CWI Syllabi

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

CWI Syllabus

Vacantiecursus 1984
Hewet - plus wiskunde

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

ISBN 90 6196 276 5

I N H O U D

HEWET en nog veel meer

De vacantiecursus gaat dit jaar, meer dan ooit, over de achtergronden van de schoolwiskunde. Verleden jaar complexe getallen, een verdwijnend keuze onderwerp uit wiskunde II, nu al het nieuwe van wiskunde A en wiskunde B. Natuurlijk, meer in het bijzonder ruimtemeetkunde voor de jongere leraren, het nieuwe stukje in wiskunde B. De oude getrouwe leraren weten nog wel van stereometrie, is dat hetzelfde, of is het hetzelfde anders? Een heerlijk echt zuiver wiskundig onderwerp, of steken ook hier de toepassingen de kop op? Zeker staat wiskunde A bol van de toepassingen (om een term uit de ruimtemeetkunde nu eens overdrachtelijk te gebruiken). En hoewel de deskundigen in de cursus de achtergronden van matrices, statistiek, besliskunde etc. apart zullen behandelen, zal de leraar op school in de dagelijkse lespraktijk dwarsverbindingen zoeken, vinden en benutten. En dan die wetenschap in snelle ontwikkeling, zelfs de naam is snel ontwikkeld en nu bij Automatische Gegevens Verwerking blijven staan. Na allerlei andere termen benut te hebben gaat nu onder deze naam de computer de A-klas in!
Heeft u er op uw school al één of meer, uit een ouderfonds, uit het honderd scholen project of zelf gemaakt?? Er moet iets mee gebeuren, want wiskunde A gebruikt simulaties, die net echt moeten zijn en dus redelijk veel getallen benutten of veel nogal slaafs rekenwerk vragen. Als van een 3 x 3 matrix de 9 getallen van het type 3,1415 of 2,7182 zijn is de berekening van de derde en vierde macht van die matrix niet zo'n leuk handwerkje! Dit soort berekeningen waren zelfs in de goede oude tijden van mijn toelatingsexamen voor de HBS in 1935 te moeilijk! Dus leve de AGV! Toch blijven we bang voor de laatste voordracht in Amsterdam: Excursie naar het Moeras van het Onberekenbare!
Dus toch? De samenvatting spreekt over een plotseling opstekende storm! Is HEWET daarmee bedoeld? Maar HEWET was toch al lang van te voren aangekondigd! En is HEWET wel een storm, of moet u alle zeilen bijzetten voor het nieuwe A programma? Er zal werk aan de winkel zijn, nieuwe onderwerpen en vooral nieuwe proefwerkopgaven. Hoe kom ik meer te weten om boven de stof te staan? Juist, door deze vakantiecursus HEWET-plus Wiskunde!
Veel plezier en veel succes en voor straks, gezond weer voor de klas!

F. van der Blij

MATHEMATISCHE BESLISKUNDE: 'THEORIE en PRAKTIJK'

prof.dr. G. de Leve

Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam

1. INLEIDING

Besliskunde is de studie die zich bezighoudt met het streven om beslissingsproblemen op zodanige wijze te vertalen in wiskundige problemen, dat de oplossing van de wiskundige versie van het beslissingsprobleem na terugvertaling de gevraagde beslissing of strategie oplevert.

Alhoewel het tijdstip waarop voor het eerst besliskunde werd bedreven, moeilijk is te bepalen, kan men toch wel zeggen dat de besliskunde zijn opkomst dankt aan de gecompliceerde beslissingsproblemen, waarvoor men zich in de Tweede Wereldoorlog gesteld zag. Daarna bleek de ontwikkelde benaderingswijze ook goed bruikbaar voor het oplossen van tal van beslissingsproblemen op het gebied van het beheer, de produktie en het transport.

In de besliskunde onderscheidt men twee typen van beslissingsituaties. Beslissingssituaties, waarin van de beslisser wordt verwacht dat hij slechts één enkele *beslissing* neemt, leiden tot zgn. *één-stapsbeslissingsproblemen.* Er zijn daarentegen ook beslissingssituaties waarin de beslisser in een al of niet begrensd tijdinterval een reeks van min of meer op elkaar afgestemde beslissingen moet nemen. De oplossing van deze *meer-stapsbeslissingsproblemen* wordt gegeven door een *strategie*; d.i. een beslissingvoorschrift dat voor ieder tijdstip vaststelt of de beslisser een beslissing moet nemen, en zo ja, welke dit zal zijn.

Het vertalen van een beslissingsprobleem in een wiskundig problem is onverbrekelijk verbonden aan het construeren van een *wiskundig model* van de te beschouwen beslissingsituatie. In een dergelijk model wordt de onderlinge samenhang en de evolutie van de verschillende voor de beslissingssituatie relevante factoren op wiskundige wijze beschreven. Bij de beschrijving wordt gebruik gemaakt van *kennis* die, hetzij door de basis-wetenschappen, zoals economie en wiskunde, wordt verschaft, hetzij uit beschikbare *gegevens* wordt gedistilleerd of uit *waarnemingen* wordt verkregen. Bovendien berust het wiskundige model op *veronderstellingen* die op het eerste gezicht redelijk lijken en niet op grond van de zojuist genoemde kennis dienen te worden verworpen. De kansrekening en de mathematische statistiek spelen een belangrijke rol bij het opstellen en testen van het wiskundige model.

Een belangrijk punt bij de de constructie van het wiskundige model is de keuze van het *kriterium* voor het onderling vergelijken van beslissingen en strategieën. In veel beslissingsproblemen houdt het kriterium nauw verband met de kosten of de gemiddelde kosten per tijdseenheid; bij sommige produktieproblemen zou men bijv. als kriterium het aantal produktiewijzingingen per tijdseenheid kunnen kiezen. Belangrijk is deze keuze van het kriterium, omdat de structuur van het kriterium dikwijls bepalend is voor de vraag of de wiskundige versie van het beslissingsprobleem al of niet met de huidige kennis en hulpmiddelen kan worden opgelost. In de praktijk betekent dit dat *vereenvoudigingen* in het model moeten worden aangebracht. Uiteraard moet dan worden nagegaan in hoeverre deze vereenvoudigingen het antwoord bepalen. Zowel nieuwe gegevens als onaanvaardbare "optimale" beslissingen leiden tot verwerping van het bestaande model en dus tot het opstellen van een nieuw.

Een groot deel van de besliskundige research heeft betrekking op het oplossen van speciale typen van wiskundige problemen. Dit onderdeel van de besliskunde wordt wel eens aangegeven met de naam *mathematische besliskunde.* De mathematische besliskunde omvat studies die betrekking hebben op de meest uitéénlopende onderdelen van de wiskunde. Het samenbindende is het dienstbaar zijn aan de analyse van beslissingssituaties. In de hierna volgende secties zullen wij daarvan voorbeelden zien.

Ieder besliskundig onderzoek begint met een inventarisatie van mogelijke beslissingen of strategieën. Om de effecten van deze beslissingen en strategieën op wiskundige wijze te kunnen bestuderen, dient de beslisingssituatie voldoende kwantificeerbaar te zijn. Aangezien in beslissingssituaties het doen van experimenten veelal is uitgesloten, komen alleen die situaties voor een besliskundig onderzoek in aanmerking, waarvoor geldt dat het vereiste model een doorzichtige structuur bezit. Dit zijn dan ook de redenen waarom besliskundige technieken tot dusver voornamelijk hun toepassing vonden bij het oplossen van bedrijfsproblemen. In principe beperkt de toepassing van besliskundige technieken zich echter niet tot één of meer probleemgroepen.

In tal van beslissingssituaties moeten beslisingen zonder dralen kunnen worden genomen. Dit betekent dat van een computer wordt verwacht dat hij, mits voorzien van optimaliseringstechnieken en beschikbare informatie, de beslisser via het beeldscherm een voorstel kan doen. In het grensgebied van besliskunde en informatica zijn interessante ontwikkelingen gaande.

2. EEN-STAPSBESLISSINGSPROBLEMEN

Bij één-stapsbeslissingsproblemen start de wiskundige modelvorming met het wiskundig beschrijven van de mogelijke *beslissingen.* Het is gebruikelijk om een beslissing, waaraan n kwantitatieve aspecten te onderscheiden zijn, aan te geven door een vector x met n componenten

$$X = (x_1, x_2, \ldots, x_n)$$

in een mengprobleem bijv. stellen deze componenten de fracties van de n samenstellende gronstoffen voor. Ook niet-kwantitatieve beslissingen laten zich dikwijls voor vectoren weergeven. Zo kan de vector $x = (0,1,0)$ uitdrukken dat de machine 2 wèl ($x_2 = 1$) en de machines 1 en 3 niet ($x_1 = x_3 = 0$) worden gebruikt bij de komende produktie. Uit de voorgaande toelichting volgt dat sommige componenten van de beslissingsvector alleen geheeltallige waarden mogen aannemen. De verzameling van indices j waarvoor x_j geheel moet zijn, wordt in het hiernavolgende steeds aangeduid met G. Omstandigheden, al dan niet typerend voor het beslissingstijdstip, beperken veelal de keuzemogelijkheden. Voor het bovengenoemde mengprobleem moet in ieder geval gelden:

$$x_1 + x_2 + \cdots + x_n = 1$$

$$x_j \geqslant 0 \quad (j = 1,2, \ldots, n),$$

maar misschien ook

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \leqslant b,$$

wanneer het mengsel, een veevoeder, hoogstens een fractie b aan ruwe celstof mag bevatten en voor de samenstellende grondstoffen grondstoffen deze fracties volgens recente analyses $a_j (j = 1,2, \ldots, n)$ bedragen. Andere kwaliteitseisen leiden tot soortgelijke voorwaarden.

In het wiskundige model wordt de verzameling van *toegelaten beslissingen* X gegeven door gelijk- en ongelijkheden waaraan de componenten van de beslissingsvector moeten voldoen. Het opstellen van deze relaties vormt dikwijls het moeilijkste onderdeel van de modelvorming en vraagt enige oefening. Zo zal de formulering van het wiskundig equivalent van de voorwaarde: "de gronstoffen 1 en 2 mogen dan en slechts dan beide in het mengsel voorkomen als de fractie van gronstof 3 groter dan of gelijk is aan die van 4" wel enige moeite opleveren. In het wiskundig model wordt deze voorwaarde gegeven door:

$$x_1 \leqslant x_{n+1}, x_2 \leqslant x_{n+2}, x_4 - x_3 \leqslant x_{n+3},$$

$$x_{n+1} + x_{n+2} + x_{n+3} \leqslant 2 \quad (n+1, n+2, n+3 \in G),$$

waarbij x_{n+1}, x_{n+2} en x_{n+3} nieuwe componenten zijn. (Ga na: $x_1 > 0$ én $x_2 > 0 \rightarrow \rightarrow x_{n+1} = x_{n+2} = 1 \rightarrow x_{n+3} = 0 \rightarrow x_3 \geqslant x_4$.)

Om een keuze te kunnen maken uit de verzameling is een *kriterium* vereist.

Als voor gronstof j de kosten per eenheid c_j bedragen dan kan wellicht de functie

$$k = c_1x_1+c_2x_2+\cdots+c_nx_n$$

als kriterium dienen. In het algemeen is het optimaliteitskriterium een functie $K(x_1,x_2,\ldots,x_n)$ van de componenten van de beslissingsvector $x \in X$.

De wiskunde versie van het één-stapsbeslissingsprobleem luidt nu als volgt: "Bepaal het maximium (minimum) van $k = K(x_1,x_2,\ldots,x_n)$ onder $x \in X$ en x_j = geheel als $j \in G$."

De studie gericht op het oplossen van problemen van dit type heet *mathematische programmering.* Als alle definiërende functies lineair zijn en de verzameling G leeg is, dan spreekt, men van *lineaire programmering* (lp). Vele toewijzings-, produktie-, transport-, maar ook financieringsproblemen laten zich vertalen als lp-problemen. Kortom, problemen waarin schaarse middelen als kapitaal, capaciteiten, arbeid etc. zodanig benut moeten worden dat de winst maximaal is of de kosten minimaal zijn. Zowel voor de algemene versie (de simplexmethode) als voor lp-problemen met een speciale structuur zijn algoritmen ontwikkeld, die problemen van grote omvang in een redelijke tijd kunnen oplossen. De wiskundige achtergrond van deze methoden wordt gevormd door de lineaire algebra. Immers voegen wij aan iedere ongelijkheid een verschilvariabele toe,

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \Leftrightarrow \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \text{ en } x_{n+i} \geqslant 0$$

$$\sum_{j=i}^{n} a_{ij}x_j \geqslant b_i \Leftrightarrow \sum_{j=i}^{n} a_{ij}x_j - x_{n+i} = b_i \text{ en } x_{n+i} \geqslant o,$$

dan wordt het lp-problem in matrixnotatie gegeven door:

$$\left.\begin{array}{r} \max z = CX \\ \text{onder } A,X = b \\ X \geqslant 0, \end{array}\right\} lp\,1$$

waarbij A een $m \times (n+s)$-matrix is wanneer de m voorwaarden een s-tal ongelijkheden bevatten
X is een vector; x_i = component van die vector
A,B,R zijn matrices

C is een vector	X_R is een vector
C_B is een vector	b is een vector
C_R is een vector	U is een vector
X_B is een vector	U_v is een vector

Laat een basismatrix B gevormd worden door m lineaire onafhankelijke kolommen uit A en laat R de overige kolommen voorstellen in A. Men kan eenvoudig nagaan dat

$$AX = b \Leftrightarrow B^{-1}RX_R + X_B = B^{-1}b$$

$$CX = C_B X_B + C_R X_R = C_B B^{-1}b + (C_R - C_B B^{-1}R)X_R,$$

waarbij

a) $X_B(X_R)$ die componenten uit X bevat welke corresponderen met kolommen uit $B(R)$

b) $C_B(C_R)$ die componenten uit C bevat welke corresponderen met componenten uit $X_B(X_R)$.

Uit bovenstaande volgt dat lp1 ook in de z.g.n. B-basisvorm geschreven kan worden:

$$\max Z = C_B B^{-1}b + (C_R - C_B B^{-1}R)X_R$$

onder

$$B^{-1}RX_R + X_B = B^{-1}b$$

$$(X_R, X_B) \geqslant 0$$

De oplossing

$$X_B = B^{-1}b \text{ en } X_R = 0$$

zullen wij een basisoplossing nemen.

Merk op dat een basisoplossing

1) toegelaten is als $B^{-1}b \geqslant 0$

2) optimaal is als $B^{-1}b \geqslant$ én $C_R - C_B B^{-1}R \leqslant 0$

3) de kriteriumwaarde $C_B B^{-1}b$ bezit

De oplossing van het lp-probleem kan o.a. worden verkregen door te starten met een B matrix waarvoor geldt: p dat als $C_R - C_B B^{-1}R \leqslant 0$ $B^{-1}b \geqslant 0$. In iedere volgende stap wordt één kolom uit B geruild voor een kolom uit R. De gekozen kolom uit R correspondeert met een hoogste positieve component in $C_R - C_B B^{-1}R$ (de simplexmethode). Merk op dat als $C_R - C_B B^{-1}R \leqslant 0$ de optimale oplossing reeds is gevonden. De kolom uit B, die moet plaats maken, volgt uit het verlangen dat

a) voor de nieuwe basis B ook geldt: $X_B = B^{-1}b \geqslant 0$ en

b) $(C_B B^{-1}b, C_B B^{-1}A - C)$ lexicografisch toeneemt.

Aan beide verlangens kan worden voldaan, zodat na een eindig aantal stappen, vanwege de eindigheid van het aantal B-matrices, een optimale basisoplossing wordt gevonden. Met lp1 is onverbrekelijk verbonden het *duale* probleem:

$$\min Z = b'u$$

onder

$$A'u \geqslant C^1$$

$$-\infty < u < +\infty$$

of met verschilvariabelen:

$$\left.\begin{array}{l} \min Z = b'u \text{ onder } A'u - U_v = c' \\ -\infty < u < +\infty \\ U_v \geqslant 0 \end{array}\right\} dlp\,1$$

De dualiteitstheorie van de lineaire programmering leert ons

1) het duale probleem van een basisvorm van het primale probleem (lp1) is een basisvorm van (dlp1) en dus
2) er is een 1-1-relatie tussen basis matrices van lp1 en dlp1.
3) er is een 1-1-relatie tussen basisoplossingen van lp1 en dpll.
4) de optimale oplossing van het primale probleem $(B^{-1}b,0)$ is gekoppeld aan de optimale oplossing van het duale probleem $(C_B B^{-1}, C_B B^{-1}A - C)$.
5) $\max Z = \min Z = C_B B^{-1} b$.

De hierboven geschetste methode levert ons niet alleen een optimale oplossing van lp1 maar tegelijkertijd ook één van dlp1. Uit punt 4) volgt van de optimale component u_i

$$u_i = \frac{\partial Z^*(b)}{\partial b_i} = \frac{\partial C_B B^{-1} b}{\partial b_i} = (C_B B^{-1})_i$$

waarbij $Z^*(b)$ de maximale waarde is van lp1 als functie van b. Conclusie: De oplossing van het primale probleem geeft antwoord op de gestelde vraag. De oplossing van het duale probleem geeft inzicht in de wijze waarop het gevonden resultaat afhangt van de rechterleden b_i van de primale voorwaarden. Als door een financieel offer het rechterlid b_i kan worden verhoogd tot $b_i + 1$, dan geeft u_i (marginaal) aan hoeveel de opbrengst toeneemt. De componenten u_i worden dan ook wel schaduwprijzen genoemd.

Lp-problemen van grote omvang kunnen in een redelijke tijd worden opgelost. Minder gunstig is de situatie als de verzameling G niet leeg is. De desbetreffende studie wordt *geheeltallige-* of *gemengde programmering* genoemd. Technieken (snedemethoden), die uitgaan van de algemene probleemstelling, zijn tot dusver niet zo succesvol geweest. Vandaar dat het onderzoek thans veel meer gericht is op specifieke structuren (branch and bound methoden). Bij de beschrijving van de beslissingssituaties kan dikwijls met succes gebruik gemaakt worden van begrippen uit netwerkanalyse en de grafentheorie. Daar vele beheers- en bedrijfsproblemen, waaronder (machine)volgorde-, vervoers-, indelings- en planningsproblemen, slechts vertaald kunnen worden in gemengde lp-problemen, is dit onderzoek zeer intensief. Ter afsluiting een tweetal voorbeelden van een geheeltallig programmeringsprobleem.

Voorbeeld 1 (een routeringsprobleem)

Gegeven een centraalmagazijn van waaruit $n-1$ vestigingen van een bedrijf worden bediend. Deze bediening geschiedt met behulp van m vrachtwagens, die het centraalmagazijn als vertrekpunt hebben en daarin ook terugkeren. De afstanden tussen de vestigingen worden gegeven door de afstandsmatrix C, waarvan de eerste rij en kolom betrekking hebben op het centraalmagazijn. De behoefte aan goederen (in tonnen) wordt van de vestiging i gegeven door q_i, terwijl vrachtwagen k een laadvermogen heeft van ϕ_k ton. Gevraagd voor iedere vrachtwagen een route te kiezen en wel zodanig dat het totaal af te leggen aantal kilometers van de vrachtwagens gezamelijk mimimaal is.

Formulering:

laat $x_{ijk} = 1$, als vrachtwagen k klant j bezoekt onmidddelijk na klant i;
$= 0$, anders.

laat $y_{ik} = 1$, als klant i wordt bezoekt door vrachtwagen k;
$= 0$, anders.

Wij beschouwen nu het volgende probleem

$$\min z = \sum_{ij} C_{ij} \sum_{k} x_{ijk} \tag{1}$$

onder

$$\sum_{k} y_{ik} = \begin{cases} 1, & i = 2,\ldots n \\ m, & i = 1 \end{cases} \tag{2}$$

$$\sum_{i} q_i y_{ik} \leqslant \phi_k, \quad k = 1, \ldots, m \tag{3}$$

$$\sum_{j} x_{ijk} = \sum_{j} x_{jik} = y_{ik}, \quad i = 1, \ldots, n; \; k = 1, \ldots, m \tag{4}$$

$$\sum_{i,j \in S} x_{ijk} \leqslant |S| - 1, \quad \forall S \subset \{2, \ldots, n\} \quad k = 1, \ldots, m \tag{5}$$

$$y_{ik} \in 0,1 \quad i = 1, \ldots, n; \; k = 1, \ldots, m \tag{6}$$

$$x_{ijk} \in 0,1 \quad i,j = 1, \ldots, n; \; k = 1, \ldots, m \tag{7}$$

De kriteriumfunctie (1) geeft het totaal af te leggen aantal kilometers aan. Voorwaarde (2) zorgt ervoor dat iedere klant één vrachtwagen op bezoek krijgt en dat er m vrachtwagen vertrekken uit het centraalmagazijn. Voorwaarde (3) dient om te voorkomen dat de vrachtwagens te zwaar worden beladen. Voorwaarde (4) zorgt ervoor dat vrachtwagen k vertrekt en aankomt in i, wanneer klant i op z'n route ligt. Voorwaarde (5) drukt uit dat er geen routes zijn zonder $i = 1$ (het centraalmagazijn).

Het behoeft geen betoog dat routeringsproblemen in de praktijk ingewikkelder zijn dat het hierboven geschetste. In ons tweede voorbeeld wordt gedemonstreerd hoe met behulp van begrippen uit de grafentheorie beslissingssituaties kunnen worden beschreven.

Voorbeeld 2 (een ontwerpprobleem)

Het architectenbureau Hut & Schuur heeft van het departement van Volkshuisvesting de opdracht gekregen een ontwerp te maken van een flatwoning dat maximaal te gemoet komt aan de wensen van de bewoner van vandaag. Om deze verlangens te leren kennen heeft Hut & Schuur een bureau voor opinieonderzoek ingeschakeld. In een onderzoek onder woningzoekenden wordt de ondervraagde een maquette getoond van een twee- kamerflat welke voldoet aan minimum eisen. Daarna wordt hem of haar verzocht de volgende items te ordenen in volgorde van belangrijkheid:

minimaal vloeroppervlak woonkamer $30m^2$	luxe keuken
maximale maandhuur f 600,-	balkon
centraal antennesysteem	tuin
fietsenbergplaats of hobbyruimte	garage
gemeenschapsruimte	geluidsisolering
een derde kamer	lift
luxe badkamer	openhaard

Hut & Schuur willen deze items ook zelf ordenen en wel zodanig dat de gekozen volgorde maximaal overeenkomt met de geopenbaarde verlangens. Laat C_{ij} ondervraagden de voorkeur geven aan j boven item i.

Wij kunnen het beslissinsprobleem van Hut & Schuur nu als volgt beschrijven:

Gegeven een volledige graaf met evenveel hoekpunten als items. Aan iedere kant (i,j) zijn twee gewichten toegekent C_{ij} en C_{ji} afhankelijk van de gekozen richting. Gevraagd een gerichte acyclische deelgraafen van maximaal gewicht.

3. Meer-stapsbeslissingsproblemen

Met enige goede wil kan men stellen dat in ieder meer-stapsbeslissingprobleem van de beslisser verwacht wordt dat hij een proces bestuurt. Zo'n proces, waarin kosten worden gemaakt en/of opbrengsten worden verkregen, speelt zich bijvoorbeeld af rond een voorraad, machine of rij wachtenden voor een loket, kortom rond een *systeem.* De verschillende toestanden waarin het systeem zich kan bevinden, laten zich wiskundig beschrijven met behulp van een (toestands)vector S. Wanneer de beslisser zich afzijdig houdt en desondanks het systeem in de loop der tijd van toestand verandert, zegt men dat het systeem onderworpen is aan een *natuurlijk proces.* Een natuurlijk proces wordt wiskundig gegeven door de (kansverdeling van de) toestanden op toekomstige tijdstippen. Een en ander zullen wij nader toelichten met het volgende beeld. Laat S de omvang (toestand) zijn van een voorraad (systeem) die door verkopen op ongeregelde tijdstippen afneemt (natuurlijk proces). Elke maandagmorgen (beslissingstijdstip) wordt nagegaan of de voorraad moet worden aangevuld en, zo ja, met hoeveel (beslissing). De maximum voorraad stelt een bovengrens aan de omvang van de bestelling (verzameling van toegelaten beslissingen). Uit bovenstaand beeld volgt dat de toelaatbaarheid van een beslissing x mede bepaald wordt door de toestand S van het systeem op het beslissingstijdstip. De verzameling van toegelaten beslissingen wordt derhalve aangeduid met $x(S)$. Beslissingen brengen in het algemeen ook toestandsveranderingen met zich mee. Het begrip toestand dient derhalve zo ruim gekozen te zijn, dat de nieuwe toestand na de beslissing kan worden aangegeven. Als in bovenstaand beeld de bestelling onmiddellijk wordt afgeleverd, is de nieuwe toestand wederom een (toegenomen) omvang van een voorraad. Is daarentegen aan een bestelling een levertijd verbonden, dan moet hoogst waarschijnlijk aan de toestand bovendien afgelezen kunnen worden hoeveel goederen nog op bestelling wachten en wellicht ook wanneer de desbetreffende orders zijn afgegeven. Ook vanuit deze toestanden als begintoestand moet het natuurlijk proces kunnen worden beschreven. Het natuurlijk proces "regelt" dan niet alleen de aankomst van de klanten maar ook de aflevering van de bestellingen. Uiteraard moet een beslissing in een toestand uiteindelijk worden beoordeeld op grond van zijn effect op bijvoorbeeld de toekomstige kosten. Dit effect kan evenwel niet onafhankelijk van toekomstige beslissingen worden vastgesteld. Bijgevolg wordt in een meer-stapsbeslisingsprobleem niet gezocht naar een enkele optimale beslissing x, maar naar een optimale strategie z. Zo'n strategie beeldt op ieder beslissingstijdstip de toestandsruimte $\mathbb{S}$ af op de verzameling X van beslissingen. Een strategie heet toegelaten als van iedere $S \in \mathbb{S}$ geldt: $z(S) \in X(S)$.

Stel dat in ons voorraadprobleem de bestellingen direct worden afgeleverd zodat wij kunnen volstaan met een toestand die slechts de omvang van de voorraad aangeeft. Een strategie van het volgende type ligt dan voor de hand

$$x = z(S) = \begin{cases} 0 \text{ als } S > m \\ M - S \text{ als } S \leqslant m \end{cases}$$

waarbij M niet groter is dan de maximale voorraadcapaciteit. Merk op dat zo'n strategie geheel wordt bepaald door de keuze van (m,M).

Het is duidelijk dat zodra een strategie wordt toegepast het natuurlijke proces vanwege de extra toestandsveranderingen niet meer geëigend is om de ontwikkelingen in de toestand van het systeem te beschrijven. Deze taak wordt nu overgenomen door het *beslissingsproces* dat voor de te beschouwen strategieën en elke begintoestand gedefinieërd moet kunnen worden.

Voor het bepalen van een optimale strategie dient men te beschikken over een optimaliteitskriterium. Laat $K_\alpha(S,z,T)$ de verwachte verdisconteerde totale kosten (opbrengst) zijn als vanuit de toestand S de strategie z gedurende een periode T wordt toegepast. De verdiscintering met $\alpha \leqslant 1$ is dikwijls ook wiskundig noodzakelijk om $K_\alpha(S,z,T)$ voor grote waarden van T begrensd te houden. Indien verdiscontering niet reeël is, dan beschikt men over het alternatieve kriterium ($\alpha = 1$):

$$y(S:z) = \lim_{T\to\infty} \frac{K_1(S,z,T)}{T},$$

de gemiddelde kosten (opbrengst) per tijdseenheid. In praktische problemen is $y(S;z)$ constant op $\mathbb{S}$ en schrijven wij dus $y(z)$ De bepaling van de gemiddelde kosten $y(z)$ geschiedt evenwel simultaan met een grootheid $v(S:z)$

$$v(S:z) = \lim_{T\to\infty} \{K_1(S,z,T) - y(z)T\} \quad (S \in \mathbb{S}).$$

die wel een waardering toekent aan de begintoestand S.

Keren wij terug tot ons voorraad probleem. Stel dat $k(S,s)$ de verwachte kosten zijn in een week die begint met een voorraad S en een bestelling van de omvang x. Als $f(y)$ de kansdichtheid voorstelt van de behoefte y in die week, dan moet $K_\alpha(S,z,\infty)$ voldoen aan

$$K_\alpha(S,z,\infty) = k(S,z(S)) + \alpha \int_0^\infty K(S+z(S)-y,z,\infty) f(y)dy.$$

De optimale strategie z^* moet derhalve voldoen aan de optimaliteits vergelijking:

$$K_\alpha(S,z^*,\infty) = \min \{k(S,x) + \alpha \int_0^\infty K_\alpha(S+x-y,z^*\infty) f(y)dy \mid x \in x(S)\}$$

Ook voor de functies $y(z)$ en $v(S:z)$ bestaan dergelijke functinaal vergelijkingen.

Een veel gebruikte methode voor het bepalen van een optimale strategie is de z.g.n strategie-verbeteringsmethode. Deze methode stelt voor iedere strategie z vast of z optimaal is, en zo niet dan wijst zij een strategie aan die beter is. Op deze wijze ontstaat een rij van strategieën $\{z_n; n = 1,2,\cdots\}$ die onder zekere voorwaarden convergeert naar een optimale strategie z^*.

Stel dat in ons voorraadprobleem in stap n van de

strategieverbeteringmethode de strategie z_n was gevonden. De stap $n+1$ verloopt dan als volgt:

1) Los op de functionaal vergelijking

$$K_x(S;z_n,\infty) = k(S,z_n(S))+\alpha \int_0^\infty K_\alpha(S+z_n(S)-y,z_n,\infty)f(y)dy.$$

2) Bepaal voor iedere S een beslissing $x \in X(S)$ waarvoor bovenstaande uitdrukking minimaal is. Als $z_n(S)$ zo'n beslissing is, kies dan $x = z_n(S)$. De strategie z_{n+1} voor de volgende ronde wordt gegeven door de op deze wijze verkregen afbeelding van $\mathbb{S}$ op X. Als $z_{n+1}(S) = z_n(S)$ voor $S \in \mathbb{S}$ dan is $z_n = z_{n+1}$ optimaal.

De studie, die gericht is op het oplossen van meer-stapsbeslissingsproblemen, heet *dynamische programmering.* Een groot aantal voorraad-, vervangings- en produktieproblemen kan met behulp van een dynamisch programmeringstechniek worden aangepakt.

Voorbeeld 3 (een schadeprobleem)

Een automobilist heeft een schadeverzekering afgesloten. In de bijbehorende polis worden o.a. de volgende voorwaarden vermeld:

1. De looptijd van de verzekering is één jaar. Aan het eind van ieder jaar kan zij worden verlengd. De premie moet aan het begin van ieder premiejaar worden voldaan.
2. De premie bedraagt f 320,--, tenzij
   a. in de voorafgaande periode van één jaar geen schade is geclaimd. In dat geval bedraagt de premie f 280,--, tenzij
   b. in de voorafgaande periode van twee jaar geen schade is geclaimd. In dat geval bedraagt de premie f 240,--, tenzij
   c. in de voorafgaande periode van drie jaar geen schade is geclaimd. In dat geval bedraagt de premie f 220,--.
3. Indien men een schade wil claimen dient dit onmiddellijk te geschieden. Slechts het verschil tussen de schade en een vast bedrag van f 80,--, het zg. eigen risico, wordt door de verzekering uitbetaald.

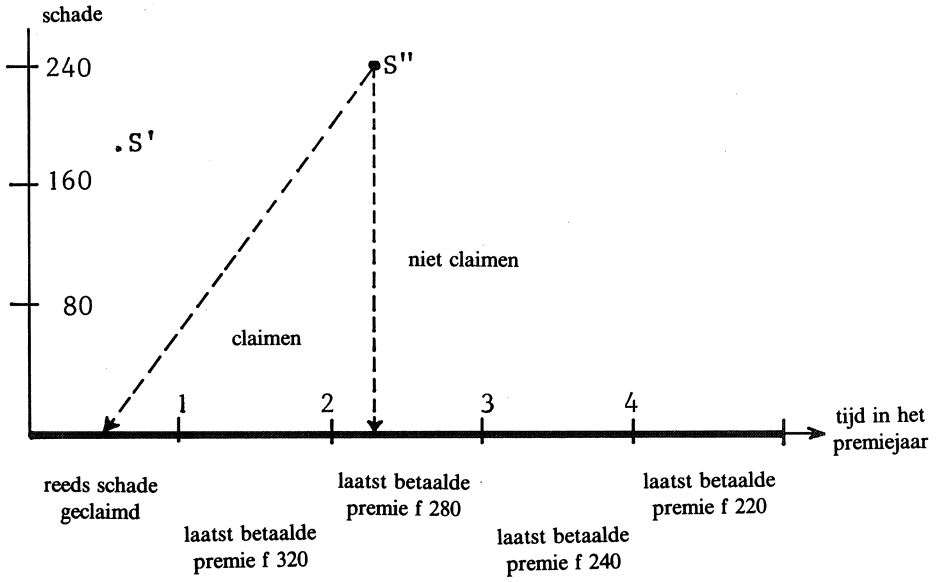
Stel dat de automobilist zijn gemiddelde kosten per tijdseenheid in de long run wil minimaliseren. Gevraagd wordt nu voor elk tijdstip aan te geven welke schaden geclaimd moeten worden.

Het is duidelijk dat de automobilist nooit een schade van minder dan f80,-- zal claimen. Het is ook duidelijk, dat hij, als nog geen schade is geclaimd dat jaar, met het oog op de premie-reducties voor schadevrij rijden geen schaden zal claimen, welke slechts een weinig hoger zijn dan het eigen risico. De vraag is nu waar precies de grens ligt tussen de schaden die wel en die niet moeten worden geclaimd. Het behoeft geen betoog, dat de grenswaarden zullen afhangen van de hoogte van de laatst betaalde premie en van het tijdstip van

de schade in het premiejaar. In dit voorbeeld wordt de toestand van het systeem bepaald door:

1. de laatst betaald premie;
2. het tijdstip in het premiejaar;
3. de eventueel te claimen schade;
4. de omstandigheid of er al eerder in het premiejaar een schade is geclaimd of niet.

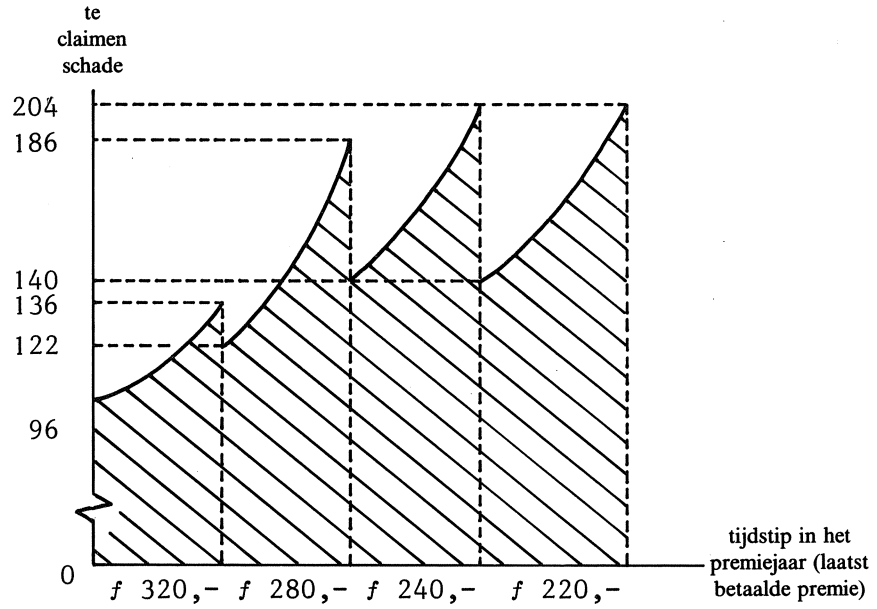
In onderstaande figuur is een afbeelding gegeven van de toestandsruimte. Op de horizontale as zijn 5 intervallen van één jaar aangegeven. Het eerste interval heeft betrekking op een jaar waarin reeds een schade werd geclaimd. Het punt S' duidt aan dat 7 maanden na de laatste premiebetaling



schadevrij gereden sinds laatste premie betaling

wederom een schade plaatsvindt en wel van f 165,--. In 1. zal opnieuw de hoogste premie moeten worden betaald. Premiebetalingen geschieden ook in 2., 3. en 4. Het punt S'' spreekt van een schade van f 240,-- 4 maanden na een premiebetaling van f 280,--, terwijl nog geen schade was geclaimd.

In de volgende figuur hebben wij een strategie aangegeven, die de beslisser adviseert *geen* schade te claimen wanneer het systeem vanwege die schade een toestand aanneeemt in het gearceerde gebied, tenzij reeds eerder in het premiejaar een schade is geclaimd. Uit het voorgaande volgt dat de oplossing van het beslissingsprobleem besliskundig gezien, een keuze is uit de verzameling van alle mogelijke gearceerde gebieden. In dit voorbeeld wordt aangenomen dat tijdens het natuurlijk proces alle schaden worden geclaimd en geen premie wordt betaald. In het natuurlijk proces komt men aan het eind van het premiejaar in de absorptietoestand (0); de schaden zijn niet meer gedekt. Een beslissing is òf de betaling van een premie òf het *niet* melden van een schade.



Als de schade S'' niet gemeld wordt dan gaat de toestand terug naar het overeenkomstige punt op de tijdas. Wordt daarentegen de schade wel geclaimd dan vindt een toestandstransformatie plaats naar het corresponderende punt in het eerste tijdsinterval (zie eerste figuur).

Indien men aanneemt dat het rijgedrag van de automobilist niet wordt beinvloed door schaden en premiebetalingen, dan kan het beslissingsproces van tal van schadeverdelingen worden gedefinieerd. De oplossing van het beslissingsprobleem wordt op een verrassend eenvoudige wijze verkregen.

CAPACITEITSMODELLEN BIJ EEN FLUKTUERENDE VRAAG

D.K. Leegwater

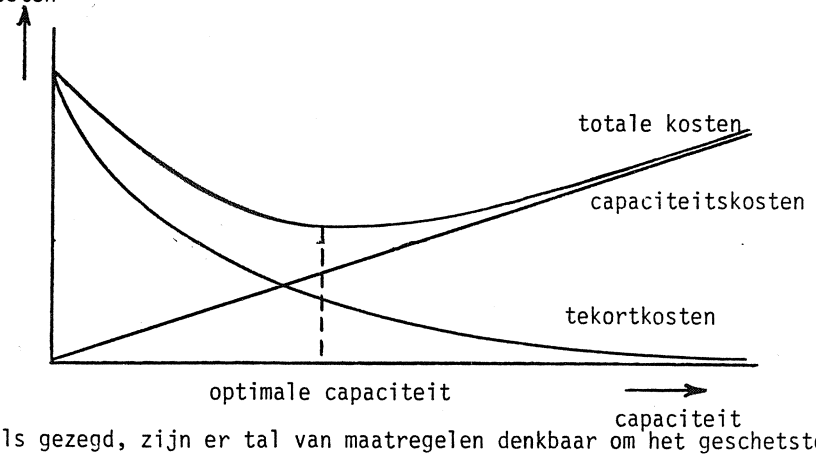
AKB (Adviseurs voor Kwantitatieve
Modellen en Bedrijfsinformatica
Erasmus Universiteit Rotterdam.

1. INLEIDING

Van oudsher is het verschijnsel van een fluktuerende vraag naar capaciteit een economisch probleem.
De ernst daarvan, uitgedrukt in kosten, c.q. opbrengstderving, is in de loop der tijden echter veranderd en dan meestal in ongunstige zin.
Met name is dat waar te nemen als het gaat om fluktuerende vraag naar arbeidscapaciteit.
Het gebruik van dagloners was vroeger een ingeburgerd middel om het aanbod van capaciteit op de vraag af te stemmen. Naast het veelvuldig gebruik van dagloners op landbouwbedrijven is vermeldenswaard het dagelijks werven van havenarbeiders, die elke dag bij de poorten van de havenbedrijven of in kroegen wachtten op werk voor één dag.
In 1889 opperde W.M. Pieters het denkbeeld in Rotterdam een corps bootwerkers op te richten om een eind te maken aan het dagelijks werven. Dit korps zou voor gezamenlijke verantwoordelijkheid en rekening van de patroons (werkgevers) dienen te komen. Er zou een garantie komen van een vast weekloon.
Het plan is toentertijd echter op dusdanig verzet gestuit dat het niet in praktijk werd gebracht. Sindsdien is men stap voor stap gekomen tot een vast dienstverband voor alle havenarbeiders, dat in 1955 tot stand werd gebracht door oprichting van een gemeenschappelijke arbeidsreserve - de CVA (Centrale voor Arbeidsvoorziening), die arbeidskrachten in vaste dienst had voor de opvang van de (fluktuerende) werkdrukte voor de havenbedrijven boven de havenarbeiders, die de bedrijven zelf in vaste dienst hadden.
De huidige organisatie (arbeidspool) opereert onder de naam SHB (Stichting Samenwerkende Haven Bedrijven).

Het oprichten van een arbeidspool of meer algemeen een pool van produktiemiddelen is echter lang niet de enig mogelijke maatregel om het probleem van een fluktuerende vraag naar de capaciteit aan te pakken. Alvorens hierop in te gaan wordt het economisch aspect van het probleem aan de hand van onderstaande grafiek nader toegelicht.
In deze grafiek zijn de kosteneffekten geïllustreerd van een toenemende vaste capaciteit van een bedrijf bij een fluktuerende vraag. Enerzijds nemen de capaciteitskosten (en hiermee de leegloopkosten) toe, anderzijds nemen de kosten ten gevolge van capaciteitstekort (bijvoorbeeld opbrengstderving) af. De optimale capaciteit ligt bij het minimumvan de som van beide kosten.

Figuur 1 : kostenverloop bij toenemende vaste capaciteit voor een bedrijf



Zoals gezegd, zijn er tal van maatregelen denkbaar om het geschetste probleem aan te pakken. Hierbij kan men onderscheiden individuele maatregelen dwz maatregelen die een bedrijf zelfstandig kan nemen en kollektieve maatregelen dwz maatregelen die een bedrijf slechts in samenwerking met een of meer andere bedrijven kan nemen.
Daarnaast zijn er enerzijds maatregelen aan de vraagkant en anderzijds maatregelen aan de aanbodkant mogelijk.

In onderstaand overzicht is een, zij het niet uitputtende,opsomming gegeven van een aantal maatregelen met voorbeelden en de relevante kenmerken.

Maatregel	Kollektief (K) Individueel(I)	Aanbodzijde(A) Vraagzijde (V)
1. Prijsdifferentiatie		
Vbd: Hogere electriciteitstarieven tijdens piekuren	I	V
2. Afspraakregeling		
Vbd: Opnameplanning ziekenhuizen (uitgezonderd spoedgevallen)	I	V
3. Vooruitwerken	I	V
Vbd: op voorraad produceren		
4. Vakantiespreiding	I	A
Vbd: Stimuleren met extra vakantiegeld van werknemers om in de winter op vakantie te gaan		
5. Gebruik van "all-round" produktiemiddelen	I	A
Vbd: een Boeing 747M voor vervoer van passagiers en vracht tegelijkertijd.		
6. Gebruik van parttime arbeidskrachten	I	A
Vbd: Schoonmaakploegen		
7. Overwerk	I	A
8. Inschakelen externe produktiecapaciteit	I	A
Vbd: huren van vorktrucks, gebruik van uitzendkrachten		
9. Onderling uitwisselen van produktiecapaciteit tussen bedrijven	K	A
Vbd: machineringen in de land- en tuinbouw		
10. Instellen van een pool van produktiemiddelen	K	A
Vbd: havenarbeidspool		

Maatregel	Kollektief (K) Individueel (I)	Aanbodzijde (A) Vraagzijde (V)
11. Onderling uitwisselen van de vraag tussen bedrijven Vbd: hotels die klanten naar elkaar verwijzen	K	V
12. Instellen van een pool van de vraag voor een aantal bedrijven Vbd: het poolen van een trajekt bijv. Amsterdam- Londen tussen luchtvaartmaatschappijen.	K	V

Drie vraagstukken dienen in feite te worden opgelost.

1. Het vaststellen welke maatregelen in aanmerking komen.
2. Het kwantificeren van de in aanmerking komende maatregelen, hetgeen inhoudt het vaststellen van de bijbehorende kosten cq opbrengsten.
3. Het vaststellen welke van de maatregelen gekozen dient te worden bijvoorbeeld door vergelijking van de verwachte kosten.

Het eerste vraagstuk houdt in een globale analyse van de praktijksituatie. Bijvoorbeeld : het overwegen van het gebruik van parttimearbeid is zinvol indien een significant patroon in de werkdrukte over de dagen van de week of over de uren binnen een dag aanwezig is; het inhuren van uitzendkrachten is het overwegen waard in een situatie van onverwachte ziekte van het vaste personeel of voor tijdelijke werkzaamheden.
Het tweede en derde vraagstuk vormen in deze paper de hoofdmoot. Met behulp van wiskundige modellen worden oplossingsmethoden aangegeven (hoofdstuk 2 voor een aantal individuele maatregelen en hoofdstuk 3 voor een aantal kollektieve) gevolgd door een aantal toepassingen (hoofdstuk 4). De paper wordt afgesloten met een aantal opgaven en de hierbij behorende oplossingen.

Deze inleiding eindigt met een opmerking over fluktuaties.Dezelfde economische problemen van enerzijds leegloop bij overschot aan capaciteit en anderzijds tekort doen zich in feite eveneens voor bij fluktuaties van het aanbod van capaciteit. Tengevolge van storingen, ongevallen,ziekte e.d. is de beschikbaarheid van de eigen capaciteit van een bedrijf of instelling niet constant.

Een sprekend voorbeeld vindt men in het onderwijs, waar de vraag naar leerkrachten in de tijd gezien constant is volgens een van te voren opgesteld rooster, terwijl het de beschikbare capaciteit tengevolge van ziekte, ongeval,wettelijk verzuim fluktueert.
Het gebruik van vervangende leerkrachten van onderwijzerspools (bijvoorbeeld in Rotterdam) is een van de maatregelen voor het basisonderwijs, die het bovengenoemde probleem moeten opvangen.
In het algemeen zal men zowel met een fluktuerende vraag als met een fluktuerend aanbod te maken hebben. In onderstaand overzicht staan een drietal bedrijven vermeld met een kwalitatieve aanduiding van de mate waarin genoemde genoemde fluktuaties optreden.

Bedrijf	Fluktuatie vraag	Fluktuatie aanbod
Produktiebedrijf	afwezig	zwak
Bouwbedrijf	zwak	zwak
Havenbedrijf	sterk	zwak

2. CAPACITEITSMODELLEN VOOR INDIVIDUELE MAATREGELEN.

In dit hoofdstuk worden een aantal capacitieitsmodellen behandeld, waarmee een deel van de in de inleiding genoemde individuele maatregelen kunnen worden ondersteund.

Als basissituatie (situatie 0) geldt de volgende:
we beschouwen een bedrijf met een aantal fulltime werknemers in vaste dienst (X) en met een normaal verdeelde vraag $\underline{w}$ per dag (uitgedrukt in mandagen werk). De gemiddelde vraag is μ en de standaardafwijking is σ. Tekort aan mankracht leidt tot omzetverlies ten bedrage van c per mandag tekort. We veronderstellen verder dat het percentage van de werknemers, die beschikbaar zijn voor werk, konstant is en gelijk aan p.
De vaste kosten voor een werknemer in vaste dienst bedragen a per mandag.
De variabele kosten voor het verrichten van een mandagwerk bedragen b.

Het probleem is X zo te bepalen dat de verwachte kosten per dag voor het desbetreffende bedrijf minimaal zijn.

De verwachte kosten per dag $\underline{EK}$ zijn als volgt te formuleren :

$$\underline{EK} = aX + b\mu + (c-b)\, E\,(\underline{w} - pX)^+$$

waarbij $(\underline{w} - pX)^+$ het omzetverlies is met

$$(\underline{w} - pX)^+ \begin{cases} \underline{w} - pX & \text{als } \underline{w} > pX \\ 0 & \text{als } \underline{w} \leq pX \end{cases}$$

en $E\,(\underline{w} - pX)^+$ de verwachting ervan.

Een andere schrijfwijze is :

$E(\underline{w} - pX)^+ = \int_{pX}^{\infty} (w - pX)\, f(w)\, dw$, waarbij $f(w)$ de kansdichtheid is van de (normaal verdeelde) grootheid $\underline{w}$.

Het probleem is nu te omschrijven als :

$$\min_{X} \; E\underline{K}$$

Dit lossen we op door de afgeleide van $E\underline{K}$ naar X gelijk te stellen aan nul en X hieruit op te lossen.
Hiervoor dienen we de afgeleide van een integraal te kennen.
De (algemene) formule hiervoor is :

$$\frac{\partial}{\partial X} \int_{g_1(X)}^{g_2(X)} h(x,X)\,dx = \int_{g_1(X)}^{g_2(X)} \frac{\partial h(x,X)}{\partial X}\,dx + h(g_2(X),X)\,\frac{\partial g_2(X)}{\partial X} - h(g_1(X),X)\,\frac{\partial g_1(X)}{\partial X}$$

Toegepast op $E\,(\underline{w} - pX)^+$ geeft dit :

$$\frac{\partial E\,(\underline{w} - pX)^+}{\partial X} = \int_{pX}^{\infty} -pf(w)dw \quad , \text{ waarbij de overige}$$

termen wegvallen hetzij door $f(\infty) = o$, hetzij door $pX - pX = o$

Resultaat :

$$\frac{dE\underline{K}}{dX} = a - (c-b)p \int_{pX}^{\infty} f(w)dw$$

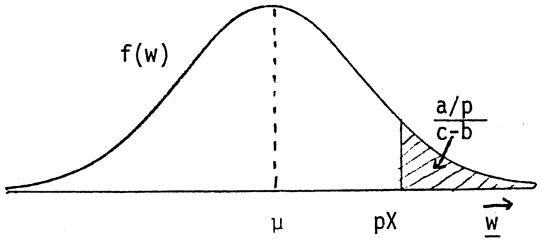
$\frac{dE\underline{k}}{dX} = o$ levert op dat de optimale waarde van

X gevonden kan worden uit:

$$\int_{pX}^{\infty} f(w)\,dw = \frac{a/p}{c-b}$$

Ofwel X is optimaal als de hierbij behorende kans op tekort aan eigen mankracht gelijk is aan de kostenverhouding $\frac{a/p}{c-b}$

Onderstaande figuur dient ter illustratie



N.B. Een meer economisch geaarde en tevens eenvoudiger afleiding van de optimale waarde van X is de volgende
Stel, we verhogen de beschikbare capaciteit (p^X) met 1, dan zijn:

$$\text{marginale kosten} = 1.\ a/p$$

$$\text{marginale baten} = (c-b)\int_{pX}^{\infty} f(w)dw$$

We zullen nu pX steeds dienen te verhogen, indien de marginale baten groter zijn dan de marginale kosten.
Ofwel het optimum wordt bereikt indien geldt :

$$\text{marginale kosten} = \text{marginale baten},$$

waaruit bovenstaande formule volgt.

De minimale verwachte kosten min $\underline{EK}$ zijn nu :

$$\min \underline{EK} = aX + b\mu + (c-b)\int_{pX}^{\infty} wf(w)dw - (c-b)pX\int_{pX}^{\infty} f(w)dw$$

$$= aX + b\mu + (c-b)\int_{pX}^{\infty} wf(w)dw - (c-b)pX\frac{a/p}{c-b} =$$

$$= \boxed{(c-b)\int_{pX}^{\infty} wf(w)dw + b\mu}$$

Nu is voor de normale verdeling gemakkelijk af te leiden dat

$$\int_{pX}^{\infty} wf(w)dw = \sigma \frac{\exp\left\{-\frac{1}{2}\left(\frac{pX-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}} + \mu \int_{pX}^{\infty} f(w)dw$$

Numeriek voorbeeld :

Stel a = 200 b = 0,05a/p c =2a/p μ = 100 σ= 30 en p = 0.80

Dan zijn de resultaten X = 125 en min EK = 16041.

In plaats van de basissituatie beschouwen we nu de volgende situaties:

1. Het bedrijf huurt externe mankracht in voor een bedrag van c = 1,8a/p per mandag (zie maatregel 8, inleiding)
2. Het bedrijf laat overwerk verrichten voor de eigen werknemers met als kosten e = 1,5a/p per mandag.
   De maximale hoeveelheid overwerk bedraagt echter γpX per dag met γ = 0,125(zie maatregel 7 inleiding)
3. Het bedrijf reduceert de fluktuatie in de vraag met 20 procent tegen kosten R per mandag (zie maatregel 2 inleiding)

De bijbehorende modellen zijn :

1. $EK = aX + b\mu + c\,E(w - pX)^+$

   met als oplossing voor de optimale waarde van X :

$$\int_{pX}^{\infty} f(w)dw = \frac{a/p}{c}$$

2. $EK = aX + b\mu + (e-b)\left\{E(w - pX)^+ - E(w - (1+\gamma)pX)^+\right\}$
   $+(c-b)\,E(w - (1+\gamma)pX)^+$

met als oplossing voor de optimale waarde van X :

$$\left(1 - \frac{c-e}{c-b}\right)\int_{pX}^{\infty} f(w)dw + \left(\frac{c-e}{c-b}\right)(1+\gamma)\int_{(1+\gamma)pX}^{\infty} f(w)\,dw = \frac{a/p}{c-b}$$

3. $E\underline{K} = R\mu + aX + b\mu + (c-b)\ E\ (\underline{w} - pX)^{+}$

De oplossing voor de optimale waarde van X wordt gegeven door

$$\int_{pX}^{\infty} f(w)\ dw = \frac{a/p}{c-b}$$ (zie basissituatie).

In onderstaande tabel staan de resultaten vermeld, waarbij de verwachte kosten voor de basissituatie op 100 zijn gesteld. De vaste kosten zijn hierbuiten gelaten, doch worden in een aparte kolom vermeld. Bovendien zijn de resultaten berekend voor de combinatie van situatie 3 (reduktie vraagfluktuatie) en de overige situaties.

Situtatie	optimale eigen capaciteit	minimale kosten	toe te voegen vaste kosten
Basissituatie	125	100	--
1.(huren externe capaciteit)	120	98,4	--
2.(overwerk)	121	97,9	--
3.(reduktie vraagfluktuatie)	125	96,4	Rμ
1,3	121	95,1	Rμ
2,3	121	94,3	Rμ

Tabel 1. Overzicht resultaten van een aantal individuele maatregelen.

3. CAPACITEITSMODELLEN VOOR KOLLEKTIEVE MAATREGELEN.

Van de in de inleiding genoemde maatregelen zullen de volgende situaties eveneens voor de produktiefaktor arbeid nader worden uitgewerkt:

4. Onderling uitwisselen van produktiecapaciteit tussen bedrijven (maatregel 9)
5. Instellen van een pool van produktiemiddelen (maatregel 10).
6. Instellen van een pool van de vraag voor een aantal bedrijven (maatregel 12)

We veronderstellen dat we te maken hebben met identieke bedrijven hoewel ook forumules voor de situatie van niet-identieke bedrijven zijn afgeleid en voor de praktijk een hanteerbare vorm hebben. Identiek houdt in dat elk bedrijf dezelfde vraagverdeling heeft, dat geen korrelatie tussen de vraag van het ene bedrijf en dat van het andere bestaat, dat het beschikbaarheidspercentage van de eigen capaciteit voor elk bedrijf gelijk is evenals de kostenfaktoren.

Wij voeren de volgende grootheden in :

n = aantal bedrijven

X = capaciteit per bedrijf

X_o = capaciteit van de pool

p = beschikbaarheidsfraktie capaciteit per bedrijf

p_o = beschikbaarheidsfraktie, poolcapaciteit

$\underline{w}_i$ = werkdrukte voor bedrijf i, normaal verdeeld met gemiddelde μ en standaardafwijking σ

W = totale werkdrukte per dag

μ_v = vraag van bedrijf i aan de pool met gemiddelde μv en standaardafwijking σ_v

$\underline{V}$ = totale vraag aan de pool met gemiddelde μ_V en standaardafwijking σ_V

a = variabele capaciteitskosten per bedrijf voor een mandag

a_o = variabele capaciteitskosten pool voor een mandag

C = vaste kosten pool (vaste overhead) per dag

b = variabele kosten bij het inzetten van een mandag

c = kosten tgv capaciteitstekort per mandag

d = variabele kosten van het uitwisselen van mankracht per mandag

D = vaste kosten van het uitwisselen van mankracht per dag

t = kosten van het toewijzen van vraag aan een bedrijf per mandag, in het geval van een pool van de vraag

T = vaste kosten per dag van een pool van de vraag.

De te gebruiken modellen zijn :

4. $E\underline{K} = \sum_{i=1}^{n} aX + b\sum_{i=1}^{n} \mu + dE\underline{r} + cE\underline{s}$,

waarbij $E\underline{r} = \sum_{i=1}^{n} E(\underline{w}_i - pX)^+ - E(\sum_{i=1}^{n} \underline{w}_i - \sum_{i=1}^{n} pX)^+$

$= nE(\underline{w} - npX)^+ + E(\underline{W} - npX)^+$

(de verwachte hoeveelheid uit te wisselen mankracht per dag)

en $E\underline{s} = E(\underline{W} - npX)^+$ (de verwachte hoeveelheid tekort per dag)

We krijgen nu :

$E\underline{K} = naX + nb\mu + dnE(\underline{w} - pX)^+ + (c-d)E(\underline{W} - npX)^+$

Door de afgeleide van $E\underline{K}$ naar X nul te stellen vinden we voor de optimale oplossing:

$$P(\underline{w} \geq pX) + \frac{c-d}{d} P(\underline{W} \geq npX) = \frac{a/p}{d}$$

waarbij $\underline{w}$ de werkdrukte voor een bedrijf is en normaal verdeeld is met gemiddelde $n\mu$ en standaard-afwijking $\sigma\sqrt{n}$

5. $E\underline{K} = \sum_{i=1}^{n} aX + b\sum_{i=1}^{n} \mu + C + a_o X_o + c\,E(\underline{V} - P_o X_o)^+$

Hierbij is $\underline{V} = \sum_{i=1}^{n} \underline{v}i = \sum_{i=1}^{n} (\underline{w}_i - pX)^+$ een bij benadering (centrale limietstelling) normaal verdeelde grootheid met gemiddelde $\mu_V = nE(\underline{w} - pX)^+$ en standaardafwijking $\sigma_V = \sigma_v\sqrt{n}$.

Door zowel de afgeleide van $\underline{EK}$ naar X als naar X_o nul te stellen vinden we :

$$P(\underline{V} \geq P_o X_o) = \frac{a_o/p_o}{c} \qquad (1)$$

$$\text{en } P(\underline{w} \geq pX) = 1 - \frac{a_o/p_o - a/p}{a_o/p_o - c\,\exp(-\frac{1}{2}u_o^2)\,\frac{\mu_v}{\sigma_v\sqrt{n}}} \qquad (2)$$

$$\text{waarbij } u_o = \frac{p_o X_o - n\mu_v}{\sigma_v \sqrt{n}} \qquad (3)$$

We vinden X en X_o als volgt :

Stap 1 Bepaal u_o uit (1)

Stap 2 Bepaal X uit (2)

Stap 3 Bepaal μ_v en σ_v gegeven X en bepaal vervolgens X_o uit (3)

$$6.\ \underline{EK} = \sum_{i=1}^{n} aX + b\sum_{i=1}^{n}\mu + T + t\sum_{i=1}^{n}\mu + c\,E\,(\underline{W} - npX)^+$$

$$= naX + n(b+t)\mu + T + CE(\underline{W} - npX)^+$$

Overeenkomstig het allereerste model vinden we de optimale X uit :

$$P(\underline{W} \geq npX) = \frac{a/p}{c}$$

Numerieke resultaten.

We beschouwen de volgende gegevens :

$\mu = 100$	$\sigma = 30$	$p = 0{,}8$	$p_o = 0{,}8$
$a = 200$	$a_o = 240$	$b = 12{,}5$	$c = 450$
$d = 25$	$t = 12{,}5$		

In onderstaande tabel staan de resultaten vermeld, waarbij tevens die uit hoofdstuk 2 zijn opgenomen.
Wederom zijn de verwachte kosten voor de basissituatie op 100 gesteld.
Ook hier is bovendien de combinatie met situatie 3(maatregel 7 : reduktie vraagfluktuatie) doorgerekend.
Het aantal bedrijven is gesteld op 5.

Situatie	Optimale capaciteit bedrijven	pool	Minimale kosten	Toe te voegen vaste kosten
Basissituatie	625	-	100	-
1(huren externe capaciteit)	600	-	98,4	-
2(overwerk)	605	-	97,9	-
3(reduktie vraag fluk.)	625	-	96,4	$5R_\mu$
4(uitwisselen capaciteit)	612	-	89,5	D
5(capaciteits-pool)	487	143	95,3	C
6(vraagpool)	620	-	93,6	T
1,3	605	-	95,1	$5R_\mu$
2,3	605	-	94,3	$5R_\mu$
4,3	612	-	88,0	$D+5R_\mu$
5,3	519	111	92,6	$C+5R_\mu$
6,3	620	-	92,0	$T+5R_\mu$

Tabel 2 Resultatenoverzicht individuele en kollektieve maatregelen.

Afhankelijk van o.a. de toe te voegen vaste kosten (in standhouden apparaat, computerkosten etc.) zal de uiteindelijke keuze welke maatregel economisch het meest zinvol is bepaald dienen te worden.
Het zal duidelijk zijn dat de economische betekenis van de situaties 4,5 en 6 zal afhangen van het aantal deelnemende bedrijven.

4. TOEPASSINGEN

Zoals uit de inleiding blijkt hebben we met name te maken met beslissingen over de grootte (dus capaciteit) van produktiemiddelen op middellange (maand tot een jaar) of lange (jaren) termijn.
Capaciteitsproblemen hebben economisch gezien te maken met de vaste kosten in de kostprijsopbouw van produkten en diensten, waarbij een fluktuerende vraag de problematiek extra accent geeft.
Zoals in de inleiding reeds is geschetst hebben de vaste kosten tgv de sociale maar ook technische ontwikkeling een steeds groter aandeel gekregen, waardoor de capaciteitsproblemen dan ook hoe langer hoe belangrijker zijn geworden.
Het is een taak van het management om in te spelen op deze problematiek, die in feite inhoudt een verschuiving van variabele kosten naar vaste.
We zien dan ook dat nieuwe maatregelen ontstaan zoals het oprichten van pools (haven, onderwijs,bouw, metaal) inclusief onderlinge uitzendbureaus (ziekenhuizen) en het onderling uitwisselen van capaciteit (machineringen).
De modellen zoals in de vorige hoofdstukken zijn beschreven hebben de rol van beslissingsondersteunend instrument om te zorgen voor een optimale inhoud van de maatregel, met name als het gaat om het bepalen van de optimale capaciteit.
Een tweetal toepassingen zullen hieronder in het kort worden behandeld.

Toepassing 1. Bepaling optimale omvang personeelsbestanden in de stukgoedsektor van de haven van Rotterdam.

We beschouwen de volgende situatie :

12 bedrijven in 4 achtereenvolgende kwartalen met elk bedrijf een in principe verschillende normale verdeling van de werkdrukte, die van kwartaal tot kwartaal kan verschillen. Daarnaast is een arbeidspool aanwezig.

De voorbeeld gegevens zijn :

a) Kostenfaktoren (zie hoofdstuk 3 voor de betekenis)
Waarde van a voor de bedrijven.

bedrijf	1	2	3	4	5	6	7	8	9	10	11	12
a	153	154	155	151	153	151	154	154	151	150	154	156

b = 1,14 c= 427 a_0 = 170 C = 2390

Daarnaast wordt van overheidswege per mandag leegloop van de pool een bijdrage gegeven ter grootte van 53.5

b) Werkdrukte en beschikbaarheidsfraktie.

Bedrijf	periode 1			periode 2			periode 3			periode 4		
	μ	σ	p	μ	σ	p	μ	σ	p	μ	σ	p
1	455	65	0.76	575	62	0.74	520	60	0.68	425	95	0.73
2	560	70	0.77	470	50	0.75	405	60	0.69	550	105	0.74
3	290	40	0.75	340	41	0.77	325	41	0.70	295	55	0.75
4	340	45	0.74	305	35	0.75	240	32	0.66	350	60	0.72
5	245	42	0.75	265	40	0.77	245	35	0.68	205	42	0.72
6	210	35	0.73	245	32	0.75	185	31	0.65	240	59	0.73
7	190	50	0.76	200	51	0.76	190	45	0.69	195	73	0.74
8	170	14	0.76	175	13	0.75	165	14	0.69	170	19	0.74
9	125	20	0.74	125	21	0.74	105	19	0.67	115	24	0.72
10	110	30	0.72	120	25	0.74	110	29	0.67	105	39	0.71
11	85	19	0.75	90	21	0.76	70	15	0.69	80	27	0.74
12	30	10	0.77	30	8	0.78	26	8	0.70	30	15	0.75
pool	-	-	0.80	-	-	0.81	-	-	0.70	-	-	0.78
totaal	2810	142	-	2940	194	-	2586	171	-	2760	202	-

De korrelatiekoëfficienten zijn o voor periode 1 en 4, voor periode 2 en 3 geldt $\rho_{ij} = 0.15$ resp. $\rho_{ij} = 0.10$, waarbij ρ_{ij} = correlatie tussen werkdrukte bedrijf i en werkdrukte bedrijf j (i,j = 1, .. , 12)

c)Er dient rekening te worden gehouden met maximale werving en maximale afvloeiing van arbeidskrachten ter grootte van 4 resp. 2 procent per periode.
De volgende situaties worden doorgerekend :

Situatie 1:in elke periode worden optimale bestanden berekend volgens model 5 in hoofdstuk 3 maar dan aangepast voor verschillende bedrijven,zonder hierbij rekening te houden met beperkingen van werving en afvloeiing.

Situatie 2 :als situatie 1 maar dan zonder arbeidspool (zie model o in hoofdstuk 2)

Situatie 3:als situatie 1, maar dan rekening houdend met de beperkte werving en afvloeiing.
Hiervoor is een apart model gemaakt gebaseerd op dynamische programmering.

Situatie 4: als situatie 3, maar dan zonder arbeidspool

De resultaten staan in onderstaande tabel :

	PERIODEN: 1	2	3	4	totaal (gemiddelde)
situatie 1					
bestand bedrijven	2838	3117	3316	2496	
bestand pool	887	796	691	1252	
totaal bestand	3725	3913	3807	3748	
verwachte kosten	605116	636791	623316	622186	621852 (1)
situatie 2					
bestand bedrijven	3769	3938	3761	3789	
verwachte kosten	649613	669223	650620	684157	663403 (2)
situatie 3					
bestand bedrijven	2816	2858	2814	2815	
bestand pool	950	988	980	970	
totaal bestand	3766	3846	3794	3785	
verwachte kosten	606218	638129	627013	624789	624037 (3)
situatie 4					
bestand bedrijven	3783	3818	3763	3784	
verwachte kosten	654435	680557	667528	691357	673469 (4)

tabel 3. Resultatenoverzicht havenarbeidspool.

Het economisch effect van pool is hieruit te berekenen.

1. Totaal economisch effect = (4) - (3) = 49432 (besparing 7,3%)
2. Economisch effect mbt dagelijkse fluktuatie = (2) - (1) = 41551 (besparing 6,2%)
3. Economisch effect mbt seizoensfluktuatie = 49432 - 41551 = 7881 (besparing 1,1%)

Toepassing 2. Onderwijzerspool in het district Los Angeles.

Bruno (zie literatuuroverzicht) heeft eensimulatiemodel gemaakt voor het bepalen van de optimale omvang van een vervangerspool van onderwijzers, die in het district Los Angeles als mogelijke maatregel voor vervanging van de afwezige vaste leerkrachten in aanmerking zou kunnen komen,

in plaats van de dagelijkse praktijk van het gebruik van minder gekwalificeerde arbeidskrachten, die meer de funktie van "kinderoppasser" vervulden dan van leerkracht. Daarnaast werd door de wel aanwezige vaste leerkrachten overwerk verricht om het onderwijspeil niet al te zeer te laten zakken.
Golberg en Moore (zie literatuurlijst) hebben voor dit probleem een analytisch model gebruikt, dat dezelfde resultaten gaf als het situatiemodel van Bruno.
Het eerstgenoemde is in feite het capaciteitsmodel 1 uit hoofdstuk 2, waarbij de kosten tgv een mandag tekort gelijk gesteld worden aan de kosten van een mandag overwerk.

We beschouwen de volgende gegevens :

a) kostenfactoren : a = 215 b = o c = 375

b) Vraag naar vervangers per dag :
De waarschijnlijkheidsverdeling ziet er als volgt uit :

Vraag	Fraktie
200-275	0,027
276-350	0,027
351-425	0,049
426-500	0,054
501-575	0,157
576-650	0,259
651-725	0,249
726-800	0.097
801-875	0,054
876-950	0,027
gemiddelde vraag	616

c) Beschikbaarheidsfraktie vervangers = 0,95
Resultaat : X = 630 min EK = 159974

N.B. Houden we rekening met de vraagverdeling voor elke dag van de week afzonderlijk, dan kunnen we met een (hier niet behandeld) model voor het bepalen van de optimale hoeveelheid parttime-capaciteit naast fulltimecapaciteit berekenen dat :
Het aantal fulltime leerkrachten is optimaal 557
Het aantal parttime leerkrachten is optimaal 129,0,75,92,144 voor resp. de dagen maandag t/m vrijdag.
De bijbehorende minimale kosten zijn 158497 per dag, zodat 1477 of 0,9% kan worden bespaard met het inschakelen van parttime krachten.

Overige toepassingen.

In principe zijn er legio overige toepassingen denkbaar :

- grootte van een parkeerplaats
- grootte van een vrachtwagenpark
- het aantal loketten van een postkantoor, bank
- grootte van een magazijn
- aantal dienstauto's.

Echter in situaties waarbij we te maken hebben met een karwei, een job, een rit die door een en hetzelfde capaciteitsgedeelte (bijvoorbeeld één programmeur of groep van programmeurs, één dienstauto) dient te worden uitgevoerd, hebben we andere modellen nodig dan de reeds genoemde. Veelal zijn wachttijdmodellen dan de meest in aanmerking komende of dient er een simulatiemodel te worden gemaakt.

De literatuurlijst bevat naast toepassingen van de hier behandelde modellen o.a. ook voorbeelden van laatstgenoemde wachttijdmodellen of simulatiemodel.

5. VRAAGSTUKKEN

Opgave 1.

De behoefte aan electrische energie per uur van een bedrijf varieert volgens een normale verdeling (gemiddeld 100 kWh, variatiekoëfficient 0,2). Men wil een eigen centrale bouwen en de piekbelastingen opvangen door het bijkopen van stroom.
De constante kosten van de eigen installatie (bij volle belasting) zijn a = 4 ct per kWh. De variabele kosten b = 3 ct per kWh.
De van buiten gekochte stroom kost 12 ct per kWh, zodat c-b = 9 ct per kWh.

a) wat is de optimale capaciteit van de eigen centrale ?
b) Als - bij hetzelfde gemiddelde verbruik - de variatiekoëfficient 0,4 zou zijn, wat is dan de optimale capaciteit van de eigen centrale ?
c) Als nu de energiebehoefte per uur een negatief exponentiële verdeling heeft(gemiddelde belasting nog steeds 100 kWh), wat is dan de optimale capaciteit van de eigen centrale ?
d) Hoe groot zijn in de drie hierboven genoemde gevallen de gemiddelde kosten per uur ?

Opgave 2.

Een wiskundig onderlegde bakker staat op vrijdagavond voor Pasen voor de beslissing hoeveel Paasbroden hij voor de komende zaterdag moet gaan bakken.

Hij maakt bij zijn berekeningen gebruik van de volgende gegevens.
De kostprijs van een Paasbrood is *f.* 2,50
De verkoopprijs is *f.* 3,40

Elk brood dat hij niet op de zaterdag voor Pasen verkoopt is dinsdag oudbakken. Een legerplaats in de omgeving is echter bereid maximaal 60 overgebleven broden tegen *f.* 2,40 per stuk af te nemen.

Zijn er meer dan 60 broden over, dan is de bakker genoodzaakt de na levering aan de legerplaats nog resterende broden weg te gooien.

De vraag naar Paasbroden op de zaterdag voor Pasen blijkt op grond van waarnemingen in voorgaande jaren bij benadering normaal verdeeld te zijn met een gemiddelde van 200 broden en een standaardafwijking van 20.

Raakt de bakker op zaterdag in de loop van de dag uitverkocht dan moet hij verder "nee" verkopen, maar hiervoor brengt hij bij zijn calculaties geen bedrag in rekening.

Gevraagd : Hoeveel broden gaat de bakker, als hij na zijn berekeningen nog tijd over heeft, voor de zaterdag voor Pasen bakken?
Ga hierbij van uit dat de bakker streeft naar winstmaximalisatie.

Opgave 3.

In een bedrijf bedragen de variabele kosten van het gebruik van capaciteit : b per eenheid per tijdseenheid.

De extra kosten bij het optreden van capaciteitstekort bedragen c-b per eenheid tekort per tijdseenheid.

De kosten van het aanhouden van X eenheden capaciteit bedragen : $a.X^{\alpha-1}$ per eenheid per tijdseenheid. ($0 < \alpha \leqslant 1$)
De vraag naar capaciteit bedraagt : x eenheden per tijdseenheid met verdelingsdichtheid $f(x)$.

Gevraagd a. Ontwikkel een formule, waaruit het optimale aantal aantehouden eenheden capaciteit kan worden afgeleid, uitgaande van minimalisatie van de totale verwachte kosten per tijdseenheid. (Behandel het vraagstuk continu)

b. Aan welke voorwaarde dient de verhouding a/c-b) te voldoen opdat de onder a. afgeleide formule voor alle waarden van α, ($0<\alpha\leqslant 1$) tot een oplossing leidt?

Opgave 4

Een commissie wil een advies uitbrengen over de wenselijke hoogte van kamerdeuren .

Als een deur hoger dan nodig is om een volwassene door te laten dan kan men spreken van overcapaciteit.
Moet iemand daarentegen bukken om zijn hoofd niet te stoten dan is er capaciteit tekort.
Beide situaties brengen een schade met zich mee die in geld kan worden uitgedrukt.
Hoe groot beide schadesoorten zijn is de commissie echter niet bekend. Er wordt daarom voorlopig verondersteld dat de schade van overcapaciteit s_1 per cm bedraagt ($s_1 \geqslant 0$) en de schade van capaciteitstekort $s_2 = \alpha s_1$ per cm ($\alpha \geqslant 0$).

Wel is bekend dat de lichaamslengte van de volwassen bevolking bij benadering normaal verdeeld is met een gemiddelde van 170 cm. en een standaardafwijking van 10 cm.

Gevraagd

1. Als wordt gestreefd naar minimalisatie van de totale verwachte schade, leid dan een formule af waaruit de optimale deurhoogte kan worden berekend als α bekend zou zijn,
2. Bereken de optimale hoogte voor $\alpha = 0, \frac{1}{2}, 1$ en 2.
3. Wanneer men in werkelijkheid deuren van 190 cm hoog maakt, hoe groot wordt α dan blijkbaar geacht ?

Opgave 5

Een bedrijf dient voor de komende n tijdseenheden het optimaal aantal eigen arbeiders X vast te stellen.
Het huidige aantal arbeiders bedraagt Y.
Bij het bepalen van het optimaal aantal X, waarvan men overigens aanneemt dat het direkt gerealiseerd wordt, dient het bedrijf rekening te houden met :

1. Ontslag van één of meer arbeiders van het huidige aantal Y kost r per arbeider.
2. Werving van één of meer arbeiders kost s per arbeider.
3. De kosten van het in dienst hebben van arbeiders bedragen a per arbeider per tijdseenheid.
4. De hoeveelheid werk ($\underline{w}$) per tijdseenheid is uitgedrukt in het aantal benodigde arbeiders en heeft een verdelingsdichtheid f (w).
5. Een capaciteitstekort wordt opgevangen door overwerk of huren van arbeiders tegen gemiddelde kosten van p.a. ($p > 1$) per eenheid capaciteitstekort per tijdseenheid.

Gevraagd

a) Leid algemene formule(s) af waarmee men het optimale aantal eigen arbeiders X kan berekenen.
b) Wat is het optimale aantal X als gegeven zijn:
   $n = 12$ $a = f\ 1500$ $p = 1{,}42$
   $r = f\ 2000$ $s = f\ 1750$ $Y = 1200$
   $\underline{w}$ is normaal verdeeld met $\mu = 1000$ en $\sigma = 100$
c) Wat gebeurt er als n naar oneindig gaat.

N.B. : Men hoeft geen rekening te houden met het feit dat een arbeider wegens ziekte of vakantie niet beschikbaar is. In dienst betekent hier dus tevens voor 100% beschikbaar.

Opgave 6

Een dienstverlenend bedrijf wil met behulp van een wiskundig model de optimale personeelsomvang X bepalen.

De vraag naar diensten is uitgedrukt in mandagen werk, dat verricht moet worden.
De vraag per dag is een stochastische variabele $\underline{x}$ met verdelingsdichtheid f(x).
De kosten per man per dag in dienst van het bedrijf bedragen a.
Voor elke gewerkte mandag wordt door het bedrijf een bedrag 2a gefaktureerd.
Als de vraag op een bepaalde dag groter is dan het aantal mensen in dienst van het bedrijf (X) wordt overwerk verricht tot een maximum van 10 procent van de normale werkdag.
De kosten per mandag overwerk bedragen 1,5 a.
Dat deel van de vraag, waaraan zelfs na het verrichten van de maximale hoeveelheid overwerk niet kan worden voldaan, wordt als verloren beschouwd hetgeen derhalve tot omzetverlies leidt.

a) Leid een formule af waarmee men op grond van bovenstaande het optimale aantal personen in dienst van het bedrijf kan bepalen zodat de verwachte winst per dag maximaal wordt.

b) Bepaal het optimale aantal personen in dienst van het bedrijf (X) indien geldt :
   $\underline{x}$ is normaal verdeeld met $\mu = 100$ en $\sigma = 25$
   $a = 200$

N.B. Het beschikbaar aantal personen wordt gelijkgesteld aan het aantal in dienst.

Aanwijzing : Men dient zich goed bewust te zijn van de betekenis van de 3 situaties, die zich kunnen voordoen, voor zowel de kosten als de opbrengst (de 3 situaties zijn :
   vraag per dag kleiner dan X
   vraag tussen X en 1,1 X (overwerk)
   en vraag groter dan 1,1 X)

Opgave 7

Het aantal per dag in een bedrijf benodigde werknemers bedraagt $\underline{x}$ en is homogeen verdeeld tussen 0 en B.
Het desbetreffende bedrijf heeft naast het beschikbaar aantal eigen werknemers X de mogelijkheid om uitzendkrachten in te schakelen, indien $\underline{x}$ groter is dan X.
Een uitzendkracht presteert echter een fraktie $^{1}\gamma_p$ minder dan een eigen werknemer en is daarnaast een fraktie γ_c duurder dan de eigen werknemer, die per dag een vast bedrag a kost.
Bepaal de optimale X zodat de totale verwachte kosten per dag minimaal worden.

Opgave 8

Voor het bepalen van het optimale aantal eigen capaciteitseenheden wordt allereerst een onderzoek verricht naar de waarschijnlijkheidsverdeling van de vraag $\underline{x}$ per dag en naar de relevante kostenfaktoren.
Het resultaat is :

1) $\underline{x}$ (uitgedrukt in eenheden capaciteit) is homogeen verdeeld tussen 0 en A ($A > 0$);
2) de kosten van eigen capaciteit per eenheid per dag bedragen a;
3) de uitbestedingskosten (uitbesteding vindt plaats als de vraag groter is dan de eigen capaciteit) per eenheid uitbesteed per dag bedragen c.

a) Bepaal het optimale aantal eigen capaciteitseenheden X zodat de totale verwachte kosten per dag minimaal worden.

b) Welke extra kosten (t.o.v. de minimale kosten) per dag worden er gemaakt indien in plaats van a de werkelijke waarde γ a ($\gamma > 0$) bedraagt.

c) Bepaal de onder b) genoemde extra kosten indien A = 25, a = 90, c = 100, γ = 0.75

ANTWOORDEN

1a Optimale capaciteit X = 102,8

1b X = 105,6

1c $X = -100\ \ell n\left(\frac{a}{c-b}\right) = 81,1$

1d min EK = 771,1 ct resp 842,2 ct resp 1424,5 ct

2 X uit : $100P(\underline{x} \geq X)+240\ P(\underline{x} \geq X - 60) = 250$
waarbij $\underline{x}$ = vraag naar broden
uitkomst: X = 220

3a $\int_X^\infty f(x)dx = \frac{a}{c-b}\alpha X^{\alpha-1}$

3b $\frac{a}{c-b} \leq \frac{1}{\alpha}$

4a $\int_L^\infty f(l)\,dl = \frac{1}{1+\alpha}$ (L=deurhoogte, $\underline{l}$= lengte volwassene)

4b α = o : geen deuren maken
$\alpha = \frac{1}{2}$: L = 165,7 x= 1: L =170 α = 2: L =174,3

4c α = 42,9

5a $\int_X^\infty f(w)dw = \begin{cases} (na-r)/npa & \text{als } X \leq Y \quad (1) \\ (na+s)/npa & \text{als } X > Y \quad (2) \end{cases}$
2 bijzondere situaties : 1) geen oplossing dan X = Y
2) twee oplossingen dan kosten vergelijken

5b (1) $\Rightarrow X_1$= 968 (< 1200) (2) $\Rightarrow X_2$ = 925, dus niet toegelaten
Resultaat X = 968

5c $\lim_{n\to\infty} \frac{(na-r)}{npa} = \frac{1}{p}$ $\lim_{n\to\infty} \frac{na+s}{npa} = \frac{1}{p}$

Resultaat : optimale X uit $P(\underline{w} \geq X) = \frac{1}{p}$

6a $0.55\ O(\underline{x} \geq 1,1X) + 1,5\quad P(\underline{x} \geq X) = 1$

6b X = 98

ANTWOORDEN

7 $X = B \frac{\gamma_c + \gamma_p}{1 + \gamma_c}$

8a $X = A(1 - \frac{a}{c})$

8b Extra kosten = $\frac{1}{2} A\gamma \frac{a^2}{c} \left(\frac{1}{\gamma} + \gamma - 2\right)$

8c Extra kosten = 62,38

Bijlage LITERATUURLIJST.

1. Stand van zaken	Gedenkboek bij het 50-jarig bestaan der Scheepvaart Vereeniging Zuid (SVZ), Rotterdam 1957
2. J.R. Virts, R.W. Garett	Weighing risk in capacity expansion Harvard Business Review (May-June 1970)
3. J. van Maare	Omvang chauffeursploeg en rijtijdenbesluit Statistica Neerlandia 13, no 3 (1959)
4. J.E. Bruno	The use of Monte Carlo techniques for determining optimal size of substitute teacher pools in large ubran school districts. Socio-Econ. Pla-Sci,4 (1970)
5. M.A. Golberg, J.Moore	The optimal size of a substitute teacher pool. Siam Review, 18 (1976)
6. J. Sittig	Pooling as an instrument of co-ordination IAEC, 2nd Seminar Munchen (1975).
7. D.K.Leegwater	Capacity Planning and the Labour Pool Proceedings in Operations Research 7, Physica Verlag (1978)
8. D.K.Leegwater	Pools and Pooling - Optimal capacities determined by mathematical methods Dissertatie, Erasmus Universiteit (1983)
9. M. Edelstein, M.Melnik	The Pool Control System, Interfaces vol 8, no 1, Part 2 (1977)
10. W.W. Williams, O.S.Fowler	Minimum Cost Fleet sizing for a University Motor Pool. Interfaces, vol 10, no 3(1980)
11. K.M.van Hee, D.K.Leegwater	Kapaciteitsplanning in het stuwadoors-bedrijf. Kwantitatieve methoden 1, Vol1,(1980)
12. B. An-Itzhak, B.A.Benn, B.A.Powell	Car Pool Systems in Railroad Transportation; Mathematical Models. Operations Research, vol 13, no. 9 (1967)

KLEINSTE-KWADRATENPROBLEMEN

R.J. Stroeker

1. INLEIDING

Een kleinste-kwadratenprobleem is een approximatie-probleem, waarbij de mate van nauwkeurigheid van de benadering wordt gemeten met behulp van een normbegrip, dat afgeleid is van het bekende euclidische afstandsbegrip. Twee wezenlijk verschillende situaties kunnen zich voordoen. In het ene geval gaat het om het benaderen van een bekende grootheid (bijv. een functie) door andere gelijksoortige grootheden van eenvoudiger type. In het andere geval wordt geprobeerd, op grond van waarnemingen, de parameters uit een wiskundig model te schatten. Betreft het een kleinste-kwadratenprobleem, dan geschiedt de wijze van benadering of schatting met de kleinste-kwadratenmethode.

Kleinste-kwadraten (afgekort k.k.) problemen kunnen in een grote verscheidenheid van situaties met succes worden toegepast. Hieraan, maar ook aan het feit, dat zij opgelost kunnen worden met behulp van matrixtechnieken, danken k.k. problemen hun grote populariteit. Dit betekent echter niet, dat voor alle approximatieproblemen de k.k. benadering de best denkbare is (zie [5] and [6]).

Carl Friedrich Gauss was de eerste die op het idee kwam om in een approximatieprobleem de fout te minimaliseren door de euclidische norm van het verschil tussen de te approximeren en de approximerende grootheid zo klein mogelijk te maken. Hij liet ook zien, dat de zo verkregen oplossing in zekere zin de meest waarschijnlijke is ([3]).

Gauss ontdekte de methode der kleinste kwadraten volgens zijn eigen zeggen in 1795 en hij gebruikte deze methode zeer intensief vanaf omstreeks 1800 bij zijn berekeningen van planetenbanen. Hij publiceerde zijn ontdekking echter pas voor het eerst in 1809. Daardoor

raakte hij in conflict met Legendre die de methode al in 1806 had gepresenteerd onder zijn eigen naam. Tegenwoordig wordt algemeen aangenomen, dat Gauss en Legendre de methode onafhankelijk van elkaar ontdekten. Voor meer informatie van historische aard verwijzen we naar [3].

2. ALGEMENE FORMULERING VAN HET PROBLEEM

We zullen in de volgende paragrafen een aantal typische k.k. problemen bespreken. Al deze problemen kunnen in eenzelfde vorm worden gegoten. De gemeenschappelijke probleemstelling kan als volgt worden omschreven. Getracht wordt om een gegeven grootheid $\underline{b}$ uit een klasse V van functies of van vectoren "zo goed mogelijk" te benaderen door een eindige lineaire combinatie van gegeven, gemakkelijk hanteerbare grootheden $\underline{b}_i$ uit een deelklasse W van V. De uitdrukking "zo goed mogelijk" betekent hier, dat die combinatie $x_1\underline{b}_1 + x_2\underline{b}_2 + \ldots + x_n\underline{b}_n$ $(x_i \in \mathbb{R})$ wordt gezocht, waarvoor het <u>residu</u>

$$\underline{r} = \underline{b} - \sum_{i=1}^{n} x_i \underline{b}_i$$

in zekere zin zo weinig mogelijk van $\underline{0}$ (de nul-functie of de nul-vector) verschilt. Dit verschil wordt in het k.k. geval gemeten met de <u>euclidische norm</u>. Een veel voorkomende situatie is dat $\underline{b}$ een vector is van waarnemingen, terwijl de vectoren $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ bekend zijn op grond van het geadopteerde (lineaire) model; de x_i zijn dan de te schatten parameters.

Zoals hierboven al wordt gesuggereerd, beschouwen we twee soorten, wezenlijk verschillende situaties: die, waarin vectoren de hoofdrol spelen (het <u>discrete</u> geval) en die, waarin deze rol wordt opgeëist door continue functies (het <u>continue</u> geval).

(2.1) HET DISCRETE GEVAL

Hier is $V = \mathbb{R}^m$, de reële m-dimensionale vectorruimte van reële m-tupels, die de vector $\underline{b}$ bevat, terwijl W een deelverzameling van V is, waaruit de vectoren $\underline{b}_i$ worden gekozen. Soms wordt voor W een n-dimensionale deelruimte van V genomen, waarbij de vectoren $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ dan een basis voor W vormen. De euclidische norm op V berust in dit geval op het gebruikelijke euclidische afstandsbegrip:

$$\|\underline{r}\| = \sqrt{\sum_{i=1}^{m} r_i^2} \text{ als } \underline{r} = (r_1, r_2, \ldots, r_m)^T.$$

Met de T wordt de transpositie-operatie aangegeven. De getransponeerde van een rijvector is een kolomvector met dezelfde kentallen. In het algemeen is de getransponeerde van een $m \times n$ matrix $A = (a_{ij})$ de $n \times m$ matrix $A^T = (b_{ij})$, waarvoor geldt dat $b_{ij} = a_{ji}$ voor elk paar (i,j).

Met

$$\langle \underline{r}, \underline{s} \rangle = \sum_{i=1}^{m} r_i s_i$$

wordt het standaard inwendig product van de vectoren $\underline{r} = (r_1, r_2, \ldots, r_m)^T$ en $\underline{s} = (s_1, s_2, \ldots, s_m)^T$ bedoeld. Men zegt, dat de euclidische norm op V door dit inwendig product wordt geïnduceerd, d.w.z. dat

$$\|\underline{r}\| = \sqrt{\langle \underline{r}, \underline{r} \rangle} \text{ voor elke } \underline{r} \in V.$$

(2.2) HET CONTINUE GEVAL

Nu is $V = C[a,b]$, de ∞-dimensionale vectorruimte van reële continue functies, gedefinieerd op het interval [a,b]. De ruimte V bevat weer de functie $\underline{b} = f(t)$, terwijl W een deelverzameling van V is, waaruit de functies $\underline{b}_i = f_i(t)$ $(i=1,2,\ldots,n)$ worden gekozen. Meestal is W een n-dimensionale deelruimte van V en vormen de functies $f_1(t), f_2(t), \ldots, f_n(t)$ een basis voor W. De euclidische norm kan in deze situatie opgevat worden als het continue analogon van de in (2.1) gedefinieerde norm, d.w.z.

$$\|\underline{r}\| = \sqrt{\int_a^b r^2(t)dt} \text{ als } \underline{r} = r(t).$$

Ook in dit geval kan de ruimte V worden toegerust met een inwendig product

$$\langle \underline{r},\underline{s} \rangle = \int_a^b r(t)s(t)dt,$$

waardoor deze norm wordt geïnduceerd, dus $\|\underline{r}\| = \sqrt{\langle \underline{r},\underline{r} \rangle}$ voor elke $\underline{r} = r(t)$.

Belangrijke voorbeelden van eindig-dimensionale deelruimten van V zijn:

(1) P_n - de (n+1)-dimensionale ruimte van veeltermen (of polynomen) van de graad ten hoogste n: $a_o+a_1t+a_2t^2+\ldots+a_nt^n$.
Als functies $\underline{b}_i = f_i(t)$ kan men bijvoorbeeld kiezen: $f_i(t) = t^i$ voor $i = 0,1,\ldots,n$.

(2) T_n - de (2n+1)-dimensionale ruimte van trigonometrische veeltermen van de orde ten hoogste n:
$a_o+a_1\cos t+b_1\sin t+a_2\cos 2t+b_2\sin 2t+\ldots+a_n\cos nt+b_n\sin nt$.
Het is in dit geval gebruikelijk voor $\underline{b}_i$ te kiezen:
$\underline{b}_i = f_i(t)$, waarbij $f_{2j}(t) = \cos jt$ $(j=0,1,\ldots,n)$ en $f_{2j-1}(t) = \sin jt$ $(j=1,2,\ldots,n)$.

Het k.k. probleem dat wij zullen bespreken en toelichten met voorbeelden kan nu precies worden geformuleerd.

(2.3) HET K.K. PROBLEEM

Laat V één van de twee reële vectorruimten $\mathbb{R}^m$ of $C[a,b]$ zijn, toegerust met de euclidische norm $\|.\|$. Zij verder W een deelverzameling van V. Gevraagd wordt, bij gegeven $\underline{b} \in V$ en $\underline{b}_i \in W$ $(i=1,2,\ldots,n)$, getallen $x_1,x_2,\ldots,x_n$ te bepalen op zodanige wijze dat

$$\|\underline{b} - \sum_{i=1}^{n} x_i\underline{b}_i\|$$

zo klein mogelijk is.

Eisen we bovendien dat W een eindig-dimensionale deelruimte van V is, dan zegt een bekende stelling uit de approximatietheorie (zie bijvoorbeeld [5]), dat er bij elke $\underline{b} \in V$ een <u>unieke</u> $\underline{a}_o \in W$ te vinden is, zodat $\|\underline{b}-\underline{a}_o\| \leq \|\underline{b}-\underline{a}\|$ voor elke $\underline{a} \in W$. Dit betekent dat probleem (2.3) onder die extra voorwaarde altijd een oplossing heeft. Kiezen we de $\underline{b}_i \in W$ ook nog zó, dat het stelsel $\{\underline{b}_1,\underline{b}_2,\ldots,\underline{b}_n\}$ een basis vormt voor W, dan heeft (2.3) zelfs een <u>unieke</u> oplossingsvector $\underline{x} = (x_1,x_2,\ldots,x_n)^T \in \mathbb{R}^n$.

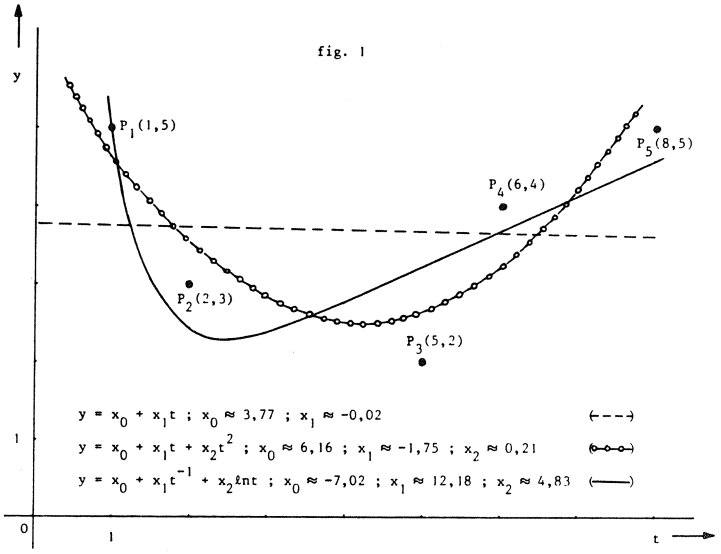
3. ENKELE TOEPASSINGEN

Wij beginnen met een voorbeeld uit de klasse van discrete k.k. problemen.

Bij het verrichten van empirisch onderzoek komt het vaak voor, dat een geschikte wiskundige relatie tussen twee variabele grootheden moet worden afgeleid uit een aantal experimenteel verkregen waarden. In een dergelijke situatie moet men over voldoende waarnemingen beschikken om ervoor te zorgen, dat het effect van meet- en afleesfouten beperkt kan blijven. Een keuze voor het type relatie, dat men hoopt te vinden - of dat aannemelijk is - wordt van te voren gemaakt en berust meestal op theoretische overwegingen. Wiskundig geformuleerd komt het hierop neer: uitgaande van een aantal waarnemingen, die als punten $(t_1,y_1),(t_2,y_2),\ldots,(t_m,y_m)$ in het platte vlak kunnen worden weergegeven, probeert men een relatie $y = g(t)$ te vinden, die zo goed mogelijk bij de waarnemingen aansluit. Anders gezegd, de functie g wordt zo gekozen, dat alle punten $(t_1,y_1),(t_2,y_2),\ldots,(t_m,y_m)$ in de buurt van de grafiek van g liggen. De functie g wordt doorgaans als meest geschikte kandidaat uit een van te voren bekende klasse van functies gekozen. Zo'n klasse bestaat meestal uit alle lineaire combinaties van een eindig aantal, willekeurige functies. Een dergelijke g heet dan een goede "fit". De beschreven procedure wordt "<u>curve fitting</u>" genoemd. In figuur 1 zijn een aantal "fits" van een vijftal

punten $P_1(1,5)$, $P_2(2,3)$, $P_3(5,2)$, $P_4(6,4)$ en $P_5(8,5)$ geschetst.

fig. 1



Omdat de punten op experimentele wijze zijn verkregen, zijn zij behept met onnauwkeurigheden. Tevens is het aantal waarnemingen meestal veel groter dan het aantal te schatten parameters. Dit heeft tot gevolg dat de functie g in het algemeen niet zo gekozen kan en behoeft te worden, dat zijn grafiek precies door de gegeven punten gaat.

Een belangrijke vraag is op welke wijze de parameters moeten worden bepaald om een zo gunstig mogelijk resultaat te verkrijgen. Elk criterium dat meet hoe goed de functie g past by de gegeven punten $(t_1,y_1),(t_2,y_2),\ldots,(t_m,y_m)$ zal op één of andere manier gebruik maken van de residuen (zie figuur 2):

$$(3.1) \qquad r_k = y_k - g(t_k) \qquad (k=1,2,\ldots,m).$$

Veronderstel dat, op grond van de ligging van de punten $(t_1,y_1),(t_2,y_2),\ldots,(t_m,y_m)$ en onze eisen en verwachtingen aangaande het gedrag van g, het gunstig lijkt voor g een gedaante te kiezen van de vorm

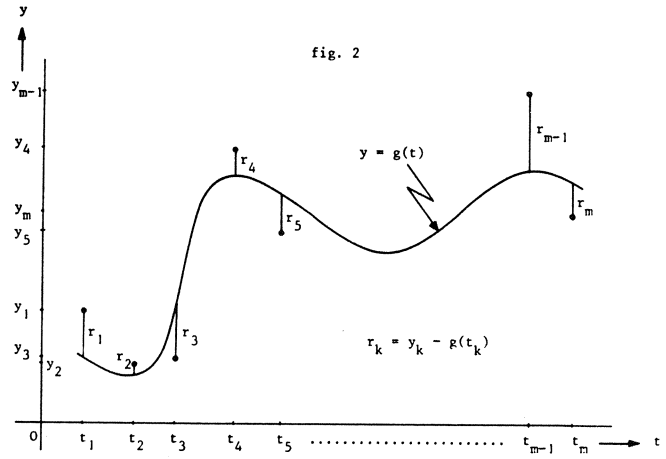


fig. 2

$$(3.2) \qquad g(t) = x_1 f_1(t) + x_2 f_2(t) + \ldots + x_n f_n(t).$$

Hierbij zijn $f_1, f_2, \ldots, f_n$ een aantal gemakkelijk hanteerbare, elementaire functies. De coëfficiënten $x_1, x_2, \ldots, x_n$ zijn de onbekende parameters, die zó gekozen moeten worden dat de functie g uit (3.2) een goede "fit" wordt. Substitutie van (3.2) in (3.1) levert het volgende stelsel vergelijkingen

$$\begin{aligned}
r_1 &= y_1 - x_1 f_1(t_1) - x_2 f_2(t_1) - \ldots - x_n f_n(t_1) \\
r_2 &= y_2 - x_1 f_1(t_2) - x_2 f_2(t_2) - \ldots - x_n f_n(t_2) \\
&\;\;\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
r_m &= y_m - x_1 f_1(t_m) - x_2 f_2(t_m) - \ldots - x_n f_n(t_m)
\end{aligned}$$

In vectorvorm geschreven ziet dit stelsel eruit als

$$(3.3) \qquad \underline{r} = \underline{y} - \sum_{i=1}^{n} x_i \underline{b}_i,$$

waarbij $\underline{r} = (r_1, r_2, \ldots, r_m)^T$, $\underline{y} = (y_1, y_2, \ldots, y_m)^T$ en

$\underline{b}_i = (f_i(t_1),f_i(t_2),\ldots,f_i(t_m))^T$ $(i=1,2,\ldots,m)$ vectoren uit $\mathbb{R}^m$ zijn. Hoewel er in dit voorbeeld steeds over functies wordt gesproken bevinden we ons toch in de discrete situatie. Immers, de functies $f_1,f_2,\ldots,f_n$ spelen in het geheel geen rol als elementen van één of andere lineaire ruimte van functies zoals C[a,b]. Maar, voor elk functie f_i vormen de functiewaarden $f_i(t_1),f_i(t_2),\ldots,f_i(t_m)$ de gegeven vector $\underline{b}_i$ uit $\mathbb{R}^m$.

We herkennen in (3.3) onmiddellijk (2.3), zodat het voor de hand ligt de getallen $x_1,x_2,\ldots,x_n$ zodanige waarden toe te kennen dat $\|\underline{r}\|$ zo klein mogelijk is.

Omdat we met vectoren te maken hebben is (3.3) ook te schrijven in matrix-vector notatie, namelijk

$$\underline{r} = \underline{y} - A\underline{x}, \tag{3.4}$$

waarbij A de m×n matrix is die gevormd wordt door de kolomvectoren $\underline{b}_1,\underline{b}_2,\ldots,\underline{b}_n$ en $\underline{x} = (x_1,x_2,\ldots,x_n)^T \in \mathbb{R}^n$. In het algemeen zal de matrix A meer rijen dan kolommen hebben (m>n). Het stelsel lineaire vergelijkingen

$$A\underline{x} = \underline{y}$$

wordt dan een overbepaald stelsel genoemd. Een dergelijk stelsel, waarin het aantal vergelijkingen dus groter is dan het aantal onbekenden, heeft meestal geen oplossingen. Om toch een aanvaardbare benadering

$$A\underline{x} \approx \underline{y}$$

te verkrijgen, kan men $\|\underline{y}-A\underline{x}\| = \|\underline{r}\|$ trachten te minimaliseren.

Dit leidt tot de volgende formulering van het lineaire k.k. probleem.

(3.5) HET LINEAIRE K.K. PROBLEEM

Zij A een gegeven reële m×n matrix en zij $\underline{y}$ een gegeven vector uit $\mathbb{R}^m$. Gevraagd wordt een vector $\underline{x} \in \mathbb{R}^n$ te vinden waarvoor de euclidische norm van de residu-vector $\|\underline{y}-A\underline{x}\|$ minimaal is. Zo'n vector $\underline{x}$ wordt een k.k. oplossing van het stelsel $A\underline{x} = \underline{y}$ genoemd.

Als in (2.3) voor V de lineaire ruimte $\mathbb{R}^m$ wordt genomen, dan valt (2.3) samen met (3.5). Immers, de gegeven vectoren $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ zijn hier de n kolomvectoren van de matrix A.

Geldt in (3.5) dat $m \leq n$, dan is het stelsel $A\underline{x} = \underline{y}$ in het algemeen oplosbaar. In dat geval is een k.k. oplossing niets anders dan een oplossing van dit stelsel. Het geval $m > n$ is echter veel interessanter en we zullen van nu af aan steeds veronderstellen dat deze laatste ongelijkheid geldt.

Ons volgende voorbeeld betreft een continu k.k. probleem.

Wil men de directe beschikking hebben over alle mogelijke waarden van een functie f(t) op een gegeven interval [a,b] - het is bijvoorbeeld nuttig in de computerpraktijk om de elementaire functies onmiddellijk bij de hand te hebben - dan heeft men een benadering voor f nodig, die in principe in elk punt t van [a,b] een goede benadering geeft van f(t) en daar gemakkelijk geëvalueerd kan worden. Afhankelijk van de specifieke situatie, worden functies $f_1(t), f_2(t), \ldots, f_n(t)$ geselecteerd, waarvoor

$$f(t) \approx \sum_{i=1}^{n} x_i f_i(t) \quad \text{voor alle } t \in [a,b]$$

bij een geschikte keuze van $x_1, x_2, \ldots, x_n$. De residu-functie

$$(3.6) \qquad r(t) = f(t) - \sum_{i=1}^{n} x_i f_i(t)$$

kan niet voor alle t de waarde nul aannemen en daarom trachten we r(t) zo klein mogelijk te maken door te eisen

dat

$$\sqrt{\int_a^b r^2(t)dt} \text{ minimaal is (zie (2.2)).}$$

Soms kan het erg lastig zijn of zelfs onmogelijk de betrokken integralen te evalueren. In dat geval is het gebruikelijk het continue probleem te discretizeren. Dit betekent dat een groot aantal punten $t_0, t_1, \ldots, t_N$ in $[a,b]$ gekozen wordt, ongeveer uniform verdeeld over $[a,b]$ - bijvoorbeeld $t_j = a + \frac{j}{N}(b-a)$ voor $j = 0,1,\ldots,N$ - en dat de continue norm vervangen wordt door de discrete norm

$$\sqrt{\sum_{j=o}^{N} r^2(t_j)}.$$

In deze laatste vorm is het probleem weer op te vatten als een lineair k.k. probleem.

4. DE NORMAALVERGELIJKINGEN

De al door Gauss aangegeven oplossing van het lineaire k.k. probleem (3.5) via de zogenaamde normaalvergelijkingen, kan op meetkundige wijze eenvoudig worden verklaard.

Beschouwen we de kolommen van A als vectoren uit $\mathbb{R}^m$, dan doorloopt $A\underline{x}$ de kolommenruimte K(A) van A als $\underline{x}$ de $\mathbb{R}^n$ doorloopt. De euclidische afstand van het punt met plaatsvector $\underline{y}$ in $\mathbb{R}^m$ tot een punt met plaatsvector $A\underline{x}$ in K(A) (zie figuur 3) is minimaal als de verschilvector $\underline{y} - A\underline{x}$ loodrecht op K(A) staat, d.w.z. als het standaard inwendig product $\langle \underline{y}-A\underline{x}, A\underline{x}' \rangle = 0$ voor elke $\underline{x}' \in \mathbb{R}^n$.
Dus ook $\langle A^T(\underline{y}-A\underline{x}), \underline{x}' \rangle = 0$ voor alle $\underline{x}' \in \mathbb{R}^n$. Dit is slechts mogelijk als $A^T(\underline{y}-A\underline{x}) = 0$, omdat geen enkele vector $\neq \underline{0}$ uit $\mathbb{R}^n$ orthogonaal kan zijn met alle vectoren uit $\mathbb{R}^n$.
Dus

$$A^T A\underline{x} = A^T\underline{y}. \qquad (4.1)$$

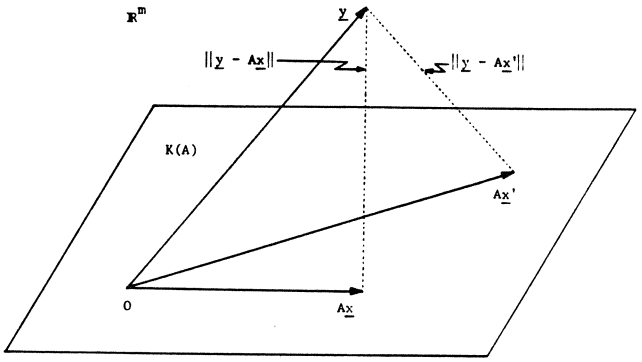


fig. 3

Dit is een stelsel van n vergelijkingen met n onbekenden; deze vergelijkingen worden de normaalvergelijkingen genoemd van het stelsel $A\underline{x} = \underline{y}$.

Ook op formele wijze zijn de normaalvergelijkingen gemakkelijk af te leiden. Uit
$\|\underline{y}-A\underline{x}'\|^2 - \|\underline{y}-A\underline{x}\|^2 = \|A(\underline{x}'-\underline{x})\|^2 - 2\langle\underline{y}-A\underline{x}, A(\underline{x}'-\underline{x})\rangle =$
$= \|A(\underline{x}'-\underline{x})\|^2 - 2\langle A^T\underline{y}-A^TA\underline{x}, \underline{x}'-\underline{x}\rangle$, vinden we onmiddellijk dat $\|\underline{y}-A\underline{x}'\| \geq \|\underline{y}-A\underline{x}\|$ voor alle $\underline{x}' \in \mathbb{R}^n$ als $\underline{x}$ voldoet aan de normaalvergelijkingen (4.1). Bedenk hierbij dat
$\|\underline{y}-A\underline{x}'\|^2 = \langle\underline{y}-A\underline{x}', \underline{y}-A\underline{x}'\rangle = \|\underline{y}\|^2 - 2\langle\underline{y}, A\underline{x}'\rangle + \|A\underline{x}'\|^2$ enz.

De vraag naar de oplossing van (4.1) is niet moeilijk te beantwoorden als de matrix A volledige rang heeft. Dit betekent dat de kolomvectoren van A dan lineair onafhankelijk zijn; we hebben immers verondersteld dat $m > n$. Het gevolg is dat de $n \times n$ matrix A^TA niet-singulier en dus inverteerbaar is. De unieke oplossing van (4.1) en dus ook van (3.5) is dan

$$(4.2) \qquad \underline{x} = (A^TA)^{-1}A^T\underline{y}.$$

De "fits" uit figuur 1 zijn tot stand gekomen door toepassing van deze formule (4.2), waarbij $\underline{y}$ de $\mathbb{R}^3$-vector

van waarnemingen is en A de 3×5 matrix met (i,j)-de element $f_j(t_i)$ (zie (3.2) en (3.3)).

En als A geen volledige rang heeft? Als rang(A) = $k < n$, hoe zit het dan? Ofschoon dit geval uit praktisch oogpunt niet zo belangrijk is, willen wij er toch enige aandacht aan schenken. Welnu, elke matrix kan geschreven worden als product van twee matrices, die volledige rang hebben. Dus, als de m×n matrix A rang k heeft, dan is er een m×k matrix A_1 en een k×n matrix A_2 zó dat $A = A_1A_2$ en $\text{rang}(A_1) = \text{rang}(A_2) = k$ (zie [11]).
De k×k matrices $A_1^TA_1$ en $A_2A_2^T$ hebben volledige rang; zij zijn dus inverteerbaar. De normaalvergelijkingen (4.1) kunnen vervangen worden door het gelijkwaardige stelsel

$$(4.3) \qquad A_2\underline{x} = (A_1^TA_1)^{-1}A_1^T\underline{y}.$$

Dit is door invulling van $A = A_1A_2$ in (4.1) eenvoudig in te zien. We definiëren de matrix

$$(4.4) \qquad A^I = A_2^T(A_2A_2^T)^{-1}(A_1^TA_1)^{-1}A_1^T.$$

De vector $\underline{x} = A^T\underline{y}$ geeft een oplossing van (4.3) en dus ook van (4.1), zoals direct volgt uit (4.4). Deze oplossing is in het algemeen niet de enige, maar van alle oplossingen is dit degene met de kleinste norm $\|\underline{x}\|$. De matrix A^I heet de gegeneraliseerde inverse (of pseudo-inverse) van A. De hier gekozen definitie (4.4) hangt niet af van de specifieke factorisatie $A = A_1A_2$, aangezien elke factorisatie van A met de gewenste eigenschappen van het type $A = (A_1B^{-1})(BA_2)$ is voor één of andere niet-singuliere k×k matrix B.

De oplossing $\underline{x} = A^T\underline{y}$ van (4.1) is als volgt op te vatten als een generalisatie van (4.2). Heeft A wèl volledige rang, dan is de factor A_2 een vierkante matrix van volledige rang en dus inverteerbaar. Na enig rekenwerk blijkt dat $A^I = (A^TA)^{-1}A^T$, zodat beide oplossingen samenvallen. Als bovendien A zelf een vierkante matrix is, dus als m = n, dan volgt onmiddellijk dat $A^I = A^{-1}$. Dit

verklaart de naamgeving "gegeneraliseerde inverse".

Samenvattend:

4.5. ALGEMENE OPLOSSING VAN HET LINEAIRE K.K. PROBLEEM

Zij A^I de gegeneraliseerde inverse van A. De vector $\underline{x} = A^T\underline{y}$ is een k.k. oplossing van het stelsel $A\underline{x} = \underline{y}$. Deze oplossing is uniek als A volledige rang heeft. Is dit laatste niet het geval, dan is $\underline{x}$ onder alle k.k. oplossingen van $A\underline{x} = \underline{y}$ de oplossing met kleinste euclidische norm.

In het continue geval zijn de normaalvergelijkingen op volkomen analoge wijze af te leiden. We behoeven ons alleen maar te realizeren dat het continue geval uit het discrete geval verkregen wordt door in de definitie van inwendig product sommen (Σ) te vervangen door integralen en door vectoren te vervangen door functies. Ook een gelijksoortige meetkundige afleiding (zie fig. 3) kan worden gegeven. Een andere schrijfwijze voor (4.1) is

$$\sum_{i=1}^{n} x_i \, \langle \underline{b}_i, \underline{b}_j \rangle = \langle \underline{y}, \underline{b}_j \rangle,$$

waarbij $\underline{b}_j$ de j-de kolom van A voorstelt (j=1,2,...,n). Verplaatst naar de continue situatie wordt dit

$$\sum_{i=1}^{n} x_i \langle f_i, f_j \rangle = \langle f, f_j \rangle \qquad j = 1,2,\ldots,n. \tag{4.6}$$

Hierbij zijn $f, f_1, f_2, \ldots, f_n$ continue functies gedefinieerd op het interval [a,b]. Ook het inwendig product heeft een (zij het niet zichtbare) wijziging ondergaan, namelijk

$$\langle f, f_j \rangle = \int_a^b f(t) f_j(t) dt.$$

De vergelijkingen (4.6) heten de <u>normaalvergelijkingen</u> van het approximatieprobleem

$$f(t) \approx \sum_{i=1}^{n} x_i f_i(t), \quad t \in [a,b]$$

Het oplossen van (4.6) wordt heel eenvoudig als de functies $f_1,f_2,\ldots,f_n$ zo gekozen worden, dat het stelsel $\{f_1,f_2,\ldots,f_n\}$ <u>orthogonaal</u> is, dus als $\langle f_i,f_j\rangle = 0$ voor elk paar (i,j) waarvoor $i \neq j$. De unieke oplossing van (4.6) kan dan worden geschreven als

$$(4.7)\quad x_j = \frac{\langle f,f_j\rangle}{\|f_j\|^2} = \frac{1}{\|f_j\|^2}\int_a^b f(t)f_j(t)dt \qquad (j=1,2,\ldots,n).$$

We veronderstellen hierbij dat geen van de functies f_j de nulfunctie is, zodat $\|f_j\| \neq 0$.

We geven een voorbeeld. Beschouw de lineaire ruimte van trigonometrische polynomen T_n (zie (2.2)). De functies $f_o,f_1,\ldots,f_{2n}$ met $f_{2j} = \cos jt$ $(j=0,1,\ldots,n)$ en $f_{2j-1} = \sin jt$ $(j=1,2,\ldots,n)$ vormen een basis voor T_n, die orthogonaal is m.b.t. het bekende inwendig product op $C[0,2\pi]$. De orthogonaliteitsrelaties

$$\int_0^{2\pi} f_i(t)f_j(t)dt = 0 \text{ voor elk paar } (i,j) \text{ met } i \neq j$$

volgen onmiddellijk uit de somformules

$$\sin(i\pm j)t = \sin it.\cos jt \pm \cos it.\sin jt$$
$$\cos(i\pm j)t = \cos it.\cos jt \mp \sin it.\sin jt,$$

en het feit dat

$$\int_0^{2\pi} \sin(i\pm j)t\,dt = \int_0^{2\pi} \cos(i\pm j)t\,dt = 0$$

als $i \neq j$. Verder is

$$\|f_o\|^2 = \int_0^{2\pi} dt = 2\pi \text{ en } \|f_j\|^2 = \int_0^{2\pi} f_j^2(t)dt = \pi \quad (j=1,2,\ldots,n),$$

zodat op grond van (4.7) de k.k. oplossing van het approximatieprobleem

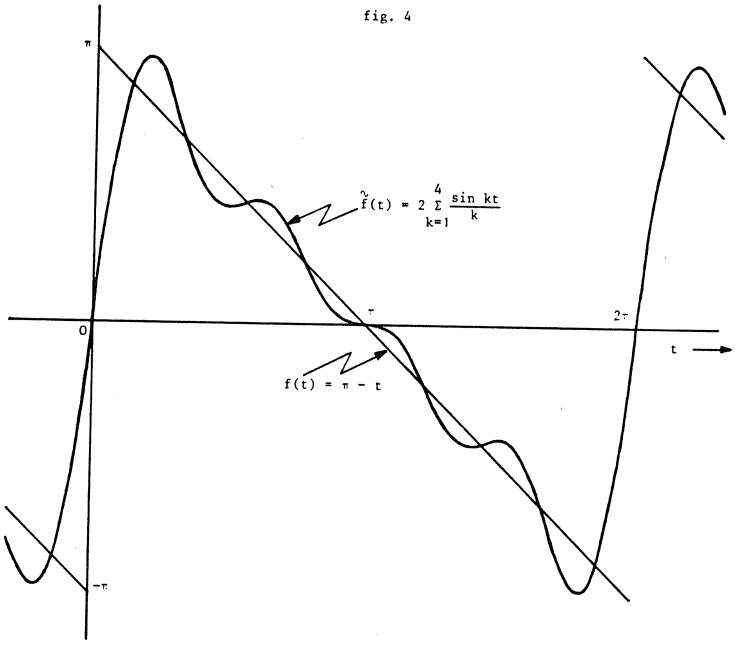
$$f(t) \approx \tfrac{1}{2}a_o + \sum_{j=1}^{n} (a_j\cos jt + b_j\sin jt),\ t \in [0,2\pi]$$

de volgende coëfficiënten heeft:

$$a_j = \frac{1}{\pi} \int_0^{2\pi} f(t)\cos jt \, dt \quad (j=0,1,\ldots,n) \text{ en}$$

$$b_j = \frac{1}{\pi} \int_0^{2\pi} f(t)\sin jt \, dt \quad (j=1,2,\ldots,n).$$

fig. 4



De reeks $\frac{1}{2}a_o + \sum_{j=1}^{\infty} (a_j \cos jt + b_j \sin jt)$ heet de <u>Fourier-reeks</u> van f op het interval $[0,2\pi]$. Voor continue functies f convergeert deze reeks in elk punt t van het interval $(0,2\pi)$ naar de waarde f(t). Om een idee te geven van het gedrag van zo'n trigonometrische benadering is in figuur 4 de k.k. oplossing geschetst voor n = 4 van de "zangtandfunctie":

$f(t) = \pi - t$ voor $t \in [0,2\pi]$, $f(t+2\pi) = f(t)$ voor $t \in \mathbb{R}$.

5. NUMERIEKE ASPECTEN VAN K.K. PROBLEMEN

Lange tijd voldeed de methode, die gebruik maakt van de normaalvergelijkingen om k.k. problemen op te lossen, uitstekend. Maar toen het na de tweede wereldoorlog door de ontwikkeling van de computer mogelijk werd steeds grotere stelsels aan te pakken, bleek al gauw dat de normaalvergelijkingen vaak slecht geconditioneerd zijn. Dit wil zeggen dat het effect van (kleine) fouten in de gegevens desastreus kan zijn voor het eindantwoord. Het kan zelfs gebeuren, dat de normaalvergelijkingen zo slecht geconditioneerd zijn, dat de oplossingen een volkomen willekeurig karakter lijken te hebben, zonder enige relatie tot het oorspronkelijke probleem.

Hoe kan men nagaan of het zinvol is een gegeven (groot) stelsel lineaire vergelijkingen met behulp van de computer op te lossen? Een begrip dat hierbij een voorname rol speelt is dat van conditiegetal van de coëfficiëntenmatrix van het stelsel. Veronderstel dat B een vierkante niet-singuliere matrix is en $\underline{a}$ een gegeven vector. Het stelsel

$$B\underline{x} = \underline{a}$$

heeft dan de unieke oplossing $\underline{x} = B^{-1}\underline{a}$. Wat is de invloed van kleine verstoringen in $\underline{a}$ en B op de oplossing van het stelsel? Preciezer geformuleerd, als $\Delta\underline{a}$ en ΔB deze verstoringen zijn en $\tilde{\underline{x}}$ is de oplossing van

$$(B+\Delta B)\tilde{\underline{x}} = \underline{a} + \Delta\underline{a},$$

hoe groot is dan de afwijking $\Delta\underline{x} = \tilde{\underline{x}} - \underline{x}$ t.o.v. het juiste antwoord $\underline{x}$? Het kan worden bewezen, dat voor deze relatieve fout in essentie de volgende ongelijkheid geldt:

$$(5.1) \qquad \|\Delta\underline{x}\|/\|\underline{x}\| \leq c(B)\{\|\Delta B\|/\|B\| + \|\Delta\underline{a}\|/\|\underline{a}\|\}.$$

In deze formule is $\|B\| = \sup_{\underline{x}\neq\underline{o}} \|B\underline{x}\|/\|\underline{x}\|$, terwijl

$c(B) = \|B\|.\|B^{-1}\|$ het conditiegetal van B is. Voor de euclidische norm (dat is nog steeds de norm die we gebruiken) betekent dit dat

$$c(B) = \frac{|\lambda|_{max}}{|\lambda|_{min}} \geq 1, \text{ waarbij } |\lambda|_{max}, |\lambda|_{min} \text{ respectievelijk}$$

de maximale modulus en de minimale modulus is van de (mogelijk complexe) eigenwaarden λ van B. Formule (5.1) wil zeggen dat in het slechst denkbare geval de relatieve fouten in de gegevens versterkt met een factor c(B) in het antwoord kunnen voorkomen. Is c(B) veel groter dan 1, dan is het stelsel $B\underline{x} = \underline{a}$ slecht geconditioneerd.

De belangrijkste reden voor een groot conditiegetal is de "bijna afhankelijkheid" van het stelsel kolomvectoren van de matrix. De 2×2 matrix

$$B = \begin{pmatrix} 10^{-6} & 0 \\ 2 & 1 \end{pmatrix}$$

heeft conditiegetal $c(B) = 10^6$.

Hoe zit het met de normaalvergelijkingen (4.1) of (4.6)? Het is helaas dikwijls zo dat het conditiegetal van de coëfficiëntenmatrix van het stelsel normaalvergelijkingen onaanvaardbaar groot is. Een treffend voorbeeld is het volgende. Laten in de normaalvergelijkingen (4.6) de functies f_i als volgt worden gekozen:
$f_i(t) = t^{i-1}$ voor $i = 1,2,\ldots,n$ en $t \in [0,1]$. Dan is

$$\langle f_i, f_j \rangle = \int_0^1 t^{i+j-2} dt = \frac{1}{i+j-1} \text{ voor } i,j = 1,2,\ldots,n.$$

De coëfficiëntenmatrix van het stelsel is dan de zogenaamde Hilbertmatrix

$$H_n = \begin{bmatrix} 1 & 2^{-1} & 3^{-1} & \dots & n^{-1} \\ 2^{-1} & 3^{-1} & 4^{-1} & \dots & (n+1)^{-1} \\ 3^{-1} & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ n^{-1} & . & . & \dots & (2n-1)^{-1} \end{bmatrix}.$$

Voor grote waarden van n verschillen de meest rechtse kolommen van H_n heel weinig van elkaar, zodat H_n voor grote waarden van n "bijna singulier" is. Het conditiegetal van H_n is dan ook zeer groot, in iedere geval neemt $c(H_n)$ exponentieel in n toe. Voor n = 10 mag men al geen enkele betekenis meer toekennen aan de via de normaalvergelijkingen berekende oplossing van het bovenstaande k.k. probleem.

Hoe moet men dan wel te werk gaan? We hebben al eerder gezien, dat de coëfficiëntenmatrix van het stelsel normaalvergelijkingen een diagonaalmatrix is als de vectoren $\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n$ een orthogonaal stelsel vormen (zie (4.7)). De normaalvergelijkingen zijn dan op triviale wijze op te lossen. Dit wekt de suggestie, dat het nuttig zou kunnen zijn te zoeken naar methoden om het stelsel $\{\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n\}$ of de matrix A te orthogonaliseren. Er zijn verschillende orthogonalisatiemethoden bekend, bijvoorbeeld die van Gram-Schmidt en het proces dat gebruik maakt van Householder transformaties. Vooral deze laatste methode is numeriek zeer stabiel en wordt daarom veel gebruikt. Het zou ons te ver voeren hier op deze methoden in te gaan. Voor verdere informatie van numerieke aard zij verwezen naar de zeer leesbare artikelen [8] en [9].

Ofschoon er nog veel interessants te vertellen is over de k.k. methode, zullen we het hierbij laten. In de literatuurlijst is een aantal boeken opgenomen, waarin uitgebreid op de k.k. problematiek wordt ingegaan ([2],[4],[10],[11]). Tenslotte willen we de aandacht vestigen op [7]. In dit boek wordt een aantal alleraardigste toepassingen van de lineaire algebra en matrixrekening besproken. De gekozen onderwerpen zijn afkomstig uit een groot aantal toepassingsgebieden, zoals de economie, de natuurkunde, de ecologie, de demografie en de genetica om er enkele te noemen.

LITERATUUR

[1] Dekker, T.J., - Numerieke Algebra, MC Syllabus nr. 12, Math. Centrum Amsterdam, 1971.

[2] Golub, G. & Charles van Loan - Matrix Computations, Johns Hopkins Univ. Press., 1983 (of: North Oxford Academic Publ. Co., Oxford 1984).

[3] Hall, Tord - Carl Friedrich Gauss, A Biography, The M.I.T. Press, Cambridge Mass., 1970.

[4] Lawson, C.L. & R.J. Hanson - Solving Least Squares Problems, Prentice Hall, 1974.

[5] Powell, M.J.D. - Approximation Theory and Methods, Cambridge Univ. Press, 1981.

[6] Rice, John R. - Numerical Methods, Software and Analysis, McGraw-Hill, 1983.

[7] Rorres, Chris & Howard Anton - Applications of Linear Algebra, John Wiley & Sons, 2nd ed. 1979.

[8] Sluis, A. van der - Stabiliteit en Oplossing van Kleinste-kwadraten-problemen, Nieuw Tijdschr.v. Wisk. 66 (1978/79), 251-264.

[9] Sluis, A. van der - Computation and Stability of Linear Least Squares Problems, In: Proceedings Bicentennial Congress of the Wisk. Genootschap, Math. Centre Amsterdam 1978, part II (1979), 331-343.

[10] Stewart, G.W. - Introduction to Matrix Computations, Academic Press, 1973

[11] Strang, Gilbert - Linear Algebra and its Applications, Academic Press, 1976.

MATRICES EN DE THEORIE VAN DYNAMISCHE SYSTEMEN

J. GRASMAN

Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam

1. INLEIDENDE OPMERKINGEN

Bij matrices denkt men onmiddelijk aan lineaire algebra. Het is echter zo dat ook in andere gebieden van de wiskunde matrices een fundamentele rol spelen, zie DE LANGE en KINDT [6]. In deze voordracht zal de funktie van matrices in de theorie van dynamische systemen behandeld worden. Bij lineaire systemen ligt het voor de hand om met matrices en vectoren te werken. Bij niet-lineaire systemen bepalen eigenwaarden van matrices de stabiliteit van oplossingen en het type oplossing dat ontstaat bij bifurcatie. We zullen zowel discrete als continue dynamische systemen beschouwen en de analogie tussen die twee benadrukken. Als voorbeeld zullen we enkele malen op een 1-dimensionaal systeem terugvallen. Hiermee komt de matrix een vector notatie te vervallen. Het is echter zo dat in vergelijkbare werkelijke problemen de andere variabelen veelal passief zijn. Voor het begrijpen van het verschijnsel van omslag van stabiliteit is dan een 1-dimensionaal voorbeeld het meest illustratief.

2. DISCRETE DYNAMISCHE SYSTEMEN

Bij het begrip systeem kunnen we denken aan een mechanisch systeem (een slinger) een electrisch systeem (een schakeling), maar ook aan een ecologisch systeem of een economisch systeem. De toestand van een systeem is vastgelegd in de waarde die aan een aantal variabelen toegekend wordt. Afgezien van externe storingen bepaalt de dynamica van het systeem de

waarden van de toestandsvariabelen voor $t > 0$ gegeven de toestand op $t = 0$. Kijken we naar de waarden op discrete tijden $t = k$, $k = 0, 1, 2, \ldots$ dan wordt het dynamisch gedrag beschreven door een stelsel differentievergelijkingen van het type

$$(1) \qquad x_i(k+1) = F_i(x_1(k),\ldots,x_n(k)), \quad i = 1,\ldots,n,$$

waarin x_i, $i = 1,2,\ldots,n$ de toestandsvariabelen zijn.

Onze kennis van de functies F_i bepaalt hoe goed we de toekomstige waarden van de variabelen kunnen voorspellen. Voor een eenvoudig systeem (een mechanische slinger) zijn deze functies exact af te leiden. Voor b.v. een economisch systeem is dit zeker niet altijd mogelijk. In eerste benadering zal men een lineair verband veronderstellen

$$(2) \qquad x_i(k+1) = \sum_{j=1}^{n} a_{ij}x_j(k) + b_i, \quad i = 1,\ldots,n.$$

Zonder verlies van algemeenheid mogen we $b_i = 0$ nemen. Vergelijking (2) wordt dan in vector vorm

$$(3) \qquad x(k+1) = Ax(k).$$

3. DYNAMICA VAN EEN BIOLOGISCHE POPULATIE

We beschouwen een populatie van een diersoort met leeftijdsopbouw: $x_i(k)$ is het aantal dieren in het i^{de} levensjaar in het jaar k. De vruchtbaarheidscoëfficient en sterfecoëfficient van dieren in leeftijdsklasse i zijn, resp., f_i en s_i. In het jaar $k + 1$ is er een geboorte

$$(4a) \qquad x_1(k+1) = \sum_{j=1}^{n} f_i\, x_i(k).$$

Het aantal dieren van leeftijdsklasse $i > 1$ in jaar $k + 1$ wordt

$$(4b) \qquad x_i(k+1) = (1-s_{i-1})\, x_{i-1}(k).$$

In de vorm (3) vinden we voor de matrix

(5) $$A = \begin{pmatrix} f_1 & \cdots & f_n \\ 1-s_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1-s_{n-1} \end{pmatrix} \cdot 0$$

In de biologische kontekst wordt dit de Leslie matrix genoemd, zie PIELOU [8]. Voor een studie van de invloed van jacht op een zeehondenpopulatie verwijzen we naar FLIPSE en VELING [2] en DE LANGE en VONK [5].

De eigenwaarden λ_i, $i = 1,\ldots,n$ van de matrix zijn bepalend voor de dynamica van de populatie. Voor biologisch realistische waarden van s_i en f_i is de eigenwaarde, welke in absolute waarde genomen het grootst is, altijd reëel en positief. Noem deze eigenwaarde λ_s. Uit de eigenwaarden kan de ontwikkeling van de populatie op lange termijn afgeleid worden. Voor $\lambda_s > 1$ is er een toename. Ook kan een uitspraak gedaan worden over de verdeling van de populatie over de leeftijdsklassen onder de aanname dat $\lambda_s > 0$ en dat de andere eigenwaarden binnen de eenheidscirkel in het complexe vlak liggen. De eigenvector ξ_s behorende bij λ_s met

$$(6) \qquad A\xi_s = \lambda_s \xi_s, \quad |\xi_s| = 1, \quad |x| = \sum_{i=1}^{n} x_i$$

geeft de stationaire leeftijdsopbouw. Voor willekeurige beginwaarden $x(0) = x_0$ gaat de populatie naar de stationaire verdeling:

$$(7) \qquad \lim_{k\to\infty} \frac{x(k)}{|x(k)|} = \xi_s, \quad x(k) = A^k x_0 .$$

Het is ook mogelijk dat meerdere eigenwaarden in absolute waarde gelijk zijn aan λ_s en voldoen aan

$$(8) \qquad \lambda_{j_m} = \lambda_s e^{2\pi i m/r}, \quad m = 1,\ldots,r.$$

In dat geval varieert de verdeling periodiek en hangt af van de beginwaarde. Dit verschijnsel doet zich voor als de matrix A de volgende vorm heeft

$$(9) \qquad A = \begin{pmatrix} 0 & & & 0 & f_n \\ 1-s_1 & \ddots & & & 0 \\ & \ddots & \ddots & & \vdots \\ 0 & & 1-s_{n-1} & & 0 \end{pmatrix} .$$

Een konkreet biologisch voorbeeld hiervan is de wisselende aanwezigheid van tweejarige planten in een gebied.

Het niet realistische aan het lineaire model (3) voor de biologische populatie is dat deze òf uitsterft òf explodeert. Dit kan ondervangen worden door A te laten afhangen van de omvang van de populatie:

$$(10) \qquad x(k+1) = \frac{1}{q} Ax(k), \quad q = 1 + \alpha|x|.$$

Hierdoor neemt ieder element van A af bij een toename van de populatie: zowel de overlevingscoëfficient $1-s_i$ als de vruchtbaarheidscoëfficient f_i worden gedeeld door de factor $1 + \alpha|x|$. We kunnen hierbij denken aan het effect als slechts een beperkte hoeveelheid voedsel voor de populatie beschikbaar is. In dat geval wordt naast de verdeling ook de omvang van de populatie stationair voor $\lambda_s > 1$:

$$(11) \qquad x_s = |x_s|\xi_s , \quad |x_s| = (\lambda_s - 1)/\alpha.$$

4. NIET-LINEAIRE DISCRETE DYNAMISCHE SYSTEMEN

Het systeem (10) is niet-lineair en is een speciaal geval van (1). In z'n algemeenheid kan men stabiliteit van stationaire en periodieke oplossingen van systemen van het type

$$(11) \qquad x(k+1) = F(x(k))$$

analyseren, zie CAPEL [1].

Definitie 1. Het punt p is een periodiek punt met periode k van (11) als

(12) $\quad p = F^{(k)}(p) \quad \text{en } p \neq F^{(j)}(p) \quad \text{voor } 1 \leq j < k,$

waarbij $\quad F^{(k)}(x) = F(F^{(k-1)}(x)) \ , \ F^{(1)}(x) = F(x).$

Definitie 2. Een stationair punt van (11) is een periodiek punt met periode 1.

Definitie 3. Een periodiek punt met periode k van (11) heet asymptotisch stabiel als

(13) $\quad \| F^{(k)}(x) - F^{(k)}(p) \| < \| x-p \|$

voor all x in een omgeving van p en als bovendien

(14) $\quad \lim_{k \to \infty} \| F^{k}(x) - F^{(k)}(p) \| \to 0 \quad (\| \cdot \| \text{ is Euclidische norm}).$

Stelling 1. Een periodiek punt p van (11) is asymptotisch stabiel als de eigenwaarden van de Jacobiaan van F in p in absolute waarde kleiner dan 1 zijn.

Stelling 2. Een periodiek punt p van (11) is instabiel als minstens één van de eigenwaarden van de Jacobiaan van F in p in absolute waarde groter dan 1 is.

De Jacobiaan van F in P is de matrix

(15) $\quad J = \left[\frac{\partial F_i}{\partial x_j} (p) \right]_{n \times n} .$

Men past de stelling toe door lokaal het niet-lineaire systeem te benaderen door een lineair systeem. Substitueer

(16) $\quad x = p + v$

in (11). Dit levert

(17) $\quad v(k+1) = J\, v(k) + R(v(k)).$

De restterm voldoet aan $R(v)/\|v\| \to 0$ voor $v \to 0$ en is te verwaarlozen. Merk op dat voor $R = 0$ systeem (17) lineair is en van het type (3). Als een eigenwaarde op de eenheidscirkel in het complexe vlak ligt dan is er iets bijzonders met het niet-lineaire systeem (11). Het kan niet zo zijn dat de parameters van een "real world" systeem exact een waarde aannemen die hiermee overéénkomt. Wel kan een parameter langzaam variëren en door zo'n kritische waarde gaan. De k-periodieke oplossing p verliest dan z'n stabiliteit en een andere oplossing takt af. De volgende mogelijkheden doen zich voor:

a. Een eigenwaarde neemt de waarde 1 aan: er ontstaat dan een tweede oplossing met periode k.

b. Een eigenwaarde neemt de waarde -1 aan: er ontstaat dan een tweede oplossing met periode 2k.

c. Twee toegevoegd complexe eigenwaarden gaan door de eenheidscirkel. Er ontstaat een quasi-periodieke oplossing of één met periode mk.

5. DE LOGISTISCHE DIFFERENTIEVERGELIJKING

Als voorbeeld nemen we het 1-dimensionale systeem

$$x(k+1) = a\, x(k)\{1-x(k)\}. \tag{18}$$

Voor $0 < a < 1$ is de stationaire oplossing $x_s = 0$ stabiel. Voor $a = 1$ neemt de eigenwaarde van het gelineariseerde systeem

$$v(k+1) = a\, v(k) \tag{19}$$

de waarde 1 aan. De triviale oplossing wordt instabiel en er ontstaat een stabiele stationaire oplossing:

$$x_s = (a-1)/a. \tag{20}$$

Linearisatie levert

$$v(k+1) = (2-a)\, v(k). \tag{21}$$

Voor $a = 3$ wordt de stationaire oplossing (20) instabiel en er ontstaat

een oplossing met periode 2 omdat de eigenwaarde door -1 gaat. Vervolgens passeert a een rij van waarden a_k waarvoor nieuwe oplossingen met verdubbelde periode 2^k ontstaan. Voorbij het punt $a_\infty \sim 3.57$ is de dynamica van het systeem zeer gecompliceerd en wordt wel "chaotisch" genoemd, zie MAY [7].

6.CONTINUE DYNAMISCHE SYSTEMEN

Indien van de variabelen een grote reeks waarden over een langere periode beschouwd wordt en als bovendien de variabelen over één discrete tijdstap weinig veranderen, dan ligt het voor de hand om op een continu systeem over te gaan. Stel

$$x_i(\tau+1) = x_i(\tau) + \delta f_i(x_1,\ldots,x_n), \quad i = 1,\ldots,n \tag{22}$$

met $0 < \delta << 1$. We voeren een tijdschaal $t = \tau\,\delta$ in, zodat

$$x_i(t+\delta) = x_i(t) + \delta f_i(x_1,\ldots,x_n). \tag{23}$$

Voor $\delta \to 0$ gaat het stelsel differentievergelijkingen (23) over in een stelsel differentiaalvergelijkingen:

$$\frac{dx_i}{dt} = f_i(x_1,\ldots,x_n), \quad i = 1,\ldots,n \quad \text{of} \quad \frac{dx}{dt} = f(x). \tag{24}$$

Voor een stationaire oplossing x_s geldt

$$f(x_s) = 0. \tag{25}$$

De stabiliteit van deze oplossing wordt onderzocht aan de hand van het gelineariseerde systeem dat gevonden wordt na substitutie van

$$x = x_s + v. \tag{26}$$

Het gelineariseerde systeem wordt

$$\frac{dv}{dt} = J\,v, \qquad J = \left[\frac{\partial f_i}{\partial x_i}(x_s)\right]_{n\times n}. \tag{27}$$

Er is een reguliere transformatie v = Hy zodanig dat

$$H^{-1}JH = D \tag{28}$$

waarin D een diagonaal matrix is met de eigenwaarden van J op de diagonaal. De vectorfunctie y(t) voldoet aan

$$\frac{dy}{dt} = Dy \quad \text{of} \quad \frac{dy_i}{dt} = \lambda_i y_i \ , \ i = 1,\ldots,n. \tag{29}$$

De componenten y_i van het nieuwe systeem zijn dus ontkoppeld. Als $\text{Re}\,\lambda_i < 0$ voor alle i dan gaat $y(t) \to 0$ voor $t \to \infty$. Een verstoring v(t) zal dus uitdempen. Als voor één (of meer) eigenwaarden λ_i geldt $\text{Re}\,\lambda_i > 0$ dan gaat $y_i(t) \to \infty$ en een verstoring v(t) zal groeien. Meer over de theorie van differentiaalvergelijkingen van het type (24) wordt gevonden in HALE [3].

7. BIFURCATIE

Bij een continu niet-lineair dynamisch systeem kan eveneens kritische afhankelijkheid van een parameter optreden, zie [4]. Beschouw het systeem (24) met een parameter μ

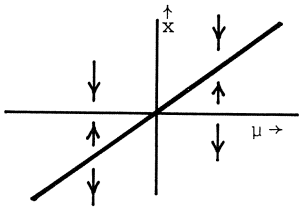
$$\frac{dx}{dt} = f(x;\mu). \tag{30}$$

We volgen een stabiele stationaire oplossing $x_s(\mu)$, waarvoor de parameter μ een kritische waarde μ_c nadert. Essentiële informatie over het systeem in de omgeving van dat punt kan verkregen worden uit de Jacobiaan J, zie (27). Het volgende is mogelijk:

a. Een reële eigenwaarde wordt positief: er takken één of twee nieuwe oplossingen af (bifurcatie).

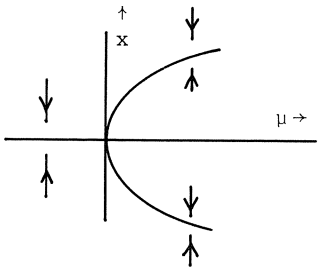
b. Twee toegevoegd complexe eigenwaarden gaan door de imaginaire as: er takt een periodieke oplossing af (Hopf bifurcatie).

Een derde mogelijkheid is dat voorbij het kritische punt geen enkele oplossing bestaat. In fig. 1 geven we een aantal eenvoudige 1-dimensionale systemen, welke de diverse mogelijkheden weergeven.



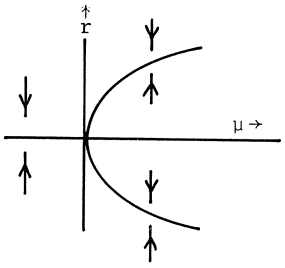
(a) $dx/dt = x(\mu-x)$,

$x_{s_1} = 0, x_{s_2} = \mu$.



(b) $dx/dt = \mu x - x^3$,

$x_{s_1} = 0, x_{s_2} = \sqrt{\mu}, x_{s_3} = -\sqrt{\mu}$.

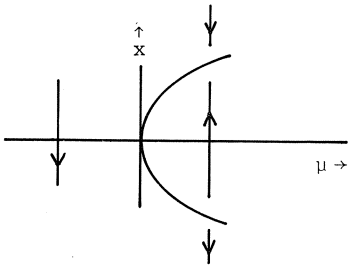


(c) $dr/dt = \mu r - r^3$,

$d\theta/t = 1$,

$x_1 = r \cos\theta, x_2 = r \sin\theta$,

$x_{s_1} = 0, r_{s_2} = \sqrt{\mu}$.



(c) $dx/dt = \mu - x^2$,

$x_{s_1} = \sqrt{\mu}, x_{s_2} = -\sqrt{\mu}$

Fig. 1. Voorbeelden van kritische afhankelijk van de parameter μ van de differentiaalvergelijking (30).

OPGAVEN

1. Geef de transformatie van toestandsvariabelen welke systeem (2) in de vorm (3) brengt.

2. Van een systeem met n variabelen hangt de toestand op het tijdstip t=k+1 af van de toestand op de tijdstippen t=k én t=k-1. Ontwerp een 2n-dimensionaal dynamisch systeem van het type (1) waarin deze afhankelijkheid tot uitdrukking komt.

3. Gegeven de Lesliematrix

$$A = \begin{pmatrix} 1 & 3/2 \\ 1/2 & 0 \end{pmatrix}.$$

Bepaal de stationaire verdeling over de twee leeftijdsklassen. Stel op t=0 zijn er 100 examplaren van leeftijdsklasse 1 en is er geen enkel exemplaar van leeftijdsklasse 2. Hoe is de verdeling na 5 jaar? Vergelijk het antwoord met de gevonden stationaire verdeling.

4. Bestudeer het gedrag van de logistische differentievergelijking (18) voor $k \to \infty$ met een programmeerbare zakrekenmachine of microcomputer. Neem waarden voor a op het interval (0,4) en een vaste startwaarde x(0) op het interval (0,1).

5. Verifieer formule (11).

6. Beschouw het discrete dynamische systeem (prooi-predator model)

$$x(k+1) = x(k)\{a - x(k) - y(k)\},$$

$$y(k+1) = b\, x(k)\, y(k)$$

voor $0 < a \leq 4$, $0 < b \leq 4/a$ en $x,y \geq 0$, $x+y < a$. Bepaal de stationaire punten en onderzoek de stabiliteit ervan. Ga na welk type oplossingen aftakken bij omslag van stabiliteit van deze oplossingen. Analyseer het systeem op een microcomputer door afbeelding van de rij punten $\{x(k), y(k)\}$ in het x,y-vlak. Varieer de parameters a en b (zie J.R. Beddington, C.A. Free and J.H. Lawton, Nature 255 (1975), 58-60).

7. Beschouw het continue dynamische systeem

$$\frac{dx}{dt} = a(y - \frac{1}{3}x^3 + x)$$

$$\frac{dy}{dt} = -x + b$$

voor $0 < a < 4$ en $0 < b < \sqrt{2}$. Bepaal het stationaire punt en onderzoek de stabiliteit ervan. Wat voor type oplossing takt af als de stabiliteit omslaat? Integreer het systeem numeriek voor a=3 en voor diverse waarden van b. Neem één vaste startwaarde en beeld de oplossingscurve in het x,y-vlak af.

8. Schrijf het systeem van Fig. 1c in de vorm van (30).

LITERATUUR

[1] CAPEL, P.H.B., *Niet-lineaire differentievergelijkingen*, doctoraalscriptie Univ. v. Amsterdam, 1977.

[2] FLIPSE, E. & E.J.M. VELING, *An application of the Leslie Matrix model to the population dynamics of the hooded seal* (Cystophora Cristata exleben), Report TN 101, Mathematisch Centrum Amsterdam, 1981.

[3] HALE, J.K., *Ordinary differential equations*, Interscience, New York, 1969.

[4] IOOSS, G. & D.D. JOSEPH, *Elementary stability and bifurcation theory*, Springer-Verlag, Berlin, 1980.

[5] DE LANGE, J. & G.A. VONK, *Klapmutsen in gevaar?* Nieuwe Wiskrant, 2e jaarg. no. 2, nov. 1982, p.44-52.

[6] DE LANGE, J. & M. KINDT, *Matrices*, OW & OC, Utrecht, 1983.

[7] MAY, R.M., *Simple mathematical models with very complicated dynamics*. Nature 261 (1976), p. 459-467.

[8] PIELOU, E.C., *An introduction to mathematical ecology*, Wiley-Interscience, New York, 1969.

STELSELS LINEAIRE ONGELIJKHEDEN,

MARKOV KETENS EN MATRIXSPELEN

S.H. TIJS

1. INLEIDING

Stelsels lineaire vergelijkingen hebben al eeuwen geleden de aandacht van onderzoekers getrokken. Resultaten in dit gebied werden al door Leibniz geboekt, zoals uit een brief van 28 april 1639 aan l'Hopital blijkt. Belangrijk voor de ontwikkeling is vooral een werkstuk van Cramer uit 1750 waarin stelsels van 3 vergelijkingen met 3 onbekenden aangepakt worden en het belang van determinanten aan het licht komt. Ook in het middelbare onderwijs is er al lang een plaats voor lineaire vergelijkingen ingeruimd. De kennis over stelsels lineaire vergelijkingen heeft ondermeer bijgedragen tot de ontwikkeling van lineaire algebra en functionaalanalyse.

Als we nu kijken naar de theorie van stelsels lineaire ongelijkheden dan moeten we constateren dat deze lange tijd door wiskundigen als een stiefkind is behandeld. Wel zijn er al belangrijke bijdragen ondermeer van Fourier [4] in 1826 en van Farkas [2] in 1902, maar deze trekken geen breed aandachtsveld. Opmerkelijk is in dit verband dat T. Motzkin, die een systematisch literatuuronderzoek heeft gedaan voor stelsels lineaire ongelijkheden, in zijn dissertatie [6] uit 1936 slechts een veertigtal artikelen kan vermelden. Na de Tweede Wereldoorlog komt hier verandering in met de opkomst van de lineaire programmeringstheorie (zie [1]). Dit is niet verwonderlijk omdat hier het probleem centraal staat om lineaire funkties te maximaliseren op gebieden die vastgelegd zijn door stelsels vergelijkingen en ongelijkheden. Mede door de vele toepassingsmogelijkheden van deze LP-theorie in economie en operations research is ook de theorie van stelsels lineaire ongelijkheden meer

centraal komen te staan en het lijkt dan ook terecht dat nu ook aandacht aan deze theorie wordt besteed in het middelbare onderwijs. Een van de redenen waarom in het verleden de lineaire ongelijkhedentheorie weinig weerklank vond ligt misschien in het feit dat er geen mooie methode was om oplossingen voor zulke stelsels te vinden. Dat ligt nu anders zoals D. Gale ([5], p. 105) opmerkt: de roemruchte simplexmethode, ontwikkeld in 1947 om LP-problemen op te lossen, is ook een uitstekende methode om oplossingen te vinden voor stelsels lineaire ongelijkheden of niet-negatieve oplossingen van stelsels vergelijkingen.

In deze bijdrage komen in paragraaf 2 enkele stellingen over het bestaan van oplossingen van stelsels van lineaire ongelijkheden aan de orde zoals de alternatiefstelling van Farkas. Met behulp hiervan wordt in paragraaf 3 aangetoond dat er voor elke vierkante stochastische matrix een stationaire verdeling bestaat. Eveneens wordt in paragraaf 4 de minimaxstelling voor matrixspelen met behulp van een alternatiefstelling bewezen. Een elementair bewijs van de stelling van Farkas wordt gegeven in de laatste paragraaf.

2. ALTERNATIEFSTELLINGEN VOOR STELSELS LINEAIRE ONGELIJKHEDEN

Bij het beschrijven van stelsels vergelijkingen zijn matrices nuttig. Dit is ook zo voor stelsels ongelijkheden. Als $A = [a_{ij}]_{i=1,j=1}^{m\ \ n}$ een $m\times n$-matrix is en $b \in \mathbb{R}^m$, dan is $Ay \leq b$ een korte schrijfwijze voor het stelsel van m ongelijkheden

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n \leq b_1 \\ \qquad\qquad \vdots \\ a_{i1}y_1 + a_{i2}y_2 + \ldots + a_{in}y_n \leq b_i \\ \qquad\qquad \vdots \\ a_{m1}y_1 + a_{m2}y_2 + \ldots + a_{mn}y_n \leq b_m \end{cases}$$

met onbekenden $y_1, y_2, \ldots, y_n$.
Hebben we ongelijkheden van verschillend type en vergelijkingen, dan zijn er voor de hand liggende manieren om zulke stelsels om te zetten in een standaardstelsel van een eenvoudige vorm. Zo is bijvoorbeeld het stelsel

$$Ay \leq b\ ,\ By \geq c\ ,\ Cy = d$$

equivalent met $Dy \leq e$, waarbij

$$D = \begin{bmatrix} A \\ -B \\ C \\ -C \end{bmatrix} \quad \text{en} \quad e = \begin{bmatrix} b \\ -c \\ d \\ -d \end{bmatrix}.$$

Evenzo komt het vinden van oplossingen van het stelsel ongelijkheden

$$Ay \leq b \ , \ y \geq 0$$

neer op het vinden van oplossingen van

$$Ay + Iz = b \ , \ y \geq 0 \ , \ z \geq 0$$

waarbij alleen tekenongelijkheden optreden. Hierbij is I de identieke $m \times m$-matrix als m het aantal rijen van A is.

Bij de beschrijving van oplossingsruimten van stelsels lineaire vergelijkingen spelen zoals bekend lineaire deelruimten en varieteiten een rol. Deze rol wordt bij stelsels lineaire ongelijkheden overgenomen door convexe kegels en convexe verzamelingen van een speciaal type. We gaan hierop niet verder in maar verwijzen naar Gale [5] en Motzkin [6].

De rest van deze paragraaf besteden we aan alternatiefstellingen. In zulke stellingen treden steeds twee verwante stelsels (on)gelijkheden op en doet zich de situatie voor dat het ene stelsel een oplossing bezit precies dan als het andere stelsel geen oplossing heeft en omgekeerd. Laten we eerst een welbekend feit uit de theorie van lineaire vergelijkingen formuleren als een alternatiefstelling.

STELLING 1. Zij A een $m \times n$-matrix en $b \in \mathbb{R}^n$. Precies één van de volgende twee uitspraken is waar.

(i) Het stelsel $xA = b$ heeft een oplossing.

(ii) Het stelsel $Ay = 0$, $b.y = -1$ heeft een oplossing.

[Hierbij is b.y het inprodukt van b en y.]

De meetkundige interpretatie van deze stelling is de volgende: òf b ligt in de lineaire ruimte V, opgespannen door de rijen van A, òf er is een vektor y in het orthoplement van V die een stompe hoek maakt met b.

BEWIJS VAN STELLING 1. Zoals bekend bezit het stelsel vergelijkingen $xA = b$ een oplossing precies dan als $\text{rang}(A) = \text{rang}\binom{A}{b}$. Evenzo bezit het stelsel $Ay = 0$, $b.y = -1$ een oplossing precies dan als $\text{rang}\binom{A}{b} = \text{rang}\binom{A \quad 0}{b \quad -1}$. Maar uit

$$1 + \text{rang}(A) = \text{rang}\binom{A \quad 0}{b \quad -1} \geq \text{rang}\binom{A}{b} \geq \text{rang}(A)$$

volgt dat precies één van de volgende gelijkheden geldt:

$$\text{rang}\begin{pmatrix} A \\ b \end{pmatrix} = \text{rang } A \quad , \quad \text{rang}\begin{pmatrix} A \\ b \end{pmatrix} = \text{rang } \begin{pmatrix} A & 0 \\ b & -1 \end{pmatrix}. \qquad \square$$

De volgende alternatiefstelling is afkomstig van Farkas en zal in de rest van het verhaal een belangrijke rol spelen. Een bewijs vindt U in paragraaf 5.

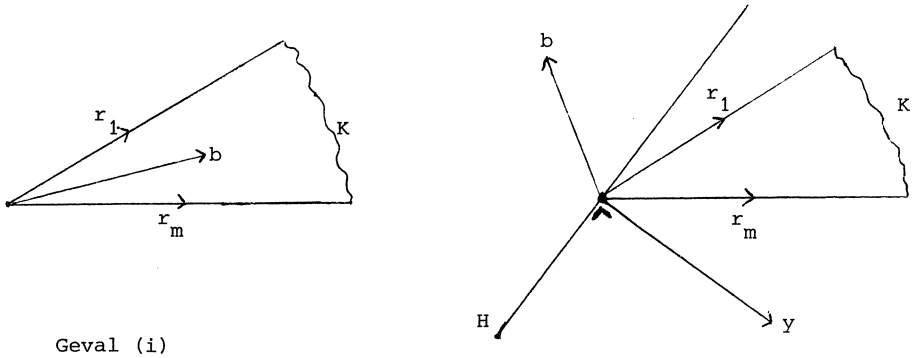
STELLING 2. Zij A een m×n-matrix en $b \in \mathbb{R}^n$. Precies één van de volgende twee uitspraken is waar.

(i) Het stelsel $xA = b$, $x \geq 0$ heeft een oplossing.

(ii) Het stelsel $Ay \geq 0$, $b.y < 0$ heeft een oplossing.

We kunnen de uitspraken in stelling 2 als volgt meetkundig interpreteren.

Fig. 1



Geval (i)

Geval (ii)

Uitspraak (i) zegt dat b een niet-negatieve combinatie is van de rijen $r_1 := e^1A$, $r_2 := e^2A, \ldots, r_m := e^mA$ van A, m.a.w. b zit in de convexe kegel, voortgebracht door de rijen van A (zie fig. 1, geval (i)). Uitspraak (ii) betekent dat er een vektor y bestaat die scherpe (d.i. niet stompe) hoeken maakt met de rijen van A en een strikt stompe hoek met b. Uitspraak (ii) kan ook als volgt vertaald worden: er is een vektor y zo dat voor het hypervlak $H := \{z \in \mathbb{R}^n : z.y = 0\}$ geldt dat de kegel K voortgebracht door de rijen van A aan één kant ligt van dit hypervlak en de vektor b strikt aan de andere kant (zie fig. 1, geval (ii)).

Een vrijwel rechtstreeks gevolg van stelling 2 is de volgende alternatiefstelling waarvan we in paragraaf 4 gebruik zullen maken.

STELLING 3. Zij A een m×n-matrix en b een vektor in $\mathbb{R}^n$. Precies één van de volgende twee uitspraken is waar.

(i) Het stelsel $xA \geq b$, $x \geq 0$ heeft een oplossing.

(ii) Het stelsel $Ay \leq 0$, $y \geq 0$, $b.y > 0$ heeft een oplossing.

BEWIJS. Het is duidelijk dat niet beide stelsels een oplossing kunnen hebben, want als $\hat{x}$ een oplossing van het stelsel uit (i) zou zijn en $\hat{y}$ een oplossing van het stelsel uit (ii), dan

$$0 < b.\hat{y} \leq \hat{x}A\hat{y} \leq \hat{x}.0 = 0$$

en dat is onmogelijk. Veronderstel nu dat het stelsel in (i) geen oplossing bezit. Dan bezit ook het stelsel

$$xA - zI = b,\ x \geq 0,\ z \geq 0$$

geen oplossing. Uit stelling 2 met $\begin{pmatrix} A \\ -I \end{pmatrix}$ in de rol van A volgt dan dat er een $u \in \mathbb{R}^n$ is met

$$Au \geq 0,\ -Iu \geq 0,\ b.u < 0.$$

Maar dan is $y := -u$ een oplossing van het stelsel in (ii). □

Meetkundig betekent uitspraak (i) van stelling 3 dat de convexe kegel K, opgespannen door de rijen van A iets gemeen heeft met het orthant $O_b := \{x \in \mathbb{R}^n : x \geq b\}$. Uitspraak (ii) betekent dat er een hypervlak H door de oorsprong is met normaal $y \geq 0$, $y \neq 0$, zo dat K aan één kant van H ligt en het orthant O_b strikt aan de andere kant.

3. STATIONAIRE VERDELINGEN VOOR MARKOV KETENS

Aan de orde komt een eigenschap voor vierkante stochastische matrices. Zo'n *stochastische matrix* $[a_{ij}]_{i=1,j=1}^{n\ \ n}$ is opgebouwd uit niet-negatieve reële getallen en elke rijsom is gelijk aan 1, dus

$$a_{ij} \geq 0 \text{ voor elke } i,j \in \{1,2,\ldots,n\} \tag{3.1}$$

$$\sum_{j=1}^{n} a_{ij} = 1 \text{ voor elke } i \in \{1,2,\ldots,n\} \tag{3.2}$$

Stochastische n×n-matrices treden op bij het beschrijven van stochastische processen welke bekend staan onder de naam *Markov ketens*.

We kunnen hierbij denken aan een systeem dat op tijdstippen $t = 0,1,2,\ldots$ geobserveerd wordt en dan steeds in één van de n toestanden $T_1,T_2,\ldots,T_n$ verkeert en waarbij de overgang van de toestand op tijdstip t naar de toestand op tijdstip t+1 stochastisch is, maar alleen afhangt van de toestand op tijdstip t. De overgangen van tijdstip t naar tijdstip t+1 worden dan beschreven door middel van de stochastische matrix $A(t) = [a_{ij}(t)]_{i=1,j=1}^{n\quad n}$ waarbij $a_{ij}(t)$ de kans is dat het systeem zich op tijdstip t+1 in toestand T_j bevindt, gegeven dat het systeem zich op tijdstip t in T_i bevindt. A(t) heet de *overgangsmatrix* op tijdstip t. Voor de berekening van de kansverdelingen van de stochasten $X_2,X_3,\ldots,X_t,\ldots$, waarbij X_t de toestand van het systeem is ten tijde t, is het voldoende de overgangsmatrices $A(1),A(2),\ldots$ te kennen en de kansverdeling van X_1. Men gaat gemakkelijk na (een aardige toepassing van matrixvermenigvuldiging) dat de j-de coordinaat van de vektor

$$x(t) := x(1)A(1)A(2) \ldots A(t-1) \qquad (t \geq 2)$$

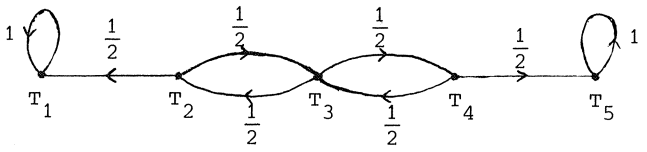
de kans is dat het systeem op tijdstip t in toestand T_j verkeert als de kansverdeling op tijdstip 1 gegeven wordt door $x(1) = (x_1(1),x_2(1),\ldots,x_n(1))$, waarbij $x_i(1)$ de kans is dat op tijdstip 1 het systeem in toestand T_i is $(i = 1,2,\ldots,n)$.

We zullen onze aandacht verder beperken tot *Markov ketens met stationaire overgangskansen*. Dit zijn Markov ketens, waarbij alle overgangsmatrices $A(1),A(2),\ldots$ gelijk zijn, zeg aan A. Voor zulke stationaire Markov ketens is het interessant te weten onder welke voorwaarden voor A de limiet $\lim_{t\to\infty} x(1)A^t$ bestaat. We gaan hier niet nader op in maar verwijzen voor meer informatie over Markov ketens naar [3] en [10]. Wel merken we op dat als zo'n limietverdeling $z = \lim_{t\to\infty} x(1)A^t$ bestaat, de limiet z een kansvektor is $(z \geq 0, \sum_{i=1}^{n} z_i = 1)$ met $zA = z$.
Wat we in deze paragraaf willen bewijzen is dat er voor elke stochastische matrix een *stationaire verdeling* bestaat d.i. een kansvektor met $zA = z$. Als dan de stochast X_1 in de Markov keten met overgangsmatrix A de kansverdeling z bezit, dan is dat ook het geval voor de kansverdelingen van $X_2,X_3,\ldots$ omdat voor de kansverdeling van X_t geldt: $zA^{t-1} = z$.

Allereerst enkele voorbeelden.

VOORBEELD 1. Een speler op een roulette zonder zero start met een beginkapitaal van *f* 200,- en zet telkens *f* 100,- op rood totdat zijn kapi-

taal aangegroeid is tot ƒ 400,- of totdat hij failliet is. Deze situatie correspondeert met een Markov keten met stationaire overgangskansen en met 5 toestanden



waarbij T_i de toestand is dat het kapitaal van de speler (i-1) 100 gulden groot is en waarbij de overgangsmatrix gegeven wordt door

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

De bijbehorende stochasten $X_1, X_2, X_3, X_4, \ldots$ hebben achtereenvolgens de kansverdeling

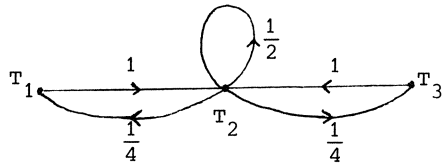
$$(0,0,1,0,0) \ , \ (0,\tfrac{1}{2},0,\tfrac{1}{2},0) \ , \ (\tfrac{1}{4},0,\tfrac{1}{2},0,\tfrac{1}{4}) \ , \ (\tfrac{1}{4},\tfrac{1}{4},0,\tfrac{1}{4},\tfrac{1}{4}) \ , \ \ldots$$

en $(\frac{1}{2},0,0,0,\frac{1}{2})$ is de limietverdeling. De verzameling van stationaire verdelingen bestaat uit de vektoren van de vorm (a,0,0,0,1-a) met $0 \leq a \leq 1$.

VOORBEELD 2. Op tijdstip 1 bevinden zich 2 witte knikkers in doos D en 2 zwarte knikkers in doos E. Op elk der tijdstippen t = 2,3,4,... wordt gelijktijdig uit elk der dozen blindelings een knikker genomen en in de andere doos gedaan. Als we geinteresseerd zijn in de inhoud van de dozen in de loop der tijd, dan is het handig om de situatie te vertalen in een Markov keten met 3 toestanden T_1, T_2, T_3, waarbij T_i de toestand is dat er zich i-1 witte knikkers in doos D bevinden (i = 1,2,3), en waarbij de overgangsmatrix wordt gegeven door

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{bmatrix} .$$

De netwerkrepresentatie van deze Markov keten is dan



Men rekent gemakkelijk na dat de enige stationaire verdeling voor deze Markov keten $(\frac{1}{6},\frac{2}{3},\frac{1}{6})$ is. Merk op dat het beeld van een vektor van de vorm (c,d,c) onder A gelijk is aan $(d,\frac{1}{2}(c+d),d)$. Met behulp hiervan kan men dan bewijzen dat voor elke $n \in \mathbb{N}$:

$$e^2A^{n+1} = \frac{1}{2}(e^1+e^2)A^n \ , \ e^1A^{n+1} = e^3A^{n+1} = e^2A^n \qquad (3.3)$$

Gebruikmakend van (3.3) volgt door volledige induktie dat voor elke $n \in \mathbb{N}$:

$$e^2A^n = e^1A + (\sum_{k=0}^{n-1} (-1)^k \frac{1}{2^k})(e^2-e^1)A.$$

Dan $\lim_{n\to\infty} e^iA^n = e^1A + \frac{2}{3}(e^2-e^1)A = (\frac{1}{6},\frac{2}{3},\frac{1}{6})$ voor elke $i \in \{1,2,3\}$.
Hieruit volgt dat voor elke beginverdeling z (dus ook voor onze beginverdeling $\hat{e}^1$) de limietverdeling $\lim_{n\to\infty} zA^n$ gelijk is aan de unieke stationaire verdeling $(\frac{1}{6},\frac{2}{3},\frac{1}{6})$.

Dan is het nu tijd voor de stelling van Markov met een bewijs dat gebruik maakt van de alternatiefstelling van Farkas.

STELLING 4. Zij A een stochastische n×n-matrix. Dan is er een kansvektor x zo dat xA = x.

BEWIJS. Een kansvektor x met xA = x bestaat precies dan als het stelsel

$$x(A-I, \ 1_n) = (0,0,\ldots 0,1) \ , \ x \geq 0 \qquad (3.4)$$

een oplossing bezit, waarbij I de identieke matrix van afmeting n×n is en 1_n de (kolom)vektor met alle n coordinaten gelijk aan 1. Veronderstel eens dat dit stelsel geen oplossing bezit. Dan volgt uit stelling 2 dat er een vektor $y = (y_1,y_2,\ldots,y_n,y_{n+1}) \in \mathbb{R}^{n+1}$ te vinden is met

$$(A-I, \ 1_n)y \geq 0 \text{ en } y_{n+1} = (0,0,\ldots,0,1).y < 0.$$

Dan geldt voor $z := (y_1,y_2,\ldots,y_n)$:

$$Az - z + y_{n+1} \ 1_n \geq 0 \text{ en dus } Az > z.$$

Maar dan volgt voor elke $i \in \{1,2,\ldots,n\}$ vanwege (3.1) en (3.2):

$$z_i < \sum_{j=1}^{n} a_{ij} z_j \leq (\max_j z_j) \sum_{j=1}^{n} a_{ij} = \max_j z_j$$

en dat is onmogelijk. Dus het stelsel in (3.4) bezit een oplossing. □

Merk op dat we stelling 4 ook als volgt kunnen formuleren:
elke stochastische matrix bezit een niet-negatieve eigenvektor bij de eigenwaarde 1.

4. DE MINIMAXSTELLING VOOR MATRIXSPELEN

Vele conflictsituaties blijken herleidbaar tot een *matrixspel*. (Zie [7], [8] en [9]) Gegeven een m×n-matrix $U = [u_{ij}]_{i=1,j=1}^{m\ \ n}$, dan verloopt een partijtje van het bijbehorende matrixspel als volgt. Speler I (de *rijenspeler*) kiest een rij van de matrix -zeg rij i, en speler II (de *kolommenspeler*) een kolom -zeg kolom j. Vervolgens krijgt speler I een uitbetaling u_{ij} van speler II. Voor het matrixspel $U = \begin{bmatrix} 0 & -1 \\ 4 & 5 \end{bmatrix}$ is het duidelijk hoe te spelen. Speler I zal rij 2 kiezen en speler II kolom 1. De uitkering 4 noemen we de waarde van dit spel en i = 2 (j = 1) een optimale zuivere strategie voor speler I (speler II). Wat de waarde (voor speler I) in het matrixspel $U = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ is, is niet direct duidelijk. We veronderstellen dat de spelers in zo'n matrixspel *gemengde strategieën* mogen gebruiken. Zo'n strategie voor speler I is bijvoorbeeld de kansvektor $(\frac{1}{3}, \frac{2}{3})$, welke correspondeert met de gemengde strategie: "kies rij 1 met kans $\frac{1}{3}$ en rij 2 met kans $\frac{2}{3}$". In bovenstaand voorbeeld blijkt de gemengde strategie $(\frac{1}{2}, \frac{1}{2})$ optimaal te zijn en de waarde van het spel 4½. Laten we één en ander preciseren.

Zij U een m×n-matrixspel. Dan heet $\Delta^m := \{p \in \mathbb{R}^m : p \geq 0, \sum_{i=1}^{m} p_i = 1\}$ de verzameling van *gemengde strategieën voor speler* I en $\Delta^n := \{q \in \mathbb{R}^n : q \geq 0, \sum_{j=1}^{n} q_j = 1\}$ de verzameling *gemengde strategieën voor speler* II. Voor $p \in \Delta^m$ en $q \in \Delta^n$ is pUq de *verwachte uitbetaling* als speler I strategie p en speler II strategie q kiest.

We willen in het volgende aantonen dat er een uniek bepaald getal v(U) bestaat waarvoor er gemengde strategieën $\hat{p}$ en $\hat{q}$ te vinden zijn zo dat

$$\hat{p}Uq \geq v(U) \quad \text{voor alle} \quad q \in \Delta^n \tag{4.1}$$

$$p\hat{U}\hat{q} \leq v(U) \quad \text{voor alle} \quad p \in \Delta^m . \tag{4.2}$$

Het getal v(U) heet dan de *waarde* van het matrixspel U en de strategieën $\hat{p}$ en $\hat{q}$ *optimale strategieën* voor achtereenvolgens speler I en speler II. De formule (4.1) drukt uit dat speler I, gewapend met strategie $\hat{p}$, zichzelf een verwachte uitkering van tenminste v(U) kan garanderen, welke strategie speler II ook kiest. Formule (4.2) vertelt dat speler II zijn verwachte betaling aan speler I met behulp van $\hat{q}$ (zwak) beneden v(U) kan houden.
Dit resultaat verkreeg John von Neumann in 1928 maar zijn eerste bewijs was erg gecompliceerd. Wij zullen, gebruikmakend van alternatiefstelling 3, een simpel bewijs geven.

Maar eerst behandelen we nog een voorbeeld.

VOORBEELD 3. Bekijk het spelletje waarbij speler II één van de getallen uit {1,2,3} opschrijft en speler I raadt welk getal opgeschreven is. Neem aan dat speler I j gulden krijgt van speler II als hij goed raadt dat speler II getal j heeft opgeschreven (j = 1,2,3) en dat in andere gevallen speler I aan speler II ƒ 1,- moet betalen. Dan correspondeert deze situatie met een 3×3-matrixspel met uitbetalingsmatrix

$$U = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix} .$$

De waarde van dit spel is $-\frac{1}{13}$ en $p := (\frac{6}{13}, \frac{4}{13}, \frac{3}{13})$ is een optimale gemengde strategie zowel voor speler I als voor speler II omdat men gemakkelijk nagaat dat $pU = (-\frac{1}{13}, -\frac{1}{13}, -\frac{1}{13})$ en $Up = -\frac{1}{13}\,1_3$.

LEMMA 5. Zij $U = [u_{ij}]_{i=1,j=1}^{m\ \ n}$ een m×n-matrix en α een reëel getal. Dan is precies één van de volgende twee uitspraken juist.
(i) Er is een $p \in \Delta^m$ met $pU \geq \alpha 1_n$
(ii) Er is een $q \in \Delta^n$ met $Uq < \alpha 1_m$.

BEWIJS. Veronderstel dat (i) onjuist is. Dan heeft het stelsel

$$p(U, 1_m, -1_m) \geq (\alpha, \alpha, \ldots, \alpha, 1, -1) \ , \ p \geq 0$$

geen oplossingen. Uit de alternatiefstelling 3 volgt dan dat er een $y = (y_1, y_2, \ldots, y_n, y_{n+1}, y_{n+2}) \in \mathbb{R}^{n+2}$ is met

$$(U, 1_m, -1_m)y \leq 0 \ , \ y \geq 0 \ , \ \alpha \sum_{j=1}^{n} y_j + y_{n+1} - y_{n+2} > 0.$$

Als $z := (y_1, y_2, \ldots, y_n) \geq 0$, dan volgt

$$Uz \leq (y_{n+2}-y_{n+1})1_m < (\alpha \sum_{j=1}^{n} y_j)1_m$$

en dan $\sum_{j=1}^{n} z_j = \sum_{j=1}^{n} y_j \neq 0$. Neem $q := (\sum_{j=1}^{n} z_j)^{-1} z$. Dan $Uq < \alpha 1_m$ en $q \in \Delta^n$. We hebben dus aangetoond dat (ii) juist is als (i) niet juist is. Het is duidelijk dat niet zowel (i) als (ii) juist kunnen zijn omdat uit $pU \geq \alpha 1_n$, $p \in \Delta^m$, $Uq < \alpha 1_m$, $q \in \Delta^n$ volgt

$$pUq \geq \alpha 1_n \cdot q = \alpha \ , \ pUq < p \cdot \alpha 1_m = \alpha$$

en dat is onmogelijk. □

We kunnen de volgende speltheoretische interpretatie aan bovenstaand lemma geven. Zij U een matrixspel en $\alpha \in \mathbb{R}$. Dan is precies één van de volgende twee uitspraken juist.

(i) Er is een gemengde strategie voor speler I die garandeert dat de verwachte uitbetaling voor speler I tenminste α is, welke strategie speler II ook kiest.

(ii) Er is een gemengde strategie voor speler II die zijn verwachte uitbetaling beneden α houdt, welke strategie speler I ook kiest.

Bekijk nu bij het m×n-matrixspel U de verzamelingen

$$L := \{\alpha \in \mathbb{R} : \text{er is een } p \in \Delta^m \text{ met } pU \geq \alpha 1_m\},$$

$$H := \{\alpha \in \mathbb{R} : \text{er is een } q \in \Delta^n \text{ met } Uq < \alpha 1_n\}.$$

Dan volgt direkt uit lemma 5 dat

(a) $L \cap H = \emptyset$ en $L \cup H = \mathbb{R}$.

Verder gelden:

(b) Als $\alpha \in L$ en $\beta < \alpha$, dan $\beta \in L$.

(c) Als $\alpha \in H$ en $\beta > \alpha$, dan $\beta \in H$.

(d) H is een open verzameling.

(e) $L \neq \emptyset$ en $H \neq \emptyset$ want $\alpha_1 := \min_{i,j} u_{ij} \in L$ en

$$\alpha_2 := 1 + \max u_{ij} \in H.$$

Uit (a) - (e) kunnen we concluderen dat er een getal $v \in \mathbb{R}$ is zo dat $L = (-\infty, v]$ en $R = (v, \infty)$.

Maar dan is er een $\hat{p} \in \Delta^m$ met $\hat{p}U \geq v 1_n$ en voor elke $k \in \mathbb{N}$ is er een $q(k) \in \Delta^n$ zo dat $Uq(k) < (v+\frac{1}{k})1_m$.

Aangezien Δ^n een compacte deelverzameling van $\mathbb{R}^n$ is bestaat er een convergente deelrij van $q(1), q(2), q(3), \ldots$. Zij $\hat{q}$ de limiet van zo'n deelrij. Dan volgt dat $\hat{q} \in \Delta^n$ en $U\hat{q} \leq v 1_m$. Hiermee is de eerste uitspraak

in de volgende stelling bewezen.

STELLING 6. Zij U een m×n-matrix. Dan geldt:

(i) Er is een $v \in \mathbb{R}$ en $\hat{p} \in \Delta^m$, $\hat{q} \in \Delta^n$ zo dat

$$\hat{p}U \geq v\,1_n \ , \ U\hat{q} \leq v\,1_m \tag{4.3}$$

(ii) $\max_{p\in\Delta^m} \min_{q\in\Delta^n} pUq = \min_{q\in\Delta^n} \max_{p\in\Delta^m} pUq$

(iii) Voor elk drietal $(v,\hat{p},\hat{q}) \in \mathbb{R} \times \Delta^m \times \Delta^n$ dat voldoet aan (4.3) geldt:

$$v = \max_{p\in\Delta^m} \min_{q\in\Delta^n} pUq$$

BEWIJS. (i) is al aangetoond. Neem v, $\hat{p}$ en $\hat{q}$ die voldoen aan (4.3). Dan

$$v \leq \min_j \hat{p}Ue_j \leq \max_{p\in\Delta^m} \min_j pUe_j \tag{4.4}$$

Het maximum in het rechterlid bestaat omdat de funktie $p \mapsto \min_j pUe_j$ als minimum van n continue (lineaire) funkties ook continu is en omdat bovendien Δ^m compact is. Uit $pUq \geq \min_j pUe_j$ voor elke $q \in \Delta^n$ volgt dat

$$\min_j pUe_j = \min_{q\in\Delta^n} pUq \ . \tag{4.5}$$

Combineren we (4.4) en (4.5) dan volgt

$$v \leq \max_{p\in\Delta^m} \min_{q\in\Delta^n} pUq \ . \tag{4.6}$$

Evenzo volgt

$$v \geq \min_{q\in\Delta^n} \max_{p\in\Delta^m} pUq \ . \tag{4.7}$$

Nu toont men gemakkelijk aan dat

$$\max_{p\in\Delta^m} \min_{q\in\Delta^n} pUq \leq \min_{q\in\Delta^n} \max_{p\in\Delta^m} pUq \tag{4.8}$$

Uit (4.6), (4.7) en (4.8) volgen dan (ii) en (iii) uit de stelling. □

Bovenstaande stelling staat vanwege (ii) bekend als de *minimaxstelling voor matrixspelen*.

Voor de waarde van U geldt dus

$$v(U) = \max_p \min_q pUq = \min_q \max_p pUq$$

en de optimale strategieënruimten

$$O_I(U) := \{\hat{p} \in \Delta^m : \hat{p}Uq \geq v(U) \text{ voor alle } q \in \Delta^n\},$$

$$O_{II}(U) := \{\hat{q} \in \Delta^n : pU\hat{q} \leq v(U) \text{ voor alle } p \in \Delta^m\}$$

zijn niet-leeg voor elke U.

5. EEN BEWIJS VAN DE STELLING VAN FARKAS

We geven nu een elementair bewijs van stelling 2.

BEWIJS VAN STELLING 2. Stel dat voor het stelsel uit (i) een oplossing $\hat{x}$ bestaat en voor het stelsel uit (ii) een oplossing $\hat{y}$. Dan $0 \leq \hat{x}.A\hat{y} = \hat{x}A.\hat{y} = b.\hat{y} < 0$ en dat is onmogelijk. Dus hoogstens één van de stelsels uit (i) en (ii) bezit een oplossing. Vanaf nu bekijken we paren (A,b) waarvoor het stelsel uit (i) geen oplossing bezit en gaan we aantonen dat het corresponderende stelsel uit (ii) wel een oplossing bezit. We onderscheiden twee gevallen.

(1) We noemen het paar (A,b) van *type* 1 als het stelsel vergelijkingen $xA = b$ geen oplossingen bezit.

(2) (A,b) heet van *type* 2 als het stelsel vergelijkingen $xA = b$ wel oplossingen bezit, maar geen niet-negatieve oplossingen.

Als (A,b) van type 1 is, dan is er volgens stelling 1 een y met $Ay = 0$ en $b.y = -1$ en bezit dus het stelsel uit (ii) een oplossing.

Als (A,b) van type 2 is en $m = 1$, dan is b een negatief veelvoud van de enige rij $r_1 \neq 0$ van A. Als we dan $y = r_1$ nemen, dan is y oplossing van het stelsel uit (ii). Voor willekeurige paren (A,b) van type 2 gaan we door induktie naar het aantal rijen m van A aantonen dat het stelsel uit (ii) een oplossing bezit. Voor $m = 1$ is dit al aangetoond.

Veronderstel nu dat het stelsel uit (ii) oplossingen bezit voor alle (A,b) van type 2 waarbij het aantal rijen van A gelijk is aan m-1. Zij (A,b) een paar van type 2 met m rijen $r_1, r_2, \ldots, r_m$ voor A. Dan is er een vektor $y_1 \in \mathbb{R}^n$ zo dat

$$r_i \cdot y_1 \geq 0 \text{ voor elke } i \in \{1,2,\ldots,m-1\} \tag{5.1}$$

$$b \cdot y_1 < 0 \,. \tag{5.2}$$

We kunnen dit als volgt inzien. Het stelsel $xB = b$, $x \geq 0$, waarbij B de $(m-1)\times n$-matrix is met rijen $r_1, r_2, \ldots, r_{m-1}$, bezit geen oplossing. Als (B,b) van type 1 is, dan gelden (5.1) en (5.2) zoals we boven gezien hebben en als (B,b) van type 2 is dan volgen (5.1) en (5.2) uit de induktieveronderstelling.

Als nu $r_m.y_1 \geq 0$, dan zijn we klaar want dan is y_1 oplossing van het

stelsel uit (ii).

Veronderstel dat $r_m.y_1 < 0$. We bekijken dan een nieuw stelsel

$$x\bar{A} = \bar{b} \ , \ x \geq 0 \tag{5.3}$$

waarbij $\bar{A}$ de $(m-1)\times n$-matrix is met i-de rij

$$\bar{r}_i = (r_i.y_1)r_m - (r_m.y_1)r_i \qquad (i = 1,2,\ldots,m-1)$$

en $$\bar{b} = (b.y_1)r_m - (r_m.y_1)b.$$

Veronderstel eens dat dit stelsel een oplossing $\bar{x}$ bezit. Dan

$$\sum_{i=1}^{m-1} \bar{x}_i(r_i.y_1)r_m - \bar{x}_i(r_m.y_1)r_i = \bar{b} = (b.y_1)r_m - (r_m.y_1)b$$

en dan

$$b = -(r_m.y_1)^{-1}\Big[\sum_{i=1}^{m-1} \bar{x}_i(r_i.y_1) - (b.y_1)\Big]r_m + \sum_{i=1}^{m-1} \bar{x}_i r_i$$

waaruit volgt met het oog op (5.1) en (5.2) dat het oorspronkelijke stelsel uit (i) een niet-negatieve oplossing bezit wat in strijd is met het feit dat (A,b) van type 2 is. Dus (5.3) bezit geen oplossingen.

Maar dan kunnen we concluderen dat er een $\bar{y} \in \mathbb{R}^n$ is met

$$\bar{r}_i.\bar{y} \geq 0 \qquad \text{voor } i = 1,\ldots,m-1 \tag{5.4}$$

$$\bar{b}.\bar{y} < 0 \ . \tag{5.5}$$

Neem $y = (r_m.\bar{y})y_1 - (r_m.y_1)\bar{y}$. Dan

$$r_i.y = (r_i.y_1)(r_m.\bar{y}) - (r_m.y_1)(r_i.\bar{y}) = \bar{r}_i.\bar{y} \geq 0 \text{ vanwege (5.4)}$$

$$b.y = ((b.y_1)r_m - (r_m.y_1)b).\bar{y} = \bar{b}.\bar{y} < 0 \qquad \text{vanwege (5.5)}$$

$$r_m.y = (r_m.\bar{y})(r_m.y_1) - (r_m.y_1)(r_m.\bar{y}) = 0.$$

Dus y voldoet aan het stelsel uit (ii). Hiermee is de induktiestap bewezen. □

LITERATUUR

[1] DANTZIG, G.B. (1963), *Linear programming and extensions*, Princeton University Press, Princeton.

[2] FARKAS, J.(1902), *Über die Theorie der einfachen Ungleichungen*, J. Reine Angew. Math. 124, p. 1-24.

[3] FELLER, W. (1950), *An introduction to probability theory and its applications*, Wiley, New York.

[4] FOURIER, J.-B. (1826), *Solution d'une question particulière du calcul des inegalités,* Nouveau bulletin des sciences par la societé philomathique de Paris, p. 99.

[5] GALE, D. (1960), *The theory of linear economic models,* McGraw-Hill, New York.

[6] MOTZKIN, T.S. (1936), *Beiträge zur Theorie der Linearen Ungleichungen,* Inaugurale Dissertation, Basel, Jerusalem.

[7] NEUMANN, J. von and O. MORGENSTERN (1944), *Theory of games and economic behavior,* Princeton University Press, Princeton.

[8] OWEN, G. (1982), *Game Theory,* Academic Press, New York.

[9] TIJS, S.H. (1975), *Graven in de speltheorie,* Vakantiecursus VC 29/75, Mathematisch Centrum, Amsterdam.

[10] WESSELS, J. en J. van NUNEN (1983), *Processen uit het dagelijks leven,* Euclides 58, p. 202-218.

S T A T I S T I E K: Het trekken van conclusies uit waarnemingen.

prof.dr. R. Doornbos

1. Inleiding.

Het opzetten en organiseren van proeven vormt een essentieel onderdeel van experimenteel wetenschappelijk onderzoek.
In een dergelijk onderzoek kunnen de volgende fasen worden onderscheiden:

1. Theorie, veronderstelling, hypothese
2. Voorspelling (deductie)
3. Waarneming
4. Analyse,

waarna eventueel door een proces van inductie de theorie wordt gewijzigd of verfijnd. Op deze wijze wordt een tweede cyclus van het iteratieproces gestart. Dit leerproces kan ook worden weergegeven als een terugkoppelingslus waarin de discrepantie tussen de waarnemingsuitkomsten en de consequenties van de hypothese H_1 leiden tot de gewijzigde hypothese H_2. Vervolgens leidt H_2 tot H_3 enzovoorts.

In figuur 1.1, ontleend aan Box, Hunter en Hunter (1978), is dit schematisch aangegeven.

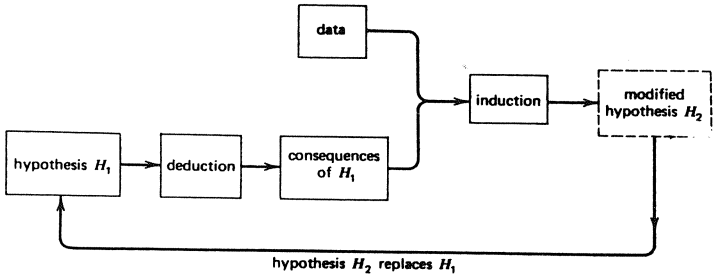


Fig. 1.1 Het leerproces als een terugkoppelingslus.

Uiteraard kan het onderzoek ook beginnen met waarnemingen in plaats van met een hypothese of theorie. Als eerste hebben we dan een verkennende of exploratieve fase die leidt tot een theorie. Vervolgens kunnen daaruit voorspellingen worden afgeleid die door nieuwe waarnemingen dienen te worden bevestigd.
Op welke punten speelt in de statistiek een rol?
Het is een kenmerk van waarnemingsuitkomsten in allerlei zeer verschillende gebieden van onderzoek dat ze variëren van proef tot proef, ook als de proefomstandigheden zo goed mogelijk constant worden gehouden. Deze variabiliteit, veroorzaakt door bekende zowel als onbekende verstorende invloeden, wordt wel de experimentele fout genoemd, hoewel echte meetfouten vaak slechts een te verwaarlozen deel er van uitmaken.
Vanwege deze variabiliteit zullen statistische methoden moeten worden toegepast om de waarnemingen te analyseren en daaruit conclusies te trekken. Als de waarnemingen dienen om de juistheid van een hypothese te controleren, zullen statistische toetsings- en schattingsmethoden worden gebruikt. Zijn we in een verkennende fase dan is er sprake van data-analyse (zie b.v. Tukey (1977)).
De verstorende effecten van de experimentele fout kunnen voor een deel worden bedwongen door een juiste opzet van de experimenten. Aan dit aspect, de theorie van de statistische proefopzetten zullen we in het volgende nader ingaan.

2. Voorbeelden van proefopzetten

2.1

Met een balans moeten twee voorwerpen (b.v. gouden sieraden) worden gewogen. Een gestileerd model van zo'n weging is het volgende:

$$X_i = \mu_i + E_i,$$

waarin μ_i het ware gewicht is, X_i de uitkomst van de meting en E_i de meetfout. Stochastische grootheden geven we aan met hoofdletters, gerealiseerde waarden met kleine letters.
Van E_i nemen we aan dat de verwachting nul is en de variantie σ^2.
Gevraagd wordt nu met twee wegingen de twee gewichten zo nauwkeurig mogelijk te bepalen.

Vanzelfsprekend kunnen we elk object afzonderlijk wegen en de gewichten, μ_1 en μ_2 schatten: $M_i = X_i$ (i=1,2).
Van beide gewichten hebben we dan een zuivere schatter met variantie σ^2.
Er is echter een betere methode. We kunnen de som en het verschil der gewichten van beide voorwerpen bepalen door ze in de eerste weging samen op één schaal te leggen en bij de tweede weging elk in een schaal.
We krijgen dan twee waarnemingen:

$$Y_1 = \mu_1 + \mu_2 + E_1$$

$$Y_2 = \mu_1 - \mu_2 + E_2$$

De schatters voor μ_1 en μ_2 zijn resp.

$$M_1 = \tfrac{1}{2}(Y_1 + Y_2)$$

en

$$M_2 = \tfrac{1}{2}(Y_1 - Y_2).$$

Beide schatters hebben als variantie:

$$\text{var. } \tfrac{1}{2}(E_1 + E_2) = \text{var. } \tfrac{1}{2}(E_1 - E_2) = \tfrac{1}{4}.2\sigma^2 = \tfrac{1}{2}\sigma^2$$

De standaardafwijking is dus met een factor $\frac{1}{2}\sqrt{2} = 0,7$ gereduceerd vergeleken met de "naieve" opzet.

2.2

Omstreeks 1930 werd in Engeland een onderzoek uitgevoerd dat later bekend is geworden als het "Lanarkshire milk experiment".
In dit onderzoek kregen ongeveer 10.000 schoolkinderen dagelijks een hoeveelheid van driekwart pint (0,35 l) melk gedurende 4 maanden. Hun toename in lengte en gewicht werd vergeleken met dat van kinderen die geen melk kregen. Ongeveer de helft van de kinderen kreeg ongekookte melk, de andere helft melk van dezelfde herkomst die was gepasteuriseerd.
De volgende methode werd toegepast om, voor iedere school, te bepalen welke kinderen de melk zouden krijgen en welke niet.
Elke school kreeg òf ongekookte òf gepasteuriseerde melk. Er werd binnen de school een indeling in twee groepen gemaakt volgens het lot of met een alfabetisch systeem. Als dit leidde tot een groep met een al te grote fractie van goed-gevoede of ondervoede kinderen vond er een uitwisseling plaats om een betere verdeling te bewerkstelligen.

Met andere woorden een aselecte toewijzing werd "verbeterd" door een subjectief oordeel.

Dit resulteerde in waarnemingen aan het eind van de proef die voor de controlegroep aanzienlijk hoger waren dan voor de behandelde groep, zowel wat het gewicht als wat de lengte betrof. Waarschijnlijk waren onderwijzers onbewust beïnvloed door de grotere behoefte aan extra voeding bij de slecht gevoede kinderen, zodat er te veel van deze kinderen bij de behandelde en te weinig bij de controlegroep terecht kwamen.

"Student" (1931), pseudoniem voor de statisticus W.S. Gosset, wees er op dat de opzet een aantal ernstige tekortkomingen vertoonde.

-Een eventueel verschil tussen het effect van ongekookte en dat van gepasteuriseerde melk zou veel nauwkeuriger kunnen worden vastgesteld als dat binnen elke school gemeten had kunnen worden.

-Zoals reeds aangegeven kwamen er relatief te veel slecht gevoede kinderen in de experimentele groep terecht. Deze kinderen en ook die van scholen uit de armere districten zouden bij elke vorm van bijvoeding een extra gewichtstoename te zien geven, onafhankelijk van de soort melk.

-Dit verschil tussen arme en rijke districten werd nog vergroot door het feit dat de eerste meting in februari plaats vond en de tweede in juni, terwijl de kinderen met kleren aan werden gewogen. Men kan namelijk aannemen dat de armere kinderen in de winter lichtere kleren droegen dan de kinderen uit meer welgestelde gezinnen, terwijl er bij de zomerkleding geen verschil in gewicht bestond.

De statistische analyse van de waarnemingen vertoonde ook verschillende fouten, zoals aangetoond door Fisher en Bartlett (1931), maar dat is hier minder interessant.

2.3

In het British Medical Journal van 1950, part 2, p. 425 wordt een onderzoek beschreven naar het effect van een anti-histamine preparaat ter bestrijding van verkoudheid. De helft van de proefpersonen onderging deze behandeling, de andere helft kreeg een dummy preparaat, een zogenaamde placebo. Dit experiment werd als een zogenaamde "double-blind test" uitgevoerd. Dat wil zeggen dat de proefpersonen niet wisten of ze anti-histaminetabletten of een placebo kregen toegediend en dat bovendien de artsen die het effect van het middel constateerden evenmin van de soort tabletten op de hoogte waren.

Dit werd bereikt door aan de deelnemende klinieken genummerde doosjes met tabletten en op de zelfde wijze genummerde formulieren te verstrekken. Doosje nr. 1 bevatte b.v anti-histamine tabletten, doosje nr.2 controle tabletten enz. Deze toewijzing was van te voren door loting bepaald.
Een typisch resultaat van het onderzoek was bijvoorbeeld dat, van patiënten die één dag verkouden waren bij het begin van het onderzoek, 13,4% werden geregistreerd als "genezen of verbeterd" op de tweede dag na toediening van het histamine preparaat. Met de placebo waren de resultaten respectievelijk 13,9 en 64,7%.
Men kan een dergelijke proef uitbreiden door naast de controlegroep meerdere groepen te nemen die verschillende doses van het te onderzoeken geneesmiddel krijgen. Eventueel kan in zo'n geval de controlegroep natuurlijk ook worden onderverdeeld om het effect van verschillende doseringen van de placebo eveneens te kunnen nagaan.
Bij het behandelen van patiënten in een controle groep met een placebo rijst de ethische vraag naar de toelaatbaarheid. Een medicus die gunstige ervaringen denkt te hebben met een middel tegen een ernstige kwaal zal er moeilijk toe kunnen besluiten om mee te werken aan een experiment waarin slechts de helft van zijn patienten het middel krijgt toegediend.
Meer over deze problematiek kan men vinden in Hill (1971) en voor een speciaal geval in Rümke (1962).

2.4

Als men wil onderzoeken of een bepaalde ziekte samenhangt met zekere persoonlijke kenmerken of gewoontes is een experimenteel onderzoek vaak uitgesloten. We moeten hier nadrukkelijk verschil maken tussen het constateren van een samenhang en het besluiten tot een oorzakelijk verband. Allerlei zogenaamde "nonsens-correlaties" illustreren dit. Dit is een verkeerde benaming: er is een echte correlatie tussen b.v. het aantal geregistreerde geestelijk gehandicapten en het aantal geregistreerde bezitters van een radio ontvangtoestel in het Verenigd Koninkrijk in de jaren 1924 - 1939 (Kendall en Yule (1950)).
Alleen de, niet uitgesproken, suggestie van causaliteit is absurd. Zoals vaker is er een derde factor die met beide gecorreleerd is, in dit geval de tijd. De eerste variabele neemt toe in de loop van de tijd, omdat diagnose- en registratiemethoden worden verfijnd en uitgebreid en omdat de bevolking toeneemt, de tweede omdat de radio-technologie in het beschouwde tijdvak een snelle ontwikkeling heeft doorgemaakt.

Als men bijvoorbeeld het verband tussen roken en longkanker wil nagaan zijn er in principe twee methoden denkbaar, te weten een patiënt-controle onderzoek (case-control study) en een cohort-onderzoek.
In het eerste geval worden twee onafhankelijke steekproeven genomen van n_1 individuen met de ziekte (cases) en van n_2 individuen zonder de ziekte (controls). Binnen beide groepen worden de aantallen individuen geregistreerd die aan de risicofactor zijn blootgesteld.
Bij een cohort-onderzoek neemt men twee onafhankelijke steekproeven van n_1^* individuen die onderhevig zijn aan de risico-factor en van n_2^* individuen die dat niet zijn. Binnen deze twee groepen worden de aantallen personen geregistreerd die de ziekte in de loop van de tijd (de zgn. follow-up periode) ontwikkelen. Het verzamelen van de gegevens kan in de loop van de tijd geschieden (prospectief cohort-onderzoek) maar het kan ook op historische gegevens zijn bebaseerd (retrospectief cohort-onderzoek).
Het case-control onderzoek is altijd retrospectief.
Om beide vormen van onderzoek kunnen de resultaten in de meest eenvoudige vorm in het volgende, zogenaamd twee-bij twee of 2x2 schema, worden samengevat:

	Individuen met de ziekte	Individuen zonder de ziekte	
Individuen onderhevig aan de risicofactor	a	b	$a + b = n_1^*$
Individuen niet onderhevig aan de risicofactor	c	d	$c + d = n_2^*$
	$a + c = n_1$	$b + d = n_2$	n

Bij het patiënt-controle onderzoek zijn n_1 en n_2 vast en n_1^* en n_2^* stochastisch, bij het cohort-onderzoek is het andersom.

Uit een onderzoek van Doll en Hill (1950) zijn de gegevens samengevat van een patiënt-controle onderzoek naar het verband tussen sigarettenroken en longcarcinoom. Een groep van 709 patiënten met longkanker uit 20 ziekenhuizen werd vergeleken met een controlegroep van eveneens 709 patiënten zonder longkanker. Tegenover iedere patiënt met longkanker werd een patiënt gekozen van hetzelfde ziekenhuis, hetzelfde geslacht en dezelfde leeftijdsgroep (ingedeeld in klassen van 5 jaar).

	MANNEN		VROUWEN	
Aantal sigaretten per dag	Longkanker	Controle	Longkanker	Controle
0	2	27	19	32
1 - 4	33	55	7	12
5 - 14	250	293	19	10
15 - 24	196	190	9	6
25 - 49	136	71	6	0
> 50	32	13	0	0
Totaal	649	649	60	60

Het enige belangrijke verschil tussen de patiënten met longkanker en de controlegroep was dat in de rookgewoonten. Dit is een goed voorbeeld voor de zorgvuldigheid die men moet betrachten om de groepen vergelijkbaar te maken. In een dergelijke studie kunnen nog altijd bedenkingen worden aangevoerd tegen de keuze van de controlegroep en bovendien wordt men niets gewaar over het verband tussen roken en eventuele andere ziektes.
Doll en Hill (1964) voerden ook een cohort-onderzoek uit door alle 59.600 artsen in het Verenigd Koninkrijk in 1951 een te enqueteren en de 68,2% die reageerden gedurende 10 jaren te volgen en doodsoorzaken te registreren. Enkele resultaten staan in de volgende tabel.

Aantal sterfgevallen per 1.000 personen

doodsoorzaak	aantal sterfgevallen	niet-rokers	sigarettenrokers aantal sig./dag 1-14	15-24	25-	ex-rokers	gemengde rokers	pijp- of sigaarrokers
longkanker	207	0,07	0,57	1,29	2,23	0,24	0,52	0,43
chronische bronchitis	111	0,05	0,34	0,64	1,06	0,38	0,33	0,15
ziekten v.d. kransslagader	1.287	3,31	4,35	4,28	4,97	3,73	3,87	3,18

3. Verloting en blokvorming

3.1
Verloting (randomization)

In een aantal van de hiervoor behandelde voorbeelden was sprake van het toewijzen van behandelingen aan proefpersonen of andere experimentele eenheden door loting. Wij gaan niet in op de techniek van het verloten onder verschillende omstandigheden, maar we beperken ons tot het waarom. De eerste reden om tot verloting over te gaan is dat op deze wijze bekende of onbekende systematische invloeden, behalve die we willen onderzoeken, worden uitgeschakeld. Zo zou men in het voorbeeld besproken in 2.3 de verbonden proefpersonen om de ander met histamine (H) en placebo (P) kunnen behandelen:
H,P ; H,P;----. Als er zich dan een trend voordoet in de ernst van de aandoening veroorzaakt dit een systematisch verschil tussen H en P. Als men om dit te voorkomen overgaat op een ander vast patroon zoals H,P; P,H; H,P; --- dan zou dit weer met een ander systematisch patroon kunnen samenvallen. Bovendien is er de mogelijkheid dat de onderzoekers het patroon kennen of denken te herkennen en in hun waarnemingen daardoor worden beïnvloed.
In de tweede plaats is het alleen mogelijk om statistische analysemethoden toe te passen en de onvermijdelijke onzekerheid in de conclusies te kwantificeren als aan een aantal voorwaarden is voldaan. Zo zal vrijwel altijd geëist moeten worden dat de waarnemingen beschouwd kunnen worden als onafhankelijke aselecte steekproeven. Bij de klassieke methoden als repressie- en variantieanalyse en de t-toets van Student wordt verder nog normaliteit van de verdelingen verondersteld.
R.A. Fisher, de grondlegger van de statistische proefopzetten, heeft er als eerste op gewezen dat de toepasbaarheid van statistische toetsingsmethoden kan worden gegarandeerd door het principe van verloting toe te passen. Het is dan in beginsel mogelijk een exacte zogenaamde permutatietoets toe te passen. Dit is echter meestal te bewerkelijk. De gangbare methoden, zoals de t-toets voor het vergelijken van twee behandelingen, kunnen als een benadering van deze exacte methoden worden beschouwd.
Voor een zeer goede en meer uitvoerige beschouwing over het verlotingsprincipe (randomization) moge worden verwezen naar Cox (1958), hoofdstuk 5 en verder naar Fisher (1966).

3.2
Blokvorming

Om een effect van een behandeling of het verschil tussen twee of meer behandelingen te kunnen vaststellen, ook als dat effect klein is t.o.v. de natuurlijke variabiliteit of experimentele fout, zijn meerdere waarnemingen nodig. Door een aangepaste proefopzet kan aan dit bezwaar tegemoet worden gekomen. Wanneer b.v. twee soorten kippevoer voor slachtkuikens moeten worden vergeleken kan de proef als volgt worden uitgevoerd.
Men neemt n paren kuikens. Van ieder paar (met dezelfde ouders) krijgt het ene kuiken gedurende een bepaalde tijd voer A en het andere voer B.
Welk van de twee kuikens A en welk B krijgt wordt uiteraard door loting vastgesteld. Na deze tijd wordt het gewicht (of de gewichtstoename) gemeten.
Uit de n verschillen in gewicht die resulteren zijn de verschillen ten gevolge van erfelijke factoren door deze opzet geëlimineerd. Op dezelfde gronden merkte Student in zijn bespreking van het Lanarkshire milk experiment op dat men, met een klein aantal identieke tweelingen veel goedkoper en preciezer het verschil in effect van de twee soorten melk zou hebben kunnen vaststellen.
Deze methode kan, althans bij kuikens uiteraard gemakkelijk worden uitgebreid voor een situatie waarbij meer dan twee behandelingen worden vergeleken.
Een homogene sub-groep van twee of meer experimentele eenheden, waarbinnen een aantal behandelingen met elkaar wordt vergeleken heet een blok en het principe wordt blokvorming (blocking) genoemd.
De eerste typen proefopzetten die hierop berusten zijn door Fisher ontworpen voor toepassing bij proeven in de land- en tuinbouw.
Ze zijn echter in allerlei andere toepassingsgebieden te gebruiken.

4. Enkele typen proefopzetten met blokvorming.

In de vorige paragrafen werd het principe van blokvorming besproken. Het doel is om de doeltreffendheid van het experiment te verhogen door vergelijkingen te maken binnen een betrekkelijk homogene groep van experimentele eenheden.
In het eenvoudigste geval worden binnen elk blok alle behandelingen vergeleken: volledige blokken (complete blocks). Een practisch bezwaar is dat niet altijd homogene blokken van voldoende omvang kunnen worden gevonden. We moeten dan overgaan op onvolledige blokken, waarbinnen niet alle behandelingen kunnen worden vergeleken.

Van dergelijke proefopzetten zijn verschillende varianten bedacht, afhankelijk van de aard van het experiment. We zullen een aantal typen bespreken en daarbij aangeven hoe de effecten van de behandelingen kunnen worden geschat, uitgaande van een additief lineair model voor de waarnemingen.

4.1
Volledige blokkenproeven (randomized block designs).

Wanneer t behandelingen worden vergeleken in r blokken kunnen we het model weergeven als

$$Y_{ij} = \mu + \beta_i + \tau_j + E_{ij} \quad , \quad \begin{matrix} i=1,---,r \\ j=1,---,t \end{matrix}$$

Hierin is Y_{ij} de waarnemingsuitkomst van de j^e behandeling in het i^e blok, en μ is het algemeen gemiddelde, β_i is het bolkeffect van het i^e blok en τ_j het effect van de j^e behandeling. Zowel voor β_i als voor τ_j geldt dat het afwijkingen betreft ten opzichte van het algemeen gemiddelde. Om dit vast te leggen en het model eenduidig te maken leggen we de volgende beperkingen op: $\sum_{i=1}^{r} \beta_i = \sum_{j=1}^{t} \tau_j = 0.$

Van de meetfouten E_{ij} wordt aangenomen dat ze verwachting 0 hebben, ongecorreleerd zijn en dezelfde variantie hebben. Om hypothesen te kunnen toetsen als $H_o : \tau_j = 0$ $(j = 1, \ldots, t)$ moet een veronderstelling worden gemaakt over de kansverdeling van de E_{ij}. Gewoonlijk wordt normaliteit verondersteld, maar op toetsingsproblemen gaan we hier niet in.

De behandelingseffecten worden geschat door

$$t_j = \frac{1}{r} \sum_{i=1}^{r} y_{ij} - \frac{1}{rt} \sum_{i=1}^{r} \sum_{j=1}^{t} y_{ij} = \bar{y}_{.j} - \bar{y}_{..} .$$

Dit is een realisatie van T_j, een zuivere schatter voor τ_j. Immers

$$E\left(\frac{1}{r} \sum_{i=1}^{r} Y_{ij}\right) = \frac{1}{r}\left(r\mu + \sum_{i=1}^{r} \beta_i + r\tau_j\right) = \mu + \tau_j$$

en

$$E\left(\frac{1}{rt} \sum_{i=1}^{r} Y_{ij}\right) = \frac{1}{rt}\left(rt\mu + t \sum_{i=1}^{r} \beta_i + r \sum_{j=1}^{t} \tau_j\right) = \mu .$$

4.2

Latijnse vierkanten.

Het principe van blokvorming om de heterogeniteit van de experimentele eenheden te verminderen kan verder worden verfijnd door een principe toe te passen dat twee-dimensionale blokvorming genoemd kan worden.
Men kan zich voorstellen dat een proefveld een systematisch verloop in vruchtbaarheid vertoont in twee richtingen of dat een rol textiel systematische verschillen in treksterkte heeft in de lengte van de rol, maar ook in de dwars richting. Door blokken te vormen in beide richtingen kunnen verschillen tussen behandelingen worden gemeten vrij van beide soorten heterogeniteit. Het schema dat op deze wijze ontstaat heet een Latijns vierkant. Een Latijns vierkant is een rangschikking van p letters in een p x p vierkant, zodat elke letter precies één keer in elke rij en precies één keer in elke kolom voorkomt. Voor p=4 bijvoorbeeld:

	kolommen			
rijen	A	B	C	D
	B	C	D	A
	C	D	A	B
	D	A	B	C

Een dergelijk schema is bijvoorbeeld ook toepasbaar in een situatie met twee factoren. B.v. rijen = 4 proefvelden (blokken), A, B, C, D = 4 tarwerassen, kolommen = 4 soorten bemesting.
Het model is:

$$Y_{ijk} = \mu + \rho_i + \kappa_j + \tau_k + E_{ijk} \; ; \; i,j,k,=1,\ldots,\rho$$

met de bijvoorwaarden:

$$\sum_1^p \rho_i = \sum_1^p \kappa_j = \sum_1^p \tau_k = 0$$

voor resp. de rij-, kolom-, en behandelingseffecten.
Men kan dit schema alleen toepassen als men er zeker van kan zijn dat het additieve model juist is en er dus geen interacties aanwezig zijn, dit zijn mengtermen als b.v. $(K\tau)_{jk}$.
De parameters worden als volgt geschat:

$$r_i = \frac{1}{p} \sum_{j,k} y_{ijk} - \frac{1}{p^2} \sum_{i,j,k} y_{ijk} = \bar{y}_{i..} - \bar{y}_{...}$$

$$\kappa_i = \frac{1}{p} \sum_{i,k} y_{ijk} - \frac{1}{p^2} \sum_{i,j,k} y_{ijk} = \bar{y}_{.j.} - \bar{y}_{...}$$

$$t_k = \frac{1}{p} \sum_{i,j} y_{ijk} - \frac{1}{p^2} \sum_{i,j,k} y_{ijk} = \bar{y}_{..k} - \bar{y}_{...}$$

Het idee van de dubbele blokvorming en de rangschikking van de behandelingen in een Latijns vierkant kan nog een stap verder worden doorgevoerd. We kunnen nog een factor op 4 niveaus invoeren, aangegeven met Griekse letters. Voor de analyse is het nodig dat deze factor weer orthogonaal is ten opzichte van de andere. Dat wil zeggen dat elke Griekse letter één keer in elke kolom, één keer in elke rij en één keer samen met elke Latijnse letter voorkomt.

Voorbeeld:

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
$B\gamma$	$A\delta$	$D\alpha$	$C\beta$
$C\delta$	$D\gamma$	$A\beta$	$B\alpha$
$D\beta$	$C\alpha$	$B\delta$	$A\gamma$

Hier zijn twee orthogonale Latijnse vierkanten gesuperponeerd. We kunnen nog een stap verder gaan en een derde vierkant orthogonaal op deze twee toevoegen. Meer is ook niet mogelijk, er zijn maximaal (p-1) onderling orthogonale p x p vierkanten. Voor p = 3,4,5,7, 8,9,11 en 13 zijn er ook inderdaad (p-1) bekend.
Een vermoeden van Euler uit 1782 luidt dat er voor p = 6 en p = 10 geen Grieks - Latijns vierkant bestaat dus zelfs geen tweetal orthogonale Latijnse vierkanten. Voor p = 6 is dit later bewezen, maar voor p = 10 is in 1959 een tweetal gevonden. Voor p = 12 waren er reeds langer twee bekend, dit is tamelijk recent tot 5 uitgebreid. In het algemeen is bewezen dat er (p - 1) onderling orthogonale vierkanten zijn als p een priemgetal is of een macht van een priemgetal.

4.3
Evenwichtige onvolledige blokkenproeven (balanced incomplete block designs).

Het komt vaak voor dat slechts een beperkt aantal behandelingen binnen één blok kunnen worden vergeleken. Zo kan men bij een smaakproef geen groot aantal monsters aan één proefpersoon voorzetten om te proeven en te vergelijken. Bij bepaalde dierproeven wil men verschillende behandelingen toepassen op dieren afkomstig uit één worp of nest. Bij sommige diersoorten zoals koeien en schapen beperkt dit de blokgrootte tot 2 of 3.
Voor deze situatie waarin het aantal behandelingen groter is dan het aantal eenheden in een blok zijn er twee soorten proefopzetten:
de evenwichtige onvolledige blokkenproeven en de gedeeltelijk evenwichtige onvolledige blokkenproeven,

in het Engels resp. balanced incomplete block designs (b.i.b.d.) en partially balanced incomplete block designs (p.b.i.b.d.) genoemd.
Op de laatstgenoemde komen we kort terug aan het eind van deze paragraaf.
Een voorbeeld van een b.i.b.d. is het volgende van 6 behandelingen in 10 blokken van 3 eenheden:

A B C A B D A C E A D F A E F
B C F B D E B E F C D E C D F

Dit schema heeft de volgende eigenschappen. Niet alleen komt elke behandeling even vaak, nl. 5 keer voor, maar elk paar behandelingen komt precies 2 keer samen in een blok voor. Meestal gebruikt men de volgende symbolen: t = aantal behandelingen, k = aantal eenheden in een blok, r = aantal herhalingen, d.w.z. het aantal keren dat elke behandeling voorkomt, b = aantal blokken en λ = aantal keren dat een tweetal binnen een blok voorkomt.
Uiteraard geldt

$$bk = tr = \text{aantal eenheden}$$

en verder

$r(k-1) = \lambda(t-1) =$ het aantal keren dat een bepaalde behandeling binnen één blok voorkomt.

Dus

$$\lambda = \frac{r(k-1)}{t-1} .$$

In ons voorbeeld:

$$\lambda = \frac{5(3-1)}{6-1}$$

Verder geldt nog de ongelijkheid van Fisher:

$$t \leqq b.$$

Een tabel van mogelijke schema's vindt men o.a. in Cochran & Cox (1957).
In dit boek wordt ook beschreven hoe de behandelingseffecten kunnen worden geschat, waarbij voor de blokeffecten wordt gecorrigeerd.
Bij gedeeltelijk evenwichtige schema's wordt de eis dat elk paar λ keer binnen een blok voorkomt verzwakt tot de voorwaarde dat er m klassen van paren zijn, die $\lambda_1,\ldots,\lambda_m$ keer voorkomen, waarbij één van de λ_i's nul mag zijn.
Een voorbeeld is het volgende
(123),(164),(175),(683),(784),(793),(285),(294).
Hierin is $m=2, \lambda_1=1$ en $\lambda_2= 0$.
Men kan gemakkelijk controleren dat iedere behandeling met 6 van de andere 1x voorkomt en met 2 van de andere in het geheel niet.

Een uitvoerige behandeling met verdere eigenschappen en generalisaties is te vinden in Kempthorne (1960).

5. Enkele slotopmerkingen.

In het voorgaande overzicht don uiteraard slechts aan enkele typen proefopzetten aandacht worden geschonken, terwijl de analyse van de waarnemingsuitkomsten vrijwel geheel buiten beschouwing moest worden gelaten.
Voor een eerste verdere oriëntatie kan het boek van Cochran & Cox (1957) worden aanbevolen.
De keuze van de voorbeelden is niet representatief voor de mogelijke toepassingsgebieden. Meer gevarieerde voorbeelden vindt men in Cox (1958), vooral technologische toepassingen in Box, Hunter & Hunter (1978).

Literatuur

G.E.P. Box, W.G. Hunter & J.S. Hunter (1978), Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building, Wiley, New York.

W.G. Cochran & G.M. Cox (1957), Experimental Design, Wiley, New York.

D.R. Cox (1958), Planning of Experiments, Wiley, New York.

R. Doll & A. Bradford Hill (1950), Smoking and Carcinoma of the lung. Preliminary report, Br. Med. J., ii, 739 - 748.

R. Doll & A. Bradford Hill (1964), Mortality in relation to smoking: ten years' observations of British doctors, Br. Med. J., i, 1399 - 1410 en 1460 - 1467.

R.A. Fisher (1966), Design of Experiments, Eighth Edition, Oliver and Boyd, Edinburgh.

R.A. Fisher & S. Bartlett (1931), Pasteurised and raw milk, Nature, 127, 591 - 592.

A. Bradford Hill (1971), Principles of Medical Statistics, Ninth Edition, Oxford University Press.

O. Kempthorne (1960), The Design and Analysis of Experiments, Wiley, New York.

M.G. Kendall & G.U. Yule (1950), An Introduction to the Theory of Statistics, Charles Griffin, London,

Chr. L. Rümke (1962), Enkele opmerkingen over de opzet van onderzoekingen naar de werkzaamheid van cytostatica in de kliniek, Statistica Neerlandica, 16, 261 - 268.

"Student" (1931), The Lanarkshire milk experiment, Biometrika 23, 398 - 406.

J.W. Tukey (1977), Exploratory Data Analysis, Addison - Wesley, Reading.

CENSORING AND SURVIVAL

Richard Gill

Centrum voor Wiskunde en Informatica
Amsterdam

In deze voordracht over statistiek ben ik juist niet van plan een indruk te geven van de wiskundige achtergronden van het Hewet materiaal, doch te laten zien waartoe de daar behandelde statistiek en waarschijnlijkheidsrekening leidt: namelijk naar toepassingen van de statistiek in het maatschappelijk leven. Dit wil ik doen aan de hand van één specifiek voorbeeld. Bovendien wil ik typerende aspecten van moderne toepassingen van de statistiek belichten. Sommige van deze (overlappende) aspecten, die overigens niet alle even sterk naar voren zullen komen, zijn:

- gecompliceerde datastructuren, o.m. veroorzaakt door gedeeltelijke waarneming van datgene waarin men geinteresseerd is
- effect van de informatica op de modellen en methoden die men aandurft
- diepe resulaten uit de kansrekening nodig om vele methoden goed te begrijpen
- veelvuldig gebruik van asymptotische resultaten
- veelvuldig gebruik van niet- en semi-parametrische modellen
- aanraking van de statistiek met vele aspecten van de werkelijkheid.

Laat ik in plaats van nog meer algemene beschouwingen te geven onmiddelijk overstappen op het voorbeeld. Ik wil een voorbeeld bespreken uit een momenteel belangrijk toepassingsgebied van de statistiek, namelijk medisch onderzoek, in het bijzonder kanker-onderzoek, waarin men op grote schaal de effectiviteit van nieuwe behandelingen probeert aan te tonen door middel van een "(randomized) clinical trial".

Patienten komen een ziekenhuis binnen met een bepaalde ziekte. Zij krijgen een korte intensieve behandeling. Men wil feitelijk twee variaties op deze behandeling vergelijken; de standaard (oude) behandeling en een nieuwe. De keuze van behandeling voor een patient wordt dan bij voorkeur door randomisatie bepaald, liefst "double blind" (Dit is echter veelal op praktische of ethische gronden onmogelijk).

Na de behandeling gaat de patient weer naar huis. Doel van de behandeling is de patient een zo lang mogelijke periode van normaal leven te geven, zodat doel van het experiment is deze lengte van tijd tussen behandeling en (eventueel) terugkeer van de ziekte, de zogenaamde *overlevingsduur* ("survival

time") in de twee behandelingsgroepen te vergelijken. Het tijdstip van terugkeer heet ook wel het *faaltijdstip* of tijdstip van falen (van de behandeling). In het meest typerende geval gaat het om enkele honderden patienten die over een periode van twee of drie jaar binnenkomen in een aantal samenwerkende ziekenhuizen. Overlevingsduren zijn van patient tot patient zeer variabel. Bovendien gaat het erom relatief kleine verschillen in overlevingsduur aan te tonen. Meer precies gezegd, het gaat erom twee kansverdelingen (van overlevingsduren) aan de hand van waarnemingen daaruit te vergelijken. Zo'n vergelijking kan men aan de hand van vele criteria maken: bijvoorbeeld de vijf jaar overlevingskans; of de mediaan overlevingsduur; of de gemiddelde overlevingsduur. In de praktijk zal het belangrijk zijn de gehele *overlevingscurve* ("survival curve") in elke groep te schatten (vergezeld van een aanwijzing van de nauwkeurigheid van deze schatting, d.m.v. zogenaamde *betrouwbaarheidsbanden*, of iets dergelijks). De overlevingscurve is één minus de cumulatieve verdelingsfunctie van overlevingsduren, oftewel de kans t eenheden van tijd te overleven uitgezet tegen t zelf:

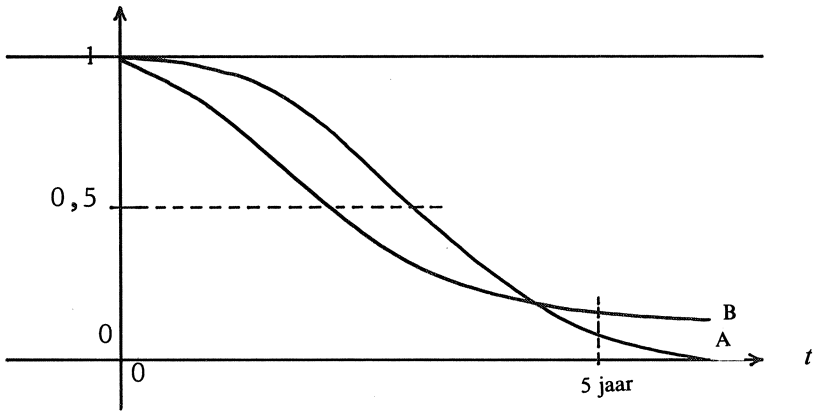


fig. 1
$S_A(t)$, $S_B(t)$ = kans op overleving $\geqslant t$
bij behandeling A resp. B.

Tegelijkertijd met de weergave van schattingen van zulke curves (tesamen met een aanwijzing van hun nauwkeurigheid) zal het toch belangrijk zijn door middel van een enkel numeriek criterium een conclusie van de vorm "A is beter dan B" of "B is beter dan A" of "er is geen aanwijsbaar verschil tussen A en B" te trekken. Beide statistische problemen: het *schatten* van de overlevingscurves, en het *toetsen* of ze gelijk zijn, zullen in het vervolg aan de orde komen; en tegelijkertijd zullen we iets van de wiskunde die erachter ligt bespreken. Allereerst zullen we beide problemen op zeer informele (onwiskundige) wijze bespreken.

Uit het bovenstaande zal duidelijk zijn dat één groot probleem de

precisering van de onderzoeksdoelstelling is. De oplossing die wij zullen geven houdt verband met het feit dat het belangrijkst voor de onderzoeker is, niet om aan te tonen dat er exact in één van die opzichten (mediaan, gemiddelde, 5 jaar overlevingskans, ...) een verschil is, doch louter het feit dát er een verschil is ten gunste van A of B. Een tweede even belangrijk probleem dat verrijst is onafhankelijk van onderzoeksdoelstelling, doch komt voort uit de aard van de gegevens die verzameld worden: men krijgt te maken met *gecensureerde* gegevens. Als men binnen een korte periode conclusies wil trekken, zal van een groot aantal patienten de overlevingsduur (tijd tot falen) nog niet geëindigd zijn: alleen bekend is dan dat de lengte van overlevingsduur groter is dan de lengte van tijd sinds binnenkomst. Van een aantal andere patienten zal eveneens de overlevingsduur op een éerder tijdstip al gecensureerd zijn, door verhuizing, door overlijden aan ongerelateerde oorzaken, in het algemeen door verlies aan "follow-up".

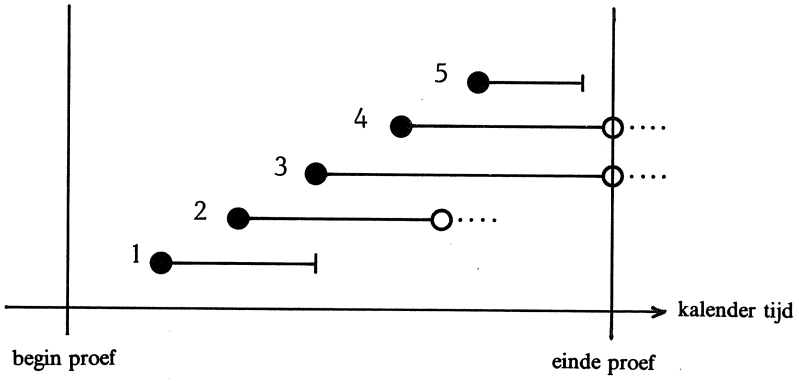


fig. 2
De overlevingsduur van patient 1 is volledig waargenomen.
Die van patient 2 is gecensureerd door verlies aan
een concurrend risico, die van patient 3 door het verstrijken
van de observatie periode.

Een eerste oplossing van dit probleem zou zijn het simpelweg te negeren: beschouw de gecensureerde overlevingsduren als gewone overlevingsduren; of laat ze geheel buiten beschouwing. Hier zijn twee grote bezwaren tegen. Ten eerste, men gooit kwistig weg een gedeelte van de met grote inspannig vergaarde gegevens die toch wel informatie verschaft over de vraag of A of B beter is. Ten tweede, men kan hierdoor volstrekt onjuiste conclusies trekken. Bij een niet-gerandomiseerde proef waarbij binnenkomst tijden van de twee groepen patienten misschien niet vergelijkbaar zijn, of in het algemeen als aan één behandeling inherent is dat verlies makkelijker voorkomt dan aan de andere, krijgt men snel ongelijke censurering in de twee groepen waardoor de vergelijking tussen groepen verkeerd kan uitvallen (als geen rekening hiermee

wordt gehouden).

Een eerste stap tot de oplossing van dit probleem is alle overlevingsduren "terug naar af" te schuiven, en dan de volgende observaties te maken:

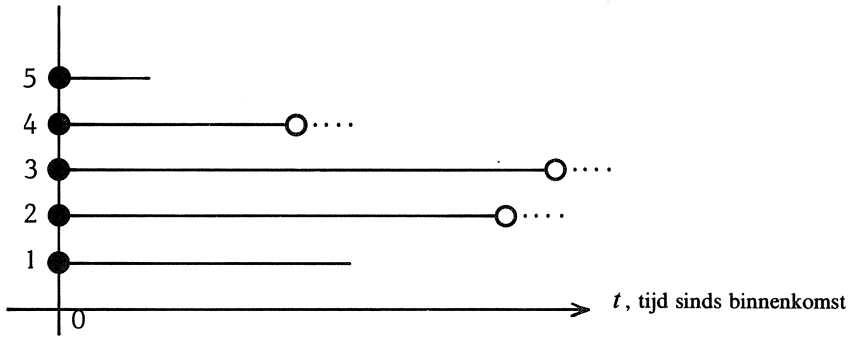


fig. 3

Als wij op (meer precies: net voor) een vast tijdstip t in de nieuwe tijdschaal (tijd sinds binnenkomst) kijken, zien wij een aantal patienten nog vóor de tijdstip van censurering of falen: de zogenaamde *risico verzameling* ("risk set") $\mathcal{R}(t)$. Dit zijn patienten die onder het risico staan, dat óp tijdstip t hun behandeling faalt (en bovendien dat dit waargenomen zou worden door de onderzoeker). De overlevingsduur voorwaardelijk van patienten in de risico verzameling is zeker groter of gelijk aan t. Als de censureringstijd van een patient werkelijk niet gerelateerd is aan de faaltijd zal de kansverdeling van de overlevingsduur voorwaardelijk op dat faaltijd T en censureringstijd C op of na tijdstip t plaatsvinden dezelfde zijn als de kansverdeling van T voorwaardelijk op, dat het alleen op of na tijdstip t valt: de voorwaardelijke verdeling van T gegeven $T \geqslant t$, $C \geqslant t$ is gelijk aan de voorwaardelijke van T gegeven $T \geqslant t$. In het bijzonder zijn de twee voorwaardelijke kansen dat $T = t$ (of dat $T \notin [t, t+dt]$) aan elkaar gelijk.

Dit is geen trivialiteit en hoeft ook in het algemeen niet waar te zijn: het is eerder een (niet controleerbare) veronderstelling over de aard van de censurering. Hoewel hij niet equivalent is met het gewone begrip van stochastische onafhankelijkheid tussen twee kansvariabelen, noemen we hem vaak wel "onafhankelijke censurering". In het vervolg zullen we er ook vanuit gaan dat hieraan voldaan is. Verder zullen we voorlopig de veronderstelling maken dat de overlevingsduren *discreet verdeeld* zijn; in het bijzonder dat ze waarden aannemen in de verzameling $t = 1,2,...$ tijdseenheden (bijvoorbeeld weken). Om te beginnen zullen we ons concentreren op de vraag, hoe de verdelingsfunctie van overlevingsduren te schatten is voor één van de behandelingen.

Laat T een stochastische variabele zijn voorstellend een overlevingsduur uit

zo'n discrete verdeling. De (cumulatieve) verdelingsfunctie F van T wordt gedefinieerd door $F(t) = P(T \leqslant t)$ voor elke t. Wij definieren ook de overlevingscurve S door $S(t) = P(T \geqslant t) = 1 - F(t-)$. Verder definieren we de (discrete) dichtheid f van F door $f(t) = P(T = t) = F(t) - F(t-)$ en tenslotte, tevens centraal in censureringsproblemen, de (discrete) *uitvalfunctie* of het *risico* ("hazard rate") λ door $\lambda(t) = P(T = t | T \geqslant t) = f(t) / S(t)$.

Deze laatste functie is zo belangrijk voor ons omdat de gecensureerde overlevingsduren op meest directe wijze informatie over λ geven, niet zo direct over f of S of F. De eerder gegeven beschouwingen over onafhankelijke censurering houden immers in, dat voor patienten in de risico verzameling $\mathcal{R}(t)$ op tijdstip t, gegeven wat zich afgespeeld heeft tot net voor tijdstip t, de kans óp tijdstip t te falen (wat dan ook waargenomen zou worden) gelijk is aan $\lambda(t)$.

Heel belangrijk is het feit dat uit de functie λ, de functie F te reconstrueren is met behulp van de formule

$$1 - F(t-) = S(t) = \prod_{s<t}(1 - \lambda(s)) \tag{1}$$

Hier zien we het gebruik van voorwaardelijke kansen in een lange keten: het rechter lid (voor gehele t) is gelijk aan

$$\prod_{s<t} P(T>s | T \geqslant s) = \prod_{s<t} P(T \geqslant s+1 | T \geqslant s) = P(T \geqslant t)$$

Op dit moment zullen we even een uitstapje maken naar het algemene (dat wil zeggen niet-discrete) geval. Het is in dit probleemgebied namelijk steeds zo dat het discrete geval bijzonder eenvoudig is en makkelijk te begrijpen. In het continue geval - dat meer op de praktijk situatie van toepassing is - zijn steeds volstrekt analoge resultaten geldig; alleen kan het bewijzen ervan vaak heel lastig zijn door grote wiskundig-technische moeilijkheden. Ook de interpretatie van de resultaten geschiedt het makkelijkst door verwijzing naar het discrete analoog. In het geval bijvoorbeeld waarin F absoluut continue is met dichtheid f, $F(t) = \int_0^t f(s)ds$, definieren we ook de "continue" uitvalfunctie $\lambda = f / S$ (als functie hoeft λ niet continu te zijn: het woord is bedoeld als onderscheiding van het discrete geval) maar nu geldt in plaats van (1) het volgende:

$$S(t) = \exp\left(-\int_0^t \lambda(s)ds\right) \tag{2}$$

waarvan de interpretatie niet zo voor de hand ligt. Zowel (1) als (2) zijn te verenigen in het algemene geval waarin F niet per se een discrete of een continue dichtheid heeft, door de *cumulatieve uitvalfunctie* Λ te definieren door $\Lambda(t) = \int_{[o,t]} S(s)^{-1} F(ds)$ (in het discrete geval geldt $\Lambda(t) = \Sigma_{s \leqslant t} \lambda(s)$, in het continue $\Lambda(t) = \int_0^t \lambda(s)ds$) en dan te schrijven

$$S(t) = \prod_{s \notin [0,t)} (1 - \Lambda(ds)), \tag{3}$$

een zogenaamde *product − integraal* (waarvan we de preciese definitie hier achterwege laten).

Terugkerend naar ons probleem, hoe kunnen we S schatten? Het antwoord zal nu bijna vanzelfsprekend zijn: als we $Y(t)$ definieren als het aantal patienten in de risico verzameling $\mathcal{R}(t)$, en $\Delta N(t)$ als het aantal daarvan dat op tijdstip t faalt, dan ligt het voor de hand de waarde van de discrete uitvalfunctie λ in het punt t te schatten door de statistische grootheid $\Delta N(t) / Y(t)$ en tenslotte analoog aan (1) S te schatten met

$$\hat{S}(t) = \prod_{s<t} (1 - \frac{\Delta N(s)}{Y(s)}). \tag{4}$$

Deze schatter heet de *product − limiet schatter* oftewel, naar de ontdekkers ervan, de *Kaplan − Meier schatter*. In verband met formule (3) kan men verwachten dat deze een goed gemotiveerde schatter van S is ook in het algemene geval. Hiervoor definieren we N door $N(t) = \Sigma_{s \leqslant t} \Delta N(s)$ (het aantal geobserveerde faal-tijden tot en met tijdstip t) en de *empirische cumulatieve uitvalfunctie* $\hat{\Lambda}$ door

$$\hat{\Lambda}(t) = \int_{s \notin [0,t]} (Y(s))^{-1} N(ds) = \sum_{s \leqslant t} \Delta N(s) / Y(s)$$

wat tenslotte leidt (cf. (3)) tot

$$\hat{S}(t) = \prod_{s<t} (1 - \hat{\Lambda}(ds)). \tag{4}$$

Wij zullen later iets verder ingaan op de mathematisch-statistische theorie van dit object en nu alleen iets opsommen over zijn eigenschappen, namelijk dat aan te tonen is dat als het aantal waarnemingen groot is en S continu, $\frac{\hat{S}(t)}{S(t)} - 1$ ongeveer verdeeld is als $W(\int_0^t \frac{N(ds)}{Y(s)^2})$ simultaan voor alle $t \geqslant 0$ waarbij W een standaard Wiener proces oftewel Brownse beweging is. Het woordje "ongeveer" zullen wij hier niet nader preciseren, alsook niet de voorwaarden van zo'n stelling. Hieruit is bijvoorbeeld af te leiden (gebruikmakend van het feit dat in de bedoelde situatie $\int_0^{\circ} dN / Y^2$ als een deterministische functie beschouwd mag worden) dat $\hat{S}(t)$ ongeveer normaal verdeeld is met verwachting $S(t)$ en variantie $S(t)^2 \int_0^t (Y(s))^{-2} N(ds)$, zodat in ons probleem van het vergelijken van twee overlevingscurves voor behandelingen A en B, als men in de $t = 5$ jaar overlevingskansen geinteresseerd zou zijn, men $S_A(t) - S_B(t)$ zou schatten met $\hat{S}_A(t) - \hat{S}_B(t)$ en de variantie van deze schatter met $\hat{\sigma}^2 = \hat{S}_A(t)^2 \int_0^t (Y_A(s))^{-2} N_A(ds) + \hat{S}_B(t)^2 \int_0^t (Y_B(s))^{-2} N_B(ds)$. Wegens de benaderende normale verdeling zou $\hat{S}_A(t) - \hat{S}_B(t) \pm 2\hat{\sigma}$ een ongeveer 95% betrouwbaarheidsinterval voor $S_A(t) - S_B(t)$ zijn. Als dit interval de waarde

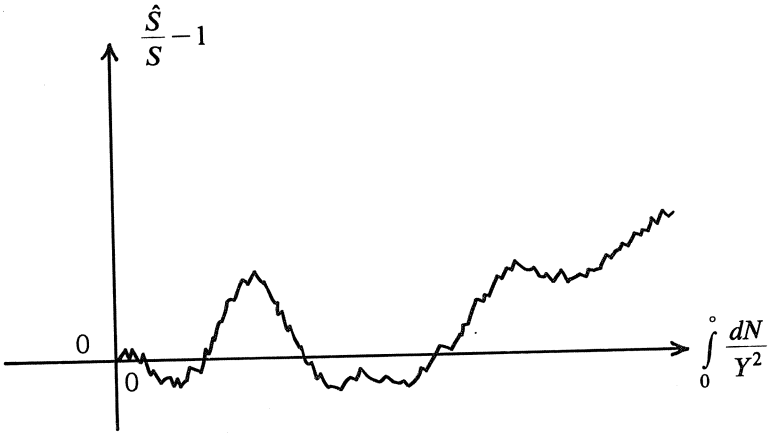


fig. 4.
Een realisatie van $\int_0^{\tau_1} S^{-1}$ uitgezet tegen $\int_0^1 dN\,/\,Y^2$

nul niet bevat zou men kunnen concluderen dat A en B niet equivalent zijn en bovendien kunnen aanwijzen welke beter is.

Uit het asymptotische (grote steekproeven) resultaat voor $\hat{S}$ kan men verder simultane betrouwbaarheids banden voor S construeren. De kansverdeling van $\sup_{0\leqslant x\leqslant a}|W(x)| \overset{\text{verdeling}}{=} a^{\frac{1}{2}}\sup_{0\leqslant x\leqslant 1}|W(x)|$ is bekend en van tabellen af te lezen. Als men voor een gekozen kans α (bijvoorbeeld 0.05) de waarde c_α (in dat geval 2.25) bepaalt zodanig dat $P(\sup_{0\leqslant x\leqslant 1}|W(x)|\leqslant c_\alpha)=1-\alpha$, dan weten we dat, met $\hat{\sigma}_\tau^2=\int_0^\tau (Y(s))^{-2}N(ds)$ voor een gekozen tijdstip τ, $P(\hat{\sigma}_\tau^{-1}\sup_{0\leqslant t\leqslant\tau}|\frac{\hat{S}(t)}{S(t)}-1|\leqslant c_\alpha) = 1-\alpha$. De hieromschreven gebeurtenis is equivalent aan de gebeurtenis

$$(1-\hat{\sigma}_\tau c_\alpha)^{-1}\hat{S}(t) \leqslant S(t) \leqslant (1-\hat{\sigma}_\tau c_\alpha)^{-1}\hat{S}(t), \quad 0\leqslant t\leqslant\tau.$$

De buitenste termen zijn statistische grootheden (dat wil zeggen uit de waarnemingen te berekenen). Dus weten we dat als we, voor gegeven waarden van τ en α, aan de hand van de data de buitenste krommen bepalen, de kans hoogstens (ongeveer) α is dat de ware maar niet bekende S niet tussen hen in ligt.

Andere banden zijn mogelijk (en zelfs wenselijk) die nauwer zijn bij kleinere t, maar degene die we hier gebruiken illustreren de onderliggende principes het makkelijkst. Het verschijnen van de Brownse beweging zullen wij later proberen uit te leggen. Laten wij eerst verder gaan met het toetsingsprobleem, het vergelijken van A en B. Het ligt nu ook voor de hand, als we op tijdstip t (discrete tijd weer) λ_A en λ_B willen vergelijken, dat te doen door

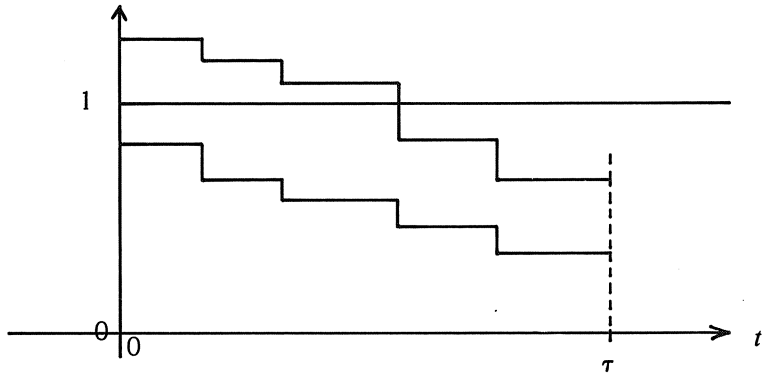


fig. 5
Betrouwbaarheidsbanden voor S

middel van

$$\hat{\lambda}_A(t)-\hat{\lambda}_B(t) = \delta N_A(t) / Y_A(t)-\delta N_B(t) / Y_A(t)$$

als schatter voor $\lambda_A(t) - \lambda_B(t)$. Het lijkt nu redelijk alle vergelijkingen bij elkaar op te tellen, gewogen met een (eventueel uit de data berekende) wegingsfactor K. We bekijken dus

$$U = \sum_t K(t)\left(\frac{\Delta N_A(t)}{Y_A(t)} - \frac{\Delta N_B(t)}{Y_B(t)}\right) = \int_0^\infty K(t)\left(\frac{N_A(dt)}{Y_A(t)} - \frac{N_B(dt)}{Y_B(t)}\right). \quad (5)$$

Als alle $K(t)$ niet-negatief zijn, dan verwachten we als in werkelijkheid

$$\hat{\lambda}_A(t)-\hat{\lambda}_B(t) = \Delta N_A(t) / Y_A(t)-\Delta N_B(t) / Y_B(t)$$

$\lambda_A(t)>\lambda_B(t)$ voor alle t een grote positieve waarde; als $\lambda_A(t)<\lambda_B(t)$ voor alle t een grote negatieve.

Maar hoe groot is groot? Hiervoor hebben we een aanwijzing nodig van de toevallige variatie van deze grootheid U (om de waarde nul als $\lambda_A \equiv \lambda_B$).

Laten we veronderstellen dat K zo geconstrueerd is dat de waarde van $K(t)$ vast ligt als de gegevens uit het tijdsinterval $[0,t)$ beschikbaar zijn. Dit is een soort "eerlijke spel" regel: we moeten $K(t)$ niet op de waarde van $\hat{\lambda}_A(t)-\hat{\lambda}_B(t)$ laten afhangen. Voorwaardelijk dan "op het tijdsinterval $[0,t)$" zijn $K(t)$, $Y_A(t)$ en $Y_B(t)$ vast en zijn $\Delta N_A(t)$ en $\Delta N_B(t)$ biomiaal verdeeld met parameters $(Y_A(t),\lambda_A(t))$ en $(Y_B(t),\lambda_B(t))$. Dus als inderdaad $\lambda_A = \lambda_B = \lambda$, heeft $K(t)(\hat{\lambda}_A(t)-\hat{\lambda}_B(t))$ voorwaardelijke verwachting nul en voorwaardelijke variantie

$$K(t)^2\lambda(t)(1-\lambda(t))\left(\frac{1}{Y_A(t)} + \frac{1}{Y_B(t)}\right).$$

Als we voor $\lambda(t)$ de "gepoolde" schatting $(\Delta N_A(t)+\Delta N_B(t)) \,/\, (Y_A(t)+Y_B(t))$ invullen, en verder vanuit gaan dat $1-\lambda(t)$ vrijwel 1 is - dit komt erop neer, dat we overstappen op het continue geval! - vinden we als schatting voor dit stukje voorwaardelijke variantie

$$K(t)^2(\Delta N_A(t)+\Delta N_B(t)) \,/\, Y_A(t)Y_B(t) \tag{6}$$

Bij redelijke grote steekproeven en bij geschikte keuze van K zal de steekproefvariantie in deze uitdrukking klein zijn in verhouding tot zijn grootte, zodat de schatting ook gezien kan worden als een schatting van de onvoorwaardelijke variantie.

Verder zijn de afzonderlijke termen in (5) ongecorreleerd. Dit kunnen we als volgt bewijzen. Noem de termen in (5) $\Delta U(t)$. Laat $\mathcal{F}_{t-}$ staan voor alles wat tot tijdstip t waargenomen is (maar niet op tijdstip t). Onze laatste redeneringen hielden dan feitelijk in, dat $E(\Delta U(t)\,|\,\mathcal{F}_{t-}) = 0$. Dus voor $s<t$ geldt $E(\Delta U(s)\Delta U(t)\,|\,\mathcal{F}_{t-}) = \Delta U(s)E(\Delta U(t)\,|\,\mathcal{F}_{t-}) = 0$, waarbij de eerste gelijkheid geldt omdat $\Delta U(s)$ een constante is gegeven $\mathcal{F}_{t-}$, $s<t$.

Dit allemaal suggereert dat de som van de termen in (6), zeg maar V, een schatting is voor de variantie van de som in (5). Aangezien de bedoeling was na te gaan of $\lambda_A = \lambda_B$ of $\lambda_A \neq \lambda_B$, is de gestandaardizeerde grootheid $U \,/\, \sqrt{V}$ een geschikte middel hiervoor: deze statistische grootheid neemt waardes aan die dicht bij 0 zijn als $\lambda_A = \lambda_B$, ver van nul als $\lambda_A \geqslant \lambda_B$ of $\lambda_B \geqslant \lambda_A$.

Dit moet nog steeds gepreciseerd worden. Hiervoor maken we gebruik van een *centraal limiet stelling.* Het zal bekend zijn dat de som van een groot aantal onafhankelijke, identiek verdeelde stochastische variablen met eindige variantie bij benadering normaal verdeeld zijn. Deze stelling blijft waar (onder de juiste voorwaarden) voor grootheden als $U = \Sigma\Delta U(t)$, een zogenaamde martingaal (Meer korrekt, het stochastisch process $U(t) = \Sigma_{s \leqslant t}\, \Delta U(s)$ is een martingaal). De verschillen $\Delta U(t)$ ("martingaal verschillen") hebben, gegeven $\mathcal{F}_{t-}$, (en als $\lambda_A = \lambda_B$) voorwaardelijke verwachting nul. Het blijkt mogelijk in deze algemeenheid een centraal limiet stelling te bewijzen onder voorwaarden die inhouden dat het aantal termen in de som groot wordt en alle afzonderlijke termen klein. Met behulp van deze stelling kan bewezen worden dat in onze situatie, voor grote steekproeven, geschikte keuze van K, en continue $S_A = S_B$, $U \,/\, \sqrt{V}$ ongeveer standaard normaal verdeeld is.

Dit levert ons de mogelijkheid $U \,/\, \sqrt{V}$ als *toetsingsgrootheid* te hanteren: als wij van te voren afspreken om A boven B te prefereren als $U \,/\, \sqrt{V} > 2$, B boven A als $U \,/\, \sqrt{V} < 2$, en anders geen voorkeur uitspreken, dan is de kans *als* er echt geen verschil tussen A en B is alleen (ongeveer) 5% om een onjuiste beslissing te nemen. (Ik hoop dat de lezer het getal 2 herkent als een benadering voor 1.960..., de waarde die met precies 5% kans door de absolute waarde van een standaard normaal verdeelde grootheid overschreden wordt).

Iets moet nog gezegd worden over de keuze van K. Hier beperken wij ons tot de opmerking dat een veelgemaakte keuze $K = Y_A\, Y_B \,/\, (Y_A + Y_B)$ is. De

toetsingsgrootheid heet dan de "log-rank test". Deze keuze blijkt zeker optimaliteits eigenschappen te hebben, d.w.z. geeft de beste kans om voor A of B te kiezen als ze niet equivalent zijn, als in werkelijkheid $\lambda_A = c\lambda_B$ voor een of ander constante c, zodat we $c = 1$ versus $c \neq 1$ willen toetsen. In de praktijk zal men nooit zeker kunnen weten dat $\lambda_A = c\lambda_B$, maar wel vaak vanuit kunnen gaan dat deze relatie bij benadering geldt. Bovendien, in deze situatie is de keuze tussen A en B eenduidig, zodat het juist dan heel belangrijk is de goede keuze te doen. Wij hebben dan een *half-parametrisch* model, waarin het fenomeen van interesse deels parametrisch (de onbekende constante c) en deels niet-paramtrisch (de onbekende functie λ_A) beschreven wordt.

De lezer zal zeker van het bovenstaande een onbehaaglijk gevoel over houden. Dit zou weggenomen kunnen worden, door diepgaand op de martingaal theorie, één van de mooiste stukken theorie in de moderne waarschijnlijkheidsrekening en één die telkens in allerlei situaties opduikt, in te gaan. Dat kan in dit korte bestek echter niet; in de literatuur lijst wordt wel een ingang in deze theorie en zijn toepassingen genoemd.

Tenslotte moeten we terugkomen op de product-limiet schatter, om te laten zien hoe de martingaal theorie ook daarbij aansluit. Een belangrijke stap is het afleiden van de vergelijking die geldig is voor continue S en voor t zodanig dat $Y(s)>0$ in $[0,t]$)

$$\frac{\hat{S}(t+)}{S(t)} - 1 = - \int_{s \in [0,t]} \frac{\hat{S}(s)}{S(s)} (\hat{\Lambda}(ds) - \Lambda(ds)).$$

Men kan deze formule varifiëren (in het geval van een *absoluut* continue S) door na te gaan dat de sprong van linker- en rechterlid in een sprong-punt t van N gelijk zijn, alsook de afgeleides in een continuiteitspunt van N, alsook de waardes in het punt $t = 0$.

Uit deze formule zien we dat ook $\hat{S}S^{-1} - 1$ te schrijven is als een integraal ten opzichte van een martingaal, waarbij de integrand alweer zo'n "eerlijke" wegingsfunctie vormt: het ligt vast op tijdstip s gegeven $\mathcal{F}_{s-}$. Een dergelijke heuristische afleiding als voor de toetsingsgrootheid U kan nu worden gevolgd om de bewering dat $\hat{S}S^{-1} - 1$ op een (in de tijd herschaalde) Brownse beweging lijkt aannemelijk te maken; de theorie van martingalen en in het bijzonder centraal limiet stellingen hiervoor maken dit exact.

Literatuur

Het standaardwerk over statistische analyse van overlevingsduren, waarbij echter de wiskundige achtergrond van deze methodes niet naar voren komt, is

Kalbfleisch, J.G. & Prentice, R.L. (1980). *The statistical analysis of failure time data,* Wiley, New York.

Een inleiding in de toepassing van martingaal theorie bij censuringsprobleem (toegespitst op een iets algemenere versie van het half-parametische model $\lambda_A = c\lambda_B$) is gegeven in

Gill, R.D. (1984), *Understanding Cox's regression model: a martingale approach,* Journal of the American Statistical Association, Vol. 79 no. 385 (June 1984).

Toepassingen zijn te vinden in

Miller, R.G. Jr., Efron, B. Brown, B.W. & Moses, L.E. (1980), *Biostatistics casebook,* Wiley, New York

en in

Kardaun, 0. (1983), *Statistical survival analysis of male larynx-cancer patients -a case study,* Statistica Neerlandica 37, 103-126.

JACKKNIFE EN BOOTSTRAP METHODEN

R. Helmers

Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ Amsterdam

1. Inleiding

Bootstrapmethoden zijn sinds de publikatie van het artikel 'Bootstrap methods: another look at the jackknife' van B. Efron in 1979 sterk in de belangstelling komen te staan. De bootstrap is, evenals de verwante meer klassieke jackknife methode, een bij veel statistische problemen toepasbare methode om de variabiliteit van statistische grootheden te bepalen. Bootstrap en jackknife technieken kunnen bijvoorbeeld gebruikt worden om de onzuiverheid en de variantie of de standaardafwijking van een schatter te schatten en om betrouwbaarheidsintervallen voor onbekende parameters te construeren.

Hoewel in principe eenvoudig, kunnen bootstrap schatters, en in minder mate ook jackknife schatters, in de praktijk meestal niet zonder hulp van een computer worden uitgerekend. Intensief computergebruik maakt het mogelijk bootstrap schattingen voor de variabiliteit van schatters te berekenen, ook in gecompliceerde statistische modellen met weinig veronderstellingen, situaties waar een traditionele statistische analyse vaak (nog) niet mogelijk is.

Diaconis en Efron (1983) en Efron en Gong (1983) geven zeer leesbare elementaire inleidingen in de bootstrap methode. Een uitvoerige behandeling van de bootstrap en de jackknife is te vinden in Efron (1982)

Na een korte inleiding in de schattingstheorie worden de jackknife en de bootstrap besproken. Ook de relatie tussen beide methoden komt aan de orde.

2. Schattingstheorie

We beschouwen een klassieke situatie in de statistiek: Gegeven is een aselecte steekproef $X_1, \ldots, X_n$ van omvang n uit een populatie met een (onbekende) verdelingsfunctie F. Laat verder

$$\theta = \theta(F) \tag{2.1}$$

een (reeëlwaardige) parameter zijn die geschat moet worden. De parameter θ wordt opgevat als een (reeëlwaardige) functie van de (onbekende) verdelingsfunctie F. Voorbeelden zijn het populatiegemiddelde of verwachte waarde $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$ en de populatievariantie $\theta = \theta(F) = \int_{-\infty}^{\infty}$

$(x-\mu)^2 dF(x)$, met $\mu = \mu(F) = \int_{-\infty}^{\infty} x dF(x)$.

Een *schatter* T_n van θ is een (reeëlwaardige) functie van de waarnemingen $X_1, \ldots, X_n$:

$$T_n = T_n(X_1, \ldots, X_n). \tag{2.2}$$

In dit model stellen de X_i's $(1 \leqslant i \leqslant n)$ onafhankelijke en identiek verdeelde (reeël- of vectorwaardige) stochastische grootheden voor met simultane verdelingsfunctie

$$P(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} F(x_i) \tag{2.3}$$

Indien

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \tag{2.4}$$

de feitelijk waargenomen waarden in de steekproef voorstellen, dan kan een *schatting*

$$t_n = T_n(x_1, \ldots, x_n) \tag{2.5}$$

voor θ berekend worden.

Een schatter T_n heet *zuiver* (unbiased) voor het schatten van θ indien zijn verwachte waarde (verwachting) ET_n gelijk is aan θ:

$$ET_n = E_F T_n(X_1, \ldots, X_n) = \theta(F) = \theta \tag{2.6}$$

Hier en elders schrijven we EX voor de verwachte waarde van een stochastische grootheid X met verdelingsfunctie F: $F(x) = P(X \leqslant x)$, $EX = \int_{-\infty}^{\infty} x dF(x)$. Soms schrijven we E_F en P_F i.p.v. E en P; P stelt de kansverdeling corresponderend met F voor.

Voorbeeld 2.1 $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$, het populatiegemiddelde; $T_n = n^{-1} \sum_{i=1}^{n} X_i = \bar{X}_n$, het steekproefgemiddelde. $E\bar{X}_n = EX_1 = \theta$; $\bar{X}_n$ is een zuivere schatter van θ.

Indien een schatter T_n niet zuiver is, dan noemt men

$$b_n(F) = E_F T_n - \theta(F) \tag{2.7}$$

of, kortweg $b_n = ET_n - \theta$, de *onzuiverheid* (bias) van T_n. De grootheid $b_n = b_n(F)$ geeft de systematische fout aan die gemaakt wordt als we T_n als schatter voor θ gebruiken.

De nauwkeurigheid van een schatter T_n wordt dikwijls gemeten met de *variantie* $\sigma_n^2 = \sigma_n^2(F)$ van T_n:

$$\sigma_n^2(F) = \sigma_F^2(T_n) = E_F(T_n - E_F T_n)^2 \tag{2.8}$$

of kortweg $\sigma_n^2 = \sigma^2(T_n) = E(T_n - ET_n)^2$. De grootheid $\sigma_n = \sqrt{\sigma^2(T_n)}$ noemt men de *standaardafwijking* van T_n; $\sigma_n = \sigma(T_n)$ geeft de grootte van de toevallige fout aan die gemaakt wordt als we T_n als schatter voor θ gebruiken.

De *gemiddelde kwadratische fout* (mean square error) van een schatter T_n van θ wordt gegeven door

$$MSE_n(F) = MSE_F(T_n) = E_F(T_n - \theta(F))^2 \tag{2.9}$$

of kortweg $MSE(T_n) = E(T_n - \theta)^2$. Er geldt

$$MSE(T_n) = \sigma^2(T_n) + (ET_n - \theta)^2 \tag{2.10}$$

Als criterium voor de kwaliteit van een schatter wordt dikwijls de MSE genomen. Hoe kleiner de MSE des te beter de schatter.

Voorbeeld 2.2. $\theta = \theta(F) = \int_{-\infty}^{\infty} (x - \mu(F))^2 \, dF(x)$ met $\mu(F) = \int_{-\infty}^{\infty} x dF(x)$, de populatievariantie, $T_n = S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, met $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, de steekproefvariantie; S_n^2 is een zuivere schatter van θ. Dit is eenvoudig in te zien: $ES_n^2 = (n-1)^{-1} E \left(\sum_{i=1}^n X_i^2 - n^{-1} (\sum_{i=1}^n X_i)^2\right) = (n-1)^{-1} (nEX_1^2 - n^{-1} \sum_{i=1}^n \sum_{j=1}^n EX_i X_j) = EX_1^2 - (EX_1)^2 = E(X_1 - EX_1)^2 = \theta$.

Een analoge berekening levert

$$MSE(S_n^2) = \sigma^2(S_n^2) = E(S_n^2 - \theta)^2 = \frac{\mu_4 - \theta^2}{n} + \frac{2\theta^2}{n(n-1)} \tag{2.11}$$

met $\mu_4 = E(X_1 - EX_1)^4 < \infty$

Men kan nagaan dat de schatter $S_n'^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2$ de zgn. *meest aannemelijke schatter* van θ is, in het klassieke geval dat de waarnemingen normaal verdeeld zijn met onbekende verwachting μ en variantie θ. Dit betekent dat $S_n'^2$ die functie van de X_i's is welke samen met $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, de simultane normale verdelingsdichtheid (de afgeleide van (2.3)) van de X_i's als functie van θ en μ maximalizeert. Merk op dat $S_n'^2$ niet zuiver is. In het geval van een normale verdeling heeft $S_n'^2$ echter een kleinere MSE dan S_n^2.

Een schatter T_n noemt men *asymptotisch raak* (consistent) voor het schatten van θ, als

$$\lim_{n \to \infty} P(|T_n - \theta| \geqslant \epsilon) = 0, \text{ voor iedere } \epsilon > 0 \tag{2.12}$$

Indien (2.12) geldt, dan zegt men dat T_n in kans (of: in waarschijnlijkheid) naar θ convergeert, voor $n \to \infty$. We noteren dit als volgt:

$$T_n \xrightarrow[p]{} \theta, \text{ voor } n \to \infty \tag{2.13}$$

In veel gevallen hebben schatters sterkere *asymptotische* ($n \to \infty$) eigenschappen dan (2.12). Voor verschillende ruime klassen van schatters T_n van θ kan men bewijzen dan T_n *asymptotisch normaal* verdeeld is, met asymptotische verwachting θ en *asymptotische variantie* $\tau^2 = \tau_\theta^2$. Meer nauwkeurig:

$$\lim_{n\to\infty} P(n^{\frac{1}{2}}(T_n - \theta) \leqslant x) = \Phi(\frac{x}{\tau}) \tag{2.14}$$

voor $-\infty < x < \infty$, waarbij

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du, -\infty < x < \infty, \tag{2.15}$$

de standaard normale verdelingsfunctie voorstelt. De functie Φ is uitvoerig getabelleerd.

Naast (2.14) geldt vaak ook nog

$$\lim_{n\to\infty} n^{\frac{1}{2}}(ET_n - \theta) = 0 \tag{2.16}$$

en

$$\lim_{n\to\infty} n\sigma^2(T_n) = \tau^2 \tag{2.17}$$

d.w.z.: voor voldoend grote steekproefomvang n geeft $n^{-1}\tau^2$ een eenvoudige benadering voor de variantie van T_n.

Indien er een *asymptotische rake schatter* $V_n = V_n(X_1,..,X_n)$ van τ te vinden is, d.w.z. een schatter V_n waarvoor geldt:

$$V_n \xrightarrow[p]{} \tau, \text{ voor } n \to \infty \tag{2.18}$$

dan volgt wegens (2.14) onmiddelijk.

$$\lim_{n\to\infty} P(n^{\frac{1}{2}} V_n^{-1}(T_n - \theta) \leqslant x) = \Phi(x) \tag{2.19}$$

voor $-\infty < x < \infty$. Wegens (2.17) is dan $n^{-\frac{1}{2}} V_n$ een bruikbare schatter voor de standaardafwijking $\sigma_n = \sigma(T_n)$ van T_n. Voor voldoend grote steekproefomvang n zal $n^{-\frac{1}{2}} V_n$ met grote kans dichtbij σ_n liggen.

Voorbeeld 2.3 (i) $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, het steekproefgemiddelde, is een asymptotische rake schatter van het populatiegemiddelde $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$. Tevens is $\bar{X}_n$ asymptotisch normaal verdeeld met asymptotische variantie $\tau^2 = \int_{-\infty}^{\infty} (x-\theta)^2 dF(x)$ (mits $< \infty$), wegens de centrale limietstelling. Ook gel-

den (2.16) en (2.17).
(ii) $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, de steekproefvariantie is, evenals $S_n'^2 = \frac{n-1}{n} S_n^2$, een asymptotisch rake schatter van de populatievariantie $\theta = \theta(F) = \int_{-\infty}^{\infty} (x-\mu)^2 dF(x)$, met $\mu = \mu(F) = \int_{-\infty}^{\infty} x dF(x)$. Ook is S_n^2, evenals $S_n'^2$, asymptotisch normaal verdeeld met asymptotische variantie $\tau^2 = \mu_4 - \theta^2$ (mits $<\infty$), met μ_4 als in voorbeeld 2.2. Ook gelden (2.16) en (2.17).
(iii) Uit (i) en het eerste deel van (ii) volgt dat ook (2.19) geldt met $T_n = \bar{X}_n$, $V_n = S_n$ of S_n', en $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$.

We kunnen het resultaat (2.19) in de praktijk gebruiken om een *betrouwbaarheidsinterval* voor de parameter θ te construeren. Relatie (2.19) impliceert dat het interval

$$(T_n - \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{V_n}{\sqrt{n}},\ T_n + \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{V_n}{\sqrt{n}}) \tag{2.20}$$

een betrouwbaarheidsinterval voor θ met asymptotische ($n \to \infty$) betrouwbaarheid $1-\alpha$ voorstelt. Hierbij is $\Phi^{-1}(1-\frac{\alpha}{2})$ het $(1-\frac{\alpha}{2})$ de percentage punt van de standaard normale verdeling. Bijvoorbeeld: $\alpha = 0.05$, $\Phi^{-1}(1-\frac{\alpha}{2}) = \Phi^{-1}(0.975) = 1.96$. Er geldt:

$$\lim_{n\to\infty} P(T_n - \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{V_n}{\sqrt{n}} < \theta < T_n + \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{V_n}{\sqrt{n}}) = 1-\alpha \tag{2.21}$$

d.w.z.: voor voldoend grote steekproefomvang n is de kans dat het interval (2.20) de onbekende parameter θ bevat bij benadering gelijk aan $1-\alpha$.

Voorbeeld 2.4. $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$, het populatiegemiddelde. Een betrouwbaarheidsinterval voor θ met asymptotische ($n \to \infty$) betrouwbaarheid $1-\alpha$ wordt gegeven door:

$$\bar{X}_n - \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{S_n}{\sqrt{n}} < \theta < \bar{X}_n + \Phi^{-1}(1-\tfrac{\alpha}{2})\frac{S_n}{\sqrt{n}} \tag{2.22}$$

met $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ en $S_n = \sqrt{(n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$ (mits de populatievariantie $<\infty$ is). Een 95%-betrouwbaarheidsinterval voor het populatiegemiddelde θ wordt dus gegeven door (2.22) met $\Phi^{-1}(1-\frac{\alpha}{2}) = 1.96$. In de praktijk wordt 1.96 vaak door 2 benaderd.

Een uitvoerige behandeling van de schattingstheorie is te vinden in

Lehmann (1983).

3. Jackknife schatters

De jackknife methode werd door M. Quenouille in 1956 voorgesteld als een techniek om de onzuiverheid van een schatter te reduceren. In 1958 heeft J.W. Tukey opgemerkt dat de jackknife methode ook gebruikt kan worden om de variantie van een schatter te schatten en om betrouwbaarheidsintervallen voor onbekende parameters te construeren.

Jackknife schatters voor de onzuiverheid en de variantie van een schatter kunnen als volgt gedefineerd worden. Indien

$$T_n = T_n(X_1, \ldots, X_n) \tag{3.1}$$

een schatter is van een parameter θ, gebaseerd op een aselecte steekproef $X_1, \ldots, X_n$ van omvang n, dan is

$$T_{n-1}^i = T_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n),\ i = 1, \ldots, n \tag{3.2}$$

dezelfde schatter berekend op basis van de steekproef van omvang $n-1$ bestaande uit $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, welke uit de oorspronkelijke steekproef $X_1, \ldots, X_n$ verkregen wordt door de i de waarneming weg te laten. We beperken ons tot schatters T_n die symmetrische functies van de waarnemingen zijn.

De *jackknife schatter* b_{nJ} voor de *onzuiverheid* $b_n = ET_n - \theta$ (vgl (2.7)) van T_n wordt gegeven door

$$b_{nJ} = (n-1)(n^{-1}\sum_{i=1}^{n} T_{n-1}^i - T_n) \tag{3.3}$$

De *jackknife schatter* T_{nJ} van θ wordt gegeven door

$$T_{nJ} = T_n - b_{nJ} = nT_n - (n-1)\overline{T_{n-1}} \tag{3.4}$$

waarbij

$$\overline{T_{n-1}} = n^{-1}\textstyle\sum_{i=1}^{n} T_{n-1}^i \tag{3.5}$$

In veel gevallen heeft de jackknife schatter T_{nJ} van θ een *kleinere* onzuiverheid dan T_n. Dit is eenvoudig in te zien. Voor ruime klassen van schatters T_n geldt:

$$b_n = ET_n - \theta = \frac{a_\theta}{n} + \frac{b_\theta}{n^2} + \frac{c_\theta}{n^3} + \cdots. \tag{3.6}$$

waarbij de functies a_θ, b_θ en c_θ niet van n afhangen. Merk op dat wegens (3.5) dan ook geldt

$$E\overline{T}_{n-1} = \theta + \frac{a_\theta}{(n-1)} + \frac{b_\theta}{(n-1)^2} + \frac{c_\theta}{(n-1)^3} + \cdots \tag{3.7}$$

zodat (vgl (3.4))

$$ET_{nJ} - \theta = nET_n - (n-1)E\bar{T}_{n-1} - \theta =$$

$$= -\frac{b_\theta}{n(n-1)} + c_\theta\left(\frac{1}{n^2} - \frac{1}{(n-1)^2}\right) + \cdots \tag{3.8}$$

Vergelijken we nu (3.8) en (3.6) dan zien we dat de onzuiverheid van de jackknife schatter T_{nJ} van de orde n^{-2} is, terwijl T_n een onzuiverheid van de orde n^{-1} heeft.

Voorbeeld 3.1. Als in voorbeeld 2.1: $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$ en $T_n = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ en $ET_n = \theta$, T_n is zuiver voor θ. Inderdaad geldt (vgl (3.3)): $b_{nJ} = 0$, de jackknife schatter voor de onzuiverheid is gelijk aan nul.

Voorbeeld 3.2. Als in voorbeeld 2.2.: $\theta = \theta(F) = \int_{-\infty}^{\infty} (x - \mu(F))^2 \, dF(x)$, met $\mu(F) = \int_{-\infty}^{\infty} x dF(x)$ en $T_n = S_n'^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Een eenvoudige berekening geeft (vgl (3.3))

$$b_{nJ} = -\frac{1}{n(n-1)}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

zodat $T_{nJ} = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = S_n^2$, de zuivere schatter van θ uit voorbeeld 2.2.

De *jackknife schatter* s_{nJ}^2 voor de *variantie* $\sigma_n^2 = \sigma^2(T_n)$ van T_n wordt gegeven door

$$s_{nJ}^2 = \frac{n-1}{n}\sum_{i=1}^n (T_{n-1}^i - \bar{T}_{n-1})^2 \tag{3.9}$$

In de praktijk is men vaak meer geinteresseerd in de standaardafwijking σ_n van T_n i.p.v. in de variantie omdat schatters voor σ_n gebruikt kunnen worden bij de constructie van betrouwbaarheidsintervallen (zie de opmerking na (2.19) en relatie (2.20)). De jackknife schatter voor de standaardafwijking $\sigma_n = \sigma(T_n)$ van T_n is natuurlijk $s_{nJ} = \sqrt{s_{nJ}^2}$.

In veel gevallen blijkt s_{nJ}^2 een bruikbare schatter voor σ_n^2 te zijn, evenals s_{nJ} dat is voor σ_n. Dikwijls geldt

$$ns_{nJ}^2 \xrightarrow[p]{} \tau^2, \text{ voor } n \to \infty, \tag{3.10}$$

met τ als in (2.14), (2.17) en (2.18). D.w.z. $n^{\frac{1}{2}} s_{nJ}$ is een schatter die de rol van V_n in (2.17), (2.18) en (2.19) kan spelen. Wegens de opmerking na (2.19) is s_{nJ} dan een bruikbare schatter voor de standaardafwijking $\sigma_n = \sigma(T_n)$ van de schatter T_n. Ook kan, analoog aan (2.20), dit resultaat gebruikt worden om een *jackknife betrouwbaarheidsinterval* voor θ op te stellen. Indien T_{nJ} (zie. (3.4)) evenals T_n, asymptotisch normaal verdeeld is, met asymptotische variantie τ^2

(vgl (2.14)), dan impliceert (3.10) (vgl (2.20)) onmiddelijk dat het interval

$$(T_{nJ} - \Phi^{-1}(1-\frac{\alpha}{2})s_{nJ}, T_{nJ} + \Phi^{-1}(1-\frac{\alpha}{2})s_{nJ}) \tag{3.11}$$

een jackknife betrouwbaarheidsinterval voor θ met asymptotische $(n \to \infty)$ betrouwbaarheid $1-\alpha$ voorstelt. Merk op dat in vergelijking met (2.20) T_n vervangen is door T_{nJ} en $n^{-\frac{1}{2}}V_n$ door s_{nJ}.

Voorbeeld 3.3. Als in voorbeeld 2.4: $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$ en $T_n = \bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$. Merk op dat $T_n = T_{nJ}$ ($b_{nJ} = 0$ zie voorbeeld 3.1) en dat $T_{n-1}^i = (n-1)^{-1}(n\bar{X}_n - X_i)$, $\overline{T_{n-1}} = \bar{X}_n$, zodat wegens formule (3.9) $s_{nJ}^2 = (n(n-1))^{-1} \Sigma_{i=1}^n (X_i - \bar{X}_n)^2 = n^{-1}S_n^2$, waarbij S_n^2 de klassieke steekproefvariantie (zie voorbeeld 2.2) voorstelt. Aangezien $\sigma^2(\bar{X}_n) = \sigma^2(n^{-1}\Sigma_{i=1}^n X_i) = n^{-1}\sigma^2(X_1)$ is $s_{nJ}^2 = n^{-1}S_n^2$ dus niets anders dan de gebruikelijke schatter voor de variantie van een gemiddelde. Het jackknife betrouwbaarheidsinterval voor θ wordt gegeven door $(\bar{X}_n - \Phi^{-1}(1-\frac{\alpha}{2})\frac{S_n}{\sqrt{n}}$, $\bar{X}_n + \Phi^{-1}(1-\frac{\alpha}{2})\frac{S_n}{\sqrt{n}})$, en is dus precies gelijk aan het gebruikelijke betrouwbaarheidsinterval (2.22) voor het populatiegemiddelde.

Voorbeeld 3.4. Als in voorbeeld 2.2: $\theta = \theta(F) = \int_{-\infty}^{\infty} (x - \mu(F))^2 dF(x)$ met $\mu(F) = \int_{-\infty}^{\infty} x dF(x)$, en $T_n = S_n^2 = (n-1)^{-1}\Sigma_{i=1}^n (X_i - \bar{X}_n)^2$. De jackknife schatter s_{nJ}^2 voor de variantie $\sigma^2(S_n^2)$ kan eenvoudig m.b.v. formule (3.9) worden berekend. We vinden, na enig rekenwerk

$$s_{nJ}^2 = \frac{n^2}{(n-1)(n-2)^2}\{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^4 - (\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2)^2\} \tag{3.14}$$

Aangezien $T_{nJ} = T_n = S_n^2$ wordt het jackknife betrouwbaarheidsinterval voor de populatievariantie gegeven door (3.11), met $T_{nJ} = S_n^2$ en $s_{nJ} = \sqrt{(3.14)}$.

Een belangrijk voorbeeld van een schatter waarvoor de jackknife methode faalt is de *steekproefmediaan.* De steekproefmediaan $M_n = M_n(X_1, \ldots, X_n)$ van een steekproef $X_1, \ldots, X_n$ van omvang n wordt gedefineerd door

$$\begin{aligned} M_n &= X_{m:n} \quad \text{indien } n = 2m-1 \\ &= \tfrac{1}{2}(X_{m:n} + X_{m+1:n}) \quad \text{indien } n = 2m \end{aligned} \tag{3.15}$$

waarbij $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ de naar opklimmende grootte geordende waarnemingen $X_1, \ldots, X_n$ voorstellen. Men kan nagaan (zie Efron (1982)) dat voor het geval $n = 2m$ de jackknife schatter s_{nJ}^2 voor de variantie $\sigma^2(M_n)$ niet

bruikbaar is: relatie (3.10) blijkt niet te gelden!

Voor ruime klassen van schatters T_n, waaronder de in de statistiek zo belangrijke klasse van *meest aannemelijke schatters,* heeft Reeds (1978) echter laten zien dat de jackknife methode asymptotisch ($n \to \infty$) correcte schatters voor de onzuiverheid, de variantie en betrouwbaarheidsintervallen geeft.

Tenslotte: Efron & Stein (1981) hebben aangetoond dat de jackknife schatter s_{nJ}^2 voor de variantie van een willekeurige schatter T_n, ruwweg gesproken, gemiddeld te grote waarden oplevert. Voor ieder steekproefomvang n blijkt te gelden

$$\frac{n}{n-1} E s_{nJ}^2 \geqslant E(T_{n-1} - ET_{n-1})^2 \tag{3.16}$$

De ongelijkheid $E s_{nJ}^2 \geqslant \sigma^2(T_n)$ is echter niet altijd juist, hoewel er verschillende klassen van schatters bestaan waarvoor ook deze ongelijkheid opgaat.

4. Bootstrapschatters

Evenals in de vorige §'en is een schatter $T_n = T_n(X_1, \ldots, X_n)$ gebaseerd op een aselecte steekproef $X_1, \ldots, X_n$ van omvang n uit een populatie met verdelingsfunctie F gegeven, en moet een parameter $\theta = \theta(F)$ geschat worden. Indien F volledig onbekend is, maar een aselecte steekproef $X_1, \ldots, X_n$ uit F beschikbaar is, dan kan F geschat worden m.b.v. de *empirische verdelingsfunctie* F_n, gebaseerd op $X_1, \ldots, X_n$:

$$F_n(x) = \frac{1}{n}\{ \text{ aantal } X_i\text{'s} \leqslant x,\ 1 \leqslant i \leqslant n \} \tag{4.1}$$

voor $-\infty < x < \infty$. Merk op dat F_n een stochastische verdelingsfunctie is - een trapfunctie - die kansmassa n^{-1} toekent aan ieder van de n waarnemingen $X_1, X_2, \ldots, X_n$. Wegens de stelling van Glivenko - Cantelli geldt

$$P\left(\lim_{n \to \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\right) = 1 \tag{4.2}$$

Dit betekent dat F_n een (sterk) consistente schatter van F is, niet alleen lokaal (voor vaste x) maar ook globaal (uniform voor alle x)

De *bootstrap schatter* b_{nB} voor de *onzuiverheid* $b_n = b_n(F) = E_F T_n - \theta(F)$ (vgl (2.7)) van T_n wordt gegeven door

$$b_{nB} = b_n(F_n) = E_{F_n} T_n - \theta(F_n) \tag{4.3}$$

Volkomen analoog wordt de *bootstrapschatter* s_{nB}^2 voor de *variantie* $\sigma_n^2 = \sigma_n^2(F) = \sigma_F^2(T_n) = E_F(T_n - E_F T_n)^2$ (vgl (2.8)) van T_n gegeven door

$$s_{nB}^2 = \sigma_{F_n}^2(T_n) = E_{F_n}(T_n - E_{F_n} T_n)^2 \tag{4.4}$$

In de meeste gevallen kunnen bootstrapschatters niet exact worden uitgerekend, maar moeten zij m.b.v. Monte-Carlo simulaties worden benaderd. De formules (4.3) en (4.4) kunnen als volgt worden geinterpreteerd: gegeven de

waargenomen waarden $x_1, \ldots, x_n$ van de steekproef $X_1, \ldots, X_n$ (vgl (2.4)), d.w.z. de waargenomen waarde van de empirische verdelingsfunctie F_n, is $E_{F_n} T_n$ de verwachte waarde van $T_n = T_n(Y_1, \ldots, Y_n)$, waarbij $Y_1, \ldots, Y_n$ een aselecte steekproef (met teruglegging) van omvang n uit F_n voorstelt, voorwaardelijk gegeven F_n. De grootheid $\theta(F_n)$ in (4.3) stelt de waarde van te schatten parameter $\theta = \theta(F)$ (opgevat als functie van de onbekende verdeling F van de waarnemingen), geevalueerd in het "punt" $F = F_n$ voor. De interpretatie van $\sigma^2_{F_n}(T_n)$ is volkomen analoog, aan die van $E_{F_n} T_n$.

De - vooralsnog fictieve - steekproef $Y_1, \ldots, Y_n$ uit F_n noemt men de *bootstrap steekproef.* In tegenstelling tot $X_1, \ldots, X_n$ wordt de stochastiek in de Y_i's niet door de "natuur" veroorzaakt, maar door de statisticus zelf gegenercerd. In de praktijk gaat men vaak als volgt te werk. Met behulp van een geschikte *pseudo aselecte getallen generator,* zie bijvoorbeeld Ripley (1983), of van Es, Gill en van Putten (1983), wordt een zgn. pseudo-aselecte bootstrap steekproef $\tilde{Y}_1, \ldots, \tilde{Y}_n$ van omvang n uit F_n getrokken. De $\tilde{Y}_i$'s zijn niet stochastisch, maar zodanig geconstrueerd dat zij voor ons doel niet of nauwelijks te onderscheiden zijn van de stochastische Y_i's. Indien nu

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \tag{4.5}$$

de feitelijk waargenomen waarden in de steekproef $X_1, \ldots, X_n$ voorstellen, kunnen bootstrapschattingen voor de onzuiverheid of de variantie van T_n met behulp van een *Monte-Carlo algoritme* bepaald worden. Het algorithme kan in principe eenvoudig op een computer worden geimplementeerd en ziet er als volgt uit:

(1) Bepaal de empirische verdelingsfunctie F_n:

$$F_n \text{ geeft kansmassa } n^{-1} \text{ aan ieder van de waargenomen waarden}$$
$$x_1, x_2, \ldots, x_n \text{ in de steekproef} \tag{4.6}$$

(2) Onafhankelijk van elkaar worden een groot aantal - zeg M - pseudo - aselecte bootstrap steekproeven

$$\begin{matrix} \tilde{Y}_{11}, \ldots, \tilde{Y}_{n1} \\ \vdots \qquad\quad \vdots \\ \tilde{Y}_{1M}, \ldots, \tilde{Y}_{nM} \end{matrix} \tag{4.7}$$

getrokken uit F_n, en M bootstrapreplica's

$$\begin{aligned} \tilde{T}_{n1} &= T_n(\tilde{Y}_{11}, \ldots, \tilde{Y}_{n1}) \\ \tilde{T}_{nM} &= T_n(\tilde{Y}_{1M}, \ldots, \tilde{Y}_{nM}) \end{aligned} \tag{4.8}$$

van T_n bepaald.

(3) Bepaal tenslotte

$$\tilde{b}_{nB} = M^{-1} \sum_{i=1}^{M} \tilde{T}_{ni} - \theta(F_n) \tag{4.9}$$

en

$$\tilde{s}_{nB}^2 = (M-1)^{-1}\sum_{i=1}^{M}(\tilde{T}_{ni} - M^{-1}\sum_{i=1}^{M}\tilde{T}_{ni})^2 \tag{4.10}$$

Voor voldoend grote waarden van M zijn $\tilde{b}_{nB}$ en $\tilde{s}_{nB}^2$ praktisch gesproken (vrijwel) gelijk aan b_{nB} en s_{nB}^2

In sommige eenvoudige gevallen kunnen b_{NB} en s_{nB}^2 echter rechtstreeks worden berekend en is het niet nodig om Monte-Carlo simulaties uit te voeren.

Voorbeeld 4.1. Als in voorbeeld 3.1.: $\theta = \theta(F) = \int_{-\infty}^{\infty} x dF(x)$ en $T_n = \bar{X}_n = n^{-1}\Sigma_{i=1}^{n} X_i$; T_n is zuiver voor θ. Inderdaad geldt (vgl(4.3)):

$$b_{nB} = E_{F_n}\bar{Y}_n - \theta(F_n) = E_{F_n}\bar{Y}_n - \int_{-\infty}^{\infty} x dF_n(x)$$

$$= E_{F_n}\bar{Y}_n - \bar{X}_n = \bar{X}_n - \bar{X}_n = 0$$

Hierbij stelt $\bar{Y}_n$ het steekproefgemiddelde $n^{-1}\Sigma_{i=1}^{n} Y_i$ van de bootstrapsteekproef uit F_n voor, Analoog vinden we:

$$s_{nB}^2 = \sigma_{F_n}^2(\bar{Y}_n) = n^{-1}\sigma_{F_n}^2(Y_1) =$$

$$= n^{-1}E_{F_n}(Y_1 - E_{F_n}Y_1)^2 = n^{-1}E_{F_n}(Y_1 - \bar{X}_n)^2$$

$$= n^{-2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = n^{-1}\frac{n-1}{n}S_n^2.$$

Op de factor $\frac{n-1}{n}$ na is de bootstrapschatter s_{nB}^2 dus gelijk aan de jackknifeschatter $s_{nJ}^2 = n^{-1}S_n^2$ voor de variantie van $\bar{X}_n$ (vgl. voorbeeld 3.3).

Voorbeeld 4.2. Als in voorbeeld 3.4: $\theta = \theta(F) = \int_{-\infty}^{\infty}(x - \mu(F))^2 dF(x)$ met $\mu = \mu(F) = \int_{-\infty}^{\infty} x dF(x)$ en $T_n = S_n^2 = (n-1)^{-1}\Sigma_{i=1}^{n}(X_i - \bar{X}_n)^2$; T_n is zuiver voor θ. Inderdaad is de bootstrapschatter b_{nB} voor de onzuiverheid gelijk aan nul:

$$b_{nB} = E_{F_n}((n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2) - \frac{n-1}{n}S_n^2$$

$$= E_{F_n}(Y_1^2 - Y_1Y_2) - \frac{n-1}{n}S_n^2 = n^{-1}\sum_{i=1}^{n}X_i^2$$

$$-(n^{-1}\sum_{i=1}^{n}X_i)^2 - \frac{n-1}{n}S_n^2 = 0$$

De bootstrapschatter s_{nB}^2 voor de variantie $\sigma^2(S_n)$ kan eveneens bepaald worden. We vinden na enig rekenwerk

$$s_{nB}^2 = n^{-1}\{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^4 - \frac{n-3}{n-1}(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2)^2\}$$

Merk op dat s_{nB}^2 vrijwel gelijk is aan de jackknife schatter s_{nJ}^2 voor grote waarden van n (vgl. voorbeeld 3.4).

De bootstrap blijkt ook te werken in tal van situaties waarin de jackknife methode faalt. We noemen als voorbeeld de *steekproefmediaan* M_n (zie (3.15): Efron (1979) heeft laten zien de *bootstrapschatter* s_{nB}^2 voor de variantie $\sigma^2(M_n)$ voor grote steekproefomvang n met grote kans dichtbij $\sigma^2(M_n)$ ligt.

Een tweede punt ten gunste van de bootstrapmethode is dat deze techniek niet alleen geschikt is voor het schatten van de onzuiverheid of variantie van een schatter, maar in principe toegepast kan worden op vrijwel elke schattingsprobleem. Laat

$$R_n = R(X_1, \ldots, X_n; F) \tag{4.11}$$

een willekeurige statistische grootheid zijn, die niet alleen van de waarnemingen $X_1, X_2, \ldots, X_n$ afhangt maar ook van de verdelingsfunctie F. We willen nu de *verdeling van* R_n - of een of ander aspect van deze verdeling, bijvoorbeeld zijn scheefheid - schatten m.b.v. de bootstrapmethode. In het voorafgaande hebben we het speciale geval $R_n = T_n(X_1, \ldots, X_n) - \theta(F)$ bekeken, en moest of ER_n (= onzuiverheid van T_n als schatter van θ) of $\sigma^2(R_n)$ (= variantie van T_n) geschat worden. Het Monte-Carlo algorithme (4.6)-(4.10) kan - met enige wijzigingen - echter ook toegepast worden bij de bepaling van de verdeling van R_n in het algemene geval (4.11). In plaats van (4.8) berekenen we M bootstrapreplica's

$$\begin{aligned} \tilde{R}_{n1} &= R_n(\tilde{Y}_{11}, \ldots, \tilde{Y}_{n1}; F_n) \\ \tilde{R}_{nM} &= R_n(\tilde{Y}_{1M}, \ldots, \tilde{Y}_{nM}; F_n) \end{aligned} \tag{4.12}$$

en (4.9) en (4.10) wordt vervangen door

$$B_n(x) = \frac{1}{M}\{\text{aantal } \tilde{R}_{ni} \leqslant x, 1 \leqslant i \leqslant M\} \tag{4.13}$$

voor $-\infty < x < \infty$, de (Monte-Carlo benadering van de) *bootstrapschatter* B_n voor de *verdelingsfunctie* van R_n. Merk op dat B_n precies gelijk is aan de empirische verdelingsfunctie gebaseerd op de bootstraprelica's $\tilde{R}_{n1}, \ldots, \tilde{R}_{nM}$.

In het belangrijke speciale geval dat, R_n van de vorm

$$R_n = n^{\frac{1}{2}}(T_n - \theta) \tag{4.14}$$

is, waarbij $T_n = T_n(X_1, \ldots, X_n)$ een schatter voor een parameter $\theta = \theta(F)$ voorstelt, kan men voor verschillende klassen van schatters T_n bewijzen (Bickel & Freedman (1981))

$$P(\lim_{n\to\infty} \sup_x |P(n^{\frac{1}{2}}(T_n - \theta) \leqslant x) - B_n(x)| = 0) = 1. \tag{4.15}$$

De *bootstrap verdeling schatter* B_n geeft, evenals de klassieke normale benadering (zie §2)), in veel gevallen een goede benadering voor de verdeling van statistische grootheden van de vorm $n^{\frac{1}{2}}(T_n-\theta)$. Voor verschillende keuzen van T_n kan men zelfs bewijzen (Singh (1981), Beran (1984)) dat de bootstrap benadering B_n in zekere zin *nauwkeuriger* is dan de normale benadering.

Relatie (4.15) kan toegepast worden bij de constructie van betrouwbaarheidsintervallen gebaseerd op de bootstrap methode. Indien $C_{n\frac{\alpha}{2}-}$ en $C_{n\frac{\alpha}{2}+}$ het $\frac{\alpha}{2}$ de en $(1-\frac{\alpha}{2})$ de percentagepunt van de (gesimuleerde) bootstrapverdeling B_n van $n^{\frac{1}{2}}(T_n-\theta)$ voorstellen, dan is het interval

$$(T_n-\frac{C_{n\frac{\alpha}{2}+}}{\sqrt{n}}\ ,T_n-\frac{C_{n\frac{\alpha}{2}-}}{\sqrt{n}}) \tag{4.16}$$

een *bootstrap betrouwbaarheidsinterval* voor θ met asymptotische $(n\to\infty)$ betrouwbaarheid $1-\alpha$. In tegenstelling tot het jackknife betrouwbaarheidsinterval (3.11) voor θ is het interval (4.16) niet symmetrisch om T_n. Dit heeft te maken met het feit dat de bootstrap benadering B_n heel nauwkeurig de scheefheid van de exacte steekproefverdeling van $n^{\frac{1}{2}}(T_n-\theta)$ in beeld brengt, terwijl dit niet het geval is bij de normale benadering die gebruikt wordt bij de constructie van het jackknife betrouwbaarheidsinterval.

Tenslotte merken we op dat de klassieke jackknife techniek in zekere zin beschouwd kan worden als een *benadering* voor de meer algemeen toepasbare bootstrap methode. Met behulp van Taylor reeks ontwikkelingen kan men laten zien (Efron (1979), (1982)) dat verwachte waarde $E_{F_n}R_n$ en de variantie $\sigma^2_{F_n}(R_n)$ van de bootstrap verdeling van R_n benaderd kunnen worden door de jackknife schatters voor deze grootheden. Dit betekent dat de bootstrap schatters b_{nB} en s^2_{nB}, voor zekere klassen van statistische grootheden $R_n = T_n-\theta$, asymptotisch $(n\to\infty)$ equivalent blijken te zijn met de jackknife schatters b_{nJ} en s^2_{nJ} voor de onzuiverheid en de variantie van T_n. Men kan dus soms terugvallen op de jackknife methode. Dit heeft als voordeel dat de berekening van de schatters met veel minder rekentijd kan gebeuren, omdat géén Monte-Carlo simulaties behoeven te worden uitgevoerd. De prijs die men hiervoor echter in veel gevallen moet betalen is dat jackknifeschatters meestal minder nauwkeurig zijn (grotere MSE) dan bootstrap schatters.

De bootstrap methode is -integenstelling tot de jackknifetechniek- een vrijwel universecl toepasbare methode om de statistische variabiliteit van de schatters te bepalen en om betrouwbaarheidsintervallen voor onbekende parameters te construeren. Niet alleen kan de bootstrap gebruikt worden in de "één steekproef modellen", waartoe we ons deze voordracht beperkt hebben, de methode vindt ook toepassing in veel gecompliceerdere statistische modellen met weinig veronderstellingen, een situatie waarin de meer klassieke methoden vaak niet

bruikbaar zijn. Een voorbeeld is de constructie van een betrouwbaarheidsgebied voor regressie coefficienten in een lineair regressie model voor het geval zowel het aantal te schatten regressie coefficienten (parameters) p, als de steekproefomvang n groot is, (Bickel & Freedman (1983)).

Literatuur.

R. Beran (1984), Jackknife approximations to bootstrap estimates, The Annals of Statistics, Vol 12, p.101-118.

P.J. Bickel en D. Freedman (1981), Some asymptotic theory for the bootstrap, The Annals of Statistics, Vol 9, 1196-1217.

P.J. Bickel en D. Freedman (1983), Bootstrapping regression models with many parameters, in: A Festschrift for Erich. L. Lehmann, in honor of his sixty-fifth. birthday, eds. P.J. Bickel, K.A. Doksum, J.L. Hodges jr.

P. Diaconis en B. Efron (1983), Computer-intensive methods in statistics, Scientific American, Vol. 248, no. 5, 96-109.

B. Efron (1979), Bootstrap methods: another look at the jackknife, The Annals of Statistics, Vol. 7, 1-26.

B. Efron en C. Stein (1981), The jackknife estimate of variance, The Annals of Statistics, vol 9, 506-596.

B. Efron (1982), *The Jackknife, the Bootstrap and Other Resampling Plans,* SIAM, monograph 38, CBMS-NSF.

B. Efron en G. Gong (1983), A Leisurely Look at the Bootstrap, the Jackknife and Cross-Validation, The American Statistician, Vol 37, no. 1, 36-48.

A.J. van Es, R.D. Gill en C. van Putten (1983), Random number generators for a pocket calculator, Statistica Neerlandica, Vol. 37, p 95-102.

D. Freedman (1981), Bootstrapping regression models, The Annals of Statistics, Vol. 9, 1218-1228.

E.L. Lehmann (1983), *Theory of Point Estimation,* Wiley, New-York.

R.G. Miller (1964), A trustworthy jackknife. The Annals of Mathematical Statistics, Vol 35, 1594-1605.

R.G. Miller (1974). The jackknife: a review, Biometrika Vol 61, 1-15.

W.C. Parr (1983), A note on the jackknife, the bootstrap and the delta method estimators of bias and variance, Biometrika, Vol 70, p 719-722.

J.A. Reeds (1978), Jackknifing maximum likelihood estimates, The Annals of Statistics, Vol 6, 727-739.

B. Ripley (1983), Computer Generation of Random Variables: A Tutorial, International Statistical Review, Vol 51, 301-319.

K. Singh (1981). On the asymptotic accuracy of Efron's bootstrap. The Annals of Statistics, Vol 9, 1187-1195.

MEETKUNDE VAN DE RUIMTE

P.W.H. Lemmens en J.J. Seidel

§1. Inleiding.

Vroeger werden op de middelbare school, althans op de HBS, de vakken stereometrie, beschrijvende meetkunde en mechanica onderwezen. Deze vakken deden een groot beroep op het ruimtelijk inzicht van de leerling. Was dat inzicht aanwezig, dan konden "stereo" en "b.m." gebruikt worden om de overige wiskunde cijfers te compenseren, maar zonder dat inzicht waren het abstracte vakken en kon men slechts door moeizaam redeneren en toepassen van de stellingen tot resultaten komen. Een van de redenen was misschien wel dat er meestal weinig verband gelegd werd met de werkelijkheid om ons heen.

In deze voordrachten zal de aandacht gevestigd worden op een paar facetten van de ruimtemeetkunde, vooral vanuit een aantal voorbeelden. Enerzijds zullen we ons bezighouden met symmetrie eigenschappen van ruimtelijke figuren, voornamelijk toegelicht aan het regelmatig 20-vlak, en anderzijds met projecties en metrische eigenschappen. De hier gepresenteerde tekst is uitgebreider dan wat er in één lesuur over verteld kan worden.

§2. Regelmatige veelvlakken.

Het eenvoudigste veelvlak is het viervlak, het oppervlak van een ruimtelijk lichaam dat begrensd wordt door vier elkaar snijdende onder-

ling verschillende vlakken. Voor elk van deze vlakken geldt dat het lichaam zich geheel aan één kant van dat vlak bevindt. Het viervlak kan ook vastgelegd worden door zijn vier hoekpunten. Vooral aan deze laatste beschrijving ziet men gemakkelijk dat de vier hoekpunten zo gekozen kunnen worden dat er regelmaat ontstaat: men kan ze zo arrangeren dat elk zijvlak een gelijkzijdige driehoek is. In rechthoekige coördinaten kan men de hoekpunten van een regelmatig viervlak heel fraai kiezen als volgt:

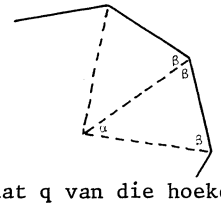
$$A = (1,-1,-1) \qquad B = (-1,1,-1) \qquad C = (-1,-1,1) \qquad D = (1,1,1)$$

Deze vectoren hebben alle dezelfde lengte ($\sqrt{3}$) en hun onderlinge inproducten zijn -1. Standaardiseren we ze op lengte 1, dat worden de onderlinge inproducten $-\frac{1}{3}$, met andere woorden: de 4 lijnen door het centrum van het tetraëder en de 4 hoekpunten vormen een gelijkhoekig lijnenstelsel in $\mathbb{R}^3$, waarvan elk tweetal met elkaar een (scherpe) hoek α maakt, waarbij $\alpha = \arccos \frac{1}{3}$.

We vragen ons af of er nog meer regelmatige veelvlakken zijn zo dat

1. de zijvlakken zijn regelmatige p-hoeken
2. in elk hoekpunt komen q van die p-hoeken samen.
3. het veelvlak begrenst een convex lichaam (dat is equivalent met te zeggen dat het lichaam aan één kant van elk van de uitgebreid gedachte zijvlakken ligt).

Van zo'n hypothetisch veelvlak (een {p,q}-polyeder) bekijken we eens nauwkeuriger wat er in een hoekpunt waar te nemen is. Blijkbaar komen daar q ribben samen als in een parapluie met q baleinen. Tussen elk tweetal naastliggende ribben zit een hoek van een regelmatige p-hoek, zeg ter grootte van 2β (zie figuur). Uit de figuur blijkt:



$$\left.\begin{aligned} p.\alpha &= 360° \\ \alpha + 2\beta &= 180° \end{aligned}\right\} \text{ dus } 2\beta = 180 - \frac{360}{p}$$

Enig meetkundig inzicht overtuigt ons ervan dat q van die hoeken 2β alleen op een convexe manier samen kunnen komen in een hoekpunt van het veelvlak als $q.2\beta < 360°$. Dit geeft een

voorwaarde voor het bestaan van een convex {p,q}-polyeder:

$$\frac{1}{p} + \frac{1}{q} > \frac{1}{2}.$$

Verder is het duidelijk dat p en q beide ≥ 3 moeten zijn. Hieruit krijgen we de te onderzoeken mogelijkheden van combinaties p, q:

$$\{p,q\} = \{3,3\}\ \{3,4\}\ \{3,5\}$$
$$\{4,3\}\ \{5,3\}.$$

Verrassend is het, dat deze mogelijkheden alle voorkomen, en bovendien nog dat elke mogelijkheid maar op één manier voorkomt. De vijf regelmatige veelvlakken zijn de z.g. Platonische lichamen:

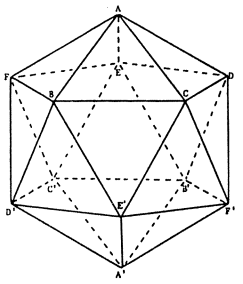
p = 3	q = 3	: tetraëder	: 4 hoekpunten, 6 ribben, 4 vlakken
p = 4	q = 3	: hexaëder	: 8 hoekpunten, 12 ribben, 6 vlakken
p = 3	q = 4	: octaëder	: 6 hoekpunten, 12 ribben, 8 vlakken
p = 5	q = 3	: dodecaëder	: 20 hoekpunten, 30 ribben, 12 vlakken
p = 3	q = 5	: icosaëder	: 12 hoekpunten, 30 ribben, 20 vlakken

Het tetraëder is het regelmatig viervlak waarmee we boven begonnen zijn, en het hexaëder is het regelmatig zesvlak, beter bekend als de kubus. Bij het octaëder komen 4 gelijkzijdige driehoeken bij elkaar in een hoekpunt, zoals bij de top van een piramide. Twee zulke piramiden met hun grondvlakken tegen elkaar geplakt vormen samen het octaëder. Ook het dodecaëder, bestaande uit 12 regelmatige vijfhoeken, is eenvoudig te beschrijven, maar het is minder inzichtelijk dat hierbij de vijhoeken netjes aan elkaar aansluiten. Daarom beschrijven we eerst het icosaëder in §3.

§3. Het icosaëder.

Bij dit 20-vlak komen in elk hoekpunt 5 driehoeken samen. Zetten we het icosaëder op een van zijn punten (hoekpunten) op tafel, dan onderscheiden we een onderkap en een bovenkap, elk de vorm hebbend van een 5-zijdige piramide met als "grondvlak" een regelmatige vijfhoek. Deze twee kappen zijn ten opzichte van elkaar iets gedraaid, zo-

dat tussen onder- en bovenkap een "ring" van 10 driehoeken past.



bovenkap A B C D E F
onderkap A'B'C'D'E'F'

Uit symmetrie overwegingen volgt dat elk tweetal tegenover elkaar liggende hoekpunten kan optreden als bovenste en onderste punt in deze beschrijving. De zes verbindingslijnen tussen deze paren hoekpunten gaan alle door het middelpunt van het icosaëder, en uit de figuur blijkt onmiddellijk dat elk tweetal van deze diagonalen met elkaar dezelfde hoek α maakt. We kunnen die hoek tamelijk gemakkelijk bepalen door nog eens naar de figuur te kijken. Denken we ons het middelpunt van het icosaëder aangegeven door de letter M en veronderstellen we dat de afstand van M tot de hoekpunten 1 is, dan zien we dat de vectoren $\overline{MA}$, $\overline{MB}$, $\overline{MC}$, $\overline{MD}$, $\overline{ME}$, $\overline{MF}$ paarsgewijs onderlinge inproducten $\pm \cos \alpha$ hebben, en wel $+ \cos \alpha$ als hun eindpunten verbonden zijn door een ribbe. Verder zien we dat $\overline{MA}$ precies centraal ligt tussen de overige vijf vectoren, zodat er een reëel getal x bestaat waarvoor geldt

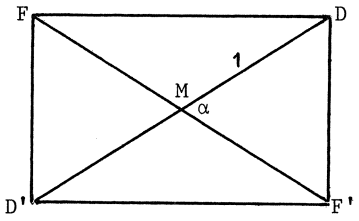
$$x.\overline{MA} = \overline{MB} + \overline{MC} + \overline{MD} + \overline{ME} + \overline{MF}.$$

Van deze vectoriële vergelijking kunen we beide kanten van het gelijkteken onderwerpen aan het inproduct met respectievelijk $\overline{MA}$ en $\overline{MB}$. Er ontstaan dan twee nieuwe vergelijkingen

$$x = 5 \cos \alpha \qquad \text{(inproduct met } \overline{MA}\text{)}$$
$$x.\cos \alpha = 1 \qquad \text{(inproduct met } \overline{MB}\text{)},$$

waaruit volgt dat $\cos \alpha = \frac{1}{\sqrt{5}}$, dus $\alpha \approx 63°26'6"$.

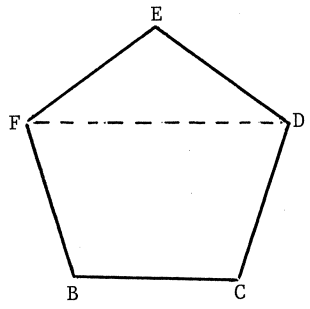
Dit heeft aardige consequenties. Allereerst lichten we de rechthoek FDF'D' (waarom is dit een rechthoek?) uit de figuur



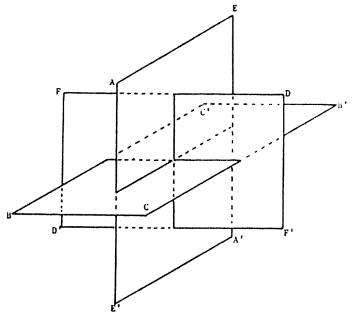
Met de cosinusregel vinden we dan $|DF'|^2 = 2 - 2/\sqrt{5}$, en vervolgens met de regel van Pythagoras $|DF|^2 = 2 + 2/\sqrt{5}$. Hieruit leiden we voor de verhouding van $|DF|$ en $|DF'|$ af

$$\frac{|DF|}{|DF'|} = \frac{1}{2}\sqrt{5} + \frac{1}{2} \quad \text{en} \quad \frac{|DF'|}{|DF|} = \frac{1}{2}\sqrt{5} - \frac{1}{2}.$$

Het getal $s = \frac{1}{2}\sqrt{5} - \frac{1}{2}$ is de bekende gulden snede: wordt een lijnstuk verdeeld in twee stukken die zich verhouden als $s : 1$, dan verhoudt de grootste van de twee zich tot het geheel ook als $s : 1$, s is de positieve wortel van de vergelijking $s^2 + s - 1 = 0$.
Kijken we nog eens naar de figuur van het icosaëder, dan zien we dat DF tevens een diagonaal is van de regelmatige vijfhoek B C D E F, en hieruit volgt dan weer dat een regelmatige vijfhoek met gegeven zijde te construeren is met uitsluitend gebruik van passer en lineaal.



We zijn nog niet uitgekeken op het icosaëder! Wegens de constructie (ABC is een gelijkzijdige driehoek, en BCDEF is een regelmatige vijfhoek) moeten de hoekpunten A en E beide in het middelloodvlak van BC liggen. Zetten we dit voort, dan blijkt dat de vlakken van de rechthoeken AEA'E', BCB'C' en DF'D'F in de ruimte loodrecht op elkaar staan! De hoekpunten van het icosaëder kunnen dus verkregen worden door drie gelijke kaarten met de gulden snede als breedte: hoogte verhouding loodrecht op elkaar te plaatsen, met samenvallende middens. De gulden snede komt als verhouding in de natuur veelvuldig voor. Men vindt deze verhouding ook in de kunst terug, o.a. bij het gezicht van de Mona Lisa. Men kan het icosaëder dus beschouwen als een soort ruimtelijke Mona Lisa.



Het aardige hiervan is dat we nu in een rechthoekig coördinatenstelsel de hoekpunten van een icosaëder gemakkelijk kunnen opschrijven als

$$\{(0,\pm 1,\pm s),\ (\pm s,0,\pm 1),\ (\pm 1,\pm s,0)\}$$

waarbij $s = \frac{1}{2}\sqrt{5} - \frac{1}{2}$.

We komen nu terug op het dodecaëder. Dit moet volgens het overzicht in §2 20 hoekpunten en 12 vlakken hebben. Het kan verkregen worden door als hoekpunten de middens van de driehoeken van het icosaëder te nemen, en als vlakken juist de vijfhoeken die gevormd worden door de middens van de driehoeken die in één hoekpunt van het icosaëder samenkomen. Het op deze manier verkregen dodecaëder staat in feite in polaire verwandschap met het icosaëder via de pool-poolvlakrelatie t.o.v. de ingeschreven bol in het icosaëder. Merk op dat iedere ribbe van het icosaëder hoort bij een ribbe van het dodecaëder, ze zijn elkaars poollijn en staan loodrecht op elkaar. Op eenzelfde wijze kan men overigens uit het hexaëder (de kubus) een polair octaëder verkrijgen, uit het octaëder een hexaëder en uit het dodecaëder weer een icosaëder! Probeert men deze constructie uit te voeren bij het tetraëder, dan ontstaat weer een tetraëder: het tetraëder is zelf-duaal.

In het icosaëder kan men ook een tetraëder construeren. De 12 hoekpunten laten zich opsplitsen in 4 groepen van 3, elk drietal de hoekpunten van een driehoek vormend. Een voorbeeld van zo'n opsplitsing is ABC, B'DF', C'EF, A'D'E'. Met behulp van de bekende inprodukten rekent men eenvoudig na dat de middens van deze 4 driehoeken

precies als de hoekpunten van een tetraëder kunnen fungeren. In totaal kunnen zo 10 tetraëders in het icosaëder gevormd worden.

§4. De groep van de draaiingen van het icosaëder.

Een icosaëder kunnen we met elk van zijn driehoeken op tafel plaatsen, en steeds biedt het dezelfde aanblik, eventueel is een kleine draaiing om de vertikale as nodig. Er zijn 20 driehoeken, en elke driehoek laat 3 draaiingen toe waarbij de driehoek, en (wegens de starre constructie) dus het gehele icosaëder weer in dezelfde positie staat. In totaal geeft dit 3x20 = 60 ruimtelijke bewegingen die het icosaëder invariant laten (maar wel de hoekpunten permuteren). Wegens de aard van de bewegingen moeten dit draaiingen zijn. Een draaiing in de ruimte (behalve de identiteit) is eenduidig bepaald door draaias en draaihoek. Welke zijn de mogelijkheden? Aan een model zien we onmiddellijk een aantal voor de hand liggende draaiassen en draaihoeken nl.

6 assen door paren tegenover elkaar liggende hoekpunten met 4 draaihoeken,
15 assen door de middens van paren tegenover elkaar liggende ribben, met 1 draaihoek
10 assen door de middens van paren tegenover elkaar liggende driehoeken, met 2 draaihoeken.

Samen leveren deze al 6x4 + 15x1 + 10x2 = 59 verschillende draaiingen, en tezamen met de identiteit hebben we nu dus alle ruimtelijke bewegingen die het icosaëder invariant laten geïdentificeerd met echte draaiingen van dat icosaëder in de ruimte. Ze vormen de draaiingsgroep R van het icosaëder.
Deze draaiingsgroep is een ondergroep van de symmetriegroep S van het icosaëder. S krijgen we door bij de draaiingen ook nog de spiegelingen te nemen die het icosaëder invariant laten. Onder spiegelingen verstaan we hierbij isometrieën van de ruimte die determinant -1 hebben. Dit is iets algemener dan wat normaal onder spiegelen in een vlak verstaan wordt, ook de z.g. draaispiegelingen (spiegelen in een vlak, gevolgd door een draaiing om de loodlijn op dat vlak) vallen

hieronder. In een volgende paragraaf komen we terug op de symmetriegroep S. Hier zullen we ons nu alleen met de draaiingsgroep R bezighouden.

Voor iemand die bekend is met permutatiegroepen (de permutatiegroep S_n is de groep van de mogelijke rangschikkingen van n verschillende elementen, en heeft dus n! elementen) zal het aantal van 60 elementen in de draaiingsgroep van het icosaëder een optrekken van de wenkbrauwen induceren: S_5, de permutatiegroep van 5 elementen heeft nl. 5! = 120 elementen, en de ondergroep A_5 van de even permutaties (permutaties waarvoor een even aantal paarverwisselingen nodig zijn) heeft 60 elementen! Hierdoor ontstaat de vraag of R misschien met A_5 isomorf is. Deze vraag kan in positieve zin beantwoord worden als er 5 "elementen" aan te wijzen zijn waarop R als een groep van permutaties werkt. Welnu, die 5 elementen zijn er! We kijken nog eens wat nauwkeuriger naar de orthogonale kaarten constellatie waarvan in §3 melding gemaakt is. Op te merken valt

1. elke constellatie bevat 6 ribben van het icosaëder, nl. de korte zijden van de drie kaarten,
2. een volledige constellatie ligt vast door één ribbe
3. elke constellatie bevat van elk hoekpunt van het icosaëder precies één van de vijf ribben die in dat hoekpunt samenkomen.

Hieruit volgt dat er precies 5 van die kaarten constellaties zijn, en met een beetje overleg kan men inzien dat deze onderling gepermuteerd worden door de rotaties in R.
Opmerking: betere namen voor S_5 en A_5 zijn respectievelijk de symmetrische groep op 5 elementen en de alternerende groep op 5 elementen. Onder een permutatiegroep verstaat men veelal een groep van niet noodzakelijk alle permutaties.

§5. De icosaëdergroep.

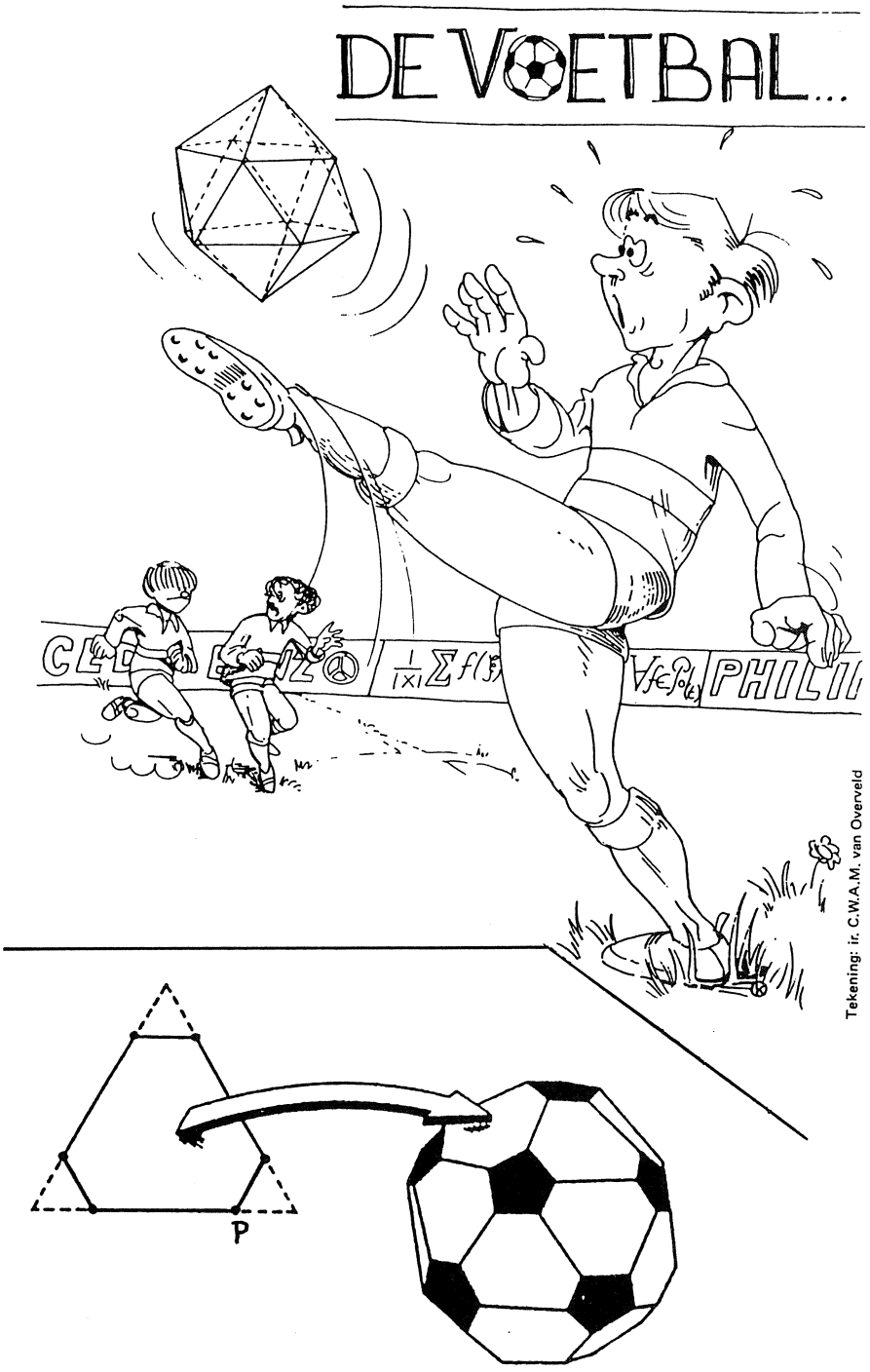
Zoals reeds opgemerkt in §4 vormt de groep R van draaiingen slechts een ondergroep van de volledige groep S van symmetrieën van het icosaeder. S heet ook wel de icosaëdergroep.
Omdat de samenstelling van twee spiegelingen een draaiing oplevert, doet zich het prettige verschijnsel voor dat voor de beschrijving van de groep S gebruik gemaakt kan worden van de groep R, met daarbij

nog één spiegeling s. Leggen we die spiegeling s vast, dan kan elke symmetrie van het icosaëder geschreven worden als een draaiing gevolgd door de spiegeling s (dus een element van s.R). Immers, zij a een symmetrie met determinant -1, dan is s^{-1}.a een draaiing r, dus a = s.r.
Bijgevolg heeft de icosaëdergroep 120 elementen, en de vraag is weer gerechtvaardigd of S wellicht isomorf is met de hele permutatiegroep S_5 van 5 elementen (in §4 hebben we gezien dat R isomorf is met A_5). Om aan te tonen dat dit niet het geval is gebruiken we een truc. In bovenstaande beschrijving van S gebruikten we de groep R met daarbij nog een vast gekozen maar niet nader gespecifieerde spiegeling s. Voor deze s kiezen we nu een zeer speciale spiegeling, en wel de centrale spiegeling in het middelpunt van het icosaëder. In feite is dit een vermenigvuldiging met -1. Daarom commuteert deze centrale spiegeling met alle elementen van S. Als nu S isomorf was met S_5, dan zou er in de permutatiegroep S_5 ook een permutatie aan te wijzen moeten zijn (ongelijk aan de identiteit) die met alle permutaties commuteert. Men kan nagaan dat zo'n permutatie niet bestaat. Dus kan S niet isomorf zijn met S_5. Voor degenen die vertrouwd zijn met groepentheorie merken we nog op dat de schrijfwijze $S = R \cup s.R$ (waarbij s nu de centrale spiegeling is) aantoont dat S de directe som is van twee groepen, nl. van R en $\{1,s\}$, in vaktermen: $S \cong R \oplus \mathbb{Z}_2$.

§6. De icosaëdergroep en andere lichamen.

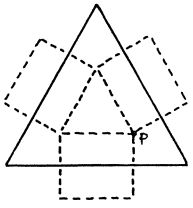
Fixeren we het icosaëder en bekijken we daarop een vast gekozen punt P (niet noodzakelijk een hoekpunt), dan kunnen de 120 symmetrieën van het icosaëder op dit punt P worden toegepast. Er ontstaat dan een eindige verzameling van punten op het icosaëder, de baan van P. Deze punten kunnen we dan weer opvatten als de hoekpunten van een nieuw lichaam, waarop ook de icosaëder groep werkt. Mooie exemplaren worden verkregen wanneer P speciaal gekozen wordt. Een aantal voorbeelden:

1. Kiezen we voor P een hoekpunt van het icosaëder, dan ontstaat weer het icosaëder zelf, elk punt van de baan van P komt 10 keer voor.

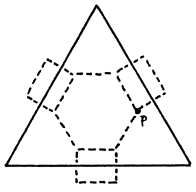


Tekening: ir. C.W.A.M. van Overveld

2. Als we voor P het midden van een driehoekje nemen, dan ontstaat als baan een verzameling van 20 punten, elk 6 keer voorkomend, en de bijbehorende figuur is uiteraard het dodecaëder.
3. De baan van het midden van een ribbe bestaat uit de 30 middens van alle ribben, elke wordt 4 keer geteld. De ontstane figuur is het zogenaamde icosidodecaëder, samengesteld uit 20 gelijkzijdige driehoekjes (liggend in de "grote" driehoeken van het icosaëder) en 12 regelmatige vijfhoeken (als het ware de afgevijlde toppen van het icosaëder.
4. Nemen we een punt op een derde van een ribbe, dan ontstaat de voetbal met zijn 20 regelmatige zeshoeken en 12 regelmatige vijfhoeken, hier komt elk punt in de baan twee maal voor.
5. Men kan ook figuren krijgen, samengesteld uit vierkanten, driehoeken en vijfhoeken, of uit zeshoeken, vierkanten en tienhoeken, steeds door andere keuze van P. We geven dit aan met een tekening.



20 3-hoeken
30 4-hoeken
12 5-hoeken



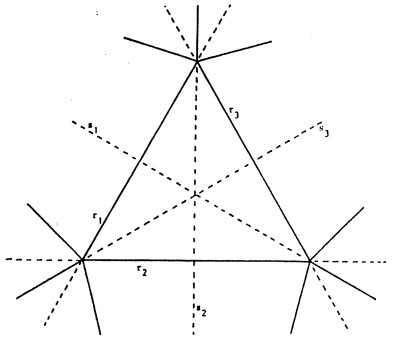
20 6-hoeken
30 4-hoeken
12 10-hoeken

In het laatste geval bestaat de baan van P uit 6x20 = 120 punten, zodat elk precies eenmaal voorkomt.

§7. Spiegelingen.

We hebben reeds gezien dat de icosaëdergroep bestaat uit draaiingen en spiegelingen. Aan de andere kant is het eenieder welbekend dat het product van twee gewone spiegelingen (elk in een vlak) een draaiing oplevert met als draaias de snijlijn van de spiegelvlakken en als draaihoek het dubbele van de standhoek tussen de spiegels. Vallen de spiegelvlakken samen dan is het resultaat de identiteit, en zijn de spiegelvlakken evenwijdig (hetgeen hier niet voorkomt) dan is

het resultaat een translatie. In onze eerdere beschrijving van de icosaëdergroep zijn we uitgegaan (§4) van draaiingen en hebben daar vervolgens spiegelingen (§5) aan toegevoegd. In deze paragraaf willen we een ander standpunt innemen en ons afvragen of de icosaëdergroep niet uitsluitend kan worden voortgebracht door een aantal spiegelingen, waarvan alle mogelijke producten de gehele groep van symmetrieën vormen. Bij het icosaëder is het duidelijk hoe we een spiegelvlak moeten aanbrengen, namelijk door het midden van het icosaëder, en bevattend een ribbe van het icosaëder (dit spiegelvlak bevat dan ook de tegenoverliggende ribbe). Eigenlijk kiezen we dan als spiegelvlak een van de kaarten uit een kaartenconstellatie zoals beschreven in §3. Aan de hand van de daar opgenomen tekening is onmiddellijk duidelijk dat zo'n spiegeling het icosaëder invariant laat. Bovendien zien we uit die tekening dat de centrale spiegeling in het midden van het icosaëder te verkrijgen is als product van drie van deze spiegelingen, immers neem als spiegelvlakken maar de drie kaarten van zo'n kaartenconstellatie! Van de in totaal 15 mogelijke spiegelvlakken wordt elke driehoek van het icosaëder "aangesneden" door 6 stuks: de drie spiegelvlakken door de zijden van de driehoek, en drie spiegelvlakken door het midden van de driehoek en door een van de hoekpunten ervan.



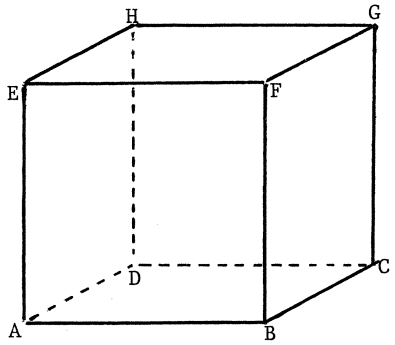
We geven deze vlakken aan met kleine letters r_1, r_2, r_3 en s_1, s_2, s_3, welke we ook zullen gebruiken om de spiegelingen in die vlakken aan te duiden.

Men ziet direct in dat r_2s_2 de draaiing van 180° is om het midden van de met r_2 corresponderende ribbe, dat r_2s_1 de draaiing van 72° is om een hoekpunt, en dat s_1s_2 een draaiing van 120° is om het midden van de driehoek. Iets moeilijker te zien is dat $s_1r_2s_1 = r_3$, daarvoor moeten we de tekening uitbreiden of beter nog de beschikking hebben over een echt model van het icosaëder. Zo kan men ook inzien dat $s_3 =$

$= s_1 s_2 s_1$ en $r_1 = s_2 r_3 s_2 = s_2 s_1 r_2 s_1 s_2$, met andere woorden: het lijkt erop dat de spiegelingen s_1, s_2 en r_2 voldoende zijn om de hele symmetriegroep van het icosaëder voort te brengen (d.w.z. elke symmetrie kan worden verkregen als productvorm van deze drie). Bij nadere uitwerking blijkt dit inderdaad het geval te zijn!
De consequentie hiervan is dat we het hele icosaëdron kunnen zien in een kaleidoskoop gevormd door drie spiegels r_2, s_1, s_2 die door één punt gaan, de spiegels r_2 en s_2 staan loodrecht op elkaar, s_1 heeft met s_2 een standhoek van 60° (= 180/3), en met r_2 een standhoek van 36° (= 180/5).
Tekent men dan bijvoorbeeld met een lipstift in deze kaleidoskoop op het spiegelvlak r_2 een lijnstuk vanuit een punt van de snijlijn van r_2 en s_1 naar de snijlijn van r_2 en s_2, loodrecht op deze laatste snijlijn, dan vormen de spiegelbeelden hiervan een compleet ruimtelijk icosaëder!

§8. Projectiemethoden.

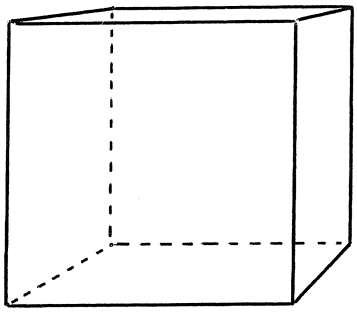
Bij het tekenen van ruimtelijke figuren op papier moet men altijd gebruik maken van een of andere vorm van projectie van de ruimte op het papier, het projectievlak. We zullen van de meest gebruikte methoden de elementaire karakteristieken behandelen.
De manier om een kubus af te beelden is meestal via de volgende tekening, waarbij de ribben AB, AE onverkort zijn, doch de ribbe AD verkort (factor ~ 1/2) wordt aangegeven. De hoek tussen AE en AB is recht, die tussen AD en AB is ongeveer 30°. In werkelijkheid evenwijdige lijnen worden ook evenwijdig getekend. We hebben hier te maken met de z.g. <u>parallelprojectie</u>.



Men stelle zich voor dat de kubus geplaatst is in het eerste octant van een rechthoekig coör-

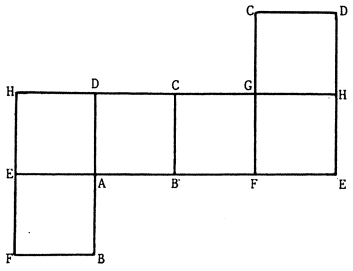
dinatenstelsel, D in de oorsprong, DA langs de x-as, DC langs de y-as en DH langs de z-as. Als vlak van tekening nemen we het yz-vlak en projecteren daarop de kubus als de schaduw onder een evenwijdige bundel lichtstralen. In het voorbeeld zouden de lichtstralen evenwijdig zijn aan de lijn door $(0,-\sqrt{3},-1)$ en $(4,0,0)$ (ga na!).
Men dient zich goed te realiseren dat de parallelprojectie nooit een afbeelding levert zoals we die (van bijvoorbeeld de kubus) in werkelijkheid waarnemen, zelfs niet als we één oog dichtknijpen.
Een andere wijze om een ruimtelijke figuur te tekenen is de centrale projectie. In wezen doen we dan hetzelfde als boven beschreven, maar laten de schaduwwerpende lichtstralen uit één punt komen, zoals bij een lamp (bij benadering). Plaatsen we de lamp in het punt $(x,y,z) = (24,5\sqrt{3},5)$ dan is uit te rekenen op welke punten van het yz-vlak de schaduw van de hoekpunten valt, en krijgt men de volgende perspectivische tekening.

*



Ook hier is het zo dat we de kubus niet zien zoals in werkelijkheid...., behalve wanneer we ons oog houden recht boven het hierboven getekende sterretje, op een afstand van 24 cm, immers dan bevindt het oog zich op de plaats van de lamp!

Als derde mogelijkheid noemen we orthogonale projectie waarbij van het voorwerp een voor-, zij-, en bovenzicht wordt getekend op respectievelijk de drie coördinaatvlakken volgens parallelprojectie in de richting loodrecht op die vlakken.
Tenslotte zullen we enige aandacht besteden aan een mogelijkheid die met name bij veelvlakken goed te gebruiken is, en wel het netwerk, of de neerslag van zo'n veelvlak, waarbij het veelvlak als het ware langs de ribben opengeknipt wordt en de zijvlakken vervolgens kunnen worden uitgevouwen tot een figuur in het tekenvlak. Voor de kubus zou zo'n netwerk op schaal 1 : 3 er uit kunnen zien als



§9. Parallelprojectie.

Met de centrale en orthogonale projectie heeft parallelprojectie gemeen dat

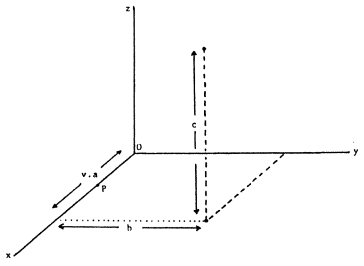
1. de projectie van een punt is een punt,
2. de projectie van een rechte lijn is een rechte lijn of een punt,
3. de projecties van elkaar snijdende rechten snijden elkaar in de projectie van het snijpunt.
4. de projectie van een punt van het projectievlak is dat punt zelf.

Bij de parallelprojectie kunnen we hier nog aan toevoegen dat de projecties van evenwijdige rechten alle bestaan uit punten of alle onderling evenwijdige rechten zijn. Bovendien geldt nog dat even lange evenwijdige lijnstukken ook in de projectie onderling gelijke lengten hebben.

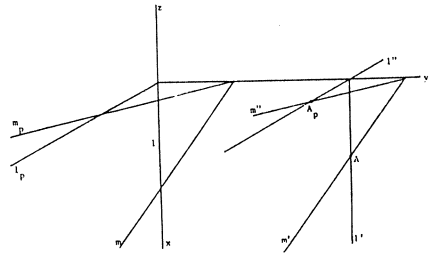
Omdat lijnstukken in het projectievlak onverkort worden geprojecteerd (immers ze zijn hun eigen projectie), volgt uit het bovenstaande dat alle lijnstukken die evenwijdig lopen aan het projectievlak onverkort worden geprojecteerd.

Om de gedachten te bepalen nemen we in een rechthoekig coördinatenstelsel het yz-vlak als projectievlak. Men gaat gemakkelijk na dat de projectie-richting vastgelegd wordt door van één punt, niet liggend in het projectievlak, de projectie te geven. Het ligt het meest voor de hand om voor dit punt (1,0,0) te nemen. Zij P het projectiepunt

hiervan, en O de oorsprong, dan heet de lengte van OP de verkortingsfactor v (hoewel het ook een verlenging kan zijn). De lijn door O en P is de projectie van de x-as. We sluiten hier uit dat O met P samenvalt, in dat geval hebben we met een orthogonale projectie te maken. Is nu in de ruimte een punt (a,b,c) gegeven, dan kan de projectie hiervan als volgt bepaald worden. Zet langs de y-as b eenheden af, en trek door het verkregen punt een lijn evenwijdig aan OP, zet hierop v.a eenheden af en trek door het nu verkregen punt een lijn evenwijdig aan de z-as en zet hierop tenslotte c eenheden af. Het laatst verkregen punt is de projectie van (a,b,c).

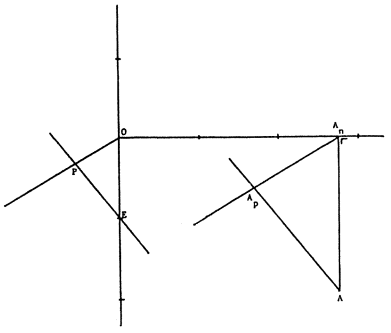


Hoewel deze methode zeer eenvoudig is, komt er erg veel meetwerk aan te pas. Een alternatieve constructie kan als volgt verkregen worden. We wentelen het werkelijke xy-vlak om de y-as tot het in het projectievlak (het yz-vlak) ligt. Nu weten we van een lijn in het xy-vlak dat zijn projectie zal gaan door het snijpunt van die lijn met de y-as. Weten we van één zo'n lijn de projectie, dan is het meteen duidelijk welke de projecties zijn van alle daarmee evenwijdige in het xy-vlak verlopen lijnen. Zouden we nu van twee elkaar snijdende lijnen l, m in het xy-vlak de projecties kennen, zeg l_p en m_p, dan kunnen we van elk punt A in het xy-vlak de projectie A_p bepalen door vanuit A lijnen l', m' te trekken evenwijdig aan l, m, en vanuit hun snijpunten met de y-as lijnen y", m" evenwijdig met l_p, m_p. Het snijpunt van l" en m" is A_p.



In de tekening hebben we voor l de gewentelde x-as gekozen. Ook voor m is een speciale lijn

voorhanden; zetten we immers op de gewentelde x-as een eenheid af, dan zal van elke lijn door dat punt E in het xy-vlak de projectie door P gaan, in het bijzonder zal de lijn door E en P in de tekening samenvallen met zijn projectie! De constructie verloopt nu als volgt: Trek vanuit A de loodlijn op de y-as met snijpunt A_n, vanuit A_n een lijn evenwijdig met OP en vanuit A een lijn evenwijdig met EP. Het snijpunt van de laatste twee lijnen is A_p. De driehoek AA_nA_p staat bekend onder de naam projectiedriehoek.



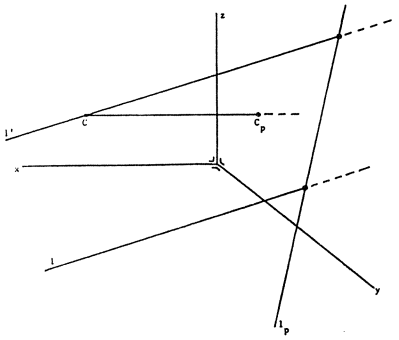
Daar het snijpunt van twee lijnen het nauwkeurigst te bepalen is als ze ongeveer loodrecht op elkaar staan, blijkt het voor deze constructie voordelig te zijn als de verkortingsfactor ongeveer gelijk is aan de sinus van de hoek die de projectie van de x-as maakt met de y-as.

Deze constructiemethode spaart vooral veel tijd als men van een in het xy-vlak op ware gedaante getekende figuur de projectie wil bepalen. Ligt een figuur in een aan het xy-vlak evenwijdig vlak, dan behoeft men slechts de y-as in de tekening te verschuiven, en kan verder dezelfde methode gevolgd worden. Op soortgelijke wijze kan te werk gegaan worden bij het projecteren van figuren in het xz-vlak. Dit vlak wentelt men dan om de z-as tot het in het tekenvlak komt te liggen.

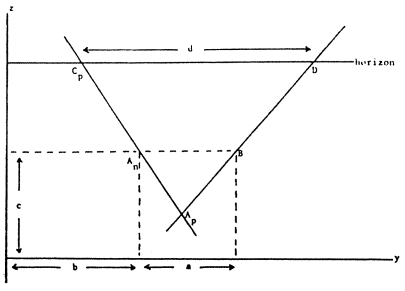
§10. Centrale projectie.

De centrale projectie is conceptueel veel moeilijker dan de parallel- en de orthogonale projectie. Weliswaar gelden ook voor de centrale projectie de vier basisregels die we aan het begin van §9 opgesteld hebben, maar een grote handicap is dat in werkelijkheid even lange stukken op één lijn niet altijd als lijnstukken van gelijke lengte geprojecteerd worden. Bovendien is er de complicatie dat onderling evenwijdige lijnen, niet evenwijdig aan het tekenvlak, in de projectie door één punt gaan (het vluchtpunt van die lijnen), terwijl

evenwijdige lijnen die bovendien nog evenwijdig aan het tekenvlak lopen ook in de projectie als evenwijdige lijnen verschijnen. In een



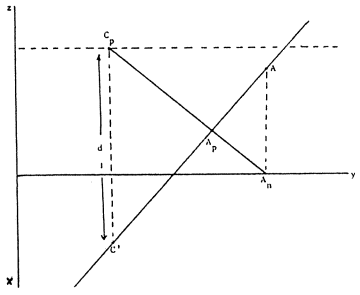
ruimtelijke tekening zien we wat er gebeurt. Laat voor het gemak het yz-vlak het tekenscherm zijn. Het centrum van de projectie zij C, met C_p het voetpunt van de loodlijn vanuit C op het scherm. C_p is het vluchtpunt van alle lijnen loodrecht op het scherm. Willen we het vluchtpunt bepalen van een willekeurige lijn 1, dan trekken we door C een lijn 1' evenwijdig met 1, en bepalen het snijpunt van 1' met het tekenscherm (hier het yz-vlak). In het bijzonder blijkt dat van alle lijnen die evenwijdig aan het xy-vlak lopen het vluchtpunt zich bevindt op de lijn in het scherm die door C_p gaat en evenwijdig aan de y-as loopt. Deze lijn is de <u>horizon</u>. Een ander belangrijk punt voor de projectie van een lijn 1 is het doorgangspunt, het snijpunt van de lijn 1 met het scherm. De projectie van 1 zal dus een lijn 1_p zijn door vluchtpunt en doorgangspunt van 1. Is nu het scherm met z-as en y-as gegeven, daarop bovendien het <u>hoofdpunt</u> C_p, en weten we bovendien de <u>distantie</u> d = $|CC_p|$, dan kan van het punt A = = (a,b,c) in x,y,z-coördinaten als volgt de projectie A_p bepaald worden.

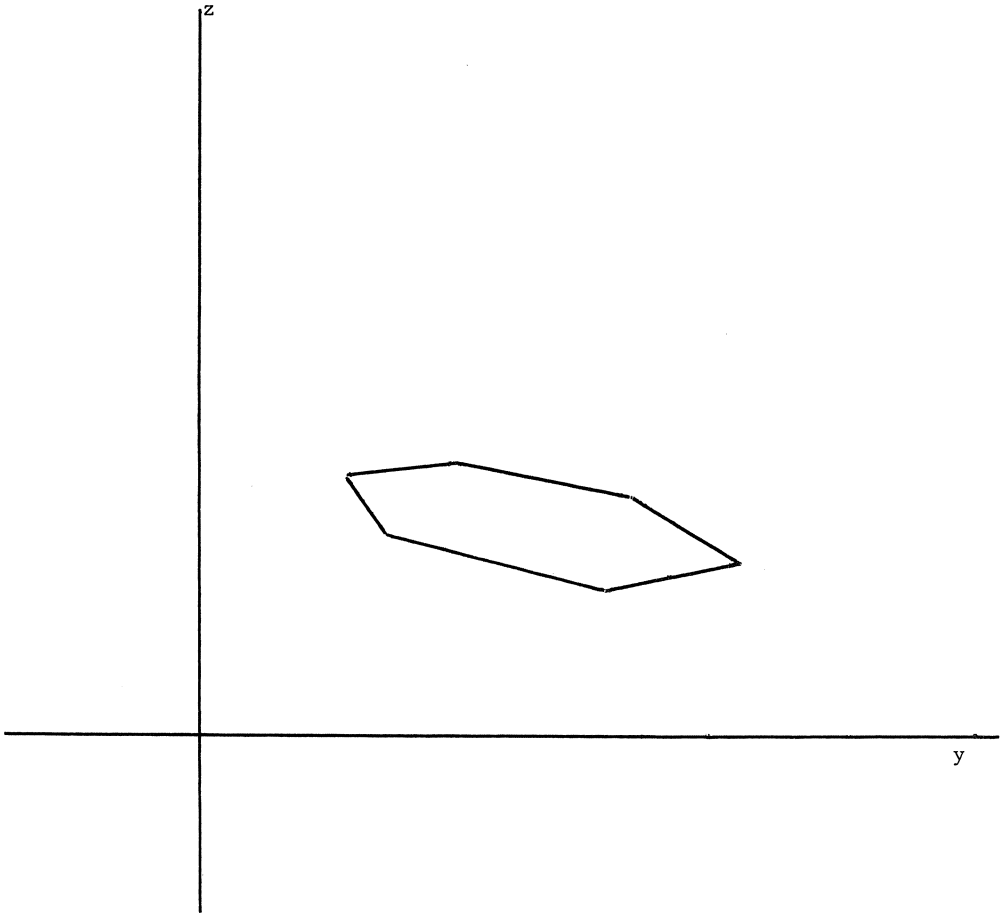


Bepaal op het tekenscherm het punt A_n met (y,z) = (b,c). Dit is het doorgangspunt van de loodlijn vanuit A op het scherm, het vluchtpunt van die loodlijn is C_p, zodat A_p ligt op de lijn door A_n en C_p. Bepaal op het scherm het punt B met (y,z) = (b+a,c) dit is het doorgangspunt van de lijn door A die met AA_n een hoek van 45° maakt, en evenwijdig aan het xy-vlak loopt. Het vluchtpunt D van die lijn ligt op de horizon, en wel op

afstand d van C_p. A_p wordt dus verkregen als snijpunt van BD en A_nC_p.

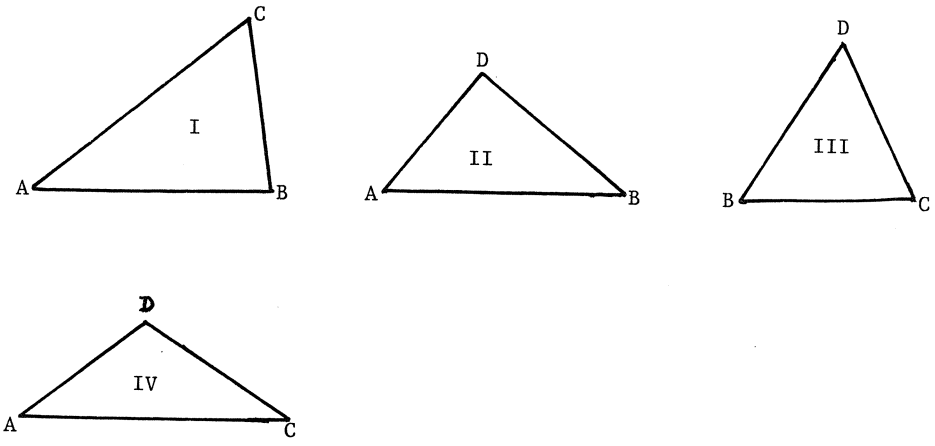
Vaak zal het voorkomen dat in het xy-vlak de ware gedaante van een figuur getekend is, welke men in perspectief wil afbeelden. Evenals bij de parallelprojectie kan het xy-vlak gewenteld worden om de y-as tot het in het vlak van tekening ligt. De projectie van een punt A in het xy-vlak kan dan als volgt geschieden. Zet op de loodlijn van C_p op de y-as de distantie d af. Dit punt C' is het om de horizon gewentelde centrum C. Laat vanuit A de loodlijn neer op de y-as, aldus het punt A_n verkrijgend. A_p is nu bepaald als snijpunt van AC' en A_nC_p. Merk op dat we in de tekening A zo gekozen hebben dat dit punt in werkelijkheid achter het scherm ligt. Deze methode is alleen goed te gebruiken als A niet te dicht bij de lijn $C'C_p$ ligt. Is dat het geval, dan is het beter om door A en C' twee evenwijdige lijnen, respectievelijk 1 en 1' te tekenen, bijvoorbeeld onder een hoek van 45° met de y-as. Zij D het snijpunt van 1 met de y-as, en D' dat van 1' met de horizon, dan ligt A_p op de lijn DD' (waarom?), en is dus bepaald als snijpunt van A_nC_p en DD'. We laten het aan de lezer over om de correctheid van deze constructies aan te tonen. In de volgende figuur is het perspectief getekend van een in het xy-vlak gelegen regelmatige zeshoek. Kunt u het hoofdpunt C_p en de distantie d bepalen? Hoe vindt u de ware grootte van de zeshoek?




z
y

§11. Netwerken.

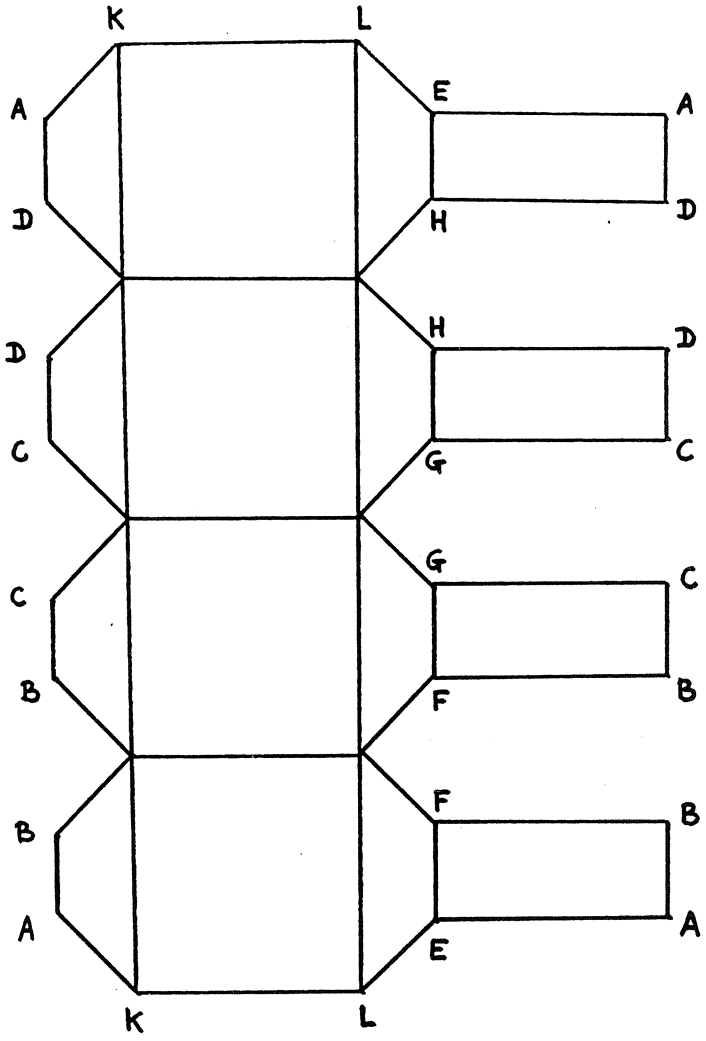
Een netwerk van een veelvlak bestaat uit de in het platte vlak getekende collectie van al zijn opbouwende vlakke veelhoeken op ware grootte, waarbij precies aangegeven is (bijvoorbeeld door de benaming van de hoekpunten) hoe zij in het veelvlak aan elkaar gehecht zijn. Zo bestaat het netwerk van een tetraëder met hoekpunten A,B,C,D, uit de volgende collectie driehoeken:

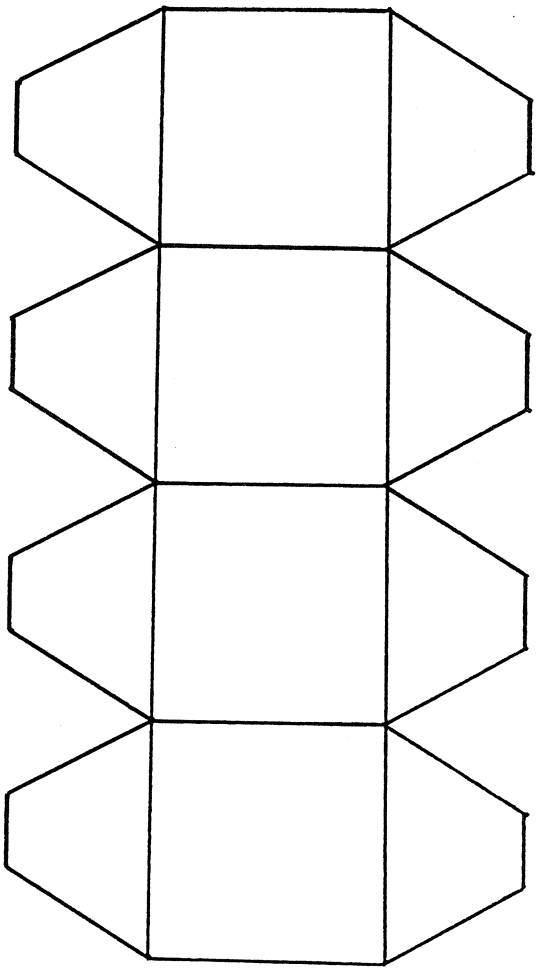


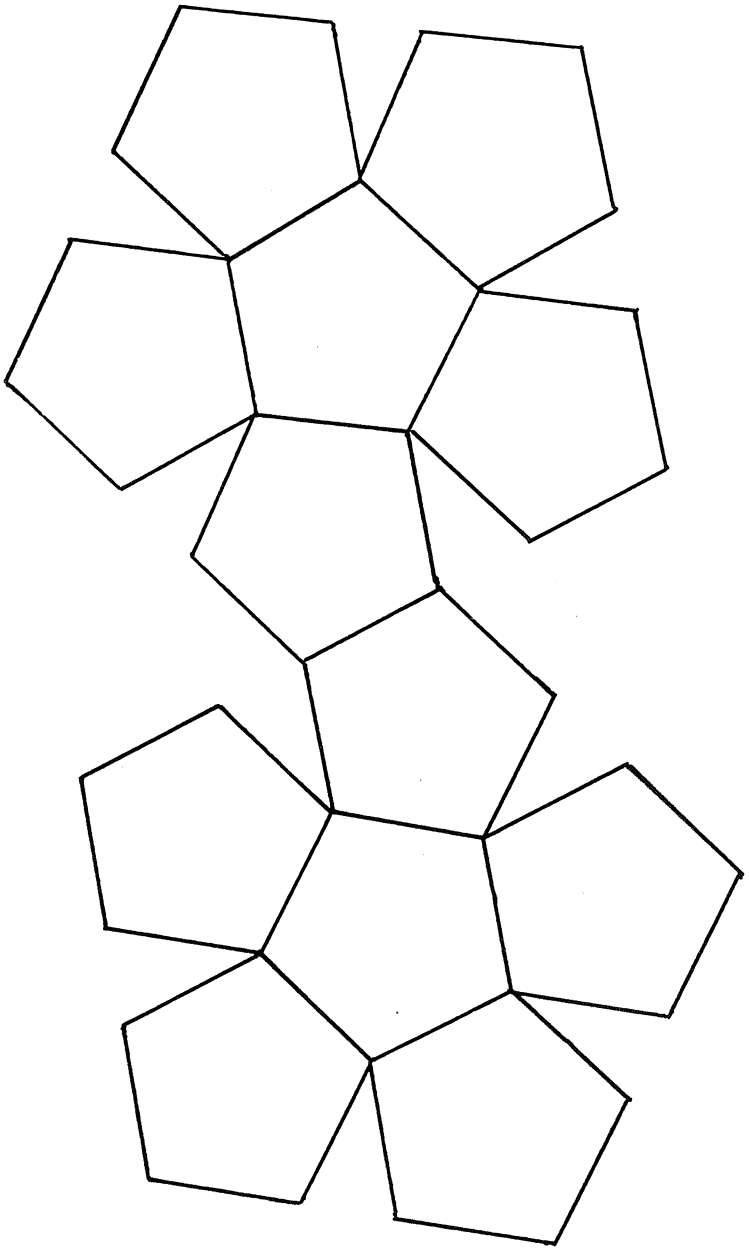
Bij de reconstructie van het tetraëder uit dit netwerk wordt de ribbe AC van I vastgemaakt aan de ribbe AC van IV, en wel zo dat de twee hoeken A op elkaar en ook de twee hoeken C op elkaar komen. Dit procédé wordt op alle ribben toegepast.

In het algemeen zal men er de voorkeur aan geven om zoveel mogelijk "in werkelijkheid" aan elkaar vast zittende veelhoeken ook in het netwerk reeds aan elkaar te tekenen, echter zonder daarbij tot overlappingen te komen!

Op de volgende bladzijden is een aantal netwerken opgenomen van meer of minder bekende veelvlakken. De geïnteresseerde lezer wordt geadviseerd ze over te nemen en uit te knippen en er daadwerkelijk een model van het veelvlak uit te construeren!


K
L
A
D
E
A
D
H
D
C
H
D
C
G
C
B
G
C
B
F
B
A
F
B
A
E
K
L



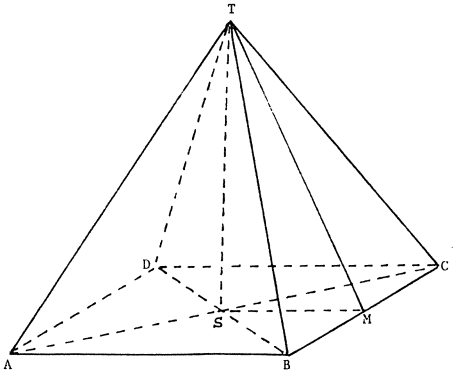


Netwerken zijn nuttig bij het construeren van veelvlakken met vooraf gegeven eigenschappen. Als voorbeelden noemen we een paar vraagstukken uit vroeger gebruikte stereometrie-leerboeken.

1. Gegeven is de zeszijdige piramide TABCDEF, waarvan het grondvlak ABCDEF een regelmatige zeshoek is met een ribbe van 3 cm, de hoogte van de piramide is 5 cm, en de loodlijn vanuit de top T treft het grondvlak in het snijpunt van AC en BD. Construeer het netwerk van deze piramide.

2. Teken het netwerk van een regelmatige vierzijdige piramide, waarvan de grondvlaksribbe 6 cm en de hoogte 4 cm is. Construeer ook de hoek die twee opstaande zijvlakken met elkaar maken. (N.B. met vierzijdige piramide wordt hier bedoeld dat er vier opstaande zijden zijn).

Het oplossen van dergelijke vraagstukken wordt voorafgegaan door een analyse van het probleem, bijvoorbeeld aan de hand van een ruimtelijke tekening. Noemen we in vraagstuk 2 de piramide TABCD, dan is ABCD blijkbaar een vierkant, de hoogtelijn uit T treft het grondvlak in het snijpunt S der diagonalen. Het vlak door T en S, loodrecht op BC treft BC in M, en SM en TM staan beide loodrecht op MC. Een rechthoekige driehoek met zijden 3 en 4 cm levert als hypothenusa de lengte van TM. Bij wenteling van de driehoek TBC komt TM in het grondvlak in het verlengde van SM te liggen.



Uit een goede analyse volgt onmiddellijk de manier waarop de constructie kan plaatsvinden. De daadwerkelijke uitvoering laten we aan de lezer over.

Andere vraagstukken waarbij een netwerk nuttige diensten kan verlenen

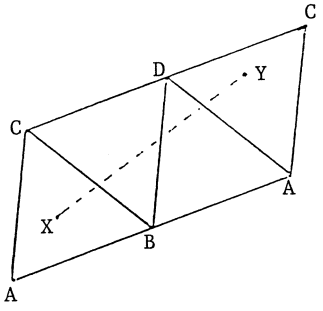
zijn die, waarbij gevraagd wordt naar de ware doorsnede van een vlak met een veelvlak. Als voorbeeld moge dienen:

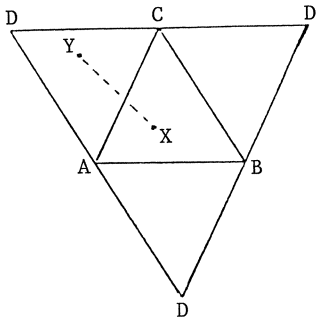
> Van een regelmatige vierzijdige piramide $\overset{T}{ABCD}$ zijn de opstaande ribben 8 cm en de ribben van het grondvlak 4 cm lang.
> Construeer de ware doorsnede van deze piramide met een vlak V dat TA, TB, TC snijdt in respectievelijk E, F, G zo dat TE = 5 cm, TF = 4cm en TG = 6 cm.

Ook hierbij kan een analyse vooraf niet gemist worden. We laten een en ander aan de lezer over!

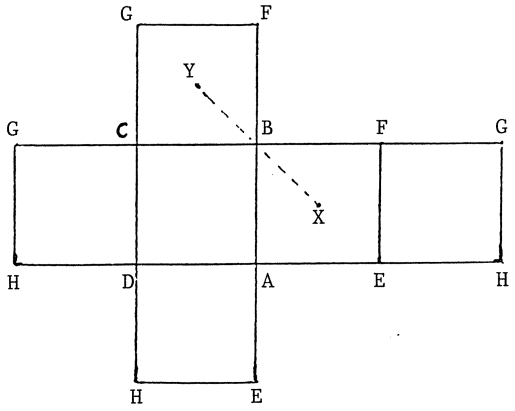
Als laatste elementaire toepassing van netwerken noemen we het bepalen van de kortste afstand, gemeten over het veelvlak, tussen twee punten van het veelvlak. Hierbij maakt men er gebruik van dat de kortste weg in een geschikte vorm van het netwerk verschijnt als een rechte lijn. Er doen zich hierbij verschillende problemen voor, zoals het nu volgende probeert te illustreren.

Als het mogelijk is in een netwerk van een veelvlak tussen twee punten X en Y een aaneengesloten recht lijnstuk XY te tekenen zodat geen hoekpunt op XY ligt, dan representeert XY een relatief kortste weg tussen X en Y over het veelvlak. Immers kleine afwijkingen worden in het netwerk gerepresenteerd door afwijkingen van de rechte lijn, die echter langer zijn. A priori kan men echter niet beweren dat XY een absoluut kortste weg is, want het is denkbaar dat in een andere configuratie van het netwerk een kortere weg mogelijk is voor te stellen door een recht lijnstuk.

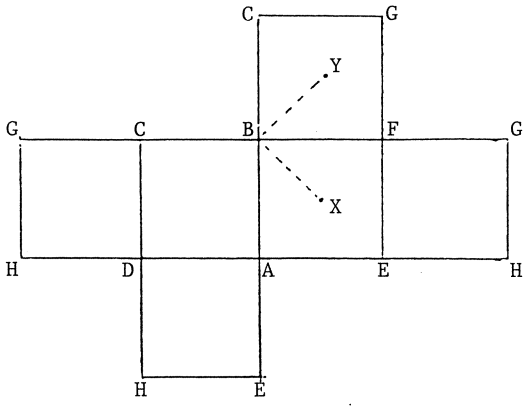




Als het lijnstuk XY in het netwerk door een hoekpunt gaat, dan kan bovenstaande redenering niet worden toegepast, omdat kleine afwijkingen ons buiten het netwerk kunnen brengen. Als illustratie bekijken we in de hiernavolgende figuur een netwerk van een kubus $\begin{matrix}EFGH\\ABCD\end{matrix}$. Voor de punten X en Y nemen we respectievelijk de middelpunten van ABFE en BCGF.



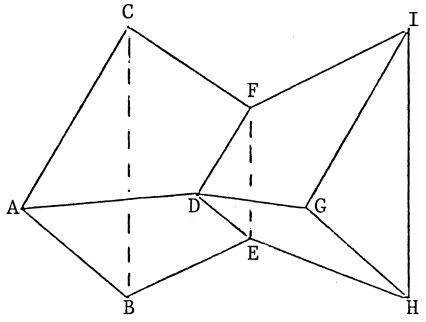
Tekenen we voor deze kubus echter een ander netwerk, waarin ABFE en BCGF aan elkaar grenzen volgens hun gemeenschappelijke ribbe BF, dan zien we dat er kleine afwijkingen (zo klein als men wil) van de bovenbeschreven weg (die nu een knik in B vertoont) zijn die korter zijn.



De redenering die hierboven opgezet is voor de kubus in het geval dat X en Y speciaal gekozen zijn, kan men veralgemenen tot:

> Een relatief kortste weg over een veelvlak kan nooit lopen via een hoekpunt B waarvoor geldt dat de som van de binnenhoeken in B van de B bevattende veelhoeken kleiner is dan 2π.

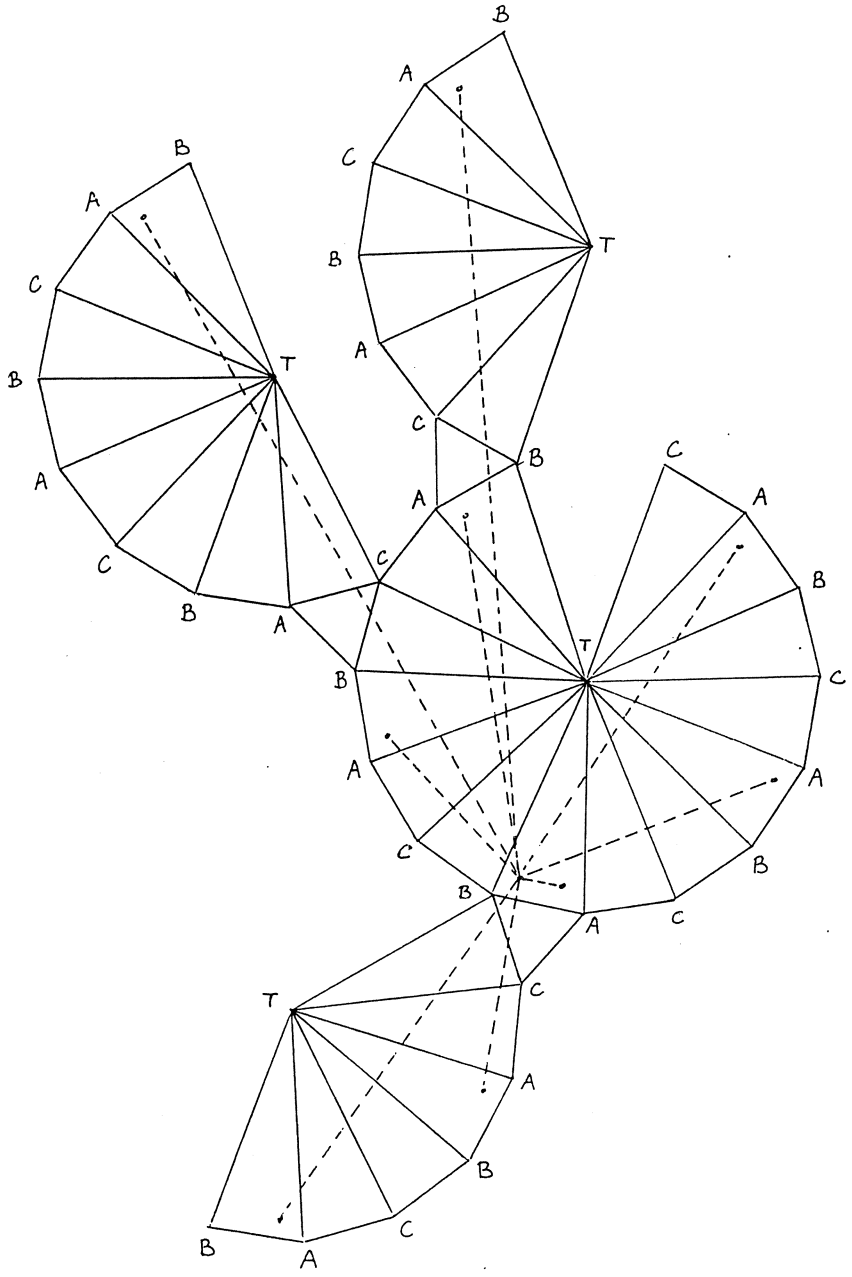
Heeft men echter een situatie zoals in onderstaande (ruimtelijke) tekening, waarin twee afgeknotte regelmatige tetraëders samengevoegd zijn,



(DEF is geen veelhoek van het veelvlak!)

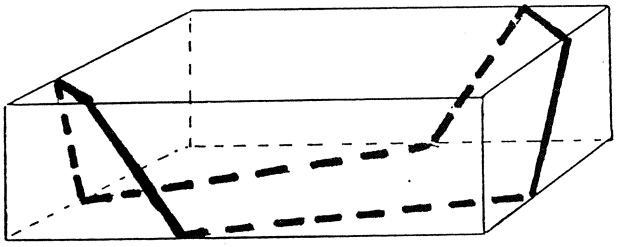
dan zijn de D bevattende veelhoeken ABED, ACFD, DGIF en DEHG en de som van hun binnenhoeken in D is 4x120° oftewel $\frac{8}{3}\pi$.
Voor een punt X op de diagonaal DI en een punt Y in ABED vormt het lijnstuk XD gevolgd door het lijnstuk DY een relatief kortste weg!
De verificatie van deze bewering wordt aan de lezer overgelaten (netwerk!).

Om een relatief kortste weg over een veelvlak als een rechte lijn in een netwerk te kunnen tekenen, moet men soms het netwerk "periodiek" uitbreiden: bepaalde veelhoeken kunnen dan vaker in de tekening voorkomen. Als illustratie bekijken we het probleem van de relatief kortste wegen tussen twee punten op een driezijdige piramide T_{ABC} met kleine tophoek. Een relatief kortste weg kan een veelhoek vaker aandoen. Op de volgende bladzijde is een aantal relatief kortste wegen tussen twee vaste punten op T_{ABC} gerepresenteerd. Probeer ze in werkelijkheid voor te stellen!


B
A
C
B
T
A
C
B
A
C
B
A
C
B
T
A
C
B
A
C
B
A
C
B
A
C
B
T
A
C
B
A
C
B
A
C
B
T
A
C
B
A
C
B
A
C
B
T
A
C
B
A

In het meetkunde nummer van Euclides (50e jaargang, 1974/75, no. 4/5) vinden we op blz. 166 de volgende opgave:

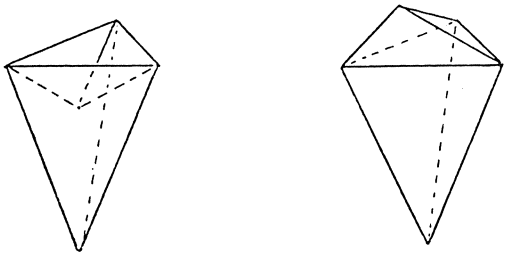
Om een doosje is een elastiekje gespannen als aangegeven in de figuur



Wat is de lengte van het gespannen elastiekje uitgedrukt in de lengten van de ribben?
Voor een uitwerking hiervan zie men ook Pythagoras, jaargang 15, No. 1 (oktober 1975).

§12. Over het verband tussen een veelvlak en zijn netwerk.

Bij het reconstrueren van een veelvlak uit een netwerk ervan, ontstaat de vraag of dan noodzakelijk weer het oorspronkelijke veelvlak ontstaat. Dat dit niet het geval hoeft te zijn, zien we bijvoorbeeld bij de volgende zesvlakken met identieke netwerken.



Hetzelfde fenomeen kan men aan een regelmatig 20-vlak (icosaëder) demonstreren. Daar kan men vijf driehoeken die in een punt samenkomen "naar binnen flippen" zonder het netwerk te veranderen. Voor interessantere "flip-flop veelvlakken" zie men Pythagoras, jaargang 20, No. 1 (1980), artikel van F. v.d. Blij.

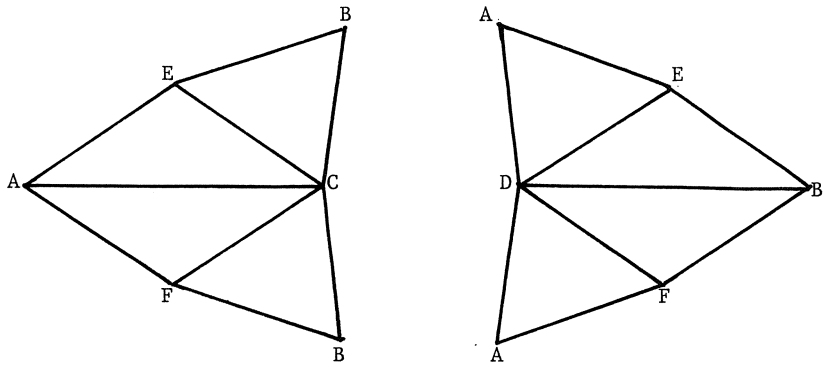
Wel bestaat ten aanzien van deze probleemstelling een resultaat van Cauchy:

Stelling: Een netwerk behoort bij ten hoogste één convex veelvlak.

Het bewijs van deze stelling is te lang om hier op te nemen. Geïnteresseerden kunnen het bijvoorbeeld vinden in [Steinitz & Rademacher].

Wanneer men echter uitgaat van een niet-convex veelvlak, dan kunnen er gekke dingen gebeuren. Zo is het bijvoorbeeld bekend dat er beweeglijke veelvlakken zijn. We bedoelen hiermee dat de samenstellende veelhoeken niet van vorm veranderen, dat ook de manier waarop de veelhoeken aan elkaar zitten niet verandert, en dat ondanks dat het veelvlak als geheel enigermate beweeglijk is.
Al lang bekend zijn beweeglijke octaëders met zelfdoorsnijding (Bricard, 1897). In de zin van de definitie zijn dit echter geen veelvlakken, maar in 1977 vond Connelly echte veelvlakken die beweeglijk zijn. Het bewijs van Connelly's vondst is te lang om hier op te nemen. Wie er meer van wil weten leze vooral zijn bijdrage in [Klarner]. Wel willen we benadrukken dat er dus blijkbaar netwerken zijn waarbij oneindig veel verschillende (niet-convexe) veelvlakken als realisering bestaan.

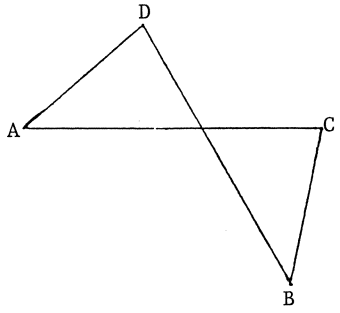
Als afsluiting van deze paragraaf geven we een beschrijving van een zichzelf doorsnijdend octaëder à la Bricard. Het netwerk van dit veelvlak ziet er als volgt uit:



Hierin zijn alle driehoeken gelijkbenig, met een en dezelfde lengte voor elk paar gelijke ribben. Bovendien is

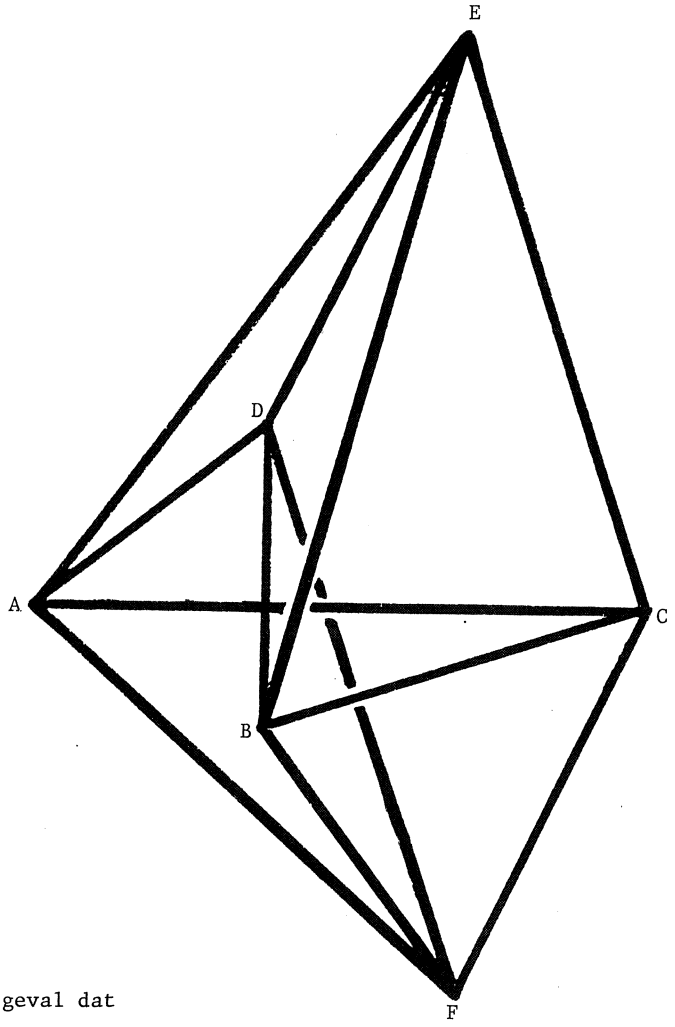
$$|AC| = |DB| \text{ en } |CB| = |AD| \text{ en } |AC| > |BC|.$$

Bij de realisering leggen we allereerst de punten A, B, C, D in één vlak V van de ruimte volgens de volgende figuur, waarbij het essentieel is dat AC en BD elkaar snijden.



Het middelloodvlak W van AB is dan tevens middelloodvlak van DC. Laten we nu de gelijkbenige driehoek ADE om AD wentelen tot E in W valt, dan is in die positie $|EA| = |ED| = |EC| = |EB|$. Vervolgens kiezen we voor F het spiegelbeeld van E t.o.v. het vlak V. Deze constructie "realiseert" het netwerk als veelvlak met zelfdoorsnijding. Omdat deze constructie mogelijk is voor variabele afstand tussen A en B, is het "veelvlak" beweeglijk!

Om een inzicht te krijgen in de ruimtelijke situatie kan men van het zichzelf doorsnijdend veelvlak een geraamte maken, uitsluitend bestaande uit de ribben. Men krijgt zo een stelsel van stangen die aan hun uiteinden draaibaar aan elkaar bevestigd zijn. Hiermee kan dan de beweeglijkheid gedemonstreerd worden, mits de stangen elkaar niet snijden.
Helaas is het bij het bovenstaande veelvlak zo dat de stangen AC en BD elkaar wel snijden, maar in de praktijk kan men dit verdoezelen door bijvoorbeeld AC een weinig gekromd te maken.
Op de volgende bladzijde is een tekening opgenomen van een stangenstelsel voor dit veelvlak.



Voor het geval dat

$|AC| = |BD| = 20$ $\qquad$ $|AD| = |BC| = 10$

$|EA| = |EB| = |EC| = |ED| = |FA| = |FB| = |FC| = |FD| = 15$

kan men als één van de mogelijke posities in $\mathbb{R}^3$ bijvoorbeeld nemen

A = (0, 0, 0) $\qquad$ D = (0,10, 0)

B = (12,-6, 0) $\qquad$ E = (10, 5, 10)

C = (20, 0, 0) $\qquad$ F = (10, 5,-10)

Stangenstelsels vormen een interessant studie-object. Behalve in wiskunde en kunst spelen starre (onbeweeglijke) stangenstelsels ook een rol in de architectuur van bijvoorbeeld koepels en bruggen (als stalen geraamten). De geïntersseerde lezer verwijzen we naar [Kenner, Geodesic Math. and how to use it] en [Pugh, Introduction to tensegrity].

LITERATUUR

A.D. ALEXANDROW, Konvexe Polyeder. BERLIN 1958.

N.L. BIGGS, E.K. LLOYD, R.J. WILSON, Graph Theory 1736-1936. OXFORD etc. 1977.

*H.S.M. COXETER, Regular Polytopes. NEW YORK 1963.

H.B. GRIFFITHS, Surfaces. CAMBRIDGE etc. 1976.

B. GRUNBAUM, Convex Polytopes. LONDON etc. 1967.

H. KENNER, Geodesic math and how to use it. BERKELEY etc. 1976.

*D.A. KLARNER, The mathematical Gardner. BELMONT 1981.

*I. LAKATOS, Proofs and refutations. CAMBRIDGE 1976.

*P. MOLENBROEK, Leerboek der Stereometrie. GRONINGEN 1923.

A. PUGH, An introduction to tensegrity. BERKELEY etc. 1976.

A. PUGH, Polyhedra, a visual approach. BERKELEY etc. 1976.

E. STEINITZ und H. RADEMACHER, Vorlesungen über die Theorie der Polyeder. BERLIN 1934.

L. FEJES TOTH, Lagerungen in der Ebene, auf der Kugel und im Raum. BERLIN etc. 1953.

L. FEYES TOTH, Regular Figures. OXFORD etc. 1964.

*H.J. VAN VEEN, Beknopt leerboek der Beschrijvende Meetkunde, GRONINGEN 1931.

*M.J. WENNIGER, Spherical Models. CAMBRIDGE etc. 1979.

B.L. VAN DER WAERDEN, Ontwakende Wetenschap. GRONINGEN 1950.

artikelen

F. VAN DER BLIJ, Met meer driehoeken bouwen. PYTHAGORAS, 20 nr. 1 (1980), 9-12.

J.M. GOETHALS, J.J. SEIDEL, The football, NIEUW ARCHIEF VOOR WISKUNDE (3), XXIX (1981), 50-58.

H.M. MULDER, De bol van Montreal. PYTHAGORAS 12 nr. 4 (1972/1973), 33-90.

Lesmateriaal

*Lessen in Ruimtemeetkunde 1, OW & OC, UTRECHT 1982.

*: Aanbevolen.

EEN WISKUNDIG MODEL VAN EEN ONZEKERE BESLISSINGSSITUATIE

W. SCHAAFSMA

Bij het wiskundig modelleren gaat het er in vele gevallen om dat de problematiek van een bepaalde klant vanuit de wiskunde wordt meebeleefd.

Dat meebeleven kan door allerlei oorzaken mislukken. Bijvoorbeeld doordat de problematiek in wezen ongeschikt is voor wiskundige behandeling. Het is dan misschien maar goed dat de samenwerking mislukt.

Het kan echter ook gebeuren dat de samenwerking mislukt terwijl de problematiek in diepste wezen wiskundig interessant en maatschappelijk relevant is. De oorzaak zal dan zijn dat de klant een verkeerde wiskundige heeft ingeschakeld. Dat gevaar is levensgroot aanwezig want velen van ons beschikken slechts over beperkte begaafdheden en hebben geen gevoel voor de wijze waarop de klant tegemoet moet worden getreden. Ook ik heb in het verleden tal van fouten gemaakt. Mijn leermeester Professor Smid heeft gebrekkige oplossingen van mij verbeterd door beter mathematische formuleringen voor te stellen. Verder heb ik mij soms, op uitnodiging van de klant, laten verleiden tot het geven van oplossingen voor detailproblemen waarbij later alles opnieuw moest omdat het beter kon. Deze ervaringen betekenen voor mij een kostbare schat. Dat men fouten mag maken, dat ze worden ontdekt en dat men ze mag verbeteren, daar leert men van en dat stemt tot dankbaarheid.

Heel belangrijk bij het wiskundig modelleren is dat *de problematiek van de klant in het middelpunt van de belangstelling* wordt geplaatst en niet de onuitgesproken wens van de wiskundige om zijn specialisatie te laten functioneren of om zelf een zeer geleerde

of buitengewoon spitsvondige indruk te maken.

Ook het menselijk aspect mag niet uit het oog verloren. In het boek "Modern Kwaliteitsbeleid" van een naamgenoot van mij en de heer Willemse, beide werkzaam bij Philips, wordt uitvoerig aandacht geschonken aan dit menselijk aspect. De kwaliteitskontroleur moet er bijvoorbeeld voor waken dat hem niet juist de goede stukken vanaf de werkvloer voor kontrole worden aangeboden. Verder moet alles zo eenvoudig worden gemaakt dat de mens of machine die de feitelijke kontrole moet uitvoeren geen enkel probleem heeft met het uitvoeren van de instructie.

Na deze inleiding mag U van mij verwachten dat ik een bepaalde concrete problematiek centraal stel en dat ik alles zodanig inricht dat U weinig moeite zult hebben met het meebeleven ervan. Alvorens U de beloofde problematiek voor te leggen wil ik echter enige algemene beschouwingen wijden aan het gebied waartoe deze problematiek behoort, namelijk dat van de wiskundige statistiek.

WISKUNDIGE STATISTIEK

Kennisvermeerdering kan op allerlei manieren geschieden. Bijvoorbeeld door overdracht, door redeneren of door waarnemen. In de wiskundige statistiek beredeneren we hoe we onze kennis kunnen vermeerderen op grond van waarnemingen. Het gaat ongeveer als geschetst in figuur 1.

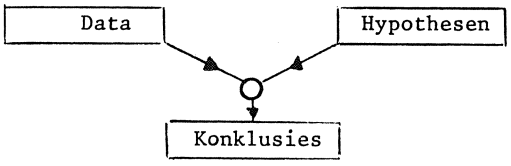


Fig. 1

Als konklusies komen in aanmerking: puntschattingen, puntvoorspellingen, beweringen omtrent de waarheid van bepaalde hypothesen, intervalschattingen, toewijzingen, beslissingen, enz. De konklusie kan natuurlijk ook zijn dat de data moet worden uitgebreid of dat de hypothesen iets anders moeten worden gekozen.

Figuur 1 laat op suggestieve wijze zien dat de konklusies kunnen afhangen van de data die men in de beschouwingen betrekt en van de

hypothesen die men aan de beschouwingen ten grondslag legt. De ideologie van de wiskundige statistiek is dat het cirkeltje in figuur 1 op zo verantwoord mogelijke wijze moet worden ingevuld. Dat voert ertoe dat koste wat het kost *onzekerheden* tot uitdrukking moeten worden gebracht. Geen enkele goede statisticus zal vrede hebben met het geven van een puntschatting of een puntvoorspelling aléén. Hij zal er minstens een of andere standaarddeviatie bij willen vermelden. Dit thema zult U straks terug zien komen.

Hoe gaat dat in zijn werk, dat tot uitdrukking brengen van onzekerheden? In de wiskundige statistiek let men vooral op de data-afhankelijkheid. Men vat de ingebrachte data x op als de uitkomst van een toevallige variabele X die waarden aan kan nemen in de z.g. steekproefruimte $\mathcal{X}$. De gevonden data x correspondeert dus met één punt in die ruimte $\mathcal{X}$. Er worden allerlei hypothesen geformuleerd ten aanzien van de kansverdeling van X. Formeel gesproken komt het er meestal op neer dat van die kansverdeling wordt vastgelegd dat het een element is uit een precies gedefinieerde klasse van kansverdelingen $\{P_\theta;\ \theta \in \Theta\}$. De uitkomst x van het experiment leert ons dan van alles ten aanzien van de werkelijke edoch ontbrekende waarde van de parameter θ. Misschien is het soms wat overdreven om te spreken van de werkelijke waarde van θ want het model $\{P_\theta;\ \theta \in \Theta\}$ bevat vaak allerlei hypothetische elementen en simplificaties.

Laten wij eens zien hoe het toegaat in de praktijk.

ONZE PROBLEMATIEK

Een ongerief waarvan sommige kinderen last hebben is de neiging tot bedplassen. Ze kunnen worden behandeld met de z.g. wekker-therapie. De werking hiervan is duidelijk. Er wordt op e.o.a. manier voor gezorgd dat de wekker begint te rinkelen zodra het bed nat wordt.
Allerlei vragen doemen op. Helpt de behandeling? Hebben bepaalde typen bedplassers er meer baat bij dan andere typen? Gegeven een kind met zijn of haar specifieke kenmerken, moet dat kind wel of niet worden behandeld? In zo'n situatie komen we geen stap verder zonder data. De goede statisticus zal zorg besteden aan elk van de drie in figuur 1 vermelde ingrediënten. Iets van *eerbied voor de data* hoop ik U bij te brengen door het volledige oorspronkelijke waarnemingsmateriaal aan U voor te leggen als

uitgangspunt voor onze discussie. Het materiaal is afkomstig van de psychologe Dr. Sylvia Dische. Ik heb het aan de literatuur ontleend. Samen met een afstudeerder, de heer Gerard Wortelboer, heb ik er menig uur aan besteed zodat ik het bijna als iets van onszelf ben gaan beschouwen. Maar dat is het dus niet. Alle rechten komen toe aan Dr. Dische.
Wat staat er allemaal vermeld in de U verstrekte tabel 1? Van elk van 113 kinderen vindt U vermeld: de uitkomst van een zevental binaire variabelen $V_1,\ldots,V_7$ en de uitkomst van een variabele C met drie verschillende mogelijke waarden n.l. 1, 2 en 3. De precieze betekenis is als volgt

V_1	gezinsmoeilijkheden (1=ja, 0=nee)
V_2	infectie aan de urinewegen (1=ja, 0=nee)
V_3	last overdag (1=ja, 0=nee)
V_4	ook poepen (1=ja, 0=nee)
V_5	leeftijd (1=8 jaar of jonger, 0=9 jaar of ouder)
V_6	w.c. binnenshuis (1=ja, 0=nee)
V_7	kamer delen met iemand anders (1=ja, 0=nee)

terwijl C het effect van de wekker-therapie beschrijft waarbij de betekenis is als volgt

C=1	geen verbetering
C=2	eerst verbetering daarna terugval
C=3	langdurige genezing

Het is duidelijk dat we tabel 1 niet goed kunnen gebruiken om na te gaan of de behandeling werkelijk helpt. De antwoorden kunnen immers niet zonder meer worden vertrouwd. Sommige geënquêteerden zouden de neiging kunnen hebben om de zaken rooskleuriger voor te stellen dan ze in werkelijkheid zijn, enz..Tabel 1 is wel geschikt om na te gaan of bepaalde typen bedplassers meer baat hebben bij de behandeling dan andere typen. Ook zou de praktische vraag kunnen worden gesteld of een kind, bijvoorbeeld met specifieke kenmerken 0000111 al of niet moet worden behandeld. Deze vragen zullen aanstonds nader worden onderzocht.
Eerst doen we echter nog iets verstandigs.

V_1	V_2	V_3	V_4	V_5	V_6	V_7	C
0	0	0	0	0	1	0	3
1	0	0	0	0	1	0	3
0	0	0	0	0	1	0	3
0	0	0	0	0	0	0	3
0	0	0	0	0	1	0	3
0	0	0	0	0	1	0	2
1	0	0	0	0	1	0	2
0	0	0	0	0	1	1	2
1	0	0	0	0	1	1	1
1	0	0	0	0	1	0	2
1	0	0	0	0	1	0	2
0	0	0	0	0	1	0	3
1	0	0	0	0	0	0	1
1	0	0	0	1	1	0	2
1	0	0	0	0	1	0	3
1	0	0	0	0	1	0	3
0	0	0	0	1	1	0	2
0	0	0	0	1	1	0	3
0	0	0	0	1	1	0	3
1	0	0	0	0	0	0	1
0	0	0	0	0	1	0	2
0	0	0	0	1	1	0	3
1	0	0	0	0	1	0	1
1	0	0	0	0	1	0	3
1	1	0	0	0	1	1	2
0	0	0	0	0	0	0	3
1	0	0	0	1	0	0	3
1	0	0	0	0	1	0	2
1	1	1	1	0	1	1	2
0	0	0	0	1	1	0	2
0	0	0	0	1	1	0	3
1	0	0	0	1	1	0	2
0	0	0	0	1	1	0	3
1	0	0	0	1	1	0	2
0	0	0	0	0	1	1	3
0	0	0	0	1	1	0	2
1	0	0	0	0	1	0	2
1	0	0	0	1	1	0	3

V_1	V_2	V_3	V_4	V_5	V_6	V_7	C
1	0	0	0	0	1	0	3
0	0	0	0	0	1	0	3
1	0	0	0	0	1	1	3
0	0	0	0	0	1	0	2
0	0	0	0	0	1	0	3
1	0	0	0	1	1	1	1
0	0	0	0	1	1	0	3
0	0	0	0	0	1	0	2
0	0	0	0	1	1	0	1
1	0	0	0	1	1	1	1
0	0	1	0	1	1	0	3
0	0	0	0	1	1	1	3
0	0	0	0	1	1	0	3
1	0	0	0	0	1	0	1
1	0	0	0	1	1	1	3
0	0	0	0	1	1	0	3
0	0	0	0	1	1	0	3
0	0	0	0	1	1	0	2
0	0	0	0	1	1	1	3
0	0	0	0	1	0	0	2
1	0	0	0	1	1	0	3
1	0	0	0	1	1	1	3
1	0	0	0	1	1	0	2
0	0	0	0	1	1	0	3
1	0	1	0	1	1	1	2
0	0	1	0	1	0	0	3
0	0	0	0	1	1	0	2
1	1	0	0	1	1	0	2
0	0	0	0	1	1	0	3
1	0	0	0	1	1	0	2
1	0	0	0	1	1	0	2
1	0	1	0	1	1	1	3
0	0	0	0	1	0	0	3
0	0	0	0	1	1	0	1
1	0	0	0	1	1	1	1
0	0	0	0	1	1	0	2

V_1	V_2	V_3	V_4	V_5	V_6	V_7	C
1	0	0	0	1	1	0	3
0	0	1	0	1	1	0	3
1	0	0	0	1	1	0	2
0	0	0	0	0	1	0	1
0	0	0	0	0	1	0	1
1	0	0	0	0	1	1	3
0	0	0	0	0	1	0	2
1	0	0	0	1	1	1	1
1	0	0	0	1	0	1	1
0	0	0	0	0	1	0	3
0	0	0	0	0	1	0	3
1	0	0	0	1	1	0	2
0	0	0	0	1	1	0	3
1	0	0	0	1	1	0	2
0	0	0	0	0	1	0	1
0	0	0	0	1	1	0	3
0	0	0	0	1	1	0	3
1	0	0	0	1	1	0	3
0	0	0	0	1	1	1	2
0	0	0	0	1	1	1	3
1	0	0	0	1	1	0	3
0	0	0	0	1	1	0	3
0	0	0	0	1	1	0	3
1	0	0	1	1	1	0	1
1	0	1	0	1	0	1	3
1	0	0	0	1	1	0	2
0	0	0	0	1	1	0	2
1	0	0	0	1	0	0	1
1	0	0	0	1	1	0	2

EEN SAMENVATTING VAN DE DATA

Een min of meer verrassend verschijnsel is als volgt. Er zijn 2^7 uitkomsten mogelijk voor $V_1,\ldots, V_7$. Als U een willekeurig rijtje van zeven nullen en enen opschrijft , ik nodig U uit om dat te doen, dan heeft U een kans van 1/128 dat Uw rijtje klopt met de uitkomst van $V_1,\ldots, V_7$ van een willekeurig van te voren gekozen kind. De verwachtingswaarde van het aantal kinderen met Uw rijtje als vector van scores is dus 113/128. Dat is kleiner dan één, U kunt dus bot vangen. Hoe vaak wordt in werkelijkheid bot gevangen? Of, iets mooier geformuleerd, voor hoeveel personen onder U zou de referentie steekproef leeg zijn? Het antwoord luidt: voor ongeveer vijf van de zes personen. Van de 128 theoretisch mogelijke uitkomsten van $V_1,\ldots,V_7$ treden in de werkelijkheid slechts 21 op! Er zijn 107 uitkomsten die nooit optreden, 7 die één keer optreden, 6 die twee keer optreden, 2 die drie keer optreden en één die 5, 8, 13, 17, 18 of 27 keer, respectievelijk, optreedt. Dat is dus een wel heel scheve verdeling. U kunt dat aflezen uit tabel 2. De betekenis

V_1	V_2	V_3	V_4	V_5	V_6	V_7	C=1	C=2	C=3	Tot	V_1	V_2	V_3	V_4	V_5	V_6	V_7	C=1	C=2	C=3	Tot
0	0	0	0	1	1	0	2	7	18	27	1	0	0	0	0	0	0	2	0	0	2
1	0	0	0	1	1	0	0	12	6	18	1	0	0	0	1	0	0	1	0	1	2
0	0	0	0	0	1	0	2	5	10	17	1	0	1	0	1	1	1	0	1	1	2
1	0	0	0	0	1	0	3	5	5	13	0	0	1	0	1	0	0	0	0	1	1
1	0	0	0	1	1	1	5	1	2	8	1	0	0	0	1	0	1	1	0	0	1
0	0	0	0	1	1	1	0	1	4	5	1	0	0	1	1	1	0	1	0	0	1
0	0	0	0	1	0	0	0	1	2	3	1	0	1	0	1	0	1	0	0	1	1
1	0	0	0	0	1	1	1	0	2	3	1	1	0	0	0	1	1	0	1	0	1
0	0	0	0	0	0	0	0	0	2	2	1	1	0	0	1	1	0	0	1	0	1
0	0	0	0	0	1	1	0	1	1	2	1	1	1	1	0	1	1	0	1	0	1
0	0	1	0	1	1	0	0	0	2	2	54	3	7	2	73	101	24	18	37	58	113

Tabel 2

van de daarin vermelde symbolen behoeft naar ik hoop geen nadere uitleg. U kunt heel gemakkelijk nagaan of het door U gekozen rijtje voorkomt. We zouden de nulhypothese kunnen toetsen dat de succeskans inderdaad gelijk is aan 21/128. Het zou best eens significant kunnen uitvallen want mijn instructie was weliswaar logisch volkomen duidelijk, n.l. schrijf een willekeurig rijtje van zeven nullen en enen op, maar sommigen onder U zullen dat niet goed hebben begrepen en een willekeurig rijtje hebben gepakt uit tabel 1. In dat geval is de kans (113-7)/113 dat er minstens nog één ander kind is met precies dezelfde uitkomst! Wil het mis gaan dan moet U n.l. beslist één van de zeven laatste kinderen uit tabel 2 hebben getrokken.

Zo ziet U maar weer hoe sterk kansen kunnen afhangen van wat in werkelijkheid wordt uitgevoerd. Het is net zoals mijn naamgenoot en de heer Willemse beweren. Het is niet voldoende dat instructies precies worden geformuleerd. Men moet er ook nog op toezien dat zij precies worden uitgevoerd.

HET WAARSCHIJNLIJKHEIDSTHEORETISCH MODEL

Wij willen de data van Dr. Dische gebruiken in andere situaties dan die welke horen bij de 113 onderzochte kinderen. Dat vereist een extrapolatie en daarin schuilen allerlei onzekerheden. In de wiskundige statistiek doen wij ons best om in ieder geval een gedeelte van die onzekerheden tot uitdrukking te brengen. We proberen dat te doen door het toeval in de beschouwingen te betrekken. In ons geval is het duidelijk dat wij zullen moeten postuleren dat *de 113 onderzochte kinderen een soort toevallige steekproef* vormen uit één of andere populatie van "bedplassende kinderen met soortgelijke kenmerken als van de door Dr. Dische onderzochte kinderen". We noemen dat de referentie populatie. Het is duidelijk dat de definitie van de referentie populatie veel te wensen overlaat. Hetzelfde geldt voor de definities van de variabelen $V_1,\ldots, V_7$ en C. Wanneer zeggen we precies dat er gezinsproblemen zijn, d.w.z. dat V_1 de waarde één aanneemt? Het speculeren daarover vertrouw ik graag aan U toe. Het is goed om op te merken dat de ene vraagstelling scherper definities eist dan de andere. Aanstonds zullen wij aantonen dat de kans op genezing, d.w.z. C=3, groter is voor kinderen uit gezinnen zonder gezinsproblemen dan voor kinderen uit

gezinnen met problemen. Deze bewering blijft natuurlijk waar als allerlei definities een beetje worden veranderd. De veel preciezer bewering dat de eerste kans ongeveer twee keer zo groot is als de tweede is natuurlijk veel gevoeliger voor verstoringen.

ENKELE HYPOTHESEN GETOETST

De gedachte dat de 113 onderzochte kinderen een steekproef vormen uit een referentie populatie vormt een prachtig kader voor onze beschouwingen. We zijn natuurlijk geïnteresseerd in de samenhang tussen de variabelen $V_1,\ldots, V_7$ en de variabele C. Iets anders uitgedrukt: wat is de predictieve betekenis van $V_1,\ldots, V_7$ voor C. Nu is deze problematiek nogal ingewikkeld en daarom redeneer ik als volgt. Beschouwen wij de kolomsommen in tabel 2 dan zien wij dat de variabelen V_2, V_3, V_4, V_6 en V_7 nauwelijks hun naam variabele verdienen want ze zijn vrijwel konstant. Waren zij echter konstant geweest dan zouden zij van nul en generlei waarde zijn geweest voor het effect van de therapie. Laten wij dus eerst maar eens de invloed van V_1 en V_5 bestuderen. In tabel 3 vindt U de bijbehorende gegevens. Zij zijn natuurlijk ontleend aan tabel 2.

	C=1	C=2	C=3	Totaal
$V_1=0,\ V_5=0$	2	6	13	21
$V_1=0,\ V_5=1$	2	9	27	38
$V_1=1,\ V_5=0$	6	7	7	20
$V_1=1,\ V_5=1$	8	15	11	34
Totaal	18	37	58	113

Tabel 3

Allerlei interessante toetsingsproblemen dringen zich nu aan ons op. Allereerst bestuderen wij de vraag of V_1 van invloed is op het resultaat van de wekker-therapie. Om dat te onderzoeken vormen wij allereerst tabel 4 uit

	C=1	C=2	C=3	Totaal
$V_1=0$	$n_{11}=4$	$n_{12}=15$	$n_{13}=40$	59
$V_1=1$	$n_{21}=14$	$n_{22}=22$	$n_{23}=18$	54
Totaal	18	37	58	113

Tabel 4

tabel 3 door variabele 5 buiten beschouwing te laten. De aantallen n_{bc} (b=1,2;c=1,2,3) in tabel 4 vormen de uitkomst van een multinomiale verdeling. Dat is een onmiddellijk gevolg van de aanname dat de 113 kinderen een lukrake steekproef vormen uit een , groot verondersteld, referentie populatie. De bij die multinomiale verdeling horende kansen p_{bc} (b=1,2;c=1,2,3) hebben de precieze interpretatie dat zij relatieve frekwenties voorstellen in de referentie populatie; p_{bc} is het aantal elementen in de referentie populatie met V_1=b en C=c, gedeeld door het totale aantal kinderen in de referentie populatie. Die kansen p_{bc} zijn natuurlijk onbekend hoewel tabel 4 allerlei informatie bevat wat betreft hun waarde.

Is V_1 van invloed op het resultaat van de wekker-therapie? Dat is de vraag die wij onderzoeken. De mathematische vertaling hiervan luidt dat wij de nulhypothese

$$H : p_{bc} = (p_{b1}+p_{b2}+p_{b3})\ (p_{1c}+p_{2c}) \quad (b=1,2;c=1,2,3)$$

van onafhankelijkheid van V_1 en C moeten toetsen tegen een alternatieve hypothese A die inhoudt dat een zekere afhankelijkheid zal bestaan. Dat men heel diep kan nadenken over de precieze mathematische formulering van die alternatieve hypothese zal U duidelijk zijn. Dat men vervolgens heel diep kan nadenken over de wijze waarop nulhypothese H moet worden getoetst tegen alternatief A zal U ook niet verbazen. In Groningen hebben wij getracht, wat deze onderwerpen betreft, een zekere reputatie op te bouwen. Enfin, ik ben niet van plan om U met de bijbehorende mathematiek te gaan vermoeien. Het is n.l. volkomen duidelijk op grond van tabel 4 dat de nulhypothese H moet worden verworpen. Het is duidelijk dat de resultaten bij V_1=0 gunstiger zijn dan bij V_1=1.

Men kan zich eenzelfde vraag stellen t.a.v. variabele V_5. In dat geval vindt men dat de nulhypothese van onafhankelijkheid tussen V_5 en C niet hoeft te worden verworpen.

Tenslotte kan men tabel 3 weer in ogenschouw nemen en zich afvragen of er invloed bestaat van V_5 op C, gegeven dat men V_1 al kent. Omdat de eerste twee rijen in tabel 3 ongeveer verhoudingsgewijs hetzelfde zijn, evenals de laatste twee rijen, zal ook hier de bijbehorende nulhypothese niet worden verworpen .

De heer Gerard Wortelboer en ik hebben verder nog wat gekeken naar de eventuele invloed van V_2, V_3, V_4, V_6 en V_7 op C, bij konstant houden

van V_1. Wij hebben niets interessants gevonden.

Onze konklusie luidt dus dat V_1 een belangrijke invloed héeft op het resultaat van de wekker-therapie en dat alle andere variabelen vrijwel waardeloos zijn.

BETROUWBAARHEIDSINTERVALLEN

Beschouwen wij de kans op C=3, d.w.z. langdurige verbetering, dan is deze kans p=P(C=3) dus gedefinieerd als de bijbehorende fractie van de referentie populatie. In de steekproef van 113 kinderen vond Dr. Dische 58 kinderen met C=3. De kans p kan dus worden geschat met behulp van de relatieve frequentie 58/113 in de refentie steekproef. De bijbehorende standaard deviatie van de gebruikte schatter kan worden geschat als

$$\{ 113^{-1} (58/113)(1-58/113) \}^{\frac{1}{2}} = .047$$

en $.513 \pm 1.960 \times .047$ levert dus bij benadering een 95% betrouwbaarheidsinterval voor p. Het wordt zoiets als [.42,.61].

We hebben net gezien dat V_1 invloed heeft op het resultaat van de behandeling. Het ligt dus voor de hand om ook te kijken naar de referentie populatie die ontstaat als je één van de restricties $V_1=0$ of $V_1=1$ invoert. De kans p=P(C=3) verandert dan in de voorwaardelijke kans $p_0=P(C=3|V_1=0)$ of de kans $p_1=P(C=3|V_1=1)$. Dat zijn natuurlijk weer de relatieve frequenties in de bijbehorende referentie populaties. Een bekende relatie is dan nog $p=p_0P(V_1=0) + p_1P(V_1=1)$. Hoe het zij, we zouden graag iets meer willen weten over de precieze waarden van de onbekende parameters p_0 en p_1. Daarvoor kunnen we in tabel 4 terecht. De 59 kinderen met $V_1=0$ laten zich opvatten als een steekproef uit de bijbehorende referentie populatie. Het gevonden aantal successen, n.l. 40, levert een schatting voor p_0 van ongeveer 2/3. Het bijbehorende betrouwbaarheidsinterval is [.54,.79]. De 54 kinderen met $V_1=1$ leverden 18 successen. De schatting van p_1 is dus 1/3 en het betrouwbaarheidsinterval [.21,.47]. Die betrouwbaarheidsintervallen heb ik gehaald uit "Documenta Geigy, Wissenschaftliche Tabellen".

DE KANSEN VAN EEN BEPAALD KIND

Bij vele stochastici heerst een zekere schroom om zich te bekommeren om de kansen van een singulier geval. Die schroom is terecht omdat in het

verleden zeer vreemde uitspraken zijn gedaan over kansen van bizondere gevallen bijvoorbeeld omtrent de "kans" dat morgen de zon zal opkomen. In ons geval is echter geen vuiltje aan de lucht.
Stel dat een moeder zich meldt bij Dr. Dische, samen met haar kind. Na enig onderzoek blijkt dat het kind kan worden gekaraktiseerd door

$(V_1,\ldots, V_7)=(0000111)$

Het is nogal logisch dat moeder en dokter willen praten over de kans dat dit kind baat zal hebben bij de wekker-therapie. De stochastici sluiten zich af voor een belangrijk toepassingsgebied van hun wetenschap als zij niet aan zulke discussies willen deelnemen, bijvoorbeeld omdat zij denken dat die kans niet goed is gedefinieerd.
Inderdaad, een intrigerende eigenschap van deze kans is dat hij afhangt van de informatie die in de beschouwing wordt betrokken. Let men helemaal niet op de specifieke kenmerken van het kind, dan is de kans gedefinieerd als de fractie van de hele referentie populatie die C=3 vertoont. Die kans hebben we eerder[P] genoemd; een schatting was 58/113 en een betrouwbaarheidsinterval [.42,.61] .

Betrekken we in de beschouwing dat V_1 voor dit kind de waarde 0 aanneemt, dan is de kans gedefinieerd als p_0 en vind je 2/3 als schatting en [.54,.79] als betrouwbaarheidsinterval.

Betrekken we in de beschouwing dat $(V_1,\ldots, V_7)$ de waarde (0000111) aanneemt dan hangt het ervan af welke hypothesen wij wensen in te voeren. Voeren we als hypothese in dat

$$P\{C=3 \mid (V_1,\ldots, V_7)=(0000111)\}=P(C=3 \mid V_1=0)$$

dan is het net vermelde resultaat voor p_0 van toepassing. Willen we helemaal geen hypothesen invoeren dan zit er niet anders op dan

$$P\{C=3 \mid (V_1,\ldots, V_7)=(0000111)\}$$

te schatten op grond van het in tabel 2 vermelde resultaat. De bijbehorende referentie steekproef bestaat uit 5 kinderen; 4 hiervan vertoonden C=3. De gezochte kans kan dus geschat met behulp van de schatting .80 of liever het betrouwbaarheidsinterval [.28,.99].

HET PROBLEEM VAN DE BESTE REFERENTIE KLASSE

Er ontstaat nu een boeiende en hoogst relevante problematiek. *Welke data en welke hypothesen zullen we in de beschouwingen betrekken?* Als wij alle data in de beschouwingen willen betrekken en geen hypothesen willen maken anders

dan dat ons kind een willekeurig element is uit Dr. Dische's referentie populatie, dan komen we weer terecht op een afschuwelijk groot betrouwbaarheidsinterval [.28,.99].

Laten we alle informatie van ons kind weg dan is het betrouwbaarheidsinterval veel kleiner, n.l. [.42,.61], maar we hebben dan niet gebruik gemaakt van de relevante informatie dat V_1=0. Gebruiken we die informatie dan vinden we betrouwbaarheidsinterval [.54,.79].

De filosoof Hans Reichenbach beschreef de problematiek aldus: gevraagd wordt "the narrowest class for which reliable statistics can be compiled". Naar mijn mening is dat in ons geval de referentie klasse die wordt gedefinieerd door V_1=0.

HET WISKUNDIG MODELLEREN VAN DE BESLISSINGSSITUATIE

De moeder die met haar kind bij de dokter komt doet dat niet uit filosofische interesse in het probleem van de beste referentie klasse, noch uit wetenschappelijke interesse in de kans dat haar kind baat zal hebben bij de wekker-therapie, gegeven de scores $(V_1,\ldots, V_7)$=(0000111). Zij, de dokter en het kind moeten een *beslissing* nemen, Er moet gekozen uit de verzameling $\{b_0, b_1\}$ waarbij

b_0 de beslissing voorstelt de therapie niet te gebruiken

b_1 de beslissing is om het kind de bel aan te binden .

In navolging van Neyman en Pearson die soortgelijke formuleringen ten grondslag legden aan hun theorie voor het toetsen van statistische hypothesen spreken we nu, stevig simplificerend, van een *fout van de eerste soort* indien beslissing b_1 wordt gekozen terwijl het resultaat is $C \leq 2$, d.w.z. de therapie wordt ten onrechte toegepast. We spreken van een *fout van de tweede soort* indien beslissing b_0 wordt gekozen terwijl toepassing van de wekker-therapie C=3 zou hebben opgeleverd. De heer Kardaun heeft mij er terecht op gewezen dat deze laatste definitie een beetje wonderlijk is want als beslissing b_0 wordt gekozen dan zou toch nog bij een latere evaluatie kunnen blijken dat C=3 en dan zou er dus eigenlijk helemaal geen fout zijn gemaakt. Hij heeft gelijk. Eigenlijk zouden we ook gegevens moeten hebben over hoe het gaat als er geen therapie wordt ingesteld. *)

Welke van de twee fouten het ergst is hangt af van de concrete situatie. Dat komt vooral doordat de narigheid voortvloeiend uit een fout van de tweede soort nogal verschillend kan worden beoordeeld. Voor een fout van de

*) Deze theorie is min of meer gebaseerd op de aanname dat C=3 onmogelijk is als niet wordt behandeld.

eerste soort, d.w.z. het toepassen van de therapie zonder gunstig resultaat, lijkt mij dat minder het geval. De dokter vindt dit vervelend omdat hij een verkeerd advies heeft gegeven, de moeder vindt het vervelend omdat alle werk voor niets is geweest, het kind is opgezadeld met een mislukte poging. Laten wij de uit zo'n fout van de eerste soort voortvloeiende narigheid begroten op 1 eenheid.

Wat is nu een redelijke invulling voor het getal a dat aangeeft het aantal eenheden narigheid voortvloeiend uit een fout van de tweede soort. Ik ben erop gebrand dat a niet moet worden beschouwd als een universele kostante maar dat a *wordt bepaald-binnen redelijke grenzen-op basis van het gesprek tussen moeder, dokter en kind.*

Dit is een tijd waarin het individu centraal staat. In de gezondheidszorg probeert men het medisch paternalisme terug te dringen door de patient te laten meebeslissen over zijn eigen wel en wee. Sommigen vinden het interessant om te filosoferen over de onontkoombaarheid van paternalistisch handelen in het geval de patient incompetent is, bijvoorbeeld doordat hij bewusteloos is of niet toerekeningsvatbaar. Veel moeilijker en interessanter lijkt het mij om op eenvoudige en toch doeltreffende wijze gestalte te geven aan de eis van "informed consent". Ons probleem van de bedplassers is zo mooi omdat allerlei complicerende factoren ontbreken. De data is eenvoudig en de inspraak van moeder en kind kan op vanzelfsprekende wijze worden gegund door hen het getal a te laten kiezen, binnen zekere op ethische of andere gronden vast te stellen grenzen. Ik kan mij heel goed voorstellen dat bij het ene gesprek a=½ wordt gekozen omdat de moeder helemaal niet zwaar tilt aan het ongerief en ook het kind niet de indruk geeft dat het eronder gebukt gaat, terwijl een ander gesprek resulteert in de keuze a=2 omdat de moeder zegt stapelgek te worden van het gedonder.
Ik zal aannamen dat onze moeder en haar kind, met scores 0000111, niet zwaar tillen aan het ongerief. Zij hebben liever dat er niets gebeurt dan dat er een vruchteloze poging wordt gedaan. Dus wordt besloten tot a=½.

WELKE BESLISSING?

Stel dat iedereen het erover eens is dat bij ons kind, met vector van scores 0000111, als referentie populatie moet worden gekozen: alle bedplassende kinderen soortgelijk aan die van Dr. Dische, maar met V_1=0 d.w.z. zonder gezinsmoeilijkheden. De fractie p_0 van C=3 gevallen in deze referentie popu-

latie is onbekend. Wel vond Dr. Dische voor de bijbehorende referentie steekproef dat deze bestond uit n=59 kinderen waarvan 40 langdurige verbetering vertoonden. Op grond hiervan werd het betrouwbaarheidsinterval [.54,.79] afgeleid voor de kans p_0 op C=3.
Wat betekent dit nu voor ons individueel kind als bovendien is besloten tot a=½? Omdat voor dit kind V_1=0, wordt dit kind opgevat als een willekeurige trekking uit dezelfde referentie populatie als de eerder vermelde referentie steekproef. Het is daarom alleszins redelijk om te stellen dat dit kind een kans p_0 heeft op genezing als het met de wekker-therapie wordt behandeld. Die kans is echter onbekend. Wel hebben we het betrouwbaarheidsinterval [.54,.79] voor p_0.
We konkluderen nu dat
(i) indien beslissing b_0 wordt gekozen, d.w.z. geen therapie wordt gebruikt, dan is er een kans p_0 dat die beslissing fout is en dus een fout van de tweede soort wordt gemaakt; dit resulteert in een verlies gelijk aan a=½. Het verwachte verlies is dus $ap_0=\frac{1}{2}p_0$ en een betrouwbaarheidsinterval is daarom [.27,.40].
(ii) indien beslissing b_1 wordt gekozen, d.w.z. de bel wordt aangebonden, dan is er een kans $1-p_0$ dat C < 3 optreedt en dus een fout van de eerste soort wordt gemaakt; dit resulteert in een verlies gelijk aan 1. Het verwachte verlies is dus $1-p_0$ en een betrouwbaarheidsinterval daarvoor is [.21,.46].
Kiezen tussen de beide intervallen [.27,.40] en [.21,.46] is een twijfelachtige zaak. Zij die grote verliezen willen vermijden zullen zich aangetrokken voelen tot het interval [.27,.40] en de bijbehorende beslissing b_0. Zij die een gokje willen wagen in de hoop dat het lot hun gunstig gezind zal zijn zullen misschien kiezen voor [.21,.46] met de bijbehorende beslissing b_1. Een risico-neutraal persoon zal echt niet weten wat hij moet doen. In mijn ogen heeft hij gelijk. Ons geval waarbij V_1=0 en a=½ is een typisch twijfelachtig geval.
In de praktijk wordt vaak geprobeerd de twijfel te elimineren door "nadere overwegingen" in de beschouwingen te betrekken of door één of andere "procedure" af te spreken.
Wat zouden hier die nadere overwegingen kunnen zijn? Men zou naast V_1=0 nog andere specifieke kenmerken van het kind kunnen beschouwen. In dat verband is tabel 5 interessant. Het eerste wat opvalt aan deze tabel is dat de

Informatie in ogenschouw	ref. steekproef n	s	betr. int. $\underline{p}$	$\bar{p}$	besl.b_0 $[\frac{1}{2}\underline{p}, \frac{1}{2}\bar{p}]$	besl.b_1 $[1-\bar{p}, 1-\underline{p}]$
geen	113	58	.42	.61	[.21,.31]	[.39,.58]
V_1=0	59	40	.54	.79	[.27,.40]	[.21,.46]
V_5=1	72	38	.41	.65	[.20,.33]	[.35,.59]
V_1=0, V_5=1	38	27	.54	.85	[.27,.43]	[.15,.46]
V=0000111	5	4	.28	.99	[.14,.50]	[.01,.72]

Tabel 5

eerste en de derde rij suggereren dat het voordeliger is om beslissing b_0 te nemen dan beslissing b_1. Het betrouwbaarheidsinterval [.21,.31] ziet er duidelijk het aantrekkelijkst uit. Het uitverkiezen hiervan is in overeenstemming met een voorstel dat ik vroeger eens op schrift heb gesteld. Het luidt dat het probleem van de referentie klasse kan worden opgelost door het betrouwbaarheidsinterval te kiezen met de kleinste bovengrens. Is dat een goed voorstel? Helemaal niet! Het relevante gegeven dat V_1=0 is dan n.l. niet in de beschouwingen betrokken. Het is als bij een rechtzaak waarbij de rechter oordeelt dat de verdachte schuldig is omdat de meeste verdachten schuldig zijn terwijl hij een alibi, hier corresponderend met V_1=0, niet in de beschouwingen heeft betrokken omdat dit hem zou hebben doen twijfelen.

Laten wij, zoals het behoort, V_1=0 als argument toe, dan gedraagt tabel 5 zich in overeenstemming met de verwachtingen. Het toevoegen van verdere overwegingen zoals V_5=1 in de voorlaatste rij, of V_2=V_3=V_4=0 en V_5=V_6=V_7=1 in de laatste rij, heeft enkel tot gevolg dat beide betrouwbaarheidsintervallen groter worden. Dat deze extra argumenten een averechtse uitwerking hebben komt doordat de referentie steekproeven kleiner worden als men deze argumenten toelaat. Daardoor neemt de ruis toe.

ENKELE BESLISSINGSPROCEDURES

Kunnen we de twijfel en onzekerheid niet uitbannen door er een rechtvaardige Salomo bij te halen? D.w.z. iemand die de knoop volgens een vaste regel door-

hakt en zich daarbij niet van de wijs laat brengen door huilende of krijsende moeders. Het probleem is dan natuurlijk welke vaste regel of procedure moet worden gehanteerd. De oplossing lijkt eenvoudig: de beste regel, de optimale procedure. Deze eenvoud is maar schijn. Het is n.l. helemaal niet duidelijk *welke optimaliteitseigenschap moet worden nagejaagd.*
Onze problematiek levert hiervan een prachtig voorbeeld. Stel dat wij het erover eens zijn dat van het kind met scorevector 0000111 enkel de score $V_1=0$ in de beschouwingen moet worden betrokken en dat wij dat willen doen door de bijbehorende referentie steekproef te beschouwen. Hier bestaat deze uit n=59 elementen waarvan s=40 een succes vertoonden, d.w.z. langdurige verbetering. Verder baseren wij ons op de keuze $a=\frac{1}{2}$.
Iets algemener geformuleerd luidt de vraag voor welke waarden van (n,s) de beslissing b_0 moet worden gekozen en voor welke waarden de beslissing b_1. We doen nu maar even net alsof de waarde n=59 een van te voren gegeven grootheid is. De waargenomen waarde s=40 is dan op te vatten als de uitkomst van een binomiaal verdeelde toevallige grootheid S waarbij de parameters dus gelijk zijn aan n en $p=p_0$, de onbekende fractie van de referentie populatie met $V_1=0$ waarbij langdurige verbetering, d.w.z. C=3,optreedt. Voor welke uitkomsten van S moet de beslissing b_0 genomen en voor welke de beslissing b_1? Het is duidelijk dat het onverstandig is een procedure te zoeken van een andere dan de volgende soort $\{\varphi_{k,\gamma}\ ;\ k=0,\ldots,n\ ;\ \gamma \in [0,1]\}$ met

$$\varphi_{k,\gamma}(s) = \begin{cases} 0 & \text{als } s < k \\ \gamma & \text{als } s = k \\ 1 & \text{als } s > k \end{cases}$$

de kans om beslissing b_1 te kiezen als de uitkomst s is gevonden. Voor $s < k$ wordt dus beslissing b_0 gekozen , voor $s > k$ beslissing b_1,en voor het twijfelgeval $s = k$ wordt beslissing b_1 gekozen met kans γ. Het is nogal natuurlijk dat zo'n randomiseringsmogelijkheid wordt ingebouwd.

Aangezien $\varphi_{k,\gamma}(s)$ de voorwaardelijke kans op beslissing b_1 beschrijft, gegeven S=s, geldt dat de onvoorwaardelijke kans op beslissing b_1 bij gebruik van $\varphi_{k,\gamma}$ gelijk is aan

$$E_p\,\varphi_{k,\gamma}(S) = \gamma\, P_p(S=k) + \sum_{s=k+1}^{n} P_p(S=s)$$

waarbij natuurlijk

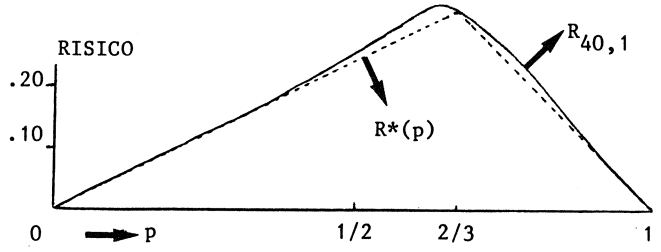
$$P_p(S=s) = \binom{n}{s} p^s (1-p)^{n-s} \qquad (s=0,\ldots,n)$$

Voor ons kind geldt dat het ook wordt opgevat als toevallig getrokken uit de referentie populatie evenals, maar tevens onafhankelijk van, de andere elementen van de referentie steekproef. Dus is zijn kans P(C=3) op langdurige

verbetering precies gelijk aan die onbekende kans p. Verder zijn C en S onafhankelijk. Als $C \leqq 2$, dan is b_1 de foute beslissing. Deze treedt op met kans $E_p\, \varphi_{k,\gamma}(s)$ en zo'n fout van de eerste soort kost 1 eenheid. Als C=3, dan is b_0 de foute beslissing. Deze treedt op met kans $1-E_p\, \varphi_{k,\gamma}(s)$ en kost a eenheden. Het resultaat van dit alles is dat het risico, d.w.z. verwachte verlies, bij gebruik van procedure $\varphi_{k,\gamma}$ gelijk is aan

$$P_p\,(C \leqq 2)\; E_p\, \varphi_{k,\gamma}(s) + aP_p\,(C=3)\;\{1-E_p\, \varphi_{k,\gamma}(s)\}$$
$$= ap + \{1-(a+1)p\}\; E_p\, \varphi_{k,\gamma}(s)$$

hetgeen moet worden opgevat als een functie $R_{k,\gamma}(p)$ van de onbekende parameter p. In figuur 2 vindt U dat weergegeven voor het geval n=59, a=2,



Figuur 2

k=40, γ=1. Het risico $R_{k,\gamma}(p)$ willen wij natuurlijk zo klein mogelijk hebben ben. Bij vaste p is dat heel gemakkelijk te verwezenlijken. Als $p < (a+1)^{-1}$ dan moeten wij natuurlijk $\varphi(s) \equiv 0$ kiezen, het risico wordt dan ap. Dat is nogal logisch want we nemen dan beslissing b_0 en dat betekent dat wij met kans p=P(T=1)een fout van de tweede soort maken die a eenheden kost. Als $p > (a+1)^{-1}$ dan moeten wij $\varphi(s) \equiv 1$ kiezen en wordt het risico $ap + \{1-(a+1)p\}=1-p$ gelijk aan de kans op een fout van de eerste soort. Het geminimaliseerde of *omhullende risico* is dus gelijk aan

$$R^*(p) = \min\{ap, 1-p\}$$

In figuur 2 vindt U dat als de gestippelde lijn ingetekend voor het geval n=59, a=2. Voor elke procedure $\varphi_{k,\gamma}$ geldt natuurlijk

$$R_{k,\gamma}(p) \geqq R^*(p)$$

De verschilfunctie

$$S_{k,\gamma}(p) = R_{k,\gamma}(p) - R^*(p)$$

kunnen we allerlei mooie namen geven zoals *tekort* (shortcoming) spijt (regret) of additioneel risico. Of we nu $R_{k,\gamma}$ proberen klein te krijgen of $S_{k,\gamma}$, dat is natuurlijk lood om oud ijzer hoewel dat wat optimaal lijkt vanuit het ene oogpunt niet optimaal hoeft te lijken vanuit het andere.

Er is een aardige optimaliteitseigenschap waarbij het niet uitmaakt of je het risico zelf of het tekort bekijkt. Vanwege

$$\int_0^1 S_{k,\gamma}(p)\,dp = \int_0^1 R_{k,\gamma}(p)\,dp - \int_0^1 R^*(p)\,dp$$

doet het er n.l. niet toe of je nu het oppervlak onder de spijtfunctie probeert te minimaliseren of dat onder de risicofunctie. Aan de laatste integraal valt n.l. niets te verhapstukken. Er zijn natuurlijk allerlei wiskundige manieren om de gevraagde oplossing te verkrijgen. Heel elegant gaat het als je subjectivistische of Bayesiaanse terminologie gebruikt, zonder natuurlijk de bijbehorende uitgangspunten te accepteren. Hoe het ook zij, de regel met minimaal oppervlak wordt als volgt vastgelegd

$s < (n-a+1)(a+1)^{-1}$ ⟶ neem beslissing b_0

$s > (n-a+1)(a+1)^{-1}$ ⟶ neem beslissing b_1

Als s onverhoopt gelijk mocht zijn aan de vorm in het rechter lid dan doet U maar waar U zin in hebt, het oppervlak verandert er niet door.
In ons geval was n=59 en $a=\frac{1}{2}$. Het gewraakte rechter lid is dan $39\frac{2}{3}$ zodat de regel met minimaal oppervlak correspondeert met k=40, $\gamma=1$ of, wat precies hetzelfde is, met k=39, $\gamma=0$. De bijbehorende risicofunctie is uitgezet in figuur 2. Voor de gevonden uitkomst s=40 wordt *de beslissing b_1 voorgeschreven door de regel met minimaal oppervlak.*

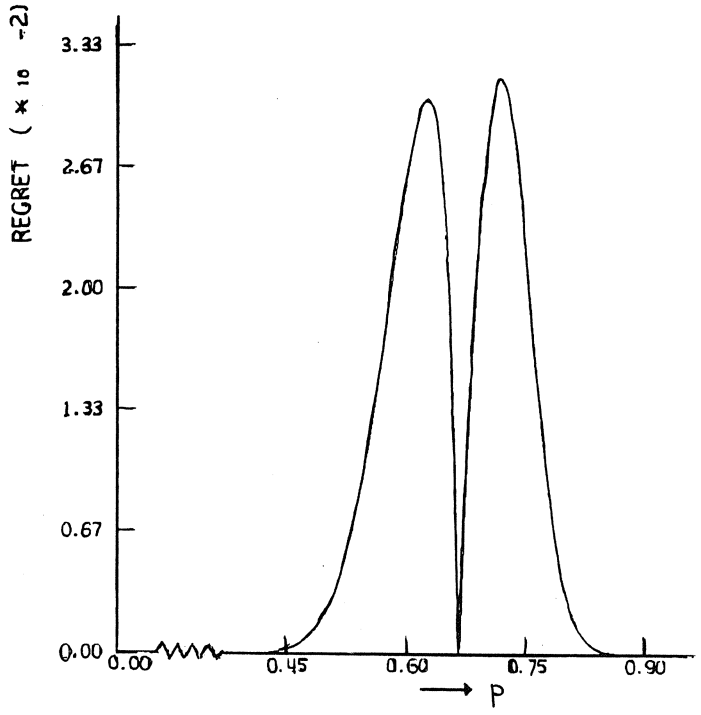
Men kan echter ook k en γ proberen te bepalen zodanig dat

$$\max_p R_{k,\gamma}(p) \qquad \text{of} \qquad \max_p S_{k,\gamma}(p)$$

zo klein mogelijk is. Men spreekt dan van de minimax risk of van de minimax regret procedure. De eerste regel is elegant te karakteriseren. Dat wil ik U niet onthouden want het sluit aan bij het U allen welbekende onderwerp van het toetsen van een kans. Het resultaat luidt dat de minimax regel wordt verkregen door de nulhypothese $H : p = (a+1)^{-1}$ te toetsen tegen $A : p > (a+1)^{-1}$ op het niveau $\alpha = a/(a+1)$. Het bewijzen van dit resultaatbetekent voor U een vette kluif die ik U niet wil ontnemen.
In ons geval was n=59, $a=\frac{1}{2}$, s=40. Het gaat dus om het toetsen van $H : p = 2/3$ tegen $A : p > 2/3$ op het significantieniveau $\alpha = 1/3$. Het resultaat s=40 is dan van verre significant en dus wordt *de beslissing b_0 voorgeschreven door de minimax risk regel.*

Aan de berekening van de minimax regret procedure is veel aandacht besteed door mijn zoon, de medisch student A. Schaafsma. Voor het onderhavige geval n=59, $a=\frac{1}{2}$ of, in zijn terminologie a=1, b=2, wist hij de computer te ontfutselen dat de minimax regret regel wordt vastgelegd door k=39 en $\gamma=1/6$. Voor het gevonden resultaat s=40 wordt dus *de beslissing b_1 voorgeschreven als je de minimax regret regel prefereert.*

Klaarblijkelijk verschilt de minimax regret regel slechts heel weinig van de regel met minimaal oppervlak. Het enige verschil is dat voor s=39 in het ene geval beslissing b_0 wordt gekozen terwijl in het andere geval wordt gerandomiseerd waarbij b_0 weliswaar niet met volstrekte zekerheid maar toch met kans 5/6 wordt gekozen. Omdat randomiseren ietwat onnatuurlijk is wordt in figuur 3 de regret functie gegeven voor het geval van de regel met minimaal oppervlak, dus de regel met k=39, γ=1. Voorts is gewerkt met a=1, b=2 i.p.v. met a=½. b=1. De afgedrukte regretfunctie is dus



Figuur 3

2 × zo groot als in onze theorie. Voor p=.62 is het regret volgens de kromme en de bijbehorende tabel gelijk aan .0305. In het kader van onze theorie zou dat dus .0152 moeten zijn. Dit wil ik even met U verifiëren door de bekende normale benadering toe te passen van de binomiale verdeling bij n=59, p=.62

$$S_{40,1}(.62) = \{1-(3/2).62\}\, P_{.62}(S \geqq 40)$$

$$= .07\, P\left(\frac{S-59\times.62}{59\times.62\times.38} \geqq \frac{39.5-59\times.62}{59\times.62\times38} \right)$$

$\approx .70\ P(U \geq 78) = .0154$

COMPUTERDIAGNOSTIEK

Het gebeurt slechts zelden dat een data verzameling zo mooi is als die van Dr. Dische. Vaak treden komtinue variabelen op zodat de definitie van referentie populatie dubieus wordt. In ieder geval krijgt men dan heel sterk te maken met de moeilijkheid dat referentie steekproeven vaak uit nul elementen zullen bestaan. Voortbordurend op figuur 1 is het duidelijk dat meer hypothesen naar binnen moeten worden gesmokkeld. Wat dat betreft staat een veelheid van paradigma's ter beschikking. Dezelfde data kan vaak op verschillende manieren worden bewerkt. Eén van deze paradigma's berust op de toepassing van de *formule van Bayes*. Stel er zijn k elkaar uitsluitende diagnostische kategorieën. Bij kategorie h heeft de vector van scores van een lukraak te trekken individu een verdeling met één of andere kansdichtheid f_h. Stel je hebt een patient met vector van scores x. Als nu à priori, d.w.z. voordat die scores zijn bepaald, de kans op kategorie t gelijk is aan p, dan is à posteriori de kans op kategorie t gelijk aan

$$p_t\, f_t\,(x) \;/\; \sum_{h=1}^{k} p_h\, f_h\,(x)$$

De moeilijheid is dat je vaak niet precies weet welke waarden moeten worden ingevuld voor $p_1, \ldots, p_k$. Ook weet je vrijwel nooit precies wat de dichtheden $f_1, \ldots, f_k$ zijn. Dus ontstaat een interessant schattingsprobleem waarbij, afgezien van de vector van scores x van het onderzochte individu, de data bestaat uit k steekproeven, de zogenaamde leersteekproeven, $X_{h1}, \ldots, X_{hn_h}$ uit f_h (h=1,..., k).

In Groningen is veel aandacht besteed aan het bepalen van de onzekerheden in de schattingen van de à posteriori kansen. Dit werk gaat schuil onder de naam POSCON. De ontwerper van het bijbehorende computerprogramma "POSCON , a decision-support system for DIAGNOSIS and PROGNOSIS heeft deze naam gekozen als samentrekking van POSterior probability en CONfidence interval. Dat is precies wat de bedoeling is: betrouwbaarheidsintervallen maken voor de achterafkansen zoals dat zo mooi lukte bij de bedplassers.

Een voorloper van POSCON is gebruikt om na te gaan of een schedel behoorde tot één van de acht Afrikaanse bevolkingen (data van de fysische anthropologen van Vark en Rightmire).

POSCON zelf is toegepast om betrouwbaarheidsintervallen te maken van de kansen van nierpatienten om, gezien hun scores op het moment van onder-

zoek, binnen 5 jaar na onderzoek dialyse te behoeven (data van nefroloog Beukhof).

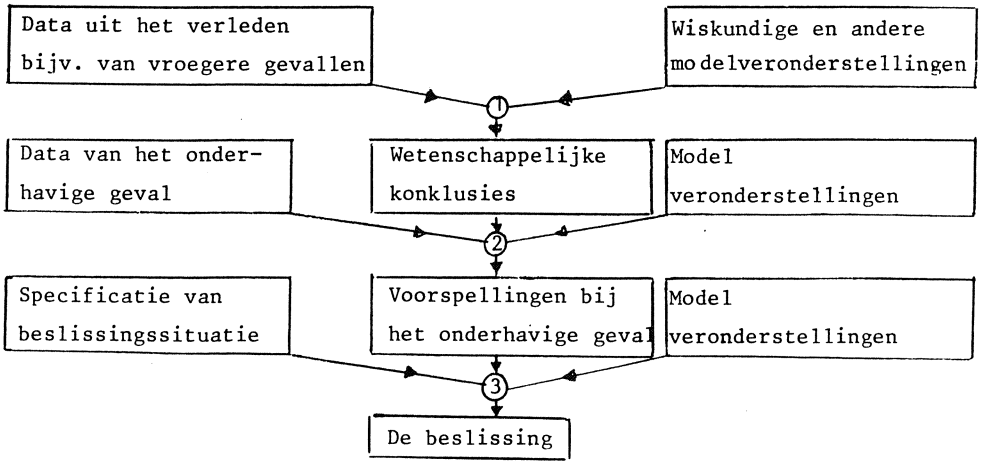
Ook is het gebruikt om de kansen te schatten dat een leerling die tot het VWO wordt toegelaten met een verdeeldadvies de eerste twee klassen haalt zonder te blijven zitten (data van psychologe Muilwijk)

POSCON is ook toegepast op allerlei ludieke voorbeelden. Het maakt een belangrijk bestanddeel uit van de cursus computerdiagnostiek van tweede jaars medische studenten (cursusleider van Vark).

EEN FILOSOFISCH SLOT

In de theorie van statistische beslissingsfuncties wordt vaak gedaan alsof het schatten van onbekende parameters en het toetsen van statistische hypothesen de meest wezenlijke beslissingsproblemen zijn die er bestaan. Daar is natuurlijk niets van waar. Echt zwaarwegende beslissingen gaan niet over parameters of hypothesen maar over individuele situaties. De tweede kamer beslist over de begroting van een komend jaar. Na de watersnoodramp heeft men beslist over de dijkhoogte. Enzovoort.

Bij zwaarwegende beslissingen wordt vaak een beroep gedaan op de wetenschap. Vaak spelen allerlei onzekerheden een rol. Sommige daarvan kunnen wetenschappelijk worden opgelost maar andere blijven bestaan. De discussie ontaardt dan vaak in een politiek steekspel waarbij de wetenschap wordt misbruikt als een soort arsenaal, een hok vol giftige pijlen, traangasgranaten en rookbommen. Niettemin herkent men vaak een structuur als aangegeven in figuur 4. Merk op dat deze figuur een soort uitwerking is van figuur 1.



Figuur 4

Tijdens een eerste puur wetenschappelijke fase wordt *statistical inference* bedreven. Het is gebruikelijk de na operatie 1 verkregen algemene konklusies op een nette wijze van allerlei onzekerheidsmarges te voorzien. Tijdens de tweede *predictieve* fase springt men vaak minder nauwkeurig met de onzekerheden om. Zodra het aan het *beslissen* toe is is vaak helemaal het hek van de dam en gaat het om de meeste stemmen of de grootste mond. Voor zulke situaties is dit verhaal niet bedoeld.

Zijn er situaties waar men bereid is of bereid hoort te zijn om tot en met het moment van de beslissing alle onzekerheden onder ogen te willen zien? Zulke situaties vindt men in de medische wetenschap. De dokter is door de *eed van Hippocrates* gebonden aan het laten prevaleren van het wetenschappelijk argument. Daar hoort bij de bereidheid om op elk moment alle onzekerheden onder ogen te willen te zien. Dat is de reden waarom bij het POSCON project wordt verwezen naar DIAGNOSE en PROGNOSE.

DANKBETUIGING

De schrijver, dezer, is dank verschuldigd aan de heren O.J.W.F. Kardaun, F. Ruddijs en G. Wortelboer voor allerlei commentaar en aan de heer A. Schaafsma voor het samenstellen van figuur 3.

LITERATUUR

D.M. van der Sluis, W. Schaafsma. POSCON, a descision-support system in diagnosis and prognosis, based on a statistical approach. Te verkrijgen bij de eerste auteur, Rekencentrum, RUG, Postbus 800, Groningen, (bevat verdere verwijzingen).

EEN WINKELMODEL

A.C.F. VORST

1. INLEIDING.

Wiskunde wordt steeds vaker toegepast om bepaalde verschijnselen te modelleren. Men neemt vaak een zeer complex verschijnsel waar en probeert dan de belangrijkste karakteristieken hiervan in een wiskundig model onder te brengen. Gegeven dit model, kan men met behulp van wiskundige theorie nagaan welke andere gebeurtenissen men zou moeten kunnen waarnemen. Komen deze redelijk overeen met de werkelijkheid, dan kan dit model gebruikt worden om gebeurtenissen te voorspellen of te beinvloeden. Een tweede voordeel van een wiskundig model is dat het de werkelijkheid zeer beknopt beschrijft, waar men anders zijn toevlucht moet nemen tot een ellenlange woordenbrij. Een nadeel daarbij is natuurlijk dat het model niet tot in de kleinste details de werkelijkheid beschrijft. Dit is minder erg dan op het eerste gezicht lijkt. Men heeft in de meeste situaties toch ook te maken met onvoorspelbare invloeden zoals bijvoorbeeld de weersgesteldheid.
De economie is één van de vele wetenschapsgebieden waar dit mathematisch modelleren wordt toegepast. Resultaten van één van de meest bekende voorbeelden in de economie kan men bijna dagelijks in de kranten vinden. We doelen hierbij op de macroeconomische modellen die het Centraal Planbureau te Den Haag gebruikt. Van iedere door kabinet of oppositie voorgestelde belastingverzwaring, lastenverlichting, etc. moet men weten wat een dergelijke maatregel oplevert of gaat kosten. Recentelijk was bijvoorbeeld aan de orde de belastingverandering voor tweeverdieners. Om de opbrengst van een dergelijke maatregel te kunnen bepalen zou men precies moeten weten wat iedere man en vrouw het komende jaar gaat verdienen. Dit is natuurlijk onmogelijk en daarom

probeert men het effect te berekenen via een model. In dit geval moet bij het zogenaamde doorrekenen van genoemde maatregel, met behulp van het gekozen model, rekening gehouden worden met neveneffecten. Bijvoorbeeld zullen sommige mensen hun baan opzeggen daar ze door de belastingmaatregel netto te weinig van hun verdiensten overhouden. Dit is natuurlijk zeer lastig in het model op te nemen, daar men het gedrag van mensen moet voorspellen. Wij zullen hier een voorbeeld geven van een wiskundig model dat gebruikt wordt in wat heet de regionale economie. Het gaat hier om de optimale verdeling van de winkelcapaciteit over een bepaalde regio

2. HET WINKELMODEL.

Voor ons model beschouwen we een bepaalde regio, bijv. een stad die bestaat uit een aantal woongebieden $X_1,\ldots,X_n$ en een aantal winkelcentra $Z_1,\ldots,Z_m$. De woongebieden kunnen bijvoorbeeld wijken zijn en onder winkelcentra verstaan we eventueel ook één à twee winkels. Sommige winkelcentra kunnen binnen bepaalde woongebieden vallen. We stellen ons nu de vraag hoe een inwoner van bijvoorbeeld woongebied X_1 zijn uitgaven over de verschillende winkelcentra zal verdelen. Een belangrijke factor hierbij is natuurlijk de afstand van de verschillende winkelcentra tot X_1. We schrijven daartoe d_{ij} voor de afstand tussen woongebied X_i en winkelcentrum Z_j. (We meten deze afstand op één of andere manier zoals bijv. afstand, kosten met openbaar vervoer.) Een tweede belangrijke factor is de attraktiviteit van een bepaald winkelcentrum. Een groot winkelcentrum dat goed gesorteerd is, trekt meer publiek dan een kleine specialistische winkel. De attractiviteit van winkelcentrum X_j gemeten in bijv. winkeloppervlakte, aantal verschillende produkten of een combinatie van deze en andere meetbare grootheden geven we aan met $W_j \geq 0$. We nemen nu aan dat als we de uitgaven van de inwoner van X_1 in de winkelcentra Z_j en Z_k met elkaar vergelijken we de volgende verhouding voor

deze uitgaven vinden

$$\frac{e^{-\beta d_{1j}} W^{\alpha}}{e^{-\beta d_{1i}} W_1^{\alpha}} \qquad (1)$$

Hierbij zijn α en β positieve constanten. Er geldt dus

$d_{1j} > d_{1k} \Rightarrow e^{-\beta d_{1j}} < e^{-\beta d_{1k}}$. α is een gedragsparameter, die aangeeft hoe sterk de bevolking reageert op attractiviteitsverschillen. Als nu O_i de totale uitgaven zijn die door de inwoners van zone X_i worden gedaan in een bepaalde periode (bijv. maand) dan ziet men eenvoudig in dat door inwoners van zone W_i het volgende bedrag in winkelcentrum Z_j wordt besteed

$$O_i \frac{e^{-\beta d_{1j}} W_j^{\alpha}}{\sum_{k=1}^{m} e^{-\beta d_{1k}} W_k^{\alpha}} \qquad (2)$$

Als wij voortaan schrijven $C_{ij} = e^{-\beta d_{ij}}$ en als we met D_j het totaal aan uitgaven in winkelcentrum Z_j weergegeven dan vinden we de volgende formule

$$D_j = \sum_{i=1}^{n} O_i \frac{C_{ij} W_j^{\alpha}}{\sum_{k=1}^{m} C_{ik} W_k^{\alpha}} \qquad (j=1,\ldots,m) \qquad (3)$$

De attractiviteit van een winkelcentrum geeft natuurlijk ook een maat voor de kosten verbonden aan de instandhouding van zo'n centrum (personeel, onderhoud, inkoop). We nemen hierbij aan dat deze kosten K_j lineair afhangen van W_j. Dus

$$K_j = kW_j \qquad (j=1,\ldots,m) \qquad (4)$$

Men kan ook $K_j = kW_j^{\gamma}$ veronderstellen, maar dit kan door een nieuwe schaling van de winkelcentra weer tot ons geval worden gereduceerd. De vraag is nu of we de winkelcentra zo

groot kunnen maken dat overal zal gelden

$$K_j = D_j \qquad (j=1,\ldots,m) \qquad (5)$$

De eis lijkt voor niet-economen vreemd aangezien D_j zo groot en K_j zo klein mogelijk de ideale toestand lijkt. Echter zou dit laatste het geval zijn, dan zullen nieuwe winkeliers, aangetrokken door de winstmogelijkheden, zich in dit centrum willen gaan vestigen. Men neemt dus aan dat in K_j altijd al een redelijk winstpercentage is meegerekend. We krijgen nu dus m niet-lineaire vergelijkingen in de m onbekenden W_j:

$$kW_j = \sum_{i=1}^{n} O_i \frac{C_{ij} W_j^{\alpha}}{\sum_{k=1}^{m} C_{ik} W_k^{\alpha}} \qquad j = 1,\ldots,m \qquad (6)$$

Het bovenbeschreven model is in de zestiger jaren ontwikkeld door Huff en tegelijkertijd door Lakshmanen en Hansen. Natuurlijk is het zo, dat de attractiviteit van de winkelcentra in de loop van de tijd kan veranderen. We voegen daarom aan W_j een tijdsindex toe, zodat we spreken van $W_j(t)$ i.p.v. W_j.
Onder de bovengenoemde aanname van een redelijk winstpercentage in K_j zal iedere winkelcentrum streven naar een verkleining van het verschil tussen $K_j(t)$ en $D_j(t)$. We nemen daarom aan, dat het verloop van $W_j(t)$ in de tijd aan de volgende differentiaal-vergelijking voldoet.

$$\frac{dW_j(t)}{dt} = \varepsilon\left(\sum_{i=1}^{n} O_i \frac{C_{ij} W_j(t)^{\alpha}}{\sum_{k=1}^{m} C_{ik} W_k(t)^{\alpha}} - kW_j(t) \right) \quad j = 1,\ldots,n \qquad (7)$$

De vergelijkingen in (7) geven n differentiaalvergelijkingen in de m onbekende functies $W_1(t),\ldots,W_m(t)$. Zo'n stelsel noemen we ook wel een dynamisch systeem. Men kan zichzelf bij zo'n dynamisch systeem een aantal vragen stellen. Wij zullen er hier enkelen geven

1. Zijn er waarden van $W_j(t)$ waarvoor het rechterlid van (7) gelijk aan nul is voor iedere $j = 1,\ldots,m$? Dit komt neer op de vraag die we via vergelijking (5) reeds stelden. Dit soort punten noemen we evenwichten.
2. Hoe hangen de evenwichten van de parameters van het model af?
3. Voor welke evenwichten geldt dat als $W_1(0),\ldots,W_m(0)$ dicht bij dit evenwicht liggen, de $W_j(t)$ naar die evenwichtswaarden gaan. Dergelijke evenwichten noemen we asymptotisch stabiel. Als een evenwicht niet asymptotisch stabiel is, dan is het niet een interessant evenwicht omdat een kleine verstoring er voor kan zorgen dat het systeem ver van dit evenwicht verwijderd raakt.
4. Is het mogelijk dat voor een j geldt dat $\lim_{t\to\infty} W_j(t) = 0$? Dit wil zeggen dat zo'n centrum op den duur verdwijnt.

In de volgende paragraaf zullen we nader op deze vragen ingaan. De antwoorden verschillen sterk voor de gevallen $0 < \alpha < 1$, $\alpha = 1$ en $\alpha > 1$. Het is dus van belang te weten wat α is in een praktische situatie. Tevens zal men natuurlijk ook β moeten weten. Voor deze α en β kan men meestal alleen maar schattingen geven. Het vinden hiervan behoort typisch tot de werkzaamheden van een econometrist. Wij zullen hierop niet nader ingaan en we veronderstellen dat α en β bekend zijn.

3. EVENWICHTEN.

Wij willen hier eerst eens het geval bekijken waarbij $0 < \alpha < 1$.
In dit geval kan men aantonen, dat er precies één evenwichtspunt bestaat. Wij gebruiken daarbij een stelling uit de differentiaalmeetkunde. Alvorens deze stelling te geven willen we hem eerst aan de hand van een eenvoudig voorbeeld duidelijk maken.

Zij

$$\frac{dx(t)}{dt} = f(x(t)),(-1 \leq x(t) \leq 1) \qquad (8)$$

een differentiaal vergelijking, waarbij f een continu differentieerbare functie is. We veronderstellen nu dat f in de randpunten naar binnen wijst. Hiermede bedoelen we dat $f(-1) > 0$ en $f(1) < 0$. Dit heeft tot gevolg dat als $-1 \leq x(0) \leq 1$ dan geldt ook voor alle $t \in [0,\infty)$ dat $-1 \leq x(t) \leq 1$.
Een evenwichtspunt is een waarde van $x \in [-1,1]$ waarvoor geldt dat $f(x) = 0$. We veronderstellen verder dat voor ieder evenwichtspunt x geldt dat $f'(x) \neq 0$. We kunnen nu aan ieder evenwichtspunt x een index getal +1 of -1 toevoegen via de volgende formule

$$\text{Ind}(x) = \begin{cases} +1 \text{ als } f'(x) < 0 \\ -1 \text{ als } f'(x) > 0 \end{cases} \qquad (9)$$

Men kan nu bewijzen dat de volgende formule geldt:

$$\sum_{\substack{x\in[-1,1] \\ f(x)=0}} \text{Ind}(x) = 1 \qquad (10)$$

Dit wil dus zeggen, dat als er n evenwichtspunten zijn waar de afgeleide van f positief is, dan zijn er n+1 evenwichtspunten waar de afgeleide negatief is. Tevens volgt dat er minstens één evenwichtspunt moet zijn. Als we een plaatje tekenen van functies f die aan de eisen voldoen zien we direct in dat dit klopt. In de figuren 1a en 1b hebben we een tweetal van deze functies getekend. Terwijl in figuur 1c niet aan de voorwaarden wordt voldaan.

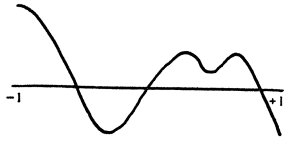


FIG. 1A

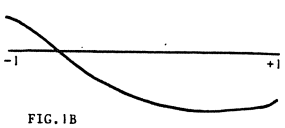


FIG. 1B

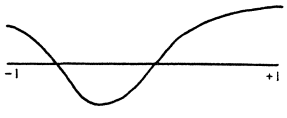


FIG. 1C

Van formule (10) bestaat ook een hoger dimensionaal analogon, welk wij nu zullen geven in een speciaal geval. Zij $S = \{(x_1,\ldots,x_n) \in \mathbb{R}^n | \Sigma\, x_j^2 \leq 1\}$. Laat

$$\frac{dx_j(t)}{dt} = f_j(x_1(t),\ldots,x_n(t)) \quad j = 1,\ldots,n \tag{11}$$

een n-tal differentiaalvergelijkingen zijn die op de rand van S naar binnen wijzen. We noemen een punt $(x_1,\ldots,x_n)$ een evenwicht als

$$f_j(x_1,\ldots,x_n) = 0 \qquad j = 1,\ldots,n \tag{12}$$

We bekijken nu voor iedere $x_1,\ldots,x_n$ de volgende nxn-matrix $A(x_1,\ldots,x_n)$ met

$$A(x_1,\ldots,x_n)_{jk} = \frac{\partial f_j(x_1,\ldots,x_n)}{\partial x_k} \tag{13}$$

We veronderstellen verder, dat voor ieder evenwichtspunt $(x_1,\ldots,x_n)$ geldt, det $A(x_1,\ldots,x_n) \neq 0$. Deze determinant vervangt de afgeleide in formule (9). We definiëren dus weer een index voor een evenwichtspunt $(x_1,\ldots,x_n)$ als volgt:

$$\mathrm{Ind}(x_1,\ldots,x_n) = \begin{cases} +1 \text{ als } \det(-A(x_1,\ldots,x_n)) > 0 \\ -1 \text{ als } \det(-A(x_1,\ldots,x_n)) < 0 \end{cases} \tag{14}$$

We hebben dan de volgende stelling, die bekend staat als de Poincaré Hopf index stelling.

Stelling. Als het dynamische systeem (11) voldoet aan de voorwaarden zoals boven gegeven dan geldt

$$\Sigma\, \mathrm{Ind}(x_1,\ldots,x_n) = 1 \tag{15}$$

waarbij we sommeren over alle evenwichtspunten.
Wij willen nu deze stelling toepassen op het winkelmodel. Hierbij zijn wij geinteresseerd in het gebied $\mathbb{R}^n_+ = \{(W_1,\ldots,W_m) \in \mathbb{R}^n | W_j \geq 0,\ j = 1,\ldots,n\}$. Het gebied S in de stelling heeft echter een andere vorm. We kunnen deze

verschillen oplossen door i.p.v. $\mathbb{R}^n_+$ een gebied $V \subset \mathbb{R}^n_+$ te nemen dat ongeveer de vorm van S heeft en dan op te merken, dat als dit aan bepaalde eisen voldoet de stelling ook geldt voor zo'n gebied V. Natuurlijk moeten we V zo kiezen, dat $\mathbb{R}^n_+ \setminus V$ geen interessante punten meer bevat.
We kunnen nu de matrix $A(W_1,\ldots,W_m)$ voor ieder evenwichtspunt bepalen als we voor de functies f_j de rechterleden van (7) nemen. Men kan dan aantonen, dat voor ieder evenwichtspunt geldt

$$\det(-A(x_1,\ldots,x_m)) > 0 \qquad (16)$$

Dus alle evenwichtspunten hebben index gelijk aan 1. Zodoende volgt uit de Poincaré-Hopf index stelling dat er precies één evenwichtspunt is. Men kan tevens aantonen, dat voor alle beginwaarden $W_1(0),\ldots,W_n(0)$ de oplossingen van de differentiaal-vergelijkingen (7) naar dit unieke evenwicht convergeren. Op ongeveer dezelfde manier kan men aantonen dat voor $\alpha \geq 1$ er altijd minstens één evenwicht bestaat, maar dat dit evenwicht niet uniek hoeft te zijn. Tevens kan men aantonen, dat er bijna altijd meer niet stabiele dan stabiele evenwichten zijn. Een volgende vraag die wij reeds gesteld hebben, is wat er gebeurt met het evenwicht als men de parameters van het model (bijv. O_i, C_{ij}, β, α) een klein beetje verandert. Dit is natuurlijk voor de economische interpretatie van belang. Men kan nu aantonen, dat zolang $\alpha < 1$ het evenwicht continu afhangt van alle parameters van het model. Dus een kleine verandering in bijvoorbeeld O_1 zal het evenwicht niet te veel verschuiven en dus zullen winkelcentra niet te veel in grootte veranderen. Geheel anders ligt dit in het geval waarbij $\alpha \geq 1$. Wij zullen dit in de volgende paragraaf aan de hand van een voorbeeld laten zien.

4. WINKELCATASTROPHES.

We bekijken nu het speciale geval van (7) waarbij $\alpha = 2$ en

$n = m = 2$ d.w.z. dat we twee gebieden hebben die zowel woon als winkelgebied zijn. Verder nemen we aan dat

$$C_{11} = C_{22} > C_{12} = C_{21} \tag{17}$$

Het ongelijkheidsteken is een logisch gevolg van de definitie van C_{ij}.
We schrijven voortaan

$$C = \frac{C_{12}}{C_{22}} \qquad 0 < C < 1 \tag{18}$$

Veronderstellen we nu verder dat $(O_1+O_2)/k = 1$ hetgeen door schaling te verkrijgen is en schrijven we

$$Q = \frac{O_1}{k} - \frac{1}{2}, \quad \frac{1}{2} \leq Q \leq \frac{1}{2} \tag{19}$$

Dan geeft Q de ongelijkheid van de inkomensverdeling over de twee woongebieden aan.
De vergelijkingen (6) van ons model worden dan

$$\frac{(Q+\frac{1}{2})W_1^2}{W_1^2+W_2^2C} + \frac{(\frac{1}{2}-Q)W^2C}{W_1^2C+W_2^2} - W_1 = 0 \tag{20}$$

$$\frac{(Q+\frac{1}{2})W_2^2C}{W_1^2+W_2^2C} + \frac{(\frac{1}{2}-Q)W_2^2}{W_1^2C+W_2^2} - W_2 = 0 \tag{21}$$

We zien hieruit onmiddellijk door het optellen van de twee vergelijkingen dat $W_1 + W_2 = 1$ en $0 \leq W_1 \leq 1$. Als we dus de W_1 van een oplossing weten dan weten we ook W_2.
Zoals in de vorige paragraaf opgemerkt is, bestaat er voor alle waarden van de parameters minstens één evenwicht en dit wordt dus gekarakteriseerd door W_1 (want $W_2 = 1-W_1$). We kunnen nu dus voor alle $-\frac{1}{2} \leq Q \leq \frac{1}{2}$ en alle $0 < C < 1$ de evenwichten zoeken. In figuur 2a en 2b laten wij zien hoe W_1

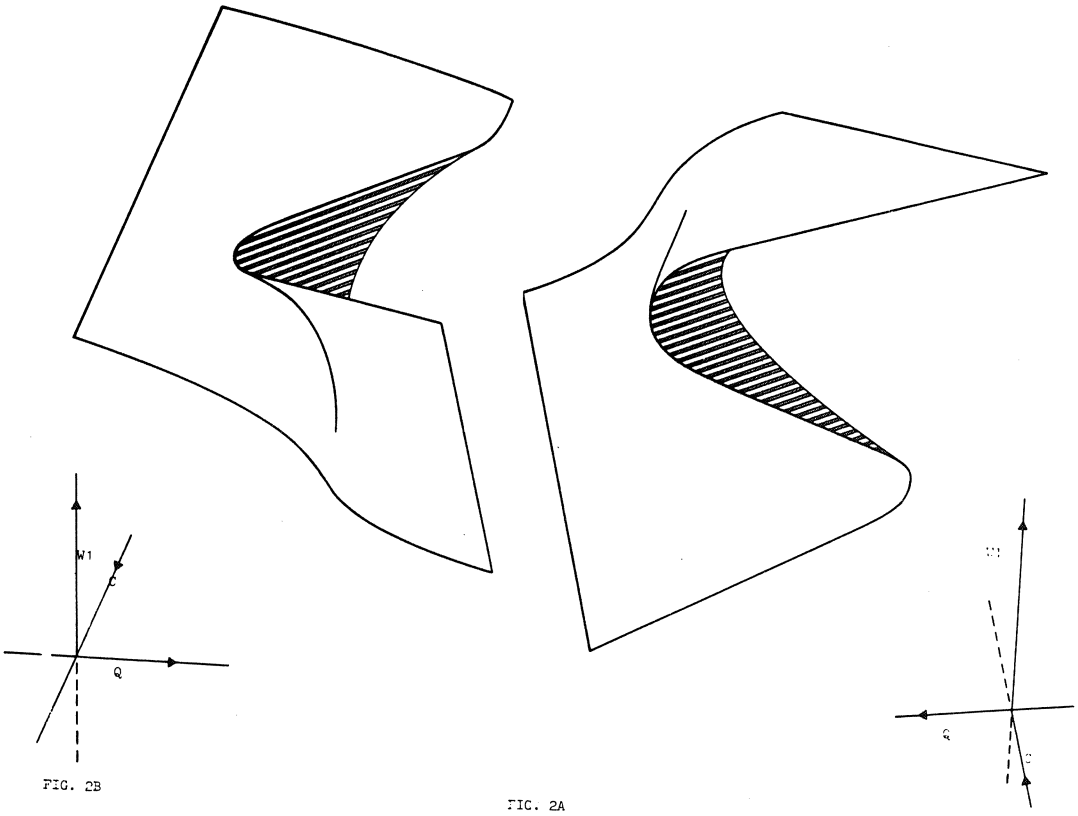


FIG. 2B

FIC. 2A

afhangt van Q en C. Dit is dus een 3-dimensionaal plaatje. We zien tevens dat voor sommige waarden van Q en C er meer dan één waarde voor W_1 is. Dit kan omdat $\alpha = 2 > 1$ terwijl we in de vorige paragraaf reeds vermeld hebben dat voor $\alpha < 1$ dit onmogelijk is. Daar de figuur 2 misschien moeilijk te interpreteren is geven wij in figuur 3 aan hoe W_1 afhangt van Q voor twee vaste waarden van C en in figuur 4 hoe W_1 afhangt van C voor twee vaste waarden van Q. Zowel de ononderbroken als gestippelde lijn geven evenwichten aan. Ook nu zien we weer dat er bij sommige waarden van C en Q drie evenwichten zijn.

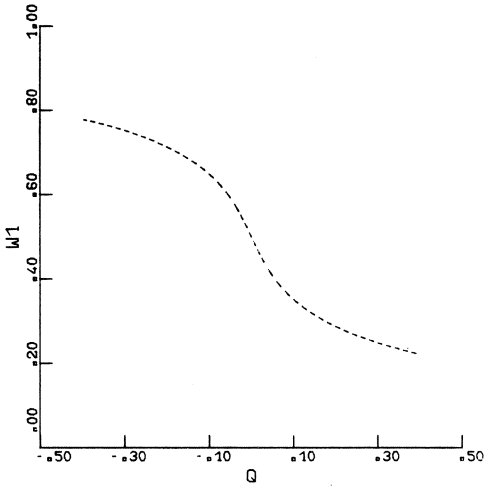


FIG.3A (C = 0.25)

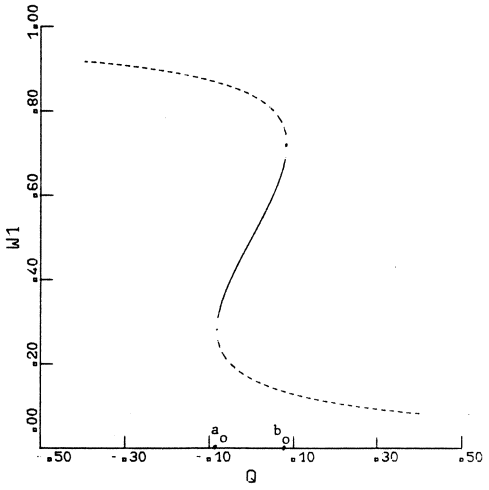


FIG.3B (C = 0.08)

Tot nog toe hebben we alleen gekeken wat de waarden van de evenwichten zijn. Maar zoals reeds eerder beschreven kunnen we ook werken met een stelsel differentiaalvergelijkingen zoals in (7). Dit kunnen we als volgt doen

$$\frac{dW(t)}{dt} = k\varepsilon\left\{\frac{(Q+\frac{1}{2})W_1(t)^2}{W(t)^2+W_2(t)^2C} + \frac{(\frac{1}{2}-Q)W_1(t)^2}{W_1(t)^2C+W_2(t)^2} - W_1\right\} = 0 \qquad (22)$$

$$\frac{dW_2(t)}{dt} = k\varepsilon\left\{\frac{(Q+\frac{1}{2})W_2(t)^2C}{W_1(t)^2+W_2(t)^2C} + \frac{(\frac{1}{2}-Q)W_2(t)^2}{W_1(t)^2(+W_2(t)^2} - W\right\} = 0 \qquad (23)$$

We kunnen nu nagaan welke van de gevonden evenwichten asymptotisch stabiel zijn. Deze evenwichten hebben wij in figuren 3 en 4 aangegeven met de ononderbroken krommen. De niet asymptotisch stabiele evenwichten worden aangegeven door de gestippelde kromme. We zien, dat er inderdaad steeds meer niet asymptotisch stabiele evenwichten dan asymptotisch stabiele evenwichten zijn. Zoals reeds eerder opgemerkt kunnen alleen de asymptotische stabiele evenwichten in de praktijk voorkomen. Het interessante is nu dat men de

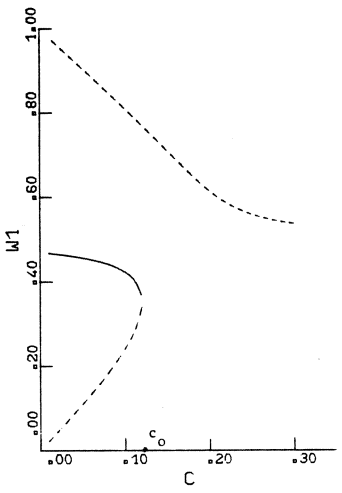


FIG. 4A (Q = -0.03)

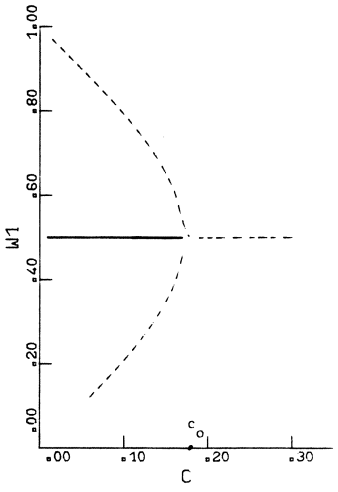


FIG. 4B (Q = 0.0)

figuren heel fraai kan interpreteren. Laten wij bijvoorbeeld naar figuur 4a kijken. We houden nu Q vast en beginnen met een $C < c_o$. We hebben dan een evenwicht W_1 op de ononderbroken lijn boven C. Stel nu eens, dat de benzineprijs langzaam gaat zakken, dit zal betekenen dat ook β zakt en daarmee zal C stijgen. We zien dan uit figuur 4a dat W_1 langzaam daalt. Echter op het moment, dat C de waarde c_o passeert verdwijnt plotseling de waarde van W_1 op de ononderbroken lijn en men kan dan bewijzen dat het eerste winkelcentrum volledig verdwijnt (d.w.z. $W_1 \to 0$). Dit kunnen we economisch als volgt interpreteren. Voor Q = -0.03 zitten we in een situatie waarbij in het tweede woongebied meer verdient wordt dan in het eerste woongebied. Bij een hoge benzineprijs zullen veel mensen in het winkelcentrum bij hun woongebied inkopen doen. Daalt nu langzaam de benzineprijs dan wordt het goedkoper om even naar het andere winkelcentrum te rijden. Dit zal voornamelijk van X_1 naar Z_2 zijn omdat Z_2 van nature groter is, daar er meer mensen dichtbij wonen. Zakt de benzineprijs nu nog verder dan zal Z_1 de concurrentie met Z_2 niet meer aankunnen en de één na de andere winkel zal sluiten. Dit voorbeeld is een goede

illustratie hoe wiskundige bewerkingen met het model bepaalde toestanden voorspellen, die in de werkelijkheid kunnen worden waargenomen. Wie kent niet tenslotte de kleine buurtwinkels die met de opkomst van de auto niet langer kunnen concurreren met de iets verder gelegen supermarkten. Soortgelijke interpretaties kan men geven bij de andere figuren. Laten wij bijvoorbeeld eens kijken naar figuur 3b. Als we beginnen met een Q tussen a_o en b_o dan hebben wij twee winkelcentra. Nemen we nu aan dat Q langzaam stijgt, d.w.z. het totale inkomen in X_1 stijgt relatief t.o.v. het totale inkomen van X_2. We zien dan dat Z_1 relatief groter wordt. Echter op het moment dat de inkomensverdeling Q het punt b_o passeert, zal winkel centrum Z_2 volledig verdwijnen daar de uitgaven van bewoners uit X_2 te klein worden.
Het fenomeen, dat een kleine verandering in de parameters eerst een geleidelijke verandering in W_1 geeft en dan plotseling een hele grote verandering, staat bekend als een catastrophe.

5. CONCLUSIES.

Wij hebben hier een voorbeeld gegeven hoe men een reële situatie kan modelleren. Vervolgens hebben wij met het model enige wiskundige exercities uitgevoerd en deze weer economisch geinterpreteerd. Het is duidelijk, dat men dit soort modellen ook in vele andere situaties kan gebruiken. Maar tevens moeten we natuurlijk opmerken dat dit model slechts een eerste benadering van de werkelijkheid geeft. Er zullen vele verfijningen moeten worden aangebracht om een dergelijk model in een specifieke situatie te kunnen gebruiken.

6. LITERATUUR.

In het volgende boek staan een uitgebreidere afleiding van het winkelmodel, alsmede toepassingen in andere situaties beschreven

Wilson, A.G. (1981), Catastrophe Theory and Bifurcations: Applications to Urban and Regional Systems, Croom Helm, London.

Voor een aardig boek over differentiaalvergelijkingen en dynamische systemen willen wij verwijzen naar:

Hirsch, M.W. & S. Smale (1974), Differential Equation, Dynamical Systems and Linear Algebra, Academic Press, New York.

Het bovenbeschrevene kan men in uitgebreidere vorm terugvinden in de volgende twee artikelen:

Rijk, F.J.A. & A.C.F. Vorst (1983), Equilibrium Points in an Urban Retail Model and their connection with Dynamical Systems, Regional Science and Urban Economics 13, 383-399.

Kaashoek, J.F. & A.C.F. Vorst (1984), The Cusp Catastrophe in the Urban Retail Model, Environment and Planning A, 16.

In de navolgende boeken kan men meer voorbeelden vinden van mathematisch modelleren:

Burmeister, E. & A.R. Dobell (1970), Mathematical Theories of Economic Growth, The Macmillan Company, London.

PROGRAMMEREN: KUNDE, KUNST OF KUNSTJE?

Konsekwenties voor het onderwijs

J.J. van Amstel
Onderafdeling Wiskunde en Informatica
Technische Hogeschool Eindhoven

PROLOOG (enkele citaten uit (1))

The programmer of today shares many of the attributes of the craftsman. He learns his craft by a short but highly paid apprenticeship in an existing programming team, engaged in some ongoing project; and he develops his skills by experience rather than by reading books or journals.
Programmers of the present day share many of the attributes of the high priest. Our altars are hidden from the profane (...) and regarded by the general public with mixed feelings of fear and awe, appropriate for their condition of powerless dependence.
We would like to claim that computer programming has transcended its origins as a craft, has avoided the temptation to form itself into a priesthood, and can now be regarded as a fully fledged engineering profession.
The most striking characteristic of an engineer is the manner in which he qualifies for entry into his profession; not only does he work out the long apprenticeship of the craftsman, not only does he undergo the brief graduation or initiation ceremonies of the high priest, but these are both preceded by many years of formal study in schools and in universities.

INLEIDING

Programmeren is een lange tijd gezien als een ambachtelijke activiteit. Ervaring was de grootste leermeester. Het enige dat nodig was, was het leren kennen en kunnen hanteren van het gereedschap: de computer en de programmeertaal. Leerboekjes begonnen dan ook met het tweetallig stelsel en het programmeeronderwijs bestond uit het onderwijs in een programmeertaal. Maar de opvattingen over de aard van het programmeren veranderen. Dit komt bijvoorbeeld tot uitdrukking in de titels van enkele boeken:

in 1968: The art of computer programming
D. E. Knuth (2)

in 1976: A discipline of programming
E.W. Dijkstra (3)

in 1981: The science of programming
D. Gries (4)

Nu moeten we natuurlijk oppassen met het trekken van conclusies op grond van de titels van boeken. Zo verscheen, evneens in 1981, een boek van J.C.Reynolds met de titel

The craft of programming

(5). Maar juist ook dit boek laat zien dat programmeren de ambachtelijke status ontgroeit. In het boek worden algoritmen op een systematische wijze afgeleid, wordt het effect van algoritmen vastgelegd zonder dat ze uitgevoerd worden op een computer en worden argumenten gegeven voor de correctheid van de algoritmen. De veranderde opvatting over het programmeren moet ook in het onderwijs zijn weerslag vinden.

De "nieuwe" manier van programmeren maakt gebruik van de predicatenrekening. In de boeken van Dijkstra en Gries gebeurt dit op een formele wijze, bij Reynolds gebeurt dit minder formeel. Er is een heel spectrum van nivo's van formaliteit, waarop het programmeeronderwijs kan worden gegeven. Het ene uiterste van het spectrum wordt gevormd door het alleen maar onderwijzen van een programmeertaal, het andere uiterste is het onderwijzen van een programmeer-calculus. Deze twee nivo's zullen we hier het "informele" respectievelijk het "formele" nivo noemen. Het gaat ons nu echter om een nivo tussen deze twee extrema; dit nivo zullen we het "semi-formele" nivo noemen. We moeten ons realiseren dat er tussen de twee uitersten vele nivo's zijn aan te brengen. Het semi-formele nivo dat hier gepresenteerd zal worden is gebaseerd op de wijze waarop het programmeren behandeld wordt in het boek "Programmeren: Het ontwerpen van algoritmen" (6), waarin Pascal als notatie wordt gebruikt voor de algoritmen. Het formele nivo is gebaseerd op "Een methode van programmeren" (7), waarin zogenaamde guarded commands worden gebruikt.

Is de semi-formele of de formele aanpak niet alleen bedoeld voor de professionele programmeurs? Kunnen we in het secundair onderwijs niet volstaan met het informele nivo? En zoniet, hoe formeel moeten we dan te werk gaan? Als we van leerlingen vragen een opstel te schrijven, dan zal dit opstel aan bepaalde eisen moeten voldoen. Zo zal het opstel moeten "passen" bij de opgegeven titel, het zal in correct Nederlands geschreven moeten zijn en ook aan de indeling zullen bepaalde eisen gesteld worden. Deze eisen worden aan alle leerlingen gesteld ondanks het feit dat hooguit een enkeling een professioneel schrijver zal worden. Voor alle vakken in het onderwijs kennen

we dit soort kwaliteitseisen, waarbij de eisen voor de verschillende vormen van onderwijs kunnen verschillen. Op veel verschillende plaatsen binnen het onderwijs wordt de differentiaal- en integraalrekening onderwezen. Voor alle plaatsen gelden bepaalde kwaliteitseisen, die niet voor alle plaatsen dezelfde zijn. Voor het programmeeronderwijs krijg je wel eens de indruk dat er bij de onderwijsgevenden niet eens een kwaliteitsbesef aanwezig is, laat staan dat er met bepaalde kwaliteitseisen gewerkt wordt. We weten echter nog niet hoe, voor de verschillende vormen van onderwijs, het programmeren onderwezen moet worden. Er zijn deskundigen op het gebied van het programmeren, die niets of nauwelijks iets weten van het onderwijs op een bepaald nivo, maar die wel menen dat zij kunnen voorschrijven hoe het programmeeronderwijs op dat nivo gegeven moet worden. Er zijn ook leraren, die niet of nauwelijks kundig zijn op het gebied van het programmeren maar wat kennis en kunde betreft iets uitsteken boven hun collega's, maar die wel menen te kunnen voorschrijven hoe het programmeeronderwijs gegeven moet worden.

EEN EERSTE VOORBEELD; DE INFORMELE, DE SEMI-FORMELE EN DE FORMELE AANPAK

De drie verschillende aanpakken zullen nu eerst gedemonstreerd worden. Dit zal voornamelijk gebeuren aan de hand van een voorbeeld. Getoond wordt wat, voor de verschillende aanpakken, de oplossing zou moeten of kunnen zijn voor een zelfde programmeerprobleem. Het probleem luidt:

> Van een array x van n (≥ 1) elementen heeft elk element de waarde 0 of de waarde 1. Gevraagd wordt een algoritme dat het aantal paren indices (i,j) bepaalt, met $1 \leq i < j \leq n$, waarvoor geldt dat $x(i) = 0$ en $x(j) = 1$.

Allereerst het informele nivo. (Hierbij zal Pascal als programmeertaal gebruikt worden.)
We kunnen het probleem verduidelijken door een voorbeeld te geven van een rij nullen en enen (bijvoorbeeld 1001010) en de daarbij behorende oplossing (5). Als we voordoen hoe we aan die 5 komen, hebben we direct ook al laten zien hoe het algoritme kan luiden. Het enige dat nu nog moet gebeuren is het formuleren van dit algoritme in Pascal. Waarschijnlijk komt dan de volgende Pascal-versie uit de bus:

```
c := 0;
for k := 1 to n-1 do
  if x(k) = 0 then for j := k+1 to n do
                     if x(j) = 1 then c := c + 1
```

Aan het vinden van de oplossing ligt geen systematiek ten grondslag. Het voorbeeld heeft geen lerend effct voor andere problemen, tenzij het om analoge problemen gaat. Het enige dat we met betrekking tot programmeren doen, is "onderwijzen door voorbeelden" en het aanleren van een programmeertaal.
We kunnen bij het voorbeeld eventueel nog een efficiëntere oplossing laten zien (om te tonen hoe goed we zelf wel zijn?), waarin het aantal nullen wordt "bijgehouden":

```
c := 0;
if x(1) = 1 then d := 0 else d := 1;
for k := 2 to n do if x(k) = 1 then c := c + d
                               else d := d + 1
```

Voor het onderwijs op dit informele nivo kan bijna elk boek gebruikt worden, als "Pascal" maar in de titel voorkomt. Er moet in het boek niet te veel nadruk gelegd worden op syntactische aangelegenheden. Anders loop je de kans dat het boek wemelt van de soort opgaven: Is .... een correcte ..... in Pascal en zoja wat is het resultaat? Wat van een informele aanpak op zijn minst gevraagd mag worden is dat de leerlingen de constructies uit de taal actief leren beheersen, dat ze leren hoe ze zich in de taal moeten uitdrukken; het lezen van andermans geschrift (vooral de vreemde brouwsels uit veel opgaven van de bedoelde soort) is veel minder belangrijk.

Nu het semi-formele nivo. (Waarbij ook gebruik wordt gemaakt van Pascal.)
In het onderwijs op dit nivo maken we gebruik van de specificaties, van invariante relaties en andere uitspraken voor de ontwikkeling van de algoritmen. De uitspraken behoeven niet in een formeel jasje te zijn gestoken, als ze maar precies en eenduidig zijn.
Het effect van de statements wordt vastgelegd door middel van transformaties van uitspraken. Het effect van statement S wordt vastgelegd in een regel van de vorm

{A} S {R}

De betekenis hiervan is: Wil statement S als resultaat R opleveren dan moet vòòr de uitvoering van S de uitspraak A gelden. Deze regels worden dan later gebruikt bij de constructie van algoritmen. We zullen er nu niet verder op ingaan, men zij verwezen naar bijvoorbeeld (6). Een zeer belangrijke regel is de zogenaamde invariantiestelling voor repetities. Deze invariantiestelling zegt dat de repetitie *while B do S* een resultaat, beschreven door een uitspraak R, oplevert als voldaan is aan de volgende condities:

- de repetitie eindigt (op een gegeven moment geldt de uitspraak not B);
- vòòr de repetitie geldt de uitspraak P met als eigenschap dat de uitspraak P and not B de uitspraak R impliceert;
- uitvoering van de statement S tast de geldigheid van P niet aan.

Als aan deze condities wordt voldaan, kunnen we het effect van de repetitie dus vastleggen met de regel:

{P} *while B do S* {P and not B}

De uitspraak P wordt de invariante relatie genoemd, omdat deze uitspraak blijft gelden tijdens de verwerking van de repetitie. Deze invariante relatie is een uitspraak over alle toestanden die gelden tijdens de verwerking van de repetitie. Ook voor de verwerking geldt P. En na de verwerking van de repetitie geldt P nog steeds; bovendien geldt dan nog not B. De regel wordt gebruikt om bij een gegeven probleem, dat met behulp van een repetitie kan worden opgelost, deze repetitie op een systematische manier te vinden.
Laten we naar het voorbeeld gaan kijken. Allereerst moet de specificatie van het probleem gegeven worden. Deze specificatie bestaat uit twee uitspraken. De eerste uitspraak, de zogenaamde beginconditie, is de uitspraak die geldt vòòr de uitvoering van het gewenste algoritme. De tweede uitspraak beschrijft de te realiseren eindtoestand. Voor ons voorbeeld luidt de specificatie:

beginconditie: de variabele x, gedeclareerd als
var x: array (1..n) of 0..1
heeft een waarde

eindconditie : c = "het aantal paren (i,j), $1 \leq i < j \leq n$, met $x(i) = 0$ en $x(j) = 1$"

Uit deze specificatie kan de volgende invariante relatie voor een repetitie afgeleid worden:

c = "het aantal paren (i,j), $1 \leq i < j \leq k$, met $x(i) = 0$ en $x(j) = 1$"

Laten we deze uitspraak aangeven met P. Het is duidelijk dat door P samen met k = n de eindconditie wordt geïmpliceerd. We kunnen P ook gemakkelijk laten gelden vòòr de repetitie door voor de variabelen in P, c en k, als waarden te kiezen 0 respectievelijk 1. (We hadden ook kunnen nemen 0 en 0, maar veel mensen hebben wat moeilijkheden bij uitspraken over lege domeinen.) Voor het algoritme hebben we tot nu toe gevonden:

```
k := 1; c := 0;
{P}
while k <> n do {P ∧ k ≠ n} .....
```

Voor de te herhalen statement(s) na *do* geldt dat voor de uitvoering hiervan steeds $P \wedge k \neq n$ geldig is. Door de statement(s) moet er voor gezorgd worden dat:
- P blijft gelden;
- de repetitie eindigt.

Voor de eindigheid kan gezorgd worden door k op te hogen; laten we voorzichtig zijn en k met 1 ophogen. Het algoritme krijgt dan de vorm:

```
k := 1; c := 0;
{P}
while k <> n do
    begin {P ∧ k≠n}
          k := k + 1;
          {P geldt niet; de uitspraak geldt voor c en k-1}
          "zorg ervoor dat P gaat gelden"
          {P}
    end
{P ∧ k=n, dus de eindconditie}
```

Om de invariant weer te laten gelden moet, voor het geval dat $x(k) = 1$, het aantal nullen bekend zijn in $x(1..k-1)$. Daarom breiden we de invariant P uit met Q:

d = "aantal nullen in $x(1..k)$"

Met de aanpassing van de initialisatie i.v.m. Q krijgen we nu:

```
k := 1; c := 0; d := 1 - x(1);
{P ∧ Q}
while k <> n do
    begin {P ∧ Q ∧ k ≠ n}
          k := k + 1;
          if x(k) = 1 then c := c + d
                      else d := d + 1
          {P ∧ Q}
    end
{P ∧ Q ∧ k = n, dus de eindconditie}
```

Voor het vinden van een repetitie hebben we nu een systematiek. De repetitie behoeft niet experimenterend vastgesteld te worden. Bij het experimenterend vaststellen van een repetitie is het aantal vrijheden zo groot dat al vlug

fouten worden gemaakt, die dan opgespoord moeten worden door het algoritme uit te voeren als een computer.

Tenslotte het formele nivo.
Op dit nivo wordt met de uitspraken formeel gemanipuleerd en de eigenschappen van de ontwikkelde algoritmen worden formeel afgeleid (bewezen). We zullen hier alleen een indruk geven van deze aanpak, men zij verder verwezen naar (7).
Eerst wordt de formele specificatie van het probleem gegeven:

```
S: |[ n: int {n ≥ 1}
   ; |[ x(p: 0 ≤ p < n): array of int
     {(A p: 0 ≤ p < n: x(p) = 0 ∨ x(p) = 1)}
     ; |[ c: int
       ; S
         {c = (N i,j: 0 ≤ i < j < n: x(i) = 0 ∧ x(j) = 1)}
       ]|
     ]|
   ]|
```

Eerst wordt getracht een invariant te vinden. Een mogelijkheid voor het vinden van een invariant is het vervangen in de eindconditie van een constante door een variabele en het opgeven van het bereik van deze variabele. We kiezen als invariant P:

$$c = (\underline{N}\, i,j: 0 \leq i < j < k: x(i) = 0 \wedge x(j) = 1) \wedge 0 \leq k \leq n$$

De eindconditie geldt als deze invariant geldt en bovendien k = n. De eindigheid van de repetitie kan gegarandeerd worden door in de repetitie bij k 1 op te tellen. Na afloop van de statement(s) in de repetitie moet P weer gelden. Onder welke voorwaarde mag k met 1 opgehoogd worden zodanig dat na deze ophoging P weer geldt. Uit de regel voor de assignment statement kan worden afgeleid dat dan voor de ophoging van k met 1 moet gelden:

$$c = (\underline{N}\, i,j: 0 \leq i < j < k+1: x(i) = 0 \wedge x(j) = 1) \wedge 0 \leq k+1 \leq n$$

Vòòr de te herhalen statement(s) in de repetitie geldt P en $k \neq n$, het tweede deel van de conjunctie is dus zeker waar. We herschrijven het eerste deel van de conjunctie als:

$$c = (\underline{N}\, i,j: 0 \leq i < j < k: x(i) = 0 \wedge x(j) = 1) + (\underline{N}\, i: 0 \leq i < k: x(i) = 0 \wedge x(k) = 1)$$

De gewenste relatie kan kennelijk gerealiseerd worden als de invariant geldt en we de beschikking hebben over het tweede deel van de bovenstaande som. We breiden daarom de invariant P uit tot een invariant die we R zullen noemen:

$$c = (\underline{N}\, i,j: 0 \leq i < j < k:\ x(i) = 0 \wedge x(j) = 1) \ \wedge\ d = (\underline{N}\, i: 0 \leq i < k:\ x(i) = 0)$$
$$\wedge\ 0 \leq k \leq n$$

Nu kan het algoritme gegeven worden. Er moet nog wel het een en ander aangetoond worden. De betreffende plaatsen in het algoritme worden aangegeven met "{noot ...}" en er wordt aan deze bewijsverplichtingen voldaan. We zullen hier na het algoritme alleen de bewijsverplichtingen opsommen, de bewijzen zelf zullen we niet geven.
De oplossing voor S luidt:

```
|( k, d: int
; k, c, d := 0, 0, 0
 {noot 1, noot 2}
; do k ≠ n → if x(k) = 0 → d := d + 1
            [] x(k) = 1 → c := c + d
           fi {noot 3}
         ; k := k + 1
  od
{noot 4}
)|
```

Bewijsverplichtingen:

noot 1: Na de initialisatie geldt de invariant.

noot 2: De eindigheid van de repetitie moet aangetoond worden. Hiertoe wordt een zogenaamde invariante functie gezocht waarvan wordt aangetoond dat deze een ondergrens heeft en dat deze in iedere slag van de repetitie kleiner wordt. (In het voorbeeld is de variante functie n-k met als ondergrens 0.)

noot 3: Na elk van de alternatieven geldt een uitspraak die na de ophoging van k met 1 de invariant garandeert.

noot 4: De invariant R samen met k = n garandeert de gewenste eindconditie.

In het vervolg komt alleen nog het semi-formele nivo aan de orde. Daarmee wil niet gesuggereerd zijn dat het formele nivo onbelangrijk zou zijn. Voor

het formele nivo zijn hier guarded commands gebruikt, we zouden ook Pascal hebben kunnen gebruiken.

HET SEMI-FORMELE NIVO. NOG TWEE VOORBEELDJES

Eerste voorbeeld.
De arrays

```
var f: array (0..p)of integer;
    g: array (0..q)of integer;
```

(waarin p en q constanten zijn, ≥ 0) hebben een waarde waarvoor geldt:
- voor alle i, $0 \leq i < p$, geldt $f(i) < f(i+1)$
- voor alle i, $0 \leq i < q$, geldt $g(i) < g(i+1)$

Schrijf een algoritme dat het aantal (i,j)-paren, met $0 \leq i \leq p$ en $0 \leq j \leq q$, bepaalt waarvoor geldt $f(i) = g(j)$.

Als je voor het vinden van een invariant een constante uit de eindconditie, zeg n, vervangt door een variabele, zeg m, dan drukt de invariant uit dat het probleem is opgelost tot (en met) m. Door m op te hogen en er voor te zorgen dat de invariant blijft gelden, wordt er op een gegeven moment voor gezorgd dat het probleem is opgelost tot (en met) n. We kiezen nu een andere soort invariant, die uitdrukt dat het gewenste resultaat gelijk is aan de waarde van een variabele plus een waarde die nog berekend moet worden. Concreet, als invariant kiezen we:

"het aantal (i,j)-paren, $0 \leq i \leq p$ en $0 \leq j \leq q$, met $f(i) = g(j)$" =
k + "het aantal (i,j)-paren, $m \leq i \leq p$ en $n \leq j \leq q$, met $f(i) = g(j)$"

Uit deze invariant P en $m = p + 1$ of $n = q + 1$ volgt de eindconditie. De invariant is gemakkelijk te realiseren vóór de repetitie door aan de variabelen k, m en n de waarde 0 toe te kennen. We hebben dan gevonden:

```
k := 0; m := 0; n := 0;
{P}
while (m ≠ p+1) and (n ≠ q+1) do {P ∧ m ≠ p+1 ∧ n ≠ q+1} ...
```

De eindigheid van de repetitie is te garanderen door m of n met 1 op te hogen. Er doen zich drie gevallen voor:
- $f(m) < g(n)$: m is op te hogen en invariant blijft gelden;
- $f(m) < g(n)$: n is op te hogen en invariant blijft gelden;
- $f(m) = g(n)$: ophoging van k, m en n zorgt ervoor dat de invariant blijft gelden.

Zo hebben we gevonden:

```
{beginconditie}
k := 0; m := 0; n := 0;
{P}
while (m <> p+1) and (n <> q+1)
    do {P ∧ m ≠ p+1 ∧ n ≠ q+1}
       if f(m) < g(n)
         then m := m + 1
         else if f(m) > g(n)
                then n := n + 1
                else begin k := k + 1; m := m + 1; n := n + 1 end;
       {P}
{P ∧ (m = p+1 ∨ n = q+1), dus de eindconditie}
```

Tweede voorbeeld.
De arrays

```
var f: array (1..m)of integer;
    g: array (1..p)of integer;
    h: array (1..n)of integer;
```

(m, p en n constanten, ≥ 1) hebben waarden waarvoor geldt:

- voor alle i, $1 \leq i \leq m$, $f(i) < f(i+1)$;
- voor alle i, $1 \leq i \leq p$, $g(i) < g(i+1)$;
- voor alle i, $1 \leq i \leq n$, $h(i) < h(i+1)$;
- er bestaan een i, $1 \leq i \leq m$, een j, $1 \leq j \leq p$, en een k, $1 \leq k \leq n$, waarvoor geldt $f(i) = g(j) = h(k)$.

Schrijf een algoritme dat de plaats bepaalt van de kleinste gemeenschappelijke waarde van de drie arrays.

Als we met i', j' en k' de indices aangeven van de kleinste gemeenschappelijke waarde dan is de eindconditie:

$$i = i' \wedge j = j' \wedge k = k'$$

Als invariant proberen we P:

$$1 \leq i \leq i' \wedge 1 \leq j \leq j' \wedge 1 \leq k \leq k'$$

Uit deze invariant volgt de eindconditie als geldt: $f(i) = g(j) \wedge g(j) = h(k)$
De invariant is waar te maken door voor i, j en k de waarde 1 te nemen.

De eindigheid van de repetitie is te garaderen door i, j of k met 1 op te hogen. De ophoging van i met 1 is mogelijk als f(i) < g(j) of f(i) < h(k) geldt. Analoog voor j en k. We krijgen dan voor de ophogingen:

```
if (f(i) < g(j)) or (f(i) < h(k))
  then i := i + 1
  else if (g(j) < f(i)) or (g(j) < h(k))
         then .....
         else .....
```

Maar de voorwaarde f(i) < g(j) alleen is natuurlijk al voldoende om i met 1 op te kunnen hogen. Analoog voor j en k. Zo krijgen we als algoritme:

```
i := 1; j := 1; k := 1;
{P}
while not((f(i) = g(j)) and (g(j) = h(k)))
    do {P ∧ (f(i) ≠ g(j) ∨ g(j) ≠ h(k))}
       if f(i) < g(j)
         then i := i + 1
         else if g(j) < h(k)
                then j := j + 1
                else {f(i) ≧ g(j) ≧ h(k)} k := k + 1
       {P}
{P ∧ f(i) = g(j) ∧ g(j) = h(k)}
```

EEN NIET ZO EENVOUDIGE OPGAVE MET EEN FRAAIE OPLOSSING (EN EEN FRAAIE AFLEIDING)

De arrays

```
var a, b: array (0..n-1) of t
```

hebben een waarde; op het type t is een ordening gedefinieerd.
Gevraagd wordt een algoritme te schrijven dat nagaat of de twee arrays, eventueel op een cyclische rotatie na, gelijk zijn. Anders gezegd: Zijn er een i en een j te vinden zodanig dat a(i) = b(j), a((i+1) mod n) = b((j+1) mod n), ..., a((n+1-i) mod n) = b((n+1-j) mod n).
We spreken af dat we nu verder in de afleiding de modulus weg zullen laten en deze pas weer in het uiteindelijke algoritme zullen toevoegen.
We definiëren voor i = 0,1,2,...,n-1 de rij a_i als a_i = (a(i), a(i+1), ..., a(i+n-1)). Zo ook voor het array b. We kunnen de eindconditie nu formuleren

als:

$$eq \equiv (\underline{E}\, i,j: 0 \leqq i < n \wedge 0 \leqq j < n: a_i = b_j)$$

Er is natuurlijk een heel eenvoudige oplossing. Er zijn n a-rijen en n b-rijen. We kunnen elke a-rij vergelijken met elke b-rij en kijken of er een keer gelijkheid optreedt. Maar dat geeft n^2 combinatiemogelijkheden. We willen kijken of het niet zuiniger kan.

De verzameling van alle a-rijen en de verzameling van alle b-rijen zijn òf gelijk òf disjunct. We zouden dus het antwoord op de gestelde vraag kunnen geven door de minima van de twee verzamelingen te vergelijken. (Voor de rijen houden we de lexicografische volgorde aan.)

Het probleem is symmetrisch in a en b. We zullen kijken of we in de oplossing de symmetrie kunnen vasthouden. Laat je de symmetrie los, dan krijg je waarschijnlijk een slechtere oplossing.

Stel dat n de waarde 8 heeft en dat de waarden van a en b zijn:

a: 1 1 8 4 1 8 4 1

b: 1 8 4 1 1 1 8 4

En stel dat we de rijen a_2 en b_1 aan het vergelijken zijn. Dan vinden we in eerste instantie drie gelijke waarden. Van deze kennis zullen we proberen te profiteren en niet zonder meer overgaan op de vergelijking van andere rijen. Dit suggereert dat we als invariant zullen nemen (noem die P):

$$(\underline{A}\, k: 0 \leqq k < h: a(i+k) = b(j+k)) \wedge 0 \leqq h \leqq n$$

Als we in staat zijn deze uitspraak invariant te houden en tegelijk ook h gelijk aan n te maken, hebben we het antwoord op de gestelde vraag gevonden. Als $a(i+h)$ gelijk is aan $b(j+h)$ kan h met 1 opgehoogd worden waarbij P invariant blijft.

Als de elementen $a(i+h)$ en $b(j+h)$ niet gelijk zijn dan weten we dat $a_i \neq b_j$, $a_{i+1} \neq b_{j+1}$, ..., $a_{i+h} \neq b_{j+h}$. Als we nu ook nog weten dat het element uit a groter is dan het element uit b dan weten we dat elke genoemde a-rij lexicografisch groter is dan de overeenkomstige b-rij. En dus dat elke genoemde a-rij groter is dan de lexicografisch kleinste b-rij (en we hebben al gezegd dat we de gelijkheid konden constateren door de lexicografisch kleinste rijen te vergelijken). We breiden daarom de invariant P uit met de stukken Q_a en Q_b:

$$Q_a: (\underline{A}\, k: 0 \leqq k < i: a_k > bb)$$

$$Q_b: (\underline{A}\, k: 0 \leqq k < j: b_k > aa)$$

waarin aa en bb staan voor de lexicografisch kleinste a-rij respectievelijk

de lexicografisch kleinste b-rij en waarin het groterteken staat voor "lexicografisch groter".
Uit Q_a volgt dat we het antwoord weten als i gelijk is aan n (ongelijkheid!), zo ook voor Q_b en $j = n$.
Hierboven stond dat we uit het feit dat $a(i+h)$ lexicografisch groter is dan $b(j+h)$ en de invariant P konden concluderen dat $a_i > bb$, $a_{i+1} > bb$, ..., $a_{i+h} > bb$. Samen met invariant Q_a volgt hieruit:

$$(\underline{A}\, k: 0 \leq k < i+h+1: a_k > bb)$$

Maar dat betekent dan weer dat i opgehoogd kan worden met h + 1 zonder dat invariant Q_a verstoord wordt.
Een analoog verhaal geldt voor Q_b als $a(i+h)$ lexicografisch kleiner is dan $b(j+h)$. Voor de invariant P betekent het wel dat deze bij de nieuwe waarde van i of j alleen nog maar kan gelden voor $h = 0$
We kunnen de invarianten P, Q_a en Q_b vóór de repetitie waar maken door voor de variabelen h, i en j de waarde 0 te nemen.
Zo hebben we gevonden:

```
h := 0; i := 0; j := 0;
{P ∧ Qa ∧ Qb}
while (h < n) and (i < n) and (j < n)
    do {P ∧ Qa ∧ Qb ∧ h < n ∧ i < n ∧ j < n}
       if a(i+h) = b(j+h)
         then h := h + 1
         else if a(i+h) > b(j+h)
                then begin i := i + h + 1;
                           h := 0
                     end
                else begin j := j + h + 1;
                           h := 0
                     end
       {P ∧ Qa ∧ Qb};
{P ∧ Qa ∧ Qb ∧ (h ≧ n ∨ i ≧ n ∨ j ≧ n)}
if h>=n then {P ∧ h ≧ n} b := true
        else (Qa ∧ i ≧ n) ∨ (Qb ∧ j ≧ n)} b := false
{b ≡ (E i,j: 0 ≦ i < n ∧ 0 ≦ j < n: a_i = b_j)}
```

Het naïeve algoritme onderzocht n^2 combinaties van a- en b-rijen. Het boven-

staande algoritme vergt ten hoogste 3n - 2 slagen van de in het algoritme voorkomende repetitie.
Het algoritme is van Y. Shiloach. De hier gepresenteerde afleiding is gebaseerd op een presentatie van het algoritme door mevr. drs. A.J.M. van Gasteren en ir. W.H.J. Feijen.

EEN COMPUTER OP SCHOOL

Het laatste jaar zijn op een honderdtal scholen computers geplaatst door het ministerie. Met zo'n computer kun je verschillende zaken doen. Allereerst kun je hem gebruiken voor de schooladministratie. Ik kan me nauwelijks voorstellen dat dat het belangrijkste doel was van het plaatsen van deze machines. De computer kan gebruikt worden om de lessen in andere vakken te ondersteunen. Maar dan moet je wel beschikken over de juiste programma's. En van een aantal leraren heb ik begrepen dat die er nu juist niet zijn. Je kunt de leerlingen ook programmaatjes laten schrijven in BASIC om "de angst voor de machine weg te nemen". Hier moeten we wel mee oppassen! Het blijkt dat mensen die op deze manier met BASIC gespeeld hebben daar last van hebben als ze leren programmeren op de hier bedoelde manier. Bovendien wordt het spelen maar al te vaak verward met programmeren. Maar voor programmeeronderwijs is nauwelijks een computer nodig, alhoewel het verwerken op een machine van het afgeleide programma wel stimulerend kan werken.

EPILOOG (een citaat uit (1))

We are like the barber-surgeons of earlier ages, who pride themselves on the sharpness of their knives, and the speed with which they can dispatch their duties, either of shaving a beard or amputation of a limb. Imagine the dismay with which they greeted some ivory-towered academic who told them that the practice of surgery should be based on a long and detailed study of human anatomy, on familiarity with surgical procedures pioneered by great doctors of the past, and that it should be carried out only in a strictly controlled bug-free environment, far removed from the hair and dust of the normal barber's shop. Even if they accepted the validity and necessity for these improvements, how are they ever to achieve them?

LITERATUUR

(1): C.A.R. Hoare
Professionalism
in: Computer Bulletin, September 1981

(2): D.E. Knuth
The art of computer programming
Volume 1: Fundamental Algorithms
Volume 2: Seminumerical Algorithms
Volume 3: Sorting and Searching

Het zijn geen programmeerboeken maar ze vormen een soort encyclopedie. De algoritmes worden gegeven in een soort symbolische machinetaal. Ze bevatten een schat aan informatie over algoritmes.

(3): E.W. Dijkstra
A discipline of programming
Prentice-Hall

Een zeer goed, maar geavanceerd, boek. Bevat prachtige algoritmen met hun afleidingen. Guarded commands worden gebruikt. De betekenis, het effect van de constructies, wordt gedefinieerd met behulp van weakest preconditions.

(4): D. Gries
The science of programming
Springer-Verlag

Het boek behandelt eerst een stuk predicatenrekening. Daarna wordt het programmeren op een formele manier behandeld. In tegenstelling tot (3) is dit een leerboek; het vormt een goede inleiding voor (3).

(5): J.C. Reynolds
The craft of programming
Prentice-Hall

Het eerste hoofdstuk van het boek behandelt het programmeren op het semi-formele nivo. Sluit verder niet zo aan bij hetgeen hier behandeld is. De algoritmen worden genoteerd in de programmeertaal ALGOL W.

(6): E.W. Dijkstra en W.H.J. Feijen
Een methode van programmeren
Academic Service

Het programmeren wordt op het formele nivo behandeld. Het kan gebruikt worden als een inleiding voor (3). In het boek wordt ook de benodigde predicatenrekening behandeld. Het bevat een groot aantal opgaven.

(7): J.J. van Amstel
Programmeren: Het ontwerpen van algoritmen
Academic Service

Het programmeren wordt behandeld op het hier beschreven semi-formele nivo. De gebruikte programmeertaal is Pascal. De nadruk ligt op het ontwerpen van algoritmen niet op Pascal. Van Pascal worden die constructies behandeld die voldoende zijn om de gekozen aanpak van het programmeren te demonstreren.

Excursie naar het Moeras van het Onberekenbare

Jan Heering
Centrum voor Wiskunde en Informatica
Amsterdam

1. Een onmogelijk programma

"*A well-known piece of folk-lore among programmers holds that it is impossible to write a program which can examine any other program and tell, in every case, if it will terminate or get into a closed loop when it is run. I have never actually seen a proof of this in print, and though Alan Turing once gave me a verbal proof (in a railway carriage on the way to a Conference at the National Physical Laboratory in 1953), I unfortunately and promptly forgot the details. This left me with an uneasy feeling that the proof must be long or complicated, but in fact it is so short and simple that it may be of interest to casual readers.*" Aldus Christopher Strachey in een ingezonden brief in *The Computer Journal* van januari 1965.

Al heeft de duidelijkheid van zijn uitleg misschien te wensen overgelaten, toch had Strachey in die treincoupé geen betere leermeester kunnen treffen dan Turing, die dit "onmogelijke programma" ruim vijftien jaar eerder zelf had ontdekt. Ongelukkigerwijs is het "korte en eenvoudige" bewijs dat Strachey in zijn brief geeft niet correct, maar het idee dat erachter zit (of dat ik erachter gezocht heb) is niettemin bruikbaar.

In het vervolg gebruik ik een informele programmeertaal L, waarvan de betekenis grotendeels duidelijk is en die slechts op een enkel punt toelichting vereist. Mocht u behoefte hebben aan meer concreetheid dan kunt u alles omzetten naar uw favoriete programmeertaal.

Stel dat $T(p)$ een in L geschreven programma is, dat aan *elk* L-programma p de waarde **true** of **false** toekent al naar gelang p bij uitvoering stopt of niet. T werkt alleen op *gesloten* L-programma's p, dat wil zeggen programma's zonder oningevulde argumenten, vrije variabelen of invoer-statements, want dat zijn de enige programma's die zonder meer uitvoerbaar zijn. T kan zijn argument inspecteren, modificeren, geheel of gedeeltelijk interpreteren, kortom er alles mee doen wat in L uitdrukbaar is, mits hij uiteindelijk maar stopt en de juiste waarheidswaarde oplevert.

De kracht van T is enorm. Neem bijvoorbeeld Goldbach's vermoeden, dat elk even getal (>2) de som is van twee priemgetallen. "*Dass ein jeder numerus par eine summa duorum primorum sey, halte ich für ein ganz gewisses theorema, ungeachtet ich dasselbe nicht demonstriren kann*" schrijft Euler in juni 1742 aan Goldbach. Sindsdien is het vermoeden bewezen noch weerlegd, maar, als het onwaar is, is het eenvoudig te weerleggen omdat in dat geval het volgende

programmaatje G het kleinste tegenvoorbeeld n_0 oplevert:

```
program G
{
 for n =4,6,8, . . .
 do
   if sum-of-two-primes(n)
    then continue
    else n0 := n ; return(n0)
   fi
 od
}
```

G gebruikt de functie *sum-of-two-primes*(k), die **true** of **false** oplevert al naar gelang k de som is van twee priemgetallen of niet. De definitie van *sum-of-two-primes*(k) is eenvoudig. Alleen de priemgetallen $\leqslant k/2$ hoeven te worden geprobeerd.

G is slechts een halve beslissingsprocedure, want als het vermoeden waar is komt G nooit met een resultaat maar blijft eindeloos doorrekenen. Dit onvermogen van G om tot een positieve beslissing te komen kan ondervangen worden door niet G maar $T(G)$ uit te voeren. $T(G)$ stopt immers altijd! In het volgende tabelletje ziet u een en ander naast elkaar:

Vermoeden van Goldbach	G	$T(G)$
Waar	Stopt niet	**false**
Onwaar	n_0	**true**

$T(G)$ kan dus het vermoeden van Goldbach beslissen, al valt er niets te zeggen over de daarvoor benodigde tijdsduur. Het enige wat zeker is, is dat er een antwoord komt. Dat is echter nog niet alles, want andere beroemde onopgeloste problemen zoals het probleem van Fermat en de Riemann-hypothese zijn, als ze onwaar zijn, op dezelfde wijze stelselmatig te weerleggen als het vermoeden van Goldbach, zodat ook hier toepassing van T tot een beslissing leidt. Zoals u ziet is T een zeer aantrekkelijk programma, maar gezien de problemen die het kan oplossen moet het wel buitengewoon ingewikkeld of zelfs onmogelijk zijn.

Dat dit laatste het geval is laat zich als volgt inzien. Als T bestaat, dan is er een programma P waarvan T onderdeel uitmaakt en dat eerst T aanroept met zichzelf (dat wil zeggen zijn eigen programmatekst, inclusief de tekst van T) als argument. Maar als P de voorspelling van T omtrent zijn eigen afloop kent, als P weet of hij volgens T zal stoppen of niet, kan hij dan niet besluiten om juist het tegenovergestelde te doen van wat T voorspeld heeft?

Een P die dat doet ziet er in eerste aanzet zo uit:

```
program P
{
 « definitie van T »;
 if T("program P  { « definitie van T » ; if . . . fi }")
  then loop: go to loop
  else return
 fi
}
```

Het argument van T is de tekst van P, dat is althans de bedoeling. Layout doet in L niet ter zake. Teksten staan tussen dubbele aanhalingstekens en deze mogen ook genest binnen teksten voorkomen. Zoals u ziet werkt deze eenvoudige opzet niet, want de stippeltjes corresponderen ten gevolge van de optredende oneindige regressie met een oneindig lange tekst.

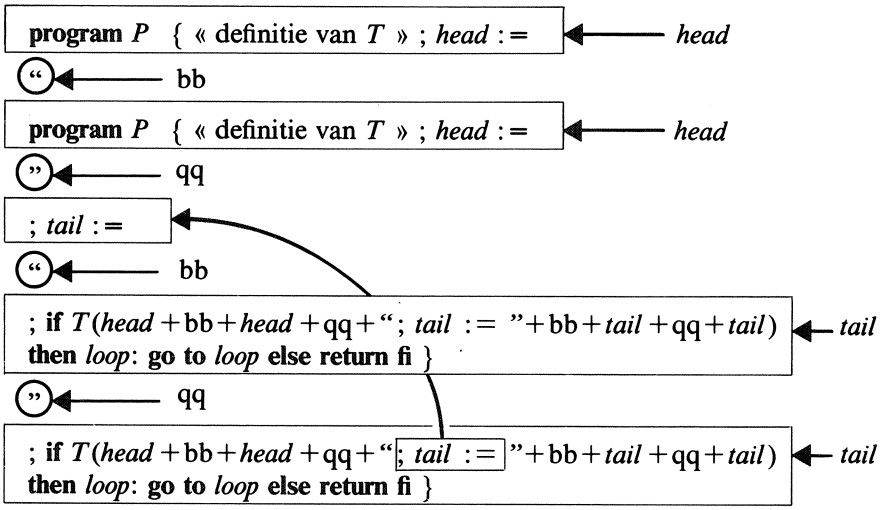
In de volgende versie* wordt de oneindige regressie omzeild door twee variabelen *head* en *tail* in te voeren die elk een gedeelte van de (nieuwe!) tekst van P als waarde krijgen en die gebruikt worden in het argument van T in plaats van de tekst van P zelf. De +-operator concateneert twee teksten ("con" + "catenatie" = "concatenatie") en bb en qq zijn twee tekstconstanten met als waarden het "dubbele aanhalingsteken openen" respectievelijk het "dubbele aanhalingsteken sluiten".

```
program P
{
 « definitie van T » ;
 head := " program P  { « definitie van T » ; head := " ;
 tail  := " ; if T(head+bb+head+qq+"; tail := "+bb+tail+qq
                  +tail) then loop: go to loop else return fi } " ;
 if T(head+bb+head+qq+"; tail := "+bb+tail+qq+tail)
  then loop: go to loop
  else return
 fi
}
```

Op de volgende pagina ziet u dezelfde programmatekst nogmaals, maar nu in een andere layout en met doosjes die aangeven hoe het argument van T evalueert naar de volledige tekst van P.

---

* Deze versie van P is gemaakt door Lambert Meertens als reactie op een niet erg geslaagde poging van mij om Strachey's bewijs te repareren.



Resumerend: P vraagt aan T zijn eigen afloop te voorspellen, maar doet, als T zijn voorspelling gedaan heeft, precies het tegenovergestelde. Daarmee is aangetoond dat een T die van *alle* programma's kan beslissen of ze wel of niet stoppen niet kan bestaan.

Welke gevolgen heeft het niet bestaan van T voor het vermoeden van Goldbach? Geen enkel. De halve beslissingsprocedure G blijft van kracht, maar de beslissingsprocedure $T(G)$ niet. T was echter zeer krachtig en kon veel meer bewijzen dan alleen het vermoeden van Goldbach. Dat werd T's ondergang, maar misschien is voor het bewijs van het vermoeden van Goldbach zoveel kracht helemaal niet nodig. Er zijn alles bij elkaar drie mogelijkheden:

- Het vermoeden is niet waar. In dat geval is het te weerleggen met G.
- Het vermoeden is waar en bewijsbaar op grond van algemeen geaccepteerde axioma's en bewijsregels.
- Het vermoeden is waar, maar niet bewijsbaar op grond van algemeen geaccepteerde axioma's en bewijsregels. In dat geval kan het vermoeden alsnog bewijsbaar blijken te zijn als nieuwe axioma's of bewijsregels ontdekt worden en geaccepteerd raken (maar dat gebeurt niet dagelijks).

Welke van de drie mogelijkheden zich in feite voordoet is een open vraag.

2. De Church-Turing these

Turing's oorspronkelijke bewijs van de onmogelijkheid van T is te vinden in zijn monumentale (maar niet gemakkelijk leesbare) artikel "On computable numbers with an application to the Entscheidungsproblem" van 1937 [1]. Toen Turing dit artikel publiceerde bestonden er nog helemaal geen "echte" computers. Hij introduceerde ze in abstracte en geïdealiseerde vorm om een precieze, bij de intuïtie aansluitende, inhoud te kunnen geven aan het begrip *effectieve procedure*, dat in het kader van het wiskundig grondslagenonderzoek om opheldering vroeg.

Turing deed daarbij een verrassende ontdekking, waarmee ieder die met

computers werkt vertrouwd is. Sommige Turing-machines (zoals zijn geïdealiseerde computers genoemd worden) bleken namelijk *universeel* te zijn, in die zin dat ze zo geprogrammeerd kunnen worden dat ze het gedrag van een willekeurige andere Turing-machine nabootsen. Een universele Turing-machine is niet veel anders dan een gewone computer, maar met een potentieel onbeperkte hoeveelheid geheugen zodat berekeningen nooit ten gevolge van geheugengebrek gestaakt hoeven te worden.

In de praktijk manifesteert het universele karakter van computers zich op vele manieren. De primitieve Intel 8080 microprocessor is, *afgezien van rekensnelheid en geheugencapaciteit*, precies even krachtig als de Cray-1 "supercomputer". De Cray-1 is immers simuleerbaar op de 8080 en omgekeerd. Anders gezegd: wat zich in 8080 machinetaal laat uitdrukken, laat zich ook in Cray-1 machinetaal uitdrukken en vice versa. Hetzelfde geldt voor hogere programmeertalen, zodat bijvoorbeeld FORTRAN niet essentieel krachtiger is dan 8080 machinetaal. Om die reden is het mogelijk FORTRAN naar 8080 machinetaal te vertalen.

Het bestaan van universele computers is volstrekt niet vanzelfsprekend en nog minder vanzelfsprekend is het dat universele computers geen bijzonder grote instructieset hoeven te hebben. Iemand die niets van computers weet zal, geloof ik, geneigd zijn te denken dat er òf steeds een essentiële toename in berekeningsmacht te verkrijgen is door nieuwe instructies aan de instructieset van een computer toe te voegen òf dat universele berekeningsmacht pas bij zeer grote instructiesets kan optreden. Geen van beide is echter het geval.

Niet lang voor het verschijnen van Turing's artikel was Church met het voorstel gekomen effectieve berekenbaarheid te definiëren als "Herbrand-Gödel general recursiveness" of "λ-definability", twee begrippen die ik hier verder onverklaard laat. Kleene had de equivalentie van deze twee, langs verschillende weg tot stand gekomen, definities aangetoond, maar niettemin stuitte Church's voorstel op verzet van Gödel en Post, die meenden dat effectieve berekenbaarheid niet "zo maar" gedefinieerd kon worden, maar dat een diepergaande analyse op basis van de algemeen aanvaarde eigenschappen van het begrip noodzakelijk was. Deze analyse werd geleverd door Turing, die bovendien aantoonde dat het uit zijn analyse voortvloeiende begrip van effectieve berekenbaarheid equivalent was met dat van Church.

Het samenvallen van deze drie verschillende definities leidde tot de algemene overtuiging dat de essentie van de begrippen "effectieve berekenbaarheid" en "effectieve procedure" hiermee onderkend was. Gödel en vooral Post bleven echter van mening dat er in dit geval geen sprake kon zijn van een gewone definitie, maar dat het om een nieuwe *natuurwet* ging die voortdurend getoetst zou moeten worden. Om dit aspect te benadrukken werd en wordt de uitkomst van deze opmerkelijke reeks analyses - waarover u meer kunt lezen in een interessant artikel van Davis [2] - geformuleerd als een these:

(Church-Turing these). *Elke effectieve procedure is programmeerbaar op een vaste willekeurig gekozen universele Turing-machine.*

Als het inderdaad een natuurwet is, in welke verhouding staat de Church-

Turing these dan tot de fysische wetten zoals die op het ogenblik bekend zijn? Aan het slot van zijn bijdrage aan de in 1981 gehouden conferentie over "Physics of Computation" vraagt Chaitin zich af: "*Is there a physical phenomenon that computes something noncomputable? Contrariwise, does Turing's thesis that anything computable can be computed by a Turing machine constrain the physical universe we are in?*" [3]. Deze vraag voert direct naar moeilijke problemen in de quantummechanica en in het bijzonder naar het befaamde "hidden variable problem". Als er, zoals algemeen wordt aangenomen, geen hidden variables zijn, dan is er een fundamentele onvoorspelbaarheid in de natuur waar geen enkel programma tegen opgewassen is [4]. Dat zou betekenen dat het eerste deel van Chaitin's vraag bevestigend beantwoord moet worden. Daarbij moet echter direct worden aangetekend dat, hoewel het onlangs gelukt is de onmogelijkheid van *lokale* hidden variable theorieën aan te tonen, de onmogelijkheid van het bestaan van hidden variables in het algemeen nooit te bewijzen is.

3. De busy beaver functie

In het vervolg neem ik aan dat in L gerekend kan worden met natuurlijke getallen in de gewone decimale representatie en verder dat *alle* berekeningen met natuurlijke getallen in L uitdrukbaar zijn. Dit laatste is geen bijzonder zware eis, want, zoals ik in de vorige paragraaf uiteengezet heb, geldt dit - als de Church-Turing these waar is - voor alle zichzelf respecterende programmeertalen.

De "busy beaver functie" laat zich eenvoudig definiëren: BB kent aan elk natuurlijk getal l het grootste getal toe dat door een L-programma met een lengte van minder dan l symbolen (letters, cijfers, etc.) berekend wordt. Strikt genomen is $BB(l)$ alleen gedefinieerd voor voldoend grote waarden van l, want in het begin kan het gebeuren dat de verzameling L-programma's met lengte $<l$ leeg is of alleen programma's bevat die niet stoppen of die geen natuurlijk getal als resultaat opleveren.

BB is net zomin berekenbaar als T uit §1. Met andere woorden, er is geen L-programma dat voor *elk* (voldoende groot) natuurlijk getal l de bijbehorende busy beaver waarde berekent. U ziet al aan de definitie dat BB een zeer snel stijgende functie moet zijn. Dit is een understatement, want BB stijgt onvoorstelbaar snel, sneller dan welke berekenbare functie ook. Neem maar een L-programma P dat een natuurlijk getal n als argument heeft en dat, als het stopt, ook weer een natuurlijk getal oplevert. Bij gegeven argumentwaarde n heeft $P(n)$ een totale lengte die de som is van de lengte van het eigenlijke programma P + de lengte van de decimale representatie van n. Deze laatste groeit echter slechts logaritmisch, zodat er een (van P afhankelijke) constante n_0 is met

$$lengte(P(n)) = lengte(P) + \lfloor {}^{10}\log n \rfloor + 1 < n \text{ voor } n > n_0(P).$$

Op voorwaarde dat $P(n)$ stopt voor de gegeven argumentwaarde geldt dus

$$P(n) \leqslant BB(n) \text{ voor } n > n_0(P).$$

Als nu BB in L uitdrukbaar was, dan zou ook BB' met $BB'(n)=BB(n)+1$ in L uitdrukbaar zijn. Maar daaruit zou volgen

$$BB(n)+1 = BB'(n) \leqslant BB(n) \text{ voor } n > n_0(BB')$$

en dit is een tegenspraak.

Impliceert de onberekenbaarheid van BB nu ook dat *individuele* busy beaver waarden niet berekenbaar zijn? In zekere zin niet, want er is voor elke busy beaver waarde natuurlijk altijd een programmaatje dat die waarde als constante bevat en deze eenvoudigweg afdrukt. Dat betekent echter niet dat iemand dat programmaatje ook werkelijk kan schrijven! Dat hangt af van de vraag of individuele waarden van BB op een of andere manier te bepalen zijn. Rado, die de busy beaver functie heeft uitgevonden, heeft het samen met Lin geprobeerd en is er inderdaad in geslaagd (op basis van een iets andere definitie van BB dan de hier gebruikte) $BB(l)$ te bepalen voor enkele kleine waarden van l [5]. Voor grotere waarden van l zagen zij er geen gat meer in. De reden daarvan is niet moeilijk te begrijpen:

Kies een vaste l en voer alle L-programma's met lengte $<l$ parallel uit, elk programma op zijn eigen computer. (Het kan ook pseudoparallel op één computer.) Als een van de programma's na enige tijd stopt kan de afgeleverde waarde vergeleken worden met de resultaten van al eerder gestopte programma's. Dat is niet het probleem. Het probleem zijn de programma's waarvan niet duidelijk is of ze ooit zullen stoppen of niet. Voor voldoend grote l is bijvoorbeeld onder de programma's met lengte $<l$ ook G uit §1. In dat geval is dus voor de bepaling van $BB(l)$ inzicht in het vermoeden van Goldbach nodig. Immers, als het vermoeden waar is dan heeft G geen invloed op $BB(\textit{lengte}(G)+1)$, maar als het onwaar is dan is $BB(\textit{lengte}(G)+1)$ tenminste gelijk aan het door G geleverde kleinste tegenvoorbeeld n_0.

De enige informatie die nodig is om $BB(l)$ te kunnen bepalen is het aantal programma's met lengte $<l$ dat stopt. Gewapend met deze kennis is het mogelijk het boven beschreven proces van parallelle uitvoering te onderbreken op het moment dat het totale aantal programma's dat stopt ook inderdaad gestopt is. De programma's die dan nog lopen stoppen blijkbaar niet en kunnen verder genegeerd worden. Het probleem is dat bepaling van het aantal stoppende programma's met lengte $<l$ al voor kleine waarden van l (afhankelijk van de keuze van L) een wiskundig inzicht vereist waarover niemand beschikt.

4. Algorithmic randomness

Vanuit fundamenteel gezichtspunt is een rij symbolen *random* als er geen wet aan ten grondslag ligt, als er geen regelmaat of structuur in te bekennen valt. Een random rij heeft geen representatie of definitie die korter is dan de rij zelf, met andere woorden hij is *niet comprimeerbaar*. Er zijn geen ezelsbruggetjes die kunnen helpen een random rij te onthouden.

In eerste benadering is een rij symbolen random als er geen definitie van bestaat die korter is dan de rij zelf. Bij nader inzien blijkt dit niet bevredigend, want het begrip "definitie" geeft in zijn algemeenheid aanleiding tot paradoxen en behoeft nadere omschrijving.

Hoe die nadere omschrijving ook uitvalt, de fractie rijen die $\geqslant k$ symbolen te comprimeren zijn zal altijd tenminste exponentieel afnemen met toenemende k. Het is voldoende alleen rijen van 0'en en 1'en (bitstrings) in de beschouwing te betrekken. Er zijn 2^l bitstrings met lengte l, maar niet meer dan

$$1+2+2^2+\cdots+2^{l-k}=2^{l-k+1}-1$$

bitstrings die als k of meer bits kortere definitie kunnen fungeren. De fractie bitstrings die $\geqslant k$ bits te comprimeren zijn is dus hoogstens

$$\frac{2^{l-k+1}-1}{2^l}<2^{-k+1}.$$

Er is altijd tenminste één bitstring met lengte l die helemaal niet te comprimeren is.

Hoe moet nu het begrip "definitie" gepreciseerd worden? Kolmogorov, Martin-Löf en Chaitin nemen in plaats van het vage en paradoxale begrip "definitie" het probleemloze begrip "effectieve definitie" tot uitgangspunt voor hun overwegingen. Een effectieve definitie van een bitstring is een programma dat die string als resultaat oplevert. Het programma drukt de precieze wetmatigheid uit die aan de geproduceerde bitstring ten grondslag ligt en wel op zo'n concrete manier dat voor paradoxen niet te vrezen valt.

In het vervolg neem ik aan dat in L gerekend kan worden met bitstrings en dat L-programma's zelf ook een bitstringrepresentatie hebben. De lengte van een L-programma zal in het vervolg steeds de lengte van de bijbehorende bitstring zijn. De lengte van het *kortste* L-programma dat stopt met een bitstring σ als resultaat heet de *algoritmische informatie-inhoud* of *Kolmogorov-complexiteit* I van σ. Verder heet een bitstring σ *algoritmisch random* als

$$I(\sigma) > \mathit{lengte}(\sigma)-K,$$

waarbij K een van te voren afgesproken constante is die de grens bepaalt tussen wat algoritmisch random is en wat niet. Deze grens is - althans in het geval van eindige bitstrings - enigszins willekeurig en afhankelijk van de eigenschappen van L. Bitstrings die niet random zijn kunnen tenminste K bits gecomprimeerd worden, terwijl random bitstrings juist minder dan K bits te comprimeren zijn.

I is afhankelijk van L. Bij overgang op een andere programmeertaal L' die eveneens universele berekeningsmacht heeft verandert I slechts in beperkte mate. Er is immers een simulator S voor L' in L, zodat bij elk L'-programma p' een L-programma $S(p')$ bestaat dat hetzelfde doet en dat lengte $\mathit{lengte}(S)+\mathit{lengte}(p')$ heeft. Omgekeerd is ook elk L-programma met behulp van een in L' geschreven simulator S' voor L om te zetten in een L'-programma dat hoogstens $\mathit{lengte}(S')$ langer is. Hieruit volgt voor alle σ

$$|I_{\mathrm{L}}(\sigma)-I_{\mathrm{L}'}(\sigma)| \leqslant \mathit{max}(\mathit{lengte}(S),\mathit{lengte}(S')).$$

Voor een gegeven bitstring σ is er altijd een L-programma p dat σ expliciet bevat. Omdat $\mathit{lengte}(p)=\mathit{lengte}(\sigma)+C$, waarbij C alleen van L afhangt, geldt dus voor alle σ dat

$$I(\sigma) \leqslant lengte(\sigma) + C(\mathrm{L}).$$

Martin-Löf heeft aangetoond dat voldoend lange bitstrings die algoritmisch random zijn door alle bestaande en überhaupt denkbare statistische tests als random geclassificeeerd zullen worden. Omgekeerd is het duidelijk dat algoritmisch comprimeerbare strings ook inderdaad niet random zijn in de intuïtieve betekenis van het woord. De programma's die random getallen genereren worden dan ook meestal pseudo-random generators genoemd, omdat de makers wel beseffen dat er iets niet in orde is.

Het eenvoudigste voorbeeld van een bitstring die niet algoritmisch random is, is een rij van l 0'en. Zo'n rij is maximaal comprimeerbaar, want hij wordt geproduceerd door een programma ter lengte $(\lfloor {}^2\log l \rfloor + 1) + c$ en dat is voor voldoend grote l veel kleiner dan $l - K$ hoe groot K ook gekozen is.

In dit extreme geval is de regelmaat van de string in één oogopslag duidelijk, maar dat is een uitzondering. Als u bijvoorbeeld, zonder dat u de oorsprong ervan kent, de eerste millioen bits van de dyadische ontwikkeling van π voorgeschoteld krijgt, zal het u moeite kosten in te zien dat dit een bitstring is die volstrekt niet algoritmisch random is! De bekende statistische tests voor randomness helpen u daarbij niet want die zijn te oppervlakkig en niet in staat de diepliggende regelmaat die in deze string aanwezig is te ontdekken.

Als een string σ niet algoritmisch random is, dan is dat (althans in theorie) te bewijzen door alle programma's met lengte $\leqslant lengte(\sigma) - K$ parallel uit te voeren (net als bij de busy beaver functie) en te wachten tot er een programma stopt met resultaat σ. Het eerste programma dat σ oplevert hoeft niet het kortste te zijn, maar dat doet er niet toe want het is niet nodig $I(\sigma)$ precies te kennen om te kunnen concluderen dat σ niet random is.

Bewijzen dat een string random is lukt in het algemeen niet. Dit stuit voor een gegeven waarde van σ op dezelfde problemen als bepaling van een individuele busy beaver waarde. Alleen gaat het nu niet om de vraag wat het grootste getal is dat door een programma met lengte $< l$ geproduceerd wordt, maar om de vraag of er een programma met lengte $\leqslant lengte(\sigma) - K$ is dat σ als resultaat heeft. In essentie maakt dit echter geen verschil en randomness-bewijzen vereisen al voor korte strings een bovenmenselijk mathematisch inzicht. Zelfs als er nieuwe axioma's en/of bewijsregels ontdekt en geaccepteerd zouden worden, dan nog zou dit niet voldoende zijn om meer dan een eindig aantal waarden van I of BB te bepalen. Dit is een fundamentele onvolledigheid, waaraan, zoals zal blijken, niet te ontsnappen valt.

T is in het vervolg een of ander formeel systeem van axioma's en bewijsregels op grond waarvan (onder andere) stellingen van de vorm "$I(\sigma) > k$" bewezen kunnen worden. Als zo'n stelling bewijsbaar is in T, dan moet hij ook waar zijn. Dat spreekt eigenlijk vanzelf, want anders was T onbetrouwbaar en onbruikbaar. Verder zijn de axioma's en bewijsregels van T niet aan enige beperking onderhevig. Chaitin's versie van de onvolledigheidsstelling van Gödel zegt nu dat er een (van T afhankelijke) constante **K** is, zodanig dat geen enkele uitspraak "$I(\sigma) > \mathbf{K}$" bewijsbaar is in T [3,6]. Op eindig veel uitzonderingen na hebben alle bitstrings echter een algoritmische informatie-inhoud

$>\mathbf{K}$! De mate van onvolledigheid van T is dus in zekere zin oneindig groot.

Het bewijs is niet moeilijk, maar wel subtiel. Bij T is een procedure $theorems_T$ te construeren die elke keer als hij geactiveerd wordt de volgende stelling van T produceert. Te beginnen met T's axioma's somt hij op deze manier de (in het algemeen oneindig vele) stellingen van T in volgorde van toenemende (niet afnemende) lengte van hun bewijs op. Kies nu een getal k en construeer op basis van $theorems_T$ een programma R_k om een bitstring met informatie-inhoud $>k$ te produceren:

```
program R_k
{
 « Definitie van theorems_T »
 for all t in theorems_T
 do
   if t = "I(" + σ + ") > k" for some σ
    then return(σ)
    else continue
   fi
 od
}
```

Voor een door R_k afgeleverde bitstring σ geldt enerzijds

$$I(\sigma) > k,$$

want dat is blijkbaar een stelling van T, maar anderzijds is

$$I(\sigma) \leqslant lengte(R_k) = lengte(theorems_T) + (\lfloor {}^2\log k \rfloor + 1) + c,$$

want σ wordt door R_k geproduceerd. De lengte van $theorems_T$ wordt bepaald door de lengte van de axioma's en bewijsregels van T. De constante c is de lengte van wat er van R_k overblijft na aftrek van de definitie van $theorems_T$ en de representatie van k. Deze laatste groeit slechts logaritmisch met k, zodat er een getal $\mathbf{K}$ afhankelijk van T is met $\mathbf{K} \geqslant lengte(R_{\mathbf{K}})$. Een door $R_{\mathbf{K}}$ geproduceerde σ heeft een informatie-inhoud die zowel $>\mathbf{K}$ als $\leqslant\mathbf{K}$ is. Dat kan niet en $R_{\mathbf{K}}$ levert dus nooit een resultaat. Zoals Chaitin zegt: "*. . . if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms*" [3].

Tenslotte dank ik Lambert Meertens en Paul Vitányi voor hun kritische opmerkingen.

Referenties

[1] A.M. Turing, "On computable numbers with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, series 2, **42**(1936-1937), pp. 230-265; corrections, **43**(1937), pp. 544-546. Tevens opgenomen in: M. Davis (Ed.), *The Undecidable*, Raven Press, New York, 1965, pp. 115-154.

[2] M. Davis, "Why Gödel didn't have Church's thesis", *Information & Control*, **54**(1982), pp. 3-24.
[3] G.J. Chaitin, "Gödel's theorem and information", *International Journal of Theoretical Physics*, **21**(1982), pp. 941-954.
[4] R.P. Feynman, "Simulating physics with computers", *International Journal of Theoretical Physics*, **21**(1982), pp. 467-488.
[5] S. Lin & T. Rado, "Computer studies of Turing machine problems", *Journal of the ACM*, **12**(1965), pp. 196-212.
[6] G.J. Chaitin, "Randomness and mathematical proof", *Scientific American*, mei 1975, pp. 47-52.

MC SYLLABI

1.1 F. Göbel, J. van de Lune. *Leergang besliskunde, deel 1: wiskundige basiskennis.* 1965.

1.2 J. Hemelrijk, J. Kriens. *Leergang besliskunde, deel 2: kansberekening.* 1965.

1.3 J. Hemelrijk, J. Kriens. *Leergang besliskunde, deel 3: statistiek.* 1966.

1.4 G. de Leve, W. Molenaar. *Leergang besliskunde, deel 4: Markovketens en wachttijden.* 1966.

1.5 J. Kriens, G. de Leve. *Leergang besliskunde, deel 5: inleiding tot de mathematische besliskunde.* 1966.

1.6a B. Dorhout, J. Kriens. *Leergang besliskunde, deel 6a: wiskundige programmering 1.* 1968.

1.6b B. Dorhout, J. Kriens, J.Th. van Lieshout. *Leergang besliskunde, deel 6b: wiskundige programmering 2.* 1977.

1.7a G. de Leve. *Leergang besliskunde, deel 7a: dynamische programmering 1.* 1968.

1.7b G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7b: dynamische programmering 2.* 1970.

1.7c G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7c: dynamische programmering 3.* 1971.

1.8 J. Kriens, F. Göbel, W. Molenaar. *Leergang besliskunde, deel 8: minimaxmethode, netwerkplanning, simulatie.* 1968.

2.1 G.J.R. Förch, P.J. van der Houwen, R.P. van de Riet. *Colloquium stabiliteit van differentieschema's, deel 1.* 1967.

2.2 L. Dekker, T.J. Dekker, P.J. van der Houwen, M.N. Spijker. *Colloquium stabiliteit van differentieschema's, deel 2.* 1968.

3.1 H.A. Lauwerier. *Randwaardeproblemen, deel 1.* 1967.

3.2 H.A. Lauwerier. *Randwaardeproblemen, deel 2.* 1968.

3.3 H.A. Lauwerier. *Randwaardeproblemen, deel 3.* 1968.

4 H.A. Lauwerier. *Representaties van groepen.* 1968.

5 J.H. van Lint, J.J. Seidel, P.C. Baayen. *Colloquium discrete wiskunde.* 1968.

6 K.K. Koksma. *Cursus ALGOL 60.* 1969.

7.1 *Colloquium moderne rekenmachines, deel 1.* 1969.

7.2 *Colloquium moderne rekenmachines, deel 2.* 1969.

8 H. Bavinck, J. Grasman. *Relaxatietrillingen.* 1969.

9.1 T.M.T. Coolen, G.J.R. Förch, E.M. de Jager, H.G.J. Pijls. *Colloquium elliptische differentiaalvergelijkingen, deel 1.* 1970.

9.2 W.P. van den Brink, T.M.T. Coolen, B. Dijkhuis, P.P.N. de Groen, P.J. van der Houwen, E.M. de Jager, N.M. Temme, R.J. de Vogelaere. *Colloquium elliptische differentiaalvergelijkingen, deel 2.* 1970.

10 J. Fabius, W.R. van Zwet. *Grondbegrippen van de waarschijnlijkheidsrekening.* 1970.

11 H. Bart, M.A. Kaashoek. H.G.J. Pijls, W.J. de Schipper, J. de Vries. *Colloquium halfalgebra's en positieve operatoren.* 1971.

12 T.J. Dekker. *Numerieke algebra.* 1971.

13 F.E.J. Kruseman Aretz. *Programmeren voor rekenautomaten; de MC ALGOL 60 vertaler voor de EL X8.* 1971.

14 H. Bavinck, W. Gautschi, G.M. Willems. *Colloquium approximatietheorie.* 1971.

15.1 T.J. Dekker, P.W. Hemker, P.J. van der Houwen. *Colloquium stijve differentiaalvergelijkingen, deel 1.* 1972.

15.2 P.A. Beentjes, K. Dekker, H.C. Hemker, S.P.N. van Kampen, G.M. Willems. *Colloquium stijve differentiaalvergelijkingen, deel 2.* 1973.

15.3 P.A. Beentjes, K. Dekker, P.W. Hemker, M. van Veldhuizen. *Colloquium stijve differentiaalvergelijkingen, deel 3.* 1975.

16.1 L. Geurts. *Cursus programmeren, deel 1: de elementen van het programmeren.* 1973.

16.2 L. Geurts. *Cursus programmeren, deel 2: de programmeertaal ALGOL 60.* 1973.

17.1 P.S. Stobbe. *Lineaire algebra, deel 1.* 1973.

17.2 P.S. Stobbe. *Lineaire algebra, deel 2.* 1973.

17.3 N.M. Temme. *Lineaire algebra, deel 3.* 1976.

18 F. van der Blij, H. Freudenthal, J.J. de Iongh, J.J. Seidel, A. van Wijngaarden. *Een kwart eeuw wiskunde 1946-1971, syllabus van de vakantiecursus 1971.* 1973.

19 A. Hordijk, R. Potharst, J.Th. Runnenburg. *Optimaal stoppen van Markovketens.* 1973.

20 T.M.T. Coolen, P.W. Hemker, P.J. van der Houwen, E. Slagt. *ALGOL 60 procedures voor begin- en randwaardeproblemen.* 1976.

21 J.W. de Bakker (red.). *Colloquium programmacorrectheid.* 1975.

22 R. Helmers, J. Oosterhoff, F.H. Ruymgaart, M.C.A. van Zuylen. *Asymptotische methoden in de toetsingstheorie; toepassingen van naburigheid.* 1976.

23.1 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 1.* 1976.

23.2 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 2.* 1977.

24.1 P.J. van der Houwen. *Numerieke integratie van differentiaalvergelijkingen, deel 1: eenstapsmethoden.* 1974.

25 *Colloquium structuur van programmeertalen.* 1976.

26.1 N.M. Temme (ed.). *Nonlinear analysis, volume 1.* 1976.

26.2 N.M. Temme (ed.). *Nonlinear analysis, volume 2.* 1976.

27 M. Bakker, P.W. Hemker, P.J. van der Houwen, S.J. Polak, M. van Veldhuizen. *Colloquium discretiseringsmethoden.* 1976.

28 O. Diekmann, N.M. Temme (eds.). *Nonlinear diffusion problems.* 1976.

29.1 J.C.P. Bus (red.). *Colloquium numerieke programmatuur, deel 1A, deel 1B.* 1976.

29.2 H.J.J. te Riele (red.). *Colloquium numerieke programmatuur, deel 2.* 1977.

30 J. Heering, P. Klint (red.). *Colloquium programmeeromgevingen.* 1983.

31 J.H. van Lint (red.). *Inleiding in de coderingstheorie.* 1976.

32 L. Geurts (red.). *Colloquium bedrijfsystemen.* 1976.

33 P.J. van der Houwen. *Berekening van waterstanden in zeeën en rivieren.* 1977.

34 J. Hemelrijk. *Oriënterende cursus mathematische statistiek.* 1977.

35 P.J.W. ten Hagen (red.). *Colloquium computer graphics.* 1978.

36 J.M. Aarts, J. de Vries. *Colloquium topologische dynamische systemen.* 1977.

37 J.C. van Vliet (red.). *Colloquium capita datastructuren.* 1978.

38.1 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part I.* 1979.

38.2 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part II.* 1979.

39 O.J. Vrieze, G.L. Wanrooy. *Colloquium stochastische spelen.* 1978.

40 J. van Tiel. *Convexe analyse.* 1979.

41 H.J.J. te Riele (ed.). *Colloquium numerical treatment of integral equations.* 1979.

42 J.C. van Vliet (red.). *Colloquium capita implementatie van programmeertalen.* 1980.

43 A.M. Cohen, H.A. Wilbrink. *Eindige groepen (een inleidende cursus).* 1980.

44 J.G. Verwer (ed.). *Colloquium numerical solution of partial differential equations.* 1980.

45 P. Klint (red.). *Colloquium hogere programmeertalen en computerarchitectuur.* 1980.

46.1 P.M.G. Apers (red.). *Colloquium databankorganisatie, deel 1.* 1981.

46.2 P.G.M. Apers (red.). *Colloquium databankorganisatie, deel 2.* 1981.

47.1 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60: general information and indices.* 1981.

47.2 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 1: elementary procedures; vol. 2: algebraic evaluations.* 1981.

47.3 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3A: linear algebra, part I.* 1981.

47.4 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3B: linear algebra, part II.* 1981.

47.5 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 4: analytical evaluations; vol. 5A: analytical problems, part I.* 1981.

47.6 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 5B: analytical problems, part II.* 1981.

47.7 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 6: special functions and constants; vol. 7: interpolation and approximation.* 1981.

48.1 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 1.* 1982.

48.2 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 2.* 1982.

49 T.H. Koornwinder (ed.). *The structure of real semisimple Lie groups.* 1982.

50 H. Nijmeijer. *Inleiding systeemtheorie.* 1982.

51 P.J. Hoogendoorn (red.). *Cursus cryptografie.* 1983.

CWI SYLLABI

1 Vacantiecursus 1984 *Hewet - plus wiskunde.* 1984.