STAtOR

Periodiek van de VVSOR jaargang 25, nummer 2, juli 2024

STAtOR
Jaargang 25, nummer 2, juli 2024

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 3 of 4 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Caroline Jagtenberg, Alex Kuiper, Miriam Loois, Guus Luijben (eindredacteur), Kerry Malone, Gerard Sierksma, Richard Starmans, Gerrit Stemerdink (eindredacteur), Vanessa Torres van Grinsven, en Inez Zwetsloot. Vaste medewerkers: Jelke Bethlehem, John Poppelaars en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, stator@vvsor.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: dr. Tsega Kahsay Gebretekle, secretaris@vvsor.nl; Penningmeester: dr. Rebecca Kuiper, penningmeester@vvsor.nl; Algemeen bestuurslid: dr. Marianne Jonker.

Voorzitters van de secties: dr. Marianne Jonker (Biometrical Section); dr. ir. Marjan van den Akker (Section for Operations Research); prof. dr. ir. Frank van der Meulen (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Iris Yocarini (Section Data Science); Luise Gummi, MSc (Young Statisticians); dr. Sanne Willems (Section Statistics Communication); dr. Stéphanie van den Berg (Section Statistics Education).

Leden- en abonnementenadministratie van de VVSOR

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

Voor advertenties

Prof. dr. J.A.S. Gromicho, stator@vvsor.nl STAtOR verschijnt in maart, juli en december.

Uitgever

ISSN 1567-3383

Actueel

We proberen als redactie meestal aan te sluiten bij de actualiteit. Tenslotte is een van de doelen van STAtOR het verspreiden van actuele kennis over praktische toepassingen in ons vakgebied. Maar ook de actualiteit van de wereld daarbuiten kan aanleiding zijn voor artikelen waarbij ons vak een rol speelt. Dit nummer lijkt dat weer goed gelukt te zijn.

Het komende Europees kampioenschap voetbal voor mannen inspireerde Henk Tijms voor een column over kansen in strafschoppenreeksen en de Amerikaanse verkiezingen van eind dit jaar brachten Jelke Bethlehem op het idee voor een column over de vele mogelijkheden om de resultaten daarvan grafisch weer te geven.

Een 'hot topic' in de statistiek is de grote aandacht voor de e-waarde. Naast een prestigieuze presentatie hierover door Peter Grünwald, Rianne de Heide en Wouter Koolen bij de Royal Statistical Society en artikelen in Nederlandse kwaliteitskranten ontving Peter Grünwald recentelijk een 2,5 miljoen ERC-grant voor verder onderzoek hiernaar. Rianne de Heide schrijft, naar aanleiding van doorlekkende luiers, een artikel hierover. Gerard Sierksma draagt een column bij met mijmeringen over deze en andere e-waarden.

Ook Mark van der Laan en Richard Starmans behandelen een nieuwe ontwikkeling, de Highly Adaptive Lasso. Zij merken hierbij terloops op dat Bayesianen en Frequentisten niet zoveel verschillen.

Twee verdere artikelen zijn gericht op heel praktische toepassingen. Maaike Sneider et al gaan in op het ontwerpen van netwerken waarbij gelijktijdig het gedrag van de gebruiker daarvan wordt meegenomen in de optimalisering. En Thomas Breugem kijkt naar de toepassing van OR bij de 'eerlijke verdeling' van taken etc. zoals dat bijvoorbeeld bij het maken van dienstroosters een rol speelt.

Natuurlijk is er verenigingsnieuws. Op de VVSOR Annual Meeting van 21 maart jongstleden werd aan Richard Post de Willem R. van Zwet Award uitgereikt voor zijn winnende proefschrift. Casper Albers schrijft hierover en ook doet hij een oproep kandidaten te nomineren voor de 5-jaarlijkse Van Dantzig Prijs die volgend jaar weer wordt toegekend. Helaas is ons een markant lid ontvallen, Ewout Steyerberg en Kit Roes gedenken René Eijkemans.

Dit nummer sluit af met een uitnodiging om volgend jaar deel te nemen aan het World Statistics Congress 2025 van het International Statistical Institute. Deze bijeenkomst zal naar verwachting met ruim 2000 deelnemers plaatsvinden in Den Haag, en is daarmee een uitgelezen mogelijkheid voor Nederlandse statistici om veel collega's te ontmoeten.

Wij wensen u een goede zomer en veel leesplezier.

De STAtOR-redactie

INHOUD

Henk Tijms

Column

Foto: FIFA World Cup 2018, Colombia vs. England, Wikimedia Commons, photo Oleg Bkhambri

Strafschoppenseries

Om het voetbalspel als schouwspel nog spannender te maken, besloot de FIFA in 1978 tot het invoeren van een strafschoppenserie als de wedstrijd een winnaar vereiste en ook na dertig minuten verlenging onbeslist was. In de FIFA wereldkampioenschappen voetbal zijn er tot nu dertig strafschoppenseries geweest. De eerste serie was op 8 juni 1982, toen West-Duitsland Frankrijk met 5-4 uitschakelde in een strafschoppenserie om door te gaan naar de finale.

De wedstrijd was geëindigd in 3-3, waarvan vier doelpunten in extra tijd, en is het best herinnerd vanwege de afschuwelijke overtreding van de Duitse doelman Harald Schumacher op de Franse speler Patrick Battiston, die nog steeds de gevolgen ondervindt van de door de Nederlandse toparbiter Charles Corver niet met rood bestrafte overtreding. Om het nog erger te maken voor het Franse team, was Schumacher de Duitse held in de strafschoppenserie. Hij redde twee strafschoppen voordat Hrubesch de winnende strafschop scoorde.

In een strafschoppenserie nemen de twee teams om de beurt een strafschop, en nadat elk team vijf pogingen heeft gedaan wint het team met de meeste succesvol genomen strafschoppen. Hierbij geldt dat de strafschoppenserie voortijdig eindigt met minder dan tien strafschoppen zodra duidelijk is dat het ene team niet meer op gelijke hoogte kan komen met het andere team. Als de stand nog steeds gelijk is nadat elk team vijf strafschoppen heeft gehad, gaat de strafschoppenserie door tot 'sudden death', waarbij de twee teams afwisselend een strafschop nemen totdat het ene team scoort en het andere team mist in dezelfde ronde.

Veel is geschreven en gezegd over de psychologie en de tactiek van het strafschoppen nemen. De toss is en blijft de belangrijkste beïnvloeder van de uitkomst van de strafschoppenserie: de Spaanse

hoogleraar Ignacio Palacios-Huerta van de London School of Economics - auteur van de bestseller *Beautiful Game Theory: How Soccer Can Help Economics* - analyseerde 1343 strafschoppen uit 129 strafschoppenseries en vond dat het team dat als eerste begon in 60,5% van de gevallen won. Overigens Marco van Basten met 92,7% en Alan Shearer met 93,5% hebben de hoogste scoring percentages van topspelers die meer dan 50 strafschoppen hebben genomen (Messi en Ronaldo hebben scoring percentages van ongeveer 78% en 84%), en dat terwijl zowel Engeland als Nederland tot de slechtst presterende landen behoren bij het nemen van strafschoppenseries. Het was niet verstandig van Gianluigi Buffon, de fameuze doelman en aanvoerder van het Italiaanse team, om na het winnen van de toss in de kwartfinale van het Europees Kampioenschap 2008 de eerste strafschop te laten aan Spanje dat uiteindelijk de strafschoppenserie won. De verklaring dat het voordeel biedt om als eerste te beginnen met het nemen van strafschoppen is dat de ploeg die begint vaak voorstaat in de serie: 1-0, 2-1, 3-2, etc. De ploeg die als tweede mag schieten voelt dus telkens de druk om gelijk te moeten maken, en dat leidt tot gemiddeld meer gemiste strafschoppen. Het stressniveau beheersen is niet zo simpel wanneer veel op het spel staat.

Deze column richt zich op een simpel kansmodel voor de berekening van de kansverdeling en de verwachtingswaarde van het aantal strafschoppen in een strafschoppenserie. De modelaanname is dat de uitkomsten van de opeenvolgende strafschoppen onafhankelijk van elkaar zijn en dat elke speler dezelfde kans p heeft om een strafschop te benutten. Voor voetbalwedstrijden op het niveau van het FIFA wereldkampioenschap of de UEFA Champions League is het redelijk voor p de waarde 0,7 te nemen. Een strafschoppenserie vereist tenminste zes strafschoppen en eindigt na precies zes strafschoppen als het ene team de eerste drie strafschoppen benut en het andere team de eerste drie strafschoppen mist. Noteren we met a_k de kans dat in een strafschoppenserie k strafschoppen nodig zijn om tot een winnaar te komen, dan geven combinatorische kansberekeningen de volgende waarden bij succeskans $p = 0{,}7$:

$$a_6 = 0{,}01852,\ a_7 = 0{,}07673,\ a_8 = 0{,}14519,$$
$$a_9 = 0{,}22835,\ a_{10} = 0{,}25964.$$

De kans dat de strafschoppenserie eindigt met 'sudden death' (geen winnaar na 10 strafschoppen) is

$$1 - (a_6 + \cdots + a_{10}) = 0{,}27157.$$

Als na de vijf rondes met 10 strafschoppen er nog geen winnaar is, dan heeft het aantal rondes tot 'sudden death' een geometrische kansverdeling met parameter $1 - (0{,}7^2 + 0{,}3^2) = 0{,}42$, oftewel het verwachte aantal nog resterende strafschoppen is $2 \times \frac{1}{0{,}42} = 4{,}76190$. Dus de verwachtingswaarde van het aantal strafschoppen in een strafschoppenserie is gelijk aan

$$\sum_{k=6}^{10} k a_k + 0{,}27157 \times (10 + 4{,}76190) = 10{,}470.$$

Dit alles onder modelaanname dat dat de de uitkomsten van de opeenvolgende strafschoppen onafhankelijk van elkaar zijn en dat elke speler dezelfde succeskans 0,7 heeft om een strafschop te benutten. Het resultaat van gemiddeld 10,470 strafschoppen bij een succeskans 0,7 kan worden uitgebreid tot een expliciete formule die voor elke waarde van de succeskans van toepassing is: bij eenzelfde succeskans p voor elke speler, wordt de verwachtingswaarde van het aantal benodigde strafschoppen in een strafschoppenserie gegeven door de formule

$$\frac{1}{r} + r[56 - r(203 - 4r(93 - 70r))],$$

waarbij $r = p(1 - p)$. In deze formule is het verwachte aantal strafschoppen symmetrisch in p en $1 - p$; bijvoorbeeld, voor een succeskans van 80% heeft het gemiddeld aantal benodigde strafschoppen dezelfde waarde 11,353 als voor een succeskans van 20%. De formule laat zien dat het gemiddeld aantal strafschoppen minimaal is voor een succeskans van 50% en dan de waarde 10,031 heeft. Anders gesteld, in het kansmodel waarin elke speler dezelfde succeskans heeft, is het gemiddeld aantal benodigde strafschoppen altijd meer dan 10.

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Recentelijk is de vierde druk van zijn boekje "*Kansrekening in werking– een moderne aanpak*", dat speciaal geschreven is voor VWO- scholieren in de bovenbouw, verschenen bij de educatieve wiskunde uitgeverij Epsilon. E-mail: h.c.tijms@xs4all.nl

Afbeelding van Freepik

De e-waarde en het ‘Safe Testing’ paper

Rianne de Heide

De e-waarde is veel in het nieuws de laatste tijd: onder andere Trouw, de Volkskrant, de NRC en de New Scientist besteedden er de afgelopen maanden aandacht aan naar aanleiding van twee gebeurtenissen.

In januari presenteerden Peter Grünwald, Rianne de Heide en Wouter Koolen hun pioniersartikel ‘Safe Testing’ (Grünwald, De Heide en Koolen, 2024) bij de Royal Statistical Society in Londen. Begin april werd bekend dat Peter Grünwald een Europese onderzoeksbeurs van 2,5 miljoen euro krijgt voor onderzoek naar de e-waarde.

De afgelopen jaren groeit het wiskundig statistisch onderzoek naar e-waardes haast exponentieel, er komt steeds meer software beschikbaar, en de eerste toepassingen komen van de grond. Een goed moment voor STAtOR om aan Rianne de Heide te vragen: wat is die e-waarde nu precies, en wat staat er in het Safe Testing paper?

Een experiment

Toen mijn zoon afgelopen februari 4 maanden oud was stapten we over van maat 1 naar maat 2 wasbare luiers. Voor de nacht kun je een extra absorberend laagje, een zogenaamde *booster* in de luier doen, en ik besloot een paar extra boosters te kopen bij de winkel waar we ook al onze wasbare luiers hadden gekocht. De eigenaar van de winkel vroeg of wij een nieuw type booster voor hen zouden willen testen. Dat leek mij wel leuk, dus wij kregen 2 normale boosters en 2 boosters van het nieuwe type opgestuurd. Ik besloot er een experiment van te maken.

We hebben twee groepen:

- Groep A: de normale boosters,
- Groep B: de nieuwe boosters.

We verzamelen de data sequentieel: na iedere nacht hebben we een nieuw datapunt. De uitkomst is binair: wel (1) of niet (0) doorgelekt. We nemen aan dat de data i.i.d. Bernoulli verdeeld zijn met een bepaalde parameter — de doorlekkans — $\theta_j \in [0,1], j \in \{a,b\}$. We hebben dus stromen data $Y_{1,a}, Y_{2,a}, \ldots$ i.i.d. ~ P_{θ_a} en $Y_{1,b}, Y_{2,b}, \ldots$ i.i.d. ~ P_{θ_b}, en we zijn benieuwd of we een significant verschil kunnen vinden tusssen θ_a en θ_b, wat wil zeggen dat de twee types boosters een verschillende doorlekkans hebben. We nemen als significatieniveau $\alpha = 0,05$. De nulhypothese en alternatieve hypothese zijn als volgt:

- $H_0 : \theta_a = \theta_b$,
- $H_1 : \theta_a \neq \theta_b$.

We verzamelen de data in paren: de nachten met de normale en de nieuwe boosters wisselen elkaar af. Na elk paar kunnen we een e-waarde berekenen: de conclusies en type I foutgaranties houden stand als we dat doen, en op basis van de resultaten besluiten of we nog meer data gaan verzamelen, of stoppen met het experiment. Dit kan eenvoudig door e-waardes te gebruiken. Hoe de e-waarde precies geconstrueerd wordt voor dit voorbeeld is te vinden in het paper (Turner, Ly en Grünwald, 2022), en er is ook een R-package: *safestats*, (Turner, Ly, Pérez-Ortiz e.a., 2022), wat ik gewoon kan gebruiken. Ik gebruik hiervoor de volgende code:

```
safe.prop.test(ya=ya, yb=yb, pilot=T)
```

In de vector ya zit de stroom data uit groep A, in de vector yb zit de stroom data uit groep B, en ik heb het argument 'pilot' op true gezet omdat ik geen enkel idee heb van de te verwachten effect size, en ik nog niet weet hoe lang ik met het experiment door wil gaan, misschien stop ik gewoon als ik geen zin meer heb in het experiment. De e-waarde die ik dan heb gevonden, kan ik interpreteren als bewijs tegen de nulhypothese, en het maakt niet uit om wat voor reden ik ben gestopt. Maar ik ga ook stoppen als ik een e-waarde krijg die groter dan $1/\alpha = 20$ is, want dat betekent 'genoeg' bewijs tegen de nulhypothese bij significantieniveau $\alpha = 0,05$. Na elke twee nachten plot ik de e-waarde in Figuur 1.

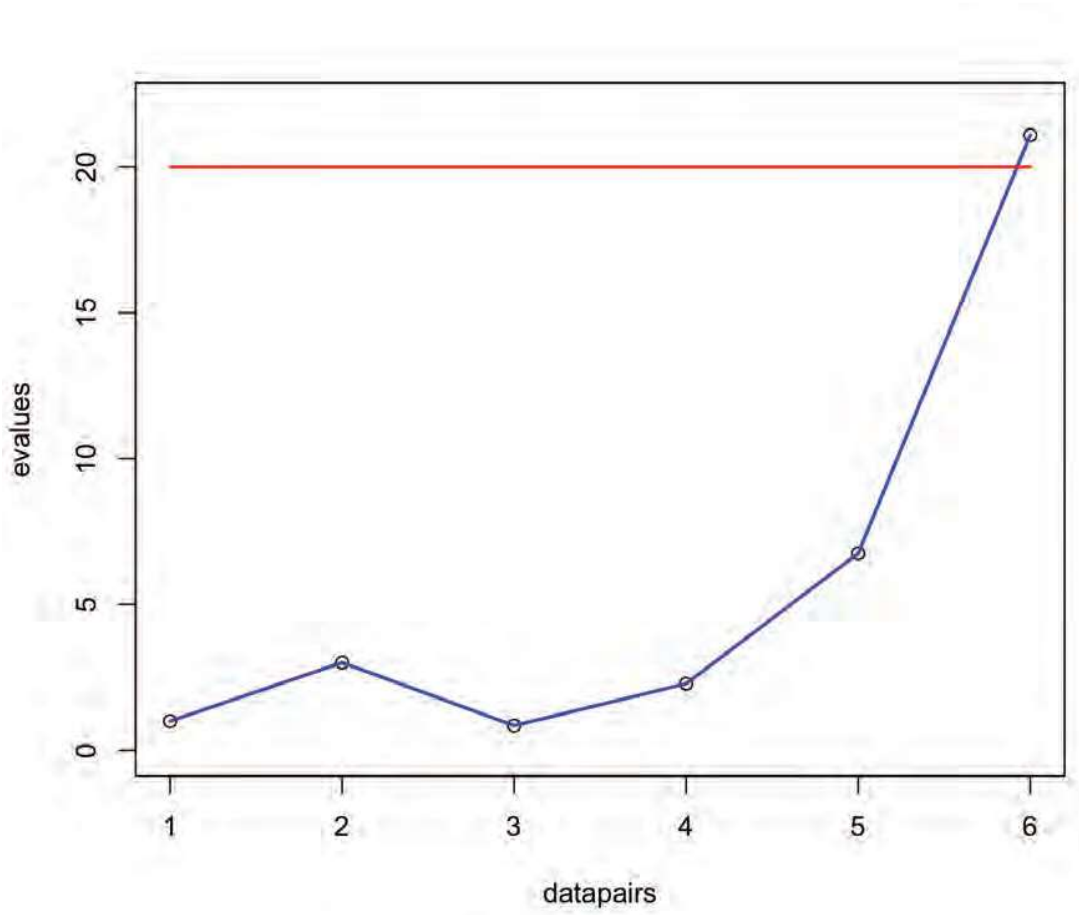
Het resultaat was overweldigend en in een richting die ik niet had verwacht. De nieuwe boosters lekten iedere nacht door, de normale alleen in de derde nacht (Tabel 1). Na 12 datapunten, 6 in iedere groep, was de e-waarde 21,088 en dus voldoende om het experiment te stoppen, de nulhypothese te verwerpen, en de luierwinkel te informeren dat die nieuwe boosters overtuigend veel slechter waren dan de normale.

Hoe moet dit met p-waardes?

Bij een klassiek onderzoek met p-waardes, had ik een schatting moeten maken van de *effect size* en met een poweranalyse berekenen wat de *sample size* zou moeten zijn. Want als het hele *sampling plan* niet van te voren vastligt, krijg je geen geldige p-waarde. Maar ik had geen flauw idee van de

Tabel 1: De data: een meting bestond uit twee nachten, een met de normale booster en een met de nieuwe booster. De uitkomst is binair: wel (0) of niet (1) doorgelekt.

	1	2	3	4	5	6
normale boosters	0	0	1	0	0	0
nieuwe boosters	1	1	1	1	1	1



Figuur 1: De blauwe lijn geeft de plot van de e-waardes weer na elke toevoeging van een datapaar: één observatie met de normale boosters en één observatie met de nieuwe boosters. De rode lijn is de grenswaarde 20, en we stoppen het experiment als we een e-waarde groter dan 20 vinden.

effect size... Ik heb toegepaste statistici om hulp gevraagd hoe dit in de praktijk wordt aangepakt, en ik kreeg een lange uitleg over een pilot studie doen om het effect te schatten, vervolgens die pilot data weggooien en een nieuwe studie doen op basis van het geschatte effect... En voor zo'n pilot studie werd me het aantal van 12 datapunten per groep aangeraden. Stel dat het dat het echte experiment ook 12 nachten had geduurd, had het mijn zoon in totaal ten minste 18 doorleknachten gekost: het lijkt me dat het experiment dan voortijdig gestopt was vanwege ethische redenen, waarna we bovendien geen p-waarde hadden kunnen rapporteren op basis van de data die tot op dat moment verzameld waren, want tussentijds stoppen invalideert immers p-waardes en hun foutganties. Werken met e-waardes in zo'n soort experiment lijkt me dus veel natuurlijker.

De e-waarde

In het Safe Testing paper leggen we een basis voor een nieuwe theorie voor hypothesetoetsen. Onze theorie is gebaseerd op e-waardes. Als we hypotheses willen toetsen, kunnen we de nulhypothese H_0 en de alternatieve hypothese H_1 beschouwen als verzamelingen van kansverdelingen. Een e-waarde S is een niet-negatieve stochastische variable (d.w.z. een functie van de data) waarvoor geldt dat voor iedere verdeling $P \in H_0$, de verwachting van de e-waarde onder P ten hoogste 1 is: $\mathbb{E}_P[S] \leq 1$. Dat betekent dat je, als de nulhypothese waar is, niet verwacht dat de e-waarde (veel) groter wordt dan 1. Is de e-waarde toch groter dan 1, dan kun je dat interpreteren als bewijs tegen de nulhypothese. Voor een significantieniveau $\alpha \in (0,1)$ kunnen we een hypothesetoets definiëren die H_0 verwerpt als de e-waarde groter is dan $1/\alpha$.

Voordelen van e-waardes

Toetsen met e-waardes biedt verschillende voordelen ten opzichte van het toetsen met p-waardes. Hier zijn de belangrijkste:

- e-waardes gedragen zich (ze behouden foutgaranties en blijven interpreteerbaar) onder *optional continuation*: wanneer je op basis van de resultaten tot nu toe besluit om nog meer data aan het experiment toe te voegen, zoals in het experiment met de boosters, of in een meta-analyse.
- Ze hebben een eenvoudige interpretatie als bewijs tegen de nulhypothese, die ook geldig blijft als je het hele concept van *significantie* afwijst (Amrhein, Greenland en McShane, 2019).
- Ze zijn flexibel: je kunt ze maken op basis van Bayesiaanse prior informatie, op basis van eerdere data, maar ook op basis van minimax garanties, en in alle gevallen behouden ze de type I foutgaranties. Je kunt ze allemaal eenvoudig met elkaar combineren: de Bayesianen en frequentisten hebben een gemeenschappelijke munt!
- Voor veel toepassingen, o.a. in meervoudig toetsen, kunnen veel afhankelijkheidsvereisten die nodig zijn voor het werken met p-waardes worden losgelaten (bijvoorbeeld, de e-waarde-Benjamini-Hochberg procedure, e-BH, biedt FDR controle onder willekeurige afhankelijkheid (Wang en Ramdas, 2022)).

De wiskundige onderbouwing hiervan, meer voordelen, de verschillende interpretaties, en meer: dat staat in het Safe Testing paper.

Wiskundige bijdrage van het paper

De belangrijkste wiskundige bijdrage van het Safe Testing paper is een algemene manier voor het construeren van goede e-waardes voor hypothesetoetsproblemen met een meervoudige nulhypothese. E-waardes en de gerelateerde toetsmartingalen bestaan al langer, alleen was het tot het verschijnen van het Safe Testing paper op Arxiv (en een half jaar later het Universal Inference paper (Wasserman, Ramdas en Balakrishnan, 2020)) onbekend hoe je e-waardes kunt construeren voor problemen met een meervoudige nulhypothese, d.w.z. problemen waarbij er meer dan één kansverdeling onderdeel uitmaakt van de verzameling H_0. De meeste praktische problemen hebben zo'n meervoudige nulhypothese, denk aan de t-toets en de chi-kwadraattoets. Daarbij is een heel belangrijke vraag: 'Wat is een *goede* e-waarde?'. Als je naar de definitie kijkt, dan is de e-waarde die altijd 1 is, ook een geldige e-waarde. We willen natuurlijk een e-waarde die snel heel groot wordt als de nulhypothese niet waar is. In het paper wordt hievoor een criterium gedefinieerd, genaamd GRO: growth-rate optimal. Je kunt het zien als een equivalent van de *power* in onze setting met het optioneel toevoegen van data en stoppen wanneer je maar wilt. In het Safe Testing paper bewijzen we dat onze algemene manier om e-waardes te construeren optimale toetsen oplevert in termen van GRO. We definiëren ook een worst-case versie (GROW) en een relatieve versie (REGROW). In het paper is verder een overzicht van gerelateerd werk te zien, waaronder sequentieel toetsen (denk aan Wald's werk), wat een stuk restrictiever is dan ons paradigma.

De toekomst

De toekomst van de e-waarde ziet er wat mij betreft heel zonnig uit: er is nog zoveel fundamenteel onderzoek dat nog moet worden gedaan, en dat wordt nu ook door meerdere geweldige groepen over de hele wereld opgepakt. Ondertussen zien we e-waardes elders ook overal opduiken: in de multiple testing community bijvoorbeeld. Bijzonder goed nieuws is de ERC Advanced grant (2,5 miljoen euro) voor Peter Grünwald waarmee hij de komende jaren PhD-studenten en postdocs gaat aanstellen voor e-waarde-onderzoek.

Literatuur

V. Amrhein, S. Greenland en B. McShane. „Scientists rise up against statistical significance". In: *Nature* 567.7748 (2019), p. 305–307.

P. Grünwald, R. De Heide en W. Koolen. „Safe Testing". In: *Journal of the Royal Statistical Society Series B* (2024). Forthcoming discussion paper.

R. Turner, A. Ly en P. Grünwald. *Generic E-Variables for Exact Sequential k-Sample Tests That Allow for Optional Stopping*. Jun 2022. eprint: 2106.02693.

R. Turner, A. Ly, M. F. Pérez-Ortiz e.a. „R-package safestats". In: *R package version 0.8.7* (2022).

R. Wang en A. Ramdas. „False discovery rate control with e-values". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.3 (2022), p. 822–852.

L. Wasserman, A. Ramdas en S. Balakrishnan. „Universal inference". In: *Proceedings of the National Academy of Sciences* 117.29 (2020), p. 16880–16890.

Rianne de Heide is assistant professor in de wiskunde aan de Vrije Universteit Amsterdam (en vanaf september aan de Universiteit Twente). Ze werkt momenteel aan haar Veni-project 'E-waardes voor meervoudig hypothesetoetsen'. https://riannedeheide.github.io

Gerard Sierksma

Column

Jakob Emanuel Handmann, Public domain, via Wikimedia Commons

De media en de nieuwe e; “een zaak van leven en dood”

Deze keer een aantal overpeinzingen en persoonlijke associaties bij een onverwachte aandacht voor wiskunde-in-de-media. Die niet verwachte aandacht betrof de Volkskrant van zaterdag 20 april met twee artikelen over wiskunde en statistiek. En dat terwijl ik twee weken daarvoor in het tijdschrift New Scientist nog de verzuchting moest lezen dat die aandacht juist zo minimaal is, ook in eigen tijdschrift. Zwaartekrachttrillingen en zwarte gaten te over, maar een verhaal over wiskunde uit de pen krijgen is in New Scientist een zeldzaamheid. Nee, van de media moet de wiskunde het niet hebben. Maar toch, wie het kleine niet eert…

Prof.dr.ir. I. Smeets; foto: Ivar Pel

Ionica antwoordt met wiskunde

Wekelijks schrijft Ionica Smeets, hoogleraar *Science Communication* aan de Leidse universiteit, een leuke en interessante column in de zaterdageditie van de Volkskrant met de ondertitel “Ionica antwoordt met wiskunde”. Daarin beantwoordt ze een lezersvraag waarbij ze het antwoord lardeert met wat wiskunde en statistiek. In de 20-aprileditie is de beurt aan ene Marjon Kok. Marjon wil weten of het goed is voor haar portemonnee om haar zonnepanelenstroom direct te gebruiken voor de wasdroger. De toegevoegde wiskunde is deze keer een karig rekensommetje met als conclusie dat “op dit moment 1 kWh zo’n 25 cent kost, (...) en één droogbeurt dus ongeveer 36 cent aan stroom”. Maar of je nou de was aan de lijn moet hangen of in de droogtrommel moet stoppen blijft onduidelijk. Ionica raadt Marjon tenslotte aan te overwegen een man in huis te nemen die de was doet. Zelf beschikt ze over zo’n behulpzame meneer, die het echter moet doen zonder panelen. Goeie vraag dus van Marjon, met ook een grappig antwoord. Het echte antwoord is trouwens kort en simpel te geven Het zit namelijk zo. In de afgelopen jaren, het gaat natuurlijk veranderen, betaalde Marjon op de jaarrekening van haar energieleverancier enkel en alleen voor de hoeveelheid kW-uren die haar zonnepanelen te kort kwamen om haar van alle genuttigde elektriciteit te voorzien. Voor een mogelijk overschot aan zelfproductie kreeg ze een paar euro uitbetaald aan het einde van het jaar. Het maakte dus op jaarbasis niet uit op welk moment Marjon de was draaide en droogde in haar machine. Ionica’s column zou deze keer gepaster zijn ondertiteld “Ionica antwoordt deze keer met slechts een vleugje wiskunde”. Maar toch, wie het kleine...

Euler en de kleine e

Als inleiding op het andere Volkskrantverhaal eerst het volgende. Voor zover ik in mijn Google-zoektocht heb kunnen nagaan is Leonhard Euler (1707–1788) ooit uit de bus gekomen als winnaar in een Duits-Schotse discussie omtrent de oorsprong van het getal e, hoewel de Schot John Napier, de man van de logaritmen, voor velen sinds 1597 de ‘uitvinder’ is. Voor mij is Euler vooral de man van het 7-bruggenprobleem van Königsberg, de stad die sinds de bezetting door de Russen in 1945, Kaliningrad heet .... (U snapt mijn puntjes). Ruim daarvoor, in de 17^{e} eeuw, was Königsberg een gewone Oost-Pruisische Duitse stad en braken de Koningsbergers (wij Nederlanders vertalen plaatsnamen) hun hoofd over de zeven bruggen van hun rivier de Pregel. Eén van die problemen luidde: Kun je een wandeling door de stad maken, waarin alle zeven Pregelbruggen precies éénmaal worden gepasseerd tijdens die wandeling. Euler maakte, na zijn aanstelling als hoogleraar wiskunde aan de Königsberg-universiteit een eind aan dit hoofdpijndossier door te bewijzen dat zo’n rondgang niet mogelijk is. Elke tweedejaars student Operations Research weet dat zo’n wandeling een Euler tour heet en weet ook dat daar meer dan een vleugje wiskunde bij nodig is. Terug naar het Eulergetal e, dat minimaal iedere eerstejaars Econometrie kent. Hopelijk ook de -misschien wel mooiste- formule in de wiskunde, $e^{i\pi} + 1 = 0$, met daarin de Grote Drie π, i en e in één formule. Ook die 0 en die 1 vind ik er fraai in staan. Hoewel e in deze formule niet ‘in de exponent zit’, heeft Euler, volgens de overlevering, toch de letter e gekozen omdat e de eerste letter is van exponentieel in *Zahl exponentiell* (Euler was een Duitse Zwitser). Niet dus de eerste letter van zijn eigen naam.

Prof.dr. P.D. Grünwald; foto: Patrick Post

2,5 miljoen voor de nieuwe kleine e

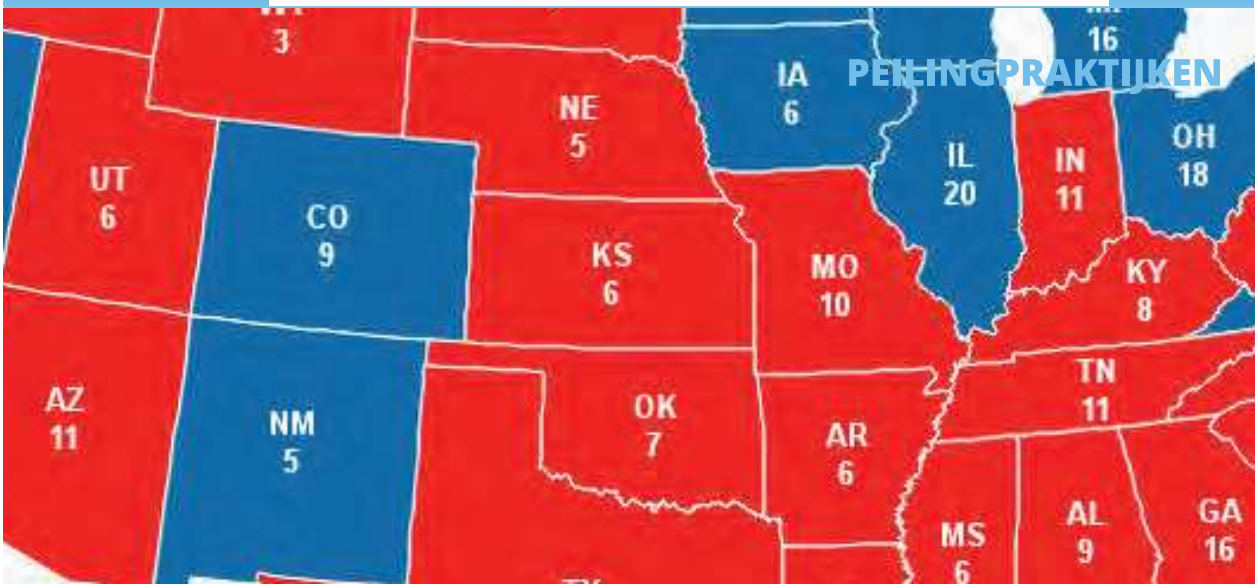
Dat andere verhaal in de Volkskrant beslaat daar twee volle pagina's, voor de helft gevuld met een afbeelding van een ontploffende e-barende letter p aan een blauw soort firmament. De artikelschrijver, Maarten Keulemans, heeft in zijn interview met de Leidse hoogleraar *Statistical Learning* P.D. Grünwald ontdekt dat "een van de hoofdpijndossiers van de wetenschap de p-waarde is" en dat de grote aanjager van het genezingsproces deze Grünwald is. Keulemans: "Wiskundige Peter Grünwald heeft een revolutionair idee en krijgt 2,5 miljoen euro voor onderzoek" naar een nieuwe p, die e dus. 2,5 miljoen euro is niet niks. Keulemans beschrijft hoe deze Leidse collega van Ionica Smeets tijdens het interview in zijn werkkamer zijn 'armen ten hemel werpt en verzucht: p is te hoekig, te bot en fraudegevoelig bovendien". Die kleine nieuwe e zou dan staan voor *evidence*, zoals de oude p stond en voorlopig staat voor *probabilty*, 'waarschijnlijkheid' zet Keulemans er tussen haakjes achter en voegt toe dat die p tegenwoordig meer voor de p van 'problemen' staat. Grünwald: Als met de p-waarde 'de werkelijkheid van het toeval wordt gescheiden', is het e-getal dan niet 'letterlijk een zaak van leven en dood'? Dit lijkt mij een hyperbool van jewelste en als u als lezer niet weet welke op-leven-en-doodstrijd bedoeld wordt, dan verandert de rest van deze column daar niets aan. Tot slot nog even Keulemans over Grünwalds uitleg van de nieuwe revolutionaire e: "Het berekenen van een p-waarde betreffende een vroegtijdig gestopt onderzoek is net zo fout als de stand in een afgebroken voetbalwedstrijd laten tellen als de einduitslag van die wedstrijd. Strikt genomen is e een maat, die wiskundig uitdrukt hoe de zekerheid van een bepaald verband zich opstapelt. Hoe hogere e, des te zekerder je weet: dit is geen toeval meer. (...) Enigszins vergelijkbaar met de voetbalwedstrijd, waarbij het publiek bij 3-0 in de 80ste minuut heus wel aanvoelt dat langer doorspelen weinig zin meer heeft." Einde citaat met 'hoe hogere e'. Ik hoop dat Grünwald geen tijd had de drukproeven te corrigeren omdat hij het te druk had met de besteding van zijn 2,5 miljoen.

De grote E

De letters e en E, klein en groot, genieten grote populariteit. Zeker ook de grote E. Zo hebben we Europese voedsel E-getallen op het etiket van onze jampotten en natuurlijk de grote E van Albert Einstein. Ook ik heb een T-shirt gedragen met de grote E van $E = mc^2$. Alleen in de sportschool trouwens: hoe meer massa, hoe meer energie. Bij mij is de waarde van c al sinds m'n geboorte aan de hoge kant en ben op tijd met het 'geboddiebilder' gestopt. De overlevering verhaalt niet over Einsteins gewetenskwestie betreffende de keuze tussen e en E, namelijk de afweging tussen enerzijds de verkwanseling van de uniciteit van Eulers kleine e en anderzijds de betichting van ijdelheid met de E van Einstein. Geen van beide volgens mij. Niemand in de wiskunde en aanpalend zou het in zijn hoofd moeten halen de hoofdletter E te gebruiken anders dan in $E = mc^2$. Heiligschennis, met terugwerkende kracht tot voorbij de oerknal, zou dat zijn. De Algemene Relativiteitstheorie en de Kwantummechanica zouden de Snaartheorie als de Grote Verzoener niet nodig hebben gehad en New Scientist te weinig kopij, zelfs als ze meer over wiskunde zouden schrijven. Euler heeft vrijwel alles omtrent de e ontdekt maar niet dat deze e 'de limiet is van het quotiënt van het aantal paden tussen twee punten in een volledige graaf en het aantal Hamilton paden tussen die twee punten in diezelfde graaf, waarbij het aantal knooppunten van die graaf in de naar oneindig gaat'. Pardon, ik vergeet erbij te zeggen dat een pad in een graaf nooit tweemaal hetzelfde knooppunt bezoekt. Ik vrees dat zelfs een stelling in deze vorm, zonder formules, de New Scientist niet zal halen, zeker niet als je ervan maakt: In een oneindige volledige graaf is het quotiënt van de grootte van de zoekruimte van kortste pad probleem en de zoekruimte van het Handelsreizigersprobleem gelijk aan 2,7182818... *ad infinitum*, inderdaad de kleine e.

Gerard Sierksma is emeritus hoogleraar operations research aan de Rijksuniversiteit Groningen.
E-mail: g.sierksma@rug.nl

Jelke Bethlehem



Bron: Esri UK (2016)

De Amerikaanse verkiezingen in kaart gebracht

Op 5 november 2024 zijn er weer presidentsverkiezingen in de Verenigde Staten (VS). De campagne voor die verkiezingen is al weer in volle hevigheid losgebarsten. En dat betekent dat er ook weer veel peilingen zijn. De uitkomsten van die peilingen, en straks ook de verkiezingsuitslag zelf, moet helder en correct in beeld worden gebracht voor het grote publiek. De media gebruiken daarvoor meestal thematische kaarten.

Een thematische kaart is een speciaal type grafiek waarmee je de verdeling van een variabele over een geografisch gebied weergeeft. Daarmee breng je een mogelijk verband tussen een variabele en de geografische locatie in beeld.

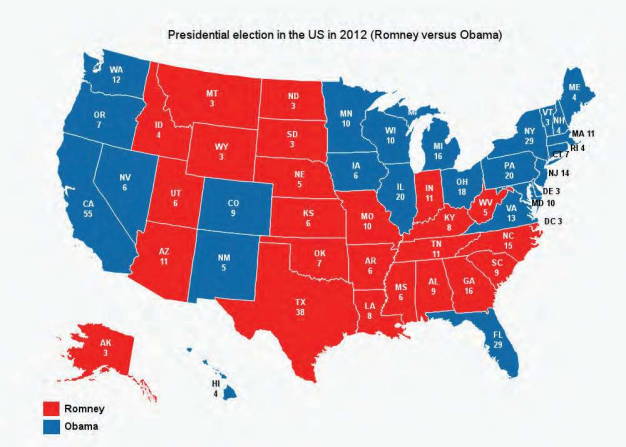
Thematische kaarten kunnen heel informatief zijn. Het in beeld brengen van de variabele moet je echter wel zorgvuldig doen, want het gevaar van verkeerde interpretatie ligt op de loer. Een specifiek probleem is de vertekening die wordt veroorzaakt door het feit dat niet alle regio's in een geografisch gebied dezelfde grootte (oppervlakte) hebben. In het Engels duiden we dit probleem aan met area bias.

Voor het weergeven van de resultaten van de Amerikaanse presidentsverkiezingen en de bijbehorende peilingen maken de media vaak gebruik van een speciale thematische kaart, en dat is een choropleet. Figuur 1 bevat een voorbeeld van zo'n kaart. Een choropleet is een kaart van een geografisch gebied dat is opgedeeld in een aantal regio's. De regio's zijn gekleurd, waarbij de kleur van elke regio de waarde van de te tonen variabele weergeeft.

De Amerikaanse presidentsverkiezingen van 2012

We keren terug naar de Amerikaanse presidentsverkiezingen van 2012. Barack Obama was de kandidaat van de Democraten en Mitt Romney was de kandidaat van de Republikeinen. Figuur 1 toont de choropleet met het resultaat van deze verkiezingen. Obama won in de blauwe staten en Romney in de rode staten. Merk op dat op het eerste gezicht de oppervlakte van de rode staten samen groter lijkt dan die van de blauwe staten. Op grond van deze kaart zou je dus kunnen concluderen dat Romney de verkiezingen heeft gewonnen.

De Amerikaanse president wordt gekozen door het Electoral College. Dit college bestaat uit 538 kiesmannen. De winnaar van de verkiezingen is de kandidaat met minstens 270 kiesmannen. Elke staat heeft een aantal kiesmannen. Het aantal per staat is gebaseerd op het aantal inwoners. Californië (CA) heeft de meeste kiesmannen (55). Verschillende klei-



Figuur 1: Choropleet met de uitslag van de verkiezingen in de VS in 2012. Bron: Esri UK (2016).

ne staten, zoals bijvoorbeeld Montana (MT), hebben slechts drie kiesmannen. De kaart in figuur 1 bevat voor elke staat de afkorting van de naam (twee letters) en het aantal kiesmannen. De kleur van een staat geeft aan welke partij de staat heeft gewonnen: blauw voor de Democraten en rood voor de Republikeinen.

In 2012 won Barack Obama de verkiezingen. Hij kreeg 332 kiesmannen. Mitt Romney kreeg slechts 206 kiesmannen. Dus Obama had een behoorlijke meerderheid van meer dan 100 kiesmannen. De uitslag van deze verkiezingen is weergegeven in figuur 1. Helaas is deze kaart misleidend. Hij leidt aan area bias. Dit probleem ontstaat doordat de geografische oppervlaktes van de staten niet overeenkomen met de waarden van de variabele (hier: de aantallen kiesmannen).

Aangezien de Democraten van Barack Obama de verkiezingen met een grote meerderheid hebben gewonnen, zou er veel meer blauw dan rood te zien moeten zijn op de kaart. Dit is echter niet het geval. Als je de aantallen blauwe en rode pixels op de kaart in figuur 1 telt, dan blijkt slechts 42% van de pixels blauw te zijn en maar liefst 58% rood. Bij de kiesmannen liggen de verhoudingen heel anders. Van de kiesmannen gaat 62% naar Obama en slechts 38% naar Romney. We kunnen dus concluderen dat de kaart een heel verkeerd beeld geeft van de uitslag van de verkiezing.

Als we deze choropleet nader bekijken, dan wordt duidelijk wat er mis is. We geven een voorbeeld. Kijk daarvoor naar de staat Montana (MT) in het noordwesten van de VS. De geografische omvang van Montana is behoorlijk groot, maar deze staat heeft toch slechts drie kiesmannen. De reden is dat er maar weinig mensen wonen. Het is een dunbevolkt gebied. Kijk vervolgens eens naar de staat Pennsylvania (PA) in het noordoosten van het land. De geografische omvang (oppervlakte) van deze staat is veel kleiner dan die van Montana. Toch heeft de staat 20 kiesmannen. Het is een dichtbevolkt gebied. Verder is het hele blauwe noordoosten ondervertegenwoordigd. Er zijn hier verschillende democratische staten met een grote bevolkingsdichtheid en een kleine geografische omvang. We moeten wel tot de conclusie komen dat de kaart te lijden heeft van area bias.

De kaart in figuur 1 is een voorbeeld van een choropleet. De naam is een combinatie van de Griekse woorden choros (gebied) en plethos (waarde). Dit is dus een kaart van een geografische gebied dat is verdeel in een aantal regio's. Elke regio heeft een kleur. Die kleur correspondeert met de waarde van de variabele die je in beeld brengt. Bij de choropleet in figuur 1 is het geografische gebied de Verenigde Staten en zijn de regio's de staten. De variabele die we in beeld brengen is een kwalitatieve variabele met twee categorieën: Democratisch (blauw) en Republikeins (rood).

Uit het voorgaande zal duidelijk zijn dat je moet oppassen met het gebruik van choropleten. Het gevaar van area bias ligt altijd op de loer. Gelukkig zijn er andere manieren om thematische kaarten te maken. Ze gaan alle uit van het principe dat je de verdeling van de variabele altijd correct moet weergeven. Dat kan echter ten koste gaan van de geografische nauwkeurigheid. Zulke kaarten noemen we ook wel cartogrammen. In het vervolg van deze paragraaf geven we enkele voorbeelden van cartogrammen.

Het Gastner-Newman-cartogram

Een eerste aanpak van area bias is het Gastner-Newman-cartogram. Dit cartogram staat in meer detail beschreven in Gastner & Newman (2004). Bij dit

Figuur 2: Gastner-Newman-cartogram van de uitkomst van de verkiezingen in de VS in 2012. Bron: Esri UK (2016).

cartogram past een computeralgoritme de grootte van de regio's aan (door uitrekken of inkrimpen) zodat hun oppervlaktes overeenkomen met de waarde van de variabele (hier: aantal kiesmannen). Daarbij zorgt het algoritme ervoor dat de grenzen tussen de regio's gehandhaafd blijven. Staten die aan elkaar grenzen, blijven dus aan elkaar grenzen.

Een Gastner-Newman cartogram vervormt de oorspronkelijke landkaart. Daarom ziet de kaart er soms wat vreemd uit en kan de interpretatie daardoor wat lastig zijn. Soms zijn de regio's moeilijk terug te vinden. Dat zou de reden kunnen zijn dat dit soort cartogrammen wat minder populair zijn in de media.

Figuur 2 bevat een voorbeeld van een Gastner-Newman cartogram. Dit cartogram toont ook de resultaten van de Amerikaanse presidentsverkiezing in 2012. Dezelfde gegevens zijn gebruikt als in figuur 1.

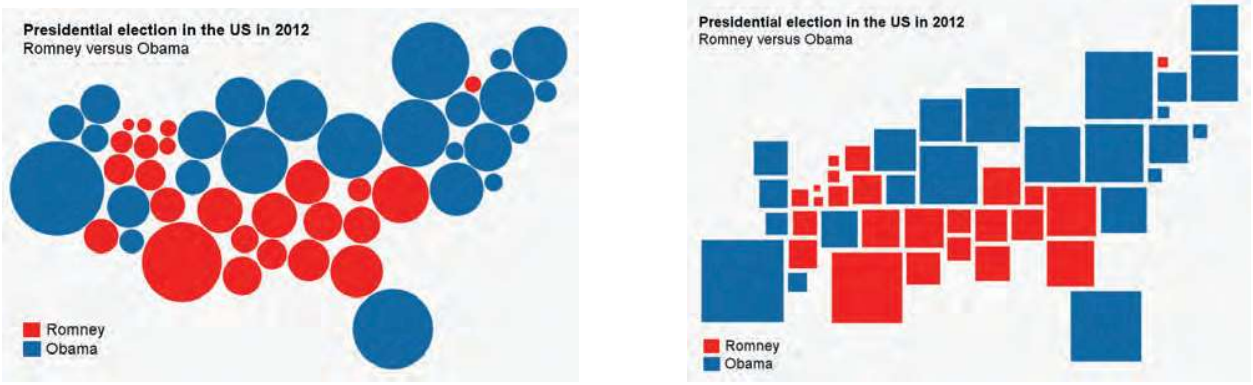
De geografie van de VS is min of meer in stand gebleven. Grote staten zoals Californië (CA), Michigan (MI), Florida (FL) en New York (NY) kun je vrij makkelijk terugvinden. Merk op dat de oppervlakte van Montana (MT) met (slechts) drie kiesmannen nu aanzienlijk kleiner is dan de oppervlakte van Pennsylvania (PA) met 20 kiesmannen.

Het Dorling-cartogram en het Demer-cartogram

Figuur 3 bevat twee andere typen cartogrammen. Het linker cartogram is voorgesteld door Dorling (1996) en noemen we daarom een *Dorling-cartogram*. Er is geen vervorming van de regio's door inkrimpen of uitrekken. De regio's blijven zoals ze waren. In plaats daarvan vervangen we elke regio door een symbool. Dat is in dit geval een cirkel. De oppervlaktes van deze cirkel maken we evenredig aan de waarde van de variabele (hier: het aantal kiesmannen).

Je wilt voorkomen dat cirkels over elkaar heen vallen en daarom minder goed zichtbaar zijn. Daarom verschuift een computeralgoritme waar nodig de cirkels. Daardoor kunnen de geografische posities van de regio's enigszins veranderen. Dat kan tot gevolg hebben dat je de regio's (staten) wat minder makkelijk kunt terugvinden.

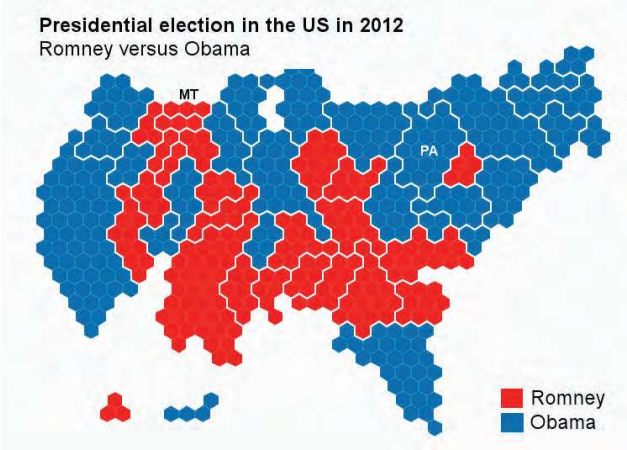
Het rechter cartogram in figuur 3 noemen we een Demer-cartogram (naar de bedenker Steve Demer). Je kunt dit cartogram zien als een Dorling-cartogram



(a) Dorling-cartogram

(b) Demer-cartogram

Figuur 3: Een Dorling-cartogram en een Demer-cartogram van de verkiezingsresultaten in de VS in 2012. Bron: Esri UK (2016).



Figuur 4: Een hexagoon-cartogram van de resultaten van de presidentsverkiezingen in de VS in 2012. Bron: Esri UK (2016).

waarin de cirkels zijn vervangen door vierkanten. Als je vierkanten gebruikt kun je efficiënter omgaan met de lege ruimte tussen de regio's. De vierkanten liggen wat dichter bij elkaar dan de cirkels.

De algoritmen voor een Dorling-cartogram en een Demer-cartogram werken op een verschillende manier. Het Dorling-cartogram probeert de afstand tussen de oude en de nieuwe positie van de regio's zo klein mogelijk te houden. Het Demer-cartogram probeert ervoor te zorgen dat de aangrenzende regio's zoveel mogelijk aangrenzend blijven. Regio's kunnen verschuiven om te voorkomen dat ze overlappen.

Het hexagoon-cartogram

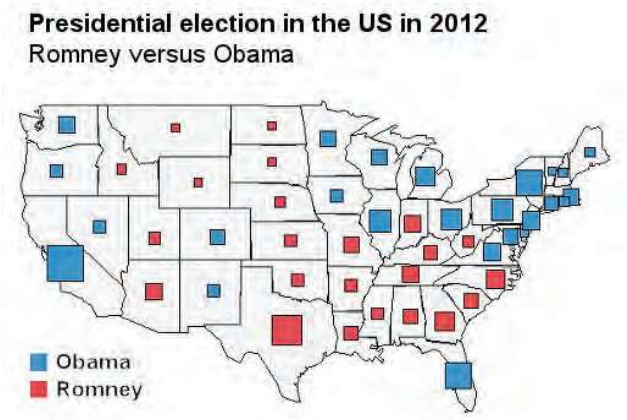
Een veel gebruikt type cartogram is het hexagoon-cartogram. Iedere regio (staat) is opgebouwd uit een aantal hexagonen (zeshoeken). Het aantal komt overeen met de waarde van de variabele (aantal kiesmannen). Figuur 4 bevat een voorbeeld van een hexagoon-cartogram. De staat Montana (MT) in het noordwesten bestaat bijvoorbeeld uit drie hexagonen omdat er drie kiesmannen zijn. En Pennsylvania heeft 20 hexagonen, wat overeen komt met 20 kiesmannen.

Volgens sommige deskundigen ogen patronen van hexagonen prettiger. Daarom trekt dit type grafiek al gauw positieve aandacht. En het is ook vrij eenvoudig om patronen van hexagonen te vormen. Een groep hexagonen is nogal flexibel. Daardoor zijn er meer mogelijkheden om zo'n groep te laten lijken op de werkelijke vorm van de regio. Dus kaarten gebaseerd op deze zeshoeken zijn herkenbaarder dan kaarten gebaseerd op cirkels of vierkanten.

De figuratieve kaart en de symbolenkaart

Er zijn nog twee typen grafische kaarten die de moeite van het vermelden waard zijn. De eerste van de twee noemen we een figuratieve kaart of een symbolenkaart. In het Engels heet dit type grafiek een *symbol map*.

Bij een figuratieve kaart blijft de geografische kaart onaangetast. We geven de waarde in een regio



Figuur 5: Een figuratieve kaart van de resultaten van de presidentsverkiezingen in de VS in 2012. Bron: Esri UK (2016).

Figuur 6: Een stippenkaart van de verdeling van de bevolking in de VS. Bron: publiek domein.

aan door middel van een symbool van een bepaalde grootte. Die grootte komt overeen met de waarde van de variabele.

Figuur 5 toont een figuratieve kaart met de uitslagen van de Amerikaanse residentsverkiezingen in 2012. De grootte van de symbolen (de vierkanten) is evenredig met het aantal kiesmannen. De kaart geeft een goed beeld van de verkiezingsuitslagen. Deze kaart laat ook een potentieel probleem zien. Er zijn een staten in het noordwesten met veel inwoners en een kleine oppervlakte. Daardoor is er amper ruimte voor al die grote vierkanten. Ze dreigen te overlappen en kunnen daardoor tot verkeerde interpretatie leiden.

Merk op dat we een vierkant hebben gebruikt in de figuratieve kaart in figuur 5. Je zou ook een ander symbool kunnen gebruiken. Een voor de hand liggende keuze zou een cirkel kunnen zijn.

Tot slot van deze paragraaf noemen we nog de stippenkaart (Eng: dot map of point map). Ook bij een stippenkaart zetten we symbolen in de regio's op de kaart. Alleen varieert de grootte van de symbolen niet. In plaats daarvan geven we de waarde in de regio aan door het symbool een aantal keren te herhalen. Meestal gebruiken we als symbool een (dikke) stip. Het aantal punten per regio nemen we dan evenredig aan de bijbehorende waarde van de variabele.

Er zijn twee versies van de stippenkaart. De simpelste versie is de één-op-één versie. Daarin correspondeert één stip met één gebeurtenis. Een voorbeeld zou een stippenkaart kunnen zijn waarin elke stip correspondeert met een fataal verkeersongeluk. Je krijgt dan een goed beeld van de verspreiding van de fatale verkeersongelukken over het hele land. Er is ook een één-op-veel versie van de stippenkaart. Daarin correspondeert elke stip met een aantal objecten.

Figuur 6 toont een één-op-veel versie van de stippenkaart. Op de kaart is de bevolking van de VS in beeld gebracht. Elke stip staat voor 50.000 mensen. Je kunt deze kaart goed gebruiken voor het zoeken naar gebieden met een grote of een juist kleine bevolkingsdichtheid. Verschillende gebieden vallen op, zoals de dichtbevolkte staten in Californië en het noordoosten en de dunbevolkte staten zoals Montana. Stippenkaarten zijn wat minder geschikt voor het aflezen van de exacte waarden van de variabele in de regio's.

Conclusie

De conclusie van dit verhaal is dat we voorzichtig moeten zijn met het gebruik van thematische kaarten, en in het bijzonder choropleten. Het gevaar bestaat dat je als gevolg van area bias een verkeerd beeld krijgt van de grafiek. Meer over thematische kaarten kun je bijvoorbeeld vinden in Bethlehem (2022, hoofdstuk 11).

Literatuur

Bethlehem, J.G. (2022), *Het Grafiekenboek*. Amsterdam University Press, Amsterdam. https://www.aup.nl/en/book/9789463720984/het-grafiekenboek

Dorling, D. (1996), Area Cartograms: Their Use and Creation. *Concepts and Techniques in Modern Geography* 59.

Esri (2016), *US Election 2016: Battle of the Maps*. Esri UK, CommunityHub, https://communityhub.esriuk.com/geoxchange/2016/11/1/us-election-2016-battle-of-the-maps.

Gastner, M.T. & Newman, M.E.J. (2004), *Diffusion-based method for producing density-equalizing maps*. Proceedings of the National Academy of Sciences of the United States of America, 101 (20), blz. 7499-7504. https://doi.org/10.1073/pnas.0400280101.

Jelke Bethlehem is expert op het gebied van steekproeven, vragenlijsten en weergave van onderzoeksresultaten.
E-mail: mail@jelkebethlehem.nl

Foto: Shanghai interchange, Wikimedia Commons

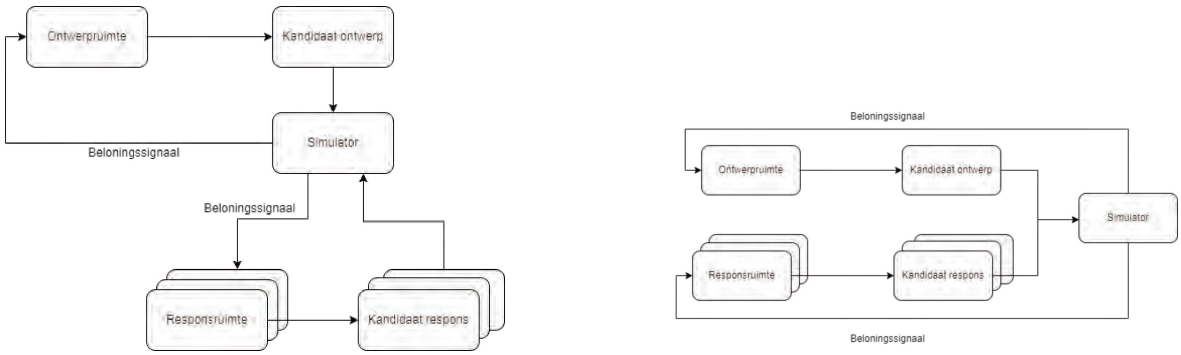
Gelijktijdige optimalisatie op twee niveaus voor netwerkontwerpproblemen

Maaike Snelder, Selmar Smit, Maarten Schadd, Erwin Walraven

Netwerkontwerpproblemen, bijvoorbeeld de optimalisatie van wegcapaciteiten, worden vaak bestudeerd met behulp van zogenaamde bi-level optimalisatie. Dat wil zeggen twee optimalisaties op verschillende niveaus: de ontwerp en de response van gebruikers. Vanwege de enorme rekenlast worden de gekozen ontwerp- en responsruimten typisch beperkt, wat doorgaans resulteert in suboptimale ontwerpen. In dit artikel introduceren we een aanpak voor gelijktijdige optimalisatie van ontwerpen en responsen. Experimenten voor de stad Delft waarin parkeerlocaties en tarieven worden geoptimaliseerd, laten zien dat gelijktijdige ontwerp- en responsoptimalisatie mogelijk is. De voorgestelde aanpak is in staat om automatisch responsen te leren, wat leidt tot ontwerpen die tot aanzienlijk beter zijn dan ontwerpen gemaakt met vantevoren gekozen responsruimte.

Op het gebied van verkeer en vervoer doen zich veel netwerkontwerpproblemen voor die gepaard gaan met grote investeringen, zoals het verhogen van wegcapaciteit. Het is van cruciaal belang om te anticiperen op hoe gebruikers zullen reageren op aanpassingen, hun zogenaamde responsen. Er zijn veel voorbeelden van ontwerpen die achteraf niet optimaal bleken, omdat gebruikers de ontwerpen op onverwachte en soms ongewenste manieren gebruikten. Olifantenpaadjes (Luckert, 2012) zijn daarbij een typisch verschijnsel, zie Figuur 1 voor een voorbeeld.

Figuur 1: Een olifantenpad voor auto's. Bron: Luckert, 2012.



(a) Sequentiele bi-level optimalisatie.

(b) Simultane bi-level optimalisatie.

Figuur 2: Een vergelijking van de klassieke aanpak (a) en de nieuwe aanpak (b).

De kwaliteit van een ontwerp wordt bepaald door hoe de gebruikers er gebruik van maken, aangezien zij hun eigen doelstellingen optimaliseren afhankelijk van het ontwerp. Mensen zijn erg creatief in het vinden van een eigen oplossing, hun eigen 'olifantenpad'. Daarom is het belangrijk om niet alleen de huidige gedragspatronen te gebruiken in een optimalisatie.

In dit artikel introduceren we om die reden een geïntegreerde aanpak voor het ontwerpen van een netwerk, waarbij rekening gehouden wordt met het toekomstige gedragspatroon van de gebruikers. Het voert een gelijktijdige optimalisatie op twee niveaus uit door ontwerpen en responsen tegelijkertijd te optimaliseren in plaats van sequentieel. Dit wordt mogelijk gemaakt door machine learning in te zetten.

Kader voor ontwerp- en responsoptimalisatie

In Figuur 2a wordt de traditionele manier van sequentiële optimalisatie weergegeven. Eerst wordt de ontwerpruimte gemaakt. Vanuit deze grote ruimte wordt een kandidaat ontwerp gekozen met behulp van een machine learning-algoritme naar voorkeur. Een simulator beoordeelt het ontwerp op bepaalde criteria en geeft een beloningssignaal aan het leeralgoritme.

Een tweede leerlus optimaliseert de respons van agenten op het ontwerp. Deze agenten kunnen, afhankelijk van de situatie, reizigers, rivaliserende bedrijven of een tegenstander zijn. De agenten optimaliseren hun eigen olifantenpad, welk meer of minder in lijn kan zijn met het ontwerp. Deze leerlus is noodzakelijk om adequate responsen te leren en het ontwerp daarop aan te passen. Door responsoptimalisatie te gebruiken, is het gevonden ontwerp robuust tegen olifantenpaden.

Er kleeft echter één nadeel aan de aanpak van Figuur 2a, namelijk de benodigde rekenkracht. Als voor elk getest ontwerp een tweede algoritme moet worden uitgevoerd voor de respons, kan het veel te lang duren om tot een oplossing te komen.

De nieuwe aanpak doet een gelijktijdige optimalisatie op beide niveaus, zodat kostbare rekentijd gespaard wordt (zie Figuur 2b). De simulator geeft een beloningssignaal voor zowel het ontwerp (KPI's) als de verschillende responsen (hoe vaak een respons gekozen is; personen willen op tijd komen en geld besparen). De responsen worden in een bibliotheek opgeslagen. Zo kunnen respons-ideeën van het ene ontwerp overgedragen worden naar het volgende ontwerp. De meest gekozen responsen houden een iteratie stand en de minst gekozen worden uit de bibliotheek verwijderd om ruimte te maken voor nieuwe respons-ideeën. Zo leert de responsbibliotheek zelfstandig welke responsen goed werken. Om te voorkomen dat goed presterende responsen worden verwijderd vanwege gebruik op slechts enkele ontwerpen, vervallen deze geleidelijk.

We demonstreren dit aan de hand van het voorbeeld in Figuur 3. Er worden meerdere ontwerpen getest (bovenaan). Voor elk ontwerp wordt de responsbibliotheek bijgewerkt (onderaan). De slechtst presterende respons wordt bij elke iteratie verwijderd. Hoe vaak een respons gekozen is, wordt weergegeven door de gekleurde balk. In Figuur 3 geven lange balken een goede prestatie weer. Deze zijn daarom groen gekleurd.

Drie voordelen van de geïntegreerde aanpak zijn:

1) de benodigde rekentijd kan lager zijn dan de sequentiële benadering;

Figuur 3: Een voorbeeld van hoe de responsbibliotheek zich aanpast aan de ontwerpen.

2) zodra een bibliotheek met goede responsen is gevuld, kan een nieuw ontwerp sneller nauwkeurig beoordeeld worden;
3) de responsbibliotheek zal zichzelf aanpassen aan de kenmerken van algemeen goed presterende responsen.

Een aantal nadelen kunnen zijn:

1) ontwerpen moeten opnieuw getest worden wanneer de bibliotheek met responsen sterk veranderd is;
2) de KPI's (Key Performance Indicators) van het ontwerp zijn moeilijk te vergelijken, omdat deze zijn gebaseerd op de inhoud van de responsbibliotheek op dat moment. Een ontwerp met goede KPI-waarden kan later slecht blijken, wanneer een andere respons aan de bibliotheek toegevoegd wordt;
3) ontwerp-specifieke responsen worden mogelijk niet gevonden.

Ontwerp- en responsoptimalisatie voor stadsplanning

Een gekozen casus betreft de plaatsing van extra parkeerplaatsen in de stad Delft en daarbij het selecteren van de bijbehorende tarieven voor parkeren. Een goed ontwerp is van groot belang om ervoor te zorgen dat de inwoners effectief gebruik kunnen maken van de parkeerfaciliteiten, zonder dat er ongewenste neveneffecten optreden. Wij illustreren dit met een futuristisch voorbeeld. Momenteel parkeren mensen hun auto doorgaans dicht bij hun bestemming. Met de introductie van volledig geautomatiseerde voertuigen (SAE-niveau 5, zie bijvoorbeeld SAE, 2021) zouden voertuigen autonoom ver weg van het stadscentrum kunnen parkeren. Aangezien niveau 5-voertuigen niet te verwachten zijn in de nabije toekomst (Shladover, 2016) zijn deze experimenten slechts bedoeld om de toegevoegde waarde van de optimalisatie-aanpak te illustreren.

Een ontwerp bestaat uit de locaties van de parkeerterreinen, en de tarieven voor parkeren, op deze terreinen of op straat. Voor parkeerterreinen en parkeren langs de straat worden verschillende tarieven gehanteerd. Er wordt echter geen onderscheid gemaakt tussen locaties. De tarieven zijn dus gelijk in de hele stad. In de praktijk kunnen de mogelijkheden en daarmee de ontwerpruimte vergroot worden door per parkeerplaats, of zone, een eigen tarief te hanteren, waardoor gebruikers wellicht massaal naar de goedkoopste parkeerplaats rijden. De respons bestaat dus hier uit reisgedrag, namelijk een actiereeks beginnend bij de oorsprong van een reis. Om precies te zijn, onderscheiden we vijf verschillende acties:

- Rijden naar de bestemming (bijvoorbeeld werk).
- Rijden naar de dichtstbijzijnde parkeerplaats. Deze actie heeft een andere uitkomst op basis van de huidige locatie van de reiziger.
- Rijden naar de herkomst (thuis).
- Eén uur parkeren. Als een parkeerplaats op dezelfde locatie aanwezig is, wordt deze gebruikt. Zo niet, dan wordt er op straat geparkeerd.
- Parkeren tot het einde van de dag.

Dezelfde actiereeks kan voor verschillende reizigers ander gedrag opleveren wanneer het begin- en eindpunt van de reis verschillen. Actiereeksen kunnen gezien worden als reispatronen en zijn een flexibel mechanisme om responsen te definiëren. Ook zijn uitbreidingen door het toevoegen van een nieuwe actie eenvoudig (bijvoorbeeld het overschakelen naar andere vervoermiddelen).
De KPI's zijn de volgende vier:

- Het aantal gereden kilometers door een zelfrijdende auto zonder passagier dient zo laag mogelijk te zijn.

- Het aantal uren dat auto's op parkeerplaatsen staan dient zo hoog mogelijk te zijn.
- Het aantal kilometers dat auto's rijden dient zo laag mogelijk te zijn.
- De parkeeropbrengsten dienen zo hoog mogelijk te zijn.

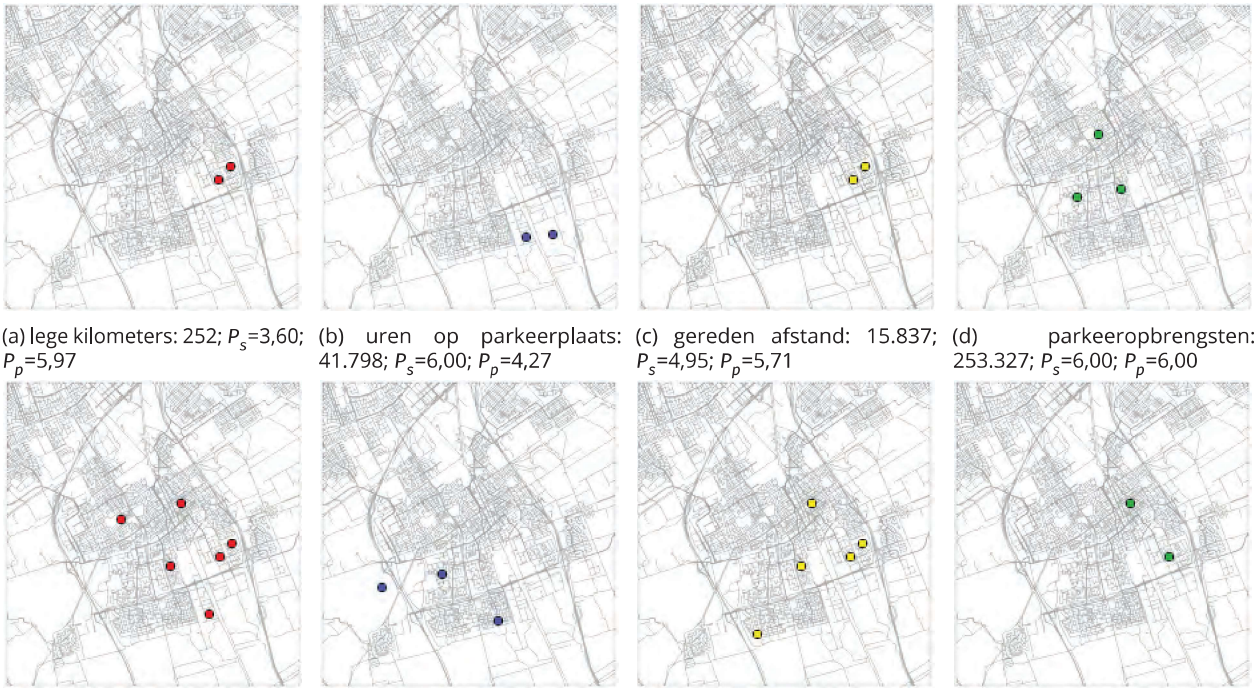
De simulator gebruikt actiereeksen van de responsbibliotheek om voor elke persoon, een zogenaamde agent, gedrag te simuleren over de dag. Na een volledige simulatiedag, worden de prestaties van de KPI's berekend. Het verkeersmodel van Delft, bestaande uit een wegennetwerk (491 knooppunten, 695 verbindingen, 25 zones) en een herkomst-bestemmingsmatrix die de reizen definieert, in dit geval een ochtendspits wordt hierbij gebruikt. De simulator kan bij drukte files laten ontstaan. Elke reiziger kiest een respons die zijn kosten minimaliseert.

Voor het optimaliseren van het ontwerp gebruiken we een genetisch algoritme (Mitchell, 1998). Het genoom bestaat uit een variabel aantal parkeerlocaties en de tarieven voor parkeerterreinen en voor parkeren op straat. Wanneer een nieuw willekeurig ontwerp wordt gemaakt, is een willekeurig aantal parkeerplaatsen willekeurig geplaatst in de stad, en de parkeertarieven worden op uniforme willekeurige wijze gekozen (3–6 euro per stuk). De initiële populatie van ontwerpen wordt gecreëerd door 100 van dergelijke willekeurige ontwerpen te maken.

Voor het optimaliseren van de responsbibliotheek wordt gekeken naar hoe vaak elke respons gekozen is. De responsbibliotheek bestaat aanvankelijk uit 100 actiereeksen. De betere helft van alle actiereeksen doorstaat een iteratie (waarbij hun score langzaam vervalt om een goede respons niet direct te moeten verwijderen). De vrijgekomen ruimte wordt opgevuld met nieuwe actiereeksen, welke mutaties zijn van de betere helft.

Resultaten

Het doel van het getoonde experiment is om te laten zien dat er slechtere ontwerpen gekozen worden wanneer er niet direct rekening gehouden wordt met de respons van de reizigers. Hiervoor wordt eerst de responsbibliotheek beperkt tot een klein aantal actiereeksen. Ten tweede is er een grote, maar statische responsset gemaakt, om zo te zien of het ontwerp bestand is tegen de creativiteit van de reizigers. En ten derde wordt de voorgestelde simultane bi-level optimalisatie toegepast. De resultaten voor de verschillende KPI's, waarbij elke keer



(a) lege kilometers: 252; P_s=3,60; P_p=5,97
(b) uren op parkeerplaats: 41.798; P_s=6,00; P_p=4,27
(c) gereden afstand: 15.837; P_s=4,95; P_p=5,71
(d) parkeeropbrengsten: 253.327; P_s=6,00; P_p=6,00
(e) lege kilometers: 27.489; P_s=3,06; P_p=6,00
(f) uren op parkeerplaats: 35.972; P_s=5,74; P_p=4,71
(g) gereden afstand: 52.835; P_s=6,00; P_p=6,00
(h) parkeeropbrengsten: 215.829; P_s=6,00; P_p=6,00

Figuur 4: Resultaten voor het optimaliseren van een enkele KPI op een statische responsset (boven) en simultane bi-level optimalisatie (beneden). P_s=parkeertarief op straat, P_p=parkeertarief op parkeerterrein. Elke kleur correspondeert met een andere KPI.

Tabel 1: Vergelijking van de nieuwe aanpak met statische responsen. KPI's zijn van het beste ontwerp (per KPI) voor de stad Delft uit 5 optimalisatie runs. Door de resultaten onder de kleine met de grote statische respons te vergelijken wordt duidelijk dat de impact van de responsbibliotheek groot kan zijn. De effect kolom toont het verschil van de simultane bi-level optimalisatie vergeleken met de grote statische responsset. Afhankelijk van de KPI is hoger of lager beter.

KPI	Kleine statische respons	Grote statische respons	Simultane bi-level optimalisatie	Effect
lege kilometers	252	31.064	27.489	-12%
uren op parkeerplaats	41.799	32.423	35.972	+11%
gereden afstand	15.837	61.167	52.835	-14%
parkeeropbrengsten	253.327	215.829	215.829	+0%

een ontwerp gezocht wordt dat een enkele KPI als focus heeft, zijn te zien in Tabel 1, en de bijbehorende ontwerpen zijn te vinden in Figuur 4. Het is belangrijk te weten dat voor de uiteindelijke KPI's van de bi-level optimalisatie dezelfde grote statische responsset gebruikt is, om een eerlijke vergelijking te maken.

Uit Tabel 1 blijkt dat er inderdaad slechtere ontwerpen gevonden worden, wanneer er onvoldoende rekening gehouden wordt met de respons. Uit de tabel blijkt bovendien dat bij de bi-level optimalisatie respons- en actiereeksen gevonden worden die lijken op de grote statische responsset, aangezien ze vergelijkbare KPI-waarden laten zien. Bij het onderzoeken van de actiereeksen blijken voertuigen vaak heen en weer te rijden tussen parkeerplaatsen, omdat dit goedkoper blijkt dan parkeren. Voor drie van de vier KPI's vindt de bi-level optimalisatie betere ontwerpen dan met de statische responsset; eenmaal wordt dezelfde oplossing gevonden.

Conclusie

Dit artikel presenteert een nieuwe aanpak voor de optimalisatie van ontwerp en respons. We tonen aan dat het mogelijk is om beide niveaus tegelijkertijd ofwel simultaan te optimaliseren. Omdat er ook rekening wordt gehouden met leergedrag in responsen biedt het potentieel betere, realistischere en robuustere ontwerpen voor grootschalige realistische netwerkontwerpproblemen.

Literatuur

E. Luckert. „Drawings We Have Lived: Mapping Desire Lines in Edmonton". In: *Constellations* 4.1 (2012), p. 1–10.

M. Mitchell. *An introduction to genetic algorithms*. MIT Press, 1998.

SAE. „Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles". In: J3016_202104 (2021).

S. E. Shladover. „The truth about "self-driving" cars". In: *Scientific American* 314.6 (2016), p. 52–57.

Maaike Snelder werkt als principal scientist bij TNO en universitair hoofddocent bij TU Delft aan het ontwerp van multimodale transportsystemen. Maaike heeft een achtergrond in Econometrie en is cum-laude gepromoveerd op het ontwerpen van robuuste wegennetwerken.
E-mail: maaike.snelder@tno.nl

Selmar Smit is manager science en technology bij TNO en coordineert de ontwikkeling van een groot Nederlands taalmodel; GPT-NL. Hij heeft een MSc en PhD in kunstmatige intelligentie, en is bij het brede publiek bekend door zijn rol binnen het televisieprogramma Hunted.
E-mail: selmar.smit@tno.nl

Maarten Schadd werkt als onderzoeker bij TNO met als zwaartepunt de kunstmatige intelligentie voor beslissingsondersteunende systemen. Hij heeft een MSc en een PhD in Kunstmatige Intelligentie en is een expert in Monte-Carlo-technieken en optimalisatie-algoritmen. Hij heeft ervaring opgedaan met het ontwikkelen van complexe interacterende systemen in de private sector.
E-mail: maarten.schadd@tno.nl

Erwin Walraven werkt als onderzoeker bij TNO aan beslissingsondersteunende systemen voor het ontwerpen van steden. Hierbij focust hij zich op de toepassing van kunstmatige intelligentie en optimalisatiemethodes in samenwerking met partijen zoals gemeentes en ministeries. Erwin heeft een achtergrond in de informatica en is gepromoveerd op het gebied van geautomatiseerd plannen voor autonome agents.
E-mail: erwin.walraven@tno.nl

Winner Van Zwet Award

Casper Albers

During the Annual Meeting on 21 March 2024, the Van Zwet Award for best PhD thesis has been handed out. The 2024 recipient of this award is Dr. Richard Post from Eindhoven University of Technology. Dr. Wouter Kool from the University of Amsterdam received an honourable mention.

Winner: Richard Post

The winner of this year's Willem R. van Zwet award is Richard Post, with his thesis "Causal Effect Heterogeneity: statistical formalization and analysis of the individual causal effect", supervised by Edwin van den Heuvel and Hein Putter. Richard has chosen a rather daring subject for his thesis. The subject of individual causality, or causal effect heterogeneity. The subject is controversial, since individual causal effects seem to require individual counterfactuals, which by definition are never observed. However, Richard describes models and assumptions under which it is possible to see that causal effects are not the same for different individuals, and develops methods to chart the resulting heterogeneity. This causal heterogeneity is relevant in cross-sectional but also in longitudinal outcomes and survival analysis, which he explores.
The jury was impressed by the broad scope encompassing both theory and applications. The exposition is very clear. The jury found Richard's work to be of the highest scientific quality. It has mathematical rigor as well as intuitive explanations, strong modelling, it is conceptually highly innovative, and illustrates all this with relevant applications. Moreover, and we find this very important, the thesis is also very well written. We congratulate Richard on a well-deserved award.

Honourable mention: Wouter Kool

The thesis "Learning and Optimization in Combinatorial Spaces with a Focus on Deep Learning for Vehicle Routing" by Wouter Kool, and supervised by Prof. dr. M. Welling, dr. H.C. van Hoof and prof. dr. J.A.S. Gromicho. This is a exiting work at the intersection of statistics, machine learning and operations research, and the jury was excited to see these fields brought together in a very creative way. In the first part of the thesis, Wouter Kool develops new methods for vehicle routing using reinforcement learning. In the second part, he develops novel statistical techniques for sampling without replacement from the search space, in order to make the reinforcement learning more efficient. The work has already had a large impact, with the chapter "Attention! Learn to solve routing problems", published 2019, already being cited over 1200 times. Let us congratulate Wouter with his thesis.

Hoe Operations Research kan helpen bij complexe eerlijkheidsvraagstukken

Thomas Breugem

Eerlijkheid speelt een belangrijke rol in veel beslisproblemen, zoals het toewijzen van humanitaire hulp aan mensen in nood of taken aan personeel. Efficiëntie criteria (bijvoorbeeld gemeten aan de hand van operationele kosten of hoeveelheid verstrekte hulp) zijn zelden in lijn met eerlijkheidsoverwegingen. Dit levert lastige afwegingen op. Operations Research kan een sleutelrol spelen in het in kaart brengen van deze afwegingen.

Figuur 1: Een vrijwilliger van Marie Stopes International geeft voorlichting. Afbeelding van: https://icfp2022.org/a-moment-to-elevate-young-womens-voices-and-family-planning-choices/.

Eerlijkheid en Operations Research

Eerlijkheidsvraagstukken komen voor bij praktisch elke toewijzing van beperkte of (on)wenselijke goederen, zoals taken, geld, of fysieke goederen. Het is dan ook geen verrassing dat eerlijkheid in optimalisatieproblemen aanzienlijke aandacht heeft gekregen in de literatuur, vooral waar eerlijkheid 'logisch' is. Denk hierbij aan volksgezondheid, niertransplantatieprogramma's, humanitaire hulpverlening, maar ook toepassingen in planning en logistiek, zoals luchtverkeersstroombeheer, luchtvaartschema's, roosters voor verpleegkundigen, en personeelsroosters in het openbaar vervoer.

Efficiëntie criteria zijn zelden in lijn met eerlijkheidsoverwegingen. Al de bovengenoemde beslisproblemen, bijvoorbeeld, bevatten een fundamentele afweging tussen eerlijkheid en efficiëntie. Operations Research (OR) kan een sleutelrol spelen in begrijpen van zulke afwegingen, omdat het de geavanceerde analytische en computationele methoden biedt die nodig zijn om de impact van eerlijkheidsoverwegingen in kaart te brengen. In dit artikel bespreek ik hier twee voorbeelden van, en sluit ik af met een discussie omtrent kansen voor impactvol toekomstig onderzoek.

Gezinsplanningdiensten in Afrika

Het eerste project (Breugem en Van Wassenhove, 2022) richt zich op een grote non-profitorganisatie

Figuur 2: Voorbeeld van rapportage van Marie Stopes International, met duidelijke focus op het helpen van de meest kwetsbaren. Figuur uit het rapport: https://www.msichoices.org/wp-content/uploads/2023/10/msi-2021-impact.pdf.

die gezinsplanningsdiensten aanbiedt in Afrika. Denk hierbij aan het verstrekken van anticonceptiva en informatie om individuen te helpen bij het behalen van hun gewenste gezinsgrootte en tijdstip van geboorten. Essentieel voor deze diensten zijn (medische) teams die naar lastig te bereiken gebieden reizen om de diensten te verlenen (Figuur 1).

Gezinsplanningsorganisaties zijn zich bewust van het belang van hun diensten voor kwetsbare gemeenschappen en proberen deze daarom voorrang te geven. Zo rapporteert Marie Stopes International, een gezinsplanningsorganisatie actief in 37 landen, het exacte percentage van hun cliënten afkomstig uit kwetsbare gemeenschappen (zie 2).

Percentages, zoals gerapporteerd door Marie Stopes International, weerspiegelen eerlijkheidsoverwegingen waarin een bepaalde (kwetsbare) groep voorrang krijgt. Echter, sturen op dit soort percentages kan 'kosten' met zich meebrengen in de vorm van verminderde cliëntenaantallen, door bijvoorbeeld logistieke uitdagingen en sociale stigma's die het bereiken van kwetsbare groepen lastiger maakt. Desondanks verwachten donoren vaak dat een organisatie het aantal cliënten vanuit kwetsbare groepen *en* het totale aantal cliënten vergroot. Het analyseren van deze 'onmogelijke opgave' is het doel van ons onderzoek.

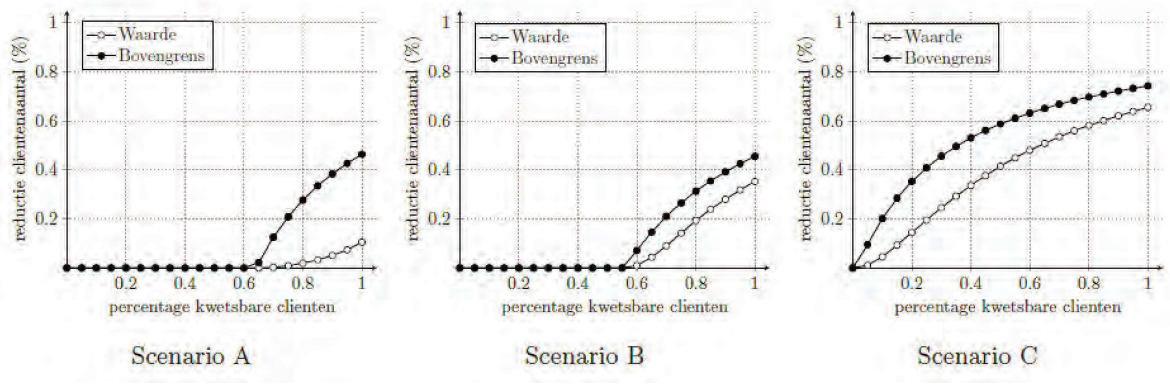
Op abstract niveau is het sturen op percentages wat we in het paper 'uitkomstbeperkingen' noemen: een bepaald percentage van de uitkomst (cliëntenaantallen) moet van een bepaald type (kwetsbare groepen) zijn. In deze specifieke context is er beperkte data en weinig ruimte voor geavanceerde optimalisatietechnieken. Daarom analyseren we de potentiële afwegingen op basis van slechts enkele belangrijke kenmerken. Voor ons algemene model bepalen we vervolgens analytische bovengrenzen op het verlies van efficiëntie door het opleggen van uitkomstbeperkingen. We baseren onze aanpak op algemene technieken uit de convexe analyse, in het bijzonder de ondersteunende hypervlakstelling.

We passen onze methodologie toe op verschillende cases (Figuur 3). Onze analyse laat zien dat uitkomstbeperkingen kunnen leiden tot een aanzienlijke daling in het aantal cliënten, soms bijna 50%. Onze resultaten bevestigen dat uitkomstbeperkingen, en dus prestatiedoelen voor organisaties, zorgvuldig moeten worden vastgesteld. Onze resultaten tonen ook aan dat de begrenzing een nuttige schatting biedt van potentieel cliëntverlies, zij het soms conservatief. De kwaliteit van deze schatting hangt af van probleemspecifieke factoren, zoals afnemende meeropbrengsten van bezoeken en het verschil in vraag tussen locaties. Onze analyse biedt organisaties een manier om aan de hand van beperkte data prestatiedoelen omtrent eerlijkheid met donoren te bespreken, die zich vaak niet realiseren wat het potentiële effect van de gestelde prestatiedoelen is.

Personeelsroostering bij de Nederlandse Spoorwegen

Het tweede onderzoek (Breugem, Dollevoet en Huisman, 2022) richt zich op personeelsroostering bij Nederlandse Spoorwegen (NS).

Over de hele wereld werken miljoenen mensen onregelmatig: ze werken niet van maandag tot en met vrijdag van negen tot vijf. In plaats daarvan werken ze in roosters waar diensten (werkdagen) op alle dagen van de week en op onregelmatige tijdstippen kunnen worden ingepland. Voorbeelden hiervan zijn verpleegkundigen, politieagenten, beveiligingspersoneel, luchtvaartpiloten, treinbemanning



Figuur 3: De begrensde en werkelijke waarde van de (relatieve) afname van het aantal cliënten als functie van het minimale percentage voor cliënten van kwetsbare groepen.

en buschauffeurs. Voor al deze werknemers is de kwaliteit van hun rooster van het grootste belang, aangezien het rooster sterk van invloed is op hun werk-privébalans. Ontevredenheid over de roosters kan leiden tot stakingen, met flinke gevolgen voor het openbaar vervoer (4).

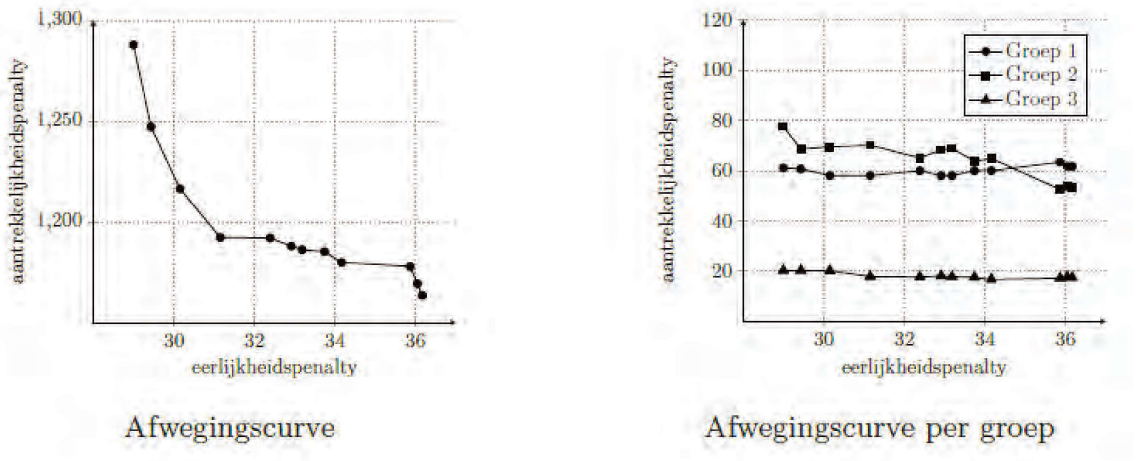
Het doel van personeelsroostering bij NS is om de taken dusdaning toe te wijzen aan het personeel dat de werk-privébalans en andere eigenschappen, waarnaar we zullen verwijzen als de aantrekkelijkheid van het rooster, gemaximaliseerd worden. Personeelsroostering is een complex en goed bestudeerd probleem in OR.

Bij NS en andere bedrijven verrichten grote groepen werknemers (zogenaamde roostergroepen) vergelijkbare taken. Hierbij zijn sommige taken leuker dan andere (denk aan werk op ouder/nieuwer materieel, bijvoorbeeld). Het is daarom belangrijk de taken op een eerlijke manier over de verschillende roosters te verdelen. Het balanceren van leuke en minder leuke taken is een typisch voorbeeld van eerlijkheidsoverwegingen. NS heeft eerlijkheid geformaliseerd aan de hand van de "Lusten-en-Lasten-Delen" regels, die een goede balans van leuk en minder leuk werk tussen roostergroepen reguleren.

We modelleren het personeelsroosteringprobleem als een bi-objectief (eerlijkheid en aantrekkelijkheid) optimalisatieprobleem. Hiervoor ontwikkelen we een nieuwe wiskundige formulering waarin eerlijkheid wordt geïntegreerd door middel van de Lusten-en-Lasten-Delen regels. Voor dit probleem ontwikkelen we een exacte Branch-Price-and-Cut-methodologie. Voor instanties die niet meer exact kunnen worden opgelost, stellen we een heuristische methode voor die de kolomgeneratie technieken van de Branch-Price-and-Cut-methodologie combineert met Variable Dept Neighborhood Search. Ook laten we zien dat de Lusten-en-Lasten regels een natuurlijke interpretatie hebben in ons optimalisatieprobleem en dat deze regels, onder schappelijke aannames, optimaliteitseigenschappen hebben en een begrensde foutmarge in termen van eerlijkheid wanneer men niet exact weet hoe het personeel denkt over eerlijkheid.

We passen onze methode toe op instanties van NS en tonen aan dat onze methodologie het mogelijk maakt om meerdere roosters te genereren met verschillende afwegingen tussen eerlijkheid en aantrekkelijkheid (Figuur 5). Onze analyse biedt twee belangrijke inzichten. Ten eerste moeten planners voorzichtig zijn om de eerlijkheid niet te "overoptimaliseren". We zien dat het versoepelen van eerlijkheidseisen de aantrekkelijkheid aanzienlijk kan verbeteren. Ten tweede zien we dat de afname van aantrekkelijkheid veroorzaakt door een strikt eerlijkheidsniveau ongelijk verdeeld kan zijn over de verschillende roostergroepen. Deze ongelijke verdeling wordt niet meegenomen in de Lusten-en-Lasten-Delen regels. Onze analyse laat dus zien dat de huidige aanpak kritisch geanalyseerd en mogelijk heroverwogen moet worden.

Figuur 4: Ontevreden personeel kan staken, met grote gevolgen voor het openbaar vervoer. Afbeelding van: https://gouweijsselnieuws.nl/2022/09/08/treinstaking-morgen-definitief-door-ontbreken-akkoord-tussen-ns-en-vakbonden/.



Figuur 5: De eerlijkheid-efficiëntie afwegingscurve voor een roosterinstantie bij NS, bestaande uit drie groepen (26 werknemers, die 115 diensten dekken). De rechterfiguur toont de gemiddelde aantrekkelijkheid per groep per werknemer. De assen zijn in termen van strafpunten, lager is dus beter.

Kansen voor impactvol onderzoek

Het inpassen van eerlijkheid in de dagelijkse besluitvorming leidt tot complexe afwegingen die diepgaande analyse vereisen. Binnen OR, en andere vakgebieden, zien we eerlijkheid dan ook een steeds grotere rol spelen. Toch staan we nog voor tal van onopgeloste problemen.

Neem bijvoorbeeld de groeiende problematiek omtrent geneesmiddelentekorten (Vries e.a., 2021). Terwijl de vraag naar essentiële medicijnen blijft stijgen, blijkt het garanderen van eerlijke toegang een steeds grotere uitdaging. Het is niet alleen een kwestie van voldoende productiecapaciteit, maar ook van eerlijke distributie en toegankelijkheid voor alle patiënten, ongeacht hun achtergrond of financiele status. Dit vereist niet alleen nauwe samenwerking tussen verschillende belanghebbenden, maar ook innovatieve oplossingen die de bevoorrading en distributie optimaliseren zonder de principes van eerlijkheid uit het oog te verliezen.

Een ander voorbeeld zijn de groeiende humanitaire crises wereldwijd (Besiou en Van Wassenhove, 2020). Volgens voorspellingen zal bijna de helft van de wereldbevolking tegen 2030 in fragiele of door conflicten getroffen gebieden leven. Tegelijkertijd blijft het tekort aan humanitaire financiering groeien, waardoor organisaties gedwongen worden om met beperkte middelen meer te doen. Dit betekent moeilijke keuzes maken: wie krijgt hulp, hoeveel hulp wordt geboden en voor hoelang? Dit zijn ethische dilemma's waarvoor geen eenvoudige antwoorden bestaan. Nieuw onderzoek kan helpen deze vraagstukken te bekijken vanuit een logistiek perspectief, en zo te zorgen dat middelen zo goed mogelijk ingezet worden en afwegingen juist in kaart worden gebracht.

Dit zijn slechts twee voorbeelden. Wanneer je het nieuws volgt zie je al snel verschillende vergelijkbare problemen, in Nederland, Europa, of wereldwijd. In dergelijke complexe situaties zijn de principes van OR van onschatbare waarde. Door geavanceerde analytische technieken toe te passen en meerdere doelstellingen in evenwicht te brengen, kunnen we betere oplossingen vinden. Dit biedt kansen voor impactvol onderzoek dat niet alleen de efficiëntie van onze operaties verbetert, maar ook bijdraagt aan een eerlijkere wereld.

Literatuur

M. Besiou en L. N. Van Wassenhove. „Humanitarian operations: A world of opportunity for relevant and impactful research". In: *Manufacturing & service operations management* 22.1 (2020), p. 135–145.

T. Breugem, T. Dollevoet en D. Huisman. „Is equality always desirable? Analyzing the trade-off between fairness and attractiveness in crew rostering". In: *Management Science* 68.4 (2022), p. 2619–2641.

T. Breugem en L. N. Van Wassenhove. „The price of imposing vertical equity through asymmetric outcome constraints". In: *Management Science* 68.11 (2022), p. 7977–7993.

H. de Vries e.a. „Short of drugs? Call upon operations and supply chain management". In: *International Journal of Operations & Production Management* 41.10 (2021), p. 1569–1578.

Thomas Breugem is Assistant Professor Operations Management aan de School of Economics and Management, Tilburg University, en maakt deel uit van het Zero Hunger Lab. Zijn onderzoek richt zich op complexe praktische vraagstukken in humanitaire logistiek, toegang tot medicijnen, en transport. Hij focust met name op het gebruik van wiskundige optimalisatiemethoden om operaties te verbeteren en strategische afwegingen te kwantificeren. Hij behaalde zijn doctorstitel aan de Erasmus School of Economics, Erasmus Universiteit Rotterdam, en is een gastonderzoeker bij INSEAD's Humanitarian Research Group.
E-mail: t.breugem@tilburguniversity.edu

In memoriam Prof. dr. René Eijkemans (1963 – 2023)

Ewout Steyerberg en Kit Roes

Op 17 april 2023 overleed René Eijkemans, nu een jaar geleden. René was een begenadigd biostatisticus. Hij was bovenal een fijn mens voor wie de inhoud en aandacht voor de mensen om hem heen voorop stonden. Daarmee heeft hij als onderzoeker, docent en collega grote impact gehad.

René studeerde wiskunde aan de TU Delft en voltooide zijn promotie aan het Erasmus MC, waar hij ruim de tijd voor nam. In zijn proefschrift paste hij geavanceerde methoden toe in de voortplantingsgeneeskunde. Een grote bijdrage van René is de (complexe) ontwikkeling van de leeftijdsverdeling van vrouwen bij geboorte van hun laatste kind als er geen enkele vorm van geboortebeperking zou worden toegepast. Deze curve is een heilige graal van de fertiliteitsgeneeskunde en vormt de basis voor het geïnformeerd adviseren van paren over gezinsplanning.

In 2009 werd René aangesteld als UHD in het Julius Centrum, UMC Utrecht, en in 2014 werd hij benoemd tot hoogleraar Biostatistiek. Voor zijn onderzoek koos hij een nieuwe richting: de complexere, hoog dimensionale modellen en hun daadwerkelijke praktische (klinische) waarde, naast zijn significante bijdrage in de voortplantingsgeneeskunde. In dit gebied van "AI & Machine Learning" koos hij niet het makkelijkste pad. De focus was niet (alleen) het ontwikkelen van een volgend slim algoritme, maar het kunnen vaststellen dat het ook "fit-for-purpose" is een klinisch relevante context om daarmee impact te hebben op de klinische praktijk. Niet altijd even "sexy", maar enorm relevant en nu een van de sleutelproblemen bij de ontwikkelingen en de toepassing van AI en Machine Learning. Zijn bijdrage in het onderzoek reikt wijd, de meest recente publicaties zijn nog van het afgelopen jaar. In het totaal heeft René aan meer dan 400 publicaties bijgedragen, in klinische toepassingen die uiteenlopen van oncologie, cardiovasculaire aandoeningen, covid-19 en de neurologie. In klassieke wetenschappelijk output maatstaven (publicaties, H index van boven de 70) hoorde René bij de top. Deze prestatie was gegrond in zijn uitzonderlijke combinatie van kwaliteiten.

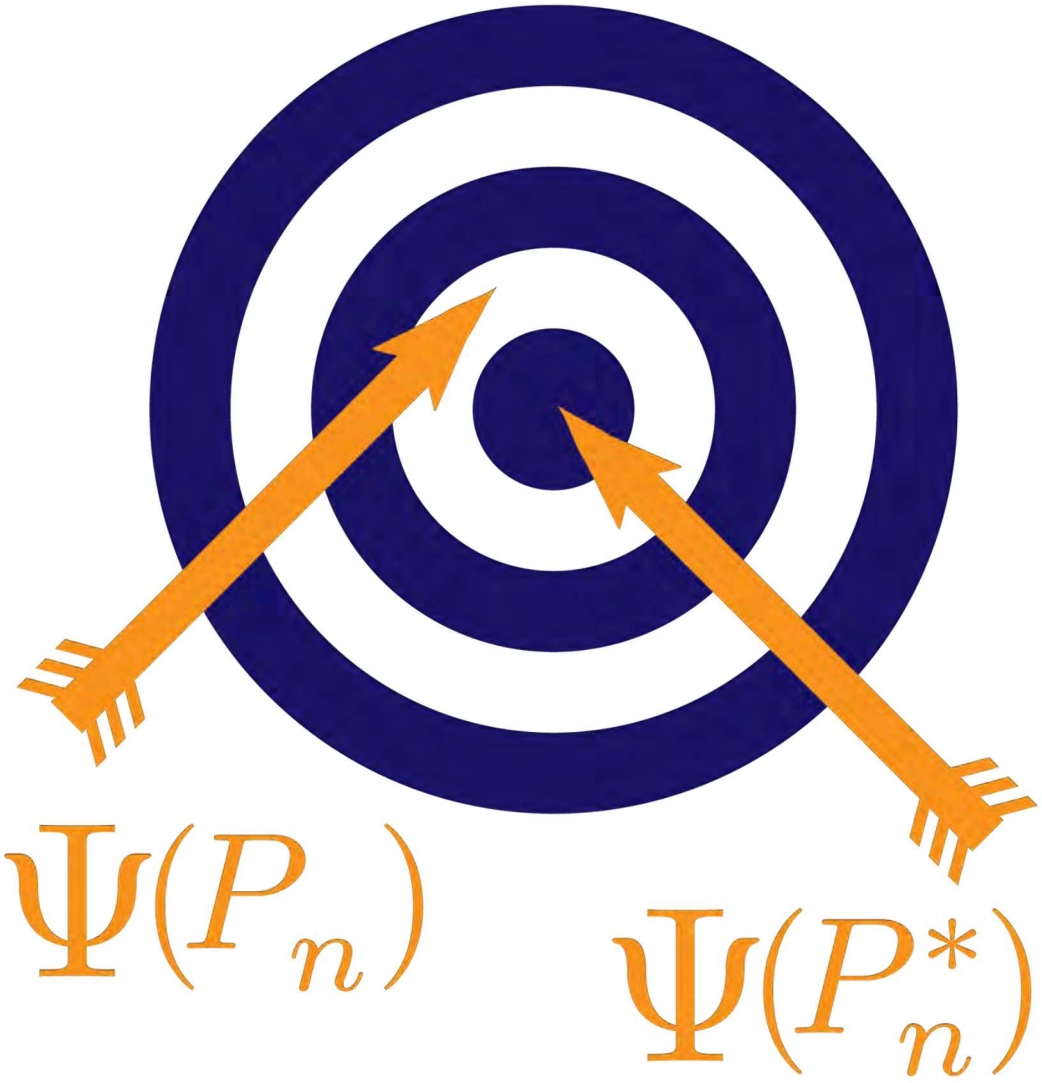
René had een enorm brede en diepe kennis van de statistiek en de toepassingen, was een uitgesproken teamspeler, positief, altijd constructief en met focus op de inhoud. Bij uitstek was hij meer verkennend en vragend, dan directief. Het onderzoek draaide niet om hem, maar om de jonge mensen en hun ontwikkeling en de bijdrage voor patiënten. Hij had een bijzondere combinatie van een zekere bescheidenheid en tegelijkertijd zelfbewust en heel goed weten welke kant hij op wilde: grenzen verleggen in de methodologie en dit ook in de praktijk brengen omdat dit de impact op de kwaliteit en waarde voor patiënten van onderzoek versterkt.

René was een begenadigd docent en in staat om de meest complexe concepten begrijpelijk te maken voor studenten, en op een wijze waardoor studenten wisten: ook dit heb ik nodig om later met de patiënt goede beslissingen te kunnen nemen in de dagelijkse praktijk. Voortvloeiend uit deze kwaliteiten stond hij aan de basis van de nieuwe track "Medical Statistics" in de Master Epidemiology in Utrecht. René was ook actief betrokken bij de onderwijs activiteiten in het Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB I) Training Program Phase I. Hij was lid van het Executive Committee en heeft via conferenties, biostatistiek workshops onderwijsontwikkeling voor MSc- en PhD-studenten, Machine Learning in het onderwijs internationaal op de kaart gezet.

Ook bestuurlijk heeft René meer dan zijn steentje bijgedragen, onder andere in de International Biometric Society en de organisatie van de jaarlijkse ISCB-conferentie in Utrecht in 2015, met zijn kenmerkende combinatie van hoge kwaliteit, teamspirit en relaxte houding.

René Eijkemans is overleden aan de vreselijke ziekte ALS – Amyotrofische Laterale Sclerose. Uitgerekend als onderzoeker heeft hij belangrijke bijdragen geleverd aan de methodologie voor klinisch onderzoek in deze ziekte. Als patiënt was hij deelnemer in een gerandomiseerde klinische studie voor een mogelijk nieuwe behandeling. Hij heeft enorm veel indruk gemaakt met de wijze waarop hij zijn ziekte droeg: met ongekend positieve grondhouding en tegelijkertijd praktische blik, realiteitszin en met professionele reflectie. Tot de laatste fase bleef zijn prioriteit de ontwikkeling van de jonge mensen, als wetenschappers, met het begeleiden van PhD- en Msc-studenten.

Met René Eijkemans zijn we een fijn mens en gewaardeerd en bevlogen collega verloren, die in bescheidenheid een grote bijdrage voor ons vak en onze nieuwe vakgenoten heeft gehad.



https://vanderlaan-lab.org/2021/05/02/tuning-the-highly-adaptive-lasso-estimator/

Voorbij regularisatie... Highly Adaptive LASSO

Mark J. van der Laan en Richard Starmans

Introductie

In 2022 bracht IBM versie 29 van het pakket SPSS uit. De inmiddels bijkans klassieke statistische software kwam in 1969 op de markt onder de naam "Statistical Software for the Social Sciences", werd later omgedoopt tot "Statistical Product and Service Solutions" en zou in 2009 door IBM worden overgenomen. Generaties sociale wetenschappers, epidemiologen, marktonderzoekers, beleidsonderzoekers en andere kwantitatief georiënteerde beroepsgroepen groeiden ermee op; voor sommigen zijn statistiek en SPSS min of meer synoniem en ook in tijden van data science en open source weet het pakket zich ogenschijnlijk moeiteloos te handhaven.

In de toelichting op de nieuwe release meldt het bedrijf met gepaste trots een aantal belangrijke verbeteringen en innovaties. Zo blijkt de bekende regressie-module, niet verrassend gelabeld als General Linear Model, te zijn uitgebreid met een nieuwe tab, te weten Regularisation, waarachter een drietal technieken schuilgaat: Ridge regression (L2-regularisatie), LASSO (L1-regularisatie) en elastic net regression, dat feitelijk een combinatie van L1 en L2 is.

De algoritmen hebben gemeen dat zij een uitbreiding van het aloude lineaire model vormen, waarbij aan de gebruikelijke verlies- of kostenfunctie een zogenaamde penalty-term wordt toegevoegd. Deze beoogt de performance op nieuwe data te verbeteren door overfitting te verminderen en dus een

betere bias-variance tradeoff te bewerkstelligen, en wil bovendien een vorm van feature selection mogelijk maken in hoog-dimensionale datasets. Dit alles geschiedt door zogenaamde "shrinkage", waarbij sommige β-coëfficienten krimpen, heel klein worden (ridge regression) of soms ook tot nul worden gereduceerd (LASSO) en de bijbehorende minder relevante variabelen uit het model (de regressievergelijking) kunnen worden verwijderd. Een en ander moet dan leiden tot zuiniger en beter interpreteerbare modellen. Bij ridge regression is de penalty term de som van de gekwadrateerde coëfficiënten, bij LASSO gaat het om de absolute waarde hiervan. De grootte van de penalty wordt beheerst door een hyperparameter lambda, waarvan de waarde door cross-validation kan worden bepaald.

Het is opmerkelijk dat een marktleider als IBM-SPSS pas recentelijk deze inmiddels vertrouwde regularisatietechnieken in het pakket heeft geïntegreerd. We beperken ons hier tot LASSO, dat staat voor "least absolute shrinkage and selection operator". Het paper van Robert Tibshirani (1956), de bedenker van de methode, verscheen in 1996 in de Journal of the Royal Statistical Society, maar was jaren daarvoor al beschikbaar en dat gold ook voor de S-PLUS software (Tibshirani, 1996). In dit artikel wordt de techniek bovendien geïntroduceerd als een verbetering ten opzichte van ridge regression, dat nog veel ouder is. Al met al duurde het dus ongeveer drie decennia voordat de marktleider in commerciële statistische software deze methoden zou integreren. Sommigen zullen opmerken dat adoptie van nieuwe technieken in een commerciële markt nu eenmaal altijd weerbarstig verloopt. Anderen zullen wellicht benadrukken dat SPSS zelden vooroploopt met implementatie van nieuwe technieken; de module neurale netwerken is daarvan een aansprekend voorbeeld. Er is evenwel een tweede reden waarom de trage adoptie van LASSO frappant is. De techniek behelst geenszins een "Umkehrung aller Werte", zij vergt geen "Gestaltswitch" van de gebruiker. Feitelijk is het een wezenlijke, maar relatief kleine uitbreiding van GLM, waarmee generaties empirische onderzoekers of "research workers", zoals Ronald Fisher ze ooit noemde, zijn grootgebracht en getraind. Het paradigma wordt niet voor niets het "werkpaard van de statistiek" genoemd en klassieke parametrische methoden als multiple lineaire regressie, logistische regressie en de Cox proportional hazard methode voor survivalanalyse worden veelvuldig en routinematig toegepast. De reikwijdte ervan is groot, de interpreteerbaarheid en transparantie worden geroemd en sommigen schrikken zelfs niet terug voor een causale duiding van de β-coëfficiënten in de regressievergelijkingen (padmodellen, SEM), ofschoon dat laatste uiteraard methodologisch en filosofisch beschouwd erg problematisch is (Starmans, 2021). Hoe dan ook, regularisatietechnieken als LASSO hielpen GLM om het tijdperk van big data te betreden, zij stellen onderzoekers in staat binnen het vertrouwde, veilige paradigma te blijven werken, en genieten dan ook een gestadig groeiende populariteit. Dit alles zou een vroegtijdiger adoptie van LASSO in commerciële software voor de hand liggend maken.

"The statistic wars"

Toch lijkt een kanttekening op zijn plaats. Juist omdat LASSO geen enorme revisie van GLM betreft, erft de techniek niet alleen de voordelen, maar ook de nadelen en beperkingen. Deze worden duidelijk tegen de achtergrond van de vele problemen en uitdagingen in de data-analytische praktijk. Men denke aan foutief gespecificeerde parametrische modellen, de erosie van het modelbegrip, de ooit door Ioannides aangekondigde tsunami aan fout-positieve resultaten, de kloof tussen statistical learning and machine learning, p-hacking, de replicatiecrisis, de oude vete tussen frequentisten en bayesianen, "the curse of causality", en meer algemeen, het in sommige kringen vigerende en historisch gewortelde wantrouwen jegens probabilistisch redeneren (Starmans, 2022). Vele problemen zijn zo oud als de in de 19e eeuw begonnen probabilistische revolutie zelf en het feit dat sommige tekstboeken incompatibele inzichten en scholen presenteren als een homogeen, verenigd veld lost dit niet op; pogingen rond de eeuwwisseling van onder meer Leo Breiman om tot een prudente boedelscheiding te komen evenmin (Breiman, 2001). De Amerikaanse filosofe Deborah Mayo sprak in haar recente boek over de voortwoedende "statistics wars" en daagde statistici uit de oplossing te zoeken in haar "error-statistical method" (Mayo, 2018). De statisticus Andrew Gelman nam de handschoen op, verzamelde wereldwijd "expert opinions", maar propageerde zelf een Bayesiaanse oplossing (Gelman, 2019).

Zo zijn er uiteraard vele initiatieven, maar tegen de achtergrond van GLM en LASSO en in het kader van dit korte essay richten we ons hier op de problematiek van verkeerd gespecificeerde parametrische modellen en de daarmee verbonden erosie van het modelbegrip. In een eerdere bijdrage in STAtOR wordt een andere benadering van de crisis in de hedendaagse data-analytische praktijk bepleit (M. v. d. Laan, 2017). In Targeted Learning; the link from

statistics to data science wordt de teloorgang van statistisch redeneren bestreden door terug te gaan naar de basisprincipes van de inferentiële statistiek. Zo is voor de statisticus kennis van het data genererende proces essentieel bij de interpretatie van de data, die worden beschouwd als realisatie van een toevalsvariabele met bijbehorende kansverdeling, de datagenererende verdeling. Om het aantal mogelijke verdelingen waardoor de data kunnen zijn gegenereerd te beperken, dient realistische kennis van het datagenererende proces te worden opgenomen in het statistische model, dat de (benadering van de) ware datagenererende verdeling moet bevatten. In dialoog met de empirische onderzoeker wordt vervolgens een "targeted estimand" gekozen, die zoveel mogelijk aansluit bij de onderzoeksvraag, een mapping van het model naar de parameterruimte behelst en voor een bepaalde verdeling het kenmerk specificeert dat moet worden geleerd van de data. Nadat op deze manier het schattingsprobleem is gedefinieerd wordt een schatter (estimator) geconstrueerd, die leidt tot een schatting (estimate) van de targeted estimand. Deze schatter is zelf een toevalsvariabele, heeft uiteraard een kansverdeling, the sampling distribution, waarvan de spreiding de onzekerheid van de schatter aangeeft. Deze hier enigszins rechttoe rechtaan geschetste principes zijn theoretisch onomstreden, maar worden in de data-analytische praktijk soms naar de achtergrond verdrongen. Dikwijls vindt misspecificatie van het model plaats, dat daardoor niet de ware data genererende functie bevat; fictieve kennis c.q. aannames over de vorm van de stochastische relaties tussen de covariaten leidt tot het arbitrair postuleren van een functioneel verband, waardoor de target estimand niet langer de onderzoeksvraag representeert en het gerapporteerde betrouwbaarheidsinterval waarschijnlijk niet langer het antwoord op die vraag verschaft. Dat dit alles dikwijls niet eens problematisch wordt geacht, vormt een belangrijke factor in de zogenaamde "erosie van het modelbegrip" (Starmans, 2011).

Juist de vele parametrische methoden die gezamenlijk het GLM-paradigma vormen, worden vaak vanuit deze benadering toegepast en GLM-LASSO zal dit niet verbeteren. Het lijkt aantrekkelijk dat een regressiemodel kan worden geschat, zelfs als het aantal covariaten het aantal observaties overtreft, maar door de misspecificatie van het model is niet eens duidelijk wat de coëfficiënten precies betekenen. Veel resultaten in de LASSO-literatuur betreffen zogenaamde "post-model selection" inferentie van een bepaalde coëfficiënt in het LASSO-model, waarbij de aanname is dat LASSO het ware model kan fitten, zodat enkel gecorrigeerd hoeft te worden voor de "data-adaptivity" van het geselecteerde model. Je kunt dan bewijzen dat, indien het beginmodel de ware regressie bevat, LASSO met grote waarschijnlijkheid de juiste niet-nul coëfficiënten weet te selecteren. Doorgaans veronderstellen deze resultaten de onrealistische aanname van "sparsity". Deze houdt in dat het gekozen regressiemodel slechts enkele ware niet-nul coëfficiënten bevat en vervolgens wordt dan bewezen dat LASSO het ware onderliggende parametrische model kan leren wanneer het aantal observaties toeneemt. De ware coëfficiënten mogen klein zijn, maar zijn evenwel zelden gelijk aan nul en ook in het licht van de overduidelijke misspecificatie van het model is het onredelijk aan te nemen dat de meeste coëfficiënten gelijk zijn aan nul. Dat alles impliceert geenszins dat LASSO geen belangrijke rol kan spelen bij prediction, het verminderen van overfitting en dimensiereductie. Voorwaarde is wel dat het wordt opgevat als een rijke verzameling kandidaat machine learning functies om bij voorbeeld het conditionele gemiddelde van de uitkomstvariabele als functie van de covariaten te schatten; inclusief interacties en transformaties van basisfuncties. Als zodanig maakt het dan deel uit van bij voorbeeld een stacking-algoritme in de context van ensemble-leren.

Highly adaptive LASSO

In het paradigma van Targeted Learning (M. v. d. Laan en Rose, 2011; M. v. d. Laan en Rose, 2018) worden de problemen in de data-analytische praktijk anders benaderd, zodat modelmisspecificatie wordt voorkomen, de erosie van het modelbegrip wordt gestopt, het schattingsprobleem precies wordt gedefinieerd en estimation weer een duidelijke rol speelt in de inferentiële statistiek. TMLE behelst een secundaire "targeting" stap, die primair de bias-variance tradeoff van de parameter of interest optimaliseert en bezit enkele eigenschappen die de methode zeer geschikt maken voor causal inference in observationele data, waarbij een methodische scheiding wordt gemaakt tussen de statistische target parameter en de causale estimand. De data zijn uiteraard een realisatie van een verzameling toevalsvariabelen uit een onderliggende verdeling P van de geobserveerde data, aangegeven met $O = (X, A, Y) \sim P$. We lichten een en ander toe aan de hand van een bekende, eenvoudige causale estimand als "the average treatment effect" (ATE), maar de methode werkt evenzeer bij complexe causale estimands met bij voorbeeld

rechts-gecensureerde survivaldata en tijdafhankelijke covariaten. Men denke aan longitudinale designs en "post-intervention"verdelingen met verschillende tijdpunt interventies, waarbij volgens een dynamische regel bij voorbeeld een therapie wordt aangepast op basis van de steeds veranderende toestand van een patiënt. Allereerst is de methode dubbel robuust. In het eenvoudige geval van een ATE betekent dit dat TMLE zuivere schattingen garandeert, indien hetzij het uitkomstmechanisme $E(Y|A,X)$, hetzij het blootstellingsmechanisme $P(A = 1|X)$ correct is gespecificeerd in het geval van parametrische regressie. ATE wordt dus zuiver geschat wanneer $E(Y|A,X)$ niet consistent is geschat, maar $P(A = 1|X)$ wel. Omgekeerd zal bij enkel een consistente schatting van $E(Y|A,X)$, de targeting stap de zuivere schatting bewaren en eindige steekproef bias verwijderen. Een tweede aantrekkelijke eigenschap betreft het feit dat TMLE ook asymptotisch efficient is, indien beide mechanismen consistent zijn geschat. In zekere zin integreert TMLE hiermee G-computation en propensity score methoden (zoals de bekende Inverse Probability Weighting techniek). Daarnaast is TMLE een substitutie-schatter, die robuuster is bij outliers en sparsity. Voor verdere details, waaronder de met TMLE geassocieerde ensembletechniek Super Learner verwijzen we naar de eerdergenoemde STAtOR-publicatie (M. v. d. Laan, 2017).

Om het beste van twee werelden te combineren werd in 2017 Highly Adaptive LASSO (HAL) voorgesteld (M. v. d. Laan en Benkeser, 2017). Ruwweg behelst dit het volgende. De gebruikelijke least squares regressie beoogt de schatting van het conditionele gemiddelde van de uitkomstvariabele als functie van de covariaten door de empirische MSE over alle lineaire combinaties van de d covariaten te minimaliseren. Wat nu indien we dit veranderen in alle functies van de d covariaten? Dat lijkt bizar en een opmaat tot enorme overfitting en dientengevolge slechte performance op nieuwe data. Een constraint is dan ook nodig en deze kan worden gevonden in zogenaamde multivariate, reëel-waardige "cadlag functies" met eindige "sectional variation norm". Deze functies van de d covariaten zijn rechtscontinu, met links overal limieten en de variatienorm van de functie, beperkt tot deelverzamelingen van de covariaten, is eindig voor alle mogelijke deelverzamelingen. Dat wil zeggen "bounded by" een constante, waarvan de waarde via cross-validation kan worden bepaald en daarmee is de constraint gedefinieerd. Een functie als $sin(1/x)$ op $[0,1]$ heeft uiteraard oneindige variatie norm, maar doorgaans zijn relaties gemeten in de werkelijkheid minder grillig en bij voorbeeld monotoon of ten minste unimodaal. Het is dan ook realistisch om aan te nemen dat de ware target functie van een multivariate X een eindige "sectional" variatie norm heeft. Daarmee fungeert deze als een maat voor de complexiteit van de functie; uiteraard is een functie met veel pieken en dalen lastiger te fitten dan wanneer deze bijna constant is. Daarom kan de begrenzing op de variatie norm fungeren als de constraint die nu juist nodig is. Zo kwantificeert de variatienorm van de univariate secties de eerste orde veranderingen van de functie en kwantificeert de variatienorm van de bivariate secties tweede orde interacties van de functie. De variatie norm sommeert (integreert) als het ware alle kleine veranderingen en pakt daarmee de eerste en tweede orde variaties van de functie op (Gill, M. J. v. d. Laan en Wellner, 1995).

De volgende stap is dan het definiëren van een non-parametrische least-square schatter, die we hier aanduiden als "Highly Aaptive LASSO (HAL) en die de empirische MSE minimaliseert over alle cadlag-functies met begrenzing M op hun sectional variatienorm. Dit levert een zeer groot model en het is dan ook realistisch aan te nemen dat de ware data genererende functie hierin is bevat. In de machine learning literatuur is het minimaliseren van het empirische risico uiteraard niet nieuw, maar bij HAL gaat het om deze specifiek gedefinieerde klasse van functies met de sectional variatie norm als voornaamste constraint. Bovendien hangt de schatter niet af van lokale smoothness-aannamen, maar respecteert globale smoothness. Het is feitelijk de meest non-parametrische schatter voor regressie en dezelfde principiële definitie kan worden toegepast op andere noties van empirisch risico, voorbij MSE. Neem bij voorbeeld een HAL-estimator van de dichtheid van de data, gegeven een non-parametrische MLE; we maximaliseren dan de empirische log-likelihood over alle dichtheden met een begrenzing op de "sectional" variatienorm. Daarmee construeren we ook HAL-schatters van (conditionele) dichtheden, gemiddelden en behandelingseffecten en feitelijk elke parameter van de verdeling waarin we geïnteresseerd zijn.

HAL staat voor "Highly Adaptive LASSO" en de verwijzing naar LASSO kan onder meer vanuit een algoritmisch perspectief eenvoudig worden gemotiveerd. Het blijkt namelijk dat de genoemde klasse van cadlag-functies kan worden geschreven als een lineaire combinatie van zero-order-splines van de vorm $I(X > u) = \prod_{k=1}^{d} I(X_k > u_k)$ geïndexeerd door een multivariate knot-point $u = (u_1, .., u_d)$. en dat de L_1-norm van de vector van coëfficiënten in de repre-

sentatie precies de "sectional" variatie norm van de functie is! Feitelijk wordt dus een LASSO gerund op een groot aantal zero-order spline basic functions $I(X > u_j)$ over vele knotpoints u_j in de d-dimensionale eenheidskubus $[0,1]^d$. In het bivariate geval wordt dit:

$$f(x_1, x_2) = f(0,0) + \int I_{u \le x_1} df_1(u) + \int I_{v \le x_2} df_2(v) + \int I_{u \le x_1} I_{v \le x_2} df(u,v).$$

Het blijkt dat HAL de ware target functie benadert met een snelheid van $n^{-1/3}$ up till a log n-factor $(logn)^{d/2}$ en relevant is daarbij vooral dat de snelheid niet afhangt van het aantal covariaten. Het theoretische bewijs betreffende de snelle convergentie is gebaseerd op empirische procestheorie, kennis van de complexiteit van de klasse van functies en standaardtechnieken voor de analyse van MLE. Daarnaast is HAL uitvoerig bestudeerd en getest middels simulaties en publiek beschikbare datasets. Alleen dat al maakt HAL een welkome uitbreiding van het Super Learner essemble. Ondanks het feit dat HAL al een asymptotisch efficiënte plugin-schatter oplevert, blijft TMLE nodig, al was het maar omdat in eindige steekproeven de "curse of dimensionality" optreedt en dan is de targeting stap van TMLE cruciaal. Andere aspecten van HAL moeten in dit korte artikel buiten beschouwing blijven, maar inmiddels zijn enkele belangrijke uitbreidingen gerealiseerd, waaronder Multitask-HAL, Meta-HAL en uitkomst-adaptieve HAL-TMLE. We hopen hierop in een volgende bijdrage te kunnen ingaan. Dat geldt ook voor de betekenis van dit alles voor causal inference en de roep om explanatory data science in tijden van deep learning en generatieve AI (Starmans, 2024).

Samenvattend kan worden gesteld dat het Targeted Learning paradigma nu is gebaseerd op een integratie van targeted maximum likelihood estimation, de ensemble-methode Super Learner en de highly adaptive lasso met uitbreidingen. De software van HAL is onder meer beschikbaar via https://cran.r-project.org/web/packages/hal9001/vignettes/intro_hal9001.html. Verdere informatie over dit paradigma in statistiek en data science is te vinden op de website van het Center for Targeted Machine Learning and Causal Inference (https://ctml.berkeley.edu/).

Literatuur

L. Breiman. „Statistical modeling: the two cultures". In: *Statistical Science* 16.3 (2001), p. 199–215.

A. Gelman. „Many Perspectives on Deborah Mayo's "Statistical Inference as Severe Testing: how to get beyond the statistics wars". In: *https://arxiv.org/abs/1905.08876* (2019).

R. D. Gill, M. J. van der Laan en J. A. Wellner. „Inefficient estimators of the bivariate survival function for three models". eng. In: *Annales de l'I.H.P. Probabilités et statistiques* 31.3 (1995), p. 545–597. URL: http://eudml.org/doc/77521.

M. van der Laan. „Targeted Learning. The link from statistics to data science". In: *STAtOR* 18.4 (2017).

M. van der Laan en D. Benkeser. „Generally Efficient Targeted Minimum Loss Based Estimator based on the Highly Adaptive Lasso". In: *International Journal of Biostatistics 13(2)* (2017).

M. van der Laan en S. Rose. *Targeted Learning; Causal Inference for Observational and Experimental Data*. Springer, 2011.

M. van der Laan en S. Rose. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer, 2018.

D. Mayo. *Statistical Inference as Severe Testing: how to get beyond the statistics wars*. Cambridge University Press, 2018.

R. J. C. M. Starmans. „Generatieve Kunstmatige Intelligentie; van dageraad tot doomsday". In: *Filosofie - Tijdschrift, 34(2)* (2024).

R. J. C. M. Starmans. „Padanalyse, structurele vergelijkingen modellen en de roep om Explainable AI en data science". In: *STAtOR, 22(2)* (2021).

R. J. C. M. Starmans. „Statistiek en Filosofie: een voortschrijdende samenspraak". In: *STAtOR, 23(2)* (2022).

R. J. C. M. Starmans. *Targeted Learning; Causal Inference for Observational and Experimental Data*. 2011. Hfdstk. Models, Inference and Truth; Probabilistic Reasoning in the Information Era, p. 1–20.

R. Tibshirani. „Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society* 58.1 (1996), p. 267–288.

Mark J. van der Laan is hoogleraar biostatistiek en statistiek aan UC Berkeley. Zijn onderzoeksinteresses omvatten statistische methoden in de genomica (computationele biologie), overlevingsanalyse, gecensureerde data, Targeted Maximum Likeihood Estatimation in semiparametrische modellen, causal inference, data adaptive loss-based super learning en multiple testing. In 2005 ontving hij de VVSOR Van Dantzig Prijs.
E-mail: laan@stat.berkeley.edu

Richard Starmans doet als statisticus en filosoof onderzoek op het snijvlak van filosofie en exacte wetenschappen, meer in het bijzonder wiskundige statistiek, informatica en data science. Hij is als associate professor "Grondslagen van AI en Data Science" verbonden aan de Universiteit van Tilburg en als managing director van een landelijke onderzoekschool op het gebied van informatie- en kennissystemen verbonden aan de Universiteit Utrecht.
E-mail: E-mail: starmans@cs.uu.nl

Van Dantzig Prijs 2025

Eens in de vijf jaar keert de VVSOR de Van Dantzig Prijs uit, de hoogste prijs in de Nederlandse Statistiek en Operations Research. Deze prijs, vernoemd naar David van Dantzig, wordt uitgekeerd aan jonge onderzoekers die de afgelopen vijf jaar een zeer noemenswaardige bijdrage hebben geleverd aan de statistiek of operations research. De vorige prijswinnaars waren Daniel Dadush en Marloes Maathuis.

De jury voor de Van Dantzig Prijs 2025 bestaat uit: Aad van der Vaart, Onno Boxma, Laura Spierdijk, Mark van de Wiel en Casper Albers. Voor deze prijs hopen wij vanuit de vereniging een (groot) aantal nominaties te ontvangen.

Kandidaten dienen op 1 januari 2025 niet ouder te zijn dan 40 jaar; waarbij we eenzelfde extensieregeling hanteren voor o.a. ouderschaps- en zorgverlof als NWO (https://www.nwo.nl/extensieregeling). Daarnaast dienen kandidaten geli-eerd te zijn aan de Nederlandse statistiek en OR; bijvoorbeeld door in Nederland werkzaam te zijn of de Nederlandse nationaliteit te hebben.

Nominaties kunnen gestuurd worden aan Casper Albers (c.j.albers@rug.nl) of een ander jurylid naar keuze. Nominaties dienen vergezeld te gaan van een motivatie, een cv van de genomineerde en twee tot vier referenties. De deadline voor nominaties is 1 december 2024, maar eerder nomineren wordt nadrukkelijk aangemoedigd.

ISI World Statistics Congress 2025

Van 5 tot en met 9 oktober 2025 vindt het 65e ISI World Statistics Congress plaats in het World Forum in Den Haag. Dit gerenommeerde congres, dat al 140 jaar bestaat, is een van de belangrijkste internationale bijeenkomsten op het gebied van statistiek en datawetenschap.

Het WSC 2025 is dé plek waar de statistische wereld samenkomt. Gedurende vier dagen zullen meer dan 2000 statistici en datawetenschappers uit de academische wereld, officiële statistiek, het bedrijfsleven, centrale banken en statistische verenigingen samenkomen. Zowel junior als senior professionals kunnen profiteren van meer dan 200 hoogwaardige sessies en lezingen van enkele van de meest invloedrijke namen in de statistiek.

Dit congres staat bekend om zijn waardevolle mix van hoogwaardige inhoud en inzichten, uitgebreide netwerk- en wervingsmogelijkheden en de vele mogelijkheden om kennis te delen en op te doen. Tevens zullen er een aantal prestigieuze awards uitgereikt worden, waaronder de International Prize in Statistics.

Het inspirerende programma biedt de mogelijkheid om de nieuwste ontwikkelingen in statistisch onderzoek en praktijk te verkennen.

De registraties voor het ISI World Statistics Congress 2025 openen op 1 juli 2024. We zien u graag in 2025 in Den Haag!

Meer info en registeren
https://www.isi-next.org/conferences/wsc2025/