Gemeenten verdelen eerste cijfers oneerlijk

Een puzzelrubriek rond de wet van Benford

M. E. Wit / A. J. Goddijn
Freudenthal instituut, Universiteit Utrecht

30, niet 11%

Doet u even mee met een klein experiment, 2000 lezers van de Nieuwe Wiskrant!
Noteer daartoe de volgende getallen:
- het aantal bakstenen van uw woning,
- het aantal letters in uw boekenkast,
- de oppervlakte van de gemeente waarin u woont (in vierkante mijlen),
- het aantal haren op uw hoofd,
- het produkt van de cijfers van uw telefoonnummer, gironummer en sofinummer, waarbij de eventuele nullen worden weggelaten.

We komen tot een verzameling van zo'n 10 000 getallen. Onze voorspelling is:

ongeveer 30% van die getallen begint met het cijfer 1.

De bedoeling van deze aflevering is deze waanzinnige bewering geloofwaardig te maken. Waanzinnig ja, want elk gezond mens verwacht dat bij zo'n chaotische verzameling de enen, de tweeën, en de enzovoorts grofweg even vaak voorkomen en dan zou toch ongeveer 11% van die getallen met het cijfer 1 moeten beginnen.[1]
Helaas, uitgebreide onderzoekingen van de Amerikaanse natuurkundige Frank Benford (1938) in de jaren dertig tonen aan dat 30% – om precies te zijn 30.103% – de regel is bij zulke dataverzamelingen. Benford greep terug op een observatie van de sterrenkundige Simon Newcomb. Deze merkte op dat de eerste bladzijden van logaritmetafels er veel beduimelder uitzagen dan de latere bladzijden. Nu is dat heel normaal bij moeilijke of vervelende romans, maar het ligt niet voor de hand bij naslagwerken. Logaritmetafels werden vroeger veelvuldig gebruikt bij vermenigvuldigen. De eerste bladzijden horen bij getallen die met een laag cijfer beginnen, en dat zijn niet bij voorkeur grote of kleine getallen.
Benford onderzocht 20 229 gegevens, zoals inwoneraantallen, oppervlaktes van zeeën, lengtes van rivieren en huisnummers van bekende mensen. Steeds vond hij die nadruk op de 1 als begincijfer.
Benford werd aanvankelijk met zijn experimenteel gevonden wet door wiskundigen niet serieus genomen. Later werd dat anders. Er bleken redelijke verklaringen te zijn voor het verschijnsel, en voor veel wiskundig vastgelegde getalrijen waarbij het bepaald niet te verwachten was gold de wet van Benford ook.

Twee controleerbare gevallen

De inwoneraantallen van de Nederlandse gemeenten staan in het Statistisch Zakboek van het CBS; in figuur 1 zijn ze gesorteerd naar eerste cijfer.

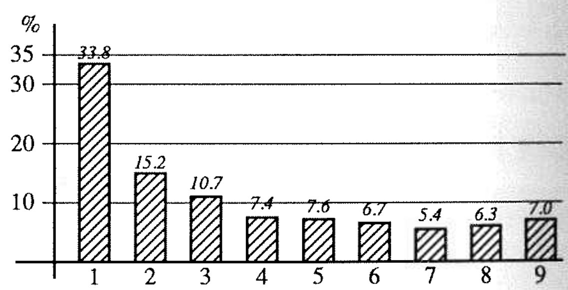

fig. 1: Inwonertallen van de 672 Nederlands gemeenten, gesorteerd naar begincijfer.

En dat is toch op zijn minst een merkwaardig plaatje. Ons tweede experiment is van meer wiskundige aard. We sorteren nu de eerste 100 000 machten van 2 naar hun eerste cijfer. Figuur 2 laat de verdeling zien.

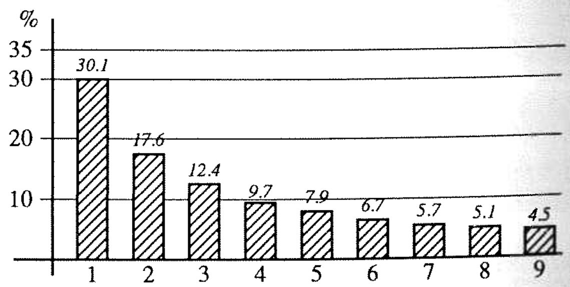

fig. 2: De eerste 100 000 machten van 2, gesorteerd naar begincijfer.

ren van voorbeelden, het verifiëren en falsificeren van vermoedens en het aanbrengen van structuur.
De samenwerking was niet overal vruchtbaar. Natuurlijk zijn er ook leerlingen die liever alleen werken. Verder bleek voor enkele leerlingen de GR nog zo aantrekkelijk te zijn dat ze er liever zelf mee speelden dan dat ze met anderen nadachten over de opdracht. Dit speeleffect zou wellicht minder zijn als de leerlingen ook thuis de GR zouden kunnen gebruiken. Een praktisch punt bij het samenwerken is verder dat, in tegenstelling tot het beeldscherm van een PC, het display van de GR lastig door buurman of buurvrouw is af te lezen.
Achteraf gezien is het jammer dat er op enkele praktische details niet geanticipeerd was. Het was beter geweest als van te voren gezegd was dat a, b, c en d constanten zijn. Verder heeft het feit dat de leerlingen de GR tot nog toe vooral bij het grafisch differentiëren hebben gebruikt remmend gewerkt bij vraag 2. Een opmerking zou ook daar wellicht van nut zijn geweest. En tenslotte was één lesuur wat te kort om alles af te krijgen.

Tot slot

Het was voor ons een boeiende les. Natuurlijk kan men uit één les geen algemene conclusies trekken. Met dat voorbehoud kunnen we toch wel enkele opmerkingen maken die uit deze praktijkervaring voortkomen.
De GR kan, door de mogelijkheid om snel veel grafieken te tekenen, een krachtig hulpmiddel zijn bij het bekijken van een verzameling functies. De voorbeeldgrafieken kunnen dan aanleiding zijn om zelf structuur te construeren, om algemene kenmerken op te sporen of om eigenschappen te onderzoeken.
De GR kan wellicht nuttig zijn bij open geformuleerde opgaven waarbij de leerlingen op eigen niveau een antwoord kunnen geven. De grafieken vormen dan het concrete vertrekpunt van het werk.
Of de GR een positieve invloed heeft op de samenwerking tussen leerlingen staat te bezien. Misschien zullen de verder experimenten daarop uitsluitsel geven.

Noten

[1] G. A. Vonk (1992): 'De Graphics Calculator in de klas, eerste jaar experiment', *Nieuwe Wiskrant 12*(1), 8-11.
[2] L.M. Doorman (1993): 'Een ander gezicht van de grafische rekenmachine', *Nieuwe Wiskrant 12*(2), 39-41.
[3] Treffers, A., E. de Moor en E. Feijs (1989). *Proeve van een Nationaal Programma voor het Reken-wiskundeonderwijs*. Tilburg, Zwijsen.

Aankondiging

De werkgroep *Vrouwen en Wiskunde* en de werkgroep *Vrouwen en Natuurwetenschappen* houden een gezamenlijke studiedag op:

zaterdag 24 april 1993 in De Poort van Kleef, Mariaplaats 7, Utrecht

Op deze studiedag bespreken we de plannen rondom de toekomstige samenwerking van de beide werkgroepen binnen de Stichting Vrouwen en Exacte vakken. Op deze studiedag wordt ook aandacht besteed aan emancipatie en exacte vakken in relatie tot de invoering van de basisvorming.
Voor meer informatie en opgave voor deze studiedag kunt u contact opnemen met het centrum van de werkgroep Vrouwen en Wiskunde, voorlopig nog gevestigd op de Tiberdreef 4 te Utrecht. Tel. 030-612806.

Hebben de Nederlandse gemeenten iets met de machten van 2? Of ligt het aan eigenaardigheden van het tientallig stelsel?
Als we nu maar een vermoeden hadden wat de rare percentages op den duur worden (want het gaat natuurlijk om een limietproces), dan kwamen we wel verder.
We krijgen meer inzicht als we naar de cumulatieve verdeling kijken; in figuur 3 geeft de stip boven de 5 aan: de fractie Nederlandse gemeenten waarvan het inwoneraantal met een 1, 2, 3, 4 of 5 begint. De kruisjes horen bij net zo'n cumulatieve verdeling van de eerste cijfers van de machten van 2.

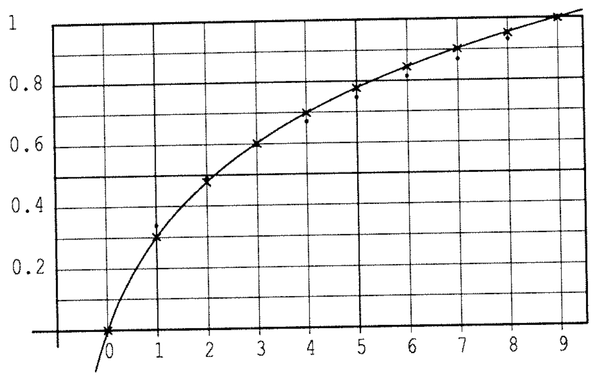

fig. 3: De data van figuur 1 en 2 cumulatief in beeld gebracht, samen met de grafiek van $^{10}\log(x+1)$.

De vloeiende lijn lijkt op de grafiek van $^{10}\log(x+1)$. Dat de grafiek van $^{10}\log(x+1)$ de juiste lijn geeft is precies wat Benford beweert. De wet van Benford wordt meestal in de niet-cumulatieve vorm geformuleerd:

In een verzameling afwijkende getallen is de kans dat het eerste cijfer van een getal het cijfer d is, juist $^{10}\log \frac{d+1}{d}$.

Het voorbehoud 'afwijkend' moet er echt bij. Het is een vaag criterium, maar tabellen met bijvoorbeeld de jaarlijkse regenval, de lengtes van personen of de geboortejaren van de nu nog levende mensen moeten we zeker uitsluiten; zulke getallen schommelen om een gemiddelde.

Vage argumenten

Hoe werkt dat nu toch? De vraag wordt nog nijpender als we wat door zouden experimenteren.
Je zou de inwoneraantallen van de Nederlandse gemeenten ook in dozijnen kunnen uitdrukken bijvoorbeeld. Dat geeft 672 andere getallen. Probeer het en merk dat ook van die getallen zo'n 30% met het cijfer 1 begint. Die horen dan (vaak) wel bij andere gemeenten dan de gemeenten waarvan het echte aantal met een 1 begint.
Of je zou de eerste 100 000 machten van 2 ook eens uit kunnen drukken in het zeventallig stelsel. Dan vinden we alleen 1 t/m 6 als begincijfers. De verdeling voldoet aan de wet van Benford die bij het zeventallig stelsel hoort, waarbij de $^{7}\log$ de rol van de $^{10}\log$ overneemt.

Hier is een eerste vaag argument.
We staan voor de getallenlijn en kijken er naar. Vóór ons liggen de getallen 10 t/m 19; van ons uit gezien ziet het stukje 90 t/m 99 er dan heel klein uit. Of staan we ongeveer bij 100? Maar dan ziet het stuk 900 t/m 999 er veel kleiner uit dan 100 t/m 199. Enzovoort. Een vaag argument, maar er zit toch wel wat in. Vanuit de cijfers bekeken, maar in zekere zin hetzelfde: een natuurlijk getal tussen 900 en 999 is door zijn derde cijfer met een veel betere relatieve nauwkeurigheid bepaald dan een natuurlijk getal van drie cijfers tussen 100 en 199. Bij een verandering van het laatste cijfer van een stapje verandert een met cijfer 9 beginnend getal maar hoogstens 0.1% terwijl een met een 1 beginnend getal dan wel 1% kan veranderen. Je zou dat zo kunnen zeggen: getallen die met het cijfer 9 beginnen krijg je alleen te pakken als je wat nauwkeuriger te werk gaat; ze worden door de semi-toevallige processen van de werkelijkheid dan ook minder vaak tevoorschijn gebracht.
Vaag argument nummer twee: we ervaren de wereld logaritmisch. Zintuigdeskundigen merkten in de vorige eeuw al op dat onze zintuigen eerder gevoelig zijn voor de verhouding tussen waarnemingen dan voor absolute grootheden. Onze zintuigen stellen zich eerst globaal op een bepaald sterktegebied in, en meten dan relatief. Denk aan gevoeligheid voor licht en geluid. Ons toonhoogte gevoel is ook sterk relatief: als we een melodietje wat hoger zingen, maken we de frequenties hoger, maar we houden de frequentieverhoudingen tussen opeenvolgende tonen dezelfde als in het lage liedje.
De DIN- en db-schalen voor lichtgevoeligheid van fotografisch materiaal en voor geluidsterkte werken net zo: een stapje op de schaal is een bepaalde factor energie of iets dergelijks meer. Er wordt als het ware naar de logaritme van de meetwaarden gekeken. Denk er ook aan dat heel veel groeiprocessen die aantallen opleveren (zoals bijvoorbeeld bacteriegroei) exponentiëel verlopen. Tot slot van dit 'logaritmisch' argument kijken we naar de ouderwetse rekenliniaal. Op de rekenliniaal nemen de getallen die met het cijfer 1 beginnen precies 30% van de liniaal in.
Iets minder vaag is de volgende verklaring. Al die gevonden getallen van zeelengtes, oppervlaktes en zo ontstaan niet zomaar, maar vaak door het maken van een aantal vermenigvuldigingen. En kijk nu eens naar de gewone tafels van vermenigvuldiging: van de 100 tafelprodukten beginnen er 21 met het cijfer 1.

Opgave 85
Zoek uw eigen intuïtieve verklaring voor Benford's wet.

In de volgende paragraaf gaan we eerst in op het begrip *gelijkverdeling modulo 1*, omdat het een belangrijk hulpmiddel is om van bepaalde wiskundig gegeven rijen getallen, zoals bijvoorbeeld de rij machten van een vast getal, de Benford-eigenschap aan te tonen. En zo'n brokje wiskundige zekerheid geeft toch wel wat meer vertrouwen in de redelijkheid van de wet van Benford in het algemeen.

De opgerolde getallenlijn

Hier zijn vier getallen met hun logaritmen:

getal	*logaritme*
83450056	7.9214266
17	1.2304489
2056	3.3130231
740221	5.8693614

Het getal voor de punt in de logaritmen is steeds één minder dan de lengte van het getal in cijfers. Het geeft de orde van grootte van het oorspronkelijke getal aan, de plaats van het getal in de rangorde der machten van tien. Na de komma komt pas de informatie die met de cijfers van het getal te maken heeft, de zogenaamde mantisse. Vandaag is alleen dat laatste deel interessant. We willen de gewone getallenlijn zodanig afbeelden, dat die gedeelten achter de komma goed naar voren komen en de gedeelten voor de komma uit zicht raken. Een goede manier is: de getallenlijn strak om een cirkel met omtrek 1 oprollen. Daardoor kunnen we geen onderscheid meer maken tussen 1000.13 en 7.13, maar nog wel tussen 7.13 en 7.14. In figuur 4 zien we het resultaat, waarbij de eerste 300 veelvouden van $\sqrt{2}$ op de opgerolde getallenlijn zijn aangestipt.

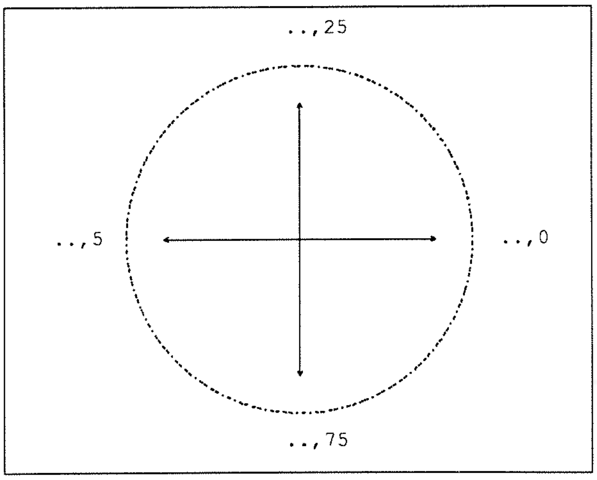

fig. 4: Opgerolde getallenlijn met de eerste 300 veelvouden van $\sqrt{2}$ aangestipt.

Een ding is jammer: de cirkel loopt snel vol en dan is van een verdeling van de punten over de cirkel niet veel meer te zien. Een verbetering is figuur 5: maak van de cirkel een wat bredere band en laat de tekenende computer via een toevalsvariabele de dwarspositie op de band bepalen. De richting vanuit het midden waarin getallen zijn afgebeeld blijft wel hetzelfde.

Als voorbeeld nemen we nu een getallenlijn met wat meer, namelijk 5000, veelvouden van $\sqrt{2}$. Zie figuur 5. De gehele getallen (als die er zouden zijn!) komen nog steeds rechts te liggen bij het pijltje dat naar ..,0 wijst en de gehele getallen plus een halfje nog steeds links bij ..,5.[1]

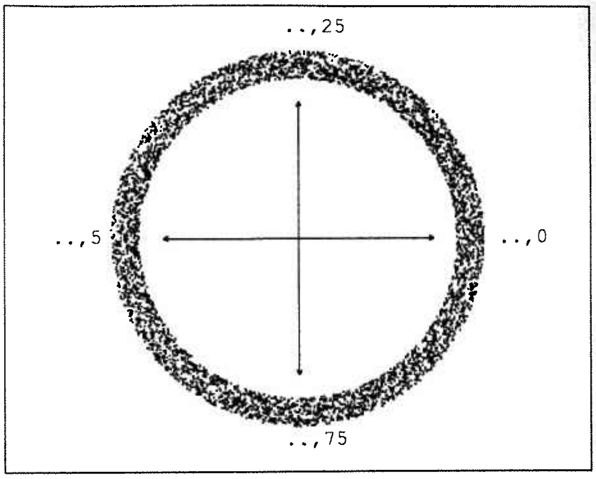

fig. 5: De eerste 5000 veelvouden van $\sqrt{2}$ op de als band afgebeelde opgerolde getallenlijn.

Bij de veelvouden van $\sqrt{2}$ ontstaat blijkbaar een egale band. Nu is wel meteen duidelijk dat er nooit verschillende veelvouden van $\sqrt{2}$ op elkaar zullen komen, want als dat gebeurde zou het verschil van die veelvouden, dat zelf ook een veelvoud van $\sqrt{2}$ is, geheel zijn en dat is in strijd met de irrationaliteit van $\sqrt{2}$.

Er valt te beredeneren dat het overal op de cirkel dringen wordt, maar dat het egaal grijs wordt is nog andere koek. De veelvouden van het nabijgelegen 1.4 gedragen zich heel anders. Zie figuur 6.

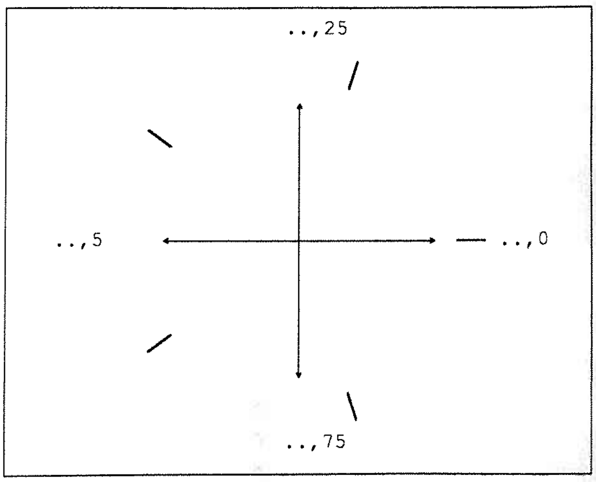

fig. 6: De eerste 5000 veelvouden van 1.4 (modulo 1).

Dat is eenvoudig te verklaren: er komen nu maar vijf verschillende staarten achter de decimale punt voor, namelijk 0.0, 0.2, 0.4, 0.6 en 0.8. In die vijf punten is alles samengebald. Het is een bepaalde manier van kijken naar rijen getallen en er hoort ook wat terminologie bij. Men zegt wel: *we bekijken alles modulo 1*. Dat wil zeggen we delen de getallen door 1 en kijken naar de resten. Zo'n rest heet ook wel: *het fractionele deel van een getal.*

De veelvouden van $\sqrt{2}$ zijn *gelijkverdeeld modulo 1*. Dat begrip moeten we eigenlijk nog preciseren en de bewering zelf bewijzen. Dat gebeurt straks in een apart paragraafje, dat u ook over kunt slaan als u tevreden bent met het intuïtieve idee van de egale grijsheid.

Figuur 7 brengt de tiende logaritmen van de getallen 1 tot en met 5000 op soortgelijke wijze in beeld.

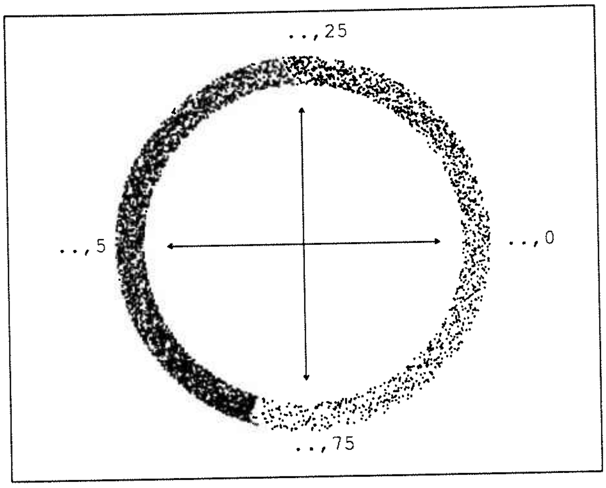

fig. 7: De fractionele delen van de logaritmen van de getallen 1 t/m 5000 afgebeeld op een cirkel met omtrek 1.

Zo op het eerste gezicht hier geen gelijkverdeling modulo 1, en dat is ook wel te verwachten: de log-functie is zo traag dat hij tussen 100 en 1000 maar één rondje maakt, en tussen 1000 en 10 000 ook. Egaal grijs zal het nooit worden.

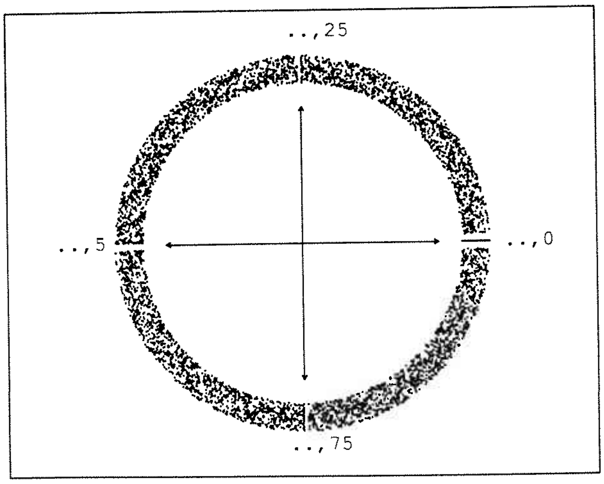

fig. 8: De wortels van de getallen 1 t/m 5000 modulo 1. Wordt het gelijkverdeling?

In figuur 8 tot slot de wortels van de getallen 1 t/m 5000. Hier valt wel het een en ander aan uit te zoeken!

Opgave 86
Er zit een gaatje rond 0 bij de wortels van de getallen 1 t/m 5000. Op den duur (als we veel verder dan 5000 gaan) loopt dat gaatje wel dicht, maar waarom blijft het toch zo lang open?

Tip: kijk naar de getallen waarvan de wortel geheel is en hun voor- of achterbuur. Vergelijk de wortels

Opgave 87
De wortel van een getal komt nooit precies op ..,5 terecht. Verklaar ook het gaatje rond ..,5.

Omdat het met eerdere opgaven uit deze puzzelrubriek te maken heeft gaan we even dieper in op het gedragsverschil tussen de log- en de wortel-functie. Met de wet van Benford heeft het niet direct te maken; u kunt het volgende paragraafje ook overslaan.

Trage log, snelle wortel

De wortel-functie loopt harder dan de log-functie. Het verschil in gedrag wat betreft gelijkverdeling heeft daar mee te maken.

Er is een algemene stelling die een verband legt tussen de afgeleide van zo'n op $(0, \infty)$ naar ∞ stijgende functie f en het gelijkverdeeld modulo 1 zijn van de getallen $f(n)$, $n = 1,2,3 \ldots$.

Die stelling zegt (informeel geformuleerd): om gelijkverdeling te krijgen moet de afgeleide van boven naar nul dalen als n naar ∞ gaat (dat wil zeggen: f stijgt wel, maar steeds trager), maar ook weer niet te snel want $n \times f'(n)$ moet naar ∞ gaan. (Zie: G. Pólya, 1925).

Bij de logaritme is dat laatste niet het geval, bij de wortelfunctie wel. Het hangt er mee samen hoe lang het duurt voor de functie één rondje over de cirkel aflegt. Naarmate de afgeleide kleiner is, duurt dat namelijk langer. Kijken we naar de wortels van natuurlijke getallen, dan zien we dat we van het ene kwadraat naar het volgende kwadraat, zeg van k^2 naar $(k+1)^2$ moeten om met de wortels van die getallen een rondje over de cirkel te lopen; een stuk van weliswaar $(2k+1)$, maar dat is geen substantieel deel van de $(k+1)^2$ getallen die we dan gehad hebben. Bij de log-functie moeten we van bijvoorbeeld k naar $10k$ om een rondje te maken, en dat is een vast deel van de $10k$ getallen die we dan gehad hebben, hoe groot k ook wordt. Vandaar dat in het plaatje ook zo'n duidelijke grens bij de 5000 te zien was: daar dringen de stipjes zich hevig samen. In het plaatje moeten ook nog eens 5000 stipjes volgen voor de 'log' weer bij de ..,0 is.

De getallen n, waarvan de wortel het cijfer 1 direct achter de decimale punt heeft, vinden we in figuur 8 tussen ..,0 en ..,1. Dat is juist het tiende deel van de cirkel en bij benadering komt dat wegens de zojuist ontdekte gelijkverdeling overeen met ongeveer het tiende deel van de getallen 1 tot en met 5000. Naarmate we meer getallen nemen, wordt de verhouding 1:10 steeds beter benaderd. Opgave 82 van deze rubriek vroeg dat te bewijzen. Het kon ook zonder de zo juist aangegeven middelen, maar dat is toch wel lastig!

Benford en gelijkverdeling modulo 1

We begonnen het onderdeel 'de opgerolde getallenlijn' met vier getallen en hun logaritmen.
Dat was niet zomaar. De combinatie van logaritme en rest modulo 1 nemen, legt namelijk juist het verband tussen een getal en zijn begincijfer.
Als de logaritme van een getal bijvoorbeeld 7.564286 is, is meteen te zien dat het getal zelf met het cijfer 3 begint. De 7 doet er namelijk niet toe. Het getal is

$$17^7 \times 10^{0.564286}$$

en het eerste cijfer wordt alleen door

$$10^{0.564286}$$

bepaald. Omdat 0.564286 tussen $^{10}\log 3$ en $^{10}\log 4$ in ligt, moet het eerste cijfer dus een 3 zijn.
Aan de rest modulo 1 van een getal is dus te zien wat het eerste cijfer van een getal is, en de grenzen liggen bij de resten $^{10}\log 2$, $^{10}\log 3$, ... , $^{10}\log 9$.
Laten we hier maar zo'n cirkelplaatje aan wijden, waar we die grenzen als getekende stralen in aanbrengen: figuur 9. De punten zijn nu dus de resten modulo 1 van de logaritmen van een rij getallen.

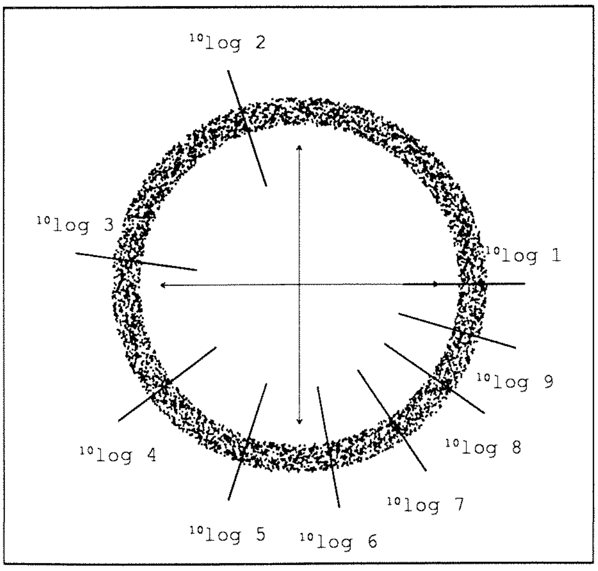

fig. 9: De logaritmen, modulo 1 genomen van een rij getallen, in gedeeld in sectoren. In elke sector horen de punten bij getallen met hetzelfde begincijfer.

Als de rij getallen zodanig is, dat de band inderdaad egaal grijs is, komt er juist de fractie $^{10}\log\frac{4}{3}$ van de punten in de sector die bij de getallen hoort die begincijfer 3 hebben.
In het algemeen geldt dus:

> *Als de logaritmen van een rij getallen gelijkverdeeld modulo 1 zijn, dan voldoet de rij aan de Wet van Benford.*

In figuur 9 zijn de machten van 2 als uitgangspunt genomen; de logaritmen daarvan zijn de veelvouden van log 2. Die zijn gelijkverdeeld modulo 1, net als de veelvouden van $\sqrt{2}$. Dus:

> *de rij van de machten van 2 voldoet inderdaad aan de wet van Benford.*

Er zit nog één gat in ons betoog: het feit dat die veelvouden van zo'n irrationaal getal inderdaad gelijkverdeeld modulo 1 zijn. Dat is niet zo simpel en we lichten het toe in een paragraafje dat weer kan worden overgeslagen door degenen die 'op den duur egaal grijs' prefereren boven een exacte definitie van gelijkverdeling modulo 1of niet houden van complexe e-machten.
Met behulp van het voorgaande en enige inventiviteit kan nu wel een aantal opmerkelijke resultaten worden geboekt. Maar let eerst nog even op dit kleinigheidje:

Opgave 88
Waarom is $^{10}\log 2$ irrationaal?

Wat met de machten van 2 lukt, kan vaker. Maar pas op:

Opgave 89
De getallenrij $a \times b^n$ is voor de meeste a en b wel een Benford-rij. Soms echter niet. Wanneer?

Opgave 90
Toon aan dat van de Fibbonacci-getallen (1, 1, 2, 3, 5, 8, 13,...) 30% met het cijfer 1 begint.

Tip: De Fibbonacci-getallen zijn van de vorm $c(\tau^n + (-\tau)^{-n})$ met $\tau = \frac{1}{2}(1 + \sqrt{5})$.

Opgave 91
De rij der kwadraten voldoet net als de rij der natuurlijke getallen zelf niet aan de wet van Benford.

Omdat het gewoon er om vraagt uitgezocht te worden:

Opgave 92
Voldoet de rij der priemgetallen aan de wet van Benford?

Tip: waag je hier niet aan, tenzij je een makkelijke manier weet om redelijk nauwkeurig te bepalen hoeveel priemgetallen er tussen 1 000 000 000 en 2 000 000 000 liggen. Het antwoord is: Nee.

Diepere gronden: Het criterium van H. Weyl

We beginnen met een nauwkeuriger omschrijving van het begrip gelijkverdeling modulo 1.
Neem een rij reële getallen:

$$\{a_n\}_{n=1}^{\infty}$$

Neem een deelinterval $[\alpha, \beta)$ van het interval $[0, 1)$. We kijken naar de resten modulo 1 van de getallen die in dat deelinterval vallen; spreek af:

$A(n, \alpha, \beta)$ = het aantal natuurlijke getallen k in $[1, n]$, waarvoor a_n(modulo 1) in $[\alpha, \beta)$ ligt.

Gelijkverdeling van de rij betekent dat al die deelintervallen met de juiste frequentie worden geraakt, dat voor al die intervallen geldt:

$$\lim_{n \to \infty} \frac{A(n, \alpha, \beta)}{n} = \beta - \alpha$$

Je kunt dat ook anders lezen, door te denken aan de periodieke functie T die op alle intervallen $[k+\alpha, k+\beta]$ juist 1 is en daarbuiten 0. Er staat dan in feite:

$$\lim_{n \to \infty} \frac{T(a_1) + \ldots + T(a_n)}{n} = \int_0^1 T(x)\, dx$$

Met andere woorden: Het gemiddelde van de functiewaarden op de rij nadert tot het gemiddelde van de functie over het interval $[0,1]$.
Omdat andere periodieke functies te benaderen zijn met samenstellingen van zulke blokfuncties als deze T, geldt die laatste relatie dan ook voor alle andere functies waarvoor de integraal in het rechterlid bestaat. Zo kun je dan voor de functie $T(x)$ ook bijvoorbeeld $sin(2\pi x)$ nemen en je vindt dan dat het gemiddelde van $sin(2\pi a_n)$ tot de integraal van $sin(2\pi x)$ over $[0,1)$ nadert, (d.w.z. tot 0) als de rij van de a_n gelijkverdeeld modulo 1 is.
Het zogenaamde criterium van H. Weyl (zie weer Pólya, 1925) zegt dat je de boel ook om kan draaien: als limiet en integraal gelijk zijn voor een 'geschikte' verzameling functies, dan geldt dat ook voor alle (integreerbare) periodieke functies, dus ook voor al die blokvormige functies T, en is er dus gelijkverdeling.
'Geschikte' verzamelingen functies zijn die, waarmee de andere periodieke functies te benaderen zijn. In aanmerking komt de verzameling functies:

$$\{\sin(2\pi kx), \cos(2\pi kx)\}_{k=0}^{\infty}$$

maar het is handiger werken met:

$$\{e^{2k\pi ix}\}_{k=-\infty}^{k=\infty}$$

omdat met die complexe e-machten veel beter te rekenen valt.
Als k geheel maar ongelijk 0 is, dan is:

$$\int_0^1 e^{2k\pi ix} dx = 0$$

Als k wel nul is, dan is de e-macht constant 1 en geldt de gelijkheid van de limiet en de integraal vanzelf.
Nu dit alles toegepast op de veelvouden van $\sqrt{2}$, van $\log 2$ of van enig ander irrationaal getal α. Blijkbaar moet voor k ongelijk 0 gelden:

$$\lim_{n \to \infty} \frac{e^{2k\pi i\alpha} + \ldots + e^{2k\pi in\alpha}}{n} = 0$$

De noemer is de som van een meetkundige reeks ....

Opgave 93
.... en werk dit zelf verder af. Ontdek daarbij waarom de irrationaliteit van α essentieel is, en dat het ook de enige eigenschap van α is die gebruikt wordt.

Nog even de tafels van vermenigvuldiging!

De eerder genoemde observatie dat van de 100 gewone tafelprodukten er al 21 met een 1 beginnen kunnen we nog wat uitbouwen.
We lieten een computer $a \times b \times c \times d \times e$ uitrekenen waarbij a, b, c, d en e onafhankelijk van elkaar de getallen 1 tot en met 9 doorlopen. We sorteerden naar eerste cijfer en Benford klopte al heel aardig.
Adhikari en Sarkar (1968) bewezen dat dit te verwachten was. Zij laten zien dat als getallen uniform uit [0,1] worden gekozen de verdeling van de begincijfers tot Benford's wet nadert, naarmate de vermenigvuldigingen langer worden.
Na dit succes maakten we het bonter. We lieten weer 100 000 van die produkten uitrekenen, maar nu werden de waarden steeds onafhankelijk door het toeval uit de rij (1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9) gekozen. Resultaat:

eerste cijfer	1	2	3	4	5	6	7	8	9
frequentie (%)	28	18	13	10	8	7	6	5	4

Zou dat bij grotere vermenigvuldigingen nog beter naar Benford neigen?
We geloven van wel, maar ook door de statistici is het laatste woord hierover nog niet gesproken. In ieder ge-

val geeft het onderbouwing voor het argument dat veel meetwaarden in werkelijkheid bereikt worden door enkele vermenigvuldigingen en dat daarom de nadruk op het begincijfer in de aard der dingen ligt.

Tot besluit

Men kan argumenteren: ja, die statistische en getaltheoretische stellingen kloppen misschien wel, maar dat zegt nog niets over de werkelijkheid waar die rare verdelingen optreden.
Dat is juist, er moet nog een schakel bij: die van werkelijkheid naar wiskundig model. Die schakel is uiteraard vager; een stukje mogelijke schakel hebben we aangegeven. Raimi (1976) vergelijkt het met de wetten van Newton die de planeetbanen 'verklaren'. Daar maakt mooie wiskunde duidelijk dat die banen ellipsvormig zijn, maar je moet dan wel een schakel aannemen: de wet over de zwaartekracht.

Nog een slotopmerking: deze aflevering van de puzzelrubriek is wat opgaven betreft afwijkend. De opgaven ondersteunen het verhaal, maar functioneren niet als onafhankelijke puzzels. Een volgende keer – in de komende meetkunde-special – wordt dat vast weer anders.

Een correctie

In de vorige aflevering (april '92) van deze rubriek zit een fout in de illustratie op bladzijde 42, rechtsboven.
In de waan dat met mooie elektronische spullen de tekening van Piet Lemmens verfraaid kon worden heb ik (ajg) de zaak verprutst.
Piet Lemmens stuurde naar aanleiding van de vorige aflevering ook nog een prachtige zelfoverlappende bouwplaat op. Het is een bouwplaat van een afgeknotte driezijdige piramide. De piramide heeft een gelijkzijdige driehoek als grondvlak en is als geheel drietallig symmetrisch. Als de oorspronkelijke piramide scherp genoeg is kan bij de afgeknotte piramide een zelfoverlappende bouwplaat horen.
Het grondvlak van de ingezonden bouwplaat speelt bij het overlappen geen rol. Daarmee is dan wel een record gevestigd: een convex vijfvlak met zelfoverlappende uitslag, gebaseerd op in feite slechts vier zijvlakken.
Omdat een viervlak nooit een zelfoverlappende uitslag kan leveren, valt het record van Piet Lemmens niet meer te verbeteren.
Natuurlijk nog dit:

Opgave 91
Wat was er eigenlijk fout aan dat plaatje en zoek uit hoe de uitslag van de afgeknotte piramide in elkaar moet zitten.

Noten

[1] We beschouwen de '0' nooit als echt eerste cijfer van een getal; in gevallen als 0.0087 kiezen we de wetenschappelijke notatie: 8.7E-3, dan is het begincijfer eenduidig bepaald. Dit is geheel in overeenstemming met het gebruik van logaritmetafels, waarbij de logaritme van 0.0087 ook tamelijk achterin het boek staat. Vandaar de misschien wat onverwachte 11% in plaats van 10%.
[2] In de cirkelplaatjes nemen we tijdelijk even de komma als decimaalscheider. Dan is de notatie ..,25 namelijk zo handig. De puntjes duiden aan dat er van alles mag staan, maar dat het wel gaat om getallen met '25' achter de decimaalscheider.

Literatuur

Adhikari, A. K., Sarkar, B. P. (1968). 'Distributions of most significant digit in certain functions whose arguments are random variables', *Indian Journal of Statistics*, *30*, 47-58.
Benford, F. (1938). The law of anomolous numbers. *Proc. Amer. Phil. Soc.*, *78*, 551-572.
CBS. *Statistisch zakboek 1991.*
Maanen, H. van (1988). *De Wet van...* Uitgeverij Boom, Meppel.
Pólya, G., Szegö, G. (1925). *Aufgaben und Lehrsätze aus der Analysis.* Springer Verlag, Berlijn.
Raimi, R. A. (1976). 'The first digit problem', *American Mathemathical Monthly*, *83*, 521-538.