Groei, grafieken en gegevensverwerking

Sir Bakx
T.H. Twente, Enschede

Samenvatting

Een van de onderdelen van Wiskunde A is het opstellen, voor een aantal 'meetgegevens', van een functievoorschrift dat als benadering kan dienen van het verband tussen die gegevens. In het artikel wordt een werkwijze behandeld die het mogelijk maakt een functievoorschrift te vinden m.b.v. grafiekenpapier en computer, waarbij de methode van de kleinste kwadraten wordt gehanteerd.

Summary

In the new curriculum for non-mathematical majors at pre-university level considerable attention is paid to find functions that fit to a given table. In this article the author presents a way to find a fitting function using graph-paper and the computer and the method of the least-squares.

Inleiding

In het door het HEWET-team van OW & OC ontwikkeld leerlingenboek 'Groei', maakt de leerling in hoofdstuk 5 kennis met het verschijnsel afgeremde groei. Nadat in voorgaande hoofdstukken groeiprocessen behandeld zijn die min of meer exponentieel zijn, wordt geconstateerd dat in het algemeen in de praktijk op een bepaald moment de groei zal afnemen. (Bij ratten vanwege voedselproblemen, bij planten vanwege een beperkt beschikbare oppervlakte, enz.) De grafische weerslag van een dergelijk proces heeft de S-vorm (fig. 1). Het soort groeiprocessen wordt genoemd: 'logistische' groei en wordt beschreven door de functie:

$$S: x \to \frac{k}{1 + c \cdot e^{-r \cdot x}}$$

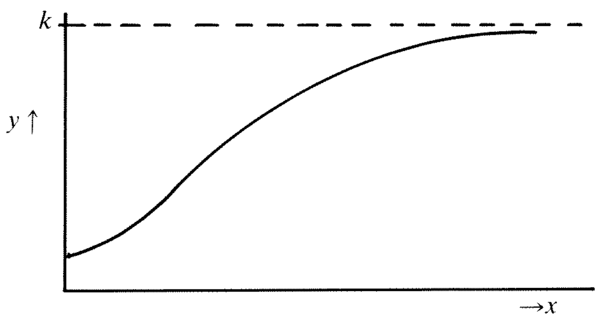


fig. 1

Een van de vaardigheden die de wiskunde A-leerling verwerft is het produceren van een (eenvoudig) functievoorschrift, waarmee een verband tussen twee grootheden zo goed mogelijk wordt benaderd. Dit verband moet worden afgeleid uit een gegeven tabel van 'meetresultaten'. Het functievoorschrift kan dan worden gehanteerd voor inter- en extrapolatie (het doen van voorspellingen, zo u wilt).

Door het OW & OC wordt ook een pakket aan standaardcomputerprogramma's uitgegeven ten behoeve van het wiskunde A-onderwijs. Eén ervan is het programma 'Logistische groei'. Dit programma produceert in eerste instantie de grafiek van een in te voeren of ingevoerde tabel. Vervolgens is het mogelijk om bijvoorbeeld door het opgeven van een aantal combinaties van parameters (zie functievoorschrift boven) een benaderingsvoorschrift te zoeken. Dit kan doordat bij elke opgegeven combinatie het programma aan de op het scherm getekende grafiek punten toevoegt, zoals die bij dat functievoorschrift horen. Wanneer een op het oog goede overlap is gerealiseerd van punten met die uit de originele tabel, heeft men een gewenst functievoorschrift gevonden.

In dit artikel nu wordt op een andere manier tegen het probleem aangekeken en daarmee een verband gelegd tussen groeiprocessen met hun functievoorschriften, grafieken c.q. het gebruik van grafiekenpapier, en automatische gegevensverwerking.
Het artikel is niet zozeer bedoeld als een voorstel voor leerstof voor leerlingen, maar meer als directe achtergrond van de leerstof voor de leraar.

Van gegevens naar functievoorschrift

Eerst een eenvoudig geval!

Een opdracht in 'Groei' luidt:

De grootte (lengte) van een blad wordt twee weken lang om 12 uur v.m. gemeten:

dag (in april)	*2*	*5*	*10*	*14*	*15*
lengte (mm)	*3.1*	*6.4*	*12.0*	*16.1*	*17.0*

a. Maak een grafiek (horizontaal tijd in dagen) en laat zien dat dit groeiproces redelijk te benaderen is door een lineaire functie.
b. Bepaal het functievoorschrift.

De grafiek van opdracht a. ziet er als volgt uit:

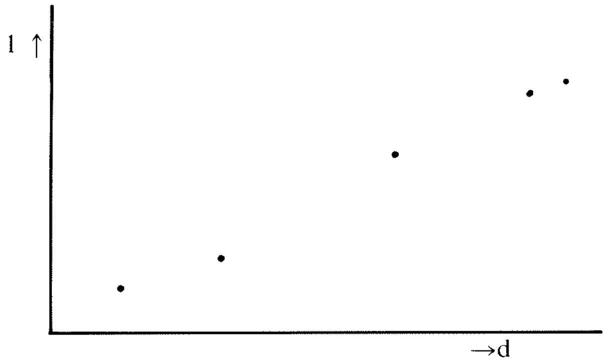


fig. 2

'Op het oog' liggen deze punten 'redelijk' op een rechte lijn.
Het tweede gedeelte van de vraag kan beantwoord worden door de gemiddelde toename van opvolgende perioden te berekenen: 1.1, 1.12, 1.025, 0.9.
Deze, ongeveer aan elkaar gelijke, getallen rechtvaardigen de conclusie dat men in het gemeten tijdsinterval te maken heeft met een te noemen lineaire groei.

Wat is nu het functievoorschrift?

Wat te denken van $l=d+3$ of $l=1.03d+3.1$ of wat leerlingen allemaal nog meer in redelijkheid (?) kunnen bedenken? Laten we het probleem wiskundig aanpakken.

Lineaire benadering (lineaire regressie)

We beschikken over p punten met coördinaten (x_i, y_i) $(1 \leqslant i \leqslant p)$. We zoeken de lineaire relatie $y=mx+n$ die deze punten kan benaderen. Dat wil zeggen, dat een verschil tussen iets van de ter beschikking staande punten en iets van de lineaire functie, zo klein mogelijk moet zijn. Een keuze, gemaakt in de methode van de kleinste kwadraten, is het minimaliseren van de som van de gekwadrateerde verschillen in y-coördinaat van een meetpunt en het punt van de functie met dezelfde x-coördinaat. (1).
Voor één punt rekenen we dus uit:

$$(mx_i + n - y_i)^2 = \\ m^2x_i^2 + n^2 + y_i^2 + 2mnx_i - 2mx_iy_i - 2ny_i$$

Gesommeerd over alle p punten levert dat:

$$\Sigma(mx_i + n - y_i)^2 = \\ m^2\Sigma x_i^2 + pn^2 + \Sigma y_i^2 + 2mn\Sigma x_i - 2m\Sigma x_iy_i - 2n\Sigma y_i$$

Dit moet minimaal zijn, dus:

$$\frac{d}{dm}(\Sigma(\bullet\bullet)^2) = 2m\Sigma x_i^2 + 2n\Sigma x_i - 2\Sigma x_iy_i = 0\text{, en}$$

$$\frac{d}{dn}(\Sigma(\bullet\bullet)^2) = 2pn + 2m\Sigma x_i - 2\Sigma y_i = 0.$$

Twee vergelijkingen met twee onbekenden, waarmee m en n uit te rekenen zijn. (2).
Het resultaat is:

$$m = \frac{p\Sigma x_iy_i - \Sigma x_i\Sigma y_i}{p\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$n = \frac{\Sigma x_i^2\Sigma y_i - \Sigma x_i\Sigma x_iy_i}{p\Sigma x_i^2 - (\Sigma x_i)^2}$$

Het intelligentste werk is gedaan, nu het domme rekenwerk. Maar daarvoor hebben we een computer. We ontwerpen een programma dat vraagt om de waarden van p en de respectievelijke x- en y-coördinaten van de meetpunten. Het moet ons vervolgens de waarden van m en n leveren en daarmee de gewenste lineaire functie.

Het structuurschema ziet er als volgt uit:

```
som-xi ← 0;  som-yi ← 0
som-xiyi ← 0;  som-xixi ← 0
p ← INVOER
p keer
    xi ← INVOER
    yi ← INVOER
    som-xi ← som-xi+xi
    som-yi ← som-yi+yi
    som-xiyi ← som-xiyi+xi*yi
    som-xixi ← som-xixi+xi*xi
noemer ← p*som-xixi-som-xi*som-xi
m ← (p*som-xiyi-som-xi*som-yi)/noemer
n ← (som-xixi*som-yi-som-xi*som-xiyi)/noemer
UITVOER m,n
```

De vertaling van dit schema naar BASIC, ECOL, COMAL, PASCAL of wat er al ter beschikking staat, levert een programma waarmee in ieder geval de leraar zijn benadering kan opstellen (ook iets voor de leerling?).

Een moeilijker geval

Wederom uit 'Groei' een tabel, die aangeeft het aantal legale abortussen in India dat in de aangegeven jaren heeft plaatsgevonden, per 1000 vrouwen tussen de 15 en 44 jaar oud.

jaar	1972	1973	1974	1975
aantal	0.2	0.4	0.8	1.5

We willen (voor het doen van voorspellingen voor welk politiek doel dan ook) een functievoorschrift hebben.

Uitzetten in de grafiek laat duidelijk (op het oog) zien dat het geen lineair verband is.

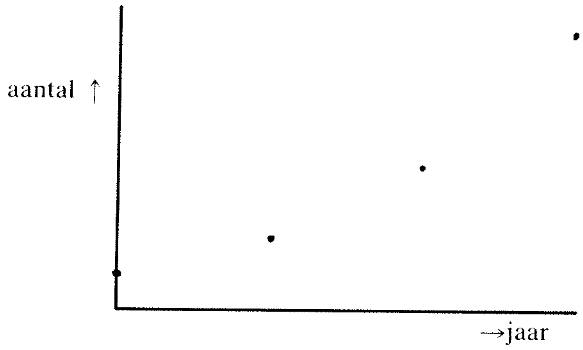


fig. 3

De leerlingen krijgen de opdracht de gegevens uit te zetten op logaritmisch grafiekenpapier:

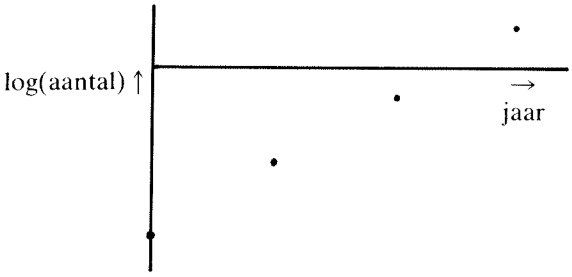


fig. 4

Nu is het (op het oog) weer een rechte lijn. (De factor 2 had u natuurlijk allang gezien in de tabel).
In feite is de transformatie $y' = \log y$ gebruikt en het resultaat is (ongeveer) een rechte lijn. Dat wil zeggen dat we met het ontworpen computerprogramma de lineaire benadering kunnen vinden van $y' = mx+n$.
Daartoe moet de transformatie opgenomen worden in het programma door in het structuurschema ná regel 6 toe te voegen:

6.	yi ← INVOER
6.a	yi ← log(yi)

Tot welk functievoorschrift leidt dit nu?
Laten we eens wat rekenen:

$$y' = mx+n$$
$$\log y = mx+n$$
$$y = 10^{mx+n}$$
$$y = (10^m)^x \cdot 10^n$$

Kennelijk vinden we een exponentiële functie:

$$y = c \cdot g^x$$

waarbij $c = 10^n$ en $g = 10^m$.
We voegen dit toe aan het computerprogramma. Na regel 13 veranderen we het structuurschema als volgt:

13a.	c ← 10 ↑ n
13b.	g ← 10 ↑ m
14.	UITVOER c,g

Hiermee beschikken we over een programma dat ons vlot, na invoering van de meetgegevens, de parameters van de benaderende exponentiële functie levert. (Men noemt dit exponentiële regressie).

Nog iets moeilijker?

De A-leerling krijgt ook de opdracht een bepaalde tabel in een grafiek om te zetten op dubbellogaritmisch grafiekenpapier.
Wat is er nu aan de hand als dit, op het oog, een rechte lijn blijkt te zijn?
We rekenen weer even wat:

$y' = \log y$ en $x' = \log x$

verder met $y' = mx' + n$

$$\log y = m \cdot \log x + n$$
$$\log y = \log x^m + \log 10^n$$
$$\log y = \log 10^n \cdot x^m$$
$$y = 10^n \cdot x^m$$

Kennelijk hebben we nu te maken met een functievoorschrift van de vorm:

$y = c \cdot x^m$ met $c = 10^n$

Het structuurschema wordt dus als volgt gewijzigd:

5.	xi ← INVOER
5a.	xi ← log(xi)
6.	yi ← INVOER
6a.	yi ← log(yi)
	. . . .
13a.	c ← 10 ↑ n
14.	UITVOER c,m

en een en ander is weer geautomatiseerd.

Het allermoeilijkste?

Nu andersom en terug naar het probleem van de inleiding van dit artikel.
We hebben het vermoeden dat bepaalde meetresultaten zullen passen in het logistische model:

$$S: x \rightarrow \frac{k}{1 + c \cdot e^{-r \cdot x}}$$

Kunnen we dit grafisch toetsen en geautomatiseerd de parameters bepalen?
Weer wat rekenwerk:

$$\frac{k}{y} = 1 + c \cdot e^{-rx}$$
$$\frac{k}{y} - 1 = c \cdot e^{-rx}$$
$$\ln\left(\frac{k}{y} - 1\right) = -rx + \ln(c)$$
$$y' = mx' + n$$

Kennelijk is de gewenste transformatie:

$y' = \ln(\frac{k}{y} - 1)$ en $x' = x$

Hieraan is te zien dat k bekend moet zijn. We moeten dus een vermoeden uiten over het niveau waarop zich het proces op den duur zal stabiliseren.

Het programma moet nu als volgt gewijzigd worden:

```
3.   p  ← INVOER
3a.  k  ← INVOER
        . . . . . . . . . .
5.       xi  ← INVOER
6.       yi  ← INVOER
6a.      yi  ← ln(k/yi−1)
        . . . . . . . . . .
13a. r  ← −m
13b. c  ← 2.718281828 ↑ n
14.  UITVOER c,r
```

We kunnen dus zelfs 'logistische regressie' plegen!

Een voorbeeld.

De groei van gistcellen werd gemeten:

t	0	1	2	3	4	5	6
y	9.6	18.3	29.0	47.2	71.1	119.1	174.6
y'	4.2	3.6	3.1	2.6	2.1	1.5	1.0

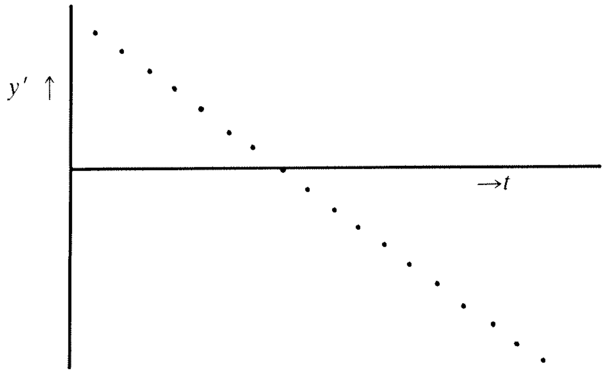
t	7	8	9	10	11	12	13
y	257.3	350.7	441.0	531.3	559.7	594.8	629.4
y'	0.5	−0.1	−0.7	−1.4	−1.7	−2.1	−2.9

t	14	15	16	17	18
y	640.8	651.1	655.9	659.6	661.8
y'	−3.3	−3.9	−4.3	−4.8	−5.3

De grafiek van y is:

We herkennen de S-vorm: logistisch model?

De grafiek van y':



Het logistische model past zo op het oog.
We gebruiken de computer:
Het vermoeden is dat $k = 665$ (zie tabel); $p = 19$;
het resultaat: $r = 0.5$; $c = 64.0$
Het model is daarmee:

$$y = \frac{665}{1 + 64.0\,e^{-0.5x}}$$

Slot

Het hier aangelegde verband tussen groei, grafieken en gegevensverwerking is de volgende. Meetresultaten van een (groei)proces kunnen door de leerlingen op de aard van groei (wiskundig model) getest worden door de resultaten in een grafiek op te nemen. Bij een bepaalde transformatie van de assen kan een (zo goed als ((3.))) rechte lijn ontstaan. Wanneer dat zo is, kan met behulp van een computerprogramma, met erin de transformatie verwerkt, gemakkelijk de parameters bepaald worden van het functievoorschrift. Dit kan vervolgens als goede benadering dienst doen voor het doen van voorspellingen van het (groei)proces.

Noten

(1) Wanneer een, 'op het oog' beste, lijn wordt getrokken, wordt een ander criterium impliciet gehanteerd. Dan worden, naar mijn mening , de loodrechte afstanden tussen punten en lijnen geschat en niet de verticale afstanden.
(2) Natuurlijk moet eigenlijk nog worden aangetoond dat hier sprake is van een minimum, maar daar ga ik nu aan voorbij.
(3) Met wat wiskunde (regressie-analyse) zijn criteria te ontwikkelen die wat meer greep op wat een 'redelijke' benadering is.

Vrouw & Wiskunde

Op zaterdag 30 maart a.s. zal de zevende landelijke dag van de groep 'Vrouwen en Wiskunde' gehouden worden in Cunera, Nieuwe Gracht 32, Utrecht, van 9.30 uur tot 17.00 uur.

Thema's van de dag zijn:
1. Terugblik op de Studiedag van de NVvW op 27 okt. 1984 die werd georganiseerd door de groep 'Vrouwen en Wiskunde'.
2. Wiskunde verplicht?

Voor verdere informatie: Marijke Melis, tel. 04950-34662.