K O N I N K L I J K   N E D E R L A N D S

M E T E O R O L O G I S C H   I N S T I T U U T

Wetenschappelijk Rapport W.R. 55-005 (III-154)

Dr C. Levert

Commentaar bij de
Methode van Craddock voor het extrapoleren van tijdreeksen

De Bilt, 1955

KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

Wetenschappelijk Rapport
W.R. 55-005 (III-154)
U.D.C. 551.545 : 519.2

Dr C. Levert

Commentaar bij de Methode van Craddock voor het extrapoleren van tijdreeksen

0 Inleiding

In Januari 1955 zond de Heer J.M. Craddock (Meteorological Office, Air Ministry; Dunstable) een brief aan Prof. Berlage, waarin hij herinnert aan wat hij in September 1954 te Rome (X U.G.G.I.-Congres) vertelde van een nieuwe methode om tijdreeksen te analyseren, terwijl hij deze methode thans volledig beschijft.

Op verzoek van Prof. Berlage hebben wij getracht de statistische achtergrond van het "Craddock-procédé" te ontdekken. Dit is niet volledig mogen gelukken. Aangezien de tijd ontbreekt om dieper op de kwestie in te gaan hebben wij gemeend er goed aan te doen de tot nu toe verkregen resultaten in een rapport samen te nemen.

Nagenoeg de gehele brief in kwestie wordt in 1 letterlijk geciteerd. Wij hebben hem van kantcijfers voorzien, waarnaar wij in 3 verwijzen. De brief munt niet steeds uit door duidelijkheid; op vele plaatsen moeten wij raden naar Craddock's bedoeling. In 2 hebben wij zijn "theorie" met eigen woorden (en de notatie ietwat logischer makende), zo scherp mogelijk, naverteld. In 3 hebben wij onze op- en aanmerkingen bij de tekst van Craddock's brief samengenomen. Tenslotte nemen wij in 4 bijeen al onze berekeningen en beschouwingen, die licht kunnen werpen op de statistische achtergrond van deze nieuwe methode voor het extrapoleren van tijdreeksen.

1 De brief van J.M. Craddock

We are considering time series of meteorological variables, in which values are given at equal intervals of time, and the unit of time will be taken as the interval between consecutive observations. Let such a series be $f_1, f_2, f_3, \ldots\ldots f_n, \ldots\ldots$ etc. Then one of our problems in forecasting is to make the best estimate we can of f_{n+1} in terms of the preceding terms up to f_n. We suppose that the series, like most of those
1 which we meet in practice, contains an element which can be ascribed to the working of definite physical processes and which can be expressed in mathematical terms, and also an element which cannot be so expressed, and
2 is effectively random. Thus even if we van identify and describe the physical processes operating, we cannot predict the series exactly, owing
3 to the random element. We suppose that the working of the physical processes at any time is expressed by a function of the form $f' = F(t)$, where f' differs from f in the omission of the random element and F(t) always has
4 the same form, but with parameters varying from time to time. Then, given the series up to term f_n, we wish to find a method of determining the
5 parameters of F(t) so as to give, on average, the best estimate of f_{n+1}.

6 That is, we wish to minimise $f_{n+1} - F\ (n+1)$ in some way. If we denote by
$F_n(t)$ the chosen function, fitted to the series up to term f_n, then we
7 propose to fit $F_n(t)$ by the method of least squares. So far there is
nothing original in this, for the steps taken include fitting by harmonic
analysis as a special case. However, we now argue that it is more important
8 to secure a close fit to the more recent terms than to the earlier terms,
and we weight the squared residuals, of which the sum has to be minimised,
with positive weights decreasing geometrically into the past. In other
words, instead of minimising

9
$$S = \sum_{j=0}^{N} \left[f_{n-j} - F_n(-j) \right]^2 \qquad - (1)$$

which is what is done in harmonic analysis, we minimise

10
$$S = \sum_{j=0}^{\infty} r^j \left[f_{n-j} - F_n(-j) \right]^2 \qquad - (2)$$

where r is a positive quantity less than unity.

This is the original feature. If $F(t)$ is made up of a sum of
11 functions of t, with arbitrary coefficients a,b,c etc, then the conditions
for minimum S are, of course, that

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0 \qquad - (3)$$

Having solved these equations to determine $F_n(t)$, we write

$$e_{n+1} = f_{n+1} - F_n\ (\ 1\ ) \qquad - - - - (4)$$

12 We then assume F_{n+1} to be known, find $F_{n+1}(t)$ and e_{n+2}, and
13 so on. Finally we form $\sum e_n^2$, which is, of course, a function of r.

The whole process then has to be repeated with different values
of r within the permissible range (about 0.5 to 1.0), and we choose the
14 value which minimises $\sum e_n^2$. Thus we have found the value of r which
15 enables the best predictions to be made, given the assumed form of the
16 function $F(t)$, and also the mean square error of forecasts using this
value of r.

Stated like this, the process sounds impossibly complicated
and laborious, but actually the forms of the function ($F(t)$ which we
17 would wish to use are very simple, and in some cases at any rate the
18 calculation reduces to a series of straightforward iterations.

We have examined the cases when F(t) is a polynomial of degree
19 0, 1 or 2, and when F(t) is the sum of a constant and a sine term.

As a practical example, if F(t) = a, a polynomial of degree 0, the quantity to be minimised is

20 $$S = \sum_{0}^{\infty} r^j (f_{n-j} - a_n)^2 \qquad -(5)$$

The solution is

21 $$a_n = (1-r) \sum_{0}^{\infty} r^j f_{n-j} \qquad --(6)$$

22 and the iteration relations are $e_{n+1} = f_{n+1} - a_n$ --(7)

and

23 $$a_{n+1} = a_n + (1-r) e_{n+1} \qquad --(8)$$

In practice, we do not use equation (6) at all, but find a suitable
24 starting value a_0, e.g. by averaging the first ten terms of the series, and then use equations (7) and (8) with n = 0.

We have applied these formulae to several time series, among
25 them the Greenwich annual mean temperatures for the years 1841-1940, with the following results.

Table 1

26

r	$\frac{1}{N}\sum e_n^2$ °F^2
·50	1·324
·67	1·219
·80	1·150
·86	1·128
·90	1·132
·95	1·154
·98	1·301

Hence for this series the best value of r for prediction purposes is about ·86 or ·87.

I should emphasise that so far we have had very little practical experience in using this method, while I have not yet had time to look into
27 the statistical background. Its main recommendations are that it is
28 physically reasonable, and capable of dealing with cases such as a temporary
29 harmonic oscillation produced by random impulses where an ordinary harmonic analysis is likely to produce negative misleading results.

2 Het "Craddock-procédé"

Gegeven zij een tijdreeks van meteorologische variabelen, waarin waarden gegeven worden na gelijke tijdsintervallen. Dit interval tussen successieve waarnemingen zal als tijdseenheid gekozen worden. Zij zulk een
(1) reeks f (1), f (2) ..... f (n) ..... f (N). Wij onderstellen dat iedere term f(i) bestaat uit een deel F(i) , dat bewerkstelligd is door een bepaald physisch proces en een van het toeval afhankelijk deel ($\mathcal{E}_i$). Het eerste deel hopen wij in een mathematische formule F(t), een functie van de tijd t (= 1, 2, 3, ...), te kunnen opnemen; het tweede deel gehoorzaamt aan toevalswetten; wel hopen wij de karakteristieken van de waarschijnlijkheids-
(2) verdeling van dit deel te vinden. Dus $f(t) = F(t) + \mathcal{E}$. In de onderstelling
(3) dat voor elke t deze $\mathcal{E}$ een normale verdeling heeft (met gemiddelde $\mu = 0$ en - onbekende - standaarddeviatie σ); verder, dat deze $\sigma_{\mathcal{E}}$ niet van t afhankelijk is en ten slotte, dat de $\mathcal{E}$'s statistisch onafhankelijk (ongecorreleerd) zijn, zal de "beste" ofwel de "meest aannemelijke" functie $F_n(t)$, berekend met behulp van de eerste n elementen, zulke parameters α_{n1}, α_{n2}, ... α_{np} hebben dat de uitdrukking

(4) $$S = \sum_{i=1}^{n} \left[f(i) - F_n\,(i) \right]^2$$

minimaal is voor deze constellatie van p parameters (omdat voor de berekening van deze p parameters gebruik gemaakt is van de eerste n elementen hebben wij de n in de index opgenomen: α_{nj} , met j = 1, 2, ... p; dezelfde betekenis heeft de aan F gehangen index n). Wij leggen t = 1 in i = 1 [1);]
dus

i = t = 1	2	3	4	5	enz.
f(1)	f(2)	f(3)	f(4)	f(5)	enz.

[1)] Deze notatie wijkt af van die van Craddock; wij vinden haar logischer en overzichtelijker.

De p parameters volgen uit de p vergelijkingen

(5) $$\frac{\partial S}{\partial \alpha_{n1}} = 0 \quad \frac{\partial S}{\partial \alpha_{n2}} = 0 \quad \ldots\ldots \quad \frac{\partial S}{\partial \alpha_{np}} = 0$$

Als de physische processen bekend zijn, d.w.z. beschreven kunnen worden in mathematische vorm, dan is F een bekende functie van i en de parameters α_j ($j = 1, 2, \ldots\ldots p$), voor welke zodanige waarden gekozen moeten worden, dat de genoemde S minimaal wordt. In de meeste gevallen echter kent men die processen niet of slecht, en neemt men functies aan, die voor ons geval een redelijke aanpassing beloven.

Men kan het van belang achten, aan de meer recente termen een grotere invloed toe te kennen dan aan de lang geledene en dus gewichten g_i hechten aan de gekwadrateerde residuen $\mathcal{E}_i$. Het residu $\mathcal{E}_i$ is gedefinieerd

(6) als $f(i) - F_n(i)$. Deze gewichten kiest men dan zodanig, dat $g_1 < g_2 < \ldots\ldots$

(7) $< g_n$. Men zou kunnen stellen $g_i = r^{n-i}$ (met $0 < r < 1$); $i = 1, 2, \ldots\ldots n$. Men moet nu minimaliseren:

(8) $$S^{*}(r, \alpha_{n1}, \alpha_{n2} \ldots \alpha_{np}) \equiv \sum_{i=1}^{n} r^{n-i} \, \varepsilon^2(i) = \sum_{i=1}^{n} r^{n-i} \left[f(i) - F_n(i, \alpha_{n1}, \ldots \alpha_{np}) \right]^2$$

De p parameters volgen nu uit de p vergelijkingen

(9) $$\frac{\partial S^{*}}{\partial \alpha_{nk}} = 0 \; ; \; k = 1, \ldots\ldots p.$$ Iedere α_{nk} is dan een functie van r (ook eigenlijk van n).

Deze α's kunnen weer gesubstitueerd worden in de $F_n(i)$, zodat deze $F_n(i)$ een functie wordt van i, r en n. Aan de grootheid r wordt voorlopig nog geen getallenwaarde toegekend; men laat haar door het verdere procédé bepaald worden. De aldus gevonden $F_n(i, r, n)$ wordt gebruikt om een beste schatting te maken van de waarde van de term op $i = n+1$ (dus $F_n(n+1, r, n)$. Wij meten $f(n+1)$, zodat het verschil tussen meting en

(10) berekening is $\mathcal{E}(n+1) = f(n+1) - F_n(n+1, r, n)$

Evenzo zouden wij twee plaatsen vooruit kunnen voorspellen, d.w.z. $F_n(n+2, r, n)$ kunnen berekenen, om vervolgens het verschil met de meting $f(n+2)$ op te maken. Dit doen wij niet. In de plaats daarvan herhalen wij het procédé, d.w.z. wij berekenen, als boven, opnieuw de F, doch nu op basis van de eerste n+1 termen. Thans moeten wij minimaliseren:

(11) $$S^{*} = \sum_{i=1}^{n+1} r^{n+1-i} \, \varepsilon^2(i) \equiv \sum_{i=1}^{n+1} r^{n+1-i} \left[f(i) - F_{n+1}(i, \alpha_{n+1,1} \ldots, \alpha_{n+1,p}) \right]^2$$

Wij moeten de p_x onbekenden $\alpha_{n+1,i}$ $(i=1, 2, \ldots p)$ oplossen uit de p vergelijkingen : $\frac{\partial S^x}{\partial \alpha_{n+1,k}}$ $(k = 1, 2, \ldots p)$.

Iedere $\alpha_{n+1,k}$ wordt een functie van r en n+1 (deze functie zal meestal een andere zijn dan $\alpha_{n,k}$). Deze substituerende in F_{n+1} komt er een F_{n+1} (i, r, n+1) [1] Nu gaan wij hiermee de beste schatting maken van de waarde van de term op t = n+2. Zij wordt F_{n+1} (n+2, r, n+1). Wij maten

(13) f(n+2) en dus is het residu $\varepsilon(n+2) = f(n+2) - F_{n+1}(n+2, r, n+1)$.

Zo gaan we door. Er komen nog achtereenvolgens de residuen (afwijkingen):

(14) $\varepsilon(n+3) = f(n+3) - F_{n+2}(n+3, r, n+2)$

$\varepsilon(n+4) = f(n+4) - F_{n+3}(n+4, r, n+3)$ enz. en ten slotte ook

$\varepsilon(N) = f(N) - F_{N-1}(N, r, N-1)$. Ten slotte moeten wij ook minimaliseren

$$(15)\quad S^* = \sum_{i=1}^{N} r^{N-i} \left\{ f(i) - F_N(i, \alpha_{N1}, \ldots \alpha_{N2}) \right\}^2$$

Zo komen wij tot F_N (i, r, N), welke wordt gebruikt om een beste schatting te maken van de waarde van de term op plaats N+1, welke term f(N+1) nog moet komen (hier dus werkelijk "voorspellen"). D.w.z.: de meest waarschijnlijke waarde, die f(N+1) zal hebben, wordt gegeven door $F_N(N+1, r, N)$. Doch deze F bevat nog de onbekende r. Eerst thans ligt het voor de hand een conditie voor deze r op te stellen, met behulp van welke de numerieke waarde van r bepaald kan worden. D.w.z. wij eisen zulk een r^* , dat

$$(16)\quad T \equiv \sum_{i=n+1}^{N} \varepsilon^2(i) \quad \text{minimaal is.}$$

Dit te eisen staat immers gelijk met ons streven de N-n voorspellingen

(17) F_{n+j} (n+j+1, r, n+j) zo dicht mogelijk te doen "liggen" bij de overeenkomstige metingen f(n+j+1), met j = 0, 1, ... N-n-1 (d.w.z. een "zo goed mogelijke" voorspelling). Deze T is een functie van r (en eigenlijk ook van n). Het gebruikelijke procédé is dus de r^* op te lossen uit $\frac{dT}{dr} = 0$. Aangezien het niet vast staat dat er wortels zullen zijn in het traject 0 tot 1 (grenzen medegerekend) (en zo zij er zijn, moet T daarbij bovendien

---

1) Meestal zal $F_n(n+1, r, n) \not\equiv F_{n+1}(n+1, r, n+1)$. Dit behoeft niet te verbazen, want in het linkerlid staat iets dat niet met f(n+1) te maken heeft, terwijl het rechterlid wel met f(n+1) samenhangt.

een minimum en niet een maximum zijn), doen wij beter te stellen: gezocht wordt naar die r^*, waarvoor T in het traject 0 t/m 1 (of, naar men verkiest, 0.3 t/m 1 of 0.5 t/m 1 enz.) een minimum bereikt.
(hierop komen wij terug in 3, blz. 11)
Als nu r^* numeriek vastligt, ligt ook de minimum-waarde T_{min} van T vast. Uit deze T_{min} volgt weer de "standaarddeviatie σ_v van de voorspelling" wanneer wij deze definiëren a.v.

(18) variantie = $\sigma_v^2 = T_{min}/(N-n-g-1)$

Hierbij is ondersteld, dat F de vorm heeft $F = \alpha_1 + \alpha_2 i + \alpha_3 i^2 + \alpha_4 i^3 + \ldots \alpha_{g-1} i^g$ (graad g; g+1 constanten). Zouden wij niet een rationale functie proberen, maar bijv. de volgende transcendente $\alpha_1 + \alpha_2 \sin(\alpha_3 i + \alpha_4)$ dan zou de noemer zijn N-n-4 (4 constanten α_1, α_2, α_3 en α_4).
Maar wij kunnen in de variantie ook de aanpassing aan de eerste n termen betrekken. In dat geval zouden wij kunnen berekenen:

(19) $$\sigma_a^2 = \frac{\sum_{i=1}^{n} \left[\varepsilon(i) - F_n(i, r^*, n)\right]^2}{n-g-1}$$

Wij zouden kunnen spreken van de voorspellings- σ_v en de aanpassings- σ_a
Tenslotte zou men ook kunnen berekenen een σ_{a+v} uit

(20) $$\sigma_{a+v}^2 = \frac{\sum_{i=1}^{n} \left[\varepsilon(i) - F_n(i, r^*, n)\right]^2 + \sum_{n+1}^{N} \left[\varepsilon(i) - F_{i-1}(i, r^*, i-1)\right]^2}{N-g-1}$$

Men kan dan zeggen

1° De gemiddeld te verwachten (tevens meest waarschijnlijke) waarde van f_{N+1} is $\hat{f}_{N+1} = F_N(N+1, r^*, N)$

2° De uitspraak, dat de komende f_{N+1} zal liggen tussen $F_N(N+1) - k.\sigma_v$ en $F_N(N+1) + k.\sigma_v$ (of wil men liever met σ_{a+v} werken?) heeft een betrouwbaarheid B%.

Enige k, B-combinaties: k=1, B=68; 1,5 87; 1.96 95; 2.58 99.

Enige opmerkingen:

a. Welke n moet men kiezen?
Craddock laat zich hierover niet uit. De kleinste n is 2.

b. Het zal bij dit "stapjes-procédé" van groot belang zijn, dat de iteratieformules eenvoudig zijn. Wij bedoelen het volgende: Uit (9) volgen zekere $\alpha_{n1} = \alpha_{n1}(r, n)$; $\alpha_{n2} = \alpha_{n2}(r, n)$ ..... $\alpha_{n,p} = \alpha_{n,p}(r, n)$. Uit (12) volgen nieuwe $\alpha_{n+1,1} = \alpha_{n+1,1}(r, n+1)$; ..... $\alpha_{n+1,p} = \alpha_{n+1,p}(r, n+1)$.

Zo ook $\alpha_{n+2,j}$; $\alpha_{n+3,j}$; enz. Omdat men (16) dikwijls slechts approximatief kan oplossen, zal het zeer aangenaam zijn als men elke $\alpha_{n+1,j}$ ($j=1,2$ .....p) kan uitdrukken in de $f(1)$, $f(2)$ ..... $f(n+1)$ én α_{ns} ($s=1,2$, ... p), liefst α_{nj} alleen. Men kan dan beginnen met (9) op te lossen; d.w.z. de α_{n1}, α_{n2} ..... α_{np} worden functies van r en n. Men vult nu een waarde r_1 in, die men dicht bij de oplossing van (16) denkt te zullen liggen. Aldus liggen α_{n1}, α_{n2} ..... α_{np} numeriek vast. Daardoor ook $\mathcal{E}(n+1)$. Met behulp der recurrente formules liggen nu ook de $\alpha_{n+1,j}$ vast. Hierdoor ligt ook $\mathcal{E}(n+2)$ vast. Nu weer, via $\alpha_{n+2,j}$, ook de $\mathcal{E}(n+3)$, etc. Zo komt men tot zekere waarde T_1 van T. Uitgaande van een wat andere r_2 komt men tot T_2, enz. Grafisch vindt men dan de r^* (in het interval 0 tot en met 1), waarvoor T minimaal is (T_{min}) (dus een approximatie). Dit alles gaat slechts vlot als bedoelde recurrente formules zeer simpel zijn.

Bijzonder geval F = constante a

Als men F = a substitueert in (8), leidt (9) tot

$$a_n = \frac{1-r}{1-r^n} \sum_1^n r^{n-i} f(i) \quad ^{1)}$$

Het Craddock-procédé volgende wordt dus met deze $F_n(i, r, n) \equiv a_n$, berekend met behulp van de eerste n elementen, een beste schatting van $f(n+1)$ gemaakt en wel a_n zelf (want de i-variabele ontbreekt, immers F = constant). Slechts als $r^n \ll 1$, (afhankelijk én van r én van n) staat er $a_n \approx (1-r) \sum_1^n r^{n-i} f(i)$. Verder wordt $\mathcal{E}(n+1) = f(n+1) - a_n$.

Daarna bepalen wij a_{n+1} uit (12). Er komt

$$a_{n+1} = \frac{1-r}{1-r^{n+1}} \sum_1^{n+1} r^{n+1-i} f(i) = \frac{(1+r+ \ldots\ldots r^{n-1})\, r\, a_n + f(n+1)}{1 + r\left[1 + r + \ldots r^{n-1}\right]} =$$

$$\frac{(1-r)\, f(n+1) + r(1-r^n)\, a_n}{1-r^{n+1}} = a_n + \frac{1-r}{1-r^{n+1}} \left[f(n+1) - a_n\right]$$

Wij vinden dus zeer simpele iteraties:

---

1) Voor $r \to 1$ gaat $a_n \to \frac{1}{n} \sum_1^n f(i)$ = rekenkundig gemiddelde.

$$a_n = \frac{1-r}{1-r^n} \sum_1^n r^{n-i} \quad f(i) \qquad \text{en} \qquad \mathcal{E}(n+1) = f(n+1) - a_n$$

$$a_{n+1} = a_n + \frac{1-r}{1-r^{n+1}} \left[f(n+1) - a_n \right] \quad \text{en } \mathcal{E}(n+2) = f(n+2) - a_{n+1}$$

$$a_{n+2} = a_{n+1} + \frac{1-r}{1-r^{n+2}} \left[f(n+2) - a_{n+1} \right] \text{en } \mathcal{E}(n+3) = f(n+3) - a_{n+2}$$

enz.

tenslotte

$$a_N = a_{N-1} + \frac{1-r}{1-r^N} \left[f(N) - a_{N-1} \right]$$

Alleen als óók nog $r^n \ll 1$, geldt in benadering

$$a_{n+k} = a_{n+k-1} + (1-r) \left[f(n+k) - a_{n+k-1} \right] \quad , \text{ met } a_n = (1-r) \sum_1^n r^{n-i} f(i)$$

$k = 1, 2, \ldots\ldots N-n$

Deze benadering is met toenemende k beter. Het is echter helemaal de vraag vanaf welke n ze geoorloofd is, omdat zulks met de waarde van r samenhangt, van welke waarde wij a priori niets weten.

3 Opmerkingen bij de tekst van de brief van Mr Craddock

/De nummers der volgende punten corresponderen met de getallen, die wij vóór Craddock's tekst plaatsten/

1 "We suppose that the series, ... , contains an element which can be ascribed to the working of a definite physical process and which can be expressed in mathematical terms and also an element which cannot be so expressed and is effectively random".

Wij dienen wel te bedenken hoe betrèkkelijk de betekenis van zulk een onderstelling is:"Wat is toeval"? Wij schrijven iets aan het "toeval" toe zolang er sprake is van een veelheid van oorzaken, zó groot en ingewikkeld (en mede daardoor onontwarbaar), dat zich de gevolgen gedragen volgens de statistische toevalswetten (beide delen, het physische en het random-deel, hebben derhalve een physische oorsprong). Een "zeer bepaald physisch proces"? Heeft de auteur zich in elk voorkomend geval zulk een scherp beeld van de physische achtergrond gevormd, dat de keuze van het type van F(t) vast ligt? Hoe komt hij tot de keuze van F(t)? Is F(t) eigenlijk wel éénduidig bepaald? Kan men criteria ontwikkelen, die, anders dan op physische gronden, helpen bij de keuze van het type van F(t)? (zie ook onder 4.5)

2 en 3 "even if we can identify and describe the physical proces operating ....." Let wel: identificeren (onderkennen) én beschrijven ..... Is de situatie niet meestal zó, dat men meer beschrijft dan werkelijk physisch verklaart?

4 "F(t) always has the same form, but with parameters varying from time to time".
Hier begrijpen wij Craddock niet. Zij $F(t)=a+bt+ct^2$. Bedoelt Craddock dan, dat a, b en c elk nog variëren met de tijd of bedoelt hij, dat a, b, c als ze bepaald zijn voor $n = 8$, anders zijn dan voor $n+1 = 9$ of $n+2 = 10$, etc.? Wij vermoeden dat Craddock het laatste bedoelt. De tekst van de brief is niet duidelijk.

5 en 6 Hier zouden wij het anders zeggen. Craddock bedoelt niet, dat hij één term, t.w. $f_{n+1} - F(n+1)$, minimaliseert, maar een som van gekwadrateerde residuen.

7 Een aanpassing m.b.v. de "methode der kleinste kwadraten" doet onmiddellijk de vraag stellen: "waarom met deze methode?". Wij doen hèt zó dikwijls op deze wijze, dat wij niet (meer) weten waarom. Hierop zouden wij willen terugkomen in de paragraaf "statistische achtergrond".

8 Deze opmerking lokt ook verdere statistische beschouwingen uit. Wanneer hechten wij gewichten aan de waarnemingen? Waarom en hoe? Wat heeft dit met "aanpassing" te maken? Zie weer onder "statistische achtergrond".

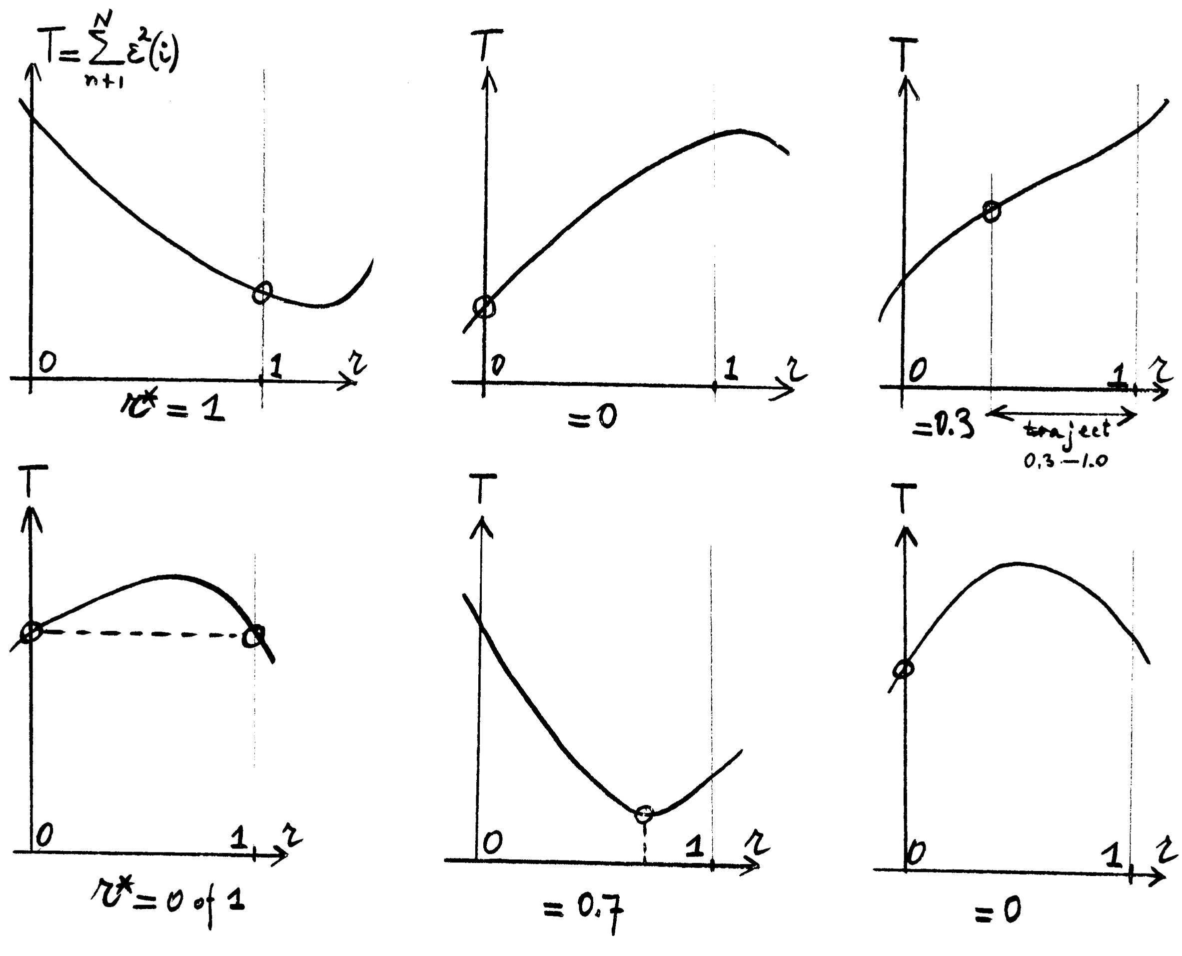
9 Is de bovengrens N juist? Hij moet zijn o.i. n, immers beschouwd worden eerst de eerste n termen; geprobeerd wordt de aanpassing $F_n(-i)$.

10 De bovengrens ∞ moet n zijn. Het enige verschil tussen S in (9) en S in (10) is gelegen in de gewichtsfactor r^i bij f_{n-i}. $(0 \leq r \leq 1)$.

11 en 12 Reeds in hoofdstuk 2 behandeld.

13 en 14 Liever schrijven $\sum_{n+1}^{N} e_i^2$. Deze uitdrukking is een ingewikkelde functie van de onbekende r. Men moet haar minimaal maken, dus men moet r oplossen uit de vergelijking $\frac{d}{dr}\sum_{n+1}^{N} e_i^2 = 0$. [Deze $\sum$ is alleen van r een functie; eigenlijk ook nog van n]. Kunnen wij bewijzen, dat deze vergelijking een wortel heeft in het traject 0 → 1? Wanneer de wortels van de vergelijking buiten het traject 0-1 liggen (bijv. negatief zijn ; maar er kunnen ook wel uitsluitend complexe wortels zijn) bedoelt C. dan, dat men onderzoekt voor welke r uit het traject 0 → 1 (of wil hij liever $\frac{1}{2}$ → 1) de uitdrukking $\sum_{n+1}^{N} e_i^2$ minimaal is? Is dat niet steeds de benedengrens of bovengrens van het traject?

Zie de volgende schetsjes.



15 "which enanbles the <u>best</u> prediction to be made". Hoezo? Vanwaar "the best"? Men heeft eigenlijk alleen die r genomen, waarvoor $\sum_{n+1}^{N} e_i^2$ - tewerkgaande volgens het Craddock-procédé - minimaal is. D.w.z. voor alle andere r-waarden (ook r=1; als r=1 past men eigenlijk de "gewone methode van het rekenkundig gemiddelde" toe) is deze $\sum$ groter. Craddock behoort hier het begrip "the best" te definiëren.

16 "also the mean square error of forecast". Hier blijkt dus de bedoeling van het berekenen van de minimale $\sum$. Maar ik zou de voorkeur geven aan $\left\{\sum_1^N e_i^2\right\}$: (N-g-1) boven $\left\{\sum_{n+1}^N e_i^2\right\}$: (n-1-g-1). Hierin is g de graad van F(t) (g=0 als F=const.). In het algemeen : g+1 = aantal te bepalen constanten in F(t). Deze g=graad van een gehele rationele functie $F = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \ldots\ldots \alpha_{g+1}\ t^g$

17, 18 "in some cases at any rate the calculation reduces to a series of straight-
en 19 forward iterations".
Heeft Craddock de drie gevallen F = a, F = a+bt F = a+bt+ct^2 en F = a+sin bt, soms ook a+b sin (ct+e)? geheel doorgerekend? Ontstonden er steeds simpele iteraties? Zo ja, welke?

20 Hier is reeds benaderd. Zie onze exacte afleiding. De moeilijkheid is, dat men - r nog niet kennende - niet tevoren weet of $1-r^n \approx 1$ is. Zo is $r^n < 0.1$ voor r = 0.60 bij $n \geqslant 5$; voor r = 0.70 bij $n \geqslant 7$; voor r = 0.80 bij $n \geqslant 11$; voor r = 0.90 bij $n \geqslant 22$; voor r = 0.95 bij $n \geqslant 45$ en voor r = 0.99 bij $n \geqslant 230$, enz.
In elk geval moet het teken ∞ in (20) door n vervangen worden.

21 t/m 24 Reeds behandeld in onze tekst.

25 Heeft Craddock dus voor Greenwich n = 10 gebruikt?

26 Waarom $\frac{1}{N} \sum e_n^2$? Moet het niet zijn: $\frac{1}{N-n-1} \sum_{n+1}^{N} e_i^2$, n.l. over de termen n+1 t/m N. Deze variantie zouden wij σ_v^2 willen noemen.
Men vraagt zich af: wat wordt nu tenslotte $\sigma_\alpha^2 + \sigma_v^2 = \frac{1}{N-1} \sum_1^N e_i^2$ (d.w.z. over alle N termen) met deze r = 0.865? Wat is $F_N \equiv a_N$?
Blijkbaar beweert Craddock, dat $\sigma_v^2 = 1.128$ ($\sigma_v = 1.06$) en wat is σ_{a+v}^2 ?
(Zie onder 2, (18),(19), (20))

27 Over de statistische achtergrond, zie bijgaande beschouwing.

28 "physically reasonable".
Hoe zo? Waar is de physica te hulp geroepen? Bij de gewichten r^i?
Craddock zegt alleen, dat wij kunnen betogen, dat het van belang is voor de gezochte F(t) een nauwere aanpassing aan de meer recente termen te

bewerkstelligen dan aan de vroegere. Hij vertelt niet op welke gronden. Physische gronden?

29 Kan Craddock voorbeelden noemen, waarin een "temporary harmonic oscillation" veroorzaakt werd door "random impulses" en een gewone harmonische analyse gemakkelijk tot "negative or misleading results" leiden zou?

4 Statistische achtergrond

4.1 Methode der kleinste kwadraten (zie punt 7 in C.'s brief)

Zij $y = f(x)+\varepsilon$, d.w.z. bij x_1 behoort één y_1, zijnde $f(x_1)$, plus een random-variabele, die aan een normale verdeling ($\mu = 0$; σ) gehoorzaamt, die niet met x samenhangt. De waarschijnlijkheidsverdeling van ε zij:

$$\varphi(\varepsilon)\,d\varepsilon = \frac{1}{\sigma\sqrt{2\pi}}\,\exp.(-\varepsilon^2/2\sigma^2)\,d\varepsilon$$

We stellen de vraag: Hoe komt men op basis van de waarnemingen $x_1, y_1; x_2, y_2 \ldots\ldots\ldots\ldots x_n, y_n$ tot de "beste" schatting van de onbekende parameters $\alpha_1, \alpha_2 \ldots\ldots \alpha_p$ in f(x)? Fisher redeneert: de waarden van de toevallige component volgen uit $\varepsilon_i = y_i - f(x_i)$; $i = 1,2 \ldots\ldots n$ (y_i gemeten en $f(x_i)$ berekend). De samengestelde kans K op het voorkomen van de waarde-intervallen $\varepsilon_1 \to \varepsilon_1 + d\varepsilon_1$; $\varepsilon_2 \to \varepsilon_2 + d\varepsilon_2 \ldots\ldots \varepsilon_n \to \varepsilon_n + d\varepsilon_n$ is $K = \varphi(\varepsilon_1)\,\varphi(\varepsilon_2) \ldots\ldots \varphi(\varepsilon_n)\,d\varepsilon_1 \ldots\ldots\ldots d\varepsilon_n$, d.i.

$$K = (2\pi\sigma^2)^{-\frac{1}{2}n}\exp.\left\{-\sum_1^n \varepsilon_i^2/2\sigma^2\right\} d\varepsilon_1 \ldots\ldots d\varepsilon_n$$

Als het type van f(x) gekozen is, moet men aan $\alpha_1, \alpha_2 \ldots\ldots$ zodanige waarden toekennen (functies van de waarnemingen), dat K maximaal is. Voor deze waarden moet dus $\sum_1^n \varepsilon_i^2$ minimaal zijn. Deze methode heet "method of maximum likelihood". Zij leidt tot schattingen van $\alpha_1, \alpha_2 \ldots\ldots$ met een middelbare fout, kleiner dan voor andere schattingen. Men moet $\alpha_1, \alpha_2 \ldots\ldots$ oplossen uit de vergelijkingen $\partial\Sigma/\partial\alpha_1 = 0$; $\partial\Sigma/\partial\alpha_2 = 0$ etc.

Als bijv. $f(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \ldots\ldots$ leidt de vergelijking

$$0 = \frac{\partial\Sigma}{\partial\alpha_1} \quad \text{tot} \quad \sum_1^n \left\{y_i - f(x_i)\right\} = 0.$$

Voor de beste f(x) is dus én $\sum\left\{y_i - f(x_i)\right\} = 0$

$$\text{én} \quad \sum\left\{y_i - f(x_i)\right\}^2 = \text{minimaal.}$$

(hetzelfde geldt voor transcendente functies).

Door de gegeven constellatie van n punten x_i, y_i is echter f(x) niet éénduidig bepaald. Wel is er één beste rechte $y = a+bx$; één beste parabool $y = a+bx+cx^2$; één beste kubische curve $y = a+bx+cx^2+dx^3$, etc. De waarde van $\sum[y_i-f(x_i)]^2$ zal afnemen met toenemende graad g van f(x), om nul te worden voor g = n. [Zie naar de waarden 7.429; 5.152 en 4.719 voor $\sum\varepsilon^2$ in tabel 4, in 4.7]. In het laatste geval namelijk kan men één parabolische kromme (gehele rationele functie) van de graad n leggen door n+1 gegeven punten.

4.2 Variantie van de voorspelde waarde

Men bedenke, dat - uitgaande van een functie F(t) met parameters α_1, α_2 ..... - de steekproef (een tijdreeks is ook een steekproef uit een universum) voor deze parameters a, b, c, ... levert. Andere steekproeven geven andere waarden. Deze a's zijn daardoor stochastische grootheden, welke aan verdelingen gehoorzamen, die meestal zeer moeilijk te overzien zijn.

Een zeer veel voorkomende fout bestaat hierin, dat de standaarddeviatie, die wij in onze tekst σ_v noemden, berekend uit de steekproef, geïnterpreteerd wordt als maat voor de standaarddeviatie van een "forecast". Deze interpretatie is alleen dan juist als de steekproef (de tijdreeks) zeer groot (lang) is. Bij kleine steekproeven moet men ook met de steekproef-standaarddeviatie van de regressie-functie zelf rekenen (d.w.z. de standaarddeviaties in de verdelingen, waaraan de a's gehoorzamen).

De variantie van een voorspelde waarde bestaat dus uit twee delen: de variantie van de regressiefunctie (in ons geval de F(t) en de variantie rondom deze regressie.

Wanneer F de vorm $\alpha_1 + \alpha_2 t$ heeft, kunnen beide variantie-delen nog wel berekend worden. Hoe hoger echter de graad van F, hoe moeilijker. Hoe kan in het bijzonder de berekening bij het typische procédé van Craddock worden uitgevoerd? In 4.6 zal blijken dat de betrouwbaarheid van een voorspelling, verricht volgens het C.-procédé, moeilijk is te berekenen.

4.3 Aanpassing en gewogen gemiddelde (punt 8 van Craddock's brief)

Craddock beweert het volgende:

Gegeven zij een tijdreeks f_1 f_2 ..... f_N (de laatste term in het heden); de termen stammen uit een gestandaardiseerde normaal verdeeld universum ($\mu=0; \sigma=1$). Er is geen persistentie[1].

---

[1] Ook voor niet gestandaardiseerde normale verdelingen geldt al het volgende. Door te standaardiseren ($\mu=0; \sigma=1$) worden de berekeningen eenvoudiger.

Normaliter zou men het gewone rekenkundige gemiddelde $\bar{f} = \frac{1}{N}\sum_{1}^{N} f_i$ aanpassen, Als men echter een betere aanpassing wenst aan de meer recente dan aan de lang geleden termen, dan doet men beter gewichten g_1, g_2 ..... g_N aan resp. f_1, f_2 ..... f_N toe te kennen, die afnemen naar het verleden: $g_1 < g_2 < \ldots\ldots < g_N = 1$. Vervolgens werke men met het gewogen gemiddelde

$$\bar{f}_g = \left\{\sum_{1}^{N} g_i f_i\right\} : \left\{\sum_{1}^{N} g_i\right\}$$

Craddock levert geen bewijs. Hier volgt ons bewijs:
Op het eerste gezicht lijkt het toekennen van gewichten alleen iets te maken te hebben met "gevoeligheid", "invloedssterkte". Wij bedoelen:

$$\partial\bar{f}_g / \partial f_j = g_j \Big/ \sum_{1}^{N} g_i \quad \text{en} \quad \partial\bar{f}_g / \partial f_k = g_k \Big/ \sum_{1}^{N} g_i \text{ , zodat}$$

"$\bar{f}_g$ gevoeliger is voor variaties in f_j dan voor variaties in f_k" als $g_j > g_k$. Als dus $g_1 < g_2 < \ldots\ldots < g_N$ dan is $\bar{f}_g$ gevoeliger voor variaties in f_s naarmate s toeneemt (de term f_s meer recent is).

Vraag: hoe komt het dat deze zaak tevens te maken heeft met "aanpassing"?
Wij moeten eerst definiëren:

Def.: de "aanpassing" A van een getal a aan de k termen

f_{p+1} , f_{p+2} ..... f_{p+k} wordt gegeven door $\frac{1}{k-1}\sum_{1}^{k} (f_{p+i} - a)^2$

Is A kleiner dan heet de aanpassing beter.

Vraag: kunnen wij bewijzen, dat A_L (voor de termen f_1 f_2 ..... f_n) groter is dan A_R (voor de termen $f_{N-(n-1)}$ ..... f_N), wanneer we $\bar{f}_g$ aanpassen?

Het bewijs

$$A_L = \sum_{1}^{n} (f_i - \bar{f}_g)^2 = \sum_{1}^{n} \left[(f_i - m) + (m - \bar{f}_g)\right]^2 = \sum_{1}^{n} f_i^2 - \frac{1}{n}\left\{\sum_{1}^{n} f_i\right\}^2 - n(m - \bar{f}_g)^2$$

als $m = \frac{1}{n}\sum_{1}^{n} f_i$

Uitwerken: $$m - \bar{f}_g = \frac{f_1 + \ldots\ldots f_n}{n} - \frac{g_1 f_1 + \ldots\ldots g_N f_N}{g_1 + \ldots g_N} =$$

$$\frac{(S - ng_1) f_1 + \ldots (S - ng_n) f_n - ng_{n+1} f_{n+1} - \ldots\ldots - ng_N f_N}{nS}$$

als $S = g_1 + g_2 + \ldots\ldots g_N$

Dus

$$\frac{(m-\bar{f}_g)^2 = S^2(f_1^2 + \dots f_n^2) - 2nS(g_1 f_1^2 + \dots g_n f_n^2) + n^2 \sum_1^N g_i^2 f_i^2 + \text{termen } f_i f_j}{n^2 S^2}$$

Dus

$$\mathcal{E}(m-\bar{f}_g)^2 = \frac{nS^2 - 2nSs_1 + n^2 \sum_1^N g_i^2}{n^2 S^2}, \quad \mathcal{E} f^2 = \frac{S^2 - 2Ss_1 + n \sum_1^N g_i^2}{nS^2} \mathcal{E} f^2$$

Met $s_1 = g_1 + g_2 + \dots\dots g_n$ en $s_2 = g_{N-\overline{n-1}} + \dots\dots\dots g_N$

N.B. $\mathcal{E} f_i^2 = \mathcal{E} f_i^2$, iedere i en j.

$\mathcal{E} f_i f_{i_{i \neq j}} = \mathcal{E} f_i \mathcal{E} f_j = 0$, omdat persistentie tussen de f's afwezig geacht wordt. $\mathcal{E} f^2 = \sigma^2 + \mu^2 = 1$ Dus:

$$\boxed{\begin{aligned} \mathcal{E} A_L &= n-1 + \frac{S^2 - 2Ss_1 + n \sum_1^N g_i^2}{S^2} \\ &\text{en} \\ \mathcal{E} A_R &= n-1 + \frac{S^2 - 2Ss_2 + n \sum_1^N g_i^2}{S^2} \end{aligned}}$$

Aangezien $g_1 < g_2 < g_3 \dots\dots < g_N = 1$ is $s_1 < s_2$ en $\mathcal{E} A_L > \mathcal{E} A_R$

m.a.w. de aanpassing aan de langer geleden termen is slechter dan aan de meer recente. Q.E.D.

N.B. De stelling geldt alleen in het gemiddelde (we namen de verwachtingswaarden $\mathcal{E}$).

Contrôle op juistheid van de formules.

1) Neem eens geen gewichten: $g_1 = g_2 = \dots\dots g_N = 1$

Dan is $S=N$; $s_1 = s_2 = n$ en $\sum_1^N g_i^2 = N$

$$\mathcal{E}_{A_L} = \mathcal{E}_{A_R} = n-1 + \frac{N-n}{N} \qquad \text{en}$$

$$\mathcal{E}\frac{A_L}{n-1} = \mathcal{E}\frac{A_R}{n-1} = 1 + \underbrace{\frac{N-n}{N(n-1)}}_{>1}, \text{ als } n < N \qquad [*]$$

Onder woorden:

Als men het algeheel gewoon rekenkundig gemiddelde (over alle N termen) aanpast dan is de verwachtingswaarde van de aanpassing over elk n-tal termen ($n < N$) groter dan over alle N termen; d.w.z. de aanpassing is slechter en wel slechter met afnemende n. Het laatste omdat: als n loopt van 2 tot en met N, verandert $[*]$ van $1 + \frac{N-2}{N} = 2 - \frac{2}{N} \approx 2$ tot 1 (ongeveer een factor 2 als $N \gg 1$)

2) Hoe verhouden zich de aanpassingen (resp. A_1 en A_2) bij gebruik van het gewogen en van het gewone gemiddelde over dezelfde N termen?

Vul n = N in

$$\mathcal{E}\frac{A-1}{N-1} = \left[1 + \frac{S^2 - 2S^2 + N\sum_1^N g^2}{S^2}\right] = \left[1 + \frac{N\sum_1^N g_i^2 - S^2}{S^2}\right]$$

en zonder gewichten (alle g's = 1)

$$\mathcal{E}\frac{A_2}{N-1} = 1 \text{ (in het algemeen de } \sigma^2\text{)}$$

Altijd is $\mathcal{E}\frac{A_1}{N-1} > \mathcal{E}\frac{A_2}{N-1}$ omdat $N\sum_1^N g_i^2 - S^2 > 0$ en wel omdat

de stelling geldt

$$N\,(a_1^2 + a_2^2 + a_2^2 \ldots a_N^2) > (a_1 + a_2 + a_3 \ldots a_N)^2$$

voor alle $a_1, a_2, a_3 \ldots\ldots a_N$

Conclusie:

<u>Betere aanpassing</u> m.b.v. <u>gewoon rekenkundig gemiddelde</u> dan met <u>een of ander gewogen gemiddelde</u>.

4.4 Hebben de gewichten $g_i = r^{n-i}$ een physische of louter statistische betekenis?

Craddock voert de gewichten in i.v.m. kwesties van aanpassing, maar hebben zij ook enige physische betekenis en zo ja, welke? Waarom afnemend naar het verleden toe volgens een meetkundige reeks? waarom niet bijv. 1, 1-x, 1-2x, 1-3x ...... (tot het kleinste positieve getal en dan verder gelijkblijvend) of nog anders?
Heeft de r misschien met persistentie te maken? En moeten wij in deze persistentie "physica" zoeken? Beschouwt C. alleen persistentie-vrije tijdreeksen of ook tijdreeksen mèt persistentie en, zo het laatste het geval is, heeft de persistentie-coëfficiënt dan iets te maken met zijn r? Het is ons niet duidelijk. Bovendien: de persistentiecoëfficiënt kan men beter op een meer directe wijze berekenen als de auto-correlatiecoëfficiënt tussen de paren f_i, f_{i+1} (i = 1, 2, 3 ....,)of f_i, f_{i+2}, enz.
Daar komt bij: als men F = constant probeert mag men niet tevens persistentie onderstellen, want met persistentie moet er een trend zijn (een sinusoïde - bijv. jaarlijkse gang - daaronder ook verstaan) en zal het proberen van F = constant zonder zin zijn. F = constant beduidt nl., dat het element van de tijdreeks op onverschillig welk tijdstip een normale verdeling volgt rondom zeker gemiddelde μ en met zekere standaarddeviatie σ (μ én σ tijd-onafhankelijk). In zulk een geval is de frequentieverdeling van f(i) volmaakt onafhankelijk van die van f(j); $i \neq j$ (persistentie 0). "Volmaakte" persistentie zou een functioneel en niet een stochastisch verband betekenen tussen successieve elementen. Maar dan is er geen random-component én geen probleem meer.
Of stelt het gewicht een "graad van belangrijkheid" ("gevoeligheidsgraad") voor, zoals ik bedoelde onder 4.3, zie punt (5)?

4.5 Is f(t) exact bepaald door "differentie-methode"?

Welk type F(t) is het beste? Welke graad? Men zegt wel, dat de z.g. differentie-methode (het bepalen van opeenvolgende verschillen) deze graad exact bepaalt. Dit lijkt ons niet exact waar te zijn. Wij willen een en ander aan een voorbeeld toelichten. Zij bekend, dat x een functie is van de tijd van de volgende vorm: $x = a+bt+ct^2+\varepsilon$; ε = random-variabele (norm. verdeling: $\mu = 0$ en $\sigma \neq 0$; deze σ is onafhankelijk van de x).
De parameters a, b en c zijn onbekend. Men beschikt alleen over metingen van x voor bijv. t=0, 1, 2, ... 9. Zodat $x(0)=a+\varepsilon_0$; $x(1)=a+b+c+\varepsilon_1$; $x(2)=a+2b+4c+\varepsilon_2$; $x(3)=a+3b+9c+\varepsilon_3$, enz. Verder de eerste differentie $\Delta_1(t)=x(t+1)-x(t)$. De tweede differentie $\Delta_2(t)=\Delta_1(t+1)-\Delta_1(t)$; de derde $\Delta_3(t)=\Delta_2(t+1)-\Delta_2(t)$ enz.
Zie tabel 1.

Tabel 1

$x_t = a + bt + ct^2 + \mathcal{E}$; $\mathcal{E}$ = random var., gehoorzamend aan norm. verd. $(\mu = 0; \sigma)$

	0	1	2	3	4
t	$x(t) = \Delta_0$	$\Delta_1(t)$	$\Delta_2(t)$	$\Delta_3(t)$	$\Delta_4(t)$
0	$a + \varepsilon_0$	$b + c + \varepsilon_1 - \varepsilon_0$	$2c + \varepsilon_2 - 2\varepsilon_1 + \varepsilon_0$	$\varepsilon_3 - 3\varepsilon_2 + 3\varepsilon_1 - \varepsilon_0$	$\varepsilon_4 - 4\varepsilon_3 + 6\varepsilon_2 - 4\varepsilon_1 + \varepsilon_0$
1	$a + b + c + \varepsilon_1$	$b + 3c + \varepsilon_2 - \varepsilon_1$	$2c + \varepsilon_3 - 2\varepsilon_2 + \varepsilon_1$	$\varepsilon_4 - 3\varepsilon_3 + 3\varepsilon_2 - \varepsilon_1$	$\varepsilon_5 - 4\varepsilon_4 + 6\varepsilon_3 - 4\varepsilon_2 + \varepsilon_1$
2	$a + 2b + 4c + \varepsilon_2$	$b + 5c + \varepsilon_3 - \varepsilon_2$	$2c + \varepsilon_4 - 2\varepsilon_3 + \varepsilon_2$	$\varepsilon_5 - 3\varepsilon_4 + 3\varepsilon_3 - \varepsilon_2$	$\varepsilon_6 - 4\varepsilon_5 + 6\varepsilon_4 - 4\varepsilon_3 + \varepsilon_2$
3	$a + 3b + 9c + \varepsilon_3$	$b + 7c + \varepsilon_4 - \varepsilon_3$	$2c + \varepsilon_5 - 2\varepsilon_4 + \varepsilon_3$	$\varepsilon_6 - 3\varepsilon_5 + 3\varepsilon_4 - \varepsilon_3$	$\varepsilon_7 - 4\varepsilon_6 + 6\varepsilon_5 - 4\varepsilon_4 + \varepsilon_3$
4	$a + 4b + 16c + \varepsilon_4$	$b + 9c + \varepsilon_5 - \varepsilon_4$	$2c + \varepsilon_6 - 2\varepsilon_5 + \varepsilon_4$	$\varepsilon_7 - 3\varepsilon_6 + 3\varepsilon_5 - \varepsilon_4$	zonder trend
5	$a + 5b + 25c + \varepsilon_5$	$b + 11c + \varepsilon_6 - \varepsilon_5$	$2c + \varepsilon_7 - 2\varepsilon_6 + \varepsilon_5$	zonder trend	
6	$a + 6b + 36c + \varepsilon_6$	$b + 13c + \varepsilon_7 - \varepsilon_6$	zonder trend		
7	$a + 7b + 49c + \varepsilon_7$	met lineaire trend			
	met kwadratische trend				
		$\bar{\Delta}$ = gemiddelde = want	2c elke $\bar{\varepsilon}_i = 0$	0 idem	0 idem
	de frequentieverdeling is		normaal	normaal	normaal
	st.deviatie		$\sigma\sqrt{4}$	$\sigma\sqrt{8}$	$\sigma\sqrt{16}$
	gemiddelde		2c	0	0

Hoe handelt men met dit procédé in de praktijk? Men berekent uit de gegeven waarden de opeenvolgende Δ's, en wel tot de eerste kolom $\Delta_g(t)$ waarin de Δ's geen trend meer lijken te bezitten (als x_t van de graad 2 is in t, is dit in kolom 2). Men behoort eigenlijk te toetsen of er al of niet een trend is, hetgeen natuurlijk moeilijker is naarmate er minder x-waarden waren. Wanneer men met de toets tot een overschrijdingskans P ver boven 5% komt, behoeft men het Δ-procédé niet voort te zetten en kan men voor x_t een functie van de graad g kiezen. Men bedenke, dat, hoe hoger het nummer van de kolom is, hoe groter de spreiding in de Δ's van deze kolom is. Zo is die in kolom $\Delta_2(t)$ (zie boven), aangezien elke ε_i onafhankelijk van elke ε_j verdeeld is,

$$\sqrt{\sigma^2(\varepsilon_i) + 2\sigma^2(\varepsilon_{i-1}) + \sigma^2(\varepsilon_{i-2})} = \sigma\sqrt{(1+2+1)} = \sigma\sqrt{4}; \text{ aldus in kolom } \Delta_3(t):$$

$$\sqrt{(1+3+3+1)\,\sigma^2} = \sigma\sqrt{8}; \text{ in kolom } \Delta_t^{(4)}: \sqrt{(1+4+6+4+1)\sigma^2} = \sigma\sqrt{16}, \text{ etc.}$$

M.a.w. de getallen in een meer rechts gelegen kolom hebben een grotere gemiddelde absolute waarde.

Hiermede is ook duidelijk geworden, dat volgens de Δ -methode de graad van f(t) niet exact vast ligt, want er rijst een nieuw probleem, n.l. dat van de toetsing van de trend. Bij zulke toetsingen behoren weer waarschijnlijkheidsuitspraken, d.w.z. nieuwe onzekerheden.

Wij herhalen: nadat de eerste kolom $\Delta_g(t)$ gevonden is, waarin een trend afwezig is, is de graad van f(t) bekend: g. Wat is f(t) nu zelf? Men kan op twee wijzen verder gaan:

1) men berekene in deze kolom de $\overline{\Delta}_g$ en de st. dev. dezer Δ 's: σ_g . Dan is ook de σ van de random-component van x bekend. Immers $\sigma_2 = \sqrt{4}$; $\sigma_3 = \sigma\sqrt{8}$; $\sigma_g = \sigma\sqrt{2^g}$; via deze $\overline{\Delta}_g$ terugwerkende naar de kolommen g - 1, g - 2 ..... 0 vindt men tenslotte

$$x_0, x_1, x_2, \ldots\ldots, \text{ d.w.z. } x_t = a_\Delta + b_\Delta \cdot t + c_\Delta \cdot t^2 + \ldots\ldots \text{ (graad g)} \quad (1)$$

Toelichting: in genoemd voorbeeld (onderstel nu, dat de graad niet bekend is, maar wel dat x van t afhangt volgens $x = a + bt + ct^2 + dt^3 + \ldots\ldots$) zullen wij wellicht vinden, dat in kolom Δ_2 geen trend meer is.

Men beschouwe nu $2c = \overline{\Delta}_2$ (dus onderstellende $\varepsilon_0 - \varepsilon_1 - \varepsilon_6 + \varepsilon_7 = 0$). Aldus ligt c vast. Daarna $b + \frac{1}{7} c\ (1 + 3 + \ldots\ldots 13) = \overline{\Delta}_1$; aldus is b bekend. Daarna $a + \frac{1}{7} b\ (1 + 2 + \ldots\ldots 7) + \frac{1}{7}(1^2 + 2^2 + \ldots\ldots 7^2)\ c = \overline{\Delta}_0$; aldus is a bekend.

2) men kan ook met behulp van de methode der kleinste kwadraten de beste waarden van de parameters in een g-de graadsfunctie berekenen en komt dan tot een ander stel a_k; b_k; c_k ..... (g + 1 stuks). (2)

Wij weten echter al bij voorbaat, dat de st. dev. van de afwijkingen tussen de gemeten x_t en de volgens (1) berekende x-waarden groter is dan die tussen de gemeten x_t en de volgens (2) berekende x-waarden (omdat juist in de methode der kleinste kwadraten de som der gekwadrateerde residuen minimaal gemaakt wordt)

4.6 Is het Craddock-procédé beter dan het Gladstrijkingsprocédé?

Gegeven: een normaal verdeeld universum (μ ; σ), waaruit men talloos vele keren een aselecte steekproef neemt van N elementen, die men steeds als een tijdreeks $f_1, f_2 \ldots\ldots f_N$ opschrijft. Men past op elk van de tijdreeksen 1° het Craddock-procédé toe, 2° het Gladstrijkings-procédé toe.

Gevraagd: Aan welke verdelingen gehoorzamen de volgens het C.-procédé berekende en de volgens het G-procédé berekende a_N? (als de verdelingen normaal mochten zijn vragen wij natuurlijk naar de μ_c; σ_c resp. μ_g en σ_g).

Toelichting:

1) Voor de betekenis van a_N, zie onder "bijzonder geval" in 2. Dat wij bij onze vraag met dit bijzondere geval te doen hebben, lijkt ons duidelijk, immers het de i^{de} keer getrokken element volgt een normale waarschijnlijkheidsverdeling, rondom μ en met st. dev. σ (onafhankelijk van i).

2) Met het gladstrijkingsprocédé wordt dit bedoeld: men neemt het rekenkundig gemiddelde van de eerste n termen ($n < N$); daarna over de termen 2 t/m n+1; daarna over de termen 3 t/m n+2, etc., de gehele reeks door. Vervolgens worden de N - n+1 aldus berekende gemiddelden weer gemiddeld tot: a_N.

Oplossing:

4.6.1 Het Craddock-procédé (C-methode):

$\mathcal{E}a_N$ = ? In elk geval is $a_N = k_1f_1 + k_2f_2 + \dots\dots k_Nf_N$, waarin k_i een functie van de r (zie het Craddock-procédé). De numerieke waarde van r wordt bepaald door te eisen, dat $T = \sum_{n+1}^{N}(f_i - a_i)^2$ een minimum is.

Natuurlijk is $\mathcal{E}a_N = k_1\mathcal{E}f_1 + k_2\mathcal{E}f_2 + \dots\dots k_N\mathcal{E}f_N = \mu\sum_1^N k_i$, omdat elke $\mathcal{E}f_i = \mu$.

Hoe vinden we $\sum_1^N k_i$?

Verwijzende weer naar 2 zien wij, dat $a_n = \frac{1 - r}{1 - r^n}\sum_1^n r^{n-i} f(i)$, zodat

$k_i = \frac{1 - r}{1 - r^N} r^{N-i}$ en dus $\sum_1^N k_i = 1$ en dus $\mathcal{E}(a_N) = \mu$.

Omdat elke f_i een normale verdeling (μ ; σ) volgt, volgt ook a_N, zijnde een lineaire combinatie der f's, een normale verdeling (met gemiddelde μ, maar met welke σ_c ?). Allereerst deze opmerking: de r wordt geval voor geval berekend; soms is hij een (toevallig tussen 0 en 1 gelegen) wortel van de hogere graads-vergelijking in r, die het resultaat is van bedoelde minimum-conditie T; soms ook is hij òf 0 òf 1, waarvoor T een relatief minimum heeft. De r is dus een stochastische grootheid met een onbekende waarschijnlijkheidsverdeling $\varphi(r)\,dr$ (traject $0 \leqslant r \leqslant 1$) Wij kunnen deze $\varphi(r)$ niet berekenen. Welke is de topwaarde (de meest waarschijnlijke r)? Dicht bij 0; dicht bij 1? Als $r = 0$ is $a_n = f_n$; $a_{n+1} = f_{n+1}$; .... $a_N = f_N$ (zie de formules voor a_r) en $\sigma_c = \sigma$. Als $r = 1$ is

$$a_n = \frac{1}{n}\sum_1^n f_i \;;\; a_{n+1} = \frac{1}{n+1}\sum_1^{n+1} f_i \;\ldots\ldots\; a_N = \frac{1}{N}\sum_1^N f_i \text{ en } \sigma_c = \sigma/\sqrt{N}$$

Wij kunnen van deze σ_c niet meer zeggen, dan dat $\sigma \geqslant \sigma_c \geqslant \sigma/\sqrt{N}$

σ_c zelf is ook een stochastische variabele.

N.B. Mocht aangetoond kunnen worden, dat de meest waarschijnlijke r dicht bij 1 ligt, dan ligt ook de meest waarschijnlijke σ_c dicht bij $\sigma/\sqrt{N}$.

Wij kunnen dit nog toelichten a.v.: de Craddock-a_N is één van vele mogelijke lineaire combinaties der waarden f_1 ..... f_N. De "beste" symmetrische lineaire combinatie [1]) van f_1, f_2, ... f_m, f_{m+1} ..... f_{2m+1} heeft de vorm:

$$f^1_{m+1} = c_1 f_1 + c_2 f_2 + \ldots\ldots c_m f_m + c_{m+1} f_{m+1} + c_m f_{m+2} + \ldots\ldots c_1 f_{2m+1}.$$

De coëfficiënten zijn $c_j = \frac{1}{2m+1}$; $j = 1, 2 \ldots\ldots m+1$

D.i. iedere f^1_{m+1}, geheten de "gladgestreken" f_{m+1}, is het gewone (rekenkundige) gemiddelde van 2m+1 f-waarden. Hierbij is er vanuit gegaan, dat elke f_i òf een normale verdeling (μ ; σ, onafhankelijk van i) volgt òf een normale verdeling: μ_i, σ, met $\mu_i = a + bi$ (lineaire trend). Wij spraken van "beste" lineaire combinatie: alleen voor deze bijzondere symmetrische lineaire combinatie is de standaarddeviatie van f^1_{m+1} zo klein mogelijk en wel $\sigma/\sqrt{2m+1}$ (natuurlijk moet ook $\mathcal{E} f^1_{m+1} = \mu$). Er zijn vele andere lineaire combinaties van de 2m+1 f-waarden te bedenken, die wij kunnen schrijven in de vorm:

$$f^*_{m+1} = \left[\sum_1^{2m+1} t_i \cdot f_i\right] : \left[\sum_1^{2m+1} t_i\right]$$

waarvoor natuurlijk $\mathcal{E} f^*_{m+1} = \mu$. Ook deze f^* volgt een normale verdeling rondom μ en met zekere σ^*.

---

1) Zie Dissertatie Groningen 1936: "Over het gladstrijken van krommen". door H.J. de Boer.

Er geldt: $\sigma^*/\sigma = \left(\sqrt{\sum_1^{2m+1} t_i^2}\right) : \sum_1^{2m+1} t_i$

Men kan bewijzen, dat $1 \geqslant \sigma^*/\sigma \geqslant \frac{1}{2m+1}$. De σ^* is alleen dan gelijk $\frac{1}{2m+1}$ als elke $t_i = \frac{1}{2m+1}$. In ons geval vervange men 2m+1 door N en men ziet direct, dat $\sigma > \sigma_c > \sigma/\sqrt{N}$

4.6.2 Het Gladstrijkings-procédé (G-methode): $\frac{1}{n}$ (1, 1, 1 ..... 1) [1]
n = oneven

Bij een doorlopende gladstrijking ontstaan N - n+1 gladgestreken termen, n.l. $f^1_{\frac{1}{2}(n+1)}$, $f^1_{\frac{1}{2}(n+3)}$ ..... $f^1_{N-\frac{1}{2}(n-1)}$

Elk van deze f^1-termen heeft een gemiddelde μ en een st. dev. $\sigma/\sqrt{n}$.
Vervolgens wordt het gemiddelde G van deze N - n+1 f^1-waarden berekend:

$$G = \left[(f_1 + f_2 + \dots f_n) + (f_2 + f_3 + \dots f_{n+1}) + \dots (f_{N-(n-1)} + \dots f_N)\right] / (N - n+1)\, n$$

Allereerst bewijzen wij, dat $\mathcal{E}G = \mu$. Er doen zich twee gevallen voor:

I) $n \leqslant \frac{1}{2}N$. In dat geval zijn de coëfficiënten van de f_i in de teller van G van links naar rechts:

$$1 \quad 2 \quad 3 \quad \dots \quad n \quad n \quad \underbrace{n \quad \dots \quad n}_{t} \quad n \quad n-1 \quad \dots \quad 2 \quad 1$$

Het aantal "tussen-f_n's" is $t \geqslant 0$; $t = N - 2n$. De som van alle coëfficiënten is $2(1 + 2 + \dots n) + (N - 2m)\, n = n\,(N - n + 1)$, zodat $\mathcal{E}G = \mu$

II) $n > \frac{1}{2}N$. In dit geval zijn dezelfde coëfficiënten van links naar rechts: $1 \quad 2 \quad 3 \quad \dots \quad q \quad \underbrace{q \quad \dots \quad q}_{t} \quad q \quad q-1 \quad \dots \quad 2 \quad 1$,

met $q = N - n + 1 < n$
$t = N - 2q = -N + 2n - 2$

Nu is $\sum$ coëff. $= 2(1 + 2 + \dots q) + (-N + 2n - 2)\, q = n(N - n + 1)$, hetzelfde als onder I en weer $\mathcal{E}G = \mu$

---

1) Deze symbolische schrijfwijze betekent: men vervangt iedere f_i door zijn gladgestreken waarde $\frac{1}{n}\sum_{j=a}^{b} f_j$, met $a = i - \frac{1}{2}(n-1)$ en $b = i + \frac{1}{2}(n-1)$.

Vervolgens berekenen wij σ_g voor elk der twee gevallen I en II.

Geval I: $n \leqslant \frac{1}{2}N$; n = oneven

$$\sigma_g = \sqrt{\frac{2(1^2 + 2^2 + \ldots n^2) + (N - 2n)n^2}{(N - n + 1)\,n}}\,\sigma = \sqrt{\frac{\frac{1}{3}n(n + 1)(2n + 1) + (N - 2n)n^2}{(N - n + 1)\,n}}\,\sigma =$$

$$\sigma\sqrt{\frac{-\frac{4}{3}n^2 + n(N + 1) + \frac{1}{3}}{n(N - n + 1)^2}}$$

Deze σ_g hangt van N en n af.
Over de keuze van n is nog niets gezegd; slechts: $1 \leqslant n \leqslant \frac{1}{2}N$ (n = oneven).
Als $n = 1$ is $\sigma_g = \sigma/\sqrt{N}$, hetgeen te begrijpen is, want dan heeft men voorlopend gladgestreken over maar één term, hetgeen rekenkundig middelen betekent.
Als $n = \frac{1}{2}N$ (indien N = 4 voud + 2) is $A = \frac{4}{3} \cdot \frac{1}{N}\,\frac{N + 1}{N + 2} > \frac{1}{N}$, omdat $N > 2$.

Als bovendien $N \gg 1$, dan $A \approx \frac{4}{3}\,\frac{1}{N}$. of $\sigma_g \approx \sigma\sqrt{\frac{4}{3} \cdot \frac{1}{N}} = 1.152\sqrt{\frac{1}{N}}$

Bovendien blijkt, dat bij $n = 1$ de A toeneemt voor toenemende n. Wij vermoeden dus, dat A continu toeneemt van $\frac{1}{N}$ (als $n = 1$) tot $\frac{4}{3}\frac{1}{N}\,\frac{N + 1}{N + 2}$ (als $n = \frac{1}{2}N$). Een exact onderzoek zou eisen, dat wij zouden bewijzen, dat $dA/dn > 0$ voor elke n tussen 1 en $\frac{1}{2}N$.

Geval II: $n > \frac{1}{2}N$; n = oneven

Nu is $\sigma_g = \sigma\,\dfrac{\sqrt{2(1^2 + 2^2 + \ldots q^2) + (N - 2q)\,q^2}}{(N - n + 1)\,n}$; met $q = N - N + 1$

$$\sigma_g = \sigma\sqrt{\frac{-\frac{4}{3}q^2 + (N + 1)\,q + \frac{1}{3}}{(N - n + 1)} \cdot \frac{1}{n^2}} = \sigma\sqrt{B}\,.$$

De q neemt af met toenemende n. De kleinste waarde, die n kan aannemen, is het eerste boven $\frac{1}{2}N$ gelegen oneven getal (voorbeelden: $N = 5$, $n \geqslant 3$; $N = 6$, $n \geqslant 5$; $N = 7$, $n \geqslant 5$; $N = 8$, $n \geqslant 5$; $N = 9$, $n \geqslant 5$; $N = 100$, $n \geqslant 51$, enz.). Als wij $n = N$ substitueren (hetgeen mogelijk is als N = oneven) is $q = 1$ en komt er:

$$B = \frac{-\frac{4}{3} + (N + 1) + \frac{1}{3}}{N^2} = \frac{1}{N} \text{ en } \sigma_g = \sigma\sqrt{\frac{1}{N}}\,.$$

Dat moet ook wel, want dan hebben wij het rekenkundig gemiddelde over alle N termen berekend. Als N even is, is de grootste n, die wij kunnen substitueren N - 1 en $B = \frac{2N-3}{2(N-1)^2} > \frac{1}{N}$ of wel $\sigma_g > \sigma / \sqrt{N}$.

Dit alles betekent (een nauwkeuriger analyse is niet nodig), dat met toenemende n (te beginnen met n = 1 en eindigende met òf N òf N - 1, al naar gelang N oneven of even is) de σ_g toeneemt van $\sigma / \sqrt{N}$ naar een maximum van ongeveer $\frac{4}{3} \cdot \frac{1}{N} \frac{N+1}{N+2} > \frac{1}{N}$ (voor grote N is dit ongeveer $\frac{4}{3} \frac{1}{N}$) en dan weer afneemt òf naar $\frac{\sigma}{2(N-1)} \sqrt{2(2N-3)}$ (als N = even), d.i. $> \sigma / \sqrt{N}$, òf naar $\sigma / \sqrt{N}$, als N = oneven.

Wij hebben dit willen illustreren voor N = 10 (fig. 1) In aanmerking komen n = 1, 3, 5, 7 en 9. De σ_c ligt ergens (volgens een mij onbekende waarschijnlijkheidsverdeling) tussen σ en $\sigma / \sqrt{10}$ (misschien ligt de meest waarschijnlijke σ_c dicht bij $\sigma / \sqrt{10}$, maar hoe dicht dan?). De σ_g heeft een waarde afhankelijk van n, doch ligt in elk geval tussen 0,316 σ en 0,349 σ. Hierbij krijgt men 0.316 door in A n = 1 te substitueren en 0.349 door n = 5 (= $\frac{1}{2}$ N) in te vullen. De σ_g hangt dus weinig af van n; als N zeer groot is, varieert, met veranderende n, de σ_g tussen $\sigma / \sqrt{N}$ en $[\sigma / \sqrt{N}] . \sqrt{4/3}$.

4.6.3 Tenslotte nog de methode der kleinste kwadraten (M.K.K.)

In dit geval worde $a_N = \frac{1}{N} \sum_1^N f_i$ en is $\mathcal{E} a_N = \mu$ en

$$\sigma_{M.K.K.} = \sigma / \sqrt{N}.$$

4.6.4

Conslusie:

Hoewel wij het niet volkomen zeker weten (aangezien wij de waarschijnlijkheidsverdeling van Craddock's r zouden moeten weten, met welke weer die van σ_c samenhangt) lijkt het ons zeer waarschijnlijk, dat de "Methode van het glijdende gemiddelde" (voortschuivend gladstrijken) tot een kleinere σ_g leidt dan het "Craddock - se procédé". De gewone methode der kleinste kwadraten leidt in elk geval tot de kleinste σ van a_N en wel $\sigma / \sqrt{N}$.

Opmerking:

Toegegeven moet worden, dat wij hier de G-methode en M.K.K. toepasten zonder aan de f-waarden gewichten toe te kennen. De vergelijking, die wij met de C-methode maakten, is dus niet geheel eerlijk.

Hoe echter anders te handelen als deze gewichten niet bekend (gegeven) zijn? Craddock geeft ze ook niet van te voren, doch laat ze van de constellatie der f-waarden zelf afhangen (hetgeen intussen ook al vreemd is). Men zou aldus kunnen handelen: de volgens het Craddock-procédé berekende gewichten (die met zijn r samenhangen) opnemen in het G-procédé en in de M.K.K. en dan weer de σ in de resulterende a_N berekenen. Hierin zijn we niet geslaagd.

4.7 Een numeriek voorbeeld (N = 7); toepassing van het procédé-Craddock, van alternatieve methodes en van de methode der kleinste kwadraten.

Wij bedachten 7 f-waarden $f_1 = 3$ $f_2 = 5$ $f_3 = 4$ $f_4 = 3$ $f_5 = 5$ $f_6 = 4$ en $f_7 = 6$ en pasten daar verschillende methodes op toe.

4.7.1 De methode van Craddock met bijv. n = 3

Geprobeerd wordt F_n (i) constant = $a_n = a_3$. De conditie $S_1 =$ $r^2 (a_3 - 3)^2 + r (a_3 - 5)^2 + (a_3 - 4)^2 =$ minimum leidt tot

$a_3 = \dfrac{3r^2 + 5r + 4}{1 + r + r^2}$. Daarna wordt a_4 bepaald uit de conditie:

$$S_1 = r^3 (a_4 - 3)^2 + r^2 (a_4 - 5)^2 + r (a_4 - 4)^2 \overbrace{+ (a_4 - 3)^2} = \text{minimum, zodat}$$

$$a_4 = \frac{3r^3 + 5r^2 + 4r + 3}{1 + r + r^2 + r^3} \quad ; \quad \text{zo ook}$$

$$a_5 = \frac{3r^4 + 5r^3 + 4r^2 + 3r + 5}{1 + r + r^2 + r^3 + r^4} \quad ; \quad a_6 = \frac{3r^5 + 5r^4 + 4r^3 + 3r^2 + 5r + 4}{1 + r + r^2 + r^3 + r^4 + r^5} \quad \text{en}$$

$$a_7 = \frac{3r^6 + 5r^5 + 4r^4 + 3r^3 + 5r^2 + 4r + 6}{1 + r + r^2 + r^3 + r^4 + r^5 + r^6}$$; vervolgens moet men r oplossen uit de conditie:

$$T = (a_3 - 3)^2 + (a_4 - 5)^2 + (a_5 - 4)^2 + (a_6 - 6)^2 = \text{minimum.}$$

Men kan $\dfrac{dT}{dr} = 0$ niet exact oplossen. Proberenderwijze komt er $r_c = 0{,}89$, zodat $a_3 = 4.04$ $a_4 = 3.73$ $a_5 = 4.05$ $a_6 = 4.04$ $a_7 = 4.42$.

Vervolgens berekenen wij - zie (17) in Craddock-procédé -

$$\sigma_c^2 = \frac{(a_7-3)^2 + (a_7-5)^2 + (a_7-4)^2 + (a_7-3)^2 + (a_7-5)^2 + (a_7-4)^2 + (a_7-6)^2}{6}$$

(n.l. $N = 7$; $g = 0$) $= 1.26$; $\sigma_c = 1.12$. Dus de voorspelling is $F(8) =$ $a_7 = 4.42$ en de variantie van deze is 1.12

4.7.2 <u>Alleen gewichten aan de eerste $n = 3$ f-waarden</u> (gewijzigde Craddock-methode) met F = constant.

Men moet eerst $S_1 = r^2 (a_3-3)^2 + r (a_3-5)^2 + (a_3-4)^2$ minimaal maken. Dus $a_3 = \dfrac{3r^2 + 5r + 4}{1 + r + r^2}$. Vervolgens moet men $r^2 (a_4-3)^2 + r (a_4-5)^2 +$ $+ (a_4-4)^2 + (a_4-3)^2$ minimaal maken, zodat $a_4 = \dfrac{3r^2 + 5r + 4 + 3}{2 + r + r^2}$; daarna

$r^2 (a_5-3)^2 + r (a_5-5)^2 + (a_5-4)^2 + (a_5-3)^2 + (a_5-5)^2$ zodat

$a_5 = \dfrac{3r^2 + 5r + 4 + 3 + 5}{3 + r + r^2}$; idem $a_6 = \dfrac{3r^2 + 5r + 4 + 3 + 5 + 4}{4 + r + r^2}$ en

$a_7 = \dfrac{3r^2 + 5r + 4 + 3 + 5 + 4 + 6}{5 + r + r^2}$.

Vervolgens moet men $T = (a_3-3)^2 + (a_4-5)^2 + (a_5-4)^2 + (a_6-6)^2$ minimaliseren, hetgeen weer tot een vergelijking $\dfrac{dT}{dr} = 0$ leidt, die alleen benaderingsgewijze opgelost kan worden: $r = 0.90$, zodat $a_3 = 4.03$; $a_4 =$ $= 3.75$; $a_4 = 4.02$ en $a_7 = 4.31$; vervolgens $\sigma^2 = \sum_1^7 (a_7-f_i)^2 / 6 = 1.235$

dus $\sigma = 1.11$ en "forecast" $F_7(8) = \hat{f}_8 = a_7 = 4.31$.

4.7.3 <u>Slechts één keer minimaliseren</u> (niet voortschuivend extrapoleren) <u>met gewichten aan de eerste $n = 3$ termen</u>; F = constant

Voor welke r en a_7 is $S = r^2 (a_7-3)^2 + r (a_7-5)^2 + (a_7-4)^2 +$ $+ (a_7-3)^2 + (a_7-5)^2 + (a_7-4)^2 + (a_7-6)^2 =$ minimum? Men moet $\dfrac{\partial S}{\partial a_7} = 0$ en $\dfrac{\partial S}{\partial r} = 0$ oplossen. Uit de eerste vergelijking volgt:

$a_7 = \dfrac{3r^2 + 5r + 4 + 3 + 5 + 4 + 6}{r + r^2 + 5}$; wanneer men deze in de tweede vergelijking substitueert komt er

$r^4 \Delta^2 + 2r^3 \Delta^2 + 2 \Delta . r^2 (2A - B) + 2rA^2 + B^2 = 0$, met

$\Delta = f_2 - f_1 = 2$; $C = f_3 + f_4 + f_5 + f_6 + f_7 = 22$; $A = C - 5f_1 = 7$; $B =$ $C - 5f_2 = -3$. Men ziet, dat de aard der 4 wortels r samenhangt met de f-waarden. Wij kunnen de volgende gevallen onderscheiden

a) als $2A > B$ (altijd zijn Δ^2, A^2 en B^2 positief) dan zijn er òf 2 stellen toegevoegd complexe wortels òf 2 reëele negatieve wortels + 1 stel toeg. complexe òf 4 reëele negatieve wortels.

Geen van deze komt in aanmerking, want zinvol is alleen $0 < r \leqslant 1$.

b) $2A < B$. Misschien een wortel tussen 0 en 1.

c) $2A = B$, als onder a.

In ons geval is toevallig $2A > B$. Wij moeten de vraag dus anders formuleren: voor welke waarde van r uit het interval 0 tot 1 is S minimaal? Antwoord: alleen voor $r = 0$; dus $a_7 = \frac{1}{5} (f_3 + f_4 + f_5 + f_6 + f_7) = 4.40$.

Vervolgens nog

$$\sigma^2 = \sum_{i=1}^{7} (a_7 - f_i)^2 / 6 = 1.253 \quad \therefore \quad \sigma = 1.12$$

4.7.4 Eén keer minimaliseren; gewichten aan alle termen; $F_N(i) =$ constant $= a_7$

Maak $S = r^6 (a_7-3)^2 + r^5 (a_7-5)^2 + r^4 (a_7-4)^2 + r^3 (a_7-5)^2 + r^2 (a_7-5)^2 + r (a_7-4)^2 + (a_7-6)^2$ minimaal. Los op $\frac{\partial S}{\partial a_7} = 0$ en $\frac{\partial S}{\partial r} = 0$.

Aangezien S als som van kwadraten beslist $\geqslant 0$ is, is de oplossing $r = 0$ en $a_7 = f_7 = 6$, met $S = 0$. Dit geval is zonder zin, ook als men een andere functie van t voor F probeert ($a + bt$; $a + bt + ct^2$, enz. enz.), want altijd wordt in het traject $0 \rightarrow 1$ het absolute minimum van S bereikt voor $r = 0$ en $F(7) = f_7$.

4.7.5 De methode der kleinste kwadraten (geen trend)

Maak $\sum_1^7 (a_7 - f_i)^2 =$ minimum. Dus $a_7 = \frac{1}{7} \sum f_i = 4.28$ en $\sigma^2 =$ $= 1.018$ of $\sigma = 1.00$

4.7.6 Volgens de Δ - methode

nulde diff.:	$x =$	3	5	4	3	5	4	6	; som 30
eerste diff.:	$\Delta_1 =$		2	-1	-1	+2	-1	+2	; gem. 0.5
tweede diff.:	$\Delta_2 =$			-3	0	3	-3	+3	; gem. 0

Met $x = a + bt + ct^2$ is $\Delta_1 =$ $b + 3c$; $b + 5c$ ..... $+ b + 13c$;

en $\Delta_2 = 2c$; $2c$ ..... $2c$. dus $c = 0$ en $b = 0.5$ en

$\sum x = 7a + b\ (1 + 2 + 3 + 4 + 5 + 6) = 30$ $\therefore$ $a = 2.8$ en $x = 2.8 + 0.5\ t.$

Tabel 2

Gemeten:	x = 3	5	4	3	5	4	6	
berekend	2.8	3.3	3.8	4.3	4.8	5.3	5.8	som = 30
verschil ε	0.2	1.7	0.2	-1.3	0.2	-1.3	0.2	som = 30
kwadr. ε^2	0.04	2.89	0.04	1.69	0.04	1.69	0.04	som = 6.43

Dus $\sigma^2 = \dfrac{\sum \varepsilon^2}{7 - 1 - 1} = 1.29$ $\therefore$ $\sigma = 1.13$. en de voorspelling is $x\ (t = 8) = 6.3$.

4.7.7 De beste lineaire trend $F = a + bt$ ($t = 1$ op plaats $n = 1$)

Men moet minimaliseren $\sum_1^7 [a + b.i - f_i]^2$; de vergelijkingen

$\dfrac{\partial \Sigma}{\partial a} = 0$ en $\dfrac{\partial \Sigma}{\partial b} = 0$ leiden tot

$$\left.\begin{array}{l} 7a + b\ (6 + 5 + 4 + 3 + 2 + 1) = \sum f_i \\ (6 + 5 + 4 + 3 + 2 + 1)\ a + b\ (6^2 + 5^2 + 4^2 + 3^2 + 2^2 + 1) = \sum (i)\ f_i \end{array}\right\} \begin{array}{l} a = 3.148 \\ b = 0.285 \end{array}$$

Dus $F = 3.148 + 0.285\ t$

Tabel 3

t	f	F	$\varepsilon = f - F$	ε^2
1	3	3.43	-0.43	0.1849
2	5	3.72	1.28	1.6384
3	4	4.00	0	0.0000
4	3	4.29	-1.29	1.6641
5	5	4.57	0.43	0.1849
6	4	4.86	-0.86	0.7396
7	6	5.14	0.86	0.7396
som	30	30.01	-0.01	5.1515

$$\sigma^2 = \frac{\sum \varepsilon^2}{7 - 1 - 1} = 1.03$$

$$\therefore \sigma = 1.01$$

De voorspelling is F (8) = 5.4

4.7.8 De beste kwadratische trend $F = a + bt + ct^2$ (t = 1 in n = 1)

Men moet nu oplossen $\sum_1^7 [a + bi + ci^2 - f_i]^2 = \min.$

Via $\frac{\partial \Sigma}{\partial a} = 0$; $\frac{\partial \Sigma}{\partial b} = 0$; $\frac{\partial \Sigma}{\partial c} = 0$ komt er a = 3.98 b = 0.704 c = 0.0699

of $F = 3.98 + 0.70t + 0.07t^2$, waarna $\sigma^2 = \frac{\sum \varepsilon^2}{7 - 2 - 1} = \frac{4.719}{4} = 1.182$

en $\sigma = 1.084$; F (1) = 6.25

4.7.9 Voortglijdend middelen over n = 3 termen

Wij berekenen achtereenvolgens $\frac{1}{3}\sum_1^3 f_i = 4$; $\frac{1}{3}\sum_2^4 f_2 = 4$;

$\frac{1}{3}\sum_3^5 f_i = 4$; $\frac{1}{3}\sum_4^6 f_i = 4$; $\frac{1}{3}\sum_5^7 f_i = 5.67$ en $\frac{1}{5}(4 + 4 + 4 + 4 + 5.67) =$

$= 4.33$. Vervolgens $\sigma^2 = \sum_1^7 (f_i - 4.33)^2 / 6 = 1.24$ en $\sigma = 1.11$.

De voorspelling is dus met deze methode (die óók nog geldt als de trend lineair is) 4.33, met een $\sigma = 1.11$.

Opmerking: een goede maat voor de goedheid van aanpassing is de correlatie-coëfficiënt R tussen de gemeten f-waarden en de theoretische φ.

De $R^2 = 1 - \dfrac{\sum (f_i - \varphi_i)^2}{(f_i - \bar{f})^2}$. De noemer is geheel gegeven door de gemeten f-waarden; de teller hangt samen met de graad van de aangepaste F en wordt kleiner met toenemende graad, om nul te zijn als deze graad gelijk aan het aantal f-waarden is (gedacht is aan een polynoom $F = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + + \ldots\ldots$). Dan is $R^2 = 1$. Men kan de waarschijnlijkheidsverdeling van R berekenen als f geen tijdfunctie is (trend afwezig). De toetsing geschiedt a.v.: men lost de k op uit $R = \dfrac{k}{\sqrt{N-g-1+k^2}}$

(R gemeten via de bovenstaande uitdrukking; N = aantal f-waarden; g = graad van F). Deze k gehoorzaamt aan de Student-distributie met ν = N - g - 1 graden van vrijheid.

Wij hebben de resultaten van dit hoofdstuk overzichtelijk samengenomen in de volgende tabel en in fig. 2.

Tabel 4

$f_1, f_2 \ldots\ldots f_7 = 3, 5, 4, 3, 5, 4, 6$

			voorspelling voor n = 8	σ van de voorspelling	$\sum \varepsilon^2$
F = constant geen trend	4.7.1	methode Craddock; n = 3	4.42	1.12	7.555
	4.7.2	alleen gewichten aan de eerste n=3 f's	4.31	1.11	7.510
	4.7.3	één keer minimaliseren; gewichten aan eerste 3	4.40	1.12	7.520
	4.7.4	één keer minimaliseren; gewichten aan alle f's	zonder zin		
	4.7.5	methode der kl.kwadr.	4.28	1.00	7.429
	4.7.9	voortglijdend middelen	4.33	1.11	7.442
F = a + bt lin.tr.	4.7.6	Δ -methode	6.3	1.13	6.43
	4.7.7	beste lineaire trend	5.4	1.01	5.152
F = a+bt+ct2 kw.tr.	4.7.8	beste kwadr. trend	6.25	1.08	4.719

De tabel leert, dat de waarden van de gemiddeld te verwachten waarde van $\hat{f}_8$, in de onderstelling, dat trend afwezig is, berekend volgens de 6 methoden (waaronder 1: Craddock-methode en 5: Methode der kleinste kwadraten), zeer weinig uiteenlopen (tussen 4.28 en 4.42); de σ's van de voorspelde waarden verschillen eveneens niet noemenswaardig (zij lopen uiteen tussen 1.00 en 1.12). Het is op basis van dit simpele voorbeeld natuurlijk niet mogelijk te beslissen of de "methode Craddock" beter (slechter) dan de andere methodes is. Daarvoor is het ook niet bedoeld. Men moet dan ook geen bijzondere betekenis hechten aan het feit, dat 1°) 4.42 (Craddock) > 4.28 (Meth. kl. kwadr.) en 2°) 4.42 (C.) > 4.33 (voortglijdend middelen). Blijkbaar is het 1° een gevolg van het feit, dat de Craddock-se r zeer dicht bij 1 ligt (0.89) in <u>dit</u> getallen-voorbeeld. Een ander getallen-voorbeeld zou een ver van 1 afgelegen r hebben kunnen leveren. Als men zich wil afvragen of de $\hat{f}_8$ volgens Craddock ook kleiner dan bijv. $\hat{f}_8$ volgens de methode van het voortglijdend middelen kan zijn, komt men weer op het onder 4.6 behandelde probleem.

De Bilt, Mei 1955.

4.8 Summary

"Comment on a new method for the extrapolation of time series proposed by Craddock".

0 In this report comment is given on a new process for the analysis of time series, developed by J.M. Craddock, especially concerning the statistical aspects.

1 In this chapter a letter is cited by Mr Craddock and addressed to Prof.Berlage. This letter gives a description of the "new theory".

2 In the second chapter the essence of the new theory is stated as briefly and concisely as possible.

3 Next, in chapter 3, remarks and observations are made with respect to the text of Craddock's letter. They refer to successive points of that letter. Among others:

1) "The working of definite physical processes", "which can be expressed in mathematical terms", and "an element which cannot be so expressed, and is effectively random". We might stress the fact that both parts of each term of the time series are physical. What is meant by the word "random"? We speak of the influence of chance if the multitude of("physical) causes is so large and so complicate that its numerical consequences satisfy the laws of probability. Craddock speaks of "definite physical processes". Why does he do so? Does he have such a clear idea of these processes, that there is no doubt whatever as to the exact type of the function F(t)? In which (physical or statistical) way should we determine the type?

7) Why do we apply the "Method of Least Squares"? What can be said of the physical and (or) statistical background of this method? (see 4.1)

13), 14) The author should write $T = \sum_{n+1}^{N} e_i^2$. This T is a complicate function of r. Solving the condition T = minimum we arrive at the equation $\frac{dT}{dr}=0$ with as many roots as the degree of the equation.
Can we show that this equation always has at least one root between 0 and 1? I don't think so. If there happens to be no root in the interval of zero to unity the question should be raised: "for what value out of this interval (or perhaps better $\frac{1}{2}$ up to 1?) has T a relative minimum value?" This proves to be either 0 or 1.

16) "also the mean square error of forecasts". One should not use the σ^2-expression

$$\frac{\sum^{N} e_i^2}{N-g-1} \quad \text{but} \quad \frac{\sum_{n+1}^{N} e_i^2}{n-2-g} \quad ; \; g = \text{the degree of the function } F(t) \equiv \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \ldots, \alpha_{g+1} t^g$$

20) This is an approximation. The exact expression is

$$a_{n+p} = a_{n+p-1} + \frac{1-r}{1-r^{n+p}}(f_{n+p} - a_{n+p-1}), \text{ with } 1 \leqslant p \leqslant N-n, \text{ and}$$

$$a_n = \frac{1-r}{1-r^n}\sum_{o}^{n} r^i f_{n+1-i}.$$ Only if $r^n \ll 1$ the following approximation holds good:

$$a_{n+p} \cong a_{n+p-1} + (1-r)(f_{n+p} - a_{n+p-1}) \text{ and } a_n \cong (1-r)\sum_{o}^{n} r^i f_{n+1-i}$$

But how do we know a priori (before having computed the numerical value of r) that $r^n \ll 1$? How great is n?

26) Why does the author write $\frac{1}{N}\sum e_n^2$ (at the top of the table)? It should be: $\frac{1}{N-n-1}\sum_{n+1}^{N} e_i^2$.

Did Craddock use n=10?

We stress the fact that the forecast - σ_o is defined as $\sigma^2 = \frac{1}{N-1}\sum_{1}^{N} e_i^2$ (here g=0).

What result does Craddock get according to the common method of least squares?

28) Why "physically reasonable"? Does this remark refer to the meaning of r? Does this r have a physical character?

4 In chapter 4 we deal with the statistical aspects of Craddock's theory.

In 4.1 stress is laid on the origin of the so-called method of least squares which may be seen in connection to the "Method of maximum likelihood".

In 4.2 we show that generally (and especially if the time series is short) the variance of the forecast consists of two parts:

i) the variance of the regression function (the uncertainty as to the values of the parameters of F) and

ii) the variance around the regression. We mean the variances in Craddock's α, β, γ .... and the variance in connection to the differences between the forecasts and the actual terms.

In 4.3 we show that it is not correct to interpret manipulating with weighted means as a matter of fitting (point 8 of C.'s letter). Introducing weights $g_1 < g_2 < \ldots\ldots < g_N = 1$ to the terms $f_1, f_2 \ldots\ldots f_N$, the weighted mean becomes $a_g = (\sum g_i f_i) : \sum g_i$, resulting from the condition $\sum g_i(a-f_i)^2$ = minimum. If $g_i < g_j$ then

$$\frac{\partial a_g}{\partial f_i} < \frac{\partial a_g}{\partial f_j}$$

In other words: the influence of f_i on a_g is less than that of f_j on a_g.

We succeeded to prove that this value a_g produces a better fit to, for instance, the last n terms ($f_{N-n+1}; \ldots\ldots f_N$) than to the first n terms ($f_1, f_2 \ldots\ldots f_n$). The goodness of fitting of certain value m to the terms $f_{p+1}, f_{p+2} \ldots\ldots f_{p+r}$ is defined by the expression $\sum_{p+1}^{p+k} (f_i - m)^2 : k$.

In 4.4 we ask whether the weights $g_i = r^{n-i}$ (introduced by Craddock) have a pure statistical meaning or, moreover, a physical one.

In 4.5 we show that the "Method of Differences": $x(i) = \Delta(i)$; $\Delta_1(i) = x(i-1)-x(i)$; $\Delta_2(i) = \Delta_1(i+1) - \Delta(i)$; $\Delta_3(i) = \Delta_2(i+1) - \Delta_2(i)$; etc. (see for instance pg 262 in the "Handbook of statistical methods in meteorology", 1953, by Brooks and Carruthers) does not furnish a way in which we can determine the degree of F(t) exactly. The reason is that **we should examine** whether a trend is absent. Of course such a test gives an answer which is not fully reliable. In practice this fact does not matter generally, but we stress this fact in connection with the question in which way we can determine exactly the degree of $F = \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \ldots\ldots$

In 4.6 a comparison is made between the "Method of Craddock" (C-method) and the "Method of smoothing by overlapping means of n terms" (G-method).

We formulate the problem as follows:
A normal distribution $(0;\sigma)$ is given from which many random samples (each of N elements) are taken; each sample is plotted as a time series: $f_1, f_2 \ldots\ldots f_N$. Next we apply three methods 1) the C-method 2) the G-method 3) the method od least squares. We want to compute the value of a_N (the forecast on place N). Of course this a_N is a stochastic variable. If its distribution proves to be a normal one we might know the mean and standard deviation respectively: μ_c, σ_c; μ_g, σ_g; μ_n, σ_n.

The G-method is applied as follows: we calculate the mean of the first n terms ($n<N$); next the mean of the terms f_2 up to and including f_{n+1}; next of f_3 ..... f_{N+2}, etc., finally of f_{N-n+1} ..... f_N. We get $N-n+1$ mean values m_i and compute the general average: $a_N = \left[\sum_1^{N-n+1} m_i\right] : (N-n+1)$

4.6.1 C-method

The expectation proves to be $\mathcal{E}\, a_N = \mu$; moreover a_N satisfies a normal universum. We show: $\sigma \geq \sigma_c \geq \sigma/\sqrt{N}$. Unfortunately it is too difficult to derive the probability function of Craddock's r, which proves to be a stochastic variable. Consequently σ_c is also a stochastic variable. We naturally want to know the most probable value (modus) in this distribution. It is only possible to show that σ_c is situated between σ and $\sigma/\sqrt{N}$. The nearer the r to unity ~~the~~ nearer σ_c to $\sigma/\sqrt{N}$.

4.6.2 G-method

Again a_N satisfies a gaussian distribution. $\mathcal{E}\, a_N = \mu$.
The σ_G depends on n and N. When the value of n (N being constant) from 1 to $\frac{1}{2}$ N is increased the σ_G increases from $\sigma/\sqrt{N}$ (if $n=1$) to nearly $(\sigma/\sqrt{N})\sqrt{\frac{4}{3}}$ (if $n = \frac{1}{2}N$ and $N \gg 1$).
For larger values of n the σ_g decreases to either $\sigma/\sqrt{N}$ ($n=N$ if N=odd) or $\frac{\sigma}{2(N-1)}\sqrt{2(2N-3)} > \frac{\sigma}{\sqrt{N}}$ ($n=N-1$ if N= even).

We have illustrated this in figure 1; N=10; n=1, 3, 5, 7, 9.
Conclusion: we cannot solve the problem in an exact way, because we cannot derive the probability distribution of the r (and in consequence that of the σ_c). Still it seems to be right to state that generally $\sigma_G < \sigma_c$, so that Craddock's method in this sence is less satisfactory than the G-method.

Finally the Method od Least Squares is the best. The value a_N satisfies a normal distribution with $\mathcal{E}\, a_N = \mu$; but $\sigma_M < \sigma_G < \sigma_c$, because $\sigma_M = \sigma/\sqrt{N}$.(We are sure of the first sign $<$; and probably in most cases the second sign $<$ also holds good).

4.7 In this section we applied seven methods to a very simple numerical example: $f_1 = 3$, $f_2 = 5$, $f_3 = 4$, $f_4 = 3$, $f_5 = 5$, $f_6 = 4$, $f_7 = 6$.
See fig.2 and table 4.
These methods are

4.7.1 The C-method. $N = 7$; $n = 3$.

We suggested F = constant = a and obtained r = 0.89. The mean value of the forecast $\hat{f}_8 = 4.42$; the forecast $-\sigma_c = 1.12$.

4.7.2 Next we gave weights only to the terms f_1, f_2, f_3 (r^2, r, 1).

The Craddock-method, but each of f_4 ..... f_7 has the weight 1. Then r = 0.90; $\widehat{f}_8$ = 4.31; σ = 1.11

4.7.3 Next we gave weights only to the terms f_1, f_2, f_3 (see 4.7.2) and minimised only one polynomium:

$$S = r^2(a_7-3)^2 + r(a_7-5)^2 + (a_7-4)^2 + (a_7-3)^2 + (a_7-5)^2 + (a_7-4)^2 + (a_7-6)^2.$$

Solving $\frac{\partial S}{\partial a_7} = 0$ and $\frac{\partial S}{\partial r} = 0$ arrive at the equation:

$$r^4\Delta^2 + 2\,r^3\Delta^2 + 2\Delta . r^2\,(2\,A-B) + 2\,rA^2+B^2 = 0 \text{ with}$$

$\Delta = f_1 - f_2 = 2$; $C = f_3+f_4+f_5+f_6+f_7 = 22$; $A=C-5f_1 = 7$; $B=C-5f_2 = -3$.

Since 2A-B $>$ 0 there are no roots within the interval 0$\rightarrow$1. We must take r = 0, bevause for all values out of the interval 0$\rightarrow$1 the S has a minimum value only for r = 0.

So $a_7 = \frac{1}{5}(f_3 + \dots\dots f_7) = 4.40$ and $\widehat{f}_8 = 4.40$ with σ = 1.12.

4.7.4 We can also give weights to all terms and minimize

$S = \sum_1^7 r^{7-i}\,(a_7-f_i)^2$. Then $S \geq 0$ and the solution of the equations $\frac{\partial S}{\partial a_7} = 0$ and $\frac{\partial S}{\partial r} = 0$ must be r = 0; $a_7 = f_7 = 6$.

There is no sense in doing so.

4.7.5 The method of least squares (no trend); all weights are 1.

We minimize $\sum_1^7 (a_7 - f_i)^2$ and come to $a_7 = \frac{1}{7}\sum_1^7 = 4.28$ and σ = 1.00. All these methods started from the assumption F(t)= constant=a. Next we also introduce F(t) = a+bt and a+bt+ct^2 (the value t=1 in place n=1).

4.7.6 The method of differences (see 4.5)

Here x=3, 5, 4, 3, 5, 4, 6; so Δ_1 = 2, -1, -1, 2, -1, 2 and Δ_2 = -3, 0, 3, -3, 3(mean = 0). Introducing x = a+bt+ct^2 we get: $\overline{\Delta}_2$ = 2c = 0 and b = 0.5 and a = 28; x = 2.8+0.5 t.

Hence $\widehat{f}_8$ = x(8) = 6.3 and σ = 1.13.

4.7.7 <u>The best linear trend</u>, computed by means of the method of least squares; $F = a+bt$.

$S = \sum_1^7 \left(a+bi-f_i\right)^2$ must be minimized. The solution of $\frac{\partial S}{\partial a} = 0$; $\frac{\partial S}{\partial b} = 0$ is $a = 5.14$; $b = 0.28$ and $F = 3.15 + 0.28\ t$.
Hence $\hat{f}_8 = F(8) = 5.4$ with $\sigma = 1.01$.

4.7.8 <u>The best trend of second degree</u> $F = a+bt+ct^2$.

We minimize $S = \sum_1^7 \left[a+(i)\ b+(i)^2c-f_i\right]^2$. The solution of

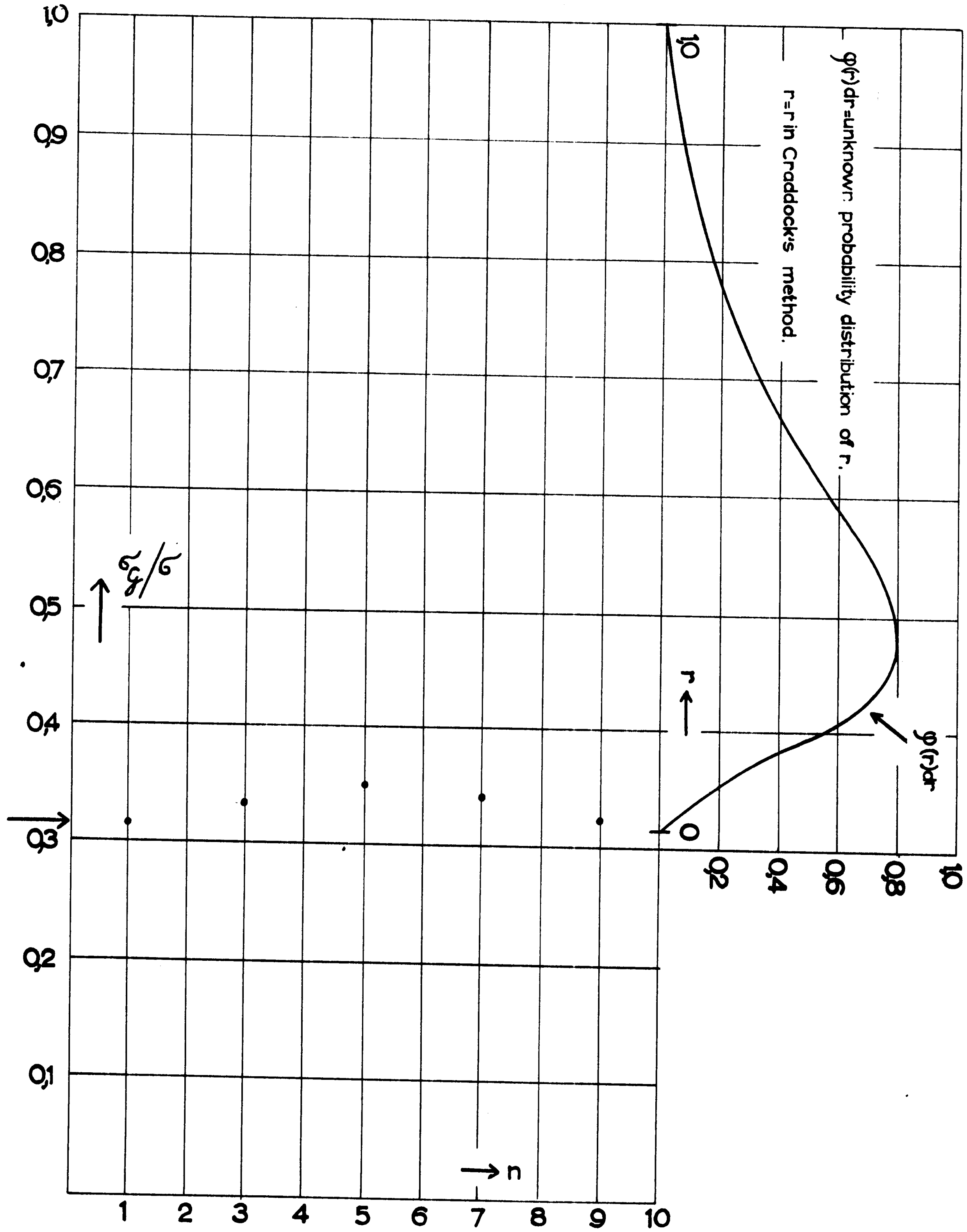
$\partial S/\partial a = 0 \quad \partial S/\partial b = 0 \quad \partial S/\partial c = 0$

becomes $a = 5.48$; $b = 0.70$; $c = 0.07$ and $F = 3.98+0.70\ t + 0.07\ t^2$.
Hence $\hat{f}_8 = F(8) = 6.2$ with $\sigma = 1.08$.

4.7.9 <u>Smoothing by overlapping means of n=3 terms</u> (see 4.6).

The successive means are 4; 4; 4; 4; 5,67 and the general mean is $\frac{1}{5}$ (4+4+4+4+5.67) = 4.33. Hence $\hat{f}_8 = 4.33$ with $\sigma = 1.11$.
Table 4 shows that the results do not differ appreciably. We conclude that there is little use in this numerical example in preferring one method to another. It is pointed out that the Craddock-forecast $\hat{f}_8$ and the Craddock-σ are a little larger than the corresponding values computed according the other methods. Perhaps the contrary occurs when another constellation of f-values is used. <u>Still we expect</u> (see 4.6) <u>that generally the Craddock-method will prove to be slightly less satisfactory than the other methods.</u>

Fig.1



φ(r)dr=unknown probability distribution of r.
r=r in Craddock's method.
φ(r)dr
r
σ_g/σ
n
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
1
2
3
4
5
6
7
8
9
10
0
0,2
0,4
0,6
0,8
1,0

Fig. 1

Let a normal distribution (μ, σ) be given from which many random samples (each of N=10 terms) are taken. Each sample is plotted as a time series f_1, f_2 ..... f_{10}. We apply 3 methods:

1. The Craddock-method (C)
2. The Method of smoothing with means of n terms (G)
3. The Method of least squares (M)

The object to examine is the distribution function of the forecast f_{11}, which proves to be a normal one.

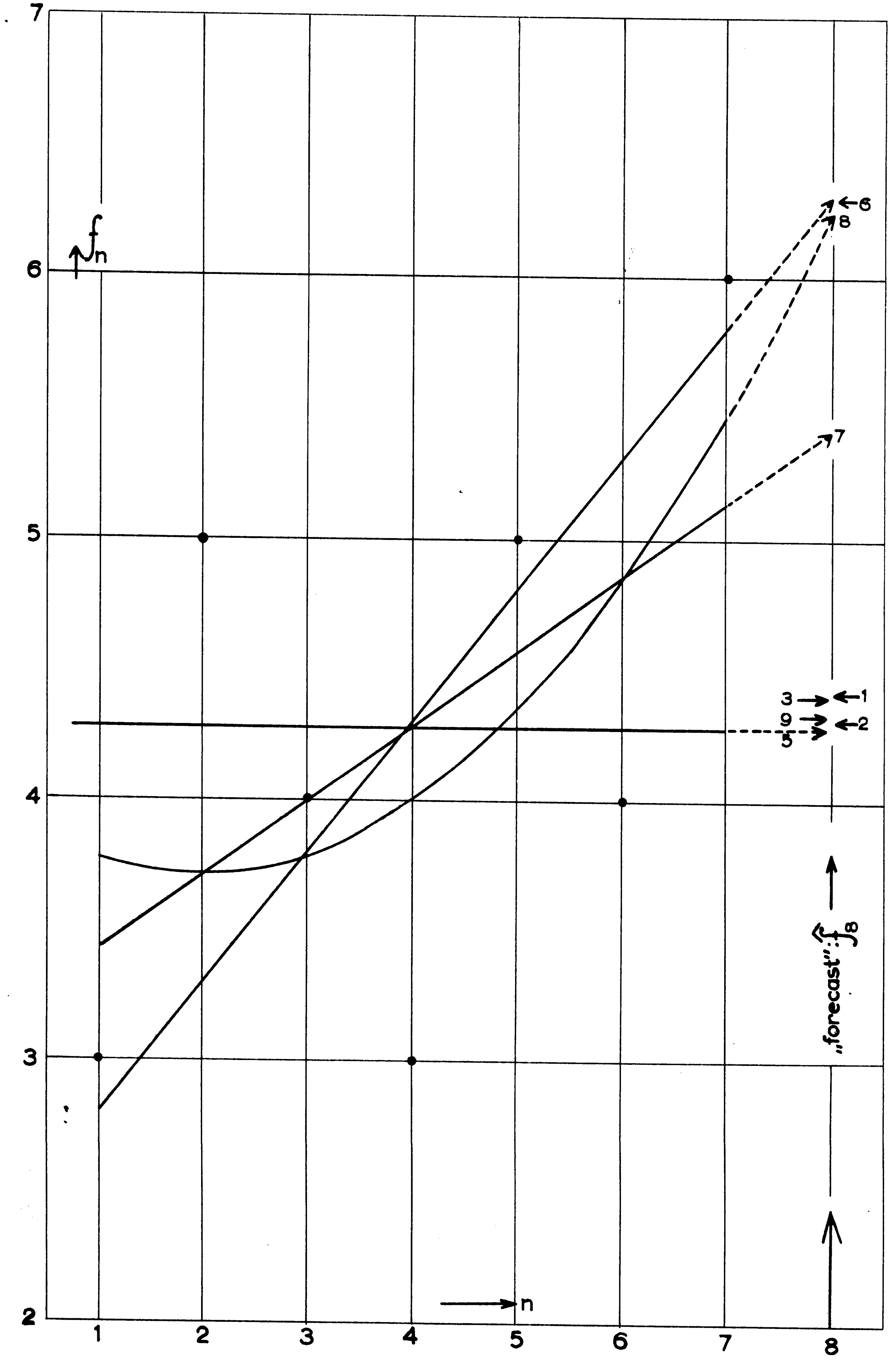
$\mathcal{E} f_{11} = \mu$, but $\sigma_M < \sigma_G < \sigma_C$ with $\sigma^2 = \mathcal{E}(f - \mathcal{E} f)^2$ = variance.

The figure illustrates the values of σ_G / σ when the method G is applied, n=1, 3, 5, 7 and 9. When method C is applied the σ_c must be situated between 0.316 σ and σ. This σ_c satisfies an unknown probability function.

Method M gives: σ_M = 0.316 (see arrow→) Here 0.316 = $1 : \sqrt{10}$.

n	σ_G/σ
1	0.316
3	0.334
5	0.349
7	0.342
9	0.324

Fig.2



7
6
5
4
3
2
$\uparrow f_n$
1
2
3
4
5
6
7
8
$\rightarrow n$
←6
8
7
3 →
←1
9 →
←2
5
„forecast": $\hat{f}_8$

Fig. 2

Numerical example $f_1=3$; $f_2=5$; $f_3=4$; $f_4=3$; $f_5=5$; $f_6=4$; $f_7=5$.

We applied 8 methods:

1. Craddock-method with $n = 3$
2. Modified C-method; only weights r^2, r and 1 to f_1, f_2 and f_3
3. Another modification. Minimizing only one time, with weights r^2, r and 1 to f_1, f_2 and f_3.
5. Method of least squares
6. Method of successive differences (linear trend)
7. The best linear trend
8. The best trend of second degree
9. Smoothing by overlapping means of 3 terms

N.B. Method 4 is left out of consideration (see text). The numbers at the right hand side of the figure belong to the values of the forecast $\hat{f}_8$ at place $n = 8$, computed according to the methods 1, ... 9.