculus die we misschien nog niet echt begrijpen te vermijden. Onvermijdelijk blijkt dan ook de verschijning van dat mysterieuze imaginaire getal i waarvoor geldt dat $i^2 = -1$. Dit getal komt ons gewoon aanwaaien.

We beginnen dus met getallen, en we eindigen er ook weer mee. In het hele boek gaan algebra, analyse, dynamica en kansrekening hand in hand. Niet altijd systematisch, maar met antwoorden op vragen die we onszelf of anderen stelden. We laten de lezer daarbij ook flink wat opgaven maken, als integraal onderdeel van de tekst, en dat alles zonder plaatjes. Die mag je zelf tekenen, bij voorkeur eerst met de hand.

Het plezier staat voorop. Niet nodig bij het lezen is het Grote RekenMonster, maar we denken soms wel aan de GRM. Tijdens de cursussen waren we het als docenten lang niet altijd eens, tot hilariteit van de deelnemers. Wiskunde geeft veel vaker aanleiding tot discussies en disputen dan de meeste mensen denken. Iets van die verschillen van mening is wel zichtbaar in de tekst. Hoe alle problemen, opgaven, feiten, argumenten en bewijzen die vanuit ons hoofd in dit boek zijn beland, eerst in ons eigen hoofd gevormd en terecht zijn gekomen is een niet-gedocumenteerd verhaal apart. We danken Cécile, Steven en Swier voor hun feedback op een eerdere versie van dit boek en wensen onze lezers veel plezier.

Joost Hulshof en Ronald Meester
Amsterdam, zomer 2015

Bij de tweede editie is het een groot genoegen om James Gerritsen[1] te mogen bedanken, die erbij was vanaf dag 1, en op vele pagina's nog vele typo's en ook een enkel foutje (grr) heeft ontdekt.

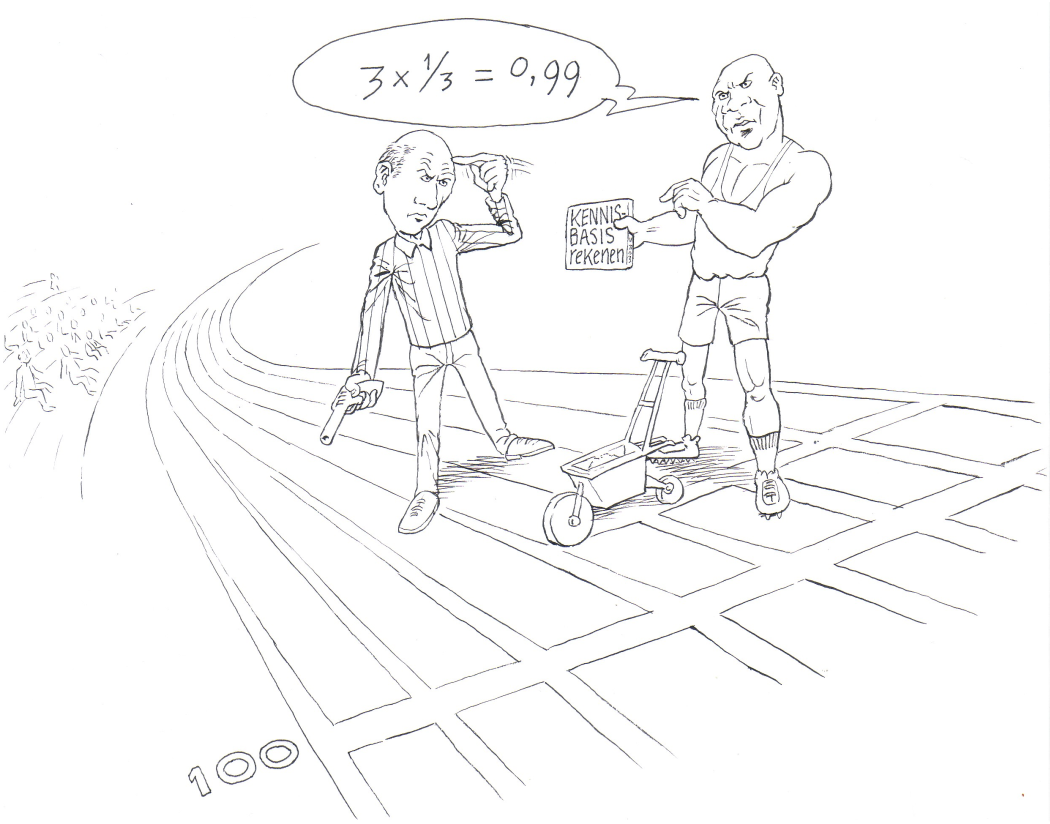
Joost Hulshof en Ronald Meester
Amsterdam, december 2016

[1] James zal nooit leraar van het jaar worden. Dat zouden meer mensen moeten doen.

Inhoudsopgave

[2] Voor de liefhebber.

[3] Voor de echte liefhebber, met een gejatte titel.
[4] Voor de liefhebber.



1 Zoveel getallen om mee te rekenen!

Ons getalbegrip begint met tellen: een, twee, drie en daarna houdt het niet meer op. Om tot oneindig te komen hoef je alleen maar te tellen. Het Hilbert[1] hotel, waarin de kamers genummerd zijn maar de laatste kamer niet bestaat, is een bron van plezier voor talrijke wiskundige denkactiviteiten[2], nog voor je kunt rekenen. Maar denken doe je met kennis, en zonder rekenen geen echte kennis. Dus beginnen we nu met rekenen, kaal, in context, en dan weer met wat afstand van die context en onszelf. Op weg naar dat hotel en denkend over wat er na oneindig komt.

[1] De eerste metawiskundige.
[2] WDA in de nieuwe examenprogramma's wiskunde.

1.1 Breuken en andere getallen

Rationale getallen zijn de breuken die we uitgaande van vingertellen via optellen, aftrekken, vermenigvuldigen en delen onvermijdelijk tegenkomen als we aan het rekenen slaan. Bij dat rekenen zijn de gebruikelijke symbolen

$$+ \quad - \quad \cdot \quad /$$

maar in de schoolrekenboekjes wordt meestal $\times$ in plaats van $\cdot$ voor keer gebruikt, en de dubbele punt $:$ voor delen. We zullen $\cdot$ en $\times$ door elkaar gebruiken, maar voor delen gebruiken we liever een horizontaal deelstreepje.

Een beetje flexibel moet je wel zijn met de notaties voor delen en breuken. Delen we drie door vier dan zijn in

$$3 : 4 = 3/4 = \frac{3}{4}$$

de twee breuknotaties zowel te lezen als drie gedeeld door vier, maar ook als twee verschillende schrijfwijzen voor de breuk drievierde[3].

Het bestaan van de rationale getallen zullen we als vanzelfsprekend aannemen. Ons andere uitgangspunt is de met 0 en 1 gemarkeerde getallenlijn, die naar twee kanten doorloopt tot zover je maar wilt. De punten op die lijn identificeren we zonder scrupules met getallen, waaronder in het bijzonder het getal 0 en het getal 1, het eerste gehele positieve getal. De getallen die aan dezelfde kant van 0 liggen als 1 noemen we positief, de getallen aan de andere kant noemen we negatief. Om de onderlinge ordening van getallen aan te geven schrijven we bijvoorbeeld $1 > 0$, $0 < 1$, $-1 < 0$, $1 < 2$, uit te spreken als 1 is groter dan 0, 0 is kleiner dan 1 enzovoorts.

Op de getallenlijn liggen niet alleen 0, 1 en de andere gehele getallen, -1, 2, -2, 3, -3, $\ldots$, geordend via

$$\cdots < -3 < -2 < -1 < 0 < 1 < 2 < 3 < \cdots$$

maar ook alle andere rationale getallen die we krijgen als we rekenen met gehele getallen. De rationale getallen zijn dan de facto alle breuken met een geheelwaardige teller en noemer, waarbij die noemer ongelijk[4] moet zijn aan 0, met de afspraak dat $\frac{4}{-6} = \frac{2}{-3} = \frac{-2}{3} = -\frac{2}{3}$ (liever de min voor de breuk dus) en $\frac{3}{4} = \frac{6}{8} = \frac{75}{100} = 0,75$. De laatste schrijfwijze is de incidenteel bruikbare decimale notatie die we moeten lezen als

$$\frac{3}{4} = \frac{7}{10} + \frac{5}{100}.$$

[3] Drie gedeeld door vier: vraag, antwoord of allebei?

[4] Sommige theoretische informatici vinden dit niet fijn.

Over de decimale schrijfwijze van rationale getallen (breuken) komen we zo nog nader te spreken maar eerst merken we op dat op de getallenlijn ook niet-rationale getallen te vinden zijn, zoals de wortel uit 2, genoteerd met $\sqrt{2}$. De vraag in welke zin dit getal bestaat is van een andere orde dan de vraag of $\frac{1}{3}$ en $\frac{2}{3}$ bestaan. Waar deze twee rationale getallen verbonden zijn aan het in drie gelijke stukken verdeelde lijnstuk tussen 0 en 1, is de aanwijzing voor het bestaan van $\sqrt{2}$ sinds Pythagoras de rechthoekige gelijkbenige driehoek met rechthoekzijde gelijk aan 1. Als we de rechthoekzijde van zo'n bijzondere driehoek gebruiken om op een getallenlijn 0 en 1 te markeren, dan is de schuine zijde te gebruiken om het punt op die getallenlijn waarmee $\sqrt{2}$ correspondeert te vinden.

Deze beschrijving van $\sqrt{2}$ als verhouding[5] tussen twee lengten in een meetkundig object is directer dan die van $\sqrt{2}$ als de zijde van een vierkant met oppervlakte 2, waarbij immers ook over eenheden van lengte en oppervlakte nagedacht moet worden[6]. We zien $\sqrt{2}$ verschijnen als een verhouding in een wellicht platonische maar toch reële context, en vervolgens als een getal dat zo reëel is dat we geen reden zien om het bestaan ervan nog in twijfel te trekken. De hogeremachtswortels[7] uit 2, en de andere niet-rationale positieve getallen, nemen we meteen ook maar voor lief. Op de getallenlijn liggen dus behalve de rationale getallen ook andere reële maar niet-rationale getallen, en dat zijn er nog veel meer dan we kunnen beschrijven met wortelconstructies zoals hierboven.

Dat $\sqrt{2}$ inderdaad een nieuw soort getal is, een getal dat niet uit de realiteit is weg te denken maar toch niet rationaal is, dus niet te schrijven als een breuk waarvan teller en noemer gehele getallen zijn, is nog niet eens zo makkelijk. Je kunt het inzien als je weet dat elk positief geheel getal even of oneven is. De volgende opgave komt dan ook in elke eerste cursus getaltheorie[8] voor. Leuk om te proberen.

Opgave 1.1. Bewijs dat $\sqrt{2}$ geen breuk is[9].

Een directe consequentie van dit inzicht is dat in een rechthoekige gelijkbenige driehoek de verhouding tussen schuine en rechte zijde geen breuk kan zijn. Een niet-rationale verhouding dus! Een conclusie zo onheilspellend voor de

[5] Zoals ook π een verhouding is: tussen omtrek en diameter van een cirkel.
[6] Al kun je je afvragen of je zonder begrip van oppervlakte tot dit inzicht komt.
[7] Maar alleen de derdemachtswortel via kubussen lijkt zo echt reëel ...
[8] Getaltheorie is iets anders dan wat door sommigen getallentheorie genoemd wordt.
[9] Stel wel, dan is 2 een quotiënt van kwadraten. Hoeveel factoren 2? Kan dat?

Griekse filosofen van weleer dat geheimhouding[10] verstandig geacht werd.

Het getal $\sqrt{2}$ is wel willekeurig goed te benaderen met breuken. We zijn zonder het misschien te weten allang vertrouwd met benaderingen, maar dan van gewone rationale getallen met bijzondere (decimaal te schrijven) rationale getallen. We weten immers, in de gebruikelijke decimale notatie van tienvingerigen, dat

$$\frac{1}{3} \neq 0{,}3\,, \quad \frac{1}{3} \neq 0{,}33\,, \quad \frac{1}{3} \neq 0{,}333\,, \quad \ldots$$

omdat nu eenmaal

$$\frac{1}{3} = 0{,}3333\ldots \tag{1.1}$$

Of we het leuk vinden of niet, er komt geen eind aan.

Niet iedereen vindt representaties als (1.1) even vanzelfsprekend[11], misschien omdat (1.1) ons nogal dwingend verzoekt te accepteren dat dan ook

$$1 = 3 \cdot \frac{1}{3} = 3 \cdot 0{,}3333\ldots = 0{,}9999\ldots \tag{1.2}$$

Opgave 1.2. Denk even na over dit verzoek[12].

Of je accepteert (1.1) en dus ook (1.2), of je verwerpt ze allebei. Het ene wel en het andere niet accepteren is geen optie. Omdat de rij breuken

$$0{,}9 = \frac{9}{10}, \quad 0{,}99 = \frac{9}{10} + \frac{9}{100}, \quad 0{,}999 = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000}, \quad \ldots$$

het getal 1 net zo goed benadert als we maar kunnen willen, wordt (1.2) in brede kring geaccepteerd[13].

We zetten nu meteen maar de onvermijdelijke volgende stap. Eigenlijk staat er in (1.2) niets anders dan

$$1 = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \frac{9}{10000} + \cdots\,,$$

een som met oneindig veel termen, waar we dus niet moeilijk over hoeven te doen: een doorlopende som[14] waar gewoon wat uitkomt. Evenzo leest (1.1) als

$$\frac{1}{3} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \cdots\,,$$

[10] Voor de verhouding π was dit nog geruststellend onbekend.

[11] De huidige kennisbasis rekenen voor de pabo vermeldt $\frac{2}{3} = 0{,}66$.

[12] Decimal notation is making us an offer we can't refuse.

[13] Zonder deze conventie is wiskunde ook leuk, zelfs nuttig (!), maar een stuk lastiger.

[14] Tot groot verdriet van sommigen ook wel reeks genoemd.

en is er geen tegenspraak meer mogelijk bij de conclusie dat

$$\frac{1}{9} = 0{,}1111\ldots = \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \frac{1}{10000} + \cdots = \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \frac{1}{10^4} + \cdots .$$

Wat hier staat is niets anders dan dat de breuk $\frac{1}{9}$ willekeurig goed te benaderen is met een decimaal geschreven getal: je kunt het verschil tussen $\frac{1}{9}$ en de som van de eerste zoveel termen zo klein maken als je maar wilt door genoeg termen uit de som mee te nemen in de optelling. Met

$$\frac{1}{9} \approx 0{,}1111 = \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \frac{1}{10000}$$

zit je met de decimale breuk dichter dan een tienduizendste bij $\frac{1}{9}$. Wil je dichterbij, dan ga je verder en neem je meer decimalen. For all practical purposes and beyond is er geen limiet aan de nauwkeurigheid die je kunt bereiken.

Opgave 1.3. Als je vingers hebt om op te tellen kun je veel leren. Denk je in dat je met n vingers geboren bent, volg de redenering hierboven en leg uit dat

$$\frac{1}{n-1} = \frac{1}{n} + \frac{1}{n^2} + \frac{1}{n^3} + \frac{1}{n^4} + \cdots \tag{1.3}$$

voor elke gehele n groter dan 1.

Voor het eerst in dit boekje staat er hierboven een formule met een letter. Die letter, hier de letter n, kun je bijvoorbeeld vervangen door 10. Het is dankzij onze tien vingers dat we op de formule met $n = 10$ zijn gekomen. Anders gezegd, met $n = 10$ heeft de formule een reële betekenis waar we eigenlijk niet om heen konden. Waren we in het Land van Oct geboren, en dus op het speelplein opgevoed met

een, twee, drie, vier, vijf, zes, zeven, acht,
wie niet weg is, staat op wacht,

en in de klas met de cijfers 1, 2, 3, 4, 5, 6, 7 en het getal acht[15], dan was dit verhaaltje begonnen met $n = 8$ in (1.3).

Kortom, acceptatie van (1.2) heeft nogal wat gevolgen. Accepteer je $n = 10$ in (1.3), dan accepteer je elke gehele positieve n. Behalve $n = 1$ natuurlijk, je moet wel kunnen tellen! Hoe moeilijk dat tellen is maakt niet uit: op Douglas Adams' planeet met duizendpoten is de deodorant weliswaar

[15] En, na invoering van het cijfer 0 voor nul, met 10 als één keer acht plus nul keer één.

eerder uitgevonden dan het wiel, maar was met $n = 1000$ uiteindelijk de gelijkheid

$$\frac{1}{999} = \frac{1}{1000} + \frac{1}{1000^2} + \frac{1}{1000^3} + \frac{1}{1000^4} + \cdots$$

niet minder reëel dan (1.2) bij ons. En mocht je het antwoord op de vraag over *life, the universe and everything* willen delen door 999, dan is het onmiddellijke antwoord:

$$\frac{42}{999} = \frac{42}{1000} + \frac{42}{1000000} + \frac{42}{1000000000} + \cdots = 0{,}042042042\ldots = 0{,}\underline{042}.$$

Ga het maar na met een staartdeling als je het niet gelooft.

Opgave 1.4. Elke breuk heeft een decimale ontwikkeling die ofwel afbreekt ofwel zichzelf herhaalt. Overtuig jezelf hiervan door wat voorbeelden uit te rekenen met een staartdeling. Omgekeerd is ook elke repeterende decimale ontwikkeling een breuk. Overtuig jezelf ook hiervan met variaties op de voorbeelden hierboven.

Natuurlijk is (1.3) ook wel zonder vingers af te leiden uit de gebruikelijke rekenregels, als je eenmaal zeker bent dat de som als getal bestaat:

Opgave 1.5. Als, voor $n > 1$, het getal A gegeven wordt door

$$\frac{1}{n} + \frac{1}{n^2} + \frac{1}{n^3} + \frac{1}{n^4} + \cdots = A,$$

dan is

$$\frac{n}{n} + \frac{n}{n^2} + \frac{n}{n^3} + \frac{n}{n^4} + \cdots = nA.$$

Bereken A door de gelijkheden van elkaar af te trekken.

Er is nog een mooie realistische context voor (1.3). Als je niet weet[16] hoe je een ronde pizza moet opdelen in gelijke stukken voor drie kinderen, maar wel weet hoe je moeten vierendelen, dan snij je de pizza eerst in vier stukken, deelt drie stukken uit, en laat ze lekker eten. Je deelt vervolgens het overgebleven stuk weer in vieren, en herhaalt[17] deze eerlijke procedure net zolang als ze honger hebben of de grootte van de pizza het toestaat. Voor het inzicht dat

$$\frac{1}{3} = \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4} + \cdots$$

is ook deze hapmethode in de praktijk overtuigend genoeg.

[16] Daarom ontbreekt Streeflands pizzamodel in eerder genoemde kennisbasis.
[17] Ook herhaaldelijk vierendelen heeft een lange traditie: google Archimedes.

Opgave 1.6. Eet een pizza in je eentje op. Hint: kies $n = 2$ in (1.3).

Zoals eerder gezegd, er is niet echt een reden om moeilijk te doen in de voorbeelden hierboven, maar het mag wel, zeker als je met een ander begrip van hetzelfde onderwerp later in andere situaties waar het allemaal niet zo makkelijk is wel uit de voeten kunt.

Wiskundigen zeggen in dit verband dat de rij breuken 0,9, 0,99, 0,999, 0,999, ... stijgend is en begrensd. De kleinste bovengrens, in dit geval 1, wordt de limiet van de rij genoemd. We zeggen dat de rij convergent is met limiet 1: de rij convergeert naar zijn limiet, en die limiet is 1.

Als je $1 = 0{,}9999999\ldots$ als vanzelfsprekend accepteert dan gaat het hier alleen maar om nieuwe naamgeving. Maar met een moeilijk af te wijzen kleinste bovengrens axioma is nu iedere decimale ontwikkeling met een eindig aantal cijfers voor de decimale komma[18] en een doorlopende rij cijfers daarachter, hoe je daar ook verder aankomt, te zien als reëel getal. Door het getal steeds verder in zijn decimale ontwikkeling af te kappen, en de limiet te nemen van de zo gevormde stijgende begrensde rij getallen, geeft dit axioma je gratis en voor niets een reëel getal!

Een reëel getal is zo altijd de limiet van een niet-dalende begrensde rij rationale getallen. Zonder erbij stil te staan werkten we altijd al zo met reële getallen. Dat onze gewoonte om breuken soms ook decimaal te schrijven tot veel inzicht leidt is duidelijk[19].

Opgave 1.7. Werken met reële getallen is bijvoorbeeld het optellen en vermenigvuldigen van zulke getallen met doorlopende decimale ontwikkelingen. Helemaal vanzelfsprekend is dat niet. Neem eens de tijd om jezelf ervan te overtuigen[20] dat som en product van twee reële getallen altijd goed gedefinieerd zijn. In de reële getallen gelden dezelfde rekenregels als in de rationale getallen. Denk nog eens over het waarom van deze uitspraak als je dit boek uit hebt.

Dat je hetzelfde getal op verschillende manieren kunt 'maken' als limiet is niet vreemd als je bedenkt dat tweevingerigen en tienvingerigen in hun praktische wiskunde, via

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots = 1 = \frac{9}{10}+\frac{9}{100}+\frac{9}{1000}+\frac{9}{10000}+\cdots,$$

[18] In het Engels een (decimal) point.

[19] Dat veel inzicht verloren gaat als we breuken alleen maar decimaal schrijven ook.

[20] Het zogenaamde kolomrekenen kan hierbij handig zijn.

uiteindelijk één zijn. En ze kunnen allebei meteen ook heel veel getallen maken! De tienvingerigen hebben bijvoorbeeld

$$0{,}1001000010000001\ldots = \frac{1}{10} + \frac{1}{10^4} + \frac{1}{10^9} + \frac{1}{10^{16}} + \cdots,$$

niet minder reëel dan 0,1111..., maar wel een getal waarbij in de decimale ontwikkeling geen periodieke herhaling zit, geen breuk dus. De bijbehorende stijgende rij van breuken is

$$\frac{1}{10}, \frac{1001}{10000}, \frac{100100001}{1000000000}, \frac{1001000010000001}{10000000000000000}, \ldots$$

Dit voorbeeldgetal is net als $\sqrt{2}$ niet rationaal. Het heeft de opmerkelijke eigenschap dat van de rij benaderende breuken de derde al 9 cijfers van de decimale ontwikkeling goed heeft, de vierde 16, enzovoorts. Kun je $\sqrt{2}$ ook zo snel en goed benaderen met breuken? Het antwoord blijkt ja te zijn. In hoofdstuk 3 zullen we een soortgelijke kwadratische progressie terugzien bij de methode van Newton toegepast op de vergelijking $x^2 = 2$.

Een andere vraag is in hoeverre de getallen $\sqrt{2}$ en 0,1001000010000001... even normaal of abnormaal zijn. Die laatste vraag komt aan de orde, als we de reële getallen in hoofdstuk 2 als verzameling gaan bekijken. Dat doen we trouwens ook al in de nu volgende sectie, waarin we onderzoeken of er 'meer' breuken dan gehele getallen zijn, en of er 'meer' reële getallen zijn dan breuken.

1.2 Tot en voorbij oneindig: telbaar en overaftelbaar

Meer of nog meer? Hoeveel getallen zijn er eigenlijk? De natuurlijke getallen geven we doorgaans aan met $\mathbb{N} = \{1, 2, 3, \ldots\}$, waarbij de $\mathbb{N}$ de N van Natuurlijk voorstelt. Je gelooft meteen dat het er oneindig veel zijn. Na de 3 komt de 4. Na de 4 komt de 5, het houdt niet op en er zit geen eind aan. Oneindig veel getallen dus. Wat er op de puntjes staat na 1, 2 en 3 geloven we wel want anders wordt dit boekje een gebed zonder einde.

Met die oneindig veel getallen in $\mathbb{N}$ maakten we nog meer getallen, als eerste de verzameling van rationale getallen, doorgaans $\mathbb{Q}$ genoemd, met de Q van quotiënt. Anders dan $\mathbb{N}$ zie je deze $\mathbb{Q}$ eigenlijk nooit uitgeschreven als $\mathbb{Q} = \{r_1, r_2, r_3, \ldots\}$. Waar $\mathbb{N}$ gegeven wordt door een simpele geordende opsomming van al haar elementen, zien we $\mathbb{Q}$ meestal aangeduid als de verzameling van breuken, de rationale getallen, net zoals $\mathbb{R}$ de verzameling van reële getallen is, reële getallen die we maakten als limieten van stijgende rijen breuken.

De vraag om nu te stellen is hoe oneindig $\mathbb{Q}$ en $\mathbb{R}$ eigenlijk zijn. Als we bijvoorbeeld $\mathbb{Q}$ inderdaad kunnen 'aftellen' met een eerste getal r_1, een tweede getal r_2, een derde getal $r_3, \ldots$, dan zien we met een beetje creatief tekstzetten aan

$$\begin{aligned} \mathbb{Q} &= \{r_1, r_2, r_3, \ldots\} \\ \mathbb{N} &= \{\,1\,,\,2\,,\,3\,,\ldots\} \end{aligned} \tag{1.4}$$

dat $\mathbb{N}$ en $\mathbb{Q}$ 'evenveel' getallen bevatten. Een verzameling V noemen we dan ook aftelbaar als we alle elementen van V een voor een kunnen aflopen, zonder er eentje over te slaan. Of nog weer anders gezegd, V heet aftelbaar als we de elementen van V kunnen nummeren met de getallen in $\mathbb{N}$. Zo is $\mathbb{N}$ nu ook per definitie aftelbaar: elk getal krijgt gewoon zijn eigen nummer. Een hele bijzonder aftelling die de natuurlijke ordening van $\mathbb{N}$ respecteert. We zouden de vraag kunnen stellen of dat ook met $\mathbb{Q}$ kan, dus of we $\mathbb{Q}$ kunnen schrijven als

$$\mathbb{Q} = \{r_1 < r_2 < r_3 < \cdots\}. \tag{1.5}$$

De lezer ziet aan de notatie hopelijk wel wat we bedoelen, al is het een kwestie van goed tussen de puntjes doorlezen. We kunnen dus ook praten over geordend aftelbaar met $\mathbb{N}$ als telverzameling.

Opgave 1.8. Makkelijk: laat zien dat $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, de verzameling van alle gehele getallen, aftelbaar is. Niet zo makkelijk: is $\mathbb{Z}$ ook geordend aftelbaar?

Aan deze opgave zie je dat $\mathbb{Z}$, die toch evident net zo veel getallen bevat als twee kopiën van $\mathbb{N}$ plus die ene 0, ook 'evenveel' getallen bevat als $\mathbb{N}$. Maar dan kan het voor $\mathbb{Q}$ niet heel anders zijn. We moeten alleen wat creatief tellen[21]. De natuurlijke ordening van de breuken kunnen we weliswaar niet respecteren[22] maar in het volgende schema zien we alle positieve getallen uit $\mathbb{Q}$ staan, inclusief 'dubbelen' zoals $\frac{1}{2}$ en $\frac{2}{4}$:

$$\begin{array}{cccccc} 1/1 & 2/1 & 3/1 & 4/1 & \cdot & \cdot\ \cdot \\ 1/2 & 2/2 & 3/2 & 4/2 & \cdot & \cdot\ \cdot \\ 1/3 & 2/3 & 3/3 & 4/3 & \cdot & \cdot\ \cdot \\ 1/4 & 2/4 & 3/4 & 4/4 & \cdot & \cdot\ \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\ \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\ \cdot \end{array} \tag{1.6}$$

[21] Zoals pabostudenten al leren met TAL: tellen en tellen is twee.
[22] Omdat er geen kleinste breuk is?

Op de eerste rij staan alle 'breuken' met noemer 1, op de tweede rij staan alle breuken met noemer 2, enzovoorts. De puntjes lopen weer door en je moet ze zelf invullen. Elke breuk komt in dit schema natuurlijk oneindig keer voor, kijk maar naar de 'lijnen' door de $\frac{0}{0}$ die er niet staat[23].

Over dit schema leggen we een ander schema

$$\begin{array}{ccccc} 1 & 2 & 4 & 7 & 11 \\ 3 & 5 & 8 & 12 & \\ 6 & 9 & 13 & & \\ 10 & 14 & & & \\ 15 & & & & \end{array}$$

waarin we het stippelwerk nu maar achterwege hebben gelaten. Je ziet dat we met $\mathbb{N}$ nu beide schema's simultaan aftellen: alle positieve breuken, die samen de verzameling $\mathbb{Q}^+$ vormen, komen zo aan de beurt! Dus $\mathbb{Q}^+$ en dan ook $\mathbb{Q}$ zijn aftelbaar.

Opgave 1.9. Alle positieve rationale getallen komen in dit schema oneindig vaak aan de beurt. Is dit een reden tot zorg over de correctheid van het argument?

Er bestaat dus een aftelling waarmee de getallen $\mathbb{Q}$ netjes in een rijtje gezet worden, en dat is nu een goede reden om te zeggen dat $\mathbb{N}$ en $\mathbb{Q}$ evenveel elementen hebben. We kunnen de getallen in $\mathbb{N}$ en $\mathbb{Q}$ immers een op een aan elkaar koppelen, zoals weergegeven in (1.4). Anderzijds lijkt het duidelijk dat er meer breuken dan gehele getallen zijn. Het begrip evenveel in de context van oneindig grote verzamelingen is dus subtiel en wellicht tegenintuïtief. De meeste wiskundigen praten dan ook niet over het aantal elementen in $\mathbb{N}$ of $\mathbb{Q}$, maar over de kardinaliteit van $\mathbb{N}$ of $\mathbb{Q}$: de kardinaliteit van $\mathbb{N}$ en $\mathbb{Q}$ heet dan hetzelfde te zijn. Een kardinale fout die je hier kunt maken is het vergeten van de ordening. Er is geen enkele manier om een schema als (1.5) te realiseren, en daar kun je jezelf misschien wel van overtuigen als het tweede onderdeel van Opgave 1.8 gelukt is.

Op deze manier bezien is $\mathbb{Q}$ dus niet 'groter' maar wel anders dan $\mathbb{N}$. De volgende vraag is nu natuurlijk hoe het zit met $\mathbb{R}$, de verzameling van alle reële getallen. We hebben de reële getallen min of meer gedefinieerd als stijgende limieten van breuken, via hun decimale representatie. Een (positief) reëel getal is voor ons dus niets anders dan een oneindig rijtje decimalen die allemaal 0, 1, 2, 3, 4, 5, 6, 7, 8 of 9 zijn, met eindig veel getallen voor de komma, en oneindig veel erna. Zou het kunnen zijn dat er ook maar aftelbaar

[23] Nul gedeeld door nul zal nog terugkomen…

veel van dat soort rijtjes zijn, met andere woorden, kunnen we ook de reële getallen aftellen zoals dat met $\mathbb{N}$ en $\mathbb{Q}$ kan?

Het aardige is dat we die vraag nu zonder wat voor poespas dan ook kunnen beantwoorden. We redeneren, net als bij de irrationaliteit van $\sqrt{2}$, uit het ongerijmde en nemen aan dat we de elementen van $\mathbb{R}$ wel kunnen aftellen. Dan kan dat dus ook met de reële getallen tussen 0 en 1, bijvoorbeeld met de rij $x_1 = 1 = 0,9999999999\ldots$, $x_2 = 0,3141592653\ldots$, $x_3 = 0,5772156649\ldots$, $x_4 = 0,2685452001\ldots$, $x_5 = 0,4669201609\ldots$, $x_6 = 0,2718281828\ldots$, et cetera. We kiezen maar wat[24].

Het schema

$$\begin{array}{cccccccccc}
\boxed{9} & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 \\
3 & \boxed{1} & 4 & 1 & 5 & 9 & 2 & 6 & 5 & 3 \\
5 & 7 & \boxed{7} & 2 & 1 & 5 & 6 & 6 & 4 & 9 \\
2 & 6 & 8 & \boxed{5} & 4 & 5 & 2 & 0 & 0 & 1 \\
4 & 6 & 6 & 9 & \boxed{2} & 0 & 1 & 6 & 0 & 9 \\
2 & 7 & 1 & 8 & 2 & \boxed{8} & 1 & 8 & 2 & 8
\end{array}$$

bevat alle getallen uit de rij. We laten de puntjes weer weg, maar het schema loopt naar rechts en naar beneden door. De decimalen van x_1 vormen de eerste rij, de decimalen van x_2 de tweede rij, enzovoorts. Om de decimalen op de diagonaal hebben we een vierkantje getekend. Deze decimalen maken zelf weer een getal, in dit geval het getal $0,917528\ldots$ en dat komt dan verderop in de rij terug als we met de rij inderdaad alle reële getallen van 0 tot en met 1 aftellen. Als het goed is.

Maar ook $y = 0,028639\ldots$ komt in de rij voor, een getal dat we uit $0,917528\ldots$ maken door de decimalen te vervangen volgens deze regel: $0 \to 1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 7 \to 8 \to 9 \to 0$. Maar zit dit gemuteerde getal y wel in de rij? Kijk naar de eerste decimaal van y en van x_1. Die zijn verschillend, dus $y \neq x_1$. Maar op de tweede decimaal verschillen y en x_2, dus $y \neq x_2$, enzovoorts. Dus y komt niet in ons rijtje voor, een ongerijmdheid.

We kunnen de reële getallen dus niet aftellen. Hoeveel wiskunde we met z'n allen ook doen en hoeveel getallen we ook ontdekken, er is altijd meer. Of we nu naar getallen of verzamelingen kijken, welke wegen we ook bewandelen op onze reis door de wiskunde, het is nooit klaar, al ligt er ergens langs de never ending road wel dat Hilberthotel. Een hotel waar iedereen altijd welkom is, de kamers genummerd met $1, 2, 3, \ldots$ Veel kennis van de getallen is er verder niet. Ook als het hotel vol is kan de vermoeide reiziger altijd terecht[25], al moet iedereen daarvoor verkassen.

[24] Niet helemaal toevallig misschien.

[25] Het argument vol is vol is ook hier ondeugdelijk.

2 Ooit een normaal getal gezien?

En, hoe beviel het? Of heb je nog nooit een normaal getal gezien? De vraag doet misschien denken aan het 'bewijs' dat alle positieve gehele getallen een bijzondere eigenschap hebben. Immers, stel maar dat er getallen bestaan die geen bijzondere eigenschap hebben. In de collectie van dat soort getallen moet een kleinste getal te vinden zijn. Maar dan heeft dat getal juist wel een bijzondere eigenschap. Tegenspraak.

Dit bewijs en deze vage definitie zijn natuurlijk niet heel serieus te nemen. In dit hoofdstuk gaat het om een concreter antwoord op de vraag wat we normaal (dus niet bijzonder) willen vinden, en ook om de vraag of dat iets te maken heeft met het willekeurig trekken van een getal. Beide auteurs hebben zo hun eigen historie met betrekking tot deze vragen.

Een gesprek op de bank leerde dat Ronald vroeger lottorijtjes invulde en daarbij steevast 123456 koos. Tot ongenoegen van de vader die vond dat die combinatie toch nooit zou optreden. De kans op een 'normaal getal' was immers veel groter. Elke zondagavond na *Studio Sport* gingen alle balletjes (genummerd 1 tot en met 45) in een grote glazen bol die na draaien via een fraai mechanisme zes balletjes in zes vakjes deed belanden. De vader had opgemerkt dat '123456 nog nooit gebeurd was'. Daar had hij gelijk in. Het aantal mogelijkheden is enorm en de kans op een trekking met alleen de eerste zes ballen in de hokjes is astronomisch klein[1]. Het argument van Ronald dat over willekeurig welk ander rijtje van zes ballen hetzelfde gezegd kan worden maakte echter geen enkele indruk.

Joost herinnert zich het Human Resource Management uit de tijd van voor de afwasmachine. 'Neem een getal tussen 0 en 10 en schrijf het op een briefje.' Vervolgens noemde moeders getallen op. Werd je getal genoemd dan moest je afdrogen. Getallen tussen 0 en 10, dat was makkelijk, e, π, $\sqrt{2}$, noem maar op. Dwars? Nee hoor. Kon hij geen normale getallen opschrijven? Ja hoor, maar wat is dat dan, een normaal getal?

In dit hoofdstuk dus twee vragen. Is er wel een eigenschap te formuleren waarmee een getal dat eraan voldoet, het predicaat 'normaal' verdient? En hoe verhoudt dit predicaat zich dan tot het trekken van willekeurige getallen, bijvoorbeeld een willekeurig reëel getal tussen 0 en 1?

Zomaar een getal getal kiezen, hoe doe je dat? Makkelijk, je schrijft maar wat op. Bijvoorbeeld het getal

$$C_{10} = 0{,}1234567891011121314151617181920212223242526272829\ldots,$$

een tamelijk niet willekeurig en dus behoorlijk bijzonder reëel getal. Je herkent het systeem[2] in de ontwikkeling wel. Ene Champernowne schreef er in 1933 al een artikel over[3]. Hij bewees dat C_{10} transcendent is.

Trancendent betekent dat het getal uitstijgt boven de oplossingen van 'gewone' vergelijkingen die we kunnen maken voor een onbekende x, uitgaande van de natuurlijke getallen, met de gewone bewerkingen $+, -, \times$. Vergelijkingen als $x^5 - x + 1 = 0$ bijvoorbeeld, of $x^2 = 2$, waar we in hoofdstuk 3 nog uitgebreid op terugkomen. Oplossingen van dit soort vergelijkingen heten algebraïsch en C_{10} is niet zo'n getal, daar is het veel te bizar voor. Een getal met een bizarre kettingbreukontwikkeling[4] ook, kortom, een heel ander getal dan bijvoorbeeld

$$0{,}12345678910123456789101234567891 0\ldots,$$

een getal dat je na het vorige hoofdstuk natuurlijk zo als breuk[5] schrijft.

Moeten we deze twee niet bepaald willekeurige getallen, die je natuurlijk nooit tegenkomt, even normaal vinden als een alledaags getal als π, de omtrek van een cirkel met diameter gelijk aan 1? Of als e, het getal van Euler dat nog uitgebreid komen gaat? Of als $\sqrt{2}$? Wie weet? Wij zeker niet, maar het is wel

[1] Leuk om te proberen die kans uit te rekenen, zie ook hoofdstuk 5.

[2] Sommige mensen kiezen zo hun passwords.

[3] Wat zou C_6 zijn? En C_{12}? En C_{14}? En C_2?

[4] Meditatie hielp de verleiding van kettingbreuken voor dit boekje te weerstaan.

[5] Wel even vereenvoudigen graag.

leuk om hier wat denkactiviteiten aan te wijden, wiskundige denkactiviteiten die ons wellicht wat leren, over reële getalen, over willekeur en over wat we normaal willen vinden.

2.1 Normale getallen

Elk reëel getal tussen 0 en 1 kunnen we schrijven als[6]

$$0{,}d_1 d_2 d_3 \ldots, \tag{2.1}$$

waarbij d_k de k-de decimaal voorstelt, een cijfer dat 0, 1, 2, 3, 4, 5, 6, 7, 8 of 9 kan zijn. Elke keuze van decimalen $d_1, d_2, d_3, \ldots$ definieert een getal. Zo hoort bij (2.1) het getal

$$a = \frac{d_1}{10} + \frac{d_2}{100} + \frac{d_3}{1000} + \cdots,$$

omdat we in het vorige hoofdstuk hebben afgesproken dat a als kleinste bovengrens van de begrensde rij

$$\frac{d_1}{10}, \frac{d_1}{10} + \frac{d_2}{10^2}, \frac{d_1}{10} + \frac{d_2}{10^2} + \frac{d_3}{10^3}, \ldots$$

bestaat, net zoals

$$\frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \frac{1}{10^4} + \cdots = \frac{1}{9}$$

bestaat.

Of we de som kunnen uitrekenen of niet doet er nu even niet toe. Reële getallen zijn doorlopende decimale ontwikkelingen. En wie niet door wil lopen hoort er nu even niet bij. Dus $\frac{1}{2} = 0{,}499999\ldots$ en niet $0{,}5000000\ldots$ Staartdelingen zonder rest zijn dus even verboden. Staartdelen is te leuk om je af te laten schepen met die nul waar je toch niet eens door mag delen. Je mag wel happen maar je moet altijd wat overlaten voor de volgende hap. Met deze afspraak is elk positief getal uniek als decimale ontwikkeling te schrijven[7]. In de inleiding van dit hoofdstuk hebben we weer twee zulke getallen gezien, het eerste was trancendent, het andere was maar een gewone breuk.

Welke getallen zouden we niet normaal, dus abnormaal willen noemen? Je zou natuurlijk kunnen zeggen dat een getal normaal is als alle cijfercombinaties oneindig vaak in de decimale ontwikkeling voorkomen. Of, als we

[6] Met de d van decimaal.

[7] Ten koste van die $0 = 0{,}00000\ldots$ weer. Jammer, maar helaas.

met drie vingers aan twee handen geboren zouden zijn, in de tertiaire ontwikkeling. Alle getallen van de vorm[8]

$$\frac{t_1}{3} + \frac{t_2}{3^2} + \frac{t_3}{3^3} + \frac{t_4}{3^4} + \frac{t_5}{3^5} + \frac{t_6}{3^6} + \frac{t_7}{3^7} + \frac{t_8}{3^8} + \frac{t_9}{3^9} + \frac{t_{10}}{3^{10}} + \cdots,$$

met de t-tjes (uiteindelijk) alleen maar 0 of 2 zouden we dan zeker niet normaal noemen. Zulke getallen zijn een feest[9] dat ons in dit boekje aan de neus voorbijgaat. En ook getallen waarin de combinatie 12321 maar eindig vaak voorkomt vinden we niet normaal. Maar blijven er zonder deze al deze niet-normale getallen nog meer getallen over die niet helemaal normaal zijn?

Een manier om meer onderscheid te maken bij de vraag over hoe vaak een eigenschap voorkomt in een oneindige rij getallen of decimalen is om eerst naar de eerste zoveel te kijken en gewoon te turven. De relatieve frequentie van de even getallen in de natuurlijke getallen $1, 2, 3, 4, \ldots$ is bijvoorbeeld achtereenvolgens

$$0, \frac{1}{2}, \frac{1}{3}, \frac{2}{4}, \frac{2}{5}, \frac{3}{6}, \frac{3}{7}, \frac{4}{8}, \ldots, \frac{42}{85}, \ldots, \frac{1024}{2047}, \frac{1024}{2048}, \ldots$$

en je ziet wel waar dat naartoe gaat als je door blijft gaan tot oneindig. Voor de vijfvouden is het

$$0, 0, 0, 0, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}, \frac{2}{10} = \frac{1}{5}, \cdots, \frac{3}{15} = \frac{1}{5}, \frac{3}{16}, \frac{3}{17}, \frac{3}{18}, \frac{3}{19}, \frac{4}{20} = \frac{1}{5}, \cdots$$

en ook dat is wel duidelijk. De relatieve frequenties van de twee- en vijfvouden komen steeds dichter bij respectievelijk $\frac{1}{2}$ en $\frac{1}{5}$ te liggen. In het geval dat zo'n limietwaarde bestaat dan noemen we die limietwaarde de relatieve frequentie of dichtheid van de eigenschap waar het om gaat, hier dus het zijn van een twee- of vijfvoud. Ook weer een getal tussen 0 en 1 trouwens, en zowel 0 als 1 komen voor.

Opgave 2.1. Wat is de frequentie van de drievouden in de natuurlijke getallen? Wat is de frequentie van getallen die zowel even zijn als een veelvoud van 3?

Opgave 2.2. Wat is de frequentie van de kwadraten? De relatieve frequenties zijn

$$1, \frac{1}{2}, \frac{1}{3}, \frac{2}{4} = \frac{1}{2}, \frac{2}{5}, \frac{2}{6}, \frac{2}{7}, \frac{2}{8}, \frac{3}{9} = \frac{1}{3}, \ldots.$$

[8] Met de t van tertiair.

[9] Cantors discontinuum, topology, and all that fine stuff.

Zou de collectie priemgetallen een frequentie hebben? Kun je een rij getallen bedenken zonder frequentie?

Bij decimale ontwikkelingen (hier van getallen tussen 0 en 1), met oneindig veel keuzemogelijkheden, maar slechts eindig veel cijfers om uit te kiezen, doen we net zoiets. Het ligt voor de hand te verwachten dat wanneer we naar de decimale ontwikkeling van een getal x kijken, de frequentie van bijvoorbeeld 7 'voor de meeste getallen' gelijk is aan $\frac{1}{10}$, en dat de frequentie van het 'blok' 69 meestal gelijk is aan $\frac{1}{100}$. Een getal noemen we daarom normaal als elk blok in de decimale ontwikkeling van dat getal de juiste frequentie heeft[10]. Dus een getal is normaal als bijvoorbeeld 9 frequentie $\frac{1}{10}$ heeft, het blok 314 frequentie $\frac{1}{1000}$, het blok 271828 frequentie $\frac{1}{1000000}$, en alle andere blokken ook allemaal de frequentie hebben die je met je boerenverstand zou verwachten.

Bestaan er zulke normale getallen? Ja, een heleboel, al hebben wij ze nog niet mogen ontmoeten. Behalve dan ... getallen als dat rare C_{10}-getal van Champernowne. Dat bleek dus normaal te zijn. Maar van gewone getallen als[11] e, π en $\sqrt{2}$ zijn ons geen uitspraken bekend dat ze normaal zijn.

Opgave 2.3. Niet makkelijk: toon aan dat de frequentie van 1 in Champernownes getal gelijk is aan $\frac{1}{10}$.

Als je ons (of Wikipedia) op ons woord gelooft hebben we met C_{10} een normaal getal. Hoera. Hoeveel zouden er nog meer zijn? Een voor de hand liggende manier om die vraag te stellen is: als we een willekeurig getal tussen 0 en 1 trekken, hoe groot is de kans dat dat getal normaal is? Is die kans kleiner of gelijk aan 1? Voor minder dan 1 willen we hier niet gaan, want willekeurig gekozen getallen zouden normaal moeten zijn.

Hoe trek je zo'n getal? Laten we die vraag beantwoorden onder de aanname dat we niet met tien, maar met zes vingers geschapen waren, drie aan elke hand: een duim, een pink, en daartussen een dikke vinger. Dan zouden we onze getallen tussen 0 en 1 (wat moeizaam) schrijven als[12]

$$x = \frac{z_1}{6} + \frac{z_2}{6^2} + \frac{z_3}{6^3} + \frac{z_4}{6^4} + \frac{z_5}{6^5} + \frac{z_6}{6^6} + \frac{z_7}{6^7} + \frac{z_8}{6^8} + \frac{z_9}{6^9} + \frac{z_{10}}{6^{10}} + \cdots, \qquad (2.2)$$

en het verhaal begonnen zijn met

$$C_6 = \frac{1}{6} + \frac{2}{6^2} + \frac{3}{6^3} + \frac{4}{6^4} + \frac{5}{6^6} + \frac{1}{6^6} + \frac{0}{6^7} + \frac{1}{6^8} + \frac{1}{6^9} + \cdots$$

[10] Dit heeft niets met de normale verdeling te maken.

[11] Eigenlijk $e-2, \pi-3, \sqrt{2}-1$.

[12] Met de z van zestallig stelsel.

als eerste poging om een normaal getal te vinden.

Nu we het er toch over hebben, het normaal zijn zou natuurlijk niet af moeten hangen van op hoeveel vingers we tellen. Of dat ook zo is? Daar is niet alles over bekend. Voorlopig kunnen we ons afvragen of een willekeurig getal met kans 1 normaal is als we decimalen gebruiken, zeg maar 10-normaal, of als we 6-vingerig rekenen, 6-normaal, zoals we nu gaan doen met dobbelstenen[13] om de z-tjes in (2.2) te trekken. Aardige vraag is ook: als voor alle vingergetallen $n > 1$ het antwoord op deze vraag ja is, is een willekeurig getal dan ook met kans 1 n-normaal voor alle $n > 1$?[14]

Kans 1, wat betekent dat eigenlijk? Dat hangt ervan af aan wie je het vraagt. Bij experimenten met eindig veel even waarschijnlijke uitkomsten heeft elke uitkomst evenveel kans. Als je afspreekt dat de kans op één zo'n uitkomst gelijk is aan 1 gedeeld door het totaal aantal mogelijkheden dan is het bij alles wat we met zuivere dobbelstenen doen wel duidelijk. De kans op drie keer 6 bij het gooien met drie dobbelstenen is[15] $\frac{1}{216}$. Kansrekening is hier tellen, combinatoriek[16] dus, waarbij het vanzelfsprekend is dat alle kansen optellen tot 1. Hieronder nemen we een limiet waarin het aantal uitkomsten naar oneindig gaat, en vertrouwen daarbij op onze intuïtie.

We gaan dus dobbelen om de z-tjes in (2.2). De eerste worp bepaalt in welk van de intervallen $[0, \frac{1}{6}]$, $[\frac{1}{6}, \frac{2}{6}]$, $[\frac{2}{6}, \frac{3}{6}]$, $[\frac{3}{6}, \frac{4}{6}]$, $[\frac{4}{6}, \frac{5}{6}]$, $[\frac{5}{6}, \frac{6}{6}]$ ons willekeurige getal gekozen wordt. De volgende worp bepaalt, samen met de eerste worp, in welk van de intervallen $[0, \frac{1}{36}]$, $[\frac{1}{36}, \frac{2}{36}]$, ... het getal ligt, enzovoorts. Elke z_n krijgen we na een worp met de hopelijk zuivere dobbelsteen. De overlap[17] in de intervallen verdwijnt met de afspraak dat getallen waarvoor vanaf zekere index alle z-tjes 0 zijn niet mee doen. De kans daarop is ook 0, net als de kans op vanaf zekere index alle z-tjes alleen maar 5.

Opgave 2.4. Overtuig jezelf ervan dat de kans dat we op deze manier een van tevoren gekozen rij z-tjes opnieuw kiezen gelijk is aan 0. Dus wat is de kans dat we een getal vinden dat we eerder gekozen hebben?

Opmerkelijk. We kiezen een getal, zoveel is duidelijk, maar voor elk van tevoren gekozen getal is de kans dat we juist dat getal weer kiezen gelijk aan 0. Vreemd, of juist niet? Wel, als de keuze echt willekeurig is, dan is elk getal even waarschijnlijk, en heeft elk getal dus dezelfde kans om gekozen te

[13] God ziet ons liever niet dobbelen, is dat de reden dat we tien vingers hebben?

[14] Antwoord is ja, maar daar branden we onze vingers hier niet aan.

[15] Omdat $6 \times 6 \times 6 = 216$ het aantal uitkomsten is.

[16] Niets minder maar ook niets meer.

[17] $\frac{1}{6}$ zit in $[0, \frac{1}{6}]$ en $[\frac{1}{6}, \frac{2}{6}]$.

worden. Maar dan kan die kans niet positief zijn. Want bij elkaar zijn alle kansen 1. En (overaftelbaar) oneindig keer iets positiefs is meer dan aftelbaar oneindig keer iets positiefs, hetgeen al meer is dan 1. En meer kans dan 1 hebben we niet[18].

De willekeurige keuze uit het interval $[0,1]$ is hierboven gerealiseerd door een oneindige rij dobbelsteenworpen. Elke worp is een worp op zich, die volkomen onafhankelijk is van elke andere worp. Alle 'toevalligheid' die er is in één willekeurig getal tussen 0 en 1, is kennelijk niets minder of meer dan de toevalligheid in een oneindige rij toevallig gekozen cijfers, op hoeveel vingers we ook rekenen. En bij alles wat je steeds opnieuw doet zonder dat het verleden het heden beïnvloedt, komt een wet om de hoek kijken die bij een breed publiek bekend staat als de wet van de grote aantallen.

Wet van de grote aantallen.[19] Als een experiment met succeskans p steeds opnieuw wordt uitgevoerd, en wel zo dat elke keer de vorige uitkomst volledig uit het systeem is verwijderd, dan geldt met kans 1 voor het aantal successen $X(n)$ na n experimenten dat

$$\frac{X(n)}{n} \to p$$

als $n \to \infty$.

Met behulp van de binomiale verdeling voor $X(n)$ kunnen we deze wet echt bewijzen als wiskundige stelling uitgaande van de axioma's voor de kansrekening[20], waarbij we overigens eerst een geschikte wiskundige vertaling moeten vinden voor 'uit het systeem verwijderen'. Maar iedere natuurkundige zal het experiment gewoon uitvoeren, turven wat er uitkomt, en bij het lezen van ons bewijs reageren met: *I told you so.* Such is life. Iedereen heeft zijn eigen gelijk, ook in het casino[21].

Hoe het ook zij, deze wet passen we toe op de keuze van de z-tjes zoals hierboven beschreven. We zeggen bijvoorbeeld dat het n-de experiment succesvol is als we als n-de decimaal een 4 kiezen. De fractie van het aantal successen (hier dus het aantal vieren) bij de eerste n decimalen convergeert volgens de wet van de grote aantallen naar $\frac{1}{6}$, de succeskans. De frequentie van een vier is dus met kans 1 gelijk aan $\frac{1}{6}$, en evenzo voor elke andere z uit $0, 1, 2, 3, 4, 5$.

De toepassing van de wet van de grote aantallen op blokjes van lengte 2 is net iets ingewikkelder omdat bij overlap de blokjes natuurlijk niet meer

[18] Waarom eigenlijk?

[19] Dit is een versie van de sterke wet, er is ook een 'zwakke' versie.

[20] Meer dan alleen combinatoriek.

[21] Na 10 keer rood bij roulette op zwart inzetten, omdat zwart rood zal 'inhalen'?

onafhankelijk zijn. Als de eerste twee z-tjes bijvoorbeeld 1 en 2 zijn, dan kan het tweede blokje alleen maar met een 2 beginnen.

Opgave 2.5. Een truc om deze complicatie te omzeilen: schrijf de blokken van lengte 2 in twee rijen op. De eerste rij bestaat uit de blokken $z_1z_2, z_3z_4, \ldots$ en de tweede uit $z_2z_3, z_4z_5, \ldots$ Op beide rijen kunnen we de wet van de grote aantallen toepassen. Geeft dat wat we verwachten? Wat verwachten we eigenlijk?

Nu we weten dat de kans op een niet normaal[22] getal gelijk is aan 0, moeten er waarachtig wel heel veel normale getallen bestaan! Maar vraag maar niet of $\sqrt{2}$ normaal is.

2.2 Hoeveel toeval verdragen we hier eigenlijk?

Terug naar onze tien vingers begrijpen we nu dat de keuze van één (al of niet toevallig) getal uit het interval $[0, 1]$ eigenlijk hetzelfde is als een oneindige rij keuzes uit $0, 1, \ldots, 9$. Maar met die rij kunnen we spelen. Bijvoorbeeld: één rij maakt ook twee rijen. Splits maar uit naar even en oneven rangnummers. Als het toevallige getal a dat we gekozen hebben als decimale ontwikkeling

$$a = 0{,}23711481258912862583 45\ldots$$

heeft, dan kunnen we hier twee getallen uit maken door de decimalen op de even posities en de decimalen op de oneven posities apart te nemen. Met de oneven decimalen maken we

$$a_o = 0{,}27182818284\ldots,$$

ook toevallig dan, en de even decimalen geven

$$a_e = 0{,}31415926535\ldots,$$

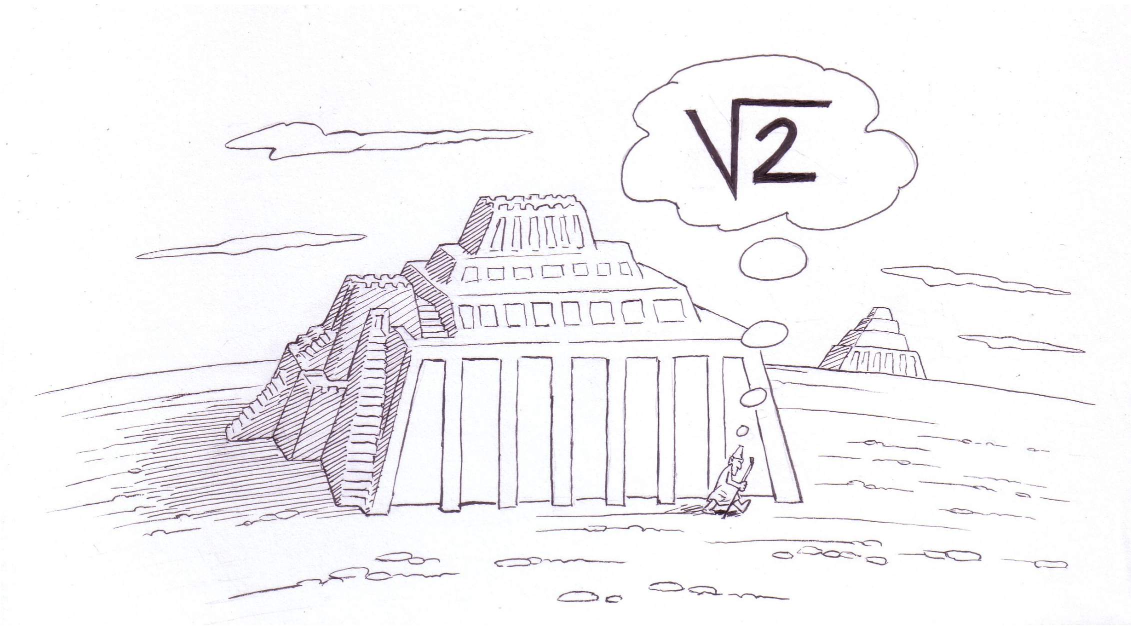
al even toevallig. Kortom, als alle decimalen willekeurig gekozen zijn uit $0, 1, \ldots, 9$ dan zijn zowel a_o als a_e ook weer willekeurig gekozen getallen in $[0, 1]$, en nog onafhankelijk van elkaar ook. Dus één (al of niet toevallig) getal maakt ook twee getallen[23]. Dat klinkt spectaculair en vreemd tegelijk. Of niet?

In de praktijk trekken we nooit een precies getal[24] en werken we bij reële getallen noodzakelijkerwijs met benaderingen. Als we de a hierboven benaderen door bijvoorbeeld de eerste tien decimalen te geven, dan hebben de a_e

[22] 6-normaal, of 10-normaal.
[23] Een eerste wiskundige versie van de wonderbaarlijke broodvermenigvuldiging?
[24] Tenzij we alleen maar met gehele getallen werken.

en a_o allebei maar vijf decimalen. Je krijgt dus twee getallen terug, maar wel met een kleinere nauwkeurigheid. Het aardige is natuurlijk dat dit verschil niet bestaat zodra je oneindig veel decimalen 'vastlegt', want de helft van oneindig is nog steeds oneindig. Als je daar een paradox in wilt zien dan is er hier nog eentje.

Opgave 2.6. Laat zien hoe je uit één toevallig gekozen getal a tussen 0 en 1 ook oneindig veel toevallig gekozen getallen kunt halen. Hint: gebruik de decimalen van a op de plaatsen $1, 3, 6, 10, 15, \ldots$ voor a_1, voor a_2 de decimalen op plaats $2, 4, 7, 11, 16, \ldots$, voor a_3 de decimalen op $5, 8, 12, 17, \ldots$, enzovoorts. Wat is de systematiek?



3 Worteltrekken met Newton

We verschuiven onze aandacht nu van getallen naar verbanden. Bij verbanden denk je aan grafieken, bijvoorbeeld in het xy-vlak. En bij grafieken denk je weer aan functies. En omgekeerd. Als de grafiek gegeven is door een functie f, dan geven de snijpunten met de horizontale as de nulpunten van f: de oplossingen van de vergelijking $f(x) = 0$, bijvoorbeeld de vergelijking $x^2 - 2 = 0$.

Vaak is het moeilijk om vergelijkingen exact op te lossen. Maar als je bij een gegeven benadering van de oplossing altijd een (veel) betere benadering kunt vinden dan kom je een heel eind. En dat is wat Newton deed, door raaklijnen aan de grafiek in het punt dat bij de eerste benadering hoort te snijden met de x-as.

We gebruiken de methode van Newton in dit hoofdstuk daarom als kapstok voor een eerste kennismaking met differentiaalrekening (lees: raaklijnen bepalen) en (discrete) dynamische systemen, met een terugblik naar de Oudheid die vanaf veilig bewaarde kleitabletten tot ons spreekt. Want $\sqrt{2}$ deden ze al heel lang geleden, met … hetzelfde algoritme als wij nu: rationale benaderingen van de wortel uit twee[1]. Daar begint dit hoofdstuk dan ook mee.

3.1 Grafieken en raaklijnen

Met de algebraïsche operaties $+, -, \times$ en $/$ maken we onze eerste functies[2]. Een veel gebruikte notatie voor functies zien we in dit voorbeeld:

$$f : x \to f(x) = x^2$$

Deze functie correspondeert met de x^2-knop op een bij voorkeur denkbeeldige rekenmachine[3]. Toets als input een x-waarde in, bijvoorbeeld $x = 3$, druk op de knop en het apparaat geeft je als output een 9 terug.

Op sommige denkbeeldige rekenmachines staat op de kwadraatknop niet x^2 maar $y = x^2$. De naam voor de outputwaarde is vaak y. Met de f in dit voorbeeld hoort bij input $x = 3$ dus een output $y = 9$. Bij iedere x maakt zo'n functie f dus een nieuwe waarde $f(x)$ met behulp van een formulevoorschrift voor f. In het voorbeeld is het voorschrift $f(x) = x^2$. Ander taalgebruik: f stuurt x naar $y = f(x)$. De x staat hier voor een vrij te kiezen getal op de getallenlijn. We spreken daarom van een functie van één reële variabele.

Het praten over functies zonder bijbehorend formulevoorschrift zullen we zo veel mogelijk vermijden. Bij voorkeur zien we iedere f als een knop op de DRM. Toets als input een x-waarde in, druk op de knop $f(x)$ of $y = f(x)$, en lees af wat de output $y = f(x)$ is. Als de knop het niet doet, dan zeggen we dat x niet in het definitiegebied van f zit. De knop $\frac{1}{x}$ doet het niet bij $x = 0$ bijvoorbeeld.

Van de knoppen e^x, $\ln x$ en ook $\sqrt{x}$ blijven we nog even af. We bekijken voorlopig alleen de rationale functies, functies die we maken met optellen, aftrekken, vermenigvuldigen en delen, zoals de denkbeeldige knop

$$F(x) = \frac{x}{2} + \frac{1}{x} = \frac{1}{2}(x + \frac{2}{x}), \tag{3.1}$$

die hoort bij een functie F die alleen $x = 0$ niet in zijn definitiegebied heeft.

Met functies heb je ook meteen vergelijkingen, bijvoorbeeld $f(x) = 0$ of, algemener, $f(x) = c$ voor een gegeven c, met de spreekwoordelijke onbekende x. Het oplossen van dit soort vergelijkingen is voor sommigen leuk, voor van alles en nog wat nuttig, en meestal lastig. Met de kwadraatfunctie en de

[1] Het daarbij gebruikte zestigtallig stelsel is een verhaal apart.
[2] Van het Latijnse fungor (*deponens:* een passieve vorm met actieve betekenis).
[3] Vanaf nu DRM genoemd.

keuze $c = 2$ krijgen we zo de vergelijking $x^2 = 2$, waarvan we al een positieve oplossing[4] gezien hebben, te weten $x = \sqrt{2}$.

Het aardige is dat $\sqrt{2}$ heel snel en heel nauwkeurig te benaderen is met rationale getallen, gebruikmakend van de F uit (3.1). De rij

$$1, F(1), F(F(1)), F(F(F(1))), \ldots$$

gaat razendsnel naar $\sqrt{2}$. Met andere woorden, voer $x = 1$ in als input op de DRM en druk steeds op de knop die bij F hoort. Dan wordt de output de nieuwe input die weer output geeft. In een paar stappen leidt dit tot verbluffend nauwkeurige rationale benaderingen van $\sqrt{2}$, die duizenden jaren geleden al werd gebruikt.

Isaac Newton[5] benaderde oplossingen van vergelijkingen van de vorm $f(x) = c$ via het herhaald trekken van raaklijnen aan de grafiek van de functie f. Wat we zullen zien is hoe deze methode bij iedere f een nieuwe F maakt die gebruikt wordt voor het numeriek[6] oplossen van de vergelijking $f(x) = 0$, hetgeen na het maken van F simpelweg neerkomt op het herhaald drukken[7] op de $F(x)$-knop van de DRM. In de praktijk is een paar keer drukken meestal voldoende.

Voor de grafiek van een functie f hebben we een assenkruis nodig. Daartoe nemen we met Descartes twee getallenlijnen, een voor x en een voor y, die elkaar loodrecht snijden in $x = y = 0$, en zo een vlak opspannen waarin we punten (x, y) kunnen tekenen. De grafiek van een functie f krijgen we dan door alle punten (x, y) in dit xy-vlak waarvoor geldt dat $y = f(x)$ te markeren. Wat slordig praten we zowel over de grafiek $y = f(x)$ als over de functie $y = f(x)$. De grafiek van de f gegeven door $f(x) = x^2$ is een parabool. Bekend verondersteld bij de lezer wordt welke functies als grafiek een recht lijn hebben, hoe de formulevoorschriften van die functies eruit zien, en hoe (makkelijk) de bijbehorende vergelijkingen worden opgelost. Deze bijzondere functies[8] noemen we lineaire functies.

Dit is misschien wel een moment om even stil te staan bij naamgeving en notatie. De letters x en y zijn vaak de namen voor de in- en outputvariabele van een functie f. De afspraak dat we bij het tekenen van de grafiek van f de x horizontaal en de y verticaal uitzetten, en in (x, y) eerst x en dan y schrijven, is maar een afspraak. Het is belangrijk dat je van elkaar weet aan welke afspraken je je houdt. Niet per ongeluk[9] de rol van x en y omdraaien.

---

[4] Probeer wellicht even te vergeten nu dat $\sqrt{2}$ al bestaat.

[5] Geniale wis- en natuurkundige, bovendien amateurtheoloog en, jawel, alchemist.

[6] Numeriek: bijvoorbeeld een decimale ontwikkeling. Hier: een rij breuken.

[7] En knoppen drukken leren we steeds vroeger.

[8] Met als grafiek een rechte lijn, al of niet door de oorsprong.

[9] Bezuidenhout Den Haag, 3-3-1945, zie de Nederlandse Wikipedia.

Ook voor functies gebruiken we letters, vaak f of g of F. Als we het over een functie met de naam f hebben, dan is bij voorkeur een formulevoorschrift van f voorhanden om te weten over welke f we het hebben.

Als de grafiek van een functie geen rechte lijn is dan kunnen we voor een gegeven punt kijken of er een rechte lijn door dat punt gaat die het 'beste' bij de grafiek past. Wat we daarmee bedoelen is dat als we de grafiek in dat punt onder de loep nemen, we vrijwel geen verschil zien tussen de rechte lijn en de grafiek. Hoe sterker de lens in de loep, hoe kleiner het verschil. Als zo'n lijn bestaat dan noemen we die lijn de raaklijn aan de grafiek door het gegeven punt op de grafiek.

De lineaire functie waarvan de grafiek die raaklijn is, heet de lineaire benadering van f rond de x-waarde van dat punt. Als dat punt $(a, b) = (a, f(a))$ is, dan is iedere niet-verticale lijn door (a, b) van de vorm

$$y = b + r(x - a) \tag{3.2}$$

en gaat het dus om het vinden van de beste r, en dat kan op een systematische manier, namelijk met de differentiaalrekening van Newton en Leibniz. De beste r is wat op school vroeger de richtingscoëfficiënt[10] RiCo heette, en RiCo wordt meestal geschreven als $f'(a)$, de afgeleide van f in a.

Differentiaalrekening voor functies wordt op school meestal uitgelegd aan de hand van limieten van differentiequotiënten. Maar limieten zijn moeilijk, moeilijker dan staartdelingen bijvoorbeeld. Om limieten te vermijden gebruiken we variaties op het merkwaardig product

$$a^2 - b^2 = (a - b)(a + b),$$

zoals

$$a^5 - b^5 = (a - b)(a^4 + a^3b + a^2b^2 + ab^3 + b^4). \tag{3.3}$$

Deze formule kan door uitvermenigvuldigen en haakjes verdrijven geverifieerd worden, maar ook afgeleid worden met een staartdeling. Het eerste stukje van die staartdeling[11] staat hieronder uitgeschreven:

$$\begin{array}{lll} a - b \quad / & a^5 + 0\,a^4 + 0\,a^3 + 0\,a^2 + 0\,a - b^5 & \backslash \quad a^4 + a^3b + \cdots \\ & \underline{a^5 - a^4b} & \\ & \quad\quad a^4b & \\ & \quad\quad \underline{a^4b - a^3b^2} & \\ & \quad\quad\quad\quad a^3b^2 & \end{array}$$

[10] Tegenwoordig hellingsgetal, van hellende lijnen en hellende vlakken.

[11] Ook wel hapjesmethode genoemd.

Opgave 3.1. Maak deze staartdeling af, en reflecteer op hoe a de rol van 10 hier heeft overgenomen. Instructief is om eerst eens 100000-1=99999 door 9 te (staart)delen, waarbij je 99999 decimaal schrijft als

$$99999 = 100000 - 1 = 1 \times 10^5 + 0 \times 10^4 + 0 \times 10^3 + 0 \times 10^2 + 0 \times 10^1 - 1 \times 10^0,$$

met andere woorden, een decimale ontwikkeling waarin de cijfers ook negatief mogen zijn.

We zullen zien hoe je voor een punt op de grafiek van een rationale functie RiCo snel kunt vinden, en dat je daar niet eerst voor hoeft te worden lastig gevallen met alle finesses van het lastige begrip limiet.

Als voorbeeldje om de smaak te pakken te krijgen nemen we de functie f met het formulevoorschrift $f(x) = x^5$. De grafiek van deze functie bestaat uit alle punten (x, y) waarvoor $y = x^5$. Als we nu voor een willekeurige waarde van x, zeg $x = a$, willen weten wat de raaklijn is aan de grafiek door het punt $(x, y) = (a, a^5)$ dan doen we dat met het merkwaardig product (3.3) hierboven. We vervangen[12] in (3.3) de letter a door x en b door a en herschrijven het resultaat als

$$x^5 = a^5 + (x^4 + ax^3 + a^2x^2 + a^3x + a^4)(x - a). \tag{3.4}$$

De grafiek van van deze vijfdemachtsfunctie f wordt dus gegeven door

$$y = f(x) = x^5 = a^5 + \underbrace{(x^4 + ax^3 + a^2x^2 + a^3x + a^4)}_{\text{reduceert tot } 5a^4 \text{ als } x=a}(x - a), \tag{3.5}$$

en de raaklijn kan dus moeilijk iets anders zijn dan de lijn gegeven door

$$y = a^5 + 5a^4(x - a), \tag{3.6}$$

en voor nu is dat ons meer dan goed genoeg. De lineaire functie L die hierbij hoort is

$$L(x) = L_a(x) = a^5 + 5a^4(x - a).$$

De a zit als parameter in de functie. We hebben de afgeleide van $f(x) = x^5$ gevonden zonder over limieten te praten, en we zullen de betekenis ook precies maken zonder limieten[13]. Daartoe moet het verschil tussen $f(x)$ en $L_a(x)$ eigenschappen hebben die voor de grafiek $y = L_a(x)$ het gebruik van het woord raaklijn in $x = a$ aan de grafiek $y = f(x)$ rechtvaardigen.

Opgave 3.2. Bereken RiCo voor de raaklijn aan $y = x^3$ door het punt $(1, 1)$ met de boven beschreven methode.

[12] We hadden ook b als naam voor de variabele kunnen kiezen.

[13] Hoofdstuk 10: verschil rechterleden (3.5) en (3.6) is deelbaar door $(x - a)^2$!

3.2 Newtons methode

Newtons methode om $\sqrt{2}$ te benaderen met breuken gaat uit van de grafiek van de functie $f : x \to x^2 - 2$ in het xy-vlak. Die grafiek is de parabolische kromme met vergelijking $y = x^2 - 2$[14]. Lees de laatste voetnoot die hier geen exponent is. Je begint met een punt op de grafiek, bijvoorbeeld $(x, y) = (1, -1)$. Neem een slim gekozen lijn door dat punt, en snijd die met de x-as. Als je lineaire vergelijkingen kunt oplossen dan lukt dat[15]. Vervolgens neem je bij de x-waarde van het snijpunt het bijbehorende punt op de grafiek, en kies je daardoor weer een slim gekozen lijn die je weer snijdt met de x-as. Zo ga je door.

Newtons slimme keuze voor de lijn die je in elke stap nodig hebt, bestaat eruit dat je steeds de raaklijn door het punt aan de grafiek neemt. Met deze keuze blijkt dat je in een paar stappen zo dicht bij $\sqrt{2}$ zit dat de fout in de benadering verwaarloosbaar is. Er zijn meer keuzes waarmee dat laatste lukt, maar die vergen veel meer stappen. Je kunt bijvoorbeeld de lijn steeds van de vorm $y = 2x + b$ nemen, met b zo gekozen dat de lijn door het punt op de grafiek gaat. Dat werkt ook, maar veel minder snel dan met Newtons dynamische keuze, die gebruikt hoe de grafiek er lokaal uitziet in het punt waardoor je steeds de lijn moet kiezen. De methode van Newton werkt in principe voor alle functies die een beetje netjes zijn.

In het speciale geval van $f : x \to x^2 - 2$ zul je zien dat Newtons methode neerkomt op het itereren van formule (3.1), waarbij je de facto steeds het gemiddelde neemt van x en $\frac{2}{x}$. Zonder te denken in termen van grafieken en raaklijnen kun je natuurlijk ook al op dit idee komen, want je zoekt immers een getal x waarvoor x en $\frac{2}{x}$ gelijk zijn. Is x te groot dan is $\frac{2}{x}$ te klein, dus ergens tussen x en $\frac{2}{x}$ moet je zoeken, en het gemiddelde van de twee ligt als eerste poging voor de hand, zoals ook blijkt uit een duizenden jaren oud Babylonisch kleitablet[16] met sexagecimale benaderingen van $\sqrt{2}$. Je vindt het tablet wel op Wikipedia. Proefondervindelijk blijkt deze eerste poging een meer dan gelukkige keuze, zoals zal blijken uit een analyse van de methode. Zoals je nog in de opgaven zult zien leidt de methode voor de vijfdemachtswortel uit 3 tot

$$F(x) = \frac{4x}{5} + \frac{3}{5x^4}, \tag{3.7}$$

niet het gewone[17], maar een gewogen gemiddelde van x en $\frac{3}{x^4}$.

[14] Maak een plaatje voor het vervolg!

[15] Tenzij je als lijntrekker de lijn horizontaal genomen hebt.

[16] Wiskunde is na het torenfiasco de enig overgebleven echt universele taal.

[17] Hetgeen toch wel verrassend is, zie Opgave 3.14.

Opgave 3.3. De functies gedefinieerd door (3.1) en (3.7) lijken op elkaar. Kun je met behulp van je DRM een schets maken van de bijbehorende grafieken? Als je een grote positieve startwaarde x_0 neemt, dan vind je proefondervindelijk in beide gevallen dat

$$x_0 > F(x_0) = x_1 > F(x_1) = x_2 > F(x_2) = x_3 > \cdots$$

Kun je inzien waarom dat zo is en wat er gebeurt? Grafisch? Algebraïsch? Heb je nodig dat de wortel waar het om gaat al bekend is?

3.3 Raaklijnen met algebra

We hebben het net al informeel over raaklijnen gehad. Nu doen we dat iets preciezer. Welke lijn door het punt $(x, y) = (1, -1)$ verdient de naam

$$\text{raaklijn aan } y = x^2 - 2 \text{ in het punt } (x, y) = (1, -1)?$$

Het Engelse woord voor raaklijn is tangent[18]. Volgens the Oxford Advanced Learner's Dictionary, 'the dictionary you can trust' als we de cover moeten geloven, is een tangent 'a straight line that touches the outside of a curve but does not cross it (geometry)'. Deze beschrijving is misschien intuïtief duidelijk, maar wiskundig precies noch correct, en biedt weinig perspectief voor lastigere zaken[19]. Wat is een raakvlak aan een torus bijvoorbeeld, om maar wat te noemen.

Zuiver of toegepast[20], goede wiskunde onderscheidt zich van slechte wiskunde met name zo door duidelijk te maken waar we het eigenlijk over hebben. Daarom de vraag: welke lijn door het punt $(x, y) = (1, -1)$ is in dat punt bijzonder met betrekking tot de grafiek $y = x^2 - 2$, en wel in die zin dat de lijn en de grafiek in de buurt van $(x, y) = (1, -1)$ het meest op elkaar lijken?

Op een na alle lijnen door $(x, y) = (1, -1)$ worden gegeven door vergelijkingen van de vorm[21]

$$y = -1 + r(x - 1), \tag{3.8}$$

met r een vrij te kiezen reëel getal. Alleen de lijn met vergelijking $x = 1$ kan zo niet weergegeven worden, want dat is de enige verticale lijn door $(x, y) = (1, -1)$. Die bijzonderheid heeft echter niets met de grafiek van f te maken en daarom laten we deze lijn verder buiten beschouwing. Voor

[18] Van tangere, Latijn voor aanraken.
[19] De opmerking die je wist dat zou komen.
[20] Vaag en vaak vals onderscheid.
[21] Nu liever niet denken in termen van $y = ax + b$ als het om lijnen gaat.

elke gegeven reële r is de lijn met vergelijking (3.8) keurig de grafiek van een functie, namelijk van

$$L:\ x \to -1 + r(x-1).$$

Als naam voor de functie gebruiken we hier L, maar let op, L hangt ook van de keuze van r af. Wat r ook is, L heeft de eigenschap dat hij als fungor met $x = 1$ hetzelfde doet als f, de output is hetzelfde:

$$L(1) = f(1) = -1.$$

De vraag voor welke r de lijn (3.8) bijzonder is met betrekking tot de grafiek van f vertaalt zich als: welke L lijkt in zijn werking als functie het meest op f als we hem loslaten op x-waarden in de buurt van $x = 1$?

Als we de gemeenschappelijke output waarde $L(1) = f(1) = -1$ aftrekken van $L(x)$ en van $f(x)$ dan zien we hun beider werking als, respectievelijk,

$$L(x) - L(1) = r(x-1), \tag{3.9}$$

$$f(x) - f(1) = x^2 - 1 = (x+1)(x-1). \tag{3.10}$$

De gemeenschappelijke factor in de rechterleden van (3.9) en (3.10) is $x - 1$. Deze factor is klein voor x-waarden in de buurt van $x = 1$. In (3.10) is de voorfactor $x + 1$ ongeveer gelijk aan 2 als x ongeveer 1 is. De overeenkomst tussen de rechterleden (3.9) en (3.10) is in de buurt van $x = 1$ dus maximaal als we $r = 2$ nemen in (3.9). Het verschil tussen de voorfactoren is dan klein, en het verschil tussen de rechterleden van (3.9) en (3.10) is nog veel kleiner, vanwege de, voor x-waarden in de buurt van $x = 1$, kleine gemeenschappelijke factor $x - 1$.

Met $r = 2$ is de grafiek van L wat in de wiskundige volksmond de raaklijn[22] aan de grafiek van f in $(x, y) = (1, -1)$ genoemd wordt. De waarde $r = 2$ heet de richtingscoëfficiënt[23] van de (raak)lijn. In dit geval is de raaklijn intuïtief de lijn die je van onderen net rakend tegen de grafiek aan legt. De Oxford Advanced Learners Dictionary heeft het dus niet altijd fout. Hoe het ook zij, de lijn die iedereen de raaklijn noemt aan onze voorbeeldgrafiek in het punt met $x = 1$, is de grafiek van de lineaire functie L, gegeven door

$$L(x) = -1 + 2(x-1).$$

Om te kijken hoe goed $L(x)$ de functiewaarden $f(x)$ benadert kijken we naar het verschil van beide functies. Met (3.9) en (3.10) vinden we

$$f(x) - L(x) = (x-1)^2$$

[22] In het Engels 'tangent line'.

[23] Eerder RiCo genoemd, in het Engels: 'slope'. Abusievelijk soms: 'gradient'.

als de fout in de benadering. Deze restterm noemen we $R(x)$. Met links alleen $f(x)$ en rechts de lineaire benadering $L(x)$ en de restterm $R(x)$ hebben we dan

$$f(x) = \underbrace{-1 + 2(x-1)}_{L(x)} + \underbrace{(x-1)^2}_{R(x)}.$$

De restterm $R(x)$ is klein in de buurt van $x = 1$. Hoe klein? Relatief klein in verhouding tot beide twee termen in $L(x)$. Hoe dichter x bij 1 komt hoe beter. Met $x = 1 + \frac{1}{10}$ staat er

$$f(1 + \frac{1}{10}) = \underbrace{-1 + \frac{2}{10}}_{L(1+\frac{1}{10})} + \frac{1}{100},$$

en met $x = 1 - \frac{1}{100}$

$$f(1 - \frac{1}{100}) = \underbrace{-1 - \frac{2}{100}}_{L(1-\frac{1}{100})} + \frac{1}{10000}.$$

Een x die dicht bij 1 zit wordt vaak geschreven als $x = 1 + h$. De rol van h wordt hopelijk inzichtelijker met af en toe een expliciete keuze voor h. Een x die dicht bij 0 zit blijven we soms liever x noemen natuurlijk[24]. Hieronder wat opgaven om met lineaire benaderingen en resttermen te oefenen. Variaties op het thema. Eerst voor monomen, dan voor polynomen en rationale functies. Als je de factorisatie niet ziet: de staartdeling doet het altijd.

Opgave 3.4. Herhaal het argument hierboven voor (de grafiek van) de functie $f : x \to x^2$. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek in $(x, y) = (1, 1)$? Hoe zit het met de restterm?

Opgave 3.5. Herhaal het argument hierboven voor (de grafiek van) de functie $f : x \to x^3$. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek in $(x, y) = (1, 1)$? Hint: factoriseer $x^3 - 1$ in twee factoren[25].

Opgave 3.6. Herhaal het argument hierboven voor (de grafiek van) de functie $f : x \to x^4$. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek in $(x, y) = (1, 1)$? Hint: factoriseer $x^4 - 1$ in (maar) twee factoren[26].

[24] Heen en weer switchen tussen x en $\frac{1}{10}$ is een leerzame rode draad in dit boekje.
[25] Gebruik (3.3) met 3 in plaats van 5.
[26] Gebruik (3.3) met $5 = 4$.

Opgave 3.7. Herhaal het argument hierboven voor (de grafiek van) de functie $f: x \to x^5$. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek in $(x, y) = (1, 1)$? Hint: factoriseer $x^5 - 1$ in twee factoren[27].

Opgave 3.8. Hiermee is de regelmaat wel duidelijk. Formuleer welk systeem je in deze rijtjessommen hebt ontdekt en behandel dezelfde vragen voor de raaklijnen aan de grafieken in de punten met $x = a$. Hint: factoriseer $x^n - a^n$, voor $n = 1, 2, 3, \ldots$. Maak je conclusies algemeen voor $f: x \to x^n$.

Opgave 3.9. Verzin een manier om de merkwaardige producten nooit meer te vergeten.

Opgave 3.10. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek van $f: x \to x^5 - 2x^3 + x + 1$ in het punt met $x = a$? Probeer het nu eens meteen met een staartdeling, eerst met $a = 2$ bijvoorbeeld.

Opgave 3.11. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek van $f: x \to \frac{1}{x}$ in het punt met $x = 1$? En in het punt met $x = a$ $(a \neq 0)$?

Opgave 3.12. Wat is de richtingscoëfficiënt van de raaklijn aan de grafiek van $f: x \to \frac{1}{x^2}$ in het punt met $x = 1$? En in het punt met $x = a$ $(a \neq 0)$?

Opgave 3.13. Als je de functie $f: x \to x^{\frac{1}{2}} = \sqrt{x}$ even voor vanzelfsprekend neemt, kun je die dan ook (differentiëren) zoals je dat hierboven deed?

Samenvattend: we kunnen in punten op grafieken van eenvoudige functies f die we maken met rationale formulevoorschriften[28], de raaklijn aan de grafiek bepalen in elk punt $(x, y) = (a, f(a))$, als $f(a)$ maar gedefinieerd is[29]. De

[27] Gebruik (3.3) met $5 = 5$.

[28] Waarin we dus alleen $+, -, \cdot, /$ gebruiken, $x \to |x|$ doet hier niet mee.

[29] Lees: als we niet door nul gedeeld hebben voor de inputwaarde $x = a$.

RiCo van de raaklijn, hierboven steeds r genoemd, wordt bepaald door a en f, genoteerd met $f'(a)$ en uitgesproken als f accent van a. De definitie van de afgeleide functie f' via

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \tag{3.11}$$

bij differentiaalrekening, die meestal pas komt nadat limieten veel uitgebreider[30] zijn behandeld voor veel algemenere functies, is hiermee consistent. We noemen de RiCo daarom vanaf nu ook $f'(a)$, de afgeleide van f in a.

Wat we nog niet kunnen, is het vinden van raaklijnen (inclusief een goede definitie) in punten op krommen in het vlak die op een andere manier gegeven zijn (door andere functies, of als vergelijking, bijvoorbeeld $x^3+y^3 = xy$). Wat we gedaan hebben, wijst wel in de juiste richting: je moet de lijn nemen die in de buurt van het punt zo veel mogelijk op de kromme lijkt. Een intuïtieve manier om dit te formuleren is dat bij inzoomen op het punt de kromme steeds rechter wordt en er in de limiet een rechte lijn verschijnt[31]. Als dit lukt heet de kromme differentieerbaar, anders niet. En natuurlijk moet de definitie ook niet afhangen van hoe we het assenstelsel op ons papier tekenen.

Dit alles is van later zorg. Voorlopig zijn we al erg blij met de definitie van eerst $f'(1)$ en daarna $f'(a)$ als RiCo van de raaklijn aan de grafiek in het punt met $x = a$ voor een willekeurige a, als f gegeven wordt door $f(x) = x^n$ met n geheel, en simpele combinaties daarvan. In het bijzonder weten en begrijpen we nu dat

$$f(x) = x^n \quad \Rightarrow \quad f'(x) = nx^{n-1}. \tag{3.12}$$

In de volgende sectie gaan we onze tot zover opgebouwde expertise over raaklijnen gebruiken om nulpunten te zoeken van functies waarmee we nu uit de voeten kunnen. In het bijzonder gaan we worteltrekken, n-de machtswortels uit positieve getallen. Anders gezegd, voor gegeven $y > 0$ lossen we de vergelijking $x^n = y$ op voor $n = 2, 3, 4, \ldots$. We maken zo de inverse functie van (3.12), en de afgeleide van de inverse functie weten we dan ook. Want raaklijnen blijven raaklijnen, de grafiek is immers hetzelfde.

3.4 Nulpunten benaderen

Newtons methode om nulpunten van een functie f te vinden gebruikt de formule

$$y = f(a) + f'(a) \cdot (x - a) \tag{3.13}$$

[30] Maar ook niet precies.

[31] Inzoomen moet recht maken wat krom is.

voor de raaklijn aan de grafiek in het punt met x-coördinaat a. We hebben hier de letter a gebruikt voor het punt waarin we de raaklijn aan de grafiek bepalen, en de letter x voor de variabele in de lineaire functie die als grafiek die raaklijn heeft. Het snijpunt van de raaklijn met de x-as is dan een vergelijking voor x en als we die oplossen kunnen we de dan gevonden x als nieuwe a zien[32]. Dat gaat als volgt.

Het snijpunt met de x-as vinden we door $y = 0$ te nemen in (3.13). Als $f'(a) \neq 0$ volgt dat

$$x = a - \frac{f(a)}{f'(a)}. \tag{3.14}$$

Het rechterlid van (3.14) gebruiken we als formulevoorschrift voor een nieuwe functie, met een mogelijk beperkter definitiegebied. We noemen deze nieuwe functie F, dus[33]

$$F : a \to a - \frac{f(a)}{f'(a)}. \tag{3.15}$$

Newtons methode kiest (soms met enig verstand) een startwaarde a, zeg $a = a_0$, gebruikt (3.14) om x uit te rekenen, en stopt het resultaat voor x als een nieuwe $a = a_1$ weer in het rechterlid van (3.14) om de volgende x uit te rekenen, die als $a = a_2$ weer in het rechterlid van (3.14) gaat. Dit kan herhaald worden zo vaak als mogelijk of gewenst is. We krijgen dus een rij a-tjes geïndexeerd door $n = 0, 1, 2, \ldots$.

Het is op dit moment dat de schrijvers bij herlezing de beslissing nemen dat ze die rij straks toch echt liever x_n willen noemen. Search and replace does the job[34] *vanaf de eerstvolgende genummerde formule.* In termen van F is Newtons methode niets anders dan het uitrekenen van

$$a_1 = F(a_0),$$

$$a_2 = F(a_1) = F(F(a_0)),$$
$$a_3 = F(a_2) = F(F(a_1)) = F(F(F(a_0))),$$

enzovoorts. Dit noemen we het itereren van F, kort opgeschreven als (met vanaf hier x_n in plaats van a_n)

$$x_{n+1} = F(x_n), \quad n = 0, 1, 2, 3, \ldots, \tag{3.16}$$

met startwaarde x_0. We noemen (3.16) een (deterministisch) dynamisch systeem met discrete tijd[35].

---

[32] De a als oude x had ook gekund.
[33] a of x in de definitie van F: what's in a name?
[34] Als dat maar goed gaat…
[35] Discreet: hier bedoeld als tegenhanger van continu.

Opgave 3.14. Laat, al of niet met behulp van de moderne kennis van de differentiaalrekening, zien dat Newtons methode voor $f(x) = x^5 - 3$ leidt tot het itereren van de functie F met formulevoorschrift

$$F(x) = \frac{4x}{5} + \frac{3}{5x^4}, \tag{3.17}$$

voor het benaderen van de vijfdemachtswortel uit 3.

Opgave 3.15. Welke F gebruik je voor de elfdemachtswortel uit 42?

Opgave 3.16. Bij een gegeven functie F en een rij punten gedefinieerd door (3.16) kunnen we in het xy-vlak de diagonaal $y = x$ en de grafiek $y = F(x)$ tekenen, alsmede de punten $P_0 = (x_0, x_0)$, $P_1 = (x_1, x_1)$, $P_2 = (x_2, x_2)$, $P_3 = (x_3, x_3)$, ... op de diagonaal. Geef een beschrijving van hoe elk volgend punt P_{n+1} bepaald wordt in termen van de diagonaal, de grafiek en het punt P_n.

Opgave 3.17. Als in Opgave 3.16 de functie F een dekpunt B heeft, dat wil zeggen, een punt B met $F(B) = B$, dan schuiven we de grafiek evenwijdig aan de diagonaal zo op dat het punt (B, B) in de oorsprong terechtkomt. Deze verschoven grafiek is weer de grafiek van een functie. Van welke functie?

Let wel, om dit laatste te kunnen doen moeten we die B wel eerst hebben. In vrijwel alle gevallen treden we dan uit de klasse van functies die gemaakt zijn met de gewone rationale getallen en de standaardbewerkingen! We gebruiken deze truc dan ook in de volgende sectie alleen maar om te laten zien dat, als we al weten dat Newtons methode het doet, hij het automatisch ook heel goed doet.

3.5 Dynamica

Elke functieknop op de rekenmachine, de kwadraatknop bijvoorbeeld, maar ook de cosinusknop (wat de cosinus ook moge zijn als functie[36]) definieert een (weliswaar numerieke benadering van) een Discreet dynamisch systeem. Het dynamisch systeem gegeven door de functie F bestuderen we door te kijken naar het gedrag van de door (3.16) bepaalde rij

$$x_0, x_1, x_2, x_3, \ldots$$

[36] Zie echter hoofdstuk 10.

in principe voor elke keuze van de startwaarde x_0. We verwachten dat de a-tjes[37] steeds dichter bij het gezochte nulpunt van de functie f komen te liggen. Newtons functie F is door middel van (3.15) immers niet zomaar gemaakt. In het bijzonder heeft Newtons F de eigenschap dat $F(a) = a$ equivalent is met $f(a) = 0$, althans voor a met $f'(a) \neq 0$. Er zijn natuurlijk bij f veel meer F-en te maken die de nulpunten van f als dekpunten hebben, zet maar een coëfficiënt voor het quotiënt in (3.15) bijvoorbeeld, of neem een andere deler dan $f'(a)$, een constante of zo, maar die andere F-en doen het niet zo goed.

Bij elke functie F, al of niet verkregen uit een f waarvan we de nulpunten zoeken, ligt de vraag wat de dekpunten van F zijn als eerste voor de hand. Die dekpunten zijn namelijk de evenwichtstoestanden en het bepalen daarvan is essentieel in de analyse van het dynamisch systeem dat bij F hoort. Gegeven zo'n evenwicht is de natuurlijke volgende vraag wat er gebeurt als je het evenwicht een klein beetje (of heftig) verstoort. Zulke stabiliteitsvragen geven aanleiding tot diverse definities van lokale en globale stabiliteit van evenwichten.

Precieze definities van de diverse begrippen laten we achterwege, maar hopelijk is in de voorbeelden wel duidelijk wat wat moet zijn en dat is voorlopig genoeg. Goede precieze definities volgen sowieso meestal pas uit inspectie van voorbeelden en wat je daarbij waarneemt, en de vraag wat Newtons F nu zo bijzonder maakt, kan aan de hand van voorbeelden goed aan de orde komen.

Daartoe kijken we eerst naar voorbeelden van functies F waarvoor $x = 0$ een dekpunt is. Op grond van Opgaven 3.16 en 3.17 is immers elke functie met een dekpunt $x = B$ te veranderen in een functie met dekpunt $x = 0$, waarvoor de stabiliteitseigenschappen moeilijk anders kunnen zijn dan die van $x = B$ voor de oorspronkelijke functie F.

Opgave 3.18. Laat $F(x) = rx$. Onderzoek voor $r = \frac{1}{2}$, $r = 1$, $r = -1$, $r = 2$, $r = -\frac{1}{3}$ de stabiliteit van het evenwicht $x = 0$. Geef steeds aan voor welke startwaarden de iteraties naar 0 convergeren.

Opgave 3.19. Dezelfde vraag voor $F(x) = x^2$, $F(x) = 2x^2$ en $F(x) = -2x^2$.

Opgave 3.20. In de vorige opgaven is de snelheid waarmee iteraties convergeren naar het evenwicht nog niet aan de orde geweest. Bestudeer in dit verband de rol van de positieve gehele exponent n als $F(x) = x^n$ met $n = 1$ en $n = 2$.

[37] Correctie, de x-jes…

3.6 Waarom doet Newtons methode het zo goed?

Laten we nog eens kijken naar Newtons methode voor het trekken van de vijfdemachtswortel uit het getal 3. De f waarmee we F maken, heeft maar één nulpunt: het positieve getal dat tot de vijfde macht verheven het getal 3 geeft. We noemen dit getal hieronder B, de unieke positieve oplossing[38] van de vergelijking $B^5 = 3$.

Opgave 3.21. Laat voor F gegeven door (3.17), $B > 0$ de unieke oplossing zijn van $B^5 = 3$. Laat met behulp van (3.3) zien dat

$$F(x) - B = x - B - \frac{x^5 - B^5}{5x^4} = (x - B)^2(\frac{4}{5x} + \frac{3B}{5x^2} + \frac{2B^2}{5x^3} + \frac{B^3}{5x^4}).$$

Het rechterlid vertelt ons meteen al dat $F(x)$ een minimum heeft in $x = B$ ter grootte van $F(B) = B$. In het bijzonder zien we ook dat de tweede factor in het rechterlid in Opgave 3.21 ongelijk nul is voor x dicht bij B, terwijl de eerst factor dan (kwadratisch) klein is. We concluderen hieruit dat de beste kwadratische benadering van F rond $x = B$ gegeven wordt door

$$B + \frac{4 + 3 + 2 + 1}{5B}(x - B)^2.$$

Deze benadering is bijzonder omdat de lineaire factor $x - B$ er niet in voorkomt. Met de opgaven hierboven kunnen we de consequenties daarvan zien. Als we de grafiek van F oppakken en zo naar de oorsprong schuiven dat (B, B) in de oorsprong uitkomt, zie Opgaven 3.16 en 3.17, dan krijgen we de grafiek van een functie die zijn minimum nul in $x = 0$ aanneemt, met als kwadratische benadering rond $x = 0$ de nu puur kwadratische term

$$\frac{4 + 3 + 2 + 1}{5B}x^2 = \frac{2}{B}x^2.$$

Zodra je bij het itereren in de buurt van $x = 0$ zit pluk je in elke stap de vruchten van het kwadraat. Dat is de reden dat Newtons methode het zo goed doet: er zit geen lineaire term in de benadering van F rond zijn dekpunt. Hiermee hebben we de methode behoorlijk goed begrepen, toch?

[38] Dat B bestaat nemen we aan; is het duidelijk dat er dan maar één zo'n B is?

3.7 Waarom doet Newtons methode het überhaupt?

Er wringt nog iets. Opgave (3.17) gebruikte het bestaan van de oplossing en dat laat een hamvraag onbeantwoord. Hoe weten we dat Newtons methode convergeert voordat we weten dat het nulpunt bestaat? De vraag is hoe moeilijk we hierover willen doen. Wellicht niet bij gewoon worteltrekken, maar deze opgave is te leuk om niet te doen.

Opgave 3.22. Laat n een natuurlijk getal groter dan 1 zijn en $y > 0$ een reëel getal. Laat zien dat Newtons methode toegepast op $f_n(x; y) = x^n - y$ als te itereren functie $x \to F_n(x; y)$ geeft met

$$F_n(x; y) = (1 - \frac{1}{n})x + \frac{y}{nx^{n-1}},$$

en laat zien dat voor $x_0 > 0$ geldt dat

$$x_0^n > y \iff F_n(x_0; y) < x_0.$$

De benaderende rij voor de oplossing van $x^n = y$ is weer $x_1 = F_n(x_0; y)$, $x_2 = F_n(x_1; y)$, $x_3 = F_n(x_2; y), \ldots$ Als x_0 te groot is om een oplossing te zijn dan is x_1 dus kleiner dan x_0. Laat zien dat x_1 nog steeds te groot is door $x_1^n > y$ aan te tonen. Hint: herschrijf de aan te tonen ongelijkheid als

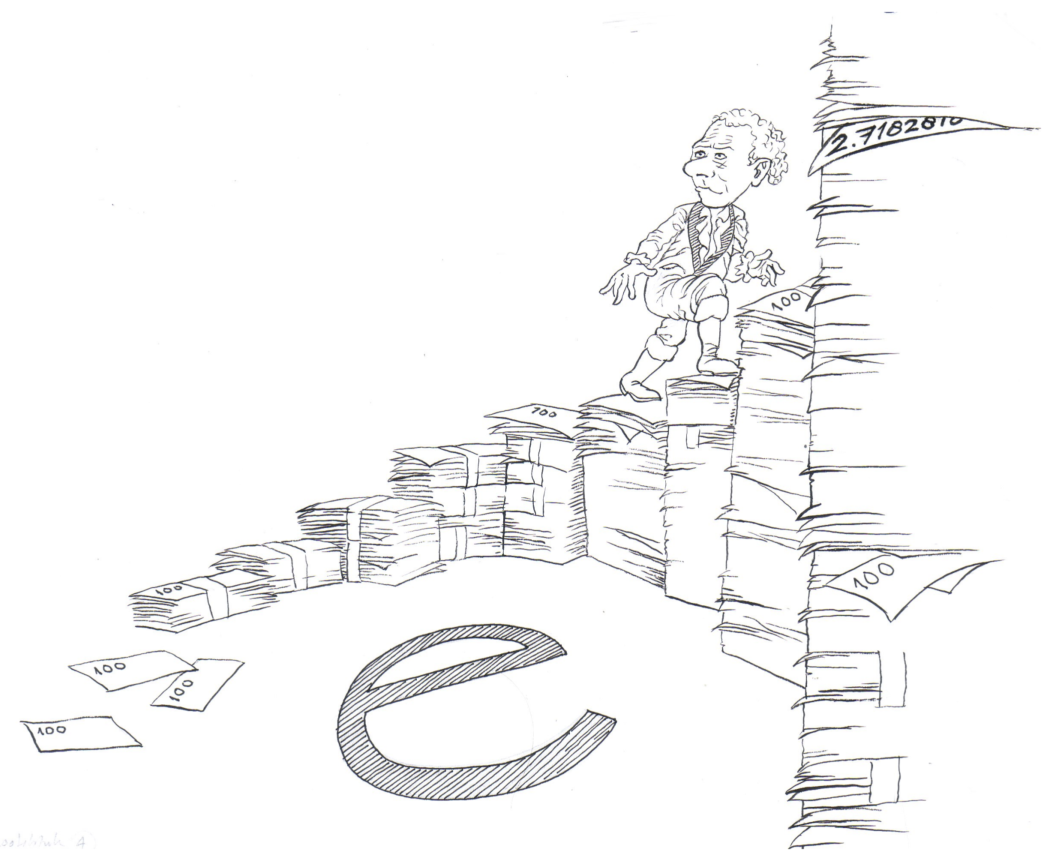
$$(n - 1 + s)^n > n^n s \quad \text{voor} \quad 0 < s = \frac{y}{x_0^n} < 1,$$

en laat zien dat deze ongelijkheid ook werkelijk geldt[39]. Vervolgens weet je dat als x_0 te groot was $x_0, x_1, x_2, \ldots$ een strikt dalende rij positieve getallen is. Deze rij heeft dus een grootste ondergrens. Noem die ondergrens $\bar{x}$ en beredeneer uitgaande van

$$x_{k+1} = (1 - \frac{1}{n})x_k + \frac{y}{nx_k^{n-1}}, \quad (k = 0, 1, 2, \ldots)$$

dat $\bar{x}^n = y$. Je hebt dus nu de x gevonden waarvoor geldt dat $x^n = y$. Hiermee is de inverse functie $y \to \sqrt[n]{y}$ van $x \to x^n$ gemaakt, met als definitiegebied de positieve reële getallen, zoals verwacht natuurlijk.

[39] Lastig. Hint: zet $s = 1 - t$ en gebruik een merkwaardig produkt voor $n^n - (n - t)^n$.



4 Rente op rente en het getal e

Banken[1] keren rente uit op het geld dat je bij hen wegzet. Bij sommige spaarrekeningen gebeurt dat op jaarbasis met een van te voren vastgestelde rentefractie. Banken gebruiken liever het woord percentage, een woord dat we hier willen vermijden. Je krijgt de (rente)fractie door het (rente)percentage door 100 te delen. Dus 30% van iets betekent een fractie $0{,}3 = \frac{3}{10}$ van dat iets, 50% van alle gevallen betekent de helft van alle gevallen. In 100% van alle gevallen waarin het woord percentage gebruikt wordt is het gebruik van dat woord overbodig[2]. Het simpelste voorbeeld, een rentepercentage van 100%,

[1] Inderdaad: and now something completely different.

[2] Met deze zin als unieke uitzondering op de regel.

betekent het verdubbelen van je saldo aan het eind van elke renteperiode. Met inleg 1 ontstaat zo de rij

$$1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, \ldots,$$

de meest voor de hand liggende eerste kennismaking met exponentiële groei, uitgaande van de pure schoonheid van de gehele getallen, met de mooiste machten die er bestaan. Hier weergegeven van twee tot de macht nul tot en met twee tot de macht tien, een rij die iedereen in het onderwijs moet zien.

Het andere grondtal dat naast deze 2 een plek verdient[3] is niet 10 maar e, de e van e^x op de rekenmachien. Met Euler zullen we zien dat dit het enige grondtal[4] is dat we met goede reden wel willen verheffen tot alle andere machten die we maar kunnen verzinnen. Voor de hand liggende vragen over rente op rente en continue rentebijschrijving wijzen ons hier de weg en een ondertitel van dit centrale hoofdstuk had kunnen zijn: hoe een banale context tot iets moois kan leiden.

4.1 Rente op rente

Als de rentefractie op jaarbasis gelijk is aan a dan wordt je tegoed elk jaar met een factor

$$1 + a$$

vermenigvuldigd. Dat oogt waarschijnlijk veel aangenamer als start voor het verhaal dan (vermenigvuldigen met) een factor $1 + \frac{A}{100}$ (bij rentepercentage A). Met rentefractie $a = 1$ wordt je geld jaarlijks verdubbeld.

Banken werken vaak met maandelijkse bijschrijvingen. Stel de bank zegt, ja onze rentefractie is misschien maar a, bijvoorbeeld[5] $a = \frac{1}{15}$, maar we keren wel elke maand uit. Als ze je niet besodemieteren[6], dan wordt aan het eind van elke maand je tegoed dus met een factor

$$1 + \frac{a}{12}$$

vermenigvuldigd. Na een maand is dat natuurlijk minder dan wat je normaal met een factor $1 + a$ na een jaar zou krijgen, want de rentefractie op maandbasis is natuurlijk twaalf keer zo klein als op jaarbasis. Maar elke volgende maand krijg je rente op het nieuwe steeds grotere bedrag. Dat kan na twaalf

[3] Joost draagt dit hoofdstuk op aan Henk Pfaltzgraff.

[4] Dat de benaming groeifactor ons maar vooral ook e zelve verder bespaard blijve.

[5] Ongeveer de rentefractie die IceSave beloofde.

[6] Dat moet je bij banken altijd maar weer afwachten.

maanden alleen maar gunstiger uitpakken, want als de elke maand bijgeschreven rente niet meegeteld zou worden bij elke volgende renteberekening, dan kom je aan het eind van het jaar op precies hetzelfde uit als wanneer je de berekening in één keer zou doen aan het einde van het jaar. Je krijgt dus in twaalf maanden in ieder geval meer dan twaalf keer de rentefractie $\frac{a}{12}$ op het startbedrag[7].

Bij rente op maandbasis is na een jaar je tegoed twaalf keer met $1 + \frac{a}{12}$ vermenigvuldigd. De eenvoudige observatie is nu dus dat voor $a > 0$,

$$\left(1 + \frac{a}{12}\right)^{12} \text{ groter moet zijn dan } 1 + a. \tag{4.1}$$

Om dit in te zien hebben we niet hoeven rekenen, en verder uitschrijven kan het inzicht voorlopig alleen maar vertroebelen, zeker met een expliciete getalkeuze voor a. Een vraag die meteen opkomt is of het eindbedrag na een jaar nog (veel) meer kan worden als we de rente vaker per jaar bijgeschreven krijgen, 52 keer, 365 keer, enzovoorts.

Ook voor de berekeningen die nu volgen is het essentieel dat we met een abstracte rentefactor a werken en de keuze van a niet meteen specificeren. Zonder die abstractie verdwijnt alle structuur uit de formules waarin we willen zien wat er gebeurt als we het aantal rentebijschrijvingen per jaar veranderen van maandelijks naar wekelijks naar dagelijks et cetera. Juist met de abstracte a die we niet hoeven te weten, houden de formules een natuurlijke ordening die de transparantie ten goede komt.

Rente op rente op maandbasis en niet op jaarbasis betekent dat aan het eind van het jaar je tegoed een factor

$$\left(1 + \frac{a}{12}\right)^{12}$$

groter is, in plaats van de factor $1 + a$ die bij rente op jaarbasis hoort. Zoals we net al opmerkten is het groter zijn van deze groeifactor zo evident dat we dat niet meer hoeven uit te rekenen, maar we doen het zo meteen toch. Wie weet is er nog meer inzicht dat we daarmee kunnen krijgen.

Dat de 12 in de formule het aantal bijschrijvingen per jaar is snappen we inmiddels. Met twee bijschrijvingen per maand krijg je zo op jaarbasis een factor

$$\left(1 + \frac{a}{24}\right)^{24},$$

dezelfde formule, maar met 24 in plaats van 12, hetgeen goed zichtbaar blijft door het gebruik van de abstracte notatie met a. Dat de factor met 24 (nog)

[7] Kijk uit, leningen met een maandelijks vastgelegd percentage komen ook voor...

meer is dan met 12 is duidelijk, want wat op maandbasis al beter is (twee keer bijschrijven per maand in plaats van één keer) is zeker ook op jaarbasis beter. Het ophakken van de renteperiode in kleinere periodes maakt je dus rijker.

In de formules zien we inmiddels waar wat staat. De oorspronkelijke rentefractie op jaarbasis is a, en het aantal renteperiodes per jaar is de in de laatste formule twee keer voorkomende 24. Rentetermijnen verder ophakken in kortere periodes is alleen maar gunstiger. Maar is het nu ook duidelijk dat rente op weekbasis, met 52 bijschrijvingen per jaar en een factor

$$\left(1+\frac{a}{52}\right)^{52}$$

op jaarbasis, ook gunstiger is?

Intuïtief wel, maar rekenen via maanden gaat niet zomaar, want 52 is niet deelbaar door 12. Wel door 13 trouwens, dus de vraag zou eerst kunnen zijn of de factor

$$\left(1+\frac{a}{13}\right)^{13} \text{ groter is dan } \left(1+\frac{a}{12}\right)^{12},$$

en de stap van 12 naar 13 schreeuwt nu als het ware om de volgende abstractie: zet een n in plaats van 12 en $n+1$ in plaats van 13.

Wat we dan willen nagaan is of de bijschrijffactor

$$e_n(a) = \left(1+\frac{a}{n}\right)^n \tag{4.2}$$

groter wordt als n groter wordt gekozen. Dat kunnen we stap voor stap doen, bijvoorbeeld van $n = 12$ naar $n = 13$, of van $n = 123456789$ naar $n = 123456790$. Met de introductie van n lukt het hopelijk om in één keer al deze stappen tegelijk[8] te zetten, door te bewijzen dat

$$e_{n+1}(a) = \left(1+\frac{a}{n+1}\right)^{n+1} > \left(1+\frac{a}{n}\right)^n = e_n(a), \tag{4.3}$$

niet alleen voor $n = 12$, maar voor elke n vanaf $n = 1$.

De a is voorlopig positief, want we praten over rentebijschrijvingen. We verwachten dan intuïtief dat (4.3) waar is: $e_n(a)$ zal groter zijn naarmate n groter is. Natuurlijk is de hamvraag of $e_n(a)$ willekeurig groot wordt als we steeds vaker rente bijgeschreven krijgen, maar dat klinkt te mooi om waar te zijn. Er zal dus wel een bovengrens zijn voor $e_n(a)$, een bovengrens die afhangt van a, maar niet van n. Dat zou betekenen dat er een grens is aan het voordeel dat we kunnen behalen door steeds frequenter rente te

[8] Bijna net als het in één keer omgooien van alle steentjes in een dominospel.

laten bijschrijven. We zullen (4.3) nog echt bewijzen, maar voor het verdere verloop van het verhaal is dat bewijs nu niet nodig, daar is (4.3) te evident voor.

4.2 exp en e

Vooruitlopend op het inderdaad stijgend maar wel begrensd zijn van de rij getallen $e_n(a)$ die je krijgt als je voor een gegeven $a > 0$ in (4.2) voor n achtereenvolgens $1, 2, 3, \ldots$ invult, kunnen we een limietwaarde introduceren voor $e_n(a)$ als n naar oneindig gestuurd wordt[9]. Voor vaste $a \geq 0$ is deze limietwaarde een positief getal dat we $\exp(a)$ noemen. En voor $a > 0$ is dat getal groter dan de 1 die we krijgen met $a = 0$.

Voorlopig is exp alleen maar een naam en (nog) niet de knop op de rekenmachine. We zeggen dat

$$e_n(a) = \left(1 + \frac{a}{n}\right)^n \text{ stijgt naar } \exp(a) \text{ als } n \to \infty, \text{ voor elke } a > 0. \tag{4.4}$$

Deze $\exp(a)$ kunnen we nu interpreteren als de maximale bijschrijffactor op jaarbasis bij rentefractie a, door de bijschrijfperiodes steeds maar korter te maken. In de banale bankpraktijk is dat misschien niet realiseerbaar, maar die praktijken laten we graag achter ons.

Voordat we in meer details duiken[10] is het goed om stil te staan bij wat we hier op de rails gezet hebben. Uitgaande van de uiteindelijk toch verheffende rente-op-rentecontext zijn we gekomen tot het idee dat dit de definitie is waar we echt niet omheen kunnen:

Definitie 4.1. *Voor $a \geq 0$ definiëren we*

$$\exp(a) := \lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n.$$

Het symbool ':=' moet je lezen als 'is per definitie gelijk aan', of iets dergelijks. Geen wortels, geen logaritmen, alleen $+, \cdot, /$ en een limiet. Strikt genomen is Definitie 4.1 nog geen definitie omdat we nog niet zeker weten of de limiet ook echt bestaat. Maar we hebben wel al een onweerlegbare intuïtie opgebouwd dat, tenminste voor $a \geq 0$, dit inderdaad een goede definitie moet zijn[11].

Wat zijn nu de eigenschappen van $\exp(a)$ als we a als variabele zien? Een a die ook t of x mag heten trouwens. Wat voor type functie

$$x \to \exp(x)$$

[9] Notatie: $n \to \infty$, het pijltje niet verwarren met het functiepijltje.
[10] Of die overslaan.
[11] Een goede definitie is per definitie een definitie.

hebben we hier eigenlijk gemaakt?

Merk op dat voor elke positieve a en b,

$$(1+\frac{a}{n})(1+\frac{b}{n}) = 1 + \frac{a+b}{n} + \frac{ab}{n^2}.$$

We zien in dit product zowel de som als het product van a en b. Als we links en rechts de n-de macht nemen krijgen we

$$\left(1+\frac{a}{n}\right)^n \left(1+\frac{b}{n}\right)^n = \left(1+\frac{a+b}{n}+\frac{ab}{n^2}\right)^n,$$

en dan zien we in de limiet aan de linkerkant $\exp(a)$ keer $\exp(b)$ verschijnen.

Rechts heeft de term met ab in de teller en n^2 in de noemer voor $n \to \infty$ steeds minder invloed op het gedrag van de n-de macht vergeleken met de term met $a + b$ in de teller en n in de noemer, simpelweg omdat $\frac{1}{n^2}$ veel kleiner is dan $\frac{1}{n}$. Als die invloed helemaal verdwijnt, dan krijgen we rechts ook een exp, en wel $\exp(a + b)$. Er moet dus wel volgen dat

$$\exp(a)\exp(b) = \exp(a+b), \tag{4.5}$$

voor alle positieve a en b.

Opgave 4.2. Wat is $\exp(0)$?

Misschien wel het belangrijkste getal in de hele wetenschap[12] wordt nu gedefinieerd door een definitie die je wellicht eerder als uitkomst van een som bent tegengekomen:

Definitie 4.3. *Het getal e is per definitie*

$$e := \exp(1) = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

Opgave 4.4. Laat met (4.5) zien dat

$$\exp(2) = e^2, \quad \exp(3) = e^3, \quad \exp(4) = e^4,$$

enzovoorts.

[12] A sweeping statement.

Opgave 4.5. Wat is $\exp(\frac{1}{2})$? En $\exp(\frac{2}{3})$?

Wat we in de opgaven hierboven hebben gezien mag wel even expliciet gemaakt worden: voor alle $p \geq 1$ geheel en $q \geq 1$ geheel geldt dat

$$\exp(\frac{p}{q}) = \sqrt[q]{e^p} = e^{\frac{p}{q}},$$

waarbij het tweede gelijkteken slechts de gebruikelijke notatie voor gebroken machten van positieve getallen betreft. Kortom, voor rationale getallen x geldt dat $\exp(x) = e^x$, spectaculair toch? En hoe zit het dan met niet-rationale x? Tja. Dat is ons inziens (nog) geen goed geformuleerde vraag, aangezien we eigenlijk nog geen definitie van e^x hebben voor alle x[13]. Maar daar is wel wat aan te doen natuurlijk. Als we e^x ook voor niet-rationale x willen invoeren, dan is er maar één zinnige keuze die het perspectief van onze vingers overstijgt[14]:

Definitie 4.6. *De enige echte one and only e-macht wordt gedefinieerd door*

$$e^x := \exp(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n,$$

in eerste instantie alleen nog voor alle $x \geq 0$.

Opgave 4.7. Wat kun je zeggen over $\exp(x)$ als $x \to +\infty$?

Nu we de e-macht voor het eerst in zijn universele gedaante zien, willen we zeker weten dat Definitie 4.1 inderdaad een goede definitie is. Daartoe moeten we het voorafgaande helaas weer even uit ons hoofd zetten. We weten weer niets, en we kijken naar de bijschrijffactor in (4.2),

$$e_n(a) = \left(1 + \frac{a}{n}\right)^n.$$

Dit is de factor op jaarbasis bij een rentefractie a die, in plaats van één keer per jaar, n keer per jaar bijgeschreven wordt (wel gedeeld door n natuurlijk). Heeft deze $e_n(a)$ inderdaad een limiet als $n \to \infty$ waarmee we $\exp(a)$ kunnen definiëren?

Wat we hiervoor nodig hebben is een combinatie van algebra en analyse. Analyse is een soort schattend rekenen met letters en dat is moeilijker dan gewone letteralgebra, net zoals goed schattend rekenen ook moeilijker is dan gewoon rekenen. Maar eerst nog wat algebra waarbij geteld moet worden, ook best lastig trouwens.

[13] Nog afgezien van de mystiek (tot nu toe) rond e zelf.
[14] Trompetgeschal.

4.3 exp(a) uitgerekend!

Die bijschrijffactor $e_n(a)$, hoe zit het daarmee voor grote waarden van n? Laten we $n = 14$ nemen ter illustratie. Kun je jezelf ervan overtuigen dat

$$(1+a)^{14} = \underbrace{(1+a)(1+a)(1+a)(1+a)(1+a)\cdots(1+a)(1+a)}_{\text{14 factoren}} =$$

$$1 + 14a + \frac{14\cdot 13}{2}a^2 + \cdots + \frac{14\cdot 13\cdot 12\cdot 11\cdot 10}{5\cdot 4\cdot 3\cdot 2}a^5 + \cdots + a^{14}?$$

Zo ja, dan kun je $\frac{a}{14}$ invullen voor a en dan zie je dat

$$(1+\frac{a}{14})^{14} = 1 + a + \frac{13}{14}\frac{a^2}{2!} + \cdots + \frac{13\cdot 12\cdot 11\cdot 10}{14\cdot 14\cdot 14\cdot 14}\frac{a^5}{5!} + \cdots\,,$$

met nog steeds slechts vijftien termen, de laatste qua typesetting te groot om op te schrijven. Zie je de regelmaat?

Opgave 4.8. Wat is de coëfficiënt van a^3? En van a^4? Doe dan ook die van a^{14} maar.

Bij $n = 99$ (om maar iets te noemen) ziet het er als volgt uit:

$$(1+\frac{a}{99})^{99} = 1 + a + \frac{98}{99}\frac{a^2}{2!} + \cdots + \frac{98\cdot 97\cdot 96\cdot 95}{99\cdot 99\cdot 99\cdot 99}\frac{a^5}{5!} + \cdots\,,$$

een som met honderd termen. Niets weerhoudt ons er van om dit nu voor algemene n op te schrijven en te concluderen dat

$$e_n(a) = 1 + a + \cdots + (1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})(1-\frac{4}{n})\frac{a^5}{5!} + \cdots\,, \tag{4.6}$$

een formule met $n+1$ termen waarvan we er eentje hebben uitgeschreven om de algemene structuur te laten zien.

Opgave 4.9. Wat is de laatste term?

Omdat

$$\frac{1}{n} \text{ daalt naar } 0$$

is het zo dat elke van de vier positieve voorfactoren in de term met a^5 in (4.6) naar 1 stijgt, en dat de term met a^5 dus stijgt naar $\frac{a^5}{5!}$. Iets dergelijks gebeurt ook bij alle andere machten van a. Dit betekent dat voor elke $a \geq 0$,

$$e_n(a) \leq 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!} + \frac{a^5}{5!} + \cdots + \frac{a^n}{n!}, \tag{4.7}$$

en dat elke term in het rechterlid van (4.6) stijgt naar de overeenkomstige term in het rechterlid van (4.7) als $n \to \infty$. Maar pas op! Niet alleen stijgen de termen, het aantal termen neemt ook toe als n groter wordt. Gelukkig zijn al die nieuwe termen steeds positief, dus dit maakt het alleen maar makkelijker voor $e_n(a)$ om stijgend in n te zijn. Dit alles bewijst dat (nog steeds voor $a > 0$) onze rente op rente uitdrukking $e_n(a)$ inderdaad stijgt als n groter wordt.

De enige vraag die nu nog overblijft, is waar het naartoe gaat met $e_n(a)$. Vooralsnog zou het ook kunnen gebeuren dat $e_n(a)$ groter wordt dan elk eindig getal. Het feit dat de term met (bijvoorbeeld) a^5 stijgt naar $\frac{a^5}{5!}$ en dat het aantal termen toeneemt, suggereert dat $e_n(a)$ misschien wel stijgt naar

$$\mathrm{EXP}(a) := 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!} + \frac{a^5}{5!} + \cdots, \tag{4.8}$$

een som met oneindig veel termen![15] Dit leidt meteen tot twee vragen: bestaat deze oneindige som eigenlijk wel, en zo ja, stijgt $e_n(a)$ daar inderdaad naartoe?

In (1.3) hebben we gezien dat er niets engs is aan sommen met oneindig veel termen. Zo hebben we gezien dat

$$\frac{1}{m} + \frac{1}{m^2} + \frac{1}{m^3} + \frac{1}{m^4} + \cdots = \frac{1}{m-1} \tag{4.9}$$

als m een geheel[16] getal is groter dan 1. Je telt wel oneindig veel termen bij elkaar op, maar die worden zo snel kleiner dat alles bij elkaar op te tellen is zonder dat de som willekeurig groot wordt[17].

In de voorbeeldsom (4.9) zien we dat elke term precies een factor m kleiner is dan de vorige. Dat helpt natuurlijk om ervoor te zorgen dat er iets uit de som komt. Hoe sneller de termen klein worden hoe beter. Hoe zit het nu met de verhouding tussen opeenvolgende termen in het rechterlid van (4.8) waarmee we $\mathrm{EXP}(a)$ proberen te maken?

De factoren die er steeds bijkomen in elke volgende term van (4.8) zijn achtereenvolgens, beginnend bij de stap van 1 naar a,

$$a, \frac{a}{2}, \frac{a}{3}, \frac{a}{4}, \frac{a}{5}, \frac{a}{6}, \cdots$$

[15] Let op het verschil tussen $\exp(a)$ en $\mathrm{EXP}(a)$, limiet respectievelijk som.
[16] Is dat eigenlijk echt nodig?
[17] Zoals Zeno's schildpad ook niet oneindig lang op Achilles voorblijft.

Hoe groot de positieve a in de tellers ook is, de rij noemers $1, 2, 3, 4, 5, 6, \ldots$ haalt a altijd in. Wil je in (4.8) dat in elke stap de volgende term minstens een factor twee kleiner is dan de vorige dan moet je de index in de som groter kiezen dan $2a$. Is dat eenmaal gelukt dan wordt elke nieuwe factor daarna alleen maar beter. Voor elke vaste a is oneindig voor het rechterlid van (4.8) buiten bereik[18] via een bovengrens in de vorm van een eindige som

$$1 + a + \frac{a^2}{2!} + \cdots + \frac{a^{n-1}}{n-1!} + \frac{a^n}{n!}\underbrace{(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots)}_{=2}.$$

Als je n zo kiest dat $n > 2a$ dan is dit een bovengrens en gegeven $a > 0$ kan dat, klaar. Dus $\mathrm{EXP}(a)$ bestaat echt, en we concluderen dat voor $a \geq 0$ de definitie van $\mathrm{EXP}(a)$ in (4.8) inderdaad betekenis heeft.

Wat ons nu nog rest is te laten zien dat $e_n(a)$ inderdaad naar $\mathrm{EXP}(a)$ stijgt. De afschatting in (4.7) vertelt ons in ieder geval dat

$$e_1(a) < e_2(a) < e_3(a) < \cdots < \mathrm{EXP}(a)$$

als $a > 0$. De limietwaarde $\exp(a)$ van $e_n(a)$ voor $n \to \infty$ kan dus alleen maar kleiner zijn dan of gelijk aan $\mathrm{EXP}(a)$, dus we krijgen

$$\exp(a) \leq \mathrm{EXP}(a) \tag{4.10}$$

gratis en voor niets.

De omgekeerde ongelijkheid is iets lastiger. We kunnen $e_n(a)$ afkappen bij zijn kwadratisch term, dan krijgen we

$$1 + a + (1 - \frac{1}{n})\frac{a^2}{2!} < e_n(a).$$

Als $n \to \infty$, dan gaat het linkerlid naar $1 + a + \frac{a^2}{2!}$ en het rechterlid naar $\exp(a)$. Er volgt dus dat

$$1 + a + \frac{a^2}{2!} \leq \exp(a).$$

Maar zo[19] volgt ook dat

$$1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} \leq \exp(a),$$

[18] The sky is not the limit!

[19] Nu ook de term met a^3 in (4.6) opschrijven.

en voor elke eindige som gaat het precies hetzelfde. Als we steeds meer termen meenemen stijgt het linkerlid zo verkregen in de rij van ongelijkheden naar $\mathrm{EXP}(a)$, terwijl het rechterlid gewoon gelijk blijft aan $\exp(a)$. We concluderen dat $\mathrm{EXP}(a)$ voldoet aan

$$\mathrm{EXP}(a) \leq \exp(a). \tag{4.11}$$

Samenvattend, (4.10) en (4.11) vertellen ons dat als $n \to \infty$, rente op rente bij rentefractie $a > 0$ een rentefactor geeft van

$$\left(1 + \frac{a}{n}\right)^n \to \exp(a) = \mathrm{EXP}(a) = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!} + \frac{a^5}{5!} + \cdots \tag{4.12}$$

in plaats van de factor $1 + a$.

We zijn nu dus zover dat $\exp(a)$ voor alle $a \geq 0$ gedefinieerd is via de stijgende rij $e_n(a)$ en gelijk is aan $\mathrm{EXP}(a)$[20]. In sommige tekstboeken wordt (4.8) gebruikt als definitie van e, met $a = 1$. Dat kan ook, maar de rente-op-rentecontext laat wellicht directer zien waarom e zo'n natuurlijk[21] getal is.

4.4 Negatieve rente

Tot nu toe hebben we aangenomen dat $a \geq 0$. Gezien het feit dat de bank werkt met rente op rente lag dat voor de hand. Maar de vraag hoe het zit met het gedrag van de factor $e_n(a)$ voor $a < 0$ is interessant, zelfs zonder economische context, ook omdat het geval $a < 0$ in sinterklaastijd aardige toepassingen heeft[22].

Opgave 4.10. Beantwoord op intuïtie: voor een negatieve a is de rij $e_n(a)$ in (4.2) monotoon[23]. Waar of niet waar? Indien waar, is de rij dan dalend of stijgend?

Opgave 4.11. Even kijken wat er gebeurt als we de grafieken van de functies $a \to e_n(a)$ plotten (ook voor $a < 0$), in één diagram, zodat we kunnen zien hoe het zit met de ordening. Voor kleine n kan dat prima met de hand.

[20] Poeh, een heel gedoe voor iets wat je eigenlijk zo ziet.
[21] Dé natuurlijke getallen zijn $1, 2, 3, 4, 5, 6, \ldots$, maar dat terzijde.
[22] Daar komen we in het volgende hoofdstuk met de Pieten op terug.
[23] Zodra n groter is dan $-a$.

Vaker rente (af)schrijven is zo te zien[24] ook bij negatieve rente gunstiger. Wie had dat meteen gedacht? Voor $a < 0$ is de vraag of er een bovengrens bestaat nu triviaal: alle $e_n(a)$ zijn kleiner dan 1, we kunnen dus sowieso nooit boven de 1 uitkomen. En door ons te beperken tot $n > -a$ is $e_n(a)$ ook positief. Als we naar de grafieken kijken ligt het misschien voor de hand om de definitie van $e_n(a)$ voor $a + n < 0$ te veranderen in $e_n(a) = 0$. Het negatieve stuk doet immers uiteindelijk toch niet mee als we voor vaste $a < 0$ de limiet voor $n \to \infty$ bekijken.

Kunnen we nu concluderen dat (4.12) ook voor $a < 0$ geldt? Voordat je verder leest wil je misschien even nadenken over wat daar nog voor nodig is. Onze intuïtie, en ook onze bewijzen tot nu toe, werken niet meer zomaar als a negatief is. Maar met wat wiskundige denkactiviteit komen we daar wel uit. Daartoe is het handig om ook het rechterlid van (4.7) een naam te geven:

$$E_n(a) := 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!} + \frac{a^5}{5!} + \cdots + \frac{a^n}{n!}. \tag{4.13}$$

Wat is nu het verschil tussen $e_n(a)$ in (4.6) en $E_n(a)$ in (4.13)? We vergelijken daartoe de eindige som met $n + 1$ termen[25]

$$e_n(a) = 1+a+(1-\frac{1}{n})\frac{a^2}{2!}+(1-\frac{1}{n})(1-\frac{2}{n})\frac{a^3}{3!}+(1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})\frac{a^4}{4!}+\cdots$$

met de nieuwe som

$$E_n(a) = 1 + a + \frac{a^2}{2!} + \cdots + \frac{a^n}{n!}.$$

De eerste twee termen in de uitdrukkingen voor $E_n(a)$ in (4.13) en $e_n(a)$ daar direct onder zijn hetzelfde en we zien dat

$$\frac{a^2}{2n} \tag{4.14}$$

als het verschil van de derde termen verschijnt. De n in de noemer van (4.14) helpt dus in ieder geval al mee om het totale verschil naar nul te trekken.

Precies (4.14) als foutafschatting[26] voor

$$F_n(a) = E_n(a) - e_n(a)$$

kunnen we misschien niet krijgen, maar misschien wel ongeveer. Daartoe hieronder nu eerst een stukje voorbereidende algebra, waarin we de termen

[24] What you see is what you get?
[25] Waarin we de laatste term voor het gemak weer niet opschrijven.
[26] Berekeningen moeten precies, foutafschattingen niet, maar wel correct.

zo ver uitschrijven als nodig is om de structuur in de formules te herkennen om ze vervolgens schattend te behandelen, een kunst apart waar in dit geval ook nog iets fraais uit komt[27]. De twee uitdrukkingen $E_n(a)$ en $e_n(a)$ van elkaar aftrekkend vinden we

$$F_n(a) = \underbrace{\left(1-(1-\frac{1}{n})\right)}_{B_2}\frac{a^2}{2!} + \underbrace{\left(1-(1-\frac{1}{n})(1-\frac{2}{n})\right)}_{B_3}\frac{a^3}{3!} + \cdots$$

$$+\underbrace{\left(1-(1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})\right)}_{B_4}\frac{a^4}{4!} + \cdots,$$

nog steeds een eindige som, nu met $n-1$ termen, want de eerste twee vallen weg. De coëfficiënten hebben een index, maar hangen ook van n af[28]. We nummeren de termen van 2 tot en met n. Goed kijkend zien we dat

$$B_4 = 1 + (1-\frac{3}{n})(B_3 - 1) = (1-\frac{3}{n})B_3 + \frac{3}{n}$$

de regelmaat al duidelijk maakt. Beginnend met

$$B_2 = \frac{1}{n},$$

$$0 < B_3 = (1-\frac{2}{n})B_2 + \frac{2}{n} < \frac{1}{n} + \frac{2}{n} = \frac{1+2}{n},$$

$$\begin{aligned} 0 < B_4 &= 1 + (1-\frac{3}{n})(B_3-1) = (1-\frac{3}{n})B_3 + \frac{3}{n} \\ &< \frac{1+2}{n} + \frac{3}{n} = \frac{1+2+3}{n}, \end{aligned}$$

zien we hoe groot de coëfficiënten kunnen zijn. Tot ons vermaak komt in de teller van de afschattingen de strafwerkwerksom die Gauss[29] op school kreeg hier ook nog langs:

Opgave 4.12. Wat is de uitkomst van $1 + 2 + \cdots + 1000$? Wat is, voor k positief en geheel, de som van de eerste k positieve gehele getallen? Laat zien hoe het antwoord voor $k+1$ volgt uit het antwoord voor k, welke $k \geq 1$ je ook neemt in deze inductiestap[30].

---

[27] Niet stiekem overslaan, maar kijk alvast naar (4.15).

[28] Eigenlijk moet je $B_{2,n}, B_{3,n}, \ldots$ lezen, twee indices dus.

[29] Wie is dat?

[30] Dominoprincipe: de '$k = 1$'-uitspraak impliceert zo de uitspraak voor elke $k > 1$.

Opgave 4.13. Laat nu voor $k = 2, \dots, n$ zien dat

$$0 < B_k < \frac{k(k-1)}{2n} \text{ ofwel } 0 < \frac{B_k}{k!} < \frac{1}{2n(k-2)!}$$

Het verschil tussen $e_n(a)$ en $E_n(a)$ is zoals we hierboven gezien hebben

$$F_n(a) = B_2 \frac{a^2}{2!} + \cdots + B_n \frac{a^n}{n!},$$

en dus volgt nu dat

$$|F_n(a)| \leq |B_2 \frac{a^2}{2!}| + \cdots + |B_n \frac{a^n}{n!}|.$$

Hiermee zijn we van negatieve termen verlost en zien we hoe groot het verschil maximaal kan zijn. Met behulp van de laatste opgave en de definitie van $E_n(a)$ volgt dat

$$|F_n(a)| \leq \frac{1}{2n} \left(|a|^2 + \cdots + \frac{|a|^n}{(n-2)!} \right) = \frac{a^2}{2n} E_{n-2}(|a|) \leq \frac{a^2}{2n} \mathrm{EXP}(|a|). \tag{4.15}$$

Dankzij de n in de noemer[31] gaat het verschil $F_n(a)$ dus naar nul als $n \to \infty$, en het maakt hierbij niet uit of a positief of negatief is.

De crux hier is dat de ongelijkheden in (4.15) ook voor $a < 0$ gelden en dat het verschil tussen

$$e_n(a) = \left(1 + \frac{a}{n}\right)^n \quad \text{en} \quad E_n(a) = 1 + a + \frac{a^2}{2!} + \cdots + \frac{a^n}{n!}$$

naar nul gaat, al weten we voor $a < 0$ misschien vanwege het alterneren van de termen nog niet goed hoe we bij de tweede uitdrukking met de overgang $n \to \infty$ moeten omgaan.

Een trucje om weg te komen met de van teken wisselende termen is om steeds twee termen bij elkaar te nemen, bijvoorbeeld

$$\frac{a^{20}}{20!} + \frac{a^{21}}{21!} = \frac{a^{20}}{20!} \left(1 + \frac{a}{21}\right) > 0 \quad \text{als } a > -21.$$

De overzichtelijkheid van alleen positieve termen, afgezien van de eerste paar termen, keert zo snel terug. Voor elke a is vanaf zekere index de som van twee opeenvolgende termen met even en oneven index positief. Trucjes die vaker werken zijn eigenlijk methodes en het trucje dat we hier gebruikt hebben is een mooi voorbeeld.

[31] Het rechterlid in (4.15) is (4.14) met de extra factor $\mathrm{EXP}(|a|)$.

Opgave 4.14. Wat komt er, als $m > 1$, uit

$$\frac{1}{m} - \frac{1}{m^2} + \frac{1}{m^3} - \frac{1}{m^4} + \cdots ?$$

(Hint: factoriseer steeds twee opeenvolgende termen.)

Opgave 4.15. Overtuig jezelf van het feit dat de som in (4.8) voor $a < 0$ ook een echte uitkomst heeft.

Dus $\mathrm{EXP}(a)$ bestaat ook als $a < 0$ en $e_n(a)$ gaat daar naartoe als $n \to \infty$. Zo kunnen we nu $\exp(a)$ definiëren voor alle a, op dezelfde manier als eerst voor $a \geq 0$. We concluderen dat (4.12) voor alle a geldt en zijn klaar voor de slotronde. De functie $\exp = \mathrm{EXP}$ die we nu gemaakt hebben heeft een eigenschap die we al hebben geraden:

$$\exp(a + b) = \exp(a)\exp(b).$$

We werken deze slotconclusie nog wat uit. Eerst maar voor $a + b = 0$, want die vraag brandde na de vorige opgave misschien al op de lippen. Nu we weten hoe het zit met $e_n(a)$ en $e_n(-a)$ kunnen we kijken naar hun product

$$\epsilon_n(a) = e_n(a)e_n(-a) = \left(1 - \frac{a^2}{n^2}\right)^n, \tag{4.16}$$

dat stijgt[32] met n als[33] $a \neq 0$.

Opgave 4.16. Laat zien dat $\epsilon_n(a)^n$ dan stijgt naar $\exp(-a^2)$, dat groter is dan 0 en op zijn hoogst 1, en combineer dit met het stijgen van $\epsilon_n(a)$ om de conclusie te trekken dat $\epsilon_n(a)$ naar 1 stijgt, gebruikmakend van het feit dat voor elke $0 < y < 1$, de n-de machtswortel $\sqrt[n]{y}$ naar 1 stijgt[34] als $n \to \infty$. Overtuig jezelf ervan dat dus

$$1 = \exp(0) = \exp(-a)\exp(a).$$

[32] Waarom?

[33] Tijd om James Gerritsen te noemen hier, met veel dank voor alle correcties!

[34] Evident? Zo nee, bestudeer de grafiek $y = x^n$ met $0 < x < 1$ en n steeds groter.

Opgave 4.17. Met de definitie $e = \exp(1)$ volgt nu dat $\frac{1}{e} = \exp(-1)$ en dat

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots\,; \quad \frac{1}{e} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots\,.$$

Opgave 4.18. We zien in Opgave 4.16 een voorbeeld van de uitspraak dat

$$\exp(a+b) = \exp(a)\exp(b), \tag{4.17}$$

namelijk het speciale geval $a + b = 0$. Voor algemene a en b geldt dat

$$(1 + \frac{a}{n})(1 + \frac{b}{n}) = (1 + \frac{a+b}{n} + \frac{ab}{n^2})$$

We zien in het product zowel de som als het product van a en b. De vorige opgaven gaven inzicht in de rol van de kwadratische term in het grondtal als we alles tot de n-de macht verheffen. Overtuig jezelf er nu van dat (4.17) waar is voor alle a en b, en dat nu ook voor elke gehele n geldt dat

$$\exp(n) = e^n.$$

Opgave 4.19. Reflecteer op de geldigheid van de formule $\exp(x) = e^x$ voor alle reële getallen x.

4.5 Vaker is altijd meer[35]

Hoewel het strikt gesproken niet meer nodig is, zijn we toch wel nieuwsgierig naar hoe het nu echt zit met de monotonie in n van $e_n(a)$ voor $a < 0$[36]. We hadden al gezien dat voor $a \geq 0$, $e_n(a)$ stijgt in n. Is dit ook waar voor $a < 0$?

Deze vraag is lastiger dan misschien op voorhand gedacht, want als we het verschil

$$\Delta_n(a) := e_{n+1}(a) - e_n(a) = \left(1 + \frac{a}{n+1}\right)^{n+1} - \left(1 + \frac{a}{n}\right)^n$$

[35] Voor de liefhebber.

[36] Als je $\ln(1 + \frac{a}{n})^n$ naar n differentieert speel je een beetje vals.

bekijken en ons afvragen of $\Delta_n(a)$ positief is, zien we dat de n op twee plaatsen in elke term voorkomt en dat maakt vergelijken lastig. De standaardtechniek om hier mee om te gaan is het invoegen van een missing link, en wel twee keer, met een plus en met een min, zodat er netto niets verandert[37]. Zo kunnen we het verschil splitsen in twee verschillen die elk op zich makkelijker zijn, omdat ze elk op maar één plek verschillen. Beide stukken leiden vervolgens hun eigen leven tot dat het moment daar is om de boel weer bij elkaar te harken. Klinkt simpel, maar is best lastig, dus goed opletten hieronder, want je moet onderweg wel iets herkennen waarmee je uit de voeten kunt, en je weet nog niet wat[38].

Met de missing links erbij gaat het om de vraag of

$$\Delta_n(a) = \underbrace{\left(1+\frac{a}{n+1}\right)^{n+1}}_{e_{n+1}(a)} \underbrace{-\left(1+\frac{a}{n}\right)^{n+1} + \left(1+\frac{a}{n}\right)^{n+1}}_{\text{missing links}} - \underbrace{\left(1+\frac{a}{n}\right)^{n}}_{e_n(a)}$$

positief is. Splitsen met in elke term een missing link geeft twee termen. In elk van de twee zien we iets:

$$\Delta_n(a) = \underbrace{\left(1+\frac{a}{n+1}\right)^{n+1} - \left(1+\frac{a}{n}\right)^{n+1}}_{\text{van de vorm (3.3) met } 5 = n+1} + \underbrace{\left(1+\frac{a}{n}\right)^{n+1} - \left(1+\frac{a}{n}\right)^{n}}_{\text{dit factoriseert}}.$$

We zien dus dat $\Delta_n(a)$ gelijk is aan

$$\underbrace{\left(\frac{a}{n+1}-\frac{a}{n}\right)\left(\left(1+\frac{a}{n+1}\right)^{n} + \cdots + \left(1+\frac{a}{n}\right)^{n}\right)}_{\text{hier is (3.3) gebruikt met } 5 = n+1} + \underbrace{\left(1+\frac{a}{n}\right)^{n}\left(1+\frac{a}{n}-1\right)}_{\text{gefactoriseerd}}.$$

De mengtermen op de puntjes bevatten elk n factoren, en als je goed kijkt naar deze nu niet opgeschreven termen, of naar (3.3), dan zie je dat elk van deze termen tussen de eerste en de laatste term in ligt. Dat zullen we straks gebruiken maar eerst brengen we in de linkerterm de voorfactor onder één noemer[39], en omdat in de tweede factor in de rechterterm de enen fijn wegvallen zien we dat

$$\Delta_n(a) = \underbrace{\frac{-a}{n(n+1)}}_{\text{was } \frac{a}{n+1}-\frac{a}{n}} \underbrace{\left(\left(1+\frac{a}{n+1}\right)^{n} + \cdots + \left(1+\frac{a}{n}\right)^{n}\right)}_{\text{ongewijzigd}} + \underbrace{\frac{a}{n}\left(1+\frac{a}{n}\right)^{n}}_{\text{versimpeld}}.$$

[37] Een welhaast universele truc.
[38] Ook wij niet toen we begonnen.
[39] Breukrekenen blijft belangrijk.

Nu kunnen de a in de tellers en de n in de noemers naar voren, en we halen ook[40] de $n+1$ in de eerste term naar buiten. De versimpelde uitdrukking rechts schrijven we nu binnen de haken als eerste op, met een factor $n+1$, want we hebben voor de haken een $n+1$ in de noemer gezet. Dit geeft

$$\Delta_n(a) = \frac{a}{n(n+1)} \left(\underbrace{(n+1)\left(1+\frac{a}{n}\right)^n}_{n+1 \text{ termen}} \underbrace{-\left(1+\frac{a}{n+1}\right)^n - \cdots - \left(1+\frac{a}{n}\right)^n}_{n+1 \text{ negatieve termen}} \right),$$

waarin de eerste uitdrukking tussen de haken geschreven kan worden[41] als som van $n+1$ identieke positieve termen. Het moment[42] is daar waarop alles weer samengenomen kan worden. Misschien zie je al hoe het verder gaat en wil je het nu zelf afmaken voor je verder leest.

We kijken alleen naar $a < 0$, want voor $a \geq 0$ weten we alles al. Met $a < 0$ is de voorfactor $\frac{a}{n(n+1)}$ negatief. Zodra $n > -a > 0$ zijn in de tweede factor alle termen met een min ook echt negatief, en de laatste is de kleinste en die is gelijk aan een positieve term die $n+1$ keer voorkomt. De negatieve termen winnen het dus van de positieve termen[43]. Wat er tussen de haken staat is in totaal negatief, dus dat betekent dat $\Delta_n(a)$ het product is van twee negatieve termen, en zelf dus positief is[44]. Klaar[45]. We hebben laten zien dat $e_n(a)$ monotoon is in n, zodra $n > -a$. En we zien dat ook voor $a > 0$ dit een bewijs geweest zou zijn, maar dan met $\Delta_n(a)$ als product van twee positieve factoren.

[40] De wedstrijd is nu echt begonnen.
[41] Combinaties op het middenveld.
[42] Aanval.
[43] Voorzet.
[44] Min keer min is plus, maar waarom eigenlijk? Zie hoofdstuk 11.
[45] Goal!

5 e-gambling

Nu we het getal e en de functie exp, gedefinieerd door middel van twee equivalente uitdrukkingen

$$\exp(x) = \lim_{n\to\infty} (1 + \frac{x}{n})^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

tot onze beschikking hebben, keren we terug naar de kansrekening en de wetten van het toeval. Waar we in hoofdstuk 2 meteen in het diepe sprongen met vragen over toevallige reële getallen en vervolgens eigenlijk alleen maar redeneerden aan de hand van de cijfers $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ gaan we nu veel systematischer mogelijkheden tellen in eenvoudige discrete contexten. Combinatoriek en kansrekening voor toevalsexperimenten met maar eindig of aftelbaar veel antwoorden liggen dicht bij elkaar. Ook hier is het weer de elementaire letteralgebra waarin de zaken bij elkaar komen. We beginnen dit hoofdstuk met de algebra in het binomium van Newton, eindigen met de driehoek van Pascal, en bespreken en passant de kansrekening van het jaarlijks terugkerende evenement van de sinterklaaslootjes.

5.1 Newtons binomium

Uitdrukkingen van de vorm

$$(a+b)^n, \tag{5.1}$$

zijn we tegengekomen aan het begin van sectie 4.3. Met $b=1$ en $n=14$ was het uitwerken van $(1+a)^{14}$ het startpunt voor het uiteindelijke inzicht dat beide uitdrukkingen voor $\exp(x)$ hierboven inderdaad dezelfde $\exp(x)$ definiëren. Het is goed om dat nog eens wat systematischer te bekijken, al heb je dat zelf misschien al gedaan in dat voorbeeld.

Als we (5.1) uitwerken door steeds uit elke factor een van de twee letters te kiezen en die met elkaar te vermenigvuldigen dan krijgen we met bijvoorbeeld $n=5$ precies $32=2^5$ termen, waaronder termen als *abbaa* en *baaba*. Omdat we bij letteralgebra rekenen met de conventie dat de volgorde bij vermenigvuldigen niet uitmaakt, kunnen we beide deze termen herschrijven als a^3b^2. We gooien dus *abbaa* en alle andere termen met hetzelfde aantal a's (en dan dus ook hetzelfde aantal b's) op één hoop, en komen zo in totaal op tien termen, die allemaal reduceren tot dezelfde a^3b^2. Bij elkaar geeft dit de term $10a^3b^2$ in de expansie hieronder. Als we alle 32 termen zo netjes herschrijven en groeperen dan vinden we dat

$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5, \tag{5.2}$$

met coëfficiënten $1, 5, 10, 10, 5$ en 1, samen 32 natuurlijk.

Er zijn verschillende manieren om erachter te komen hoe de coëfficiënten precies tot stand komen. Als voorbereiding voor uitdrukkingen als

$$(a+b+c+d+f+g+h+i+j+k)^{42}$$

gaan we daar nu wat omslachtig op in. Je kunt bijvoorbeeld de uitdrukking eerst wat algemener maken en kijken naar

$$(a_1+b_1)(a_2+b_2)(a_3+b_3)(a_4+b_4)(a_5+b_5).$$

Als je dit uitwerkt krijg je echt 32 verschillende termen, waaronder

$$a_1b_2b_3a_4a_5 = a_1a_4a_5b_2b_3 = a^3b^2,$$

met de laatste gelijkheid geldig mits $a_1 = a_4 = a_5 = a$ en $b_2 = b_3 = b$. De geordende middelste vorm bevat de extra informatie die we nodig hebben. Aan de subscripts kun je zien dat we deze term krijgen als we de drie a's respectievelijk uit de eerste, de vierde en de vijfde factor kiezen en vervolgens de twee b's uit de tweede en de derde factor. Bij de eerste a hadden we vijf keuzemogelijkheden, bij de tweede nog vier, en bij de derde nog drie om in totaal drie verschillende factoren te kiezen waaruit we de a nemen. In totaal $5 \times 4 \times 3$ mogelijkheden dus.

Met andere keuzes van de factoren krijgen we andere termen, tenzij we dezelfde drie (verschillende) factoren (in dit geval de eerste, de vierde en de vijfde) in een andere volgorde kiezen. Hoeveel van zulke volgorden voor dezelfde drie zijn er? Wel, je kunt weer kiezen welke van factor nummer 1, 4 en 5 je als eerste kiest, drie mogelijkheden dus, daarna twee mogelijkheden voor de tweede factor, en dan nog maar één mogelijkheid voor de laatste. We krijgen de term $a_1a_4a_5$ dus $3 \times 2 \times 1$ keer. Voor elke andere keuze van drie uit vijf factoren gaat het precies hetzelfde. Elke term kunnen we in $3 \times 2 \times 1$ verschillende volgorden kiezen.

Met $5 \times 4 \times 3$ keuzes voor de eerste, de tweede, en de derde a tellen we elke term dus niet één keer maar $3 \times 2 \times 1$ keer. We concluderen daarom dat we de term a^3b^2 in totaal

$$\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = \frac{5!}{3!2!}$$

keer krijgen. In de teller staat het aantal keuzes, in de noemer het aantal dubbele keuzes. We noteren dit getal als $\binom{5}{3}$, dus

$$\binom{5}{3} = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10.$$

Als we eenmaal bepaald hebben welke a-termen we kiezen, dan liggen de b-termen natuurlijk ook vast. Daar hoeven we ons dus geen zorgen meer over te maken. Als je dat toch doet, bedenk dan dat we na de $5 \times 4 \times 3$ mogelijkheden voor de drie a's nog 2×1 mogelijkheden hebben voor de twee factoren waaruit we de b's kiezen. Maar elke van die mogelijkheden kan ook weer in 2×1 volgordes. Dus het aantal mogelijkheden om a^3b^2 te kiezen met verschillende a's en b's is ook gelijk aan

$$\frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1}$$

en je ziet dat de conclusie hetzelfde is. Het is een kwestie van smaak aan welke van de twee beschrijvingen je de voorkeur geeft, maar voor $(a+b+c)^5$

is de laatste uitdrukking het natuurlijke opstapje. Alleen naar de a's kijken is dan niet meer voldoende.

Bij $(a+b)^5$ hadden we natuurlijk ook alleen naar de b's kunnen kijken. De uitkomst zou dan

$$\binom{5}{2} := \frac{5\times 4}{2\times 1} = \frac{5!}{2!3!}$$

geweest zijn, en (weer) zien we dat $\binom{5}{3} = \binom{5}{2}$.

De uitdrukkingen $\binom{5}{3}$ en $\binom{5}{2}$ worden binomiaalcoëfficiënten genoemd. Voor gehele getallen $0 \le k \le n$ gedefiniëerd door

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{5.3}$$

Aardig in dit verband is dat '5\choose3' het LaTeχ-commando is waarmee de binomiaalcoëfficiënt gezet wordt. Op de GRM gaat $\binom{5}{3}$ het daarentegen weer via $5nCr3$[1]. Kies 3 uit 5, op hoeveel manieren kan dat? Op

$$\binom{5}{3} = 10$$

manieren, zoals we hierboven hebben gezien. Het verband met de kansrekening zie je als je de factoren in $(a+b)^5$ aftelt en elke keer met een munt gooit om te bepalen of je de a of de b kiest. De kans op a^3b^2 (dus de kans op drie keer a en twee keer b) is dan gelijk aan

$$\frac{10}{32},$$

want van het totale aantal $2\times 2\times 2\times 2\times 2 = 32$ mogelijkheden zijn er tien die a^3b^2 geven. Als alle mogelijkheden even waarschijnlijk zijn[2], dan is de kans op a^3b^2 het quotiënt van deze twee aantallen: $\frac{10}{32} = \frac{5}{16}$. Hoe meer je de breuk vereenvoudigt hoe minder goed je ziet hoe je aan het antwoord gekomen bent.

Wat voor 3 kan, kan ook voor 2. Ook a^2b^3 heeft coëfficiënt 10, zoals we al zagen. Voor 1 en 4 zie je waarschijnlijk vrijwel onmiddellijk dat a^4b en ab^4 precies vijf keer voorkomen. De term $a_1a_2a_3a_4a_5$ lijkt een buitenbeentje, net als de term $b_1b_2b_3b_4b_5$. Beide termen kun je op $5\times 4\times 3\times 2\times 1 = 5!$ verschillende manieren kiezen maar elke volgorde geeft dezelfde term. Dat a^5 en b^5 elk precies één keer voorkomen zal geen verbazing wekken, maar het is

[1] Daarmee heeft nCr zich in het jargon een plaats verworven.

[2] Als de munt zuiver is.

goed om even na te gaan dat het voor dit geval wat ingewikkelde argument met de indices tot dezelfde conclusie leidt. We concluderen dat

$$(a+b)^5 = a^5 + 5a^4b + \binom{5}{3}a^3b^2 + \binom{5}{2}a^2b^3 + 5ab^4 + b^5,$$

en dat met de afspraak $0! = 1! = 1$ ook de buitenbeentjes in het systeem passen. Omdat

$$\binom{5}{5} = \binom{5}{0} = \frac{5!}{5!0!} = 1,$$

krijgt voor $n = 5$ het binomium van Newton zo de vorm

$$(a+b)^5 = \binom{5}{5}a^5b^0 + \binom{5}{4}a^4b^1 + \binom{5}{3}a^3b^2 + \binom{5}{2}a^2b^3 + \binom{5}{1}a^1b^4 + \binom{5}{0}a^0b^5.$$

Opgave 5.1. Waarom geldt $\binom{5}{2} = \binom{5}{3}$ veel algemener? Dus waarom is

$$\binom{n}{k} = \binom{n}{n-k}$$

als k en n geheel zijn met $0 \le k \le n$?

Er zijn nu twee manieren om het binomium in een standaardsomnotatie te zetten. Met n in plaats van 5 hebben we

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

en met een aparte naam voor $j = n - k$ ziet dat er zo nog net iets symmetrischer uit:

$$(a+b)^n = \sum_{\substack{k,j\ge 0\\ k+j=n}} \binom{n}{k\ j} a^k b^j, \tag{5.4}$$

waarbij

$$\binom{n}{k\ j} = \frac{n!}{k!j!}.$$

De coëfficiënten in (5.4) hebben we bepaald voor $n = 5$ door combinaties te tellen[3], k keer een a en $j = 5 - k$ keer een b, met $k = 0, 1, 2, 3, 4, 5$ en $j = 5, 4, 3, 2, 1, 0$. Wat we met 5 kunnen, kunnen we met alle $n = 1, 2, 3, 4, 5, 6, 7, \ldots$, dus de stap van 5 naar een willekeurige gehele $n \ge 1$ is klein. Maar ook, wat we met a en b kunnen, kunnen we ook met a, b en c, en (5.4) verklapt al wat we dan krijgen.

[3] Slim tellen is het vakgebied dat combinatoriek heet.

Opgave 5.2. Laat zien dat

$$(a+b+c)^{11} = \sum_{\substack{k,j,l \geq 0 \\ k+j+l=11}} \binom{11}{k\ j\ l} a^k b^j c^l,$$

waarin

$$\binom{11}{k\ j\ l} = \frac{11!}{k!j!l!}.$$

Hint: bekijk

$$(a_1+b_1+c_1)(a_2+b_2+c_2)\cdot\cdots\cdot(a_{10}+b_{10}+c_{10})(a_{11}+b_{11}+c_{11}).$$

Opgave 5.3. Overtuig jezelf ervan dat voor gehele $n, m \geq 1$ geldt dat

$$(a_1+\cdots+a_m)^n = \sum_{\substack{k_1,\ldots,k_m \geq 0 \\ k_1+\cdots+k_m=n}} \binom{n}{k_1\ \ldots\ k_m} a_1^{k_1}\cdots a_m^{k_m}$$

met

$$\binom{n}{k_1\ \ldots\ k_m} = \frac{n!}{k_1!\cdots k_m!}.$$

Schrijf bijvoorbeeld $(a+b+c+d)^5$ helemaal uit.

De kansrekening blijft dichtbij. We hadden het al over zuivere munten. Met a voor kop en b voor munt is bij vijf keer gooien de kans op drie keer kop (a) en twee keer munt (b) af te lezen uit het binomium voor $(a+b)^5$. Die kans krijg je door coëfficiënt van a^3b^2 te delen door 32, maar dat is hetzelfde als $\binom{5}{3}$ vermenigvuldigd met

$$\frac{1}{32} = \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2.$$

In deze wat merkwaardige factorisatie voor $\frac{1}{32}$ is de eerste $\frac{1}{2}$ de kans op kop en de tweede $\frac{1}{2}$ de kans op munt. Nemen we een munt die met kans p kop en met kans $1-p$ munt geeft, dan is de kans op drie keer kop bij de eerste drie worpen en twee keer munt bij de laatste twee worpen gelijk aan

$$p^3(1-p)^2.$$

De p is hier te zien als een parameter en we nemen natuurlijk aan dat $0 < p < 1$. De gevallen $p = 0$ en $p = 1$ zijn flauw, en kansen groter dan 1 of kleiner dan 0 kunnen niet. Vragen we bij zo'n munt naar de kans op drie keer kop en twee keer munt, zonder belang te hechten aan de volgorde van de uitkomsten dan moeten we de verschillende volgorden tellen. Maar dat hebben we zojuist in detail gedaan. Dus dit is nu makkelijk:

Opgave 5.4. Leg uit waarom de kans op drie keer kop en twee keer munt bij vijf keer gooien gelijk is aan $\binom{5}{3}p^3(1-p)^2$.

Als je er even over nadenkt dan zijn de kansen op respectievelijk $5, 4, 3, 2, 1$ en 0 keer munt precies de termen in het binomium voor $(a+b)^5$ als je $a = p$ en $b = 1 - p$ invult en zo de totale kans 1 schrijft als som van de kansen op de verschillende gebeurtenissen. Hier is een wat ingewikkeldere vraag.

Opgave 5.5. Laten we als experiment nu eens 21 keer gooien met een niet per se eerlijke dobbelsteen, met kansen $p_1, p_2, p_3, p_4, p_5, p_6$ op respectievelijk $1, 2, 3, 4, 5$ en 6. Kun je met behulp van Opgave 5.3 de kans op een keer een 1, twee keer een 2, drie keer een 3, vier keer een 4, vijf keer een 5 en zes keer een 6 bepalen?

Hoe vaker we gooien hoe ingewikkelder het wordt. Onverwachts leuk wordt het als we de parameters af laten hangen van het aantal keren dat we gooien.

5.2 Sinterklaaslootjes

Het ankerpunt in deze sectie is de uitdrukking

$$\left(1 - \frac{1}{n}\right)^n, \tag{5.5}$$

die we inmiddels kennen als benadering voor $\frac{1}{e}$, ongeveer[4] 0,368. Voor elke $n = 1, 2, 3, \ldots$ heeft deze uitdrukking een voor de hand liggende interpretatie in de kansrekening rond het jaarlijkse sinterklaaslootjesgebeuren waarbij door het hele land groepjes mensen hun namen opschrijven en in een bakje doen om vervolgens bij voorkeur niet hun eigen naam te trekken. Als in zo'n groepje de eerste persoon, laten we hem Joostje noemen, een briefje trekt, dan is de kans dat hij zichzelf niet trekt $1 - \frac{1}{n}$. Zo meteen gaan we kijken wat

[4] Voor een keertje handig om de rekenmachine bij de hand te hebben.

er gebeurt als Joostje het briefje niet terugdoet in de hoed, maar we houden het eerst nog even eenvoudig, en nemen aan dat Joostje, na op het briefje gekeken te hebben, het briefje weer terugdoet in de hoge hoed. Nu trekt de volgende persoon, laten we zeggen oom Ronald, een briefje. Natuurlijk is ook de kans dat oom Ronald zichzelf niet trekt gelijk aan $1 - \frac{1}{n}$. De kans dat Joostje en oom Ronald allebei zichzelf niet getrokken hebben, is gelijk aan $(1 - \frac{1}{n})^2$, als we er tenminste vanuitgaan dat de trekking van Joostje die van oom Ronald niet beïnvloedt, met andere woorden dat beide trekkingen onafhankelijk van elkaar gebeuren[5].

Wanneer we op deze manier alle personen in de groep steeds een briefje laten trekken, lezen en terugleggen, dan is de kans dat niemand zichzelf trekt gelijk aan $(1 - \frac{1}{n})^n$. Dit is enerzijds precies de 'negatieve-rente-op-negatieve-rente'-uitdrukking $e_n(-1)$, maar ook een uitdrukking zoals bestudeerd in de vorige subsectie met een succeskans $\frac{1}{n}$ die van n afhangt. Met $n = 6$ is dit de kans dat je geen enkele keer 3 gooit als je zes keer gooit met een dobbelsteen en met $n = 52$ is dit de kans dat je geen enkele keer de schoppenaas trekt als je 52 keer een kaart trekt (en weer terugsteekt en schudt) uit een pak met 52 kaarten[6].

Opgave 5.6. Die eis van onafhankelijkheid tussen de trekkingen moet er wel bij! Als bijvoorbeeld iedereen hetzelfde briefje trekt (dat is natuurlijk verre van onafhankelijk), wat is dan de kans dat niemand in de groep zichzelf trekt?

De kans dat niemand zichzelf trekt (als we de lootjes elke keer terugdoen in de hoge hoed) voor grote groepen convergeert dus kennelijk naar $\frac{1}{e} = e^{-1}$, en is voor een grote n dus ongeveer gelijk aan 0,368 en dus niet, zoals je misschien verwacht zou hebben, bijna 1 of bijna 0.

De procedure bij sinterklaaslootjes[7] is net iets anders: een briefje dat getrokken wordt, gaat niet meer terug in de hoge hoed. Sinterklaas laat zijn Pieten trekken zonder teruglegging. Wie dit vaker geprobeerd heeft, weet dat het regelmatig gebeurt dat de hele trekking opnieuw moet omdat iemand zichzelf [8] heeft getrokken.

Een interessante vraag is wat de kans is dat er iemand in het gezelschap zichzelf trekt. Met andere woorden, hoe vaak (gemiddeld gesproken) moet de trekking over?[9] Het antwoord op deze vraag hangt natuurlijk af van de

[5] Het product van de twee kansen, zo bepaalt de kanswet.
[6] Waarin de schoppenaas precies een keer voorkomt.
[7] Ook Zwarte Piet erbij natuurlijk.
[8] Oom Ronalds zus trok eens zichzelf, zei niets, cadeau voor tante viel een beetje op.
[9] Helemaal opnieuw dus.

grootte van de groep, en net als bij het eerste voorbeeld waarbij we met teruglegging werkten, zijn we benieuwd naar de kans dat niemand zichzelf trekt wanneer het om grote groepen gaat. We kijken dus weer wat er met die kans gebeurt als n willekeurig groot genomen wordt. Daartoe moeten we eerst een uitdrukking voor die kans hebben. Met teruglegging is de kans bij $n = 4$ gelijk aan

$$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4},$$

maar zonder teruglegging is het lastiger. Het wellicht voor de hand liggende antwoord

$$\frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} \times \frac{0}{1}$$

kan niet goed zijn.

Opgave 5.7. Als je met twee mensen Sinterklaas viert, wat is dan de kans dat niemand zichzelf trekt? En bij drie personen? En bij vier personen?

Zoveel mogelijkheden zijn er niet, dus deze opgave is vast wel gelukt. Maar wat als n groot is?

Opgave 5.8. Is de kans voor steeds grotere n bij trekken zonder teruglegging echt anders of ongeveer hetzelfde als de kans bij trekken met teruglegging? Wat zegt je intuïtie?

Of je intuïtie goed is moet blijken. Mogelijkheden tellen bij trekken zonder teruglegging is lastig, omdat de eerste trekking gevolgen heeft voor de tweede, die weer gevolgen heeft voor de derde, en dat werkt zo door. Het principe van inclusie-exclusie, een slimme telmethode biedt hier uitkomst.

Als A en B twee mogelijke gebeurtenissen zijn met kans $P(A)$ op A en kans $P(B)$ op B, wat is dan de kans dat A of[10] B optreedt? Lees de voetnoot! Als je $P(A)$ en $P(B)$ bij elkaar optelt, dan heb je de kans op zowel A als B[11] twee keer meegenomen. Dat kun je goedmaken door de kans op de doorsnede van A en B er weer van af te trekken, en dat vertelt ons dat de kans op A of B gelijk zou moeten zijn aan

$$P(A) + P(B) - P(A \text{ en } B).$$

[10] Hiermee bedoelen we de kans dat ten minste een van beide gebeurtenissen optreedt.

[11] De doorsnede van A en B.

Opgave 5.9. Bereken met deze formule[12] de kans dat je bij het werpen met een dobbelsteen een even getal gooit of een getal groter dan 3.

Op dezelfde manier is bij drie gebeurtenissen A, B en C de kans op A of B of C gelijk aan

$$\begin{gathered} P(A) + P(B) + P(C) \\ -P(A \text{ en } B) - P(A \text{ en } C) - P(B \text{ en } C) \\ +P(A \text{ en } B \text{ en } C), \end{gathered}$$

en zijn de termen met een minteken net als eerder bedoeld om alles wat dubbel geteld is er weer van af te trekken. Alleen trekken we er dan weer te veel van af, en dat wordt weer goedgemaakt met de laatste term.

Opgave 5.10. Neem bij het gooien met een dobbelsteen voor A de gebeurtenis van een even uitkomst, voor B de gebeurtenis van een uitkomst minstens 3, en voor C de gebeurtenis dat je 4 gooit. Pas de regel toe en verifieer dat het antwoord hetzelfde is als wat je rechtstreeks uitrekent (alleen met uitkomst 1 is het mis).

Opgave 5.11. Nu je begrijpt hoe het principe werkt, kun je een formule opschrijven voor de kans dat minstens een van de gebeurtenissen $A_1, A_2, \ldots, A_n$ optreedt. Doe dat. De formule wordt het principe van inclusie-exclusie genoemd. Overtuig jezelf ervan dat je de goede formule hebt.

We passen het principe van inclusie-exclusie nu toe op het sinterklaaslootjes-probleem met vier personen, laten we zeggen Diederik, Koeno, Marja en Anne. Als achtereenvolgens Anne, Diederik, Koeno en Marja hun briefjes trekken dan zetten ze zichzelf in een volgorde. Je moet er misschien even over nadenken, maar in welke volgorde ze dat doen maakt niet uit. We noemen de gebeurtenissen nu ook Anne, Diederik, Koeno en Marja. Anne is dus de gebeurtenis dat hij, Anne, zichzelf trekt. De kans daarop is $\frac{1}{4}$, en bij eerlijk spel geldt hetzelfde voor de kans op Diederik[13]. Ook Koeno en Marja zijn hun eigen gebeurtenis, elk met kans $\frac{1}{4}$. Dus de som van de kansen op respectievelijk Anne, Diederik, Koeno en Marja is gelijk aan 4 keer $\frac{1}{4}$, en dat is 1. Duiden we de gebeurtenissen alleen met initialen aan[14] dan is

$$P(A) + P(D) + P(K) + P(M) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1,$$

[12] Dat kan natuurlijk makkelijker maar daar gaat het nu even niet om.
[13] Die altijd al zijn eigen gebeurtenis was.
[14] Dat lag misschien al voor de hand.

en dat is natuurlijk meer dan de kans dat iemand zichzelf trekt.

Evenzo zijn uit deze vier personen

$$\binom{4}{2} = \frac{4 \cdot 3}{2} = 6$$

tweetallen te maken. De kans dat binnen zo'n tweetal, zeg Koeno en Marja, ieder zichzelf trekt (Koeno trekt Koeno, Marja trekt Marja) is gelijk aan

$$\frac{1}{4} \cdot \frac{1}{3},$$

en met zes tweetallen is de som van al deze kansen gelijk aan

$$\frac{4 \cdot 3}{2} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{2}.$$

Let op, ook dit is weer meer dan de kans dat twee mensen zichzelf trekken.

Met de vier drietallen gaat het hetzelfde. De som van de kansen op drietallen[15] die ieder zichzelf trekken is nu

$$\frac{4 \cdot 3 \cdot 2}{3 \cdot 2} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2 \cdot 3},$$

en voor het ene viertal is de kans

$$\frac{4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{2 \cdot 3 \cdot 4}.$$

Die laatste wist je natuurlijk al.

Het principe van inclusie-exclusie voor vier gebeurtenissen leidt nu tot een kans van

$$\binom{4}{1} \cdot \frac{1}{4} - \binom{4}{2} \cdot \frac{1}{4} \cdot \frac{1}{3} + \binom{4}{3} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} - \binom{4}{4} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 1 \tag{5.6}$$

dat er iemand is die zichzelf trekt. Als je (5.6) zoals we al deden term voor term vereenvoudigt dan staat er

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}.$$

Dat is natuurlijk te mooi om toevallig te zijn: het lijkt wel het begin van de reeks voor EXP(-1), maar het klopt net niet. Trek het van 1 af en je krijgt

[15] Het Centraal Eindexamen Wiskunde B stopte hier de correcties door het thuisfront.

de eerste '$1+4$' termen van de reeks voor $\mathrm{EXP}(-1)$. De kans dat niemand zichzelf trekt bij trekken zonder terugleggen is dus gelijk aan[16]

$$E_4(-1) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!},$$

en dat is (4.13) met $n = 4$ en $a = -1$[17].

We hadden al gezien dat $n = 4$ en $a = -1$ in (4.6) de kans $(1 - \frac{1}{4})^4$ geeft dat bij trekken met terugleggen niemand zichzelf trekt. Is dat geen toeval?[18] Het verschil tussen wel of niet terugleggen is precies het verschil tussen de rente-op-renteformule[19] die tot de benadering van $\exp(-1)$ leidt en de directe benadering van $\mathrm{EXP}(-1)$.

Opgave 5.12. Overtuig jezelf ervan dat de regelmaat zich inderdaad doorzet: de kans dat tenminste één persoon zichzelf trekt bij een groep van n personen is gelijk aan

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots \pm \frac{1}{n!}.$$

Of de laatste term een plus of een min heeft, hangt af van of n even of oneven is.

De onontkoombare conclusie is nu dat, hoe groot de groep ook is, de kans dat je niet opnieuw moet trekken altijd ongeveer 0,368 is. We concluderen ook dat het voor grote n weinig uitmaakt of het loten met of zonder teruglegging gebeurt, want zowel $E_n(-1)$ als $e_n(-1)$ liggen steeds dichter bij $\frac{1}{e}$ als n groter genomen wordt. Had je dat verwacht? Hoe dan ook, het is het slim tellen van de mogelijkheden bij het trekken met terugleggen dat tot deze conclusie leidt.

Dat slimme tellen hebben we via het principe van inclusie-exclusie gedaan in de berekening van de kansen, met getallen tussen 0 en 1 die in de juiste combinatie tot onze verrassing een benadering voor $\frac{1}{e}$ geven. Anders dan bij bijvoorbeeld het gooien van een onzuivere munt of dobbelsteen, of het trekken van lootjes uit een hoed die niet allemaal even groot zijn, zijn alle kansen daarbij eigenlijk fracties: steeds het quotiënt van het aantal mogelijkheden voor de gebeurtenis gedeeld door het totaal aantal mogelijkheden, te weten $4 \cdot 3 \cdot 2 = 4!$ bij $n = 4$. Je kunt die mogelijkheden ook rechtstreeks tellen, zonder over kansen of fracties te praten, maar de antwoorden ogen net iets minder mooi en het verband met $\frac{1}{e}$ zie je minder direct.

[16] We gebruiken nu de notatie uit het vorige hoofdstuk.
[17] De eerste twee termen vallen weg.
[18] Toeval bestaat niet in de kansrekening.
[19] Met een op jaarbasis dodelijk negatieve rente van honderd procent.

Opgave 5.13. Laat door slim tellen zien dat het aantal mogelijkheden dat met vier personen iemand zichzelf trekt bij trekken zonder terugleggen gelijk is aan $4 \cdot 3 \cdot 2 - 4 \cdot 3 + 4 - 1$ en dat het aantal mogelijkheden dat niemand zichzelf trekt gelijk is aan $4 \cdot 3 - 4 + 1$. Hint: je weet al dat deze uitkomsten juist zijn maar het gaat om de structuur in het antwoord. Wat zijn de aantallen bij n personen?

5.3 Van discreet naar continu

We hebben gezien dat $p = \frac{1}{n}$ een voor de hand liggende natuurlijk keuze is wanneer we n identieke experimenten doen. Dat maakt nieuwsgierig. Wat gebeurt er bijvoorbeeld als we dit iets algemener maken en niet $p = \frac{1}{n}$ kiezen maar $p = \frac{\lambda}{n}$, met λ een positief getal[20]? De kans op geen enkel succes is dan natuurlijk[21]

$$\left(1 - \frac{\lambda}{n}\right)^n,$$

en als n groter wordt, convergeert dit zoals we inmiddels wel weten naar $e^{-\lambda}$.

Kennelijk is het zo dat bij een p die precies of ongeveer evenredig[22] is met $\frac{1}{n}$ het aantal experimenten dat hier 'natuurlijk' bij past, ongeveer evenredig is met n. Dat is niet zo gek. Als we n experimenten doen met een kleine kans p op succes, dan is het redelijk te verwachten dat je ongeveer np successen zult hebben. We verwachten dan in totaal ongeveer $\frac{\lambda}{n} \times n = \lambda$ successen, en dat getal hangt niet van n af!

We kunnen de kans op precies k successen opschrijven met behulp van de binomiale verdeling. Deze kans noemen we even $p_n(k)$ en is gelijk aan (gebruik de binomiale verdeling!)

$$\begin{aligned} p_n(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^n\left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{n^k(n-k)!}\left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Opgave 5.14. Laat met behulp van je kennis over e zien dat deze uitdrukking voor

[20] Alleen zinvol als $\lambda \leq n$, maar n wordt willekeurig groot, we nemen dus aan dat $n \geq \lambda$.
[21] Net als bij de sinterklaaslootjes met teruglegging.
[22] p is van dezelfde orde als n.

vaste k, als n groter wordt, convergeert naar

$$p_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{5.7}$$

De derde factor met de faculteiten is het lastigst, hoe zie je dat deze voor vaste k naar 1 gaat? Schrijf er eentje uit en kijk nog eens goed naar (4.6) en de formules daarboven.

Dit resultaat illustreert opnieuw de samenhang van de binomiale verdeling met e. Niet alleen hebben we gebruikgemaakt van het feit dat $(1+\frac{1}{n})^n \to e$, maar ook komt de al eerder besproken doorlopende som EXP(1) voor e weer om de hoek kijken.

Opgave 5.15. Laat zien dat

$$p_\lambda(0) + p_\lambda(1) + p_\lambda(2) + \cdots = 1.$$

De kansverdeling die door (5.7) beschreven is, wordt de Poisson verdeling genoemd met parameter λ.

Wat we hierboven gezien hebben is dat als we p op de 'juiste' manier van n laten afhangen, we een zogenaamde limietstelling ontdekken: als p 'van de orde $\frac{1}{n}$ is', dan convergeert de kans op precies k successen naar een prachtige limietwaarde als n groter wordt. Gewapend met deze kennis kunnen we als kansrekenaar ook naar zogenaamde continue processen kijken. Zulke processen kom je onherroepelijk tegen in het dagelijks leven als je vragen stelt als: wat is de kans dat een lamp stukgaat als ik hem aan laat staan? Het antwoord op die vraag is natuurlijk dat de lamp vroeg of laat stukgaat, dus die kans is 1. Maar dat antwoord voegt weinig toe aan de kennis die we hebben, namelijk dat lampen nu eenmaal stukgaan.

Een herformulering van de vraag is dus nodig om wat toe te voegen aan onze oppervlakkige praktische kennis. Daartoe beperken we onze aandacht eerst tot een vast tijdsinterval, zeg tussen nu, tijdstip nul, en straks, tijdstip t. We kunnen dit tijdsinterval opdelen in kleine intervalletjes[23] ter lengte $\frac{1}{n}$, en poneren dat de kans dat wanneer de lamp aan het begin van zo'n klein interval nog niet stuk is, de kans dat hij in dat gegeven intervalletje wel stukgaat[24] gelijk is aan $p = \frac{\lambda}{n}$.

[23] We doen net of dit past.

[24] Deze kans is proportioneel met de intervallengte, met proportionaliteitsconstante λ.

Waarom we die aanname zouden mogen doen mag Joost weten. We gaan ervan uit dat de lamp geen geheugen heeft en niet slijt. Het enige ingrediënt is dat op elk moment er een vaste kans $p = \frac{\lambda}{n}$ is dat de lamp tussen dan en iets $(\frac{1}{n})$ later stukgaat. Tot aan tijdstip t zijn er precies nt intervallen, dus de kans dat de lamp voor die tijd niet stukgaat is $(1-\frac{\lambda}{n})^{nt}$. Als dit argument je overtuigt dan ligt via een limietovergang met $n \to \infty$ de conclusie voor de hand dat de kans dat de lamp op tijdstip t nog brandt gelijk is aan $e^{-\lambda t}$. Deze formule definieert de exponentiële verdeling.

Opgave 5.16. Leuk om over na te denken: hoe groot is de gemiddelde wachttijd tot het stukgaan van de lamp bij deze exponentiële verdeling?

Opgave 5.17. De samenhang tussen binomiale, exponentiële en Poisson verdeling wordt zichtbaar door de juiste keuzes van de parameters. Reflecteer hierop, in het bijzonder op de rol van de parameters λ en p.

5.4 De driehoek van Pascal

We hebben hem niet echt nodig, die driehoek van Pascal[25], maar als hulpmiddel om bij het binomium van Newton, bijvoorbeeld bij

$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5, \tag{5.8}$$

het rekenwerk te reduceren, is hij te mooi om hier niet te bespreken. Een andere reden om er nu iets over te zeggen is dat we een met deze driehoek vergelijkbaar schema nog zullen tegenkomen in het volgende hoofdstuk.

Iedereen heeft Pascals driehoek wel eens gezien, dit is hem:

$$\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1
\end{array}$$

[25] Pascal? Ook in China, Italie, ...

$$1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1$$

Deze driehoek is het unieke schema van rijen getallen waarin elke volgende rij steeds een getal meer heeft, het meest linker- en meest rechtergetal steeds 1 zijn, en elk ander getal steeds de som is van de twee getallen er direct boven. Je kunt de driehoek[26] naar beneden door laten lopen zover je wilt. Elke rij volgt uit de vorige rij. Via het uitwerken van bijvoorbeeld

$$(a+b)^5 = (a+b)(a+b)^4 = (a+b)(a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4) \tag{5.9}$$

zie je hoe de vijfde de zesde rij maakt en wat het verband is met het binomium van Newton.

Opgave 5.18. Ga na dat als je (5.9) uitwerkt, dat je $10a^3b^2$ krijgt als de som van a keer $6a^2b^2$ en b keer $4a^3b$, en geef aan welke drie getallen in de Pascal-driehoek hierbij betrokken zijn. Conclusie?

De conclusie is dat alle termen in de driehoek van Pascal precies alle binomiaalcoefficiënten zijn: Het k-de getal in de n-de rij is gelijk aan $\binom{n}{k-1}$[27].

Opgave 5.19. Laat met behulp van de driehoek van Pascal zien dat voor $1 \leq k \leq n-1$ geldt dat

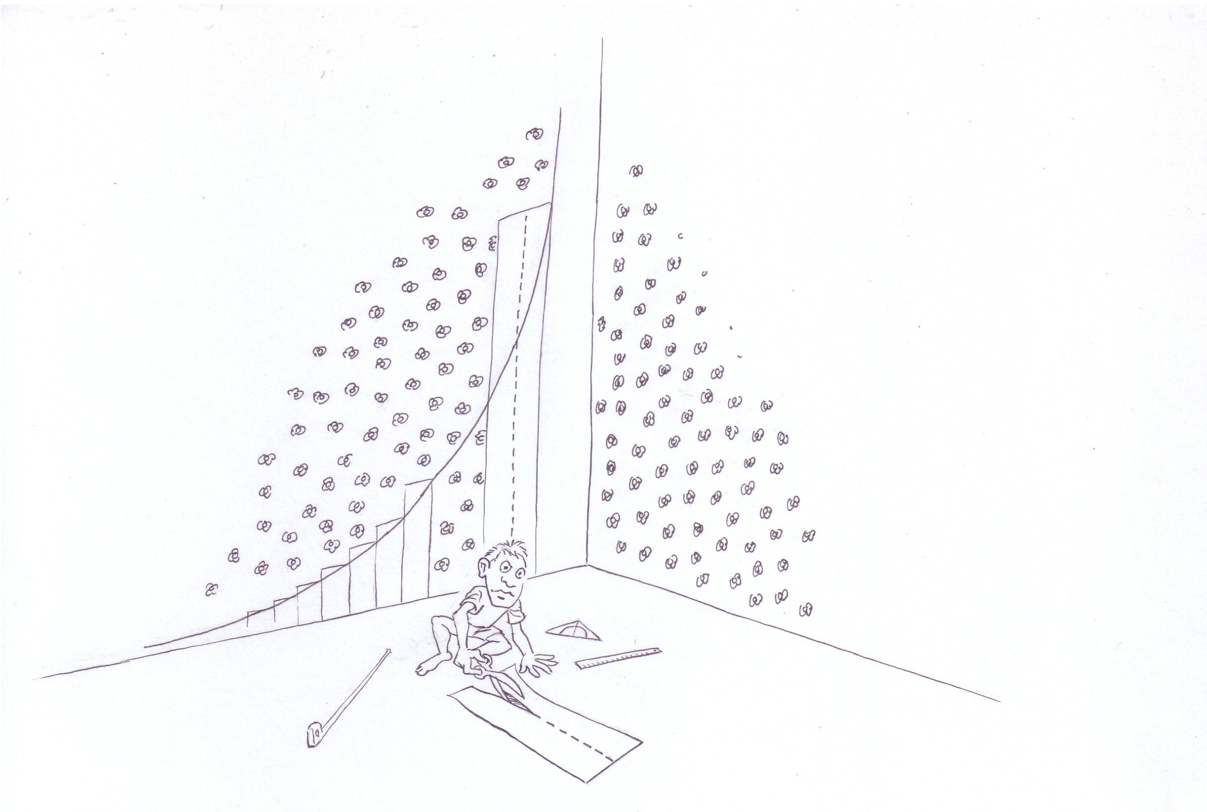
$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \tag{5.10}$$

Het aardige van wiskunde is dat je veel mooie gelijkheden (of identiteiten) op verschillende manieren kunt zien en ontdekken. De gelijkheid in (5.10) is weer een fraai voorbeeld. In de laatste twee opgaven van dit hoofdstuk nog twee andere manieren om deze gelijkheid ook te bewijzen[28].

Opgave 5.20. Bewijs (5.10) rechtstreeks vanuit (5.3) met breuken en faculteiten door de noemers gelijknamig te maken en gewoon stug door te rekenen.

[26] Die al lang voor Pascal in China bekend was.
[27] Wel even nagaan natuurlijk.
[28] Nieuwe bewijzen van oude stellingen is een wat ondergewaardeerde sport.

Opgave 5.21. Bewijs (5.10) met behulp van de combinatorische interpretatie van $\binom{n}{k}$ als het aantal deelverzamelingen van k elementen uit een verzameling van n elementen. Splits deze deelverzamelingen op in twee soorten: zij die het laatste element wel bevatten, en zij die dat dat niet doen. Laat zien dat (5.10) hieruit volgt.



6 Oppervlakterekening zo makkelijk mogelijk!

In het vervolg van dit boek spelen integralen en oppervlakten een grote rol, en voordat we ons wagen aan allerlei ingewikkelde integralen ligt het voor de hand om eerst eens te kijken naar de oppervlakten van gebieden beschreven met polynomen.

We beginnen ons verhaal met het gebied in het xy-vlak begrensd door

$$x = 1, \quad y = 0 \quad \text{en} \quad y = x^n.$$

Met de moderne integraalrekening is de oppervlakte van dit gebied gelijk aan

$$\int_0^1 x^n \, dx = \left[\frac{x^{n+1}}{n+1}\right]_0^1 = \frac{1}{n+1},$$

maar dat kunnen we ook veel rechtstreekser[1] en zonder de machinerie van Newton en Leibniz begrijpen.

6.1 Van rechthoeken naar oppervlakten

Het gaat ons om te beginnen om de vraag hoe we zien dat

$$A_n = \text{ de oppervlakte van } \{(x, y) : 0 \leq y \leq x^n \leq 1\} = \frac{1}{n+1}.$$

Als uitgangspunt nemen we dat over de oppervlakte van een rechthoek geen twijfel bestaat, waarbij we zonder eenheden (dus geen vierkante centimeters of zo) werken. De oppervlakte van een rechthoek die parallel ligt aan de assen wordt eenvoudig bepaald door de schaalverdeling op de assen.

Voor $n = 1$ gaat het om het gebied in het xy-vlak begrensd[2] door

$$x = 1, \quad y = 0 \quad \text{en} \quad y = x,$$

een driehoek met oppervlakte $\frac{1}{2}$, precies de helft van de oppervlakte van het vierkant met hoekpunten $(0, 0)$, $(1, 0)$, $(1, 1)$ en $(0, 1)$. Dat de oppervlakte kleiner is voor $n = 2$ dan voor $n = 1$ is evident, hopelijk ook voor de wiskundige die eerst de vraag stelt hoe en of de oppervlakte eigenlijk precies gedefinieerd is. Het is ook duidelijk[3] dat A_2 in ieder geval groter moet zijn dan de oppervlakte van een rechthoek die geheel in het gebied ligt, bijvoorbeeld de rechthoek met hoekpunten $(\frac{1}{2}, 0)$, $(1, 0)$, $(1, \frac{1}{4})$ en $(\frac{1}{2}, \frac{1}{4})$. We concluderen hieruit dat A_2 minimaal $\frac{1}{2} \times \frac{1}{4}$ moet zijn.

Evenzo kunnen we het stuk op de x-as tussen $(0, 0)$ en $(1, 0)$ in drie gelijke stukjes knippen. Door nu zo groot mogelijke rechthoekjes tussen $(\frac{1}{3}, 0)$ en $(\frac{2}{3}, 0)$ enerzijds en tussen $(\frac{2}{3}, 0)$ en $(1, 0)$ anderzijds binnen A_2 te maken, zien we dat moet gelden dat

$$\frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{4}{9} < A_2.$$

Opgave 6.1. Laat, door drie rechthoekjes te tekenen die samen A_2 bevatten zien dat ook geldt[4]

$$A_2 < \frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{4}{9} + \frac{1}{3} \cdot \frac{9}{9}.$$

[1] Toch een soort alchemie, deze formule.
[2] Voor $n = 0$ is de beschrijving onjuist, waarom?
[3] Maak een plaatje.
[4] Nu even niet $\frac{9}{9}$ door 1 vervangen!

We hebben nu dus een ondergrens en een bovengrens voor A_2, grenzen waarin het getal 3 (de 3 van de drie rechthoekjes voor de bovengrens) goed herkenbaar is. Als we de x-as tussen $(0,0)$ en $(1,0)$ niet in drie maar in N stukjes opknippen, met N een willekeurig[5] geheel positief getal, dan gaat het precies zo:

Opgave 6.2. Laat zien dat voor elke gehele positieve N geldt dat

$$A_2 > \frac{1}{N}\left(\frac{1}{N}\right)^2 + \frac{1}{N}\left(\frac{2}{N}\right)^2 + \cdots + \frac{1}{N}\left(\frac{N-1}{N}\right)^2$$

en

$$A_2 < \frac{1}{N}\left(\frac{1}{N}\right)^2 + \frac{1}{N}\left(\frac{2}{N}\right)^2 + \cdots + \frac{1}{N}\left(\frac{N}{N}\right)^2.$$

Dit herschrijven we als twee ongelijkheden

$$\frac{1^2 + 2^2 + \cdots + (N-1)^2}{N^3} < A_2 < \frac{1^2 + 2^2 + \cdots + N^2}{N^3}, \tag{6.1}$$

die moeten gelden voor elke $N \geq 2$. Als we tot de conclusie komen dat er precies één zo'n getal A_2 is, dan moet A_2 wel de oppervlakte zijn die we zochten.

Of de ondergrenzen steeds groter en de bovengrenzen steeds kleiner worden met N is niet onmiddelijk duidelijk, maar wat we toch zeker nodig zullen hebben is dat het verschil tussen de ondergrens en bovengrens klein te maken is door N groot te kiezen. Als we even nadenken[6] over hoe we deze grenzen hebben gevonden dan is onmiddelijk duidelijk dat het verschil precies $\frac{1}{N}$ is, en dat kunnen we zo klein maken als we willen door N groot te nemen.

Mooi, maar wat we echt willen weten is de getalswaarde van A_2 zelf, en daartoe moeten we weten hoe groot $1^2 + 2^2 + \cdots + N^2$ is in verhouding tot N^3. Dat is niet meteen duidelijk, maar de vraagstelling is vergelijkbaar met de strafwerk[7]som van Gauss, namelijk hoe groot is $1^1 + 2^1 + \cdots + N^1$? Die zag meteen dat

$$1 + 2 + \cdots + N = N\,\frac{N+1}{2} = \frac{1}{2}N^2 + \frac{1}{2}N,$$

[5] Oeps, bestaat dat willekeurige positieve gehele getal wel?
[6] Lees: even naar de formules kijken.
[7] Zoek op!

namelijk N keer het gemiddelde, een tweedegraads polynoom in N. Met voor eerste machten een kwadratische uitdrukking ligt een vermoeden voor tweede machten voor de hand: zou het een derdegraads in N kunnen zijn? Dat zou wel erg mooi zijn, want je gaat natuurlijk eenvoudig na dat er in dat geval maar één zo'n derdegraads kan zijn, en welke dat is[8]. Dit leidt tot de vraag of het volgende waar is: voor alle N positief en geheel geldt

$$1^2 + 2^2 + \cdots + N^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}. \tag{6.2}$$

Let wel, er is a priori geen enkele garantie dat (6.2) waar zou zijn, want raden is maar raden[9]! Maar als het waar is, dan is de bovengrens voor A_2 gelijk aan

$$\frac{1}{3} + \frac{1}{2N} + \frac{1}{6N^2}$$

en dat daalt prachtig naar $\frac{1}{3}$, de breuk die we natuurlijk op basis van de moderne integraalrekening al verwacht hadden. En evenzo is de ondergrens voor A_2 gelijk aan (niet opnieuw uitrekenen maar gewoon $\frac{1}{N}$ eraf[10])

$$\frac{1}{3} - \frac{1}{2N} + \frac{1}{6N^2},$$

dat juist prachtig naar $\frac{1}{3}$ stijgt[11].

Maar ja, alles staat of valt dus met (6.2). We doen nu een keertje voor hoe je een uitspraak als (6.2) voor alle N kunt bewijzen. Anders gezegd, we kijken gewoon of (6.2) waar is voor alle $N \geq 1$. Eerst voor de eerste N waarvoor je de uitspraak wil weten, in dit geval $N = 1$. Dat is meestal niet moeilijk, en hier al helemaal niet, want we hadden het derdegraadspolynoom juist geraden door naar de eerste paar waarden van N te kijken.

De truc is nu om te laten zien dat als de uitspraak juist is voor een vaste maar willekeurig[12] gekozen $N \geq 1$, dat dan ook de uitspraak voor $N + 1$ volgt. Deze implicatie heet de inductiestap. Omdat de uitspraak waar is voor $N = 1$, is dat dankzij (steeds weer) dezelfde inductiestap ook zo voor $N = 2$, $N = 3$, $N = 4, \ldots$ Dit heet het principe van volledige inductie[13]. Als je elke uitspraak ziet als een dominosteentje dat je omgooit als op het steentje een juiste uitspraak staat, dan gooi je als het ware met het eerste steentje in

[8] Hoe?
[9] Maar goed raden is een kunst.
[10] Spreek uit: bovengrens min $\frac{1}{N}$.
[11] Ga maar na. Of je de monotonie echt nodig hebt is een goede vraag.
[12] Wat willekeurig ook is...
[13] De naamgeving is helaas een poging tot volledige intimidatie.

één keer de hele oneindige rij om. Als ergens in de rij de inductiestap fout gaat dan houdt het op.

Hoe ziet de inductiestap er hier uit? Wel, je kijkt wat je wil hebben voor $N + 1$ en herleidt dat tot wat je al weet voor N als alle steentjes daarvoor omgevallen zijn. In het geval van (6.2) kom je dan tot de conclusie dat

$$\begin{aligned} 1^2 + 2^2 + \cdots + N^2 + (N+1)^2 &= (1^2 + 2^2 + \cdots + N^2) + (N+1)^2 \\ &= (\frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}) + (N+1)^2 \\ &= \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6} + N^2 + 2N + 1. \end{aligned}$$

Even apart uitrekenen wat

$$\frac{(N+1)^3}{3} + \frac{(N+1)^2}{2} + \frac{N+1}{6}$$

geeft als je alles uitvermenigvuldigt, en je ziet dat we gekregen hebben wat we moesten hebben, en dat met het N-de steentje ook het N plus eerste steentje omvalt. Ad infinitum.

We zijn er dus uit voor $n = 2$. De prachtige identiteit in (6.2) geldt voor alle N en bijgevolg is $A_2 = \frac{1}{3}$. Hoe zit het nu met A_3, $A_4, \ldots$? Gaat het daarvoor net zo? Het antwoord is ja, en dit doen we aan de hand van een aantal opgaven waarvan we denken dat die gezien het bovenstaande goed te doen zullen zijn. Na $A_1 = \frac{1}{2}$ en $A_2 = \frac{1}{3}$ is raden wat eruit moet komen weer makkelijk. Wat wel aardig en nuttig is om op te merken is dat we van de derdegraads in het rechterlid van (6.2) alleen de kopcoëfficiënt nodig hadden voor het eindantwoord.

Opgave 6.3. Laat zien dat

$$\frac{1^3 + 2^3 + \cdots + (N-1)^3}{N^4} < A_3 < \frac{1^3 + 2^3 + \cdots + N^3}{N^4},$$

waarin het verschil tussen ondergrens en bovengrens gelijk is aan $\frac{1}{N}$, en dus willekeurig klein wordt als N groot wordt.

Opgave 6.4. Laat inductief zien dat

$$1^3 + 2^3 + \cdots + N^3 = \frac{N^4}{4} + \frac{N^3}{2} + \frac{N^2}{4}$$

en leg uit waarom hieruit volgt dat $A_3 = \frac{1}{4}$.

Opgave 6.5. Laat zien dat

$$1^4 + 2^4 + \cdots + N^4 = \frac{N^5}{5} + \frac{N^4}{2} + \frac{N^3}{3} - \frac{N}{30}.$$

Leg uit waarom $A_4 = \frac{1}{5}$.

Zo is de regelmaat wel duidelijk, toch? Algemener geldt kennelijk voor gehele positieve n dat

$$1^n + 2^n + \cdots + N^n = \frac{N^{n+1}}{n+1} + \cdots = P_n(N),$$

een polynoom van graad $n+1$ in N, met op de puntjes lagere-orde-termen in N. Over die lagere-orde-termen is veel interessants[14] te vertellen, maar voor de oppervlakte A_n heb je dat niet nodig. Aangenomen dat onze algemene conclusie juist is, volgt dat

$$A_n = \frac{1}{n+1},$$

zoals we aan het begin van dit hoofdstuk al zagen, maar toen met moderne integraalrekening die we nu helemaal niet hebben gebruikt. Eigenlijk veel leuker zo[15].

Over twee zaken kunnen we ons nog zorgen maken. Stijgen de ondersommen en dalen de bovensommen ook echt wel? Wel, misschien niet, maar wel vanaf zekere N. Je kunt jezelf ervan overtuigen dat het de tweede coëfficiënt in het polynoom is die bepaalt hoe het uiteindelijk zit met de monotonie. Of die tweede coëfficiënt ooit nul kan zijn? Eerst raden en dan bewijzen. Het verschil tussen ondersom en bovenom is steeds $\frac{1}{N}$ dus allebei nul kan niet. Maar één misschien wel? Niet? Wel? Wat dan? Hoef je alles wel in zoveel detail te weten? Wat je ook kunt doen is de hele monotonievraag vergeten en simpelweg twee getallen invoeren, te weten $\underline{A}_n$ als de kleinste bovengrens voor alle ondergrenzen, en $\overline{A}_n$ als de grootste ondergrens voor alle bovengrenzen, en vervolgens heel secuur beredeneren dat we zowel $\underline{A}_n \leq \overline{A}_n$ als $\overline{A}_n \leq \underline{A}_n$ moeten hebben, en dus ook $\underline{A}_n = \overline{A}_n = A_n = \frac{1}{n+1}$. Ook leuk.

De andere kwestie is of we inderdaad steeds de goede kopcoëfficiënt vinden. Dat is onderwerp voor een aparte sectie hieronder.

---

[14] Ook aan de schrijvers.

[15] De schrijvers zijn het hier met elkaar eens.

6.2 $1^n + 2^n + \cdots + N^n$

Die $\frac{1}{n+1}$ willen we dus echt zeker weten. Hoe is snel te zien dat voor elke n deze kopcoëfficiënt verschijnt? Dan gaat het dus eerst om de vraag waarom

$$1^n + 2^n + \cdots + N^n$$

een polynoom is in N van graad $n+1$, en daarna waarom dit polynoom van de vorm

$$P_n(N) = \frac{N^{n+1}}{n+1} + \cdots$$

is.

We willen dit weer doen met volledige inductie, dus met dezelfde dominostenentechniek als hierboven, en dan is de eerste vraag of we dat naar n of N moeten doen. Omdat we voor elke n een uitspraak doen over een polynoom in N, kiezen we nu voor n als inductieparameter en zoeken een manier om hogere machten van k, de somindex, om te zetten in lagere machten van k. Vervolgens nemen we aan dat de uitspraak al bewezen is voor $1, 2, \ldots, n-1$ en laten we zien dat de uitspraak dan ook waar is voor n[16]. Daarbij gebruiken we dan dat we de lagere machten al kunnen optellen tot polynomen in N, waarvan we de graad weten.

Voor de reductie van de exponenten gebruiken we het binomium van Newton in de vorm

$$(k+1)^{n+1} - k^{n+1} = (n+1)k^n + \frac{(n+1)n}{2}k^{n-1} + \cdots + (n+1)k + 1.$$

Sommeren we dit over k lopend van 1 tot en met N, dan zien we de sommen van machten verschijnen waar het ons om gaat:

$$\begin{aligned}
\sum_{k=1}^{N}\left((k+1)^{n+1} - k^{n+1}\right) &= \sum_{k=1}^{N}((n+1)k^n + \frac{(n+1)n}{2}k^{n-1} + \cdots + \\
&\quad + \cdots + (n+1)k + 1) \\
= \sum_{k=1}^{N}(n+1)k^n + \sum_{k=1}^{N}\frac{(n+1)n}{2}k^{n-1} &+ \cdots + \sum_{k=1}^{N}(n+1)k + \sum_{k=1}^{N}1 \\
= (n+1)\sum_{k=1}^{N}k^n + \frac{(n+1)n}{2}\sum_{k=1}^{N}k^{n-1} &+ \cdots + (n+1)\sum_{k=1}^{N}k + \sum_{k=1}^{N}1,
\end{aligned}$$

[16] Elk dominosteentje vergt een tikje van al zijn voorgangers. Dat zie je nooit op tv.

waarin we inderdaad de sommen zien staan van het type dat we beschouwen. De eerste, met exponent n, willen we bepalen. Met de voorfactor $(n+1)$ moet hier een polynoom in N van graad $n+1$ met kopcoëfficient 1 uitkomen. Is dat zo? Wel, gebruikmakend van de aanname dat voor exponenten $1, 2, \ldots, n-1$ al bekend is dat er polynomen uitkomen, die allemaal graad hoogstens n hebben, zijn we klaar als

$$\sum_{k=1}^{N} \left((k+1)^{n+1} - k^{n+1}\right)$$

een polynoom in N is van graad $n+1$ met kopcoëfficient 1, en dat blijkt inderdaad het geval. Immers, deze zogenaamde telescoopsom[17] is gelijk aan

$$\sum_{k=1}^{N} \left((k+1)^{n+1} - k^{n+1}\right) = (N+1)^{n+1} - 1 = N^{n+1} + \cdots .$$

Hiermee is het bewijs geleverd. Opgemerkt kan nog worden dat we in de inductiestap om de somformule voor exponent n te bewijzen alleen gebruiken dat voor lagere exponenten de somformule een polynoom is van graad hoogstens n, zonder verdere informatie over de coëfficienten, ook niet over de kopcoëfficienten. We hadden dus wel alle eerdere dominosteentjes nodig, maar niet alles wat er op die steentjes stond. In elke stap zien we vervolgens de juiste kopcoëfficient verschijnen[18].

6.3 Van sommen naar integralen

We hebben hierboven A_n uitgerekend zonder ons druk te maken over de vraag of de oppervlakte wel goed gedefinieerd was[19]. De vanzelfsprekende insluiting tussen onder- en bovensommen, waarvan het verschil zo klein genomen kan worden als we maar willen, lijkt voldoende overtuigend voor alle praktische doeleinden, en in dit expliciete voorbeeld volgde vervolgens ook meteen het antwoord op de vraag over de numerieke waarde van A_n. Twee vliegen in één klap. De eerste vlieg lukt voor elke monotone functie. Teken maar een plaatje van de grafiek van een positieve functie f op een interval $[a, b]$, neem een N zo groot als je wilt, teken rechthoekjes onder en boven de grafiek die in de hoeken net passen, allemaal met breedte $\frac{b-a}{N}$. Schuif je de bovenrechthoekjes een rechthoekje opzij, naar rechts als de functie stijgt, dan kiepert een rechthoekje

[17] Even nadenken over deze naamgeving.

[18] Altijd goed om na te gaan welke informatie je wel en niet hebt gebruikt.

[19] Vrees wel: dat maken we nog goed in het volgende hoofdstuk.

van het interval af. Aan de andere kant zet je er eentje bij, en je ziet makkelijk in dat het verschil tussen boven- en ondersom gelijk is aan

$$|f(b) - f(a)| \cdot \frac{b-a}{N},$$

en dat is zo klein als je maar wilt dankzij de N in de noemer.

De zo gedefinieerde oppervlakte noteren we met

$$\int_a^b f = \int_a^b f(x)\,dx,$$

en noemen we de integraal van f over $[a,b]$. Monotone functies zijn dus integreerbaar over begrensde intervallen, hoe lelijk ze verder ook zijn! We benaderen de integraal simpelweg met

$$h\left(f(a) + f(a+h) + f(a+2h) + \cdots + f(b)\right), \quad \text{met} \quad h = \frac{b-a}{N}.$$

Hoe groter N, hoe dichter de benadering bij de integraal zit. Zonder de eerste term is dit een ondersom of bovensom, en zonder de laatste term ook, maar dan andersom. En die ene term maakt in de limiet niets uit.

Ook continue functies zijn integreerbaar over begrensde intervallen, maar dat is veel lastiger om overtuigend te brengen, omdat de definitie van continuïteit nu eenmaal echt lastig is[20], in tegenstelling tot de voorwaarde voor integreerbaarheid zelf: functies zijn integreerbaar als je ondersommen en bovensommen net zo dicht bij elkaar kunt krijgen als je maar wilt[21]. In dat geval kan er maar een getal tussen onder- en bovensommen zitten en dat is dan per definitie de integraal van de functie over het interval.

Uit deze definitie volgt rechtstreeks dat als f en g op $[a,b]$ integreerbare functies zijn, dat ook de somfunctie $f+g$ integreerbaar is, met

$$\int_a^b (f(x) + g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

Ook veelvouden gaan goed. Als f integreerbaar is en λ een reëel getal, dan is

$$\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx.$$

Ten slotte volgt ook dat

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx,$$

[20] Zie voor dit begrip sectie 10.2.

[21] Het probleem hebben we nu eigenlijk weggedefinieerd.

als c tussen a en b ligt. Te lezen als: is de rechterkant goed, dan ook de linkerkant en omgekeerd. Er zijn overigens veel begrensde functies die niet integreerbaar zijn:

Opgave 6.6. Laat $f(x) = 2$ voor x rationaal en $f(x) = 1$ voor alle overige x. Is deze f integreerbaar op $[0, 1]$?

6.4 Integraalrekening voor polynomen

De functie $x \to x^n$ is over elk begrensd interval $[a, b]$ integreerbaar en we kunnen de integraal ook als getal bepalen. De volgende opgave is na het begrijpen van sectie 6.1 goed te doen.

Opgave 6.7. Verifieer, eerst voor $0 < a < b$, dat

$$\int_a^b x^n \, dx = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1}.$$

Met de regels in sectie 6.3 volgt nu dat alle polynomen inderdaad worden geïntegreerd zoals op school verteld: bij een polynoom $p(x)$ kies je een nieuw polynoom $P(x)$ met de eigenschap dat $P'(x) = p(x)$ en $P(0) = 0$. Dat kunnen we als we weten hoe we polynomen differentiëren. Voortbouwend op Opgave 6.7 is het niet moeilijk in te zien dat

$$\int_a^b p(x) \, dx = [P(x)]_a^b = P(b) - P(a).$$

Deze primitieve P van p is, dankzij de keuze $P(0) = 0$, uniek bepaald. Dit is de hoofdstelling van de integraalrekening voor polynomen.

Een veelvoorkomende integraal van dit type[22] is

$$\beta(n, m) = \int_0^1 x^{n-1}(1 - x)^{m-1} \, dx,$$

die op grond van wat we nu weten bestaat voor alle positieve gehele m en n. Uitrekenen met de hoofdstelling is tamelijk vervelend, en wij zijn net als veel wiskundigen van nature lui. We gaan dus niet iets uitrekenen als het

[22] Google eens naar bètafunctie.

niet hoeft, en denken liever eerst na voordat we stug gaan rekenen. Over dit probleem hebben andere mensen natuurlijk ook nagedacht en er is een mooie truc om $\beta(n,m)$ toch vrij gemakkelijk te bepalen. Dat begint met de volgende opgave.

Opgave 6.8. Laat zien dat

$$\beta(n+1,m)+\beta(n,m+1)=\beta(n,m)=\beta(m,n).$$

Omdat we weten dat $\beta(n,1)=\frac{1}{n}$ voor elke n kun je zo in het (n,m) vlak op de roosterpunten in het eerste kwadrant de waarden van $\beta(n,m)$ eenvoudig invullen. Een soort driehoek van Pascal op zijn kant, die de andere kant opgaat[23]. Ook in deze driehoek is de numerieke waarde van elke $\beta(n,m)$ makkelijk te bepalen, maar dat is natuurlijk nog iets anders dan een formule ervoor vinden. Maar met een beetje ervaring en een beetje geluk kun je uit Opgave 6.8 wel inzien dat er een formule voor $\beta(n,m)$ is. We helpen de lezer een beetje, en geven het antwoord, maar laten het aan de lezer over om te controleren dat dit antwoord inderdaad aan de betrekking in Opgave 6.8 voldoet.

Opgave 6.9. Laat zien dat

$$\beta(n,m)=\frac{(n-1)!(m-1)!}{(n+m-1)!},$$

waarbij $n! = 1\times 2\times 3\times\cdots\times n$.

De integraal

$$\beta(n,m)=\int_0^1 x^{n-1}(1-x)^{m-1}\,dx=\frac{(n-1)!(m-1)!}{(n+m-1)!}$$

vertelt ons wat de oppervlakte is van

$$\{(x,y):\ 0\le y\le x^{n-1}(1-x)^{m-1},\ 0\le x\le 1\}.$$

Vooruitlopend op hoofdstuk 8 veranderen we de schaal op de x-as door de eenheden met een factor m te schalen. Waar 1 stond zetten we m en daarna laten we m groot worden:

[23] Naar de top toe in plaats van ervanaf.

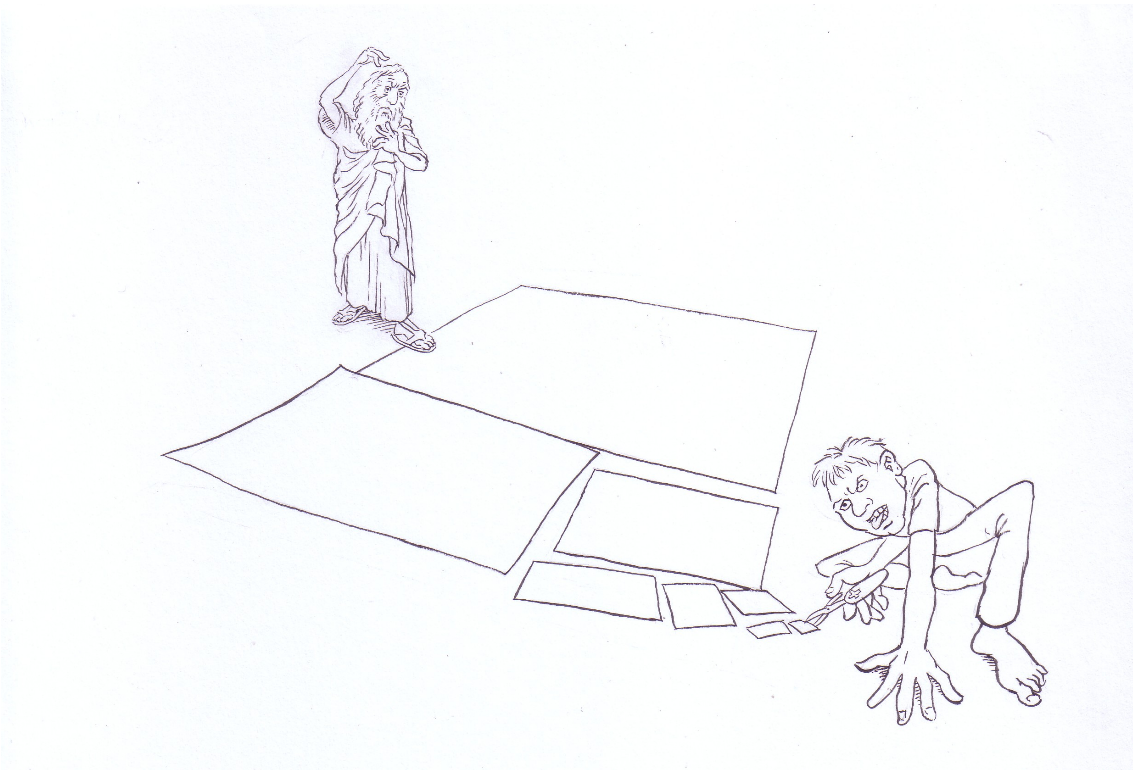
Opgave 6.10. Laat zien dat

$$\int_0^m x^n(1-\frac{x}{m})^m\,dx = \frac{m^{n+1}\,m!\,n!}{(m+n+1)!}.$$

Een interessante formule, waar we even iets nauwkeuriger naar willen kijken. Als m groter wordt dan gaat de bovengrens in de integraal naar ∞. Bovendien, zoals we inmiddels wel weten, gaat de term $(1-\frac{x}{m})^m$ in de integraal naar e^{-x}. Het zal dus geen verbazing moeten wekken dat als m groter wordt, het linkerlid convergeert naar $\int_0^\infty x^n e^{-x}dx$. Als je goed kijkt zie je dat voor $m \to \infty$ in het rechterlid precies $n!$ overblijft. Zo te zien zou dus moeten gelden dat

$$n! = \int_0^\infty x^n e^{-x}dx, \tag{6.3}$$

en dat willen we zeker weten. Hoog tijd voor wat oppervlakteberekening voor onbegrensde gebieden die niet per se mooi zijn. Een intermezzo. En daarna gaan we aan de slag met (6.3) om te zien of we $n!$ kunnen uitrekenen en hoe groot $n!$ eigenlijk is.



7 Oppervlakterekening zo moeilijk mogelijk?

In hoofdstuk 6 hebben we integraalrekening gezien als de techniek waarmee je oppervlakten uitrekent van niet al te rare gebieden in het xy-vlak, bijvoorbeeld de oppervlakte van het gebied A ingesloten door de grafiek $y = f(x)$ van een positieve functie f, twee x-waarden $a < b$ en de x-as. Voor monotone niet-negatieve functies hebben we gezien dat

$$\int_a^b f(x)dx$$

zo altijd een getal definieert dat we met recht associëren met de oppervlakte van A. Daarbij maakt het niet uit hoe vreemd we f kiezen[1]. Als f maar monotoon is.

Uitgangspunt was dat over de oppervlakte van een rechthoek in het xy-vlak geen twijfel kan bestaan: vier hoekpunten $(0,0),(B,0),(0,H),(B,H)$

[1] Google tip: Devil's Staircase.

met B en H allebei positief[2] maken een rechthoek met oppervlakte BH, breedte keer hoogte. De fysische eenheid van oppervlakte is L in het kwadraat, waarbij L de fysische eenheid van lengte is, maar in het xy-vlak is het de gewoonte om die eenheid niet te specificeren. De hoekpunten $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$ maken dus een vierkant met oppervlakte gelijk aan 1 en niet 1 vierkante el, om maar een eenheid te noemen. Op de coördinaatassen staan immers ook geen eenheden.

Het lag in hoofdstuk 6 voor de hand om met rechthoeken te beginnen. Als de grafiek van f in ligt tussen twee staafdiagrammen dan kan de oppervlakte als numerieke waarde alleen maar een getal zijn dat tussen onder- en bovensommen ligt die we uitrekenen met BH-tjes, waarbij (horizontaal) verschoven rechthoekjes (staafjes) natuurlijk dezelfde oppervlakte hebben als rechthoekjes die met een punt in de oorsprong staan. Bij verfijning van de staafdiagrammen bleef zo maar één kandidaat voor die numerieke waarde over en dat moest dan wel de gezochte oppervlakte van A zijn. Geen reden om daar verder moeilijk over te doen.

We werkten tot nu dus alleen maar met oneindig veel eindige sommen van getallen om oppervlakten uit te rekenen, met

$$\int_0^1 x^n dx = \frac{1}{n+1}$$

als eerste wapenfeit, en vervolgens een elegante theorie voor integralen van polynomen. Van BH rechtstreeks naar integraalrekening voor mooie functies

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

Het zal geen verbazing wekken dat wiskundigen zich afvragen of we niet meteen met sommen met oneindig veel termen kunnen werken. Een fundamentele vraag die vraagt om bezinning op waar we mee bezig zijn.

Aan welke axioma's of regels zou onze oppervlakterekening moeten voldoen? Van wat voor gebieden willen we de oppervlakte weten? Denk aan een stuk van een landkaart in een atlas, begrensd door natuurlijke of getrokken grenzen. Daarbij kun je het zo gek maken als je wilt[3]. De vragen die we daarom stellen zijn:

1. Heeft elke verzameling in het vlak een oppervlakte?
2. Hoe bereken je die oppervlakte (rechtstreeks of via een benadering)?

Het antwoord op deze vragen heeft natuurlijk ook te maken met wat we met oppervlakten en verzamelingen[4] willen kunnen doen. Wat willen we, wat

[2] Een beperking die vanwege de algebra niet nodig is, maar dat is een ander verhaal.
[3] Zoom maar eens in op de kust van Engeland.
[4] Dus niet alleen maar mooie gebieden.

kunnen we en wat denken we te kunnen? En wat zijn de gevolgen? In dit hoofdstuk laten we zien dat met deze vragen een ware doos van Pandora wordt geopend. Maar eerst nog een amusant stapje terug.

7.1 Van vierkanten naar rechthoeken

Ieder vierkant is een rechthoek, maar een rechthoek is bijna nooit een vierkant. Wat als om wat voor reden[5] dan ook de formule BH tot problemen leidt in de klas? Wel, als $H > B$ dan knippen we net zolang vierkantjes van de rechthoek (gezien als staaf) van boven af tot de hoogte kleiner is geworden dan B. Na n_1 keer knippen is de overgebleven nieuwe hoogte dan $h = h_1 = H - Bn_1 < B$, tenzij we een heel speciale staaf hadden en precies uitkwamen, maar dat gebeurt natuurlijk nooit in de knippraktijk.

Vervolgens knippen we van het overgebleven stuk van de staaf vanaf rechts vierkantjes af, net zolang tot we na m_1 keer knippen een kleinere breedte $b = b_1 = B - h_1 m_1 < h_1$ over houden. En zo knippen we vrolijk verder, van boven, van rechts, van boven, van rechts, $\ldots$, met $h_2 = h_1 - b_1 n_2 < b_1$, $b_2 = b_1 - h_2 m_2 < h_2$, enzovoorts. Met $b_0 = B$ en $h_0 = H$ is het zo evident dat

$$\begin{aligned} b_0 h_0 = n_1 b_0^2 + m_1 h_1^2 + n_2 b_1^2 + m_2 h_2^2 + n_3 b_2^2 + m_3 h_3^2 + \cdots = \\ \underbrace{b_0^2 + \cdots + b_0^2}_{n_1 \text{ vierkantjes}} + \underbrace{h_1^2 + \cdots + h_1^2}_{m_1 \text{ vierkantjes}} + \underbrace{b_1^2 + \cdots + b_1^2}_{n_2 \text{ vierkantjes}} + \cdots, \end{aligned}$$

een prachtige som van kwadraten, de oppervlakten van steeds kleinere vierkantjes. De som is het geheel der delen en over kniprandjes zeuren we niet. Die zien we niet, en als we ze wel zien dan hebben ze toch geen oppervlakte en dus zien we ze niet.

7.2 Heeft elke vlakke verzameling een oppervlakte?

Ons uitgangspunt is nu dus dat als we een gebied in stukken knippen de oppervlakten van de nieuwe stukken samen de oppervlakten van het gebied maken. Net zoals op een kaart in een atlas de oppervlakte van de Benelux gelijk is aan de oppervlakte van Nederland plus de oppervlakte van België plus de oppervlakte van Luxemburg. Een belangrijk punt is dat we dit ophakken in kleinere stukken willen en kunnen herhalen ad infinitum. De vierkantsoefening hierboven lijkt wellicht wat kunstmatig[6], maar voor minder kunnen en willen we het echt niet doen[7].

[5] Hangt wellicht van de leeftijdsgroep af.
[6] Maar Euclides zou het wel kunnen waarderen. Waarom?
[7] Bij welke gouden rechthoek is het knipwerk minimaal?

Immers, bij de formule voor $n!$ krijgen we te maken met zogenaamde oneigenlijke integralen van positieve functies. We willen dan natuurlijk dat

$$\int_0^\infty f(x)dx = \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^3 f(x)dx + \int_3^4 f(x)dx + \cdots ,$$

waarmee de term links betekenis krijgt via de individuele termen rechts die hun wiskundige betekenis al eerder hadden gekregen.

In beide voorbeelden, de rechthoek en het gebied tussen $y = 0$ en $y = f(x)$ in het rechterhalfvlak, is het zo dat we het hele gebied als in oneindig veel stukjes geknipt zien waarbij de kniprandjes bestaan uit lijnstukjes die allemaal bij elkaar ook een stuk van het oorspronkelijke gebied vormen. Of we dat stuk weer een gebied willen noemen is een kwestie van smaak en afspraak. Daarom spreken we hier liever van verzamelingen. Alle punten op de randstukjes van de oneindig veel rechthoekjes waarin we het vierkant verknippen vormen samen een verzameling. Ook aan deze wat rare verzameling willen we een oppervlakte toekennen en als je er even over nadenkt dan moet die oppervlakte wel nul zijn:

Opgave 7.1. Beargumenteer dat je door een willekeurig klein vierkantje in stukjes te knippen in beide voorbeelden de verzameling van alle kniprandjes helemaal kunt bedekken met rechthoekjes die samen hoogstens de oppervlakte van het kleine vierkantje hebben, en dat de oppervlakte van alle kniprandjes bij elkaar dus alleen maar 0 kan zijn.

De eerste vraag in de inleiding van dit hoofdstuk lezen we dus met in het achterhoofd de regel dat (de oppervlakte van) een geheel altijd de som der (oppervlakten van de) delen moet zijn, niet meer en niet minder[8]. En zoals gezegd, bij deze regel willen we ons niet hoeven te beperken tot eindige sommen. Ook de aftelbare versie van de regel moet betekenis hebben.

Het ligt wellicht voor de hand om te verwachten dat bij iedere deelverzameling A van punten in het vlak een getal $\mu(A)$ hoort, de maat van A, dat aangeeft wat de oppervlakte van A is. Dat getal moet dan ook 0 of ∞ kunnen zijn, want de oppervlakte van een punt moet wel nul zijn, en de oppervlakte van het hele vlak[9] is natuurlijk oneindig.

Rekenen met ∞ is een verwarrend onderwerp dat nauw verwant is met de vraag of je door 0 kunt delen, maar ook zonder oneindige oppervlakten is de

[8] Een deel kan dus ook nooit meer zijn dan het geheel.

[9] Is er een fysisch model waarin dat nog bestaat?

eerste vraag al fundamenteel. Het rekenen blijft daarbij beperkt tot het optellen van gewone getallen, inclusief 0, met vanzelfsprekende optelsommetjes als

$$0+0=0;\ 0+0+0=0;\ 0+0+0+0=0,$$

maar ook

$$0+0+0+\cdots=0.$$

Denk aan:

$$0=0,000000\cdots=0+\frac{0}{10}+\frac{0}{100}+\frac{0}{1000}+\frac{0}{10000}+\frac{0}{100000}+\frac{0}{1000000}+\cdots.$$

Om geen last te hebben van ∞ kijken we nu eerst naar oppervlakten van deelverzamelingen van het boloppervlak van een driedimensionale bol, zeg maar een globe van de aarde. In onze perceptie heeft op die globe de deelverzameling die Afrika heet een welgedefinieerde oppervlakte, en het ligt voor de hand om te verwachten dat dat voor elke deelverzameling A van het hele boloppervlak het geval moet zijn. We kiezen er voor de fysische eenheid van lengte zo te kiezen dat de totale oppervlakte van de globe gelijk is aan 1.

Laten we ons voorstellen dat een gekromde landkaart als een rigide vlies om de globe is aangebracht waarop we de continenten hebben geschilderd. Als we dat randje draaien over de bol, dan schuift het gebied A naar een andere plek en definieert zo een ander gebied op de bol dat we B kunnen noemen. Natuurlijk heeft B in onze perceptie dezelfde oppervlakte als A[10]. En als A en B geen overlap hebben dan is de gezamenlijke oppervlakte van A en B precies twee keer de oppervlakte van A.

Hoe we de werkelijkheid ook abstraheren, twee disjuncte deelverzamelingen op een boloppervlak die precies over elkaar heen geschoven kunnen worden hebben samen twee keer de oppervlakte van elk van de twee apart. Denk aan het noordelijk en het zuidelijk halfrond. Laat je niet in verwarring brengen door de gemeenschappelijke rand die evenaar heet, want die heeft oppervlakte 0: de optelsom is dus eigenlijk $\frac{1}{2}+\frac{1}{2}+0=1$. Ook als we niet twee, maar drie disjuncte nette gebieden hebben, zeg A, B en C, die door draaiingen uit elkaar verkregen zijn en die gezamenlijk het hele boloppervlak vullen, moet het geheel de som der delen zijn[11]. Omdat we de eenheden zo gekozen hebben dat het hele boloppervlak oppervlakte 1 heeft, moet elk van de verzamelingen A, B en C wel oppervlakte $\frac{1}{3}$ hebben, want $1=\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+0$, waarbij de nul weer voor de oppervlakte van de randen staat.

Het ongelooflijke nu is dat er een ogenschijnlijk volstrekt helder argument voor het bestaan van een wel heel bijzondere verzameling A is[12] waarvoor

[10] Tektonische effecten verwaarlozen we.

[11] Denk maar aan een meloen die in drie stukken wordt gesneden.

[12] Of die verzameling echt bestaat is de vraag. Daar komen we zo op terug.

we B en C als zojuist kunnen vinden, maar waarvoor ook nog *drie* andere verschoven versies E, F en G van A bestaan die ook allemaal disjunct zijn, en die met zijn *vieren* de bol bedekken. Rara hoe kan dat? Want hieruit volgt toch dat de oppervlakte van A tegelijkertijd $\frac{1}{3}$ en $\frac{1}{4}$ moet zijn, hetgeen wel een wiskundige crisis genoemd mag worden. Een mogelijke uitweg is dat de oppervlakte van A simpelweg niet bestaat. Een andere uitweg is dat A niet bestaat omdat je de regels waarmee we A gemaakt hebben niet accepteert. Als je gelooft dat A bestaat zeggen we dat A niet meetbaar is[13]. Deze fantastische paradox heet de Banach-Tarski paradox, genoemd naar twee Poolse wiskundigen die dit fenomeen beschreven hebben.

Het is onvermijdelijk dat de filosofie van de wiskunde nu aan de orde komt. Over wat wiskunde werkelijk is maken ook wiskundigen zich vaak liever niet zo druk, maar in dit geval komt niemand er onderuit. De zojuist besproken paradoxale verzameling A kan alleen maar gemaakt worden met behulp van het zogenaamde *keuzeaxioma*. Het keuzeaxioma zegt dat het mogelijk is om een verzameling te vormen door uit een overaftelbare collectie verschillende verzamelingen uit elke verzameling precies één element te kiezen. Het klinkt vanzelfsprekend, maar is het dat wel? Verbeelden we ons niet te veel? De enige manier om daarop antwoord te geven is spelen met die verbeelding en de consequenties onder ogen zien.

Kun je zeggen dat het keuzeaxioma waar is of onwaar? Natuurlijk kun je dat zeggen maar of je argumenten kunt aanvoeren is vraag twee. Die zijn er eigenlijk niet. Je kunt het axioma vanzelfsprekend vinden. Of niet. Je kunt het accepteren. Of niet. De keuze van Ronald[14] leidt tot een andere wiskunde dan die van Joost[15]. Ronald accepteert dat A wel bestaat maar geen oppervlakte heeft, terwijl Joost de paradoxale en niet-meetbare verzameling A meer als een bijzonder amusant hersenspinsel ziet.

De constructie van Banach en Tarski is ingewikkeld, en de paradox extreem. Er is echter een veel simpelere constructie van een veel minder paradoxale verzameling A, die tot een vergelijkbare conclusie leidt wat betreft meetbaarheid. De A die we aanstonds maken is geen deelverzameling van een boloppervlak maar van een cirkel, en kan dus ook weer gedraaid worden. Het bijzondere is nu dat er met deze A oneindig maar wel aftelbaar veel andere A's gemaakt kunnen worden, allemaal disjunct, allemaal hetzelfde in de zin dat ze via draaiing uit elkaar worden verkregen, zo dat met al deze A's de hele cirkel wordt bedekt.

Waar door Banach en Tarski in een boloppervlak een A werd gemaakt

[13] Een uitspraak die hier betrekking heeft op de (bol)oppervlaktemaat.
[14] Wel keuzeaxioma.
[15] Geen keuzeaxioma.

die niet meetbaar kan zijn omdat zijn boloppervlakte a zou moeten voldoen aan $3a = 4a = 1$, is het nu zo dat we een A in de cirkel hebben die met aftelbaar oneindig veel gedraaide versies van zichzelf de hele cirkel vult, zonder onderlinge overlap: de cirkel is zo dus de vereniging van verzamelingen

$$A = A_0, A_1, A_2, A_3, A_4, \ldots$$

en wel zo dat elk punt van de cirkel in precies één verzameling A_n uit deze rij ligt. We laten nu eerst zien hoe dat in zijn werk gaat.

Neem een cirkel en beschrijf de punten op de cirkel met een hoek die je in graden schrijft als α keer 360 graden[16]. Die α is op veelvouden van 1 na uniek bepaald. We noemen twee punten met bijbehorende α en β equivalent als $\alpha - \beta$ een breuk is (een rationaal getal). De cirkel wordt hiermee onderverdeeld in equivalentieklassen: zo'n klasse[17] krijg je door een punt op de cirkel te kiezen en dat punt met alle punten die equivalent zijn met dat punt in één klasse te stoppen. Elke klasse heeft dan aftelbaar oneindig veel punten. De cirkel is zo onderverdeeld in equivalentieklassen, en dat moeten er wel overaftelbaar oneindig veel zijn!

Kies[18] nu uit elke klasse precies één punt om met die punten een deelverzameling A van de cirkel vormen. Deze A kan gedraaid worden over alle mogelijk rationale veelvouden van 360 graden. Dat correspondeert met aftelbaar veel α's tussen 0 en 1. Al deze gedraaide A's zijn onderling disjunct[19] en vormen samen de hele cirkel.

De natuurlijke maat voor deelverzamelingen van de cirkel is booglengte. Kan A een booglengte a hebben? Zo ja, dan volgt dat $A = A_0$ en A_1 samen maat $2a$ hebben, en $A = A_0$, A_1 en A_2 samen $3a$, enzovoorts. De totale lengte van de cirkel is de omtrek, en die is eindig en gelijk aan 1 als we de eenheid van booglengte handig kiezen. Maar dan volgt dat $a < 1$, $2a < 1$, $3a < 1$, $\ldots$, en dat gaat alleen maar met $a = 0$. Omdat de totale booglengte van de cirkel gelijk is aan 1 leidt dit tot de paradox dat

$$0 + 0 + 0 + \cdots = 1,$$

tenminste, als we de lengtes ook bij dit soort aftelbaar oneindige partities op mogen tellen. We concluderen dat we op de cirkel vanwege het bestaan van A geen booglengtemaat kunnen hebben die rotatie-invariant is en voldoet aan de eigenschap dat de maat van een aftelbare disjuncte vereniging altijd gelijk is aan de som van individuele maten.

[16] Of welke andere hoekmeting je ook wilt gebruiken.
[17] Gewoon een verzameling.
[18] Ga er maar aan staan, overaftelbaar veel keer.
[19] Wel even nagaan waarom. Hint: bewijs dit uit het ongerijmde.

Met dit voorbeeld krijgen we nu overigens ook een verzameling die geen oppervlakte kan hebben. Elk punt op de cirkel correspondeert immers met een lijnstuk[20] dat het middelpunt van de cirkel met dat punt verbindt. Alle lijnstukken bij de punten in A vormen een deel van de platte schijf dat we $\tilde{A}$ noemen en we kunnen ons weer afvragen of $\tilde{A}$ een oppervlakte heeft. Omdat alle gedraaide versies van $\tilde{A}$ samen weer de hele disk opvullen[21] worden we geconfronteerd met hetzelfde probleem. Als $\tilde{A}$ een oppervlakte a zou hebben dan moet onvermijdelijk weer gelden dat $a + a + a + \cdots = \frac{1}{4\pi}$, omdat een cirkel met omtrek 1 oppervlakte $\frac{1}{4\pi}$ heeft[22]. Dat kunnen we in onze wiskunde met geen enkele a voor elkaar krijgen[23].

De constructie van de A van Banach en Tarski is enigszins vergelijkbaar met de constructie van A op de cirkel, maar veel ingewikkelder. Joost haalt wat betreft beide A's de schouders op, omdat je nu eenmaal niet overaftelbaar veel keuzes kunt maken; Ronald ziet dat anders, en accepteert dat sommige dingen niet te meten zijn. Over veel intuïtie voor oppervlakten zijn ze het wel eens. Een lijn bijvoorbeeld heeft oppervlakte 0, hoe lang[24] hij ook is. Ook de grafiek van een (voldoende nette) functie zelf (dus niet het gebied eronder) heeft oppervlakte 0.

7.3 Meer bizarre wiskunde

Het keuzeaxioma zegt dus dat het mogelijk is om een verzameling te vormen door uit een overaftelbare collectie verschillende verzamelingen uit elke verzameling precies één element te kiezen. Een onuitputtelijke bron van vreemde noties is de verzameling van alle rijen van nullen en enen. Je kunt daarbij denken aan tweezijdig bedrukte bordjes wel of niet aankloppen die de gasten aan de buitenkant van hun deur hebben gehangen voor de ochtendschoonmaakploeg in Hilberts hotel. We denken aan de problemen die ontstaan als alle gasten 's ochtends vergeten zijn hoe ze hun bordje hebben gehangen, maar wel door hun raam oneindig ver de oneindige gang in kunnen kijken en zien hoe de bordjes van de hogere kamernummers hangen. Hun eigen bordje en de bordjes van de lagere kamernummers kunnen de hotelgasten echter niet zien. Je kunt het je voorstellen, toch?

Meteen schoonmaken representeren we met een 1 en met rust gelaten worden met een 0. Voor de schoonmaakploeg is het prettig als 's ochtends alle gasten de juiste keuze maken, vroeg opstaan of blijven liggen met de deur

[20] We kunnen het middelpunt zelf eventueel weglaten.

[21] Samen met het middelpunt dat oppervlakte nul heeft.

[22] Wat we overigens nog niet echt weten trouwens.

[23] Wat deze a is wil je, maar kun je niet weten.

[24] Google tip: space filling curves...

op slot. Alle gasten worden 's ochtends tegelijkertijd even wakker en kijken op een briefje onder hun kussen wat ze moeten doen nu ze vergeten zijn hoe hun bordje hangt. Alle briefjes zijn hetzelfde. De gasten zijn namelijk toevallig allemaal wiskundige en de avond tevoren hebben ze gezamenlijk over dit probleem nagedacht. En daar zijn ze uitgekomen!

Op elk briefje staat nu hoe ze de informatie die ze kunnen vergaren, te weten, hoe de bordjes bij de hogere nummers hangen, kunnen gebruiken om het hotelpersoneel zo min mogelijk tot last te zijn, door goed te raden hoe het bordje aan de buitenkant van hun eigen kamerdeur hangt, en daar naar te handelen. Zo min mogelijk tot last zijn is wat de gasten betreft dat iedereen goed raadt, op een paar[25] gasten na. Het wiskundige raadsel is nu: wat staat er op de briefjes?

Wel, de gasten hebben nagedacht over rijen met nullen en enen en bedacht dat ze twee zulke rijen equivalent kunnen noemen als de rijen op eindig veel rangnummers na hetzelfde zijn. En daarmee hoort bij elke rij een equivalentieklasse van alle rijen die met deze rij equivalent zijn. Zo wordt de verzameling van rijen nullen en enen ingedeeld in equivalentieklassen en uit al die equivalentieklassen hebben de gasten gezamenlijk een vertegenwoordigend rijtje gekozen (dat kon vanwege het keuzeaxioma) en die keuze op het briefje vastgelegd. Dat was wat werk, maar de moeite waard, want 's ochtends kijkt iedereen even via het raam naar de bordjes op de hogere kamernummers en stelt vast in welke equivalentieklasse het hele rijtje met nullen en enen zit. De lagere kamersnummers zijn daarvoor niet nodig.

Op het briefje staat nu welke rijtje uit die klasse uitverkoren was. Elke gast handelt nu naar de nul of een die in dat uitverkoren rijtje bij zijn kamernummer staat. Op een paar na doen ze dus allemaal wat ze vergeten waren, en het ongenoegen bij de schoonmaakploeg blijft zo beperkt als mogelijk[26].

Bizar? Hier is er nog eentje, nu met maar drie hotelgasten, zeg Adri, Jan en Marja, die spelletjes spelen bij Marja op de kamer. Eén spelletje hebben ze voor het laatst bewaard. Het leukste spelletje waarbij ieder van de drie uit twee rollen moet kiezen. Voor de zekerheid hebben ze die keuze op hun voorhoofd geschreven. Maar na alle andere spelletjes zijn ze vergeten wat ze gekozen hebben en alledrie proberen ze aan het eind van de avond te raden welke rol ze gekozen hadden, en daarnaar te handelen. Aan elkaar vragen wat op het eigen voorhoofd staat kan niet meer. Ten minste één iemand moet handelen, maar niet noodzakelijk allemaal. Maar als iemand verkeerd handelt leidt dat tot hommeles. Ook deze drie hebben voor het kiezen van hun rol in het laatste spelletje nagedacht over een gemeenschappelijk strategie bij het

[25] Een paar is twee, drie, vier, ... maar niet oneindig.

[26] Hoe is het mogelijk?

begin van het laatste spel. Ze kennen elkaar, en het keuzeaxioma hebben ze niet nodig. De strategie moet leiden tot tevredenheid zonder hommeles. Dat wil zeggen, iedereen die handelt moet juist handelen. Als bijvoorbeeld van te voren is afgesproken dat alleen Marja handelt, en Marja handelt door random een rol te kiezen, dan is de succeskans $\frac{1}{2}$. Het verrassende is dat het veel beter kan:

Opgave 7.2. Laat zien dat er een strategie bestaat die een winstkans van $\frac{3}{4}$ heeft. Reflecteer hierop.

7.4 Terug naar de werkelijkheid

De theorie die alles rond oppervlakten wiskundig precies maakt, heet maattheorie, een prachtig maar vaak als moeilijk ervaren vak. Integreren is in dat vak eigenlijk niets anders dan het bepalen van een oppervlakte van een verzameling. We hebben gezien bij het uitrekenen van de oppervlakte onder de grafiek van een polynoom dat we de oppervlakte onder zo'n grafiek benaderen door steeds meer rechthoekjes te nemen, die beter en beter gaan passen in het gebied waarin we zijn geïnteresseerd. Helaas, het blijkt zo te zijn dat deze methode voor veel te weinig verzamelingen werkt.

In de meeste verzamelingen[27] kun je helemaal geen rechthoekjes leggen. Maar je kunt elke verzameling wel met rechthoekjes bedekken. Als je aan een A een 'oppervlakte' of 'maat' $\mu(A)$ wil toekennen, kun je proberen dat te doen door te zeggen dat elke aftelbare bedekking van A met rechthoekjes een totale oppervlakte groter dan $\mu(A)$ moet hebben. Je probeert A dus zo zuinig mogelijk met rechthoekjes te bedekken. Iedere begrensde A heeft dan een grootste waarde $\mu(A)$ waar je niet onder kunt komen en deze $\mu(A)$ zou moreel gesproken de oppervlakte van A moeten zijn.

Heeft deze $A \to \mu(A)$ toekenning de eigenschappen die we willen hebben? Hoe zit het met de maat van de vereniging van twee verzamelingen A en B? Als A en B allebei begrensd zijn dan kun je voor zowel A als B onder respectievelijk $\mu(A)+\frac{1}{n}$ en $\mu(B)+\frac{1}{n}$ komen, voor elke n die je maar wilt. Door bedekkingen van A en B te combineren kom je voor de vereniging $A \cup B$ dus altijd onder $\mu(A)+\frac{1}{n}+\mu(B)+\frac{1}{n}$. Maar dan is $\mu(A \cup B) < \mu(A)+\mu(B)+\frac{2}{n}$, met n zo groot als je wil. Kortom, er geldt altijd dat

$$\mu(A \cup B) \leq \mu(A) + \mu(B),$$

[27] Bijvoorbeeld in het eenheidsvierkant zonder de rationale punten.

met naar je verwacht een strikte ongelijkheid alleen als A en B te veel overlap hebben. Met andere woorden, je zou verwachten dat $\mu(A \cap B) = \mu(A) + \mu(B)$ als A en B geen punten gemeen hebben, dat wil zeggen als $A \cap B$ de lege verzameling is. Helaas, aan de A van Banach-Tarski zien we dat dat kennelijk niet altijd waar kan zijn, als die A echt bestaat.

Kortom, de maat μ die we zo gedefinieerd hebben voldoet niet aan onze wensen. Het geheel kan minder[28] zijn dan de som der delen en zulke delen willen we niet. Welke delen willen we dan wel? Het antwoord is simpel: precies die verzamelingen A die we als mal kunnen gebruiken om elke andere verzameling B in twee stukken te verdelen waarbij het geheel niet minder is dan de som der delen. Die twee stukken zijn alles in B dat wel in A zit en alles in B dat niet in A zit, respectievelijke genoteerd als $B \cap A$ en $B - A$. Anders gezegd, de som der delen kan nooit meer[29] zijn dan het geheel als we A gebruiken om een verzameling B in twee stukken te verdelen. We noemen zulke A's meetbaar.

Als we alleen dit soort A's gebruiken dan gelden alle regels die we zo graag wilden hebben. Bij ieder college maattheorie wordt bewezen dat de meeste verzamelingen die je zonder overaftelbare operaties kunt maken, inderdaad de eigenschap hebben dat ze niet gebruikt kunnen worden als mal om van oppervlakte meer oppervlakte te maken dan waar je mee begonnen bent. In het bijzonder kan voor $a < b$ de integraal $\int_a^b f(x)dx$ van een positieve (of liever een niet-negatieve) functie f rechtstreeks als oppervlakte gedefinieerd worden van het gebied tussen de grafiek $y = f(x)$, de horizontale as $y = 0$ en de grenzen $x = a$ en $x = b$ van de integraal, als dat gebied maar meetbaar is. En functies waarvoor dat het geval is noemen we natuurlijk meetbaar. Als de maat van het gebied eindig is dan heet f integreerbaar op het interval (a, b). Voor functies met tekenwisselingen kun je zelf wel een definitie verzinnen.

Als al het werk gedaan is dan volgen er een paar krachtige stellingen die zeggen dat alle uitspraken die je zou verwachten waar zijn. We formuleren er een paar hieronder, en bij elke uitspraak kun je jezelf afvragen of je die uitspraak ook had voor gewone integralen, gemaakt met boven- en ondersommen. Toepassing van deze convergentiestellingen komen aan de orde in hoofdstuk 8.

De monotone convergentiestelling 1. Voor niet-negatieve integreerbare functies f_1, f_2, $f_3, \ldots$ die puntsgewijs stijgen naar een limietfunctie[30] f geldt

[28] Een beetje als het vak Algemene Natuurwetenschappen.
[29] Denk aan de wonderbaarlijke broodvermenigvuldiging.
[30] $f_n(x) \uparrow f(x)$ voor elke x.

dat

$$\int_a^b f_n(x) \to \int_a^b f(x)dx.$$

Dus de oppervlakten onder de grafieken stijgen naar de oppervlakte onder de grafiek van f. De puntsgewijze convergentie hoeft niet eens voor alle x te gelden, als de eendimensionale maat van alle punten waar het misgaat, maar niet positief is.

Een soortgelijke uitspraak geldt als de linkerintegratiegrens daalt met n, dus als $a = a_n$ met a_n een dalende rij. Essentieel is dat we het hebben over integralen die de oppervlakten van steeds groter wordende gebieden beschrijven. Het geval dat $a_n \to -\infty$ is daarom nadrukkelijk toegestaan, maar je krijgt dan wel een oneigenlijke[31] integraal, met linkergrens $-\infty$ in de limiet. Of ∞ als rechtergrens in het geval dat de rechtergrenzen b_n stijgen met $b_n \to \infty$. De integraalformule (6.3) voor $n!$ volgt zo via de intuïtief aannemelijke limietovergang in Opgave 6.10. En die limietovergang kan dus met maattheorie precies gemaakt worden. Gelukkig maar.

Je zou kunnen zeggen dat de monotone convergentiestelling al gebruikt wordt bij het bepalen van de integraal van x^n, ons eerste wapenfeit bij integreren. Hoe intuïtief duidelijk de stelling ook moge zijn, de uitspraak is *niet* langer waar is als we de monotonie niet aannemen. Zelfs als de functies f_n *dalen* naar f kan het fout gaan met de limiet. Een illustratief eenvoudig en flauw voorbeeld wordt gegeven door $f_n(x) = \frac{1}{n}$ voor alle x te nemen. De grafiek van deze functie is de horizontale rechte lijn $y = \frac{1}{n}$. De oppervlakte tussen deze grafiek en de x-as moet wel oneindig zijn, voor alle n, want je kunt er een rechthoek met hoogte $\frac{1}{2n}$ en breedte $2n^2$ tussen leggen. De limietfunctie is de functie gegeven door $f(x) = 0$ voor alle x. De grafiek van deze functie is precies de x-as, en de oppervlakte onder deze grafiek kan niet anders dan 0 zijn.

De oppervlakten dalen dus niet mee naar nul, ze waren immers allemaal oneindig. Beetje flauw inderdaad. Door te eisen dat de oppervlakte onder elk van de f_n's eindig is worden dit soort flauwiteiten uitgesloten door middel van een monotone convergentiestelling voor dalende functies:

De monotone convergentiestelling 2. Voor niet-negatieve integreerbare functies $f_1, f_2, f_3, \ldots$ die puntsgewijs dalen naar een limietfunctie f geldt dat

$$\int_a^b f_n(x) \to \int_a^b f(x)dx,$$

als $\int_a^b f_n(x)dx$ eindig is vanaf zekere n. Zowel $a = -\infty$ als $b = \infty$ zijn hier toegestaan.

[31] Eindige bovensommen zijn dan niet meer bruikbaar in de definities.

Je kunt natuurlijk ook denken aan situaties waarbij de positieve functies $f_1, f_2, f_3, \ldots$ helemaal niet monotoon zijn in n, en 'alleen maar' convergeren naar een limietfunctie f. Convergeren de oppervlakten onder de grafiek nu altijd mee als we eisen dat de oppervlakte onder elke van de grafieken van de f_n eindig is? Helaas, dat kan niet waar zijn. Oppervlakte kan weglopen naar oneindig en dan zie je die oppervlakte niet meer in de limiet. Schuif je favoriete grafiek maar op.

Nog een instructief voorbeeld: neem f_n gegeven door $f_n(x) = n$ voor $x \in (0, \frac{1}{n})$, en $f_n(x) = 0$ voor alle andere x. De oppervlakte onder de grafiek van f_n is $\frac{1}{n} \times n = 1$ voor alle n, maar $f_n(x)$ convergeert naar 0 voor alle x en $1 \to 0$ als $n \to \infty$ is ongerijmd. Om oppervlakten onder grafieken mee te laten lopen naar de limietoppervlakte is dus een extra voorwaarde nodig, bijvoorbeeld:

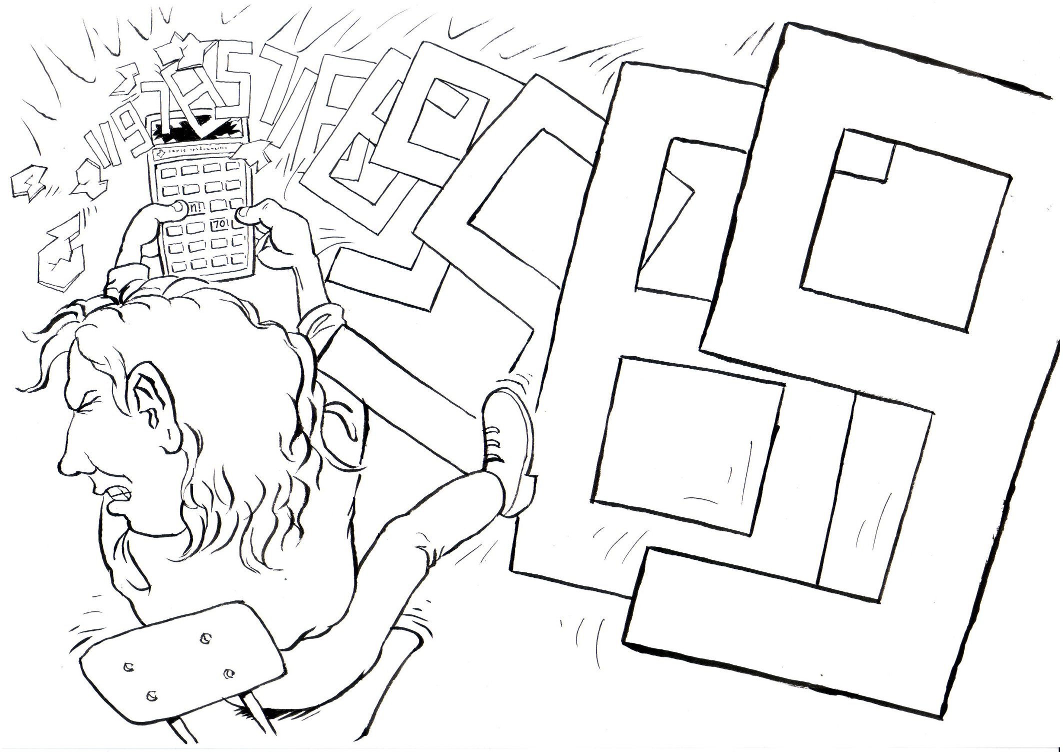
De gedomineerde convergentiestelling. Als $f_1, f_2, f_3, \ldots$ op (a, b) puntsgewijs convergeren naar een limietfunctie f en er een vaste functie g is zodanig dat $f_n(x) \leq g(x)$ voor alle x met $\int_a^b g(x)dx$ eindig, dan geldt dat

$$\int_a^b f_n(x) \to \int_a^b f(x)dx.$$

Ook dit geldt weer in de 'oneigenlijke' gevallen $a = -\infty$ en/of $b = \infty$.

Het is te veel werk om deze stellingen in dit boekje in detail te bewijzen, ook omdat ze bewezen worden zonder enige aanname op de limietfunctie zelf, behalve dat die functie bijna overal de puntsgewijze limiet is van de integreerbare functies f_n. De essentie voor de bewijzen is dat in de klasse van meetbare verzamelingen alle uitspraken over maten van verzamelingen die je krijgt met aftelbare operaties en limieten van de formules, waar zijn. Als je overlap tussen verzamelingen weg wil strepen om geen bijdragen dubbel te tellen moet je een beetje uitkijken dat je geen oneindig van oneindig aftrekt, maar binnen verzamelingen met eindige maat in het xy-vlak kan er weinig fout gaan.

In het bijzonder werkt de theorie ook voor integralen van functies als in Opgave 6.6. We kunnen nu dus veilig verder met de vraag hoe groot $n!$ nu eigenlijk is als n groot is, en je kunt jezelf bij elke stap afvragen hoeveel van de theorie je echt nodig vindt om overtuigd te worden van de conclusies die gaan komen.



8 Hoe groot is n! vraagteken

We hebben al gezien dat de faculteit van een positief geheel getal n per definitie gelijk is aan het product van alle positieve gehele getallen tot en met n. Dus

$$1! = 1, \quad 2! = 1 \times 2 = 2, \quad 3! = 1 \times 2 \times 3 = 6, \quad 4! = 1 \times 2 \times 3 \times 4 = 24,$$

enzovoorts. De eerste factor 1 doet niets, maar wordt er meestal bijgeschreven, bijvoorbeeld omdat je in binomiaalcoëfficiënten als

$$\binom{7}{3} = \frac{7!}{3! \times 4!} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1}$$

dan net zoveel factoren in de teller als in de noemer hebt staat.

De vraag hoe snel dat groot wordt is er een die tot de verbeelding spreekt. De laatste rekenmachine van Joost[1] kon bij benadering $n!$ tot en met $n = 69$

[1] Aangeschaft voor het natuurkundepracticum in zijn eerste jaar in Leiden.

aan. Dat zat nog onder 10 tot de macht 100, een 1 met honderd nullen[2]. Groter kon niet en dus was 70! al te groot want 70! = 11 978 571 669 969 891 796 072 783 721 689 098 736 458 938 142 546 425 857 555 362 864 628 009 582 789 845 319 680 000 000 000 000 000[3]. Combinaties tellen is dus met relatief kleine aantallen al een probleem. Grote dobbelsteenproblemen waarin faculteiten van grote getallen voorkomen zijn al snel lastig. En de werkelijkheid is nog erger. Als de temperatuur van een liter gas gerelateerd is aan het totaal van de energietoestanden van de individuele moleculen[4] die elk net als een dobbelsteen alleen discrete waarden hebben, dan komen voor n al snel getallen zo groot als het getal van Avogadro[5] in beeld, dus een relevante vraag is hoe groot $(6 \times 10^{23})!$ is!

Kortom, er is alle reden om ons te verdiepen in de rij getallen $n!$ die voor $n = 0, 1, 2, 3, \ldots$ inductief gedefinieerd wordt door

$$0! = 1 \quad \text{en} \quad n! = n \times (n-1)! \quad \text{voor} \quad n \geq 1 \quad \text{geheel.}$$

De facto is $n!$ dus het product van n factoren, namelijk van de eerste n positieve gehele getallen.

Bij afspraak begint de rij van faculteiten met $0! = 1$, een product van nul factoren. Omdat 1 het neutrale element van de bewerking $\times$ is, zeggen we dat een leeg product gelijk is aan 1, net zoals een lege som van getallen als 0, het neutrale element van de bewerking $+$, wordt gedefinieerd. In plaats van $\times$ schrijven we trouwens meestal $\cdot$, en als we letterrekenen meestal niets. Dus

$$n \times (n-1)! = n \cdot (n-1)! = n(n-1)!\,,$$

met voorrang[6] voor het uitroepteken dat gebruikt wordt in de notatie van faculteiten.

8.1 De formule van Stirling

Faculteiten komen overal in de wiskunde en haar[7] toepassingen voor, en daarmee ook de vraag hoe groot $n!$ is als n groot is. Het antwoord is GROOT. Hoe groot GROOT is wordt voor $n!$ gegeven door de uitspraak dat

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad \text{voor} \quad n \to \infty. \tag{8.1}$$

[2] 10^{100}, ook wel googol, zie via google Wikipedia.

[3] Waarom dit getal op zestien nullen eindigt is wellicht een aardige vraag.

[4] Google op Boltzmann.

[5] Wie was dat?

[6] Faculteiten wachten niet op antwoord.

[7] ‘Die Mathematik macht traurige Leute’ stond er in 1976 in de studiebrochures.

Deze benadering van $n!$ met deze formule van Stirling is het leidende onderwerp van dit hoofdstuk met vragen als: wat staat hier eigenlijk, waarom is het waar, en wat komt er allemaal bij kijken als we dit precies willen maken?

Lezend van links naar rechts valt als eerste het teken $\sim$ op. Dit teken moet gelezen worden als een zwakkere vorm van het gelijkteken, in combinatie met de toevoeging voor $n \to \infty$: (8.1) is een uitspraak over het gedrag van twee rijen getallen, namelijk de rij[8]

$$1, 2, 6, 24, 120, 720, \ldots$$

en een zo op het oog totaal andere rij[9]

$$\frac{1}{e}\sqrt{2\pi}, \frac{4}{e^2}\sqrt{4\pi}, \frac{27}{e^3}\sqrt{6\pi}, \frac{64}{e^4}\sqrt{8\pi}, \frac{3125}{e^5}\sqrt{10\pi}, \frac{46656}{e^6}\sqrt{12\pi}, \ldots$$

De uitspraak zegt niets over het vierde element in de ene rij met betrekking tot het vierde element in de andere rij, en ook niet over het googolste element van beide rijen. Als a_n en b_n twee rijen positieve getallen zijn dan zeggen we dat

$$a_n \sim b_n \quad \text{voor} \quad n \to \infty$$

als het quotiënt

$$\frac{a_n}{b_n} \quad \text{naar 1 gaat als } n \to \infty.$$

Slordig gezegd: als $n \to \infty$ dan is in de limiet de verhouding tussen a_n en b_n gelijk aan 1. Precies gezegd[10]: gegeven iedere gewenste foutmarge $\epsilon > 0$ is vanaf zekere n, zeg vanaf $n = N$, voldaan aan de foutafschatting

$$\left| \frac{a_n}{b_n} - 1 \right| < \epsilon.$$

Twee zulke rijen heten asymptotisch gelijk, hoe groot het verschil tussen a_n en b_n ook blijft. Het gaat alleen om de verhouding voor steeds grotere n: als de getallen a_n en b_n in paren voorbijkomen, dan vallen de eerste paren buiten de gewenste foutmarge, maar vanaf $n = N$ is aan de gewenste precisie voldaan. Precisie is hier een kwestie van geduld: hoe kleiner $\epsilon > 0$ hoe langer je moet wachten[11].

Opgave 8.1. Ga na dat de uitspraak $a_n \sim b_n$ voor de rijen $a_n = n(n+1)$ en $b_n = n^2$ equivalent is met de uitspraak dat gegeven een willekeurige $\epsilon > 0$ er een index N is

[8] Om interessante verwarring te vermijden laten we beide rijen nu beginnen met $n = 1$.

[9] Waarin naast de vierkantswortel ook e en π voorkomen, waarover later meer.

[10] We schrijven hier de formele definitie van $A_n \to A = 1$ als $n \to \infty$ over met $A_n = \frac{a_n}{b_n}$.

[11] Als $a_n = b_n$ voor alle n dan hoef je helemaal niet te wachten.

zo dat voor alle gehele $n \geq N$ geldt dat $\frac{1}{n} < \epsilon$. Deze eigenschap van de positieve getallen wordt de archimedische eigenschap genoemd. We zullen deze eigenschap als zijnde vanzelfsprekend waar beschouwen[12]: dat $\frac{1}{n}$ naar 0 gaat als $n \to \infty$ is de basis voor alles wat we doen in de analyse[13].

Opgave 8.2. Kun je bewijzen dat als

$$a_n \sim b_n \quad \text{voor} \quad n \to \infty, \quad \text{dat dan ook} \quad b_n \sim a_n \quad \text{voor} \quad n \to \infty?$$

Opgave 8.3. En als

$$a_n \sim b_n \quad \text{voor} \quad n \to \infty \quad \text{en} \quad b_n \sim c_n \quad \text{voor} \quad n \to \infty,$$

dat dan ook

$$a_n \sim c_n \quad \text{voor} \quad n \to \infty?$$

Op grond van deze twee opgaven spreken we over $a_n \sim b_n$ voor $n \to \infty$ als een equivalentierelatie tussen rijen positieve getallen. De relatie is immers symmetrisch (Opgave 8.2), transitief (Opgave 8.3), en ook reflexief: iedere rij positieve getallen is asymptotisch gelijk aan zichzelf.

Natuurlijk moet het zo zijn dat als je bijvoorbeeld een irrationaal getal op twee manieren met rijen rationale getallen benadert, dat dan die twee rijen asymptotisch gelijk zijn[14]. Maar in (8.1) gaat het zeker niet om benaderingen van getallen. De getallen in beide rijen worden immers steeds (en heel snel) groter. Of je kunt spreken over de formule van Stirling als een benadering van $n!$ is niet eens zo duidelijk. Maar als je een groot probleem wilt oplossen en in moet schatten hoe lang dat duurt, dan is een formule als (8.1) al snel relevant.

Dus hoe groot is $n!$ volgens Stirlings formule? Wel, in het rechterlid van (8.1) komt naast de factor $\sqrt{2\pi n}$ de n-de macht van $\frac{n}{e}$ voor, en die gaat hard, erg hard. Gezien de relatief bescheiden factor

$$\sqrt{2\pi n}$$

[12] Maar daar is over te twisten. Google op non-standard analysis.
[13] Het deelgebied van de wiskunde waarin limieten een rol spelen.
[14] Hoe kun je dit uit de opgaven hierboven halen?

gaat $n!$ nog net ietsje harder.

Beide factoren in Stirlings formule zijn mysterieus. Waarom hebben π en e überhaupt iets met $n!$ te maken? We hebben al gezien dat

$$e = \lim_{n\to\infty}(1+\frac{1}{n})^n = \sum_{n=0}^{\infty}\frac{1}{n!},$$

dus de e is misschien niet helemaal vreemd in relatie tot faculteiten, maar waar komt de wortel van π vandaan?

Wat verder opvalt is het verschil tussen de formules. Beter gezegd, $n!$ is niet eens als formule in n gedefinieerd. Want om $n!$ uit te rekenen heb je naast n ook $n-1$, $n-2$, et cetera nodig. Dit in tegenstelling tot Stirlings formule

$$\left(\frac{n}{e}\right)^n\sqrt{2\pi n},$$

waarin je gewoon n kunt invullen[15]. Stirlings formule is gebaseerd op de uitdrukking waar we hoofdstuk 6 mee afsloten,

$$n! = \int_0^\infty x^n e^{-x}\,dx,$$

en strikt genomen is er nu geen reden meer om in het rechterlid n alleen geheel[16] te nemen.

8.2 Een bewijs voor in de klas?

Met de integraalformule voor $n!$ is $n!$ dus te zien als de oppervlakte van het onbegrensde gebied G_n in het xy-vlak gedefinieerd door

$$G_n = \{(x,y) : 0 \le y \le f_n(x)\}, \quad \text{met} \quad f_n(x) = x^n e^{-x} = x^n\exp(-x),$$

waarbij we $n=0$ buiten beschouwing laten omdat we n zo meteen toch groot gaan nemen.

Indachtig de volkswijsheid dat de meeste vis tussen kop en staart zit, kijken we naar het maximum van de functie $f_n(x)$ op het x-interval $[0,\infty)$. Dit maximum vinden we door $f_n(x)$, een uitdrukking die voor iedere x als machtreeks is gedefinieerd, te differentiëren[17] en gelijk aan nul te stellen, een vast item in het functieonderzoek[18] van een functie en het tekenen of schetsen van de grafiek.

[15] $n =$ googol bijvoorbeeld
[16] Google op Gammafunctie.
[17] We komen nog te spreken over differentiaalrekening in hoofdstuk 10.
[18] Waar je meer van leert dan plotten met de GRM.

Opgave 8.4. Gebruik de regels van de differentiaalrekening om af te leiden dat er maar een nulpunt is van $f_n'(x)$, namelijk $x = n$, waar f_n zijn enige maximum $f_n(n) = (\frac{n}{e})^n$ aanneemt.

We merken alvast op dat nu de belangrijkste factor in Stirlings formule al geïdentificeerd is! Hoe nu verder te gaan? We zorgen eerst maar eens dat het maximum precies in het midden bij $x = 0$ ligt door de grafiek van f_n horizontaal naar links op te schuiven. Waar eerst een n stond op de x-as, zetten we nu dus een 0 neer, maar de eenheden op de x-as blijven verder hetzelfde.

Vervolgens schalen we eenheden op de y-as precies zo dat het maximum niet meer met $y = (\frac{n}{e})^n$, maar met $y = 1$ correspondeert. In de nieuwe coördinaten hebben we dan de grafiek van de functie g_n gedefinieerd door

$$g_n(x) = \frac{f_n(x+n)}{(\frac{n}{e})^n}, \tag{8.2}$$

voor $x \geq -n$. De noemer in het rechterlid van (8.2) is de verticale schalingsfactor waarmee het maximum gelijk wordt aan 1, en aan de $x+n$ in f_n zie je de horizontale verschuiving naar links waarmee het maximum in $x = 0$ komt te liggen[19].

Uitwerken van (8.2), met behulp van het functievoorschrift van f_n, leert ons dat g_n gegeven wordt door

$$g_n(x) = (1 + \frac{x}{n})^n e^{-x} = \frac{(1+\frac{x}{n})^n}{\exp(x)} = \frac{e_n(x)}{\exp(x)} \tag{8.3}$$

voor $x \geq -n$. Een quotiënt van twee oude bekenden: $e_n(x)$ gedefinieerd door (4.2) en de limiet voor $n \to \infty$ van diezelfde (stijgende) rij $e_n(x)$. Voor elke x geldt dus dat

$$\lim_{n\to\infty} g_n(x) = 1.$$

De oppervlakte tussen de grafiek van g_n en de x-as die door het horizontale schuiven niet, maar door het verticale schalen wel veranderd is, is nu gelijk aan

$$\int_{-n}^{\infty} g_n(x)dx = n!\,(\frac{e}{n})^n\,, \tag{8.4}$$

en dat gaat nog steeds (monotoon stijgend[20]) naar oneindig als $n \to \infty$ omdat de grafiek van g_n steeds meer gaat lijken op de horizontale lijn gegeven door $y = 1$.

---

[19] Maak zelf een plaatje, ook bij wat nog volgt.
[20] Waarom?

De enige manier om een oneindige limiet hier te vermijden is door ook horizontaal te schalen en wel op zo'n manier dat in de limiet de geschaalde functie (wel) een strikt maximum heeft in $x = 0$. Ook deze schaling, waarmee we de grafiek als het ware naar de y-as toetrekken, moet dus afhangen van n. Trekken we te hard dan blijft er geen oppervlakte over onder de grafiek, en trekken we niet hard genoeg dan is de oppervlakte niet begrensd als $n \to \infty$. Het is dus zaak om een schaling te vinden waarmee we tussen nul en oneindig laveren.

Als je goed kijkt is zo'n schaling af te lezen uit de formules, maar daar gaan wat denkactiviteiten aan vooraf, want wat voor limietfunctie verwachten we eigenlijk, als we op de juiste manier schalen? De limietfunctie g moet een globaal strikt maximum 1 hebben in $x = 0$ en positief zijn voor elke x, en waarschijnlijk wel strikt positief, want het nulpunt $x = -n$ van g_n zal in de schaling die we zoeken nog steeds wel weglopen naar $-\infty$. Althans, als we zoeken naar een uitdrukking die (8.4) koppelt aan

$$\int_{-\infty}^{\infty} g(x)dx$$

voor $n \to \infty$ via een evenredigheidsconstante die afhangt van n. Als zo'n functie g een gewoon maximum (ter grootte van 1 nog steeds) heeft in $x = 0$ en in de vorm van een machtreeks is te vinden, dan moet wel gelden dat

$$g(x) = 1 - ax^2 + \cdots ,$$

waarbij de factor $a > 0$ dan door modificatie van de schaling ook wel gelijk aan 1 te kiezen is. Of aan $\frac{1}{2}$, als je denkt in termen van $g''(0)$ zo simpel mogelijk[21]. Het ligt dus voor de hand om zo te schalen dat de geschaalde versie van g_n in $x = 0$ een tweede afgeleide krijgt die gelijk is aan -1. Probeer het maar en kijk of je vervolgens iets met de limiet voor $n \to \infty$ kunt. Niet eenvoudig.

De slimme truc[22] nu is om de limietfunctie g te zoeken met een formulevoorschrift van de vorm

$$g(x) = \exp(-\psi(x)),$$

met $\psi(x)$ zo simpel mogelijk. Dat vraagt voor g_n dus om de introductie van functies ψ_n via

$$\frac{e_n(x)}{\exp(x)} = g_n(x) = e^{-\psi_n(x)} = \frac{1}{\exp(\psi_n(x))}$$

[21] Welke parabool vind je mooier, $y = x^2$ of $y = \frac{1}{2}x^2$?
[22] Van Laplace vermoeden we.

met $\psi_n(0) = 0$ en $\psi_n(x) > 0$ voor $0 \neq x > -n$. De facto komt dit neer op

$$\psi_n(x) = -\ln(g_n(x)),$$

mits we de natuurlijke logaritme ln als de inverse functie van exp tot onze beschikking hebben, hetgeen vanzelfsprekend lijkt na hoofdstuk 4.

We smokkelen nu een beetje en lopen in het vervolg van deze sectie vooruit op wat we in hoofdstuk 10 precies zullen maken, namelijk dat de functie ln wordt gedefinieerd door

$$\exp(x) = y \quad \Leftrightarrow \quad x = \ln(y),$$

en ook dat we inmiddels wel kunnen raden hoe het zit met de afgeleide functies van exp en ln. Uit hoofdstuk 4 weten we immers dat

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots ,$$

en als we het rechterlid term voor term mogen differentiëren dan vinden we dat $\exp'(x) = \exp(x)$, probeer het maar. Met behulp van de relatie tussen de grafieken van exp en ln moet het dan wel zo zijn dat

$$\ln'(y) = \frac{1}{\exp(x)} = \frac{1}{y}$$

de afgeleide is van $\ln(y)$. De grafiek van ln hadden we immers al: klap de grafiek $y = \exp(x)$ in het xy-vlak maar om door hem als een schilderij met twee handen vast te pakken en verkeerd om met je rechterhand boven en je linkerhand onder weer terug te hangen. Als het papier doorzichtig is dan kijk je nu naar de grafiek $x = \ln(y)$ in het yx-vlak. Raaklijnen zijn daarbij raaklijnen gebleven. Je moet alleen de richtingscoëffiënten op hun kop zetten, en dat is precies wat in de formule hierboven is gebeurd. Hiermee is ln nu een differentieerbare functie gedefinieerd op de positieve reële getallen met als afgeleide de rationale functie met formulevoorschrift

$$\ln'(x) = \frac{1}{x}. \tag{8.5}$$

Opgave 8.5. Laat nu met de bekende rekenregels[23] voor ln zien dat $\psi_n(x)$ gegeven wordt door

$$\psi_n(x) = x - n\ln(1 + \frac{x}{n}) = n(\frac{x}{n} - \ln(1 + \frac{x}{n})) = n\psi_1(\frac{x}{n}).$$

[23] Waarom geldt met de gegeven definitie van ln dat $\ln(ab) = \ln(a) + \ln(b)$?

We zien hier dat de rij functies ψ_n door schalen ontstaat uit een en dezelfde functie $\Psi = \psi_1$, en die functie heeft als formulevoorschrift

$$\Psi(x) = x - \ln(1+x), \tag{8.6}$$

het verschil van x en $\ln(1+x)$.

Hoe ziet de grafiek van Ψ eruit? Je weet al dat $\Psi(0) = \Psi'(0) = 0$. Wat de afgeleide betreft, je ziet via schuiven[24] dat de afgeleide van $\ln(1+x)$ gelijk moet zijn aan $\frac{1}{1+x}$ voor $x > -1$, en dus is

$$\Psi'(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x} \sim x$$

als $x \to 0$, omdat de noemer $1+x$ naar 1 gaat. Je moet er misschien even over nadenken, maar voor Ψ zelf betekent dit dat

$$\Psi(x) \sim \frac{1}{2}x^2$$

en de grafiek van Ψ wordt verder gekenmerkt door een verticale asymptoot bij $x = -1$ en bijna-lineaire groei voor $x \to \infty$ (maar geen scheve asymptoot).

Wel een schetsje waard deze grafiek[25] maar belangrijker is dat voor ψ_n nu geldt dat

$$\psi_n(x) = n\Psi(\frac{x}{n}) \sim \frac{x^2}{2n},$$

niet alleen voor $x \to 0$, maar ook voor $\frac{x}{n} \to 0$, hetgeen steeds grotere x toelaat, maar waarbij in de limiet $n \to \infty$ niets overblijft. Een schaling waarbij wel wat overblijft lezen we onmiddellijk af. We moeten x zo schalen dat de n in het rechterlid verdwijnt, en dus substitueren we $x = s\sqrt{n}$. Met deze schaling volgt uit (8.4) deze schitterende formule, waaronder we met twee vraagtekens aangeven wat de laatste twee vragen zijn:

$$n! = (\frac{n}{e})^n \sqrt{n} \underbrace{\int_{-\sqrt{n}}^{\infty} \exp(-n\Psi(\frac{s}{\sqrt{n}}))ds}_{\substack{\downarrow ? \\ \int_{-\infty}^{\infty} \exp(-\frac{1}{2}s^2)ds = ?}}. \tag{8.7}$$

In de integrand zien we een nieuwe anders geschaalde versie van Ψ, die we zo hebben gekozen dat voor elke n geldt dat

$$n\Psi(\frac{s}{\sqrt{n}}) \sim \frac{1}{2}s^2, \tag{8.8}$$

[24] De kettingregel zou hier het spreekwoordelijke kanon voor de mug zijn.
[25] Eerst met de DRM bij de hand.

nu niet alleen voor $s \to 0$, maar ook voor $\frac{s}{\sqrt{n}} \to 0$. Het fraaie is dat met grotere n deze geschaalde $\Psi(s)$ steeds meer op zijn kwadratische benadering bij $s = 0$ gaat lijken, zoals we zo meteen zullen laten zien. We zoomen als het ware in op de oorsprong en in de limiet blijft alleen de parabool over, juist omdat we (de grafiek van) Ψ verticaal en horizontaal anders schalen. De schaling is zo gekozen dat $\frac{1}{2}s^2$ van schalen niets merkt en in de limiet $n \to \infty$ wordt het linkerlid in (8.8) gelijk aan het rechterlid. De linkergrens van de integraal schuift daarbij op naar $-\infty$. Precies wat we wilden. Dus wat moeten we nu nog doen om dit precies te maken opdat we kunnen concluderen dat

$$n! \sim (\frac{n}{e})^n \sqrt{n} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}s^2)ds \text{ uitroepteken?}$$

Een gewetensvraag. De lezer die hier zegt: 'Ik geloof het nu verder wel', kan het onderstaande wat globaler lezen of overslaan, maar na alle inspanningen willen we de andere lezer de precieze toedracht hier niet onthouden. Het simpele feit is namelijk dat het linkerlid in (8.8) dalend is in n voor $s < 0$ en stijgend voor $s > 0$, met als limiet $\frac{1}{2}s^2$. Het vergt wat rekenwerk om dit netjes aan te tonen, maar de uiteindelijke conclusie is nu dat (voor $n \uparrow \infty$)

$$\int_{-\sqrt{n}}^{0} \exp(-n\Psi(\frac{s}{\sqrt{n}}))ds \uparrow \int_{-\infty}^{0} \exp(-\frac{1}{2}s^2)ds,$$

zo je wilt vanwege de eerste monotone convergentiestelling in sectie 7.4, en ook, vanwege de tweede monotone convergentiestelling, dat

$$\int_{0}^{\infty} \exp(-n\Psi(\frac{s}{\sqrt{n}}))ds \downarrow \int_{0}^{\infty} \exp(-\frac{1}{2}s^2)ds,$$

indien voor $n = 1$ de integraal eindig is. En dat is inderdaad het geval omdat de bijna lineaire groeit van $\Psi(s)$ impliceert dat

$$\int_{0}^{\infty} \exp(-\Psi(s))ds < \infty.$$

We concluderen dus dat de hele integraal in (8.7) als som van deze beide integralen convergeert naar de som van de limietintegralen en die som is

$$\int_{-\infty}^{\infty} \exp(-\frac{1}{2}s^2)ds,$$

de integraal die we in de volgende sectie nog zullen uitrekenen. Voor nu is het resultaat dat voor $n \to \infty$ geldt dat

$$n! \sim (\frac{n}{e})^n \sqrt{n} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}s^2)ds \quad \text{uitroepteken!} \tag{8.9}$$

We besluiten deze sectie met een precieze analyse van de functie Ψ en zijn geschaalde versie in het linkerlid van (8.8), en we doen dit aan de hand van de grafieken gegeven door

$$y = n\Psi(\frac{s}{\sqrt{n}}), \quad \Psi(s) = s - \ln(1+s)$$

in het sy-vlak. Zoals we gezien hebben lijken al deze grafieken in de buurt van de top $(0,0)$ op dezelfde dalparabool $y = \frac{1}{2}s^2$. Dat we in de limiet $n \to \infty$ die dalparabool zelf krijgen is een gevolg van het inzoomen, en heeft verder weinig met de precieze vorm van de grafiek te maken. Elke functie Ψ met $\Psi(s) \sim \frac{1}{2}s^2$ als $s \to 0$ zal dit gedrag vertonen en de standaardmanier om dit aan te tonen is met de bekende regel[26] voor $\frac{0}{0}$-limieten, die in

$$\lim_{s\to 0} \frac{\Psi(s)}{\frac{1}{2}s^2} = \lim_{s\to 0} \frac{\Psi'(s)}{s} = \lim_{s\to 0} \frac{\Psi''(s)}{1} = \Psi''(0) = 1$$

twee keer is toegepast, gebruikmakend van de standaardrekenregels voor de differentiaalrekening die we tot nu toe zo veel mogelijk hebben vermeden, maar nu even niet. Uit deze limiet volgt voor elke s onmiddellijk dat

$$\lim_{n\to\infty} n\Psi(\frac{s}{\sqrt{n}}) = \frac{1}{2}s^2$$

en om te zien dat de grafieken aan de rechterkant de dalparabool monotoon van onderen, en aan de linkerkant monotoon van boven benaderen, moet je weten of afgeleide naar n van de geschaalde Ψ-functie links en rechts het juiste teken heeft, *that's all.* Die afgeleide rekenen we maar meteen uit met de bekende regeltjes, waaronder nu ook de kettingregel. We vinden

$$\frac{d}{dn} n\Psi(\frac{s}{\sqrt{n}}) = \Psi(\frac{s}{\sqrt{n}}) + n\Psi'(\frac{s}{\sqrt{n}})(-\frac{1}{2})\frac{s}{n^{\frac{3}{2}}} = \Psi(\frac{s}{\sqrt{n}}) - \frac{1}{2}\frac{s}{\sqrt{n}}\Psi'(\frac{s}{\sqrt{n}}). \quad (8.10)$$

Kijk nou, daar zien we

$$\Psi(x) - \frac{1}{2}x\Psi'(x) = x - \ln(1+x) - \frac{1}{2}\frac{x^2}{1+x}$$

verschijnen, met voor x weer een geschaalde s ingevuld. Het uitgewerkte rechterlid definieert een prachtige stijgende functie die 0 is in 0. Dat zie je zo[27]. Hetzelfde geldt dus voor (8.10). Het was even dom rekenen, maar we we weten het nu zeker. De geschaalde functies Ψ in (8.7) hebben precies de eigenschappen die we wilden. Ze gaan monotoon naar $\frac{1}{2}s^2$, dalend aan de linker- en stijgend aan de rechterkant.

[26] Voor machtreeksen een kwestie van wegdelen en invullen.

[27] Bereken zijn afgeleide maar, die is $\frac{1}{2}\frac{x^2}{(1+x)^2}$, hoe is het mogelijk?

8.3 Een cirkelredenering voor een integraal

We kiezen nu weer voor de notatie met e als grondtal en de variabele in de exponent[28], en wat ons nog rest is het uitrekenen van de integraal in (8.9). Een primitieve vinden lukt niet, we hebben het geprobeerd, maar er zijn wiskundigen die ons uitleggen hoe dom dat was. In plaats daarvan geven we nu een rechtstreekse methode die aansluit bij onze voorkeur om in dit boekje essentiële zaken zo basaal mogelijk te laten zien. Let wel, in de berekening moet ook duidelijk worden hoe π in het antwoord terecht komt.

Allereerst een mooie truc, een slim idee waar je maar op moet komen. Het *volume* van de met een xyz-assenstelsel beschreven verzameling van punten in de driedimensionale ruimte waarvoor geldt dat

$$0 \leq z \leq e^{-\frac{1}{2}(x^2+y^2)},$$

is het kwadraat van de oppervlakte A tussen $y = e^{-\frac{1}{2}x^2}$ en $y = 0$ in het xy-vlak.

Hoe zien we dat? Wel, de even functie f met functievoorschrift

$$f(x) = e^{-\frac{1}{2}x^2}$$

geeft ons A als

$$A = \int_{-\infty}^{\infty} f(x)dx = 2\int_0^{\infty} f(x)dx,$$

en gegeven deze f kunnen we ook kijken naar het volume B van het gebied in de xyz-ruimte tussen $z = 0$ en $z = f(x)f(y)$, en dat is vanwege de symmetrie weer vier keer de inhoud van het gebied waar $x \geq 0$ en $y \geq 0$. In de volgende opgaven laten we zien dat $B = A^2$ door te laten zien dat $\frac{B}{4} = (\frac{A}{2})^2$ en dat doen we weer in de geest van wat we eerder deden bij integralen van polynomen: we halen het resultaat dat we willen uit berekeningen met de benaderingen van A en B.

We benaderen $\frac{A}{2}$ daarom door onder- en bovensommen. De ondersommen voor $\frac{A}{2}$ krijgen we als de totale oppervlakte van de rechthoekjes met breedte $\frac{1}{n}$ en hoogte $e^{-\frac{1}{2}(\frac{k}{n})^2}$, waarbij we, nu we vertrouwd zijn met wat maattheorie, k wel door kunnen laten tellen, dus $k = 1, 2, \ldots$ De oppervlakte van de individuele rechthoekjes noemen we even $a_k^{(n)}$, dus de totale ondersom voor $\frac{A}{2}$ wordt

$$\sum_{k=1}^{\infty} a_k^{(n)} = \frac{1}{n}\sum_{k=1}^{\infty} e^{-\frac{1}{2}(\frac{k}{n})^2}.$$

[28] Hetgeen de lezer ons wellicht in dank zal afnemen.

Als je liever met eindige sommen werkt dan neem je bijvoorbeeld

$$\sum_{k=1}^{n^2} a_k^{(n)},$$

waarmee je op steeds grotere x-intervallen $[0, n]$ de oppervlakte onder de grafiek benadert. De monotone convergentiestelling garandeert dat we zo in de limiet de totale oppervlakte $\frac{A}{2}$ krijgen.

Ook zonder alle theorie is dat wel duidelijk. De aftelbare ondersom is een ondergrens en de aftelbare bovensom, die begint bij $k = 0$, is een bovengrens. Het verschil is precies $\frac{1}{n}$. Kortom, zodra er een bovensom[29] eindig is doen (ook) de (eindige) ondersommen wat we willen, ze benaderen $\frac{A}{2}$ zo goed als we maar willen.

Opgave 8.6. Waarom kunnen we het volume tussen $z = 0$ en $z = f(x)f(y)$ van onderen benaderen met het product

$$\left(\sum_{k=1}^{n^2} a_k^{(n)}\right) \times \left(\sum_{m=1}^{n^2} a_m^{(n)}\right)?$$

Hint: je kunt $a_k^{(n)} a_m^{(n)}$ meetkundig interpreteren. Concludeer dat $\frac{B}{4} = (\frac{A}{2})^2$.

We hebben nu dus rechthoekjes en 'blokjes' gebruikt om aan te tonen dat $B = A^2$. Om B ook daadwerkelijk uit te rekenen benaderen we het volume ook met cilinders, of beter gezegd, met de dikke randen van cilinders. Waar de blokjes op vierkantjes met zijde $\frac{1}{n}$ stonden, staan de cilinders op annuli in het xy-vlak met binnenstraal $\frac{k}{n}$ en buitenstraal $\frac{k+1}{n}$, waarbij k loopt van nul tot oneindig. De randen van de annuli zijn de concentrische cirkels met middelpunt in de oorsprong en straal $\frac{k}{n}$.

We noemen die cirkels even $C_{n,k}$ en gebruiken de speciale eigenschap die $f(x)f(y)$ voor onze keuze van f heeft, namelijk dat de productfunctie radiaal symmetrisch en dalend is en op de annulus tussen $C_{n,k}$ en $C_{n,k+1}$ dus inzit tussen

$$e^{-\frac{k^2}{2n^2}} \quad \text{en} \quad e^{-\frac{(k+1)^2}{2n^2}}.$$

Het volume tussen de grafiek $z = f(x)f(y)$ en $z = 0$ is dus minder dan de benaderende bovensom

$$\sum_{k=0}^{\infty} (\text{oppervlakte tussen } C_{n,k+1} \text{en } C_{n,k}) \cdot e^{-\frac{k^2}{2n^2}}.$$

[29] Overtuig jezelf ervan dat dat hier zo is, de termen gaan keihard naar nul.

De oppervlakte tussen twee zulke cirkels $C_{n,k+1}$ en $C_{n,k}$ is

$$\pi\left(\frac{k+1}{n}\right)^2-\pi\left(\frac{k}{n}\right)^2=\pi\left(\frac{2k}{n^2}+\frac{1}{n^2}\right)=\frac{2\pi k}{n^2}+\frac{\pi}{n^2},$$

en de bovensom is dus gelijk aan

$$\sum_{k=0}^{\infty}\left(\frac{2\pi k}{n^2}+\frac{\pi}{n^2}\right)e^{-\frac{k^2}{2n^2}}=2\pi\underbrace{\frac{1}{n}\sum_{k=0}^{\infty}\frac{k}{n}e^{-\frac{1}{2}(\frac{k}{n})^2}}_{\to\int_0^\infty re^{-\frac{1}{2}r^2}dr}+\frac{\pi}{n}\underbrace{\frac{1}{n}\sum_{k=0}^{\infty}e^{-\frac{1}{2}(\frac{k}{n})^2}}_{\to\int_0^\infty e^{-\frac{1}{2}x^2}dx},$$

waarin we benaderende sommen voor integralen herkennen, zoals aangegeven onder de accolades. De eerste benaderende som convergeert[30] naar de integraal

$$\int_0^\infty re^{-\frac{1}{2}r^2}dr,$$

en die kunnen we uitrekenen[31], (zo je wilt) door $-e^{-\frac{1}{2}r^2}$ als primitieve van de integrand (eerst als een machtreeks) te herkennen, en omdat die primitieve naar 0 gaat als $r\to\infty$ en gelijk aan -1 is in $r=0$ is deze integraal gelijk aan 1! De tweede benaderende som convergeert naar de integraal die we juist willen uitrekenen, maar gelukkig staat er nog een $\frac{\pi}{n}$ in, die de hele tweede term naar 0 trekt. We concluderen dat het volume B waar we naar zochten gelijk is aan 2π. En dus volgt nu dat

$$\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2}dx=\sqrt{2\pi}.$$

Dat er een π verschijnt in de uitkomst is vanwege de zeer bijzondere integrand die een puur kwadratische term in de exponent heeft. Deze kwadratische term zorgt ervoor dat wanneer we de functie als het ware met 'zichzelf' vermenigvuldigen, maar wel met een andere variabele, de functie van twee variabelen die we dan krijgen alleen van x en y afhangt via de afstand tot de oorsprong. Dat maakt dat we het volume onder de grafiek kunnen benaderen met dikke cilinders, en daarbij verschijnt de π waarvan je wist dat die ging komen.

8.4 Multinomiaalcoëfficiënten en entropie

We weten nu dat

$$n!\sim(\frac{n}{e})^n\sqrt{2\pi n},$$

[30] Misschien wil je onderscheid maken tussen $r<1$ en $r>1$ vanwege monotonie.

[31] $\int_0^\infty f(x)dx=F(\infty)-F(0)$, niet waar voor polynomen, voor machtreeksen soms wel!

hetgeen dus betekent dat het quotiënt van linker- en rechterterm naar 1 gaat als $n \to \infty$. Zoals we al opmerkten aan het begin van dit hoofdstuk, faculteiten kunnen faculteiten van grote getallen zijn, waarmee $n!$ EXTREEM groot wordt en van dat soort grootheden neem je dan vaak de logaritme. Mogen we die links en rechts nemen als er een $\sim$ in plaats van een $=$ staat? Zoja, dan is

$$\ln(n!) \sim n \ln(\frac{n}{e}) + \frac{1}{2}\ln(2\pi n) = n \ln n - n + \frac{1}{2}\ln(n) + \frac{1}{2}\ln(2\pi) \sim n \ln n,$$

waarin na de tweede $\sim$ alle fijnere details zijn weggegooid. Dat de tweede $\sim$ correct is kun je eenvoudig nagaan, maar hoe zit het met de eerste $\sim$?

Die vraag laten we hier voor wat hij is. Dat de eindconclusie

$$\ln(n!) \sim n \ln(n)$$

correct is kun je echter wel aanvoelen. Onze expansie is van de vorm GROOT (n^n) keer minder groot (e^n keer de rest). Maar

$$\ln(\text{GROOT} \times \text{groot}) = \ln(\text{GROOT}) + \ln(\text{groot})$$

en deel je dat door ln(GROOT) dan krijg je 1 plus een quotiënt van in dit geval zoiets als $e \ln(n)$ en $n \ln(n)$ en dat gaat wel naar 0. Onze bescheiden conclusie voor $\ln n!$ zal dus wel correct zijn, want de fout die we maken is ongetwijfeld kleiner dan het stuk van de Stirling-benadering zelf dat we weggooien.

We kunnen de formule van Stirling nu gebruiken om iets meer te weten te komen over oude bekenden: de multinomiaalcoëfficiënten. Stirling zegt dat

$$\begin{pmatrix} & n & \\ k_1 & \dots & k_m \end{pmatrix} = \frac{n!}{k_1! \cdots k_m!} \sim \frac{(\frac{n}{e})^n \sqrt{2\pi n}}{(\frac{k_1}{e})^{k_1}\sqrt{2\pi k_1} \cdots (\frac{k_m}{e})^{k_m}\sqrt{2\pi k_m}}$$

(omdat $k_1 + \cdots + k_m = n$)

$$= \frac{n^n \sqrt{2\pi n}}{k_1^{k_1}\sqrt{2\pi k_1} \cdots k_m^{k_m}\sqrt{2\pi k_m}} = (2\pi)^{\frac{1-m}{2}} \sqrt{\frac{n}{k_1 \cdots k_m}} (\frac{n}{k_1})^{k_1} \cdots (\frac{n}{k_m})^{k_m}.$$

Nu pas nemen we links en rechts de logaritme, delen door n en vinden de formule

$$\frac{1}{n} \ln \left[\begin{pmatrix} & n & \\ k_1 & \dots & k_m \end{pmatrix} \right] \sim \frac{k_1}{n}\ln(\frac{n}{k_1}) + \cdots + \frac{k_m}{n}\ln(\frac{n}{k_m}) = \sum_{i=1}^{m} p_i \ln(\frac{1}{p_i}),$$

als we p_i gelijkstellen[32] aan $p_i = \frac{k_i}{n}$. Dit is een uitdrukking waar veel over te vertellen is[33]. Het rechterlid wordt de entropie genoemd van de p_i-tjes, en doorgaans[34] aangegeven met H en genoteerd als

$$H = -\sum_{i=1}^{m} p_i \ln(p_i).$$

Hoe groter de entropie, hoe groter de multinomiaalcoëfficiënt, en dus hoe groter het aantal combinatorische mogelijkheden om een verzameling n elementen in stukken ter grootte $k_1, \ldots, k_m$ te hakken. Het is leuk om daar een beetje mee te spelen.

Opgave 8.7. Neem $m = 2$ en laat zien dat de entropie maximaal is als $p_1 = p_2 = \frac{1}{2}$.

Entropie speelt een belangrijke rol in de statistische fysica, waar kanstheoretische wetten ons vertellen welke toestanden we wel of juist niet zullen zien en waarbij toename van entropie een soort basisprincipe is. Ook in de informatietheorie wordt dit begrip gehanteerd. Daar is entropie een maat voor de onzekerheid over het volgende te ontvangen bit als er m opties zijn met kansen $p_1, \ldots, p_m$. Ten slotte speelt entropie een belangrijke rol in de dynamica. Ook dat vergt een heel boek op zichzelf.

[32] Merk op dat de som van de p_i-tjes gelijk is aan 1.
[33] Maar dat doen we niet allemaal hier.
[34] Behalve door Joost die liever geen positief getal wil laten beginnen met een minteken.

9 De normale verdeling

Waar komt die normale verdeling toch vandaan? De vraag werd gesteld tijdens een cursus over het eerste deel van dit boekje toen we het hadden over de wet van de grote aantallen bij de vraag hoe groot de kans is dat een willekeurig gekozen getal normaal is. In het speciale geval dat we getallen binair representeren en met een zuivere munt gooien om de bits te trekken zegt die wet dat het gemiddelde aantal enen naar $\frac{1}{2}$ gaat.

Preciezer gezegd, met kans 1 geldt dat

$$\frac{X_n}{n} \to \frac{1}{2}$$

als $n \to \infty$, waarbij

$$X_n = \sum_{j=1}^{n} B_j,$$

en $B_1, \ldots, B_n$ onderling onafhankelijke stochastische grootheden zijn, iedere B_j met twee mogelijke uitkomsten, 0 en 1, allebei optredend met kans gelijk aan $\frac{1}{2}$. Met zulke B-tjes maak je natuurlijk alle grootheden die met gelijke kans twee verschillende uitkomsten hebben, $U_j = 2B_j - 1$ bijvoorbeeld, met waarden -1 en 1, en deze U-tjes maken

$$S_n = \sum_{j=1}^{n} U_j, \tag{9.1}$$

de samengestelde stochastische grootheid die ons nu interesseert.

Je ziet eenvoudig dat $S_n = 2X_n - 1$ en dus ook dat

$$\frac{S_n}{n} \to 0$$

met kans 1. Een vraag om nu te stellen is of een andere schaling van S_n een limiet geeft die meer of andere informatie geeft dan deze nul die eigenlijk maar weinig[1] informatie bevat. Is er misschien een a priori argument voor wat die schaling zou moeten zijn? Die schaling moet natuurlijk iets te maken hebben met de spreiding van de uitkomsten en die spreiding hangt ongetwijfeld af van n, maar hoe weten we nog niet. We hebben nog weinig ervaring met kansverdelingen voor stochastische grootheden die ontstaan door aan uitkomsten van steeds meer experimenten waarden toe te kennen via formules zoals hierboven. De kennis op basis waarvan we een *judicious guess* kunnen doen, ontbreekt ons. Daarom kiezen we nu voor een down to earth aanpak: we rekenen gewoon uit wat de kansverdeling van S_n is en kijken hoe de resulterende formules van n afhangen, en dan zien we wel. Daar moeten we vast bij tellen, dus om de faculteiten waar grote getallen in voorkomen kunnen we niet heen. Dat is mooi, want daar weten we inmiddels wel het een en ander van.

9.1 Tellen en rekenen met faculteiten

De somnotatie voor S_n in (9.1) is elegant, maar als we het aantal keren dat we kop gooien K_n noemen, en het aantal keren munt M_n, dan oogt

$$S_n = K_n - M_n$$

toch wat overzichtelijker. We hebben hier dan afgesproken dat kop $+1$ geeft, zeg maar winst, en munt -1, verlies. De mogelijk uitkomsten voor S_n lopen dan van $-n$ tot $+n$, met stapjes van twee, en dat brengt ons op de gedachte

[1] Nul als limiet is maar niks.

om eerst naar even n te kijken, omdat 0 zelf dan ook een mogelijke uitkomst is. We kijken daarom eerst naar S_{2n}, met n zelf niet per se even maar de mogelijke uitkomsten van S_{2n} wel.

Laten we eens beginnen met S_4. Wat is bijvoorbeeld de kans dat $S_4 = 0$? Wel, daartoe moeten we twee keer kop en twee keer munt gooien. Van alle $2^4 = 16$ uitkomsten zijn er precies $\binom{4}{2}$ met twee keer kop en de rest munt. De kans op $S_4 = 0$ is dus gelijk aan

$$P(S_4 = 0) = \binom{4}{2}\left(\frac{1}{2}\right)^4 = \binom{4}{2}2^{-4},$$

wat we herkennen als de kans op 2 bij een binomiale verdeling met parameters $N = 4$ en $p = \frac{1}{2}$.

Op dezelfde manier berekenen we de kans op drie keer kop en één keer munt (de kans op $S_4 = 2$) en als we wat algemener kijken zien we ook wel hoe het zit met de kans op de gebeurtenis dat $S_{2n} = K_{2n} - M_{2n} = 2k$. Omdat $K_{2n} + M_{2n} = 2n$ moet je om op $S_{2n} = 2k$ uit te komen $n + k$ keer kop gooien en $n - k$ keer munt, en de kans daarop is

$$P\left(S_{2n} = 2k\right) = \binom{2n}{n+k}2^{-2n}, \tag{9.2}$$

en voor grote n doet Stirling nu de rest (hopen we).

Met

$$n! \sim (\frac{n}{e})^n\sqrt{2\pi n}$$

als $n \to \infty$ leidt (9.2) tot een formule waaraan we hopelijk wat kunnen zien. We halen diep adem en rekenen rechtuit door. Let op hoe daarbij de e-tjes en de tweetjes in drie Stirling-formules wegvallen. Voor de kans op $S_{2n} = 2k$ vinden we

$$\begin{aligned} P\left(S_{2n} = 2k\right) &= \binom{2n}{n+k}2^{-2n} = \frac{(2n)!}{(n+k)!(n-k)!}2^{-2n} \\ &\sim \frac{(2n)^{2n}2^{-2n}}{(n+k)^{n+k}(n-k)^{n-k}}\frac{\sqrt{2\pi(2n)}}{\sqrt{2\pi(n+k)}\sqrt{2\pi(n-k)}} \\ &= \left(\frac{n}{n+k}\right)^{n+k+\frac{1}{2}}\left(\frac{n}{n-k}\right)^{n-k+\frac{1}{2}}\frac{1}{\sqrt{\pi n}} \end{aligned} \tag{9.3}$$

voor $n \to \infty$.

Nu is het moment daar om te kijken wat er gebeurt als we n naar oneindig sturen. Voor vaste k zien we dat de grondtallen in de eerste twee factoren van (9.3) naar 1 gaan, en daarmee verdwijnen in de limiet de bijdragen van k en $\frac{1}{2}$ in de exponenten. Het gaat voor vaste k dus alleen maar om de n in beide exponenten. Enigzins vermomd verschijnt hier onze standaardlimiet voor de e-macht. Met

$$\lim_{n\to\infty} \underbrace{\left(\frac{n+x}{n}\right)^n}_{e_n(x)} = e^x \tag{9.4}$$

als equivalente schrijfwijze herkennen we met de exponent n in de eerste factor $e_n(k)$, en in de tweede $e_n(-k)$, allebei op hun kop. In de limiet geeft dat $1 = e^k e^{-k}$ op zijn kop. Maar dat is ook 1 zodat de derde factor voor elke vaste k de kans op $S_{2n} = 2k$ naar 0 trekt. We moeten dus wel concluderen dat eindige k's geen kans maken in de limiet $n \to \infty$.

Merk op dat de uitkomsten voor S_{2n} lopen van $-2n$ tot en met $+2n$. Als we niet schalen houden we voor vaste k geen kansen over, en als we schalen met $2n$ zelf dan gaat met kans 1 alles naar 0 als we de wet van de grote aantallen mogen geloven die we in hoofdstuk 2 hebben besproken. Tussen deze twee uitersten moeten we het dus zoeken.

De situatie is nu vrijwel identiek met wat we zagen in de afleiding van de formule van Stirling. Kijk nog eens naar Opgave 8.5. De schaling van x met n volgde uit de berekeningen, maar bleek nog niet optimaal omdat de integrand in (8.4) zich daarbij verspreidde over de hele reële rechte en in de limiet overal gelijk werd aan 1. Uiteindelijk bleek uit de lokale analyse dat we x beter met $\sqrt{n}$ konden schalen met als resultaat de schitterende formule (8.7).

In onze analyse van S_{2n} zien we nu vrijwel hetzelfde! De eerste schaling van S_{2n} met n was niet goed omdat alle kans zich dan in 0 concentreert, precies het tegenovergestelde[2] van wat we bij ψ_n zagen. En schalen we S_{2n} niet, en dus k (eigenlijk $2k$) ook niet, dan verspreiden de uitkomsten zich over alle even gehele getallen en gaan in de limiet alle kansen naar 0.

Dus hoe zouden we k moeten schalen? Met de wortel uit n soms? Dan moeten we $k = s\sqrt{n}$ nemen met s als nieuwe (discrete) variabele. Voor we dat doen kijken we nog even naar de formules die we hier in deze sectie voor onze neus hebben. In (9.3) ging het om de bijdrage van de n in de exponenten, de rest deed er niet toe, dus wat moeten we met

$$\left(\frac{n}{n+k}\right)^n \left(\frac{n}{n-k}\right)^n = \left(\frac{n^2}{n^2-k^2}\right)^n = \cdots ?$$

Voor vaste k ging dit naar 1 vanwege (9.4) en daar hadden we niks[3] aan,

[2] Belletje: dit doet ons aan Heisenberg denken, maar dat terzijde.

[3] Merk op: $\ln 1$ is niks.

maar met $k = s\sqrt{n}$ gaat het zo verder:

$$\cdots = \left(\frac{n^2}{n^2 - s^2 n}\right)^n = \left(\frac{n}{n - s^2}\right)^n \to e^{s^2},$$

Een mooie *educated guess*, maar nog niet helemaal perfect. Waarom? Wel, eigenlijk omdat Ronald zo graag die $\frac{1}{2}$ voor het kwadraat in de exponent van de normale verdeling wil. Na Opgave 8.5 zagen we met een natuurlijke schaling die $\frac{1}{2}$ al verschijnen in (8.7), en daar moet ook de kansrekening nu aan geloven. Met de cosmetisch net iets fijnere schaling

$$k = s\sqrt{\frac{n}{2}}$$

verschijnt in de limiet

$$e^{\frac{1}{2}s^2}$$

de zo graag geziene $\frac{1}{2}$.

Klaar? Nee, er ontbreekt nog een min... Voordat dit tot imaginaire toestanden leidt kijken we voor de zekerheid ook naar de bijdragen van de k in de exponenten in (9.3). Die is nu

$$\left(\frac{n}{n+k}\right)^k \left(\frac{n}{n-k}\right)^{-k} = \left(\frac{n}{n + s\sqrt{\frac{n}{2}}}\right)^{s\sqrt{\frac{n}{2}}} \left(\frac{n}{n - s\sqrt{\frac{n}{2}}}\right)^{-s\sqrt{\frac{n}{2}}}$$

$$= \left(\frac{\sqrt{\frac{n}{2}}}{\sqrt{\frac{n}{2}} + \frac{s}{2}}\right)^{s\sqrt{\frac{n}{2}}} \left(\frac{\sqrt{\frac{n}{2}}}{\sqrt{\frac{n}{2}} - \frac{s}{2}}\right)^{-s\sqrt{\frac{n}{2}}} =$$

$$\left(\left(\frac{\sqrt{\frac{n}{2}}}{\sqrt{\frac{n}{2}} + \frac{s}{2}}\right)^{\sqrt{\frac{n}{2}}} \left(\frac{\sqrt{\frac{n}{2}}}{\sqrt{\frac{n}{2}} - \frac{s}{2}}\right)^{-\sqrt{\frac{n}{2}}}\right)^s \to \left(e^{-\frac{s}{2}} e^{-\frac{s}{2}}\right)^s = \left(e^{-s}\right)^s = e^{-s^2},$$

waarbij we in de één na laatste stap heel sneaky (9.4) gebruikt hebben, met $\sqrt{\frac{n}{2}}$ in plaats[4] van n, en $\pm\frac{s}{2}$ in plaats van x, en in de laatste stap een rekenregeltje[5] voor machten van machten. De factoren in (9.3) die eerder geen rol speelden, doen nu wel mee! Want de exponent k groeit nu mee met n, langzamer weliswaar, maar hard genoeg.

Samen met de bijdrage van de n's zien we nu dat de factor voor de wortel in (9.3) naar

$$e^{-\frac{1}{2}s^2}$$

[4] In (9.4) zie je nu wel dat de monotonie in n ook voor reële n waar is, toch?

[5] Waarvan je jezelf even moet overtuigen.

gaat, en alles bij elkaar geeft dit dus

$$P(\frac{S_{2n}}{\sqrt{2n}} = s) \sim \frac{1}{\sqrt{\pi n}} e^{-\frac{1}{2}s^2} = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}}_{\text{kansdichtheid}} \sqrt{\frac{2}{n}}, \tag{9.5}$$

met links in (9.5) de kans op de discrete s-waarden die onze met $\sqrt{2n}$ geschaalde grootheid S_{2n} kan aannemen. Die s-waarden zijn

$$s = k\sqrt{\frac{2}{n}},$$

met steeds kleinere stapjes ter grootte $\sqrt{\frac{2}{n}}$, in totaal $2n + 1$ waarden, die steeds dichter bij elkaar komen te liggen, maar die tegelijkertijd ook een steeds groter s-interval, van $-\sqrt{2n}$ tot $+\sqrt{2n}$, doorlopen. In de limiet zien we hier dus een continue stochastische grootheid verschijnen met een kansdichtheid[6], zoals al aangegeven in (9.5).

9.2 Kansdichtheden en de centrale limietstelling

Kansdichtheid, een nieuw woord. Hoe komen we van discrete kansen op kansdichtheden? Ongeveer net zoals we van benaderende sommen naar integralen gingen. Bij onze met $\sqrt{2n}$ geschaalde stochastische grootheid S_{2n} kun je voor elke vaste n een staafdiagram maken met blokstaafjes die als oppervlakte $P(\frac{S_{2n}}{\sqrt{2n}} = s)$ hebben, $\sqrt{\frac{2}{n}}$ breed zijn en gecentreerd in s, waarbij s de genoemde $2n + 1$ waarden doorloopt.

Het hoogste staafje heeft hoogte

$$\sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = 0\right)$$

en zit in het midden, rustend op het s-interval

$$\left[-\sqrt{\frac{1}{2n}}, \sqrt{\frac{1}{2n}}\right].$$

Direct links en rechts van dit blokstaafje staan twee blokstaafjes met hoogte

$$\sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = -\sqrt{\frac{2}{n}}\right) = \sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = \sqrt{\frac{2}{n}}\right),$$

[6] Zie ook (5.7), waarin we met de kennis van nu...

en direct links en rechts daarvan staan weer twee blokstaafjes met hoogte

$$\sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = -2\sqrt{\frac{2}{n}}\right) = \sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = 2\sqrt{\frac{2}{n}}\right),$$

enzovoorts.

Alles is keurig symmetrisch rond $s = 0$. Naar buiten lopend zien we dat het blokstaafje bij $\pm(k+1)$ als hoogte

$$\frac{n-k}{n+k+1}$$

keer de hoogte van het blokstaafje bij $\pm k$ heeft, een factor die steeds dichter bij 1 zit naarmate n groter wordt. De hoogtes van de blokjes veranderen voor grotere n dus steeds geleidelijker en zijn gelijk aan

$$\underbrace{\sqrt{\frac{n}{2}}\, P\left(\frac{S_{2n}}{\sqrt{2n}} = \underbrace{k\sqrt{\frac{2}{n}}}_{s}\right)}_{\sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}}, \quad k = -n, \ldots, n,$$

waarin we de rechterkant van (9.5) gebruikt hebben. Dit is het resultaat van onze berekening uitgaande van $P\left(S_{2n} = 2k\right)$ en de formule van Stirling in de geschaalde variabele s.

De upshot van dit alles is dat voor grote n de bovenkant van het staafdiagram steeds dichter bij de grafiek van de functie

$$s \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}$$

ligt en dat is de kansdichtheid van de standaard normale verdeling. Voor $a < b$ is de kans op een uitkomst S_{2n} tussen $a\sqrt{2n}$ en $b\sqrt{2n}$ in de limiet $n \to \infty$ simpelweg gelijk aan de integraal

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}\, ds,$$

en kennelijk is S_{2n} dus ‘typisch van de orde’ $\sqrt{2n}$.

De analyse voor een oneven aantal worpen geeft natuurlijk in de limiet precies hetzelfde. Zowel voor even als oneven n is het verschil tussen het aantal keer kop en aantal keer munt na n keer gooien van de orde $\sqrt{n}$, een conclusie die vrijwel rechtstreeks uit de formule van Stirling volgt in de berekeningen hierboven. De volgende stelling formuleert dit precies.

Centrale limietstelling (CLS). Voor elke $a < b$ geldt dat

$$P\left(a \leq \frac{S_n}{\sqrt{n}} \leq b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx,$$

als $n \to \infty$.

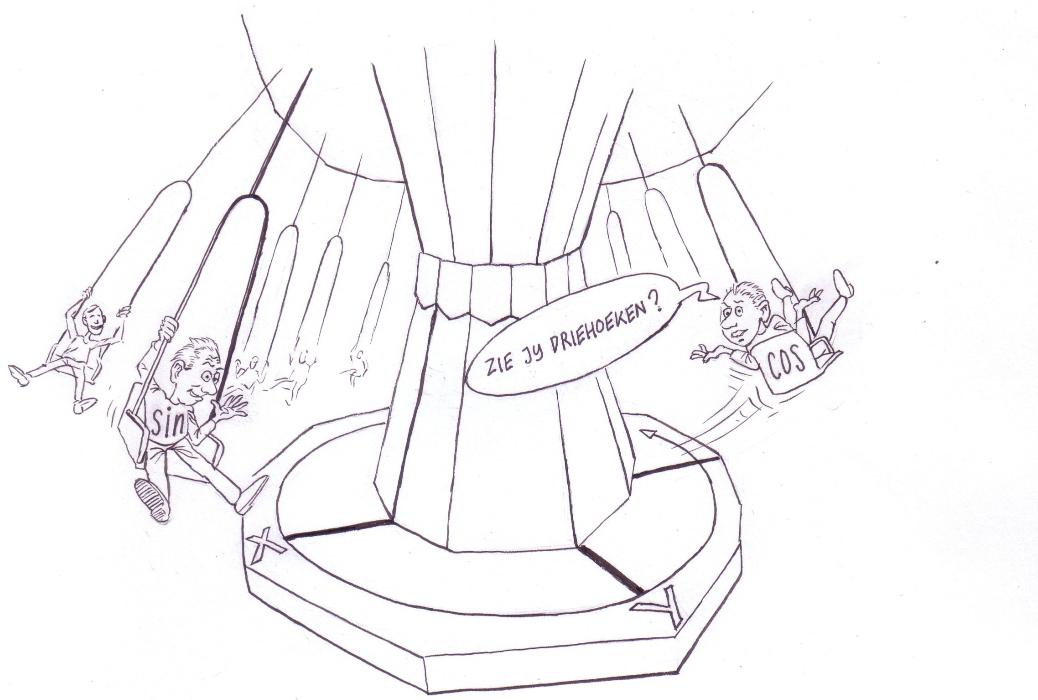
De functie die geïntegreerd wordt kenden we al uit het vorige hoofdstuk, met ook in de kansrekening de factor $\frac{1}{2}$ in de exponent. De grafiek van deze functie wordt de Gauss kromme genoemd[7], en staat bekend als de Bell[8] Curve. Niet alle belvormige curves zijn Gauss krommen natuurlijk[9]. En niet alleen in dit speciale geval verschijnt deze speciale belkromme. De CLS die we hier hebben afgeleid en (bijna) bewezen kan veel algemener geformuleerd en bewezen worden.

In alle situaties waarin we onafhankelijke stochastische grootheden, die min of meer van dezelfde ordegrootte zijn, bij elkaar optellen, verschijnt na de schaling met het juiste veelvoud van $\sqrt{n}$ de universele normale kansdichtheid met de factor $\frac{1}{2}$ in de exponent. Zelfs de aanname dat alle stochastische grootheden onafhankelijk zijn kan nog afgezwakt worden. De bewijzen worden wat anders en diverser, dat wel[10].

Natuurlijk is de (CLS) een bron voor leuke opgaven. Hier is er eentje. We houden ons in. Een getallenvoorbeeldje zoals ze wel vaker bedacht worden voor een toets.

Opgave 9.1. Bertje gooit duizend maal met een munt waarvan hij zegt dat deze zuiver is. Van deze duizend worpen blijken er zeshonderd kop te zijn. Paultje vertrouwt Bertje niet. Is Paultjes wantrouwen terecht?

[7] Dezelfde belhamel als van de strafwerksom $1 + 2 + \cdots + n$.
[8] Google Pietje.
[9] Niet alle belhamels heten Pietje.
[10] Druk verkeer rond Rome met al die wegen.



10 Polynomen van graad oneindig

We hebben in hoofdstuk 4 de limiet

$$\exp(x) = \lim_{n\to\infty} (1 + \frac{x}{n})^n \tag{10.1}$$

zien verschijnen in de context van rente op rente. Binnen de limiet staat een formule die voor elke positieve geheelwaardige n kan worden uitgerekend in termen van x met behulp van de standaard rekenkundige operaties.

Met $n = 1$ hebben we de x in $1+x$ gezien als de fractie van je spaarinleg bij de bank die je na afloop van de renteperiode van de bank als rente uitbetaald krijgt. Keert de bank de rente meerdere keren per renteperiode uit, zeg n keer, dan is aan het eind van renteperiode je inleg door de bank niet met $1 + x$ maar met $(1 + \frac{x}{n})^n$ vermenigvuldigd. Oneindig rijk zul je niet worden als de bank vervolgens de hypothetische limiet n naar oneindig neemt, maar bepaald meer dan $1 + x$ is het wel.

Bij het nadenken over de vraag of (10.1) wel betekenis had, zijn we beloond met de ontdekking dat $\exp(x)$ ook te schrijven is als een machtreeks, te weten

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots , \tag{10.2}$$

en dit is een veel betere manier om, gegeven x, $\exp(x)$ daadwerkelijk uit te rekenen. Door de faculteiten in de noemers worden de termen zo snel klein dat voor elke x deze som met oneindig veel termen niet alleen bestaat, maar ook met grote nauwkeurigheid is te benaderen met relatief weinig termen, als je x niet te groot neemt[1].

Waarom is deze functie exp zo fundamenteel? Als knop zit de functie op elke rekenmachine, en natuurlijk ook op onze DRM. Onder de knop van onze DRM zit bij voorkeur de representatie van $\exp(x)$ als machtreeks in (10.2), maar $\exp(x)$ is met (10.1) als $n \to \infty$ limiet toch veel elementairder en realistischer gedefinieerd via de rente-op-renteformule

$$e_n(x) = (1 + \frac{x}{n})^n.$$

We hebben gezien dat de grafiek van de door $e_n(x)$ gedefinieerde functie e_n een verschoven en geschaalde versie is van de functie gedefinieerd door de formule x^n. We hebben ook gezien, in hoofstuk 3, dat de afgeleide van x^n gelijk is aan nx^{n-1}, uitgaande van de definitie van de afgeleide aan de hand van de raaklijk aan de grafiek in een gegeven punt, zeg het punt met x-coördinaat $x = a$. Via schuiven en schalen zou de volgende opgave daarom te doen moeten zijn zonder gebruik te maken van de gebruikelijke rekenregels uit de differentiaalrekening.

Opgave 10.1. Laat rechtstreeks vanuit de definities met verschuivingen en schalingen zien dat de afgeleide van $e_n(x)$ wordt gegeven door

$$e_n'(x) = (1 + \frac{x}{n})^{n-1} = \frac{e_n(x)}{1 + \frac{x}{n}}.$$

Als je stiekem wel de rekenregels heb gebruikt kwam je natuurlijk meteen tot deze conclusie. Deze gelijkheid in Opgave 10.1 moet gevolgen hebben voor de limietfunctie $\exp(x)$. Als we differentiëren en limiet nemen mogen verwisselen dan kan het niet anders zijn dan dat de functie exp differentieerbaar is met

[1] Hoe groter x hoe meer termen je nodig hebt.

$\exp'(x) = \exp(x)$. De $\frac{x}{n}$ in de noemer gaat immers naar 0 als $n \to \infty$ en $e_n(x)$ gaat naar $\exp(x)$.

Mogen is een rotwoord. Nu we het vermoeden hebben dat $\exp'(x) = \exp(x)$ differentieerbaar is[2], is de leukste en misschien wel leerzaamste manier om dat te begrijpen via de machtreeks in (10.2). Algemene beschouwingen over het differentiëren van machtreeksen zijn daarom nu de volgende stap en we weten al dat er bijzondere dingen gaan gebeuren. Want sommige functies reproduceren zichzelf bij het nemen van afgeleiden. Je bent vast bekend met $\sin'(x) = \cos(x)$ en $\sin''(x) = \cos'(x) = -\sin(x)$. De goniometrische functies cos en sin lijken dus kennelijk op exp. Machtreeksen maken de gelijkenis compleet en nog verassender.

10.1 Differentiaalrekening met puntjes

Met een beetje goede wil kun je de machtreeks

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

zien als een polynoom van graad oneindig, en vervolgens rekenen met dit soort 'polynomen' alsof het echte polynomen zijn. Net als in polynomen kunnen we in zulke machtreeksen getallen (x-waarden) invullen, maar ook, en dat gaan we zien in dit hoofdstuk, kunnen we machtreeksen term voor term differentiëren en zo de afgeleide functie bepalen.

Het nemen van de afgeleide is een van de belangrijkste voorbeelden van wat je kunt en mag[3] doen met machtreeksen, als x maar echt binnen in het definitiegebied zit: de waarden van x waarvoor er iets uit de machtreeks komt als je x invult. Voor de machtreeks hierboven is dat definitiegebied de hele verzameling van reële getallen, maar bij machtreeksen met coëfficiënten die minder snel (of niet[4]) klein worden kan dat definitiegebied een begrensd interval $(-R, R)$ zijn, met 0 precies in het midden. Voor $|x| < R$ gaan de termen hard naar 0, hard genoeg om alle termen bij elkaar op te kunnen tellen, maar voor $|x| > R$ worden de termen groot en gaat alles mis. De kritieke waarde van R wordt bepaald door de coëfficiënten van de machtreeks. Alle waarden van R komen voor, ook $R = 0$ en $R = \infty$.

Opgave 10.2. Kun je machtreeksen bedenken met $R = 1$? En met $R = 0$?

[2] Dit zijn twee uitspraken waaruit volgt dat $\exp''(x) = \exp(x)$, et cetera.
[3] In de zin van dat wat je denkt te krijgen dan ook goed is.
[4] Kun je een machtreeks verzinnen die alleen voor $x = 0$ een uitkomst geeft?

Op de rand van het interval $(-R, R)$ is niet meer te verwachten dat alles zo maar mag, zelfs als de rand in het definitiegebied ligt. De afgeleide functie in een punt $x = a$ wordt immers gedefinieerd aan de hand van inputwaarden voor x in de buurt van dat punt, en op de rand leidt dat onvermijdelijk tot problemen.

Misschien is het goed om ten overvloede nog maar eens te benadrukken dat we het vaste punt a waarin we de afgeleide bepalen, steeds ook variabel maken en dan weer met x aanduiden. De variabele x varieerde zo eerst in de buurt van een a die vervolgens weer x heet. De oude x is dus de nieuwe x plus een klein beetje: $x + h$. Verwarrend? Wellicht, dus neem even de tijd om je verwarring te ontwarren. Hieronder werken we soms met x en $x + h$ in plaats van met a en x, maar blader nog even terug naar hoofstuk 3 om te zien hoe met a en x de staartdeling werd gebruikt om de afgeleide van x^n in $x = a$ te bepalen als na^{n-1}. En daarmee nx^{n-1} als de afgeleide van x^n, het uitgangspunt voor het vervolg hieronder.

Als je in het rechterlid in (10.2) term voor term differentieert, dan komt er hetzelfde uit als wat er stond, probeer het maar. Als dat term voor term differentiëren mag[5] dan is $\exp(x)$ kennelijk zijn eigen afgeleide. Het is deze eigenschap die de functie $\exp(x)$ zo uniek maakt, nog voordat $\exp(x)$ is omgedoopt tot e^x. De functie exp is de enige door een machtreeks gedefinieerde functie die in $x = 0$ gelijk is aan 1 en die zijn eigen afgeleide is[6].

Dat $\exp(x) = e^x$ differentieerbaar is en gelijk is aan zijn eigen afgeleide hebben we al aangekondigd in sectie 8.2, waarbij we alvast net deden of je de machtreeks term voor term mag differentiëren. Ook via het gebruikelijke differentiequotiënt

$$\frac{e^{x+h} - e^x}{h}$$

en de limietovergang $h \to 0$ kun je wel aan zien komen dat e^x zijn eigen afgeleide is. Via de multiplicatieve eigenschap is dit quotiënt immers een product van twee factoren:

$$\frac{e^{x+h} - e^x}{h} = e^x \times \frac{e^h - 1}{h}.$$

De laatste factor in het rechterlid is zelf weer te schrijven als een machtreeks, namelijk als

$$\frac{e^h - 1}{h} = 1 + \frac{h}{2!} + \frac{h^2}{3!} + \cdots$$

Als je $h = 0$ invult krijg je gewoon 1, dus moet de afgeleide van e^x wel e^x zijn. Na het uitdelen van de factor h in het differentiequotiënt is de limiet

[5] Dat rotwoord weer.

[6] Dat zo'n functie alleen maar een machtreeks kan zijn laten we even liggen.

nemen immers een kwestie van invullen[7], net als bij het differentiëren van x^n.

We willen limieten echter zo veel mogelijk vermijden, zelfs bij limieten die uit te rekenen zijn door invullen. Om te begrijpen dat we met machtreeksen kunnen rekenen zoals met polynomen bekijken we daarom eerst de componenten x^n waaruit een machtreeks is opgebouwd. Vervolgens maken we de stap van deze monomen naar polynomen op zo'n manier dat de stap van monomen naar machtreeksen nauwelijks anders is!

De basisgedachte blijft dat differentiëren eigenlijk hetzelfde is als lineair benaderen, en dat dat laatste makkelijk is voor polynomen als je goed kan staartdelen of anderzins handig bent met algebra. Via het herschreven differentiequotiënt

$$\frac{x^7 - a^7}{x - a} = x^6 + x^5 a + x^4 a^2 + x^3 a^3 + x^2 a^4 + x a^5 + a^6$$

moet door invullen van $x = a$ de afgeleide van x^7 in $x = a$ gelijk zijn aan $7a^6$. Omdat a willekeurig was, is de afgeleide functie van x^7 zo onvermijdelijk (goed) gedefinieerd en gelijk aan $7x^6$.

Het voorbeeld, niet te makkelijk en niet te moeilijk, laat zien hoe het zit. Als je snapt hoe x^7 moet, dan snap je ze allemaal. Het differentiëren van polynomen als

$$P(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5$$

is dus makkelijk. De exponenten komen steeds naar voren en op de plek waar ze stonden gaan ze eentje omlaag. Automatiseren maar[8].

In het natuurlijk niet zomaar gekozen voorbeeld krijgen we

$$P'(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4,$$

hetgeen roept om de overgang naar differentiaalrekening voor functies die eigenlijk machtreeksen zijn, zoals de hele bijzondere functie gedefinieerd door

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots .$$

Het hele punt van dit hoofdstuk is dat wat we in sectie 3.2 hebben gedaan toen we de methode van differentiëren via lineaire benaderingen precies maakten door in voorbeeldjes de fout in detail te bekijken en af te schatten, net zo

[7] De lezer begrijpt dat we meer van invullen houden dan van limieten.

[8] Automatiseren kun je doen voor- of nadat het begrip komt. We parkeren dat begrip!

moeilijk[9] is voor polynomen als voor machtreeksen. Een beetje systematiek in wat we doen is daartoe voldoende[10].

We beginnen weer met het voorbeeld en scherpen de identiteit voor het differentiequotiënt eerst nog wat aan door middel van

$$\frac{x^7 - a^7}{x - a} - 7a^6 = (x - a)(x^5 + 2x^4a + 3x^3a^2 + 4x^2a^3 + 5xa^4 + 6a^5). \quad (10.3)$$

Opgave 10.3. Leid eerst een vergelijkbare uitdrukking af voor x^3 en overtuig jezelf van de gelijkheid in (10.3), bij voorkeur met behulp van een staartdeling.

Een andere manier om (10.3) te schrijven is als

$$x^7 = \underbrace{a^7 + 7a^6(x - a)}_{\text{lineaire benadering}} + \underbrace{(x^5 + 2x^4a + 3x^3a^2 + 4x^2a^3 + 5xa^4 + 6a^5)(x - a)^2}_{\text{restterm}},$$

ofwel

$$x^7 = a^7 + 7a^6(x - a) + R_a(x).$$

De lineaire benadering van x^7 rond $x = a$ staat er nu al. De $7a^6$ lezen we af als de (richtings)coëfficiënt in de tweede term. De enige vraag is of de restterm zich goed gedraagt en die is gelijk aan

$$R_a(x) = \underbrace{(x^5 + 2x^4a + 3x^3a^2 + 4x^2a^3 + 5xa^4 + 6a^5)}_{\substack{x=a \text{ geeft} \\ \underbrace{(1+2+3+4+5+6+7)a^5}_{\frac{7\cdot 6}{2}a^5}}}(x - a)^2.$$

De voorfactor in deze restterm reduceert met Gauss' strafwerksom tot $\frac{7\cdot 6}{2}a^5$ als je $x = a$ invult. Ook mooi. Belangrijker echter nu is de vraag hoe groot $R_a(x)$ kan zijn voor x in de buurt van a en om dat uit te zoeken kiezen we een vast interval $[-\rho, \rho]$ waarbinnen we $R_a(x)$ bekijken. Voor alle x en a in $[-\rho, \rho]$ geldt nu dat

$$|R_a(x)| \leq \frac{7 \cdot 6}{2}\rho^5\,(x - a)^2,$$

simpelweg omdat de absolute waarde van een som op zijn hoogst gelijk is aan de som van de absolute waarden[11].

Dat kan natuurlijk weer algemener. Vul in de volgende opgave eerst nog even $n = 7$ in om het verband met het voorgaande te zien.

[9] Of makkelijk, maar niet -ker.
[10] Lees: behandel alle x^n tegelijk.
[11] Waarom is dat zo? Denk aan de driehoeksongelijkheid.

Opgave 10.4. Voor elke gehele positieve n is de lineaire benadering van x^n rond $x = a$ van dezelfde vorm. Er geldt dat

$$x^n = a^n + na^{n-1}(x-a) + R_a(x)$$

met

$$|R_a(x)| \leq \frac{n(n-1)}{2}\rho^{n-2}\,(x-a)^2$$

voor x en a in $[-\rho, \rho]$. Laat dit zien, eerst voor $n = 3, 4, \ldots$, en daarna ook voor $n = 2, 1$ en zelfs (als $a \neq 0$ James) voor $n = 0$.

Wat voor monomen x^n kan, kan ook voor polynomen. Als

$$P_n(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n = \sum_{k=0}^{n} \alpha_k x^k$$

een n-de graadspolynoom is, dan is de afgeleide van P_n in $x = a$ gelijk aan

$$P_n'(a) = \alpha_1 + 2\alpha_2 a + \cdots + n\alpha_n a^{n-1} = \sum_{k=1}^{n} k\alpha_k a^{k-1}.$$

Dat wisten we al zo'n beetje, maar nog niet met een expliciete schatting voor het verschil tussen differentiequotiënt en limiet, hetgeen nu wel kan als een opgave:

Opgave 10.5. Laat net als hierboven zien dat voor x en a in $[-\rho, \rho]$ geldt dat

$$P_n(x) = P_n(a) + P_n'(a)(x-a) + R_a(x)$$

met

$$|R_a(x)| \leq (x-a)^2 \sum_{k=2}^{n} \frac{k(k-1)}{2}|\alpha_k|\rho^{k-2}.$$

Een klein stapje voor een tekstzetter, maar een grote sprong voorwaarts in de wiskunde is dat voor

$$P(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots = \sum_{k=0}^{\infty} \alpha_k x^k,$$

met

$$P'(a) = \alpha_1 + 2\alpha_2 a + \cdots = \sum_{k=1}^{\infty} k\alpha_k a^{k-1},$$

eenzelfde lineaire benadering met bijbehorende foutafschatting geldt als in de zojuist gemaakte Opgave 10.5, maar nu met

$$|R_a(x)| \leq (x-a)^2 \sum_{k=2}^{\infty} \frac{k(k-1)}{2} |\alpha_k| \rho^{k-2}. \tag{10.4}$$

Het enige verschil is dat de som nu tot oneindig loopt in plaats van tot n.

Is dit allemaal wel correct? Het antwoord is: 'niet altijd, maar wel als de som van positieve getallen

$$S_\rho = \sum_{k=2}^{\infty} \frac{k(k-1)}{2} |\alpha_k| \rho^{k-2} = |\alpha_2| + 3|\alpha_3|\rho + 6|\alpha_4|\rho^2 + 10|\alpha_5|\rho^4 + \cdots$$

eindig is'. In dat geval hebben ook de formules voor $P(x)$ en $P'(a)$ betekenis en dankzij de afschatting[12]

$$|R_a(x)| \leq S_\rho (x-a)^2$$

voor x, a in $[-\rho, \rho]$ is $P'(a)$ inderdaad de afgeleide van P in $x = a$. Of P nu een polynoom is of een power series[13], het principe is hetzelfde. Zolang S_ρ eindig is, gaat alles goed.

Opgave 10.6. Leg uit waarom bij de functie

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

elke ρ een eindige S_ρ geeft en dat bijgevolg voor alle x geldt dat

$$\exp'(x) = \exp(x).$$

Nu we de afgeleide van $\exp(x)$ kennen (en begrijpen), is het aardig om nog even terug te grijpen naar hoofdstuk 3. We hebben in dat hoofdstuk de methode van Newton besproken voor het benaderen van nulpunten van een functie f. De bijbehorende iteratiefunctie F had de vorm

$$F(a) = a - \frac{f(a)}{f'(a)},$$

en deze uitdrukking bevat de afgeleide f' van f. Nu we weten wat de afgeleide van $\exp(x)$ is, kunnen we de methode van Newton ook op $\exp(x)$ toepassen.

[12] Dit is gewoon (10.4).

[13] Engels voor machtreeks.

Opgave 10.7. Laat zien dat voor vaste y, de methode van Newton voor de functie $f(x) = \exp(x) - y$ neerkomt op het itereren van de functie F_y gedefinieerd door

$$F_y(x) = x - \frac{\exp(x) - y}{\exp(x)} = x - 1 + y\exp(-x).$$

Controleer dat $F_y'(x) = 1 - y\exp(-x)$ zodat

$$F_y'(x) > 0 \iff y < \exp(x).$$

Start nu met zo'n x en noem die x_0. Deze x_0 is te groot om een oplossing te zijn. De eerst Newtonbenadering $x_1 = F_y(x_0) = x_0 - 1 + y\exp(-x_0) < x_0$ is dan kleiner maar nog steeds te groot, want

$$\exp(x_1) = \exp(x_0 - 1 + y\exp(-x_0)) = \exp(x_0)\exp(\underbrace{y\exp(-x_0) - 1}_{\star}) >$$

$$\exp(x_0)(1 + \underbrace{y\exp(-x_0) - 1}_{\star}) = y.$$

Voor het met een $\star$ gemarkeerde stuk is gebruikt dat $\exp(x) > 1 + x$ voor $x \neq 0$, hetgeen je al weet na hoofdstuk 4. Bijgevolg geeft Newtons methode een dalende rij $x_0 > x_1 > x_2 > \cdots$ die naar beneden begrensd is (waarom?) en dus een grootste ondergrens $\bar{x}$ heeft. Gebruik ten slotte de gelijkheid

$$x_{n+1} = x_n + 1 - y\exp(-x_n)$$

om te concluderen $\exp(x_n)$ als grootste ondergrens y heeft en overtuig jezelf ervan dat $\exp(\bar{x}) = y$.

Zoals gezegd, het zijn de faculteiten in de noemers waardoor alles zo goed gaat met de machtreeks voor $\exp(x)$. Hier is een voorbeeld van een machtreeks die zich alleen voor $|x| < 1$ goed gedraagt. Laat

$$P(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

Als je hier $\rho = 1$ kiest dan zie je dat S_1 niet eindig kan zijn[14], maar elke kleinere ρ geeft wel een eindige S_ρ. En dus geldt voor x tussen -1 en 1 dat

$$P'(x) = 1 - x + x^2 - x^3 + \cdots = \frac{1}{1+x}$$

en die functie hebben we eerder gezien, als afgeleide van $\ln(1 + x)$. Blader maar terug naar sectie 8.2. In $x = 0$ zijn $P(x)$ en $\ln(1 + x)$ allebei gelijk aan 0. Dan moet wel gelden dat $P(x) = \ln(1 + x)$voor x in $(-1, 1)$. Toch?

[14] Wel even doen natuurlijk.

10.2 Analysis now![15]

Toch? Het is hier dat de algebra ophoudt en de analyse begint. Laten we nog even terugkijken op onze insteek tot nu toe. Uitgaande van de rationale getallen hebben we de reële getallen gezien als vanzelfsprekend gedefinieerd door hun al of niet doorlopende decimale ontwikkeling. Gezeur over limieten hebben we zo veel mogelijk vermeden.

Op dezelfde manier hebben we de stap van polynomen naar machtreeksen

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$$

gemaakt, sommen met oneindig veel termen die zo hard naar nul gaan dat alles bij elkaar optellen geen probleem is en je het verschil met polynomen

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N$$

nauwelijks ziet. Sterker, met de informele slordige puntjesnotatie lijkt het juist makkelijker en typografisch is het dat ook. Rekenen met machtreeksen is meestal net zo makkelijk. Bijvoorbeeld:

Opgave 10.8. Neem aan dat de machtreeks $f(x)$ hierboven een positieve R heeft. Bepaal $f(0)$, $f'(0)$, $f''(0)$, en alle hogere-ordeafgeleiden in $x = 0$.

Echter, waar we met de stap van rationale getallen naar reële getallen wel klaar[16] zijn, is dat bij de stap van polynomen naar machtreeksen allerminst het geval. Niet alle functies zijn machtreeksen, hoe graag we dat ook zouden willen. Als we een functie beschrijven door bepaalde eigenschappen die die functie moet hebben, dan kan het zo zijn dat we onmiddellijk een unieke machtreeks herkennen die de gegeven eigenschappen heeft, maar is het niet vanzelfsprekend dat er geen andere functies zijn met dezelfde eigenschap.

De functie $f : (-1, 1) \to \mathbb{R}$ gedefinieerd door $f(x) = P(x) - \ln(1 + x)$ met

$$P(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

heeft bijvoorbeeld de eigenschap dat $f(0) = 0$ and $f'(x) = 0$ voor alle x in $(-1, 1)$. Bijgevolg zijn alle hogere-ordeafgeleiden van $f(x)$ ook nul en de enige machtreeks met deze eigenschap is de machtreeks met alle coëfficiënten gelijk aan 0, kijk nog maar even naar de antwoorden van de opgave die je

[15] Voor de echte liefhebber, met een gejatte titel.

[16] We komen wel nog over de wortel uit -1 te praten.

zojuist gemaakt hebt. Maar dat bewijst nog niet dat elke differentieerbare functie $f : (-1, 1) \to \mathbb{R}$ waarvoor geldt dat $f(0) = 0$ en $f'(x) = 0$ voor alle x in $(-1, 1)$ ook de nulfunctie is.

Hier ligt dus nog werk om te doen. De wiskundige stelling die bewezen moeten worden luidt:

Niet-triviale Stelling. Een functie die differentieerbaar is op een open interval is constant op dat interval als de afgeleide functie overal nul is op dat interval.

Als we vervolgens willen concluderen dat het met deze stelling inderdaad zo is dat voor alle x in $(-1, 1)$ geldt dat

$$f(x) = -\ln(1 + x) + x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots ,$$

de nulfunctie is, dan moet ook wel even precies gemaakt worden waarom $f'(x)$ voor alle x in $(-1, 1)$ überhaupt bestaat met afgeleide identiek gelijk aan nul. We gebruikten daartoe dat

$$\ln'(1 + x) = \frac{1}{1 + x},$$

maar hoe wisten we dat zo zeker?

De functie ln is ingevoerd door middel van de equivalentie

$$y = \exp(x) \iff x = \ln(y)$$

voor $x \in \mathbb{R}$ en $y > 0$. Aan het eind van hoofdstuk 3 hebben we de suggestie gewekt dat dit verband precies te maken is via

$$y = e_n(x) \iff x = n(\sqrt[n]{y} - 1)$$

en de limiet overgang $n \to \infty$, maar hierboven heb je in Opgave 10.7 het verband rechtstreeks gedaan, zonder limietovergangen. Twijfel over exp en ln als elkaars inverse functies is er dus niet meer, en hun grafieken zijn nog steeds hetzelfde. En de raaklijnen aan die grafieken dus ook. Vanwege

$$\frac{dy}{dx} = \exp'(x) = \exp(x) = y$$

volgt daarom dat

$$\ln'(y) = \frac{dx}{dy} = \frac{1}{y}.$$

In hoeverre dit raaklijnenargument wiskundig echt precies is blijft natuurlijk de vraag. Dat het limietargument voor de in n monotone functies $e_n(x)$ niet per se ook een correct limietargument is voor de grafieken is na Opgave 10.7 voor het bestaan van ln van mindere zorg. Maar waar de raaklijnbeschrijving natuurlijk symmetrisch in x en y, is de foutafschatting voor de lineaire benadering is dat zeker niet. Kortom, de correspondentie tussen y en x via functies als e_n en exp vraagt om een stelling die ons van tevoren vertelt hoe het zit, niet alleen met het bestaan van de inverse functie, maar ook met de afgeleide van de inverse functie.

Inverse Functie Stelling. Een functie f die differentieerbaar is op een open interval, met een afgeleide functie die overal positief is op dat interval, heeft een inverse functie g met dezelfde eigenschappen. Als $y = f(x)$ dan is $f'(x)g'(y) = 1$.

Voor beide stellingen is de zogenaamde middelwaarde-eigenschap voor differentieerbare functies cruciaal. Die eigenschap zegt dat de vergelijking

$$f(b) - f(a) = f'(x)(b - a)$$

altijd een oplossing x tussen a en b heeft als f differentieerbaar is op een open interval dat a en b bevat. De naam van de eigenschap is wat vaag. De uitspraak is dat het differentiequotiënt

$$\frac{f(b) - f(a)}{b - a}$$

wordt aangenomen door de afgeleide $f'(x)$ in een x-waarde tussen a en b.

Opgave 10.9. Bewijs de Niet-triviale Stelling uitgaande van de middelwaarde-eigenschap.

Opgave 10.10. Laat zien dat het voldoende is om de uitspraak in de middelwaarde-eigenschap te bewijzen voor het speciale geval dat $f(b) = f(a) = 0$ en dat het dan dus gaat om het vinden van een nulpunt van $f'(x)$.

Het gaat voor het bewijs van de Niet-triviale Stelling dus uiteindelijk om het vinden van nulpunten van afgeleide functies. Vergelijkingen van de vorm $f'(x) = 0$ los je voortdurend op in de wiskunde als je maxima of minima van functies zoekt. Dat differentieerbare functies hun extremen alleen maar op de rand van een interval of in nulpunten van $f'(x)$ kunnen aannemen volgt rechtstreeks uit de definitie van differentieerbaarheid.

Uitspraken over het bestaan van maxima en minima zijn van een andere moeilijkheidsgraad. In de opgave hierboven wil je weten dat als $f(a) = f(b) = 0$ en $f(x)$ tussen a en b ook een positieve waarde aanneemt, dat dan ergens tussen a en b door $f(x)$ ook een maximale positieve waarde wordt aangenomen. En dat is weer een uitspraak over eigenschappen van continue functies.

We hebben het begrip continuïteit lang vermeden, maar nu kan dat niet langer. Een functie heet continu als in elk punt a van het definitiegebied (meestal een interval) geldt dat $|f(x) - f(a)|$ kleiner te maken is dan iedere gewenste foutmarge $\epsilon > 0$ alleen maar door $|x - a|$ kleiner te maken dan een $\delta > 0$ (die dus bij de gegeven $\epsilon > 0$ gevonden moet kunnen worden).

Kortom, $f(x)$ moet $f(a)$ benaderen, en de fout in de benadering moet klein zijn als x dicht genoeg bij a ligt. Net zo klein als je wilt. Neem je ϵ kleiner, dan moet je δ kleiner te kiezen om aan de gewenste nauwkeurigheid voldoen met alle x die dichter dan δ bij a liggen. De belangrijkste uitspraken over continue functies[17] zijn de volgende twee uitspraken die we als één stelling formuleren.

Bereik Stelling. Als $f : [a, b] \to \mathbb{R}$ een continue functie is op een begrensd gesloten interval dan neemt $f(x)$ voor x in (a, b) alle waarden aan tussen $f(a)$ en $f(b)$, en heeft $f(x)$ zowel een maximum M als een minimum m op $[a, b]$. Het bereik van f is het interval $[m, M]$.

De tweede uitspraak in deze stelling is essentieel voor de middelwaarde-eigenschap en de gewenste uitspraak over functies waarvan de afgeleide overal nul is. De derde uitspraak is een gevolg van de eerste twee uitspraken. De eerste uitspraak heet ook wel de tussenwaarde-eigenschap en is belangrijk voor het bestaan van oplossingen van gewone vergelijkingen.

Opgave 10.11. Laat zien dat de vergelijking $\exp(x) = y$ voor elke $y > 0$ een unieke oplossing heeft. Hint: gebruik zowel de Niet-triviale Stelling als de tussenwaarde-eigenschap.

Met deze opgave is de functie ln nu meteen als inverse functie van exp gedefinieerd, zonder dat Opgave 10.7 nog nodig is. Opgave 10.11 is eigenlijk een stukje van het bewijs van de Inverse Functie Stelling, die ons nog extra vertelt dat ln inderdaad differentieerbaar is met de afgeleide die we in inmiddels wel kennen. En in het bijzonder is $\ln(1 + x)$ daarmee differentieerbaar

[17] Differentieerbare functies zijn altijd continu, waarom?

en via de Niet-triviale Stelling gelijk aan de mooie machtreeks $P(x)$, als x maar tussen -1 en 1 in zit[18].

De gewetensvraag nu is wat we hier nog willen en moeten bewijzen van en voor de lezer. Het gaat nog om de Bereik Stelling en een deel van de Inverse Functie Stelling, namelijk de uitspraak over differentieerbaarheid van de inverse functie als eenmaal duidelijk is dat deze bestaat, en de formule voor de afgeleide van de inverse functie als de inverse van de afgeleide van de functie zelf[19].

Opgave 10.12. Neem aan dat g de inverse functie is van de f in de Inverse Functie Stelling. Stel dat 0 in het interval ligt, dat $f(0) = 0$ en dat $f'(0) = 1$. Om het bewijs van de Inverse Functie Stelling af te maken moeten we laten zien dat de inverse functie $g(y)$ goed benaderd wordt door y in de buurt van $y = 0$. Laat zien dat de uitspraak dat $f(0) = 0$ en $f'(0) = 1$ equivalent is met de uitspraak dat er voor elke $\epsilon > 0$ een $\delta > 0$ is zo dat de grafiek van f tussen de twee rechte lijnen

$$y = (1+\epsilon)x \quad \text{en} \quad y = \frac{x}{1+\epsilon}$$

in zit voor $|x| < \delta$. Maak vervolgens een tekening van de situatie. Kun je een $\eta > 0$ vinden in termen van δ en ϵ, zodanig dat de grafiek van de inverse functie voor $|y| < \eta$ in zit tussen

$$x = (1+\epsilon)y \quad \text{en} \quad x = \frac{y}{1+\epsilon}?$$

Zo ja, dan heb je bewezen dat g differentieerbaar is in $y = 0$ en dat $g'(0) = 1$.

Het bijzondere geval $f(0) = 0$ en $f'(0) = 1$ is via schuiven en schalen weer representatief voor het algemene geval, dus zodra we de inverse functie hebben is differentieerbaarheid niet heel moeilijk. Het echte werk zit in de Bereik Stelling. Merk op dat als M en m eenmaal 'gemaakt' zijn, met bijbehorende originelen $\overline{x}$ en $\underline{x}$, elke c tussen m en M een functiewaarde is op grond van de tussenwaarde-eigenschap. Ook de tussenwaarde-eigenschap heeft weer een speciaal geval, namelijk de uitspraak dat $f(a)f(b) < 0$ impliceert dat $f(x)$ een nulpunt heeft tussen a en b, en die uitspraak kunnen we toepassen op $g(x) = f(x) - c$ met $g(\overline{x})g(\underline{x}) < 0$ om alle $c = f(x)$ terug te vinden.

Opgave 10.13. Als $f : [a, b] \to \mathbb{R}$ een continue functie is op een begrensd gesloten interval met $f(a)f(b) < 0$ dan heeft $f(x) = 0$ een oplossing in (a, b). Hint: neem aan

[18] Beter (anders) kijken geeft de gelijkheid ook voor $x = 1$, dus $\ln 2 = \ldots$?

[19] De ene in x de andere in y.

dat $f(a) < 0$, beschouw de kleinste bovengrens s van alle x waarvoor $f(x) < 0$, en laat zien dat zowel $f(s) > 0$ en $f(s) < 0$ tot een tegenspraak leiden.

Je ziet in deze opgave dat we het nulpunt niet echt maken. In het geval dat f een machtreeks is hebben we directere methoden tot onze beschikking, zoals in Opgave 10.7. Ook voor het bestaan van M en m zijn in concrete gevallen wel directe methoden te verzinnen. Maar als van f niets bekend is behalve zijn continuiteit, dan is het een ander verhaal.

Opgave 10.14. Om voor een willekeurige f het bestaan van M te bewijzen nemen we een rij punten x_n in $[a, b]$ met de eigenschap dat $f(x_n)$ zo groot mogelijk wordt als maar kan. Bekijk daartoe de verzameling van alle functiewaarden $f(x)$. Noem de kleinste bovengrens daarvan M. Als dat niet kan dan zeggen we dat $M = \infty$. In beide gevallen kan de rij zo gekozen worden dat $f(x_n) \uparrow M$ als $n \to \infty$. Waarom? Neem nu aan dat de rij x_n convergeert naar een limiet $\overline{x}$ in $[a, b]$. Bewijs dat $f(\overline{x}) = M < \infty$ het maximum is van $f(x)$ op $[a, b]$. Hint: gebruik de definitie van continuïteit van f in $\overline{x}$ en de definitie van convergente rijen hieronder.

Definitie. Een rij getallen x_n met $n = 1, 2, 3, 4, \ldots$ heet convergent als er een limiet $\overline{x}$ bestaat waarvoor geldt dat bij elke foutmarge[20] $\delta > 0$ een index N te vinden is zo dat voor alle $n \geq N$ geldt dat $|x_n - \overline{x}| < \delta$.

Kortom, als de rij x_n in de laatste opgave convergent gekozen kan worden met limiet $\overline{x}$ in $[a, b]$ dan is het bewijs van de Bereik Stelling rond. We besluiten daartoe met het volgende feit: iedere rij in $[0, 1]$ kan convergent gemaakt worden met limiet in $[0, 1]$ door maar genoeg waarden weg te laten.

Een bewijs? Misschien leuk om zelf te proberen. Je kunt de rij ook aan een fysicus geven en hem of haar vertellen dat het meetwaarden zijn.

10.3 cos, sin, exp en de wortel uit -1

Hopelijk vinden we op dit punt[21] al onze lezers weer terug. In deze sectie wordt het weer concreet en hopelijk spannend. We hebben gezien dat we een machtreeks mogen differentiëren als we maar binnen het definitiegebied van de machtreeks blijven. Grosso modo gelden alle rekenregels voor polynomen ook voor machtreeksen. Bijvoorbeeld, als $f(x)$ en $g(x)$ allebei machtreeksen zijn, dan is $f(x)g(x)$ gedefinieerd als x in allebei de definitiegebieden zit en

[20] Nu even niet ϵ genoemd, met goede reden!

[21] The headaches are all gone and it's morning in this song.

voor die waarden van x is $f(x)g(x)$ zelf ook weer een machtreeks, die je kunt uitrekenen door uitvermenigvuldigen van de machtreeksen, net zoals je dat bij polynomen doet. De afgeleide van $f(x)g(x)$ is wat je verwacht:

Opgave 10.15. Laat voor polynomen zien dat de afgeleide van $f(x)g(x)$ gelijk is aan $f'(x)g(x) + f(x)g'(x)$. Overtuig je ervan dat ook voor machtreeksen deze regel geldig is, zolang x maar echt binnen beide definitiegebieden ligt.

De algebra voor vermenigvuldigen en differentiëren van machtreeksen is zo hetzelfde als die voor polynomen, als je maar weet wat alle coëfficiënten zijn van de machtreeks, bijvoorbeeld via een suggestieve schrijfwijze met puntjes zoals bij de machtreeks voor $\exp(x)$. Twee andere voorbeelden van dit type zijn

$$c(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

en

$$s(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots ,$$

voorbeelden die we nu bewust uit de lucht laten vallen, met coëfficiënten die net zo hard naar 0 gaan als de coëfficiënten in de machtreeks voor $\exp(x)$. Dat ook deze twee machtreeksen voor elke x een $c(x)$ en $s(x)$ definiëren is dus niet verrassend.

Je ziet natuurlijk dat $c(0) = 1$ en $s(0) = 0$. Maar ook zien we, omdat we term voor term mogen differentiëren, dat de functies c en s voldoen aan

$$c'(x) = -s(x) \quad \text{en} \quad s'(x) = c(x)$$

Dit zijn gelijkheden van functies en hun afgeleiden. Zulke vergelijkingen heten differentiaalvergelijkingen. De functies c en s voldoen aan de gekoppelde vergelijkingen

$$c' = -s \quad \text{en} \quad s' = c. \tag{10.5}$$

De belangrijkste reden om (x) weg te laten is dat we zo meteen liever een andere letter in plaats van x gebruiken. De notatie is ook wat cleaner. En consistent met de wiskundige opvatting dat de functie, in dit geval c of s, aan een variabele, hier x, de waarde $c(x)$ of $s(x)$ toevoegt. Maar hoe netjes we ook proberen te zijn, het blijft verleidelijk en soms ook gewoon praktisch om over de functie $c(x)$ te praten. Of $c(t)$ als we de variabele liever t noemen.

Welke notatie we ook kiezen, komt dit alles je misschien bekend voor[22]? Inderdaad, deze eigenschappen van c en s komen overeen met de veelal niet

[22] Even nadenken voor je verder leest.

bewezen eigenschappen dat

$$\cos' = -\sin \quad \text{en} \quad \sin' = \cos,$$

en natuurlijk ook met $\cos(0) = 1$ en $\sin(0) = 0$. Zou het kunnen dat de functies c en cos dezelfde functies zijn? En dat dan dus ook $s = -c' = -\cos = \sin$? Vast wel, anders zouden we dit voorzetje niet geven.

Alvorens hier nader op in te gaan is het aardig om voor c en s wat van de bekende[23] goniometrische identiteiten te controleren.

Opgave 10.16. Bereken met zo min mogelijk rekenwerk de machtreeks voor de afgeleide van $c(x)c(x) + s(x)s(x)$[24]. Wat is je conclusie voor $c(x)^2 + s(x)^2$?

Gezien deze opgave ligt het voor de hand om aan de eenheidscirkel in het vlak te denken, en dat vlak beschrijven we bij voorkeur met x en y. De x in $\cos(x)$ en $\sin(x)$ vervangen we daarom ook door t, en we denken aan t als aan tijd[25]. Dat met $x = \cos(t)$ en $y = \sin(t)$ een parametervoorstelling van de cirkel gegeven door $x^2 + y^2 = 1$ wordt gegeven zal bekend zijn. De opgave vertelt ons dat ook $x = c(t)$ en $y = s(t)$ op die cirkel liggen. Voor $t = 0$ hebben we in beide gevallen het punt $(x, y) = (1, 0)$.

Hoe kunnen we nu inzien dat $c(t) = \cos(t)$ en $s(t) = \sin(t)$? Als we ervan overtuigd zijn[26] dat ook $x = x(t) = \cos(t)$ en $y = y(t) = \sin(t)$ inderdaad voldoen aan

$$x'(t) = -y(t),\ y'(t) = x(t),\ x(0) = 1,\ y(0) = 0, \tag{10.6}$$

net als $x = x(t) = c(t)$ en $y = y(t) = s(t)$, dan is het voldoende om te laten zien dat er precies één paar functies x, y is dat voldoet aan (10.6).

Daartoe kijken we eerst naar een eenvoudigere vraag: welke functies $x = x(t)$ voldoen aan $x'(t) = x(t)$? Om deze vraag te beantwoorden gebruiken we een bekende truc. Voor $x(t)$ moet gelden dat $x'(t) - x(t) = 0$ en dus ook dat $e^{-t}(x'(t) - x(t)) = 0$. Die factor $\exp(-t) = e^{-t}$ hebben we erbij gehaald omdat er nu de afgeleide staat van $e^{-t}x(t)$ naar t.

De afgeleide van $e^{-t}x(t)$ is dus 0, maar dan moet de functie $t \to e^{-t}x(t)$ wel constant zijn, en we concluderen dat $e^{-t}x(t) = c$, ofwel $x(t) = ce^t$. De constante c kan dan bepaald worden door bijvoorbeeld $t = 0$ in te vullen. Op deze manier vinden we dus alle oplossingen. En gegeven $x(0)$ is er precies één oplossing.

[23] Ken je ze? En hoe weet je ze?

[24] Gebruik de productregel. We schrijven expres niet $c(x)^2 + s(x)^2 = 1$.

[25] Nadat hierboven de fysica al zo behulpzaam was.

[26] Hoe eigenlijk?

Opgave 10.17. Los op dezelfde manier de vergelijking $x'(t) = ax(t)$ op voor gegeven a in $\mathbb{R}$. Hint: vermenigvuldig met $\exp(-at)$.

Deze truc met een zogenaamde integrerende factor werkt algemener. Met een iets geavanceerdere invulling van dit idee kun je ook laten zien dat er inderdaad maar precies één paar functies is dat voldoet aan (10.6). Voor de liefhebber(s) geven we de details daarvan aan het einde van dit hoofdstuk in sectie 10.4, maar nu vervolgen we het grotere verhaal met de onvermijdelijke conclusie dat (na t weer vervangen te hebben door x) de volgende machtreeksontwikkelingen voor elke x correct zijn:

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots ;$$

en

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots ,$$

de goniometrische functies cos en sin, gedefinieerd via meetkunde op een cirkel, zijn dus ook te maken als machtreeksen. Wie had dat kunnen denken?

Meetkunde of niet, geen ogen kunnen nu missen dat de machtreeksen voor $\cos(x)$ en $\sin(x)$ een schitterende overeenkomst vertonen met de machtreeks voor $\exp(x)$. We zien dat $\cos x + \sin x$ zelfs nog meer op $\exp(x)$ lijkt. Optellen en sorteren geeft

$$\cos x + \sin x = 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} - \frac{x^6}{6!} + \cdots ,$$

afgezien van de minnen precies de reeks voor $\exp(x)$.

Kunnen we met $\cos(x)$ en $\sin(x)$ niet ook een reeks voor $\exp(x)$ zelf maken, met andere woorden, is er een echt verband tussen $\cos(x), \sin(x)$ en $\exp(x)$? Kijk eens naar de plussen en minnen. De termen $x^2, x^3, x^6, x^7, x^{10}, x^{11}, \cdots$ hebben een min, de andere coëfficiënten hebben een plus. Met een extra min voor de x krijgen we een andere machtreeks die op $\exp(x)$ lijkt, maar hem ook weer niet is. Als we echter bij exp beginnen en

$$\exp(ax) = 1 + ax + a^2\frac{x^2}{2!} + a^3\frac{x^3}{3!} + a^4\frac{x^4}{4!} + a^5\frac{x^5}{5!} + a^5\frac{x^6}{6!} + \cdots$$

uitwerken als

$$1 + a^2\frac{x^2}{2!} + a^4\frac{x^4}{4!} + a^6\frac{x^6}{6!} + \cdots + a(x + a^2\frac{x^3}{3!} + a^4\frac{x^5}{5!} + \cdots)$$

en vergelijken met de reeksen voor $\cos x$ en $\sin x$ dan is de keuze duidelijk. Met

$$a^2 = -1$$

krijgen we

$$\exp(ax) = \cos x + a \sin x.$$

Helaas, kwadraten zijn altijd positief, tenzij we even net doen of onze neus bloedt en een getal i in het leven roepen met de eigenschap dat $i^2 = -1$. Met dit bijzondere getal[27] lukt het wel. Dus, vervang a door i en gebruik dat $i^2 = -1, i^3 = i^2 i = -i, i^4 = i^2 i^2 = 1, i^5 = i^4 i = i, \ldots$ en het staat er:

$$\exp(ix) = \cos x + i \sin x. \tag{10.7}$$

Die i laten we ons niet meer afpakken.

In veel wiskundeboeken geeft (10.7) aanleiding tot de mooiste formule uit de wiskunde. Als je $x = \pi$ invult, dan krijg je

$$\exp(\pi i) = -1$$

ofwel

$$\exp(\pi i) + 1 = 0, \tag{10.8}$$

doorgaans geschreven als

$$e^{\pi i} + 1 = 0, \tag{10.9}$$

een formule waarin de vijf belangrijkste symbolen uit de wiskunde met elkaar in verband lijken te worden gebracht:

$$0 \quad 1 \quad e \quad \pi \quad i$$

Helaas, helaas, helaas, dit is meer dan een beetje misleidend. Het machtsverheffen in (10.9) is helemaal geen machtsverheffen. Het getal e komt zelf niet eens voor in de definitie van $\exp(\pi i)$. Wij houden het in dit boekje dan ook op (10.8). Maar hoe het ook zij, met het imaginaire getal i zijn zo $\cos(x)$, $\sin(x)$ en $\exp(x)$ onder een gemeenschappelijke noemer gebracht. En tot slot, mocht je vinden dat ook

$$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots$$

en

$$x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$

heel mooie reeksen zijn dan heb je volkomen gelijk en wordt de cirkel hierboven vervangen door een tak van een hyperbool[28].

[27] Bestaat zo'n getal i wel?

[28] Welke hyperbool?

10.4 Met cos en sin de cirkel rond[29]

We zijn ermee vertrouwd dat de eenheidscirkel $x^2 + y^2 = 1$ in het xy-vlak wordt geparametriseerd met $x(t) = \cos(t)$ en $y(t) = \sin(t)$, twee functies die als paar een oplossing zijn bij beginwaarden $x(0) = 1$ en $y(0) = 0$ van het stelsel differentiaalvergelijkingen

$$x'(t) = -y(t) \quad \text{en} \quad y'(t) = x(t),$$

waarvan we nu de algemene oplossing bepalen. In het bijzonder laten we zien dat de beginvoorwaarden voor $x(0)$ en $y(0)$ de oplossing vastleggen. Dus als je op de een of andere manier twee functies gemaakt hebt die via hun afgeleiden aan elkaar gekoppeld zijn zoals cos en sin, en die beginnen in 1 en 0 op $t = 0$ dan zijn het dezelfde functies, ook als de variabele θ heet en om wat voor reden dan ook hoek genoemd wordt.

Deze differentiaalvergelijkingen kunnen we in matrixnotatie opschrijven als

$$\begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \tag{10.10}$$

consistent met de gebruikelijke matrixvermenigvuldiging als we vectoren ook als matrices zien, en het gaat ons nu om de matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Schrijven we

$$z(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \quad \text{en} \quad z'(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix},$$

dan leest (10.10) als

$$z'(t) = Az(t), \tag{10.11}$$

hetgeen er veel strakker uitziet[30], en vergeten we even dat A een matrix is en geen getal, dan is de oplossing $z(t)$ van deze vergelijking na het maken van Opgave 10.17 gelijk aan

$$\exp(tA)c = (1 + tA + \frac{t^2A^2}{2!} + \frac{t^3A^3}{3!} + \cdots)c.$$

Heel sneaky hebben we hier de t al voor de A geschreven en de c na de $\exp(tA)$, omdat we natuurlijk niet helemaal vergeten waren dat A een matrix is en $z(t)$ een kolomvector. Als we c vervangen door

$$c = z(0) = \begin{pmatrix} x(0) \\ y(0) \end{pmatrix},$$

[29] Voor de liefhebber.

[30] Goede notatie is belangrijk, maar onthoud wel waar de notatie voor staat.

dan past alles qua matrixnotatie precies.

Elke A^n is het n-voudig matrixproduct van A met zichzelf[31], en e^{tA} is een gewogen som van al die matrixproducten, zelf dus weer een matrix die we kunnen loslaten op c met als resultaat

$$z(t) = \exp(tA)z(0).$$

Als het goed is wordt deze conclusie gerechtvaardigd door het via een integrerende factor equivalent zijn van de vergelijking (10.11) met

$$\frac{d}{dt}(\exp(-tA)z(t)) = 0, \tag{10.12}$$

en dus $\exp(-tA)z(t) = c$, met een unieke c die volgt uit de beginwaarde.

Kortom, we hebben naïef een formule opgeschreven die lijkt op wat we gewend zijn voor de vergelijking $x'(t) = ax(t)$. Kan het resultaat goed zijn? Als we weten hoe we matrices moeten vermenigvuldigen dan kunnen we gewoon kijken of het klopt. Als uitgangspunt nemen we dat we weten hoe A op een kolomvector met entries ξ en η werkt. We hebben dan

$$A\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}-\eta\\ \xi\end{pmatrix}.$$

Maar dan is

$$A^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} = AA\begin{pmatrix}\xi\\ \eta\end{pmatrix} = A\begin{pmatrix}-\eta\\ \xi\end{pmatrix} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}-\eta\\ \xi\end{pmatrix} = \begin{pmatrix}-\xi\\ -\eta\end{pmatrix},$$

en dat is waar we net mee begonnen zijn, op een minteken na[32], dus nu is

$$A^3\begin{pmatrix}\xi\\ \eta\end{pmatrix} = AA^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} = A\begin{pmatrix}-\xi\\ -\eta\end{pmatrix} = \begin{pmatrix}\eta\\ -\xi\end{pmatrix},$$

en

$$A^4\begin{pmatrix}\xi\\ \eta\end{pmatrix} = A^2A^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\xi\\ \eta\end{pmatrix},$$

precies waar we net mee begonnen zijn. Hierna herhaalt het zich dus en kunnen we de beoogde algemene oplossing uitrekenen als

$$z(t) = \exp(tA)\begin{pmatrix}\xi\\ \eta\end{pmatrix} = (1 + tA + \frac{t^2A^2}{2!} + \frac{t^3A^3}{3!} + \cdots)\begin{pmatrix}\xi\\ \eta\end{pmatrix}$$

[31] Waarbij $n = 0$ voor A^0 de identiteitsmatrix I geeft.

[32] Grappig: $A^2 = -I$. Doet je dat ergens aan denken?

$$= \begin{pmatrix}\xi\\ \eta\end{pmatrix} + tA\begin{pmatrix}\xi\\ \eta\end{pmatrix} + \frac{t^2}{2!}A^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} + \frac{t^3}{3!}A^3\begin{pmatrix}\xi\\ \eta\end{pmatrix} + \cdots$$

$$= \begin{pmatrix}\xi\\ \eta\end{pmatrix} + t\begin{pmatrix}-\eta\\ \xi\end{pmatrix} + \frac{t^2}{2!}\begin{pmatrix}-\xi\\ -\eta\end{pmatrix} + \frac{t^3}{3!}\begin{pmatrix}\eta\\ -\xi\end{pmatrix} + \frac{t^4}{4!}\begin{pmatrix}\xi\\ \eta\end{pmatrix} + \frac{t^5}{5!}\begin{pmatrix}-\eta\\ \xi\end{pmatrix} + \cdots$$

$$= \begin{pmatrix}\xi - \frac{t^2}{2!}\xi + \frac{t^4}{4!}\xi + \cdots - t\eta + \frac{t^3}{3!}\eta - \frac{t^5}{5!}\eta + \cdots\\ t\xi - \frac{t^3}{3!}\xi + \frac{t^5}{5!}\xi + \cdots + \eta - \frac{t^2}{2!}\eta + \frac{t^4}{4!}\eta + \cdots\end{pmatrix} = \begin{pmatrix}\xi c(t) - \eta s(t)\\ \xi s(t) + \eta c(t)\end{pmatrix}.$$

De vector

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix}$$

in de zo gevonden (oplossings?)formule

$$z(t) = \exp(tA)\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\xi c(t) - \eta s(t)\\ \xi s(t) + \eta c(t)\end{pmatrix} = \begin{pmatrix}c(t) & -s(t)\\ s(t) & c(t)\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix} \tag{10.13}$$

speelt de rol van de constante c in de algemene oplossing

$$x(t) = ce^{at} = e^{ta}c = \exp(ta)c$$

van $x'(t) = ax(t)$, en in (10.13) zien we dat

$$\exp(tA) = \begin{pmatrix}c(t) & -s(t)\\ s(t) & c(t)\end{pmatrix},$$

hetgeen dan ook geldt met t vervangen door $-t$.

Natuurlijk is (10.13) niets anders dan de matrixnotatie voor

$$x(t) = \xi c(t) - \eta s(t) \quad \text{en} \quad y(t) = \xi s(t) + \eta c(t), \tag{10.14}$$

en deze oplossing zou gerechtvaardigd moeten worden door (10.12), dus laten we meteen maar zien of dat ook zo is. De uitdrukking in (10.12) waarvan de tijdsafgeleide wordt genomen is gelijk aan

$$\exp(-tA)z(t) = \begin{pmatrix}c(-t) & -s(-t)\\ s(-t) & c(-t)\end{pmatrix}\begin{pmatrix}x(t)\\ y(t)\end{pmatrix} = \begin{pmatrix}c(t)x(t) + s(t)y(t)\\ -s(t)x(t) + c(t)y(t)\end{pmatrix}.$$

Komt daar inderdaad de nulvector uit als je naar t differentieert?

Opgave 10.18. Laat zien dat $c(t)x(t) + s(t)y(t)$ en $-s(t)x(t) + c(t)y(t)$ naar t gedifferentieerd nul geven, gebruikmakend van de vergelijkingen voor zowel c en s als x en y.

Beide uitdrukkingen zijn dus constant, en de constanten noemen we ξ en η. We concluderen dat

$$c(t)x(t) + s(t)y(t) = \xi \qquad -s(t)x(t) + c(t)y(t) = \eta.$$

Opgave 10.19. Los de laatste twee vergelijkingen op naar $x(t)$ en $y(t)$. Maak gebruik van de identiteiten voor $c(t)$ en $s(t)$ en leid zo (10.14) af.

Kortom, de unieke oplossing van van (10.6) wordt gegeven door $x = c(t)$ en $y = s(t)$. Uit de beginvoorwaarden volgt immers dat $\xi = x(0) = 1$ en $\eta = y(0) = 0$. Er is maar een zo'n oplossing, en die hebben we al, gegeven door de machtreeksen voor $c(t)$ en $s(t)$ die zo op onze exp-machtreeks lijken.

De volgende opgave illustreert nog eens een keer de kracht van deze uitspraak. De somformules voor de cosinus en sinus die je al kende of geloofde, zonder plaatjes, zonder hoeken, rechtstreeks verkregen vanuit de uniciteit van oplossingen beginnende op $t = s$ met een bijzonder gekozen ξ en η.

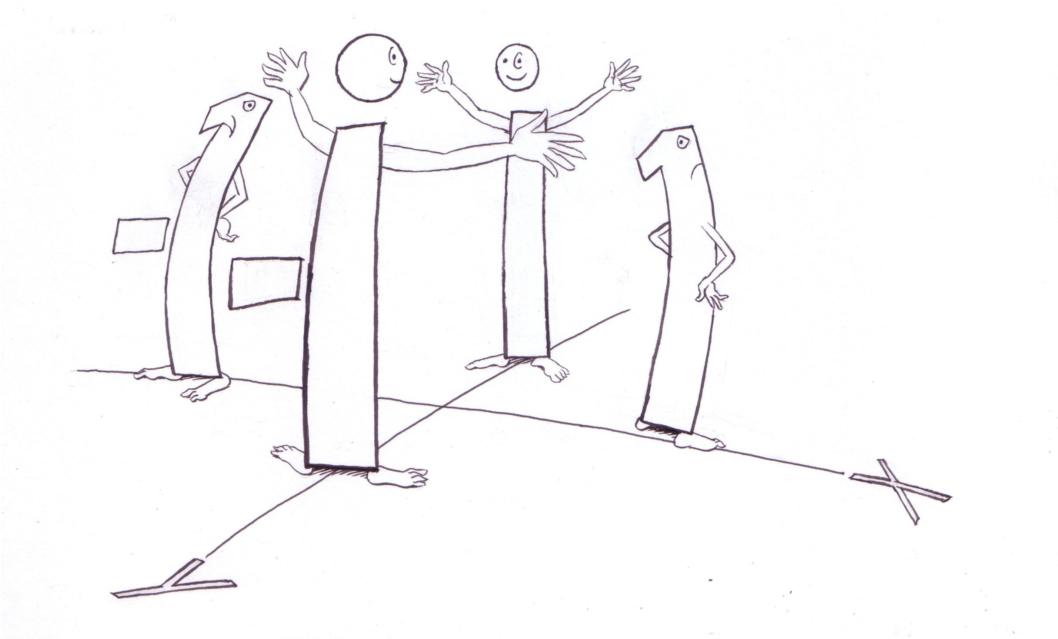
Opgave 10.20. Als je op $t = s$ begint met de oplossing $x(s) = \cos(s)$ en $y(s) = \sin(s)$ dan hoef je de oplossing niet meer uit te rekenen, want je hebt de oplossing al met $x(t) = \cos(t)$ en $y(t) = \sin(t)$. Schrijf nu $t = s + \tau$ en verifieer dat

$$\cos(s + \tau) = \cos(s)\cos(\tau) - \sin(s)\sin(\tau);$$

$$\sin(s + \tau) = \sin(s)\cos(\tau) + \cos(s)\sin(\tau),$$

door te laten zien dat zowel de linker- als de rechterleden als functie van τ (in de rol van nieuwe t) een oplossing definiëren die op $\tau = 0$ begint in $x = \cos(s)$ en $y = \sin(s)$. Daar is er maar eentje van en dus weten we nu zeker dat de somformules[33] voor $\cos$ en $\sin$ gelden, zonder plaatjes of wat dan ook.

[33] Vergelijk: $\exp(s+\tau) = \exp(s)\exp(\tau)$ via uniciteit voor $x'(t) = x(t)$ met $x(0) = \exp(s)$.



11 Complexe getallen

In het vorige hoofdstuk hebben we drie functies waarmee we dachten wel een beetje vertrouwd te zijn vanuit het 'niets' opnieuw geïntroduceerd. De belangrijkste van die drie is de functie exp, die samen met het getal e ons in de schoot viel na wat alledaagse denkactiviteiten over rente op rente en het rekenen met producten van eenvoudige breuken die steeds dichter bij 1 liggen maar waarvan je er wel steeds meer neemt. Wat we gedaan hebben is de enen in zinloze producten als

$$1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1$$

een stukje groter (of kleiner) maken. Een verhaal dat vervolgens begon bij sommetjes als

$$\frac{9}{8} \times \frac{9}{8} \times \frac{9}{8} \times \frac{9}{8} \times \frac{9}{8} \times \frac{9}{8} \times \frac{9}{8} \times \frac{9}{8}$$

leidde eerst tot

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3 \times 2} + \frac{x^4}{4 \times 3 \times 2} + \cdots = (\exp(1))^x = e^x,$$

en vervolgens tot de differentiaalrekening voor dit soort polynomen die we in de wiskunde machtreeksen[1] zijn gaan noemen. Met deze $\exp(x)$ als de functie die zijn eigen afgeleide is.

Een beetje spelen[2] met varianten op deze $\exp(x)$ leidde ook tot paren van functies die elkaars afgeleiden zijn, soms met een min teken, zoals $\sin(x)$ en $\cos(x)$. Een begin van een verhaal op zich, dat bij de simpelste vragen al aanleiding gaf tot de vergelijking

$$a^2 = -1$$

voor een a waarmee we sin en cos aan exp zouden kunnen koppelen. Die a bestond niet, maar we hadden er al mee gerekend en hem daarom maar i van imaginair gedoopt. We denken ons i in. Dus $a = i$ voldoet aan $a^2 = -1$, net zoals $a = \sqrt{2}$ voldoet aan $a^2 = 2$.

Die $\sqrt{2}$ kenden de Grieken al als platonisch getal bij een imaginaire perfecte rechthoekige driehoek met rechte zijden allebei precies van lengte 1. Dat zulke driehoeken in de werkelijkheid niet bestaan was wellicht de werkelijke reden om het niet rationaal zijn van $\sqrt{2}$ maar liever geheim te houden.

Niet $\sqrt{2}$ maar i wordt nu imaginair genoemd, het is maar hoe je ertegenaan kijkt. Als je op het idee komt om in plaats van met $\sqrt{2}$ eerst liever met i te rekenen, wie houdt je dan eigenlijk tegen? Voor rekenen met i zijn wortels niet eens nodig. En laten we wel wezen, als je eenmaal wortels aan het trekken bent, dan wil je gewoon worteltrekken, en niet door negatieve discriminanten worden tegengehouden. Hier zijn een paar opgaven voor wortelliefhebbers, en daarna gaan we rekenen met i.

Opgave 11.1. Los de vierkantsvergelijking $x^2 + px + q = 0$ op door kwadraatafsplitsen en worteltrekken.

Opgave 11.2. Probeer de differentiaalvergelijking $f''(x) = f(x)$ op te lossen door $f(x) = \exp(\mu x)$ in te vullen, met een geschikte keuze van μ. Probeer hetzelfde bij $f''(x) + f(x) = 0$. En bij $f''(x) + \gamma f'(x) + f(x) = 0$ voor $\gamma \in \mathbb{R}$ vast gekozen.

Opgave 11.3. De vergelijking $x^3 + px + q = 0$ kun je oplossen voor x met de volgende truc. Zoek x in de vorm $x = y + z$ en vul dit in in[3] de vergelijking. Je

[1] Power series they are!
[2] Alleen even of oneven termen, plus-min-plus-min invoegen.
[3] Twee keer in, net als twee keer aan bij aanzitten aan het banket.

ziet twee pure derde machten verschijnen. De andere termen met y en z bij elkaar genomen factoriseren in twee factoren waarvan je er eentje op nul kunt stellen. Mag dat? Waarom niet? Bij een van de twee heeft het niet veel zin, maar met de andere gelijk aan nul krijg je twee prachtige vergelijkingen voor (en symmetrisch in) y en z, in één keer opgeschreven als $y^3 + z^3 + q = 0 = 3yz + p$. Dat geeft na even rekenen een vierkantsvergelijking voor y^3 (of z^3) en die kun je wel aan met Opgave 11.1. In een simpel voorbeeldje met $p = -6$ en $q = -4$ moet dat toch goed gaan verder? Onderzoek met je DRM de grafiek van $x^3 - 6x - 4$ maar. Drie nulpunten, één wortel raad je wel en dan lukken de andere twee ook wel. Maar via y en z? Brrr.

Opgave 11.4. Ook $x^3 + x - 2 = 0$ is een leuke. Behalve via raden lukt het nu ook via y en z. Maar klopt het antwoord ook?

11.1 Rekenen met complexe getallen

Kortom, die wortels uit de negatieve getallen hebben we gewoon nodig, net zo hard als $\sqrt{2}$. Maar een beetje anders zijn ze wel, zoals we zullen zien bij het rekenen en tekenen. Waar $\sqrt{2}$ en $-\sqrt{2}$ gewoon twee (heel) verschillende plekjes op de getallenlijn hebben is diezelfde lijn vol voor i. Noch i noch $-i$ zijn welkom in de reële getallen. Vol is vol. Maar dat is bepaald geen reden om i weg te denken uit het bestaan. Een stellige uitspraak die we willen bewijzen[4].

The proof of the pudding is in the eating en eten is rekenen zoals we gezien hebben bij de pizza's. Rekenen is wat we dus nu gaan doen, met i, en we zien wel of dat rekenen goed gaat of niet, al of niet in betekenisvolle context. Maar eerste geven we i een plek. De lijn die we begonnen met 0 en 1 is vol, dus wat is een goede plek voor i? Wel, een beetje rekenen met de regels zoals we die gewend zijn geeft met $i^2 = -1$ dat $i^3 = i^2 \cdot i = -1 \cdot i = -i$ en $i^4 = i^2 \cdot i^2 = -1 \cdot -1 = 1$.

Meer getallen dan $1, i, -1$ en i krijgen we niet als we i met zichzelf blijven vermenigvuldigen en dus zijn deze vier getallen samen met 0 bepalend voor de grafische voorstelling die we bij deze nieuwe getallen gebruiken. Net als -1 en 1 willen we i en $-i$ allebei op afstand 1 van 0 tekenen, en in het platte vlak is maar één manier om dat mooi te doen. Teken maar de getallenlijn, markeer $-1, 0, 1$ en kies i en $-i$ op een cirkel met straal 1 en middelpunt 0 en wel zo dat allevier de getallen daar mooi symmetrisch omheen liggen.

Bijvoorbeeld door die cirkel te beschrijven als $x^2 + y^2 = 1$ in het xy-vlak en bij de punten $(0,0)$, $(1,0)$, $(0,1)$ de getallen 0, 1 en i neer te zetten. Je ziet

[4] Een stelling over wiskunde eigenlijk.

wel waar je -1 en $-i$ moet zetten en als we verder rekenen en meer nieuwe getallen vinden dan ligt waarschijnlijk nu ook wel voor de hand waar we de uitkomsten van onze simpele complexe sommetjes een plek willen geven in het xy-vlak dat we vanaf nu het complexe vlak noemen.

In dat complexe vlak zien we de vier bijzondere getallen $1, i, -1, -i$ die de eigenschap hebben dat ze oplossingen zijn van de vergelijking $z^4 = 1$ voor de variabele[5] z die kennelijk zowel reëel als imaginair moet kunnen zijn. Best wel complex, maar eigenlijk ook weer niet, want onze vierdegraadsvergelijking heeft nu vier oplossingen, net als de twee tweedegraadsvergelijkingen $z^2 = 1$ en $z^2 + 1 = 0$ die elk twee oplossingen hebben. Niet zo gek als je bedenkt dat $z^4 - 1 = (z^2 + 1)(z^2 - 1)$.

Rekenen is optellen (hetgeen leidt tot sommen), vermenigvuldigen (producten) en het bijbehorende aftrekken (verschillen) en delen (quotiënten). De producten met de vier bijzondere getallen hebben we gehad. Als we alleen maar producten nemen in onze sommenmakerij dan zijn we met deze vier getallen tevreden. Ze vormen samen een groepje met één rekenkundige bewerking dat als prachtige simpele wiskundige entiteit bestaansrecht heeft. Net als het deelgroepje met -1 en 1 trouwens.

Zodra we ook gaan optellen worden er meer nieuwe getallen voortgebracht, want nu doen uitgaande van $1, i, -1, -i$ ook als eerste 0 zelf mee via $0 = 1 + (-1) = i + (-i)$, maar ook $i + 1 = 1 + i$, dat we niet verder kunnen vereenvoudigen, en dat in ons complexe vlak als plaats het punt $(1, 1)$ krijgt, waar anders. Op dezelfde manier is duidelijk waar $i + (-1) = -1 + i$, $-i + 1 = 1 - i$ en $-1 + (-i) = -1 - i$ thuishoren, maar ook $1 + 1 = 2$ en $i + i = 2i$, en vervolgens $1 + 2i$, $2i - 1 = -1 + 2i$, enzovoorts. Als je eenmaal aan het optellen bent houdt[6] het niet op. Op elk roosterpunt in het xy-vlak staat nu een complex getal van de vorm $a + bi$ met a en b allebei geheel. Al deze getallen bij elkaar worden de gehele getallen van Gauss genoemd en daar kan een boek op zich over worden geschreven, bijvoorbeeld omdat $5 = (1 + 2i)(1 - 2i)$ nu geen priemgetal meer is.

De volgende stap is dat we alle punten in het xy-vlak als complex getal zien, dus bij $(x, y) = (a, b)$ hoort het complexe getal $a + bi$. En net zoals we $x = a$ vaak zien als de waarde van een reële variabele x zien we $a + bi$ als de waarde van een complexe variabele $z = x + iy$ die we ook gebruiken om complexe functies te maken, bijvoorbeeld de functie f met formule voorschrift $f(z) = z^4 - 1$. Elk gewoon polynoom met x als variabele en gewone reële coëfficiënten geeft zo een bijbehorende complexe functie gedefinieerd door dezelfde formule met $z = x + iy$ in de plaats van x.

[5] De spreekwoordelijke onbekende x heeft zijn bevoorrechte positie verloren.

[6] Het is bij de konijnen af.

Je ziet dat de notatie een beetje flexibel is. We schrijven meestal iy en bi vanwege het alfabet. En liever niet $i2$ maar $2i$. En $x + iy$ in plaats van $iy + x$, maar dat is een kwestie van smaak. Het complexe getal $a + bi$ wordt vastgelegd door een geordend paar reële getallen a en b waarmee behalve het punt (a, b) ook het complexe getal $a + bi$ correspondeert. De overgang van (a, b) naar $a + bi$ is vanuit typografisch oogpunt gezien niet meer dan een kwestie van tekstzetten waar geen wiskundig begrip voor nodig is.

Als Descartes om wat voor reden dan ook had besloten om zijn punten niet als (x, y) maar als $x+iy$ te noteren dan hadden we het verschil niet eens gezien. We noemen daarom a het reële deel, en b het imaginaire deel van het complexe getal $a + bi$. Als $b = 0$ hebben we een 'gewoon' reëel getal en als $a = 0$ hebben we een puur imaginair getal. Samen met alle andere $a + bi$ is het complexe vlak nu te zien als de verzameling van complexe getallen, en in dat vlak liggen alvast de vier nulpunten van $f(z) = z^4 - 1$, twee die we al hadden op de reële x-as en twee op de imaginaire y-as, die helemaal niet imaginairder is dan de x-as.

Het optellen (en aftrekken) van zulke getallen gaat als vanzelf. Voor de som van $a + bi$ en $c + di$ (met a, b, c, d in $\mathbb{R}$ natuurlijk) volgt met de gebruikelijke rekenregels

$$(a + bi) + (c + di) = a + bi + c + di = a + c + bi + di = (a + c) + (c + d)i.$$

Vermenigvuldigen is wat ingewikkelder maar als we ook daarbij gewoon rekenen zoals we gewend zijn dan werken we $(a + bi) \times (c + di)$ uit als

$$\begin{aligned}(a + bi)(c + di) &= ac + adi + bic + bidi\\ &= ac + i^2bd + adi + bci\\ &= ac - bd + (ad + bc)i,\end{aligned}$$

en vinden de regel dat

$$(a + bi)(c + di) = (ac - bd) + i(ad + bc). \tag{11.1}$$

Als $b = d = 0$, dus als we gewoon met een reëel getal van doen hebben, dan is het product ac, precies zoals het hoort, en ook de som $(a + bi) + (c + di)$ reduceert dan tot $a + c$. Dezelfde regels vinden we als we met variabele getallen werken, die we meestal letters aan het eind van het alfabet als naam geven.

De basisoperaties zijn dus niet erg ingewikkeld, maar er is nog een andere schrijfwijze van de complexe getallen die het vermenigvuldigen een nieuwe interpretatie geeft die bij berekeningen soms handiger is. Een punt (a, b) in het vlak, en daarmee het complexe getal $a + bi$, wordt vastgelegd door de twee coördinaten a en b, maar ook door zijn afstand r tot de oorsprong, en de hoek θ die de lijn door de oorsprong en het punt (a, b) maakt met de positieve x-as.

We noemen r ook wel de absolute waarde van z, en noteren r als $r = |z|$. De hoek θ heet het argument van z en wordt genoteerd als $\arg z$. Je kunt afspreken dat θ tussen de 0 en 2π ligt, of dat θ juist tussen $-\pi$ en π ligt. Handiger vaak is de keuze onbepaald te laten. Dan mag θ elke reële waarde aannemen, maar is het verband tussen r, θ en $z = x + iy$ niet meer een op een. Immers, als je 2π bij θ optelt krijg je dezelfde z. Let op, $r = 0$ is een beetje een bijzonder geval, en niemand die hier met $r < 0$ rekent gelukkig.

Opgave 11.5. Overtuig jezelf ervan dat in bovenstaande notatie geldt dat $a = r\cos(\theta)$ en $b = r\sin(\theta)$.

Gezien deze opgave schrijven we vaak $z = r(\cos(\theta) + i\sin(\theta))$. De punten waarvoor geldt dat $|z| = r = R$ vormen de cirkel met middelpunt de oorsprong en straal gelijk aan R.

Opgave 11.6. Verifieer met deze notatie en regel (11.1) dat voor $z = x + iy = r(\cos(\theta) + i\sin(\theta))$ en $w = u + iv = s(\cos(\phi) + i\sin(\phi))$ geldt dat

$$\begin{aligned} wz &= rs(\cos(\theta) + i\sin(\theta))(\cos(\phi) + i\sin(\phi)) \\ &= rs(\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)) + i(\sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi)) \\ &= rs(\cos(\theta + \phi) + i\sin(\theta + \phi)). \end{aligned}$$

De laatste gelijkheid volgt uit de somformules voor cosinus en sinus die je in het vorige hoofdstuk opnieuw gezien hebt. Het resultaat van het vermenigvuldigen van w en z is zo een nieuw complex getal waarvan de afstand tot de oorsprong het product rs van de oorspronkelijke afstanden is, en waarbij de hoek tussen wz en de positieve x-as de som $\theta + \phi$ is van de hoeken van w en z apart. Hiervan kun je naar believen veelvouden van 2π aftrekken.

Opgave 11.7. Ga na dat dit consistent is met het vermenigvuldigen van reële getallen zoals we dat gewend zijn.

Opgave 11.8. Verifieer de formule van de Moivre:

$$(\cos(\theta) + i\sin(\theta))^n = \cos(n\theta) + i\sin(n\theta).$$

Als je een complex getal z steeds met zichzelf vermenigvuldigt om de rij getallen z^n met $n = 1, 2, 3, \ldots$ te maken, dan 'spiraliseert' de rij steeds verder van de oorsprong weg als $r > 1$, naar de oorsprong toe als $r < 1$, en blijft voor $r = 1$ wilde rondjes draaien over de eenheidscirkel. Tenzij uiteindelijk z reëel is natuurlijk. Waar in de reële getallen rijen als x^n met $n = 1, 2, 3, \ldots$ betrekkelijk saai zijn, geven de complexe getallen een rijker gedrag met nieuwe verschijnselen en interpretaties.

Ontdekt vanuit de reële getallen, die we weer ontdekt hadden vanuit de rationale getallen, zijn de complexe getallen dé verzameling van getallen waar we in de wiskunde mee werken en de sommetjes die we hierboven ter motivatie hebben gemaakt laten zien dat we net zo min om die complexe getallen heen kunnen als om de reële getallen. Wat we al hadden, blijft echter bewaard en het leukste is natuurlijk als een rationaal probleem via een irrationaal of zelfs complex uitstapje weer een rationaal antwoord geeft.

Daarvan zijn voorbeelden te over, zoals de gulden snede die gebruikt wordt in de formule voor de rij van Fibonacci, of de complexe functietheorie die onmisbaar is in de getaltheorie en de moderne natuurkunde. Dankzij de complexe getallen kunnen we meer dan zonder de complexe getallen, veel meer dan we hier kunnen bespreken. En soms is de complexe oplossing gewoon mooier, bijvoorbeeld bij de eenvoudige vergelijkingen die we oplossen als we worteltrekken.

11.2 Complexe wortels trekken, de vergelijking $z^n = c$

Binnen de reële getallen heeft de vergelijking $x^2 = -1$ geen oplossing. Tenminste, dat hebben we allemaal zo geleerd, want min keer min is plus, en plus keer plus ook, dus een kwadraat kan nooit negatief zijn. Maar waarom is dat eigenlijk zo? Je kunt je afvragen wanneer je het als vanzelfsprekend bent gaan zien dat min keer min plus is. Wie dwong je ertoe? Was het uit vrije wil?

Er zijn natuurlijk meer regels die je hebt geaccepteerd, gelukkig maar. Deze twee regels bij het rekenen met reële getallen bijvoorbeeld:

1. $a \cdot (b + c) = a \cdot b + a \cdot c$ voor alle reële getallen a, b en c.
2. $a + (-a) = 0$ voor alle reële getallen a.

Daar kun je een beetje mee spelen, en dat doen we hieronder, met zowel ronde als rechte haakjes om niet in de war te raken bij deze simpele exercities[7]. Uit

$$\begin{aligned}(-a)\cdot b + a\cdot b &\overset{(1)}{=} [(-a)+a]\cdot b\\ &\overset{(2)}{=} 0\cdot b = 0.\end{aligned}$$

concluderen we door er links en rechts $-(a\cdot b)$ bij op te tellen dat

$$(-a)\cdot b = -(a\cdot b).$$

Maar nu geldt ook dat

$$\begin{aligned}(-a)\cdot(-b) + [-(a\cdot b)] &\overset{\text{zojuist}}{=} (-a)\cdot(-b) + (-a)\cdot b\\ &\overset{(1)}{=} (-a)\cdot[(-b)+b]\\ &\overset{(2)}{=} (-a)\cdot 0\\ &= 0.\end{aligned}$$

Als we nu links en rechts $a\cdot b$ optellen dan vinden we dat

$$(-a)\cdot(-b) = a\cdot b,$$

kortom, min keer min is plus. Die eigenschap kiezen we dus niet, hij wordt ons opgedrongen door een algebra waarin we (1) en (2) wensen te accepteren en gebruiken[8]. De huis, tuin, en keukeneigenschappen die we voor de reële getallen willen hebben, laten ons geen andere keus.

We hebben dus goede redenen om de vergelijking $z^n = c$ te bestuderen beginnende met de constatering dat er voor $c < 0$ en n even geen oplossingen zijn in de reële getallen. Complexe oplossingen voor z vinden we echter vrijwel onmiddellijk met de formule van de Moivre uit Opgave 11.8 door naar de absolute waarden en argumenten te kijken.

Kijk eerst maar naar de absolute waarde. Als de n-de macht van z gelijk moet zijn aan c dan moet de absolute waarde van z dus $|c|^{1/n}$, de gewone n-de machtswortel uit $|c|$ zijn. Dat is makkelijk. Voor het argument van z (de hoek die de lijn door z maakt met de positieve x-as) ligt het iets subtieler en het is juist hier dat onze rijkdom met de complexe getallen zichtbaar wordt. Noem het argument van c even θ. Aangezien argumenten optellen bij

[7] Elk jaar hetzelfde gelazer verzuchtte Joosts wiskundeleraar bij rumoer in de klas.
[8] Samen met nog wat vanzelfsprekende regeltjes.

vermenigvuldigen[9] kunnen we door $\arg z = \frac{\theta}{n}$ te kiezen een oplossing maken. Dan komen we immers precies goed uit als we de n-de macht van z nemen.

Met andere woorden, het complexe getal

$$z_1 = |c|^{1/n} \left(\cos(\frac{\theta}{n}) + i \sin(\frac{\theta}{n}) \right)$$

is een oplossing van de vergelijking $z^n = c$. Maar het is niet de enige[10]! Want als we in plaats van $\frac{\theta}{n}$ het argument $\frac{\theta+2\pi}{n}$ kiezen lukt het ook. Immers, als we n keer dit argument nemen dan komen we uit of $\theta + 2\pi$, maar dat is hetzelfde als θ. Dus

$$z_2 = |c|^{1/n} \left(\cos(\frac{\theta + 2\pi}{n}) + i \sin(\frac{\theta + 2\pi}{n}) \right)$$

is ook een oplossing van $z^n = c$. En wat voor 2π geldt, geldt ook voor 4π, enzovoorts. De volgende opgave vat het samen.

Opgave 11.9. Laat zien dat alle oplossingen van de vergelijking $z^n = c$ gegeven worden door $z_1, z_2, \ldots, z_n$, met als laatste oplossing in deze opsomming

$$z_n = |c|^{1/n} \left(\cos(\frac{\theta + 2(n-1)\pi}{n}) + i \sin(\frac{\theta + 2(n-1)\pi}{n}) \right).$$

Maak een plaatje om te concluderen dat deze n oplossingen liggen op een regelmatige veelhoek in de cirkel met straal $|c|^{1/n}$.

Opgave 11.10. Kijk even naar het bijzonder geval $z^2 = -4$. Welke twee oplossingen krijgen we nu?

Nu hebben we de smaak te pakken. Hoe zit het met nulpunten van algemene polynomen van graad n? Waar er binnen de reële getallen polynomen P zat zijn waarvoor $P(x) = 0$ geen oplossing heeft, heeft iedere tweedegraads vergelijking in de complexe getallen altijd een oplossing. Dat kun je nu zelf laten zien. Bijvoorbeeld:

Opgave 11.11. Gebruik Opgave 11.1 om $z^2+2z+2 = 0$ op te lossen in de complexe getallen, door de complexe wortels te trekken uit het negatieve kwadraat.

[9] Waar heb je deze elegante eigenschap eerder gezien?
[10] Tenzij $n = 1$.

Je ziet dat de tweedegraadsvergelijking zo altijd twee oplossingen heeft, die soms samen vallen als je de wortel uit nul moet trekken in de laatste stap. Een bijzonder geval van de hoofdstelling van de algebra, de stelling die zegt dat elk polynoom minstens één nulpunt a heeft in de complexe getallen. Met algebra niet te bewijzen trouwens. Maar als je die oplossing eenmaal hebt dan deel je een factor $z - a$ uit met een staartdeling en pas je de stelling nog een keer toe. En nog een keer zolang de graad van het quotïent groter is dan 1. Zo krijg je alle nulpunten.

Een bewijs van deze stelling gaat hier te ver, en we zouden ook maar moeilijk een keuze kunnen maken welk bewijs we zouden moeten geven. De stelling kan bijvoorbeeld ook kanstheoretisch worden bewezen, via een proces dat lijkt op muntjes werpen, maar als je een willekeurig bewijs kiest is de kans groter dat het anders gaat[11].

[11] Via Theo de Jong.

NAWOORD

Dit boek is geen systematische inleiding in de wiskunde geworden. Onze vingers fungeerden als startpunt voor een reisje door de klassieke wiskunde met als kapstok een paar vragen uit het publiek tijdens een cursus op de Vrije Universiteit voor een clubje enthousiaste leraren. Complexe getallen zijn voor nu het eindpunt. We moeten ergens ophouden. Een volgende doorlopende leerlijn dient zich natuurlijk aan. Iedere formule in dit boek waarin een reële x voorkomt heeft ook betekenis als de x door een complexe z vervangen wordt. Eerst maar eens kijken wat daar dan zoal uitkomt en wat je daar dan vervolgens aan hebt zou een mooi startpunt kunnen zijn voor een volgende excursie. Met wederom de vraag op welke moment de letteralgebra de epsilons echt nodig heeft en wanneer de volgende abstractie onmisbaar wordt. Want geen complexe analyse zonder topologie. En geen toepassingen zonder complexe analyse. Ook in de kansrekening. Maar eerst maar weer eens wat vragen uit het publiek.

Dit boek heeft geen literatuurlijst. Geen beginnen aan. Bovendien, als je eenmaal weet wat je moet zoeken dan vind je je weg wel. Of niet? Maak je eigen literatuurlijst en koop eens een boek. Toevallig kwam het thuisfront terug van de Nationale Wiskunde Dagen met het welbekende ϵ-analyseboek van een kleurrijke hoogleraar uit Nijmegen. Zeer genietbaar, ook vanwege de historische kanttekeningen, maar met een andere doelgroep en uit een andere tijd, van voor het worldwide web waar alles nu zo te vinden is. Het is interessant om te zien of een boek als het onze de lezer daarbij op weg helpt. Ons doel was inzicht in de schoonheid en samenhang van de wiskunde, met minimale middelen, hier en daar wat filosofische overwegingen, en alles op basis van intuïtie en gezond verstand. Meer weten? Suggesties voor een cursus over dit boek of dingen die er niet in staan maar er wel mee samenhangen? *Just let us know.*

Index