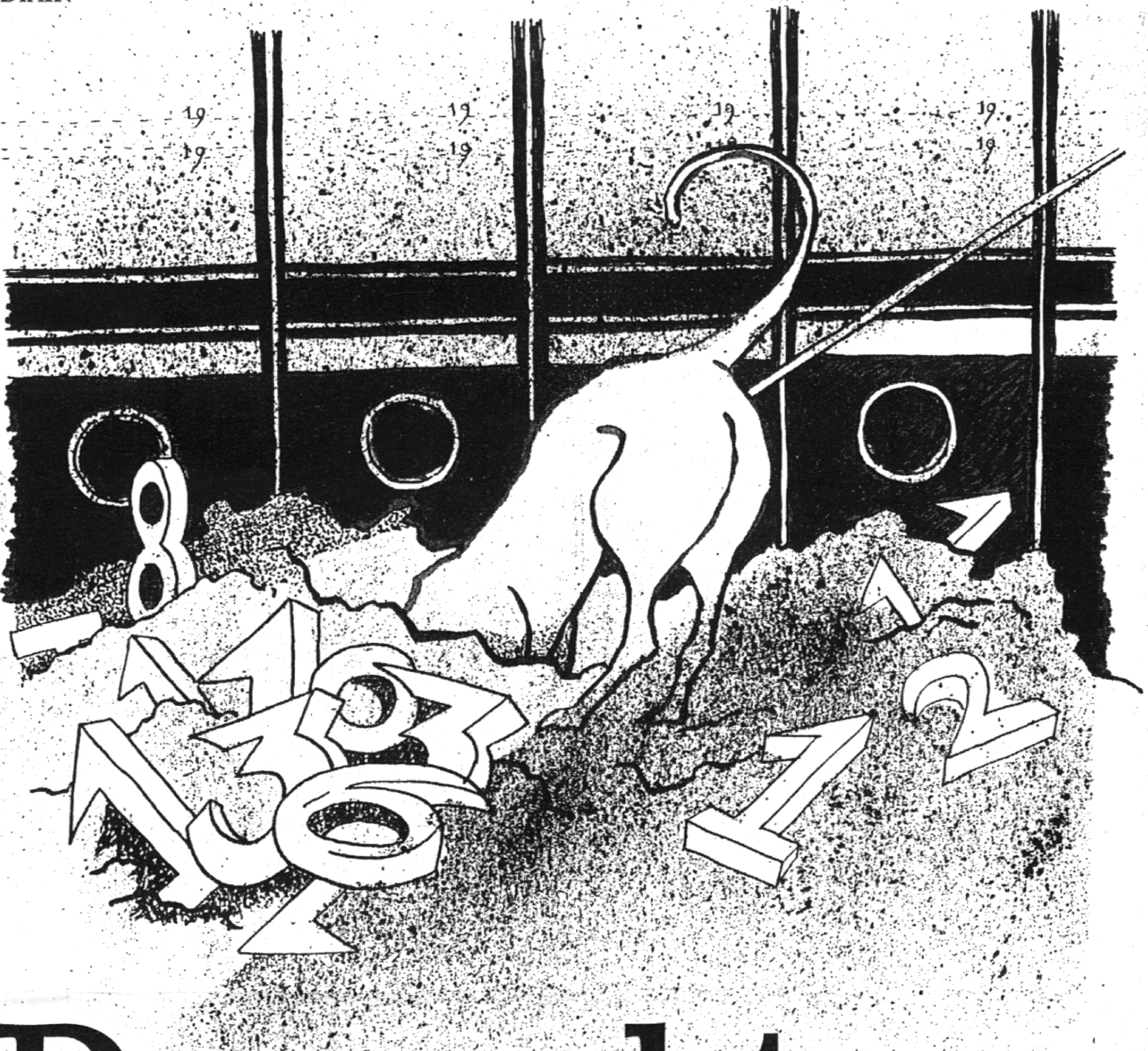


De macht van het getal één

Dertig procent van de grote getallen begint met een 1. Waarom dat zo is, is punt twee. Het primaat van het getal 1 helpt in ieder geval om frauderende boekhouders op te sporen.

ROBERT MATTHEWS - © NEW SCIENTIST

Alex had geen idee op welk akelig geheim hij zou stuiten toen hij vroeg of hij zijn zwager mocht helpen bij de boekhouding. Voor een scriptie in het kader van zijn accountantsopleiding aan de Universiteit van Saint Mary in Halifax, Nova Scotia moest Alex echte bedrijfsgegevens analyseren, en daarvoor koos hij zijn zwagers ijzerwarenwinkel.

Hij liep de verkoopcijfers van dat jaar door en kon daar niets vreemds aan ontdekken. Toch deed hij braaf wat hij voor zijn scriptie doen moest en voerde het vreemde ritueel uit dat Mark Nigrini, zijn hoogleraar boekhoudkunde, verplicht had gesteld. Hij ging netjes na hoeveel van die verkoopcijfers met een 1 begonnen, en kwam uit op 93 procent. Hij leverde zijn werk in en dacht er verder niet meer over na.

Toen Nigrini later het werk van zijn studenten nakeek, hoefde hij maar even naar die uitkomst te kijken om te beseffen dat hier een ongemakkelijke situatie dreigde. Zijn wantrouwen groeide toen hij Alex' analyse van zijn zwagers boekhouding verder doorkeek. Geen van de verkoopcijfers begon met de cijfers 2 tot en met 7, slechts vier begonnen er met een 8, en 21 met een 9. Nog een paar controles, en Nigrini's laatste twijfel verdween: Alex' zwager was een oplichter die systematisch met z'n gegevens knoeide om de bank en de belastingen op een dwaalspoor te brengen.

De zwager had het aardig geprobeerd. Op het eerste gezicht was er weinig verdachts aan de verkoopcijfers. Maar dat was het 'm nou net: ze waren veel te regelmatig, en daarom vielen ze door de mand. Want wat Nigrini wist – en Alex' zwager overduidelijk niet – was dat de cijfers waaruit de getallen van de verkoopgegevens zijn opgebouwd, moeten voldoen aan een wiskundige regel die meer dan een eeuw geleden per ongeluk werd ontdekt. Deze regel, die bekend staat als de wet van Benford, geldt voor een verbijsterend aantal verschijnselen: van aandelenkoersen via volkstellingscijfers tot gegevens over de warmtecapaciteit van chemicaliën. Zelfs cijfermateriaal uit een willekeurige krant zal voldoen aan de eis van deze wet, die luidt dat ongeveer dertig procent van de getallen begint met een 1, achttien procent met een 2, tot maar net 4,6 procent met een 9.

Vieze bladzijden

Die regel is zo onverwacht dat veel mensen weigeren te geloven dat hij waar is. Een echt degelijke, wiskundige verklaring voor deze wet is ook pas enkele jaren bekend. Maar na lang te zijn beschouwd als niet meer dan een wiskundige curiositeit, staat de wet van Benford opeens volop in de belangstelling. Het Amerikaanse Instituut voor Interne Accountants organiseert speciale cursussen over de toepassing van de wet van Benford bij onderzoek naar fraude, een terrein waarop die wet als de grootste vooruitgang sinds jaren wordt beschouwd.

Het verhaal over de ontdekking van de wet is al even vreemd als ▸▸

intermediair *19 augustus 1999*
35e jaargang nummer 33

Belastingfraudeurs en creatieve boekhouders opgepast!

►► de wet zelf. In 1881 stuurde de Amerikaanse astronoom Simon Newcomb een berichtje naar het *American Journal of Mathematics* over een eigenaardigheid die hem was opgevallen in de boeken met logaritmetafels die toen nog veelvuldig voor wetenschappelijke berekeningen werden gebruikt. De eerste bladzijden van zulke boeken leken veel eerder vies te worden dan de laatste.

De voor de hand liggende verklaring was verbijsterend. Om de een of andere reden begonnen de getallen waarmee men berekeningen uitvoerde vaker met een 1 dan met een 8 of een 9. Newcomb bedacht een formule die het patroon aardig dekte: kennelijk mag de natuur cijfers graag zo ordenen dat het deel van de getallen dat met cijfer C begint, gelijk is aan $^{10}\log (1 + 1/C)$.

Er waren geen overtuigende argumenten waarom die formule van toepassing zou zijn, en daarom viel Newcombs bericht niemand echt op. Het vieze-bladzijdeneffect raakte voor meer dan een halve eeuw in de vergetelheid. Maar in 1938 ontdekte Frank Benford, een natuurkundige die werkte voor General Electric in Amerika, dat effect opnieuw, en hij kwam tot dezelfde formule als Newcomb. Benford ging verder. Hij verzamelde meer dan twintigduizend getallen – variërend van gegevens over de grootte van stroomgebieden van rivieren tot getallen uit oude tijdschriftartikelen – en toonde aan dat ze allemaal dezelfde grondregel volgden: ongeveer dertig procent begon met het cijfer 1, achttien procent met een 2, enzovoort.

> *De prijzen van flesjes bier voldoen natuurlijk niet aan de wet van Benford*

Net als Newcomb had ook Benford geen echt goede verklaring voor het bestaan van die regelmatigheid. Toch is zijn naam, dankzij de enorme hoeveelheid bewijsmateriaal die hij aandroeg, sindsdien wel met die wet verbonden gebleven.

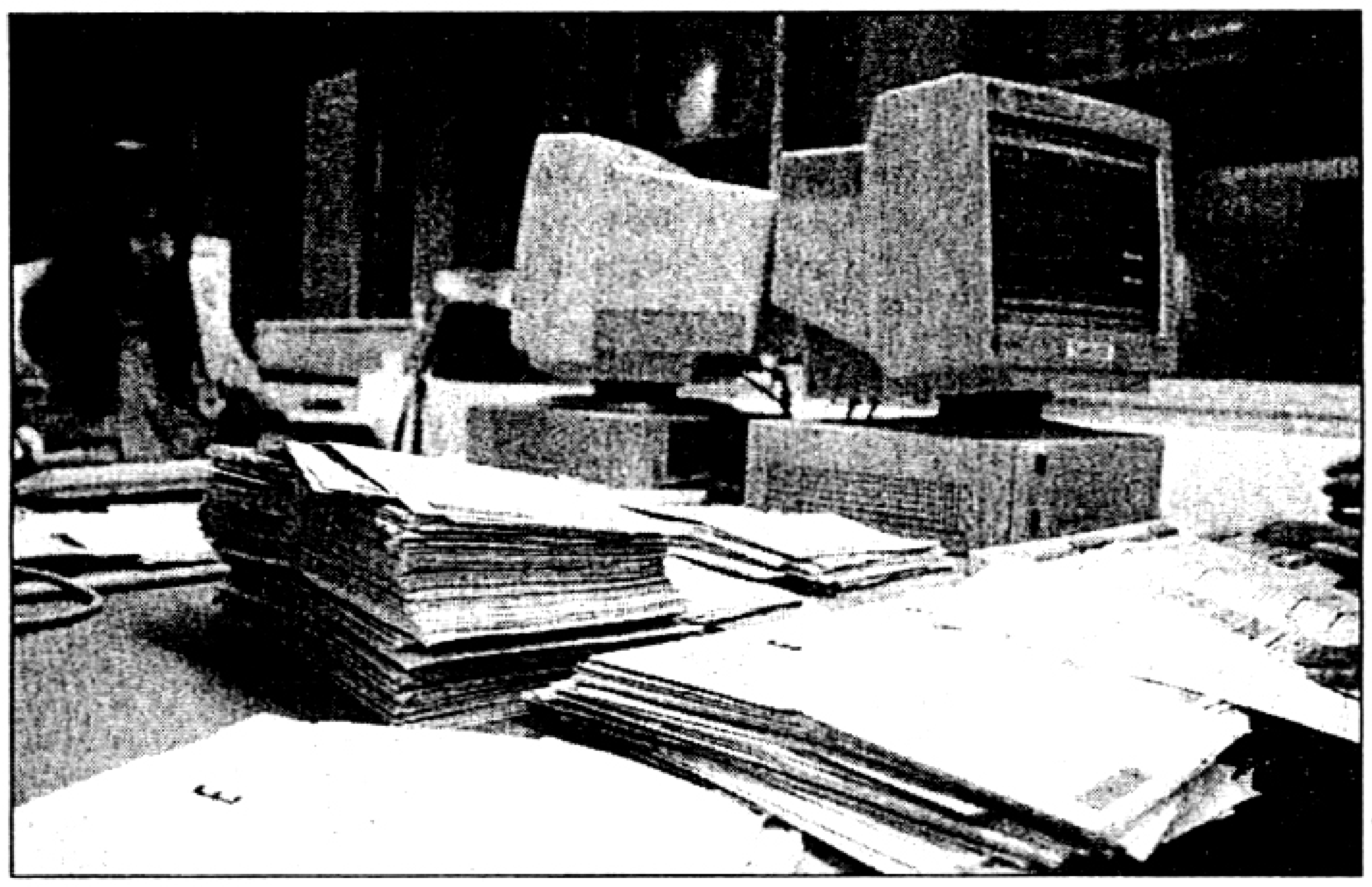
De planeet Zob

Het duurde nog bijna een kwart eeuw voordat iemand met een aannemelijk antwoord kwam op de cruciale vraag waarom die wet in hemelsnaam zou opgaan voor getallen uit zoveel verschillende domeinen. De eerste grote doorbraak kwam in 1961 dankzij enig fraai onorthodox denkwerk van Roger Pinkham, een wiskundige van de Rutgers Universiteit in New Jersey. Stel je voor, zei Pinkham, dat er echt een wet bestaat die vastlegt uit welke cijfers de getallen die natuurlijke verschijnselen als de stroomgebieden van rivieren en de eigenschappen van verbindingen beschrijven, zijn opgebouwd. Dan moet zo'n wet onafhankelijk zijn van de gebruikte eenheden. Zelfs de bewoners van de planeet Zob, die land meten in grondeki's, zouden bij stroomgebieden precies dezelfde verdeling van cijfers moeten vinden als wij bij onze metingen in hectaren. Maar hoe kan dat als een grondeki overeenkomt met 87.331 hectare?

Het antwoord, vervolgt Pinkham dan, is dat je ervoor zorgt dat de verdeling van de cijfers zich niets aantrekt van een verandering van grootheid. Stel dat je het stroomgebied van een miljoen verschillende rivieren in hectaren weet. Als je die waarden omrekent naar grondeki's dan zal elk individueel getal beslist veranderen. Maar de verdeling van de cijfers zou, over het geheel genomen, hetzelfde patroon moeten volgen. Die verdeling zou dan wat heet 'schaalonafhankelijk' zijn.

Pinkham toonde langs wiskundige weg aan dat de wet van Benford inderdaad schaalonafhankelijk is. Maar – en dat is het cruciale punt – hij toonde ook aan dat deze wet de enige manier beschrijft waarop je cijfers schaalonafhankelijk kunt verdelen.

Het werk van Pinkham gaf flink wat steun aan de geloofwaardigheid van de wet, en bracht anderen ertoe op zoek te gaan naar mogelijke toepassingen. Maar één serieuze vraag bleef: wat voor soort getallen zouden voldoen aan de wet van Benford? Twee vuistregels stonden al snel vast. Allereerst moest het aantal getallen groot genoeg zijn om de voorspelde verhoudingen de kans te geven tot uitdrukking te komen. Ten tweede zouden de getallen niet aan kunstmatige grenzen onderhevig moeten zijn: ze zouden zo goed als elke gewenste waarde moeten kunnen aannemen. Het is bijvoorbeeld duidelijk onzin om te verwachten dat de prijzen van tien verschillende merken bier voldoen aan de wet van Benford. Niet alleen is dat aantal gewoon te klein, maar – belangrijker nog – marktwerking zorgt ervoor dat die prijzen binnen een vast en nauw bereik blijven.



JIRI BÜLLER

Wilt u frauderen met aftrekposten? Gebruik voldoende enen!

De Moeder aller Verdelingen

Aan de andere kant kunnen écht willekeurige getallen niet aan de wet van Benford voldoen: het aandeel van de verschillende cijfers in zulke getallen is, per definitie, gelijk. De wet van Benford geldt dus voor het gebied tússen strak beperkte en volledig ongebonden getallen.

Wat dat precies betekent, bleef een raadsel tot de wiskundige Theodore Hill van het Georgia Instituut voor Technologie in Atlanta, nu drie jaar geleden, ontdekte wat kennelijk echt de oorsprong is van de wet van Benford. Het zit 'm in de verschillende manieren waarop verschillende soorten metingen gespreid zijn. Uiteindelijk is alles wat we in het heelal kunnen meten het resultaat van een of ander proces: de willekeurige botsingen van atomen, bijvoorbeeld, of de wetmatigheden der genetica. Wiskundigen weten al lang dat de spreiding van de waarden bij elk van die processen de een of andere basale wiskundige regel volgt. De lengte van bankdirecteuren, bijvoorbeeld, volgt de klokvormige kromme van Gauss, de dagelijkse temperatuur stijgt en daalt in een golfvormig patroon, terwijl er tussen kracht en frequentie van aardbevingen een logaritmisch verband bestaat.

Stel nu dat je willekeurige handjes vol getallen uit een cocktail van zulke verdelingen haalt. Hill bewees dat naarmate je meer van zulke getallen vergaart, de cijfers daarin steeds nauwer beantwoorden aan één enkele, heel specifieke wet. Die wet is een soort uiteindelijke verdeling, de 'Verdeling van alle Verdelingen'. En hij toonde aan dat die de wiskundige vorm heeft van ... de wet van Benford.

Geen grapje

Mark Nigrini, Alex' voormalige docent en nu hoogleraar boekhoudkunde aan de Zuidelijke Methodistenuniversiteit in Dallas, beschouwt het theorema van Hill als een cruciale doorbraak. Dankzij Hill kon Nigrini anderen er ook van overtuigen dat de wet van Benford veel meer is dan een wiskundig grapje. De afgelopen jaren ontwikkelde Nigrini zich tot de drijvende kracht achter een alles behalve frivole toepassing van de wet: het opsporen van fraude.

In zijn baanbrekende proefschrift uit 1992 toonde Nigrini aan dat veel cruciale boekhoudkundige gegevens, van verkoopcijfers tot aftrekposten, de wet van Benford volgen, en dat je afwijkingen van die wet met eenvoudige statistische toetsen snel kunt constateren. Nigrini noemt zijn techniek om fraude te achterhalen *digital analysis*, en het succes daarvan krijgt steeds meer aandacht in de zakenwereld en daarbuiten.

Enkele van de eerste gevallen – met inbegrip van de gehaaide manipulaties van Alex' boekhoudende zwager – kwamen boven water tijdens door Nigrini opgezet studentenonderzoek. Maar al gauw gebruikte hij zijn cijferanalyse om veel aanzienlijker fraudegevallen te ontmaskeren.

Een recent geval betrof een Amerikaanse touroperator met een eigen motelketen. Met behulp van cijferanalyse ontdekte de interne accountant van het bedrijf dat er iets niet klopte aan de opgaven van het hoofd ziektekosten van het bedrijf. 'Men controleerde of de eerste twee cijfers van de uitkeringen in verband met ziektekosten voldeden aan de wet van Benford, en daarbij kwam aan het licht dat een onverwacht groot aantal opgaven begonnen met 65', vertelt Nigrini. 'Bij nader onderzoek kwamen dertien betalingen aan het licht van bedragen tussen de 6.500 en 6.599 dollar op vervalste facturen voor hartoperaties die het hoofd had verwerkt, en waarbij de cheques uiteindelijk in haar handen terechtkwamen.'

Ze werd gepakt dankzij de wet van Benford, hoe zorgvuldig ze ook had geprobeerd die declaraties zo plausibel mogelijk te laten lijken. 'Ze zag er met zorg op toe dat ze op naam kwamen te staan van werknemers bij motels met een groter dan gemiddeld aantal oudere medewerkers', vertelt Nigrini. 'De analyse bracht ook andere frauduleuze vorderingen voor in totaal zo'n miljoen dollar aan het licht.' ◆

Robert Matthews werkt als wetenschapsjournalist voor de *Sunday Telegraph*.
Digital Analysis Tests and Statistics, geschreven en gepubliceerd door Mark Nigrini, is te bestellen via mark_nigrini@msn.com
Om redenen van privacy kreeg Alex een gefingeerde naam.

Vertaling: Bart Voorzanger