STAtOR

periodiek van de VVSOR jaargang 22, nummer 2, juni 2021

STAtOR

Jaargang 22, nummer 2, juni 2021

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 Stemerdink (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), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, j.a.s.gromicho@uva.nl

Bestuur van de VVSOR
Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Algemeen bestuurslid: Thomas Wise MSc, db@vvsor.nl; Webmaster: Eugenio Traini 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); Marije Sluiskes 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

ISSN 1567-3383

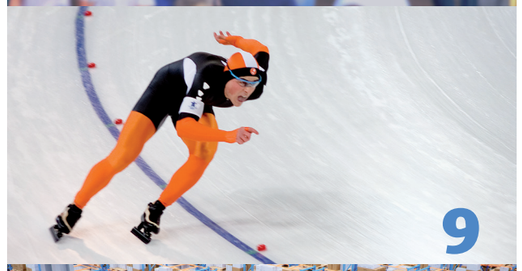
What goes up will go down

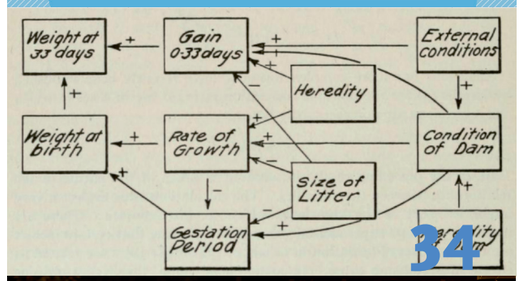
Een bekende kreet, te vinden in zowel beursrubrieken als in mental-life-style-artikelen. De afgelopen tijd ging het met de COVID-cijfers lange tijd *up*, nu zien we ze overal een *down* beweging maken. Een triviale vaststelling uit de afgelopen anderhalf jaar kan zijn dat het exponentiële toe- en afnemen weinig wordt begrepen. Velen verbazen zich over de plotseling sterk dalende cijfers, zich niet realiserend dat dit hetzelfde exponentiële mechanisme is als de snel oplopende cijfers van eerder dit jaar. Kortom, wij kunnen niet genoeg nadruk leggen op een goede wiskundige ondergrond. Met ons blad hopen we daar een steentje aan bij te dragen.

Dit nummer is weer goed gevuld en zeer gevarieerd van inhoud. Van het voorspellen van dreigingen van een drone aanval door Laura Middeldorp tot een gedegen overzicht van meer dan 20 jaar technieken voor voorraadbeheersing door Ton de Kok. In de hopelijk warme sportzomer die ons te wachten staat brengt Gerard Sierksma verkoeling door uitgebreid in te gaan op de rol van statistiek en OR bij het optimaal samenstellen van schaatsploegen. Frank Bais en Joost van der Neut beschrijven de vergelijking van gedragsprofielen en Kim van den Houten cs demonstreren hoe de capaciteit van een endoscopie afdeling maximaal kan worden benut. De LISREL techniek van Jöreskog en de opvolgers daarvan zijn bij de meeste van onze lezers wel bekend, maar Richard Starmans gaat in op de oorsprong van deze techniek, dit jaar 100 jaar geleden! Anton Meijburg maakt duidelijk hoe belangrijk het is sensitiviteit en specificiteit van testen te benoemen. Hij doet dit aan de hand van COVID-cijfers, maar dat had ook geïllustreerd kunnen worden met totaal andere testen.

Helaas ook triest nieuws in dit nummer: een coryfee is ons ontvallen, het overlijden van Arie Hordijk wordt herdacht. *STAtOR* is niet compleet zonder columns, ditmaal van John Poppelaars en een lichtelijk aangeschoten Gerrit Stemerdink. En tot slot een felicitatie voor het 75-jarige CWI, nieuws over de toekenning van de Franz Edelman Award aan het World Food Program en over het aanstaande virtuele World Statistics Congress.

Wij wensen u veel leesplezier.





INHOUD

HET VOORSPELLEN VAN DE DREIGING VAN EEN DRONE MET BEHULP VAN EEN BAYESIAANS NETWERK

Stelt u zich voor, de COVID-pandemie is voorbij en u bezoekt een openluchtfestival. Plots hoort u een zoemend geluid boven u. U kijkt omhoog en ziet een drone. U besluit de politie in te schakelen. Die komt en moet ter plaatse beslissen of en hoe er actie ondernomen moet worden. Om de juiste actie te bepalen, is het van belang dat de intentie van de drone wordt achterhaald: gaat de drone een aanslag plegen of maakt de drone slechts foto's van het evenement? Voor mijn werkstage bij TNO heb ik een Bayesiaans netwerk opgesteld, dat gegeven de observaties die gemaakt worden van de drone en zijn omgeving, de meest waarschijnlijke dreiging van de drone kan voorspellen.

Laura Middeldorp

Een drone is een onbemand luchtvaartuig. Hij kan zich voortbewegen met behulp van vleugels, propellers of straalmotoren. Drones maken steeds vaker deel uit van het straatbeeld, mede omdat de modellen die voor de consument zijn bedoeld almaar betaalbaarder worden. Een drone geraakt mogelijk in handen van een individu of groep met minder goede intenties die kunnen variëren van privacyschendingen tot gerichte aanslagen. Dit vormt een uitdaging op het gebied van veiligheid. Het is daarom cruciaal dat veiligheids- en beveiligingsdiensten zoals de politie zich wapenen tegen dronedreigingen. De meest waarschijnlijke dreiging kan achterhaald worden met behulp van een Bayesiaans netwerk.

Wat is een Bayesiaans netwerk?

Een Bayesiaans netwerk is een gerichte acyclische graaf, waarin de nodes stochastische variabelen beschrijven en de pijlen de causale relaties tussen de stochasten weergeven. Een simpel voorbeeld van een Bayesiaans netwerk ziet u in figuur 1. Die laat zien dat node A en B een causale relatie hebben. Dit betekent dat de kansverdeling van node B afhangt van de uitkomst van node A.

Een Bayesiaans netwerk werkt aan de hand van de regel van Bayes: gegeven de a priori kansen van node A en de voorwaardelijke kansen van node B gegeven node A kunnen de a posteriori kansen van node A gegeven de uitkomst van de node B worden achterhaald. Stel dat node A en B ieder twee mogelijke uitkomsten kunnen aannemen: Ja en Nee. Als we erachter komen dat node B gelijk is aan Ja, dan kan de kans dat node A gelijk is aan Ja gegeven dat node B gelijk is aan Ja als volgt bepaald worden:

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

Een Bayesiaans netwerk is in staat om de kansverdeling van nodes te updaten gegeven geobserveerde informatie.

Hoe is een Bayesiaans netwerk toe te passen in deze context?

In het Bayesiaanse netwerk voor dronedreigingen worden de mogelijke dreigingen gemodelleerd door hypothesen. De waarnemingen die gemaakt kunnen worden van de drone en zijn omgeving worden beschreven door indicatoren. Een hypothese en een indicator hebben een causale relatie als het optreden van de hypothese het waarschijnlijk maakt dat de indicator optreedt.

Een mogelijke dreiging is bijvoorbeeld:

De drone gaat een bomaanslag plegen.

Een mogelijke indicator is:

Er hangt een pakketje onder de drone.

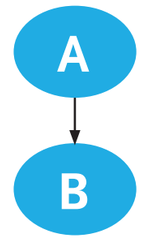
Deze hypothese en indicator hebben een causale relatie:

gegeven dat de drone een bomaanslag gaat plegen, is het waarschijnlijk dat er een pakketje onder de drone hangt, waarbij het pakketje mogelijk de bom bevat.

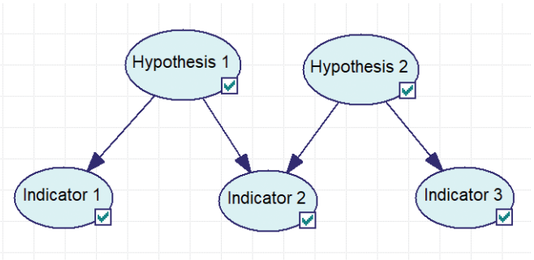
Figuur 2 laat een voorbeeld zien van het model bestaande uit 2 hypothesen en 3 indicatoren. Gegeven de a priori kansen op de hypothesen, de voorwaardelijke kansen op de indicatoren gegeven de hypothesen en de geobserveerde indicatoren kunnen de a posteriori kansen op de hypothesen gegeven de indicatoren worden berekend. De meest waarschijnlijke dreiging is de hypothese met de grootste a posteriori kans gegeven de geobserveerde indicatoren.

Het opstellen van het Bayesiaanse netwerk

Om het Bayesiaanse netwerk op te kunnen stellen, moet er eerst een scenario worden gedefinieerd. Het scenario is de omgeving waarin de drone zich bevindt. De hypothesen en indicatoren in het Bayesiaanse netwerk hangen namelijk af van het scenario: als een drone in een oorlogsgebied vliegt in Afghanistan zijn de mogelijke dreigingen anders dan wanneer een drone in een drukke stad in Nederland wordt geobserveerd. Voor het scenario is gekozen voor een druk evenement in de stad waarbij er geen strenge toegangscontrole is.



Figuur 1. Een simpel voorbeeld van een Bayesiaans netwerk bestaande uit 2 nodes A en B



Figuur 2. Een voorbeeld van een Bayesiaans netwerk bestaande uit 2 hypothesen en 3 indicatoren

De hypothesen en indicatoren behorende bij dit scenario zijn in samenwerking met experts binnen de politie en defensie bepaald. In het uiteindelijke Bayesiaanse netwerk zijn 8 hypothesen en 23 indicatoren gedefinieerd, waarbij iedere indicator twee mogelijke uitkomsten heeft: Ja en Nee.

Om de meest waarschijnlijke dreiging gegeven de observaties van de drone te achterhalen, zijn de kansen op de hypothesen en de kansen op de indicatoren gegeven de hypothesen benodigd. Gezien er onvoldoende data beschikbaar waren om deze kansen in te schatten, is er gebruik gemaakt van meningen van experts om de kansen te bepalen. Hiervoor is een expert-sessie georganiseerd waaraan experts binnen zowel politie als defensie hebben deelgenomen.

Robuustheid van het resultaat van het Bayesiaanse netwerk

Een nadeel aan het gebruik van meningen van experts voor het inschatten van de kansen is dat het onzekerheid met zich meebrengt. Als een andere groep experts zou zijn ingeschakeld om de kansen in te schatten, zouden de kansen in het Bayesiaanse netwerk mogelijk anders zijn waardoor het resultaat zou verschillen. Dit brengt ons op het vraagstuk van robuustheid: hoe zeker is het dat de meest waarschijnlijke dreiging volgens het Bayesiaanse netwerk ook daadwerkelijk de meest waarschijnlijke is?

Een van de mogelijke methoden die gebruikt kan worden om de robuustheid van het resultaat te bepalen is de methode van rangverschuivingen (Kipersztok & Wang, 2001). Deze methode is gebaseerd op een Monte Carlo simulatie. In iedere simulatie wordt er noise toegevoegd aan de kansen in het Bayesiaanse netwerk waarna de a posteriori kansen van de hypothesen berekend worden. Vervolgens wordt de rangorde van de hypothesen in de simulatie vergeleken met de originele rangorde. Op deze wijze wordt het aantal rangverschuivingen dat optreedt met de hypothesen inzichtelijk en wordt er meer duidelijkheid verkregen over de robuustheid van het resultaat.

Net als in Kiperzstok & Wang (2001) wordt er aan de voorwaardelijke kansen noise toegevoegd uit de log-odds normale verdeling. Hiervoor wordt de voorwaardelijke kans eerst getransformeerd naar de log-odds waarna er een noise component uit de normale verdeling wordt toegevoegd:

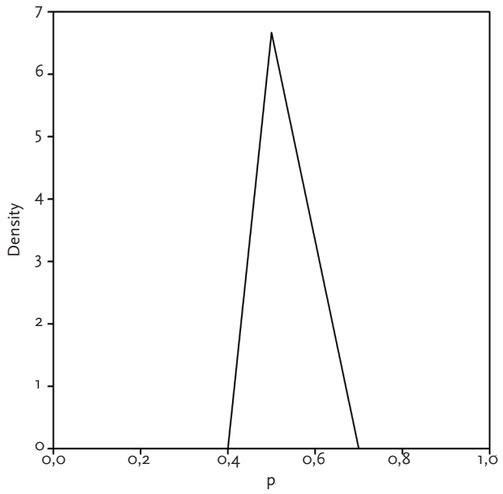
$$Y = \log_{10}\left(\frac{p}{1-p}\right) + \epsilon, \quad \epsilon \sim N(0, \sigma)$$

Vervolgens wordt de verstoorde voorwaardelijke kans verkregen door de inverse te nemen van bovenstaande:

$$\tilde{p} = \frac{10^{Y}}{1 + 10^{Y}}$$

Voor iedere voorwaardelijke kans moet er een standaarddeviatie worden gekozen. Hiervoor is gebruik gemaakt van de resultaten van de expert-sessie: hoe meer de experts het eens waren over de waarde van de voorwaardelijke kans, hoe kleiner de standaarddeviatie van de normale verdeling.

Naast de voorwaardelijke kansen moet er ook noise worden toegevoegd aan de a priori kansen van het Bayesiaanse netwerk. Aanvankelijk wilde ik ook de log-odds normale verdeling gebruiken voor de a priori kansen, maar dit gaf enkele problemen. Uiteindelijk bleek de driehoeksverdeling de meest geschikte verdeling om noise toe te voegen aan de a priori kansen. Figuur 3 toont een voorbeeld van een driehoeksverdeling.



Figuur 3. Plot van de dichtheid van een driehoeksverdeling.

Voor iedere a priori kans is het interval waarop de driehoeksverdeling positieve massa heeft bepaald aan de hand van de expert-sessie: voor iedere hypothese zijn de kleinste en grootste a priori kans ingeschat door de experts gebruikt als grenzen voor het interval. De mode van de driehoeksverdeling is gelijk aan de a priori kans van de hypothese.

De robuustheid van het resultaat van het Bayesiaanse netwerk wordt bepaald door 2 dingen: de kansen in het Bayesiaanse netwerk en de mate waarin de experts het met

CATEGORIE 1		
I1	☒ Ja	☐ Nee
I2	☐ Ja	☐ Nee
I3	☐ Ja	☒ Nee
I4	☐ Ja	☐ Nee
CATEGORIE 2		
I5	☐ Ja	☐ Nee
I6	☐ Ja	☐ Nee
CATEGORIE 3		
I7	☒ Ja	☐ Nee
I8	☐ Ja	☐ Nee
I9	☐ Ja	☒ Nee
I10	☐ Ja	☐ Nee
I11	☒ Ja	☐ Nee
I12	☐ Ja	☐ Nee
I13	☐ Ja	☐ Nee
CATEGORIE 4		
I14	☐ Ja	☐ Nee
I15	☐ Ja	☐ Nee
I16	☐ Ja	☐ Nee

Instructie voor het gebruik van de tool:

Voor iedere geobserveerde indicator kunt u de uitkomst aanvinken in de checkboxes. Om een checkbox aan te vinken, is het voldoende om met uw muis te klikken op de cel waarin de checkbox staat. Indien de uitkomst van een indicator onbekend is, kunt u de checkboxes behorende bij deze indicator leeg laten.

Als u alle geobserveerde indicatoren heeft aangevinkt, kunt u op een van de onderstaande knoppen drukken. De knop 'Bereken het resultaat' geeft u alleen de kansen op de hypothesen. Wanneer u op de knop 'Bereken de robuustheid van het resultaat' klikt, krijgt u, naast de kansen op de hypothesen, ook de robuustheid van de uitkomst te zien.

Wanneer u op een van de twee knoppen drukt, ziet u een scherm verschijnen wat betekent dat het programma aan het runnen is. Als het scherm verdwijnt kunt u de resultaten aflezen in de sheet genaamd 'Resultaat'.

In het geval dat u extra indicatoren observeert, kunt u deze toevoegen door de desbetreffende states aan te vinken en opnieuw op een van de twee knoppen drukken.

Het runnen van de robuustheidsanalyse duurt ongeveer 25 à 30 seconden. In het geval dat een snel resultaat gewenst is, is het aangeraden om op de 'Bereken het resultaat' knop te drukken

BEREKEN HET RESULTAAT **BEREKEN DE ROBUUSTHEID VAN HET RESULTAAT**

Figuur 4. Prototype van de tool voor het voorspellen van drone dreigingen

	Kans(%)	1	2	3	4	5	6	7	8
H5	62,6	80,6	16,6	2,7	0,1	0	0	0	0
H1	23,6	17,7	67,5	13,3	1,4	0,1	0	0	0
H2	8,8	1,7	13,8	59,6	21,8	2,7	0,4	0	0
H6	3,4	0	2	22,6	56,6	14,1	4,3	0,4	0
H7	1	0	0,1	1,7	12,6	50,9	29,9	4,2	0,6
H8	0,6	0	0	0,1	7,5	29,7	53	8,4	1,3
H3	0	0	0	0	0	0,5	3,5	42,4	53,6
H4	0	0	0	0	0	2	8,9	44,6	44,5

Figuur 5. Output van het prototype inclusief robuustheid van het resultaat

elkaar eens zijn over de waarden van de kansen. De kansen in het Bayesiaanse netwerk bepalen de a posteriori kansen. Hoe dichter de a posteriori kansen bij elkaar liggen, hoe groter de kans dat er een rangverschuiving optreedt als er noise wordt toegevoegd aan de kansen. Verder geldt dat de noise die wordt toegevoegd aan de kansen groter is als de experts het minder met elkaar eens zijn. Hierdoor treedt er meer variatie op in de a priori en voorwaardelijke kansen en als gevolg de a posteriori kansen waardoor er meer rangverschuivingen zullen optreden.

Het communiceren van het resultaat van het Bayesiaanse netwerk naar de gebruiker

Het uiteindelijke doel van het Bayesiaanse netwerk is om de politie te ondersteunen bij de keus welke actie er ondernomen moet worden tegen de drone. Om de resultaten van het Bayesiaanse netwerk te communiceren naar de politie toe, is er een prototype van de tool ontwikkeld in Excel. Dit prototype is een eerste opzet voor de interface die de politie mogelijkerwijs in de toekomst kan gaan gebruiken. Figuur 4 toont het prototype.

In figuur 5 is te zien dat de meest waarschijnlijke hypothese H5 in 80,6 procent van de gevallen ook de meest waarschijnlijke hypothese was in de Monte Carlo simulatie. Dit impliceert dat het resultaat redelijk robuust is.

Conclusie

Een Bayesiaans netwerk biedt een goede manier om de dreiging van een drone te voorspellen. Een nadeel echter is dat het gebruik maken van mening van experts onzekerheid veroorzaakt in het resultaat van het netwerk. De robuustheid kan worden vergroot als er meer zekerheid is over de waarden van de kansen. Hierbij kan een drone incident database uitkomst bieden: gegeven voldoende data kunnen de a priori en voorwaardelijke kansen accurater worden ingeschat. Een ander nadeel is dat het huidige Bayesiaanse netwerk opgesteld is rondom een scenario. Als er een grote verandering optreedt in het scenario, moeten er nieuwe hypothesen en indicatoren gedefinieerd worden als ook de kansen opnieuw ingeschat worden wat erg veel werk is. Meer onderzoek naar het efficiënt uitbreiden van het Bayesiaanse netwerk voor meerdere scenario's is gewenst.

Literatuur

Kipersztok, O., & Wang, H. (2001). Another look at sensitivity of Bayesian networks to imprecise probabilities. In *Proceedings of Machine Learning Research, R3* (pp.149–155)

Laura Middeldorp is masterstudent Applied Mathematics aan de Technische Universiteit Delft. Tijdens haar stage bij de afdeling Military Operations van TNO heeft ze een Bayesiaans netwerk opgesteld. E-mail: lmiddeldorp@xs4all.nl

Sven Kramer tijdens de Olympische Spelen in 2010. Foto: David Rosen CC

DE MATRIX VAN OLYMPISCH SCHAATSGOUD

Gerard Sierksma

De hoge prestatiedichtheid aan de top van veel traditionele sporten heeft geleid tot een aantal interessante uitdagingen betreffende onder meer het vergelijken van prestaties van atleten en het beslissen over winnaars. Zo liggen de onderlinge verschillen steeds vaker binnen de foutmarges van de meetsystemen en als het dan gaat om 'goud' is het in die gevallen onmogelijk de winnaar te bepalen. Die uitdagingen betreffen niet alleen het vergelijken van wedstrijdresultaten, ook het bepalen van eenduidige selectieprocessen, waar wordt vastgesteld op welke wijze atleten worden geselecteerd voor belangrijke toernooien, is, gegeven die grote prestatiedichtheid aan de top, een tamelijk precaire aangelegenheid. In het geval van het Nederlandse schaatsen is de poel van competitieve topatleten groot en zullen potentiële medaillewinnaars niet altijd kunnen worden geselecteerd. In het geval van de Olympische Winterspelen gelden strikte quota, waardoor slechts een beperkt aantal atleten het eigen land mag vertegenwoordigen. Geen wonder dat in zulke gevallen de selectieprocessen, die in het verleden werden uitgevoerd door goedwillende experts, met argusogen werden gevolgd en de beslissingen juridisch werden aangevochten.

De top-10 ambitie

In aanloop naar de Olympische Winterspelen 2010 in Vancouver werden in de media de selectieperikelen van de Koninklijke Nederlandsche Schaatsenrijdersbond

(KNSB) breed uitgemeten. Die perikelen inspireerden schrijver dezes en zijn toenmalige afstudeerstudent Yori Zwols tot een idee voor een beslissingsondersteunend systeem, dat zou kunnen helpen bij het bepalen van een 'optimale' selectie met de grootste 'kans' op het winnen van Olympische medailles. We zochten contact met technisch directeur Arie Koops van de schaatsbond om onze ideeën uiteen te zetten en toe te lichten.

Wat toen gebeurde was tamelijk uniek in de, zeker toen nog, conservatieve topsportwereld. Het kostte meer tijd om Koops e-mailadres te vinden dan hem warm te krijgen voor onze ideeën. Hij was onmiddellijk enthousiast, ook omdat hijzelf al zat te broeden op meer data-gedreven beslissingen. Vanaf dat moment zijn we nauw betrokken bij het Olympische selectieproces van de Nederlandse schaatsers. In 2014 werd Bertus Talsma bij ORTEC aangesteld om de ontwikkeling van het beslissingsondersteunende systeem verder gestalte te geven en de bijbehorende berekeningen uit te voeren.

De samenwerking tussen de Rijksuniversiteit Groningen, ORTEC-sports en KNSB heeft geresulteerd in een selectiesysteem dat nu breed gedragen wordt door zowel de sporters als de ondersteunende staf. Juridische procedures van schaats(st)ers en coaches behoren vrijwel helemaal tot het verleden. Tijdens de laatste twee Olympische Winterspelen, Sotsji (2014) en Pyeongchang (2018), bereikte de Nederlandse ploeg de doelstelling van technisch directeur Maurits Hendriks van het Nederlands Olympisch Comité (NOC*NSF) om de top-10 op de, weliswaar niet-officiële, Medaillespiegel Olympische Winterspelen te behalen, met in 2014 maar liefst 23 schaatsmedailles van de in totaal 24 Nederlandse medailles; in 2018 waren dat er 16 van de 20. In beide gevallen behaalde Nederland zelfs plek 5 in de Medaillespiegel (zie figuur 1).

SOCHI 2014

Rank	Country	Gold	Silver	Bronze	Total
1	Russia	13	11	8	33
2	Norway	11	5	10	26
3	Canada	10	10	5	25
4	United States	9	7	12	28
5	Netherlands	8	7	9	24
6	Germany	8	6	5	19
7	Switzerland	6	3	2	11
8	Belarus	5	0	1	6
9	Austria	4	8	5	17
10	France	4	4	7	15

PYEONGCHANG 2018

Rank	Country	Gold	Silver	Bronze	Total
1	Norway	14	14	11	39
2	Germany	14	10	7	31
3	Canada	11	8	10	29
4	United States	9	8	6	23
5	Netherlands	8	6	6	20
6	Sweden	7	6	1	14
7	South Korea	5	8	4	17
8	Switzerland	5	6	4	15
9	France	5	4	6	15
10	Austria	5	3	6	14

Figuur 1. Medaillespiegels Olympische Winterspelen

Knellende quota

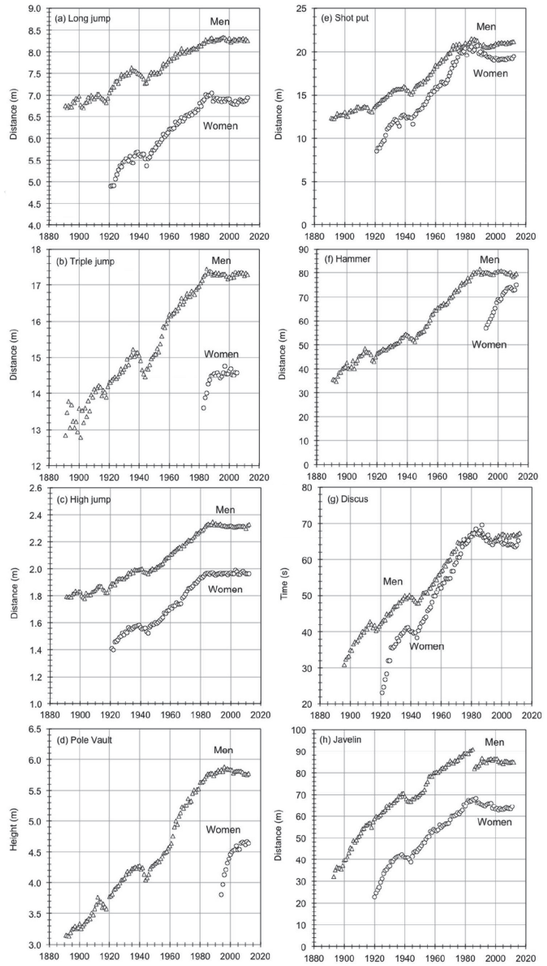
Om duidelijk te maken waarom 'quota' bij Olympische Winterspelen een oorzaak zijn van de schaatsselectieproblemen, schetsen we de situatie van de laatste Winterspelen, die van 2018. Nederland krijgt dan van het Internationaal Olympisch Comité (IOC) een totaal van 38 startposities bij het langebaanschaatsen toegewezen: 19 voor de vrouwen en 19 voor de mannen. Die tweemaal 19 startposities waren verdeeld over zeven disciplines, namelijk voor de vrouwen drie startposities op de 500 meter, drie op de 1.000 meter, drie op de 1.500 meter, drie op de 3.000 meter en twee startposities op de 5.000 meter. Voor de mannen gold 500(3), 1.000(3), 1.500(3), 5.000(3) en 10.000(2); tussen haakjes staat het aantal startposities. De in 2018 geïntroduceerde massastartcompetitie, waarbij de schaats(st)ers tegelijk starten, kende twee startposities per deelnemend land voor zowel de vrouwen als de mannen. Daarnaast mochten nog drie schaatsers en drie schaatssters deelnemen aan de beide ploegenachtervolgingen. Voor bijna elke discipline had Nederland een grote poel van potentiële medaillewinnaars. Daarbij kwam nog de extra complicerende IOC-beperking op het totaalaantal Nederlandse deelnemers, namelijk maximaal tien per sekse, zodat dus 10 atleten 19 startposities moesten 'vullen'. Omdat dus het aantal startposities behoorlijk groter was dan het aantal in te zetten schaats(st)ers, moesten meerdere schaats(st)ers op meerdere disciplines starten en was het daarom niet mogelijk om voor elke startpositie en elke discipline de beste atleet te selecteren. (Een klein voordeeltje in 2018 was dat we ons konden beperken tot zes disciplines, omdat de ploegachtervolgers zich eerst moesten kwalificeren voor een afstandsonderdeel.) Duidelijk is dat, om de top-10 in de Olympisch Medaillespiegel te bereiken, het 'optimaal' was om 'generalisten' te selecteren in plaats van 'specialisten'. Voor de komende Winterspelen is de selectie nog lastiger, omdat dan slechts 9 in plaats van 10 schaat(st)ers mogen meedoen.

Het hierboven geschetste quota-regiem bij de Olympische Winterspelen geldt niet voor Wereldkampioenschappen Schaatsen voor Afstanden (WSDC). Dit toernooi kent vrijwel hetzelfde programma als de Winterspelen, maar het aantal startposities is bij het WSDC gelijk aan het aantal deelnemende schaats(st)ers. Hier is het selectieprobleem dan tamelijk recht-toe-recht-aan op te lossen: selecteer voor elke afstandsdiscipline de hoogst gerangschikte schaats(st)ers van het er aan voorafgaande kwalificatietoernooi.

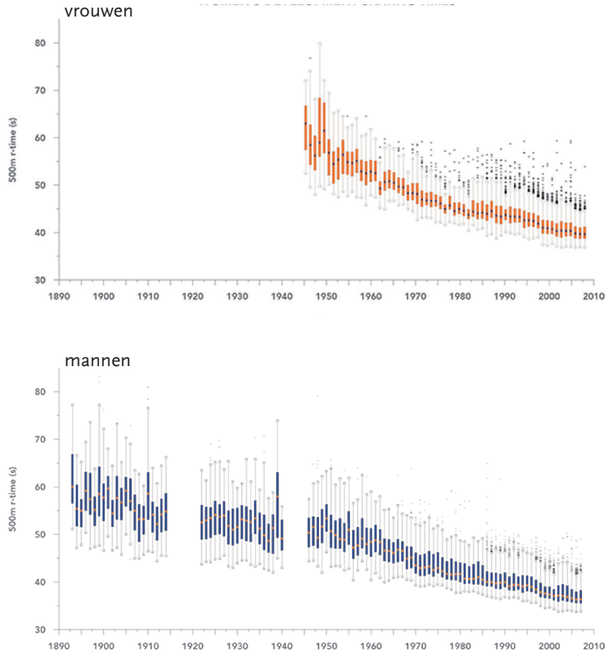
Simpel dus bij de WSDC. De Winterspelen kennen grotere en, leuk voor ons, interessantere uitdagingen vanwege dus die knellende quota. Duidelijk is dat een oplossing voor het selectieprobleem afhangt van een operationele vertaling van de doelstelling 'de top-10 bereiken', met vanzelfsprekend inachtneming van de quotabeperkingen. We hebben ervoor gekozen om medaille-winnende-kansen te berekenen voor de door de KNSB aangewezen potentiële schaats-Olympiërs en daaruit een 'optimale' selectie te bepalen. Voor ons lag de uitdaging om van deze, nog steeds tamelijk vage doelstelling, chocola te maken. Maar juist op zulke momenten begint het ware *STAtOR*-hart sneller te kloppen met evenwel een onverwachte wending in het verschiet. Maar eerst geven we een korte analyse van de hierboven vermelde prestatiecongestie aan de top.

Een evolutiebioloog over prestatiecongestie

Hoewel we ongetwijfeld allemaal topsporters kennen die sneller rennen of hoger springen dan ooit (zie Kuper & Sterken, 2003), is het wellicht minder bekend dat onderlinge verschillen tussen topprestaties in de loop der jaren kleiner en kleiner worden. Een interessant artikel van Haake, James en Foster (2015), bevat grafieken (figuur 2) die de evolutie weergeven van topprestaties op acht klassieke atletiekonderdelen: verspringen, kogelstoten, hinkstap-sprong, hamergooien, hoogspringen, discuswerpen, speerwerpen en polsstokhoogspringen, voor zowel de mannen als de vrouwen. Wat blijkt? Direct na de Tweede Wereldoorlog is er sprake van, zoals ze dat zelf zeggen, dramatische prestatieverbeteringen, die vervolgens aan het einde van de twintigste eeuw tot stilstand komen. Zien we dit verschijnsel ook bij het schaatsen?



Figuur 2. De evolutie van atletiek topprestaties. Overgenomen uit 'An improvement index to quantify the evolution of performance in field events' door S. Haake, D. James & L. Foster, 2015, *Journal of Sports Sciences, 33*(3)



Figuur 3. *Box plots* van schaatsresultaten van vrouwen en mannen, herleid naar 500-metertijden

Figuur 3 geeft per schaatsseizoen *box plots* van alle schaatsresultaten over een periode van meer dan honderd jaar. Hierbij zijn alle resultaten omgerekend naar 500-metertijden. Zo zijn de 1500-metertijden gedeeld door drie en de 10-kilometertijden door 20. Naast de evident dalende trend in beide grafieken, is ook te zien dat de verschillen tussen de beste en de slechtste eindtijden per seizoen geleidelijk afnemen.

Wat vertelt de Amerikaanse evolutiebioloog Stephen Jay Gould ons over dit onderwerp? Vrij naar Gould (1996): Wanneer de prestaties van individuen van 'complexe systemen' zich verbeteren in de loop der tijd en wanneer de gehanteerde 'spelregels' niet veranderen, dan worden de marges van geleverde topprestaties steeds kleiner en nemen ook hun onderlinge verschillen af. Er doet zich, volgens Gould, een voortdurende verbetering voor van het algemene prestatieniveau, puur ten gevolge van bezig zijn en trainen. Hij noemt dat het *maturation proces*. Daardoor worden, volgens Gould, meer en meer de grenzen bereikt van wat menselijkerwijs mogelijk is, namelijk een prestatieplafond dat leidt tot *nivellering van prestaties aan de top en tot het schaarser worden van uitzonderlijke prestaties*. Zodra de spelregels worden veranderd kan het plafond hoger komen te liggen en, volgens Gould, zal het ook steeds sneller gaan dat de toppers het nieuwe plafond benaderen.

Mij dunkt dat het ogenblikkelijk bestempelen van 'vals spel', als een atleet een zeldzame en uitzonderlijke prestatie levert, een derhalve tamelijk voorbarige conclusie is. Atleten als Lance Armstrong, Usain Bolt en Sven Kramer hebben niet bij voorbaat andere spelregels gehanteerd dan hun concurrenten, ze waren gewoon een forse 'snok' beter dan hun concurrenten en hebben met hun uitzonderlijke klasse het oude plafond opgehoogd.

Gould ontdekte het congestiefenomeen bij zijn favoriete sport *American baseball*. Hij observeerde de verdwijning van 40% *hitters* in de Major League. Dergelijke *hitters* zijn spelers die in één seizoen minimaal 40% van hun officiële slagbeurten raakslaan. Dergelijke slagpercentages waren tamelijk gebruikelijk in de jaren 1920 tot 1930, maar hebben zich sinds 1941 niet meer voorgedaan. Commentatoren gebruikten het niet meer voorkomen ervan om te betogen dat het algemene niveau van de honkballers was gedaald. Gould kwam echter met de tegenovergestelde verklaring, namelijk dat het algemene niveau juist was gestegen. Volgens Gould leidt zo'n prestatieplafond dus tot congestie naar dat plafond toe en een toenemende zeldzaamheid van extreme gebeurtenissen. Het is dan ook geen wonder dat de verschillen tussen finishtijden, ook bij het langebaanschaatsen, uiteindelijk zo klein kunnen worden, dat ze binnen de foutmarges van de meetsystemen komen te liggen en dan in feite onvergelijkbaar zijn.

En dit is en was zeker niet alleen maar theorie. Zo verloor de Nederlandse schaatser Koen Verweij in 2014 Olympisch goud op de 1.500 meter in zijn race tegen Zbigniew Bródka uit Polen met een verschil van 0,003 seconden, ruim binnen de foutmarge van de officiële meetsystemen. Een typische ex aequo-situatie dus: beide atleten hadden goud moeten krijgen. Voorlopig zijn wij als sportliefhebbers in zulke gevallen gedoemd tot een treurig geloof in beslissingen van finishfoto-experts. Geen wonder dat er ook onrust is in de gelederen van menig topsporter. De topprestatiecongestie maakt het vergelijken en dus ook het selecteren van atleten inderdaad tot een precaire aangelegenheid.

Algoritmen, prestatiematrices en academici

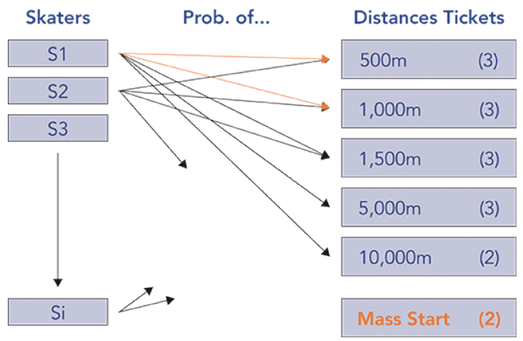
Zoals eerder vermeld, was een van de doelstellingen van het Nederlands Olympisch Comité het bereiken van de top-10 in de Olympische Winterspelenmedailletabel. Daarbij waren de volgende drie randvoorwaarden van belang:

1. Draagvlak. De betreffende atleten en coaches zijn akkoord met de selectieprocedure;
2. Controleerbaar en herhaalbaar. Wanneer de selectieprocedure (later) wordt herhaald met dezelfde data, dan is de uitkomst onveranderd;
3. Juridisch waterdicht.

Aan ons de taak om onder deze voorwaarden een systeem te ontwerpen waaruit een selectie rolt met de hoogste kans op een klassering in die top-10. Om deze doelstelling binnen de quota-restricties te halen deden we wat OR-professionals eigenlijk altijd doen, namelijk: data verzamelen, modellen en algoritmen ontwerpen, en scenarioanalyses uitvoeren onder meer ter beoordeling van de stabiliteit van de oplossing (kleine variaties in mogelijk onzekere parameterwaarden kunnen namelijk grote gevolgen hebben voor de uitkomst.) Een cruciaal moment betrof de beantwoording van de vraag: Wat bedoelen we precies met 'winkansen' en hoe bepalen we ze? In overleg met de KNSB is ervoor gekozen de resultaten te gebruiken van internationale schaatstoernooien tijdens de twee seizoenen voorafgaand aan de Spelen, waarbij we meer gewicht moesten leggen bij de recente toernooien.

Een van onze trucs werd het gebruikmaken van zogenaamde AV5-waarden. Voor elke afstandswedstrijd is de AV5-waarde van een schaats(st)er het verschil tussen de werkelijke finishtijd van die schaats(st)er en het gemiddelde van de eerste vijf finishtijden op die afstand. (De keuze voor AV5-waarden in plaats van bijvoorbeeld AV4 of AV6-waarden was tamelijk arbitrair, hoewel onze scenarioanalyses met AV5-waarden de meest stabiele uitkomsten lieten zien (zie Talsma, 2013). Deze 'normalisatie' van de racetijden corrigeerde voor de mogelijk ongelijke omstandigheden tijdens de toernooien. Zo waren op de hooggelegen ijsbanen van Salt Lake City en Calgary de racetijden in onze dataset gemiddeld genomen veel sneller dan die van de laaggelegen banen zoals Heerenveen. Voor alle schaats(st)ers in de dataset zijn per afstand statistische verdelingen gemaakt van zijn/haar AV5-waarden. Met de als lognormale distributies gemodelleerde verdelingen zijn vervolgens race-simulaties uitgevoerd, namelijk telkens 5.000 simulaties. Voor elke run zijn de simulatieresultaten eerst gerangschikt. Uit het percentage van de 5.000 runs, dat de betreffende Nederlandse schaats(st)er als eerste, tweede of derde finishte, is een winkans op die afstand bepaald, rekening houdend met de prioritering 'goud-boven-zilver-boven-brons'. De matrix (één voor de vrouwen en één voor de mannen) met daarin de Nederlandse schaats(st)ers met hun winkansen, werd in die tijd bekend als de Prestatiematrix.

De berekening van de Prestatiematrices betrof de statistische kant van het proces, het ontwerp van een model waarmee een 'optimale' selectie kon worden berekend de OR-zijde. Met een geheeltallig lineair optimalisatiemodel is dit klusje geklaard. De basis-ingrediënten ervan zijn schematisch weergegeven in figuur 4. De 'volledige bipartite graaf' in deze figuur bevat de namen van de schaats(st)ers en de te schaatsen disciplines, waarbij de getallen tussen haakjes het aantal startposities op die discipline is. De beperkingen van het model modelleren de verschillende quotumbeperkingen op de aantallen startposities en op de omvang van de beide selecties (10 mannen of 10 vrouwen). Het model heeft twee logische beperkingen die worden gebruikt om de 'afstandsparticipatie' te koppelen aan de 'teamparticipatie'. Het maximaliseren van de doelstellingsfunctie, met de getallen van de Prestatiematrix als parameters, leidde tot een team



Figuur 4. Bipartite graaf met inputdata

Sven Kramers onverwachte verlies tijdens de Olympische Spelen in 2010 (foto: David Rosen CC) en Esmee Vissers onverwachte winst tijdens het OKT in Heerenveen in 2017 (foto: Bart Noort CC)

(een voor de vrouwen en een voor de mannen dus) met een grootste kans op Olympische medailles, zoals gezegd prio goud-zilver-brons. Klus geklaard, eenvoudig en duidelijk. Althans, dat dachten we toen. De praktijk bleek weer eens weerbarstiger.

Algoritmen in een weerbarstige praktijk

De acceptatie van onze aanpak zou het einde betekenen van het gebruikelijke en zeer aantrekkelijke Olympisch kwalificatietoernooi (OKT) met volle stadions. Dat OKT vindt plaats in Heerenveen gedurende vijf dagen in december voorafgaand aan de Spelen in februari. Onze statistische simulaties en optimalisatiealgoritmen zouden niet de gewaardeeerde pecunia in het KNSB-laadje brengen. Gelukkig voor ons bleef Arie Koops optimistisch over onze bijdrage en kwam snel met een nieuwe uitdaging: Hoe kunnen we het populaire en lucratieve OKT in de selectieprocedure opnemen? En hij opperde het idee om rangschikkingen te maken van de tweemaal 16 startposities met bovenaan de afstandpositie met de grootste winkans, enzovoort tot plaats 16 met de afstandpositie met de kleinste winkans (Koops, Sierksma, 2010). We besloten om de volgorde van de zo geheten Selectievolgordes (SeVo's) te baseren op de waarden van de Prestatiematrices. Deze keuze was en is zeker discutabel omdat de Prestatiematrix 'persoons'waarden bevat, terwijl de beide SeVo's zijn gebaseerd op 'team'waarden, namelijk op de winkansen voor Nederland, gemaakt door ongeacht welke Nederlander. (De vraag is, hoe hier toch het 'landsbelang'argument kan worden bepaald en gehanteerd.)

Direct na afloop van het OKT en op basis van de uitslagen werden de beide SeVo's van namen voorzien. De eerste positie van een SeVo werd ingevuld met de naam van de schaats(st)er die als eerste eindigde op de betreffende afstand tijdens het OKT. En zo voort. Zodra er tien verschillende namen waren ingevuld, werden de andere posities van het SeVo ingevuld met reeds ingevulde namen, rekening houdend met de OKT-uitslagenvolgordes. Daarnaast had de KNSB de mogelijkheid om gemotiveerd af te wijken van de uitkomsten van beide procedures. Zo'n ingreep zou zich kunnen voordoen als een 'zeer waardevolle' schaats(st)er, voor bijvoorbeeld de Ploegenachtervolging met een grote goudkans, vlak voor het OKT ziek zou zijn geworden. Daarmee kan de ontworpen selectieprocedure inderdaad worden beschouwd als een echt beslissingsondersteunend systeem.

Hoge verwachtingen en teleurstellende resultaten

Selectiebeslissingen zijn meestal gebaseerd op huidige verwachtingen omtrent toekomstig functioneren. Niemand weet op het moment van de selectiebeslissing hoe de geselecteerde atleet over twee maanden zal gaan presteren. Het hoort bij de topsport dat hoge verwachtingen tot grote belangstelling van de media, de sponsoren en natuurlijk ook de fans leiden. En dat betekent volle buidels en volle stadions. Maar er is ook een keerzijde. Die hoge verwachtingen kunnen leiden tot grote teleurstellingen, terwijl lage verwachtingen verrassende resultaten in zich kunnen bergen. De volgende twee gebeurtenissen mogen deze uitersten illustreren.

Olympische Spelen van Vancouver 2010. De grootste Nederlandse schaatser aller tijden, Sven Kramer, was voorbestemd om de 10km, de langste olympische schaatsafstand, te winnen. Hij had niet alleen de hoogste winstkans (100%) in onze Prestatiematrix, ook was hij de nummer één op die afstand tijdens het OKT. Met nog een kwart van de Olympische race te gaan was Kramers voorsprong op de, op dat moment, nummer twee staande schaatser praktisch onoverbrugbaar. Het duurde slechts een fractie van een seconde om de torenhoge verwachtingen de grond in te boren. Kramer nam op dat moment de verkeerde afslag (tijdens elke ronde wisselen schaatsers van rijbaan) en werd hij gediskwalificeerd voor het rijden in de verkeerde baan tijdens de laatste ronden. Ondanks twee nieuwe pogingen is het Kramer nooit gelukt goud te winnen op zijn geliefde Olympische 10 kilometer.

OKT in Heerenveen 2017. Het gebeurde op de laatste dag tijdens de allerlaatste rit van dat OKT, op de 5km voor vrouwen. We zaten op de KNSB-tribune omringd door 'koninklijke' schaatsbestuursleden. Om ons heen heerste de overtuiging dat deze laatste race overbodig was met het oog op de SeVo: die zou niet meer veranderen door de uitslag van deze allerlaatste rit. De Olympische vrouwenselectie lag vast. De zich voorbereidende schaatsvrouwen werden kansloos geacht. Op zo'n moment kunnen uitkomsten van kansberekeningen leiden tot een geheel andere kijk op de werkelijkheid: zeker als de verwachtingen afwijken van wat alom om je heen wordt voorspeld. Onze berekeningen gaven aan dat een van de aan-de-start-staande vrouwen, namelijk Esmee Visser, een hoge waarde had in de Prestatiematrix en zou worden geselecteerd volgens ons lineaire optimaliseringsmodel. Twee maanden later was Esmee de glorieuze winnares van Olympisch goud op de 5 kilometer.

Samenwerking, succes en zorg

In de hectische wereld van topsporten met kleine onderlinge prestatieverschillen en beperkte aantallen startplaatsen, dienen selectieprocedures vanzelfsprekend extra zorgvuldig te worden uitgevoerd. De drie uitgangseisen, zoals die zijn geformuleerd in sectie 5, dienen in Nederland als benchmark bij de evaluatie van het schaatsselectiebeleid. Was die brede steun van de atleten, hun begeleidende teams, de media en de fans er inderdaad? Waren er geen juridische procedures en was de gehele selectieprocedure doorzichtig en 'controleerbaar'? Tijdens de nu meer dan acht jaar samenwerking met ORTEC en de Rijksuniversiteit Groningen is het NOC*NSF nauwelijks geconfronteerd met juridische aanvechtingen. De tribunes waren overvol tijdens de vijf dagen durende OKT's. En de verwachting is dat dat de komende jaren zo blijft, zeker omdat de verwachtingen op olympische oranjesuccessen hoog zullen blijven. Hansje Brinkers *finger in the dyke* (Wikipedia, z.d.) zal Nederland niet beschermen tegen de gevolgen van het smeltende ijs op de polen en de bij ons zachter wordende winters. Het kunstijs van de maar liefst 19 ijsbanen in ons land achter hoge dijken is het Nederlandse antwoord op dit aspect van klimaatverandering. Onze bijdrage was in dit opzicht klein, maar zeker effectiever dan *the finger* van Hansje Brinker.

De vraag blijft hoe we moeten omgaan met het toenemend aantal ex aequo-situaties in de topsport. Zelfs als nano-nauwkeurig de sportuitslagen bepaald gaan worden en de toekijkende sportliefhebber niet meer ziet wie heeft gewonnen en dus de apparatuur moet geloven, moeten we ons afvragen of het niet verstandiger is met tolerantiemarges en ex aequo's te werken. Of moeten we, om met Stephen Jay Gould te eindigen, de (spelregels van de) sport aanpassen, zodat nieuwe ruimtes ontstaan. Ik denk dit laatste. Innoveren is inspireren, ook de toeschouwer.

Noot

Bovenstaande tekst is gebaseerd op het artikel 'The Dutch Approach in Selecting Olympic Speed Skaters' in *OR/MS Today* van Gerard Sierksma en Bertus Talsma, met als ondertitel 'How the Netherlands deals with a large pool of highly competitive athletes and strengthens its place as the greatest speed skating nation on Earth by forging a sustainable partnership between the Dutch Olympic Committee, ORTEC, and the University of Groningen'.

Literatuur

Kuper, G. H., Sterken, E. (2003). Endurance in speed skating; The development of world records. *European Journal of Operational Research 148*, 293–301.

Haake, S. James, D. Foster, L. (2015). An improvement index to quantify the evolution of performance in field events, *Journal of Sports Sciences, 33*(3), 255–267.

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin*. Three Rivers Press

Wikipedia. (z.d). https://ps:en.wikipedia.org/wiki/hans_brinker_or_the_silver_skates

Koops, A. Sierksma, G. (2010). De prestatiematrix als hulpmiddel bij het selectiebeleid van de KNSB. *Sport Knowhow XL*

Robinson, J. (2018, 23 februari). The secret of Dutch speed skating; It's not what you think. *The Wall Street Journal.*

Sierksma, G. (2017). Introduction: Olympics track & field. In J. J. Cochran, J. Bennett & J. Albert (Eds.), *The Oxford Anthology of Statistics in Sport, Vol. 1.* Oxford University Press.

Sierksma, G., & Talsma, B. G. (2021). The Dutch approach in selecting olympic speed skaters. *OR/MS Today, 48*(2), https://doi.org/10.1287/orms.2021.02.09.

Talsma. B., Sierksma, G., & Turkensteen, M. (2017). The growing problem of comparing elite sport performances: The olympic speed skating case. *Journal of Human Sport and Exercise, 13,* 892–907.

Talsma, B. G. (2013). *Performance analysis in elite sports* (PhD thesis, SOM Research School). University of Groningen.

Gerard Sierksma is emeritus-hoogleraar Operations Research aan de Rijksuniversiteit Groningen. Sierksma adviseert, in samenwerking met ORTEC-Sports, het NOC*NSF over Olympische selectieprocedures.
E-mail: g.sierksma@rug.nl.

Voorraadbeheersing in de praktijk: over foute formules, niet-toegelaten oplossingen en wiskunde op drijfzand

Ton de Kok

Met een titel als bovenstaande hoop ik natuurlijk de aandacht te trekken. Die aandacht zou moeten leiden tot belangstelling voor een vakgebied, voorraadbeheersing, waar nog veel te halen is als het gaat om het uitvinden van de wiskunde die het mogelijk maakt om heuristieken met een gelimiteerde bruikbaarheid te vervangen door exacte analyses, die diepere inzichten mogelijk maken. Die inzichten kunnen dan weer leiden tot numerieke methodes die door hun implementatie kunnen leiden tot betere operationele beslissingen, die uiteindelijk een hogere winstgevendheid van bedrijven impliceren, maar ook minder verspilling door het verminderen van het aantal spoedorders, het verminderen van verschrotting van materialen, en betere benutting van schaarse middelen. Het goede nieuws is dat de empirische validiteit van wiskundige modellen voor voorraadbeheersingsvraagstukken is aangetoond, min of meer zoals we dit doen in de natuurkunde. Maar de wiskundige analyse van deze modellen is verre van af. Ook goed nieuws, dus.

Dit verhaal is als volgt gestructureerd. We introduceren het begrip materiaalcoördinatie en formuleren hiervoor een basismodel. Dat basismodel maakt het mogelijk om de doelstelling van materiaalcoördinatie te formuleren, en de te nemen wiskundige hordes aan te duiden. De essentie van deze hordes zijn de stochastische processen waarmee we te maken hebben. Zelfs voor simpele voorraadbeheersingssituaties zijn exacte methodes al numeriek complex en daarmee tijdsintensief. Dat maakt grootschalige toepassing nog altijd lastig, ook al worden computers steeds sneller. We bespreken de typische *shortcuts* die op dit moment gebruikt worden binnen de zogenaamde Enterprise Resource Planning (ERP) systemen, die bij vrijwel ieder bedrijf gebruikt worden. Hier duiken dan de foute formules en niet-toegelaten oplossingen op. Deze hebben implicaties voor het dagelijks werk van planners, schedulers en managers. We betogen dat er in wezen oplossingen voorhanden zijn, die de proof-of-concept fase al geruime tijd voorbij zijn, maar nog altijd niet gemeengoed zijn. Vervolgens bespreken we een aantal wetmatigheden die nuttig zijn bij het zoeken naar nieuwe resultaten. En tot slot de wiskundige uitdagingen die de moeite waard zijn om verder te exploreren.

Materiaalcoördinatie

Voorraadbeheersing betreft het monitoren van de hoeveelheid op voorraad van een item op een locatie, zodat aan de vraag van klanten kan worden voldaan en tijdig orders kunnen worden geplaatst bij de leverancier van het item om de voorraad aan te vullen. Dit laatste noemen we de ordervrijgave. Materiaalcoördinatie is het coördineren van de ordervrijgavebeslissingen van items die worden gebruikt bij de productie en distributie van producten. Denk hierbij aan grondstoffen, componenten, ingredienten, halffabricaten en eindproducten op verschillende locaties om de marktvraag tijdig te kunnen voldoen. Coördineren suggereert afstemming. Deze afstemming dient eruit te bestaan dat ordervrijgavebeslissingen uitvoerbaar zijn en dat de beslissingen die vandaag genomen worden, voorbereiden op wat morgen en daarna besloten wordt. Je zou kunnen zeggen: de beslissing vandaag moet kunnen, maar die van morgen ook. De huidige materiaalcoördinatieconcepten maken dit zichtbaar door voor ieder beheerst materiaal het verloop van haar voorraad in de tijd aan de planner te laten zien, evenals de ordervrijgavebeslissingen in de tijd die leiden tot dit voorraadverloop. Hierbij is het meestal zo dat een ordervrijgavebeslissing twee gezichten heeft. Het genereert een toekomstige aanvulling in het ontvangende voorraadpunt en een onmiddellijke afname bij de leverende voorraadpunten. Dit laatste is belangrijk om te onthouden: één ordervrijgavebeslissing betekent dat van meerdere materialen de gewenste hoeveelheid beschikbaar moet zijn. Juist daaruit hoort een belangrijk deel van de coördinatiefunctionaliteit van een concept te bestaan, want meestal zijn door allerlei oorzaken de werkelijk beschikbare hoeveelheden materiaal niet precies op elkaar afgestemd. En dan is het handmatig bepalen van de juiste hoeveelheden praktisch onuitvoerbaar. Vooral ook omdat het oplossen van een probleem op de ene plaats lijkt te leiden tot een nieuw probleem op een andere plaats. Bij materiaalcoördinatie lijkt alles met alles samen te hangen. En dat lijkt niet alleen zo, het is ook echt zo. Om dat te verduidelijken, formuleren we een basis-

N	Aantal items
E	Verzameling items met exogene vraag
a_{ij}	Aantal items i nodig om één item j te maken, $i=1,...,N$, , $j=1,...,N$
T	Beslissingenhorizon
$D_i(t)$	Exogene vraag naar item i in periode t, $t=1,...,T$, $i=1,...,N$
$F_{t,i}(t+s)$	Voorspelling gemaakt aan begin van periode t van de exogene vraag naar item i in periode $t+s$, $t=1,...,T$, $s=0,...,T-t$, $i=1,...,N$
L_i	Levertijd van item i, $i=1,...,N$
$X_i(t)$	Netto voorraad van item i aan het eind van periode t, $t=0,...,T$, $i=1,...,N$
$r_i(t)$	Hoeveelheid vrijgegeven van item i aan het begin van periode t, $t=1,...,T$, $i=1,...,N$
h_i	Voorraadkosten per item i in voorraad aan het eind van een periode, $i=1,...,N$
P_k	Boetekosten per item k tekort aan het eind van een periode, $K \in E$
v_i	Veiligheidsvoorraad van item i, $i=1,...,N$

Tabel 1

model voor materiaalcoördinatie. Hierbij veronderstellen we dat capaciteit altijd in voldoende mate aanwezig is om na ordervrijgave van een hoeveelheid voor een item, deze hoeveelheid binnen een vastgestelde tijd, de levertijd, af te leveren in het voorraadpunt van het item.

Basismodel

In ons basismodel gaan we uit van discrete tijd. Het basismodel wordt gedefinieerd door de items en hun onderlinge relaties, de norm voor de tijd nodig om een item uit haar onderdelen te produceren, het exogene vraagproces van de eindproducten en eventuele reserveonderdelen, en de kostenstructuur. De beslissingsvariabelen betreffen de ordervrijgaven. De netto voorraad is gedefinieerd als de fysieke voorraad minus naleveringen (zie tabel 1).

Dan komen we tot een basismodel (model 1). We willen de som van voorraadkosten en boetekosten over een eindige horizon T minimaliseren, waarbij we over deze horizon de juiste beslissingen $\{r_i(t)|1\le i\le N, 1\le t\le T\}$ willen nemen, rekening houdend met de beschikbaarheid van zogenaamde *child items*, de items die nodig zijn om een item te maken. En rekening houdend met de aanname dat wat nu wordt vrijgegeven na de levertijd beschikbaar is. Hierbij is het belangrijk dat de levertijd de *normtijd* is tussen het moment van ordervrijgave, *nadat alle benodigde child items aan de order zijn gealloceerd*, en het moment van ontvangst van de order in het voorraadpunt van het item.

Dit basismodel blijkt in de praktijk al bruikbaar te zijn, omdat eventuele capaciteitsbeperkingen kunnen worden opgelost met korte-termijn-maatregelen, zoals aanpassingen van routeringen en prioriteiten van reeds vrijgegeven orders, inzet van extra personeel en productiemiddelen, uitbesteding, en overwerk. De bruikbaarheid is gebaseerd op twee aanpassingen:

1. Het vervangen van $D_i(t+s)$ door de voorspelling van de vraag $F_{t,i}(t+s)$ op tijdstip t. Immers, bovenstaand model is niet oplosbaar zonder wiskundige veronderstellingen over de **toekomstige vraag** $D_i(t+s)$. Met deze aanpassing veronderstellen we dat we de toekomstige vraag exact kennen!
2. Het introduceren van een veiligheidsvoorraad v_i die onzekerheid in de vraag moet opvangen.

$$\min_{\{r_i(t)|1\le i\le N,1\le t\le T\}} \sum_{t=1}^{T}\sum_{i=1}^{N} h_i X_i^+(t) + \sum_{t=1}^{T}\sum_{k\in E} p_k X_k^-(t)$$

$$s.t.$$

$$\sum_{j=1}^{N} a_{ij} r_j(t) \le X_i(t-1), i=1,...N, t=1,...,T \qquad \text{materiaalbeschikbaarheid}$$

$$X_i(t) = X_i(t-1) + r_i(t-L_i) - \sum_{j=1}^{N} a_{ij} r_j(t) - D_i(t), i=1,...N, t=1,...,T \qquad \text{voorraadbalansvergelijking}$$

$$r_i(t) \ge 0$$

Model 1

Dit leidt tot het LP-model (model 2). Hoewel planningsproblemen vaak als (MI)LP worden geformuleerd, waarna er oplossing wordt berekend met een optimalisatiemethode of heuristiek, is dit niet de manier waarop het materiaalcoördinatieprobleem doorgaans in de praktijk wordt opgelost. Hiervoor wordt een veel eenvoudiger algoritme gebruikt, dat bekend staat als het Material Requirements Planning (MRP) algoritme. Hierbij wordt wederom voor ieder item een veiligheidsvoorraad verondersteld, die als doelvoorraad wordt gekozen aan het eind van de periode, waarin de bestelling binnenkomt, dat wil zeggen:

$$X_i(t+L_i+s) = v_i, i=1,...,N, s=0,...,T-t-L_i. \qquad (1)$$

Dan wordt voor tijdstip t een rollend plan berekend door recursief de ordervrijgave te berekenen, beginnend met de items, die geen opvolgers hebben, en vervolgens de directe voorgangers. Ook nu wordt weer gebruikgemaakt van de voorspelling van de vraag, ofwel de hoeveelheden die men van de items met exogene vraag wil maken om de exogene vraag te voldoen. Dit laatste wordt het Hoofdproductieplan (HPP) genoemd, in het Engels Master Production Schedule (MPS). Het rollend plan wordt dus bepaald als volgt.

$$r_i(t) = \sum_{s=0}^{L_i} F_{ti}(t+s) + \sum_{s=0}^{L_i}\sum_{j=1}^{N} a_{ij} r_j(t+s) + v_i - \sum_{s=0}^{L_i-1} r_i(t+s-L_i), i=1,...,N \qquad (2)$$

Het is niet moeilijk om in te zien dat dit algoritme tot *niet-toegelaten oplossingen* leidt, omdat geen rekening wordt gehouden met de materiaalbeschikbaarheidsrandvoorwaarden. Er zijn twee redenen dat dit algoritme toch het meest gebruikte planningsalgoritme in de industrie is:

1. De rekencomplexiteit is lineair in het aantal items, waardoor het toegepast kan worden bij elk bedrijf.
2. Rond dit algoritme en de systemen die zorgen voor de invoer en uitvoer is een professioneel opleidingsprogramma opgezet door de American Production and Inventory Control Society (APICS), waarmee sinds 1970 vrijwel iedere logistieke professional is opgeleid, direct of indirect (de auteur dezes incluis).

Toen het MRP-systeem in de jaren 1970 werd ingevoerd bij grote bedrijven als Philips, was het een grote sprong vooruit, doordat de computer een groot deel van de administratie overnam, die bij materiaalbeheer nodig is. Met het toenemen van de rekenkracht van computers werd de frequentie van het runnen van het MRP-algoritme opgevoerd van wekelijks naar dagelijks naar real-time. Eigenlijk heeft het MRP-algoritme zich hiermee tegen de gebruiker gekeerd. Immers, iedere keer als een klantenorder binnenkomt, wordt de 'evenwichtssituatie'verstoord: de vergelijking (1) geldt niet meer. En dan zal het MRP-systeem onmiddellijk weer aangepaste ordervrijgaven genereren voor heel veel items. Maar omdat vergelijking (2) geen rekening houdt met de materiaalbeschikbaarheid van *child items*, is het voorgestelde ordervrijgaveplan vaak niet toegelaten en moet er handmatig worden ingegrepen. En heeft men het eindelijk op orde, dan komt de volgende klantenorder binnen, of wijzigen de verkoopplannen. Dit voortdurend herberekenen van het plan, leidend tot afwijkingen ten opzichte van het oorspronkelijke plan, noemt men planningsnervositeit. Tot op de dag van vandaag lijkt het niet toegestaan het

$$\min_{\{r_i(t)|1\le i\le N,1\le t\le T\}} \sum_{s=0}^{T-1}\sum_{i=1}^{N} h_i \left(X_i(t+s) - v_i\right)^+ + \sum_{s=1}^{T-1}\sum_{k\in E} p_k \left(X_k(t+s) - v_k\right)^-$$

$$s.t.$$

$$\sum_{j=1}^{N} a_{ij} r_j(t+s) \le X_i(t+s-1), i=1,...N, s=0,...,T-1$$

$$X_i(t+s) = X_i(t+s-1) + r_i(t+s-L_i) - \sum_{j=1}^{N} a_{ij} r_j(t+s) - F_{t,i}(t+s), i=1,...N, s=0,...,T-1$$

$$r_i(t+s) \ge 0, i=1,...,N, s=0,...,T-1$$

Model 2

probleem bij de wortel aan te pakken: het vervangen van een te simplistisch algoritme door een algoritme dat wel 'werkt'. Al het handmatig ingrijpen wordt beschouwd als normaal werk van een planner.

Het moet gezegd, (MI)LP is nog geen goed alternatief, omdat dit voor de grootte van de problemen niet snel genoeg rekent: voor planning moet herberekenen binnen 10 seconden mogelijk zijn, want men wil nagaan of handmatige aanpassingen, waar we niet onderuit komen (we gebruiken immers een model van de werkelijkheid), effectief zijn. Typische aanpassingen betreffen het versnellen of vertragen van al uitstaande productie- en klantenorders, verhogen of verlagen van verkoopplannen. Maar er blijkt wel degelijk een alternatief te zijn, dat ongeveer evenveel rekentijd vraagt als het MRP I algoritme, en zich al bewezen heeft als een effectief planningsalgoritme in de praktijk, zie De Kok et al. (2005). Dit algoritme is een bijproduct van onderzoek gericht op het bepalen van optimale strategieën voor ordervrijgave in multi-echelon voorraadsystemen, zoals hierboven beschreven, onder stochastisch stationaire vraag. In De Kok (2018) wordt inzicht gegeven in de zogenaamde Synchronized Base Stock (SBS) strategieën, die het mogelijk maken om algemene waardenetwerken te besturen, waarbij in een *split second* een toegelaten plan voor een realistisch waardenetwerk wordt gegenereerd, en waarvan via numerieke studies is aangetoond dat deze plannen tot aanzienlijk lagere (ca. 10%) kosten leiden dan de (MI) LP-gebaseerde plannen. Schouten (2018) heeft aangetoond dat de SBS-strategieën ook kunnen worden aangepast voor laag-volume productie, zoals in high-tech, waar we niet om geheeltalligheid van ordervrijgaven heen kunnen. In een *real-life* experiment toont ze aan dat de SBS-strategieën zonder enig menselijk ingrijpen tot minstens dezelfde prestatie leiden als in de praktijk met veel menselijk ingrijpen.

Praktische inzichten in de besturing van waardenetwerken

Vraag op item niveau is niet van stationair te onderscheiden

De basis voor de toepassing van stochastische modellen is, dat de vraag naar items vrijwel nooit te onderscheiden is van stationaire vraag door de relatief hoge waarde van de standaarddeviatie ten opzichte van het gemiddelde. Hierdoor wordt een mogelijk onderliggend patroon overstemd door ruis. Toepassing van stochastische modellen vraagt wel zorgvuldige modellering en modelanalyse, waar het helaas in de praktijk van voorraadbeheersingssoftware aan schort: *foute formules*.

Stochastische multi-echelon voorraadmodellen zijn empirisch valide

Voor de analyse en 'optimalisatie' van waardenetwerken is de ChainScope software ontwikkeld, ondersteund door een STW Valorisation Grant. Hoewel het bedrijf ChainScope niet van de grond is gekomen, is de software sinds 2009 veelvuldig ingezet bij MSc projecten (cf. De Kok (2018)). De basis van de software bestaat uit de berekening van 'optimale' SBS-strategieën, maar om de software te kunnen inzetten bij ieder bedrijf zijn *heuristieken* toegevoegd om rekening te kunnen houden met seriegroottebeperkingen, maximale voorraadbeperkingen (bijvoorbeeld in geval van tanks en silo's, of instabiliteit van een tussenproduct), en productafkeur (bijvoorbeeld in de semiconductorindustrie). Hoewel bij de wiskundige analyse uitgebreid gebruik is gemaakt van *discrete event simulation*, bleek dat het bouwen van een simulatiemodel voor algemene waardenetwerken met al de genoemde kenmerken, uitermate complex is en voor zover bekend nog niet beschikbaar. Deze problematiek werd ondervangen door empirische validatie.

Klantenservice wordt bepaald door gemiddelde voorraad en bestelfrequentie

Nu hebben we boven al aangegeven dat er door mensen veelvuldig wordt ingegrepen in de ordervrijgave voorstellen van systemen. Toch vinden we empirische validiteit van de gebruikte modellen. De enige verklaring voor deze bevinding is, dat de klantenserviceprestatie van een waardenetwerk wordt bepaald door de gemiddelde voorraad en de gemiddelde ordervrijgavehoeveelheid (of gemiddelde ordervrijgavefrequentie). In De Kok (2018) wordt deze bevinding ondersteund door de prestatie van verschillende één-product-één-locatie voorraadstrategieën te vergelijken, die allen dezelfde gemiddelde voorraad en gemiddelde bestelhoeveelheid hebben. Onder de aanname van stationaire vraag blijken deze strategieën vergelijkbare klantenservice op te leveren, tenzij de gemiddelde bestelgrootte groot is ten opzichte van de gemiddelde vraag per tijdseenheid.

*Optimale besturing creëert een goederen**stroom***

Bij het berekenen van de *optimale voorraadkapitaalverdeling* in waardenetwerken blijken er drie belangrijke inzichten naar voren te komen in alle praktische gevalsituaties:

1. Het grootste deel van het voorraadkapitaal ligt op het zogenaamde klantenorderontkoppelpunt, de meest stroomafwaarts gelegen, op vraagvoorspellingen bestuurde, voorraadpunten.
2. Van items met lage waarde, maar lange levertijden, wordt ook een aanzienlijke voorraad *in tijd* aangehouden.
3. Van het grootste deel van de items wordt niet of nauwelijks voorraad aangehouden, waarmee een groot deel van de items in het waardenetwerk nooit stilliggen, tenzij als *work in progress*.

Het is inzicht 3. dat de grootste uitdaging stelt bij de implementatie van de verkregen inzichten in concrete besturingsparameters voor bestaande materiaalcoördinatiemethoden als MRP-I. We weten dat het MRP-I algoritme niet-toegelaten oplossingen genereert, die handmatig moeten worden aangepast. Het MRP-I algoritme synchroniseert de orders van child items niet, noch heeft het een allocatiemechanisme. En juist dat laatste mechanisme speelt een cruciale rol wanneer we optimale strategieën willen implementeren. Dus zolang we vasthouden aan het MRP-I algoritme, zijn we gedwongen om te hoge voorraden in de keten aan te houden op plaatsen, waar ze niet direct bijdragen aan de echte klantenservice.

Open problemen

Ik heb hierboven aangegeven dat er op dit moment slechts heuristieken bestaan in de vorm van *proprietary knowledge* in de ChainScope software voor waardenetwerken met onder meer seriegroottebeperkingen en productafkeur. Het ontwikkelen van strategieën die op een effectieve manier kunnen worden ingezet in praktische situaties, waar onzekerheid in vraag en aanbod de norm is, vormt een grote uitdaging. Hoewel de huidige modellering en analyse 'werkt', mag gerust worden gezegd dat relaxaties van randvoorwaarden en onterechte aannames van stochastische onafhankelijkheid, leiden tot wiskundig drijfzand.

Een belangrijk ontbrekend aspect is capaciteitsbeperking. Deze kan de vorm aannemen van een maximale output op een productiemiddel per tijdseenheid, of van een maximale hoeveelheid onderhanden werk. Dit laatste type beperking komt veel voor in de high-tech industrie, bijvoorbeeld in het geval van clean rooms en andere typen zogenaamde opstelplaatsen waar een kapitaalgoed wordt geassembleerd en getest.

Natuurlijk is het mogelijk om MILP-formuleringen te gebruiken om rekening te houden met alle mogelijke generieke en specifieke randvoorwaarden. Maar we weten dus dat SBS strategieën tot betere prestaties leiden dan LP-formuleringen binnen een rollend plan aanpak. Het endogeniseren van de onzekerheid in het model van de werkelijkheid is blijkbaar cruciaal. Daarnaast kan ook de rekencomplexiteit van de ontwikkelde strategieën, net als bij SBS-strategieën ten opzichte van LP, veel lager zijn dan bij MILP, terwijl wellicht besturingsparameters kunnen worden geoptimaliseerd. Het zal ongetwijfeld veel inspanning en geduld vragen om de uitdagingen om te zetten in bruikbare planningssoftware, maar de bijdrage aan effectiever en efficiënter gebruik van schaarse middelen en materialen voor de eeuwigheid lijkt mij geen slechte beloning.

Noot

Van dit artikel is een uitgebreide versie, waarin meer uitleg wordt gegeven over de analyse van stochastische modellen, beschikbaar op http://home.kpn.nl/tondekok/

Literatuur

De Kok, T. (2018). Inventory management: Modeling real-life supply chains and empirical validity. *Foundations and Trends® in Technology, Information and Operations Management, 4*(11), 343–437. http://dx.doi.org/10.1561/0200000057

De Kok, T., Janssen, F., Van Doremalen, J., Van Wachem, E., Clerkx, M., & Peeters, W. (2005). *Philips electronics synchronizes its supply chain to end the bullwhip effect. Interfaces. 35*(1), 37–48. https://doi.org/10.1287/inte.1040.0116

Schouten, T., 2018. *Material availability planning in a low volume environment; A case study at ASML.* (Master thesis, Operations Management & Logistics). Eindhoven University of Technology. https://pure.tue.nl/ws/portalfiles/portal/113711086/

Ton de Kok is hoogleraar Operations Management aan de Technische Universiteit Eindhoven en bij de capaciteitsgroep Operations Planning, Accounting, and Control (OPAC) binnen de faculteit Industrial Engineering and Innovation Sciences. Hij is sinds oktober 2020 directeur van het CWI.
E-mail: ton.de.kok@cwi.nl

EEN AANPASSING OP DE NON-PARAMETRISCHE EFFECTMAAT CLIFFS DELTA VOOR DE VERGELIJKING VAN GEDRAGSPROFIELEN

Cliffs Delta is een non-parametrische effectmaat die is gebaseerd op data-observaties. We passen Cliffs Delta aan om er gedragsprofielen mee te vergelijken. Gedragsprofielen zijn dichtheidsverdelingen waarin antwoordgedrag van surveys is samengevat voor specifieke groepen respondenten of items. De aangepaste Cliffs Delta kan twee groepen respondenten vergelijken (bijvoorbeeld hoog- versus laagopgeleide respondenten) op de frequentie van specifiek antwoordgedrag (bijvoorbeeld het antwoorden van 'weet niet'). De aangepaste Cliffs Delta is een solide en conservatieve effectmaat die zowel bruikbaar als gunstig is om gedragsprofielen met elkaar te vergelijken.

Frank Bais & Joost van der Neut

Kwaliteit van surveydata kan worden beïnvloed door respondentkenmerken en itemkenmerken. Zulke kenmerken, bijvoorbeeld opleidingsniveau van de respondent of moeilijk taalgebruik in een item, kunnen leiden tot ongewenst antwoordgedrag, zoals het antwoorden van 'weet niet'. We nemen aan dat elke individuele respondent een latente of 'ware' kans (uitgedrukt in proportie) heeft om bepaald antwoordgedrag te laten zien. Een uitdaging is hoe om te gaan met de onzekerheid die gepaard gaat met de relatieve gedragsfrequentie of gedragskans van respondenten. Hoe kleiner het aantal items dat een respondent invult voor een schatting van bepaald antwoordgedrag, hoe onzekerder het is dat de resulterende gedragskans naar de 'ware' gedragskans van die respondent verwijst. Daarnaast bestaan veel surveys deels uit filtervragen, waardoor sommige respondenten minder items invullen dan andere respondenten. Dit betekent dat respondenten onderling veelal niet makkelijk te vergelijken zijn op hun gedragskans vanwege een verschillend aantal ingevulde items. Om met deze onzekerheden rekening te houden, zodat we groepen respondenten kunnen vergelijken, kunnen we gedragsprofielen construeren.

Een gedragsprofiel is een dichtheidsverdeling die een visuele en statistische samenvatting van antwoordgedrag geeft voor een groep respondenten of een groep items met bepaalde kenmerken, respectievelijk het respondentprofiel en het itemprofiel. Een gedragsprofiel representeert de relatieve kans van een groep respondenten (bijvoorbeeld hoogopgeleide respondenten) of een groep items (bijvoorbeeld items die moeilijk taalgebruik bevatten) op het vertonen van specifiek antwoordgedrag (bijvoorbeeld het antwoorden van 'weet niet') voor alle mogelijke kansen van 0 tot 1. Zie figuur 1 voor enkele voorbeelden van gedragsprofielen. Behalve de technische notatie en de uitwisselbaarheid van de termen 'respondent' en 'item' verloopt het construeren van respondentprofielen en itemprofielen hetzelfde. Voor de overzichtelijkheid behandelen we daarom vanaf hier alleen respondentprofielen.

Het optreden van antwoordgedrag voor een enkele respondent r wordt bepaald door het aantal ingevulde items I_r waarvoor dat gedrag had kunnen worden vertoond en het aantal items G_r waarvoor het gedrag daadwerkelijk is vertoond:

$$\lambda_r(p) = \binom{I_r}{G_r} p^{G_r}(1-p)^{I_r-G_r} \quad \text{en} \quad \tilde{\lambda}_r(p) = \frac{\lambda_r(p)}{\int_{p=0}^{1} \lambda_r(p)dp}. \qquad (1)$$

Hier is λ de kanswaarschijnlijkheid, λ_r is het individuele gedragsprofiel voor respondent r, p is de kans tussen 0 en 1, en $\tilde{\lambda}_r$ is het genormaliseerde individuele gedragsprofiel voor respondent r opdat elk individueel profiel een vergelijkbare oppervlakte krijgt van 1. Vervolgens kan voor een bepaalde groep respondenten het gedragsprofiel worden berekenend door het gemiddelde van alle genormaliseerde individuele profielen te nemen:

$$\bar{\lambda}(p) = \frac{1}{R}\sum_{r=1}^{R} \tilde{\lambda}_r(p) \quad \text{en} \quad \bar{E} = \int_{p=0}^{1} p\bar{\lambda}(p)dp \ . \qquad (2)$$

Hier is $\bar{\lambda}$ het gedragsprofiel, R is het aantal respondenten in de groep, en $\bar{E}$ is de verwachte waarde of kans voor de groep op het gedrag. Zie Bais (2021) voor een uitvoerige verhandeling over gedragsprofielen en hun eigenschappen.

Om twee gedragsprofielen (voor groepen van bijvoorbeeld hoogopgeleide en laagopgeleide respondenten) met elkaar te kunnen vergelijken (op bijvoorbeeld het antwoorden van 'weet niet'), hebben we een statistische maat nodig. In dit artikel introduceren we daarvoor een aanpassing op de bestaande effectmaat Cliffs Delta.

Cliffs Delta voor gedragsprofielen

Cliffs Delta δ is ontwikkeld door Norman Cliff voor het gebruik met ordinale data (Cliff, 1993, 1996ab). De effectmaat berekent de kans dat een willekeurige data-observatie x_a van een groep A groter is dan een willekeurige data-observatie x_b van een groep B, min de omgekeerde kans (Hess & Kromrey, 2004; Rousselet, Foxe, & Bolam, 2016; Rousselet, Pernet, & Wilcox, 2017):

$$\delta = p(x_a > x_b) - p(x_a < x_b) \ . \qquad (3)$$

De steekproefschatting $\hat{\delta}_O$ van Cliffs Delta wordt verkregen door elke data-observatie in groep A te vergelijken met elke data-observatie in groep B:

$$\hat{\delta}_O = \frac{\sum_{a=1}^{R_A}\sum_{b=1}^{R_B} \operatorname{sgn}(x_a - x_b)}{R_A R_B} \ . \qquad (4)$$

De functie $\text{sgn}(x_a - x_b)$ resulteert in 1, 0, of -1 wanneer respectievelijk $x_a > x_b$, $x_a = x_b$, of $x_a < x_b$. Het totaal aantal vergelijkingen is het product van de groepsgroottes R_a en R_b van respectievelijk groep *A* en *B*. Hoe kleiner de overlap tussen de verdelingen van twee groepen, hoe meer de groepen verschillen. Een $\hat{\delta}_o$ van -1 of 1 betekent afwezigheid van overlap tussen twee groepen en een $\hat{\delta}_o$ van 0 betekent gelijkheid van groepen (Hess & Kromrey, 2004).

Om $\hat{\delta}_o$ te transformeren in een aangepaste Cliffs Delta die gedragsprofielen kan vergelijken, realiseren we ons dat elke observatie van een groep *A* precies één keer met elke observatie van een groep *B* wordt vergeleken voor de berekening van $\hat{\delta}_o$. Voor elke afzonderlijke vergelijking hebben beide observaties aldus een 'frequentie' of 'gewicht' van 1. Als we deze frequenties toevoegen aan formule (4) verkrijgen we

$$\hat{\delta}_O = \frac{\sum_{a=1}^{R_A}\sum_{b=1}^{R_B}\text{sgn}(x_a - x_b)(w_a w_b)}{\sum_{a=1}^{R_A}\sum_{b=1}^{R_B}(w_a\, w_b)}, \qquad (5)$$

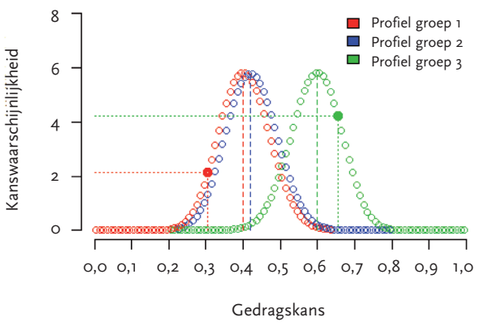
waar w_a en w_b de frequenties van respectievelijk de dataobservaties x_a en x_b van de groepen *A* en *B* zijn. Wanneer we dit idee toepassen op een aangepaste Cliffs Delta $\hat{\delta}_p$ voor gedragsprofielen, dan zouden we de gedragskansen van 0 tot 1, met een bepaalde stapgrootte, als onze 'observaties' kunnen beschouwen en de waarschijnlijkheden voor elke kans als hun bijbehorende 'frequentie' of 'gewicht':

$$\hat{\delta}_P = \frac{\sum_{a=1}^{A}\sum_{b=1}^{B}\text{sgn}(p_a - p_b)\bar{\lambda}_A(p_a)\bar{\lambda}_B(p_b)}{\sum_{a=1}^{A}\sum_{b=1}^{B}\bar{\lambda}_A(p_a)\bar{\lambda}_B(p_b)}, \qquad (6)$$

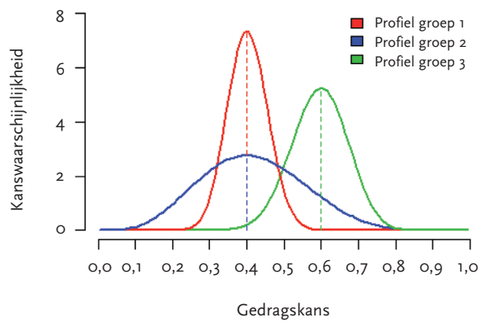
waar p_a en p_b de kansen van 0 tot 1, en $\bar{\lambda}_A$ en $\bar{\lambda}_B$ de gemiddelde kanswaarschijnlijkheden voor groep A en B respectievelijk zijn. Voor de gedragskansen kiezen we een stapgrootte van 0,01, zodat we voor elk profiel 100 intervallen verkrijgen. De middelste waarden van elk interval, {0,005, 0,015, 0,025, ..., 0,995}, kunnen we als onze 'observaties' beschouwen die alle hun eigen gewicht hebben in de vorm van een kanswaarschijnlijkheid.

Zie grafiek 1 in figuur 1 voor een voorbeeldpaar van observaties in het vergelijken van groep 1 en groep 3. Beschouw de gedragskansen van 0,305 en 0,655 en hun respectievelijke waarschijnlijkheden van ongeveer 2,1 en 4,2 (zie de gestippelde lijnen). Het plaatsen van deze waarden in de teller van formule (6) geeft sgn(0,305-0,655) (2,1*4,2) en levert een negatieve contributie aan de totale som van de teller. Alle positieve en negatieve contributies van alle mogelijke paren gedragskansen worden opgeteld. Het resultaat wordt gedeeld door de som van alle mogelijke paren waarschijnlijkheden. Evenals de originele $\hat{\delta}_o$ zal de aangepaste $\hat{\delta}_p$ altijd tussen de -1 en 1 liggen.

In het vergelijken van gedragsprofielen houdt $\hat{\delta}_p$ rekening met zowel de locatie van de profielen ten opzichte van de gedragskansen als met de vorm van de profielen. Zie grafiek 1 in figuur 1. Wanneer twee groepen en daarmee hun verwachte gedragskans dichtbij elkaar liggen, is hun overlap groot en ligt δ_p relatief dichtbij 0 (zie groep 1 en 2). Wanneer twee groepen en hun verwachte gedragskans verder van elkaar vandaan liggen, is hun overlap kleiner en ligt $\hat{\delta}_p$ verder bij 0 vandaan (zie groep 1 en 3).



1. Respondentprofielen voor groep 1, 2 en 3: invloed van profiellocatie



2. Respondentprofielen voor groep 1, 2 en 3: invloed van profielvorm

Figuur 1. Voorbeelden van respondentprofielen met verschillende locaties (grafiek 1) en verschillende vormen (grafiek 2)

Bij het vergelijken van groep 1 met groep 3 vinden we een grote negatieve $\hat{\delta}_p$, wat betekent dat de gedragskans hoger is voor groep 3 dan voor groep 1.

Zie grafiek 2 in figuur 1. Wanneer een profiel relatief 'uitgestrekt' is over de breedte van de gedragskansen (zie groep 2) betekent dit dat er onzekerheid over de verwachte gedragskans voor de groep bestaat. In het algemeen betekent dit dat respondenten weinig items hebben ingevuld. Een uitgestrekt profiel heeft over het algemeen veel overlap met een ander profiel, waardoor $\hat{\delta}_p$ relatief dichtbij 0 ligt (zie groep 2 en 3). Wanneer een profiel relatief 'samengeperst' is (zie groep 1) betekent dit dat de verwachte gedragskans relatief zeker is. In het algemeen betekent dit dat respondenten relatief veel items hebben ingevuld. Een samengeperst profiel heeft over het algemeen weinig overlap met een ander profiel, waardoor $\hat{\delta}_p$ verder bij 0 vandaan ligt (zie groep 1 en 3).

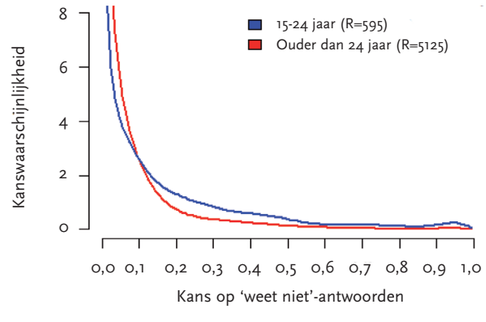
Het vergelijken van twee gedragsprofielen met $\hat{\delta}_p$ heeft veel voordelen. Zo maakt $\hat{\delta}_p$ geen assumptie over de vorm van de onderliggende verdelingen (Cliff, 1993, 1996ab; Goedhart, 2016; Vargha & Delaney, 2000) en is $\hat{\delta}_p$ robuust in het geval van outliers en scheve of andere niet-normale verdelingen (Goedhart, 2016). De effectmaat is makkelijk te berekenen en interpreteren, en gestandaardiseerd, wat betekent dat verschillende categorieën van effectgrootte kunnen worden onderscheiden (Goedhart, 2016). Voor onze $\hat{\delta}_p$ zijn relatief kleine of ongelijke groepsgroottes niet problematisch.

Twee voorbeelden

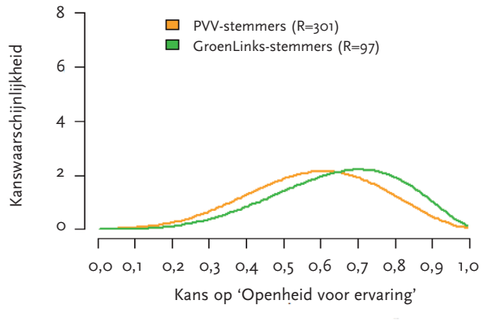
Nu laten we twee voorbeelden met surveydata zien die elk uit twee respondentprofielen bestaan waarvoor we $\hat{\delta}_p$ berekenen. We gebruiken data van het door CentERdata geadministreerde LISS Panel. We hanteren een bootstrapprocedure om 99% betrouwbaarheidsintervallen voor $\hat{\delta}_p$ te berekenen.

Zie grafiek 1 in figuur 2. We hebben respondenten van 15–24 jaar vergeleken met respondenten ouder dan 24 jaar op het geven van 'weet niet'-antwoorden voor items van de survey 'Politiek en Waarden' (wave 6, 2013). Vrijwel alle respondenten in beide groepen vulden 64 tot 67 items in waarbij een 'weet niet'-antwoord een optie was. De resulterende $\hat{\delta}_p$ van 0,28 kan worden gecategoriseerd als een 'medium' effect (Goedhart, 2016; Vargha & Delaney, 2000). We concluderen dat respondenten van 15–24 jaar meer 'weet niet'-antwoorden geven dan respondenten ouder dan 24 jaar op items over politieke inhoud. Zie Bais (2021) voor een eerste praktische implementatie van $\hat{\delta}_p$ voor het exploreren van de relatie tussen respondentkenmerken en antwoordgedrag.

Zie grafiek 2 in figuur 2 voor een andersoortig voorbeeld om de brede bruikbaarheid van $\hat{\delta}_p$ te illustreren. We onderzoeken de vraag of individuen die stemmen op GroenLinks mogelijk hoger scoren op de persoonlijkheidstrek 'openheid voor ervaring' dan individuen die stemmen op de PVV. Op basis van antwoorden op de vraag uit de survey 'Politiek en Waarden' op welke



1. Leeftijd: 'weet-niet'-antwoorden; Survey 'Politiek en Waarden'; Geschatte Cliffs Delta = 0,28; 99% Betrouwbaarheidsinterval = (0,22, 0,34)



2. Openheid voor ervaring voor PVV-stemmers vs. GroenLinks-stemmers; Geschatte Cliffs Delta = 0,21; 99% Betrouwbaarheidsinterval = (0,10, 0,32)

Figuur 2. Respondentprofielen en hun statistische eigenschappen voor leeftijdsgroepen voor het antwoorden van 'weet niet' voor de survey 'Politiek en Waarden' (grafiek 1) en voor PVV-stemmers versus GroenLinks-stemmers en hun score op 'Openheid voor Ervaring' (grafiek 2)

politieke partij men op 12 september 2012 heeft gestemd, hebben we een groep GroenLinks-stemmers en een groep PVV-stemmers geconstrueerd. Uit de survey 'Persoonlijkheid' (wave 6, 2013) beschouwen we de 10 items die 'openheid voor ervaring' meten op een 5-punts Likertschaal. We transformeren de respondentscores van 1 tot en met 5 op elk item in respectievelijk 0, 0,25, 0,50, 0,75 en 1. Daarna tellen we deze getransformeerde scores op en ronden het totaal af op een geheel getal voor elke respondent. Deze totalen worden gebruikt als G_r in het linkerdeel van formule (1) met I_r=10. De respondentprofielen voor beide politieke partijen worden vervolgens berekend met het rechterdeel van formule (1) en linkerdeel van formule (2). Vergelijking van de twee profielen geeft een $\hat{\delta}_p$ van 0,21, wat een 'klein' maar duidelijk aanwezig effect betekent. We concluderen dat individuen die op GroenLinks stemmen tot op zekere hoogte meer 'openheid voor ervaring' vertonen dan individuen die op de PVV stemmen.

Conclusie

We concluderen dat $\hat{\delta}_p$ een solide en conservatieve statistiek is die zowel nuttig als gunstig is om twee gedragsprofielen mee te vergelijken. In essentie is $\hat{\delta}_p$ een onderschatting van de 'ware' δ en convergeert $\hat{\delta}_p$ naar deze δ naarmate het aantal items toeneemt waarop profielen zijn gebaseerd. De motivering voor dit onderzoek was het gebruiken van een effectmaat voor het vergelijken van twee gedragsprofielen. Met zulke vergelijkingen kunnen relaties tussen respondent-/itemkenmerken en ongewenst antwoordgedrag worden ontdekt. Deze relaties kunnen richting geven aan het ontwerpen van surveys die ongewenst antwoordgedrag minimaliseren en datacollectie optimaliseren.

Dankwoord
Graag bedanken we CentERdata voor de beschikbaarheid van LISS Panel data en Vanessa Torres van Grinsven voor haar hulp bij publicatie van dit artikel.

Literatuur

Bais, F. (2021). *Constructing behaviour profiles for answer behaviour across surveys* (Dissertation). Utrecht University. https://doi.org/10.33540/538

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114*(3), 494-509. https://doi.org/10.1037/0033-2909.114.3.494

Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research, 31*(3), 331–350. https://doi.org/10.1207/s15327906mbr3103_4

Cliff, N. (1996b). *Ordinal methods for behavioral data analysis*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315806730

Goedhart, J. (2016). *Calculation of a distribution free estimate of effect size and confidence intervals using VBA/Excel*. BioRxiv. https://doi.org/10.1101/073999

Hess, M. R., & Kromrey, J. D. (2004). *Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's Delta under non-normality and heterogeneous variances* (Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California). American Educational Research Association.

Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience, 44*(9), 2647–2651. https://doi.org/10.1111/ejn.13400

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience, 46*(2), 1738-1748. https://doi.org/10.1111/ejn.13610

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL Common Language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics, 25*(2), 101–132. https://doi.org/10.2307/1165329

Frank Bais heeft in 2013 een research master Psychology aan de Universiteit van Amsterdam afgerond. De afgelopen jaren deed hij promotie-onderzoek in survey-methodologie aan de Universiteit Utrecht en bij het Centraal Bureau voor de Statistiek. Hij gebruikte gedragsprofielen en de aanpassing van Cliffs Delta om de relatie tussen respondentkenmerken en antwoordgedrag te onderzoeken. Op 31 maart jongstleden is hij gepromoveerd.
E-mail: frank_bais@hotmail.com.

Joost van der Neut heeft een BSc en MSc in Applied Geophysics van de Technische Universiteit Delft. Hij promoveerde in 2012 cum laude op seismische beeldvormingstechnieken bij TU Delft en ontving vervolgens een Veni beurs van NWO. Tevens heeft hij de laatste jaren onderzoek gedaan naar medische beeldvormingstechnieken met ultrageluid.
E-mail: jrvanderneut@yahoo.com.

Samenwerken voor duurzaamheid

Bijna tien jaar geleden schreef ik een blog (Poppelaars, 2012) over mijn bezoek aan het Wereldvoedselprogramma (WFP), de grootste humanitaire hulporganisatie gericht op het bestrijden van honger in de wereld. In samenwerking met het programma Moving the World van TNT Express had mijn toenmalige team al verscheidene projecten[2] (in WFP-termen missies) om de voedseldistributie in onder andere Liberia, Ethiopië en Mali te verbeteren met goed resultaat afgerond (Plat, van Dijk & Poppelaars, 2009). Ik was bij het WFP in Rome om te verkennen hoe we onze samenwerking verder konden uitbreiden.

WFP wint Edelman Award

Mijn bezoek viel samen met de viering van het 50-jarig bestaan van het WFP. Tijdens de algemene vergadering hoorde ik Josette Sheeran, directeur van het WFP, over het werk en de uitdagingen van het WFP spreken. De kille statistieken die ze in haar speech aanhaalde, gaven aan hoe ernstig de problemen op dat moment waren. Een op de zeven mensen op aarde lijdt chronisch honger, elke 10 seconden sterft er een kind aan de gevolgen van ondervoeding. Gruwelijke feiten, zeker als je beseft dat het onnodig leed is aangezien er voldoende grondstoffen en technologie beschikbaar zijn om iedereen op de wereld te voeden (FAO, z.d.).

Tien jaar later zijn de statistieken wat beter, één op de elf mensen op aarde lijdt chronisch honger. Kindersterfte is echter nog steeds een groot probleem. De gevolgen van klimaatverandering en sinds vorig jaar COVID maken het werk van het WFP niet eenvoudiger. Er is nog een lange weg te gaan om honger uit te bannen maar de bereikte verbetering biedt hoop. Het WFP heeft deze verbetering mede weten te bereiken door de innovatieve inzet van data, statistiek en operations research. Voor deze impactvolle inzet van operations research heeft het WFP in april van dit jaar de INFORMS Franz Edelman Award mogen ontvangen (Informs, 2021). Deze Award heeft ook

deels een Nederlands tintje door de bijdrage van het Zero Hunger Lab[5] aan de ontwikkeling van Optimus, de supply chain optimalisatie tool van het WFP (Tilburg University, z.d.). De Edelman award is een bevestiging dat, ondanks de bijzondere omstandigheden waarin en de snelheid waarmee het WFP moet werken, Statistiek en Operations Research waardevolle, impactvolle ondersteuning kunnen bieden.

Duurzame ontwikkeldoelen

Honger is niet het enige probleem in de wereld dat aandacht nodig heeft. In 2015 formuleerde de Verenigde Naties (VN) zeventien ontwikkelingsdoelstellingen[6] voor 2030 (United Nations, z.d.). Elk van die zeventien ontwikkeldoelen adresseert een probleem dat een mondiale aanpak nodig heeft. Met die doelen vraagt de VN aandacht voor onder meer de groeiende ongelijkheid, de rechten van vrouwen en meisjes, vrede, veiligheid en klimaatverandering. Het spreekt voor zich dat er een afhankelijkheid is tussen de ontwikkeldoelen. Armoede aanpakken kan niet zonder ook ongelijkheid aan te pakken en toe te werken naar een vreedzame en inclusieve samenleving. Er zijn ook beperkingen waar nadrukkelijk rekening mee moet worden gehouden. Immers, we kunnen geen duurzame economische groei of voedselzekerheid bereiken zonder rekening te houden met de impact van onze activiteiten op het milieu.

De ontwikkeldoelstellingen zijn *wicked* problemen. Ze zijn slecht gedefinieerd, hebben vele belanghebbenden en hebben sterke sociale, politieke en juridische componenten. Het is zelfs de vraag of je kunt vaststellen of en wanneer je het probleem hebt opgelost. Ackoff (1974) zou ze een *mess* noemen. Omdat ze zo veel facetten hebben kun je je afvragen of Statistici of Operations Researchers zinvolle ondersteuning kunnen bieden bij het aanpakken van deze problemen. Hoe modelleer je politieke geschillen? Zijn mensen wel zo rationeel als we in onze modellen veronderstellen? Hoe om te gaan met de fundamentele onzekerheid in de wereld? Het voorbeeld van de bestrijding van honger door het WFP bewijst dat veel mogelijk is, echter de vele facetten van een ontwikkelingsdoelstelling maken dat een enkele discipline nooit voldoende kan zijn.

Interdisciplinaire aanpak

De missies die mijn team voor het WFP heeft uitgevoerd leerden me dat modellen en data een te eenzijdige benadering zijn om de vraagstukken van het WFP aan te pakken. Veel van de uitdagingen waar het WFP voor staat zijn niet te vangen in data of modellen en vragen om andere disciplines. Dit geldt ook voor de duurzame ontwikkeldoelen. De sleutel ligt in een interdisciplinaire aanpak en samenwerking. Zo gebruikt het WFP inzichten uit de klimaatwetenschap om te bepalen waar voorraden van noodhulpgoederen moeten worden aangehouden om zo veel mogelijke passende hulp te kunnen bieden aan hen die dat nodig hebben. Ook als door extreem weer de aanvoerwegen onbegaanbaar zijn geworden of als extra noodhulp nodig is door aanhoudende droogte.

Voor de aanpak van de duurzame ontwikkeldoelen is een combinatie van disciplines nodig. Die combinatie zal leiden tot innovatieve oplossingen die het realiseren van de ontwikkeldoelen een stukje dichter bij zal brengen. Als Statistici en Operations Researchers moeten we de samenwerking zoeken met klimatologen, ecologen, politicologen, economen en vele andere disciplines om een impactvolle bijdrage te kunnen leveren aan het realiseren van de VN ontwikkelingsdoelstellingen.

Literatuur

Ackhoff, R. L. (1974). *Redesigning the future: a systems approach to societal problems*. Wiley.

FAO. (z.d.). *Sustainable development goals*. http://www.fao.org/sustainable-development-goals/goals/goal-2/en/.

Informs. (2021). *Edelman winner: UN World Food Programme*. https://www.informs.org/Resource-Center/Video-Library/Edelman-Competition-Videos/2021-Edelman-Competition-Videos/2021-Edelman-Winner-UN-World-Food-Programme.

Plat, T., Dijk, T. van, & Poppelaars, J. (2009). De wereld geoptimaliseerd! Waarom Operations Research essentieel is bij voedselhulp. *STAtOR, 10*(4). 4–8.

Poppelaars, J. (2012). *A billion in need*. https://john-poppelaars.blogspot.com/2012/02/billion-in-need.html.

Tilburg University. (z.d.). *Zero Hunger Lab*. https://www.tilburguniversity.edu/nl/onderzoek/impact/creating-value-data/zero-hunger-lab.

United Nations. (z.d.). *Sustainable Development Goals*. https://sdgs.un.org/goals.

John Poppelaars, Doing the Math
E-mail: john@doingthemath.nl

ROOSTEREN VANUIT ENDOSCOPISCH PERSPECTIEF

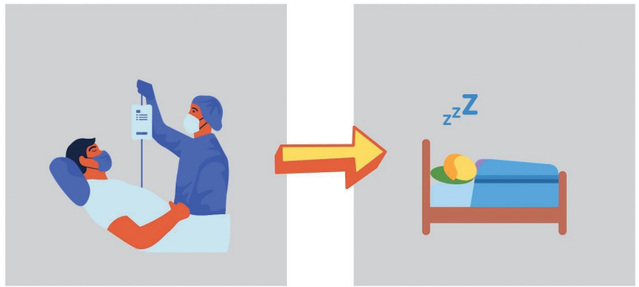
Dit artikel belicht een heuristiek en een MIP-model voor het roosteren van endoscopieën. Op de endoscopieafdeling van het Maag Darm Levercentrum van het LUMC vinden dergelijke kijkoperaties plaats, met een aansluitend herstel van 90 minuten op een uitslaapkamer. Een optimaal rooster maximaliseert het aantal endoscopieën per dag en houdt tevens het aantal herstellende patiënten gedurende de dag zo constant mogelijk.

Kim van den Houten, Marelot de Vos, Erik van der Sluis, Thomas Schneider & Nico van Dijk

Het uitvoeren van scopieën, oftewel kijkoperaties, is een hoogwaardige vorm van moderne operatietechnieken. Een dankbaar toepassingsgebied betreft maag-, darm- en leveroperaties (MDL). Maar hoeveel patiënten kunnen we aan en hoe plannen we operaties met verschillende duren op dagbasis in?

Diverse beperkingen spelen een rol. Een beperkt aantal operatiekamers, een beperkte uitslaapkamer, een verplichte uitslaapduur, beperkte werktijden, een gegeven mix van operaties en vanzelfsprekend: personeel dat enigszins op tijd naar huis wil zonder veel uitloop.

Het inplannen van deze scopieën is ogenschijnlijk simpel, maar dat blijkt allesbehalve zo te zijn. Noch in literatuur noch in ziekenhuispraktijk is een algemene



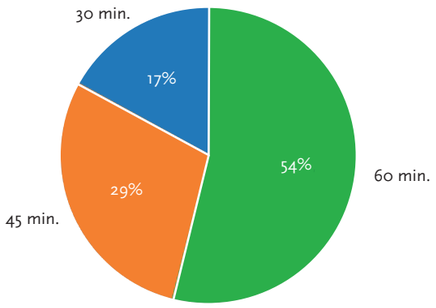
Van een endoscopiekamer naar de uitslaapkamer

oplossingsmethode voorhanden. Er wordt enkel gewerkt met vuistregels. Hoe goed zijn deze? OR biedt mogelijkheden tot verbeteringen.

Op endoscopisch bezoek

Wij bezochten het Maag, Darm en Levercentrum van het Leids Universitair Medisch Centrum (LUMC). Het LUMC is de trotse eigenaar van vier endoscopiekamers. In deze kamers worden met veel deskundigheid endoscopieën uitgevoerd. Hoe verloopt zo'n endoscopie? Tijdens een afspraak wordt de patiënt op een bed naar een endoscopiekamer gebracht, waar het onderzoek plaatsvindt. Na afloop van het onderzoek geldt een verplichte hersteltijd van 90 minuten. Hiervoor is een gemeenschappelijke 'uitslaapkamer' ingericht, met een capaciteit voor 8 patiënten. Tussen de scopie en de rustperiode mag geen wachttijd zitten, een patiënt kan immers niet zomaar op de gang worden geparkeerd.

De endoscopieën in het LUMC zijn globaal in 3 groepen te verdelen op basis van geplande duur (zie figuur 1). In het LUMC werden deze patiënten per kwartier nog op volgorde van binnenkomst ingeroosterd. De planafdeling



Figuur 1. Verhouding patiënten per endoscopieduur

is verantwoordelijk voor het maken van een rooster. Het LUMC spreekt twee wensen uit:

1. zoveel mogelijke endoscopieën en
2. een zo constant mogelijk aantal patiënten op de uitslaapkamer.

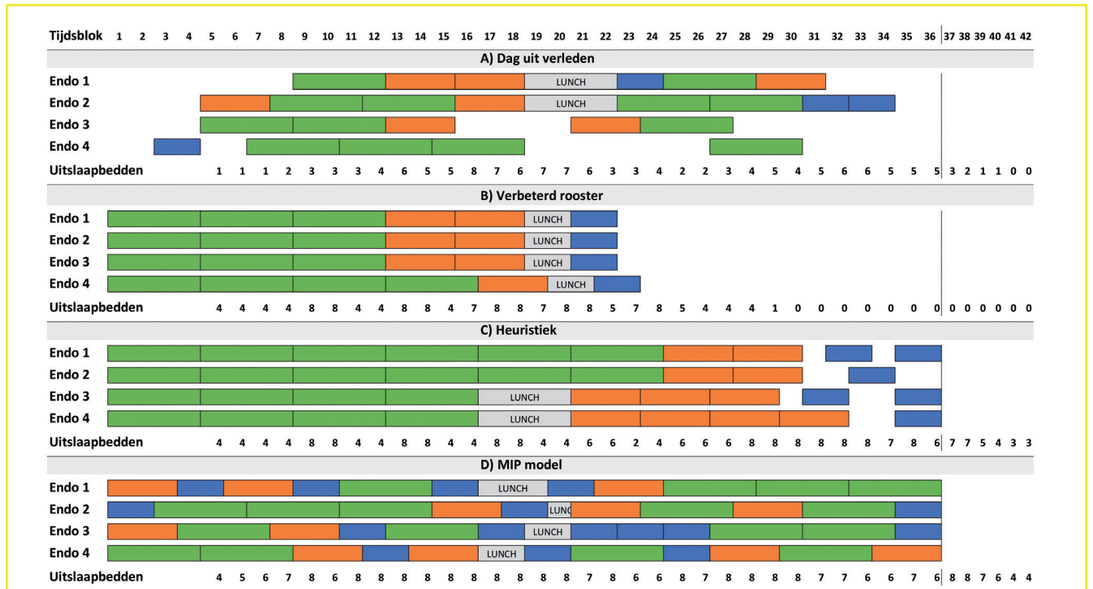
De eerste spreekt voor zich, de tweede beoogt piekdrukte en leegstand te vermijden.

Een dag in het rooster van het LUMC

Om de huidige situatie in beeld te brengen groeven wij in het roostergeheugen van het LUMC. Figuur 2A visualiseert een representatieve dag. De werkdag van 08:00-17:00 is opgedeeld in 36 tijdsblokken van ieder een kwartier. Uiteraard moet het personeel kunnen lunchen. Hiervoor moeten in totaal 8 tijdsblokken worden gereserveerd. De verticale as representeert de 4 endoscopiekamers (1 rij per kamer) en de bezetting van de uitslaapkamer. De gekleurde blokjes visualiseren rasters waarbinnen 1 patiënt is ingepland. De kleuren variëren met de duur van de behandeling (30, 45 of 60 minuten). Het rooster volgt de verhouding van figuur 1. De teller in de onderste rij houdt het aantal 'uitslapers' bij. Het mogen er dus nooit meer dan 8 zijn. Met het huidige rooster werd er een benutting van 60% behaald. Het aantal aanwezigen in de uitslaapkamer fluctueerde behoorlijk. Rond 12 uur een piek van 8 uitslapers, terwijl dit aantal 2 uur later gedaald was tot 2. Voor de inzet van verpleegkundigen was dit allesbehalve ideaal. Ons doel was een rooster te creëren dat beter met de beschikbare capaciteiten omgaat.

Wat kan een vuistregel bieden?

Een simpele heuristiek om het aantal uitslapers constanter te houden en de endoscopiekamers beter te benutten is 'van lang naar kort'. Hiermee konden we eerder op de



Figuur 2. Dagroosters
A. Willekeurige dag uit verleden LUMC.
B. Verbetering op rooster A: dezelfde afspraken, maar aangepast rooster volgens rooster regel 'lang naar kort'.
C. Heuristiek 'lang naar kort' met een verhoudingsgewijze maximalisatie van het aantal geplande patiënten.
D. MIP-model met maximalisatie van het aantal geplande patiënten als doelfunctie.

dag klaar zijn met zodoende over een kortere tijdsperiode een hogere benutting, zoals zichtbaar in figuur 2B om 14:00 uur, zonder de maximale capaciteit te overschrijden! Dit schiep ruimte om de endoscopiekamers beter te benutten, meer patiënten inplannen dus! Werkend 'van lang naar kort' kon vervolgens het aantal ingeplande scopieën worden verhoogd, in figuur 2C gevisualiseerd. De heuristiek leverde een benutting van 92% (zie ook figuur 3) op. Een mooi resultaat, maar er is nog ruimte voor extra patiënten.

De kunst van het roosteren

Roosteren is iets waar de OR-loep al vaak boven heeft gehangen. Een bekend besliskundig model is het probleem van Flexible Flow Shop (FFS). Een basismodel is genomen op basis van FF2 | k_1, k_2, nowait | C_{max} (zie Blazewicz et al., 2007). In een FFS staan er meerdere groepen van dezelfde machines in serie achter elkaar, waar vervolgens taken voor worden gepland. Dit model is direct toepasbaar op de endoscopieplanning van het LUMC. De endoscopiekamers en de uitslaapkamers zijn vertaalbaar naar deze twee groepen 'machines'. De patiënten zijn te vertalen als 'taken', die gepland moeten worden. Na het endoscopisch onderzoek moet een patiënt direct kunnen worden overgebracht naar de uitslaapkamer. Dit is in de literatuur bekend als een *nowait* voorwaarde. Het toevoegen van deze voorwaarde aan het model zorgt ervoor dat er altijd voldoende ruimte is op de uitslaapkamer. De FFS kan een bepaalde doelfunctie optimaliseren gegeven het aantal taken en gegeven de beschikbare capaciteit. De FFS diende voor ons als geschikt uitgangspunt voor het uiteindelijke LUMC-roostermodel. Uniek aan dit LUMC-roostermodel is de doelfunctie die het aantal operaties maximaliseert, terwijl in de klassieke FFS wordt uitgegaan van een vast aantal taken. Hierdoor is dus niet vooraf vastgelegd hoeveel patiënten per type (op basis van geplande duur) er zullen worden ingepland. Omdat we te maken hebben met een vaste verhouding van type patiënten (zie figuur 1), moet ook voor de uiteindelijke verhoudingen in het rooster restricties kunnen worden opgegeven. Het uiteindelijke LUMC-roostermodel is te zien in Model 1.

BASISMODEL

FF2 | *k*1, *k*2, nowait | *Cmax*

PATIËNT MODEL

max	totaal aantal operaties	(1)
s.t.	elke patiënt krijgt een eindtijd op de endoscopiekamer	(2)
	elke patiënt krijgt een eindtijd op de uitslaapkamer (90 min na scopie)	(3)
	duur onderzoek ≤ eindtijd onderzoek ≤ eindtijd laatste onderzoek van dag	(4)
	aantal patiënten tegelijk in endoscopiekamers ≤ 4	(5)
	0 ≤ aantal patiënten tegelijk in uitslaapkamer ≤ 8	(6)
	aantal geplande operaties per type ≥ minimale aantal type	(7)
	percentage aantal operaties per type ≥ minimaal percentage type	(8)

Model 1. Roostermodel LUMC

Plannen maar

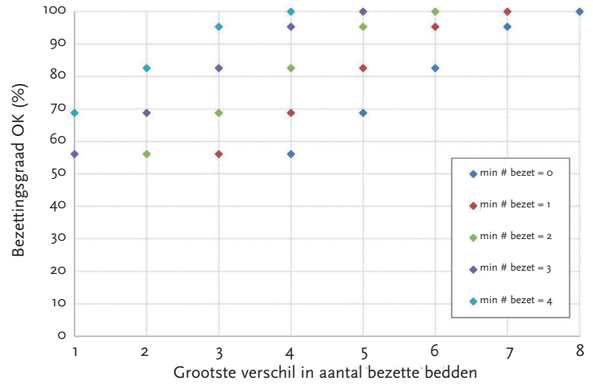
Omdat het hoofddoel van het LUMC-roostermodel het inplannen van zoveel mogelijk patiënten is, tonen wij een oplossing waarin het aantal uitslapers tussen 0 en 8 mag liggen. Hieronder volgt een korte toelichting van de mogelijkheden dat het begrenzen van het aantal uitslapers biedt. Met behulp van een solver (AIMMS) bleek het model makkelijk oplosbaar. Wij hebben de solver gebruikt voor het maken van een weekrooster en hiervan 1 dag gevisualiseerd in figuur 2D. Hierdoor kan de verhouding tussen de verschillende soorten endoscopiesoorten in rooster D afwijken van roosters A, B en C. Wat leverde het ons op? De gemiddelde bezettingsgraad van de endoscopiekamers kwam uit op 100% (lunchpauzes niet meegerekend). Dat is een mooie score! Zeker in vergelijking met de gemeten benutting volgens de andere methoden. Een samenvattingstabel is gegeven in figuur 3.

Model kan meer

Maar hoe zit het met de tweede doelstelling: drukte op de uitslaapkamer als we de verschillende roostermethodieken met elkaar vergelijken? Op de momenten dat

Roostermethode	A	B	C	D
Aantal patiënten	24	24	37	45
# rasters 30 min	4 (17%)	4 (17%)	6 (16%)	16 (36%)
# rasters 45 min	7 (29%)	7 (29%)	11 (30%)	12 (27%)
# rasters 60 min	13 (54%)	13 (54%)	20 (54%)	17 (38%)
Benutting	60%	60%	92%	100%
Laagst gemeten aantal bedden	1	0	2	4
Hoogst gemeten aantal bedden	8	8	8	8

Figuur 3. Behaalde KPI-waarden per roostermethode



Figuur 4. Uitwisseling benutting en verschil

er 8 uitslapers zijn, is er maximale verpleegkundige inzet nodig op de afdeling. We zagen bij elk rooster dat het hoogst aantal bezet gemeten bedden 8 is; zelfs bij roosters A en B, ondanks de lage benutting van 60%. Het laagst gemeten aantal bezette bedden was gedurende de werkdag (gemeten vanaf tijdsblok 5, na de opwarmperiode) lager bij roosters A, B en C vergeleken met rooster D. Omdat de heuristiek en het wiskundig model als hoofddoel het maximaliseren van het aantal geplande endoscopieën hanteert, is het de vraag of er nog verbeterruimte zit in het aantal uitslapers op de verpleegkamer. Voorwaarde (6) uit Model 1 biedt de mogelijkheid om het aantal toegestane uitslaapbedden in te stellen. Bij het roosteren kunnen dus striktere onder- en bovengrenzen voor de uitslaapkamer worden opgegeven, waardoor piekdrukte en leegstand in de uitslaapkamer worden voorkomen. Dit kan ook per dag worden ingesteld. Deze mogelijke uitwisseling tussen bezetting en fluctuaties in bedden is, zonder in details te treden, gevisualiseerd in figuur 4.

Slotwoord

Ondanks dat ziekenhuizen hun eigen vaste manieren gebruiken om patiënten in te plannen, blijkt het nuttig een OR-bril op te zetten. Wiskundig modelleren maakt het mogelijk aan verschillende wensen van een ziekenhuis te voldoen. Met het voorgestelde MIP-model kan het aantal patiënten per werkdag significant worden verhoogd, terwijl te allen tijde de beschikbare capaciteiten in acht worden genomen. Indien gewenst, kan ook het aantal patiënten in de uitslaapkamer begrensd worden. Door deze tactische manier van plannen, kan het LUMC piekdrukte en leegstand voorkomen. Vooralsnog is een brug voor verdere implementatie benodigd, waarvoor dit artikel een mooi startpunt biedt.

Literatuur

Blazewicz, J., Ecker, K., Pesch, E., Schmidt, G., & Weglarz, J. (2007). *Handbook of Scheduling*. Springer-Verlag.

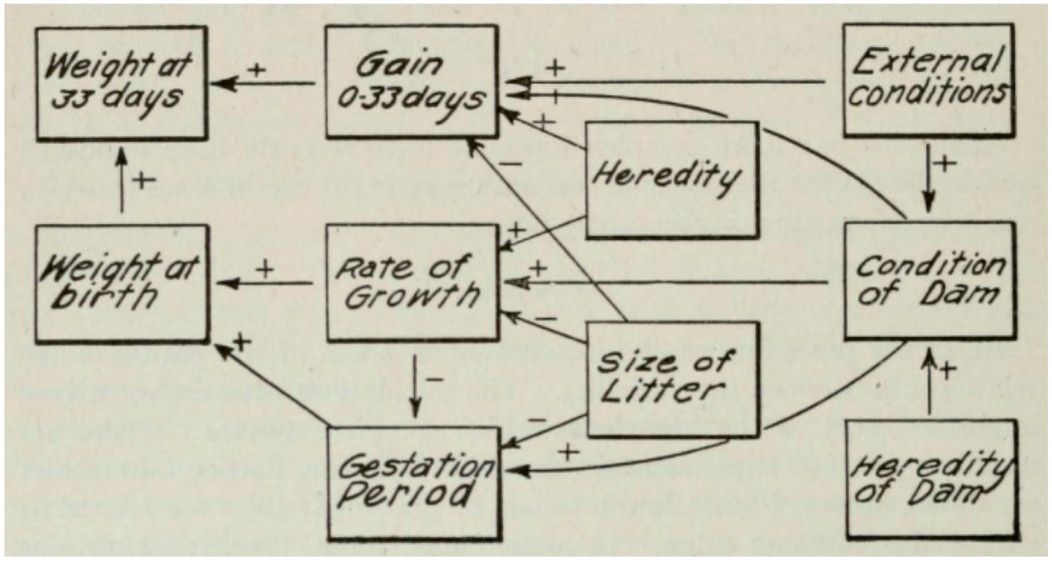
Kim van den Houten is masterstudente Operations Research aan de Vrije Universiteit van Amsterdam en onderwijsassistente voor de bachelor Artificial Intelligence. Tijdens haar scriptie is haar interesse gewekt voor toepassing van OR binnen gezondheidszorg, waarna zij 1 jaar werkte als student-consultant bij Rhythm B.V. (capaciteitsmanagement in de zorg).
E-mail: kimvandenhouten97@live.nl.

Marelot de Vos is masterstudente Operations Research and Quantitative Logistics aan de Erasmus Universiteit Rotterdam. Zij is aan het afstuderen op een Vehicle Scheduling probleem voor elektrische bussen.
E-mail: marelotdevos@hotmail.com.

Erik van der Sluis is hoofddocent Operations Research aan de Universiteit van Amsterdam.
E-mail: h.j.vandersluis@uva.nl.

Thomas Schneider is gepromoveerd aan de Universiteit Twente op integraal capaciteitsmanagement en -planning in ziekenhuizen. Hij heeft ruime ervaring met capaciteitsmanagement en is manager van de afdeling integraal capaciteitsmanagement in het OLVG. Zijn onderzoeksinteresse liggen in de vertaling van praktische problemen naar onderzoek en onderzoeksresultaten naar praktische implementatie.

Nico van Dijk is hoogleraar Stochastische Operations Research aan de Universiteit Twente en de Universiteit van Amsterdam. Tevens is hij verbonden aan het Center for Healthcare Operations and Improvement Research (CHOIR) aan de Universiteit Twente.
E-mail: N.M.vanDijk@uva.nl.



Het padmodel laat zien welke factoren er van invloed zijn op het geboortegewicht van cavia's en de gewichtstoename tijdens de speentijd van 33 dagen. Overgenomen uit 'Correlation and causation' door Sewall Wright, 1921, *Journal of Agricultural Research*

Over padmodellen, structurele vergelijkingen en de roep om Explainable AI

Richard Starmans

De Amerikaanse geneticus en statisticus Sewall Green Wright (1898–1988) speelde begin vorige eeuw samen met Ronald A. Fisher en J. B. S. Haldane een belangrijke rol bij de totstandkoming van de Moderne Synthese in de biologie, ruwweg de integratie van de evolutietheorie van Wallace en Darwin en de mendeliaanse genetica. Juist in deze periode was de kruisbestuiving tussen de nog prille wetenschap der statistiek enerzijds en 'variatie-en-verandering' rijke disciplines als biologie, landbouwwetenschap, (populatie)genetica en evolutietheorie anderzijds enorm. Sommige pioniers slaagden erin op beide terreinen grensverleggend onderzoek te verrichten. Dat gold allereerst voor Fisher zelf, maar Sewall Wright deed in menig opzicht niet voor hem onder. Het is dit jaar precies 100 jaar geleden dat Wright in de *Journal of Agricultural Research* zijn befaamde artikel *Correlation and Causation* (1921) publiceerde, waarin hij zijn 'method of path coëfficiënts' introduceerde, de weg effende voor wat later als padanalyse en structurele vergelijkingsmodellen bekend zou worden en tevens de basis legde voor hedendaagse (grafisch georiënteerde) benaderingen van *causal inference* binnen AI en informatica. De auteur legt in de openingszin van het artikel direct zijn kaarten op tafel: 'The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated.' Helaas blijken in de weerbarstige praktijk deze oorzaken van variatie dikwijls *beyond control* en met name biologen worstelen met metingen c.q. variabelen die gecorreleerd zijn 'because of a complex of interacting, uncontrollable and often obscure causes', aldus de auteur, zonder daarbij overigens direct naar observationele data te verwijzen. De correlatiecoëfficiënten van Francis Galton en Karl Pearson vormen bij dit alles slechts een soort brutoresultaat van 'all connecting paths of influence'. Met zijn artikel beoogt Wright een methode te demonstreren, die elke directe invloed 'along each separate path' specificeert en meet, om zo te kunnen bepalen hoe de totale variatie in de uiteindelijke afhankelijke variabele 'is determined by each particular cause'. Dat ontrafelen van de gemeten correlaties leidt dan tot (stelsels van) vergelijkingen waarbij directe en indirecte effecten, alsmede endogene en exogene variabelen een rol spelen.

De wijze waarop Wright beklemtoont hoezeer in de genetica correlatiecoëfficiënten en 'platte' regressiemodellen tekortschieten is zonder meer saillant, maar de auteur gaat nog een stap verder door allereerst het woord 'causation' nadrukkelijk in de titel op te nemen en aansluitend in de eerste twee alinea's van het artikel maar liefst zesmaal naar de woorden 'cause' en 'causal' te verwijzen. In de rest van de dertig pagina's tellende publicatie wordt niet alleen het empirisch onderzoek naar geboortegewicht en draagtijd van cavia's en de wateropname en transpiratie van planten causaal geduid en geïnterpreteerd, maar ook de statistische analyse van dit alles. Deze historisch en filosofisch belangwekkende aanpak van Wright leidde al spoedig tot opmerkelijke reacties en een grillige receptie van zijn werk, die decennia voortduurde en – belangrijker nog – waarvan de implicaties tot op de dag van vandaag zichtbaar zijn. Enige aspecten daarvan brengen we hier kort voor het voetlicht.

Statistiek en causaliteit

De *causale mark-up* van Wright is om diverse redenen frappant. Allereerst kan met enig gevoel voor pathos worden gesteld dat de tijdgeest en het intellectuele klimaat in het eerste kwart van de twintigste eeuw sterk anti-causalistisch waren. Causaliteit werd geassocieerd met een achterhaalde wetenschapsopvatting, met metafysica in de wijsbegeerte en vooral met een deterministisch wereldbeeld, dat sedert het laatste kwart van de negentiende eeuw op zijn retour was. In de (wetenschaps-)filosofie kreeg dit onder meer gestalte in het strenge empirisme van Ernst Mach in Duitsland, het utilitarisme van John Stuart Mill in Engeland en het positivisme van Auguste Comte in Frankrijk. Allen waren op hun beurt schatplichtig aan David Hume, die in de achttiende eeuw het empirisme in Engeland voortstuwde, zich kritisch uitliet over de status van oorzaak-gevolg relaties en ook vandaag de dag nog door velen als een scharnier- en ijkpunt in het denken over causaliteit wordt beschouwd. Die kritiek op causaliteit werd in het begin van de twintigste eeuw nog krachtiger verwoord door Bertrand Russell in zijn beroemde *On the notion of cause* uit 1905 en zou in de jaren na Wrights publicatie culmineren in het logisch-positivisme van de Wiener Kreis, dat de zintuigelijke waarneming als de enige kenbron van de waarheid beschouwde, metafysica als betekenisloos bestempelde en causaliteit in het gunstigste geval obsoleet achtte, een overbodige kantiaanse categorie uit de hoogtijdagen van het achttiende-eeuwse determinisme. Die houding typeerde mutatis mutandis ook de net opgekomen statistiek, die een stevige opmars maakte in de wetenschappen en daar een probabilistische wending in gang had gezet (Kruger, 1987, 1990). Statistici als Galton en Pearson waren biometrici van het eerste uur en met name Pearson rekende in zijn filosofische *The Grammar of Science* hardhandig af met causaliteit en determinisme (Pearson, 1892).

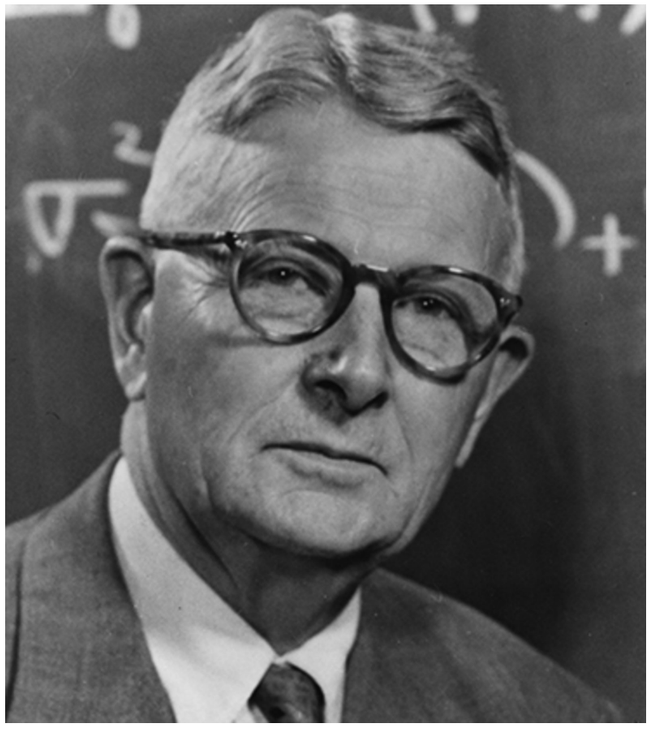
Uiteraard bleef de biometrische aanpak niet onweersproken. Met name genetici hekelden een reductie van biologie tot statistiek en meenden dat hiermee de aandacht voor (fysiologische) werkingsmechanismen en (deel)processen in wetenschappelijke beschrijvingen en verklaringen naar de achtergrond werd gedrongen. Om recht te doen aan deze mechanismen, zochten de meeste genetici evenwel toch vooral hun heil in het streven

naar een volwaardige experimentele methodologie in plaats van terug te grijpen op een archaïsche, verdachte en methodologisch nauwelijks ontwikkelde notie van causaliteit. Die experimentele methodologie, door Wright zelf beschouwd als de 'ideale methode', was op dat moment nog volop in ontwikkeling en zou onder meer leiden tot variantieanalyse en *experimental design*. Ook hier had Fisher het voortouw genomen en hij zou de eerste resultaten optekenen in zijn populaire, op wiskundig ongeschoolde empirische onderzoekers gerichte *Statistical Methods for research workers* uit 1925 en later vooral in *The Design of Experiments* uit 1935. Ook Wright was in de eerste plaats empirisch onderzoeker en het publiek dat hij moest overtuigen was veeleer geïnteresseerd in realistische toepassingen van een constructieve methodologie en uiteraard minder in *armchair philosophy* of epistemische beschouwingen over oorzakelijkheid. Dat causaliteit het domein van de speculatieve filosofie verre oversteeg en bovendien allerminst noodzakelijkerwijs met determinisme verbonden was, zou voor velen pas veel later duidelijk worden, toen probabilistische benaderingen van oorzaak-gevolg relaties wel degelijk mogelijk bleken. Weliswaar publiceerde de wiskundige Hans Reichenbach (1891–1953) – paradoxaal genoeg zelf logisch-positivist van het eerste uur – reeds in 1923 een belangrijk artikel met de omineuze titel *The principle of causality and the possibility of its empirical confirmation*, maar hij gaat daarin toch vooral in op de causale duiding van natuurwetten, die in 'symmetrische' wiskundige formules zijn beschreven. Pas in zijn beroemde, postuum gepubliceerde *The Direction of Time* uit 1956 wordt de probabilistische, 'asymmetrische' benadering van causaliteit volledig uitgewerkt. In diezelfde periode verscheen *A theory of causality* (1959) en *A causal calculus I and II* (1961) van de hand van de statisticus Irving John Good (1916–2009), die een fysische conceptie van toeval hanteerde, gebaseerd op Poppers propensity-benadering van het kansbegrip. De belangrijkste pionier vanuit filosofisch perspectief is wellicht Patrick Suppes (1922–2014) wiens *A probabilistic theory of causality* uit 1970 een moderne klassieker werd. We zijn dan wel ongeveer een halve eeuw verwijderd van Wrights vroege werk en het feit dat vandaag de dag nagenoeg alle formele benaderingen van causaliteit probabilistisch zijn, is in het licht van het voorgaande wellicht ironisch, maar toch zeker veelzeggend.

Duidelijk is dat Wright een statement wilde maken en wellicht heeft hij de uiteenlopende reacties ook onderschat (Denis, 2006). Zo ontstond onmiddellijk een forse en onaangename polemiek met de bioloog Henry E. Niles, die stelde dat de aanpak om a priori een causaal schema of paddiagram te tekenen geen filosofische basis had, de methode van padcoëfficiënten daarmee *faulty* was en toepassing ervan enkel kon leiden tot resultaten die *wholly unreliable* waren (Niles, 1922, 1923), (Wright, 1923). Tegelijkertijd werden Wrights ideeën al spoedig opgemerkt buiten de lifesciences, onder meer door de enigszins vergeten ontwikkelingspsycholoog Barbara Burks (1902–1943), die vooral onderzoek deed in het kader van het klassieke nurture-versus-nature debat. In haar statistische werk bouwde zij voort op Wright en toonde onder meer aan dat de partiële correlatie en de determinatie-coefficiënt tekortschoten bij het analyseren van causale verbanden (Burks, 1926). Het is geen toeval dat Judea Pearl (1931), de meest uitgesproken hedendaagse protagonist van causal inference nog vrij recentelijk in zijn *The Book of Why; the new science of cause and effect* (2018) naast Sewall Wright vooral ook Barbara Burks roemt als één van de pioniers van wat hij zelf beschouwt als de Causale Revolutie. Voor Sewall Wright was een dergelijke rehabilitatie niet nodig, al zou het lang duren voordat zijn methode echt werd onderkend en toegepast. Met name de vraag of de padmodellen causale uitspraken toelieten zorgde voor vele debatten en misverstanden, waarvan Pearl er een aantal in zijn boek de revue laat passeren. Hoe dan ook, de problematische receptie van Wrights werk past in de moeizame samenspraak tussen statistiek en causaliteit in de afgelopen honderd jaar (Starmans, 2018).

Sociale feiten en latente variabelen

Toch zouden de ontwikkeling en toepassing van padmodellen na de Tweede Wereldoorlog in een stroomversnelling geraken buiten de lifesciences en – opmerkelijk genoeg – vooral in de economie en sociale wetenschappen, dikwijls in het licht van observationele studies en de mogelijkheden en onmogelijkheden daarin causale relaties te postuleren of aan te tonen. Dat begon in de vroege jaren vijftig met het werk van onder meer de economen Herman Wold (1908–1992) en Herbert Simon (1916–2001). Wold publiceerde in 1954 zijn paper *Causality and Econometrics*, waarin hij een bijkans wijsgerige fundering van de problematiek schetst en enige jaren later onder meer *Econometrics as Pioneering in Nonexperimental Model Building* (1969). Daarbij bouwde hij voort op de padmodellen van Wright, die hij en anderen typeerden als structural equations models en systems of simultaneous equations, waaraan de econoom Trygve Magnus Haavelmo al had gewerkt. Herb Simon publiceerde eveneens in 1954 zijn beroemde *Spurious Correlation; a causal interpretation* en ontwikkelde, voortbouwend op Wold een procedure waarbij telkens wanneer een pijl uit het paddiagram wordt weggelaten een testvergelijking ontstaat, die uitdrukt dat een partiële correlatiecoëfficiënt gelijk is aan nul. Ofschoon Simon pionierswerk verrichtte in economie, informatica, AI en cognitive science (Simon, 1957) was het vooral de methodoloog Herbert Blalock die in de vroege jaren zestig in het kielzog van Simon een en ander toepaste binnen de sociale wetenschappen, waar eveneens de sluimerende causaliteitsproblematiek inzake observationele data speelde. Zijn *Causal inferences in nonexperimental research* uit 1961 geeft een diepgaande analyse van de problematiek en de zogenaamde Simon-Blalock methode vormde de basis voor vele uitbreidingen, waaronder de populaire Baron-Kelly methode voor mediation uit de jaren tachtig, volgens Pearl vaker geciteerd dan Albert Einstein en Sigmund Freud. Hoe dan ook, padanalyse duikt op in nagenoeg elke inleiding tot de methodologie van de sociale wetenschappen, wanneer de invloed van een derde variabele wordt onderzocht als eerste stap bij de bespreking van confounding (*spurious correlation*), van effectmodificatie c.q. statistische interactie en uiteraard van mediation analysis, waarbij directe en indirecte effecten in het belang van het zoeken naar een (causaal mechanisme) worden beschreven.



Sewall Wright

Een volgende impuls voor de ontwikkeling van padmodellen vond plaats in de jaren zeventig en komt eveneens voort uit de sociale wetenschappen, waar dikwijls wordt geredeneerd met niet-gemeten grootheden, theoretische (multidimensionale) constructen en latente variabelen, die met behulp van multiple indicatoren worden gerepresenteerd. Dit ging reeds terug tot de negentiende-eeuwse sociologie, toen Emile Durkheim van de sociologie een exacte wetenschap wilde maken, gebaseerd op zogenaamde 'sociale feiten', die niet reduceerbaar zijn tot individuele entiteiten, hun eigenschappen, attitudes, preferenties of handelingen. Het ging dus om generalisaties en abstracties van die entiteiten, die bovendien doorgaans zijn ingebed in groepen, hetgeen onder meer leidde tot multilevel benaderingen. Nu waren uiteraard Pearsons Principale Componenten Analyse (1905), Spearmans factoranalyse (1904) en Hotelings canonische correlatie (1936) als modellen voor analyse en constructie van latente variabelen al genoegzaam bekend, maar nog nauwelijks geïntegreerd met padmodellen. Vooral met het werk van Karl Gustav Jöreskog zouden de padmodellen worden uitgebreid tot echte SEM-modellen, dankzij zijn Linear Structural Relations Sytem, (Jöreskog, 1973) kortweg LISREL, dat ook de naam zou vormen van het meest gebruikte statistische pakket voor SEM-analyse, later gevolgd door AMOS en EQS. Dat leidde tot technieken als *confirmatory factor analysis* en *partial least squares path modeling*, waarbij allerlei beperkingen van de klassieke padmodellen werden ondervangen; latente variabelen, cyclische, niet-recursieve verbanden en correlaties tussen *errorterms* konden nu eenvoudig worden gemodelleerd en maakten SEM na de introductie van grafische interfaces toegankelijk voor alle *researchworkers*, tot wie Fisher zich indertijd ook had gericht. Naast Blalock waren het onder meer Lazersfeld, Coleman en vele anderen die hiermee furore maakten, niet in de laatste plaats de Franse socioloog Raymond Bourdon, die mede vanuit deze padmodellen zijn dependency analyse ontwikkelde. Het vormde een opmaat tot verschillende, veelal grafische methoden

voor de conceptuele analyse van complexe, abstracte begrippen en processen, waarbij afhankelijkheden, interacties en andere relaties tussen variabelen worden ontrafeld (sociale actortheorie, netwerkanalyse, conceptuele grafen en semantische netwerken in informatica en AI, etc). Het *Verstehen* en *Erklären* daarvan – om met Wilhelm Dilthey te spreken – kan dan verlopen volgens een triptiek van allereerst *transparantie*, vervolgens *interpreteerbaarheid* en tot slot *verklaring*, waarmee een belangrijke hedendaagse toepassing van Wrights oorspronkelijke ideeën in het vizier komt in tijden van data science en AI.

Explainable AI en de menselijke maat

Het streven van AI een 'verklarende' wetenschap in de ruimste zin van het woord te maken was al aanwezig bij de ontwikkeling van traditionele regel-gebaseerde expertsystemen in de jaren tachtig (Buchanan, 1984), maar ook in de jaren negentig bij integratieve statistische benaderingen 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 (Cohen, 1995). Hedendaagse autonome systemen moeten daarbij ook meta-redeneren, kunnen verklaren, beargumenteren en rechtvaardigen (en dus in ruimere zin communiceren) waarom een bepaalde conclusie is getrokken, een bepaalde beslissing is genomen, een bepaalde keuze is gemaakt of een bepaalde actie is uitgevoerd. Idealiter verloopt dit eveneens volgens het genoemde drieluik van eerst transparantie, dan interpreteerbaarheid en vervolgens Explainable AI, die uiteindelijk moet leiden tot *Trust*, *Fairness* en *Responsible AI*, de zorgen in de samenleving wegneemt en dus een essentiele voorwaarde vormt voor het welslagen van het project van de AI. In 2019 constateerde de Australische informaticus Tim Miller in zijn artikel *Explanation in artificial intelligence; Insights from the social sciences* dat de focus op verklaringen in de AI weliswaar een heropleving beleeft, maar een louter technologische invalshoek kent (Miller, 2019) waarbij de omvangrijke sociaalwetenschappelijke literatuur grotendeels wordt genegeerd. Wie het gedrag van artificiële *agents* wil harmoniseren met het gedrag van menselijke *agents*, doet er goed aan te rade te gaan bij sociale wetenschappers die studie hebben gemaakt van de wijze waarop 'people define, generate, select, evaluate and present explanations', welke bias daarbij optreedt, welke verwachtingen er worden gewekt en welke sociale interactie daarbij optreedt. De auteur voerde daarom een uitvoerige literatuurstudie c.q. survey uit naar de 'philosophical, cognitive and social foundations of explanation withan emphasis on everyday explanation'. De belangrijkste bevindingen laten zich evenwel kort samenvatten. Zo zijn verklaringen volgens Miller in de eerste plaats contrastief, dat wil zeggen dat de vraag 'waarom A', eigenlijk impliceert 'waarom niet B en niet C'. In de tweede plaats zijn verklaringen dikwijls selectief en biased. Men stelt zich tevreden met een of twee oorzaken, die als 'de verklaring' worden aangemerkt en overtuigend heten te zijn zonder het gehele proces of mechanisme c.q. overige variabelen in ogenschouw te nemen. Daarnaast zijn volgens de auteur probabilistische aspecten ondergeschikt, mensen lijken waarschijnlijkheidsaspecten ten onrechte te negeren of niet goed te begrijpen, maar bij het geven of beoordelen van verklaringen terug te vallen op kwalitatieve (causale) benaderingen en kwalificaties. Tot slot benadrukt hij het sociale karakter van verklaringen, het gaat in feite om een bepaald discourse, een dialoogvorm, of taalspel en een linguïstisch, argumentatief perspectief is daarbij essentieel.

Deze focus op de *homo mensura* of menselijke maat door Miller is belangwekkend en lovenswaardig, al zijn enige kanttekeningen op zijn plaats. Zo vermijdt de auteur bewust de omvangrijke literatuur over causaliteit. Dat is vreemd omdat verklaringen, causaliteit en wetmatigheden/natuurwetten nauw verbonden zijn en dikwijls in hun onderlinge samenhang worden geïntroduceerd en besproken. Meer in het bijzonder negeert hij belangrijke filosofische inzichten over verklaringen, een notoir en grillig begrip, dat vele verschijningsvormen kent en een lange geschiedenis heeft doorlopen. Daarnaast moet worden opgemerkt dat veel van Millers bevindingen al genoegzaam bekend waren door het werk van psychologen als Kahnemann en Tversky uit de jaren zeventig, taalfilosofen als Wittgenstein, Austin en Searle. Dat geldt ook voor het door de reeds genoemde Reichenbach gemaakte onderscheid tussen de *logic/context* of *discovery* en de *logic/context* of *justification*. Wetenschapsfilosofen moesten zich primair met de tweede categorie bezighouden. Kort gezegd is de wijze waarop c.q. het pad waarlangs wetenschappelijke resultaten tot stand komen schimmig, wisselvallig, onsystematisch niet te doorgronden (gebaseerd op deels onbewuste processen, dromen of vergissingen) en in wezen irrelevant. De context of justification betreft de uiteindelijk verantwoording, rechtvaardiging en geaccepteerde presentatie van de resultaten en dit is voor de voortgang van de wetenschap en het begrip daarvan relevant. Dit principe is ook van toepassing op menselijke experts, die veeleer hun keuzes, beslissingen en acties moeten rechtvaardigen en toelichten, dan de indruk te wekken volledig toegang te hebben tot de gevolgde denkprocessen, intenties of redeneerstappen. Waarom zou men de artificiële agent dan niet toestaan in eenzelfde taalspel deel te nemen als zijn menselijke tegenpool? (Starmans, 2021)

Belangrijker dan deze kritiek is evenwel het feit dat bij de reconstructie van verklaringen in het licht van Millers inzichten de methodiek die met de padmodellen van Wright in gang is gezet een belangrijke rol speelt, eerst wat betreft de transparantie (de selectie van relevante variabelen en begrippen), dan de interpretatie van verwante concepten en hun relaties en tot slot verklaringen die recht doen aan de genoemde randvoorwaarden van de menselijke informatieverwerking. Daarbij kan men nog een stap verder gaan. Volgens Pearl, wiens methode feitelijk een uitbreiding is van de padanalyse en SEM met innovaties als backpropagation, do-operator en een uiteindelijke counterfactual interpretatie van causaliteit, is deze aanpak noodzakelijk om de problemen waarmee de AI worstelt het hoofd te bieden. Wil men de schaduwzijde van Deep Learning compenseren en uiteindelijk het project van de Sterke AI redden, dan biedt deze grafische methode voor causal inference uitkomst, aldus de auteur, die hierover in het slothoofdstuk van *The Book of Why* geen enkel misverstand laat bestaan.

Een tweetal opmerkingen tot slot. Allereerst blijkt dat het erfgoed van Sewall Wright honderd jaar later nog niets aan relevantie heeft ingeboet. Klassieke technieken als PCA en regressieanalyse mogen dan nog steeds hoog op ranglijsten van meest populaire machine learning algoritmes staan, de 'methode der padcoëfficiënten' speelt in tijden van data science en big data een andersoortige, maar eveneens belangrijke rol. Ook illustreert de hier uiteraard summier geschetste ontwikkelingsgang van Wrights methode hoezeer vooruitgang in de wetenschap dikwijls toch veeleer evolutionair dan revolutionair verloopt. Dat blijkt ook uit het proefschrift dat Jerzy Neyman in 1923 verdedigde, twee jaar na het verschijnen van Wrights artikel, waarin de ruwe contouren van de *potential outcome/counterfactual* benadering van Donald Rubin uit de jaren negentig al bespeurbaar zijn. Daarmee is het fundament van twee dominante hedendaagse probabilistische benaderingen van causaliteit (Pearl en Rubin) reeds in de vroege jaren twintig van de vorige eeuw gelegd.

Literatuur

Blalock, H. M. (1961). Correlation and causality: The multivariate case. *Social Forces, 39*, 246–251.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. University of North Carolina Press.

Buchanan, B., & Shortliffe, E. (1984). *Rule-based expert systems; The MYCIN experiments in the Stanford Heuristic Programming Project*. Addison-Wesley.

Burks, B. S. (1926). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology, 17*(8), 532–540.

Cohen, P. (1995). *Empirical Methods for Artificial Intelligence*. MIT Press.

Daniel J. D., & Legerski, J. (2006). Causal Modeling and the Origins of Path Analysis. *Theory and Sience, 7*(1).

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Models in the social sciences* (pp. 85–112). Academic Press.

Krüger, L., Daston, L., & Heidelberger, M. (Eds.). (1987). The probabilistic revolution; Volume 1: Ideas in history. MIT Press.

Krüger, L., Gigerenzer, G., & Morgan, M. (Eds.).(1990). The probabilistic revolution; Volume 2: Ideas in the sciences. MIT Press.

Niles, H. E. (1922). Correlation, causation and Wright's theory of 'path coefficients'. *Genetics, 7*, 258–273.

Niles, H. E. (1923). The method of path coefficients: An answer to wright. *Genetics, 8*, 256–260.

Pearson, K. (1892). *The Grammar of Science*. Walter Scott. Dover Publications.

Simon, H. A. (1957). *Models of man*. Wiley.

Starmans, R.J.C.M. (2018). Statistiek en Causaliteit; voortgang van een moeizame samenspraak. *STAtOR, 19*(4).

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.

Starmans, R. J. C. M. (2021). Over de breekbaarheid van het Goede; toeval, moraal en de reikwijdte van een verklarende wetenschap. *Tijdschrift Filosofie, 31*(2).

Wright, S. (1923). The theory of path coefficients: A reply to Niles's criticism. *Genetics, 8*, 239–255.

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

Sensitiviteit, specificiteit en COVID-19

Testen wordt genoemd als een van de middelen om het verspreiden van het SARS-CoV-2 virus te stuiten. Hoe effectief medische testen zijn hangt nauw samen met de sensitiviteit, specificiteit en prevalentie van het virus in de geteste populatie. Alleen die prevalentie in testgroepen is vanuit beleid en testadvies te beïnvloeden. Afhankelijk van de testuitslagen is veel patiënten in de testgroep opnemen wenselijk of juist niet.

Anton Meijburg

Nederland is nu een jaar de ban van het SARS-CoV-2 virus, dat de ziekte COVID-19 veroorzaakt, en het stoppen ervan. Naast vaccins wordt ook testen veelvuldig genoemd als manier om de verspreiding te stuiten. Het percentage positieve uitslagen is niet gelijk aan het percentage COVID-19-patiënten in je testgroep. Vanuit individueel perspectief is de kans dat de testuitslag juist is ook niet constant. Prevalentie, sensitiviteit en specificiteit beïnvloeden deze kansen. Dit komt in de media nauwelijks aan bod. Reden genoeg om het testen nader te bekijken.

Sensitiviteit, specificiteit en prevalentie

Een fout-positieve uitslag noemt de statistiek een type I fout, een fout-negatieve uitslag een type II fout. Bij medisch-diagnostische tests spreekt men vaak over de verwante begrippen ‘sensitiviteit’ en ‘specificiteit’. Wanneer een uitslag positief is, heeft de patiënt in deze casus COVID-19 :

> Sensitiviteit: Het **aantal juist-positieve** testuitslagen **als percentage van alle COVID-19-patiënten**
>
> Specificiteit: Het **aantal juist-negatieve** testuitslagen **als percentage van alle niet COVID-19-patiënten**.

In het document *Over de betrouwbaarheid van de PCR-test voor SARS-CoV-2* geeft het RIVM aan dat de sensitiviteit van de testen tussen de 67% en 98% ligt. Dat is nogal brede schatting die samenhangt met de hoeveelheid virusuitscheiding, moment van monsterafname en de kwaliteit van het monster. De specificiteit is preciezer bekend, die ligt tussen de 96,0% en 99,5%. De prevalentie is het percentage COVID-19-patiënten in de testgroep.

Rekenvoorbeeld week 2 2021

Op de website van het RIVM is te zien hoeveel testen wekelijks uitgevoerd worden en wat de uitkomsten zijn. In week 2 lieten 294.712 mensen zich testen en daaruit kwamen 32.535 positieve resultaten. Maar wat zegt dit nu precies? Uitgaande van de hoogste schatting in sensitiviteit en specificiteit (respectievelijk 98% en 99,5%) leidt dit tot de cijfers in tabel 1, waarbij EC staat voor ‘Echt COVID-19’ en EN voor ‘Echt Niet COVID-19’:

We zien 2 vergelijkingen met EC en EN als onbekenden, oplossen geeft de cijfers die tabel 2 laat zien.

Met de laagste schattingen sensitiviteit en specificiteit (respectievelijk 67% en 96,0%) zijn de cijfers zoals weergegeven in tabel 3.

Met deze testuitslagen loopt de schatting voor het totaal aantal werkelijke patiënten in de groep niet heel erg uiteen. Hoeveel patiënten juist geïdentificeerd worden wel.

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	0,98*EC	0,02*EC	0,98*EC+0,02*EC
Geen COVID-19	0,005*EN	0,995*EN	0,005*EN+0,995EN
Totaal	32.535	262.177	294.712

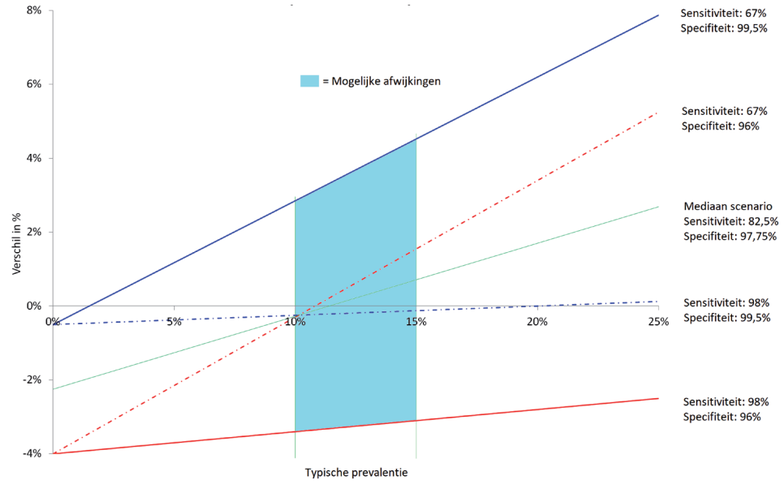
Tabel 1. Berekeningen week 2 met hoogste sensitiviteit en specifiteit

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	31.221	637	31.858
Geen COVID-19	1.314	261.540	262.854
Totaal	32.535	262.177	294.712

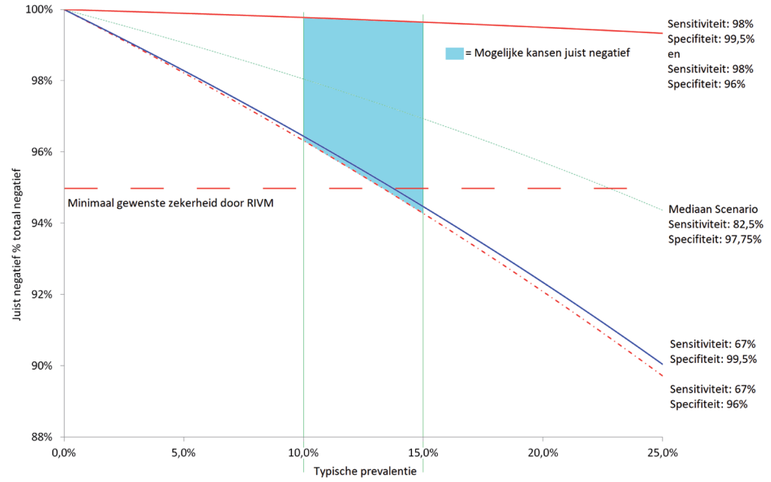
Tabel 2. Uitkomsten week 2 met hoogste sensitiviteit en specifiteit

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	22.064	10.867	32.931
Geen COVID-19	10.471	251.310	261.781
Totaal	32.535	262.177	294.712

Tabel 3. Uitkomsten week 2 met laagste sensitiviteit en specifiteit



Figuur 1. Verschil prevalentie en percentage positief testresultaat



Figuur 2. Juist-negatieve uitslagen als percentage van alle negatieve uitslagen

De schattingsfout in prevalentie

Het percentage positieve uitslagen ligt doorgaans tussen de 10% en 15% in de testgroepen maar niet elke patiënt krijgt dus een positieve uitslag. Het verschil tussen het aantal positieve uitslagen als percentage van het totaal aantal uitslagen en de werkelijke prevalentie in de testgroep staat in figuur 1 voor verschillende scenario's voor sensitiviteit en specificiteit. Het mediane scenario zit qua sensitiviteit en specificiteit in het midden van de schattingen, op respectievelijk 82,5% en 97,75%.

Bij een prevalentie tussen 10% en 15% is het percentage positieve uitslagen maximaal 3,4% minder tot 4,5% meer dan die prevalentie. Op een totale prevalentie van maximaal 15% zijn dit grote marges. Bij een sensitiviteit van 98% en een specificiteit van 99,5% zijn de onzekerheden uiteraard het kleinst.

De grafiek laat zien dat een sensitiviteitsverschil van 31% een grote variatie geeft in de schattingsfout voor de werkelijk prevalentie. Het verschil in schattingsfout bij een gelijkblijvende sensitiviteit maar een 3,5%-verandering in specificiteit is al de helft van de inschattingsfout bij een sensitiviteitwijziging van 31%. Dus een kleine verandering in specificiteit heeft een veel grotere invloed op het misschatten van de prevalentie dan een *gelijke* verandering in sensitiviteit.

Dit komt omdat 4% van de groep met een negatieve uitslag (die fout-blijkt) in aantal uitslagen al de helft is van het aantal juist-positieve uitslagen bij typische testgroepen. Wanneer dit percentage fout-negatieven daalt naar 0,5% gaat het meteen om 8 keer zo weinig uitslagen. Het aantal fout-negatieve uitslagen veranderd dan van de helft naar minder dan een tiende van het aantal juist-positieve uitslagen waardoor een veel groter deel van de positieve uitslagen juist is.

De inschattingsfout is bij een gegeven sensitiviteit en specificiteit ook niet constant voor verschillende prevalenties. De prevalentie is het aantal juist-positieve uitslagen opgeteld bij het aantal fout-negatieve uitslagen. Het aantal positieve uitslagen is het aantal juist-positieven opgeteld bij het aantal fout-positieven. Bij de prevalenties in de grafiek zijn zelfs 3 van de 4 scenario's bij een bepaalde prevalentie helemaal juist. De grafiek geeft de indruk dat voor de testgroepen het scenario met de laagste sensitiviteit en specificiteit het op een na beste scenario is.

Op die punten waarin geen voorspellingsfout gemaakt wordt in het aantal patiënten is het aantal fout-positieven gelijk aan het aantal fout-negatieven. Zo is het totaal aantal positieve uitslagen gelijk aan het aantal werkelijke COVID-19-patiënten. Het zijn alleen niet allemaal de juiste mensen die de positieve uitslagen krijgen. Dus voor 3 van de 4 test scenario's is er wel een punt waarin het juiste aantal mensen geïdentificeerd wordt en in quarantaine zit, maar een flink deel daarvan is niet de juiste persoon. In tabellen 2 en 3 doet dit fenomeen zich ook voor.

Groepsperspectief versus individueel perspectief

De vraag 'Welk deel van de COVID-19 patiënten wordt gevonden?' heeft een 'groepsperspectief'. Mensen die zelf een uitslag krijgen willen het antwoord op de vraag: 'Wat is de kans dat mijn uitslag juist-positief of juist-negatief is?'. Deze vraag heeft een 'individueel perspectief'. Daarbij is de hoeveelheid juist-positieven en juist-negatieven *als percentage van alle of positieve of negatieve uitkomsten* van belang.

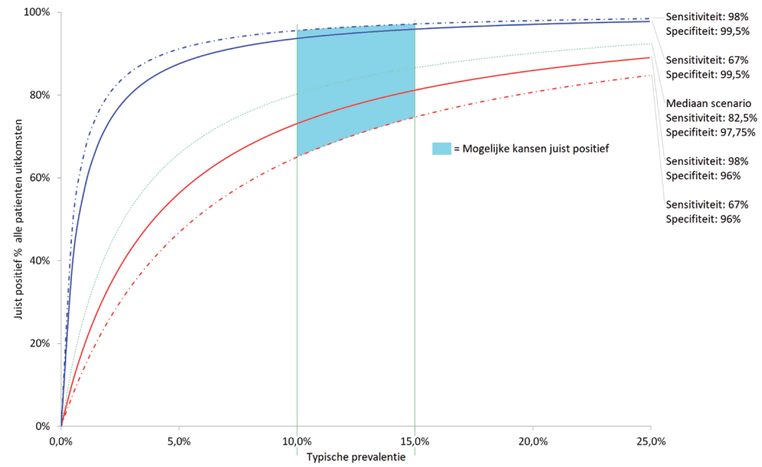
Figuur 2 toont de cijfers voor een negatieve uitkomst: De kans dat een negatieve uitslag juist is wanneer je zo'n uitslag krijgt ligt tussen de 94% en 100%. Het RIVM wenst dat de kans op een juist-negatieve uitslag minimaal 95% is, dit is een zekerheid die in de praktijk nagenoeg altijd gehaald zal worden. Deze keer heeft de sensitiviteit een grote invloed. Wanneer die hoog is zijn er relatief veel mensen met een juist-positieve uitslag. Dan zijn er nauwelijks patiënten over die een negatieve uitslag kunnen krijgen, het aantal fout-negatieven is noodzakelijker wijs klein.

Ook is te zien dat hoe hoger de prevalentie is, hoe groter de kans is dat een negatieve uitslag onjuist is. Het aantal patiënten dat geen juiste uitslag krijgt wordt namelijk groter en de groep niet-patiënten kleiner. Als percentage van de niet-patiënten neemt het aantal patiënten met een fout-negatieve uitslag hard toe.

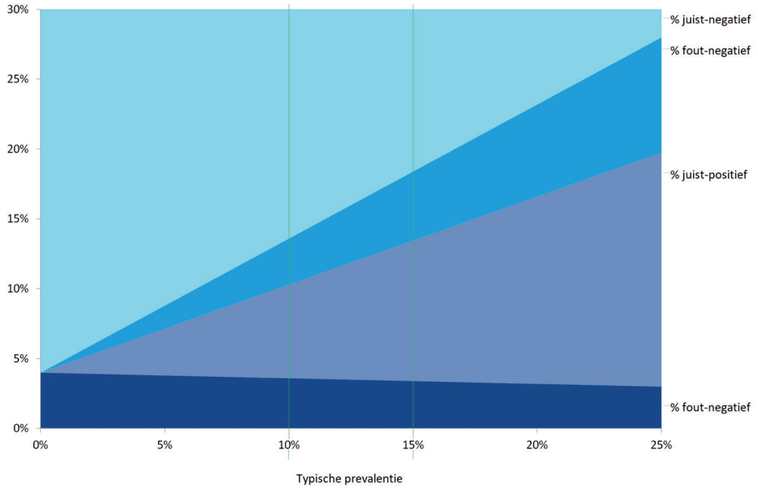
Belang van individu met negatieve uitslag tegen gesteld aan belang van individu met positieve uitslag

Voor het testen op COVID-19 ligt de kans dat een positieve uitslag juist is tussen de 65% en 97,2%. Voor de kans op een juist-positieve uitslag heeft de kleine verandering in specificiteit van 3,5% een veel groter effect dan enorme verandering in sensitiviteit van 31%. Dit is weer het gevolg van het feit dat het merendeel van de geteste mensen geen COVID-19 heeft. Een net wat groter deel van die mensen als correct negatief identificeren is relatief veel mensen ten opzichte van het aantal COVID-19-patiënten. Deze mensen kunnen dan niet meer een fout-positieve uitslag krijgen. Zie figuur 3.

Wanneer er gekozen wordt om ook te testen zonder



Figuur 3. Juist-positieve uitslagen als percentage alle positieve uitslagen



Figuur 4. Percentage alle uitslagen, mediaan scenario

klachten zal de prevalentie in de testgroep flink afnemen en de kans op een juist-positieve uitslag daalt dan ook snel. Hoe hoog de specificiteit dan ook is, bij heel weinig patiënten is een kleine hoeveelheid fout-positieven al snel een groot aantal vergeleken met het aantal patiënten. De groep die een positieve uitslag ontvangt bevat dat veel niet-patiënten.

Meer COVID-19-patiënten in de testgroep is in het belang van de mensen die een positieve uitslag ontvangen. De kans dat die dan daadwerkelijk klopt is groter. Het gaat juist tegen het belang in van mensen die een negatieve uitslag krijgen, de kans dat die klopt wordt juist kleiner. De sensitiviteit en specificiteit zijn met beleid niet goed te beïnvloeden. Maar hoeveel COVID-19 patiënten in de testgroep zitten kan wél worden beïnvloed door adviezen te geven over wanneer wel en niet te testen. Een beleidsmaker moet kiezen tussen meer mensen met COVID-19 rond laten lopen of meer mensen zonder COVID-19 isoleren.

Het totale plaatje

Inzicht in de aandelen juiste en onjuiste uitkomsten van negatieve en positieve testen voor de gehele testgroep is wenselijk. Voor het mediane testscenario is de opbouw van alle testuitslagen zoals figuur 4 toont

De absolute hoeveelheid fout-negatieven neemt sneller toe dan de hoeveelheid fout-positieven afneemt bij een hogere prevalentie. Dus het aantal onjuiste uitslagen is groter bij een toenemende prevalentie. Dat komt omdat voor deze testen de sensitiviteit doorgaans lager is dan de specificiteit. Al met al blijft de hoeveelheid juiste uitslagen (negatief en positief) groot.

Testbeleid in de praktijk

Het testen om COVID-19-patiënten op te sporen is steeds effectiever naarmate het percentage patiënten in de testgroep toeneemt. Hoe hoger de prevalentie, hoe groter het aandeel patiënten dat je vindt en hoe groter de kans dat een positieve uitslag juist is. Het testbeleid is er ten tijde van dit schrijven dan ook op gericht om het aandeel COVID-19-patiënten in test groepen zo hoog mogelijk te maken. De kans dat een negatieve uitslag juist is neemt dan wat af maar blijft doorgaans boven de 95%. Een minimale wijziging in de specificiteit heeft een grote invloed op het aantal gevonden patiënten. Om mensen meer vrijheid te geven wordt gedacht over continue testen. De prevalentie van COVID-19 in de testgroep is dan laag en dan zal een positieve uitslag zal weinig toegevoegde waarde hebben. De kans op een fout-positieve uitslag is immers erg groot.

Literatuur

RIVM. (z.d.) *Over de betrouwbaarheid van de PCR-test voor SARS-CoV-2.* RIVM.

Anton Meijburg studeerde Toegepaste Wiskunde aan de Stenden NHL Hogeschool Leeuwarden en is Statistisch Analist A. Hij werkt bij PGGM Vermogensbeheer als medewerker bestuurlijke informatie.
E-mail: antonmeijburg@gmail.com.

Onlangs bevatte *STAtOR* een artikel waarin werd beschreven hoe 19e-eeuwse statistici zich inzetten tegen drankmisbruik. Toen ik het las moest ik meteen denken aan een van de zeldzame keren in mijn leven dat ik behoorlijk aangeschoten ben geraakt, en dat nog wel in functie!

Het was aan het eind van de jaren 90 van de vorige eeuw, tijdens een van de tweejaarlijkse ISI-congressen. Een gebruikelijk onderdeel van die congressen is het diner dat de Directeur-Generaal voor de Statistiek van het gastland geeft voor al zijn collega's die het congres bezoeken en nog wat gasten van vergelijkbare statuur. Afhankelijk van budget, luim van de gastheer en ruimte in de zaal worden dan, naast de directeur van het Permanent Office van het ISI, ook enkele medewerkers van dat Office uitgenodigd. Ik heb gedurende de ruim 25 jaar dat ik als vrijwilliger bij dat Permanent Office werk vrijwel alle congressen bezocht en regelmatig aan zo'n diner deelgenomen.

Ook ditmaal was ik uitgenodigd en men had mij tegenover de DG van een Balkanland geplaatst, een land dat nauw betrokken was bij de vele vijandigheden die volgden op het uiteenvallen van het voormalige Joegoslavië. De goede man sprak amper Engels en werd daarom vergezeld van een assistente die als tolk fungeerde. Na het voorgerecht ging het mis, hij was tot de ontdekking gekomen dat ik uit Nederland kwam, een van de landen die participeerden in het bombarderen van Sarajevo. Hij hield mij zo ongeveer persoonlijk medeverantwoordelijk voor die bommen. Met een verontschuldigende glimlach vertaalde zijn assistente zijn woedende opmerkingen in mijn richting. Daarbij vermoedde ik dat ze de ál te scherpe kantjes wegliet. Nu heb ik altijd weinig interesse en inzicht gehad in politieke ontwikkelingen, en de ingewikkelde verhoudingen in de Balkan waren al helemaal onbegrijpelijk voor me. Maar ik kon de heftige verbale aanvallen moeilijk negeren, ik zat daar tenslotte niet alleen en allengs begonnen meer mensen mee te luisteren.

Ik besloot me totaal te onthouden van enig politiek commentaar, dat was niet moeilijk want zoals gezegd wist ik daar niets van. Ik heb verteld dat ik me verontschuldigde voor mijn totale politieke ignorantie en zei dat dit een wetenschappelijke bijeenkomst was waarbij ik aanwezig was als wetenschapper. Ik zei dat hij dat ongetwijfeld ook was en stelde daarom voor om, als collega-wetenschappers, samen het glas te heffen op onze gezamenlijke interesse en de politiek er verder buiten te laten. Waarschijnlijk was de man op voornamelijk politieke gronden in zijn belangrijke functie benoemd want hij was zichtbaar gevleid dat ik hem als wetenschapper beschouwde. Tot mijn grote opluchting sloeg zijn eerst zo vijandige houding om als een blad aan een boom, hij omhelsde me, greep zijn glas, leegde het in een teug en schonk onze glazen nog eens in.

De politiek kwam die avond verder niet meer ter sprake, wel moest er vele malen op onze wetenschappelijke vriendschap worden geklonken. Ondanks de uiterst zuinige nipjes die ik nam was ik aan het eind van het diner toch meer aangeschoten geraakt dan me lief was. Maar gelukkig schoot ook daar die assistente weer te hulp, zij gaf me een steuntje op de smalle loopplank naar de boot die ons terug zou brengen van het eiland waar het diner plaatsvond. Eenmaal op de vaste wal, buiten gehoorbereik van haar baas, verontschuldigde ze zijn eerdere aanvallen en zei 'You handled that awful situation very well'.

Tsja, er zullen niet veel wetenschappers zijn die kunnen zeggen dat ze uit hoofde van hun functie aangeschoten zijn geraakt.

Gerrit Stemerdink is eindredacteur van *STAtOR*.
E-mail: gjstemerdink@hotmail.com

Van 11 tot en met 16 juli 2021 vindt WSC2021 plaats, het 63e ISI World Statistics Congress. Oorspronkelijk zou dit een 'traditioneel' congres zijn dat plaats zou vinden in Den Haag. Maar vanwege de Coronacrisis zal deze editie een virtuele zijn.

Het thema is *Statistics and Data Science for a Better World*. Binnen dit thema zullen de ontwikkelingen en bijdragen van statistiek en data science in alle aspecten van de samenleving worden belicht. Het Scientific Programme Committee heeft een programma samengesteld met meer dan 216 Invited Paper Sessions. Het zal dus een omvangrijk programma zijn, met zowel vooraf opgenomen presentaties als live sessies. Voor de laatste is er dagelijks, van maandag t/m vrijdag, minstens één (dinsdag en donderdag twee) op een prime time tijdstip gepland waarbij ook rekening is gehouden met de verschillende tijdzones. Het WSC behoudt hierdoor ook in het virtuele format zijn internationale karakter. Sterker nog, door het online format, de lagere registratiekosten en het wegvallen van reiskosten verwacht men aanzienlijk meer deelnemers dan gewoonlijk te krijgen.

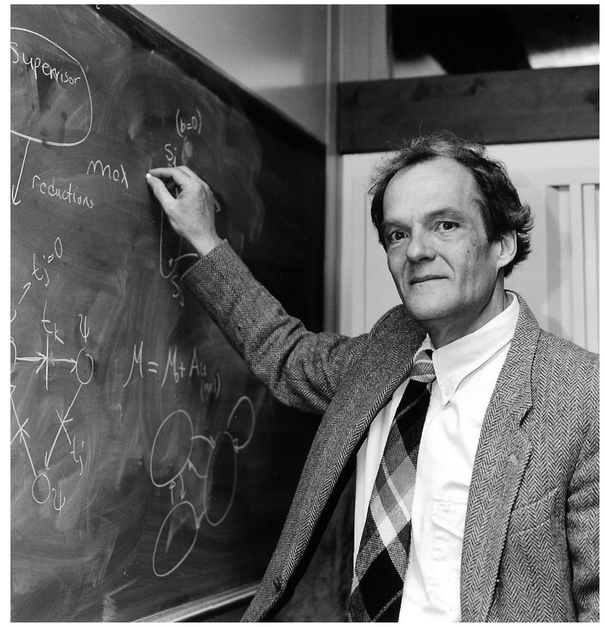
Prominente sprekers zoals Kerrie Mengersen (ISI President's Invited Speaker) en Nan Laird (winnaar van de International Prize in Statistics 2021) maken deel uit van het programma. De International Prize in Statistics 2021 zal aan Nan Laird worden uitgereikt tijdens de openingsceremonie op zondag 11 juli. Behalve de gebruikelijke toespraken staan ook enkele culturele verrassingen gepland voor de openingsceremonie.

De interactie tussen deelnemers, sprekers, organisatoren en sponsoren is een belangrijk onderdeel van het congres. In de virtuele expositieruimte zullen zowel organisaties als bedrijven zichzelf presenteren. Ook netwerk sessies voor deelnemers maken onderdeel uit van het programma. Deelnemers worden uitdrukkelijk uitgenodigd om gebruik te maken van de extra mogelijkheden die het virtuele platform biedt om actief te netwerken.

Het WSC2021 wordt georganiseerd door het International Statistical Institute (ISI) en haar zeven Associations: Bernoulli Society (BS), International Association for Official Statistics (IAOS), International Association for Statistical Computing (IASC), International Association for Statistical Education (IASE), International Association of Survey Statisticians (IASS), International Society for Business and Industrial Statistics (ISBIS) en The International Environmetrics Society (TIES).

WSC2021 is actief op sociale media: **Twitter** & **Instagram** @isi_wsc_2021,
Facebook @63rd-ISI-WSC-2021, **LinkedIn** /international-statistical-institute-nl/. **Tag het WSC 2021**: #isiwsc2021.
Website https://www.isi2021.org/ wordt dagelijks geactualiseerd.

IN MEMORIAM



Arie Hordijk 1940 — 2021

Lodewijk Kallenberg, Ger Koole & Floske Spieksma

Op 6 april 2021 is Arie Hordijk, emeritus-hoogleraar Matematische Besliskunde van de Universiteit Leiden, overleden. Hij is 81 jaar geworden. Zijn wetenschappelijk werk lag op het terrein van de stochastische besliskunde, een vakgebied waarin hij internationaal een zeer vooraanstaand onderzoeker was.

Opleiding

Arie Hordijk werd op 18 februari 1940 in Rockanje geboren. Na het HBS-B diploma aan de Christelijke HBS te Rotterdam in 1958 te hebben behaald, wilde hij graag wiskunde gaan studeren. Zijn ouders stuurden hem echter naar de Technische Hogeschool Delft, waar hij – bij gebrek aan een wiskundeopleiding – technische natuurkunde ging studeren. Binnen een jaar ging hij zijn eigen weg en vertrok uit Delft om aan de Vrije Universiteit Amsterdam alsnog wis- en natuurkunde te gaan studeren. De colleges die hem het meest boeiden waren die van de hoogleraren Baayen (topologie) en Van Rooijen (numerieke wiskunde). Zijn eerste publicatie stamt uit 1966 en gaat over topologische halflichamen.

In 1967 behaalde Arie cum laude zijn doctoraal en stond voor de keuze 'wat nu'? Hij besloot naar het Mathematisch Centrum in Amsterdam (het huidige CWI) te gaan, waar hij, na overleg met Baayen, koos voor de groep Statistiek en Operations Research, die onder leiding stond van de statisticus Jan Hemelrijk.

Amsterdam

Op het Mathematisch Centrum ontmoette hij Gijs de Leve, hoofd van de sectie Operations Research. De Leve deed als eerste in Nederland onderzoek op het jonge gebied van de Markov beslissingstheorie. Dit onderwerp boeide Arie enorm en hij zou dit onderzoeksgebied zijn hele leven trouw blijven. Hij ging ook andere onderdelen van de stochastische besliskunde intensief bestuderen en verrijken met 'diepe' resultaten. Diepgang, niet alleen in wiskundig onderzoek maar ook op andere terreinen, was een kenmerkende eigenschap.

Naast zijn werkkring op het Mathematisch Centrum werd Hordijk in 1971 aangesteld als wetenschappelijk medewerker aan de Vrije Universiteit.

Promotieonderzoek

Voor eindige toestandsruimtes was de Markov beslissingstheorie al deels ontwikkeld, maar voor oneindige toestandsruimtes lag het terrein grotendeels nog open. Samen met Henk Tijms en Awi Federgruen stortte Arie

zich bij het Mathematisch Centrum op dit onderzoek. Ook in Eindhoven werd hieraan, onder leiding van Jaap Wessels, gewerkt. Zo deed Nederland internationaal in de frontlinie mee op dit gebied en Hordijk werd daarin een van de leidende figuren.

Het onderzoek leidde tot zijn proefschrift *Dynamic programming and Markov potential theory,* dat hij onder supervisie van prof. Th. Runnenburg (Universiteit van Amsterdam) en van prof. J.A. Bather (University of Sussex) in 1974 voltooide.

Stanford

Arie's onderzoeksresultaten betroffen onder andere een generalisatie van het onderzoek van A. F. Veinott Jr. uit Stanford. Na zijn promotie werd Arie door hem uitgenodigd om enige maanden in Stanford als *visiting associate professor* te komen werken. Dit leidde tot een verdere verdieping van zijn werk aan Lyapunov-functies voor regeneratieve Markov beslissingsproblemen.

Leiden

In 1976 werd Arie Hordijk, als opvolger van Guus Zoutendijk, benoemd tot hoogleraar in de mathematische besliskunde aan de Rijksuniversiteit van Leiden. Naast colleges in de stochastische besliskunde (Wachttijdtheorie en Stochastische Dynamische Programmering) gaf hij ook colleges op het gebied van de deterministische besliskunde (Grafentheorie en Discrete Wiskunde). Deze colleges worden nog steeds gegeven.

Promovendi

Onderzoek deed hij zowel alleen, met promovendi, waarvan hij er elf heeft gehad, en met andere Nederlandse en buitenlandse onderzoekers. Zijn onderzoek was breed en diepgaand, en betrof het hele spectrum van de stochastische besliskunde, met de nadruk op Markov (beslissings) ketens, Markov spelen en wachttijdtheorie. De onderliggende processen waren zowel discreet als continu, met zowel eindige als oneindige toestandsruimte. Zijn brede wiskundige achtergrond in de zuivere en de toegepaste wiskunde kwamen daarbij uitstekend tot hun recht. Ook wist hij snel nieuwe gebieden te doorzien en zich nieuwe kennis eigen te maken.

Het samenwerken met één andere onderzoeker was hem het liefst en zijn promovendi hebben hier dankbaar gebruik van kunnen maken. Wekelijkse sessies van enkele uren waren geen uitzondering. Met zijn promovendi werkte hij aan continue Markov beslissingsprocessen (Frank van der Duijn Schouten [1979]), lineaire programmering voor eindige stochastische beslissingsproblemen (Lodewijk Kallenberg [1980]), tijdsdiscretisatie en netwerken van wachtrijen (Nico van Dijk [1983]), gevoelige criteria voor problemen met aftelbare toestandsruimte (Rommert Dekker [1985]), stochastische ongelijkheden voor wachtrijen (Ad Ridder [1987]), ergodische Markov ketens en optimaal sturen van wachtrijen (Floske Spieksma [1990]), routering en scheduling in netwerken van wachtrijen (Ger Koole [1992]), Markov beslissingsketens met gedeeltelijke informatie (Anneke Loeve [1995]), Markov spelen en wachtrijen (Olaf Passchier [1996]), optimale routering (Dinard van der Laan [2003]) en random walks (Nicolay Popov [2003]).*

Andere samenwerking

Hij onderhield ook vele contacten met buitenlandse collega's, wat leidde tot een groot aantal publicaties. Enkele van deze personen zijn (min of meer in tijdsvolgorde): Schweitzer, Iglehart, Schassberger, Sladký, Hill, Puterman, Chang, Righter, Weiss, Lasserre, Altman, Gajrat, Malyshev, Gaujal, Yushkevich, Liu, Towsley, Borovkov, Shwartz en Weisshaupt.

Behalve met zijn promovendi had Arie ook met andere Nederlanders regelmatig contact, wat gezamenlijke publicaties opleverde. We noemden al Henk Tijms en Awi Federgruen. Daarnaast waren dit (ook weer in chronologische volgorde): Kees van Hee, Piet Holewijn, Jan van der Wal, Koos Vrieze, Gerard Wanrooij, Sandjai Bhulai en Bernd Heidergott. Met deze laatste heeft Arie nog lang na zijn pensionering intensief samengewerkt.

Uit het bovenstaande blijkt duidelijk het brede onderzoeksgebied waarop Arie Hordijk werkzaam en productief is geweest. Dit heeft geleid tot een viertal boeken, waaronder 1 monografie, en 150 publicaties. Van deze publicaties staan er 13 op zijn eigen naam, 76 samen met promovendi en 61 met andere co-auteurs.

Van Dantzigprijs

Voor zijn indrukwekkend onderzoek op het gebied van de Operations Research werd hem in 1980 de Van Dantzig Prijs toegekend, de hoogste prijs in de Nederlandse Statistiek en Operations Research, die eenmaal in de vijf jaar wordt toegekend.

Onderwijs

Naast de al eerdergenoemde colleges die hij gaf en het opleiden van zijn promovendi, heeft hij vele andere studenten begeleid bij het schrijven van hun bachelor- of doctoraal scripties. Het moeten er tussen de 150 en 200 zijn geweest.

Een van de schilderijen gemaakt door Arie Hordijk

Redactiewerk

Arie Hordijk was een zeer gewaardeerd redacteur van de tijdschriften *Mathematics of Operations Research, Applied Probability*, *Mathematical Methods of Operations Research* en *Probability in the Engineering and Informational Sciences*. Hij heeft dit werk vele jaren en consciëntieus gedaan.

Organisatorische zaken

Hij had een hekel aan organisatorische zaken; hij vond de meeste eigenlijk zonde van zijn tijd. Hij zocht dit daarom zeker niet op, maar als op hem een beroep werd gedaan en hij dat redelijk achtte, dan was hij bereid dit werk op zich te nemen. Organisatorische zaken op het gebied van de stochastische besliskunde deed hij wel met genoegen. Hij was projectleider van de Stochastische Besliskunde binnen het Stieltjes Instituut, medeorganisator van drie Oberwolfach workshops en van drie workshops die in Leiden werden gehouden.

Privéleven

De diepgang die Arie aan de dag legde in de wiskunde was ook typerend voor de activiteiten die hij privé ontplooide. Hij was een enthousiast fietser, wandelaar en liefhebber van de prachtige duinen, zee en strand in zijn woonomgeving. Kunst, met name zingen, was zijn echte passie. Vrijwel zijn hele leven zong hij in een koor, samen met zijn vrouw Marina.

Rond zijn pensionering in 2005 werd Arie helaas getroffen door de ziekte van Parkinson. Dit belemmerde hem niet vanuit zijn zo geliefde huis in Castricum actief te blijven in de wetenschap. Ook het zingen was een grote steun. Toen hij niet meer naar het koor kon, liet hij een professionele zangeres bij hen thuis komen.

De laatste jaren van zijn leven ging Arie steeds meer op in zijn hobby's schilderen en dichten. Vele schilderijen heeft hij gemaakt, de meeste hadden als onderwerp 'duinen en zee'. Op de overlijdensannonce stond zijn gedicht 'Op het stille strand'.

Met zijn heengaan verliezen wij een markante persoonlijkheid en een gedreven en excellent wetenschapper.

Voedselhulp in het noorden van Ouagadougou in Burkina Faso, 2020. Foto: WFP/Marwa Awad

Wereldvoedselprogramma wint Franz Edelman Award

Koen Peters

Het Wereldvoedselprogramma (WFP) van de Verenigde Naties heeft met inbreng van Tilburg University's Zero Hunger Lab de internationale Franz Edelman Award 2021 gewonnen. Dat is de meest prestigieuze prijs ter wereld voor toepassingen van analytics en optimalisatie, waarbij niet alleen naar vernieuwing wordt gekeken maar vooral naar de impact.

INFORMS, de internationale vereniging voor operations research and analytics professionals, had zes finalisten geselecteerd voor de 50e jaarlijkse Franz Edelman Award for Achievement in Advanced Analytics, Operations Research and Management Science. Dit jaar, net als vorig jaar, digitaal. Bedrijven en universiteiten lieten zien welke impact ze hebben met analytics en optimalisatie. De concurrenten waren niet de minste – onder meer Amazon en Alibaba dongen mee naar de prijs – maar de Franz Edelman Award 2021 ging naar het Wereldvoedselprogramma (WFP) bereikt elk jaar meer dan 100 miljoen mensen met voedselsteun. '*Their achievements epitomize how O.R. and Analytics are saving lives, saving money, and solving problems. Nothing could be more resoundingly human.*' De toekenning was dit jaar extra speciaal, want de helft van het winnende team werkt in Nederland![1]

Humanitaire crisissen worden elk jaar complexer – denk aan Jemen of Syrië – en veel landen zijn sterk getroffen door de COVID-pandemie. Zelfs vóór COVID waren er al bijna 700 miljoen mensen ondervoed, en zo'n drie miljard mensen die een gezonde maaltijd niet kunnen betalen.[2] Het afgelopen jaar hebben we deze cijfers alleen maar zien groeien, en het einde is nog niet in zicht. Het is cruciaal om een oog in het zeil te houden, te voorspellen waar de meeste hulp nodig gaat zijn, en vervolgens de juiste beslissingen te maken. En dit is natuurlijk precies waar ons vakgebied om draait! De laatste tien jaar heeft het WFP dan ook flink geïnvesteerd in analytics, zoals dashboards en optimalisatiemodellen.

De juryleden vonden het prachtig om te zien wat WFP met analytics heeft bereikt. Zo is ondertussen al meer dan 150 miljoen dollar bespaard – genoeg om twee miljoen mensen een jaar lang van voedsel te voorzien! Daarnaast stonden analytics het afgelopen jaar centraal om WFP's COVID-operaties wereldwijd te coördineren en te optimaliseren, niet alleen voor WFP zelf maar ook voor tientallen van haar partners waaronder de Wereldgezondheidsorganisatie en het Rode Kruis.

Interdisciplinaire samenwerking stond centraal in WFP's presentatie. Zo werkt het WFP al jaren samen met universiteiten en de private sector om nieuwe prototypes te ontwikkelen en op te schalen. Op het gebied van optimalisatie werd de Universiteit van Tilburg in het zonnetje gezet – al sinds 2011 werken zij samen met het WFP aan Optimus, een optimalisatie-app die WFP onder andere helpt om de meest kosteneffectieve voedselpakketten samen te stellen. Deze app heeft het WFP al meer dan 50 miljoen dollar bespaard!

WFP's winst laat zien dat wij met onze wiskundeknobbels een positieve impact kunnen hebben op de wereld. Humanitaire organisaties kampen met ontzettend lastige beslissingen, en hebben vaak niet de tijd of de kunde om door alle data te doorploegen om de beste oplossingen te vinden. Met onze wiskunde en analyses kunnen wij inzicht brengen, en ervoor zorgen dat elke euro aan ontwikkelingssteun zo nuttig mogelijk ingezet wordt. En dit redt levens. Meer en meer overheden en humanitaire organisaties bekeren zich nu tot analytics, en met nieuwe initiatieven zoals het Zero Hunger Lab in Tilburg (belicht in de vorige editie van *STAtOR*) en Analytics for a Better World in Amsterdam tonen wij als Nederlanders aan dat wij paraat staan. Het succes van WFP heeft laten zien dat er een plaats is voor analytics aan de humanitaire tafel, en nu is het aan ons om te laten zien wat we in onze mars hebben.

Noten

1. Het projectteam bestaat onder andere uit Koen Peters (afgestudeerd in Operations Research en promovendus bij het Zero Hunger Lab in Tilburg), Nina van Ettekoven (afgestuurd in Econometrie bij de Universiteit van Amsterdam), Tim Sergio Wolter (afgestuurd in Econometrie bij de Universiteit van Tilburg), Hein Fleuren (hoogleraar Toegepaste Operations Research en mede-oprichter van Zero Hunger Lab in Tilburg) en Dick den Hertog (hoogleraar Operations Research in Amsterdam en mede-oprichter van Analytics for a Better World).
2. SOFI 2020, http://www.fao.org/publications/sofi/2020/en/

CWI

STAtOR feliciteert het 75-jarige CWI

Dit voorjaar vierde het CWI haar 75-jarig bestaan. De redactie van *STAtOR* feliciteert haar daarmee van harte.

De geschiedenis van het CWI en de VVSOR loopt gelijk op, zeker in de beginjaren. MC en VVS, zoals ze heetten bij hun beider oprichting in de naoorlogse jaren, hadden deels dezelfde initiatoren. Zonder iemand tekort te willen doen noemen we David van Dantzig als belangrijke gemeenschappelijke factor.

Gedurende het gehele bestaan van beide organisaties zijn er verbanden geweest. Een groot deel van de Nederlandse statistici heeft ooit bij het CWI gewerkt of is opgeleid door een hoogleraar met CWI-achtergrond. Zelfs de administratie van de VVSOR is enige tijd ondergebracht bij het CWI, evenals het drukken van *Statistica Neerlandica* en andere periodieken.

In een van de volgende nummers gaan we meer aandacht besteden aan de rol van het CWI. Maar voor nu: VAN HARTE!

Redactie STAtOR

Lineaaltje als reclame voor de VVS-cursussen met daarop het CWI adres