

STATOR

periodiek van de VVSOR jaargang 20, nummer 3, september 2019

Math to the rescue; Optimalisatie in radiatietherapie

Automatisch corrigeren van inconsistenties in de officiële statistiek

Personal Health Train; Het analyseren van gefragmenteerde data

Examining cause-effect relations in the social sciences; A structural causal modelling approach

Inconsistenties in de Arbeidstijdenwet; Wiskunde toegepast in de weerbarstige praktijk

Over het nut van grondslagencrises; Filosofie en praktijk van data science



STATOR

Jaargang 20, nummer 3, september 2019

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

Redactie

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

Kopij en reacties richten aan

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

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl; Secretaris: dr. Laura Zwep MSc, db@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Algemeen: Nikky van Buuren MSc, webmaster@vvsor.nl

Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); prof. dr. Casper Albers (Social Sciences Section); dr. Michel van de Velden (Economics Section); Jonas Haslbeck MSc (Young Statisticians) Sanne Willems MSc (Section Statistics Communication).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Postbus 1058, 3860 BB Nijkerk, telefoon 033 2473408, admin@vvsor.nl. Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

Voor advertenties

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

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

HERFST

De hete zomer is kennelijk niet van invloed geweest op de productiviteit van onze auteurs. Voor u ligt het herfstnummer van STATOR dat maar liefst 40 pagina's telt met een grote verscheidenheid aan artikelen.

Wij hadden in het verleden al eens een artikel over 'slim plannen' van bestralingen, Koos van Amerongen schrijft nu over de nieuwste ontwikkelingen. Hij kreeg voor zijn werk op dit gebied tijdens de afgelopen Annual Meeting de Jan Hemelrijk Award uitgereikt.

Inconsistenties komen in dit nummer twee maal aan bod. Jacco Daalmans schrijft over zijn promotie-onderzoek naar het automatisch corrigeren van inconsistenties in de officiële statistiek en Laurens Fijn van Draat en anderen laten zien dat er inconsistenties voorkomen in de Arbeidstijdenwet.

Interessant is het artikel van Lianne Ippel en Johan van Soest over problemen die kunnen ontstaan als een onderzoeker zijn gegevens combineert met die uit verschillende bronnen, zoals het CBS.

Guillaume Wunsch et al. presenteren een door hen ontwikkeld framework over recente ontwikkelingen in Structural Causal Modelling. Wederom een artikel in het Engels, u hebt enkele nummers geleden kunnen lezen dat de redactie daar in voorkomende gevallen voor open staat.

Klaas van Harn en Piet Holewijn († 4 februari 2019) hebben onlangs een belangrijk boek gepubliceerd over de maattheoretische uitgangspunten van de waarschijnlijkheidsrekening. Dat was voor Richard Starmans aanleiding tot een zeer lezenswaardig artikel over grondslagen crises.

Naast informatie over allerlei verenigingszaken en een verslag van de Verolog Solver Challenge bieden wij u ter verstrooiing een drietal columns met zeer verschillende onderwerpen.

Rest ons te melden dat u vanaf dit nummer Johan van Leeuwen niet meer zult vinden in het colofon. Hij was in het verleden een actief columnist. Johan bedankt voor je vele fraaie bijdragen

Tot slot roepen wij alle lezers op actief mee te denken over de vernieuwing van het curriculum van het basis en middelbaar onderwijs. In het onderdeel Rekenen & Wiskunde wil men meer aandacht voor statistiek; het spreekt vanzelf dat wij als vereniging daar ook een inbreng in willen hebben.

De redactie wenst u veel herfstig leesplezier.



INHOUD

- 2 Redactioneel
- 4 Math to the rescue; Optimalisatie in radiatietherapie | KOOS VAN AMERONGEN
- 8 Oproep voor Nominaties Van Dantzig Prijs 2020
- 9 Automatisch corrigeren van inconsistenties in de officiële statistiek | JACCO DAALMANS
- 12 Personal Health Train; Het analyseren van gefragmenteerde data | LIANNE IPPEL & JOHAN VAN SOEST
- 16 Sinterklaas en de laatste lootjes – column | HENK TIJMS
- 18 Examining cause-effect relations in the social sciences; A Structural Causal Modelling approach | GUILLAUME WUNSCH, MICHEL MOUCHART & FEDERICA RUSSO
- 23 Statistiek vanaf de basisschool in het lespakket | JUDITH TER SCHURE
- 24 Impact – column | JOHN POPPELAARS
- 26 Inconsistenties in de Arbeidstijdenwet; Wiskunde toegepast in de weerbarstige praktijk | LAURENS FIJN VAN DRAAT, STAN DE GREEF, GERHARD POST & EGBERT VAN DER VEEN
- 30 Over het nut van grondslagen crises; Filosofie en praktijk van data science | RICHARD STARMANS
- 36 Nieuwe sectie: Statistics Communication
- 37 Een Zegen of een Ramp – column | GERRIT STEMERDINK
- 38 Verrassende ontknopning VeRoLog Solver Challenge 2019 | PIM VAN 'T HOF & JOAQUIM GROMICHO
- 39 Young Statisticians
- 39 Hemelrijk en Van Zwet Awards 2020: oproep
- 40 63rd ISI World Statistics Congress 2021 in Den Haag



MATH TO THE RESCUE

Optimalisatie in radiatietherapie

Kanker is wereldwijd een van de meest voorkomende doodsoorzaken. Dankzij snellere detectie en betere behandelingen is de kans om aan kanker te overlijden in de afgelopen decennia drastisch gedaald. Toch overlijdt nog altijd een op de zes mensen aan kanker. Op diverse vlakken worden wiskundige technieken onderzocht in een poging het gedrag van kankercellen beter te begrijpen en de behandeling van tumoren te verbeteren. *Math to the rescue!*

KOOS VAN AMERONGEN

Meer dan 50% van de met kanker gediagnostiseerde patiënten wordt behandeld met radiatietherapie, ook wel bestraling genoemd. Doorgaans gebeurt dit in combinatie met een of meerdere chirurgische ingrepen. Bij radiatietherapie worden tumorcellen gesteriliseerd door toediening van ioniserende straling. Meestal gebeurt dit door van buitenaf bundels straling op de patiënt te 'schie-

ten'. Deze stralingbundels worden gegenereerd door een lineaire accelerator (*linac*). De linac is gemonteerd aan een *gantry*, welke 360 graden rond de patiënt kan bewegen. Figuur 1 laat een voorbeeld zien van een moderne behandelmaschine.

Helaas beschadigt de ioniserende straling niet alleen de tumorcellen: alle cellen die in de lijn van de straal lig-

gen raken beschadigd. Dat leidt tot conflicterende doelen: het voldoende schade toebrengen aan de tumor en het ontzien van de gezonde weefsels daaromheen. De belangrijkste uitdaging bij het opstellen van een bestralingsplan is het in balans houden van deze doelen. Het bestralingsplan wordt opgesteld door een team van medisch specialisten, meestal bestaande uit een oncoloog, een dosimetrist en een fysicus.

Om tot een bestralingsplan te komen worden verschillende stappen doorlopen. Aan de hand van diverse scans wordt gezond en tumorweefsel geïdentificeerd en weergegeven in een anatomisch model van de patiënt. Op basis hiervan worden doelen en voorwaarden voor de hoeveelheid straling (dosis) voor elk risico lopend orgaan en de tumor vastgesteld. Ook wordt bepaald over hoeveel behandelingen (fracties) verspreid deze wordt toegediend en met welke frequentie. Het verspreiden over meerdere behandelingen is nuttig omdat gezond weefsel doorgaans sneller herstelt van de radiatieschade dan tumorweefsel. Vervolgens wordt gekeken hoe de linac en gantry ingesteld dienen te worden om tot een daadwerkelijk leverbaar en wenselijk bestralingsplan te komen.

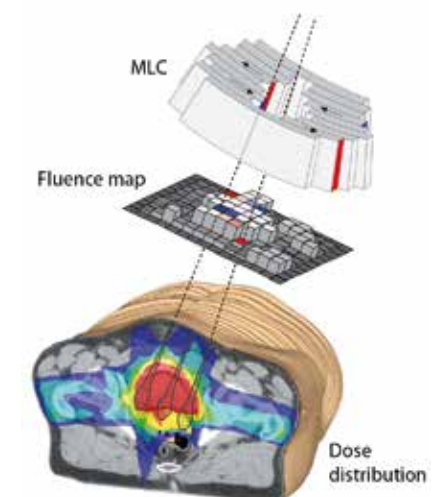


Figuur 1. De Clinac iX™ System Linear Accelerator van Varian Medical Systems, een van de machines die wordt gebruikt om een patiënt die op de horizontale plaat ligt, te bestralen; de straling komt vanuit de kop van de gantry.

Tot slot wordt het resulterende bestralingsplan uitvoerig getest en waar nodig aangepast.

Optimaliseren machine instructies

Er zijn verschillende methoden om van het gewenste plan naar een set van machine-instructies te komen. Tot eind vorige eeuw werd dat op basis van *trial-and-error* gedaan. Dankzij de ontwikkeling van wiskundige modellen kunnen de instructies voor de machine nu ook worden geoptimaliseerd. De belangrijkste instelbare variabelen zijn de vorm, de intensiteit en de richting van de straal op elk moment gedurende de behandeling. De straal wordt gevormd met behulp van een *multi-leaf collimator* (MLC), bestaande uit meerdere paarsgewijs tegenover elkaar staande, even dikke bladen (figuur 2). De bladen – meestal van wolfram gemaakt – laten geen straling door



Figuur 2. De MLC-bladen (ver)vormen de straal door een opening te creëren. Door de positie van de bladen en de intensiteit van de straal te variëren wordt de toegediende dosis over het behandelgebied verdeeld. De uiteindelijke geleverde dosisverdeling wordt middels een *fluence map* gemodelleerd. Bron: RaySearch Laboratories

en kunnen onafhankelijk van elkaar horizontaal bewegen. Samen kunnen de bladen de straal op vele manieren vormen, zie figuur 2.

Het is voor het menselijk brein schier onmogelijk om in te beelden hoeveel dosis er waar zou worden geleverd als de instelbare variabelen continu veranderen. Daarom maakte men voorheen gebruik van de *step-and-shoot* methode, waarbij de straal werd onderbroken bij het herpositioneren van de linac en de MLC-bladen. Dat maakte het mogelijk om vooraf voor elk stukje van het menselijk lichaam (voxel) te berekenen wat de totale hoeveelheid ontvangen straling (dosis) zou zijn voor een gegeven bestralingsplan. Een nadeel was dat veel vrijheidsgraden onbenut bleven.

Met de komst van geavanceerde dosiscalculatiemodellen werd het mogelijk om de dosis per voxel uit te rekenen voor de dynamische bestralingsvorm, waarbij de linac aan blijft gedurende het herpositioneren van de MLC-bladen en de gantry. Deze dynamische vorm wordt *Volumetric Modulated Arc Therapy* (VMAT) genoemd. De oplossingsruimte van de instelbare variabelen is door vele non-convexe afhankelijkheden vele malen groter en veel complexer gevormd voor deze behandelvorm dan dat voor de *step-and-shoot* methode. Daarom wordt het optimalisatieproces vaak opgesplitst in meerdere stappen. Dat kan op diverse manieren en leidt tot interessante subproblemen en relaties daartussen.

Een belangrijk subprobleem binnen de zogenaamde *beamlet based approach* is het *leaf sequencing* probleem: gegeven een te leveren intensiteitsprofiel (*fluence map*), hoe kunnen we alle linker en rechter MLC-bladen dusdanig manoeuvreren en de intensiteit van de straling dusdanig variëren dat het te leveren intensiteitsprofiel zo goed mogelijk wordt nagebootst, rekening houdend met de beperkingen van de machine? Een intensiteitsprofiel geeft aan hoeveel straling er uiteindelijk overal net achter de MLC-bladen is doorgelaten, wat representeert hoe de dosis over de behandelzone wordt verdeeld, zie figuur 2. Het intensiteitsprofiel wordt vaak gediscrediteerd in bixels, waarvan de hoogte overeenkomt met de dikte van de bladen en de breedte gekozen wordt aan de hand van de complexiteit van de tumor. De machinebeperkingen waar rekening mee moet worden gehouden zijn de maximale straalintensiteit van de linac, de maximale beweegselheid van de MLC-bladen en de voorwaarde dat linker- en rechterbladen niet mogen botsen.

Er bestaan verschillende algoritmes die dit optimaliseringsprobleem optimaal kunnen oplossen. Een van de eerste en nog altijd meest populaire algoritmes is *het sliding-window leaf-sweep* (SWLS) algoritme (Stein, Bortfeld,

Dörschel, & Schlegel, 1994). In dit algoritme wordt de straalintensiteit gefixeerd op het maximale niveau, waarna het resterende probleem – het optimaliseren van de MLC-bladen – paarsgewijs ontkoppelt. Door de bladen vervolgens in één richting te *sweepen*, zeg van links naar rechts, kan het bij dat paar horende deel van de fluence map perfect worden nagebootst. Het fijne aan dit algoritme is dat er een analytische formule voor de positie van elk blad en de stralingsintensiteit op elk moment uit volgt. Ook voor de totaal benodigde behandeltijd (T_{SWLS}) bestaat een analytische formule.

Verkorten behandeltijd

Een van de belangrijkste tekortkomingen van deze formulering van het *leaf sequencing* probleem is dat het geen waarde hecht aan de levertijd van de stralingsbehandeling, terwijl dat om uiteenlopende redenen wel een belangrijk aspect is (Kelly, van Amerongen, Balvert, & Craft, 2019). Zo heeft de patiënt baat bij een korte levertijd, aangezien het negatieve effect van diens bewegen op de accuraatheid van de behandeling stijgt in de tijd van blootstelling. De kliniek heeft baat bij het verkorten van een behandeling, aangezien zo meer mensen geholpen kunnen worden op de aanwezige machines. Dat is in het bijzonder relevant voor klinieken in ontwikkelingslanden, omdat één machine tot wel twee miljoen euro kan kosten. In het algemeen is er een *trade-off*-relatie tussen de behandeltijd en behandelkwaliteit. Gezien het belang van beiden en de schaal waarop radiatietherapie wordt toegepast, is het waardevol om ervoor te zorgen dat een bestralingsplan Pareto-optimaal is.

Door een maximale levertijd mee te nemen in de formulering van het *leaf sequencing* probleem en deze vervolgens voor meerdere levertijden op te lossen kan de relatie tussen levertijd en -kwaliteit worden benaderd. Balvert en Craft (2017) tonen aan dat het in één richting bewegen van de bladen en het fixeren van de straalintensiteit op het maximale niveau, zoals het SWLS-algoritme doet, kan leiden tot suboptimaliteit. Daarom formuleren zij een model dat deze restricties niet oplegt, welke in Van Amerongen (2017) uitvoerig wordt besproken. Deze formulering wordt gekenmerkt door twee bijzondere eigenschappen: de oplosruimte is lineair terwijl de doelfunctie om twee redenen non-convex is: de vermenigvuldiging van de bladposities en bestralingsintensiteit en de non-convexe relatie tussen bladposities en bixelblootstelling.

Balvert en Craft (2017) pakken dit probleem aan door vanuit diverse startpunten middels de *interior point*

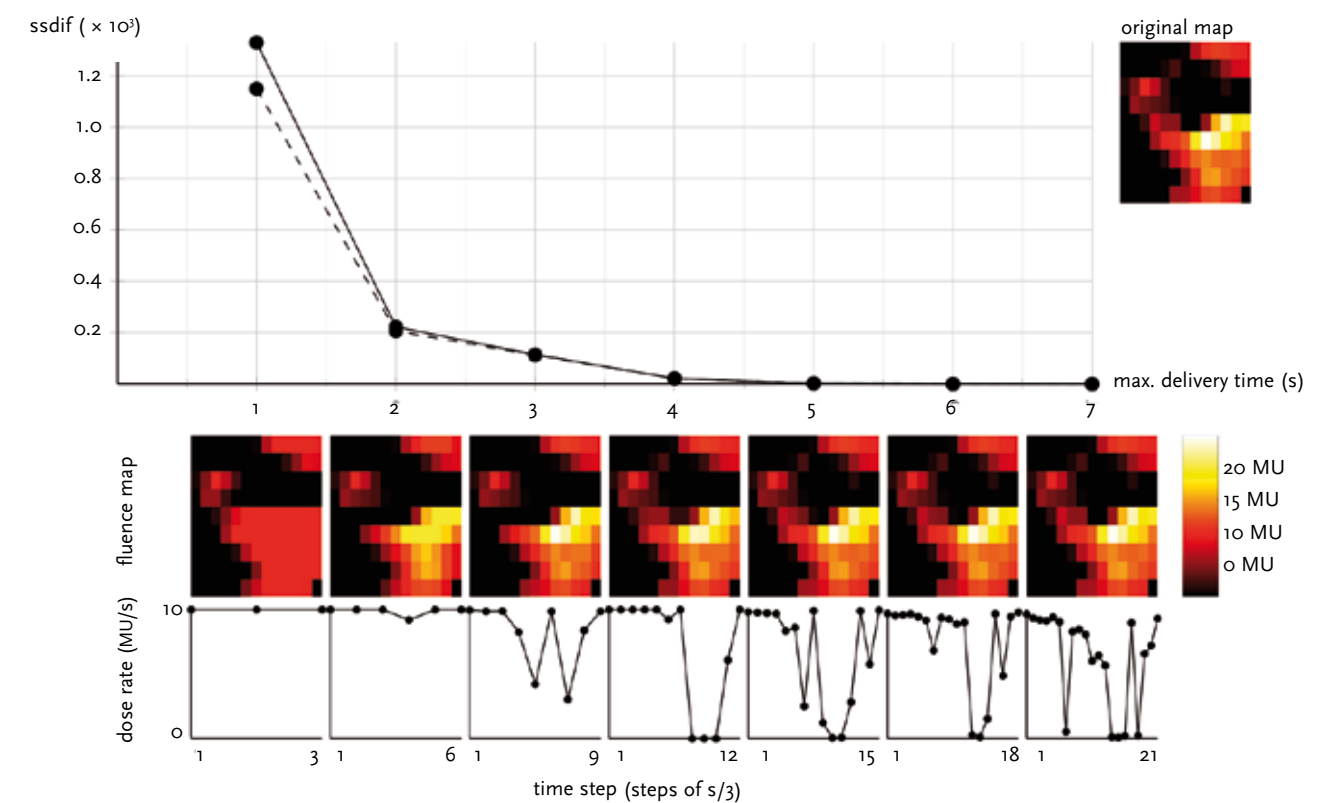
method in MATLAB's *fmincon* functie naar lokale optima af te dalen. Hierbij maken ze gebruik van een analytisch bepaalde gradiënt en Hessiaan. Aangezien de doelfunctie stabiliseert naarmate het probleem vanuit meerdere startpunten wordt opgelost, vermoeden zij dat de uiteindelijke kwaliteit dicht bij het optimum ligt. Door de afwezigheid van goede ondergrenzen is onbekend hoe dichtbij. Het grootste nadeel van deze methode is dat de berekentijd te groot is om in de praktijk bruikbaar te zijn. Dit komt doordat het lang duurt voordat *fmincon* vanuit een startpunt een lokaal optimum heeft gevonden.

Versnellen optimalisatie

Voor mijn masters thesis heb ik onderzoek gedaan naar hoe deze optimalisatieprocedure versneld kan worden. Hierbij heb ik eerst gekeken naar globale zoektechnieken zoals *simulated annealing*, maar deze deden er ook te lang over om een globaal optimum te vinden, als ze daar al bij in de buurt kwamen. Uiteindelijk bleek het

vinden van goede startoplossingen de meeste tijdsbesparing op te leveren. Hiervoor heb ik gebruik gemaakt van probleemspecifieke eigenschappen. Een voorbeeld daarvan is dat met een constante stralingsintensiteit en met vaste start- en eindpunten van alle MLC-bladen het probleem zodanig opgesplitst en omgeschreven kan worden dat een reeks eenvoudig tot optimaliteit oplosbare lineaire problemen overblijft. Over een verzameling van op echte patiëntendata gebaseerde *fluence maps* leverde dat berekentijdsbesparingen op van tussen de 97,3 en 99,5 procent, voldoende om het in de praktijk toepasbaar te maken. Daarnaast zijn er passant een aantal vondsten gedaan die hebben bijgedragen aan het begrijpen van (de complexiteit van) het leaf sequencing probleem.

Ter illustratie een voorbeeld van hoe de uiteindelijke trade-off tussen levertijd en leverkwaliteit, de bijhorende gerealiseerde fluence maps en stralingsintensiteitspatroon er uit zien voor een fluence map die gebaseerd is op een casus van een patiënt met prostaatkanker. Figuur 3 laat zien dat het mogelijk is om in minder dan T_{SWLS} (6,3 sec voor deze map) een bestralingsplan te



Figuur 3. Gevonden relatie tussen de maximale levertijd van een gewenste *fluence map* (rechtsboven) en de mate waarin deze kon worden nagebootst. De solide en gestreepte lijnen corresponderen met Van Amerongen (2017) respectievelijk Balvert en Craft (2017). Onder de maximale levertijden staan voor eerstgenoemde de geleverde *fluence maps* en bijhorende straalintensiteit versus tijd

krijgen dat nauwelijks slechter is dan het optimale plan. Voor kleine levertijden is het belangrijk om zo veel mogelijk dosis af te leveren, is de straalintensiteit (*dose rate*) maximaal en wordt alleen de contour van de gewenste map zichtbaar. Bij middelgrote levertijden is de geleverde map accurater en gaat de straalintensiteit af en toe omlaag zodat de MLC-bladen kunnen herpositioneren zonder te hoeven sluiten om geen straling door te laten.

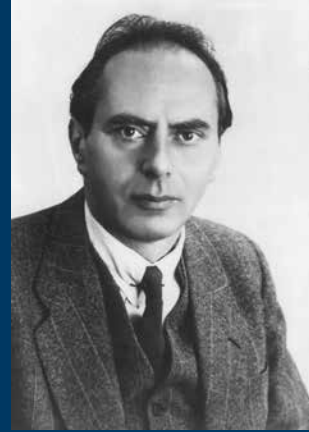
Dankzij dit onderzoek is het nu mogelijk om de tijd dat een patiënt is blootgesteld aan radiatie mee te nemen als een van de aspecten in het optimaliseren van het bestralingsplan van een met externe straling behandelde kankerpatiënt. Meer onderzoek is gaande om dit in commerciële software te kunnen integreren. Het leaf sequencing probleem is slechts een van de vele integrerende optimalisatieproblemen die zich binnen de radiatietherapie voordoen. Velen daarvan zijn nog onopgelost. Ben je benieuwd welke en heb je interesse om er een te tackelen? Neem dan gerust contact met me op.

LITERATUUR

- Balvert, M., & Craft, D. (2017). Fast approximate delivery of fluence maps for IMRT and VMAT. *Physics in Medicine and Biology*, 62(4), 1225–1247.
- Kelly, M., Van Amerongen, J. H. M., Balvert, M., & Craft, D.L. (2019). Dynamic fluence map sequencing using piecewise linear leaf position functions. *Biomedical Physics & Engineering Express*, 5(2), doi:10.1088/2057-1976/aaffe7
- Stein, J., Bortfeld, T., Dörschel, B., & Schlegel, W. (1994). Dynamic X-ray compensation for conformal radiotherapy by means of multi-leaf collimation. *Radiotherapy and Oncology*, 32(2), 163–173.
- Van Amerongen, J. H. M. (2017). *Fast approximate delivery of fluence maps in volumetric modulated arc therapy* (unpublished master's thesis). Tilburg: Tilburg University.

KOOS VAN AMERONGEN werkt als analist bij de innovatieve onderwijssteuner Lyceo. Hij heeft zijn master Operations Research en Management Science gedaan aan Tilburg University. Dit artikel is geschreven op basis van zijn afstudeeronderzoek aan Massachusetts General Hospital / Harvard Medical School, waarvoor hij de Jan Hemelrijk Award 2019 kreeg. Hij is gespecialiseerd in optimalisatie en geïnteresseerd in sociaal relevante uitdagingen, zoals voorkomend in de gezondheidszorg en in het onderwijs. E-mail: koosvanamerongen@hotmail.com

Oproep voor nominaties VAN DANTZIGPRIJS 2020



De VVSOR reikt eenmaal in de vijf jaar de Van Dantzig Prijs uit, de hoogste prijs in de Nederlandse Statistiek en Operations Research. Deze prijs is ingesteld ter nagedachtenis aan prof. dr. David van Dantzig (1900 – 1959), de grondlegger van de mathematische statistiek in Nederland. De eerstvolgende uitreiking zal plaatsvinden tijdens de Annual Meeting van de VVSOR in 2020. De prijs bestaat uit een in brons gegoten medaille met de beeltenis van Van Dantzig in reliëf.

Genomineerden voor de Van Dantzigprijs 2020 zijn op 1 januari 2020 niet ouder is dan 40 jaar en hebben de afgelopen vijf jaar een zeer noemenswaardige bijdrage – hetzij theoretisch, hetzij praktisch van aard – geleverd aan de statistiek of operations research.

Eerdere winnaars van de Van Dantzigprijs zijn Willem van Zwet (1970), Ton van Meurs (1975), Arie Hordijk (1980), Alexander Rinnooy Kan (1985), Richard Gill (1990), Geert Ridder (1995), Aad van der Vaart (2000), Sem Borst & Mark van der Laan (2005), Peter Grünwald & Harry van Zanten (2010) en Bert Zwart (2015).

Voel u vrij om deze oproep te verspreiden naar mogelijk geïnteresseerden. Nominaties en vragen over het proces kunnen worden gericht aan het dagelijks bestuur van de VVSOR, c.q. de jury voor de prijs, via [<vandantzig@vvsor.nl>](mailto:vandantzig@vvsor.nl).



AUTOMATISCH CORRIGEREN van inconsistenties in de officiële statistiek

Officiële statistieken moeten consistent zijn. Dat betekent dat statistische resultaten hetzelfde moeten zijn in verschillende publicaties. Consistente uitkomsten ontstaan niet vanzelf. Vaak zijn er meerdere bronnen beschikbaar voor één statistiek. Afhankelijk van de gekozen bron kan er iets anders uitkomen. Het proces om tot consistente uitkomsten te komen heet ook wel macro-integratie. In dit artikel wordt uitgelegd dat in toenemende mate wiskundige methoden worden ingezet voor macro-integratie.

JACCO DAALMANS

Stel je bezoekt met vrienden een café. De rekening wordt gesplitst. Iedereen legt een bedrag op tafel dat hij of zij denkt te hebben besteed. Na telling blijkt dat er te weinig geld ligt. Of anders gezegd: de hoeveelheid geld is niet 'consistent' met het bedrag op de rekening. Meestal wordt dan eerst gecontroleerd of er een fout is gemaakt. Heeft de ober te veel drankjes aangeslagen of zijn de be-

zoekers een schaal bitterballen vergeten? Als de oorzaak niet wordt gevonden dan kan een ingewikkelde discussie ontstaan, waarna alsnog één bedrag wordt afgerekend.

In de officiële statistiek speelt een vergelijkbaar probleem. Met 'officiële statistiek' worden cijfers bedoeld die door een officiële instantie, zoals het Centraal Bureau voor de Statistiek (CBS), worden gepubliceerd. Officië-

le statistieken moeten hetzelfde zijn in alle publicaties waarin deze voorkomen. Het aantal werklozen mag niet 300.000 zijn in de ene publicatie en 400.000 in een andere, uitgaande van een gelijke definitie van werkloosheid. Een verschil in uitkomsten zorgt voor verwarring en dit strookt niet met het doel van statistische bureaus om onbetwistbare statistiek te leveren.

Meer in het algemeen betekent consistentie dat statistieken aan onderlinge relaties voldoen. Een voorbeeld is dat twaalf maandcijfers optellen tot één jaarcijfer. Als hier niet aan wordt voldaan, kan een gebruiker naast het gegeven jaarcijfer, een alternatief jaarcijfer afleiden door maandcijfers op te tellen. Zo ontstaat verwarring over het 'ware' jaarcijfer.

Consistentie van statistische output ontstaat niet vanzelf. Data die een statistisch bureau verzamelt komen uit tal van bronnen, ieder met hun eigen onzuiverheden. Zo komen steekproeffouten, nonresponsfouten en meetfouten voor. Macro-integratie is een proces om de verschillende weg te werken. De uitkomsten van verschillende statistieken worden gecorrigeerd, om ze beter op elkaar af te stemmen. Men corrigeert bijvoorbeeld maandcijfers, om te zorgen dat twaalf maandcijfers optellen tot één jaarcijfer. Men zou kunnen denken dat door macro-integratie 'fouten' ontstaan, omdat de resultaten afwijken van de afzonderlijke bronnen. Dit is echter niet het geval, omdat we te maken hebben met bronnen met onzuiverheden. Door het confronteren van de verschillende bronnen, kan juist een nauwkeuriger beeld ontstaan dan door bronnen in afzondering te beschouwen.

Een toepassing van macro-integratie heeft gelijkenissen met het gesplitst betalen van een rekening. Eerst worden grote verschillen, met een aanwijsbare oorzaak, gecorrigeerd, daarna de overgebleven, kleinere verschillen. Hieronder gaat de aandacht alleen uit naar de tweede stap. We zullen zien dat wiskundige methoden hiervoor van grote waarde zijn en dat deze steeds meer worden toegepast in de officiële statistiek.

Nationale Rekeningen

Een traditionele toepassing van macro-integratie vindt plaats bij de Nationale Rekeningen. De Nationale Rekeningen bestaan uit een aantal zeer gedetailleerde tabellen, die het economisch proces van een land beschrijven.

In die tabellen worden macro-economische indicatoren, zoals productie, consumptie, investeringen, import en export gepubliceerd voor een groot aantal goederen en diensten. Dit gebeurt zowel op kwartaal- als op jaarbasis. De tabellen worden samengesteld uit tal van bronnen. De integratie van de data is lastig, vanwege de grote samenhang tussen cijfers. Wanneer één cijfer moet worden aangepast, betekent dit doorgaans ook dat veel gerelateerde cijfers moeten worden gewijzigd.

Van oudsher werden informele methoden toegepast om Nationale Rekeningen consistent te maken. Dat betekent dat experts correcties aanbrengen in grote tabellen op basis van vakinhoudelijke kennis. Hoewel zo'n aanpak meestal goed werkt, is de toepassing arbeidsintensief. Bovendien zijn de uitkomsten lastig te reproduceren. Als alternatief voor informele methoden, zijn er ook formele methoden beschikbaar, gebaseerd op een wiskundige methode. Formele methoden hebben als voordeel dat deze reproduceerbare uitkomsten geven. Het mechanisme van de correcties ligt immers vast.

Wiskundige methoden

Een bekende wiskundige methode voor integratie van Nationale Rekeningen is die van Stone et al. (1942). Wiskundig gezien is dit een gewogen kwadratisch optimaliseringsprobleem met lineaire restricties. De lineaire restricties geven de eisen weer waar cijfers aan moeten voldoen. De doelfunctie minimaliseert een (gewogen) som van correcties, die men moet uitvoeren om aan de restricties te voldoen.

Een belangrijke eigenschap van Stone's methode is dat betrouwbaarheids gewichten mee kunnen worden genomen. Dit maakt het mogelijk om de meest nauwkeurige cijfers het minst aan te passen en vice versa. Niet alleen sluit zo'n aanpak het best aan bij onze intuïtie, maar ook wiskundig is te bewijzen dat op die manier de meest nauwkeurige resultaten worden verkregen. Dit betekent echter wel dat een inschatting moet worden gemaakt over de nauwkeurigheid van de broncijfers.

Naast de methode van Stone bestaan er nog veel andere methoden in de literatuur, met verschillende toepassingsmogelijkheden. Voor het consistent maken van reeksen kwartaal- en jaarcijfers is bijvoorbeeld de methode van Denton (1971) beschikbaar.

De methoden, die in de vorige eeuw zijn ontwikkeld, zijn niet meteen grootschalig toegepast op statistische bureaus. Eén van de redenen hiervoor is de stand van de IT. Toepassing van een automatische data-integratie methode op de aanbod- en gebruik tabellen, zoals die momenteel op het CBS worden samengesteld, betekent bijvoorbeeld dat een optimaliseringsprobleem met meer dan 500.000 variabelen moet worden opgelost. Dit stelt eisen aan de software. Tegenwoordig is efficiënte software beschikbaar voor het oplossen van grote kwadratische optimaliseringsproblemen. Deze software kan relatief eenvoudig op een standaard PC worden toegepast. Een optimaliseringsprobleem met 500.000 variabelen vormt geen enkel probleem. Voorbeelden van commerciële software voor kwadratische optimalisering zijn: CPLEX, XPRESS, GARUBI en AIMMS.

Een tweede beperkende factor voor toepassing van automatische methoden bestaat uit de begrensde mogelijkheden van de methodologie. Zo kunnen de meest eenvoudige methoden alleen lineaire gelijkheidsrestricties aan. Voor de toepassing bij de Nationale Rekeningen is het nodig om een breder scala van relaties te kunnen modelleren. Een analyse van het integratieproces wees uit dat in ieder geval de volgende drie soorten van functionaliteiten toepasbaar moeten zijn.

Ten eerste is het nodig om ongelijkheidsrestricties mee te nemen, zoals de niet-negativiteitsrestrictie. Veel economische cijfers kunnen geen negatieve waarde aannemen. Dit moet men dan ook kunnen afdwingen in een automatische methode.

Ten tweede moet het mogelijk zijn om zogenaamde ratio-restricties op te stellen. In economische rekeningenstelsels zijn verhoudingen tussen twee cijfers vaak erg belangrijk. Een voorbeeld is de verhouding tussen productie van een bepaald goed en het verbruik van de benodigde grondstoffen. Voor de productie van 1 kg kaas is bijvoorbeeld 10 liter melk nodig. Het is nodig dat men restricties op kan leggen aan de uitkomsten van een ratio.

Ten derde moet de mogelijkheid bestaan om zogenaamde zachte relaties mee te nemen. Het gaat namelijk om relaties die sturend zijn, dus niet bindend. Of in andere woorden, die bij benadering moeten gelden. Een voorbeeld is dat verhouding tussen lonen en productie ongeveer hetzelfde moet zijn als vorig jaar.

Nieuwe methode

In een hoofdstuk van mijn proefschrift (Daalmans 2019) beschrijf ik, met Reinier Bikker en Nino Mushkudiani, een uitbreiding van de oorspronkelijke Denton methode (Bikker et al. 2013). De nieuwe methode combineert verschillende functionaliteiten van bestaande methoden en is specifiek gericht op de behoeften van de Nationale Rekeningen. De drie soorten van relaties, die hierboven zijn genoemd, kunnen bijvoorbeeld worden meegenomen.

De voorgestelde methode wordt al enige tijd daadwerkelijk toegepast voor de integratie van Nationale Rekeningen op het CBS. Daarmee is een groter deel van het productieproces geautomatiseerd en is de reproduceerbaarheid van de officiële statistiek vergroot. Hoewel de uitgebreide Denton specifiek is ontworpen voor de Nationale Rekeningen, is de methode ook toepasbaar voor andere statistieken waarin consistente tabellen moeten worden samengesteld. Het CBS past een vergelijkbare methode bijvoorbeeld ook toe voor energiestatistieken. De methodologie voor macro-integratie wordt steeds verder uitgebreid. Hierdoor ontstaan ook steeds nieuwe toepassingsmogelijkheden. Het splitsen van een caférekening zullen we echter zelf moeten blijven doen.

LITERATUUR

- Bikker, R. P., Daalmans, J. A., & Mushkudiani, N. (2013). Benchmarking large accounting frameworks: a generalized multivariate model. *Economic Systems Research*, 25, 390–408.
- Daalmans, J. A. (2019). *Pushing the boundaries for automated data reconciliation in official statistics* (PhD Thesis). Tilburg University, Tilburg.
- Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An Approach based on quadratic minimization. *Journal of the American Statistical Association*, 66(333), 99–102.
- Stone, J. R. N., Champerowne, D. G., & Meade, J. E. (1942). The Precision of National Income Estimates. *Reviews of Economic Studies*, 9, 111–135.

JACCO DAALMANS is methodoloog bij het Centraal Bureau voor de Statistiek. In maart 2019 is hij gepromoveerd aan de Universiteit van Tilburg, op een proefschrift dat onder andere gaat over het onderwerp van dit artikel.
E-mail: j.daalmans@cbs.nl

PERSONAL HEALTH TRAIN

Het analyseren van gefragmenteerde data

Een voorwaarde voor een correcte schatting van een effect is dat alle relevante factoren zijn meegenomen in de analyse. Wanneer er een factor is weggelaten, zullen de schattingen van een effect vertekend zijn. Bijvoorbeeld in een analyse naar de effectiviteit van een medicijn, kan een vergeten factor ertoe leiden dat een medicijn veel effectiever lijkt dan werkelijk het geval is. Echter, als onderzoeker heb je niet altijd alle relevante factoren tot je beschikking. In zo'n situatie kan een onderzoeker ervoor kiezen om de data van bijvoorbeeld het Centraal Bureau voor de Statistiek (CBS) te koppelen aan de data van de onderzoeker om op deze manier meer relevante factoren op te nemen in de analyse. Om zo'n koppeling tot stand te brengen moet er eerst een aantal hobbels worden genomen. In dit artikel zullen de uitdagingen en oplossingen worden besproken.

LIANNE IPPEL & JOHAN VAN SOEST

Over het algemeen spreken we van twee verschillende soorten van fragmentatie: horizontaal of verticaal. Horizontale fragmentatie betekent dat meerdere instellingen, bijvoorbeeld ziekenhuizen, dezelfde gegevens verzamelen over hun patiënten, zoals in tabel 1. Een analyse gebaseerd op deze gefragmenteerde data is eenvoudig omdat de analyseresultaten van elk ziekenhuis achteraf samengevoegd kunnen worden, zonder de privacy van de patiënten te schenden.

Zo eenvoudig is het echter niet wanneer we de gegevens van Ziekenhuis A willen koppelen aan gegevens van het CBS, zoals in tabel 2. In dit geval spreken we van verticaal gefragmenteerde data: Ziekenhuis A en het CBS hebben beide een ander deel van de gegevens van dezelfde mensen. Verticaal gefragmenteerde data zijn complex omdat we eerst moeten achterhalen welke gegevens van het CBS gekoppeld kunnen worden aan de gegevens

van het ziekenhuis. Dit wordt nog extra bemoeilijkt omdat bij wet bepaald is dat de wetenschap geen gebruik mag maken van het Burgerservicenummer. Hierdoor is een unieke identificatie tussen de ziekenhuisgegevens en de CBS-gegevens onzeker. Daar komt bij dat het CBS gegevens bezit van veel meer individuen dan het ziekenhuis waardoor de kans van een foutieve koppeling groter is. Als we bijvoorbeeld geboortedatum gebruiken om de gegevens van een persoon van het ziekenhuis te koppelen aan de gegevens van het CBS, is de kans groot dat er meer mensen bekend zijn bij het CBS die op dezelfde datum geboren zijn. We hebben dan geen unieke koppeling en meer informatie is nodig om de juiste en unieke koppeling te vinden. Ten slotte staat de CBS-wet niet toe dat CBS-data uit hun faciliteiten gehaald worden, wat betekent dat de analyse altijd op de infrastructuur van het CBS gedaan zou moeten worden.



Het combineren van medische gegevens van een ziekenhuis met gegevens van het CBS kan dus alleen op een veilige manier gebeuren als de juiste juridische en ethische kaders aanwezig zijn die rekening houden met bijvoorbeeld de wet Algemene Verordening Gegevensbescherming (AVG), de CBS-wet, ethische aspecten van het onderzoek en daarnaast een technische infrastructuur die het mogelijk maakt om factoren van verschillende instellingen te combineren voor het onderzoek. Vervolgens moeten de data niet alleen juridisch maar ook praktisch herbruikbaar zijn voor onderzoek.

De FAIR-richtlijnen

In een notendop staat FAIR voor het bevorderen van het hergebruiken van data, voor de juiste doeleinden en

onder de juiste voorwaarden. FAIR staat voor *Findable*, *Accessible*, *Interoperable* en *Reusable* (vindbaar, toegankelijk, voor meerdere doelen geschikt, en herbruikbaar) (Wilkinson et al., 2016).

Findable betekent dat alle digitale objecten een uniek identificatielabel krijgen waardoor er aan zo'n object gerefereerd kan worden. Een voorbeeld is DOI (Digital Object Identifier) een unieke code waarmee we digitale objecten kunnen vinden. De data moeten daarnaast een duidelijke beschrijving hebben zodat anderen ook begrijpen wat voor data ze gevonden hebben.

Accessible betekent dat er een protocol bestaat dat beschrijft hoe iemand toegang kan krijgen tot een digitaal object. Dit wil *niet* zeggen dat alle data al-

ZIEKENHUIS A					
Patiënt nr.	Leeftijd	Geslacht	Diabetes Type 2	Behandeling met diëtist	Roken
90123	45	M	Nee	Ja	Nee
90124	66	M	Ja	Ja	Ja
90125	51	V	Ja	Nee	Ja
90126	47	M	Nee	Nee	Ja
90127	63	V	Ja	Nee	Ja

ZIEKENHUIS B					
Patiënt nr.	Leeftijd	Geslacht	Diabetes Type 2	Behandeling met diëtist	Roken
298751	85	V	Nee	Nee	Ja
298752	71	M	Nee	Ja	Nee
298753	76	V	Ja	Nee	Nee
298754	41	M	Nee	Ja	Ja
298755	76	M	Ja	Nee	Ja
298756	54	V	Nee	Ja	Nee

Tabel 1. Horizontale fragmentatie van data: Ziekenhuis A en Ziekenhuis B hebben verschillende patiënten maar beide ziekenhuizen verzamelen dezelfde gegevens over hun patiënten

ZIEKENHUIS A						CENTRAAL BUREAU VOOR DE STATISTIEK				
Patiënt nr.	Leeftijd	Geslacht	Diabetes Type 2	Behandeling met diëtist	Roken	Id nr.	Leeftijd	Geslacht	Jaarinkomen euro	Eigenaar koopwoning
90123	45	M	Nee	Ja	Nee	9020983	45	M	16.000-18.000	Nee
90124	66	M	Ja	Ja	Ja	9020936	66	M	40.000-42.000	Ja
90125	51	V	Ja	Nee	Ja	9027503	51	V	34.000-36.000	Nee
90126	47	M	Nee	Nee	Ja	9024947	47	M		Ja
90127	63	V	Ja	Nee	Ja
					
					
						9054903	63	v	44.000-46.000	Ja

Tabel 2. Verticaal gefragmenteerde data: Ziekenhuis A en het CBS hebben gegevens van dezelfde mensen, maar ze verzamelen andere gegevens

tijd en voor iedereen beschikbaar zijn. De eigenaar van de data is te allen tijden verantwoordelijk dat de verzamelde data geen schade aan zullen richten aan degene van wie de data verzameld zijn. Daarom kan de eigenaar van de data beslissen om delen van de data af te schermen.

Interoperable betekent dat de data ook voor een ander doel gebruikt kunnen worden dan waar ze oorspronkelijk voor verzameld zijn. De data moeten daarom op zo'n manier gecodeerd worden dat de beschrijving van de data 'machine readable' is. Dit maakt het makkelijker om de computer een zoekopdracht te geven en alle relevante data te vinden. Daarnaast moeten de data opgeslagen zijn in een formaat wat te openen is door meerdere programma's zoals '.csv' in plaats van '.xlsx' omdat deze laatste alleen door het programma MS Excel te openen is.

Ten slotte *Reusable*, de herbruikbaarheid van de data, is de combinatie van de bovenstaande aspecten, met de juiste licentie zodat een andere onderzoeker weet wat toegestaan is om met de data te doen.

Data-stations, Treinen en Spoorwegen

Wanneer de juiste data gevonden zijn, en het ook juridisch en ethisch toegestaan is om de data te koppelen, kan het nog steeds voorkomen dat de data niet te downloaden zijn zoals bij CBS-data. Door middel van een nieuw initiatief, *Personal Health Train*, kunnen analyses naar verschillende data-partijen gestuurd worden zonder dat onderzoekers de data op hun eigen computer laden.

Het idee van de *Personal Health Train* is om in plaats van de data te verplaatsen naar de analyse, de analyse naar de data te brengen (Dutch tech center for life sciences, 2017). Dit heeft een aantal voordelen, bijvoorbeeld het beter beschermen van privacygevoelige informatie. Onderzoekers sturen namelijk alleen hun onderzoeksvraag en methode naar de data, en krijgen een antwoord

terug, in plaats van dat zij de data van individuele personen inzien. Daarnaast is het verplaatsen van grote databestanden niet efficiënt: er ontstaan vele kopieën van hetzelfde databestand. Dit maakt het uitvoeren van de nieuwe AVG-wet nagenoeg onmogelijk. Als er echter maar één beschikbare versie van een databestand is, wat in een FAIR data-station staat, kunnen gegevens gemakkelijk aangevuld worden. Er zijn 'spoorwegen' die toegang kunnen verlenen tot deze stations. De spoorwegen zorgen ervoor dat onderzoekers alleen toegang tot de gegevens kunnen krijgen waar zij toestemming voor hebben. Wanneer iemand toegang heeft om de gegevens in het FAIR-data-station te analyseren kan de 'analysetrein' bij het station komen. Het station is daarmee ook verantwoordelijk dat alleen de toegestane informatie (resultaten) naar de onderzoeker worden verstuurd. Op deze manier kan de dataeigenaar zelfs na analyse toestemming voor toekomstig gebruik intrekken.

Applicatie

In het project FAIRHealth (Sun et al., 2018) – onderdeel van de route Verantwoordelijk Waardecreatie met Big Data (VWdata), een startimpulsprogramma van de Nationale Wetenschapsagenda – ontwikkelen we zowel technische als beleidsmatige oplossingen om gegevens afkomstig van De Maastricht Studie en het CBS in combinatie te analyseren. De Maastricht Studie verzamelt medische gegevens van participanten uit de regio Zuid-Limburg (Schram et al., 2014). We combineren deze medische gegevens met CBS-gegevens om de relatie tussen leefstijl, diabetes type 2 en zorgkosten te analyseren. Het combineren van deze gevoelige gegevens gebeurt in een veilige, afgeschermd, en privacybeschermende omgeving, waar niemand toegang toe heeft: de analyse wordt op de gecombineerde dataset geautomatiseerd uitgevoerd. Na de analyse wordt de gecombineerde dataset vernietigd en zullen alleen resultaten (bijvoorbeeld schattingen van de effecten) teruggestuurd worden naar de onderzoekers. Deze resultaten bevatten geen informatie die de identiteit van individuele patiënten kan onthullen. Zo waarborgen we de privacy van patiënten terwijl we toch een volledig

beeld verkrijgen van de leefomstandigheden van de patiënten.

De technische en beleidsmatige oplossingen die ontwikkeld zijn in dit project zullen in de toekomst onderzoekers ondersteunen in het combineren van data; niet alleen van De Maastricht Studie en het CBS, maar ook data van andere partijen en in andere sectoren (bijvoorbeeld agricultuur of overheid). Hiervoor worden ook pilots uitgevoerd waarbij kennis uit dit project wordt hergebruikt, bijvoorbeeld met Vektis, de Nederlandse Zorgautoriteit en Zorginstituut Nederland.

Conclusie

Door databronnen met elkaar te koppelen kunnen we een beter totaalbeeld van de context krijgen. Om dat voor elkaar te krijgen en om zulke analyses op een verantwoordelijke en correcte manier mogelijk te maken, moeten juristen, IT'ers en methodologen samenwerken met onderzoekers. Verticale fragmentatie blijft dan ook een *hot topic* voor methodologen en computerwetenschappers, want naast de juiste infrastructuur om data te combineren zijn ook correcte analytische instrumenten nodig om tot valide uitkomsten te komen wanneer de data verticaal gefragmenteerd zijn.

LITERATUUR

- Dutch Tech Center For Life Sciences (2017). *Manifesto of the Personal Health Train consortium*. <https://www.dtls.nl/wp-content/uploads/2017/12/PHT_Manifesto.pdf>.
- Schram, M. T., Sep, S. J. S., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., et al. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European Journal of Epidemiology*, 29, 439–451.
- Sun, C., Ippel, L., Wouters, B., van Soest, J., Malic, A., Adekunle, O., et al. (2018). *Analyzing Partitioned FAIR Health Data Responsibly*. <<http://arxiv.org/abs/1812.00991>>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

LIANNE IPPEL is postdoc bij het Institute of Data Science aan de Universiteit Maastricht.

E-mail: lianne.ippel@maastrichtuniversity.nl

JOHAN VAN SOEST is postdoc bij het Institute of Data Science aan de Universiteit Maastricht en bij Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, E-mail: johan.vansoest@maastrichtuniversity.nl



Sinterklaas en de laatste lootjes

Wie kent niet het Sinterklaaslootjesprobleem. Om de beurt trekt elk van n personen random een Sinterklaaslootje. De naam van elk van de personen staat op precies één lootje. Wat is de kans dat niemand zijn eigen lootje trekt? Een opmerkelijk en welbekend resultaat is dat bij niet al te kleine waarden van n de kans dat niemand zijn eigen lootje trekt praktisch gesproken niet van n afhangt en gelijk is aan $1/e \approx 0,368$. Deze benadering is in zeven of meer decimalen accuraat al vanaf $n = 10$. Wat wordt de oplossing van het probleem bij de volgende aanpassing van het probleem? Stel nu dat ieder die zijn eigen lootje trekt het lootje teruglegt en opnieuw random een lootje trekt totdat een ander lootje verkregen wordt. Wat is de kans dat de persoon die als laatste een lootje trekt met het eigen lootje blijft zitten? Deze kans is niet zo simpel te berekenen. Het antwoord is niet

$$\frac{n-1}{n} \times \frac{n-2}{n-1} \times \dots \times \frac{1}{2} \times 1 = \frac{1}{n}.$$

De reden is dat deze berekening geen rekening houdt met het feit dat als iemand een lootje gaat trekken zijn eigen lootje al eerder getrokken kan zijn. Wel is $1/n$ een bovengrens op de gezochte kans. De sleutel tot de juiste oplossing is als volgt. Zonder beperking mag de aannahme gemaakt worden dat elke keer door loting bepaald wordt wie als volgende een lootje trekt. De kans dat de laatste persoon blijft zitten met het eigen lootje verandert daardoor niet. De gemaakte aannahme maakt het echter wel mogelijk een recursie op te stellen voor de berekening van de kans. Om de recursie op te stellen, is het inzichtelijk om met toestand (a,b) de situatie aan te geven dat a personen nog een lootje moeten trekken en dat voor b van deze a personen de lootjes met hun namen al getrokken zijn. In toestand (a,b) is de volgende persoon die een lootje gaat trekken met kans b/a een persoon wiens naam al eerder getrokken is en met kans $1 - b/a$ een persoon wiens naam nog niet getrokken is. Als in toestand (a,b) een persoon uit de groep van b personen wier namen al

getrokken zijn een lootje trekt, dan is de kans $(a-b)/a$ dat deze persoon de naam van één van de andere $a-b$ personen trekt en gaat toestand (a,b) over in $(a-1,b)$; anders gaat met kans $1 - (a-b)/a = b/a$ de toestand (a,b) over in $(a-1,b-1)$. Evenzo, als in toestand (a,b) een persoon uit de groep van $a-b$ personen wiens namen nog niet getrokken zijn een lootje trekt, dan is de kans $(a-b-1)/(a-1)$ dat deze persoon de naam van één van de andere $a-b-1$ personen trekt wiens namen niet eerder getrokken waren en gaat toestand (a,b) over in $(a-1,b+1)$; anders gaat met kans $1 - (a-b-1)/(a-1) = b/(a-1)$ de toestand (a,b) over in $(a-1,b)$. Definieer nu $P(a,b)$ als de kans dat de laatste persoon het eigen lootje zal trekken wanneer de huidige toestand (a,b) is. De gezochte kans is $P(n,0)$. Deze kans wordt berekend met de recursie

$$P(a,b) = \frac{b}{a}Q(a,b) + \left(1 - \frac{b}{a}\right)R(a,b),$$

waarbij

$$Q(a,b) = \frac{a-b}{a}P(a-1,b) + \frac{b}{a}P(a-1,b-1)$$

$$R(a,b) = \frac{a-b-1}{a-1}P(a-1,b+1) + \frac{b}{a-1}P(a-1,b).$$

De startwaarden voor de recursie zijn $P(1,1) = 0$ en $P(1,0) = 1$. Intuïtief is het duidelijk dat $P(n,0)$ naar $1/n$ gaat als n heel groot wordt. Dit wordt bevestigd door numerieke berekeningen. De kans dat de laatste persoon met het eigen lootje blijft zitten heeft de waarden 0,07565, 0,03451, 0,01832 en 0,00950 voor $n = 10, 25, 50$ en 100 .

Tot slot, nog even terug naar het klassieke Sinterklaaslootjesprobleem. Dit probleem komt in vele verschijnningen voor. Eén van de aardigste toepassingen betreft de ontmaskering van de 'babyfluisteraar' Derek Ogilvie door James Randi. De Amerikaan James Randi was een beroemde goochelaar die zich later in zijn leven richtte op het bestrijden van psychische mediums, kwakzalvers

en andere beurzensnijdners. De James Randi Foundation, speciaal opgericht voor dit doel, loofde een prijs van 1 miljoen dollar uit aan een ieder die kon aantonen over paranormale vermogens te beschikken. Uiteraard moest dit wel aangetoond worden onder controleerbare testomstandigheden, zo mocht iemand die beweerde lepels te kunnen buigen zonder daar kracht op uit te oefenen niet zijn eigen lepels meenemen. Uri Geller en James Randi waren bepaald geen vrienden van elkaar. Verschillende personen namen de 1-miljoen dollar uitdaging aan. Onder deze personen Derek Ogilvie die beweerde in staat te zijn tot buitenzintuiglijke waarnemingen op afstand. Hij werd aan de volgende test onderworpen. Ogilvie mocht zelf een kind kiezen waarmee hij dacht telepathisch contact te hebben. Ogilvie kreeg tien verschillende stukken speelgoed getoond die achtereenvolgens aan het kind in een random volgorde zouden worden gegeven. Het kind werd geplaatst in een geïsoleerde kamer en elke keer dat het kind een stuk speelgoed kreeg moest Ogilvie zeggen welk speelgoed dit was. Als hij het zes of meer keer goed had, dan zou hij de miljoen dollar winnen. Ogilvie raadde slechts één keer goed en verloor de uitdaging, evenals anderen die de handschoen hadden opgenomen. James Randi hoefde zich van te voren weinig zorgen te maken dat hij de miljoen dollar zou moeten uitbetalen. De kansverdeling van het aantal goede antwoorden is – net zoals bij Sinterklaas – vrijwel hetzelfde als de Poisson-verdeling met verwachtingswaarde 1 en dit betekent dat de kans op zes of meer goede antwoorden gelijk is aan 0,00059, oftewel ongeveer 0,06%. Ogilvie dacht vooraf blijkbaar vijf goed mag ik verwachten en dan zal zes goed niet onwaarschijnlijk zijn, dus ik waag het erop. In kansproblemen kan intuïtie je snel op het verkeerde been zetten.

HENK TIJMS is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Email: h.c.tijms@xs4all.nl

EXAMINING CAUSE-EFFECT RELATIONS IN THE SOCIAL SCIENCES

A Structural Causal Modelling Approach

The objective of this paper is to present a short overview of the *Structural Causal Modelling* (SCM) framework developed by the present authors in a series of articles spanning the last decade or so. The text is based on a presentation given at Statistics Netherlands in Heerlen on December 4, 2018 (Russo, Wunsch, Mouchart, 2019). The purpose is to explain how the SCM framework provides the tools to hypothesize, model, and test explanatory mechanisms. Our framework proves particularly useful in social science contexts, since it allows us to adopt an explicit causal perspective even when analyzing observational data. Social science experiments are notoriously difficult to carry out for ethical or practical reasons, and our approach allows social scientists to go beyond mere description and to propose a causal explanation even in the absence of experiments and interventions.

GUILLAUME WUNSCH, MICHEL MOUCHART, FEDERICA RUSSO

As indicated by the name itself, structural models aim for an analysis of the structural relationships among variables and are based on field knowledge and on theoretical contributions. In econometrics or in social sciences, structural models have typically the form of a set of equations. In general, these approaches specify a statistical model without developing a detailed analysis in terms of recursivity and therefore do not end up with an explicit view of the underlying mechanism and sub-mechanisms. The SCM framework is not based on a system of equations, but on an analysis of multivariate distributions. Adopting an SCM approach means endorsing a particular view on modelling in general (the hypothetico-deductive methodology), and a specific stance on exogeneity, namely as a condition of separability of inference, on the one hand, and in interpreting marginal-conditional decompositions as sub-mechanisms, on the other hand. The construction of the statistical model is then deduced from the above approach.

In this paper, we focus on SCM as one possible perspective in quantitative social science research. There are of course other ways to analyze social phenomena, such as, for example, a systemic approach (Loriaux, 1994), agent-based modelling (Billari et al., 2007), or qualitative designs such as the case study. These other ways will not be discussed here.

A Structural Causal Modelling (SCM) Framework

The *Structural Causal Modelling* approach may be viewed as a chapter in the domain of statistical modelling, where a statistical model is considered as a set of “reasonable” hypotheses concerning the *data generating process* (DGP) represented as a probability distribution. A probabilistic representation of the DGP is used to explain a phenomenon of interest. Such an explanation involves two compo-



nents: (i) a stochastic element that embodies what is not explained in the working of the DGP (*i.e.* of the mechanism), and (ii) a non-stochastic element, the characteristics – or parameters – of the probability distribution, that provides the nature of what is explained by the statistical model. Said differently, the statistical model provides only a partial explanation of the mechanism of the DGP. For more on the relationship between statistics, causality, and explanation, see e.g. the interesting paper published in *STAtOR* by Richard Starmans (2018).

Compared to structural models in econometrics or in social sciences, the present framework takes distance from the latter in several aspects. To begin with, our structural approach is based on a *hypothetico-deductive* (H-D) methodology. This means that a hypothesis is first formulated, a model developed and tested, and the results interpreted in order to confirm or disconfirm the initial hypothesis. H-D methodologies are

widely used in science and are often associated with the falsificationist view of Karl Popper (1934, English translation 1959). However, in philosophy of science, hypothetico-deductivism has been developed much beyond the original Popperian approach detailing, among others, the role of background knowledge at the hypothesis formulation stage or the fact that we learn also from disconfirmed hypotheses—so models can be iteratively improved on, and we do not start each time from scratch. Other important methodological features are the following: Causal and structural; Recursive decomposition and DAG; Exogeneity and causation; Distributions rather than equations; Explanation and parametrization; Stability and invariance.

Causal and structural

Focusing on causal analysis, SCM depends upon reliable background information and evidence for assessing puta-

tive causes of outcomes and evaluating effects of causes, and more generally on the structure of relations between causes and outcomes. Background knowledge plays a crucial role at each stage of the H-D methodology. Firstly, causal attribution is often quite a difficult issue once a system becomes complex. Secondly, background knowledge typically involves theories concerning the domain of analysis, but also embraces a much wider scope, in particular involving previous results and preliminary analysis of data. It is on this basis that a preliminary hypothesis is formulated. Background knowledge is likewise involved in the process of developing a specific statistical model, where one makes important choices about parametrization, testing methods, etc. Finally, the results of tests are interpreted against available background knowledge.

Recursive decomposition and DAG

'Explaining' essentially means representing and decomposing a complex and global mechanism in terms of a set of simpler sub-mechanisms. The explanation is based here on a recursive decomposition of the joint distribution of the variables entering the statistical analysis. This recursive decomposition is equivalent to a systematic marginal-conditional decomposition according to a specific ordering of the variables. For example, if one considers a vector of variables, the joint distribution can be written as: $P(X_1, \dots, X_p) = P(X_1)P(X_2 | X_1) \dots P(X_p | X_1 \dots X_{p-1})$. Thus, the joint distribution is written as a product of conditional distributions where the conditioning variables form an increasing sequence and where each factor of this product represents a sub-mechanism. For this reason, *directed acyclic graphs* (DAGs) provide a privileged tool of representation (Pearl, 2000), though a DAG does not allow representing all particularities of a multivariate distribution nor of a recursive decomposition.

Exogeneity and causation

Associations among variables are not necessarily causal. They can be due to the presence of one (or more) *confounders*, the latter being common factors of the putative causes and outcomes. One should therefore control for confounders, in order to avoid making false causal claims. Under a suitable *exogeneity* condition of non-confounding, one can then view the conditioning variables as causing variables in the sub-mechanism where they appear. This is the reason why the structural model is called a causal model, because causation is relative to a particular model built with the purpose of eliciting causes.

Distributions rather than equations

The basic objects of analysis are sets (in product form)

of distributions rather than sets of equations. Equations are related, at best, to conditional expectations, although effects of causes may take other ways. For instance, in actuarial applications, the effect of some contracts may be more in the tails of the distributions than in the expectations. To give another example, the analysis of the determinants of fertility should not only focus on the average number of children per woman but also e.g. on women having no children and on those having large families.

Explanation and parametrization

In SCM, explanation is based on a recursive decomposition. As mentioned before, representing a DGP by a probability distribution implies that this representation leaves unexplained some part of the DGP, namely the stochastic component of the model. Therefore, the statistical explanation concerns the characteristics, or parameters, of the probability distributions. This fact raises the issue of the specification of the parametrization. Once a conditional distribution is deemed to represent a specific sub-mechanism, the role of the parametrization is to identify the operation of the sub-mechanism (more information in Mouchart and Orsi, 2016).

Stability and invariance

Considering as structural a mechanism underlying the workings of a DGP requires that the model enjoys suitable properties of stability, or invariance, under a specific class of interventions and of modifications of the environment. Indeed, a model that would be different for, say, each observation should not be considered as structural. Said differently, the issue here is to look for a proper separation between the incidental and the structural aspects of the DGP. From a statistical point of view, this issue is also that of properly defining the *population of reference*. One reason for this is that no model in the social sciences can pretend to be universal in time and in space. There are no *laws* here. It should be stressed that this stability, or invariance, regards both the ordering of the variables and the value of the parameters.

An example

Consider a study on the recourse to contraception in urban Africa, the example being taken from Gourbin et al. (2017). The cities are characterized by different levels of contraceptive prevalence, but also by the different effectiveness of the methods used. Several questions may be raised, the following two amongst others. Firstly, what is

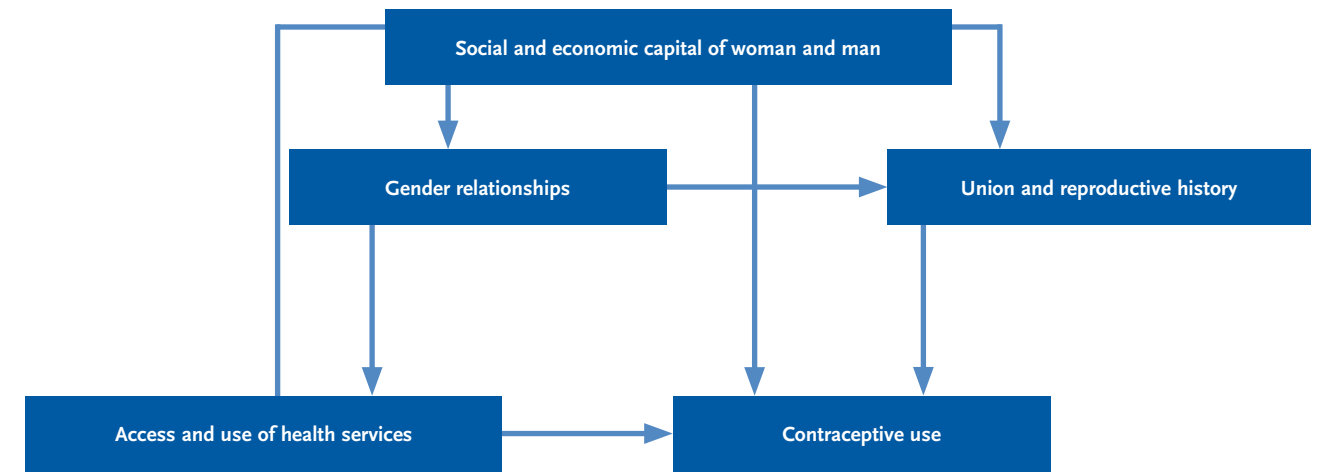


Figure 1. Conceptual model of contraceptive use in Africa (reproduced from Gourbin et al., 2017)

the hierarchical ordering of causal relationships among the individual factors involved in the use of contraception in the urban populations considered? Secondly, as education is a major factor of fertility transition, are two main indirect pathways that have been proposed in the literature (a union-reproductive path and a socio-cultural one) – leading from women's education to contraceptive use – confirmed by the data?

Most analyses of contraceptive use have had resort to statistical methods that do not take into account a possible causal ordering among the variables, implicitly assuming that all the putative determinants just have a direct effect on the dependent variable. However, the impact of these various factors on the use of contraception

can be direct or indirect, meaning in the latter case that the effect of some putative causes can be *mediated* by one or more intermediate factors. To answer the questions raised above, the SCM approach allows researchers to propose an *explanatory mechanism* for the outcome of interest, composed of various sub-mechanisms, and subsuming in particular the distinction between mediators, moderators, and confounding variables.

Based on background knowledge relating to contraceptive use and fertility, the following conceptual framework can be proposed (see Figure 1).

To test this conceptual model, one needs to obtain measurable indicators for each of the concepts in Figure 1. Using existing survey data* for the cities concerned,

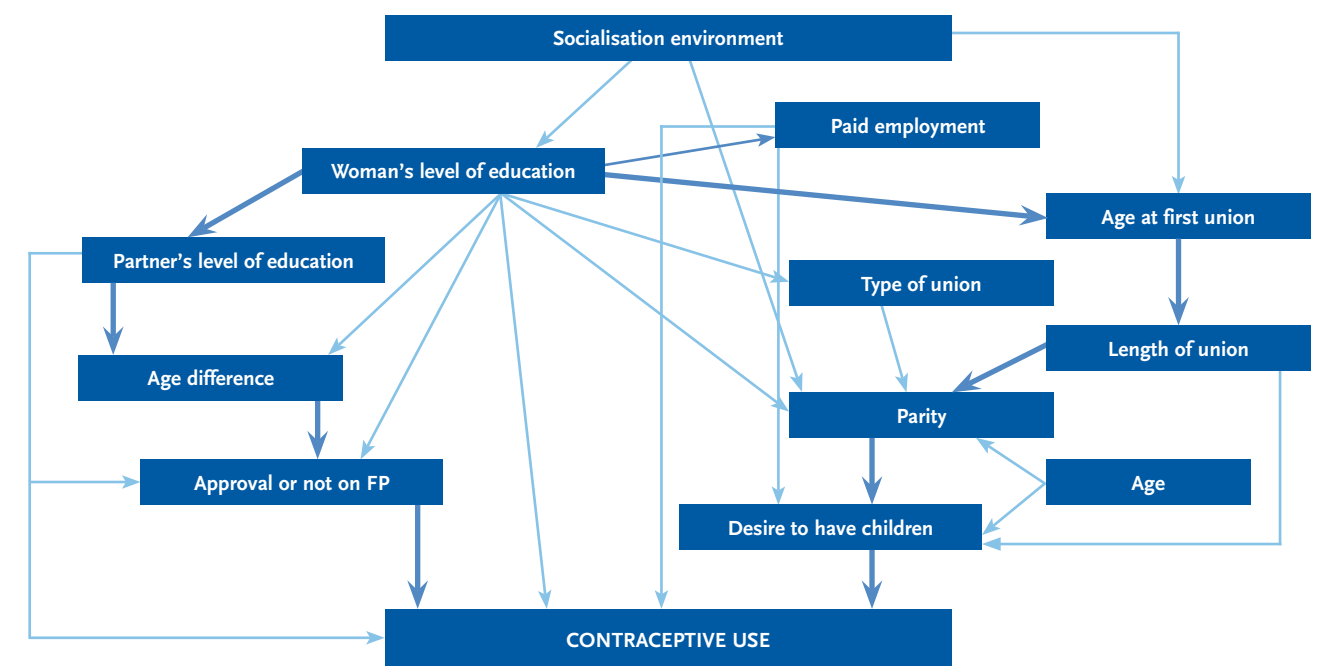


Figure 2. Operational model of contraceptive use in Africa (reproduced from Gourbin et al., 2017)

the authors have proposed the operational model presented in Figure 2. This figure is actually a *directed acyclic graph* (DAG) where the variables are ordered according to the various putative sub-mechanisms. All variables being in categorical format, the operational model has then been tested with the survey data available, using sequential logistic regressions. Concerning the two main indirect paths (in addition to the direct path) leading from education to contraceptive use proposed in the literature and drawn in **bold** in Figure 2, the data have confirmed the union-reproductive indirect path (on the right in the graph) but not the socio-cultural one (on the left). One should point out, however, that the individuals concerned were interviewed at a same moment in time and not over their lifetime. Results refer therefore to inter-individual differences obtained at the time of the surveys providing the data, and not to life-course differentials as could be derived from retrospective or prospective data.

Some remarks on data, to conclude

As the structural model presented here aims at representing causal relations among variables, the latter should be ordered according to their causal priority, which implies *inter alia* a temporal ordering of causes and effects. This can be based, for example, on retrospective or prospective population surveys, on specific registries and other forms of permanent registration of individual events, such as a national register, etc. Many sources of data actually refer to the same individuals. If each individual receives at birth a personal identification number (PIN), data from multiple sources can be linked together. One can thus examine for an individual e.g. the move from good health to ill health, then to chronic disease, disability, and finally to death, possibly also taking into account various characteristics of the individual (such as education and employment) and their change over time. If individual longitudinal data are available, the causal model can also take into account reverse causation and feedback effects, by time-ordering the variables. One must however consider the fact that an event is often the result of a temporally prior decision-making process, based on the preferences, values, beliefs, emotions, of the agents in possible interaction with others. Data on the decision-making process are unfortunately most often unavailable. Contrary to the time-ordering of events, that of the various decision-making processes is thus difficult to specify.

Due to insufficient background knowledge or to a lack of information on the temporal sequence of events, it

happens that the variables cannot be causally ordered. In this case, an exploratory analysis of the data and especially of the so-called *Big Data* (in the sense of very large structured and unstructured data sets) can possibly be helpful in revealing changing characteristics over time and suggesting the temporal sequence of events. An exploratory data approach is never a substitute for sound causal modelling, such as the framework presented in this paper, but it can usefully inform it, especially when background knowledge on the topic of interest is scant.

ACKNOWLEDGMENTS

The authors wish to thank Sofie De Broe for inviting us to give a talk on the SCM approach at Statistics Netherlands in Heerlen, in December 2018, and Joep Burger for asking us to write a short article on this topic for *STATOR*.

* No indicators were however available for the use of health services.

REFERENCES

- Billari, F. C., Prskawetz, A., Aparicio Diaz, B., & Fent T. (2007). The "Wedding-Ring": An agent-based marriage model based on social interaction. *Demographic Research*, 17/3, 59–82, doi: 10.4054/DemRes.2007.17.3.
- Gourbin, C., Wunsch, G., et al. (2017). Direct and indirect paths leading to contraceptive use in urban Africa. *Revue Quetelet / Quetelet Journal*, 5(1), 33–71.
- Loriaux, M. (1994). Des causes aux systèmes: la causalité en question, chap. 2 in R. Franck (Ed.), *Faut-il chercher aux causes une raison?*, Vrin, Paris, 41–86.
- Mouchart, M., & Orsi R. (2016). Building a Structural Model: Parameterization and structurality. *Econometrics*, 4(2), doi: 10.3390/econometrics4020023.
- Pearl, J. (2000). *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press (revised and enlarged in 2009).
- Popper, K. (1934, English translation 1959). *The Logic of Scientific Discovery*. London: Hutchinson.
- Russo, F., Wunsch, G., & Mouchart, M. (2019). Causality in the Social Sciences. A structural causal modelling framework. *Quality & Quantity*, 53(5), 2575–2588.
- Starmans R. (2018). Statistiek en causaliteit: voortschrijdende liaison of moeizame samenspraak, *STATOR*, 20(4), 28–34.

GUILLAUME WUNSCH is Emeritus Professor of Demography at the University of Louvain (UCLouvain) and a member of the Royal Academy of Belgium.
E-mail: g.wunsch@uclouvain.be

MICHEL MOUCHART is Emeritus Professor of Statistics at the University of Louvain (UCLouvain).
E-mail: michel.mouchart@uclouvain.be

FEDERICA RUSSO is Professor of Philosophy of Science at the University of Amsterdam.
E-mail: federica.russo@gmail.com



Statistiek vanaf de basisschool in het lespakket

‘Het lesprogramma moet op de schop: statistiek al vanaf de basisschool [...]’ Zo luidde afgelopen mei de kop van een persbericht over het conceptvoorstel van Curriculum.nu voor de herziening van de kerndoelen en eindtermen voor het primair en voortgezet onderwijs. Het voorstel Rekenen & Wiskunde wil namelijk zowel in het primair als het voortgezet onderwijs meer aandacht voor statistiek. Een mooi moment om ons als vakvereniging te laten horen.

Het onderwijs leeft binnen onze vereniging. De VVSOR acht voortdurende aandacht voor verbetering en vernieuwing van het onderwijs van groot belang. Zo ontstond het idee om te reageren op de het voorstel Rekenen & Wiskunde van Curriculum.nu. Verschillende leden waren betrokken bij de totstandkoming van het voorstel via de vakorganisaties voor wiskundeonderwijs. De onderwijs-technische kant van het voorstel heeft al op veel belangstelling mogen rekenen (zoals in de discussie over breuken), daarom is de reactie van de VVSOR juist gericht op de statistische inhoud van het voorstel.

De VVSOR is enthousiast over de doelen van het voorstel Rekenen & Wiskunde, met name om scholieren op te leiden tot kritische cijferconsumenten. We snijden in onze brief wel wat kritische punten in ons vakgebied aan waar het voorstel rekening mee moet houden.*

- Hoe kans en toeval in het basisonderwijs het best hun plek krijgen met een vaste set voorbeelden (alledaagse kansen als ‘de kans op regen’ zijn lastiger dan ze lijken).
- Dat p-waarde significantietoetsen vraagt om extra voor-

zichtigheid, vanwege de recent opgelaaide discussies en het empirische bewijs dat het vaak verkeerd wordt begrepen.

- En hoe ook zonder de statistische cyclus plat te slaan tot vuistregels er prachtig materiaal voorhanden is op websites als Health News Review, Peiling Praktijken, Risk Litteracy en Understanding Uncertainty.

Met name voor de bovenbouw van de middelbare school moeten de doelen van Curriculum.nu nog verder uitgewerkt worden. Als vereniging bieden we daarbij onze hulp aan en stellen we voor om met (wiskunde)leraren en geïnteresseerden uit het onderwijsveld bijeen te komen op een conferentie.

Ideeën, reacties en conferentie

De brief is totstandgekomen in overleg met sectievoorzitters en betrokken leden die hebben gereageerd op onze oproepen begin juni en eind juli. Vanwege de vakantieperiode zullen alleen niet alle ideeën verzameld zijn. Heb je extra voorbeelden, bronnen voor lesmateriaal of een ander belangrijk punt dat we moeten vermelden? Of zou je graag meedenken over de conferentie die we over het onderwerp willen organiseren? Laat het ons dan voor 1 oktober weten en mail naar penningmeester@vvsor.nl. We hopen eind september alle stemmen verzameld te hebben zodat we onze reactie een groter podium kunnen geven.

JUDITH TER SCHURE,
penningmeester VVSOR

* De volledige VVSOR-reactie is te lezen op de ledenpagina van de website: <https://www.vvsor.nl/members/>. Deze brief is op 9 augustus gedeeld met Curriculum.nu (voor de deadline van 11 augustus).



IMPACT

Als ik in het buitenland ben, probeer ik altijd een plek te bezoeken die belangrijk is geweest voor de muziek of de wiskunde, twee onderwerpen die mijn grote interesse hebben. Zo bezocht ik in Berlijn de Hansa Tonstudio waar David Bowie *Low* en *Heroes* opnam, liep ik in Londen over de oversteekplaats van Abbey Road, bezocht ik het graf van Lagrange in het Pantheon in Parijs en legde bloemen bij het beeld van Alan Turing in de Sackville Gardens in Manchester. Een paar weken geleden was ik in Dublin, een stad die in alles muziek ademt, maar ook de stad van William Rowan Hamilton, naamgever van onder andere de Hamilton cycle en ontdekker van het Quaternion¹.

Ik was in Dublin voor de conferentie EURO-K 2019 als jurylid voor de EURO Excellence in Practice Award², een competitie die tot doel heeft de bekendheid van ons vakgebied te bevorderen en de impact van Operations Research (OR) op de dagelijkse praktijk te erkennen door de beste praktische OR-toepassing te belonen. Wat is er mooier dan op de plek waar Hamilton gewerkt heeft de impact van OR op onze samenleving te laten zien? Werk aan de winkel, de muziek moest maar even wachten.

Voorafgaand aan de conferentie had de EEPA-jury al tientallen inzendingen beoordeeld en op basis van de aangeleverde achtergrondinformatie, aanbevelingen en publicaties zes kandidaten geselecteerd die tijdens de conferentie hun project zouden presenteren. Aan diversiteit van toepassingen van OR geen gebrek, van meer traditionele OR-toepassingen als het opstellen van roosters van treinconducteurs en de *scheduling* van gespecialiseerde servicemonteurs ook opvallende toepassingen zoals het verbeteren van de productie van prototypes van personenauto's en het verbeteren van de biochemische fabricage van medicijnen. Stuk voor stuk voorbeelden van toegepast OR-werk. Of het nu wel of niet een traditionele OR-toepassing betreft, de presentaties lieten zien dat het nog steeds niet eenvoudig is een goed en meetbaar resultaat te boeken. Het vraagt om vakmanschap en doorzettingsvermogen.

Praktische toepassingen van OR zijn vaak erg uitdagend. Ondanks dat we in een big-data-tijdperk leven is er aan betrouwbare data nog steeds een gebrek. Ook zit *the Devil* vaak in *the details*, wat is relevant en wat laat je

buiten beschouwing om zo toch een goed oplosbaar en effectief model te maken? Het zijn niet alleen 'harde' uitdagingen, ook de zachte kant van ons vak speelt vaak in de praktijk een grote rol. Hoe overwin je bijvoorbeeld de weerstand in de organisatie om de ontwikkelde oplossing te gaan gebruiken. Dat vraagt om meer dan alleen een goed model. Er zijn veel uitdagingen die een impactvolle praktische toepassing van OR in de weg kunnen staan, maar de zes geselecteerde kandidaten waren er prima in geslaagd die uitdagingen aan te gaan.

Een van de meest in het oog springende deelnemers dit jaar was voor mij het project dat de plaatsing van leerlingen op middelbare scholen in Chili heeft geoptimaliseerd. In 2015 werd in Chili een hervormingswet voor het schoolsysteem ingevoerd, de School Inclusion wet die een einde moest maken aan de vaak ondoorzichtige en oneerlijke selectie van leerlingen door scholen. Deze praktijken leidden tot polarisatie en segregatie in het Chileense schoolsysteem. Minderbedeelde kinderen kregen nauwelijks toegang tot beter onderwijs met als gevolg dat ze geen kans kregen zich te ontwikkelen en de armoede waarin ze leefde te ontstijgen. Om deze misstanden aan te pakken werd de wet School Inclusion ingevoerd, maar die creëerde meteen ook een probleem. De wet vraagt ouders meerdere voorkeursscholen voor hun kinderen op te geven waarna, op basis van een aantal criteria, een plaatsing wordt bepaald. Maar hoe zorg je voor een eerlijke verdeling van aanvragen van 250.000 leerlingen over de beschikbare plekken van 6400 middelbare scholen? Sommige scholen zullen immers overvraagd worden terwijl andere veel minder aanvragen krijgen. Daarnaast zijn er allerlei praktische condities, bijvoorbeeld kinderen van hetzelfde gezin moeten bij voorkeur bij dezelfde school worden geplaatst. Of dat je je huidige plek niet verliest als je probeert om op een andere school te worden geplaatst. Kortom een groot en complex vraagstuk met een hoog maatschappelijk belang. Een typische uitdaging waar OR uitkomst biedt. In dit geval bracht een voor velen bekend algoritme de uitkomst, het Deferred Acceptance algoritme van Gale en Shapley. Een algoritme uit 1962 dat met enige aanpassingen geschikt was om de uitdaging aan te gaan. Technisch gezien misschien niet een supercom-

plexe oplossing, maar het lost wel een heel belangrijk maatschappelijk probleem in Chili op. En daar gaat het in de EEPA om, Impact!

De EEPA-award werd dit jaar toegekend aan Yves Lucet en Warren Hare voor hun onderzoek naar het gebruik van optimalisatietechnieken in het ontwerp en de aanleg van wegen. Ook dit project heeft een hoge maatschappelijke impact. Door hun onderzoek stelden Lucet en Hare bedrijven en overheden in staat miljoenen aan kosten te besparen bij het ontwerp en de aanleg van wegen. De modellen van Lucet en Warren zijn verwerkt in software waarmee ingenieurs een instrument in handen krijgen dat hen niet alleen in staat stelt sneller wegen te ontwerpen, maar dit ook nog eens slimmer (= kosteneffectief) en beter (= veilig) te doen. Gezien de reikwijdte van hun onderzoek zijn Lucet en Hare daarmee een te rechte winnaar van de competitie dit jaar.

De volgende editie van de EEPA is in 2021. Dus mocht je op dit moment werken aan een project waarin OR wordt toegepast met een meetbaar praktisch resultaat, schroom niet om je aan te melden. Dat kan vanaf november 2020.

Was er nog tijd om iets met die interesses van me te doen zul je wellicht afvragen? Die was er gelukkig, zij het beperkt. Tussen de bedrijven door een kort bezoek gebracht aan Trinity College³, een van de plekken waar Sir William Rowan Hamilton zijn onderzoek deed. Op weg naar het vliegveld nog langs Windmill Lane geweest, de plek waar U2 albums als *Boy*, *October* en *War* opnam, over impact gesproken! Na een laatste Guinness in Bruxelles⁴, Phil Lynotts favoriete pub, met zijn bastonen in mijn oren teruggereisd naar Nederland. Op weg naar het volgende, hopelijk impactvolle, OR-consultingproject.

1. <https://nl.wikipedia.org/wiki/Quaternion>
2. <https://www.euro-online.org/web/pages/209/excellence-in-practice-award-eepea>
3. <https://www.tcd.ie/library/old-library/long-room/>
4. <https://www.dublintonline.ie/story-bruxelles-bar-star-signs-phil-lynott/>

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



William Rowan Hamilton

Sir William Rowan Hamilton (1805 – 1865) was een Ierse wiskundige, natuurkundige en astronoom die belangrijke bijdragen leverde aan de ontwikkeling van de optica, dynamica en algebra. Hamilton werd in het bijzonder bekend door de door hem uitgedachte quaternionen.

Op zijn vijftiende begon hij de klassieke werken van Newton en Laplace te bestuderen. In 1822 ontdekte hij een belangrijke fout in Laplaces *Mécanique céleste*, wat hem onder de aandacht bracht van John Brinkley, op dat moment Astronoom Royal van Ierland. Brinkley zei over Hamilton: 'Deze jongeman zal niet alleen de beste wiskundige van zijn tijd zijn, hij is het al.'

Hamilton heeft een icosiaans puzzelspel ontwikkeld – bekend als Hamilton's puzzle – waarvan hij de rechten in 1859 voor £ 25 had verkocht aan speelgoedfabrikant. Commercieel was het icosiaanse spel van Hamilton een flop, maar een origineel exemplaar (zie afbeelding) is heden ten dage een *collector's item*.

De laatste jaren van zijn leven besteedde Hamilton aan het schrijven van het ultieme werk over quaternionen. *Elements of Quaternions* werd in 1866 door zijn zoon William Edwin Hamilton uitgebracht. Hijzelf was op 2 september van het jaar ervoor overleden. (Bron: Wikipedia)

INCONSISTENTIES IN DE ARBEIDSTIJDENWET

Wiskunde toegepast in de weerbarstige praktijk

Dat de praktijk vaak weerbarstiger is dan de theorie, zien we terug bij het maken van personeelsroosters. Het breed bestudeerde personeelsplanningsprobleem blijkt onder invloed van inconsistenties in wetgeving lastig wiskundig te formuleren. Daarnaast dragen de gestelde regels in de huidige vorm niet altijd bij aan de beoogde doelen van de wetgeving. In deze bijdrage gaan wij in op een aantal van de inconsistenties in de Nederlandse Arbeidstijdenwet.

LAURENS FIJN VAN DRAAT, STAN DE GREEF, GERHARD POST EN EGBERT VAN DER VEEN

'Met het oog op de veiligheid, de gezondheid en het welzijn in verband met de arbeid [...] en mede ter bevordering van de combineerbaarheid van arbeid en zorgtaken, alsmede andere verantwoordelijkheden buiten de arbeid, [is] het noodzakelijk wettelijke regelen te stellen inzake arbeids- en rusttijden.'

De zin hiernaast komt uit de aanhef van de Arbeidstijdenwet (ATW)¹. Naast de Arbeidstijdenwet is er ook nog een Arbeidstijdenbesluit² (ATB) met aanvullende, sectorgebonden wetten. Doel van deze wetten is het stellen van grenzen aan arbeid ten behoeve van de veiligheid, gezondheid en het welzijn van werknemers.

Aan de hand van voorbeelden zullen wij laten zien dat voor tenminste vier regels in de wet geldt dat deze niet altijd bijdragen aan dit doel. Daarbovenop zijn verschillende categorieën werknemers uitgezonderd van de ATW en ATB, zoals zzp-ers, wetenschappelijk onderzoekers en militairen. Is het welzijn van deze mensen niet belangrijk?

Hoewel deze laatste vraag ook een interessant debat oplevert, richten wij ons verder in dit artikel op de regels zelf. Om de inconsistentie tussen het beoogd doel van de wetgeving en een deel van de regels aan te kunnen tonen, introduceren wij in de volgende paragraaf eerst een aantal definities en een tweetal consistentieregels. In de opvolgende paragrafen gaan wij in op de inconsistenties.

Consistentieregels en definities

We definiëren het volgende, zoals ook in artikel 1:7 van de ATW:

Dienst. Een aaneengesloten periode waarin arbeid wordt verricht zonder rusttijden van ten minste acht uren. Hierbij is arbeid elke activiteit waarbij gewerkt wordt, of waarbij de werkgever bepaalt waar de werknemer verblijft om voor werk beschikbaar te zijn. Zo zijn brandweerlieden in de kazerne 's nachts niet aan het werk (als er geen oproep is), maar verrichten sinds 2005 wel arbeid. Deze activiteit (ter plekke aanwezig zijn om op afroep werk te doen) wordt *aanwezigheid* genoemd.

Aanwezigheidsdienst. Een dienst is een aanwezigheidsdienst als deze ten minste een activiteit heeft van het type 'aanwezigheid', hoe kort dan ook. De maximale duur van een aanwezigheidsdienst is 24 uur (ATB, artikel 1:1).

Nachtdienst. Een dienst waarin langer dan een uur tussen 0:00 uur en 6:00 uur wordt gewerkt;

Pauze. Een rustperiode van ten minste 15 minuten tijdens een dienst. Een dienst met meer dan 5 ½ uur arbeid dient volgens artikel 5:4 een pauze te bevatten.

Week. De periode van zondag 0:00 uur tot de volgende zondag 0:00 uur.

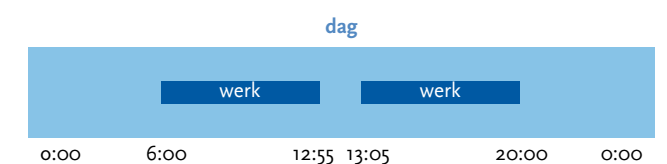
We stellen twee consistentieregels voor. Het doel van deze regels is ervoor te zorgen dat een schending niet kan worden opgelost door de medewerker meer te belasten.

1. Een rooster met teveel arbeid kan niet worden gerepareerd door arbeid toe te voegen;
2. Een rooster met te weinig rust kan niet worden gerepareerd door arbeid toe te voegen.

Met deze definities en consistentieregels zullen wij ingaan op vier voorbeelden van inconsistenties in regelgeving met betrekking tot de duur van (aanwezigheids) diensten, het aantal aanwezigheidsdiensten in een periode, dagelijkse rust voor jeugdigen (< 18 jaar) en rust na serie nachtdiensten. In elk van de voorbeelden starten we met een rooster waarin zich schendingen voordoen. Het doel is het rooster te repareren, zodat er geen schendingen meer optreden.

Inconsistentie 1: Duur van aanwezigheidsdiensten

Het ATB stelt dat bepaalde regels, die volgens de ATW gelden voor diensten met arbeid, niet gelden voor aanwezigheidsdiensten. Dit leidt tot de eerste inconsistentie. We geven een voorbeeld, uitgaande van een dienst van 14 uur lang, bestaande uit twee blokken van 6 uur en 55 minuten werk (werk is wel arbeid maar geen aanwezigheid), met daartussen een onderbreking van 10 minuten (figuur 1). Deze dienst overtreedt twee regels: de



Figuur 1. Werkdienst in twee blokken

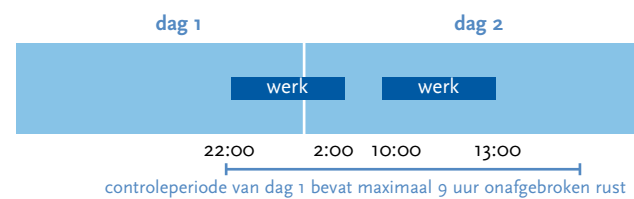
dienst bevat meer dan 12 uur arbeid (artikel 5:7, lid 2a), en de dienst heeft geen pauze van 15 minuten (artikel 5:4, lid 3).

Als we nu de rustperiode van 10 minuten vervangen wordt door een aanwezigheidsactiviteit, dan worden de overtredingen opgeheven. De dienst is nu namelijk een aanwezigheidsdienst geworden, en daarmee is een pauze niet langer verplicht en de duur van de dienst kan nu 24 uur zijn. Dus, behalve het opvullen van de 10 minuten rust, mag nu zelfs nog 10 uur extra arbeid toegevoegd worden. Draagt het toevoegen van extra arbeid (in de vorm van aanwezigheid) aan deze dienst bij aan het welzijn van werknemers?

Inconsistentie 2: Aantal aanwezigheidsdiensten in een periode

De ATW vertelt niet bij welke week een dienst hoort die over de weekgrens gaat. Gelukkig is de Europese wetgeving op dit gebied nauwkeuriger. Bepalend is het startmoment van de dienst: een dienst die op zaterdag voor middernacht begint en op zondag eindigt, hoort bij de week van de zaterdag.

Deze regel leidt tot inconsistenties, omdat sommige regels zijn uitgedrukt in een maximaal aantal diensten in een periode van een aantal weken, zoals de regel die zegt dat maximaal 52 aanwezigheidsdiensten in 26 weken toegestaan zijn (ATB, artikel 4.8:1). Veronderstel een situatie van 26 weken, met een aanwezigheidsdienst vanaf middernacht op de eerste zondag en verder in alle weken twee aanwezigheidsdiensten, bijvoorbeeld op woensdag en donderdag. Deze 26 weken bevatten nu 53 aanwezigheidsdiensten, en dat is één te veel. Voeg vervolgens een



Figuur 2. Niet voldoende rust

blokje werk toe aan de eerste aanwezigheidsdienst van zaterdag 23:00 uur tot middernacht. Daarmee schuift de aanwezigheidsdienst naar week 'o' en is de overtreding verdwenen. Draagt het verlengen van deze dienst bij het welzijn van werknemers?

Inconsistentie 3: Dagelijkse rust voor jeugdigen (<18)

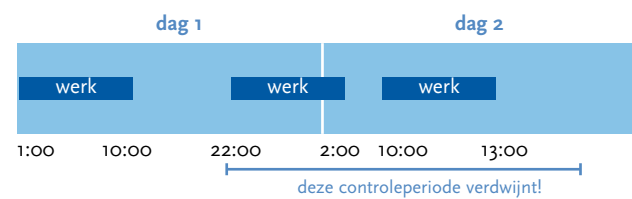
Er zijn verschillende regels rond verplichte rustperiodes, zoals de dagelijkse rust (11 uur), de dagelijkse rust met één of meer uitzonderingen, de wekelijkse rust die ook gerealiseerd kan worden in twee weken, en de verplichte rustperiodes in het geval er aanwezigheidsdiensten in de periode vallen. Deze laatste regel kwam eerder³ in STA+OR aan de orde. Hier bekijken we de regel voor dagelijkse rust, zoals deze geldig is voor jeugdigen (<18):

Artikel 5:3: *De werkgever organiseert de arbeid zodanig, dat de jeugdige werknemer een onafgebroken rusttijd heeft van ten minste 12 uren in elke aaneengesloten periode van 24 uren.*

[...] *De hier genoemde periode vangt aan op het eerste tijdstip van de dag, waarop de werknemer arbeid verricht.*

Lees deze regel gerust nog een keer. Het klinkt al vreemd: 'elke aaneengesloten periode' (dat zijn er dus heel veel) en vervolgens 'de hier genoemde periode' (dat is er dus maar één!). En dat is ook precies wat er mis gaat met deze regel. We bekijken de situatie afgebeeld in figuur 2. Omdat jeugdigen geen aanwezigheidsdiensten mogen doen, bestaat de arbeid alleen uit werkactiviteiten.

Volgens de regel begint de controle periode op dag 1



Figuur 3. Voldoende rust door extra arbeid

om 22:00 uur en loopt tot 22:00 uur op dag 2. Deze periode bevat twee rustperiodes (van 2:00 uur tot 10:00 uur, en van 13:00 uur tot 22:00 uur), waarvan de langste 9 uur duurt, en dus korter is dan de gevraagde 11 uur (figuur 3).

Als we nu op dag 1 een dienst van 1:00 uur tot 10:00 uur toevoegen, dan loopt de controleperiode van 1:00 uur op dag 1 tot 1:00 uur op dag 2. De periode van 10:00 uur tot 22:00 uur op dag 1 is nu de dagelijkse rustperiode op dag 1 (figuur 3). De eerdere controleperiode is verdwenen en is er geen overtreding meer. Draagt het toevoegen van een extra dienst bij aan het welzijn van werknemers? Waar de toevoeging: 'De hier genoemde periode vangt aan op het eerste tijdstip van de dag waarop de werknemer arbeid verricht' vandaan komt is voor ons onduidelijk. Een formulering als 'De hier genoemde periode vangt aan op ieder tijdstip, waarop de werknemer begint met arbeid te verrichten' zou wél het beoogde doel (rust voor jongeren) beter dekken.

Volledigheidshalve moet gezegd worden dat deze situatie voor jeugdigen eigenlijk niet kan optreden, omdat zij geen nachtwerk mogen doen. Toch is dit voorbeeld wel relevant omdat de dagelijkse rust regel voor volwassenen, waarin eens per zeven dagen een inkorting van de rust toegestaan is, dezelfde definitie van controleperiode hanteert.

Inconsistentie 4: Rust na serie nachtdiensten

De ATW (artikel 5:8, lid 5) bepaalt dat na drie of meer opeenvolgende nachtdiensten, een rust van 46 uur genoten moet worden. Volgens deze regel is het toegestaan om vijf opeenvolgende nachtdiensten te werken, maar als de vierde hiervan uitgepland wordt, ontstaat er een overtreding (de serie wordt onderbroken, maar er is minder dan 46 uur rust). Op zich is dat nog geen inconsistentie ten opzichte van onze gestelde consistentieregels, maar kijk naar het volgende voorbeeld. Stel we plannen vier nachtdiensten van 18:00 – 2:00 uur met daaropvolgend één dienst van 18:00 – 1:00 uur, welke volgens de definitie van de ATW geen nachtdienst is. Volgens artikel 5:8, lid 5 is dit niet toegestaan, maar dit kan opgeheven worden

door deze laatste dienst een uur langer te maken. Draagt dit extra uur bij aan meer rust en beter welzijn en gezondheid van de medewerker?

Conclusie

De ATW en ATB zijn overduidelijk niet geschreven door wiskundigen en zijn, gegeven de vele uitzonderingen, het resultaat van een flinke dosis polderen en stapelen van nationale en internationale (Europese) wetgeving. Het goed toepassen is daarmee niet eenvoudig en uitzonderingen op regels staan (waarschijnlijk) onbedoelde situaties toe. Een herziening van wetten, met inachtneming van goede consistentieregels lijkt op zijn plaats. Hiermee wordt voorkomen dat het toevoegen van arbeid of het verminderen van rust bijdraagt aan het 'verbeteren' van het welzijn van medewerkers.

Maar let op, wij klagen niet. Het bestaan van dit soort wetten maakt het handmatig controleren van roosters onmogelijk en zorgt dat wij voorlopig werk genoeg hebben.

NOTEN

1. Arbeidstijdenwet, zie <https://wetten.overheid.nl/BWBR0007671/2018-01-01>.
2. Arbeidstijdenbesluit, zie <https://wetten.overheid.nl/BWBR0007687/2018-11-14>.
3. Fijn van Draat, L., Poppelaars, J., & Post, P. (2007). De complexiteit van de Vereenvoudigde Arbeidstijdenwet. STA+OR, 2(8), p. 20–24.

LAURENS FIJN VAN DRAAT is optimalisatie-expert van de afdeling Workforce bij ORTEC te Zoetermeer.
E-mail: laurens.fijnvandraat@ortec.com

STAN DE GREEF is senior consultant van de afdeling Health Care bij ORTEC te Zoetermeer.
E-mail: stan.degreef@ortec.com

GERHARD POST is optimalisatie-expert bij PCA mobile te Raalte en universitair docent aan de Universiteit Twente.
E-mail: g.f.post@utwente.nl

EGBERT VAN DER VEEN is Business Analyst Maritieme IT bij het ministerie van Defensie te Den Helder.
E-mail: E.vd.Veen.o4@mindef.nl



OVER HET NUT VAN GRONDSLAGECRISES filosofie en praktijk van data science

RICHARD STARMANS

De praktijk van Big Science

Wie zich erop toelegt hedendaagse ontwikkelingen in wetenschap en techniek te duiden vanuit een *historisch-wijsgerig* perspectief kan de laatste dertig jaar bezwaarlijk om het oeuvre van de Amerikaanse fysicus en wetenschapshistoricus Peter Galison (1955) heen. In vele opzichten neemt hij in het ideeënhistorische discours een merkwaardige positie in. Ruim een kwart eeuw geleden verscheen de door Galison geredigeerde essaybundel *Big Science: the growth of large-scale research* (1992). De titel was allerm minst een al dan niet bedoelde vooruitwijzing naar onze huidige preoccupaties met Big Data en Data Science. Door deeltjesversnellers, ruimtetelescopen en satellieten beschikken natuurkundigen en astronomen al veel langer over terabytes en petabytes aan data en er

was anno 1992 dan ook geen enkele reden dit middels de titel van een nieuwe publicatie te bekrachtigen of te benadrukken. Veeleer beoogde Galison de invloed van schaalgrootte en organisatie van het moderne fysische onderzoek op de experimentele praktijk te analyseren en tevens de wetenschapsfilosofische implicaties hiervan te doorgronden. Anno 2019 zullen velen beamen dat het wetenschappelijk metier – ook buiten de fysica – in dit opzicht een forse metamorfose heeft ondergaan. Onderzoek behelst dikwijls allang niet meer de noeste arbeid, creatieve invallen of diepzinnige overpeinzingen van individuele geesten of zelfs maar kleine onderzoeksgroepen, die in hun eigen vertrouwde laboratorium zelfstandig *hypothesen bedenken*, eensgezind *experimenten opzetten*, deze tussentijds naar eigen inzicht of bevinden aanpassen en die uiteindelijk zelf bepalen wat als *resultaat* geldt

en welke conclusies daaraan verbonden mogen worden. 'Big Science' schetst de wereld van internationale samenwerking, politieke spanningsvelden en lobby's, mondiale subsidieprogramma's, peperdure deeltjesversnellers en onderzoeksgroepen van soms honderden wetenschappers. Maar ook de protocollen en conventies, minutieuze bestudering van logboeken, de veelvuldige meetings tussen experts en stakeholders, inclusief onderhandelingen, debatten en bijbehorende besluitvormings- of consensusmechanismen.

Zelfs de ogenschijnlijk eenvoudige vraag wanneer en op welke wijze in de praktijk een experiment als beëindigd of voltooid kan worden beschouwd blijkt verre van triviaal, zoals Galison al indringend had uiteengezet in zijn *How Experiments End* (1987), een studie die hem internationale bekendheid bracht. In dit boek bespreekt de auteur een drietal experimentele perioden uit de fysica, die zich kenmerken door een toenemende graad van complexiteit en organisatie. Allereerst gaat Galison in op de kleinschalige, min of meer traditionele, macroscopische studies van onder meer Einstein en De Haas naar gyromagnetische effecten. Vervolgens besteedt hij aandacht aan experimenten met kosmische straling, die in de jaren 30 zowel door Amerikaanse als Europese onderzoeksgroepen werden uitgevoerd en die zouden leiden tot de ontdekking van het mu-meson (muon). Tot slot analyseert Galison in ruim honderd pagina's de ontdekking van zogenoemde *weak electric current* als resultaat van mega-onderzoeksprojecten met langdurige experimenten, waarbij honderden onderzoekers waren betrokken. Is er bij dit alles nog ruimte voor het romantische of faustische 'genie-begrip', het idiosyncratische of hyperindividuele in het wetenschappelijke denken? Galison illustreert in zijn historische casestudies overtuigend hoe precair het kan zijn heden ten dage een nieuwe uitvinding of ontdekking te verbinden met één enkele persoon, plaats of tijdstip. De beroemde Law of Eponymy, zoals in 1980 verwoord door de historicus van de statistiek Steven Stigler (1941) lijkt op zo'n moment a fortiori in werking te treden.

Opmerkelijk is dat Galison met zijn focus op de onderzoekspraktijk, het wetenschappelijke bedrijf en op 'het collectief' ten koste van 'het individu', een invalshoek kiest die in sommige opzichten lijkt aan te sluiten bij de vroege sociologische studies van Bruno Latour (1947), zoals diens fameuze *Science in Action: How to Follow Scientists and Engineers Through Society* uit 1987, en bij

het zogenaamde Strong Program van Barry Barnes en David Floor en andere wegbereiders van de latere Science Studies. Er lijkt zelfs een raakvlak met aanhangers van het sociaal-constructivisme, dat de gehele werkelijkheid, inclusief wetenschap en techniek opvat en bestudeert als een door mensen gestuurde en beregelde constructie met groepsinteracties, instituties en conventies, die wetenschap niet wezenlijk doen verschillen van vele andere menselijke activiteiten. Voornoemde filosofen doen dit overigens doorgaans vanuit een externalistisch perspectief, zonder kennis van de materie of inhoudelijke veranderingen en vanuit een emancipatoir, maatschappijkritisch vertrekpunt. Daarmee tekent zich een groot verschil af met Galison, die in zijn casestudies technische en soms minutieuze details niet schuwt, de natuurkunde vooropstelt en met klem afstand neemt van voornoemde traditie. Sterker nog, Galison fulmineert tegen sociaal-constructivisten als de fysicus en socioloog A. R. Pickering, in wiens bijna 700 bladzijden tellende studie *History of particle physics: a sociological analysis* uit 1983 de eerste contouren van een sociaal-constructivistisch paradigma zich aftekenen en dat kan worden opgevat als een opmaat tot zijn bekende *Constructing Quarks: A Sociological History of Particle Physics* (1984). Ondanks alles omarmt Galison in dit opzicht de wijsgerige positie van het wetenschappelijk realisme. Hij beschouwt het experiment als een 'real encounter with the world'; een werkelijkheid die objectief bestaat, onafhankelijk is van de menselijke geest en kenbaar is, al is die kennis veeleer gemedieerd door dan geconstrueerd vanuit sociale principes en mechanismen. Maar Galison gaat verder, ook in later studies. Hij hekelde in zijn ogen monomane gerichtheid van filosofen op thema's als de structuur van wetenschappelijke *theorieën*, abstracte modellen van verklaring, causaliteit en natuurwetten, principes van confirmatie, falsificatie en rationaliteit en zeker ook de daarmee gepaard gaande 'discontinuïteiten' en 'crises': paradigmawisselingen, revoluties, gepostuleerde grondslagen crises en de zoektocht naar filosofische fundamenten. Van de weeromstuit wil hij het experiment wederom een min of meer autonome status geven, ongeacht paradigmawisselingen op theoretisch vlak of wijsgerige objecties als de Duhem-Quine thesis. Het was paradoxaal genoeg evenwel de wetenschapsfilosoof en historicus van kansrekening en statistiek Ian Hacking (1936), die in *Representing and Intervening* (1983) langs filosofische weg reeds tot vergelijkbare conclusies was gekomen en beargumenteerde dat experimenten

‘a life of their own’ bezitten. Hoe dan ook, voor de filosoof Robert J. Ackermann was het werk van Ian Hacking, Nancy Cartwright en andere Stanford School-philosophers aanleiding in 1989 een nieuwe stroming te proclameren, The New Experimentalism, waartoe hij overigens ook de fysici Allan Franklin en Peter Galison rekende. (Starmans, 2017a)

Grondslagen, crises en praktijk

Met Galison lijken we mijlenver verwijderd van een romantische conceptie van kennis waarbij de wetenschap per figureert als gedreven eenling, pionier en zoeker naar waarheid, die eigenhandig een omwenteling probeert te bewerkstelligen, die tracht inzicht te verwerven in de ware aard van de verschijnselen en hun onderliggende (eerste) beginselen. Of, eloquenter verwoord door Faust in Goethe's beroemde drama, die proclameert ‘*daß ich erkenne was die Welt am innersten zusammenhält, Schau alle Wirkenskraft und Samen und tu nicht mehr in Worten kramen*’. Als zodanig dient kennis betrekking te hebben op de Great Chain of Being, een idee afgeleid van Plato en Plotinos, verder ontwikkeld in de middeleeuwse filosofie en nieuw leven ingeblazen door Arthur O. Lovejoy (1873–1962) in zijn gelijknamige ideeënhistorische studie (Lovejoy, 1938). Daartoe is het bovendien noodzakelijk de ‘oorzaken der dingen’ te kennen, zulks indachtig het befaamde dictum ‘*felix qui potuit rerum cognoscere causas*’ (‘gelukkig is hij die de oorzaken van de dingen begrijpt’) verwoord door Vergilius in zijn Georgica. Daarmee treedt het aloude filosofische thema van de causaliteit opnieuw op de voorgrond, een thema dat volgens Judea Pearl (1936) ruim 100 jaar lang stelselmatig door statistici is miskend en gedwarsboomd (Pearl, 2018).

Belangrijker zijn echter enkele principiële kritiekpunten op het overigens imposante oeuvre van Galison. Uiteraard heeft de auteur gelijk dat afgebakende, gedetailleerde casestudies een diep inzicht kunnen geven in de natuurkundige praktijk en dat filosofen niet vanuit een ivoren toren of *ex cathedra* kunnen verordnen hoe wetenschap dient te worden begrepen of gewaardeerd. Maar zo'n voorstelling van zaken is niet minder dan een false dilemma, een *stroman-fallacy*. In Galisons minutieuze casestudies en daaruit volgende wetenschapsfilosofische bespiegelingen overheerst een monomane focus op de lopende gang van zaken, ontbreekt een werkelijk diachroon perspectief, bestaat door rigide afbakening van disciplines nauwelijks aandacht voor grensoverschrijdend of interdisciplinair onderzoek en wordt bovenal de kloof

tussen praktijk en theorie zeer eenzijdig en gechargeerd uitgewerkt, waarbij elk normatief kader of zelfs maar metaperspectief als ‘hobbyhorses’ van filosofen dreigt te worden afgedaan. Hierbij staat de pejoratieve duiding van algemene methodologische en epistemische kwesties, zoals abstracte verklaringsmodellen en de hiervoor genoemde discontinuïteiten en crises voorop. Het punt is dat voor veel disciplines, diezelfde ‘dagelijkse praktijk’ nu juist wel door deze kwesties wordt bepaald en zonder reflectie daarop niet kan worden begrepen. Het gaat dan wellicht om jongere, interdisciplinaire wetenschapsgebieden, waarbij verschillende paradigmawisselingen of juist een methodenpluralisme aan de orde van de dag zijn, waarbij grondslagen minder zijn uitgekristalliseerd, die als het ware in een permanente crisis verkeren, of – milder geformuleerd – hun wetenschappelijke status moeten verdedigen en steeds op zoek zijn naar methodologische rechtvaardiging en algemene filosofische fundering van hun wetenschapsgebied. Disciplines, waarbij bovendien niet altijd duidelijk is welke inhoud in academische curricula nu moet worden gedoceerd en op welke wijze, of hoe de verworven kennis moet worden toegepast in de beroepspraktijk.

Tot deze disciplines behoort in menig opzicht ook de statistiek. Dat blijkt als we de gekozen historisch-filosofische invalshoek voortzetten; van vroege studies van Ian Hacking als *The Emergence of Probability* (1975) en *The Taming of Chance* (1989) tot Steven Stiglers meer recente *The Seven Pillars of Statistical Wisdom* (2016); van Deborah Mayo's *Error and the Growth of Knowledge* tot Andrew Gelmans pleidooi voor wijsgerige verankering in *Philosophy and the Practice of Bayesian Statistics* uit 2011; van Gerd Gigerenzer's vroege *The Empire of Chance; how probability changed science and everyday life* (1989) tot zijn meer recente studies over *decision-making* en *risk-communication*. Daarbij komt dat statistische handboeken dikwijls een coherente en niet ter discussie staande methodologie suggereren, terwijl het vaak gaat om onderling tegenstrijdige methoden, naast elkaar bestaande scholen, iets dat al sinds het vroege werk van vader en zoon Pearson, Fisher, Neyman en Bayesianen als Lindley en later Savage manifest bleek (Starmans, 2018a). De traditionele problemen rond de interpretatie van het kansbegrip komen daar nog eens bij. In data science wordt dit alles nog versterkt en staan statistical learning, gebaseerd op de verre van eensgezinde (inferentiële) statistiek en machine learning, gebaseerd op computational intelligence en algorithmic data analysis in menig opzicht tegenover elkaar. Dergelijke grondslagen crises leiden doorgaans tot meer alge-

mene wijsgerige of epistemische reflecties en dikwijls ook tot het ontstaan van subdisciplines als de Filosofie van het betreffende Wetenschapsgebied, zoals economie, psychologie, sociologie of biologie en meer recentelijk de informatica (Floridi, 2014). Filosofie van data science is dan een voor de hand liggende volgende schrede. We zullen hier niet eens ingaan op het feit dat data science en AI in het huidige AI-science debat nauw met elkaar verweven zijn en dat ook historisch beschouwd AI en de Filosofie van AI altijd twee kanten van dezelfde medaille hebben gevormd.

Hoe dan ook, wie zich op deze problematiek wil toeleggen kan het beste de probabilistische revolutie eind 19e en begin 20e eeuw als vertrekpunt nemen, die ontstond tegen de achtergrond van de opkomst van de moderne wetenschap, een proces van historisering van het wereldbeeld, de proliferatie van nieuwe disciplines, fragmentatie van kennis, grondslagen crises en spectaculaire vooruitgang in wiskunde en natuurkunde, en vervolgens terugkerende grondslagen crises in ‘nieuwe’ disciplines zoals psychologie, sociologie, economie, genetica, landbouwwetenschap en vele meer. Juist deze zochten fundering in de opkomende probabilistische wetenschap, kansrekening en statistiek en doen dat in zekere zin nog steeds (Krüger, 1981, 1987). Een kleine, zeer onvolledige bloemlezing:

- Pogingen binnen de psychologie om hardnekkige problemen rond replicerbaarheid, geldigheid en generaliseerbaarheid het hoofd te bieden (Paschler, 2012), (Open Science Collaboration, 2015);
- Sociale wetenschappers worstelen met toepassen en interpreteren van significance testing, p-waarden en betrouwbaarheidsintervallen (Meehl, 1990), (Gigerenzer, 2018);
- Publicatiebias en verwante problemen in de epidemiologische en medische literatuur, waarmee Ioannidis zich al vele jaren bezighoudt (Ioannidis, 2005);
- Cathy O'Neils aanklacht tegen slimme, maar biased algoritmen in *Weapons of Math Destruction* (2016) en de roep om Responsible and Explainable data science;
- Ethische problemen en morele dilemma's die intrinsiek lijken aan het gebruik van statistische methoden (Starmans, 2018a);
- Judea Pearl's pleidooi voor causaliteit om het project van de sterke AI te redden en de in zijn ogen desastreuze invloed van de statistiek op wetenschappelijke vooruitgang een halt toe te roepen (Pearl, 2018);
- Het feit dat veertig jaar na het werk van Kahneman en Tversky experts nog steeds moeite hebben op een con-

sistente manier met onzekere en onvolledige kennis te redeneren;

- Juridische besluitvorming in de zaak van Lucy de B. en andere rechtzaken;
- Discussies over interpretatie en extrapolaties van meetgegevens in klimaatdiscussies;
- De gemengde gevoelens, verdeelde standpunten en geconstateerde complicaties bij recente plannen voor invoering van statistiek in lager onderwijs, waarbij VVSOR terecht een mediërende rol wil spelen.

AI deze kwesties zijn zuiver methodologisch van aard en worden voor een belangrijk deel bepaald door de stand van zaken in de moderne statistiek. Daarnaast kunnen we constateren dat op vergelijkbare manier bepaalde richtingen binnen management science, accountancy, communicatiewetenschap hun beproefde methodologie overboord zetten ten faveure van een data-driven benadering. Kort gezegd, de crisis van statistiek en data science kan dan de crisis van de wetenschap worden, waarop de filosoof Edmund Husserl in een geheel andere context ongeveer 100 jaar geleden reeds preludeerde (Starmans, 2018c).

Boole, Shannon en Lebesgue

In weerwil van Galisons focus op het collectief en Stiglers Wet van de Eponymie is het zoeken naar specifieke ankers en fundamenten vanuit de hier gekozen aanpak wel degelijk zinvol. Het vormt een wezenlijk onderdeel van elke aanzet tot een filosofie van data science. We beperken ons vooralsnog tot een drietal wetenschappers die op heel verschillende wijze bijdroegen aan de fundering van data science, waarbij wel het probabilistische aspect op de voorgrond treedt. Toen George Boole (1815–1864) in 1854 de beginselen van zijn formele methode publiceerde in *An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities* werd het spoedig menigeen duidelijk dat hij een visionair was met verstrekkende ambities. Zoals uiteengezet door Starmans (2017b) zou de sterk wijsgerig georiënteerde Boole de latere 20e eeuwse analytische filosofie, logica en wetenschapsfilosofie ingrijpend beïnvloeden en laten zijn bijdragen aan de wetenschap zich als een triptiek duiden: formalisering van de (wijsgerige) logica in een 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,

waarmee hij een van de wegbereiders van de AI werd. Alereerst gaf Boole stevast een epistemische duiding aan het waarschijnlijkheidsbegrip; een kansuitspraak moet worden gezien tegen de achtergrond van de beschikbare kennis en de 'verwachting' van de persoon die zich aan de kansuitspraak committeert. Hij stond daarmee dicht bij Laplace's epistemische kansbegrip, maar hekelde diens 'principle of insufficient reason', waarbij gebeurtenissen even waarschijnlijk worden geacht indien er onvoldoende reden is het tegengestelde aan te nemen. Kansuitspraken betroffen bij Boole logische relaties tussen proposities. Ofschoon vormen van redeneren met onzekerheid doorgaans inductief zijn en dus buiten de deductieve logica vallen, vormde dit voor Boole geen reden logica en waarschijnlijkheid principieel te scheiden of de laatste uit te sluiten van de 'nieuwe' symbolische orde. Dit alles geschiedde bijna 80 jaar voordat de Rus Andrej Kolmogorov (1903–1987) de kansrekening een solide axiomatische basis zou geven.

Om diverse redenen heeft Amerikaanse wiskundige, elektrotechnisch ingenieur en uitvinder Claude Shannon (1916–2001) eveneens een onwrikbare plaats in de wetenschapsgeschiedenis ingenomen. Het was de jonge Shannon, die als student wiskunde en elektrotechniek aan MIT – nota bene – in zijn doctoraalscriptie *A Symbolic Analysis of Relay and Switching Circuits* (1937) zou wijzen op het belang van de Boolese algebra voor een systematisch ontwerp van elektronische schakelingen en circuits. Het ontwerpen en bouwen van telefooncentrales geschiedde indertijd vooral op basis van ad hoc principes, hetgeen door de sterke opkomst van de telefonie gaandeweg problematischer werd. Velen onderkenden de noodzaak van een 'design-science', maar Shannon zag als een van de eersten in dat Boole's werk hiertoe een solide wiskundige grondslag bood. Zonder twijfel vormde Shannons thesis een preambule tot de latere micro-elektronica-revolutie en de komst van de elektronische computer. Het uitgangspunt van zijn latere, bijkans klassieke *Mathematical Theory of Communication* uit 1948 behelsde de praktische vraag hoe een signaal (bericht) door een zender via een kanaal naar een ontvanger verstuurd wordt, hoe dit gecodeerd en gedecodeerd wordt, welke rol ruis en redundantie spelen. Diverse kwantitatieve, contextuele en dynamische aspecten van informatie werden op een elegante wiskundige, probabilistische wijze geanalyseerd c.q. van een maat voorzien: hoeveelheid, dichtheid, opslag, transport, waarde, variatie en spreiding, verwachting, onzekerheid, binaire codering, et cetera. Hierdoor werd in de wetenschappen steeds vaker de fundamentele rol van informatie in het grondslagenonderzoek zichtbaar

werd, waardoor Shannons bijdrage zou uitgroeien tot een universele probabilistische theorie voor communicatie met toepassingen in de elektrotechniek, de biomedische wetenschappen, de natuurwetenschappen en ook de sociale wetenschappen en de humaniora. Overal staan informatie en informatieprocessen centraal, die vanuit verschillende invalshoeken worden bestudeerd, geformaliseerd of gesimuleerd. Saillante voorbeelden hiervan zijn terug te vinden in de hedendaagse wetenschapsfilosofie en informatica (Kolmogorov-complexiteit), fysica (Gibbs-entropie), biologie (DNA-codering) en economie (speltheorie) (Floridi, 2014). In zekere zin zette Shannon hiermee de kroon op de probabilistische revolutie (Starmans, 2017b).

Zonder kansrekening geen statistiek, zonder statistiek geen data science. Deze dikwijls gezongen mantra maakt, gelet op de perikelen bij statistiek, de waarschijnlijkheidsrekening een welhaast vanzelfsprekend fundament van data science. Aan de reeds genoemde axiomatisering van Kolmogorov ging uiteraard het een en ander vooraf. Hier beperken we ons tot de bijdrage van Henri Lebesgue (1875–1941) die anders dan Boole en Shannon nauwelijks bekend is buiten de wiskunde. Net als Shannon zou hij met een vroege publicatie min of meer eigenhandig de basis leggen voor een nieuw onderdeel binnen de wiskunde. Zijn proefschrift *Intégrale, longueur, aire* (1902) vormde een synthese en generalisatie van veel versnipperd werk op het gebied van de integraalrekening, dat ruim 2000 jaar eerder een aanvang had genomen met Archimedes, vervolgens een belangrijke impuls kreeg in de 17e eeuw met de infinitesimaal rekening van Newton en Leibnitz, verder werd ontwikkeld in de 19e eeuw met de fundering van de analyse door Weierstrass, de epsilon-delta-definitie van limieten door Cauchy en uiteraard de Riemann integraal. Lebesgue breidde de toepassing ervan uit tot klassen van minder fraaie functies en introduceerde ook de Lebesgue-maat, waarmee de basis werd gelegd voor de maattheorie, ook in de moderne wiskunde een belangrijke subdiscipline. Lebesgue's maat generaliseerde het toekennen van bijvoorbeeld een lengte aan intervallen tot een grotere groep verzamelingen, die dan meetbaar kunnen worden genoemd. Zo kan aan uiteenlopende wiskundige objecten een lengte, oppervlakte, maar ook bijvoorbeeld een kans worden toegekend. Lebesgue's benadering werd een standaardmethode om van een maat over te gaan op een integraal. Maat- en integraaltheorie zijn dan ook nauw met elkaar verbonden en de maattheorie vormt ook historisch beschouwd een belangrijk fundament van de waarschijnlijkheidsrekening, dat daarmee ook als serieuze wiskundige discipline werd erkend.

Wie in het licht van de onderhavige problematiek op zoek gaat naar rotsvaste fundamenten voor data science kan terecht bij het recentelijk verschenen *Waarschijnlijkheidsrekening: maat theoretische uitgangspunten en fundamentele eigenschappen* van de Nederlandse wiskundigen Klaas van Harn en Piet Holewijn. Het boek doet in vele opzichten recht aan de hier geschetste betekenis van de kansrekening. De auteurs geven een bijkans volledig overzicht van de belangrijkste grondslagen van de waarschijnlijkheidsrekening. Vele elementaire onderdelen worden zeer grondig uitgewerkt; variërend van convergentiebegrippen, limietverdelingen voor extrema en voor sommen, karakteristieke functies tot stochastische wandelingen. De grote stellingen worden besproken, inclusief bewijzen. Kennis van maattheorie wordt voorondersteld, maar een uitvoerige en heldere appendix verschaft de benodigde basiskennis. Voor menig data scientist, statisticus of methodoloog zal dat geen overbodige luxe zijn, aangezien maattheorie doorgaans alleen in wiskundige of econometrische curricula een vaste plaats inneemt. Wie echter met het oog op de hier geschetste problematiek rond grondslagen en fundering de koninklijke weg wil volgen wordt in dit boek op zijn wenken bediend.

Epiloog

Thomas Kuhn benadrukte met zijn paradigmatheorie het belang van 'revoluties' en discontinuïteit in de geschiedenis van de natuurwetenschappen. Michel Foucault deed met zijn 'archeologische' methode ruwweg hetzelfde voor de geesteswetenschappen. Beide lieten zien dat de geschiedenis cruciale gebeurtenissen of perioden van radicale omwentelingen kent en trachten deze fundamentele cesuren aan te wijzen en te benoemen. Galison, Latour en Pickering hebben gelijk dat met dergelijke grote filosofische abstracties geen recht wordt gedaan aan de praktijk van onderzoek, maar hun eenzijdig historisch-sociologische benaderingen zijn evenmin toereikend. Tegelijkertijd toont die praktijk binnen diverse disciplines, waaronder statistiek en data science, aan dat vaak permanente methodologische crises optreden die allerminst op voorhand negatief hoeven te worden geïdentificeerd, juist een wezenlijk onderdeel vormen van die praktijk vragen, maar wel om een normatieve analyse en grondslagenonderzoek vragen die de discipline kunnen overstijgen. Het nut van dit soort grondslagen crises is getuige de talrijke hier genoemde historisch-filosofische statistische studies evident: ze kunnen leiden tot een zoektocht naar fundering van het vak, kruisbestuiving tussen de betreffende discipline en de fi-

losophie, een proeftuin voor de wetenschapsfilosofie en vormen soms zelfs een opmaat tot een zelfstandige status van de Filosofie van dat specifieke wetenschapsgebied.

LITERATUUR

- Floridi, L. (2014). *The Fourth Revolution; How the infosphere is reshaping human reality*. Oxford: Oxford University Press.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
- Harn, K. van, & Holewijn, P. J. (2019). *Waarschijnlijkheidsrekening: Maattheoretische uitgangspunten en fundamentele eigenschappen*. Amsterdam: VU University Press.
- Krüger, L., Daston, L., & Heidelberger, M. (Eds.). (1981). *The Probabilistic Revolution. Volume I: Ideas in History*. Cambridge, MA: MIT Press.
- Krüger, L., Daston, L., Heidelberger, M., Gigerenzer G., & Morgan, M. S. (Eds.). (1987). *The Probabilistic Revolution. Volume II. Ideas in the Sciences*. Cambridge, MA: MIT Press.
- Lovejoy, A. O. (1938). *The Great Chain of Being; A study of the history of an idea*. Harvard University Press (ed. 1990).
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological reports*, 66(1), 195–244.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530.
- Pearl, J., & MacKenzie, D. (2018). *The book of Why: The new science of cause and effect*. New York: Basic Books.
- Starmans, R. J. C. M. (2017a). Het nieuwe huis van Salomon: Peter Galison en de empirische traditie. *Filosofie*, 27(4).
- Starmans, R. J. C. M. (2017b). Van Heraclitus tot Shannon: de fluwelen revolutie van data in context en flux. *STATOR*, 18(1), 22–31.
- Starmans, R. J. C. M. (2018a) The Predicament of Truth: on Statistics, Causality, Physics and the Philosophy of Science. In Mark J. Van der Laan & Sherri Rose (Eds.), *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics, Springer.
- Starmans, R. J. C. M. (2018b). Een eigentijds Eutyphro-dilemma: over Deep Learning en de kolommen van orakeltaal. *Filosofie*, 28(3).
- Starmans, R. J. C. M. (2018c). Edmund Husserl: over de eenheid en de crisis van de wetenschappen. *Filosofie*, 28(6).
- Starmans, R. J. C. M. (2019). Prometheus unbound or Paradise regained - the concept of causality in the contemporary AI-data science debate. *Journal of the French Statistical Society*, Special Issue on Causality, Antoine Chambaz (ed).

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



EEN ZEGEN OF EEN RAMP

Bracht een routinematig uitgevoerde analyse een staatssecretaris ten val?

Na de oertijd waarin bijkans iedere onderzoeker zijn eigen computerprogramma's voor statistische berekeningen schreef of liet schrijven, en waarin standaardisatie ver te zoeken was, begon halverwege de jaren zeventig van de vorige eeuw de snelle opkomst van standaardprogrammatuur, met name van SPSS.

Dit waren de hoogtijdagen van de SSP, de voormalige Sectie Statistische Programmatuur van de VVSOR. De sectie had diverse subgroepen die zich bezighielden met programmatuur voor regressie- of variantieanalyse, met programmatuur in het algemeen of met pakketten als SAS en SPSS. Het was de tijd dat er nog een grote verscheidenheid aan computersystemen in gebruik was bij de Nederlandse universiteiten, onderzoeksinstituten en grote bedrijven. Alleen al bij de universiteiten konden men mainframes vinden van onder meer IBM, CDC, DEC en ICL en dan ook nog verschillende Operating Systems binnen één fabrikant. Een belangrijk aspect van het werk van de SSP was het onderzoeken van de juiste werking en de verschillen tussen de implementaties van een pakket op al deze verschillende systemen.

Naast dit zeer nuttige vergelijkingswerk werd er in de SSP ook fundamenteel nagedacht over taak en functie van standaardprogrammatuur. Een breed gedeeld gevoel van onbehagen was dat, met het wijd verspreiden van gemakkelijk toegankelijke programmatuur, ook statistisch niet-geschoolde onderzoekers geavanceerde analyses konden gaan verrichten. Veel bijeenkomsten hadden dan ook het thema 'Statistische Programmatuur: een zegen of een ramp?'. Heden ten dage zet een doorsnee-onderzoeker zijn data in een Excel-spreadsheet en gebruikt wat standaardfuncties om simpele tabellen en grafieken te produceren. En wil men verdergaande analyses dan dient dat spreadsheet als invoer voor bijvoorbeeld SPSS. Als men dit op een verstandige manier doet is daar niets op tegen, maar als men zonder meer de default instellingen gebruikt kan het wel eens mis gaan.

Het zou me niet verbazen als het recente aftreden van staatssecretaris Harbers mede is veroorzaakt door zo'n op de automatische piloot geproduceerde rapportage. Even de herinnering opfrissen: er was een lijst gerapporteerd van door asielzoekers gepleegde overtredingen en daarin waren een paar zeer ernstige gevallen onder de grote categorie 'overig' gerangschikt. Achteraf bleek overigens dat het merendeel van deze gevallen bij nadere inspectie veel minder ernstig te zijn dan in eerste instantie was aangenomen. Maar een groot deel van de Tweede Kamer schreeuwde moord en brand en betichtte het COA en Justitie van het misleiden van de Kamer en het verdoezelen van ernstige feiten. De staatssecretaris kon niet anders dan aftreden.

Maar wat is hier nu eigenlijk misgegaan? Allereerst zijn de gegevens niet goed gescreend alvorens ze te gebruiken. Als dat wel was gebeurd was er amper sprake geweest van de ernstige feiten die ons parlement in rep en roer brachten. Maar de feitelijke fout is dat men waarschijnlijk niet heeft nagedacht over de vorm van de rapportage. Een frequentietabel kan namelijk op verschillende manieren worden samengesteld: de categorieën alfabetisch, of van veel naar weinig voorkomend, óf naar gelang de inhoud van de categorie. Is het onderwerp iets als het favoriete biermerk dan zet men meestal het favoriete bier bovenaan en het minst aantrekkelijke als laatste. Maar als schooltypes het onderwerp vormen is het gebruikelijk de categorieën van laag naar hoog, of omgekeerd, te noemen. Bij rangschikken op frequentie zet men meestal de meest voorkomende categorie bovenaan en is het gebruikelijk categorieën die zelden scoren samen te vatten onder de noemer 'overig'. Dit om lange onoverzichtelijke tabellen te voorkomen.

Het moge duidelijk zijn dat bij het gewraakte onderzoek een inhoudelijke rangschikking van 'zwaar' naar 'licht' de aangewezen manier zou zijn geweest. Maar ik vermoed dat de samenstellers van de lijst ervoor hebben gekozen de meest voorkomende categorie bovenaan te zetten. Het zou zelfs kunnen dat men dit niet bewust zo heeft gekozen, maar dat dit de default instelling is van het programma dat men gebruikte. Tja, dan valt een ernstig feit dat maar een hoogst enkele keer voorkomt vanzelf onder 'overig'. En dan hebben we de poppen (of in dit geval de populist) aan het dansen.

En laat dit nu precies het soort automatisme zijn waar al zo'n 45 jaar geleden de leden van de SSP zich onder het motto 'Zegen of Ramp' zorgen over maakten. Kortom, werk aan de winkel voor onze nieuwe sectie Statistics Communication. GERRIT STEMERDINK is eindredacteur van STA&OR. E-mail: gjstemerding@hotmail.com

Nieuwe sectie **STATISTICS COMMUNICATION**

Tijdens de ALV op 20 maart 2019 is de VVSOR-sectie *Statistics Communication* opgericht die zich zal richten op het verbeteren van de communicatie van statistiek vanuit de statistici.

Data nemen in het dagelijks leven een steeds belangrijkere plek in. Algoritmen adviseren je over films en aankopen, je overlevingskans wordt berekend met behulp van modellen en in de krant krijgen we regelmatig veranderend advies over voedsel gebaseerd op de zoveelste studie. Maar ondanks dat statistische resultaten zo'n grote rol spelen zijn ze voor veel mensen erg lastig te begrijpen.

Het is daarom van belang dat statistiek begrijpelijker wordt, en dat kan vanuit twee kanten komen. Natuurlijk kan het publiek zich weren door zich te informeren om zo met een meer deskundige blik statistieken te benaderen. Maar ook wij als statistici kunnen helpen. Zo kunnen wij resultaten helder visualiseren en verduidelijken met voorbeelden, uitleggen op basis van welke karakteristieken een voorspelling is gebaseerd, en kansen in hun context plaatsen.

De nieuwe sectie Statistics Communication wil meer bewustzijn creëren over hoe lastig het eigenlijk is om statistiek te communiceren. Het doel is een platform te worden waar leden hun ervaringen en strategieën met elkaar kunnen delen. We zullen bijvoorbeeld symposia organiseren met onder andere sprekers die onderzoek doen op het gebied van communicatie of ervaringsdeskundigen zijn. Of er wordt besproken hoe "black box" modellen begrijpelijker gemaakt kunnen worden, of welke alternatieve modellen makkelijker te interpreteren zijn. Verder willen

we leden helpen bij het verbeteren van hun communicatievaardigheden. Zo zullen wij workshops organiseren waarbij deelnemers vaardigheden leren die ze meteen in de praktijk toe kunnen passen. De sectie tracht ook samenwerkingen aan te gaan met andere organisaties die vergelijkbare doelen nastreven om zo van elkaar te leren.

Tot nu toe heeft de sectie al twee drukbezochte activiteiten georganiseerd. Het eerste symposium (oktober 2018) met als onderwerp het communiceren van statistiek naar een breed publiek met impact met sprekers als Risk Literacy hoogleraar Gerd Gigerenzer en Thomas Ruigrok van de communicatieafdeling van het CBS. Daarnaast heeft in mei 2019 de eerste workshop plaatsgevonden waarbij deelnemers leerden om R Shiny applicaties en presentaties met interactieve visualisaties te maken.

We zijn erg blij met het enthousiasme dat tot nu toe voor de sectie is getoond en hopen dat veel VVSOR-leden deel blijven nemen aan onze activiteiten. Houd de website in de gaten om op de hoogte te blijven!

Bestuur van de sectie Statistics Communication
Sanne Willems
Niky van Buuren
Nynke Krol

Contact
www.vvsor.nl/statistics-communication
StatisticsCommunication@vvsor.nl



PIM VAN 'T HOF & JOAQUIM GROMICHO

Het softwarebedrijf ORTEC mocht dit jaar voor de tweede keer de VeRoLog Solver Challenge (VSC) organiseren. Zoals eerder in *STATOR* te lezen was (nummer 3 van 2017, pagina 14), was de vorige challenge een groot succes. De nieuwe editie was minstens zo succesvol, met lof alom van deelnemers en van de VeRoLog board.

Een nieuw team, onder leiding van Pim van 't Hof, had een probleem opgesteld waarbij distribueren en 'vervolgbezoekritten' moeten worden gecombineerd. Een dergelijke problematiek komt voor bij het leveren en installeren van apparatuur. (De probleembeschrijving is te vinden via https://verolog2019.ortec.com/pdf/VSC2019_info.pdf en via <https://doi.org/10.1007/s41604-019-00011-8>.) Dezelfde puzzel moet ook worden opgelost bij het inzetten van *merchandisers* die na het leveren van goederen de visuele opstelling in de winkels gaan organiseren.

Net als de vorige keer bestond ook deze challenge uit twee onderdelen: een *all-time-best challenge* en een *restricted resources challenge*.

Bij de all-time-best challenge konden de deelnemers hun oplossingen – verkregen op welke manier dan ook – uploaden via de Challenge-website. Op de website werden per instantie de tussenstanden bijgehouden, zodat ieder team te allen tijde kon zien hoe goed ze presteerden ten opzichte van de andere teams. Zo was voor alle deelnemers te zien dat team mijg, winnaar van de vorige VSC, al redelijk snel aan de leiding ging bij alle instanties. Slechts op één van de 25 instanties werd team mijg na

een paar maanden voorbijgestreefd door team UOS, een VSC-debutant; voor alle andere instanties gaf team mijg de koppositie niet meer uit handen.

Bij de restricted resources challenge moesten de teams, naast hun oplossingen voor een (nieuwe) set instanties, ook hun *solver* uploaden. Na validatie van de ge-uploadde oplossingen werden op basis van de resultaten acht finalisten geselecteerd. De solvers van de finalisten werden vervolgens gerund op een set verborgen instanties om de uiteindelijke uitslag te bepalen. De top-3 van de restricted resources challenge werd bekend gemaakt op 4 juni 2019 tijdens het galadiner van de VeRoLog-conferentie in Sevilla, waar ook de bijbehorende prijzen werden uitgedeeld.

Op basis van de resultaten bij de all-time-best challenge lag het in de lijn der verwachting dat team mijg -net als vorig jaar- met de hoofdprijs aan de haal zou gaan. Menig 'ingewijde' was dan ook verrast toen team mijg als winnaar van de tweede prijs naar voren werd geroepen. De solver van team UOS bleek op 19 van de 25 verborgen instanties de solver van team mijg te hebben verslagen. Daarmee werd team UOS de afgetekende winnaar van de VeRoLog Solver Challenge 2019. Het podium werd gecompleteerd door team COKA coders uit Nieuw-Zeeland, met in de gelederen *STATOR*-redactielid Caroline Jagtenberg. Caroline verruilde in april vorig jaar ORTEC voor een baan aan de University of Auckland (zie *STATOR* nummer 1 van 2018, pagina 4) en mocht als zodanig deze keer wél deelnemen aan de door ORTEC



Van links naar rechts: Caroline Jagtenberg vanuit Nieuw-Zeeland als vertegenwoordiger van team COKA coders (3e prijs); Benjamin Graf uit Osnabrück van team UOS (1e prijs); Joaquim Gromicho als vertegenwoordiger van de organisatie; en Martin Geiger uit Hamburg van team mijg (2e prijs).

georganiseerde VeRoLog Solver Challenge.

Dat er verschillende manieren zijn om een ingewikkeld routeringsprobleem als dat van de VSC 2019 'te lijf te gaan' is welbekend. Het was dan ook erg interessant om van de drie prijswinnaars te horen welke tools, technieken en trucs zij hadden toegepast om hun goede oplossingen te vinden. Deelnemers aan de VeRoLog-conferentie in Sevilla kregen daartoe de gelegenheid tijdens de aan de VSC 2019 gewijde sessie op dinsdagmiddag 4 juni. (Merk op dat het tijdens die sessie nog niet bekend was welke teams in de prijzen waren gevallen; dat werd pas 's avonds bekendgemaakt.) ORTEC organiseerde een mini-seminar op haar hoofdkantoor in Zoetermeer op 30 augustus waar de drie prijswinnaars hun winnende aanpak presenteerden.

Wij feliciteren alle winnaars met hun prestaties, en bedanken alle deelnemers en bij de organisatie betrokken ORTEC-collega's voor hun bijdrage aan het succes van deze VeRoLog Solver Challenge.

PIM VAN 'T HOF deed als PhD-student en postdoc in respectievelijk Durham (Engeland) en Bergen (Noorwegen) onderzoek in de algoritmische grafentheorie. Bij ORTEC werkt hij als Senior OR Engineer en team lead in het team dat verantwoordelijk is voor de ontwikkeling van ORTEC's routingsoftware. E-mail: Pim.vantHof@ortec.com

JOAQUIM GROMICHO is Scientific and Education Officer bij ORTEC en hoogleraar Toegepaste Optimalisatie aan de VU in Amsterdam. Hij is hoofdredacteur van *STATOR*. E-mail: joaquim.gromicho@ortec.com



Young Statisticians are looking forward to the new season

At this moment not much can be announced for the Young Statisticians. However the Board is busy composing an interesting program for the new season.

One exiting event was a recent visit to Philips on September 10, 2019. From lightbulbs to televisions to improving sleep with deep learning. We learned about the latest developments happening at the Philips High Tech Campus!

Please visit our website at <https://www.vvsor.nl/young-statisticians/events/> for more info about this visit and all other plans for the near future.



OPROEP

Hemelrijk en Van Zwet Awards

Op de Annual Meeting van volgend jaar zullen, zoals gebruikelijk, weer de jaarlijkse prijzen voor de beste MSc scriptie (Hemelrijk Award) en het beste proefschrift (Van Zwet Award) worden uitgereikt.

Een oproep om te nomineren zal in het volgende nummer van *STATOR* verschijnen, maar u kunt nu vast actief om u heen kijken naar potentiële kandidaten voor deze prestigieuze Awards.



Na het 100-jarig jubileum congres in 1985 in Amsterdam, dat werd geopend in aanwezigheid van koningin Beatrix, keert het World Statistics Congress weer terug naar Nederland. Het 63e WSC zal van 11 – 15 juli 2021 in het World Forum in Den Haag plaatsvinden. Het National Organising Committee, onder co-voorzitterschap van Eric Schulte Nordholt en John Bailer, is al geruime tijd bezig dit voor te bereiden. Ook de VVSOR is hierbij actief betrokken.

World Statistics Congresses hebben hun oorsprong in het werk van de Belgische geleerde Adolphe Quetelet die in 1853 een eerste Europees congres over statistiek organiseerde. Na nog enkele van deze congressen werd in 1885 het International Statistical Institute opgericht om de continuïteit van deze congressen en de vele internationale contacten te bewerkstelligen. In de beginjaren lag de focus vooral op harmonisatie van methoden en technieken tussen de verschillende Europese landen. Op die manier wilde men waarborgen dat economische en demografische cijfers vergelijkbaar zijn. Het ISI beschikt al sinds 1913 over een Permanent Office dat in Den Haag is gevestigd. De reeks tweejaarlijkse WSC's, die oorspronkelijk ISI Sessions hetten, is slechts door de beide Wereldoorlogen onderbroken. Vanaf de jaren '50 is het ISI uitgebreid met Associations, momenteel zeven in getal, die zich bezighouden met specifieke onderwerpen zoals Mathematical Statistics, Statistical Education of Statistical Computing.

De WSC's kenmerken zich door een open en gastvrij karakter. Beginnende wetenschappers en *hot shots* men- gen zich zonder drempels in discussies en vele langdurende vriendschappen zijn hier ontstaan. Dit congres is daarom een unieke kans voor Nederlandse statistici om zonder hoge reiskosten nieuwe contacten op te doen en veel van de grootste experts ter wereld te ontmoeten. U bent welkom, niet alleen als deelnemer, ook uw ideeën over onderwerpen of sponsoring zijn waardevol.

INFORMATIE

ISI Permanent Office (P.O. Box 24070, 2490 AB The Hague, phone +31 70 33 75 737)

e-mail: isi@cbs.nl

ISI website: isi-web.org

Congres website: isi2021.org



Het 62e World Statistics Congress vond van 18 – 23 augustus 2019 plaats in Kuala Lumpur, Maleisië. Meer dan 2500 deelnemers konden kiezen uit honderden presentaties. Ook presenteerde Maleisië zich met een uitgebreid cultureel programma.

Een groot aantal deelnemers uit Nederland was in Kuala Lumpur aanwezig. Samen met ISI president John Bailer nodigen zij op deze foto alle statistici wereldwijd uit om deel te nemen aan het 63e WSC in 2021 in Den Haag.