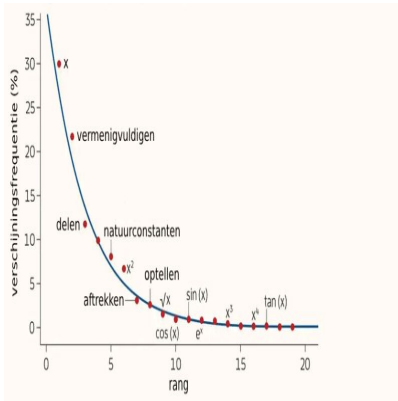




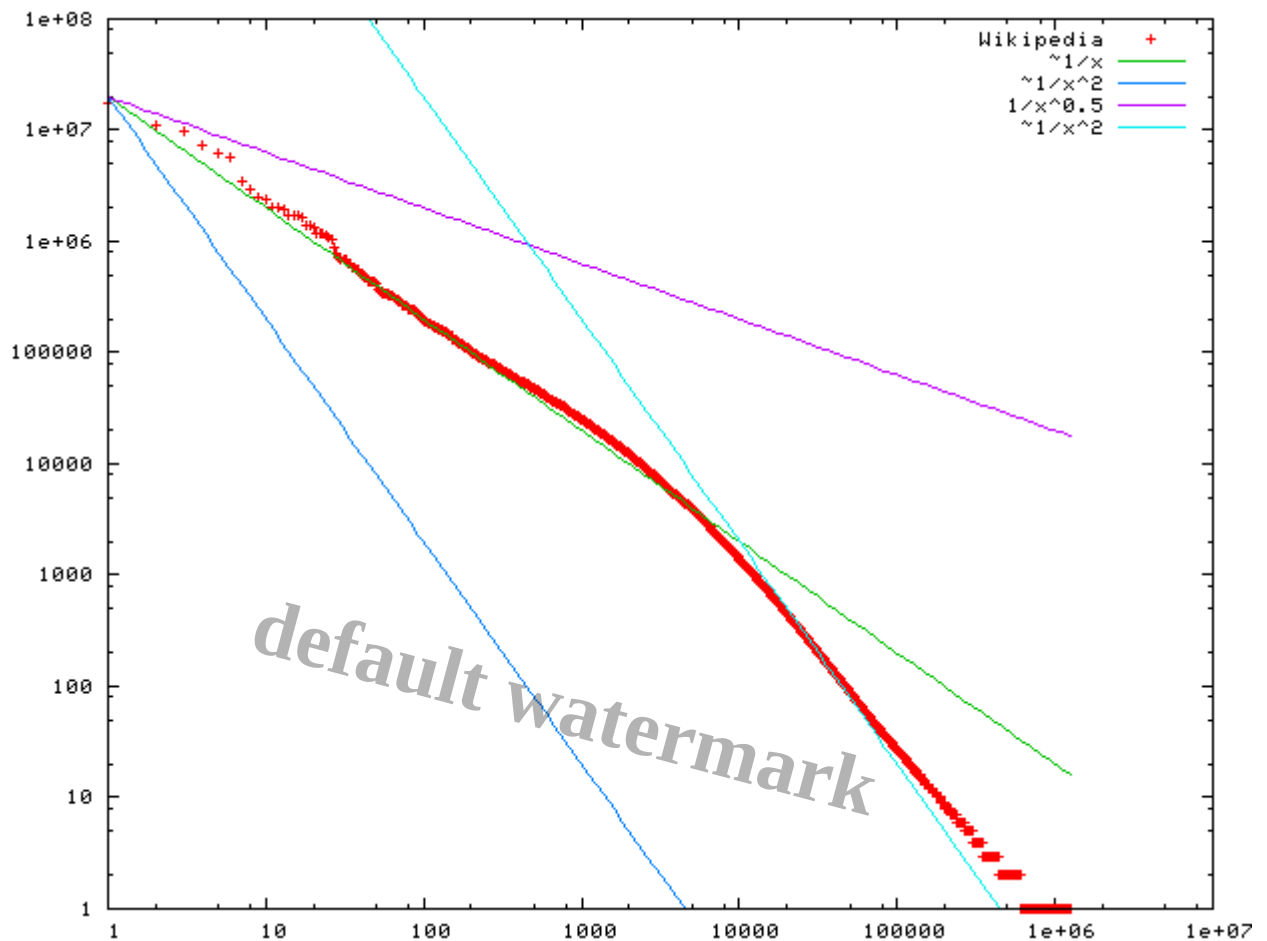
De wet van Zipf in de natuurkunde

Beschrijving

De **wet van Zipf** is een door *George Kingsley Zipf*¹ geconstateerde, en naar hem genoemde, wetmatigheid in de taalkunde. Hij stelde vast dat het meest gebruikte woord in een taal maar liefst twee keer zo vaak voorkomt als het tweede meest voorkomende woord, drie keer zo vaak als het derde, enzovoort. Met andere woorden het product rf is (ongeveer) gelijk aan 1 (r is rangvolgorde en f is frequentie).

Deze wetmatigheid bleek ook op te gaan voor andere talen en voor allerlei andere situaties. In het Engels, bijvoorbeeld, bestaat een lap tekst voor ongeveer 7 procent uit het woord *the*. Het volgende meest gebruikte woord, *of*, vormt ongeveer 3,5 procent van elke tekst.

Op 27 november 2006 werd een grafiek van de woordfrequentietabel van de Engelstalige Wikipedia, opgesteld (de rode lijn in onderstaande grafiek). Ook daar volgt de verdeling de wet van Zipf (de groene lijn) voor de acht- ã negenduizend meest voorkomende woorden.



Door Victor Grishchenko, <https://commons.wikimedia.org/w/index.php?curid=1415459>

De constatering door Zipf markeert het begin van de kwantitatieve linguïstiek. Buiten de linguïstiek duikt dit verband evenwel ook in uiteenlopende andere domeinen op. Zo blijkt het verband tussen de rang en de bevolking van de grootste steden in de Verenigde Staten en verschillende andere landen overeen te komen met de wet van Zipf. De wetenschappelijke productiviteit (aantal publicaties of referenties naar publicaties van wetenschappers) en het aantal maal per maand dat de webpagina's van een website worden bezocht zijn andere voorbeelden van data die de wet van Zipf volgen.

Natuurlijk patroon

In [NewScientist](#) lees ik dat de natuurkundige *Andrei Constantin* van de *Universiteit van Oxford* en zijn collega's hebben ontdekt dat een soortgelijke wet ook geldt voor de symbolen die we in natuurkundige vergelijkingen gebruiken.

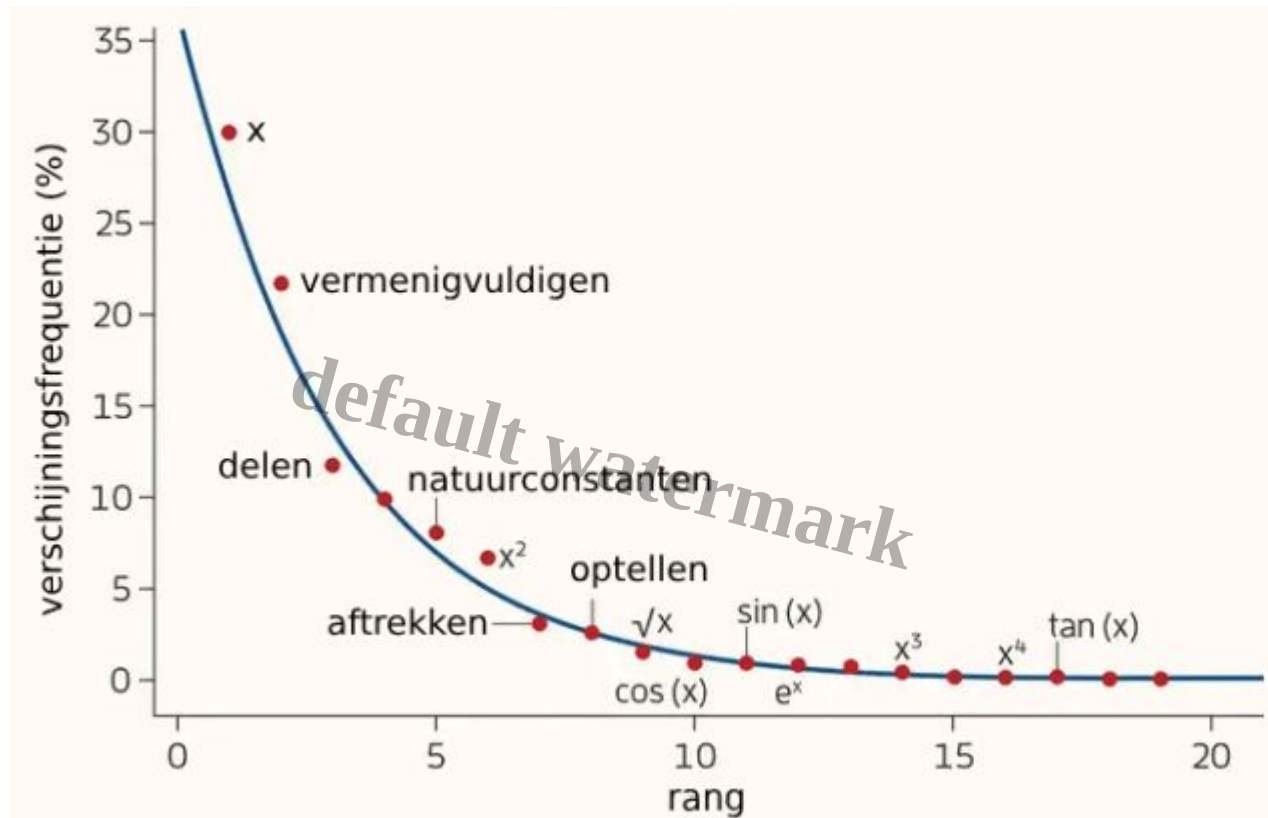
Ze analyseerden drie bronnen vol vergelijkingen:

- de bekende natuurkundeboeken *The Feynman Lectures on Physics*,
- een Wikipedia-overzicht van vergelijkingen die zijn vernoemd naar mensen, en
- een verzameling vergelijkingen die de inflatie van het vroege universum beschrijven.

De onderzoekers behandelden elk symbool en elke wiskundige bewerking in de vergelijkingen als een woord. Ze turfden welke het meest voorkwamen en welke het minst, en keken zo of ook in de

natuurwetten een wet van Zipf bestaat.

“Je zou verwachten dat de [verdeling] behoorlijk verschilt tussen de drie verschillende verzamelingen van vergelijkingen, omdat ze van verschillende bronnen komen”, zegt teamlid Deaglan Bartlett van de Sorbonne-universiteit in Frankrijk. Tot hun verbazing was dat echter niet het geval. In alle drie de sets verscheen hetzelfde patroon. Dit is wat ze vonden:



Als je de symbolen en wiskundige bewerkingen in natuurkundige vergelijkingen turft, blijkt hun verschijningsfrequentie een patroon te volgen. De “x”, die staat voor een onbekende in een vergelijking, staat met stip op ∞ .

¹ **George Kingsley Zipf** (1902-1950) was een Amerikaans linguïst en filoloog die woordfrequenties in verschillende talen onderzocht. Naar hem is de *wet van Zipf* en de Zipfdistributie vernoemd.

Datum aangemaakt
2024/10/24