De wet van Benford

In 1881 ontdekte Simon Newcomb een bijzondere eigenschap van sommige reeksen getallen. Hij keek steeds naar het *begincijfer* van de getallen in zo'n reeks en constateerde dat daarbij lage begincijfers veel vaker voorkomen dan hoge begincijfers.
Een voorbeeld van zo'n reeks getallen zijn de beurskoersen die elke dag in de krant verschijnen. De lijst met beurskoersen van vrijdag 10 september 2004 begon met de getallen:

17,75 9,15 5,30 28,07 11,02 6,66 39,67 18,73 1,59 1,53 24,29 41,00

20,37 42,31 6,32 5,03 26,08 19,33 10,77 19,39 50,15 1,54 21,86 13,64

Je kunt hier bijvoorbeeld zien dat bij deze reeks getallen het begincijfer 1 veel vaker voorkomt dan het begincijfer 6 (tien keer tegen twee keer).

In 1938 deed Frank Benford uitgebreid onderzoek naar dit verschijnsel. Hij bekeek onder andere de wateroppervlakte van een groot aantal rivieren. Het resultaat van zijn onderzoek vind je in tabel 2.

tabel 2 **wateroppervlakte van rivieren**

begincijfer	1	2	3	4	5	6	7	8	9
frequentie	104	55	36	38	24	29	18	14	17

Je ziet dat lage begincijfers vaker voorkomen dan hoge.

3p **9** ☐ Bereken in hoeveel procent van de waarnemingen in tabel 2 het begincijfer 1, 2 of 3 is.

Het blijkt dat dit verschijnsel bij veel reeksen van getallen optreedt, bijvoorbeeld ook bij de lengte van rivieren of bij inwoneraantallen van gemeenten.
Benford heeft een formule opgesteld waarmee je in dergelijke situaties de relatieve frequentie van de verschillende begincijfers kunt benaderen. Deze formule, die bekend staat als *de wet van Benford*, ziet er als volgt uit:

$$F(n) = 100 \cdot \log\left(\frac{n+1}{n}\right)$$

In deze formule is $F(n)$ het percentage getallen met het begincijfer n ($n = 1, 2, 3, \ldots, 9$).
Bij getallenreeksen die voldoen aan de wet van Benford zal bijvoorbeeld ongeveer 17,6% van de getallen met het cijfer 2 beginnen, want $F(2) = 100 \cdot \log\left(\frac{3}{2}\right) \approx 17{,}6$. En in ongeveer 60,2% van de gevallen zal het begincijfer 1, 2 of 3 zijn, want $F(1) + F(2) + F(3) \approx 60{,}2$.

Benford heeft aangetoond dat ook de oneindige rij opeenvolgende machten van het getal 2, de rij 1, 2, 4, 8, 16, … , voldoet aan deze wet.
Maar het is de vraag of de wet van Benford ook geldt wanneer je je beperkt tot een beginstuk van deze rij. Om dat na te gaan, tellen we hoe vaak de cijfers 1, 2 of 3 als begincijfer voorkomen bij de eerste twaalf getallen uit deze rij.

4p **10** ☐ Onderzoek of het percentage getallen met begincijfer 1, 2 of 3 bij deze twaalf getallen in overeenstemming is met de wet van Benford.

We bekijken in de rest van deze opgave een reeks getallen die voldoet aan de wet van Benford.

Uit deze reeks kiezen we aselect 160 getallen. Zoals we hierboven hebben gezien, is voor elk van deze getallen de kans dat het begincijfer 1, 2 of 3 is, gelijk aan 0,602.

4p **11** ☐ Bereken in vier decimalen nauwkeurig de kans dat meer dan 100 van deze getallen als begincijfer 1, 2 of 3 hebben.

De getallen die voorkomen in een groot financieel overzicht, voldoen meestal bij benadering aan de wet van Benford. Accountants maken daar tegenwoordig vaak gebruik van om mogelijke fraude bij het opstellen van dergelijke overzichten op te sporen. Zij gaan er daarbij van uit dat het bewust manipuleren van getallen door fraudeurs een andere verdeling van begincijfers oplevert dan de wet van Benford voorspelt. Als bijvoorbeeld 7% van de getallen in een financieel overzicht met het cijfer 9 begint, zal men dit overzicht vrijwel zeker nader onderzoeken. Immers: $F(9) \approx 4{,}6$ en het geconstateerde percentage is ongeveer anderhalf keer zo veel.
Bij het accountantskantoor Levrek & Partners neemt men uit een financieel overzicht een steekproef van 800 getallen. Er zal nader onderzoek worden verricht als het aantal getallen met begincijfer 8 meer dan 20 afwijkt van het door de wet van Benford voorspelde aantal. In de steekproef van 800 worden 62 getallen gevonden met begincijfer 8.

4p **12** ☐ Onderzoek of dit voor Levrek & Partners voldoende aanleiding is voor nader onderzoek.