# STAtOR

periodiek van de VvS+OR jaargang 16, nummer 3, december 2015

**STAtOR**
Jaargang 16, nummer 3, december 2015

STAtOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VvS+OR). STAtOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 3 keer per jaar.

**Redactie**
Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Ana Isabel Barros, Kristiaan Glorie, Johan van Leeuwaarden, Guus Luijben (eindredacteur), Richard Starmans, Gerrit Stemerdink (eindredacteur), Hilde Tobi en Vanessa Torres van Grinsven. Vaste medewerkers: Fred Steutel en Henk Tijms.

**Kopij en reacties richten aan**
Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, telefoon 020-5986010, mobiel 06-55886747, <j.a.dossantos.gromicho@vu.nl>.

**Bestuur van de VvS+OR**
Voorzitter: prof. dr. Jacqueline Meulman <president@vvs-or.nl>
Secretaris: dr. Fetsje Bijma <bestuur@vvs-or.nl>
Penningmeester: dr. Ad Ridder <penningmeester@vvs-or.nl>
Overige bestuursleden: prof. dr. Eric Cator (SMS), prof. dr. Jeanine Houwing-Duistermaat (BMS), Maarten Kampert MSc., prof. dr. Albert Wagelmans (NGB), dr. Michel van de Velden (ECS), dr. Jelte Wicherts (SWS), Nynke Krol (Young Statisticians)

**Leden- en abonnementenadministratie van de VvS+OR**
VvS+OR, Postbus 244, 6700 AE Wageningen, telefoon 0317 - 419572, e-mail <admin@vvs-or.nl>.
Raadpleeg onze website over hoe u lid kunt worden van de VvS+OR of een abonnement kunt nemen op *STAtOR*.

**VvS+OR-website**
www.vvs-or.nl

**Sociale media**
Wilt u uw vakgenoten ontmoeten en wilt u discussiëren over actuele thema's, volg dan de VvS+OR en de Young Statisticians via LinkedIn, Facebook, Twitter en Flickr.
Sluit u aan bij de LinkedIn-groep van VvS+OR of Young Statisticians; bekijk foto's op <www.flickr.com/photos/vvs-or/sets>; like onze Facebook-pagina; volg de President van VvS+OR op <https://twitter.com/#!/dutchstat>.

**Advertentieacquisitie**
M. van Hootegem <hootegem@xs4all.nl>
STAtOR verschijnt in maart, juli en december.

**Ontwerp en opmaak**
Pharos, Nijmegen

**Uitgever**

ISSN 1567-3383

# INHOUD





## Over geld, geld en nog eens geld ...

Soms zie je ineens een verband opduiken tussen heel verschillende zaken. Zo kwam bij het schrijven van de redactionele inleiding bij dit nummer van *STAtOR* als vanzelf de gedachte aan Geld op. Geld lijkt namelijk een leidende rol te spelen bij veel van de artikelen. Natuurlijk kan men stellen dat geld per definitie een rol speelt bij *alle* artikelen op OR-gebied. Daar gaat het immers om optimale oplossingen en die zijn vrijwel altijd synoniem met minder kosten.

Een goed voorbeeld is het openingsartikel van Evelot Duijzer en haar co-auteurs. Zij beschrijven hoe men een beperkte hoeveelheid vaccin optimaal kan inzetten om een uitbraak van een infectieziekte maximaal te bestrijden.

Peter Grünwald kreeg in 2010 de Van Dantzigprijs voor zijn werk. Zijn artikel is langer en technischer dan gebruikelijk is voor *STAtOR*. Wij hebben toch voor publicatie gekozen omdat het een heel nieuwe ontwikkeling schetst. Hij vertelt over een alternatief voor de vele problemen bij het gebruik van *p*-waarden. Daarbij 'vertaalt' hij kansen naar een opbrengst in geld. Dat geeft een verrassend nieuwe kijk op oude problemen. Maar denkt u nu niet dat zo'n financiële vertaling nieuw is: Christiaan Huygens deed dat al in 1660, in de eerste publicatie ooit over statistische problemen! In *Van Rekeningh in Spelen van Geluck* schrijft hij iets als 'laat ons deze uitkomst A waard zijn' en hij vergelijkt en weegt uitkomsten van het gooien met dobbelstenen alsof het om geldbedragen gaat.

Caroline Jagtenberg en Rob van der Mei beschrijven een manier om niet meer met vaste standplaatsen voor ambulances te werken en daardoor de aanrij-tijden zo kort mogelijk te maken. Rutger Kerkkamp schrijft over nagenoeg hetzelfde probleem, hij kreeg voor zijn MSc scriptie hierover dit voorjaar de Jan Hemelrijkprijs van de VvS+OR.

Jolien Ketelaar kijkt samen met Jolien Cremers, Kees Mulder en Irene Klugkist naar de analyse van gegevens die niet alleen een grootte, maar ook een richting hebben en als punten op een cirkelomtrek kunnen worden gezien.

George Boole wordt meestal vereenzelvigd met de naar hem genoemde algebra die ten grondslag ligt aan schakelproblemen in computers. Richard Starmans zet deze pionier in een groter licht dan alleen het werken met 0/1-notaties.

In dit nummer start een nieuwe rubriek waarvan wij hopen dat veel lezers daar aan zullen bijdragen. In deze rubriek vertelt iemand welk boek of artikel een belangrijke rol in zijn/haar loopbaan heeft gehad. Wij zijn verheugd dat Ivo Molenaar een boeiende eerste bijdrage heeft willen leveren met een verhaal over een boek van Lehmann. Schrijft u een volgende bijdrage?

Geld komt weer terug in de column van Henk Tijms, hij schrijft over de waan die gokkers bekruipt wanneer zij niet nadenken over de wetten van de kansrekening. De column van Fred Steutel bevat wat wiskundige aspecten van Sudoku's, en gelegenheidscolumnist Gerrit Stemerdink doet een poging een mopperende oude heer te vertolken wanneer hij een pleidooi doet voor meer, en meer wiskundig, nadenken bij het opzetten van onderzoek.

Ten slotte vindt u ook nog diverse mededelingen over bijeenkomsten op ons vakgebied. Noteert u vast 17 maart 2016 als de datum voor de volgende Dag voor Statistiek en OR?

Wij wensen u een goede jaarwisseling en hopen op veel mooie artikelen voor 2016.

*De STAtOR-redactie*

# VACCINATIEALLOCATIE EN DE KRACHT VAN OPTIMALISATIE

**Voor veel infectieziekten is vaccinatie de meest effectieve manier om een epidemie te voorkomen. Bij onverwachte uitbraken zijn er echter zelden genoeg vaccins beschikbaar om de hele bevolking te vaccineren. Dit resulteert in het uitdagende probleem van vaccinatieallocatie. Door gebruik te maken van technieken uit de besliskunde verkrijgen we inzichten in de structuur van dit allocatieprobleem.**

Evelot Duijzer, Willem van Jaarsveld, Jacco Wallinga & Rommert Dekker

Door de eeuwen heen hebben infectieziekten miljoenen doden veroorzaakt. Verschillende uitbraken van de pest kostten de levens van velen en ook griepepidemieën eisten hun slachtoffers. Vandaag de dag komen zulke immense aantallen slachtoffers gelukkig nauwelijks meer voor en sommige ziekten zijn zelfs geëlimineerd. Niettemin kunnen uitbraken van een nieuw griepvirus of van ziekten als SARS, MERS of Ebola op de loer liggen, zoals we ook recent nog hebben gezien. Onderzoek naar de preventie van een epidemie, maar ook naar de juiste handelwijze in het geval van een uitbraak, heeft daarom hoge prioriteit. Een medische of epidemiologische insteek ligt hierbij het meest voor de hand. Tegelijk komen er ook allerlei logistieke problemen om de hoek kijken. Moeten er bijvoorbeeld voorraden aangelegd worden van vaccins? Zo ja, hoe groot moeten die voorraden zijn en waar moeten ze opgeslagen liggen? Wanneer de middelen beperkt zijn, brengt dat allocatieproblemen met zich mee: hoe kunnen de beschikbare gezondheidswerkers, medicijnen en vaccins het beste worden ingezet? Het zijn precies deze vragen waarbij de kennis en technieken van besliskunde van grote betekenis kunnen zijn.

## Vaccinatieallocatie

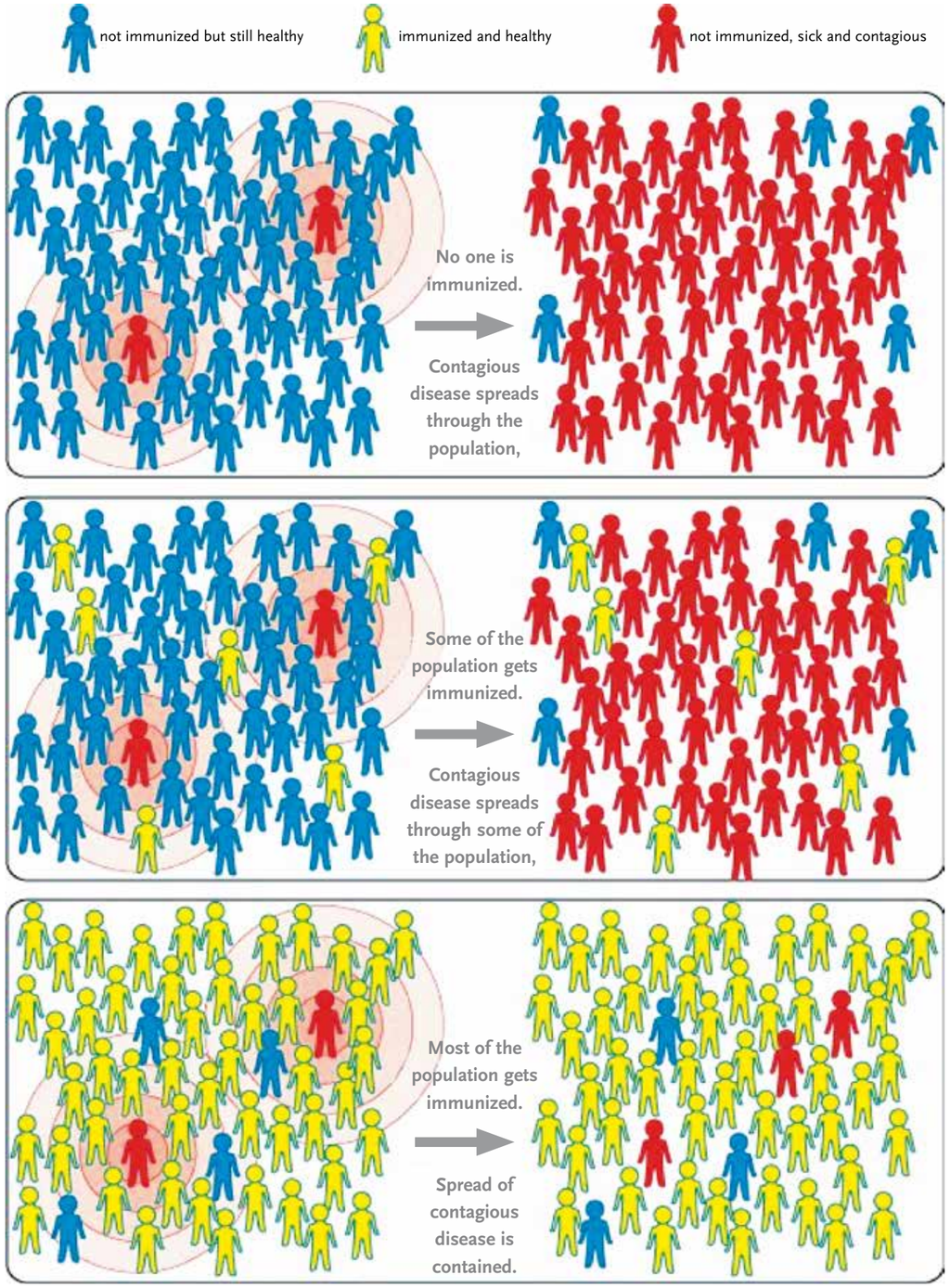
Eén van de meest effectieve manieren om een uitbraak van een infectieziekte te voorkomen, is door mensen te vaccineren. Vaccinatie leidt ertoe dat iemand niet of minder vatbaar is voor de ziekte. Wanneer een deel van de bevolking is gevaccineerd, treden er dus minder besmettingen op. Hierdoor kan een infectieziekte zich minder snel verspreiden en wordt een uitbraak voorkomen of ingedamd. Echter, niet voor alle infectieziekten bestaat er een vaccin. Als er wel een vaccin is, is de beschikbare hoeveelheid bij een onverwachte uitbraak zelden toereikend om de gehele bevolking te vaccineren. Dit brengt een allocatieprobleem met zich mee: hoe kunnen de beschikbare vaccins het best verdeeld worden? Wij bestuderen deze allocatie van vaccins over verschillende groepen in een totale bevolking. Deze bevolkingsgroepen noemen we populaties. Voorbeelden van populaties zijn verschillende steden en dorpen, maar ook de verschillende leeftijdsgroepen kunnen als populaties worden beschouwd.

In de literatuur is het vaccinatieallocatie probleem uitvoerig onderzocht. Veelal wordt met gedetailleerde simulatiemodellen bepaald wat het effect is van een vaccinatieallocatie. Verschillende allocaties worden vervolgens vergeleken en men trekt conclusies over de beste allocatie. Deze aanpak, gebaseerd op scenarioanalyse, is weliswaar informatief maar resulteert niet in een duidelijk verklaring *waarom* bepaalde allocaties beter zijn dan andere. Dit levert vooral problemen op, aangezien de allocaties die goed blijken te zijn vaak tegenintuïtief en oneerlijk zijn (Keeling & Shattock, 2012). Wij stellen daarom voor dit probleem vanuit de optimalisatie te benaderen en analytische methoden te gebruiken om inzichten te krijgen in de structuur van het probleem.

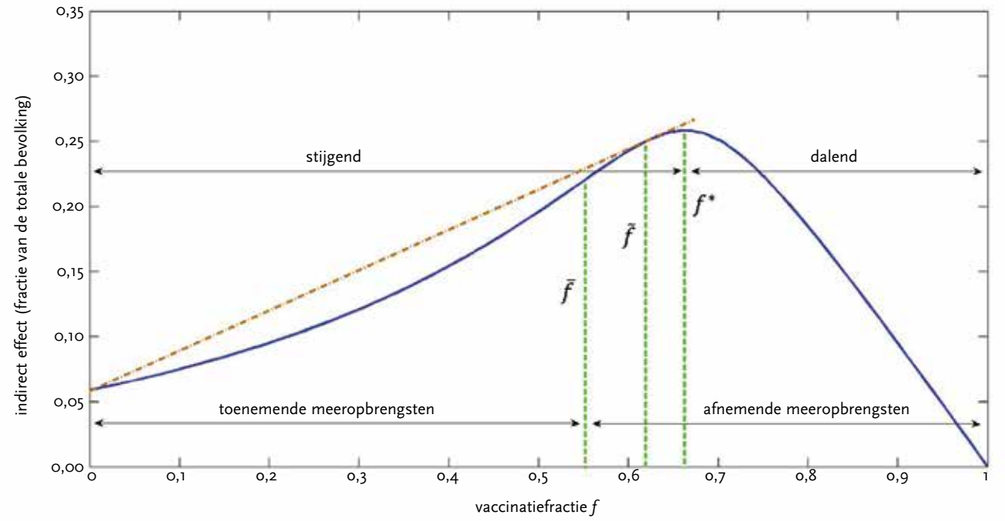
## Groepsimmuniteit en het indirecte effect van vaccinatie

Om verschillende vaccinatieallocaties met elkaar te kunnen vergelijken, bekijken wij in ons onderzoek het effect van vaccinatie. Vaccinatie leidt tot een bepaald niveau van immuniteit, waardoor een gevaccineerd individu een kleinere kans heeft om geïnfecteerd te worden. Hierdoor zijn de gevaccineerde individuen beter beschermd, wat we het *directe effect* van vaccinatie noemen. Daarnaast heeft vaccinatie ook een indirect effect dat te maken heeft met het begrip *groepsimmuniteit*. Door vaccinatie wordt het niet-gevaccineerde deel van de bevolking omringd door gevaccineerde mensen die deels immuun zijn. Mensen die zelf niet gevaccineerd zijn, komen dus minder snel in aanraking met besmette mensen en hebben daardoor zelf ook een lagere kans op besmetting. Dit mechanisme noemen we *groepsimmuniteit*, wat wordt geïllustreerd in figuur 1. Door vaccinatie worden dus twee groepen mensen beschermd tegen de infectieziekte: allereerst zij die zelf zijn gevaccineerd en daarnaast ook diegenen die indirect baat hebben bij de groepsimmuniteit.

Een effectieve vaccinatieallocatie maakt zo goed mogelijk gebruik van het indirecte effect. Op die manier krijg je het meeste waar voor je geld. We zoeken daarom naar de vaccinatieallocatie die het indirecte effect maximaliseert. Daarbij nemen we aan dat we een beperkte




Figuur 1. Als een kritieke fractie van de bevolking immuun is voor een besmettelijke ziekte, zijn de meeste mensen in die bevolking beschermd. Dit verschijnsel heet 'groepsimmuniteit' en is geïllustreerd in de onderste van de drie afbeeldingen. [Bron: National Institute of Allergy and Infectious Diseases (NIAID)]




Figuur 2. De relatie tussen de gekozen vaccinatiefractie en het indirecte effect in een populatie. De verticale gestreepte lijnen geven de drie belangrijke vaccinatiefracties weer: $f^*$ (rechts), $\bar{f}$ (links) en de dosis-optimale vaccinatiefractie $\tilde{f}$ (midden).

hoeveelheid vaccins tot onze beschikking hebben. Om dit indirecte effect te kwantificeren maken we gebruik van het bekende epidemiologische SIR model. Voor de precieze modelspecificaties verwijzen we naar Duijzer et al. Het SIR model bestaat uit een set niet-lineaire differentiaalvergelijkingen waaruit we een impliciete functie afleiden voor het indirecte effect van vaccinatie. Een grondige analyse van deze impliciete functie levert een aantal interessante inzichten op die zijn samengevat in figuur 2. Zoals te zien in de figuur zijn er drie belangrijke vaccinatiefracties:

- $f^*$ de vaccinatiefractie die het indirecte effect maximaliseert,
- $\bar{f}$ de vaccinatiefractie waarvoor de meeropbrengsten van toenemend naar afnemend gaan,
- $\tilde{f}$ de vaccinatiefractie die resulteert in het grootste gemiddelde effect *per* vaccin.

In het bijzonder deze laatste vaccinatiefractie is van belang, omdat vaccineren met deze fractie op de meest efficiënte manier gebruik maakt van de beschikbare vaccins. Deze vaccinatiefractie noemen we daarom de *dosis-optimale* vaccinatiefractie. Afhankelijk van de eigenschappen van een populatie en van de te bestuderen infectieziekte hebben de drie belangrijke vaccinatiefracties verschillende getalswaarden. De volgorde van de drie, zoals gepresenteerd in figuur 2, is echter wel altijd hetzelfde.

## De optimale allocatie

De dosis-optimale vaccinatiefractie speelt een cruciale rol in de optimale allocatie. We kunnen deze optimale allocatie in grote lijnen als volgt karakteriseren: kies een aantal populaties en verdeel alle vaccins hierover zodanig dat de vaccinatiefractie in ieder van de gekozen populaties zo dicht mogelijk ligt bij de *dosis-optimale* vaccinatiefractie van die populatie. Aangezien de dosis-optimale vaccinatiefractie per vaccin het hoogste indirecte effect heeft, probeert de optimale allocatie daar zo goed mogelijk gebruik van te maken. Dit resulteert echter wel in allocaties die mogelijk de ene populatie hogere prioriteit geven dan de andere. Dit kan als volgt worden verklaard. Een kleine hoeveelheid vaccins is in een grote populatie wellicht een druppel op de gloeiende plaat. Dezelfde hoeveelheid kan echter in een kleine populatie het verschil maken tussen het explosief groeien en het direct uitdoven van een uitbraak. Hierdoor is het eerlijk verdelen van de vaccins over alle populaties in sommige gevallen een slecht idee, aangezien er dan in geen enkele populatie daadwerkelijk iets bereikt kan worden.

Bij een kleine hoeveelheid beschikbare vaccins zal de prioriteit daarom liggen bij enkele kleine populaties of bij populaties met een lage dosis-optimale vaccinatiefractie. Voor grote hoeveelheden beschikbare vaccins is het juist mogelijk de dosis-optimale vaccinatiefractie te

bereiken in meerdere populaties of in grote populaties. Hoewel de keuze voor de populaties dus verschilt per hoeveelheid beschikbare vaccins, is de intuïtie achter die keuze wel steeds hetzelfde. Namelijk, het kiezen van de populaties is erop gericht zo goed mogelijk de dosis-optimale vaccinatiefractie te benutten. Voor een gedetailleerde en meer technische beschrijving van de optimale allocatie verwijzen we naar Duijzer et al.

## Conclusie

De allocaties die optimaal zijn met betrekking tot het maximaliseren van het indirecte effect, zijn soms uiterst oneerlijk. Sommige populaties krijgen wel vaccins, terwijl andere populaties volledig ongevaccineerd blijven. Ook in eerder onderzoek, waarbij simulatiemodellen of numerieke analyses gebruikt zijn, kwamen dergelijke oneerlijke resultaten naar voren. Echter, vanuit zulke niet-analytische aanpakken is het lastig deze resultaten uit te leggen. De analytische benadering in ons onderzoek leidt tot het bewijs van het bestaan van drie interessante vaccinatiefracties, waarvan de *dosis-optimale vaccinatiefractie* het belangrijkst is. Dit inzicht draagt eraan bij dat we kunnen uitleggen hoe de oneerlijke, maar optimale allocaties tot stand komen.

In de praktijk spelen eerlijkheidsoverwegingen een belangrijke rol bij het opstellen van een vaccinatieallocatie. Het is daarom niet mogelijk, of zelfs niet wenselijk, onze optimale allocaties direct te implementeren. Verschillende studies constateren hetzelfde probleem: optimaal en eerlijk liggen soms ver uit elkaar. Eén van de oplossingen die hiervoor wordt aangedragen is om de beschikbare hoeveelheid vaccins te splitsen in twee delen. Het eerste deel kan dan op een eerlijke manier verdeeld worden, terwijl het tweede deel gebruikt kan worden om optimaal het indirecte effect van vaccinatie te benutten. Het onderzoek dat wij gedaan hebben, richt zich op deze optimale allocatie. Wij hopen dat onze analytische resultaten, die leiden tot meer inzicht in de effecten van vaccinatie, waardevol zijn in het opstellen van vaccinatieprogramma's. Bovendien laat ons onderzoek zien dat er, naast de klassieke logistieke problemen in de gezondheidszorg, zeker andere interessante terreinen zijn waar besliskunde en optimalisatie van toegevoegde waarde kunnen zijn.


Literatuur

Keeling, M. J., & Shattock, A. (2012). Optimal but unequitable prophylactic distribution of vaccine. *Epidemics, 4*(2), 78–85.

Duijzer, L. E., van Jaarsveld, W. L., Wallinga, J., & Dekker, R. (submitted). Dose-optimal vaccine allocation over multiple populations.



Evelot Duijzer is promovenda bij het Econometrisch Instituut van de Erasmus School of Economics.
E-mail: <duijzer@ese.eur.nl>

Willem van Jaarsveld is universitair docent bij de faculteit Industrial Engineering & Innovation Sciences van de Technische Universiteit Eindhoven.
E-mail: <W.L.v.Jaarsveld@tue.nl>

Jacco Wallinga is hoofd afdeling Modellering van Infectieziekten bij het Rijksinstituut voor Volksgezondheid en Milieu en bijzonder hoogleraar Mathematische Modellering van Infectieziekten bij het Leids Universitair Medisch Centrum.
Email: <Jacco.Wallinga@rivm.nl>

Rommert Dekker is hoogleraar Operations Research, Quantitative Logistics en IT bij het Econometrisch Instituut van de Erasmus School of Economics.
E-mail: <rdekker@ese.eur.nl>


Klassiekers

### VERZOEK

Velen van ons zullen dierbare herinneringen hebben aan een boek dat een grote invloed heeft gehad op hun loopbaan. De redactie denkt dat het delen van zulke herinneringen met andere lezers interessante artikelen kan opleveren.

Wij vragen u daarom eens goed na te denken of er ook in uw loopbaan zo'n boek met grote impact is geweest.

Voor een eerste bijdrage hebben we Ivo Molenaar benaderd; hij reageerde enthousiast en blikt in dit nummer terug op zijn ervaringen met *Testing Statistical Hypotheses* van E.L. Lehmann. Wij hopen op veel vervolgen in de rubriek 'Klassiekers'!

Ons redactielid Richard Starmans ontvangt graag uw reactie: <r.j.c.m.starmans@uu.nl>!

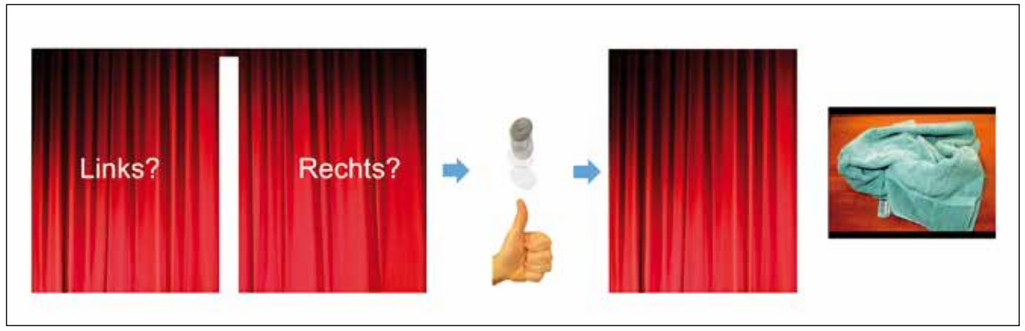


# PARANORMALE STATISTIEK

## over de vele problemen met *p*-waarden, en een redelijk alternatief


Peter D. Grünwald


Het tijdschrift *Science* heeft er een speciale themauitgave aan gewijd en ook *The Economist* vond het een hoofdartikel waard: een schrikbarend hoog percentage wetenschappelijke resultaten is niet reproduceerbaar – met andere woorden: blijkt niet te kloppen. Met name in de geneeskunde en psychologie wordt onderkend dat er sprake is van een *reproducibility crisis*. In de geneeskunde is dit al een aantal jaren duidelijk, vooral door het kritische artikel *Why most published research findings are false* van John Ioannidis in *PLOS Medicine* (2005). Voor de psychologie werd het afgelopen augustus nog eens bevestigd, alweer in *Science*: een grote groep onderzoekers probeerde een groot aantal psychologische studies zo zuiver mogelijk te reproduceren. Zoals in vrijwel alle Nederlandse kranten stond te lezen bleek minder dan de helft van de onderzoeken reproduceerbaar (Open Science Collaboration, 2015). De echte eye-opener binnen de psychologie kwam echter al eerder, toen in 2011 het artikel *Feeling the Future* van Daryl Bem in het prestigieuze *Journal of Personality and Social Psychology* werd gepubliceerd. Daarin toonde hij aan dat mensen in de toekomst kunnen kijken. Bem liet proefpersonen twee gordijntjes zien (figuur 1). Op een gegeven moment zou er aan één kant een plaatje verschijnen, en men moest van te voren raden of dat links of rechts was. De werkelijke positie werd pas na het raden bepaald door een worp



Figuur 1: Het Bem-Experiment

met een munt, en vervolgens onthuld. Dit werd voor iedere proefpersoon vele malen herhaald, met plaatjes van verschillende aard en onafhankelijke muntworpen. In het geval van erotische plaatjes deden proefpersonen het in 53% van de gevallen goed – een onzinnig resultaat, maar gezien de steekproefgrootte toch statistisch significant. Deze paranormale statistiek werd binnen de kortste keren ontkracht (o.a. door Wagenmakers et al. aan de Universiteit van Amsterdam, 2011), maar het zette psychologen wel aan het denken.

## *p*-waarden vs. supermartingalen

Als hoofdoorzaak voor de reproducibility crisis wordt vaak publicatiebias aangegeven.[1] Publicatiebias is zeker belangrijk, maar hier wil ik mij richten op een andere belangrijke, maar minder bekende reden: er kleeft een aantal fundamentele bezwaren aan de standaardmethode voor statistisch onderzoek, het *p*-waarde-gebaseerde nulhypothese toetsen (pHT). De meeste van deze bezwaren zijn al sinds ongeveer 1960 bekend,[2] maar tegen die tijd was pHT al zo wijdverbreid geraakt in de toegepaste wetenschappen dat alle – herhaalde – pogingen tot verandering van paradigma op niets uitliepen. Ik zal hieronder de belangrijkste problemen opsommen en kort aangeven dat er wel degelijk een veelbelovend alternatief bestaat – de *supermartingaalmethode*.

Een supermartingaal meet feitelijk de opbrengst van een gokstrategie. In standaard pHT staat een *kleine p-waarde* voor veel evidentie tegen de nulhypothese; in de supermartingaalmethode staat een *grote (virtuele) financiële winst* voor evidentie tegen de nulhypothese. Dit idee heeft zijn wortels in meer en minder recent werk van V. Vovk (toevallig of niet Kolmogorov's laatste promovendus) en bouwt zelf weer verder op het werk van onder andere de vroeg gestorven Jack Kiefer (1924–1981) (Brownie & Kiefer, 1977; Vovk, 1993; Shafer et al., 2011). Er moet echter nog veel gebeuren om de methode in de praktijk uit te werken – en hier wil ik mij de komende jaren op richten.

## Problemen met de *p*-waarde

Kleine opfrisser. Stel we observeren data $X^N \equiv X_1, \ldots, X_N$. We hebben twee mogelijke verklaringen voor deze data: de nulhypothese $H_o$, die de status quo representeert, en daarnaast de alternatieve hypothese $H_1$. In medische toetsen staat $H_o$ vaak voor 'nieuw geneesmiddel werkt niet', en $H_1$ voor 'geneesmiddel is werkzaam'. In het Bem-voorbeeld hierboven staat $H_o$ voor 'de kans dat iemand raadt waar het plaatje zit is 50%' en $H_1$ voor 'de kans is groter dan 50%' (hoewel er ook iets te zeggen valt voor $H_1'$: 'de kans is ongelijk aan 50%', want als de kans veel kleiner dan 50% is, is er ook iets magisch aan!). Zowel $H_o$ als $H_1$ worden gerepresenteerd als verzamelingen van kansverdelingen. Voor het gemak beperken we ons in deze opfrisser tot het geval dat $H_o = \{P_o\}$ *enkelvoudig* is, dat wil zeggen slechts een enkele verdeling $P_o$ bevat. Een *p*-waarde is een functie $p$ die de verzameling van mogelijke uitkomsten afbeeldt op [0,1], en waarvoor geldt dat

Voor elke $0 \le \alpha \le 1$: $P_o\,(p\,(X^N) \le \alpha\,) \le \alpha$ (1)

De meer gangbare indirecte definitie (bv. op Wikipedia) komt neer op de striktere eis dat (1) met gelijkheid geldt, dus $P_o\,(p\,(X^N) \le \alpha) = \alpha$, maar dit is eigenlijk niet noodzakelijk voor de klassieke toetsingstheorie. Om de definitie verder toe te lichten gebruiken we het *B-voorbeeld* waarbij de $X_i \in \{0, 1\}$ binair zijn, de nulhypothese $H_o$ zegt dat $X_i$ onafhankelijk Bernoulli(1/2)-verdeeld zijn, en de alternatieve hypothese $H_i = \{P_{1,\theta} \mid \theta > 1/2\}$ zegt dat de $X_i$ onafhankelijk Bernoulli ($\theta$) verdeeld zijn voor $\theta > 1/2$. 'B' kan dus staan voor Bem, Bernoulli of Biased Coin.

*p*-waarden kunnen gedefinieerd worden aan de hand van verschillende statistieken van de data, de zogenaamde toetsingsgrootheden. Laten we eerst even aannemen dat het aantal datapunten $N$ vast ligt. In het B-voorbeeld kunnen we bijvoorbeeld als toetsingsgrootheid $T\,(X^N) = \sum_{i=1}^{N} X_i$ het aantal enen in $X^N$ nemen. We zetten $\alpha_t := P_o\,(T(X^N) \ge t)$, en we definiëren vervolgens $p(X^N) := \alpha_{T(X^n)}$ als de kans op een uitkomst die minstens zo extreem is als de uitkomst die we daadwerkelijk hebben waargenomen. Aangezien voor elke $t$ geldt $\{X^N : T(X^n) \ge t\} = \{X^N : \alpha_{T(X^n)} \le \alpha_t\}$ volgt dan ook

$$P_o\,(p\,(X^n) \le \alpha_t) = P_o\,(\alpha_{T(X^n)} \le \alpha_t) = P_o\,(T\,(X^n) \ge t) = \alpha_t$$

zodat (1) geldt met gelijkheid voor alle $\alpha$ die gelijk zijn aan $\alpha_t$ voor een $t$ in het bereik van $T$; dan is eenvoudig in te zien dat (1) zelf (met ongelijkheid) geldt voor alle $0 \le \alpha \le 1$, zodat $p\,(X^N)$ een valide *p*-waarde is. Met behulp van een toetsingsgrootheid $T$ kunnen we dus een *p*-waarde definiëren, en om de toets compleet te maken hebben we nu nog een *onbetrouwbaarheidsdrempel* $\alpha$ nodig; vaak wordt deze op 0,05 gezet. We 'verwerpen' dan de nulhypothese als de waargenomen *p*-waarde $p(X^N) \le \alpha$. De rationale hierachter is dat met deze procedure de *Type-I fout kleiner* is dan $\alpha$. Dit wil zeggen dat, als we een leven lang hypothesen toetsen en steeds dezelfde drempel $\alpha$ aanhouden, we zelfs in het geval dat de nulhypothese altijd waar is, op de lange termijn maar een fractie $\alpha$ van de keren de nulhypothese verwerpen.

Na deze opfrisser kunnen we de nu de grootste problemen met *p*-waarden de revue laten passeren:

### *PROBLEEM 1*
### *Beperkte toepasbaarheid, wegens onbekende kansen*

Om *p*-waarden te gebruiken moet je een toetsingsgrootheid $T$ hebben zodat je de kans $P_o(T(X^N) \ge t)$ kunt uitrekenen.

Je kunt pHT niet gebruiken in vele eenvoudige scenario's waarbij toetsen intuïtief heel wel mogelijk is, terwijl zo'n toetsingsgrootheid niet valt te definiëren. Een extreem simpel voorbeeld is de weersvoorspelling. Tot een aantal jaren geleden gaven de weerman op RTL 4 en de weervrouw op NOS elke dag een 'kans dat het morgen regent'. Je zou willen toetsen wie van de twee dit beter doet. Als de RTL-man consequent hoge regenkans geeft voor dagen waarop de zon blijkt te schijnen en de NOS-vrouw dat niet doet, zou je willen concluderen dat de NOS-vrouw beter is. Het is echter onmogelijk om een *p*-waarde te berekenen onder de nulhypothese 'ze zijn even goed'. Daarvoor moet je namelijk niet alleen weten hoe goed de voorspellers het doen op de data die echt hebben plaatsgevonden, *maar ook hoe goed ze het hebben gedaan in situaties die zich niet hebben voorgedaan*! Je moet dus weten wat de RTL-man voor kans zou voorspellen voor vandaag als het eergisteren geregend zou hebben, ook al scheen in werkelijkheid eergisteren de zon. Dit soort *counterfactual* voorspellingen is intuïtief irrelevant, maar ze zijn nodig om *p*-waarden te berekenen (Dawid, 1984; Grünwald, 2007).

### *PROBLEEM 2*
### *Nog beperkter toepasbaarheid, alsmede onvermijdelijk gesjoemel, wegens onbekende stopregel*

Om *p*-waarden uit te kunnen rekenen moet het experimentele protocol ofwel de *stopregel* volledig bekend zijn bij begin van het experiment. Voorbeelden van zo'n protocol in het B-voorbeeld zijn 'zet $N$ vast op 100' of 'ga door met data verzamelen totdat je voor het eerst achterelkaar drie enen hebt gezien' (dan is $N$ zelf een stochast). *p*-waarden hangen af van de stopregel – in ons B-voorbeeld hangt de waarde van $\alpha_t$ bijvoorbeeld af van $N$; als $N$ van te voren niet bekend is kunnen we $\alpha_t$ dus niet uitrekenen. Maar in de praktijk weten we vaak niet van te voren wat de stopregel is, of willen we het protocol aanpassen aan de hand van onvoorziene omstandigheden; met standaard pHT mag dat niet – terwijl het in de praktijk voortdurend gebeurt. Dit bezwaar wordt vaak weggewuifd – onderzoekers die het protocol aanpasssen worden dan gezien als onkundig of frauduleus, en tegen onkunde en fraude kun je je nu eenmaal niet beschermen. Maar nadere analyse leert dat dit nogal cru is: stel bijvoorbeeld dat je van plan bent een nieuw geneesmiddel aan 30 proefpersonen te toetsen. Je voert de toets uit en vindt een p-waarde van 0,10 – hoopvol maar niet echt overtuigend. Je vertelt dit aan je baas en, volkomen onverwacht, zegt zij dat ze wat extra geld heeft en dat je nog eens 30 proefpersonen tot je beschikking krijgt. Je gaat je geneesmiddel ook op deze 30 mensen te toetsen, maar hoe moet je vervolgens je nieuwe *p*-waarde

uitrekenen? Je zult de neiging hebben ervan uit te gaan dat $N = 60$ van te voren vast lag. Dit is echter niet goed: om de nieuwe $p$-waarde uit te rekenen moet je feitelijk de precieze stopregel weten waarop $N$ gebaseerd is. Je moet dus ook weten of je baas je extra geld zou hebben gegeven als je in eerste instantie $p = 0{,}2$ had geobserveerd; of dat je misschien nog meer geld had gekregen als je $p = 0{,}06$ had geobserveerd. Jij weet dit soort zaken niet en je baas waarschijnlijk ook niet. *Strikt genomen kun je gewoon geen p-waarde meer uitrekenen in bovenstaand scenario!* Varianten van dit scenario treden echter voortdurend op in de psychologische praktijk (John et al., 2012). Zou het niet veel beter zijn om een methode te hebben waarvan het resultaat geldig blijft, hoe je de stopregel ook aanpast ten gevolge van onverwachte ontwikkelingen?

Zowel probleem 1 als probleem 2 komen erop neer dat je, om een $p$-waarde uit te rekenen, kennis moet hebben van *counterfactuals*: wat zou de weerman voorspeld hebben voor $X_t$ als $X_{t+1}$ anders was dan het daadwerkelijk was? En: bij welke $N$ zouden we gestopt zijn met data vergaren als de eerste paar uitkomen anders waren dan ze daadwerkelijk waren? De supermartingaalmethode heeft hier geen last van.

*PROBLEEM 3*
*Geen eenduidige manier om experimenten te combineren*

Stel een nieuw geneesmiddel is in twee ziekenhuizen getoetst – de nulhypothese is, zoals altijd, dat het geneesmiddel niet beter werkt dan het placebo. Ziekenhuis A rapporteert $p$-waarde $p_A$ en ziekenhuis B rapporteert $p$-waarde $p_B$. Natuurlijk zijn de twee patiëntenpopulaties verschillend, maar we zouden toch graag een indicatie willen krijgen van de gezamenlijke bewijskracht van de twee experimenten. Hoe doen we dit? Het meest voor de hand ligt de $p$-waarden te vermenigvuldigen, maar zoals Fisher (1925) al aangaf is dit niet goed: omdat $p$-waarden kleiner dan 1 zijn wordt het resultaat altijd kleiner, wat de oorspronkelijke $p$-waarden ook waren. In werkelijkheid moet je een correctie toepassen. Er zijn meerdere 'correcte' correcties, die echter allemaal een ander antwoord geven. Welke correctie moet je gebruiken? Zou het niet fijner zijn om een methode te hebben waarbij er een uniek-optimale manier (bijvoorbeeld vermenigvuldiging) is om resultaten te combineren?

*PROBLEEM 4*
*Interpretatie*

Naast deze technische problemen is er nog een ander probleem: $p$-waarden zijn notoir moeilijk te interpreteren, zo blijkt bijvoorbeeld uit het onderzoek *What Do Doctors Know about Statistics?* (Wulff et al., 1987) dat ongeveer de helft van alle artsen denkt dat $p < 0{,}05$ betekent dat de kans dat de nulhypothese waar is kleiner is dan 5%, hetgeen betekent dat zij de zogenoemde *Prosecutor's Fallacy* begaan. We zullen hier verder niet op ingaan; zie bijvoorbeeld Wagenmakers (2007) voor een overzicht van deze en vele andere interpretatieproblemen. Zou het niet beter zijn om een methode te hebben met een concretere interpretatie, die zich minder tot drogredenen leent?

## Een redelijk alternatief: de supermartingaalmethode

In dit korte stukje moeten we noodgedwongen afzien van een algemene definitie van supermartingaal toetsen. In plaats daarvan lichten we het idee toe aan de hand van het $B$-voorbeeld: volgens de nulhypothese $H_0 = \{P_0\}$ zijn $X_1, X_2, \ldots$ onafhankelijke Bernoulli(1/2)-variabelen; we nemen nu voor het gemak aan dat volgens de alternatieve hypothese $H_1'$ geldt dat $X_1, X_2, \ldots$ onafhankelijke Bernoulli ($\theta$)-variabelen zijn voor willekeurige $\theta \neq 1/2$.

Stel nu dat we mogen gokken op de uitkomsten $X_1, X_2, \ldots$ op de volgende manier: we hebben een bepaald beginkapitaal $K_0$ en mogen dat vrijelijk verdelen over uitkomst '$X_1 = 0$' en '$X_1 = 1$'. Nadat $X_1$ gerealiseerd is, zeg $X_1 = x$, krijgen we twee keer onze inzet op $x$ uitbetaald; de inzet op de andere uitkomst zijn we kwijt. Dus als we bijvoorbeeld een fractie $r(1)$ van ons kapitaal op $X_1 = 1$ inzetten en een fractie $r(0) := 1 - r(1)$ op $X_1 = 0$, en de uitkomst is 1, dan is ons nieuwe kapitaal $K_1$ na ronde 1 gelijk aan $K_1 = 2r(X_1)K_0$. We mogen nu het nieuwe kapitaal $K_1$ vrijelijk verdelen over de twee mogelijke uitkomsten van $X_2$ en krijgen opnieuw 2 keer onze inzet op de echte uitkomst uitbetaald, terwijl we de inzet op de andere uitkomst verliezen. Als we bijvoorbeeld een fractie $r'(1)$ inzetten op $X_2 = 1$ en $r'(0) := 1 - r'(1)$ op $X_2 = 0$, dan is ons eindkapitaal na ronde 2 gelijk aan $K_2 := 2r'(X_2)K_1 = 4r'(X_2)r(X_1)K_0$. Zo gaat het spel door: we kunnen steeds $K_j$ herverdelen over uitkomst $X_{j+1}$, en krijgen steeds twee keer onze inzet op de daadwerkelijke uitkomst terug, en dit gaat zo door tot aan ronde $N$ als er geen data meer zijn, en we eindigen met eindkapitaal $K_N$. Deze manier van sequentieel-gokken-met-herinzet heet *Kelly gambling* in de economische literatuur. Een *gokstrategie* is een functie die elk initieel segment $X_1, \ldots, X_t$, van elke lengte $t \geq 0$ afbeeldt op een reëel getal $r(1 \mid X^t) \in [0, 1]$ dat aangeeft, gegeven verleden $X_1, \ldots, X_t$, welke proportie van ons tot nog toe verzamelde kapitaal $K_t$ we inzetten op uitkomst $X_{t+1} = 1$. We schrijven dus voortaan $r(X_2 \mid X_1)$ in plaats van $r'(X_2)$, om duidelijk te maken dat de strategie die we hanteren af mag hangen van het verleden. Ons eindkapitaal bij data $X^N$ wordt dan gegeven door $K_N = M(X^N) \cdot K_0$, waarbij

$$M(X^N) := 2^N \prod_{t=1}^{N} r(X_t \mid X^{t-1}) \qquad (2)$$

We noemen de functie $M$, gedefinieerd op data sequenties van willekeurige lengte, de *toets-supermartingaal behorend bij gokstrategie r*. We merken eerst op dat de uitbetaling (een factor 2 van de inzet) neerkomt op een *eerlijke gok* als de nulhypothese waar is. Immers, bij deze uitbetaling geldt dat, wat voor gokstrategie we ook hanteren, ons *verwacht* eindkapitaal onder de nulhypothese precies gelijk is aan ons beginkapitaal.

Als we met een bepaalde gokstrategie juist wél veel geld winnen, is dat dus een indicatie dat de nulhypothese onjuist is. Dit is in het kort waar de supermartingaalmethode op neer komt: bij klassiek toetsen leggen we een *test statistic* (toetsingsgrootheid) vast, die een $p$-waarde bepaalt; hoe kleiner de p-waarde, hoe groter de indicatie dat de nulhypothese onjuist is. In de martingaalmethode leggen we een *gokstrategie* vast; hoe meer geld we daarmee winnen, hoe groter de indicatie dat de nulhypothese onjuist is.

In het Bernoulli-voorbeeld zouden we bijvoorbeeld op tijdstip $t$ kunnen kijken naar de waargenomen frequentie van enen tot nu toe, $\hat{\theta}_t := \sum_{i=1}^{t} X_i/t$. Als dit sterk afwijkt van 1/2 is dit, intuïtief, een indicatie dat de nulhypothese onjuist is en dat er meer winst te behalen valt als we een proportie van ongeveer $\hat{\theta}_t$ van ons geld op 1 zouden zetten. Nu is het gevaarlijk om *precies* $\hat{\theta}_t$ in te zetten: als $\hat{\theta}_t$ gelijk is aan 0 (alleen nullen gezien) of 1 (alleen enen gezien) zouden we al ons geld op 0 resp. 1 zetten, en dus met een bepaalde kans ook al ons geld verliezen. We kunnen ons indekken voor dit risico door, net iets minder agressief, op tijdstip $t$ een proportie van $\breve{\theta}_t = (\sum_{i=1}^{t} X_i + 1) / (t + 2)$ op uitkomst 1 in te zetten. Dit blijkt een zeer effectieve gokstrategie te zijn onder het alternatief $H_1'$: als deze alternatieve hypothese juist is, winnen we met bovenstaande gokstrategie exponentieel veel geld. Meer specifiek, is het eenvoudig om aan te tonen dat ons verzamelde kapitaal $K_N$ op tijdstip $N$ gegarandeerd tenminste

$$M(X^N) = (N + 1)^{-1} \exp(2N(\hat{\theta}_N - 1/2)^2) \qquad (3)$$

keer $K_0$ bedraagt (Grünwald, 2007). Volgens de wet van de grote aantallen zal $\hat{\theta}_N$ naar $\theta$ convergeren, zodat (3) exponentieel stijgt als $\theta \neq 1/2$. De ongelijkheid van Hoeffding impliceert zelfs dat voor elke vaste $\varepsilon > 0$, de kans $P_\theta(|\hat{\theta}_N - \theta| > \varepsilon)$ exponentieel klein is in $N$; als $\theta \neq 1/2$ maken we dus exponentieel veel winst met kans vrijwel 1.

*Opmerkingen*

We mogen dus gokstrategieën gebruiken waarbij onze inzet op tijdstip t van het verleden $X_1, \ldots, X_{t-1}$ afhangt; dit is zelfs cruciaal om winst te kunnen maken als de nulhypothese niet klopt. Maar de gokstrategie *zelf* mag niet afhangen van de data; deze moet feitelijk vaststaan voordat we de data gezien hebben. Als we de gokstrategie namelijk achteraf bepalen kunnen we altijd de triviale strategie gebruiken die al het geld op tijdstip $t$ inzet op de daadwerkelijke uitkomst $X_{t+1}$; dat zou natuurlijk op bedrog neerkomen.

Verder is van belang dat, als we deze methode gebruiken voor een daadwerkelijke hypothese toets, we niet *echt* hoeven te gokken en we dus ook niet *echt* iemand hoeven te vinden die bereid is onze inzetten te accepteren en eventuele winsten uit te betalen:

> *Het gaat om een puur virtueel spel, waarbij we kijken hoeveel winst we zouden maken als we volgens een bepaalde gokstrategie zouden gokken, onder uitbetalingen die in verwachting geen winst (of zelfs verlies) op zouden leveren als de nulhypothese waar was.*

De methode kan zonder problemen worden uitgebreid naar meervoudige nulhypothesen, aftelbaar oneindige en continue uitkomstenruimten en willekeurige kansverdelingen. Om dit te doen moeten we het begrip van een 'eerlijke gokstrategie' veralgemeniseren naar willekeurige $H_0$ en $H_1$. Hoewel dit niet heel moeilijk is voert het te ver voor deze korte inleiding.

We kunnen nu een hoge waarde van een super-martingaal direct interpreteren als 'sterke indicatie dat nulhypothese onjuist is'. Maar als we, zoals in klassiek pNHT, graag grenzen op onze Type-I fout willen, dan kunnen we weer een geld-drempel $A$ instellen, en de nulhypothese verwerpen als $M(X^N) \geq A$. Het mooie is nu dat deze procedure toch weer gerelateerd kan worden aan klassiek pHT: de Type-I fout van zo'n procedure is op zijn hoogst $\alpha := 1/A$, maar in dit geval geldt dit onafhankelijk van de stopregel die $N$ bepaalt:

*STELLING*
*Supermartingaal groot ⇒ Robuuste p-waarde klein*

Laat $M$ een willekeurige supermartingaal zijn (zodat er dus een gokstrategie bestaat zodanig dat, voor alle $n$, $M(X^n)$ het verzamelde kapitaal is op tijdstip $n$ bij beginkapitaal 1 en 'eerlijke' uitbetalingen). Dan geldt (vergelijk met (1)!):

Voor elke $0 \leq \alpha \leq 1$:

$$P_0\left(\text{Er bestaat een } n \text{ met } \frac{1}{M(X^n)} \leq \alpha\right) \leq \alpha. \qquad (4)$$

Deze stelling geldt voor algemene toetsmartingalen en willekeurige $H_1$; hij kan worden uitgebreid naar meervoudige, willekeurige $H_0$. Hij wordt bewezen door Shafer et al. (2011); zie ook Van der Pas & Grünwald (2014). Ze impliceert dat de kans dat er *überhaupt* een $n$ bestaat waarbij de supermartingaal over $A$ heengaat kleiner is dan $1/A$. We kunnen $1/M(X^n)$ dus als een *robuuste* $p$-waarde zien: verwerpen als $1/M(X^n)$ kleiner is dan $1/A$ leidt tot een Type-I fout die onafhankelijk is van de gebruikte stopregel!

## De ontknoping: zijn de *p*-problemen opgelost?

Gewapend met deze kennis kunnen we nu teruggaan naar onze vier problemen hierboven. We beginnen met probleem 1, de beperkte toepasbaarheid. We zien onmiddellijk dat dit probleem verdwijnt als we ons realiseren dat we een weersvoorspelling 'de kans op $X_t = 1$ (regen) is $r_t$' kunnen identificeren met het inzetten van een fractie $r_t$ van het kapitaal $K_{t-1}$ op tijdstip t op de uitkomst $X_t = 1$. Om het eindkapitaal $K_N$ en de martingaal $M(X^N)$ te berekenen hoeven we niet te weten welke data de weersvoorspeller heeft gebruikt om $r_t$ te produceren. Dit kunnen eerdere uitkomsten $X_{t'}$, $t' < t$ zijn, maar ook extra informatie als luchtdruk en temperatuur op heel veel plaatsen in de wereld op tijdstippen $t' < t$. Het is voldoende om $r_1, \ldots, r_t$ te weten! We kunnen dus zonder problemen een supermartingaal-toets uitvoeren.

Wat betreft probleem 2 zien we dat, in tegenstelling tot de $p$-waarde $p(X^N)$, de hoeveelheid geld $M(X^N)$ niet afhangt van de wijze waarop $N$ bepaald is. Ook laat (4) hierboven zien dat een grens op de Type-I fout gegeven kan worden onafhankelijk van hoe $N$ bepaald is. Het probleem is daarmee feitelijk opgelost. Als een onderzoeker extra geld krijgt mag hij dat vrijelijk gebruiken om meer data te vergaren!

Wat betreft probleem 3: stel $M$ is een supermartingaal voor data $X_1, X_2, \ldots$ en $M'$ is een supermartingaal voor data $X'_1, X'_2, \ldots$ Dan is $M''$, met $M''(X_1, \ldots, X_{N_1}, X'_1, \ldots, X_{N_2}) := M(X_{N_1}) \cdot M'(X_{N_2})$ een supermartingaal voor data van de vorm $(X_1, \ldots, X_{N_1}, X'_1, \ldots, X'_2 \ldots)$. We kunnen dit onmiddellijk zien als we ons realiseren dat we het eind-kapitaal $K_{N_1} = K_0 M(X^{N_1})$ na gokken op de eerste data sequentie kunnen gebruiken als beginkapitaal $K'_0$ voor het gokken op de tweede sequentie. We mogen supermartingalen dus gewoon vermenigvuldigen! – anders dan $p$-waarden kunnen ze zowel groter als kleiner dan 1 zijn, zodat ze niet 'vanzelf' naar 0 of oneindig gaan.

Wat betreft probleem 4: in tegenstelling tot de $p$-waarde, die zo'n beetje elke interpretatie verliest als de stopregel niet bekend is, heeft de supermartingaal altijd een hele duidelijke interpretatie in termen van *geld*. Er valt hier nog veel meer over te zeggen, maar dat zal ik hier niet doen.

## Tot slot – de context: toetsmartingalen vs. standaardmartingalen vs. Bayes

De martingaal als wiskundig object werd geïntroduceerd door Ville (1939). Martingalen zijn niet weg te denken uit de moderne kansrekening, maar het gaat dan meestal om additieve martingalen; de multiplicatieve toetsmartingalen worden eigenlijk vrij weinig bestudeerd. Interessant genoeg was Ville (1939) zelf wel degelijk in multiplicatieve toets-achtige martingalen geïnteresseerd. De algemene toets-(super-)martingaalmethode is voor het eerst in volledige vorm geformuleerd door Vovk (1993), maar pas de laatste jaren wordt – door vooralsnog een klein groepje mensen – ingezien hoe krachtig zij is. Vovk zelf gebruikt gokstrategieën niet alleen voor statistiek, maar ook als hoofdbouwsteen voor een fascinerende alternatieve grondslag van de kansrekening, die niet op maattheorie maar op *speltheorie* is gebaseerd (Shafer & Vovk, 2001). Hierbij worden gebeurtenissen $E$ over oneindige rijen (bijv. 'de frequentie van 1en gaat naar 1/2') fundamenteel anders behandeld: een uitspraak als 'gebeurtenis $E$ treedt met kans 1 op' wordt vervangen door het naar mijn smaak meer intuïtieve 'ofwel gebeurtenis $E$ treedt op ofwel met gokstrategie $S$ word je oneindig rijk'. Je doet dus uitspraken die voor *alle* oneindige uitkomstenrijen gelden in plaats van *bijna alle*. Bayesiaanse statistici zullen wellicht vermoeden dat de supermartingaalmethode feitelijk neerkomt op Bayesiaans toetsen. De hierboven gegeven gokstrategie komt neer op Laplace's *rule of succession* en leidt inderdaad tot een Bayesiaanse herinterpretatie van de supermartingaal als de zogenoemde *Bayes factor* in een Bayesiaanse analyse met homogene a priori verdeling over de verdelingen in $H_1$. Inderdaad kan bij enkelvoudige (maar niet meervoudige!) $H_0$, de Bayes factor altijd geherinterpreteerd kan worden als een supermartingaal; omgekeerd is echter niet het geval. Zo vormt bijvoorbeeld de *switching strategy* van Van Erven et al. (2011)en Van der Pas & Grünwald (2014) een gokstrategie die, ook met enkelvoudige $H_0$, niet direct een Bayesiaanse interpretatie heeft, en ten opzichte van Bayesiaanse aanpakken een significant kleinere Type-II fout oplevert. Een precieze vergelijking van Bayes met supermartingalen zou meerdere pagina's vergen!

Deze tekst is gebaseerd op een lezing gehouden tijdens de Nederlandse Wiskunde Dagen 2015. Dank aan mijn co-auteurs Stèphanie van der Pas en Eric-Jan Wagenmakers.

1. In plaats van uit te leggen wat publicatiebias is, verwijs ik liever naar de cartoon https://xkcd.com/882/ van het onvolprezen xkcd.com: een plaatje zegt hier meer dan 1000 woorden.
2. Bayesiaanse statistici publiceerden een aantal invloedrijke kritieken op pHT begin jaren '60 (Edwards et al., 1963, Pratt, 1962); ze richtten zich daarbij vooral op probleem 2 en 4 hieronder. Probleem 1 staat centraal in het werk van A.P. Dawid (1984). Probleem 3 is al bekend sinds het werk van Fisher (1925), maar die zag het niet als een probleem.

Literatuur

Brownie, C., & Kiefer, J. (1977).The ideas of conditional con_dence in the simplest setting. *Communications in Statistics – Theory and Methods*, 6(69), 691–751.

Dawid, A. P. (1984). Present position and potential developments: Some personal views, statistical theory, the prequential approach. *Journal of the Royal Statistical Society, Series A*, 147(2), 278–292.

Edwards, W., Lindman, H., & Savage. L. J. (1963).Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193–242, 1963.

Fisher, R. (1925). *Statistical Methods for Research Workers*. Guildford UK: Genesis Publishing, 1925.

Grünwald, P. (2007). *The Minimum Description Length Principle*. Cambridge: MIT Press.

Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8). DOI:10.1371/journal.pmed.0020124. PMC 1182327.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science, 23*(5).

Open Science Collaboration, (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Pratt, J. W. (1962).On the foundations of statistical inference: Discussion of Birnbaum's paper. *Journal of the American Statistical Association*, 57, 314–315.

Shafer, G., Shen, A., Vereshchagin, N., & Vovk, V. (2011). Test martingales, Bayes factors and p-values. *Statistical Science*, 26(1), 84–101.

Shafer, G., & Vovk, V. (2001). *Probability and Finance; It's Only a Game!* New York: Wiley.

Pas, S. van der, & Grünwald, P. D. (2014). *Almost the best of three worlds: Risk, consistency and optional stopping for the switch criterion in single parameter model selection*. Preprint available on arXiv as arXiv:1408.5724.

Erven, T. van, Grünwald, P., & and Rooij, S. de (2011). Catching up faster by switching sooner: A predictive approach to adaptive estimation with an application to the AIC-BIC dilemma. *Journal of the Royal Statistical Society*, *Series B*, 74(3), 361–397 (with discussion).

Ville. J. (1939). Étude critique de la notion de collectif. *Monographies des Probabilités, 3*.

Vovk. V. G. (1993). A logic of probability, with application to the foundations of statistics. *Journal of the Royal Statistical Society, series B, 55*, 317–351 (with discussion).

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin and Review, 14*(5), 779–804.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Maas, H.L.J. van der (2011). Why psychologists must change the way they analyze their data: The case of Psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*, 426–432.

Wulff, H. R., Andersen, B., Brandenhoff, P., & Guttler, F. (1987). What do doctors know about statistics? *Statistics in medicine, 6*(1), 3–10.

Peter Grünwald is senior onderzoeker aan het Centrum Wiskunde & Informatica te Amsterdam en hoogleraar Statistisch Leren aan de Universiteit Leiden. In 2010 ontving hij, samen met Harry van Zanten, de Van Dantzigprijs van de VvS+Or, de hoogste Nederlandse onderscheiding op het gebied van statistiek en OR. Hij is auteur van *The Minimum Description Length Principle* (MIT Press, 2007), een boek over data analyse met behulp van data compressie, gerelateerd aan de supermartingaalmethode.
E-mail: <pdg@cwi.nl>

# DYNAMISCH AMBULANCE MANAGEMENT

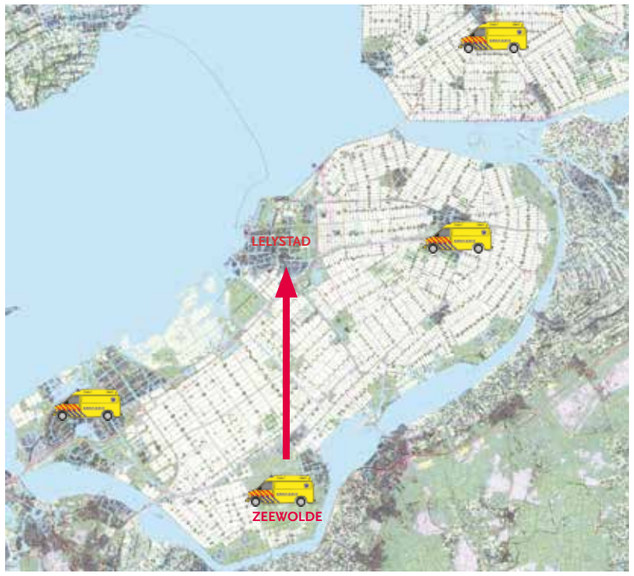
## Mensenlevens redden met operations research

**In noodsituaties waarin elke seconde telt, is het tijdig ter plaatse zijn van een ambulance van levensbelang. Een veelbelovende manier om de aanrijtijden van ambulances te verkleinen is Dynamisch Ambulance Management, waarbij ambulances geen vaste standplaats hebben, maar slim en dynamisch over de regio kunnen worden verspreid afhankelijk van de voertuig- en incidentlocaties. In het kader van het project 'Van reactieve naar proactieve planning van ambulancediensten', kortweg REPRO, hebben CWI en TU Delft nieuwe methoden ontwikkeld voor een optimale dynamische spreiding van ambulances over een verzorgingsgebied. De resultaten zijn veelbelovend en worden momenteel uitgetest in een pilot in samenwerking met GGD Flevoland.**

Caroline Jagtenberg & Rob van der Mei

In levensbedreigende situaties waarin elke seconde telt, kan het al dan niet op tijd ter plaatse zijn van een ambulance het verschil maken tussen leven en dood. In deze gevallen is het streven om zo snel mogelijk en bij voorkeur binnen 15 minuten op de plaats van de patiënt te zijn. Om zulke korte responstijden te realiseren is een uitgekiende planning van ambulanceritten noodzakelijk. Deze planning kan zowel op strategisch, tactisch als operationeel niveau worden gemaakt. In het geval van operationele planning gaat het om de volgende vraag: hoe kunnen we op elk moment van de dag een goede dekking van ambulances over een regio realiseren door slimme dynamische en proactieve *real-time* herpositionering van ambulances?

Binnen het STW-project REPRO, een samenwerking tussen CWI, TU Delft, CityGIS en een vijftal ambulancediensten, wordt niet alleen antwoord gegeven op deze vraag, maar wordt het antwoord ook toegepast in de praktijk. In de traditionele ambulancedienstverlening heeft elk ambulancevoertuig een vaste standplaats: elke wagen rijdt, wanneer deze vrijkomt, terug naar zijn eigen thuisbasis. Tegenwoordig wordt dit echter vaak gezien als een verouderde aanpak, en beseft men dat het terugsturen van de ambulance naar zijn thuisbasis niet altijd de beste keuze is. Afhankelijk van de situatie op dat moment, bijvoorbeeld de locaties en status van de huidige incidenten en van de andere voertuigen, kunnen de vrije ambulances wellicht efficiënter over de standplaatsen worden verdeeld. Dit is de kern van het Dynamisch Ambulance Management, kortweg DAM. Zie figuur 1.




Figuur 1. Illustratie van Dynamisch Ambulance Management in Flevoland: een ambulancevoertuig wordt van Zeewolde naar Lelystad gestuurd.

### De huidige situatie

In Nederland is een verandering in paradigma merkbaar: steeds vaker wordt de klassieke 'statische' planning voor een, al dan niet door de ambulancevoorzieningen zelf ontworpen, dynamische planning ingeruild. Meestal wordt daarbij gebruik gemaakt van een zogenaamde schuifregeltabel: een overzicht dat voor elk aantal (vrije) wagens aangeeft op welke basislocaties zij zich dienen te bevinden. Vervolgens is het aan de meldkamercentralist om één of meer ambulancevoertuigen zodanig te dirigeren dat de gewenste opstelling bereikt wordt.

De schuifregeltabellen zijn veelal gebaseerd op enkelvoudige dekking van de regio, waarbij ambulances zodanig zijn verspreid over de regio dat de kans dat het eerstvolgende incident binnen een straal van 15-minuten valt is gemaximaliseerd. Het probleem is echter dat in de praktijk in veel gevallen meerdere incidenten tegelijkertijd moeten worden afgehandeld, waardoor een dergelijke eenvoudige schuifregeltabel niet goed functioneert. Langzaam maar zeker worden ambulance aanbieders zich er van bewust dat een meervoudige dekking gewenst is, maar de vraag hoe deze gerealiseerd kan worden hebben zij daarmee nog niet beantwoord.

De vraag is: zou een schuifregeltabel met meervoudige dekking dan de beste oplossing zijn? Verrassend genoeg is het antwoord 'nee'. Het gebruik van een schuifregeltabel in het algemeen brengt beperkingen met zich mee. Schuifregels houden namelijk geen rekening met de actuele locaties van de voertuigen. Bovendien geldt voor gebieden met veel incidenten dat het aantal vrije ambulances zeer snel kan wijzigen, vaak zo snel dat de situatie al wijzigt voordat de ambulances de nieuwe gewenste configuratie hebben bereikt. Daarnaast vertelt een schuifregel niet hoe de wagens moeten worden verplaatst om de gewenste configuratie te bereiken.

### Voor- en nadelen

Het grote voordeel van DAM is dat door de proactieve verplaatsingen van voertuigen veel flexibeler kan worden geanticipeerd op toekomstige incidenten, waardoor de responstijden kort kunnen worden gehouden. Verbeterde responstijden betekenen een reductie in de mortaliteit en morbiditeit. DAM leidt dus direct tot een betere kwaliteit van zorg. De verbetering is bovendien van substantieel formaat. Om dezelfde prestaties met een statische planning te bereiken, zouden meerdere extra ambulances (en bijbehorend personeel) bekostigd moeten worden.

Een veelgenoemd nadeel van DAM is dat ambulances

vaker moeten verplaatsen, waardoor de werkdruk verhoogt. Ook heeft personeel geen vaste standplaats meer en moet daar aan wennen.

Ondanks de genoemde nadelen is de verwachting dat DAM de toekomst heeft. Natuurlijk kan er wel rekening worden gehouden met bijvoorbeeld de werkdruk. Binnen REPRO is een model ontwikkeld waarbij ambulances in eerste instantie niet vaker moeten rijden dan in het statische geval, maar enkel naar andere bases Dit model kan op verzoek van een ambulance-aanbieder uitgebreid worden met extra ambulancebewegingen, die de prestaties nog verder ten goede komen.

## Wat een toestand(en)

Om een geschikte DAM-methode te vinden, wordt binnen het REPRO-project onderzoek verricht naar online optimaliseringsalgoritmen, waarbij optimale relocaties worden bepaald, bijvoorbeeld voor iedere mogelijke 'toestand' van het systeem (zoals incident- en voertuiglocaties). Het probleem hierbij is dat het aantal mogelijke toestanden van het systeem exponentieel snel stijgt in het aantal voertuigen en incidentlocaties, waardoor de rekentijden voor het bepalen van 'optimale' relocaties van voertuigen veel te lang worden voor praktische toepassingen.

Idealiter wil men natuurlijk een algoritme dat in real time met een zinnig antwoord komt, en dat bovendien inzetbaar is voor uiteenlopende regio's, elk met hun eigen karakteristieken. Uit onderzoek binnen REPRO blijkt dat het inderdaad mogelijk is om dergelijke online optimaliseringsalgoritmen te ontwikkelen. Een voorbeeld is het Dynamic Maximum Expected Coverage Location Problem (D-MEXCLP) algoritme, geïnspireerd door het MEXCLP-model (Daskin, 1983), waarin de bezettingsgraad q wordt meegenomen voor optimaliseren van meervoudige dekking.

Het D-MEXCLP-algoritme (Jagtenberg, Bhulai & Van der Mei, 2015) beperkt zich in eerste instantie tot het verplaatsen van een ambulance wanneer deze zich vrij meldt. Hierbij wordt slechts een beperkt deel van de realtimedata meegenomen, namelijk enkel de bestemmingen van andere vrije ambulances. Daarmee wordt als het ware een soort snapshot van de toekomst genomen: de rijdende ambulances zullen daar waarschijnlijk binnen afzienbare tijd aankomen. Dit snapshot wordt afgezet tegen de verwachte behoefte aan ambulances.

We bepalen de gewenste locatie voor de zojuist vrijgekomen ambulance door te berekenen op welke standplaats de ambulance de grootste toename in dekking zou geven volgens het MEXCLP model. MEXCLP modelleert ambulances als onafhankelijke voertuigen die met kans q bezet zijn. De verzorgingsregio wordt opgedeeld in een discrete set punten V. Merk op dat de k-de ambulance die vraagpunt $j \in V$ op tijd kan bereiken, pas waarde toevoegt als de overige k-1 ambulances bezet zijn, terwijl de k-de ambulance vrij is – oftewel met kans $(1-q)q^{k-1}$. Deze kans wordt vermenigvuldigd met $d_j$ de grootte van de vraag in punt j, $j \in V$.

De zo gedefinieerde strategie $\pi(n)$, met $n = (n_1, n_2, \ldots, n_{|\mathbb{W}|})$ kan genoteerd worden aan de hand van het aantal vrije ambulances $n_x$ dat bestemming x heeft. Merk op dat vrije ambulances altijd stilstaan op een basis, of daar naar onderweg zijn, dus het volstaat om $x \in \mathbb{W}$ te beschouwen, waarbij $\mathbb{W}$ de set basislocaties is. Voor een gegeven voertuigconfiguratie n stuurt het D-MEXCLP-algoritme een vrijkomende ambulance naar de bestemming $w \in \mathbb{W}$ waarvoor de verwachte toename in de totale dekkingsgraad

$$\sum_{j \in V} d_j (1-q) q^{k(j,w,n)-1}$$

maximaal is, waarbij

$$k(j,w,n) := \sum_{i \in W} n_i \mathbb{1}_{\{t_{ij} \leq r\}} + \mathbb{1}_{\{t_{wj} \leq r\}}$$

het aantal voertuigen is dat knoop j kan bereiken binnen r minuten, voor gegeven voertuigconfiguratie n, aannemende dat de vrijkomende ambulance instantaan op de knoop van bestemming $w \in \mathbb{W}$ is. Doordat de keuzemogelijkheden beperkt zijn (in de praktijk is $|W|$ meestal niet groter dan 30 of 40), kunnen we de optimale bestemming *brute force* uitrekenen.

## Verificatie

Het is niet direct evident of het hierboven beschreven algoritme succesvol zal zijn. Er zijn namelijk veel aspecten van het probleem waarvan het niet triviaal is hoe ze meegenomen dienen te worden (denk bijvoorbeeld aan gemaakte keuzes zoals het negeren van bezette wagens, het modelleren van vrije wagens als onafhankelijk, en het nemen van een snapshot van de toekomst). Alvorens het beschreven model in de praktijk wordt toegepast, dient geverifieerd te worden of de gemaakte keuzes leiden tot een succesvol DAM-algoritme.

Aan de hand van simulaties kunnen uitspraken gedaan worden over hoe goed het D-MEXCLP-algoritme zal presteren in de praktijk. De simulaties zijn gedaan voor een realistische dataset van de provincie Utrecht. De resultaten laten zien dat door het gebruik van D-MEXCLP het aantal rijtijdoverschrijdingen met 15-20% kan worden gereduceerd ten opzichte van de statische oplossing (Jagtenberg, Bhulai & Van der Mei, 2015).

## Toepassing in de praktijk

Het hierboven beschreven model voor DAM wordt momenteel uitgetest in een pilot die wordt uitgevoerd in samenwerking met GGD Flevoland, CityGIS en het CWI. Omdat Flevoland een relatief dunbevolkte regio is, is in overleg met GGD Flevoland besloten om extra relocatiemomenten toe te voegen: wanneer een ambulance naar een incident gestuurd wordt, mag bovendien één vrije wagen naar een andere basis gestuurd worden. Er wordt dan op een vergelijkbare manier bepaald welke verplaatsing de grootste verbetering in dekking oplevert – alleen moet nu naast de bestemming, ook de ambulance gekozen worden.

De relocatievoorstellen worden als advies getoond aan de centralisten. Het is dan aan de centralist om een afweging te maken tussen de toename in bedekking die met de relocatie kan worden bereikt, en eventuele andere belangen die op dat moment spelen. Om deze afweging goed te kunnen maken, wordt naast het relocatievoorstel ook het bijbehorende 'belang' getoond: een getal dat aangeeft in hoeverre de dekking van de regio er op vooruit zal gaan als deze relocatie wordt uitgevoerd. Daarmee krijgt het advies nuance en kunnen keuzes onderbouwd worden gemaakt.

Caroline Jagtenberg is promovendus, werkzaam aan het CWI, en onderzoekt verschillende aspecten van de operationele ambulance planning. Daarnaast is ze in deeltijd software engineer bij ORTEC, waar zij werkt aan personeelsroosteren.
E-mail: <C.J.Jagtenberg@cwi.nl>

Rob van der Mei is Manager Research & Development aan het CWI, hoogleraar aan de Vrije Universiteit Amsterdam en specialist op het gebied van de toegepaste wiskunde. Hij is de projectleider van REPRO.
E-mail: <R.D.van.der.Mei@cwi.nl>

Literatuur

Daskin, M. S. (1983). A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, 17(1), 48–70.

Jagtenberg, C. J., Bhulai, S., & Van der Mei, R. D. (2015). An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care, 4,* 27–35.



Fred Steutel **column**

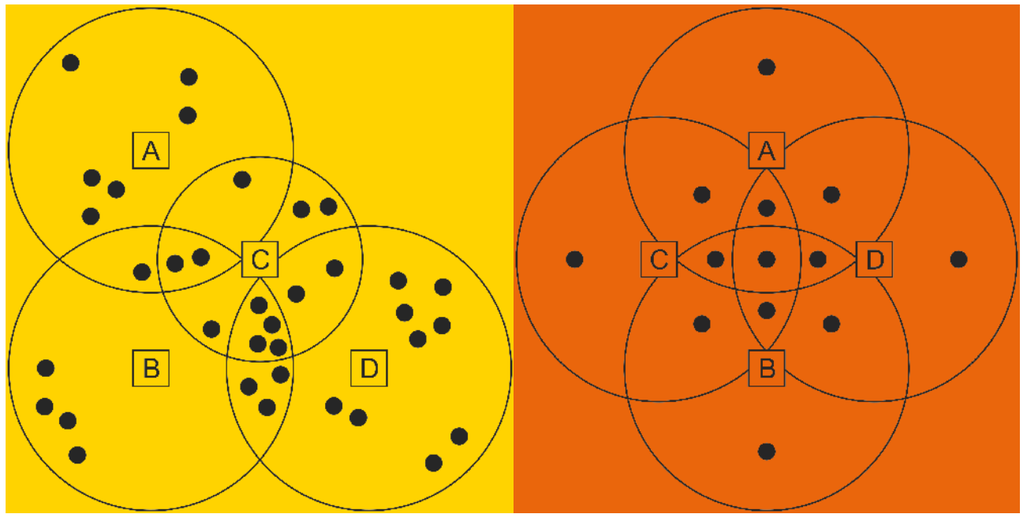
# Sudokustatistiek

'Sudoku's hebben niets met wiskunde te maken' lees ik op een website. 'Dat maken wij wel uit', denk ik dan. Ten eerste kun je het aantal mogelijke sudoku's tellen. Dat is gelukkig al gedaan, en dat blijken er 6,671 maal tien tot de eenentwintigste te zijn; dit getal is iets kleiner dan het getal van Avogadro, maar toch heel groot. Van de *NRC*-sudoku in de NRC die ik elke dag maak, zijn er wat minder: elke *NRC*-sudokuoplossing is ook een gewone oplossing, maar omgekeerd niet. Q.E.D.

Ze zeggen ook dat je de cijfers 1 tot en met 9 best kunt vervangen door andere symbolen, bijvoorbeeld door negen verschillende bloemetjes of vlinders. Dat is maar zeer gedeeltelijk waar, want de cijfers hebben een *ordening*. Als ik met de *NRC*-sudoku klaar ben, kijk ik altijd of er opeenvolgende drietallen in voorkomen, horizontaal of verticaal. Dat is bijna altijd het geval, dikwijls meer dan eens. Ik schat dat het aantal opvolgende drietallen per sudoku ongeveer Poissonverdeeld is met een gemiddelde van drie. De kans op geen drietal zou dan ongeveer 0,05 zijn, maar lijkt in de praktijk kleiner: je treft (vrijwel?) nooit een sudoku zonder opeenvolgend drietallen. De Poisson-verdeling zal dus niet helemaal kloppen. Misschien valt te bewijzen dat (niet?) elke *NRC*-sudoku minstens één opeenvolgend drietal bevat. Opdracht voor het volgende nummer?

Een bijzondere eigenschap van de *NRC*-sudoku is de volgende: de negen plaatsen gevormd door de vier hoekpunten, de vier middens van de zijden en het centrum, bevatten ook altijd de cijfers van 1 tot en met 9. Niet iedereen schijnt dat te weten, maar je kunt daar bij het oplossen gebruik van maken. Gewone sudoku's hebben deze eigenschap niet.

Fred Steutel is emeritus hoogleraar kansrekening aan de TU Eindhoven.
E-mail: <fsteutel@xs4all.nl>



Figuur 1. Twee instanties van het locatieprobleem: (links) a. asymmetrische instantie; (rechts) b. symmetrische instantie.

# IS SYMMETRIE SLECHT?

Rutger Kerkkamp

Als u zou moeten kiezen welk figuur mooier is: figuur 1a of figuur 1b, wat zou dan uw keuze zijn? Hoogstwaarschijnlijk heeft u voor figuur 1b gekozen, vanwege de eenvoud en symmetrie. Als mens zijn we voortdurend op zoek naar patronen in ons leven, in het bijzonder naar regelmaat en symmetrie. Symmetrie wordt vaak beschouwd als een goede eigenschap, een schoonheid. Wiskundigen zijn over het algemeen gevoeliger voor deze (abstracte) schoonheid, zij het aangeboren of aangeleerd door onze universele taal. Maar is symmetrie eigenlijk wel gewenst? Is het altijd 'goed' (gewenst) of kan het ook 'slecht' (ongewenst) zijn?

Om deze vraag te beantwoorden, of een poging daartoe te doen, beschouwen we een bekend optimalisatieprobleem in de besliskunde: het plaatsen van ambulancebases (standplaatsen) om een zo groot mogelijk gebied op tijd te kunnen bereiken. In Nederland groeit de belangstelling naar en het gebruik van wiskundige optimalisatietechnieken om inzicht te verkrijgen in de (toekomstige) situatie van de Nederlandse ambulancezorg. Dit betreft ook de keuze van de locaties van ambulancebases, die vaak door historie of door niet-wiskundige expertise bepaald zijn.

## Het kiezen van ambulancestandplaatsen

Vanwege de medische urgentie bij ongelukken bestaat er een afgesproken maximale tijd tussen de melding van een ongeluk en de aankomst van de ambulance ter plaatse. Als deze tijd niet gehaald wordt, dan wordt de respons van de ambulancevoorziening als 'te laat' beschouwd. In Nederland geldt voor meldingen met hoge spoed een maximale tijd van 15 minuten.

Dit leidt ertoe dat elke ambulancebasis een gebied heeft dat op tijd bereikt kan worden vanuit die basis. In figuur 1 worden de bases weergegeven door de vierkanten met een letter. De open cirkels om elk vierkant geven het op tijd bereikbare gebied aan en de zwarte punten zijn locaties waar ongelukken kunnen plaatsvinden. Het locatieprobleem vraagt om een gegeven aantal bases te kiezen zodanig dat de meeste punten op tijd bereikt kunnen worden. De punten kunnen ook een bepaald gewicht hebben, waardoor sommige punten belangrijker zijn dan andere. In dat geval moet het totale gewicht van de bereikbare punten gemaximaliseerd worden.

Wiskundig gezien is dit locatieprobleem moeilijk om op te lossen (NP-moeilijk), maar in de praktijk zijn de problemen relatief eenvoudig oplosbaar met behulp van een scala aan methodes. Wij lichten een methode uit: de Swap Local Search, een heuristiek die over het algemeen goed presteert. Dit algoritme begint met een (willekeurige) startoplossing en verbetert deze iteratief door een deel van de geopende bases te vervangen met andere bases. Als de huidige oplossing niet verbeterd kan worden op deze wijze stopt het algoritme.

Hoeveel bases er per iteratie maximaal uitgewisseld kunnen worden, is een belangrijke ontwerpkeuze. In de praktijk kiest men vaak om 1 tot 3 bases per iteratie te veranderen. Merk op dat deze methode niet gegarandeerd het optimum vindt, tenzij alle geopende bases tegelijk vervangen kunnen worden. Het is daarom van belang om te bepalen hoe goed de Swap Local Search kan presteren. Oftewel, voor welke instanties is de oplossing van dit algoritme het slechtst ten opzichte van de optimale oplossing?

## Kwaliteitsgaranties voor de oplossing

Stel we passen de Swap Local Search-methode toe op het locatieprobleem in figuur 1a met bepaalde gewichten voor de punten. Om a priori garanties te kunnen geven voor de kwaliteit van de Swap Local Search-oplossing moet bepaald worden wat er in het slechtste geval kan plaatsvinden. Dit blijkt mogelijk te zijn door eerst een gerelateerde instantie op te lossen.

We kunnen bewijzen dat er een symmetrische instantie bestaat waar de heuristiek slechter presteert dan voor de instantie in figuur 1a. Het zal u waarschijnlijk niet verbazen dat deze symmetrische instantie de instantie in figuur 1b is. Deze instantie wordt geconstrueerd door middel van een transformatie die een specifiek gedefinieerd gemiddelde neemt van alle gewichten. Het idee van de transformatie is dat elk punt geclassificeerd kan worden aan de hand van de bases van waaruit dit punt op tijd bereikt kan worden. Bovendien zijn de namen van de bases arbitrair en kunnen gepermuteerd worden. De permutaties leveren meerdere equivalente instanties op, waarvan we een bepaald gemiddelde kunnen nemen dat rekening houdt met de classificatie van de punten. Het resultaat is een symmetrische instantie die niet meer afhangt van de namen van de bases, maar van de classificatie van de punten.

Het bepalen van de slechtst mogelijke prestatie van de heuristiek komt neer op het oplossen van een lineair programmeringsprobleem, waarbij de restricties modelleren dat we Swap Local Search-oplossingen beschouwen. Voor algemene instanties is deze aanpak niet schaalbaar vanwege een exponentieel aantal variabelen en restricties. Voor symmetrische instanties is dit aantal echter polynomiaal in de instantiegrootte.

Kortom, vanwege de symmetrie van de nieuwe instantie is eenvoudig te bepalen hoe slecht de heuristiek maximaal kan presteren. Het feit dat de symmetrische instantie gerelateerd is aan de originele instantie leidt tot a priori garanties voor de oplossing van het originele probleem.

## Goed of slecht?

Het bewijs van deze garantie toont aan dat de Swap Local Search het slechtst presteert voor een familie van symmetrische instanties. Om terug te keren naar onze vraag of symmetrie slecht kan zijn, is dit een voorbeeld waar symmetrie ongewenst is.

De oplettende lezer zou opmerken dat we geen eisen hebben gesteld aan de startoplossing voor de heuristiek. Een logische gedachte is om deze startoplossing te vinden met een heuristiek die complementair is aan de Swap Local Search. Dat wil zeggen, ze vangen elkaars zwakheden op. Een bekend algoritme vervult deze rol: de Greedy Search, waar bases een voor een geopend worden zodanig dat per iteratie de toename van het bereikbare gebied gemaximaliseerd wordt. In ons geval vindt de Greedy Search direct het optimum voor deze symmetrische instanties. Misschien is symmetrie toch niet zo slecht als gedacht ...

### Meer informatie

De details van het bewijs voor de kwaliteitsgarantie zijn subtieler dan hier beschreven. Zie voor het volledige bewijs paragraaf 4.4 van *Facility location models in emergency medical service* van Kerkkamp (2014). Vanwege de simpliciteit van Swap Local Search en Greedy Search zijn er geregeld kwaliteitsgaranties te bepalen als deze methodes toegepast worden op optimalisatieproblemen, zie bijvoorbeeld Williamson & Shmoys (2011).

Dit onderzoek is uitgevoerd als onderdeel van het REPRO project, een samenwerking tussen CWI en TU Delft geleid door Rob van der Mei. Dit project richt zich onder andere op de plaatsing van ambulancebases (Van den Berg & Aardal, 2015), de dynamische bijsturing van ambulances (Jagtenberg et al., 2015) en de simulatie van de ambulancezorg (Van Buuren et al., 2012).

### Literatuur

Jagtenberg, C., Bhulai, S., & Van der Mei, R. D. (2015). An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care, 4*, 27–35.

Kerkkamp, R. B. O. (2014). *Facility location models in emergency medical service: robustness and approximations.* Delft University of Technology. Beschikbaar op <repository.tudelft.nl>.

Van den Berg, P. L., & Aardal, K. (2015). Time-dependent MEXCLP with start-up and relocation cost. *European Journal of Operational Research, 242*(2), 383–389.

Van Buuren, M., Aardal, K., Van der Mei, R. D., & Post, H. (2012). Evaluating dynamic dispatch strategies for emergency medical services: TIFAR simulation tool. In *Proceedings of the Winter Simulation Conference (*pp. 509–519).

Williamson, D., & Shmoys, D. (2011). *The design of approximation algorithms*. New York: Cambridge University Press.

Rutger Kerkkamp is promovendus aan de Erasmus Universiteit Rotterdam en onderzoekt voorraadbeheerstrategieën als informatie asymmetrisch gedeeld wordt onder de betrokken partijen. In maart 2015 heeft de VvS+OR hem de Jan Hemelrijkprijs uitgereikt voor zijn MSc-scriptie aan de TU Delft.
E-mail: <kerkkamp@ese.eur.nl>

KLASSIEKERS

# LEHMANN OVER HYPOTHESETOETSING

Ivo W. Molenaar

In 1959 publiceerde Wiley *Testing Statistical Hypotheses* van Erich Leo Lehmann (1917-2009). Dat boek heb ik een kleine tien jaar later intensief bestudeerd in een groepje van een tiental statistici, vooral afkomstig van het Mathematisch Centrum (later CWI) en de Universiteit Utrecht. Dat vond ik leuk en leerzaam, zoals ik hieronder ga toelichten. Pas veel later heb ik de auteur persoonlijk ontmoet, en ook dat was leuk en leerzaam.

Lehmann werd geboren in Straatsburg en groeide op in Frankfurt am Main, tot het Joodse gezin in 1933 uitweek naar Zwitserland. Hij studeerde twee jaar in Cambridge UK en emigreerde in 1940 naar de VS, waar hij in 1946 een PhD behaalde bij Jerzy Neyman in Berkeley. Hoewel hij wel eens korte tijd elders verbleef (Guam, New York, Princeton, Stanford) speelde het overgrote deel van zijn lange leven zich af op de campus van de University of California te midden van veel beroemde collega's. Zo'n levensroute van continentaal Europa via Engeland naar de VS was in die tijd niet ongewoon. De machtsgreep van Hitler in 1933 speelde daarbij vaak een directe rol. In Duitsland was je niet meer veilig, vervolgens elders in Europa ook niet. Toen Lehmann in 1940 zonder diploma uit Cambridge via New York naar Berkeley vertrok was de Battle of Britain in volle gang, met onzekere afloop. Maar de Britten hielden stand, en Erich wist ook zonder de juiste diploma's de doctorstitel te verwerven.

De volgende generatie kende geen nazi-dreiging meer, maar ervoer de verarming van Europa en de aantrekkelijkheid van het werkklimaat aan de Amerikaanse universiteiten. Toen ik in 1962 afstudeerde en in 1970 promoveerde was een tijdelijke aanstelling in de VS de droom van menig Nederlandse student. Willem van Zwet, Jaap Fabius en Willem Schaafsma verbleven zo in Berkeley, Rob Mokken in Michigan en ikzelf in State College Pennsylvania.

Lehmann schreef acht boeken, waarvan twee samen met J. L. Hodges Jr. *Testing Statistical Hypotheses* was zijn eerste, en voor mij zijn beste. Wiley gaf het uit in 1959, maar de *lecture notes* die er de kern van vormen gaan terug tot 1949. Elk van de acht hoofdstukken kent een aantal paragrafen gevolgd door een lijst van opgaven (uiteenlopend van 14 bij hoofdstuk 8 tot 49 bij hoofdstuk 6) en een deels geannoteerde literatuurlijst van enkele tientallen verwijzingen per hoofdstuk.

Laat ik proberen aan te geven waarom ik dit een mooi boek vond. Uitgangspunt is een collectie waarnemingen, zeg X, waarvan de kansverdeling bepaald wordt door een parameter, zeg $\theta$. We beginnen met de eenvoudige situatie waarin de nulhypothese en de alternatieve hypothese beide precies één parameterwaarde specificeren. Hoe vind ik de toets die onder $H_0$ met hoogstens kans alfa verwerpt en onder $H_1$ een zo groot mogelijke verwerpingskans (onderscheidingsvermogen, power) heeft? Het lemma van Neyman-Pearson stelt dat het optimaal is het verwerpingsgebied te laten bestaan uit die punten X=x waarvoor de verhouding van de verwerpingskansen onder $H_1$ resp. $H_0$ maximaal is, terwijl de verwerpingskans onder $H_0$ niet meer dan alfa mag zijn. Zo koop je zo veel mogelijk power voor de alfa-cent die je bereid bent uit te geven.

Vanuit dit principe brengt Lehmann successievelijk allerlei uitbreidingen en verfijningen aan. Om te beginnen is het bij discrete verdelingen voor de wiskundige eenvoud beter op de rand van het verwerpingsgebied te randomiseren, dat wil zeggen een loting uit te voeren waarbij de kans op verwerping een getal tussen 0 en 1 is zodanig gekozen dat de kans op de fout van de eerste soort precies alfa bedraagt. Bij een echte hypothesetoetsing in opdracht van een echte klant wordt deze hulploting voor de klant verzwegen: het zou ongepast zijn een onderzoeker te melden 'het was op het randje, we hebben voor U gedobbeld, maar U heeft pech gehad en we kunnen Uw nulhypothese niet verwerpen'.

Vervolgens bestaat de alternatieve hypothese vrijwel altijd uit een interval van parameterwaarden, en de nulhypothese dikwijls ook. Nu zoeken we bijvoorbeeld een eenzijdige toets van $H_0$: $\theta \leq a$ tegen $H_1$: $\theta > a$, voor een voor de onderzoeker interessante vaste waarde van a. In het algemeen kan de beste toets (most powerful) nu afhangen van de ware waarde van $\theta$ binnen het alternatief. Maar een UMP toets (Uniformly Most Powerful) bestaat als de bestudeerde familie van kansverdelingen Monotone Likelihood Ratio heeft, dat wil zeggen er bestaat een functie T(x) van de waarnemingen zodanig dat de verhouding van de kansdichtheden voor twee waarden van $\theta$ steeds een monotone functie van T is. Lehmann geeft

De Lehmannclub bijeen in 1969 in Utrecht. V.l.n.r. Peter Sander, Ivo Molenaar, Edo van der Velden, Martin van Montfort(achter), Rob Verdooren (voor), Paul van der Laan, Kobus Oosterhoff en Theo de Vries.

al direct een aantal voorbeelden waar dit principe goed werkt en kondigt aan dat bij veel vaak gebruikte kansverdelingen zo'n functie T bestaat. Wat later in het boek krijgt T de naam *sufficient statistic*, hij bevat alle nuttige informatie over het toetsingsprobleem. Het was voor mij bij lezing nieuw dat je met deze theorie veel bestaande toetsingsproblemen gezamenlijk kunt aanpakken omdat hun kansdichtheden kunnen worden herschreven in het format van de 'exponentiële familie'.

Wat te doen als het ingewikkelder ligt dan bij een eenzijdige toets van $H_0$: $\theta \leq a$ tegen $H_1$: $\theta > a$? Verderop in zijn boek verkent Lehmann deel-oplossingen voor bijvoorbeeld tweezijdig toetsen, of voor toetsen waarbij een 'nuisance parameter' van invloed is op de kansverdelingen (zeg je toetst een locatieparameter maar een onbekende variantieparameter speelt een rol). Soms kan een optimale toets worden gevonden binnen een zinvolle deelklasse (Uniformly Most Powerful Unbiased of Uniformly Most Powerful Invariant). Ook komen verwante zaken aan de orde zoals kruistabellen, permutatietoetsen, lineaire hypothesen, betrouwbaarheidsintervallen en minimax toetsen.

Wat de leden van de Lehmann-club een halve eeuw geleden aansprak was de ontwikkeling van een kader van algemene begrippen en criteria waarmee problemen die we tot dusver hoogstens ad hoc hadden bekeken onderling gerelateerd bleken te kunnen worden. Omdat de oefenopgaven soms vrij lastig waren en het wiskundig abstractieniveau nogal pittig, was het overleg met de vakgenoten erg nuttig. Er staan heel wat potloodaantekeningen in mijn exemplaar die daarvan getuigen.

Enkele jaren later kreeg ik een leerstoel voor Statistiek en Meettheorie aan de Sociale Faculteit van de RUG. Ik werd lid van de Psychometric Society, een gezelschap van een kleine duizend leden in vooral Noord-Amerika, waar statistische modellen werden ontwikkeld voor bijvoorbeeld meting van mentale vaardigheden. Rond 1995 ontmoette ik op een congres van de Society Juliet Popper Shaffer uit Berkeley, die vergezeld werd door haar tweede echtgenoot. Deze hield zich bescheiden op de achtergrond. Maar het was niemand minder dan Erich Lehmann, inmiddels de tachtig gepasseerd. Toen ik in 1998 tot President of the Society was verkozen mocht ik de *invited speaker* voor het jaarlijkse congres kiezen. Ik was zo brutaal Erich te benaderen en zijn reactie was 'I do not immediately say no'. Zo hebben de psychometrici zijn inzichten over asymptotiek kunnen vernemen nog voor zijn boek daarover in druk kwam.

Ziehier mijn eerbetoon aan een prominent statisticus, wiens verdiensten door de Leidse Universiteit in 1985 met een eredoctoraat zijn gewaardeerd.

Ivo W. Molenaar is emeritus hoogleraar statistiek en meettheorie.
E-mail: <molenaarivo@gmail.com>

# STATISTIEK OP DE CIRKEL

Wassily Kandinsky, Color Study, Squares with Concentric Circles, 1913. Collectie Lenbachhaus in München.

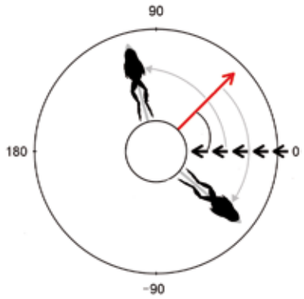
**Onderzoekers gebruiken vaak numerieke data zoals lengte, of categorische data zoals geslacht. Data kunnen echter ook gemeten worden in hoeken en worden dan circulaire data genoemd. Onderzoekers negeren in hun analyse soms het periodieke karakter van circulaire data en gebruiken een analyse voor lineaire data. Voor een goede analyse van circulaire data zijn echter speciale statistische methoden nodig, waarbij het periodieke karakter van de data wel kan worden meegenomen. We geven drie voorbeelden van circulaire data uit verschillende onderzoeksdisciplines. Daarna volgt een korte beschrijving van beschikbare circulaire statistische methoden en van de ontwikkelingen hierin.**

Jolien Ketelaar, Jolien Cremers, Kees Mulder & Irene Klugkist

Niet iedere onderzoeker zal dagelijks circulaire data tegenkomen, maar het komt wel degelijk voor in vakgebieden zoals biologie, politicologie en geologie. Meetinstrumenten die circulaire data opleveren zijn een kompas, klok, en circumplex. We geven voorbeelden van onderzoek met elk van deze instrumenten.

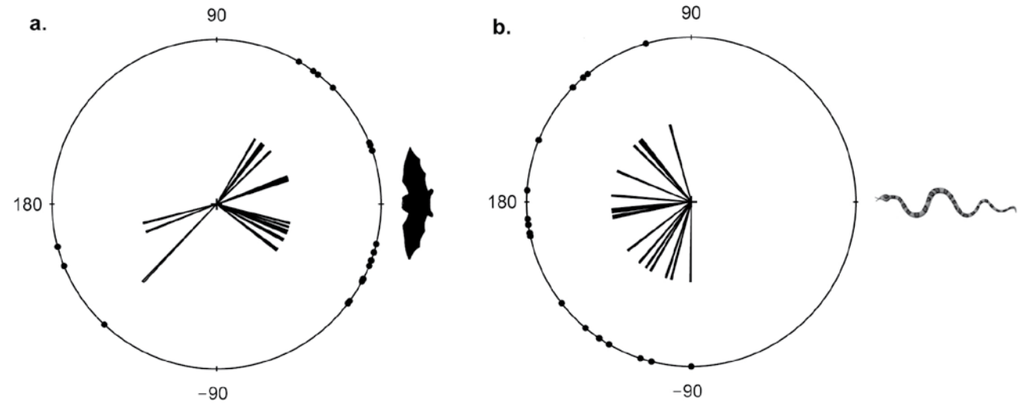
*Kompas*

Een voorbeeld van data gemeten in hoeken op het kompas is de vluchtrichting van een kikker ten opzichte van een roofdier. Bulbert, Page, & Bernal (2015) onderzochten of de vluchtrichtingen van kikkers voor een vleermuis anders waren dan voor een slang. Hiervoor werd een kikker op het middelpunt van een cirkel gezet, en had deze de mogelijkheid om alle kanten op te springen, dat wil zeggen 360° om zich heen (zie figuur 1).




Figuur 1. Vluchtrichting van een kikker. Aangepast van Bulbert et al. (2015), p. 6.






Figuur 2. Vluchtrichtingen van kikkers ten opzichte van: (a) een vleermuis, en (b) een slang. Aangepast van Bulbert et al. (2015), p. 6.

De onderzoekers plaatsten meerdere keren kikkers in een omgeving met daarin een vleermuis- of een slangmodel en bekeken welke kant de kikkers op vluchtten. In figuur 2 zijn de geobserveerde vluchtrichtingen afgebeeld voor beide prooidieren. Er is te zien dat de kikkers vaak in de richting van de vleermuis sprongen, terwijl ze bij een slang juist vooral de andere kant op vluchtten.

*Klok*

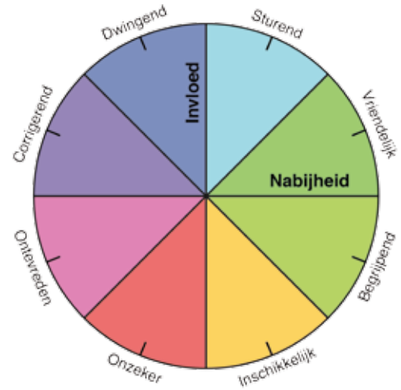
Naast data in richtingen, kan ook tijd beschouwd worden als circulair. Kirschner en Tomasello (2009) onderzochten bij kinderen van verschillende leeftijden in welke mate zij synchroon mee konden trommelen met een voorbeeldritme. Hierbij was de slag van het voorbeeldritme het nulpunt op de cirkel en werd de slag van het kind gemeten ten opzichte van dit voorbeeld. Wanneer deze slag te laat kwam, gaat de score ten opzichte van het nulpunt op de cirkel de ene kant op, drumde het kind te vroeg dan valt de score juist aan de andere kant van nul. Door circulair te meten is er geen direct onderscheid tussen te vroeg of te laat: als een kind veel te laat is voor een slag, is deze eigenlijk te vroeg voor de volgende slag. Bij een lineaire methode moet de onderzoeker een keus maken of een slag te vroeg is of te laat en dit kan problemen opleveren wanneer er zeer asynchroon wordt mee gedrumd, en dus niet duidelijk is bij welke slag in het voorbeeldritme een slag van het kind hoort.

*Circumplex*

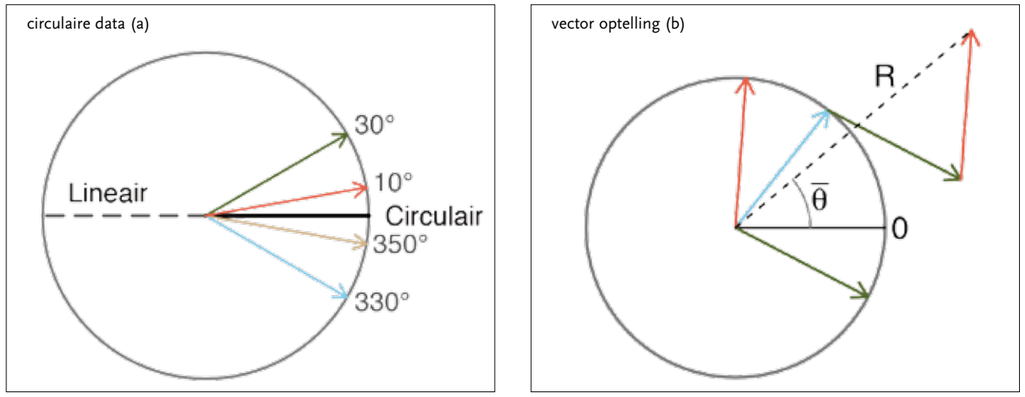
In onder andere de psychologie en de onderwijskunde wordt gebruik gemaakt van circumplex modellen. Deze modellen leveren circulaire data op, waarbij iemands positie op twee orthogonale assen wordt weergegeven door een positie op de cirkel. Het bekendste voorbeeld van een circumplex model is de Roos van Leary, met de dimensies invloed en nabijheid (zie figuur 3). Mainhard, Brekelmans, Den Brok en Wubbels (2011) onderzochten hiermee het sociale klimaat in schoolklassen. Leerlingen scoorden meerdere malen tijdens het schooljaar hun leerkracht op dit instrument. Aan de hand van deze scores werd gekeken hoe het sociale klimaat in de klas veranderde gedurende het jaar. Door het gebruik van circulaire methoden kunnen beide dimensies met elkaar worden gecombineerd en interpreteren we de richting op de cirkel in plaats van twee aparte componenten.

## Circulaire statistiek

Door het periodieke karakter van circulaire data zijn zowel voor descriptieve als inferentiële doeleinden speciale statistische technieken nodig. Stel dat we een dataset met vier observaties hebben, namelijk 10°, 30°, 330° en 350° (zie figuur 4a). Het lineaire gemiddelde van deze data is 180°. Echter op de cirkel bekeken is 180° ver weg van de data. Ook de spreiding wordt niet goed weer-




Figuur 3. Voorbeeld van een circumplex model.




Figuur 4. Een voorbeeld van circulaire data (a). De gekleurde pijlen zijn de geobserveerde datapunten gemeten in graden. De gestreepte lijn geeft het lineaire gemiddelde aan en de ononderbroken lijn het circulaire gemiddelde. Een voorbeeld van vector optelling (b), waarbij de verschillende richtingen zijn weergegeven in verschillende kleuren.

gegeven door de lineaire standaard deviatie, omdat deze een grote waarde (185°) heeft terwijl de punten dicht bij elkaar liggen.

Daarom kunnen we beter de gemiddelde richting gebruiken, die we krijgen door het optellen van de vectoren richting de datapunten op de cirkel (zie figuur 4b). De lijn die resulteert uit deze optelling wordt de resulterende vector genoemd. De richting van deze lijn is het circulaire gemiddelde ($\theta$). Door de lengte van de lijn ($R$) te delen door het aantal observaties kan de gemiddelde resulterende lengte ($R$) worden gevonden, die informatie geeft over de spreiding in de data (Fisher, 1995; Mardia, & Jupp, 2009). Deze manier van werken met circulaire data staat aan de basis van de meeste circulaire tests, zoals de circulaire ANOVA of regressie.

## Werken met circulaire data

Er bestaan momenteel statistische toetsen voor circulaire data uit relatief eenvoudige designs. Statistische methoden voor meer ingewikkelde designs zijn volop in ontwikkeling. Dit kan gaan om het analyseren van circulaire variabelen in de context van factoriële variantie analyses, herhaalde metingen analyses, een multipele regressie of multilevel analyse. Er bestaan drie verschillende benaderingen om te werken met circulaire data, de *intrinsic*, *embedding* en *wrapping approach*.

### *Intrinsic approach*

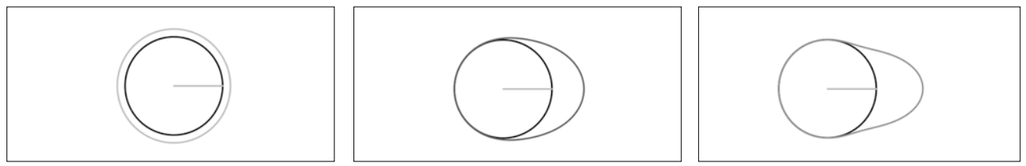
De *intrinsic approach* is een van de meest gebruikte benaderingen. Hierbij wordt een verdeling direct gedefinieerd op de cirkel. Vaak wordt er gebruik gemaakt van de Von Mises-verdeling, de circulaire analoog van de Normaalverdeling:

$$f(\theta|\mu,\kappa) \propto \exp(\kappa\cos(\theta-\mu)).$$

De cosinus-functie zorgt ervoor dat de data op de cirkel gezet worden. In figuur 5a, b, en c, staan drie Von Misesverdelingen met toenemende kappa-waarden, wat overeenkomt met afnemende spreiding. Deze aanpak is erg natuurlijk, en de Von Misesverdeling heeft een aantal voordelige statistische eigenschappen. Aan de andere kant is het moeilijk om de distributies van teststatistieken onder de nulhypothese af te leiden, waardoor het niet altijd eenvoudig is om nieuwe statistische methoden te ontwikkelen binnen deze benadering, vooral in een multivariate context.

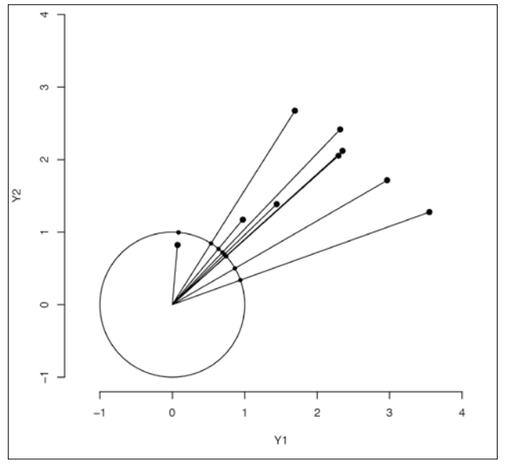
### *Embedding approach*

Bij de *embedding approach* wordt aangenomen dat de circulaire variabele een geprojecteerde bivariate lineaire verdeling heeft. Dit houdt in dat een gegeven variabele op de cirkel ($\theta$) wordt gezien als projectie van een niet geobserveerde, onderliggende variabele in een tweedimensionaal vlak ($Y$). Deze twee variabelen kunnen aan elkaar worden gerelateerd door het schatten van een latente variabele ($R$) die de afstand van de onderliggende lineaire



Figuur 5. Voorbeelden van Von Misesverdelingen, met kappa 0 (a), 3 (b) en 6 (c). Het nulpunt wordt aangegeven met de grijze lijn.






Figuur 6. Bivariate datapunten geprojecteerd op de eenheidscirkel. $R$ is de afstand van de datapunten tot het nulpunt, de snijpunten van deze lijnen met de cirkel geven circulaire datapunten.

datapunten tot het nulpunt representeert (zie figuur 6). In plaats van direct een model voor te schatten, zoals in de intrinsic approach, wordt er een model geschat voor $Y$ en $R$. Schattingen voor parameters van deze modellen krijgen we echter op twee componenten, zoals ook in lineaire bivariate modellen (bijvoorbeeld in een regressiemodel zijn er twee $\beta$'s voor een onafhankelijke variabele, een voor component $Y_1$ en een voor component $Y_2$). Dit betekent dat we voor een circulaire interpretatie van de parameters een manier moeten vinden om twee parameters samen te voegen tot één circulaire parameter.

### *Wrapping approach*

Bij de *wrapping* benadering wordt een bestaande lineaire verdeling om de cirkel heen gewikkeld. Hierbij komt de dichtheid van de verdeling op de cirkel tot stand komt door meerdere lagen van het wikkelen bij elkaar op te tellen. De geobserveerde variabele op de cirkel ($\theta$) wordt hier dus gezien als een op de cirkel gewikkelde onderliggende, niet geobserveerde variabele in de lineaire ruimte ($y$):

$$\theta = y(mod\ 2\pi)$$

Ook hier hebben we te maken met een extra latente variabele (veelal aangeduid met $k$), namelijk het aantal wikkelingen dat heeft plaatsgevonden. Ferrari (2010) laat zien hoe in het Bayesiaanse framework identificatieproblemen die ontstaan door de latente variabele $k$ relatief eenvoudig overbrugd kunnen worden.

Een verdeling die veel wordt gebruikt binnen de wrapping benadering is de normaalverdeling, die na het wikkelen de wrapped normal wordt genoemd. Een belangrijk voordeel van deze en de embedding benadering is dat beide zeer flexibel zijn. Je kunt bijvoorbeeld ook een asymmetrische verdeling om de cirkel wikkelen als je ziet dat de data asymmetrisch zijn.

### Onderzoek in Utrecht

Hoewel er wereldwijd onderzoek wordt gedaan om de beschikbare methoden voor circulaire data verder uit te breiden, loopt het veld nog achter ten opzichte van statistiek voor lineaire data. Net als bij veel andere relatief complexe datastructuren lijkt de Bayesiaanse aanpak veel te bieden te hebben. Om die reden houdt een onderzoeksgroep in Utrecht zich bezig met het ontwikkelen en onderzoeken van Bayesiaanse methoden voor circulaire data. De mogelijkheden, en voor- en nadelen van de drie benaderingen (intrinsic, embedding, wrapping) worden bestudeerd voor modellen die we in de praktijk vaak tegenkomen, zoals bijvoorbeeld covariantie-analyse, multipele regressie, en mixed (multilevel) modellen. Behalve op verbeteringen en uitbreidingen van bestaande methoden wordt vooral gefocust op de ontwikkeling van interpretatiemethoden die gebruikt kunnen worden door toegepaste onderzoekers. Door samenwerkingen met sociale en gedragswetenschappers zorgen we voor relevantie en bruikbaarheid van de methoden voor met name deze doelgroep. Voor meer informatie zie ook de website <cdm.sites.uu.nl>.

### Literatuur

Bulbert, M. W., Page, R. A., & Bernal, X. E. (2015). Danger Comes from All Fronts: Predator-Dependent Escape Tactics of Túngara Frogs. *PLoS ONE, 10*(4): e0120546. doi: 10.1371/journal.pone.0120546

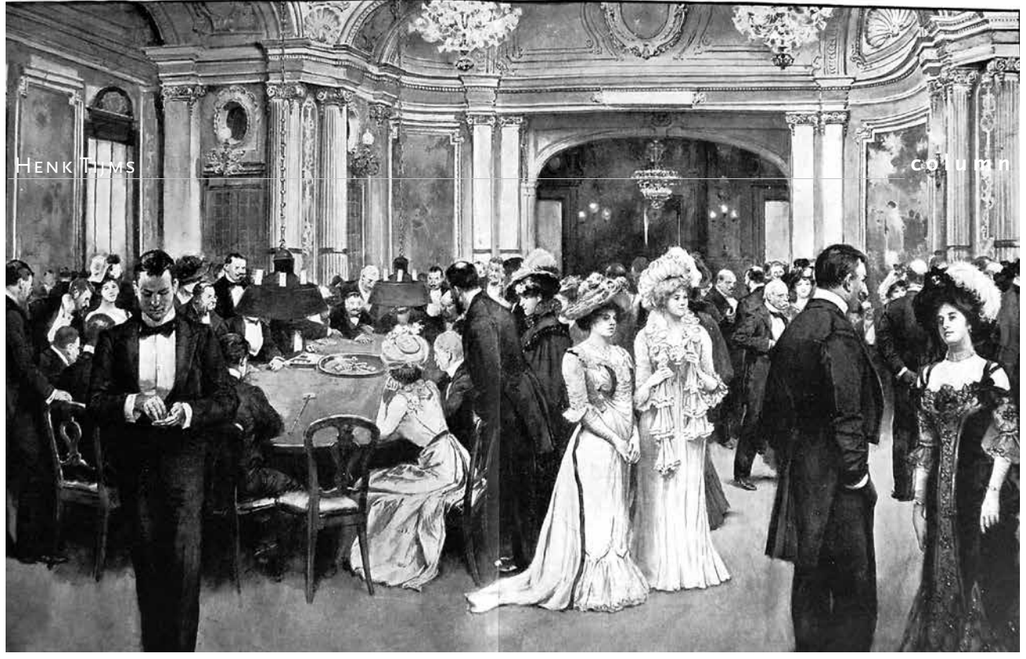
Ferrari, C. (2010). *Circular Data Bayesian Modeling: The Wrapping Approach.* Saarbrücken: VDM Verlag Dr. Müller.

Fisher, N. I. (1995). *Statistical analysis of circular data.* New York: Cambridge University Press.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of experimental child psychology, 102*(3), 299–314.

Mainhard, M. T., Brekelmans, M., den Brok, P., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary educational psychology*, *36*(3), 190–200.

Jolien Ketelaar is research master student Methodology & Statistics of the Behavioural, Biomedical and Social Sciences, Universiteit Utrecht. Dit artikel is geschreven naar aanleiding van haar bachelor thesis onder begeleiding van Jolien Cremers en Kees Mulder, beiden promovendi binnen het VIDI project (NWO 452-12-010) 'A different angle: new tools for circular data', en Irene Klugkist, hoogleraar Methoden & Statistiek aan de Universiteit Utrecht en bijzonder hoogleraar Onderzoeksmethodologie Meetmethoden en Data-Analyse, Universiteit Twente.
E-mail: <j.ketelaar@students.uu.nl>



De speelzaal van het casino in Monte Carlo begin twintigste eeuw

# DE WAAN VAN DE GOKKER

De Amerikaanse wiskundige Ted Hill had de gewoonte om bij zijn eerste college kansrekening de studenten te vragen thuis een experiment uit te voeren en de resultaten mee te nemen naar het volgende college. De studenten werd gevraagd 200 opeenvolgende uitkomsten van een daadwerkelijk opgeworpen zuivere munt te noteren of net te doen alsof de munt 200 keer opgegooid was en 200 opeenvolgende uitkomsten op te schrijven van de zogenaamd uitgevoerde muntworpen. Tot grote verbazing van de studenten haalde Ted Hill vrijwel altijd de gefingeerde resultaten er feilloos uit. Een snelle blik volstond om de inzendingen van de studenten op waarde te schatten. In een interview met de *New York Times* legde Ted Hill uit dat het 'geheim' van zijn aanpak heel simpel was: elk resultaat dat niet een ononderbroken reeks van tenminste zes keer kop of zes keer munt bevatte, werd door hem als gefingeerd bestempeld. De kans op een dergelijke reeks is verrassend hoog en is 96,53 % (de kans op een run van lengte 4 of meer is zelfs 99,99999 % en de kans op een run van lengte 8 of meer is nog steeds 54,19 %). In overeenstemming met deze kansen is er een nuttige vuistregel die voor de situatie van n worpen met een zuivere munt stelt dat de kansverdeling van de lengte van de langste run van alleen kop geconcentreerd is rond de waarde $\log_2(n)$-1, terwijl de kansverdeling van de lengte van de langste run van alleen kop of alleen munt is geconcentreerd rond de waarde $\log_2(n)$ voor n voldoende groot. Bij 200 worpen met een zuivere munt is 7 de meest waarschijnlijke waarde voor de lengte van de langste run. Vanzelfsprekend was het doel van Ted Hill de studenten aan den lijve te laten ervaren dat verreweg de meeste mensen zich geen goede voorstelling kunnen maken van *randomness*. In de psychologie zijn in het verleden diverse studies uitgevoerd die laten zien dat de meeste mensen niet in staat zijn data te genereren die random zijn. Niet zo lang geleden werd door Amerikaanse psychologen aangetoond dat de *hot hand* in basketbal een misvatting is. Lange runs in random data treden op basis van toeval veel vaker op dan men gewoonlijk geneigd is te veronderstellen.

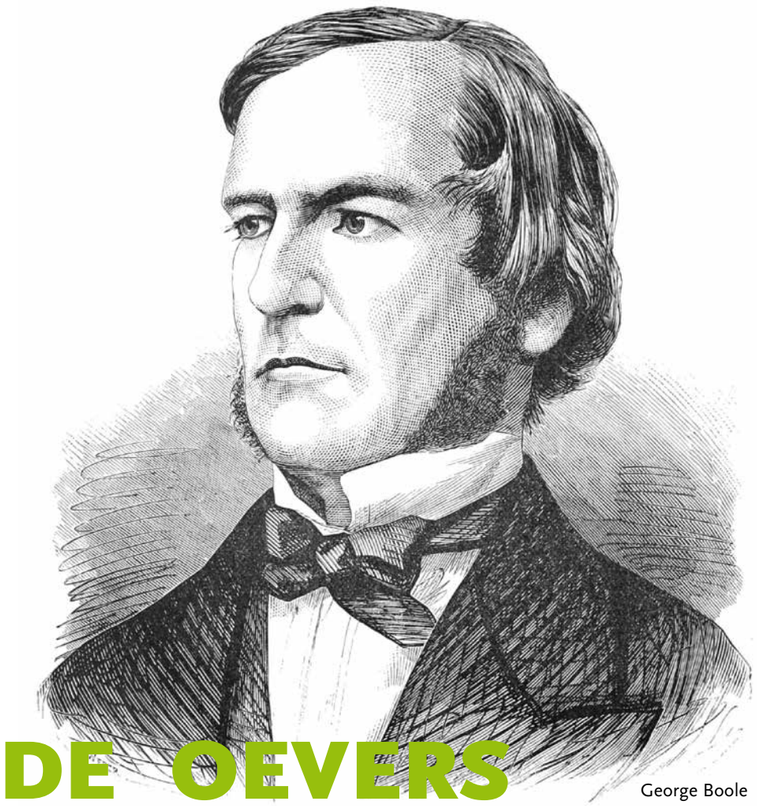
Bovengenoemde misvattingen over *randomness* in een reeks van onafhankelijke Bernoulli-experimenten kunnen geschaard worden onder het kopje 'waan van de gokker'. Dit slaat op een gokker die gelooft dat als een bepaalde gebeurtenis minder vaak dan gemiddeld optreedt in een zekere periode, de gebeurtenis daarna wel vaker zal optreden. Een aansprekende illustratie van dit verschijnsel is het gebeuren dat plaatsvond op 18 augustus 1913 in het casino van Monte Carlo toen het roulette balletje 26 keer achter elkaar op zwart viel. Steeds grotere bedragen werden toen op rood ingezet nadat het balletje een aantal keren achterelkaar op zwart was gevallen en dat maar bleef doen. Hoe uitzonderlijk is een dergelijke gebeurtenis over een langere periode bezien? Deze vraag kunnen we beantwoorden met de vuistregel die stelt dat de kansverdeling van de lengte van de langste run van alleen zwart of alleen rood in n draaiingen van het roulette wiel sterk geconcentreerd is rond de waarde $\log_{1/p}(n(1-p))+1$ voor n groot, waarbij p = 18/37 in Europees roulette. Dit betekent dat in 100 miljoen draaiingen van het roulette wiel de kansverdeling van de lengte van de langste run sterk geconcentreerd is rond de waarde 26, precies de waarde die optrad in 1913 in het casino van Monte Carlo. In het licht van tientallen jaren roulette spel en de vele casino's in de wereld is een run van de lengte 26 niet onvoorstelbaar lang.

Was het gevolg van de waan van de gokker nog tamelijk onschuldig in het casino van Monte Carlo, dit was zeker niet het geval voor de 'Venetië-53' waanzin die in een landelijke Italiaanse loterij uitbrak eind 2004 en begin 2005. In deze loterij vinden iedere week twee trekkingen plaats in elk van tien steden over het land waaronder de stad Venetië. Bij elke trekking in de tien steden worden vijf getallen getrokken uit de getallen 1 tot en met 90. In de loterij kun je op 1, 2, 3, 4 of 5 getallen gokken met een uitbetaling van ongeveer 10 keer je inzet als je op één getal juist gegokt hebt en een uitbetaling van 1 miljoen keer je inzet als je op vijf getallen juist gegokt hebt. Hoewel het getal 53 in verschillende andere steden herhaaldelijk viel, bleef dit getal uit in Venetië tussen mei 2003 en februari 2005 gedurende 182 trekkingen. In deze periode werd voor meer dan 3,5 miljard euro vergokt op het getal 53, waarbij complete familiekapitalen er doorheen gejaagd werden. In januari 2005 alleen al werd voor 672 miljoen euro op het getal 53 vergokt. Verschillende hoogleraren wiskunde verschenen op de Italiaanse televisie om mensen uit te leggen dat de loterijballetjes geen geheugen hebben, om zo te waarschuwen voor onverantwoordelijk hoge inzetten. Tevergeefs, vele Italianen bleven rotsvast ervan overtuigd dat het getal 53 op het punt stond te vallen en bleven grote geldbedragen inzetten. In de maand januari 2005 vonden diverse tragische gebeurtenissen plaats die het gevolg waren van het met enorm grote geldbedragen gokken op het vallen van het getal 53 in de loterij in Venetië. Een huisvrouw verdronk zichzelf in de Toscaanse zee, een verzekeringsagent uit Florence schoot zijn vrouw en zijn zoon dood alvorens hij de hand aan zichzelf sloeg, en een man uit Sicilië werd gearresteerd vanwege zware mishandeling uit frustratie voor de gokschulden die hij opgelopen had. Vele andere incidenten vonden plaats: mensen die in handen vielen van genadeloze geldwoekeraars, anderen die uit hun huis werden gezet vanwege hoog opgelopen gokschulden. Na 182 opeenvolgende trekkingen niet verschenen te zijn, viel uiteindelijk bij de trekking op 9 februari 2005 in Venetië het getal 53 en kwam de gokwaanzin die Italië zo lang in zijn greep gehouden had tot een eind. De schatting is dat de loterij ongeveer 600 miljoen euro heeft uitgekeerd aan degenen die op het getal 53 gewed hadden. Een groot bedrag maar in wezen een kleinigheid in het licht van wat de loterij aan de Venetië-53 goklust heeft verdiend. Deze episode was de ergste vorm van gokwaanzin in Italië sinds in 1941 het getal 8 gedurende 201 opeenvolgende trekkingen niet verschenen was in de loterij in Rome. Het sterke vermoeden is dat Mussolini destijds de trekkingen gemanipuleerd heeft om de inkomsten van de loterij te verhogen waarmee dan oorlogsinspanningen gefinancierd konden worden.

Het is een kwestie van tijd dat de geschiedenis van Venetië-53 zich herhaalt. De kans is groot dat dit binnen 10 jaar of 25 jaar gebeurt. De loterij vindt plaats in 10 Italiaanse steden en elke week zijn er twee trekkingen. De kans dat in de komende 10 jaar een of ander getal tussen 1 en 90 wegblijft in één van de 10 Italiaanse steden bij 182 of meer opeenvolgende trekkingen is ongeveer gelijk aan 50 %, terwijl bij een periode van 25 jaar deze kans ongeveer gelijk is aan 91 %. Hoe komen we aan deze kansen? Neem daartoe eerst een specifiek getal (zeg 53) en een specifieke stad (zeg Venetië). Je kunt dan met een Markov-keten de exacte waarde p = 0,0007754 berekenen voor de kans dat over een periode van 10 jaar in de 1040 trekkingen er 182 of meer opeenvolgende trekkingen zijn waarin het specifieke getal niet valt in de genomen stad (deze kans is 0,0027013 bij 2600 trekkingen over een periode van 25 jaar). De vijf te trekken winnende loterijgetallen zijn niet onafhankelijk van elkaar, maar de afhankelijkheid is 'licht genoeg' om te kunnen stellen dat bij goede benadering de kans dat in een gegeven stad *niet* sprake zal zijn van een loterij getal dat in 182 of meer opeenvolgende trekkingen wegblijft ongeveer gelijk is aan $(1-p)^{90}$. In totaal zijn er 10 steden waarin onafhankelijk van elkaar de loterij zich afspeelt. Dit leidt tot de conclusie dat de kans dat in de komende 10 jaar *wel* sprake zal zijn van een loterijgetal dat in 182 of meer opeenvolgende trekkingen wegblijft in één van de 10 Italiaanse steden ongeveer gelijk is aan $1-(1-p)^{900}$, oftewel een kans van ongeveer 50 %. Bij een periode van 25 jaar is deze kans zelfs ongeveer 91 % en de genoemde kansen worden bevestigd door simulatie.

Mogelijk kan deze column in het Italiaans vertaald worden en dan voor de Italianen als les voor de toekomst dienen.

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening.
E-mail: <tijms@quicknet.nl>



George Boole

# AAN DE OEVERS VAN DE RUBICON

Richard Starmans

De precies 200 jaar geleden geboren Britse wiskundige George Boole (1815-1864) publiceerde in 1854 zijn thans klassieke studie *An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities*. De titel laat aan duidelijkheid weinig te wensen over en de inleidende paragraaf geeft evenmin aanleiding tot misverstanden over het oogmerk van de auteur: '*The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus, and upon this foundation to establish the science of Logic and construct its method; to make that method itself the basis of a general method for the application of the mathematical doctrine of Probabilities; and, finally, to collect from the various elements of truth brought to view in the course of these inquiries some probable intimations concerning the nature and constitution of the human mind.*' In weerwil van dit alles staat Boole in de annalen van de wetenschapsgeschiedenis vooral geboekstaafd als pionier van de informatica, die met zijn landgenoot Charles Babbage (1791-1871) het pad effende voor het informaticatijdperk, waarvan in de jaren dertig van de 20e eeuw door Turing, Church, Shannon en Von Neumann de fundamenten zouden worden gelegd. Zijn naam is vandaag de dag vooral verbonden aan de zogenaamde Boolese algebra, een zuiver wiskundige theorie, waarvan de voornaamste beginselen in zijn *An Investigation of the Laws of Thought* zijn terug te vinden. Pas in 1937 zou in de doctoraalscriptie van Claude Shannon worden gewezen op het belang ervan voor telefooncentrales, elektronische schakelingen en circuits, een preambule tot de latere micro-elektronica revolutie en de komst van de elektronische computer. De herdenking van Boole's geboortejaar is in 2015 in de academische wereld mondiaal op gepaste wijze luister bijgezet met symposia, congressen en diverse nieuwe publicaties over zijn werk.



## Opmaat tot een symbolische logica

Boole's werk oversteeg echter het belang van de informatica in engere zin. Hij was evenals Babbage sterk wijsgerig georiënteerd en zou de latere 20e-eeuwse analytische filosofie en wetenschapsfilosofie ingrijpend beinvloeden. Allereerst was hij zonder twijfel de wegbereider van de linguïstisch-logische revolutie, die met Frege begon en in het werk van Russell en Wittgenstein in het begin van de 20e eeuw een hoogtepunt beleefde. In zijn pamflet *Mathematical Analysis of Logic* (1847) betoogde Boole dat logica niet langer op de traditionele filosofische wijze in natuurlijke taal benaderd moest worden, maar aansluiting moest zoeken bij de wiskunde. De (deductieve) wetten van het denken moesten in wiskundige taal worden uitgedrukt. Als opmaat tot een daadwerkelijk symbolische logica liet hij in voornoemde publicatie zien dat de aristotelische syllogismen met behulp van een eenvoudig, elegant algebraïsch systeem konden worden gerepresenteerd en dat de geldigheid van conclusies formeel kon worden bewezen. Daarmee plaveide hij de weg voor Gottlob Frege, die later aan de wieg van de predicatenlogica zou staan.

Belangwekkend is daarnaast Boole's bijdrage aan het redeneren met onzekerheid, inductie en vooral kansrekening en statistiek. Allereerst gaf hij steevast een epistemische duiding aan het waarschijnlijkheidsbegrip; een kansuitspraak moet worden bezien tegen de achtergrond van de beschikbare kennis en de 'verwachting' van de persoon die zich aan de kansuitspraak committeert. Daarbij koos Boole overigens allerminst voor een subjectivistische of mentalistische aanpak en duiding van de notie van verwachting. Kansuitspraken betroffen bij Boole logische relaties tussen proposities en daarmee werd hij een belangrijke voorloper van de logische interpretatie van het kansbegrip, zoals die door Carnap zou worden bepleit. Hij omarmde evenals Laplace een epistemische notie van waarschijnlijkheid, maar hekelde diens *principle of insufficient reason*, waarbij gebeurtenissen even waarschijnlijk worden geacht als er onvoldoende reden is het tegengestelde aan te nemen. Anders dan de deductieve logica waren vormen van redeneren met onzekerheid doorgaans inductief, maar dit vormde voor Boole geen reden logica en waarschijnlijkheid principieel te scheiden en de laatste uit te sluiten van de 'nieuwe' symbolische orde. Met zijn logisch/probabilistische calculus treedt hij nadrukkelijk in de voetsporen van Leibniz en geeft hij tevens een aanzet tot het programma van het logisch positivisme in de jaren dertig van de 20e eeuw.

## Mechanisering van de kennis

Deze twee hier summier geschetste aspecten van Boole's werk culmineren in een derde, waarin wellicht zijn grootste betekenis ligt: zijn bijdrage aan het door filosofen eeuwenlang nagejaagde ideaal kennis langs mechanische weg te produceren om zo ware kennis te verwerven en deze te relateren aan de structuur van de menselijke geest. We vinden dit ideaal van mechanisering van de kennis reeds terug in speculatieve geschriften van Raymond Lull uit de 13e eeuw, maar het kreeg nadrukkelijker gestalte bij 17e eeuwse denkers als Bacon, Descartes, Hobbes, Pascal en Leibniz, die nadachten en speculeerden over de mogelijkheden van een mechanische, universele methode van kennisverwerving, waarbij fouten en dwalingen op voorhand werden geëlimineerd en waarbij redeneren werd gereduceerd tot rekenen, zodat bij voorbeeld juridische conflicten konden worden opgelost op dezelfde wijze waarop boekhouders dat kunnen, zonder emoties of retorica. De problematiek van het reduceren van kennis en denken tot elementaire rekenstappen, die automatisch kunnen worden uitgevoerd, werd evenwel op de meest saillante wijze door Thomas Hobbes verwoord in zijn politieke traktaat Leviathan (1651): '*When man reasoneth, he does nothing else but conceive a sum total, from addition of parcels; or conceive a remainder, from subtraction of one sum from another: which, if it be done by words, is conceiving of the consequence of the names of all the parts, to the name of the whole; or from the names of the whole and one part, to the name of the other part. And though in some things, as in numbers, besides adding and subtracting, men name other operations, as multiplying and dividing; yet they are the same: for multiplication is but adding together of things equal; and division, but subtracting of one thing, as often as we can. These operations are not incident to numbers only, but to all manner of things that can be added together, and taken one out of another. For as arithmeticians teach to add and subtract in numbers, so the geometricians teach the same in lines, figures (solid and superficial), angles, proportions, times, degrees of swiftness, force, power, and the like; the logicians teach the same in consequences of words, adding together two names to make an affirmation, and two affirmations to make a syllogism, and many syllogisms to make a demonstration; and from the sum, or conclusion of a syllogism, they subtract one proposition to find the other. Writers of politics add together pactions to find men's duties; and lawyers, laws and facts to find what is right and wrong in the actions of private men.*' Hobbes onderstreept zijn 'redenering als berekening'-benadering door het uit-

werken van diverse voorbeelden van menselijke redeneren en andere vormen van complex cognitief gedrag, die kunnen worden herleid tot de bovengenoemde elementaire acties. Hobbes' preoccupaties en overwegingen vormden voor de filosoof John Haugeland voldoende reden hem uit te roepen tot *grandfather of Artificial Intelligence*. Hobbes had evenwel nog geen enkele conceptie van berekenbaarheid, was niet of nauwelijks op de hoogte van de beginselen der kansrekening en beschikte evenmin over een voldoende rijke symbolische taal om deze vorm van denken gestalte te geven. De inlossing van de Hobbesiaanse belofte kreeg pas vorm met de symbolisch-logisch-probabilistische resultaten van George Boole, die daarbij van mening was dat de (onderliggende) wetten van het menselijke denken langs deze weg moesten worden beschreven en konden worden gekend. Zoals gezegd, wilde hij op die manier inzicht verschaffen in *the nature and constitution of the human mind*. Zijn ideeën vormden dan ook niet minder dan een aanzet tot latere concepties van berekenbaarheid en intelligentie, zoals die door Alan Turing werden ontwikkeld in zijn wellicht twee beroemdste artikelen, respectievelijk *On Computable Numbers, with an Application to the Entscheidungsproblem* uit 1936 en later *Computing Machinery and Intelligence* uit 1950.

## Integratie van kansrekening en logica

In 49 voor Christus stak Julius Caesar de rivier de Rubicon over, nadat hij volgens Suetonius de gevleugelde woorden *alea iacta est* (de teerling is geworpen) had uitgesproken. Deze uitspraak was evenzeer omineus, als de daad zelf omstreden en onomkeerbaar. Caesar was zich terdege bewust dat het bestaande equilibrium in een klap werd verstoord en er geen weg terug zou zijn. De hierboven geschetste context mag dan minder dramatisch zijn, na Boole's werk zou de relatie tussen logica en filosofie nooit meer dezelfde zijn. Geldige redeneringen en de zoektocht naar waarheid vereenzelvigen met het oplossen van algebraïsche vergelijkingen, een formele, inhoudsloze duiding in plaats van interpretatie van redeneringen; de breuk was even onvermijdelijk als onomkeerbaar. Filosofen als Hegel en later Husserl zouden van de weeromstuit hun eigen, 'extravagante' concepties van logica ontwikkelen en Boole's werk markeerde wel degelijk een cesuur, die zou bijdragen aan het latere, nog steeds bestaande schisma tussen de analytische / Angelsaksische filosofie en de continentale wijsbegeerte. Boole kon dat alles natuurlijk niet geheel anticiperen, maar gaat in *The Laws of Thought* wel uiterst omzichtig en piëteitsvol om met de eeuwenoude logische traditie, waarbij hij sterk de continuïteit benadrukt. Anders dan Immanuel Kant, die stelt dat er sedert Aristoteles weinig voortgang in de logica is geboekt, noemt en roemt Boole bijdragen van laat-klassieke denkers als Proclus en Plotinus' leerling Porphyrus, van scholastieke denkers als Anselmus van Canterbury en Pierre Abelard, tot 17e-eeuwse vernieuwers als Descartes, Locke, Pascal en de Logica van Port Royal. Het hoeft evenwel geen betoog dat Boole's probabilistische ambities in wijsgerige kring op nog minder bijval konden rekenen. Waar de logica kon bogen op een oude en eerbiedwaardige traditie en met de queeste naar zekere kennis werd geassocieerd, kende kansrekening een veel bescheidener genealogie; zij ging terug tot de 17e eeuw, kwam deels voort uit de studie van kansspelen, levensverzekeringen en andere praktische en profane zaken, en was vooral gericht op het managen van onzekerheid. Waarschijnlijkheid behelsde derhalve volgens velen geen kernbegrip in de kennisleer. Maar ook buiten de filosofie was en is de integratie van kansrekening en logica, die Boole evenals zijn tijdgenoot De Morgan propageerde, allerminst evident en werd dikwijls niet begrepen. Het visionaire karakter van Boole blijkt ook vandaag de dag. Zo manifesteerde zich in de kunstmatige intelligentie vele decennia een heuse scheiding der geesten; een symbolische op logica en kennisrepresentatie gerichte invalshoek tegenover een subsymbolische probabilistische benadering. Er tekent zich het laatste decennium enige convergentie af, maar beide onderstromen zijn nog steeds bespeurbaar.

Hoe dan ook, Boole kende de traditie waaruit hij voortkwam en besefte dat velen weliswaar aan de oevers van de Rubicon hadden gestaan en daar langere tijd hadden vertoefd zonder deze te kunnen of willen oversteken. Met enig gevoel voor pathos kan worden gesteld dat Boole dit kunststukje wel voltooide, juist door de hier geschetste drieslag in zijn werk: formalisering van de logica in algebraïsche calculus, integratie van logisch en probabilistisch redeneren, en dit alles gefundeerd in en corresponderend met een aanzet tot een computationele theorie van de menselijke geest.

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






# ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ

Deze tekst stond boven de ingang van de academie waar Plato discussieerde met zijn leerlingen. Slordig vertaald komt het er op neer dat zij die geen benul van wiskunde hebben hier niet op hun plaats zijn. Nu was wiskunde, eigenlijk staat er geometrie, in die tijd een begrip dat breder was dan tegenwoordig. Je zou het kunnen zien als 'systematisch denken' en dan is die tekst ineens veel logischer: iemand die dat niet kan heeft weinig te zoeken in het wetenschappelijk debat. Soms denk ik stiekem dat zo'n vereiste nu nog steeds heel welkom zou zijn. We moeten veel beter en systematischer nadenken, ook in ons vak. Regelmatig zie ik onderzoeken die fantastische dingen doen met grote hoeveelheden data, maar die data zijn dan twijfelachtig. Je kunt nu eenmaal geen fatsoenlijke analyse doen op gegevens die niet kloppen.

Omdat ik een ver en schimmig verleden heb in het marktonderzoek kan ik het niet over mijn hart verkrijgen te weigeren als ik weer eens een verzoek krijg een vragenlijst in te vullen. Na meer dan 50 jaar weet ik nog steeds hoe moeilijk het was voldoende respondenten te krijgen. Maar dikwijls erger ik me geweldig aan de knullige opzet en de slechte vragen. Zo schijnt het tegenwoordig mode te zijn enkele groepjes foto's van mensen aan de ondervraagde voor te leggen en hen dan te vragen welk groepje mensen het best bij welk merk bank/supermarkt/etc. hoort. Kennelijk kan men aan die foto's een bepaalde levensstijl herkennen. Maar ik zie totaal geen verschillen tussen die groepen foto's, voor mij als 70-plusser zijn het allemaal hetzelfde soort jonge mensen. Zo'n vraag heeft voor mij dus geen enkele zin. Ook zeer geliefd is het vragen welke van een lange lijst merken auto/tijdschrift/etc. men kent, 'al is het alleen maar van naam'. Op zich is dat best een goede vraag, maar als men dan vervolgens doorvraagt over het beeld dat ik heb over een bedrijf dat ik alleen maar van naam ken kan dat niet anders dan flauwekul gegevens opleveren. Een van mijn grootste ergernissen betreft de vraag hoe groot de kans is dat ik mijn verzekeraar/energieleverancier/etc. aan anderen zou aanbevelen. Die vraag wil ik beantwoorden met iets als 'dat doe ik nooit', want ik weiger altijd dit soort adviezen te geven. Maar helaas biedt de vragenlijst die mogelijkheid niet, ik kan zelfs niet 'geen antwoord' of 'weet niet' invullen, ik moet een keus maken uit een vijf- of zeven-puntsschaal. Het resultaat wordt dan ook dat ik zeer tevreden ben over mijn leverancier (als ik dat niet was nam ik wel een ander), maar desondanks een uiterst negatieve score produceer op het punt van aanbevelen.

Het is met die vragenlijsten ongeveer als het domweg vragen naar een oordeel over het laatste boek van Giphart zonder eerst te vragen of men wel eens boeken leest en vervolgens welke schrijvers dat dan zijn. Als daar dan al Giphart bij zou zitten moet je eerst ook nog vragen welke boeken de ondervraagde van hem kent, want het zou heel goed kunnen dat voor hem 'het laatste boek' een ander werk is dan dat wat de vragensteller in gedachten heeft. Denk nu niet dat dit een gekunsteld voorbeeld is, dergelijke slecht doordachte vragen worden regelmatig op het Nederlandse volk losgelaten en de resultaten leveren menig sappig stukje in de pers op.

Kortom, die Plato zag het zo slecht nog niet. Ga eerst maar eens fatsoenlijk en systematisch nadenken over de vraag die je wilt stellen. Ondanks de vele goede methodologen die Nederland kent en gekend heeft ontbreekt nog veel te vaak een goede onderzoeksopzet. Statistiek is niet alleen maar een spreadsheet met getallen vullen en daar SPSS op loslaten. Het is óók een zaak van systematisch onderzoeken of de vragen die men gaat stellen wel de achterliggende onderzoeksvraag dekken en of die vragen wel logisch en consequent geformuleerd zijn. Een goed ontwikkelde vorm van wiskundig denken kan daar zeker bij helpen: ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ !

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

## NEDERLANDS PLATFORM COMPLEXE SYSTEMEN OPGERICHT

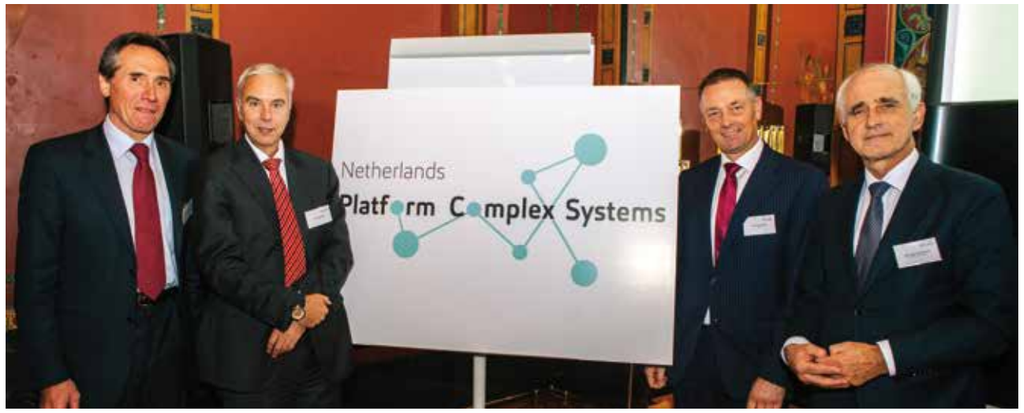
**Op 27 november 2015 vond in Utrecht de kick-off van het Nederlands Platform Complexe Systemen (NPCS) plaats. Tijdens deze speciale bijeenkomst gaven de initiatiefnemers (penvoerder Universiteit Utrecht, TNO en NWO) het officiële startsein voor het NPCS. Daarna volgden de voordracht van Marten Scheffer (WUR) en een video boodschap van Alexander Rinnooy Kan (UvA). Ten slotte vond er een matchmakingsessie plaats om vertegenwoordigers van bedrijfsleven en wetenschap nauw te laten samenwerken aan komende calls for proposals op het gebied van complexiteitsonderzoek over logistiek en agri-food.**

Complexiteitsonderzoek bestudeert verschijnselen die voortkomen uit het collectieve gedrag van een verzameling van veel interacterende agenten (objecten, onderdelen). Complexe systemen bestaan uit veel onderdelen, die allemaal op elkaar reageren. Hoewel die onderlinge interacties simpel en voorspelbaar zijn, ontstaat op de schaal van het hele systeem soms 'nieuw', niet te voorspellen gedrag. Door invloed van buitenaf kan een complex systeem, soms vrij plotseling, bovendien in heel nieuwe toestanden raken. Complexe systemen zijn overal om ons heen: grote zoals het klimaat, het internet en de wereldeconomie, en kleinere zoals lokale vervoersnetwerken en ons eigen lichaam en brein.

Een beter begrip van complexiteit kan ons leren signalen van omslagpunten in complexe systemen eerder te herkennen, en onvoorspelbare systemen op slimme manieren toch te beïnvloeden. Het zal ons in staat stellen complexe systemen te ontwerpen met uiterst nuttige toepassingen, zoals slimmere materialen of stabiele stroomnetwerken met duizenden kleine en wisselvallige producenten. Vanwege de complexiteit en de verschillende initiatieven is er een sterkte behoefte om de nationale afstemming op dit multidisciplinaire onderzoeksgebied in Nederland te stimuleren.

Daarom is het Nederlands Platform Complexe Systemen (NPCS) opgericht. Het platform moet er voor zorgen dat de onderzoekers die werken aan complexiteitsvraagstukken meer als *community* gaan opereren; wetenschappers van verschillende disciplines elkaar beter leren kennen, van elkaar leren en gaan samenwerken: dus een platform als voorbeeld van een complex systeem. Ook gaat het platform terreinen van complexiteitsonderzoek, waar Nederland bijzonder sterk in is en waar de maatschappelijke en industriële behoeften groot zijn, identificeren. Bovendien zal het platform een goede toegang bieden aan (potentiële) partners in het bedrijfsleven en de maatschappij tot de onderzoeksgemeenschap van complexiteitswetenschappers. Ten slotte zet het platform complexe systemen meer als een onderzoeksveld neer, en zorgen voor betere coördinatie zowel binnen Nederland als voor de inbreng van Nederland in het Europese complexiteitsonderzoek. Zo treedt het platform op als medeorganisator van de grootste internationale conferentie op het gebied van complexiteit, Conference on Complex Systems CCS'16 in Amsterdam. Voor meer informatie zie de website www.npcs.nl.

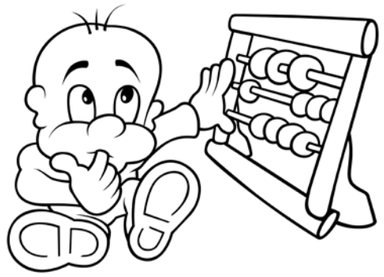
Wij wensen het platform veel succes.




(vanaf links) Prof. Peter Vervest (NWO programme committee Complexity, Erasmus University Rotterdam), Prof. Jos Keurentjes (Chief Scientific Officer TNO), Dr. Louis Vertegaal, (directeur NWO Physical Sciences) en Prof. Bert van der Zwaan (Rector Magnificus Utrecht University)



# Young Statisticians



### YOUNG STATISTICIANS DENMARK

We would like to introduce you to Young Statisticians Denmark, founded this year and partly inspired by the Dutch Young Statisticians as two of our members co-founded the organization. Just like us, they are also organizing interesting activities for young statisticians, for example statistical pub quizzes. So if you are thinking of a PhD or other position in Denmark, make sure to sign up with them! And maybe we will visit them some time...

### UPCOMING ACTIVITIES

➔January 21: Science Cafe
➔March/April: Pub Quiz
More information will follow soon!

### CONFERENTIE

### Dag voor de Statistiek & OR 2016
### 17 maart 2016

**De jaarlijkse Dag voor Statistiek en OR 2016 zal plaatsvinden op 17 maart 2016. Het thema zal zijn: THE ROLE OF STATISTICS IN DATA SCIENCE**

**Verdere bijzonderheden zult u kunnen vinden in het eerstvolgende nummer van *STAtOR* en op de website van de VvS+OR.**

**Reserveer alvast deze datum!**

EURO
The Association of European Operational Research Societies

## MAKING AN IMPACT
## EURO2016 FOR PRACTITIONERS

**In July 2016 up to 3000 OR professionals, from every part of the European OR community, will be gathering in the historic city of Poznan in Poland to take part in Europe's premier OR conference, EURO2016. Special emphasis will be laid on the importance of practical OR, with a strong 'Making an Impact' programme to provide plenty of value for practitioners.**

'Making an impact' gives you the opportunity to grow your professional skills, build your local and international network of contacts, be inspired by others, and contribute to the success of European OR practice, through activities expected to include case studies, promotion of academic-practitioner collaboration opportunities, workshops, mentoring, facilitated networking, and a careers exposition.

More information about 'Making an Impact' planned activities is on the website: http://www.euro2016.poznan.pl/making-an-impact/

**If your day-job is using OR, analytics, or anything similar to help organisations become more effective, put these dates in your diary now: 3rd – 6th July 2016 in Poznan.**


## BACK TO SCHOOL, LEARN ABOUT THE LATEST DEVELOPMENTS IN OPERATIONS RESEARCH

*LNMB Conferentie Lunteren, 14 januari 2016*

We leven in een 'genetwerkte' wereld: we zijn met elkaar verbonden via verschillende netwerken, terwijl de technologie die ons daartoe in staat stelt ook gebaseerd is op netwerken. Vanwege de overweldigende schaal van dergelijke netwerken is hun ontwerp, onderhoud en realisatie een niet-triviale taak. Daarnaast is in veel situaties de onderliggende structuur zeer dynamisch – denk aan de groei van de sociale netwerken.

Deze ontwikkeling levert nieuwe uitdagingen op het vakgebied operations research. De problemen die moeten worden aangepakt hebben vaak een combinatorisch element, als het gaat om ontwerpproblemen, of zijn van een meer algoritmische aard als het gaat om de activiteiten van het netwerk. Tegelijkertijd zijn de omstandigheden waarbinnen het netwerk opereert in hoge mate stochastisch (denk aan het gedrag van gebruikers, maar ook factoren zoals het weer, vraag en aanbod, en allerlei willekeurige gebeurtenissen). Reden waarom stochastische operations research hierbij een cruciaal instrument is.

Beide zijden van de OR zijn dan ook prominent aanwezig in het grote NWO-project *NETWORKS*, dat vorig jaar werd gelanceerd. Voor het komende LNMB / NGB seminar is een programma opgesteld met experts uit de academische wereld en de industrie, die OR-gereedschappen toepassen op verschillende soorten netwerken.

Michel Mandjes, coördinator van het *NETWORKS* project, opent de conferentie met het schetsen van het project en de programma-onderdelen. De sprekers zijn Rommert Dekker (Erasmus Universiteit), Ana Barros (TNO), Maaike Snelder (TNO en TUD), Jan van Doremalen (CQM), Richa Malhotra (SURFnet) en Pieter Cornelisse (KLM). Zij behandelen de op OR-gebaseerde onderzoeken van verschillende soorten netwerken, met toepassingen variërend van wegverkeer, communicatie en logistiek tot sociale netwerken. Zie voor meer informatie en inschrijving de website www.lnmb.nl/conferences/2016/.

**Tot ziens op 14 januari 2016 in Lunteren!**