STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 168 (M 64)

Het aanpassen van een exponentiële regressielijn

H. Kesten en J.Th. Runnenburg

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
A m s t e r d a m - O.

Rapport S 168(M 64)
door H. Kesten en
J. Th. Runnenburg.

Het aanpassen van een exponentiële regressielijn.[1]

1. Het probleem.

Laat ξ en η twee variabelen zijn, die voldoen aan de betrekking

(1) $$\eta = \theta + \alpha\, 10^{-\beta\xi},$$

waarin θ, α en β onbekende constanten zijn. Laat verder een aantal (N) aequidistante waarden van ξ gegeven zijn en wel

(2) $$x_i = c + i d \qquad (c \text{ en } d \text{ bekend; } i = 1, \ldots, N).$$

Bij iedere x_i behoort dan, volgens (1), een waarde η_i van η :

(3) $$\eta_i = \theta + \alpha\, 10^{-\beta x_i} \qquad (i = 1, \ldots, N).$$

Deze η_i worden nu waargenomen, waarbij echter toevallige afwijkingen optreden. Noemen wij de waarnemingen $\underline{y}_i$, dan geldt dus

(4) $$\underline{y}_i = \eta_i + \underline{u}_i \qquad (i = 1, \ldots, N).$$ [2]

Over de waarschijnlijkheidsverdeling der $\underline{u}_i$ is vaak wel iets bekend. Meestal wordt ondersteld, dat de $\underline{u}_i$ onderling onafhankelijk verdeeld zijn met verwachting 0. Soms ook dat zij gelijke spreidingen bezitten en normaal verdeeld zijn.

Het probleem is nu, een methode aan te geven, om de onbekende parameters θ, α en β te schatten op grond van de waarnemingen $(x_1, y_1), \ldots, (x_n, y_n)$. In dit memorandum wordt een eenvoudige methode daartoe beschreven. De eigenschappen van de verkregen schattingen, die afhangen van de over de $\underline{u}_i$ gemaakte onderstellingen, worden hier niet besproken.

----

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Onderstreping van een letter duidt aan, dat de grootheid stochastisch is; dezelfde letter zonder onderstreping wordt gebruikt voor een door de stochastische grootheid aangenomen waarde.

2. Schatting van β.

Laat N een even getal zijn

(5) $$N = 2n.$$

(Is N oneven, dan wordt het laatste punt niet gebruikt voor de schatting van β).

Beschouw nu, voor $k \leqq n$, het verschil

$$\eta_k - \eta_{n+k} = \alpha\left[10^{-\beta(c+k\alpha)} - 10^{-\beta\{c+(n+k)\alpha\}}\right] =$$
$$= \alpha \cdot 10^{-\beta(c+k\alpha)}\left\{1 - 10^{-\beta n\alpha}\right\},$$

en definieer:

$$\zeta_k \stackrel{def}{=} {}^{10}\log(\eta_k - \eta_{n+k}) = {}^{10}\log\alpha - \beta(c+k\alpha) + {}^{10}\log\left\{1 - 10^{-\beta n\alpha}\right\} =$$

(6) $$= C - \beta(c+k\alpha) = C - \beta x_k,$$

met $$C = {}^{10}\log\alpha + {}^{10}\log\left\{1 - 10^{-\beta n\alpha}\right\}.$$

Stel verder

(7) $$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \qquad \left(\sum_{k=1}^{n}(x_k - \bar{x}) = 0\right).$$

en vorm:

(8) $$\sum_{k=1}^{n}(x_k - \bar{x})\zeta_k = C\sum_{k=1}^{n}(x_k - \bar{x}) - \beta\sum_{k=1}^{n}(x_k - \bar{x})x_k =$$
$$= -\beta\sum_{k=1}^{n}(x_k - \bar{x})^2.$$

Dan is dus

(9) $$\beta = -\frac{\sum_{k=1}^{n}(x_k - \bar{x})\zeta_k}{\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$

Op grond hiervan ligt het voor de hand, als schatting voor β te nemen:

(10) $$\boxed{b \stackrel{def}{=} -\frac{\sum_{k=1}^{n}(x_k - \bar{x})z_k}{\sum_{k=1}^{n}(x_k - \bar{x})^2} = -\frac{6\sum_{k=1}^{n}\{2k - (n+1)\}z_k}{n(n^2-1)\alpha},}$$

met

(11) $$\boxed{z_k \stackrel{def}{=} \log(y_k - y_{n+k}).}$$

3. Schattingsmethode voor θ en α.

Na bepaling van b kunnen wij θ en α met behulp van de methode der kleinste kwadraten schatten. Wij zoeken daartoe de waarden t en a van θ en α, die

$$\sum_{i=1}^{N} (y_i - \theta - \alpha\, 10^{-bx_i})^2 \tag{12}$$

minimaal maken, waarbij, ook als N oneven is, alle waarnemingen gebruikt kunnen worden. Derhalve moet gelden:

$$\sum_{i=1}^{N} y_i - Nt - a \sum_{i=1}^{N} 10^{-bx_i} = 0 ,$$

en

$$\sum_{i=1}^{N} y_i\, 10^{-bx_i} - t \sum_{i=1}^{N} 10^{-bx_i} - a \sum_{i=1}^{N} 10^{-2bx_i} = 0 .$$

Oplossing van deze vergelijkingen geeft als schattingen t en a:

$$t = \frac{\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} 10^{-2bx_i}\right) - \left(\sum_{i=1}^{N} y_i\, 10^{-bx_i}\right)\left(\sum_{i=1}^{N} 10^{-bx_i}\right)}{\left(N \sum_{i=1}^{N} 10^{-2bx_i}\right) - \left(\sum_{i=1}^{N} 10^{-bx_i}\right)^2}$$

of

$$\boxed{t = \frac{\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} 10^{4-2bx_i}\right) - \left(\sum_{i=1}^{N} y_i\, 10^{2-bx_i}\right)\left(\sum_{i=1}^{N} 10^{2-bx_i}\right)}{\left(N \sum_{i=1}^{N} 10^{4-2bx_i}\right) - \left(\sum_{i=1}^{N} 10^{2-bx_i}\right)^2}} \tag{13}$$

en

$$a = \frac{\left(N \sum_{i=1}^{N} y_i\, 10^{-bx_i}\right) - \left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} 10^{-bx_i}\right)}{\left(N \sum_{i=1}^{N} 10^{-2bx_i}\right) - \left(\sum_{i=1}^{N} 10^{-bx_i}\right)^2}$$

of

$$\boxed{a = 10^2\, \frac{\left(N \sum_{i=1}^{N} y_i\, 10^{2-bx_i}\right) - \left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} 10^{2-bx_i}\right)}{\left(N \sum_{i=1}^{N} 10^{4-2bx_i}\right) - \left(\sum_{i=1}^{N} 10^{2-bx_i}\right)^2}} \tag{14}$$

De teller en noemer zijn, om (13) en (14) te verkrijgen, met 10^4 vermenigvuldigd in de formules voor t en a om te voorkomen dat negatieve machten van 10 (i.c. 10^{-bx_i}) opgezocht moeten worden. De exponent 4 is aangepast aan ons getallenvoorbeeld. In andere gevallen zal misschien een hogere exponent gewenst zijn.

4. Opmerkingen.

De betrekking $\eta = \theta + \alpha\, 10^{-\beta \xi}$ is aequivalent met

$$\eta = \theta + \alpha \,.\, e^{-\beta \,.\, {}^{e}\log 10 \,.\, \xi} = \theta + \alpha \,.\, e^{-2{,}3026\, \beta \xi}. \tag{15}$$

In dit memorandum is alleen de eerste uitdrukking beschouwd, omdat tabellen van ${}^{10}\log x$ beter beschikbaar zijn dan tabellen van ${}^{e}\log x$. Het is verder aan te bevelen de x-schaal zo te kiezen, dat

$$x_1 = 0 \quad (c = -d) \text{ en } x_i = (i-1)\, d \text{, eventueel met } d = 1 . \tag{16}$$

In dat geval is α precies het verschil tussen de eerste waarde van η en de asymptotische waarde ($\xi \to \infty$) van η , en a is hiervan de schatting.

De aangepaste waarden (of: regressiewaarden) Y_i worden gegeven door

$$Y_i = t + a \,.\, 10^{-b x_i} \qquad (i = 1, \ldots, N). \tag{17}$$

5. Voorbeeld.

Waarnemingen: (met $c = -1$, $d = 1$ $N = 11$)

i	x_i	y_i
1	0	46,5
2	1	4[illegible],5
3	2	40,0
4	3	38,0
5	4	36,5
6	5	36,0
7	6	35,5
8	7	34,5
9	8	33,5
10	9	32,5
11	10	32,5

Berekening van b.

z_k $\qquad\qquad\qquad\qquad\qquad\qquad$ $2k-(n+1)$ $\qquad$ $n=5$, $n(n^2-1)=120$

$z_1 = {}^{10}\log\,(46{,}5 - 36{,}0) = 1{,}021$ $\qquad$ -4

$z_2 = {}^{10}\log\,(42{,}5 - 35{,}5) = 0{,}845$ $\qquad$ -2

$z_3 = {}^{10}\log\,(40{,}0 - 34{,}5) = 0{,}740$ $\qquad$ 0

$z_4 = {}^{10}\log\,(38{,}0 - 33{,}5) = 0{,}653$ $\qquad$ 2

$z_5 = {}^{10}\log\,(36{,}5 - 32{,}5) = 0{,}602$ $\qquad$ 4

$$\sum_{k=1}^{5} z_k\,[2k-(n+1)] = -\,2{,}060,$$

$$b = +\,\frac{6 \,.\, 2{,}060}{120} = 0{,}103.$$

Indien z_k tot in r decimalen nauwkeurig bepaald wordt, is de maximale fout in b tengevolge van afrondingen bij de berekening $\approx \frac{3}{2nd}\,.\,10^{-r}$. De stochastische fout $(= b-\beta)$ kan echter groter zijn.

Berekening van t en a.

x_i	$2-bx_i$	10^{2-bx_i}	y_i	$Y_i = t + a \,.\, 10^{-bx_i}$
0	2,000	100,0	46,5	45,9
1	1,897	78,9	42,5	42,8
2	1,794	62,2	40,0	40,4
3	1,691	49,1	38,0	38,5
4	1,588	38,7	36,5	36,9
5	1,485	30,6	36,0	35,7
6	1,382	24,1	35,5	34,8
7	1,279	19,0	34,5	34,0
8	1,176	15,0	33,5	33,4
9	1,073	11,8	32,5	33,0
10	0,970	9,3	32,5	32,6

$$\sum_{i=1}^{N} 10^{2-bx_i} = 438{,}7\ ,$$

$$\sum_{i=1}^{N} 10^{4-2bx_i} = 26331{,}45\ ,$$

$$\sum_{i=1}^{N} y_i = 408{,}0\ ,$$

$$\sum_{i=1}^{N} y_i\, e^{2-bx_i} = 17570{,}5\ ,$$

$$t = \frac{408{,}0 \times 17570{,}5 - 17570{,}5 \times 438{,}7}{11 \times 26331{,}45 - 438{,}7 \times 438{,}7} = 31{,}23,$$

$$a = \frac{11 \times 17570{,}5 - 408{,}0 \times 438{,}7}{11 \times 26331{,}45 - 438{,}7 \times 438{,}7} = 14{,}70 .$$

Indien θ , α en β ongeveer de waarden hebben die hier gevonden zijn, en de berekening gemaakt wordt zoals hier is aangegeven, zal de fout in t en a , die door afronding veroorzaakt wordt, $\leqq$ 0,05 zijn. De fout in de aanpassing is dan $\leqq$ 0,1.

Literatuur, waarin andere methoden behandeld worden:

D.J. Cowden, Simplified methods of fitting certain types of growth curves. Journ.Am.Stat.Ass. 42 (1947) 585-590.

E.S. Keeping, A significance test for exponential regression. Ann.Math.Stat. 22 (1951) 180-198.

D.S. Villars, A significance test and estimation in the case of exponential regression. Ann.Math. Stat. 18 (1947) 596-600.