

De hyperbolische tangens schaal Th: Van Verhulst naar McFadden's Nobelprijs

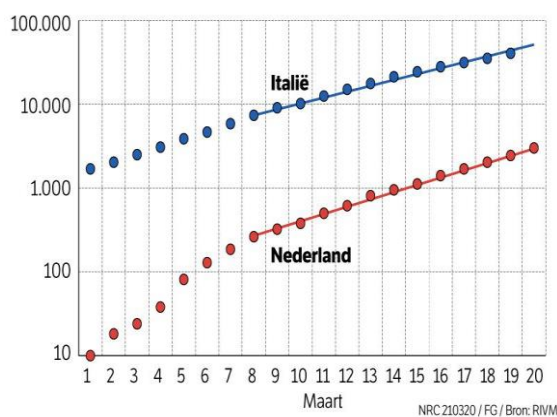
Simon van der Salm

Introductie

In de wetenschapbijlage van de NRC van zaterdag 21 maart 2020, bladzijde W5, onder de titel *Kromme corona en rechte corona*, geschreven door Karel Knip [1], staat een intrigerend plaatje (figuur 1) met twee *groeicurven*, bij verticaal een *logaritmische schaalverdeling*. Die figuur toont het *cumulatieve* aantal *bevestigde* coronabesmettingen in Nederland en Italië, gedurende 1 tot en met 20 maart 2020.



Bevestigde besmettingen sinds 1 maart (logaritmische schaal)

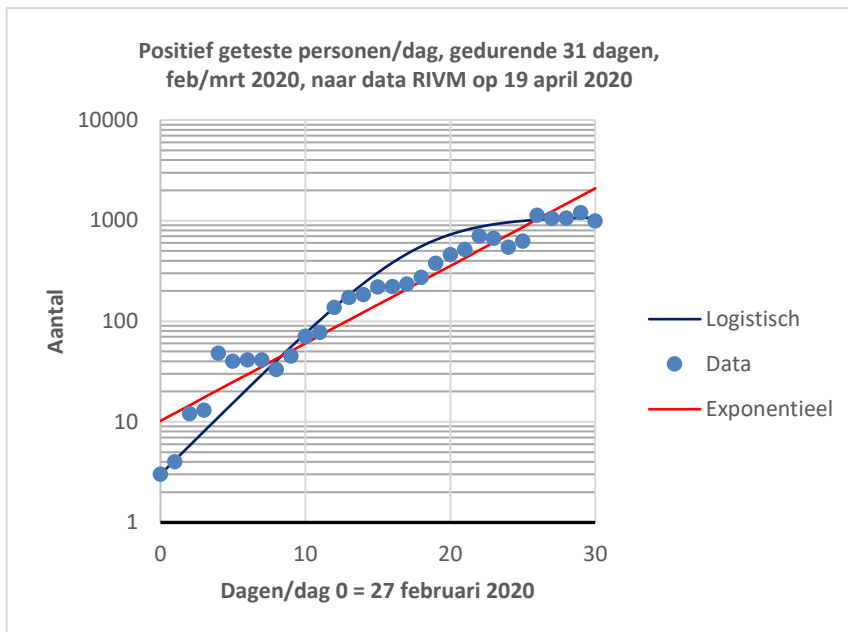


Met rechte corona wordt *exponentiële groei* bedoeld. De kromme corona is te beschrijven als *logistische groei*, afgeremde exponentiële groei, waarvan de curve een tangens hyperbolicus (Th of tanh) is, een functie die ook wel op rekenlijnen voorkomt. Aardig om eens te kijken hoe deze schaal gebruikt kan worden in tijden van corona.

Fig. 1. Grafiek uit de NRC van 21 maart 2020. Artikel van Karel Knip. De rechte lijnstukken van 8 t/m 20 maart duiden op exponentiële groei. Na het schrijven van dat artikel zijn de cijfers van het RIVM enigszins gecorrigeerd. Zie figuur 2.

Figuur 2 toont, iets pregnanter, logistische ontwikkeling van het aantal *bevestigde* besmettingen *per dag* in *Nederland* o.g.v. de gegevens van het RIVM [2], over 27 februari t/m 28 maart 2020, peildatum 19 april (lichtblauwe bolletjes). De exponentiële groei in de data (rode, rechte lijn) volgens lineaire regressie, wordt vergeleken met de logistische groei in de data (donkerblauwe curve), eveneens via lineaire regressie verkregen.

Fig. 2 op de volgende bladzijde. Ontwikkeling van het aantal bevestigde coronabesmettingen per dag in Nederland gedurende de maanden februari/maart 2020. Dag 0 = 27 februari; dag 30 = 28 maart. De data (lichtblauwe bolletjes) zijn afkomstig van de website van het RIVM op 19 april. Zie referentie [2]. De rode regressielijn is de exponentiële groei in de data; de donkerblauwe curve de logistische groei.



Elektronici zullen de Th-functie kennen i.v.m. transmissielijnen (lengte veel groter dan de golflengte van het signaal). Als je op een kabel een sinusvormig signaal aanbiedt, hebben de spanning- en stroomgolven daarin merkwaardigerwijs geen sinusvorm maar een hyperbolische vorm. Een soortgelijk verschijnsel zie je als je naar de branding kijkt: als een golf ketst op een strand ontstaat een hyperbolisch golfpatroon op de oppervlakte.

Hier kijken we naar een andere toepassing van een hyperbolische functie, n.l. naar de *logistische groeifunctie van Verhulst*. De grafieken in figuur 2 zijn d.m.v. lineaire regressie bepaald. Daarvoor werd tot ongeveer 1970 ook wel een rekenliniaal gebruikt, zij het zeer beperkt. Zie [4]. Inmiddels is het vakgebied van *logistische regressie*, gebaseerd op de functie van Verhulst, zo uitgebreid en de hoeveelheid rekenwerk zo enorm toegenomen, dat het ondoenlijk is daarvoor nog een rekenliniaal te willen gebruiken.

Rekenlinialen met een Th-schaal en logistische rekenschijven

Logistische groei wordt uitgedrukt in een Th. Er zijn betrekkelijk weinig rekenlinialen met een Th-schaal. Opvallend is dat Chinese rekenlinialen relatief vaak een Th-schaal hebben. In dit artikel bestuderen we de Th-schaal van de Flying Fish 1015 (FF-1015). (De FF-1003 heeft een identieke Th-schaal).

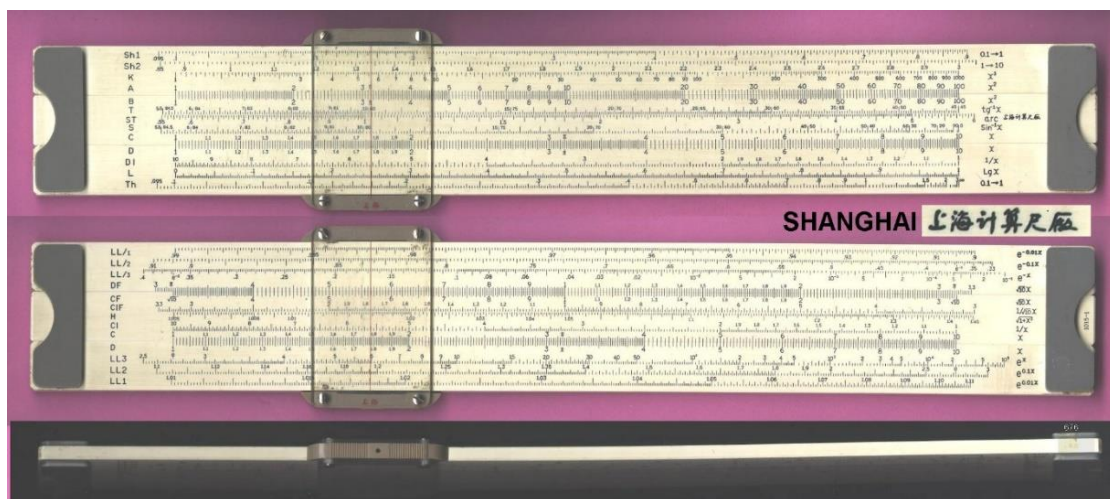


Fig. 3. De Flying Fish 1015: www.sliderulemuseum.com/orient.htm [6].

Dit artikel betreft de onderste schaal Th op de voorzijde van de liniaal in relatie tot schaal D.

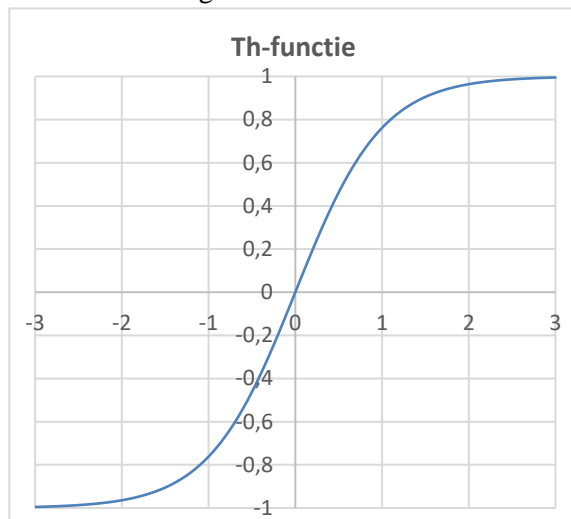
De website van Bill Robinson [3] vermeldt een groot aantal toepassingen van hyperbolische functies i.v.m. rekenlinialen. De *logistische functie van Verhulst* (1845) als model voor afgeremde exponentiële groei (zie figuur 2) en de op die functie gebaseerde statistische techniek van *logistische regressie* worden

niet vermeld. Bij logistische regressie wordt de logistische functie van Verhulst toegepast als een statistisch *beslismodel*. Dat is zeer actueel omdat tegenwoordig het vele rekenwerk daarvan gemakkelijk door een digitale computer kan worden uitgevoerd.

Logistische regressie wordt al heel lang toegepast, in met name het actuariaat en de verzekeringswiskunde. Daarbij kwam ook de rekenliniaal wel eens van pas, hoewel de mogelijkheden daarvan beperkt bleken [4], p.11. Door de toegenomen rekenkracht van computers wordt het toepassingsgebied van logistische regressie nog steeds verder uitgebreid. In het jaar 2000 werd zelfs de Nobelprijs voor economie toegekend aan Daniel L. McFadden, voor toepassingen van logistische regressie in econometrische beslismodellen. Zie [4]. Dit artikel schetst de geschiedenis en de wiskunde van de logistische functie in relatie tot de Th-schaal op rekenlinialen. Die geschiedenis begint bij Verhulst in 1845 en eindigt (voorlopig) bij McFadden in 2000. Een uitgebreidere geschiedenis is te vinden in een artikel van J.S. Cramer, van het Tinbergeninstituut (UvA), een bron die *Andries de Man* vond [4]. Maar Andries heeft nog een ontdekking gedaan: er zijn ook rekenschijven ontwikkeld voor het uitvoeren van logistische regressie. Dit artikel toont in figuur 13 als voorbeeld de *Dundee-rekenschijf*; voor een beschrijving daarvan is een volgend artikel nodig. Zie [5]. Hier zullen we ons beperken tot de logistische functie en de Th-schaal.

De Flying Fish 1015 rekenliniaal (FF-1015)

Th-functiewaarden kunnen we berekenen m.b.v. sommige rekenlinialen, bijvoorbeeld met de onderste schaal op de voorkant van de Flying Fish 1015 (FF-1015). Zie figuur 3. Een eigenaardigheid daarvan is dat steeds *extra* gedeeld moet worden door 10.



Naast de getallen x op D is het daarom handig om daarbij (denkbeeldig) getallen $y = \frac{x}{10}$ langs D voor

ogen te hebben. Getallen op de Th-schaal zullen we aanduiden met t . De *hyperbolische tangens* is

$$y = \text{Th}(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \quad (1). \text{ Zie de sigmoïde (S-}$$

vorm) in figuur 4. I.t.t. tot de goniometrische tangens heeft deze tangens *horizontale asymptoten*, namelijk $y = -1$ en $y = +1$.

Fig. 4. Grafiek van $y = \text{Th}(t)$, met horizontale asymptoten $y = -1$ en $y = +1$.

De Th-schaal van de FF-1015, met t -waarden, is gekoppeld aan D, met x -waarden die $10x$ te groot zijn.

Als x op D, dan geeft de Th-schaal $t = \text{Th}^{-1}\left(\frac{x}{10}\right)$, een waarde van de inverse functie! We zullen zien

dat dit feit ook nog tot een extra, foutengevoelige handeling leidt. Feitelijk is Th een verkeerd opschrift.

Dat er *extra* gedeeld moet worden door 10 is rechts van de Th-schaal aangeduid met de cryptische aanduiding $0.1 \rightarrow 1$. Dus voor t op de Th-schaal vinden we $x = 10 \cdot \text{Th}(t) = 10y$ op D.

Figuur 5 toont het verband tussen de schalen D en Th. D is uiteraard logaritmisch. Omdat voor kleine waarden van t geldt dat $\text{Th}(t) \approx t$, wordt het recht evenredige verband tussen t en x alleen duidelijk op een log-log-schaalverdeling, vandaar dat in figuur 5 beide assen een logaritmische schaalverdeling hebben.

Aan de rechterkant van de Th-schaal van de FF-1015 staat het symbool ∞ . Figuur 5 maakt duidelijk waarom.

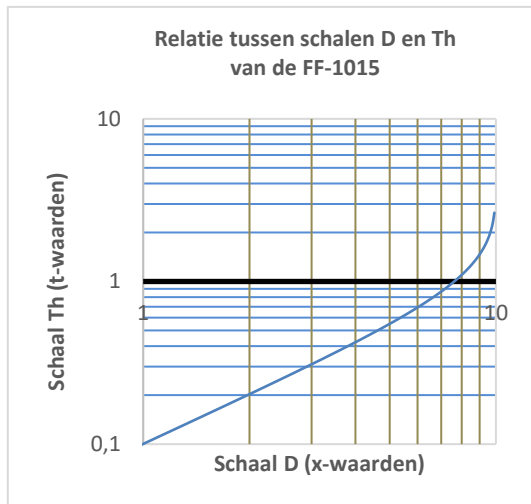


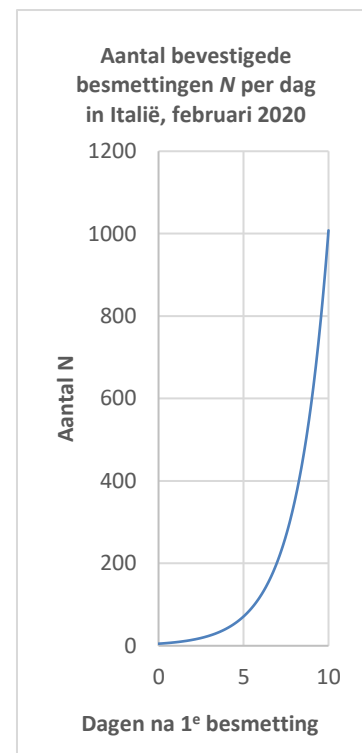
Fig. 5. Relatie tussen de D-schaal en Th-schaal van de FF-1015 op een log-log-schaalverdeling. De getallen x op D lopen van 1 tot 10; y -waarden dus van 0,1 tot 1; t -waarden op Th lopen van ongeveer 0,1 tot ongeveer 3. Bijbehorende Th-waarden zijn $y = x/10$.

De Th-functie is, zoals figuur 4 toont, symmetrisch ten opzichte van het buigpunt O; we kunnen ons dus beperken tot positieve waarden van t en x . Praktische t -waarden liggen tussen 0 en +3, maar 0 komt uiteraard op D niet voor. Figuur 5 toont dat de kleinste waarde op Th 0,1 is en niet 0. Met de Th-schaal komen we daarom niet bij het buigpunt O van de Th-functie, terwijl dat buigpunt vaak wel belangrijk is. Voor (positieve) t -waarden $< 0,1$ geldt echter $Th(t) \approx t$, dus heel problematisch is dat niet.

Exponentiële groei

Naast alle nare berichtgeving over het coronavirus is het interessant om te zien hoe de media in de verslaggeving van COVID-19 met wiskundige en statistische begrippen worstelen. Zo gonsde ineens, voornamelijk tijdens de eerste dagen (begin maart 2020) van de crisis, het begrip *exponentiële groei*, te pas, maar vooral te onpas, rond. Lezers en tv-kijkers werd de stuipen op het lijf gejaagd met snel stijgende grafieken, zoals in figuur 6. O. g. v. slechts enkele datapunten (na een dag of vijf na de eerste besmetting) werden de in Italië waargenomen data driftig exponentieel geëxtrapoleerd en werd er oeverloos gepraat over de betekenis daarvan voor de mogelijke ondergang van de wereld.

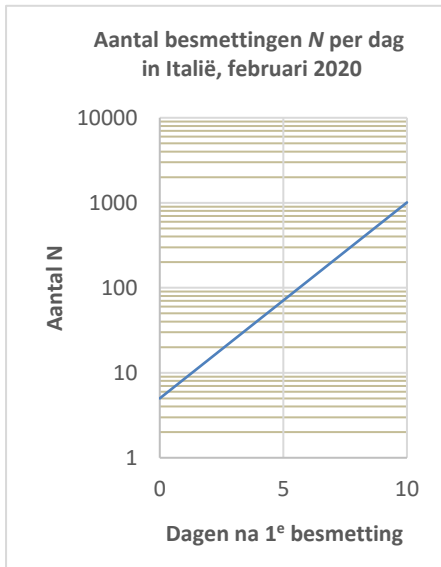
Fig. 6. Voorbeeld van een exponentiële extrapolatie tot en met dag 10 o.g.v. de waarnemingen gedurende de eerste 5 dagen. De (niet zichtbare) beginwaarde is 5; de groeifactor is 1,7 per dag (Italië, laatste week van februari 2020, uit diverse bronnen afgeleid). Verdubbelings-tijd ongeveer 1,3 dagen.



Exponentieel extrapoleren op grond van slechts enkele datapunten is natuurlijk niet erg verstandig, vooral ook omdat geen enkele groei in de fysische werkelijkheid exponentieel kan zijn, maar wordt afgeremd door externe factoren, zoals gebrek aan ruimte of beperkte voedingsbronnen.

Wat opvalt aan een grafiek zoals in figuur 6 is dat voor een leek met geen mogelijkheid op te maken valt of er wel of niet sprake is van exponentiële groei. Is de grafiek een parabool, of een derdegraads functie, of een functie van nog hogere graad? Of misschien nog iets anders? Iedere willekeurige, stijgende, gebogen grafiek kan er exponentieel groeiend uitzien. Of er daadwerkelijk sprake is van exponentiële groei is het best te zien als we de verticale as een *logaritmische* schaalverdeling geven. Zie figuur 7. Een stijgende rechte lijn (gemakkelijk te herkennen) duidt op exponentiële groei. Helaas, een log-schaalverdeling gaat het begripsvermogen van de meeste media te boven.

Fig. 7, volgende bladzijde. Dezelfde exponentiële functie als in figuur 6. Beginwaarde is 5 (nu wel herkenbaar); de groeifactor is 1,7. Een stijgende rechte lijn bij verticaal een logaritmische schaalverdeling duidt op exponentiële groei.



Logistische groei

In de fysische werkelijkheid kan geen enkele groei gedurende langere tijd exponentieel zijn. Groei van een populatie $N = N(t)$ volgt vaak een Th-groecurve met een verzadigingsniveau Ω , zoals in figuur 8. In die figuur is sprake van *logistische groei*, voor het eerst in 1845 beschreven door de Belgische wiskundige *Pierre-François Verhulst* (1804 -1849). In figuur 8 is $\Omega = 1000$ de horizontale asymptoot van een geschaalde en verschoven Th-grafiek. Het verzadigingsniveau 1000 is de bovengrens die de variabele $N = N(t)$ benadert als de tijd langs de horizontale as alsmear toeneemt.

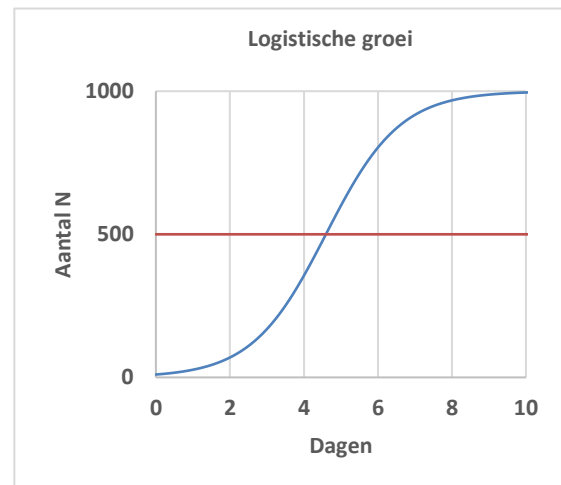


Fig. 8, rechts. Voorbeeld van logistische groei met horizontale asymptoot $\Omega = 1000$ (verzadigingsniveau) en buigpunt bij $1/2 \Omega = 500$, het snijpunt van oranje lijn en de grafiek. De beginwaarde is $N_0 = 10$. Verticaal een lineaire schaalverdeling.

Het buigpunt van de logistische curve ligt bij de verticale waarde $1/2 \Omega = 500$, dus bij precies de helft van het verzadigingsniveau. Dat betekent dat er van buigpunt naar verzadigingsniveau van slechts *één verdubbeling* sprake is! Een logistische curve groeit aanvankelijk exponentieel, maar die groei wordt, als de tijd voortschrijdt, steeds meer afgeremd. De exponentiële groei duurt ongeveer tot aan het buigpunt. Dat is in figuur 8 niet zo goed te zien, maar wel in figuur 9, waar de verticale as een logaritmische schaalverdeling heeft. Zie tevens ook nog een keer figuur 2.

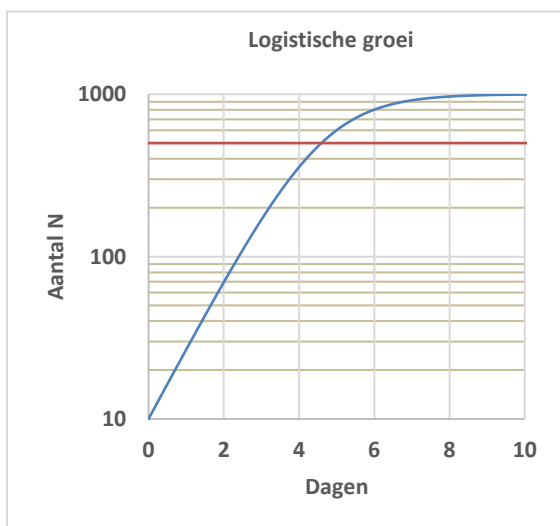


Fig. 9. Logistische groei met verzadigingsniveau $\Omega = 1000$ en buigpunt bij $1/2 \Omega = 500$. Hier verticaal een logaritmische schaalverdeling. Duidelijk te herkennen is exponentiële groei van beginpunt $N_0 = 10$ tot aan circa het buigpunt. N verdubbelt van buigpunt tot verzadigingsniveau.

De grafiek stijgt aanvankelijk (circa) lineair, tot ongeveer het buigpunt, wat duidt op exponentiële groei; rechts van het buigpunt wordt de groei steeds meer afgeremd. Van buigpunt naar verzadigingsniveau is de groeifactor 2. In figuur 8 is de beginwaarde $N_0 = 10$ nauwelijks te herkennen; in figuur 9 wel. Hier zien we een voordeel van een logaritmische schaalverdeling:

zowel kleine als grote waarden van N zijn zichtbaar.

De populatievergelijking van Verhulst (1845)

Pierre-François Verhulst (1804 -1849), leerling van *Adolphe Quetelet* (van de BMI = Body Mass Index), is het meest bekend als wiskundig demograaf. Hij realiseerde zich dat door allerlei begrenzende

factoren de grootte N van een populatie niet exponentieel kan blijven stijgen, ook als die stijging aanvankelijk wel exponentieel is. Zie [11]. Wat we tegenwoordig exponentiële groei noemen werd in de tijd van Verhulst *logaritmische groei* of *geometrische groei* genoemd. Zie figuur 10. Verhulst noemde, mogelijk (er zijn diverse verklaringen, zie [4]) zijn groeifunctie met verzadigingsniveau Ω *log-geometrisch* en maakte daarvan *logistisch*. Dat toont ook het karakter van zijn groeifunctie: eerst geometrisch en daarna afbuigend, zo ongeveer als een logaritme. Een aardige bijkomstigheid is dat het Griekse *logistikós* ook nog *bedreven in het rekenen* betekent.

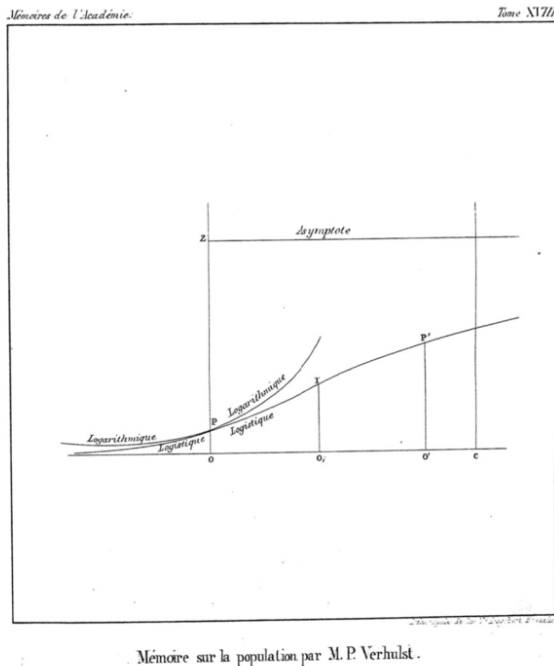


Fig. 10. Tekening uit *Recherches mathématiques sur la loi d'accroissement de la population* (1845) van Verhulst. Te zien is dat hij exponentiële groei logaritmische groei noemt. Verder toont de tekening dat hij voor geremde groei het woord *logistique* gebruikte [7].

Bij *exponentiële* groei is de gradiënt van de populatiegrootte N recht evenredig met de grootte van die populatie, dus

$$\frac{dN(t)}{dt} = \rho N(t) \rightarrow N(t) = N_0 e^{\rho t} \quad (2).$$

De parameter $\rho > 0$, de *groeiconstante*, bepaalt de snelheid van de groei per tijdseenheid en daarmee de verdubbelingstijd. De *groeifactor* is $g = e^\rho$; de

verdubbelingstijd is $t_{dub} = \frac{\ln 2}{\rho}$. De parameters N_0

en ρ van de exponentiële groei zijn eenvoudig te schatten d.m.v. van *lineaire regressie*:

$$\ln N(t) = \ln N_0 + \rho t = \alpha + \beta t \quad (3), \text{ met}$$

$\hat{N}_0 = e^\alpha$; $\hat{\rho} = \beta$ (4). Aldus is de rode regressielijn in figuur 2 bepaald, met intercept $a = 2,32$ als schatting van α en gradiënt $b = 0,18$ als schatting van β .



Fig. 11. Portret van Verhulst. Bron: Wikipedia.

De vraag die Verhulst zich stelde is: “Welke remmende factor moet ik aan vergelijking (2) toevoegen voor een realistischer modelbeschrijving van de groei van een populatie?”. Om die vraag te kunnen beantwoorden verrichtte hij monnikenwerk. Na bestudering van grote hoeveelheden bevolkingsdata concludeerde hij dat (2) moet

worden aangevuld tot $\frac{dN(t)}{dt} = \rho N(t) \left(1 - \frac{N(t)}{\Omega}\right)$ (5). Dat was

een gelukkige ontdekking, want (5) heeft een bekende oplossing:

$N(t) = \Omega \left(\frac{1}{2} + \frac{1}{2} \text{Th} \frac{\rho(t - t_B)}{2} \right)$ (6). Hier is t_B de t -waarde van het

buigpunt: $t_B = \frac{1}{\rho} \ln \left(\frac{\Omega}{N_0} - 1 \right)$ (7). (6) is, geschaald en horizontaal

verschoven de *standaard logistische functie van Verhulst*

$L(t) = \frac{1}{2} + \frac{1}{2} \operatorname{Th} \frac{t}{2} = \frac{1}{1 - e^{-t}}$ (8). Formule (6) wordt daarom ook wel geformuleerd als

$N(t) = \frac{\Omega}{1 + \exp(-\rho(t - t_b))}$ (9), een formule die als nadeel heeft dat de puntsymmetrie van de logistische curve ten opzichte van het buigpunt minder goed herkenbaar is. Als we logistische waarden willen uitrekenen met een rekenliniaal, doen we dat met (6).

De vergelijkingen (5) t/m (9) van Verhulst werden vergeten en in 1911 opnieuw opgesteld door *Anderson Gray McKendrick* [9], die overigens zelf niet het woord *logistisch* gebruikte, omdat hij het werk van Verhulst niet kende. Als gevolg van het onderzoek van Pearl en Reed in 1920, werd het fundamentele werk van Verhulst herontdekt en vanaf toen op grote waarde geschat. Zie [4].

Terzijde: De Schotse epidemioloog *A.G. McKendrick* stelde, samen met *W.O. Kermack*, het SIR-infectiemodel (Susceptible, Infectious, Recovered) op voor de verspreiding van infectieziekten. Dat model is de moeder van alle wiskundige infectiemodellen. De termen r_0 -getal, groepsimmunitet en afvlakken van de curve, die momenteel i.v.m. COVID-19 zo in het nieuws zijn, zijn gerelateerd aan dat Kermack-McKendrick-SIR-model. Zie [10] voor de biografie van McKendrick.

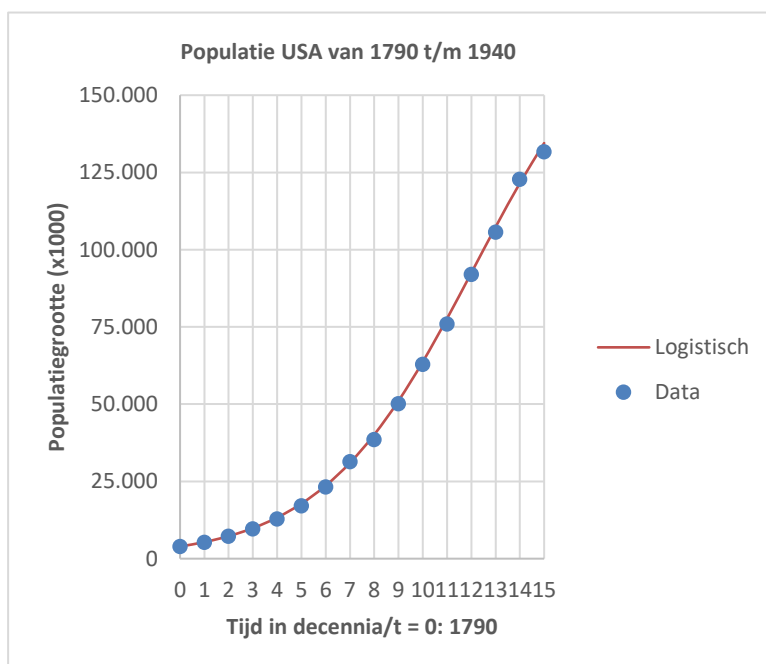
De bevolking van de USA (Verhulst 1845)

Zoals bij alle wiskundige modellen van verschijnselen in de fysische werkelijkheid moeten de waarden van de *parameters* van het model worden geschat op grond van empirische gegevens. Verhulst vond een verbluffend goede schatting voor ρ en Ω voor de groei van de Amerikaanse bevolking tijdens de negentiende en het begin van de twintigste eeuw. Op grond van slechts 5 empirische waarden (volkstellingen van 1790, 1800, 1810, 1820 en 1830) wist hij die twee parameters zodanig nauwkeurig te schatten dat daarmee een logistische extrapolatie tot en met 1940 gerechtvaardigd bleek.

De waarden van de parameters ρ (groeiconstante) en Ω (verzadigingsniveau) kunnen we schatten door

te bedenken dat uit (5) het lineaire verband $\frac{dN(t)}{N(t)} = \rho \left(1 - \frac{N(t)}{\Omega} \right) = \alpha + \beta N(t)$ (10) volgt. Uit dat

verband volgen schattingen van α en β : $\alpha = 0,3179$ en $\beta = -1,691 \cdot 10^{-6}$. Daaruit volgen weer de schattingen van de parameters: $\hat{\rho} = 0,3179$ en $\hat{\Omega} = 1,880 \cdot 10^5$. Substitueren we die waarden in (6) en tekenen we



de bijbehorende grafiek, uitgaande van de bekende N_0 in 1790, dan zien we hoe verbluffend goed de logistische extrapolatie van Verhulst (1845!) was. Die nauwkeurigheid maakte Verhulst, lang na zijn dood, alsnog beroemd.

Fig. 12. De werkelijke ontwikkeling van de bevolking van de USA (blauwe bolletjes) van 1790 t/m 1940 per decennium en de logistische benadering volgens Verhulst (rode grafiek). 1 tijdseenheid = 10 jaar. Buigpunt bij circa 12 tijdseenheden (120 jaar). De grafiek begint als exponentiële groei met groeifactor 1,37.

Op overeenkomstige wijze is de logistische (donkerblauwe) curve in figuur 2 van de RIVM-gegevens bepaald. Voor die curve vinden we de schattingen $\hat{\rho} = 0,330$ en $\hat{\Omega} = 1089$, bij $N_0 = 3$.

Logistische berekening met de FF-1015

Vervolgens rijst de vraag of waarden van $N(t) = \Omega \left(\frac{1}{2} + \frac{1}{2} \text{Th} \frac{\rho(t-t_B)}{2} \right)$ (6), met bijvoorbeeld de schattingen $\hat{\rho} = 0,3179$ en $\hat{\Omega} = 1,880 \cdot 10^5$ (USA, Verhulst), een beetje handig zijn uit te rekenen met de FF-1015 rekenliniaal. Voor de beantwoording van die vraag onderzoeken we twee voorbeelden.

Voorbeeld 1:

Neem $t = 15$, rechts van het buigpunt. De afstand tot het buigpunt is 3. We berekenen $\frac{0,3179 \cdot 3}{2} = 0,477$

(4,77 op D). Het vervelende is dat dit getal op D staat en de Th-functie juist andersom, van Th-schaal naar D, werkt. Dit laat zien dat de Th-schaal eigenlijk een inverse Th-schaal is, dus precies verkeerd om vergeleken bij wat je nodig hebt bij deze berekening. Er zit niets anders op dan het getal 0,477 ‘met het oog’ over te brengen naar de Th-schaal. Delen door 10 gaat daarbij automatisch. We lezen de waarde van de Th-functie af op D: 0,444 (4,44). Delen door 2 is eenvoudig 0,222 op D (2,22). De optelling met 0,5 moet met de hand plaatsvinden: 0,722 (7,22). Dat moeten we naar D brengen en vermenigvuldigen met 1,880 op C. We vinden 1,36 op D en dus $N = 136.000$, acceptabel dicht bij de werkelijke waarde.

Voorbeeld 2:

Neem $t = 11$, links van het buigpunt. De afstand tot het buigpunt is 1. We berekenen $\frac{0,3179 \cdot 1}{2} = 0,1590$

(1,59) en brengen dat van D over naar Th. Op D lezen we de Th-waarde af: 0,158 (1,58). (Merk op, nagenoeg gelijk aan 0,1590). Delen door 2 is gemakkelijk: 0,0788 (7,88 op D). Maar omdat het een punt links van het buigpunt betreft, moeten we dat als -0,0788 denken en daarbij op papier 0,5 optellen. We stellen de som 0,421 (4,21) in op D. Vermenigvuldigen met 1,88 geeft 0,792 (7,92). We vinden een populatiegrootte N van ruim 79.000, iets groter dan de 76.000 in figuur 2.

We zien, dat tamelijk veel stappen moeten worden genomen voor het berekenen van één functiewaarde. Voor de 16 punten in figuur 12 zou de bovenstaande groep handelingen 16x moeten worden uitgevoerd. Het berekenen van een logistische curve m.b.v. de Th-schaal op de FF-1015 is dus weliswaar mogelijk, maar flink bewerkelijk. Met een hedendaagse grafische rekenmachine, Excel-rekenblad of eenvoudig computerprogramma is het werk gemakkelijker, sneller en aanzienlijk nauwkeuriger uit te voeren.

Nobelprijs voor logistische regressie

Logistische regressie is een statistische techniek die gebruik maakt van de logistische functie van Verhulst. Bovenstaande voorbeelden zijn relatief eenvoudig, omdat er van een betrekkelijk kleine dataset en slechts één onafhankelijke variabele (t) en maar twee parameters (ρ en Ω) sprake is. Bij wat ingewikkelder logistische modellen neemt het rekenwerk al snel duizelingwekkend toe. Met behulp van computers is dat geen probleem, maar dat was in de jaren vijftig en zestig, toen men de logistische functie steeds meer ging toepassen voor het oplossen van praktische, statistische beslisproblemen (heeft u recht op een levensverzekering?, slaagt u voor een examen?, lijdt u aan een bepaalde ziekte?, enz.) een enorme klus. Cramer [4] schrijft in zijn artikel over de geschiedenis van logistische regressie het volgende:

Until the advent of the computer and the pocket calculator, some twenty years later, all numerical work was done by hand, that is with pencil and paper, sometimes aided by

*graphical inspection of 'freehand curves', fitted by the eye...Numerical work was supported rather feebly by the **slide rule** and by the mechanical calculating machines, driven by hand or powered by a small electric motor, which were capable of addition and multiplication; punched card equipment was helpful if numerous data had to be analysed.*

Vanaf de jaren zeventig, toen digitale computers het tijdrovende en foutengevoelige rekenwerk begonnen over te nemen, is de logistische regressie met behulp van de *logit-functie* (logische unit, inverse van de standaard logistische functie van Verhulst) uitgegroeid tot basisgereedschap van de statisticus. De ontwikkeling van dat gereedschap resulteerde in het jaar 2000 in een Nobelprijs voor Economie voor de econometrist Daniel L. McFadden, een ontwikkeling die begon bij de logistische functie van Verhulst ruim 150 jaar eerder. Zie [4].

Rekenslijven voor logistische regressie

Andries de Man ontdekte dat er ook rekenslijven zijn ontwikkeld voor logistische regressie. Een voorbeeld is de *Dundee coronary risk-disk*, die met vier onafhankelijke variabelen en 5 parameters werkt. Zie figuur 13. Het is niet te zien, maar deze rekenslijf hanteert de logistische functie van Verhulst (dus



een Th) als beslisfunctie. Een onderwerp voor een volgend artikel.

Fig. 13. De Dundee coronary risk-disk for management of change in risk factors. Bron: <http://journals.rcni.com/nursing-standard/using-the-dundee-coronary-riskdisk-ns.7.30.50.s63>.

Dank

Mijn dank gaat in de eerste plaats uit naar Andries de Man, die voor mij bronnen vond met de geschiedenis van

de logistische curve en –regressie, bronnen waarvan ik dankbaar gebruik heb gemaakt. En bronnen (o.a. over de Dundee schijf) die misschien tot nader onderzoek leiden.

Verder wil ik Chris Hakkaart en Leo van der Lucht bedanken voor hun waardevolle opmerkingen over versie 0 van dit artikel.

Referenties

- [1] <https://www.nrc.nl/nieuws/2020/03/20/exponentiele-groei-inzichtelijk-gemaakt-a3994425>
- [2] <https://www.rivm.nl/coronavirus-covid-19/actueel>
- [3] <https://osgalleries.org/hyperbolic/>
- [4] Cramer, J.S., *The Origins of Logistic Regression*, Tinbergen Institute, Universiteit van Amsterdam, 2002
- [5] Tunstall-Pedoe, H., *The Dundee coronary risk-disk for management of change in risk factors*, British Medical Journal, September 1991
- [6] www.sliderulemuseum.com/orient
- [7] https://en.wikipedia.org/wiki/Logistic_function
- [8] Idem
- [9] Thieme, H.R., *Mathematics in Population Biology*, Princeton University Press, 2003
- [10] <http://mathshistory.st-andrews.ac.uk/Biographies/McKendrick.html>
- [11] <http://mathshistory.st-andrews.ac.uk/Biographies/Verhulst.html>
- [12] Veaux, R. de, Velleman, P.F., Bock, D.E., *Stats: Data and Models*, pp. 574 – 577, Pearson Education, 2005.
- [13] Miles, J., Shevlin, M., *Applying Regression and Correlation*, Sage Publications, Ltd, 1e druk, 2001.