

STAtOR

periodiek van de VVSOR jaargang 21, nummer 4, december 2020

Het vullen van theaters in coronatijd

Het gebruik van een tijd-ruimte netwerk voor het *ship-to-shore* probleem

Het temmen van het toeval; kanttekeningen bij de probabilistische wending in de AI

Kerstpuzzels en de verschrikkelijke kerstman

Nieuwe statistiek voegt wereldwijd corona-onderzoek samen

Fabels

Für Elise; problemen met wachttijden bij de GGZ

Geloven tussen 3D-lijnen

Goede contacten



STATOR

Jaargang 21, nummer 4, december 2020

STATOR is een uitgave van de Vereniging voor Statistiek en Operations Research (VVSOR). STATOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. Verschijnt 4 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Joep Burger, Kristiaan Glorie, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stermerdink (eindredacteur), Vanessa Torres van Grinsven en Sanne Willems. Vaste medewerkers: John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, mobiel 06 55886747, j.a.dossantos.gromicho@vu.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl; Secretaris: Laura Zwep MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Bestuurscoördinator: Pieter Jongsma MSc, db@vvsor.nl; Webmaster: Mitra Ebrahimpoor MSc, Nikky van Buuren MSc, webmaster@vvsor.nl
Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); prof. dr. Casper Albers (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Jonas Haslbeck MSc (Young Statisticians); dr. Sanne Willems (Section Statistics Communication).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Maarsbergseweg 20, 3956 KW Leersum, admin@vvsor.nl. Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

Voor advertenties

M. van Hootegem, hootegem@xs4all.nl
STATOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

WAT EEN JAAR!

Het jaar is bijna ten einde, een jaar zoals we dat nooit hadden kunnen vermoeden. Onze Annual Meeting die, zonder dat we het toen wisten, op de laatste dag voor de lock-down werd gehouden, was de afsluiting van 'life as it used to be'. Van de ene dag op de andere waren restaurants, winkels en kappers dicht en moest het onderwijs in al zijn geledingen zich al proefondervindelijk hervinden in een online variant. Maar we zijn kennelijk inventief, binnen enkele weken waren er honderdduizenden thuiswerkplekken gecreëerd en stond het merendeel van de docenten, van basis tot hoger onderwijs, als vol-leerde acteurs voor de camera les te geven alsof het nooit anders was geweest.

Ook STATOR pakte de draad moeiteloos op en ging door met het samenstellen van boeiende en afwisselende nummers. Onze redactievergaderingen vonden niet meer plaats rond een grote hoektafel in een Utrechts restaurant maar vanuit zolderkamers, keukens en studeerkamers. In nummer 2 van dit jaar kondigden we aan na te denken over een geheel aan COVID-19 gewijd nummer. Maar al snel verschenen er via allerlei kanalen zoveel gespecialiseerde publicaties dat het ons duidelijk werd dat wij met onze driemaandelijkse uitgave niet zouden kunnen tippen aan die actualiteit. We richten ons daarom nu op een meer retrospectieve *special* tegen het eind van 2021 die voornamelijk gewijd zal zijn aan de 'lessons learned'. Wij hebben Gerrit Timmer en Casper Albers bereid gevonden als gast-samenstellers voor zo'n nummer te fungeren.

In dit nummer zijn enkele artikelen gerelateerd aan COVID-19. De optimale bezetting van theaterzalen door Danny Blom en zijn collega's bijvoorbeeld. Judith ter Schure en Peter Grünwald hebben een methode ontwikkeld om uitkomsten van veel kleine corona-onderzoeken te combineren. Zij zijn hierover geïnterviewd door Fenna van der Grient voor *New Scientist* en wij mochten dat artikel overnemen. Twee van onze columnisten zijn ook deels geïnspireerd, John Poppelaars schrijft over complottheorieën en Gerrit Stermerdink over goede contacten dankzij conferenties die voorlopig wel niet meer gehouden zullen worden. Maar het is niet een-en-al COVID-19. Mette Wagenvoort beschrijft de planning van *ship-to-shore* operaties, Richard Starmans de probabilistische wending in de AI en Bob van Limburg de problemen rond wachttijden in de GGZ. Gerard Sierksma wijdt zijn column aan de precisie van de VAR in het voetbal en Henk Tijms trakteert ons op enkele smakelijke puzzels voor onder de kerstboom.

Het kost ons dat ook geen moeite u het traditionele 'veel leesplezier' toe te wensen.



INHOUD

- 2 Redactioneel – Wat een jaar!**
- 4 Het vullen van theaters in coronatijd**
| DANNY BLOM, RUDI PENDAVINGH & FRITS SPIEKSMa
- 10 Het gebruik van een tijd-ruimte netwerk voor het *ship-to-shore* probleem**
| METTE WAGENVOORT
- 14 Het temmen van het toeval; kanttekeningen bij de probabilistische wending in de AI**
| RICHARD STARMANS
- 19 Kerstpuzzels en de verschrikkelijke kerstman – column**
| HENK TIJMS
- 21 Nieuwe statistiek voegt wereldwijd corona-onderzoek samen**
| FENNA VAN DER GRIENT
- 24 Fabels – column**
| JOHN POPPELAARS
- 26 Für Elise; Problemen met wachttijden bij de GGZ**
| BOB VAN LIMBURG
- 31 Geloven tussen 3D-lijnen – column**
| GERARD SIERKSMA
- 34 LNMB/NGB conferentie 22 januari 2021**
- 34 Verslag NGB pubquiz**
- 35 Goede contacten – column**
| GERRIT STERMERDINK
- 36 VVSOR Annual Meeting 18 maart 2021
Going viral – Modeling spread**
- 36 Oproep om kandidaten te nomineren voor de Jan Hemelrijk en Willem R. van Zwet Awards 2020**

Het vullen van theaters in coronatijd

Foto: Mark Williams via Unsplash



De coronapandemie en de bijbehorende maatregelen hebben grote gevolgen voor vele aspecten van het dagelijkse leven. In het bijzonder is het bezoeken van een theater geen vanzelfsprekendheid meer. Door de beperkende maatregelen omtrent het aantal mensen in een ruimte en door de anderhalvemetermaatregel is het voor veel theaters puzzelen om zo veel mogelijk bezoekers te kunnen verwelkomen in hun zalen. Hoewel de sector middels creatieve ideeën zoveel mogelijk mensen probeert te bereiken, blijft een ouderwets fysiek bezoek aan het theater een geliefde optie. Met wiskundige technieken analyseren we wat er mogelijk is binnen het ‘nieuwe normaal’.

DANNY BLOM, RUDI PENDAVINGH & FRITS SPIEKSMA

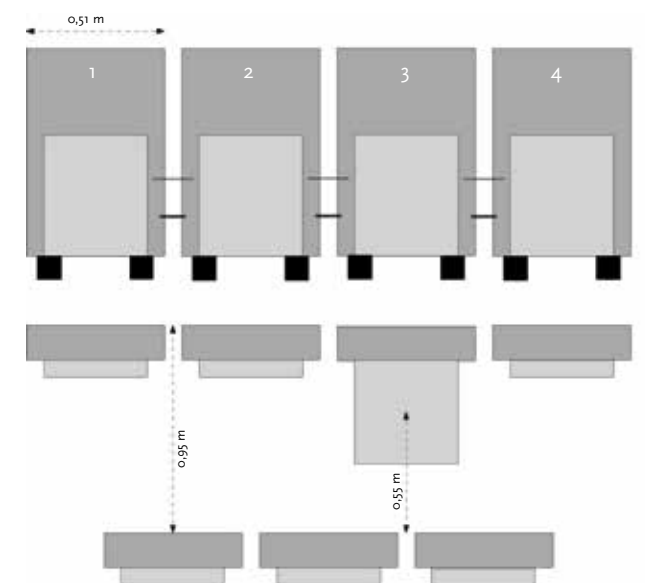
Sinds SARS-CoV-2 eind 2019 zijn ongewenste entree maakte op het wereldpodium, zijn er verschillende pogingen geweest om de verspreiding van het virus tegen te gaan. In het beleid van verschillende landen kwam dit neer op het beperken van sociale contacten, bijvoorbeeld door het sluiten van horecagelegenheden en het drastisch inperken van de capaciteit van theaters. Tijdens de eerste golf van de coronapandemie in Nederland waren theatervoorstellingen zelfs verboden, totdat de restricties in juni 2020 werden versoepeld tot maximaal 30 bezoekers per zaal. Dit was echter in geen geval genoeg voor theaters om rendabel voorstellingen te kunnen laten opvoeren.

Gaandeweg zijn de maatregelen dusdanig versoepeld dat theaters alleen nog maar gebonden waren aan de onderlinge anderhalve meter afstand tussen bezoekers die tot verschillende huishoudens behoren. De onderstaande wiskundige technieken zijn gebaseerd op deze setting waarin dus alleen *social distancing* een beperkende rol speelt.

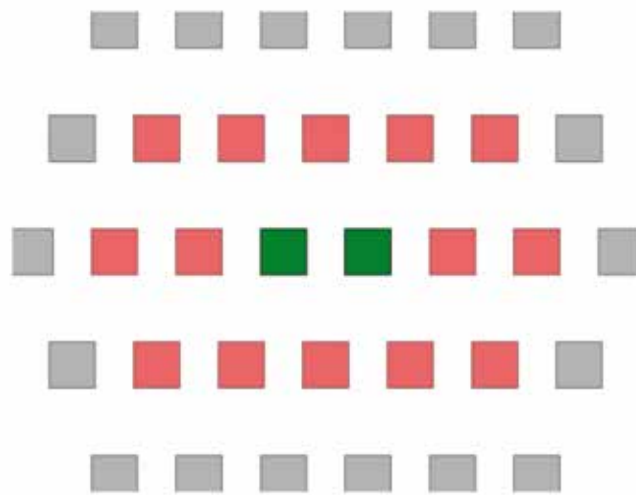
Onze analyse is gebaseerd op de zalen van het Muziekgebouw Frits Philips Eindhoven, een theater met twee zalen: de Grote Zaal (1.250 stoelen), vooral geschikt voor klassieke concerten en symfonieorkesten, en de Kleine Zaal (400 stoelen) die zich vooral leent voor kamermusiekconcerten in een intieme setting.

Zaaleigenschappen Muziekgebouw Frits Philips Eindhoven

Voordat we kunnen gaan kijken naar wiskundige technieken om het aantal bezoekers te optimaliseren binnen de coronaregels, moeten we eerst eigenschappen van de stoelenconfiguratie in de zaal bestuderen. Figuur 1 geeft een schematische weergave van het boven- en vooraanzicht van stoelen in het Muziekgebouw Eindhoven, met de bijbehorende afmetingen van de stoelen. De stoelen in opeenvolgende rijen staan in een zigzagpatroon, zodat het zicht van bezoekers verder achterin de zaal zo min



Figuur 1. Voor- en bovenaanzicht van stoelen in Muziekgebouw Eindhoven



Figuur 2. Als een familie van twee de groene stoelen bezet, mogen andere families niet op de rode stoelen plaatsnemen

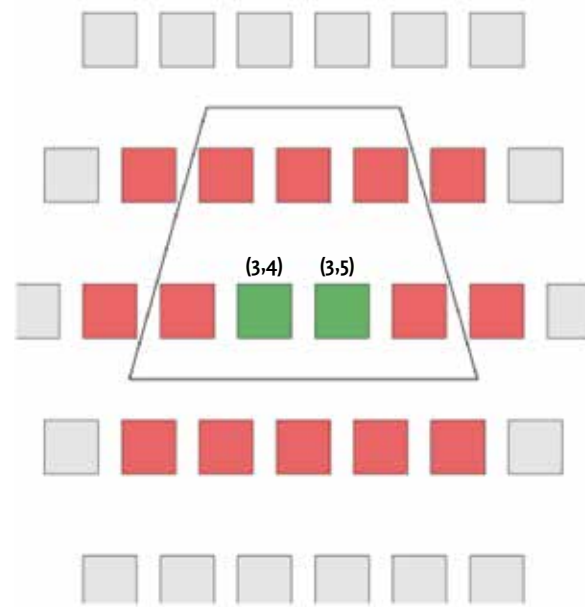
mogelijk wordt geblokkeerd.

Personen die tot hetzelfde huishouden behoren, noemen we een familie; die hoeven niet anderhalve meter uit elkaar te zitten. We gaan ervan uit dat leden van een familie opeenvolgende stoelen bezetten in dezelfde rij. Het aantal leden van een familie noemen we de grootte van de familie. Met behulp van de gegevens uit figuur 1 kunnen we voor iedere bezoekende familie bepalen welke stoelen ‘verboden gebied’ zijn voor bezoekers van een andere familie (zie figuur 2 voor een familie van twee).

Trapezoides

Het doel van het optimaliseringsprobleem dat we hier willen oplossen is om zoveel mogelijk bezoekers toe te kunnen wijzen aan stoelen in de theaterzaal. Dit probleem is een zogenaamd *packing* probleem; in dit geval gaat het om een gewogen *packing* van families, waarbij het gewicht van elke familie de familie grootte is. Voor het Muziekgebouw Eindhoven levert dit abstracte probleem een mooie meetkundige interpretatie op. Met elke familie en elke verzameling stoelen die door een familie bezet kan worden, kan een unieke trapezoïde geassocieerd worden; zie figuur 3 voor een illustratie.

We introduceren deze trapezoides om de volgende reden: een toewijzing van families aan stoelen die voldoet aan de anderhalvemeterregel, is equivalent met een *packing* van de bijbehorende trapezoides (Blom et al., 2020). Dit inzicht stelt ons in staat om het probleem op heldere wijze te formuleren: het doel van ons probleem is dus om een verzameling trapezoides te vinden zodanig



Figuur 3. De unieke trapezoïde behorende bij een paar op stoelen (3,4) en (3,5)

dat elk paar trapezoides niet overlapt en waarbij de som van de familie groottes behorende bij deze trapezoides zo groot mogelijk is.

Een andere manier om naar dit probleem te kijken is door het te vertalen naar een graafprobleem. In deze graaf vertegenwoordigt iedere plaatsing van een familie een knoop in de graaf, en er bestaat een kant tussen twee knopen bestaat wanneer deze twee families niet tegelijkertijd geplaatst kunnen worden. Het doel komt nu overeen met het vinden van een maximum gewogen stabiele verzameling van knopen (maximum *weight stable set problem*).

Klantprofielen

Wanneer men, zonder verdere informatie over de grootte van families, het trapezoidemodel zou doorrekenen voor de zalen van Muziekgebouw Eindhoven, zullen er in een optimale zaalbezetting voornamelijk relatief grote families voorkomen. De reden hiervoor is duidelijk: de ratio van het aantal bezette stoelen door een familie en het aantal stoelen in de verboden zone, is groter voor grote families dan voor kleinere families. Maar aangezien de grootte van de families die een theater bezoeken niet door het theater bepaald wordt, introduceren we het begrip klantprofielen. Een klantprofiel specificeert voor elk familie grootte het percentage families van die grootte die de voorstelling komen bezoeken. Zo kunnen we bij elke voorstelling een geschikt klantprofiel kiezen, en geven we aan het algoritme een parameter mee die ervoor zorgt de gevonden stoelbezetting bij benadering voldoet aan een geschikte verdeling van familie groottes.

Implementatie

Het wiskundige model (gebaseerd op het vinden van een maximum gewogen stabiele verzameling in een graaf), kan door middel van een lineaire doelfunctie en lineaire beperkingen als een geheeltallig programmeringsprobleem (integer programming, ook wel IP) worden beschreven (Blom et al., 2020). Met behulp van het commerciële softwarepakket Gurobi hebben we het bovenstaande trapezoidemodel geïmplementeerd, en doorerekend voor de Grote en Kleine Zaal van het Muziekgebouw Eindhoven. Hierbij is uitgegaan van vier verschillende scenario’s voor het klantprofiel, namelijk:

- MGE1: 18% singles, 70% tweetallen, 6% drietallen, 6% viertallen
- MGE2: 100% tweetallen
- MGE3: 20% singles, 80% tweetallen
- MGE4: 50% tweetallen, 50% viertallen.

Een standaardformulering van het stable set probleem is met behulp van Gurobi binnen enkele seconden op te lossen voor input groottes die horen bij deze twee zalen. Het toevoegen van de klantprofielen aan deze formulering maakt het probleem echter moeilijker. Een *stable set* probleem is op te lossen als een lineair programmingsprobleem wanneer er een volledige beschrijving beschikbaar is van het convexe omhulsel van alle binaire vectoren die een *stable set* karakteriseren, in termen van lineaire ongelijkheden. De standaard IP-formulering bevat slechts een klein deel van al deze lineaire ongelijkheden. Door het toevoegen van extra lineaire ongelijkheden is het mogelijk dat het resulterende geheeltallige programma makkelijker oplosbaar is, bijvoorbeeld omdat er sterkere bovengrenzen worden gevonden in het branch-and-bound algoritme. Naast de standaardimplementatie hebben we daarom een subklasse van zogenaamde *clique constraints* toegevoegd aan het geheeltallig programma. De toegevoegde ongelijkheden houden concreet in dat voor elk

drietal stoelen geldt dat van alle trapezoides die twee van deze drie stoelen bevatten, er slechts één trapezoïde kan worden gekozen. De correctheid van deze ongelijkheden volgt uit het duiventilprincipe.

Meerdere opeenvolgende shows

In tabel 1 geven we de resultaten weer voor het trapezoïde model. De kolom ‘% cap’ bevat het percentage gevulde stoelen in vergelijking met de volledige capaciteit (de bezettingsgraad), en de kolommen ‘standaard (s)’ en ‘versterkt (s)’ respectievelijk bevatten de rekentijden (in seconden) voor de standaard en de versterkte formulering.

Zoals te zien in tabel 1, ligt de bezettingsgraad rond de 33%, waarbij voorstellingen met klantprofielen met relatief veel grote families tot 37% scoren. Echter, de meeste theaters kunnen met deze bezettingsgraden niet rendabel functioneren. Vandaar dat er gekeken is naar de mogelijkheid om een voorstelling meerdere malen achter elkaar op te voeren, telkens voor een verschillend publiek. Het in goede banen leiden van binnenkomend en weggaand publiek, vergt dan wel een grote inzet van het theaterpersoneel, en het schoonmaken van de stoelen tussen twee van zulke voorstellingen is niet meer mogelijk. We zijn dus op zoek naar oplossingen waarbij elke stoel slechts tijdens één van deze voorstellingen op een avond bezet kan worden. Hierdoor hoeven stoelen niet tussentijds te worden schoongemaakt door het personeel. Deze beperking blijkt relatief eenvoudig te modelleren met behulp van lineaire ongelijkheden.

Door het uitbreiden van het model vanwege het toelaten van meerdere opeenvolgende voorstellingen, worden symmetrieën toegevoegd aan het model. Als er twee stoelbezettingen gevonden worden voor voorstelling 1 en voorstelling 2 onder de juiste restricties, dan zou men deze bezettingen ook kunnen omwisselen en dan zou dit nog steeds een toelaatbare oplossing zijn. Dit noemen

KLANTPROFIEL	Grote Zaal			Kleine Zaal		
	% cap	standaard (s)	versterkt (s)	% cap	standaard (s)	versterkt (s)
MGE1	32	3,39	1,50	34	1,19	0,41
MGE2	29	0,28	0,10	31	0,02	0,02
MGE3	30	1,39	0,97	31	0,22	0,07
MGE4	36	5,29	1,10	37	0,08	0,11

Tabel 1. Resultaten van het trapezoïde model met een standaard en versterkte IP-formulering

KLANTPROFIEL	Grote Zaal			Kleine Zaal		
	% cap	standaard (s)	versterkt (s)	% cap	standaard (s)	versterkt (s)
MGE1	63	532,69	48,28	64	8,18	13,82
MGE2	56	6,67	2,49	58	0,30	0,35
MGE3	58	2107,68	6,05	59	2,19	0,89
MGE4	70	4485,33	726,11	70	5,46	9,17

Tabel 2. Resultaten van het trapezoïde model voor twee voorstellingen met een standaard en versterkte IP-formulering

we symmetrische oplossingen. Gelukkig bestaan er zogenaamde technieken om symmetrieën te breken, zoals opnieuw het toevoegen van lineaire beperkingen waaraan slechts één van de symmetrische oplossingen voldoet. In tabel 2 beschrijven we gelijksoortige resultaten voor het geval van twee opeenvolgende voorstellingen. De versterkte formulering maakt hier gebruik van zowel de extra clique constraints als de symmetriebrekende technieken.

De resultaten in tabel 2 laten zien dat de bezettingsgraden ruwweg verdubbelen bij het introduceren van een extra show. Dit is een belangrijke observatie omdat hiermee aangetoond wordt dat commerciële exploitatie van een theater in principe mogelijk is, ook in tijden van corona, wanneer een show twee keer opgevoerd kan worden.

We zien wel dat de rekentijden voor deze situatie snel toenemen, met name voor de Grote Zaal. Dit is niet verrassend, omdat het aantal variabelen en beperkingen snel groeit wanneer het aantal voorstellingen toeneemt. We zien ook, in de kolom ‘versterkt (s)’, dat het toevoegen van extra ongelijkheden aan het model leidt tot een significante afname in de rekentijd. Het belang van een korte rekentijd volgt uit het feit dat voor verschillende muziekgenres zeer waarschijnlijk verschillende klantprofielen gebruikt moeten worden, omdat de samenstelling van het publiek heel verschillend kan zijn. Daarnaast is het met korte rekentijden ook eerder mogelijk om zoge-

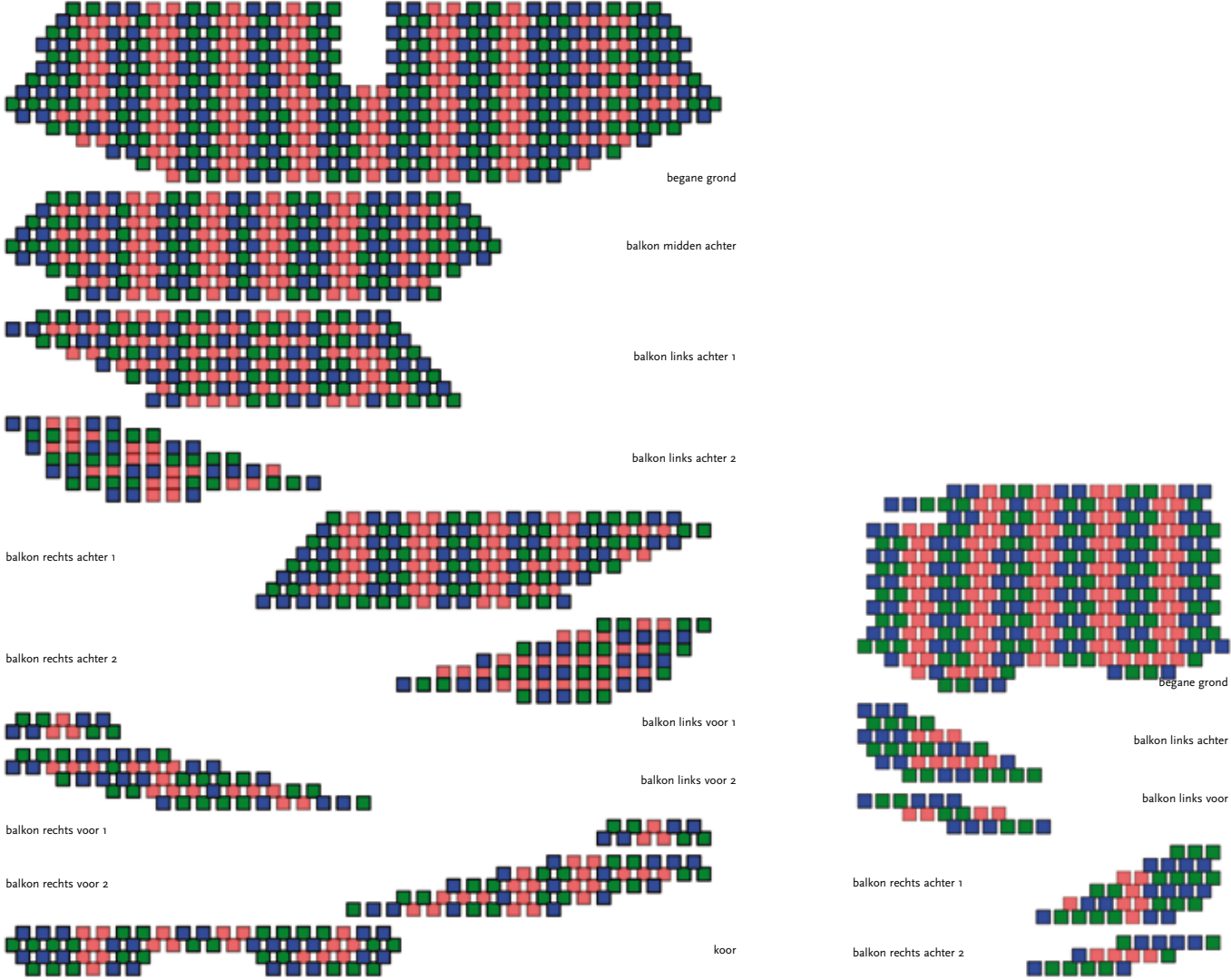
naamde online algoritmes te kunnen implementeren. Bij dergelijke methodes hoeft een klantprofiel niet als input gebruikt te worden, maar kunnen de continu binnenkomende reserveringen gebruikt worden om het klantprofiel dynamisch aan te passen, om zo tot een optimale bezetting te komen.

Wat als we niet alle rijen gebruiken?

Als we de stoelbezettingen uit figuur 4 bekijken, dan zien we dat voor beide voorstellingen, in bijna elke rij stoelen wel minstens één familie plaats kan nemen. Dit leidt tot een complexe logistieke operatie: hoe kan het theaterpersoneel ervoor zorgen dat de families in de juiste volgorde de zaal binnenkomen zonder dat ze binnen anderhalve meter van andere families moeten komen? Als alternatief beschouwen we een situatie waarbij de rijen alternerend gebruikt worden, bijvoorbeeld er mogen alleen families plaats nemen in rijen 1, 3, 5, etc.. De resultaten in tabel 3 geven weer dat de resulterende bezettingsgraad met ongeveer 5% daalt wanneer deze strategie wordt gebruikt in een theater. Daarentegen is het natuurlijk wel zo dat het naar binnen en buiten leiden van bezoekers veel eenvoudiger is, omdat de theaterzaal nu makkelijk van binnen naar buiten gevuld kan worden met bezoekers.

KLANTPROFIEL	Grote Zaal			Kleine Zaal		
	% cap	% verlies	rekentijd (s)	% cap	% verlies	rekentijd (s)
MGE1	57	-9,4	0,33	58	-9,8	0,11
MGE2	52	-7,5	0,05	52	-9,6	0,01
MGE3	52	-9,8	0,07	52	-11,8	0,08
MGE4	65	-6,0	0,12	66	-5,7	0,03

Tabel 3. Resultaten van het trapezoïde model voor twee voorstellingen met een standaard en versterkte IP formulering in geval van alternerende rijen



Figuur 4. Stoolbezettingen voor de zalen van Muziekgebouw Eindhoven met twee voorstellingen; de groene stoelen worden gebruikt tijdens de eerste voorstelling (Grote Zaal 391, Kleine Zaal 129) de blauwe tijdens de tweede (Grote Zaal 393, Kleine Zaal 126)

Conclusie

Al met al concluderen we dat, als een voorstelling twee keer op een avond gespeeld wordt, er een bezettingsgraad van 60-70% mogelijk is, wanneer we rekening moeten houden met de anderhalvemetermaatregel. Verder leidt het gebruik van alternerende rijen tot een (bescheiden) negatief effect op de maximale bezettingsgraad. Deze analyse geeft het Muziekgebouw Eindhoven de mogelijkheid om een weloverwogen besluit te kunnen nemen over de tradeoff tussen het maximaliseren van de bezettingsgraad en het minimaliseren van de complexiteit van het begeleiden van bezoekersstromen in de theaterzalen. Hierbij heeft het Muziekgebouw Eindhoven gekozen om de strategie van alternerend gebruik van rijen te implementeren in hun zalen. Daarnaast zijn er zeker mogelijkheden voor vervolgonderzoek, voornamelijk voor het nog realistischer in acht nemen van daadwerkelijke verdeling

van familiegroottes, bijvoorbeeld door te kijken naar onlinemethodes voor stoelbezettingen.

We danken de medewerkers van Muziekgebouw Eindhoven voor de prettige samenwerking.

Zie voor een uitgebreide versie van deze bijdrage het artikel *Filling a theatre in times of corona* van D. Blom, R. Pendavingh en F. C. R. Spieksma. arXiv:2010.01981, 30 september 2020.

DANNY BLOM is promovendus Combinatorial Optimization at Eindhoven University of Technology
E-mail: d.a.m.p.blom@tue.nl

RUDI PENDAVINGH is assistent professor Combinatorial Optimization at Eindhoven University of Technology
E-mail: r.a.pendavingh@tue.nl

FRITS SPIEKSMa is professor Combinatorial Optimization at Eindhoven University of Technology
E-mail: f.c.r.spieksma@tue.nl

HET GEBRUIK VAN EEN TIJD-RUIMTE NETWERK VOOR HET SHIP-TO-SHORE PROBLEEM

In het *ship-to-shore* probleem is het doel goederen en personeel van amfibische schepen zo snel mogelijk naar de kust te vervoeren om daar vervolgens ingezet te worden bij taken op de wal. In dit artikel wordt het gebruik van een tijd-ruimte netwerk voorgesteld om het *ship-to-shore* probleem op te lossen. Bij het maken van de planning van een amfibische operatie komen verschillende voorwaarden kijken. Om de interactie van tijd en ruimte in dit probleem te modelleren, kan een tijd-ruimte netwerk gebruikt worden waarna het probleem opgelost kan worden als een pad-probleem in dit netwerk.



Figuur 1. Schematische weergave van een amfibische operatie

METTE WAGENVOORT

Stel je hebt de leiding over een operatie waarin goederen en personeel vanaf een groot schip aan land moeten worden gebracht. Voor dit vervoer heb je kleinere schepen en helikopters, ook wel connectoren genoemd, tot je beschikking. Deze hebben echter een beperkte capaciteit en daarom moeten ze meerdere trips van je schip naar het land maken. Het is dus belangrijk om te bepalen welke goederen en welk personeel in welke trip vervoerd moeten worden. Dit moet gedurende de nacht gebeuren zodat je onopgemerkt blijft, dus moet je voor het licht wordt alle benodigde goederen en al het benodigde personeel aan land hebben gebracht. Dit is een voorbeeld van een *amfibische operatie*.

In amfibische operaties worden goederen en personeel niet alleen naar het land vervoerd om daar ingezet te

worden, maar ook teruggehaald nadat een operatie aan land heeft plaatsgevonden. Dit kan verschillende doelen hebben, bijvoorbeeld een aanval, terugtrekking, overval, demonstratie of voor ondersteuning bij een crisissituatie, zoals een natuurramp (Maritime Warfare Centre, 2019).

In het *ship-to-shore* probleem is het doel om goederen en personeel naar het strand te vervoeren, waarna de operatie aan land wordt voortgezet, in plaats van direct naar het doelwit te gaan (*ship-to-target*).

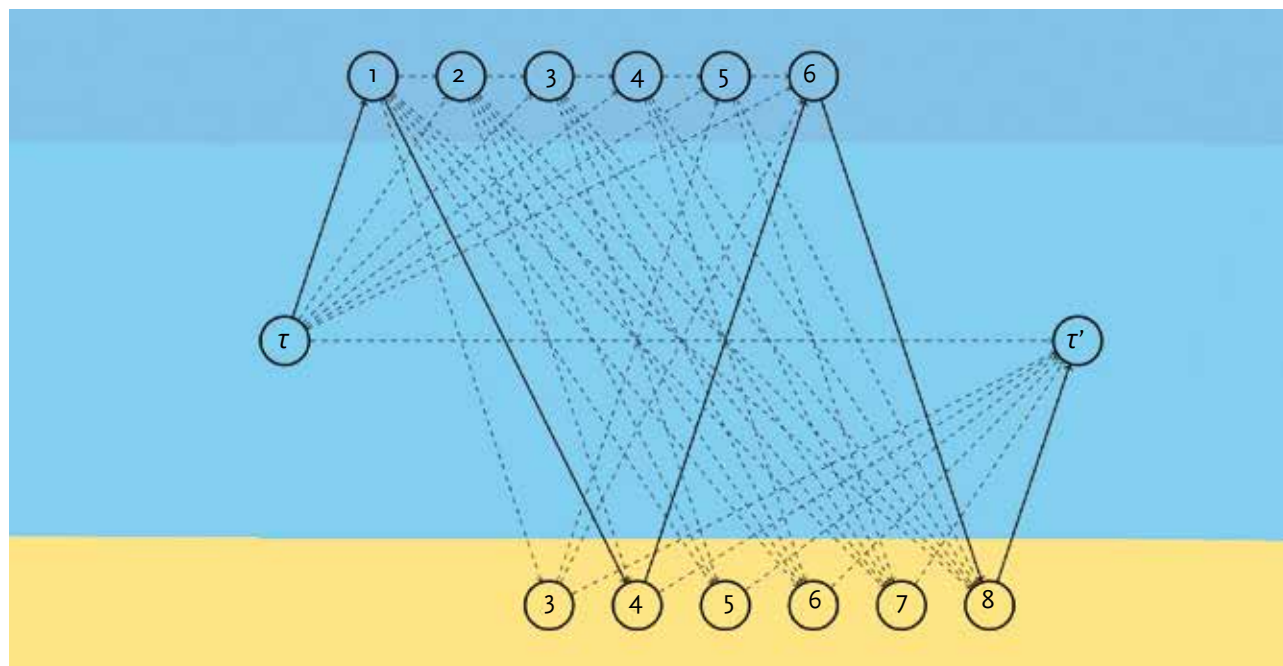
Bij zulke operaties komen verschillende voorwaarden kijken, wat het maken van de planning moeilijk maakt. In dit artikel worden eerst de voorwaarden aan zulke operaties toegelicht. Daarna wordt aangegeven hoe een operatie als een pad-probleem in een tijd-ruimte netwerk omschreven kan worden.

Ship-to-shore probleem

In figuur 1 is een schematische weergave van een amfibische operatie te zien. Hier zijn twee amfibische schepen weergegeven, de grotere schepen aan de bovenkant van de afbeelding, vanaf waar goederen en personeel vervoerd moeten worden. Het vervoeren van de goederen en het personeel wordt gedaan met connectoren. Connectoren zijn helikopters en kleinere schepen, ook wel landingsvaartuigen genoemd, die het strand kunnen bereiken. De locaties op het strand waar goederen en personeel afgeleverd moeten worden landingszones genoemd. De connectoren maken meerdere trips van de schepen naar de landingszones om alle goederen en al het personeel aan land te krijgen.

Er zijn verschillende voorwaarden en eigenschappen van het probleem:

- **Capaciteit bij het beladen**
De connectoren worden geladen vanaf de amfibische schepen. Helikopters kunnen landen op een landingsplatform op het schip, vanaf waar ze geladen kunnen worden. Landingsvaartuigen kunnen geladen worden via een aanlegplaats in het schip middels een klep aan de voorzijde van het schip. Er is echter een beperkte hoeveelheid aanlegplaatsen en landingsplatformen. Het aantal landingsplatformen en aanlegplaatsen is dan ook bepalend voor het aantal connectoren dat tegelijkertijd beladen kan worden.
- **Volgorde van het afleveren van de goederen en het personeel**



Figuur 2. Voorbeeld van een tijd-ruimte netwerk met één amfibisch schip en één landingszone

Er zijn verschillende typen goederen die aan land gebracht moeten worden. Sommige van deze goederen zijn aan het begin van de operatie nodig. Een voorbeeld hiervan is het *beach armoured recovery vehicle* welke gebruikt wordt om de connectoren aan te duwen als ze bij het verlaten van het strand vast zitten in het zand. Een ander voorbeeld is een 'wiellader' die een mat kan uitrollen op het strand waardoor het makkelijker is voor andere voertuigen om vanaf de landingsvaartuigen het strand op te rijden. Andere goederen zijn pas later nodig en lopen een risico als ze al vroeg aan land worden gezet, omdat ze bewaakt moeten worden als ze je niet direct nodig hebt. Aan de goederen en het personeel zijn daarom verschillende prioriteiten toegekend. Eerst moeten alle goederen en het personeel met de hoogste prioriteit afgeleverd worden en pas als deze allemaal aan land zijn mogen de goederen en het personeel met het volgende prioriteitsniveau aan land gebracht worden.

- **Groepen complementaire goederen en personeel**
Naast de prioriteiten zijn er ook goederen en personeel die complementair zijn aan elkaar. Daarom is het noodzakelijk dat deze tegelijkertijd of vlak na elkaar op het strand afgeleverd worden. Dit kan bij

voorbeeld personeel zijn dat tot dezelfde unit horen en daarom de operatie samen uitvoeren, maar ook voertuigen die gezamenlijk met een unit de operatie voortzetten op land.

Er moet aan deze voorwaarden voldaan worden bij het bepalen wat er in elke trip vervoerd wordt. Er zijn veel goederen en er is veel personeel op een amfibisch schip en dit moet allemaal vervoerd worden met behulp van connectoren met een beperkte capaciteit. Er moet dus zowel bepaald worden wanneer een connector van welk amfibisch schip naar welke landingszone gaat, alsook welke goederen en welk personeel er in welke trip vervoerd moeten worden.

Het tijd-ruimte netwerk

Om het *ship-to-shore* probleem te modelleren werd er gebruik gemaakt van een tijd-ruimte netwerk. In deze paragraaf wordt deze keuze beargumenteerd.

Door de beperkte capaciteit bij het laden van de connectoren, kan er steeds maar een maximum aantal connectoren tegelijkertijd beladen worden. Daarnaast is de tijd waarop een connector uitgeladen wordt op het strand van belang voor de prioritering en levering van goederen

en personeel per groep. Daarom speelt tijd een rol in dit probleem. Er moet dus voor elke trip bepaald worden wanneer een connector op een locatie arriveert, wanneer die weer vertrekt en welke goederen er in de trip vervoerd worden.

Deze interactie tussen tijd en ruimte kan gemodelleerd worden in een tijd-ruimte netwerk. Een voorbeeld van een tijd-ruimte netwerk is weergegeven in figuur 2. In dit voorbeeld is er één amfibisch schip en één landingszone. De tijdshorizon is in discrete intervallen opgedeeld, waarbij de lengte van een tijdsperiode gelijk is aan de maximumtijd voor het beladen van een connector, waarna voor elke combinatie van locatie en tijd een knoop aangemaakt wordt. Er worden pijlen toegevoegd als er een haalbare transitie door zowel tijd als ruimte mogelijk is. Als de afstand tussen twee locaties d mijl is en de connector vaart v mijl per uur, dan heeft de connector minstens $t = \frac{d}{v}$ uur nodig. Dit staat gelijk aan minstens $\lceil \frac{t}{l} \rceil$ tijdsperiodes, waarbij l de lengte van een tijdsperiode in uren is. In het voorbeeld in figuur 2 heeft het landingsvaartuig minstens één tijdsperiode nodig om van het amfibisch schip naar de landingszone te varen. Als de connector in tijdsperiode 1 geladen wordt op het schip, kan de connector dus op zijn vroegst in tijdsperiode 3 de landingszone bereiken. In figuur 2 zijn alle mogelijke pijlen voor de connector weergegeven.

Gegeven de weergave van het probleem als een pad-probleem in een tijd-ruimte netwerk, kan een route voor een connector omschreven worden als een pad van de start locatie, τ , naar de eind locatie, τ' , van de connector waarbij in elke trip van het schip naar de landingszone een groep met goederen en personeel vervoerd wordt. De zwarte lijnen in figuur 2 geven een pad weer door het netwerk. Hierin maakt de connector twee trips van het schip naar de landingszone, namelijk van tijdsperiode 1 naar tijdsperiode 4 en van tijdsperiode 6 naar tijdsperiode 8. Voor elk van deze trips moet een belading toegewezen worden.

Met deze definitie van een route kan een planning voor een operatie omschreven worden als een aantal routes, één per connector, waarbij alle goederen en al het personeel van de schepen naar het strand vervoerd worden, terwijl ze aan de voorwaarden voldoen en de lengte van de operatie geminimaliseerd wordt. Deze voorwaarden zijn nu toe te voegen doordat elke knoop in de graaf gelinkt is aan een bepaalde tijdsperiode op een locatie.

Discussie

Door het probleem te omschrijven als een pad-probleem in een tijd-ruimte netwerk is het mogelijk om het probleem te formuleren en aan de voorwaarden te voldoen. De uitkomst van dit model is een planning waarin gegeven is wanneer een connector vertrekt en welke goederen en welk personeel er in de trip mee moet. Op het schip kunnen voorbereidingen getroffen worden zodat het beladen van de connectoren gesmeerd verloopt. Ook ondersteunt deze formulering herplanning tijdens een operatie door de huidige situatie als invoer te geven. Tevens kan dit netwerk uitgebreid worden door het gebruik van verschillende routes van en naar de landingszone te modelleren door het toevoegen van extra knopen. Ook kan het een vereiste zijn dat bepaalde landingsvaartuigen gezamenlijk naar de landingszone varen, wat ook te implementeren is door middel van het toevoegen van extra knopen in de graaf.

Echter, als het aantal locaties groeit, groeit ook de grootte van het tijd-ruimte netwerk. Dit heeft effect op het aantal knopen en pijlen in de graaf en daarmee de berekeningstijd. Dit is inherent aan NP-moeilijke problemen. Met behulp van algoritmes en heuristieken kan het probleem opgelost worden.

Naast het aantal locaties, heeft ook de lengte van een tijdsperiode invloed op het aantal knopen. Hoe kleiner de lengte van een tijdsperiode, hoe meer knopen er zijn, maar ook hoe nauwkeuriger de planning gemaakt kan worden. Omdat de schepen relatief ver van de kust af liggen, is de reistijd van de connectoren in verhouding tot de tijd die nodig is om de connectoren te beladen, relatief hoog. Daarom is de beladingstijd als lengte voor de tijdsperiodes gebruikt (Amrouss et al., 2017).

LITERATUUR

Amrouss, A., El Hachemi, N., Gendreau, M., & Gendron, B. (2017). Real-time management of transportation disruptions in forestry. *Computers & Operations Research*, 83, 95-105.
Maritime Warfare Centre (2019). *Manual Surface Assault*.

METTE WAGENVOORT heeft haar masterscriptie voor de studie Econometrics and Management Science in de specialisatie Operations Research and Quantitative Logistics aan de Erasmus Universiteit bij TNO gedaan over het oplossen van het *ship-to-shore* probleem. Ze is afgelopen september begonnen als promovenda bij de Erasmus Universiteit, TNO en de Nederlandse Defensie Academie.
E-mail: wagenvoort@ese.eur.nl



Illustratie: Ahmed Gad via Pixabay

HET TEMMEN VAN HET TOEVAL

kanttekeningen bij de probabilistische wending in de AI

RICHARD STARMANS

Volgens een gangbare opvatting onder wetenschapshistorici heeft de afgelopen eeuw in vele wetenschappen een probabilistische revolutie plaatsgevonden. Dikwijls betekent dit niet alleen dat de onderzoeksmethoden en technieken binnen die disciplines sterk zijn gebaseerd op kansrekening en statistiek, maar ook het kennisdomein zelf, de gehanteerde begrippen, concepten en theorieën, ja zelfs het achterliggende wereldbeeld. Dat laatste wordt niet altijd toegejuicht. De probabilistische wending in de wetenschap gaat volgens menigeen gepaard met een verdere teloorgang van de aanschouwelijkheid van het wereldbeeld en de vertrouwde categorieën van denken en handelen, juist omdat gehanteerde concepten en methoden nog slechts probabilistisch te duiden zijn (Starmans, 2018). De rationalisering van de samenleving en bijbehorende instituties, de modellen van beleid, organisatie en

technologie, die geënt zijn op deze principes om greep te krijgen op onzekerheid, hebben a fortiori gevolgen voor concepties die direct de morele ervaring betreffen of een sterke ethische dimensie bezitten: verantwoordelijkheid, rechtvaardigheid, redelijkheid, betrouwbaarheid, vertrouwen, macht, democratie, zorgvuldigheid, veiligheid en risico. Deze blijken dikwijls steeds minder in overeenstemming met de individuele morele ervaring, de beproefde imperatieven voor moreel handelen en het daarmee verbonden morele houvast.

Wie een minder hooggestemd en meer *down-to-earth*-gezichtsveld prefereert, kan terecht bij tal van historici en filosofen van kansrekening en statistiek, die de afgelopen dertig jaar onderzoek hebben gedaan naar de mate waarin en de wijze waarop die probabilistische wending zich voltrok binnen de afzonderlijke wetenschapsgebieden.

Het tweedelige standaardwerk *The Probabilistic Revolution* (Kruger, 1987; 1990), waarin de eerste bevindingen van deze onderzoeken staan geboekstaafd, fungeerde als een opmaat tot vele latere studies. Uiteenlopende disciplines passeren hierin de revue, waaronder natuurkunde, biologie, genetica, landbouwwetenschap, psychologie, sociologie en economie. Niet toevallig zijn dit juist de disciplines die rond de eeuwwisseling diverse grondslagen crises doorliepen (natuurkunde) of – nog saillant – die in de late 19e eeuw opkwamen in een tijdsgewricht waarin gaandeweg het inzicht ontstond dat variatie en verandering in de (levende) natuur niet pejoratief dienen te worden geïdentificeerd of moeten worden ‘wegverklaard’ (zoals bij Laplace en Quetelet), maar intrinsiek aan die werkelijkheid zijn. Tegen de achtergrond van deze door Ian Hacking gepostuleerde ‘erosie van het determinisme’, zouden Francis Galton (1822–1911), Karl Pearson (1857–1937), Ronald Fisher (1890–1962) en anderen de fundamenteën leggen voor de moderne statistiek. Juist bij al deze – indertijd nog jonge en – bovenal ‘variatie-en-verandering-rijke’ wetenschappen, die bovendien nog geen onderzoekstraditie kenden, wetenschappelijke status moesten verwerven en worstelden met de grondslagen van hun vak, zouden de probabilistische methoden van voornoemde pioniers van de statistiek een vruchtbare voedingsbodem vinden.

Ook voor jongere disciplines die na de Tweede Wereldoorlog ontstonden, kan de vraag worden gesteld of en in welke mate zich een probabilistische wending heeft voltrokken. Een voorbeeld betreft de *Artificial Intelligence* (AI) die amper 70 jaar geleden ontstond en de laatste jaren in rap tempo lijkt te zijn opgeschoven van een marginale positie op het palet van de wetenschappen naar het zenit van het wetenschappelijk onderzoek. AI heeft de volle aandacht van politiek en bedrijfsleven, staat bovenaan internationale onderzoekagenda's, leidt tot verhitte debatten over kansen en bedreigingen en trekt enorme aantallen studenten. De vraag of en in welke mate zich binnen de AI een probabilistische wending voltrok heeft tot nu toe weinig aandacht gekregen. Enkele aspecten van deze problematiek worden hier kort voor het voetlicht gebracht.

AI, empirie en statistiek

Het is dit jaar precies een kwart eeuw geleden dat de Amerikaanse informaticus en AI-wetenschapper Paul R. Cohen zijn boek *Empirical Methods for Artificial Intelligence* (1995) publiceerde. Achter deze neutrale en ogenschijnlijk weinig pretentieuze titel, die een doorsnee tekstboek

suggereert, gaat wel degelijk een opmerkelijk en ambitieus project schuil. Cohen laat daarover geen misverstand bestaan en gaat met grote voortvarendheid van start. Zo stelt de auteur onomwonden dat er in de informatica en de AI geen methodologische traditie bestaat, geen gangbaar paradigma dat richtinggevend is in onderzoek en onderwijs. Hij kritiseert de opvattingen over empirisch onderzoek van AI-pioniers als Herb Simon en Allen Newell en stelt dat in een tijd van complexe, ingebedde en geavanceerde systemen er dringend behoefte is aan een basis voor *new, more powerful research methods*. De oplossing voor dit alles kan volgens Cohen worden gevonden in het bouwwerk van de statistiek, dat bij hem geschraagd wordt door verschillende pijlers. Allereerst omarmt hij de erfenis van John Tukey (1915–2000). De exploratieve data-analyse helpt het menselijk oog in het detecteren van patronen en structuren en biedt daarvoor de *statistical microscopes* en datalenzen. Daarnaast kan inferentiële statistiek bewerkstelligen dat *data speak convincingly*. Dit is aan de orde op het moment dat ideeën dienen te worden onderworpen aan de *blunt interrogation of hypothesis testing*, bij voorbeeld wanneer *generalisation* van bevindingen nodig is en het AI-systeem wordt aangepast om te kunnen worden ingezet en beoordeeld indien omgeving of taak is gewijzigd. Opmerkelijk is dat Cohen bij dit alles reeds in 1995 niet alleen de klassieke parametrische statistiek (*parameter estimation, hypothesis testing, p-values* en betrouwbaarheidsintervallen) behandelt, maar ook uitvoerig de moderne computer-intensieve technieken. Zo bespreekt hij achtereenvolgens de niet-parametrische *bootstrap sampling distributions* en betrouwbaarheidsintervallen, de Monte Carlo methoden en tot slot Randomisation Tests. Het gaat bij de laatste categorie vooral om permutation tests en exacte tests, die feitelijk teruggaan tot Fishers befaamde exact test voor 2 bij 2 kruistabellen. Daarin worden exacte p-waarden berekend, die Fisher reeds in *The Design of Experiments* uit 1935 beschrijft. Hij doet dit aan de hand van zijn beroemde *the-lady-tasting-tea-experiment* over de vraag of een Engelse dame het verschil kan proeven tussen thee waaraan melk is toegevoegd en melk waaraan thee is toegevoegd, uiteraard in vaste verhoudingen (Salsburg, 2009). Een laatste pijler van Cohens statistische bouwwerk wordt gevormd door klassieke modelleermethoden, variërend van variantieanalyse voor experimenteel design tot path-analyse en structurele vergelijkingsmodellen, dit alles om causale, verklarende modellen van het gedrag van planningsalgoritmen, *information retrieval* systemen of robots te kunnen ontwikkelen.

Voor Cohen dus geen *Querelle des anciens et des*

modernes, maar een integratieve benadering binnen de statistiek, die nodig is om het gedrag van complexe AI-systemen te beschrijven, te beoordelen, te voorspellen, maar bovenal *causaal* te verklaren en zeker ook te *generaliseren* naar nieuwe contexten. AI is een wetenschap waarin AI-programma's en AI-systemen de objecten dienen te zijn van empirische studie. Zij zijn te vergelijken met mechanische, chemische, biologische en psychologische processen en zijn als object van studie niet veel anders dan intelligente dieren als ratten, aldus Cohen. AI-programma's dienen te worden bestudeerd in een dynamische omgeving, waarin de invloed van een verandering in taak T, in omgeving O, of structuur van het systeem S op de *performance* moet worden onderzocht. Dat kan volgens Cohen al lang niet meer door het apparaat open te maken en eigenschappen van het systeem of algoritme rechtstreeks te relateren aan het gedrag om vervolgens één enkele run van het programma uit te voeren en op te vatten als een soort *crucial experiment* dat een antwoord verschaft op de vraag die de onderzoeker aan de natuur stelt. 'It is a comforting fiction or delusion that building a program confers much understanding of its behavior.' Vooruitgang in de AI is een monotoon stijgende functie in een coördinatenstelsel waarbij de x-as achtereenvolgens beschrijving, voorspelling en uiteindelijk causale verklaring representeert en de y-as generalisatie van specifieke taken naar generieke, algemeen toepasbare systemen aanduidt.

Duidelijk is dat Cohen hiermee niet alleen een eigen invulling geeft aan de noties van interne en externe validiteit in een design science context, maar vooral dat hij AI beschouwt als een 'variatie-en-verandering-rijke' discipline, zoals eerder beschreven. Dat laatste wordt nog versterkt door een tweede eigenschap van het boek: de stelselmatige toepassing van de agent-metafoor, waarin systemen geen passieve objecten zijn, maar actieve subjecten, die reactief en proactief handelen in een veranderende omgeving en uitgerust zijn met *high-level cognitive functions*. Cohen schetst een prospectieve visie op de AI middels een soort parabel van Megan, een vrouwelijke agent engineer, die getraind is in human computer interaction, agent-pathologie en agent-ethiek en werkt in een wereld waarin agents zelfstandig functioneren, redeneren met onzekerheid en beslissingen nemen. Daarmee toonde hij zich in 1995 zonder meer een visionair, aangezien autonome systemen (*selfdriving cars, drones, trading systems, autonomous weapons*) al jaren het debat bepalen in AI en Human Computer Interaction. AI met al betoont de AI zich bij de bestudering van de samenwerking tussen menselijke en artificiële actoren nadrukkelijk ten dele een



Paul R. Cohen

sociale wetenschap, hetgeen het 'variatie-en-verandering-rijke' karakter ervan versterkt en a fortiori het belang van de probabilistische wending onderstreept.

Statistiek en AI-controversen

Die *agent-metafoor* vormt eveneens het uitgangspunt van een ander in 1995 verschenen boek, Peter Norvicks en Stuart J. Russells *Artificial Intelligence; a modern approach*, dat inmiddels is uitgegroeid tot een mondiale bestseller, het *state-of-the-art*-tekstboek, dat in meer dan honderd landen wordt gebruikt. Anders dan bij Cohen staat niet de methodologie centraal, maar het kennisdomein zelf: de deelgebieden, paradigma's, toepassingsdomeinen, kernthema's en controversen. Het boek kent inmiddels vier drukken, waarvan de laatste in 2020 verscheen. De drukken verschillen sterk, weerspiegelen de inhoudelijke verschuivingen binnen het vakgebied in de afgelopen 25 jaar en zijn dan ook verplichte kost voor eenieder die de geschiedschrijving van de contemporaine AI ter hand wil nemen. Bovenal verschaft de opeenvolging van vier drukken een uitmuntende bron om de vraag naar de probabilistische wending binnen de *AI als kennisdomein* te beantwoorden. Dat geschiedt hier niet alleen op basis van de traditionele, enigszins geforceerde tegenstelling tussen sterke en zwakke AI, maar vooral tegen de achtergrond van enkele fundamentele controversen die onder meer door (Franklin, 2014) zijn beschreven: symbolische versus subsymbolische AI, representatie versus niet-representatie, gesitueerde versus niet-gesitueerde AI, en tot slot *high level* redeneren versus *low level* perceptie. Opmerkelijk genoeg speelt de betekenis van probabilistische methoden in al deze controversen een belangrijke rol. We beperken ons hier tot de saillante tegenstelling *logic* versus *probability*, die als een rode draad door de geschiedenis van de AI loopt. Dat begon al met het pionierswerk van McCulloch en Pitts uit 1943

op het gebied van neurale netwerken, *A logical calculus of the ideas immanent in nervous activity*. Niet alleen de titel spreekt boekdelen, ook de literatuurlijst is opmerkelijk en kent slechts drie iconische referenties: *The logical syntax of language* van de logisch-positivist Rudolph Carnap, *Grundzüge der theoretischen Logik* door de wiskundige David Hilbert en het monumentale *Principia Mathematica* van A. N. Whitehead en Bertrand Russell. Wellicht wilden McCulloch en Pitts hun nieuwe aanpak een zekere status verschaffen door te verwijzen naar deze vier boegbeelden van de logische traditie en expliciet aansluiting te zoeken bij de wiskundige logica. Deze had immers in de jaren dertig grote vooruitgang geboekt en genoot onder wiskundigen een hoger aanzien dan kansrekening en statistiek. Wellicht ook wilden zij in de voetsporen treden van de 17e eeuwse filosoof Gottfried Wilhelm Leibniz (1646–1716) en diens oude ideaal van een *calculus ratiocinator* of poogden zij voort te bouwen op de 19e eeuwse AI-pionier George Boole (1815–1864), die 'de wetten van het denken' en een theorie van de geest expliciet koppelde aan een taal waarin zowel logica als kansrekening verdisconteerd zijn. Toch lijken beiden de probabilistische betekenis van hun neurale netwerk enigszins te negeren.

Hoe dan ook, in 1958 zou Frank Rosenblatt op basis van hun beider werk zijn perceptron ontwikkelen, een neurale netwerk dat door hemzelf expliciet werd benoemd als een probabilistisch model. De jaren daarop zou Rosenblatt sterke weerstand ondervinden van de logische traditie binnen de AI. Deze domineerde het vak al sedert de oprichtingsconferentie te Dartmouth in 1956 ondanks de aanwezigheid van Claude Shannon aldaar. Diens revolutionaire 'probabilistische' werk betrof niet alleen zijn befaamde informatietheorie, maar ook zijn fundamentele artikel 'Programming a Computer for Playing Chess' dat het pad effende voor veel later AI-onderzoek. Met name Marvin Minsky bestreed op allerlei manieren het perceptron en de geestelijk vader ervan. In 1969 leidde dat tot het boek *Perceptrons; an introduction to computational geometry* geschreven door Minsky en Papert, waarin deze aantoonde dat het perceptron alleen lineair scheidbare data kon classificeren en onder meer de XOR niet aankon. Hun deels onbewezen bewering dat de getoonde bezwaren evenzeer voor multi-layer netwerken en toekomstige netwerken zouden gelden, leidde tot de befaamde 'neurale winter' of 'neural eclipse' (Sejnowski, 2018) die tot eind jaren 80 zou duren. Het prille onderzoek naar neurale netwerken werd de das omgedaan, financiering ervan bleek vrijwel onmogelijk.

Tegen de achtergrond van dit alles vormt de publicatie van Peter Cheesemans polemische artikel 'In Defence of

Probability' in de *Proceedings van de Ninth International Joint Conference on AI* in 1983 een cesuur. De auteur beoogt hierin onder meer dat kansrekening en probabilistische methoden toereikend zijn om automatisch redeneren met onvolledige en onzekere kennis en het in de AI beoogde *common sense* redeneren te bewerkstelligen. Hij hekelt de dominante logische traditie binnen de symbolische AI en stelt dat alle kritiek op probabilistische benaderingen voortkomt uit misverstanden en onkunde. De *sources of error* worden vervolgens ragfijn en vilein geanalyseerd; verwarring over een frequentistisch kansbegrip versus *measures of belief*, verwarring over absolute en relatieve waarschijnlijkheid, over waarschijnlijkheid en de onzekerheid van die waarschijnlijkheid. Ook gispt hij de critici over hun – in zijn ogen – evidente onbegrip van de Bayesiaanse grondslagen. Bovenal betreurt Cheeseman de *proliferation of representation languages with associated inference procedures*, allemaal extensies van de klassieke logica, die voor het realiseren van het ultieme Project van de AI ongeschikt en onnodig zijn. 'Probability is all that is needed', aldus de auteur. Daarmee formuleerde Cheeseman een sterke antithese van de positie van Patrick Hayes die in zijn evenzeer polemische *In Defense of Logic* (1977) onomwonden stelde dat de hoge ambities van de sterke AI het beste met op klassieke logica, taal en semantiek gebaseerde symbolische AI konden worden bewerkstelligd. Een theorie van de menselijke geest stond daarbij centraal, *high level cognitive functions* moesten worden gerepresenteerd, symboolmanipulatie gold als een voldoende en noodzakelijke voorwaarde voor intelligentie en redeneren; een rijke logische taal vormde daarbij het fundament. Hayes' positie was vooral salonfähig bij cognitieve psychologen, *philosophers of mind* en anderen die het ideaal van sterke AI koesterden. Neurale netwerken waren uit den boze en meer in het algemeen speelde de subsymbolische AI, waartoe – terecht of niet – ook de probabilistische methoden werden gerekend, een tweede viool.

Cheesemans positie was onmiskenbaar gedurfd, te meer daar anno 1983 probabilistisch redeneren in de AI nog grotendeels in de kinderschoenen stond. Grote wapenfeiten ontbraken vooralsnog. Er waren noch vooraanstaande paradigma's, noch killer-applicaties (Starmans, 2019). Daar zou spoedig verandering in komen en die ontwikkelingsgang van de probabilistische AI is in de vier opeenvolgende drukken van Norvig en Russells boek goed zichtbaar: de introductie van de Bayesiaanse netwerken, methoden voor *probabilistic causal inference*, de opkomst van *machine learning* in de jaren 90, AI en *information theory*, het spanningsveld tussen *statistical*

learning en machine learning, big data met als voorlopige climax de opkomst van deep learning.

Epiloog

Omdat dit korte essay uiteraard geen definitieve conclusies toelaat en zelfs een voorzichtige tussenbalans prematuur lijkt, beperken we ons tot enkele observaties. Paradoxaal genoeg is het werk van Cohen nauwelijks nog verkrijgbaar, het kent geen grondig herziene herdrukken en schrijver dezes is geen recenter vergelijkbaar tekstboek bekend. Ook Cohens integratieve statistische aanpak is niet de de facto standaard geworden. Dat neemt niet weg dat de auteur op belangrijke onderdelen de plank allerminst misloeg en niet alleen wat betreft zijn prospectieve visie op de AI. Zo lijkt EDA in tijden van big data een eigen paradigma te zijn geworden en vervullen computer-intensieve technieken een belangrijke rol in de verzoening tussen statistical learning en machine learning. Wanneer het gaat om zaken als *overfitting*, *bias-variance trade-off*, het *train-and-test paradigm* en *cross-validation* is die tegenstelling duidelijk minder relevant. Dat geldt ook voor pogingen het toeval te 'temmen' (Hacking, 1990) of beter nog als 'werkpaard' in te zetten (bootstrap, Monte Carlo, et cetera). Toch lijkt de pijler van de klassieke, inferentiële statistiek minder zichtbaar in het moderne AI-onderzoek, hetgeen deels te maken heeft met de kritiek op de standaardaanpak van *hypothesis testing*, die door sommigen wordt beschouwd als een curieuze mengeling van Fishers *significance* testen en de aanpak van Jerzy Neyman (1894–1981) en Egon Pearson (1895–1980). Deels heeft het ook te maken met methodologische oriëntatie; de op *estimation theory* gerichte, wiskundig geschoolde statistici versus het algoritmen- en datastructuren-paradigma van de informatici die de eerste viool spelen in machine learning en data science. Daar staat tegenover dat Cohens klassieke pijler van de verklarende modellen een sterke opmars maakt. Dit niet eens zozeer omdat zij de basis vormen van de grafisch-georiënteerde causale modellen van bijvoorbeeld Judea Pearl, maar met name tegen de achtergrond van de eisen van *responsible en explainable AI* (Starmans, 2020) en de in de inleiding geschetste wijsgerige bezwaren.

Daarnaast lijkt het oude ideaal van Boole in ere te zijn hersteld in de hybride AI, waarin symbolische en subsymbolische methoden en in het kielzog *logic* en *probability* worden geïntegreerd. Vele psychologische studies hebben aangetoond dat klassieke logica geen realistisch model voor redeneren biedt en dat de mens evenmin een

intuïtieve statisticus is. De opkomst en betekenis van het Bayesianisme is ook in Norvig en Russell goed zichtbaar. Dit betreft niet alleen statistiek en informatica, maar ook de in de cognitiewetenschap bestudeerde Bayesiaanse leertheorie als adequaat model voor het representeren van cognitieve functies en processen. Ook de andere genoemde AI-controversen zijn allerminst beslecht.

AI met al is de band tussen AI en statische methoden in de ruimste zin van het woord alleen maar sterker geworden. Oude scheidslijnen, hoezeer ook conceptueel en filosofisch van belang, lijken steeds minder zichtbaar, zeker tegen de achtergrond van het maatschappelijke AI-data science debat, maar ook voor generaties onderzoekers die zijn opgevoed in een paradigma van data science. Onderzoeksprogramma's, onderwijscurricula en beroepsprofielen onderstrepen dit. Tegen deze achtergrond kunnen polemieken over al dan niet rigide demarcaties licht als een enigszins vermoeiend achterhoedegevecht worden opgevat.

REFERENTIES

- Cohen, P. R. (1995). *Empirical Methods for Artificial Intelligence*. MIT Press.
- Franklin, S. (2014). History, motivation and core themes. In K. Frankish & W. M. Ramsey (Eds.), *Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.
- Kruger, L., Daston, L., & Heidelberger, M. (Eds.) (1987). *The probabilistic revolution, Volume 1: Ideas in history*. MIT Press.
- Krüger, L., Gigerenzer, G., & Morgan, M. S. (Eds.) (1990). *The probabilistic revolution, Volume 2: Ideas in the sciences*. MIT Press.
- Hacking, I. (1990). *The taming of chance*. Cambridge University Press.
- Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. Henry Holt and Company.
- Sejnowski, T. J. (2018). *The deep learning revolution; Artificial intelligence meets human intelligence*. MIT Press.
- Starmans, R. J. C. M. (2018). The Predicament of truth: On statistics, causality, physics and the philosophy of science. In M. J. Van der Laan & S. Rose (Eds.), *Targeted learning in data science: Causal inference for complex longitudinal studies*. Springer.
- Starmans, R. J. C. M. (2019). De apologie voorbij; kansrekening en het feilbare denken. *STATOR*, 19(1).
- Starmans, R. J. C. M. (2020). Prometheus unbound or paradise regained; The concept of causality in the contemporary AI-data science debate. *Journal of the French Statistical Society*, 161(1), 4-41. Special Issue on Causality, Antoine Chambaz (Ed).

RICHARD STARMANS is verbonden aan de Faculteit Bètawetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht en aan Tilburg University. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica.
E-mail: starmans@cs.uu.nl

HENK TIJMS

column

Foto: Thanasis Papazacharias via Pixabay



KERSTPUZZELS EN DE VERSCHRIKKELIJKE KERSTMAN

Een beroemd probleem is het honderd-gevangenen-probleem dat eerder onderwerp van een column was. Waar ging dit probleem over? Honderd gevangenen worden door een bewaker bijeengeroepen. Hen wordt meegedeeld dat besloten is om ze over te plaatsen. De overplaatsing is naar een berucht strafkamp waar hen een vrijwel zekere dood wacht. Als doekje voor het bloeden doet de gevangenisdirecteur de gevangenen een voorstel dat tot hun vrijlating kan leiden. Op een lange tafel staat een rij van 100 afgesloten, genummerde dozen. In willekeurige volgorde is de naam van elke gevangene in precies één doos opgenomen. De gevangenen mogen één voor één hooguit 50 dozen openen om hun naam te vinden. De gevangenen mogen tijdens de zoektocht geen onderling contact hebben en ze mogen ook niet dozen verplaatsen of namen in de dozen verwisselen. Als het iedere gevangene lukt zijn naam te vinden, dan worden de gevangenen vrijgelaten. Het lijkt een voor de gevangenen kansloze missie. Als iedere gevangene 50 willekeurig gekozen dozen onderzoekt, dan is de kans op vrijlating van de groep gelijk $(\frac{1}{2})^{100}$. Eén van de gevangenen, een aan lager wal geraakte investeringsbankier, komt echter met de mededeling dat ze niet geheel kansloos zijn, maar dat met een kans van iets meer dan 30% de vrijheid lonkt als de gevangenen samenwerken en zijn instructies opvolgen. Wat zijn deze? De dozen worden één op één aan de gevangenen

toegewezen en elke gevangene noteert op een papiertje welke doos bij welke gevangene hoort. Een gevangene gaat eerst naar de aan hem toegewezen doos. Treft hij daar een andere naam dan de zijne, dan gaat hij vervolgens naar de doos die aan de andere naam toegewezen is. Dit gaat voort totdat de gevangene zijn eigen naam vindt of vruchteloos 50 dozen heeft geïnspecteerd. De investeringsbankier rekent de gevangenen voor dat de kans op vrijlating dan de verrassend hoge waarde van 31,2% heeft. De berekeningen zijn vrij ingewikkeld en minder geschikt om onder de kerstboom uit te leggen. Dit ligt anders voor de volgende twee kanspuzzels die ook de kerstboodschap 'samenwerken loont' uitdragen en waarmee je onder de kerstboom een slim neefje of nichtje kan uitdagen.

Een TV-spel

Omroep Laks, de omroep voor jongeren, heeft wekelijks het educatieve tv-programma *Wiskunde kun je leren*, waarin ook aandacht wordt besteed aan kansrekening. Aan het eind van het programma wordt altijd een kansspelletje gespeeld waarmee 5000 euro valt te winnen. Twee personen worden willekeurig gekozen uit het publiek. Deze twee personen worden elk in een apart hokje geplaatst. In het hokje gooit elk van hen een zuivere munt op. De uitkomst van de worp

is niet zichtbaar voor de andere persoon maar wel voor het publiek. Daarna moet elk van hen raden wat de ander heeft geworpen, kop of munt. Hebben ze het beiden juist, dan krijgen ze elk 2500 euro. Niet een bedrag waarmee de omroep zich een buil valt, zeker niet omdat tot nu toe de omroep gemiddeld slechts één in de vier keer tot uitbetaling hoefde over te gaan. Jij bent bij de uitzending van het programma aanwezig en tot je verrassing word jij met een ander naar voren geroepen om het spel te spelen. Terwijl jullie naar voren lopen, fluister je de metgezel een aanpak van samenwerking in die met een kans groter dan 25% de 5.000 euro oplevert. Wat is deze aanpak en wat is de winkans?

Monty Hall 2.0

In een tv-show staat een vader met zijn dochter voor drie gesloten deuren genummerd als 1, 2 en 3. Eén deur verbergt een auto, een andere deur de autosleutel en de overgebleven deur een fopprijs. Elk van hen mag achter maximaal twee deuren kijken, één deur tegelijk, buiten het zicht van de ander en zonder deze een aanwijzing te verschaffen. Alleen als de vader de auto treft en de dochter de sleutel, dan winnen ze de auto. Wel mogen ze vooraf een strategie afspreken over hoe te werk te gaan bij het openen van de deuren. Wat is een optimale strategie en wat is de maximale winkans?

Als het slimme neefje of nichtje de bovenstaande twee puzzels fluitend heeft opgelost, dan mag hij of zij de tanden zetten in de volgende puzzel.

Het laatste chocolaatje

In een kerstboom hangen tien chocolaatjes, waarvan twee wit en acht puur. Je pakt chocolaatjes random uit de boom, één tegelijk, en eet deze op totdat je een chocolaatje van de andere kleur pakt. Dit chocolaatje hang je weer terug en daarna zet je het procedé voort met de overgebleven chocolaatjes. Wat is de kans dat het laatste chocolaatje dat je opeet wit is?

In de volgende puzzel, waarin de kerstgedachte ver te zoeken is, berust elke overeenkomst met een bestaande situatie of bestaande personen op louter toeval.

De boosaardige Kerstman

In het Vrije Universum leefden lange tijd Sinterklaas en de Kerstman in betrekkelijke vrede samen. Maar op een

dag krijgen Sinterklaas en zijn medewerkers onverwacht te horen dat ze geen dobbel- en kaartspellen meer in hun pakjes mogen stoppen, hoewel ze altijd een voortreffelijke en door de kinderen hogelijk gewaardeerde keuze hebben gemaakt uit het ruime assortiment. Het geven van deze spellen is voortaan voorbehouden aan de niets ontziende, op macht beluste Kerstman die zo met het argument van grotere belasting extra rendieren wil aanschaffen om voor zijn slee te spannen. Deze kwalijke gang van zaken leidt tot een scherpe reactie van een gerespecteerde emeritus Sinterklaas die stelt dat het algemeen belang van een goed surprisepakket ondergeschikt gemaakt wordt aan de platte koopmansgeest van een Kerstman voor wie alleen omzet telt en kwaliteit er niet toe doet. Lichtelijk geschrokken door deze felle reactie doet de voorzitter van de Bond voor het Educatieve Kansspel op gereformeerde grondslag de toezegging, na overleg met de kerstman, de Sint-groep een kans te bieden om het geschenkenrecht tot het geven van dobbel- en kaartspellen te behouden. Zonder dat ze van tevoren op de hoogte zijn, worden de tien stafleden van de Sint-groep bijeengeroepen en elk apart in een kamer gezet. Ieder van hen krijgt een zuivere munt in de handen gedrukt en moet binnen vijf minuten beslissen of hij/zij deze munt wel of niet opgooit, waarbij overleg tussen de stafleden niet mogelijk is. De Sint-groep krijgt het geschenkenrecht op dobbel- en kaartspellen alleen dan terug als tenminste één van de tien stafleden de munt opgooit en als daarbij elk van de opgegooide munten kop geeft. De stafleden van de Sint-groep zijn echter bij uitstek getraind in logisch denken, dat moet ook wel als je jarenlang de gevaarvolle tochten over steile daken moet organiseren. Voor hen is het dan ook direct duidelijk dat ze ieder volgens eenzelfde beslisregel moeten handelen. Wat is deze beslisregel en wat is de kans dat ze het geschenkenrecht op dobbel- en kaartspellen terug krijgen? Laat ik verklappen dat randomisatie het sleutelwoord tot de oplossing is. De kans dat de tien stafleden van de Sint-groep succesvol zijn in het terughalen van het geschenkenrecht voor dobbel- en kaartspellen is 26,3%, waar de Kerstman en zijn kliek de kans als verwaarloosbaar hadden ingeschat. Zelfs als honderd en niet tien stafleden van de Sint aan de test zouden zijn onderworpen, zou de kans op succes nog steeds iets meer dan 25% zijn geweest!

HENK TIJMS is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Voor vwo-scholieren schreef hij het Zebraboekje *Kansrekening in Werking – een moderne aanpak* dat verschenen is bij de wiskundeuitgeverij Epsilon in Amsterdam.
E-mail: h.c.tijms@xs4all.nl

Foto: Polina Tankilevitch via pexels



Nieuwe statistiek voegt wereldwijd corona-onderzoek samen

FENNA VAN DER GRIENT

Een veelbelovende troef in de strijd tegen corona is het vaccin tegen tuberculose: het BCG-vaccin. Dit vaccin geeft het afweersysteem een boost, waardoor het niet alleen beschermt tegen tuberculose, maar het lichaam ook weerbaarder maakt tegen andere infecties. Dit maakt de gevolgen van een coronabesmetting mogelijk minder ernstig.

Op tal van plaatsen in de wereld wordt onderzocht welke rol een tuberculosevaccin kan spelen bij de bestrijding van COVID-19. Twee Nederlandse statistici gebruiken een nieuwe methode om al die onderzoeken met elkaar te combineren.

Onomstotelijk bewijs is echter nog niet geleverd. Zo'n 15 onderzoeksgroepen wereldwijd hebben studies opgezet om de invloed van het BCG-vaccin op het ziekteverloop van corona te onderzoeken. Dat levert een gigantische berg gegevens op. Er is alleen één probleem: reguliere statistische methoden maken het heel lastig om deze onderzoeksresultaten tussentijds met elkaar te



Judith ter Schure



Peter Grünwald

combineren. En dat terwijl het in deze tijd van essentieel belang is om zo snel mogelijk tot een conclusie te komen. Daarom gebruiken statistici Peter Grünwald en Judith ter Schure, werkzaam aan het Amsterdamse Centrum Wiskunde & Informatica (CWI), een compleet nieuw soort statistiek. Hiermee kunnen zij de gegevens van alle onderzoeken combineren en aan de bel trekken zodra er voldoende bewijs is voor de effectiviteit van het vaccin.

Toeval of niet?

In een medische studie vergelijk je meestal twee groepen met elkaar. In dit geval heeft één groep het BCG-vaccin toegediend gekregen en de andere groep, de controlegroep, een placebo. Vervolgens ga je kijken of je een verschil kunt ontdekken tussen deze twee groepen. Er komt echter altijd een toevalsfactor bij kijken. In de ene groep kunnen ook door puur toeval net iets minder mensen ziek worden. Wanneer vind je het verschil tussen de twee groepen groot genoeg om te zeggen: het vaccin doet iets? Om die vraag te beantwoorden heb je statistiek nodig.

Het nadeel van klassieke statistische methoden is dat je van te voren altijd precies moet vastleggen hoeveel patiënten je meeneemt in je onderzoek. 'Je kunt niet eerder stoppen, of bijvoorbeeld langer doorgaan als het er veel-

belovend uit ziet', zegt Grünwald. 'Af en toe zal het door puur toeval lijken alsof een vaccin werkt. Als je alleen in die gevallen verder gaat met het onderzoek, of juist precies na een toevallige reeks goede resultaten stopt, dan lijkt het bewijs groter dan het eigenlijk is.'

Te weinig patiënten

Bij het combineren van meerdere studies vormt dit ook een groot probleem. Je wilt kunnen stoppen zodra je genoeg gegevens hebt verzameld om te zeggen: het werkt. Op dat moment kun je het vaccin gaan inzetten om de gevolgen van een coronabesmetting te beperken. Maar met klassieke statistiek zou je van te voren precies moeten bepalen hoeveel patiënten elke studie meeneemt en moeten wachten tot alle studies voltooid zijn.

En dat kan wel even duren. 'Afgelopen zomer waren er heel weinig covidpatiënten,' vertelt Ter Schure. 'Heel veel studies konden daardoor niet worden afgerond. Dan kun je de resultaten heel moeilijk samenvoegen. Onze methode, die we *safe testing* noemen, maakt dit wel mogelijk. Het was dus echt een vooruitziende blik van Henri van Werkhoven, de statisticus van de Nederlandse BCG-studies, om *safe testing* in te zetten.'

Roulette

Met *safe testing* mag je wel midden in een studie stoppen of langer doorgaan. Dat maakt het ook mogelijk om de resultaten van alle onderzoeken met elkaar te combineren, zonder dat je van te voren de details van alle afzonderlijke studies weet. 'Onze methode werkt eigenlijk hetzelfde als een weddenschap', vertelt Grünwald. Uit elke studie komt een getalletje rollen dat aangeeft hoeveel bewijs er is dat het vaccin effect heeft: de E-waarde. 'Die E-waarde kun je zien als virtueel geld. Je kunt er geld op inzetten dat de mensen die het echte vaccin kregen minder vaak ernstig ziek worden dan mensen in de placebogroep. Als je dan veel geld wint, dan doet het vaccin waarschijnlijk inderdaad iets (of je hebt ongelofelijk veel geluk gehad).'

Grünwald: 'Je kunt het vergelijken met het inzetten in een casino. Stel dat je er achter probeert te komen of het roulettewiel eerlijk is of niet. Als het wiel eerlijk is, is de kans altijd heel klein dat je veel geld verdient, wanneer je ook stopt. Als je dus begint met 100 euro, en op een moment 10.000 euro in handen hebt, dan is het wiel waarschijnlijk niet eerlijk. Het maakt dan niet uit wanneer je stopt.' In plaats van zelf een week lang in het casino te vertoeven, kun je ook verschillende mensen achter elkaar het roulettewiel laten testen. Dan heb je dezelfde situatie als het samenvoegen van meerdere studies.

'Maar', benadrukt Ter Schure, 'Het is wel belangrijk dat je door blijft spelen met hetzelfde geld. De tweede tester gaat dus door met het bedrag waarmee de vorige is geëindigd. Zo voeg je de onderzoeken daadwerkelijk samen. Als je 150 mensen tegelijk met honderd euro aan de roulettetafel laat spelen is er altijd wel toevallig één iemand die veel geld wint. Dat hoeft niet te betekenen dat het casino niet eerlijk is.'

Dashboard

Een hoge E-waarde zegt hetzelfde als een hoog gewonnen bedrag in het casino: er is meer aan de hand dan

je op grond van puur toeval zou verwachten. Het vaccin doet iets. Ter Schure houdt op een dashboard het verloop van de resultaten in de gaten. Passeert de gezamenlijke E-waarde een bepaalde drempel, dan is er voldoende bewijs dat het vaccin effect heeft.

Ter Schure: 'Het meten van zo'n gezamenlijke E-waarde brengt een grote verantwoordelijkheid met zich mee. Wat doe je op het moment dat de E-waarde de drempel passeert? Ga je dan direct iedereen vaccineren? We gaan nu nog met alle deelnemende studies afspreken hoe we die beslissing willen maken. Het is belangrijk om dit vooraf vast te leggen, zodat iedereen vertrouwen heeft in de drempel en weet hoe er gehandeld wordt als het bewijs de drempel haalt. Gelukkig is de bijdrage van iedere studie aan het totaal heel inzichtelijk. Stel dat één onderzoek een gigantisch hoge E-waarde heeft vergeleken met alle andere onderzoeken, dan zie je dat meteen.'

Andere mindset

De coronacrisis is de vuurdoop voor de *safe testing*-methode. 'Het is uniek dat er zoveel studies op exact hetzelfde moment lopen', zegt Ter Schure. 'Daarnaast heeft iedereen natuurlijk veel haast. Ik hoop ook eigenlijk dat dit kan zorgen voor een verandering van mindset in de wetenschap. Nu belanden veel onderzoeken die net niet overtuigend genoeg zijn in een la. Omdat onze combinatie methode zo simpel is, wordt het veel makkelijker om samen te werken. Zo kan iedere studie bijdragen, hoe klein ook. Ik hoop dit de wetenschap wat minder 'ieder voor zich' zal maken.'

Het interview is met dank overgenomen; het verscheen eerder in september als blog op de website van *New Scientist* en in *Het Parool* van 19 september 2020. Judith ter Schure heeft een verder, meer gedetailleerd, artikel over dit onderwerp toegezegd voor een van de volgende nummers van *STAtOR*.

FENNA VAN DER GRIENT is freelance wetenschapsjournalist en schrijft voor onder andere *New Scientist* en *Het Parool*. Ze studeerde Technische Natuurkunde aan de Universiteit Twente. E-mail: fenna.vd.grient@gmail.com

*All lies and jests
still a man hears what he wants to hear
and disregards the rest*

Paul Simon (The Boxer, 1968)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Fabels

Het is 1889, het jaar waarin de *Wall Street Journal* voor het eerst verschijnt, de Eiffeltoren tijdens de Wereldtentoonstelling in Parijs wordt geopend en in Japan, Nintendo Koppai (later Nintendo Company, Limited) wordt opgericht. De wereld is volop in beweging, nieuwe technologische uitvindingen versnellen de stedelijke groei, waardoor efficiëntere fabrieken en beter transport mogelijk worden. Een van de meest dramatische verbeteringen is de uitvinding van de elektrische lamp door Thomas Edison. Zijn uitvinding maakt verlichting praktischer en verandert het leven volledig. Het noodlot slaat echter toe als aan het einde van het jaar de eerste griepandemie uitbreekt, de Russische griep. In korte tijd eist het virus meer dan een miljoen slachtoffers. De medische wetenschap tast in het duister over de oorzaak, een remedie is er niet. Als gevolg van deze grote onzekerheid groeit het geloof in samenzweringstheorieën sterk en gaan de meest absurde verhalen over de oorzaak van de ziekte rond. Naast sterrenstof en vulkanische as zitten ook elektriciteit (Knap, 2020), elektrisch licht en het toenemend aantal telegraafpalen op het beklagdenbankje. Als je de (sociale) media van vandaag bekijkt en leest hoe in sommige groepen in onze maatschappij over de oorzaak van COVID-19 wordt gedacht lijken we niet veel vooruitgang geboekt te hebben in de afgelopen 130 jaar om feit van fabel te leren onderscheiden.

Complotten zijn van alle tijden

Met verbazing hoor en zie ik hoe mensen de huidige corona-pandemie proberen te verklaren. Sommige zien

de plaatsing van 5G-zendmasten als de boosdoener. Anderen zien Bil Gates, de Chinese overheid of de *Deep State*, als de kwade genius. Samenzweringstheorieën en bijgeloof doen het altijd goed in tijden van crisis. De geschiedenisboeken staan er vol mee. Zo ging in de 14de eeuw het gerucht rond dat de uitbraak van builenpest in delen van Europa het gevolg was van een internationale samenzwering (Fordham University, 2020) van het jodendom om het christendom te vergifigen. In 1918, de Spaanse-griepandemie is dan in volle gang, ging het gerucht dat de aspirinetabletten (Cohut, 2020) van Bayer de veroorzaker waren. Ondanks onze enorme wetenschappelijke vooruitgang blijven samenzwerings-theorieën hardnekkig de kop opsteken, hoe ongeloofwaardig ze soms ook zijn. Neem Pizzagate (Wikipedia). Ondanks dat deze samenzweringstheorie allang naar het land naar fabelen is verwezen en bovendien een VS aangelegenheid is, staan er mensen op het Museumplein te protesteren in de overtuiging dat het Pizzagate-verhaal waar is en dat ook de Nederlandse elite er onderdeel van is.

Onwrikbare overtuiging

Een samenzwering opzetten en geheimhouden is niet eenvoudig, zeker niet als het een grootschalige is, zoals het in scene zetten van een maanlanding of het in het geheim ontwikkelen en verspreiden van een dodelijk virus. David Robert Grimes, schrijver van *The irrational ape*, ontwikkelde een model (Grimes, 2016) dat kan worden gebruikt om de houdbaarheid van een samen-

zweringstheorie te onderzoeken. Met zijn op een Poissonproces gebaseerd model maakte hij schattingen van hoe lang het duurt voordat een samenzwering uitkomt. Het model laat zien dat samenzweringen waar veel mensen bij betrokken zijn, met hoge kans snel worden onthult, voor de maanlanding binnen 4 jaar. We hoeven echter niet 4 jaar te wachten om uit te vinden of de verspreiding van COVID-19 een samenzwering is. De hele tijd komen nieuwe feiten aan het licht die we kunnen gebruiken om ons geloof in een samenzwering bij te stellen. Het mechanisme waarmee we dat gestructureerd kunnen doen is de formule van Bayes. Vanuit welke beginsituatie we ook vertrekken (al of niet geloven in een complot) formaliseert de formule van Bayes hoe nieuwe feiten leiden tot het bijstellen van onze overtuiging. In de praktijk blijkt echter dat we dat nauwelijks doen omdat we als gevolg van de *confirmation bias* alleen die feiten die onze huidige overtuiging bekrachtigen in ogenschouw nemen. De door de (sociale) media veroorzaakte filterbubbels lijken dit effect verder te versterken.

De regel van Cromwell

Dogmatisme staat vooruitgang in de weg, de formule van Bayes leert ons dat direct. Een onverzettelijk geloof in het al of niet bestaan van een complot, wat zich vertaalt in een prior van 0 of 1, zorgt ervoor dat geen enkel nieuw feit invloed heeft op onze overtuiging, wat absurd zou zijn. Laat dus ruimte en accepteer, met een kleine waarschijnlijkheid, dat je het misschien bij het verkeerde eind kunt

hebben. Deze regel wordt ook wel de Regel van Cromwell genoemd en komt voort uit de oproep van Cromwell (Lindley, 1985) in 1650 aan de Church of Scotland om 'te denken dat het mogelijk is dat u zich vergist'. Met andere woorden, blijf openstaan voor andere opvattingen en nieuwe inzichten. Een krachtig idee dat zonder moeite in het achterhoofd kan worden gehouden, ook door de complotdenker, en voorkomt een hoop ongenueanceerd geschreeuw tot zelfs de bedreiging van politici. Met een beetje gezonde twijfel in het achterhoofd kunnen we ieder nieuw feit wegen en zo fabel van werkelijkheid onderscheiden.

LITERATUUR

- Knapp, A. (2020). The Original Plandemic: Unmasking the eerily parallel conspiracy theories behind the Russian flu of 1889. *Forbes*, 15 mei 2020. <https://bit.ly/statorknapp39mccGQ>.
- Fordham University. *Jewish history sourcebook: The Black Death and the Jews 1348–1349 CE*. Fordham University. <https://bit.ly/statorsourcebooks3fEriIP>
- Cohut, M. (2020). The flu pandemic of 1918 and early conspiracy theories; The controversy around aspirin. *Medical news today*, 29 september 2020. <https://bit.ly/statorcohut301QE6b>
- Wikipedia. Pizzagate. <https://nl.wikipedia.org/wiki/Pizzagate>
- Grimes, D. R. (2016). On the viability of conspiratorial beliefs. *PLoS ONE*, 11(1): e0147905. <https://doi.org/10.1371/journal.pone.0147905>
- Lindley, D. (1985). *Making decisions*. Wiley. (p.104). <https://archive.org/details/makingdecisionsoodvli/page/104/mode/2up>

JOHN POPPELAARS, Directeur LITIC B.V.
E-mail: john.poppelaars@litic.com



BOB VAN LIMBURG

Case

Elise wacht al twee jaar op hulp bij haar eetstoornis. Margriet is haar zus en is precies een jaar ouder dan Elise. Steeds weer wordt Elise afgewezen op zoek naar hulp en belandt ze op een nieuwe wachtlijst. Het argument voor deze afwijzingen was vaak dat men Elise te ingewikkeld vond. Zeker, ze was nogal angstig geworden en bij de GGZ vond men dat ze eerst daar maar aan moest gaan werken. ‘Wij behandelen maar één klacht tegelijk’ had men haar steeds gezegd. Jaren terug was Elise fors afgevallen en toch kreeg ze complimenten: ‘Wat zie je er goed uit!’ Ze was onzeker, en de complimenten brachten haar op een idee: blijven afvallen. Elise lijdt nu al vijf jaar aan anorexia. Elise die in Eindhoven woont en studeert, stond al twee jaar op verschillende wachtlijsten. Ze meldde zich begin 2017 bij de GGZ die haar anorexia zou kunnen behandelen. Na vier maanden wachten kreeg ze te horen: je bent te ingewikkeld voor ons, ga eerst maar aan je trauma werken. Dat ging ze doen, en werd opgenomen in een kliniek. Intussen viel Elise weer negen kilo af. Elise: ‘Ik kreeg

In dit artikel worden problemen met wachttijden bij de GGZ geanalyseerd. De GGZ wordt een complex adaptief systeem genoemd. Uit de analyse blijkt dat het probleem niet de te hoog veronderstelde gemiddelde wachttijd is, maar juist de spreiding daarvan. Die spreiding is weer het gevolg van de complexiteit binnen de GGZ. Met behulp van een simulatie wordt aangetoond dat de combinatie van een laag gemiddelde en een hoge spreiding een hoog gesimuleerd gemiddelde kan opleveren.

Für Elise WACHTTIJDEN BIJ DE GGZ

tijdens die opname paniekaanvallen, en de enige manier die ik ken om ermee om te gaan is: Honger. En dat is nog steeds zo.’

Na de opname meldde ze zich hoopvol bij de GGZ-instelling die haar de eerste keer had afgewezen. Het bleek dat die haar nog steeds ‘te ingewikkeld’ vond: weer afgewezen.

Structuur van de GGZ

De GGZ weergegeven als een systeem op het eerste niveau van de hiërarchie van Boulding (het raamwerk) is even absurd als de mens weergegeven als een systeem van niveau twee (het uurwerk). (Wikipedia)

De GGZ is een complex systeem. Dat komt omdat iedere behandeling maatwerk is, en er veel terugkoppelingen zijn, maar ook terugverwijzingen. Belangrijke vraag: wat is de structuur van een GGZ? Er zijn twee wachtlijsten achter elkaar, die van de intake en van de behandeling. Zie figuur 1.



Foto: goodluz

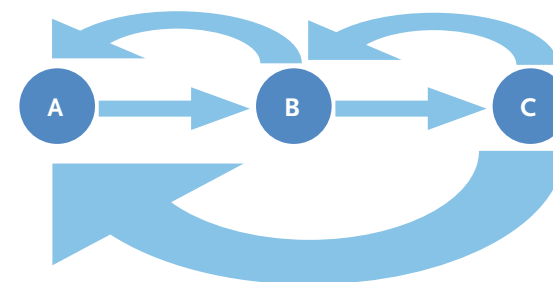
In A bevinden zich de mensen die om hulp vragen, bij B vindt de intake plaats. De pijl tussen A en B staat voor de wachtrij voor de intake. In C vindt de behandeling plaats, de pijl tussen B en C is de wachtrij voor de behandeling. Er wordt veel heen en terug verwezen, en dat veroorzaakt complexe bewegingen. Er kan tijd worden gewonnen bij de administratieve rompslomp die 10 tot 33 procent van de tijd opslokt.

In augustus 2018 had Elise zich weer gemeld bij de kliniek. Tevergeefs, zo bleek zes weken later, nu was de reden:

‘We mogen je niet helpen, ons budgetplafond is bereikt.’ Die instelling krijgt tot het einde van het jaar geen vergoeding meer voor de behandeling van anorexia patiënten.

Statistische analyse I

Margriet heeft twee katten, Mijnheer en Janssen. Ze woont op de begane grond en Mijnheer wil nog wel eens een ommetje maken. Dat kan veilig, het appartement ligt in een verkeersluw gebied. Janssen is een binnenkat en wil niks met buiten te maken hebben. Margriet las in de krant: ‘15.000 patiënten wachten minstens 19 weken op hulp. En 20.000 wachten minimaal 15 weken.’ (NRC, 5 december 2018). Margriet heeft informatica gestudeerd en wist het een en ander van statistiek. Ze besloot de data te gaan bewerken: ‘Allereerst delen we de aantallen wachtenden door de totale populatie in de wachtrij, $80.000: 15 \div 80 = 19\%$. $20 \div 80 = 25\%$.’ Ze pakte er statistische tabellen bij en noteerde: ‘We nemen een normaalverdeling aan, de z waarden zijn 0,88 respectievelijk 0,67.’



Figuur 1. Structuur van de GGZ

‘Eerst de standaarddeviatie berekenen,’ dacht Margriet. Ze noteerde: ‘ $X - \mu = z\sigma$ ’, en voorts: ‘ $15 - \mu = 0,67\sigma$, $19 - \mu = 0,88\sigma$. Aftrekken levert $-4 = -0,21\sigma$ dus $\sigma = 4 \div 0,21$ dus $\sigma = 19,1$ week en een μ van 2,2 week.’

‘Oei die standaarddeviatie is hoog!’, constateerde Margriet. ‘Een zo’n hoge standaarddeviatie en lage μ betekent dat we te maken hebben met turbulentie of chaos in het systeem.’

Bij die hoge spreiding moeten we even bij stilstaan. Mensen die bij de GGZ werken zijn erg sociaal, en nemen graag diensten waar, bijvoorbeeld bij ziekte. Medewerkers sturen een cliënt vaak naar een collega voor een second opinion. Er wordt dus veel terugverwezen, en dat veroorzaakt de turbulentie. GGZ-instellingen zijn termietenhopen waar alles kriskras door elkaar loopt, en als je er van bovenaf naar kijkt zie je een bepaald patroon.

De wachttijden ontstaat door de turbulentie. Het werkelijke gemiddelde, μ , is gelijk aan 2,2 week, en dat is veel lager dan de veertien weken die wordt getolereerd. De verwachtingswaarde is dus prima, maar er is een grote spreiding in het systeem. Met bovengenoemde gegevens is de kans dat 14 weken wordt overschreden gelijk aan $14 - 2,2 = 11,8 \div 19,1$; deze berekening geeft een z score van 0,62. De bijbehorende overschrijdingskans levert 27% kans, en dat is niet uitzonderlijk. De ‘schuldige’ is

* Opgegeven gemiddelde	2,2
* Opgegeven standaarddeviatie	19,1
* Aantal simulaties	1000
* Berekend overall gemiddelde	15,9
* Langste wachttijd	65,8
* Kortste wachttijd	0,2
* Variatiebreedte wachttijd	65,6
* Gemiddelde wachttijd intake	2,21
* Gemiddelde wachttijd behandeling	5,27
* Overall gemiddelde	7,48
* Gesimuleerd overall gemiddelde	15,90
* Verschil	8,40 weken
Deze ruim acht weken zijn onverklaarbaar en moeten worden toegewezen aan emergenties van het complex adaptief systeem dat de GGZ is. Het probleem in de wachttijden is dus bovenal een organisatieprobleem. Niet alleen bij de GGZ treffen we deze problematiek aan. Ook in de jeugdzorg treffen we het opmerkelijke verschijnsel dat tussen 2015 en 2018 het volume aanvragen met achttien procent steeg, en de kosten met 43%. Er zit dus een gat tussen kosten en volume van 25%, dat niet kan worden verklaard.	

Kader 1. Simulatiemodel van de Nederlandse zorg (Bob van Limburg met good old visual basic 6, 15-11-2019)

Aantal simulaties	1000
Gemiddelde wachttijd intake	2,21
Gemiddelde wachttijd behandeling	5,27
Overall totaal gemiddelde	7,48 (verklaard), afgerond tot 7,50
Gesimuleerd empirisch gemiddelde	15,90
Onverklaard verschil	8,40 weken
Onverklaard verschil in procent	52,8%'

Kader 2. Resultaten van de simulaties

de hoge σ , en niet het gemiddelde, terwijl we daar wel op gefocused zijn. In kader 1 zijn de resultaten van een aantal computersimulaties weergegeven. Indien we aannemen dat de wachttijden van de intake (maximaal vier weken) en van de behandeling na de intake (maximaal tien weken) uniform verdeeld zijn, dan komen we na duizend simulaties op een gemiddelde wachttijd van $2,21 + 5,27 = 7,48$ weken. Dit zou het *moeten* zijn. Indien we echter een duizend simulaties uitvoeren met $\sigma = 19,1$ week en een μ van 2,2, waarbij we aannemen dat deze σ en μ de parameters van een normale verdeling zijn, dan komen we op een gesimuleerd gemiddelde van 15,90 weken. Dat is het *empirisch* gemiddelde van de wachttijden. Zie kader 2, een samenvatting van de resultaten van de simulatie.

Mijnheer kwam klaaglijk miauwend binnengelopen, en Margriet stopte met typen: ‘Nee Mijnheer, het is nog lang geen etenstijd. Nog dik een uur, je wacht maar even. Je hebt al een half uur van de wachttijd afgesnoept, eerst aten jullie om zeven uur, nu om half zeven. En trouwens, ik ben met wachttijden bezig die je niet zo makkelijk kunt sturen.’ Mijnheer droop af, hij leek het te hebben begrepen. Janssen kwam intussen binnengelopen en groette Mijnheer even met een spinnend geluid, nam een aanloop en sprong op de schoot van Margriet. Margriet typte verder:

‘De hoge waarde van σ van ruim 19 weken toont de enorme verscheidenheid in de wachttijden. Dat blijkt ook uit de bijlage, waar de variatiebreedte in de wachttijden maar liefst 65,6 weken bedraagt. We hebben te maken hebben met een Complex Adaptief Systeem (CAS) dat onder andere wordt gekenmerkt door diversiteit en constante uitwisselingen van informatie tussen specialisaties binnen de GGZ die weer wordt teruggekoppeld waardoor een niet lineaire spiraalvormige flow ontstaat.’ Het geheel doet chaotisch aan (Gleick, 1996).

Lange wachttijden zijn bij een hoge standaarddeviatie

volkomen normaal. Stel we berekenen het 95% eenzijdige betrouwbaarheidsinterval. Met 95% betrouwbaarheid kan het dan zo zijn dat de maximale wachttijd dan bedraagt: $X = \mu + z\sigma \Rightarrow X = 2,2 + 1,645 \times 19,1 \Rightarrow 2,2 + 32,9 = 35,1$ weken, dus maar liefst bijna 9 maanden, en dit past bij het 95 procent eenzijdig betrouwbaarheidsinterval.

De GGZ is een complex adaptief systeem. Maar wat is complexiteit?

Complexe problemen

All things considered, the only way to solve these wicked problems is to create an equally wicked but wise solution.

In 1959 stelde Lindblom in zijn artikel ‘The science of muddling through’ dat complexe en veranderlijke problemen zich alleen incrementeel laten oplossen, in kleine stappen die niet planbaar zijn. De term *wicked problems* werd voor het eerst gebezigd in een publicatie van Rittel en Webber (1973) over onhanteerbare cases bij ontwerp- en planningsvraagstukken. Zij benoemden tien eigenschappen van *wicked* (ook wel venijnig genoemd) *problems*:

- een complete omschrijving van het probleem is niet mogelijk;
- een oplossing van een *wicked problem* is nooit definitief;
- elke gekozen oplossing roept weer (externe) effecten op die om actie vragen;
- er is een veelheid aan oplossingen;
- ze vragen om een unieke benadering en oplossing;
- een *wicked problem* is een symptoom van achterliggende problemen;
- *wicked problems* zijn onbegrensd, ze blijven altijd in enigerlei vorm bestaan;
- oplossingen van *wicked problems* zijn alleen maar subjectief als goed of slecht te benoemen;
- de oorzaken van *wicked problems* zijn op veel manieren te duiden;
- de probleemoplosser heeft geen speelruimte maar moet er toch voor zorgen dat hij het goed doet.

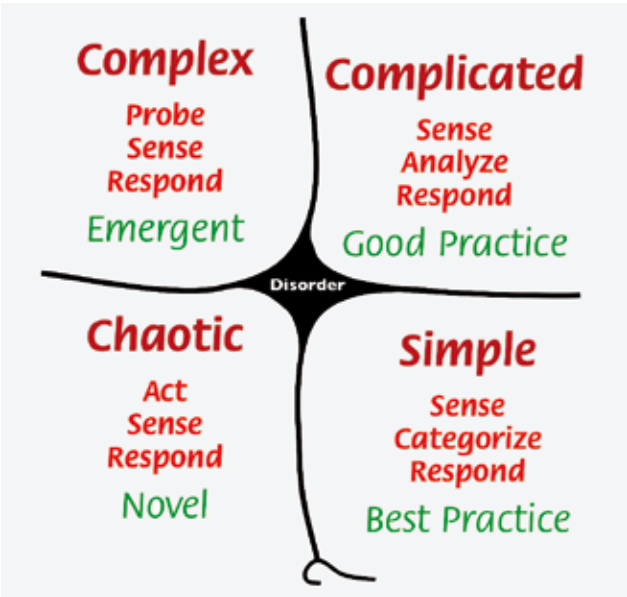
Complexe problemen kunnen omslaan ofwel in chaos (het op hol slaan van het systeem) ofwel leiden tot rigiditeit en stagnatie (figuur 2). Een CAS past bij de organische wereldoriëntatie, het is een levend en lerend systeem. In de levende natuur zijn veel voorbeelden te vinden van zelf-scheppende (autopoiesis) systemen. Voorbeelden zijn: een vlucht spreuwen, een V-formatie van ganzen, een

termietenhoop. spontane orde in een verkeerschaos van voetgangers in de binnenstad en bij golfbewegingen (waves) van voetbalsupporters in een stadion. Bij een zelf-scheppend systeem ontbreekt centrale sturing. Verval of slijtage (entropie, chaos) scheppen ruimte voor adaptatie en het ontstaan van nieuwe levensvormen. De omgeving (wereld) van een CAS is dynamisch, contingent (wisselvallig), onzeker en ambigu. Een CAS bestaat uit actoren, die zichzelf sturen en elkaar sturen. Het feit dat de GGZ een complexe organisatie is, en het oplossen van problemen zoals wachtrijen daarin een venijnig probleem is, betekent dat de problematiek niet met lineaire methoden kan worden geschetst, laat staan opgelost.

Wicked problems gelden als ongrijpbaar; de probleemstelling formuleren is al een probleem. Tamme problemen kunnen rechtlijnig worden opgelost door ze op te delen in deelproblemen en deze afzonderlijk op te lossen. Venijnige problemen kennen geen lijn. Onderstaand een classificatie. Op de x-as staat de mate van ordening, op de y-as de mate van planbaarheid bij het handelen. Oplossing van complexe en ingewikkelde problemen kan gepland worden, bij chaos is de beste tactiek meteen handelen en zien wat er gebeurt.

Statistische Analyse II

Per 100.000 personen hebben er 500 behoefte aan hulp van de GGZ (bron: CBS online). Dat betekent op 17 miljoen ongeveer 85.000 personen. De totale rij wachten bedraagt 80.000. Margriet keek op een velletje, met daarop de woorden ‘formules wachttijden’. Daarbij wordt



Figuur 2. Cynefin Framework (D. Snowden en M.Boon, 2007)

μ	86000	87000	88000	89000	90000	100000
ρ	98,8%	98,7%	98,3%	97,7%	94,4%	85%
Lq	85000	42000	27000	20000	16000	5000

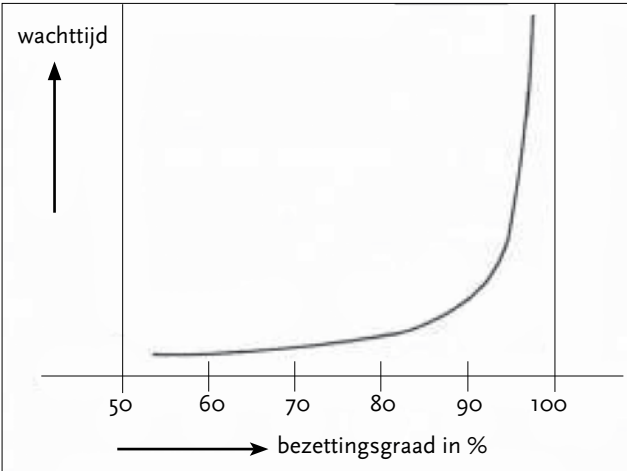
Tabel 1. Resultaten van variaties van de verwerkingscapaciteit (eigen onderzoek)

de GGZ gezien als een systeem met een beperkte verwerkingscapaciteit en een bepaalde aankomstintensiteit. Daarnaast wordt verondersteld dat zowel de aankomst als de verwerkingscapaciteit een negatief exponentiële verdeling volgen.

Margriet vult de formules in. Ze begint met formule 2:

- $85^2 / (\mu(\mu-85)) = 80(000)$. Dat betekent dat $(\mu(\mu-85))$ gelijk is aan 90.
- $\mu^2 - 85\mu - 90 = 0$. Dit levert een μ van 86 of één van min één.

Deze laatste komt natuurlijk te vervallen. We werken met een μ van 86(000) cliënten. Het gemiddeld aantal cliënten in het systeem is $85 + (86-85)$ is dus 85.000 cliënten. De bezettingsgraad, $\rho = (\lambda / \mu)$, $(85/86) \cdot 100\%$, bedraagt bijna 99 procent.



Figuur 3. De relatie tussen bezettingsgraad en wachttijden

staat uit intake, diagnose en therapie. Onder de indirecte tijd vallen de administratietijd, reistijd en overleg.

Slot

‘Zo’, zei Margriet. Een klein traantje waggelde over haar linkerwang terwijl ze door het raam keek. Janssen zat voor het raam en wilde binnen. Het regende, en Janssen had daar een hekel aan. Margriet opende het raam, en Janssen kwam binnen. De analyse was klaar. Ze schreef erop: ‘Fur Elise. Voor zusje.’

LITERATUUR
Lindblom, C. (1959). The Science of ‘Muddling Through’. In *Public Administration Review*, 19(2), 7988. Blackwell Publishing. <http://www.jstor.org/stable/973677>
Gleick, J. (1996). *Chaos*. Vintage publishing.
Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Science*, 4, 155–169. <https://doi.org/10.1007/BF01405730>
Wikipedia. https://nl.wikipedia.org/wiki/Algemene_systeem-theorie

BOB VAN LIMBURG is afgestudeerd als algemeen, kwantitatief en sociologisch econoom en filosoof. Hij is gepromoveerd in de ruimtelijke wetenschappen en verbonden aan de faculteit Medische Wetenschappen van RUG en Wittenborg University of Applied Sciences, Apeldoorn. E-mail: l.j.w.van.limburg@gmail.com



Diego Maradona's goal 'Met de hand van God'; muurschildering van Lean Frizzera, Emy Mariani en Martin Ron in Palermo

Geloven tussen 3D-lijnen

Met het toenemen van de miljoenenbelangen in de topsport en de verkleining van de onderlinge verschillen aan de top ervan neemt ook de impact van onjuiste beslissingen toe. Vooral in de voetballerij is de druk op scheidsrechters om geen fouten te maken groot. Gelukkig hebben we tegenwoordig de VAR, de Video Assistant Referee. Waren uitvergrotingen van ‘millimeterovertredingen’ voorheen voorbehouden aan het tv-kijkende publiek, nu beschikken ook scheidsrechters over die gedetailleerde informatie. Ook voor de beoordeling van buitenspel is de VAR onlangs ‘supernauwkeurig’ gemaakt. De presentatie van de beelden is echter zo complex geworden dat zelfs de VAR-scheidsrechters het spel minutenlang moeten stilleggen om tot een juiste beoordeling te komen met als gevolg dat de doorsnee-toeschouwer er helemaal geen biet meer van snapt. Moesten wij als toeschouwers eerder de scheidsrechter geloven op zijn fluit, nu moeten we abracadabra computerbeelden geloven. Wat is er aan de hand? Eerst een terugblik met een persoonlijke noot.

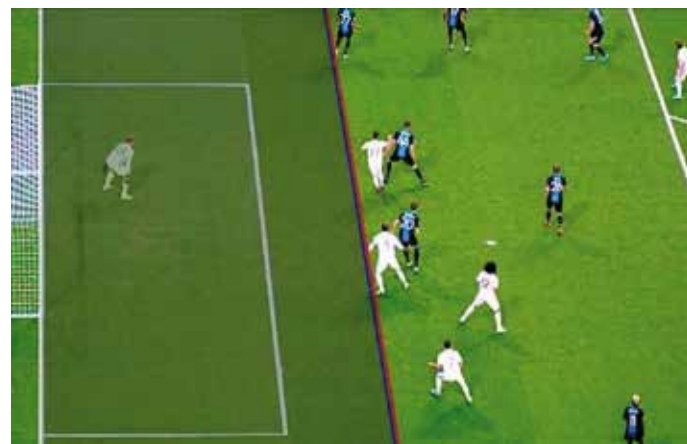
Kampen, april 1960. Terwijl het kooppubliek de binnenstad verlaat, zingt het gerucht reeds rond kerken en door stegen: Go Ahead K. heeft de uitwedstrijd gewonnen. Solexen en zwarte fietsen, veel, verzamelen zich in de

late zaterdagmiddag bij de winkelruit van het *Kamper Nieuwsblad* aan de Oudestraat, de Kalverstraat van Kampen. Zo tegen zessen wordt het cellotape van de overlidenspagina van de krant losgepeuterd van het raam en wordt op de opengevallen plek de uitslagenlijst tegen het lentelicht gehouden: SVZW – Go Ahead (wij zeggen kowet): 1-2. Het geloof in en de hoop op de overwinning slaan om in ‘zeker weten’. Zondagochtend, tijdens de preek van de eerste ochtenddienst in de Nieuwe Kerk aan de Broederweg, reken ik uit op welke plek in de ranglijst Kowet over drie weken zou kunnen staan. Nog drie wedstrijden. Eerst zien, dan geloven. ‘G’leuv’m doe’j in de kärke’, op z’n Kampers; niet op de Oudestraat en niet op het voetbalveld.

Mexico-City, 22 juni 1986. De 51ste minuut. Zondagavond. De persconferentie na afloop is hilarisch. Argentinië heeft net Engeland met 2–1 naar huis gestuurd in de kwartfinale van het WK-voetbal. De Argentijnse spits Diego Maradona tikt, met zijn net boven zijn gekrulde kruin uitstekende rechterhand, de bal, tussen de wanhopig omhoog gestoken armen van de Engelse doelman Peter Shilton door, in het doel. Miljoenen tv-kijkers zien de handsbal in slow motion terug. De scheidsrechter wijst naar de middenstip, hij heeft de handsbal niet gezien: 1–0



HawkEye-beeld



VAR2.0-beeld

voor de Argentijnen. Na afloop verklaart Maradona: het was 'half met het hoofd en half met de hand van God'. Een iconisch sportmoment is een feit.

Een herhaling van zoiets als Maradona's *finest second* zal met de komst van de VAR wel nooit meer plaatsvinden. De VAR legt de geringste overtredingen bloot. Overtredingen, die eerder nooit waargenomen werden, worden heden ten dage langs de zijlijn online getoond. Ook de *handgoal* van Thierry Henry, die Frankrijk op 18 november 2009 naar het WK-miljoenenbal in Zuid-Afrika bracht ten koste van de Ieren, zou door de VAR afgekeurd zijn. Iedere tv-kijker kon de handsbal zien. Alleen de scheidsrechter zag het niet en wees naar de middenstip.

De VAR kwam dus als geroepen. Met dank aan ziensers als Marco van Basten die er al of niet binnen de FIFA in een vroeg stadium op wees dat voetbal waar het vernieuwingen betreft nogal eens achterloopt op andere sporten. Zoals tennis. Tennis maakt sinds 2007 gebruik van HawkEye, een videosysteem dat met behulp van 3D-camera's de baan van de bal in beeld brengt. Op basis van een *virtual reality*-simulatie (zo heet dat) wordt vervolgens nauwkeurig bepaald of de bal de lijn wel of niet is gepasseerd. Niet alleen de scheidsrechter ziet de computerbeelden, ook de tv-kijker kan meekijken en de situatie beoordelen. Na tennis volgden volleybal, cricket en snooker in rap tempo met de introductie van HawkEye. Het duurde nog tot 2012 totdat de Wereldvoetbalbond, FIFA, de cameratechniek inzette. Echter voornamelijk om vast te stellen of een bal over de doellijn was of niet. Zaken als *hands* en buitenspel moesten nog acht jaar wachten op enige computerassistentie.

Inmiddels is de VAR nauwelijks meer weg te denken en ondanks de nodige kinderziekten een veel geraadpleegd instrument. Vaak leidt dat tot goede correcties, zoals bij

penalty's of kaarten voor elleboogstoten. Maar dat wil zeker niet zeggen dat de beeldscheidsrechter de oplossing is geworden die Van Basten voor ogen had. Van het begin af aan lagen vooral de buitenspelbeoordelingen onder vuur. Een voetbalschoenneus over een, niet nauwkeurig digitaal te tekenen, buitenspellijn betekende: bal niet naar de middenstip, goal afgekeurd. Te vaak ging het over millimeters met soms miljoenenconsequenties, denk maar aan de afgekeurde goal van Ajacied Quincy Promes thuis tegen Chelsea vorig jaar. Ook uitvergrotingen van zulke situaties leverden zelden eenduidige inzichten op. En omdat die millimetertjes binnen de foutmarges van het meetsysteem van de VAR lagen, waren ze feitelijk ook irrelevant (zie *STATOR* van maart 2020). De eerste ideeën om foutmarges te vertalen in een tolerantie-marge, ten faveure van de schoenneus-buitenspel-staande aanvaller, werden door de binair-denkende pers weggehoond.

En toen kwam in maart van dit jaar COVID-19. De bijgeleverde crisis heeft ervoor gezorgd dat alle competities stil kwamen te liggen. De KNVB greep de vrijgekomen tijd aan om na te denken over een update van de VAR. Die update, laten we zeggen VAR2.0, zou het einde moeten betekenen van de lange spelonderbrekingen en de toeschouwer begrijpelijke beelden geven van de discussiemomenten. Niet alleen de tv-toeschouwer ook de supporter op de tribune zou op grote matrixborden kunnen meekijken. Nadat de eerste golf van de corona-crisis bijna voorbij was, kondigde de KNVB aan dat de nieuwe VAR een tolerantiezone van 10 cm gaat gebruiken die recht doet aan de *margin-of-error* die ontstaat bij de bepaling van de buitenspellijn. Problemen met 'schoenneus-buitenspel' en met de lange spelonderbrekingen zouden daarmee verleden tijd zijn. En zo is met de nieuwe competitie in september ook VAR2.0 van start gegaan. Helaas

moeten de tribunes voorlopig leeg blijven en vervliegt de informatie op de matrixborden in het niets. Voorlopig zitten we thuis.

Het is 25 september, FC Twente speelt tegen FC Groningen. De twaalfde minuut. Vaclav Cerny scoort voor Twente. De grensrechter vlagt voor buitenspel. Uit de herhaald getoonde televisiebeelden zou je kunnen opmaken dat het geen buitenspel is. De wedstrijd wordt stilgelegd. Het VAR2.0-collectief wordt in stelling gebracht. Dat heeft vervolgens een dikke vijf minuten nodig om erachter te komen dat de VAR-beelden geen duidelijkheid bieden. De officiële afspraak is dat bij grote onzekerheid alsnog de genomen beslissing van de veldscheidsrechter leidend is en zijn oordeel staan blijft. Doelpunt afgekeurd, Twente boos. Wat was er in Enschede werkelijk aan de hand? Waarom kostte het beoordelen met VAR2.0 zoveel tijd en waarom ging het fout? De verwachting was toch dat VAR2.0 aan de volgende twee eisen zou voldoen: a. begrijpelijke beelden voor de scheidsrechters (zodat in korte tijd de beslissing kan worden genomen); en b. begrijpelijke VAR-beelden voor de toeschouwers!

Een poging tot uitleg. In het VAR2.0-beeld verschijnen twee lijnen: een rode behorend bij de vermeend buitenspel staande aanvaller en een blauwe behorende bij de voorlaatste verdediger. Die lijnen zijn volgens KNVB-scheidsrechterdocent Robin van 't Hof (KNVB-persconferentie van 19 augustus jl.) '2-dimensionaal' en kunnen zelfs '3-dimensionaal' zijn. Een '2-dimensionale digitale lijn' komt in beeld als een lichaamsdeel (uitgezonderd de armen) van de betreffende speler zich bevindt op of in het (kunst)gras; de '3-dimensionale digitale lijn' wordt gebruikt als die speler, afgezien van zijn beide armen, zweeft boven het veld. Hoe de 2- en 3-dimensionale lijnen precies worden getrokken gaat mij vooralsnog boven de pet. Wel vind ik de namen mooi: De 2-dimensionale lijn-methode heet de *gridlines*-methode en de 3-dimensionale lijn-methode heeft de prachtige naam *crosshair*-methode.

Beide 'lijnen' zijn vijf centimeter breed als ze echt op het veld zouden worden geprojecteerd. Die vijf centimeter komen, naar verluidt, overeen met een breedte van tien pixels in het camerabeeld. Snel uitgerekend betekent dat een halve centimeter per pixel. De foutmarges bij deze berekening laat ik achterwege. Maar die zijn er zeker

en zelfs tweemaal, voor de blauwe en de rode namelijk.

En nu komt de clou. Zodra de VAR-scheidsrechters, na volledig inzoomen, ruimte zien tussen de rode en de blauwe lijnen (tussen de twee stroken derhalve), dan is het onverbiddelijke advies aan de veldarbiters: buitenspel. Overlappen beide stroken, dan geen buitenspel. De poppen zijn dus pas aan het dansen als er onduidelijkheid bestaat over het al dan niet overlappen en elkaar dus net wel of net niet raken van beide stroken. Wat dat 'raken' betreft heb ik zo'n ruim ingezoomd plaatje van het 'buitenspeldoelpunt' van Cerny bekeken. Ik zag twee zaagtandlijnen, die lepelte-lepelte lagen. Maar of ze elkaar raken? Ook de VAR-scheidsrechters tastten kennelijk in het duister over het wel of niet aanwezig zijn van 'licht' tussen de blauwe en rode lijnen. Zoals gezegd is in zo'n dubieuze situatie de regel dat de genomen beslissing van de veldscheidsrechter niet wordt teruggedraaid. Precies wat er in de wedstrijd Twente – Groningen gebeurde. En dat is raar.

Immers, eerst is er twijfel over de beslissing van de veldscheidsrechter. Dan wordt VAR2.0 geraadpleegd, die toch juist ontworpen is om te helpen beslissen in dit soort onzekere situaties. Na lang beraad is vervolgens de conclusie dat de 2D- en zelfs de 3D-beelden geen duidelijkheid bieden. En tot slot blijft de beslissing van de scheids in het veld gewoon staan en wordt de wedstrijd hervat. De problemen met schoenneus-buitenspel zijn dus niet verdwenen. Ofwel, zelfs de meest geavanceerde technologie biedt als het er echt op aan komt geen soelaas. Conclusie: ondanks de dikke anderhalve ton die VAR2.0 heeft gekost zijn de beide doelstellingen niet gehaald.

En dat had anders gekund. Mooi van de nieuwe VAR-versie is dat rekening is gehouden met een foutmarge bij de bepaling van buitenspel. Maar waarom die is vertaald in twee tolerantiezones, een rode en blauwe, is mij een raadsel. De VAR had kunnen volstaan met alleen de blauwe strook die bij de virtuele buitenspellijn behoort. In plaats van vijf centimeter zou die blauwe strook dan tien centimeter breed kunnen zijn. Net als bij de tennisbal 'op of over de lijn' van HawkEye, kijk je dan naar de positie van de aanvaller ten opzichte van één 'lijn', de blauwe zone: erbinnen (of ervoor natuurlijk), dan geen

buitenspel geven. Een essentieel kenmerk van zo'n tolerantiezone is dat je bij onzekerheid op de achterste grens (in de richting dus van de aanval) ervan precies kunt afspreeken wat je moet doen. Als die zone relatief breed is gekozen – en dat is met 10 cm zeker zo – dan is het bij onzekerheid op de achterste grens zeker buitenspel. Je bent dan namelijk ver genoeg verwijderd van de theoretische buitenspellijn, de lijn die getrokken zou moeten worden exact op het moment dat de bal wordt gespeeld naar de van buitenspel staan verdachte aanvaller. Bij de minste onzekerheid van de VAR over het wel of niet over die grens staan kun je dan onmiddellijk buitenspel geven, zonder minutenlang het spel te onderbreken. Omdat 10 cm breed genoeg is heb je ook geen 3D-lijnen nodig van de *crosshair* methode. De geringe fout die je maakt zonder *crosshair* zit dan in de tolerantiebreedte van die 10 cm blauwe zone. Belangrijk bijkomend voordeel is dat de VAR-beelden ook voor het publiek inzichtelijk zijn: een heldere blauwe strook met daarin geprojecteerd het relevante deel van de speler, precies zoals bij HawkEye de tennisbal op de lijn is weergegeven. Beslissingstechnieken kunnen inderdaad ook te nauwkeurig zijn en daardoor hun doel voorbijschieten. Voor gecompliceerde problemen werken simpele oplossingen vaak het best. En wat de VAR betreft, niet alleen de scheidsrechter ook de toeschouwer is 'gebruiker' van die oplossing.

P.S. Een opmerkelijke lezer van een eerdere versie van deze column vroeg me of ik nog wel fiducia heb in de toekomst van de topsport nu de onderlinge prestatieverschillen zo vaak onmeetbaar zijn en de toeschouwer niets snapt van de uitkomsten van de algoritmen en technologieën waarmee de winnaars worden aangewezen. Goed punt. Ik weet het echt niet. Ik denk trouwens wel dat de FIFA de huidige versie van de VAR gaat aanpassen. Het komt vast goed met de VAR. Maar ... zeker weten? Eerst zien, dan geloven.

LITERATUUR

- Abbenhuijs, M. (2020). Buitenspellijntechnologie: 'De aanvaller moet het voordeel van de twijfel krijgen'. *Algemeen Dagblad*, 29 september 2020.
- Sierksma, G. (2019) De VAR kan bij buitenspel sneller en eerlijker. *Sport Knowhow XL*, december 2019.

GERARD SIERKSMA is emeritus hoogleraar Kwantitatieve Logistiek en Sportstatistiek aan de Rijksuniversiteit Groningen. E-mail: g.sierksma@rug.nl

LNMB/NGB conferentie

22 januari 2021

De jaarlijkse LNMB/NGB-conferentie wordt ditmaal niet in Lunteren maar geheel online gehouden.

Keynote speakers

Adam Wierman (California Institute of Technology)
Pinar Keskinocak (Georgia Institute of Technology).

Zie voor het hele programma de websites van LNMB en NGB: www.lnmb.nl/pages/conferences/2021/

VERSLAG NGB PUBQUIZ

Er werd fanatiek gestreden en een hoop gelachen

Het Nederlands Genootschap voor Besliskunde organiseerde op 12 november een online pubquiz. Ten tijde van een pandemie waarin fysieke activiteiten moeten worden afgezegd, leek ons dit een laagdrempelige manier om toch nog in contact te blijven met collega's en oud-studiegenoten. Een deel van het NGB-bestuur wierp zich op als quizmasters, de overige bestuursleden verenigden zich in een deelnemend team. Onder het motto 'hoe meer zielen, hoe meer vreugd' was de quiz was ook voor niet-leden toegankelijk. Er volgden aanmeldingen van de VU, CWI, Universiteit Twente, Erasmus Universiteit, ORTEC, TNO en vriendenteams. Dit leverde 14 teams met in totaal 61 deelnemers op.

De muziekrunde werd door velen een hoogtepunt genoemd, met vragen die niet zozeer gingen om het benoemen van de artiest of titel, maar om een cijfer of rekensom. Bijvoorbeeld, bij het draaien van 'I'm Gonna Be' van The Proclaimers: hoeveel kilometer zou deze man bereid zijn te lopen? Dat is 1000 mijl (let op, niet 500!) vertalen naar kilometers.

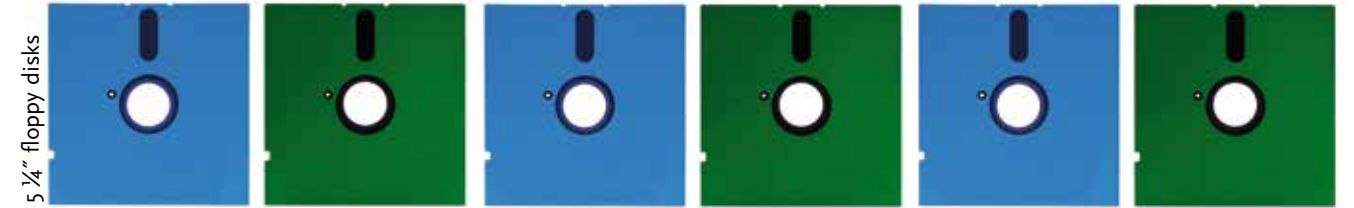
Uiteindelijk ging de hoofdprijs naar team Mathketels bestaande uit Kamiel Cornelissen, Fabian Akkerman, Thomas Visser en Pim van 't Hof. Wij feliciteren deze heren, allen (ex)ORTEC'ers, met hun prestatie. Ook danken wij alle deelnemers voor hun enthousiaste bijdrage en de gezelligheid.

Het NGB bestuur (secretariaat@ngb-online.nl)

Ben je wel NGB-lid maar heb je geen uitnodiging voor de quiz ontvangen? Controleer dan met welk e-mailadres je geregistreerd staat.

GERRIT STEMERDINK

column



Goede contacten

Er wordt nu, als gevolg van de corona-crisis, alom nagedacht over hoe het verder moet met grote internationale bijeenkomsten. Hebben wetenschappers nog wel zin in en tijd over voor dagenlange conferenties met verre en dure reizen? Ongetwijfeld zal de animo hiervoor, zeker de eerste tijd, flink zijn gedaald. En persoonlijk denk ik niet dat de 'oude tijden' compleet zullen terugkeren. Men heeft, veelal noodgedwongen, ontdekt hoeveel er mogelijk is via virtuele bijeenkomsten. En er zijn zelfs voordelen ontdekt: de bijeenkomsten zijn voor véél meer deelnemers haalbaar door het wegvallen van de vaak hoge reis- en verblijfkosten.

Op zich is zo'n verandering niet nieuw, we hebben het vaker gezien. In de tijd van Huygens en Newton onderhielden wetenschappers contact via uitgebreide briefwisselingen. Pas zo halverwege de 19^e eeuw kwamen de eerste wetenschappelijke conferenties tot stand. Toen werden ook organisaties als de Royal Statistical Society, de American Statistical Association en het International Statistical Institute opgericht. De betere reismogelijkheden, veelal mogelijk door het aanleggen van spoorwegen, zullen hieraan debet zijn geweest. Vanaf de jaren '60 van de 20^e eeuw begonnen vliegvluchten betaalbaar te worden en groeide de omvang van conferenties. De tweejaarlijkse ISI Sessions hebben regelmatig tegen de drieduizend deelnemers gehad.

Een belangrijk aspect van bijeenkomsten waar je liefelijk aanwezig bent is het gemakkelijk tot stand komen van persoonlijke contacten. Natuurlijk, je kunt altijd iemand een bericht sturen met een vraag. Maar als je tijdens een koffiebreek of receptie terloops een praatje hebt gemaakt heb je dikwijls gemerkt dat zo'n grote naam toch maar een gewoon mens is. En dan wordt de drempel om te benaderen lager. Het is mijn vaste overtuiging dat persoonlijke contacten het allerbelangrijkste product zijn van conferenties, hoe indrukwekkend het lezingenprogramma ook moge zijn.

Zelf heb ik erg veel profijt gehad van zo'n contact. Halverwege de jaren 80 van de vorige eeuw werkte ik bij het Rekencentrum van de Landbouwniversiteit Wageningen. Ik was daar verantwoordelijk voor alle statistische

software. Op een dag kreeg ik de vraag van een hoogle- raar of ik kans zag een klein jaar lang een promovendus uit Sri Lanka te assisteren bij de verwerking van zijn data. Het leek me een mooie gelegenheid een nieuwe module van een pakket in de praktijk te leren kennen en ik stemde toe. Enkele weken later kwam de promovendus langs met twee 5.25" floppy's met een grote hoeveelheid data die hij tijdens zijn veldwerk bij Sri Lankese rijstboeren had verzameld. Helaas waren die aangemaakt op een weinig gangbare computer. Die floppy's kon ik niet lezen, in die tijd had ieder merk zijn eigen afwijkende format, de latere standaardisatie met MS-DOS moest nog tot stand komen. Ik had via de VVS veel contacten met mijn collega's bij andere rekencentra, maar niemand kon mij helpen. Toen viel mij in dat ik dat merk computer wel eens in Engelse bladen geadverteerd had gezien. Ik had via de Compstat-conferenties een goed contact gekregen met een collega in Hull. Ik vroeg of hij kans zou zien deze floppy's te lezen en te converteren naar een meer gangbaar format. Enkele dagen later had hij zo'n computer gevonden in een kelder vol afgedankte apparatuur en die met behulp van een speciaal gesoldeerde kabel via een MS-DOS computer verbonden aan een VAX-systeem. Ik heb de kostbare floppy's aangetekend verzonden, ze bevatten tenslotte drie jaar veldwerk, en er was geen enkele back-up behalve een grote stapel ingevulde formulieren aan de andere kant van de wereld! Binnen een week stonden de data op de Wageningse computer en tien maanden later vond de promotie plaats. Als ik niet toevallig, dankzij liefelijk bijgewoonde conferenties, dit contact had gehad zou de promotie een enorme vertraging hebben opgelopen. Zó belangrijk kunnen goede contacten zijn.

De oude Grieken wisten het al: zij hielden symposia. Die hebben wij ook, maar weinigen realiseren zich dat de betekenis van dit woord iets is als 'samen drinken'. En ja, dan komen goede contacten gemakkelijk tot stand. Helaas zal dat voor de hedendaagse wetenschappers nog wel even duren.

GERRIT STEMERDINK (eindredacteur van STATOR)
E-mail: gjstermerk@hotmail.com



Modeling spread

VVSOR ANNUAL MEETING 2021

thema

Going viral – Modeling spread

donderdag 18 maart 2021 | Online

De Annual Meeting (AM) van 2021 zal in het teken staan van verspreiding. Denk daarbij niet alleen aan de verspreiding van virussen, maar ook aan informatie, meningen, goederen of mensen.

De sprekers zijn:

Prof. dr. D.J. (Dimitris) Bertsimas

Prof. dr. J. (Jacco) Wallinga

Prof. dr. N.V. (Nelly) Litvak

Prof. dr. G. (Gavin) Shaddick

De sprekers zullen vanuit verschillende perspectieven uit de Statistiek en OR hun licht laten schijnen op het modelleren van verspreiding.

Omdat de verspreiding van COVID-19 in maart 2021 nog voort zal duren, zal de AM dit keer als online streaming event plaatsvinden. De middag zal starten met de ALV.

De AM 2021 zal Engelstalig zijn. Het definitieve programma en informatie over hoe u de annual meeting kunt bijwonen kunt u enkele weken van tevoren vinden in de volgende editie van *STA&OR* en op de website van de VVSOR.

E-mail: annualmeeting@vvsor.nl



UITMUNTENDE MASTERS OF PH.D. THESIS BEGELEID?

*OPROEP OM KANDIDATEN TE NOMINEREN
VOOR DE JAN HEMELRIJK
EN WILLEM R. VAN ZWET AWARDS 2020*

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs/hoger beroepsonderwijs looft de VVSOR al vanaf 1989 een scriptieprijs uit. In 2014 kreeg deze de naam *Jan Hemelrijk Award*. Sinds 2012 is er ook een prijs voor dissertaties: de *Willem R. van Zwet Award*.

De VVSOR roept op tot nominaties voor deze prijzen. Beide prijzen bestaan uit een oorkonde en een geldbedrag van 1.000 euro. De prijzen zullen worden uitgereikt tijdens de Annual Meeting van de VVSOR, op 18 maart 2021. Genomineerd kunnen worden personen die van september 2018 tot en met augustus 2020 zijn afgestudeerd respectievelijk gepromoveerd en die nog niet eerder zijn genomineerd.

Hierbij worden supervisors (begeleiders) opgeroepen om een uitmuntende afstudeerscriptie (Master) of dissertatie (Ph.D.) te nomineren voor de Jan Hemelrijk dan wel Willem R. van Zwet Award 2020.

De indiening van een nominatie dient vergezeld te gaan van een aanbevelingsbrief van de supervisor van de genomineerde. De precieze procedure voor beide prijzen, alsmede de reglementen en het nominatieformulier zijn te downloaden via de website van de VVSOR, www.vvsor.nl. De nominatie dient uiterlijk 24 januari 2021 binnen te zijn.

Namens de VVSOR,

Dr. Ad Ridder, juryvoorzitter Jan Hemelrijk Award

Prof. dr. Jelle Goeman, juryvoorzitter Willem R. van Zwet Award

Dr. Sander Scholtus, Secretaris der beide jury's