

STAtOR

Periodiek van de VWSOR jaargang 25, nummer 3/4, december 2024

**Het fundamentele verschil tussen
individuele en geïndividualiseerde
causale effecten**

Vertrouwen op een planningstool

Eerlijker serveren bij squash

De munt is vol verrassingen

**Wapenen tegen onzekere parameters
bij optimalisatie?**

3D-Beleggen

**Statistiek, astronomie en de
constructie van data**

The UK Polling Disaster

**Oproep voor het insturen van
abstracts voor de Annual Meeting**



STAToR

Jaargang 25, nummer 3/4, december 2024

STAToR is een uitgave van de Vereniging voor Statistiek en Operations Research (VWSOR). 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 Stemerding (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@wsor.nl

Bestuur van de VWSOR

Voorzitter: prof. dr. Casper Albers, db@wsor.nl; Secretaris: dr. Tsega Kahsay Gebretekle, secretaris@wsor.nl; Penningmeester: dr. Rebecca Kuiper, penningmeester@wsor.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 VWSOR

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

Voor advertenties

Prof. dr. J.A.S. Gromicho, stator@wsor.nl

STAToR verschijnt in maart, juli en december.

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

Alweer een jaar verder

Met dit nummer maken we 25 jaar STAToR vol. Begin 2000 werd een proefnummer uitgebracht en de reacties waren zó positief dat men besloot door te gaan. Onder het hoofdredacteurschap van Dick den Hertog en met redacteuren uit alle VWSOR secties werd begonnen aan iets waarvan we ook nu nog vinden dat het waardevol is. We hopen in ons volgende nummer hier uitgebreider bij stil te staan.

Dit nummer kent weer een afwisseling aan artikelen. Richard Post vertelt over het verschil tussen individuele en geïndividualiseerde Causale Effecten, iets dat belangrijk is bij het bepalen van het effect van een interventie zoals een medische behandeling.

Planningen worden tegenwoordig met een tool gemaakt, maar dan moet er wel vertrouwen zijn in zo'n instrument. Ignace De Croix en zijn co-auteurs lichten dit toe. Mocht u het taalgebruik ongewoon vinden: de redactie heeft Vlaamse woorden en zinsconstructies geëerbiedigd.

Mirjam Loois beschrijft waarom het om-en-om serveren bij squash eerlijker is dan de gebruikelijke volgorde en Pieter Knops legt uit hoe je om kunt gaan met onzekerheid omtrent de parameters bij optimalisatie.

Grote beleggers, zoals pensioenfondsen, kunnen focussen op de lange termijn en dat geeft de mogelijkheid om aspecten als duurzaamheid mee te nemen in hun beslissingen. Over dit 3D beleggen vertelt Rogier Swierstra.

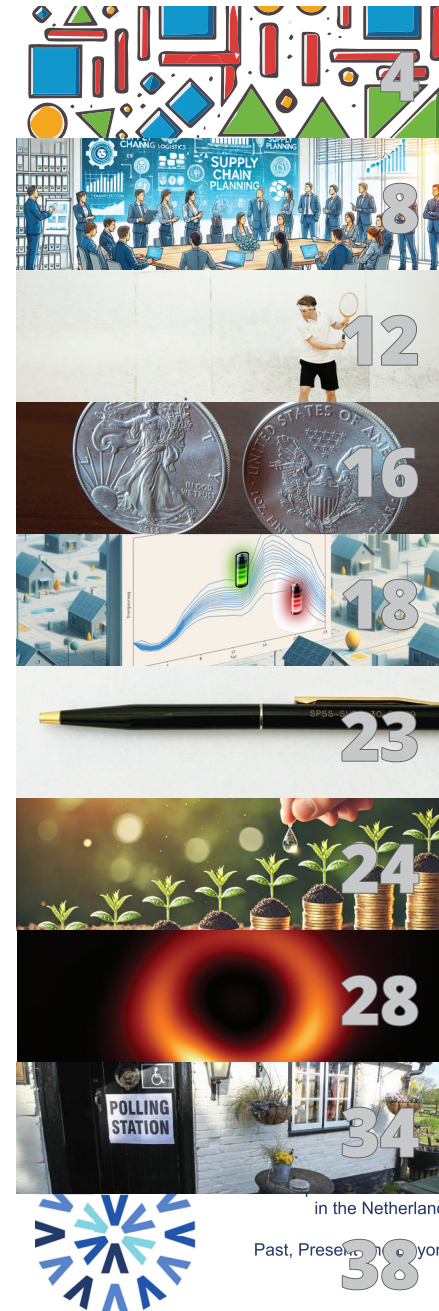
De moderne astronomie berust op astronomische hoeveelheden data. Het artikel van Richard Starmans gaat over de belangrijke rol van data en statistiek in deze wetenschap.

Ook de columns zijn divers. Als u dit nummer leest weten we in hoeverre de verschillende voorspellingen Trump vs. Harris juist waren. Maar Jelke Bethlehem probeert te achterhalen waarom peilingen in het Verenigd Koninkrijk steeds weer misgaan. Henk Tijms is geboeid door de vaak foute ideeën die mensen hebben over kansen en Gerrit Stemerding verklaart waarom het feit dat de WUR in de jaren '80 vrijwel alleen DEC computers had leidde tot de massa-licentie van SPSS voor heel universitair Nederland.

Verder verenigingszaken zoals nieuws van de Young Statisticians, de oproep om kandidaten aan te bevelen voor de jaarlijkse VWSOR-prijzen, de vraag om topics voor de komende Annual Meeting en de samenwerking met Instats. Het tweejaarlijkse World Statistics Congress van het ISI, dat al stamt uit de 19e eeuw, beleeft in oktober 2025 haar 65e bijeenkomst, ditmaal in Den Haag. Hierover het laatste nieuws.

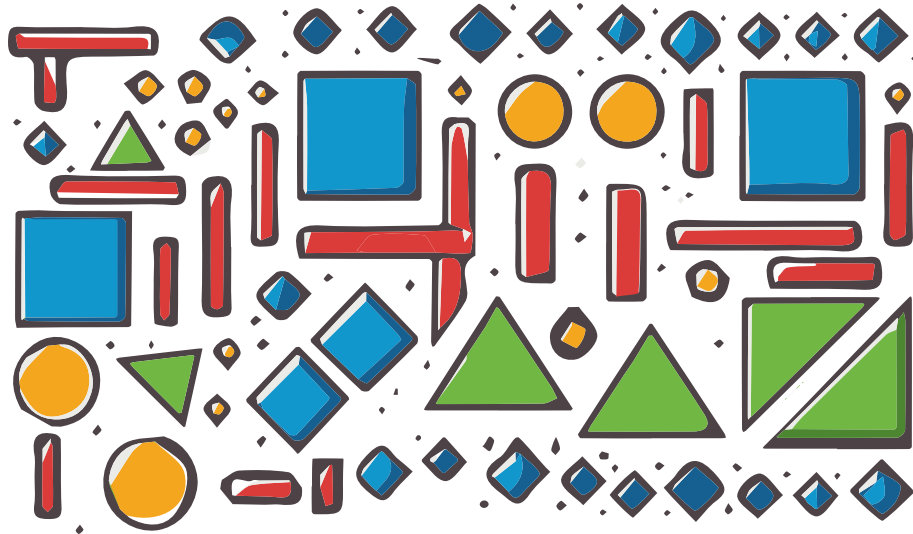
Wij wensen u een goede winter en veel leesplezier met dit dubbelnummer.

De STAToR-redactie



INHOUD

- 2 Alweer een jaar verder**
- 4 Het fundamentele verschil tussen individuele en geïndividualiseerde causale effecten** | Richard Post
- 8 Vertrouwen op een planningstool: valkuilen en succesfactoren** | Ignace De Croix, Ann Vereecke, Karlien Vanderheyden en Shari De Baets
- 12 Om en om serveren bij squash is eerlijker** | Miriam Loois
- 16 De munt is vol verrassingen** | Henk Tijms
- 18 Kansverdelingsvoorspellingen: De manier om je te wapenen tegen onzekere parameters bij optimalisatie?** | Pieter Knops
- 22 Young statisticians**
- 23 Kleine oorzaken ... – column** | Gerrit Stemerding
- 24 3D-Beleggen: Portefeuilleconstructie voor beleggers met duurzaamheidsdoelen** | Rogier Swierstra
- 28 Statistiek, astronomie en de constructie van data** | Richard Starmans
- 33 Samenwerking VWSOR met Instats**
- 33 65th ISI World Statistics Congress**
- 34 The UK Polling Disaster – column** | Jelke Bethlehem
- 39 Oproep voor het insturen van abstracts voor de Annual Meeting**
- 39 Uitmunten master's of PhD thesis begeleid?**



Het fundamentele verschil tussen individuele en geïndividualiseerde causale effecten

Richard Post

Steeds vaker is men geïnteresseerd in het schatten van het effect van een interventie voor een specifiek individu om zo tot een optimale behandel keuze te komen. In dit artikel definieer ik het individuele causale effect en zet ik de (on)mogelijkheden om de heterogeniteit in causale effecten te schatten met observationele data uiteen. In het bijzonder ga ik in op de verschillen tussen conditionele gemiddelde effecten en daadwerkelijke individuele causale effecten.

Inleiding

De toenemende hoeveelheid beschikbare gegevens in de gezondheidszorg heeft er voor gezorgd dat veel onderzoek in de geneeskunde zich richt op *personalized medicine* (Kosorok en Laber, 2019). Daarbij is het ultieme doel om voor ieder individu op basis van diens unieke eigenschappen de beste behandeling beslissing te nemen. Het overgrote deel van de beschikbare data is van observationele (routinematig verzameld) aard waarmee correlaties tussen de individuele eigenschappen en uitkomsten geïdentificeerd kunnen worden. Op basis van die correlaties kunnen voorspellingen gedaan worden om

patiënten te identificeren die in het huidige zorgsysteem een verhoogd risico hebben. Echter, het is een veel grotere uitdaging om daadwerkelijke causale relaties af te leiden om zo te kunnen voorspellen hoe het risico verandert ten gevolge van een specifieke behandeling (Hernán, Hsu en Healy, 2019). De afgelopen 30 jaar zijn er grote stappen gezet op het gebied van theoretisch en methodologisch onderzoek naar causale inferentie met observationele data (Hernán en Robins, 2020). Dit geeft inzicht in de causale interpretatie van geschatte effecten en heeft aan het licht gebracht welke aannames nodig zijn om causale relaties af te leiden uit observationele data.

Het individuele causale effect

In dit artikel beperk ik mij tot een setting met een dichotome interventie ($a \in \{0, 1\}$) en een continue uitkomst. De geobserveerde uitkomst, en de toegevoegde behandeling, voor een willekeurig individu i kan beschreven worden met de stochastische variabelen Y_i en A_i . Potentiële uitkomsten, voor een willekeurig individu i , in de universa waar niemand ($a = 0$) of iedereen wordt behandeld ($a = 1$), kunnen beschreven worden met de tweedimensionale stochastische variabele (Y_i^0, Y_i^1) . De *consistency*-aannname impliceert dat behandeling uniek is gedefinieerd zodat $Y_i = Y_i^{A_i}$. Het individuele causale effect van de behandeling voor individu i is gelijk aan het verschil tussen de potentiële uitkomsten met en zonder behandeling, oftewel $Y_i^1 - Y_i^0$. Idealiter heeft een behandeling een groot (gunstig) effect en is dit effect zo homogeen mogelijk in de populatie. Echter, in de praktijk zijn effecten vaak heterogeen: er zijn individuen die meer baat hebben bij de behandeling dan gemiddeld terwijl er ook individuen kunnen zijn waarvoor de behandeling schadelijk is zoals in het voorbeeld weergegeven in Figuur 1. Het fundamentele probleem van causale inferentie is dat per individu maar één potentiële uitkomst geobserveerd kan worden, Y_i^1 als $A_i = 1$ en Y_i^0 als $A_i = 0$. De verdeling van individuele causale effecten is daarom, net zoals de gezamenlijke verdeling van de potentiële uitkomsten, niet identificeerbaar. De focus van causale inferentie is daarom vaak op gemiddelde effecten, $E[Y^1 - Y^0]$. Het afleiden van dit gemiddelde effect is al een hele uitdaging als gevolg van het fundamentele probleem. Onder *exchangeability*: $A \perp Y^a$, (en *consistency*) is het gemiddelde causale effect gelijk aan het verschil in de verwachting van de geobserveerde uitkomsten in de behandelde en onbehandelde groep

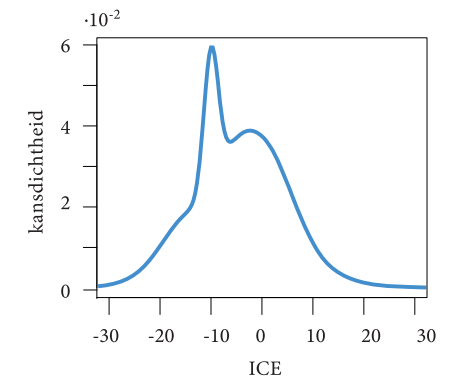
$$E[Y^1] - E[Y^0] = E[Y | A = 1] - E[Y | A = 0].$$

Als *exchangeability* niet geldt, verschillen behandelde en onbehandelde individuen gemiddeld genomen ook in andere kenmerken (bijvoorbeeld leeftijd). Een geobserveerd verschil in de gemiddelde uitkomsten kan dan het gevolg zijn van het laatstgenoemde verschil in plaats van het gevolg van de behandeling. Een gerandomiseerd experiment is zo opgezet dat *exchangeability* altijd geldt zodat verschillen in de uitkomst toegeschreven kunnen worden aan de behandeling en is daarom de gouden graal voor causaliteit-onderzoek.

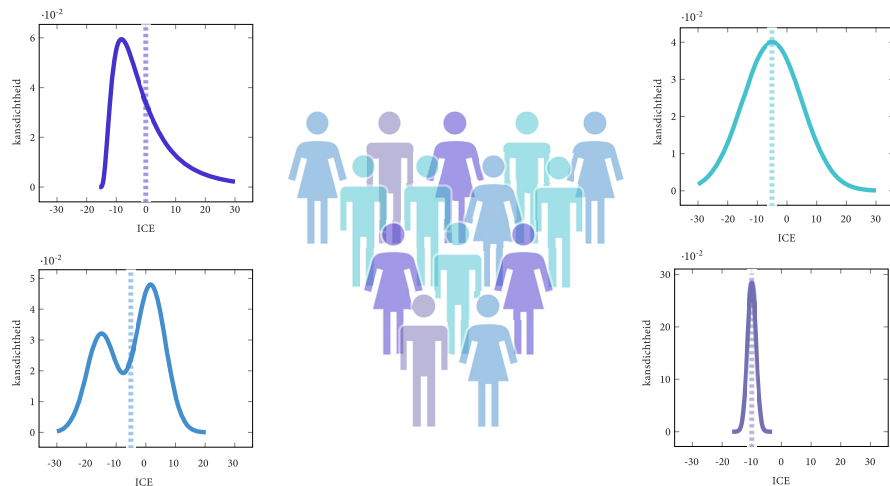
Kwantificeren van effect heterogeniteit op basis van data

In veel gevallen is het gemiddelde causale effect niet genoeg om een behandel beslissing te nemen voor een specifiek individu. De focus verschuift daarom vaak naar het gemiddelde conditionele (op gemeten kenmerken X) causale effect, beter bekend als het *conditional average treatment effect* (CATE), $E[Y^1 - Y^0 | X]$. Wanneer men geïnteresseerd is in de CATE, moet deze allereerst gelinked worden aan een correlatiemaat welke uit geobserveerde data afgeleid kan worden. Hiervoor moet, naast *consistency*, *conditional exchangeability* gelden, $A \perp Y^a | X$, zodat de CATE (voor $X = x$) gelijk is aan

$$E[Y | A = 1, X = x] - E[Y | A = 0, X = x].$$



Figuur 1: Hypothetisch voorbeeld van een verdeling van individuele causale effecten waarbij een negatief effect gunstig is. Denk bijvoorbeeld aan het effect van een medicijn op de systolische bloeddruk bij patiënten met hypertensie.



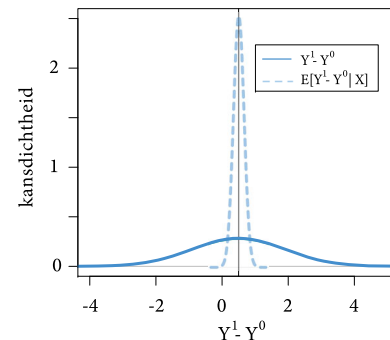
Figuur 2: De populatieverdeling in Figuur 1 zou opgesplitst kunnen worden in vier conditionele causale effect-verdelingen met verschillende CATEs (verticale stippellijnen).

$E[Y | A = a, X = x]$ is een maat van correlatie en kan dus geschat worden op basis van geobserveerde data. Gerandomiseerde experimenten zijn vaak relatief klein en zijn daardoor ongeschikt om deze conditionele verwachtingen te schatten met een redelijke maat van onzekerheid. Daarom wordt er voor het schatten van de CATEs vaak gebruik gemaakt van grote observationele datasets. Door het fundamentele probleem van causale inferentie is het onmogelijk om data te gebruiken om de *conditional exchangeability*-aanname te verifiëren. De validiteit van deze aanname berust dus volledig op *expert knowledge*.

Zelfs wanneer we ervan uit kunnen gaan dat *exchangeability* van toepassing is, bijvoorbeeld na het uitvoeren van een gigantisch gerandomiseerd experiment, is het belangrijk om te blijven beseffen wat we kunnen schatten. De geïndividualiseerde CATE kan verschillen voor twee individuen wanneer zij verschillen in kenmerken X . Echter, binnen subpopulaties gevormd door individuen met vergelijkbare X , kan er ook nog steeds heterogeniteit in individuele causale effecten bestaan. In Figuur 2 wordt een hypothetisch voorbeeld geschetst waarbij op basis van de beschikbare X de populatie kan worden

opgedeeld in vier subpopulaties met elk een andere causale effect verdeling. De CATE is inderdaad verschillend voor iedere groep. Er is één groep waarin het behandelings-effect relatief homogeen is waarvoor je kunt stellen dat de CATE praktisch gelijk is aan het individuele effect. Echter, in de overige drie populaties is het duidelijk dat een individueel effect nog altijd serieus kan afwijken van de CATE.

Steeds meer onderzoek richt zich op de ontwikkeling van machine learning-algoritmes om CATEs te schatten voor hoog dimensionale X om zoveel mogelijk van de heterogeniteit in behandelings-effecten te verklaren. Denk bijvoorbeeld aan het *causal random forest* beschreven in (Wager en Athey, 2018) en geïmplementeerd in het R package *grf*. Zeker in een setting waar de relatie tussen A en Y op een ingewikkelde manier verandert per X , kan de focus van de data-analyse geheel verschuiven naar het afleiden van deze complexe correlaties. Het gevaar is dan dat men vergeet wat er daadwerkelijk wordt geschat. De geïndividualiseerde CATE wordt dan misschien geïnterpreteerd als het individuele causale effect. In het geval van resterende heterogeniteit, zal de variabiliteit in CATEs de heterogeniteit in individuele effecten onderschatten (Figuur 3).



Figuur 3: De verdeling van individuele causale effecten gebruikt in (Post e.a., 2024) en de verdeling van CATEs die in dit geval de effect heterogeniteit flink onderschat.

Identificeerbaarheid van de conditionele variantie

We zouden dus graag kennis nemen van de resterende heterogeniteit in de causale effecten na het conditioneren op kenmerken X . Echter, de conditionele variantie $\text{var}(Y^1 - Y^0 | X)$ is gelijk aan

$$\text{var}(Y^1 | X) - \text{var}(Y^0 | X) - 2\text{cov}(Y^1 - Y^0, Y^0 | X),$$

en door het fundamentele probleem van causale inferentie is de conditionele covariantie van het effect en de potentiële uitkomst zonder behandeling niet identificeerbaar.

Er kunnen toepassingen zijn waarin het beargumenteerd kan worden dat het effect onafhankelijk zou moeten zijn van de potentiële uitkomst zonder behandeling. Denk hierbij bijvoorbeeld aan een medicijn waarvan de effectiviteit afhangt van de hoeveelheid van een bepaald enzym (Post e.a., 2024). Het enzym zelf heeft geen invloed op de uitkomst zonder behandeling waardoor $Y^1 - Y^0 \perp Y^0 | X$. Dan is de conditionele variantie van het causale effect gelijk aan

$$\text{var}(Y^1 | X) - \text{var}(Y^0 | X),$$

wat onder *consistency* en *conditional exchangeability* op zijn beurt gelijk is aan

$$\text{var}(Y | A = 1, X) - \text{var}(Y | A = 0, X).$$

Beide laatstgenoemde varianties kunnen geschat worden op basis van de geobserveerde data. In (Post e.a., 2024) wordt er gedemonstreerd hoe, onder de onafhankelijkheidsaanname, het eerder genoemde *causal random forest* algoritme voor het schatten van CATEs uitgebreid kunnen worden met het addi-

oneel schatten van de conditionele variantie van het effect. Daarbij is het belangrijk om te beseffen dat wanneer voor bepaalde X de individuele effecten deterministisch zijn (en dus gelijk aan de CATE), de onafhankelijkheidsaanname ook van toepassing is. Als de conditionele variantie (betrouwbaar en) relatief groot wordt geschat kan er dus geconcludeerd worden dat er nog resterende heterogeniteit aanwezig is. Een behandelkeuze zou dan niet puur op basis van de CATE gemaakt moeten worden. Wanneer $\text{var}(Y | A = 1, X = x) \approx \text{var}(Y | A = 0, X = x)$ zal de conditionele variantie wel relatief dicht bij 0 worden geschat. Het is dan belangrijk te beseffen dat dit nog niet impliceert dat er geen resterende effect heterogeniteit is voor individuen met kenmerken X tenzij het aannemelijk is te maken dat de onafhankelijkheidsaanname geldt zodat $\text{cov}(Y^1 - Y^0, Y^0 | X) \approx 0$.

We moeten ons te allen tijde bewust blijven van de interpretatie van de correlaties die worden geschat. Het nadenken over die causale interpretatie mag niet ondersneeuwen door de complexiteit van de correlaties en inherent daaraan de complexiteit van de data-analyse.

Literatuur

- M. A. Hernán, J. Hsu en B. Healy. „A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks“. In: *CHANCE* 32.1 (2019), p. 42–49. DOI: 10.1080/09332480.2019.1579578.
- M. A. Hernán en J. M. Robins. *Causal Inference: What If*. 1st. Boca Raton, Florida: Boca Raton: Chapman & Hall/CRC, 2020. URL: <https://www.hsph.harvard.edu/miguel-herman/causal-inference-book/>.
- M. R. Kosorok en E. B. Laber. „Precision Medicine“. In: *Annual Review of Statistics and Its Application* 6.1 (2019), p. 263–286. DOI: 10.1146/annurev-statistics-030718-105251.
- R. A. J. Post e.a. „Flexible Machine Learning Estimation of Conditional Average Treatment Effects: A Blessing and a Curse“. In: *Epidemiology* 35.1 (2024). ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000001684.
- S. Wager en S. Athey. „Estimation and Inference of Heterogeneous Treatment Effects using Random Forests“. In: *Journal of the American Statistical Association* 113.523 (2018), p. 1228–1242. DOI: 10.1080/01621459.2017.1319839.

Richard Post is als universitair docent verbonden aan de groep biostatistiek van het Erasmus MC. Hij promoveerde in 2023 op zijn proefschrift *Causal Effect Heterogeneity*. Voor dit proefschrift ontving hij in 2024 de Willem R. van Zwet Award van de VWSOR. Het onderzoek van Richard richt zich voornamelijk op de identificeerbaarheid van de heterogeniteit in causale effecten en op de causale interpretatie van effectmaten in survival analysis. E-mail: r.a.j.post@erasmusmc.nl



Vertrouwen op een planningstool: valkuilen en succesfactoren

Ignace Decroix, Ann Vereecke, Karlien Vanderheyden en Shari De Baets

Het gebruik van geavanceerde software binnen de werkomgeving lijkt vanzelfsprekend anno 2024. Doch is het aanvaarden van deze tools door werknemers alles behalve dat. In dit onderzoek werd gepolst naar de valkuilen en succesfactoren om een goede relatie tussen supply chain planners en planningsoftware te bewerkstelligen.

Bedrijven schakelen gaandeweg over op een nieuwe generatie informatiesystemen, zo ook om hun supply chain management en, de focus van dit onderzoek, planning te optimaliseren. Een rijk mozaïek aan planningsoftware (e.g., *Advanced Planning Systems (APS)*) wordt ondertussen aangeboden door een veelvoud aan ontwikkelaars. Deze software ondersteunt bedrijven op korte-, middellange-, en lange termijn op het vlak van *procurement, production and distribution, en sales*. Desondanks deze systemen al een aantal jaren toegankelijk zijn en geïmplementeerd werden, staat ook hier de technologische

voortgang niet stil. De meer recente vooruitgang op het vlak van Artificiële Intelligentie (AI), bijvoorbeeld, biedt nieuwe capaciteiten aan. Zo kunnen werknemers nu veel sneller stringen in de keten opmerken en reageren, en kunnen levertijden verder geoptimaliseerd worden (Abedinnia, Glock en Schneider, 2017; Awan e.a., 2021). In vergelijking met voorgaande systemen, maken deze nieuwe, vaak ietwat 'slimme', systemen intenser gebruik van data. Daarnaast zijn zij meer dan ooit geautomatiseerd en handelen zij gedeeltelijk autonoom (Ulfert e.a., 2024). De mogelijkheden en beloftes zijn talrijk. Toch botsen deze tools op een aantal grenzen.

Het blijft vaak onduidelijk welke data nu wel of niet gebruikt worden door het systeem om tot een bepaalde beslissing of advies te komen (i.e., het *black box* probleem). Bovendien kampen deze systemen regelmatig met vooroordelen (i.e., *biases*) en worden beslissingen genomen of voorstellen gemaakt op basis van incomplete datasets. Bedrijven komen als volgt alsmear vaker tot de conclusie dat hun werknemers cruciaal blijven binnen het beslissingsproces. Het is echter hier, in de verhoopte samenwerking tussen mens en technologie, dat het



Figuur 1: Vijf dimensies om de vertrouwensband te stimuleren

schoentje spant.

Werknemers zijn namelijk verdeeld over het al dan niet gebruiken van technologie, zo ook supply chain planners. Sommigen redeneren – of vrezen – dat deze nieuwe generatie technologie bepaalde taken en misschien zelfs de volledige job zal overnemen. Maar net zoals data *bias* bevatten, zo ook mensen. Wanneer een algoritme een kleine fout maakt, zal men deze harder afstraffen dan wanneer een mens een veel grotere fout maakt. Wanneer een systeem advies geeft, zal men sneller geneigd zijn om dit advies af te breken, te negeren, of te verwerpen (Dietvorst, Simmons en Massey, 2015). Ook binnen de context van supply chain management (SCM) zijn deze vooroordelen aanwezig. Voor bedrijven is dit uiteraard een ongewenste realiteit: enerzijds wordt geld gepompt in software die niet optimaal gebruikt wordt; anderzijds verhogen eindgebruikers onbewust hun werklast.

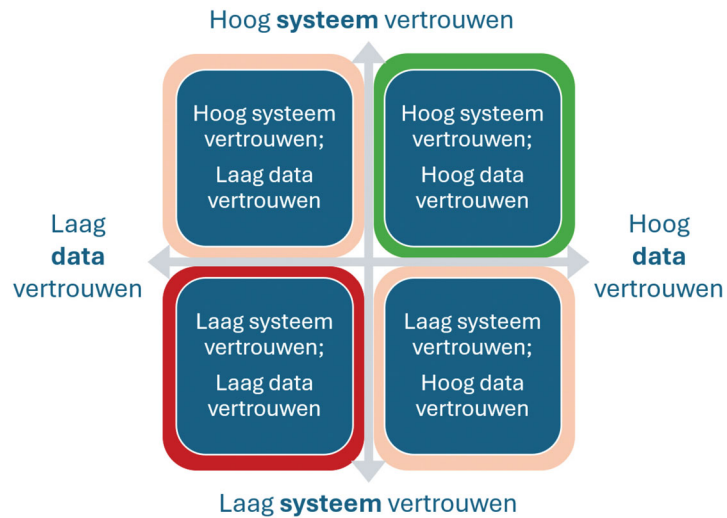
Menselijke gedragsfactoren liggen aan de basis van het succesvolle aanvaarden en gebruiken van technologie. Volgens Wickens e.a. (2021) is vertrouwen (i.e., *trust*) de meest belangrijke variabele in dit verhaal. Kortweg, waar vertrouwen ontbreekt zal zelden succesvol of optimaal omgegaan worden met technologie. SCM en planning kennen hun eigen unieke kenmerken en uitdagingen die mogelijk palend kunnen zijn voor een optimale relatie tussen de supply chain planners en het planningssysteem.

Om dit fenomeen beter te begrijpen, namen we interviews af met 26 systeem-*stakeholders*. Zij waren actief in oa. België, Zwitserland, Ierland, Nieuw-Zeeland, en Puerto Rico, met verschillende functies (e.g., *master data planner, master and detail planner, director end-to-end product planning, en end-to-end product planning specialist*). We identificeerden hierbij verscheidene grote 'dimensies' om de vertrouwensband te stimuleren (zie Figuur 1):

- het profiel van de eindgebruiker;
- de data in het systeem;
- het functioneren van het systeem;
- de institutionele ondersteuning;
- de betrouwbaarheid van betrokken partijen.

Het profiel van de eindgebruiker

De eindgebruiker is de voornaamste stakeholder in dit gehele verhaal. Van zodra het planningssysteem werd aangekondigd, bleken zij reeds bepaalde verwachtingen te verwerven over wat het systeem al dan niet zou kunnen. Deze verwachtingen vertaalden zich vaak naar angstgevoelens. Zo vreesde men dat, op basis van wat verteld en getoond werd, het systeem niet zou zijn zoals gewenst. Desondanks bleef men ook nuchter: de planners waren er zich van bewust dat bepaalde onderliggende redenen aan de basis lagen voor elke gemaakte keuze.



Figuur 2: Het belang van systeem vertrouwen en data vertrouwen

Vanaf het moment dat het systeem in gebruik werd genomen, bleven zij echter alert over de context waarin het systeem handelde. Denk bijvoorbeeld aan de sector en het type producten. De planners gooiden hierbij niet zomaar hun eigen kennis en ervaring overboord. Zo werd geopteerd voor een eerder voorzichtige, soms zelf sceptische, houding ten opzichte van het planningssysteem: het systeem hielp hen weliswaar bij het uitvoeren van taken, maar bleek zeker niet in staat om elke taak op zich te nemen, noch had het de vereiste contextgevoeligheid.

De data in het systeem en het functioneren van het systeem

Kijken we vervolgens naar het planningssysteem zelf, dan maakten we een interessante observatie: de eindgebruikers reflecteerden voortdurend op ofwel hun vertrouwen in het systeem (en de daaraan gelinkte aspecten, zoals bv. fouten) ofwel hun vertrouwen in de data (en de daaraan gelinkte aspecten, zoals bijvoorbeeld kwaliteit). Zetten we even de bril van de planners op, dan kan het systeem niet liegen of fouten maken. Het systeem genereert louter haar output op basis van de input die het krijgt (i.e., *garbage in, garbage out*). Met andere woorden,

de eindgebruikers koppelden continue hun data-perceptie los van hun systeem-perceptie tijdens de interactie: *"ik heb veel stress omdat mijn data niet correct zijn, maar dat heeft niks te maken met het systeem. Zelfs al zou je Excel gebruiken, als je ook daar niet de juiste waarden hebt, zullen er fouten zijn. Dit, opnieuw, heeft niks te maken met het systeem"*. De planners legden namelijk de verantwoordelijkheid bij hun collega's, *"het is altijd omwille van menselijke input dat het systeem een fout maakt. Maar ik vertrouw het systeem op zich wel"*. Houden we deze observatie vast, dan komen we tot de volgende vier situaties en kwadrant (zie Figuur 2):

1. De eindgebruiker vertrouwt het systeem, maar wantrouwt de data
2. De eindgebruiker wantrouwt het systeem, maar vertrouwt de data
3. De eindgebruiker vertrouwt het systeem, en vertrouwt de data
4. De eindgebruiker wantrouwt het systeem, en wantrouwt de data

Kortweg, een bedrijf lijkt te zullen moeten streven naar een ideale situatie waar het vertrouwen in zowel het systeem (o.a. capaciteiten) als in de data (o.a. kwaliteit) hoog genoeg is. Om dit vertrouwen te verhogen schaalden de geïnterviewden zich quasi

unaniem achter één woord: valideren. Enerzijds willen eindgebruikers garantie hebben dat zij met kwaliteitsvolle data werken. Indien dit onduidelijk blijft, zullen zij zelf stappen ondernemen om de data te valideren. Daarnaast willen zij ook garantie hebben over de functionaliteiten van het systeem, *"Indien het systeem op een juiste manier geprogrammeerd werd, zal ik het systeem vertrouwen, maar we moeten dit zorgvuldig valideren om de betrouwbaarheid te weten"*. Ook hier bleken sommige planners het heft in eigen handen te nemen en gingen zelf over tot (pogingen tot) validering. Dit deden zij, bijvoorbeeld, door in parallel in het oude systeem te blijven werken en te bekijken of de output al dan niet identiek bleek. Uiteraard verhoogde dit hun werklast aanzienlijk daar zij nu in meerdere systemen werkten om dat garantiegevoel te verkrijgen.

De institutionele ondersteuning en de betrouwbaarheid van betrokken partijen

Los van het bovengenoemde, kaartten interviewees het belang van training, steun, hun betrokkenheid aan, alsook de betrouwbaarheid van andere partijen. Een planningssysteem ontwikkelen en uitrollen vereist de kunde van verschillende spelers, zoals het management, key users, consultants, en de ontwikkelaar. De planners wikten en wogen regelmatig de betrouwbaarheid van deze partijen en spraken hierbij over de invloed van deze partijen op hun vertrouwen in het planningssystemen. Een grote verantwoordelijkheid wordt gelegd bij de ontwikkelaar, *"indien ik kijk naar het afgelopen jaar, dan vertrouw ik de developer's manier van werken minder dan het systeem. Ze hebben oude bugs geherintroduceerd. Nu wordt het gevaarlijk want als zij terug een upgrade doorvoeren zou het kunnen dat er terug oude bugs aanwezig zijn"*. Daarnaast reflecteerden de planners ook op het belang van de betrouwbaarheid van andere eindgebruikers – en hoe ook zij het vertrouwen in zowel systeem als data beïnvloedden op zowel negatieve manier, *"bij bepaalde collega's weet je dat je meer correcte data zult krijgen dan bij andere collega's, en dat heeft een impact op hoeveel vertrouwen je hebt"*, als positieve manier, *"aangezien de data die je gebruikt ingevoerd zijn door collega's, elk van hun met hun eigen specialiteit, vertrouw je op wat in het systeem zit"*.

De relatie tussen werknemers en technologie versterken is geen sinecure. Vertrouwen speelt hierbij een belangrijke rol. In dit onderzoek brachten we verschillende factoren in kaart die van belang zijn

in een supply chain planning context. Dit laat toe aan bedrijven om actief in te spelen op verschillende succesfactoren en valkuilen. Een extra uitdaging schuilt echter in de observatie dat planners gescheiden redeneren over hun vertrouwen in het systeem en in de data. Indien u een goede relatie beoogt, zullen beide uitdagingen vakkundig aangepakt moeten worden. Zo kan u aan de hand van heldere communicatie tijdig inspelen op de verwachtingen van eindgebruikers, hun extra duidelijkheid geven aan de hand van transparante validering, en dient men stil te staan bij de betrouwbaarheid van elke betrokken partij.

Literatuur

H. Abedinnia, C. H. Glock en M. D. Schneider. „Machine scheduling in Production: A content Analysis“. In: *Applied Mathematical Modelling* 50 (2017), p. 279–299.

U. Awan e.a. „Artificial intelligence for supply chain success in the era of data analytics“. In: *The fourth industrial revolution: Implementation of artificial intelligence for growing business success* (2021), p. 3–21.

B. J. Dietvorst, J. P. Simmons en C. Massey. „Algorithm aversion: people erroneously avoid algorithms after seeing them err.“ In: *Journal of experimental psychology: General* 144.1 (2015), p. 114.

A.-S. Ulfert e.a. „Are we ahead of the trend or just following? The role of work and organizational psychology in shaping emerging technologies at work“. In: *European Journal of Work and Organizational Psychology* 33.2 (2024), p. 120–129.

C. D. Wickens e.a. *Engineering psychology and human performance*. Routledge, 2021.

Ignace Decroix is doctoraal onderzoeker aan Vlerick Business School en Universiteit Gent, België. Zijn onderzoek focust zich op de interactie tussen werknemers en technologie, zoals IT-systemen, AI, en drones.
E-mail: ignace.decroix@vlerick.com.

Ann Vereecke is professor in operations management en partner aan Vlerick Business School en professor aan Universiteit Gent, België. Ze is Directeur van het Onderzoekscentrum Smart People in the Smart Supply Chain.
E-mail: ann.vereecke@vlerick.com.

Karliën Vanderheyden is professor in organisatiegedrag en partner aan Vlerick Business School. Ze is Directeur van het Onderzoekscentrum Smart People in the Smart Supply Chain.
E-mail: karlien.vanderheyden@vlerick.com.

Shari De Baets is doctor in de toegepaste economische wetenschappen aan de Universiteit Gent en is universitair Docent aan de Open Universiteit Nederland en gastprofessor aan de KULeuven, België. Ze is redacteur van Foresight voor de judgment column.
E-mail: Email:shari.debaets@ou.nl.



Om en om serveren bij squash is eerlijker

Miriam Loois

Wie bij squash een rally wint mag het punt erna serveren. Hierdoor heeft de toss invloed op de kans om een game te winnen. Het maakt squash niet eerlijk. Twee even sterke spelers hebben geen gelijke kans om te winnen. De puntentelling kan eenvoudig eerlijk gemaakt worden door om en om te serveren, zoals in de tiebreak bij tennis.

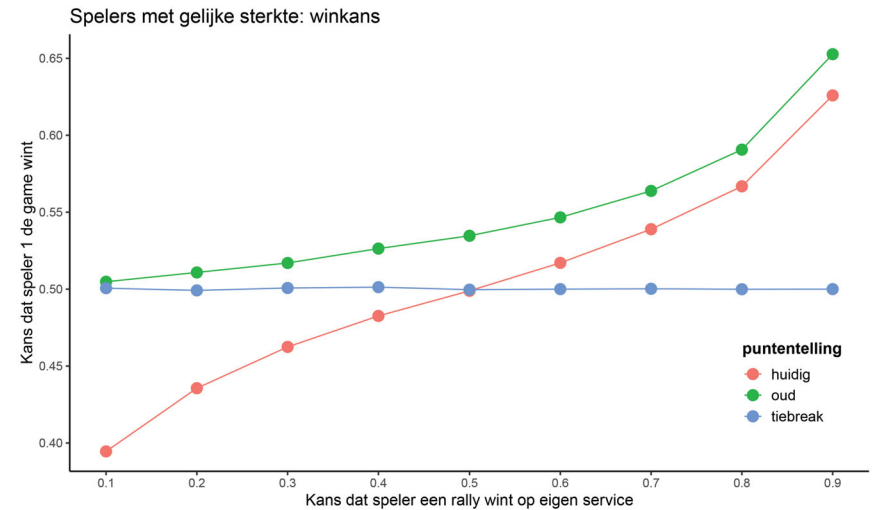
Bij topsquashers levert serveren geen uitgesproken voor- of nadeel op, in tegenstelling tot bijvoorbeeld tennis. Op lager niveau, bijvoorbeeld onder recreanten, zien we wel regelmatig spelers die een reeks punten pakken met een sterke service. In dit artikel onderzoeken we wat de invloed van servicedominantie (de mate waarin de speler die serveert de overhand heeft in de rally) is op de kans om een game te winnen. We vergelijken meerdere puntentellingen. Ook onderzoeken we de hypothese dat een sterke service minder belangrijk is geworden sinds de overstap op de huidige puntentelling is

gemaakt, 15 jaar geleden.

Puntentelling

Sinds 2009 wordt bij squash het zogenaamde par-11 (point a rally to 11) systeem gebruikt. Elke gewonnen rally levert een punt op. Als je een punt wint mag je de volgende rally serveren. Bij verlies gaat de opslag naar de tegenstander. Wie als eerste 11 punten heeft, met minimaal 2 punten verschil, heeft de game gewonnen. Mogelijke eindstanden zijn dus 11-3, 11-9, 13-11 of 20-18. Wie 3 games heeft gewonnen wint de wedstrijd. De toss bepaalt wie er in de eerste game mag beginnen met serveren. Voor 2009 kon je alleen een punt maken als je zelf serveerde, en ging een game tot 9, met maar 1 punt verschil. Bij een score van 8-8 mocht de ontvanger kiezen of de game tot 9 of tot 10 ging. Spelers met een sterke service mogen dus in beide systemen blijven serveren en kunnen zo veel punten achter elkaar maken. Dan is het voordelig om te mogen beginnen met serveren.

Bij (beach)volleybal wordt een vergelijkbare puntentelling gehanteerd als bij squash. Ook hier gaat



Figuur 1: Kans dat speler 1, de speler die begint met serveren, de game wint, afhankelijk van de servicedominantie. De huidige en de oude puntentelling zijn niet eerlijk, om en om serveren wel.

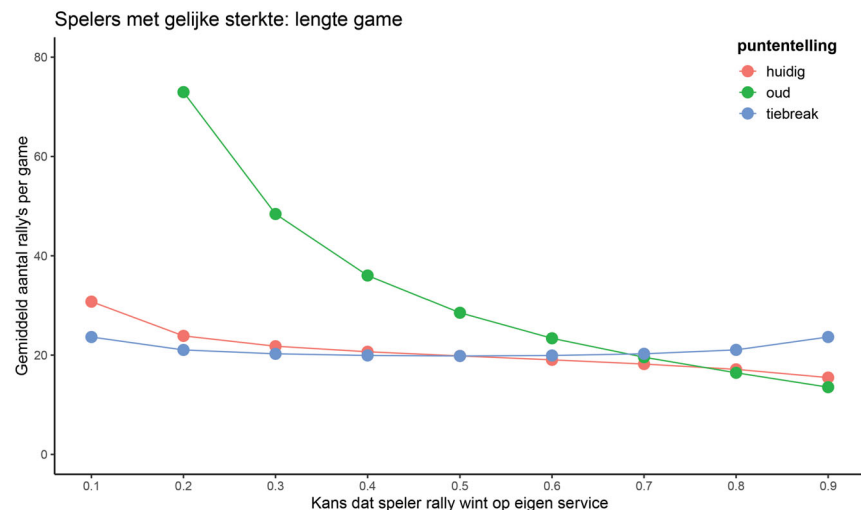
de service naar het team dat het laatste punt heeft gewonnen. Serveren leidt daar echter meestal tot puntverlies. Hier zul je dus eerder zien dat de teams om en om een punt winnen. Toch heeft de toss hier ook invloed. Frits Spijksma heeft in Volley Techno laten zien dat dit niet eerlijk is (Spijksma, 2016). Een kort gedachtenexperiment. Bij beachvolleybal gaat de beslissende set tot 15. Team 1 begint met serveren. Als beide teams steeds het punt op de eigen service verliezen, is het op een gegeven moment 14-13 in het voordeel van team 2. Team 2 is dan aan de beurt om te serveren. Als het één keer een punt wint op de eigen service heeft het de set gewonnen. Stel, dit gebeurt niet. Het is 14-14, Team 1 serveert weer en het houdt de eigen service. Dan is het 15-14. Ze moeten dan nog een keer de eigen service houden om de set te winnen. Dit kan Nederland zomaar de wereldtitel hebben gekost in 2015. Nederland begon toen met serveren in de beslissende derde set, die uiteindelijk met 22-20 naar Brazilië ging. Als beide teams even sterk zijn, is de winkans niet 50-50. Er is ook een eenvoudige oplossing: serveer om en om.

Toss heeft invloed

Data over servicedominantie bij squash zijn nauwelijks te vinden. Ondanks gebrek aan data is dus wel

duidelijk dat servicedominantie invloed heeft op de winkans. Als serveren een nadeel is, dan zal degene die begint met serveren een keer extra een punt moeten winnen op de eigen service. Als serveren een voordeel is, dan kan degene die mag beginnen met serveren een reeks aan punten maken, tot de service overgaat naar de tegenstander. Maar hoe groot is het effect? Er zijn enkele papers waarin wordt beschreven hoe winkansen exact kunnen worden berekend. We gebruiken in ons onderzoek een simulatie, omdat aan een simulatie eenvoudig nieuwe tellings-systematieken kunnen worden toegevoegd. Ook kun je eenvoudig andere statistieken zoals de lengte van een game bijhouden. Elke spelsituatie wordt een miljoen keer gesimuleerd in R. We gaan er vanuit dat in de oude puntentelling de ontvanger altijd kiest voor een game tot 9. Speler 1 begint altijd met serveren. De kans om de rally te winnen hangt alleen af van wie serveert, en uitkomsten van de rally's zijn onafhankelijk van elkaar verondersteld. We vergelijken drie systemen: De huidige en de oude puntentelling, met daarnaast een systeem vergelijkbaar met de tiebreak bij tennis. Hier serveert speler 1 eerst één keer. Daarna serveren ze om en om steeds twee keer (vanaf links en vanaf rechts), tot 11 met minimaal 2 punten verschil.

Eerst kijken we naar twee spelers die even sterk zijn. De kans om een punt te maken op de eigen



Figuur 2: Gemiddelde lengte van een game in de drie systematieken, beide spelers zijn even sterk.

service varieert tussen de 10% en 90%, zie figuur 1. Als er om en om wordt gereserveerd blijkt het systeem eerlijker te zijn, en is de winkans altijd 50%. In het huidige systeem zien we wat we verwachten.

Als serveren nadelig is, is het nadelig om te beginnen en vice versa. Zelfs bij kleine afwijkingen van een neutrale service verschuiven de winkansen al met enkele procenten. In de wat extremere gevallen verschuift de kans van 50-50 naar 60-40. Het resultaat van het oude systeem (de groene lijn in figuur 1) is op het eerste gezicht vreemd. Bij een lage kans om de service te houden nadert de winkans de 50%. Dit komt omdat de service heel vaak heen en weer zal gaan, en er maar af en toe een punt gescoord wordt op de eigen service. Het effect van wie er begint is daardoor verwaarloosbaar, en beide spelers moeten even vaak de eigen service houden om de game te winnen. Omdat je alleen een punt kunt maken op je eigen service, zal het altijd een

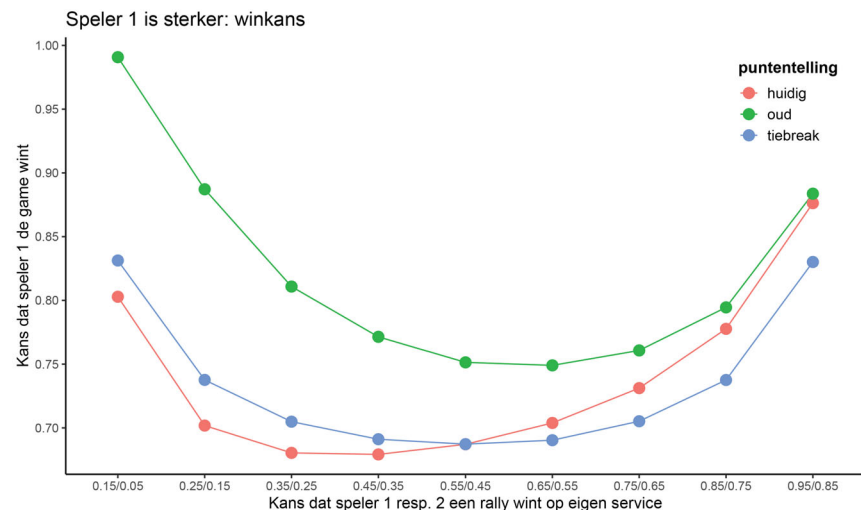
voordeel opleveren om te beginnen. De winkans loopt op van net iets boven de 50% naar ongeveer 65% bij een kans van 90% om de rally te winnen op eigen service. De nieuwe puntentelling heeft het professionele squash, waarin serveren geen uitgesproken voordeel oplevert, dus wel degelijk eerlijker gemaakt!

In figuur 2 zien we hoe lang een game gemiddeld duurt. Logischerwijs duurt een game in het oude systeem erg lang als de service steeds heen en weer gaat. In het huidige systeem en bij een tiebreak tot 11 duurt een game gemiddeld zo'n 20 rally's. De wedstrijdduur in een tiebreak is iets stabielier dan in het huidige systeem.

Wat nu als speler 1 sterker is dan speler 2? Wiskundig gezien hoop je dat de beste speler een zo groot mogelijke kans heeft om te winnen. Je kunt een wedstrijd zien als een soort hypothesetoets

Tabel 1: De kans dat een speler met een sterkere service 1 game wint. In het oude systeem leverde dit iets meer voordeel op.

Kans dat speler 1 game wint	Winkans eigen service 0,5/0,5	Winkans eigen service 0,6/0,5	Toename
Huidige systeem	50,0%	69,4%	19,4%
Oude systeem	53,5%	74,8%	21,3%



Figuur 3: Kans dat speler 1, de speler die begint met serveren en de sterkere speler is, de game wint, afhankelijk van de servicedominantie. De kans om de rally te winnen op eigen service van speler 1 resp. 2 varieert van 0,15/0,05 tot 0,95/0,85.

waarbij de kans op een foute conclusie zo klein mogelijk moet zijn. De toeschouwer kan hier anders over denken. Het is een stuk spannender als degene die achter staat nog kans heeft om terug te komen. In figuur 3 zien we wat er gebeurt als speler 1 steeds 10 procentpunt meer kans heeft om de rally te winnen op de eigen service. De betere speler was in het oude systeem beter af. De groene lijn ligt altijd boven de andere twee lijnen. De kans dat de betere speler ook als winnaar uit de bus komt is dus in het oude systeem het grootst. Niet helemaal een eerlijke vergelijking, want de wedstrijd duurde vaak ook langer. We zien ook dat de betere speler beter af is als serveren heel nadelig of juist heel voordelig is. Dit komt omdat het verschil in absolute kansen steeds 10% is. Maar bij een kans van 5% om 15% heeft speler 1 wel drie maal zoveel kans om een punt te winnen op de eigen service, dus is het verschil relatief groter. De mindere speler hoeft de handdoek zeker niet bij voorbaat in de ring te gooien. In het huidige systeem en bij een tiebreak heeft hij typisch nog steeds 25% tot 30% kans om de game te winnen.

Vervolgens gaat speler 1 flink trainen op zijn service, en stijgen zijn kansen naar 60%. In tabel 1 zien we dat zijn kans om de game te winnen in het oude systeem iets meer stijgt, maar het ontloopt elkaar niet veel. Tijd om de servicetraining onder het stof vandaan te halen!

Conclusie

We stellen voor om de puntentelling nogmaals aan te passen. Door om en om te serveren is squash niet alleen eerlijker voor de toppers, maar ook voor alle liefhebbers. Je zult moeten zorgen dat je goed kunt serveren en ontvangen, vanaf links en vanaf rechts. De wedstrijdduur is ook nog eens stabielier. Wie vaker wil winnen doet er goed aan zijn service te trainen. O ja, en squash is echt veel leuker dan padel.

Literatuur

F. Spijksma. „Eerlijke kans om beach volleybalfinale te winnen: om en om serveren in de derde set, please!“ In: *Volley Techno 2* (2016).

Miriam Loois is docent Toegepaste Wiskunde aan de Hogeschool van Amsterdam en schrijft artikelen over sport, wiskunde en statistiek. Ze is lid van Squash Stats, een internationale groep van liefhebbers van squash en data, bestaande uit vijf mensen. Wie sluit aan? www.miriamloois.nl.

Trainen op de service

En dan de vraag of de service inderdaad belangrijker was in het oude systeem. We vergelijken twee situaties. In het uitgangsscenario hebben beide spelers 50% kans om het punt op de eigen service te winnen.



Bron: Pixabay

De munt is vol verrassingen

Henk Tijms

De meeste mensen hebben over toeval een voorin- genomen standpunt dat niet met de werkelijkheid overeenstemt en kunnen zich geen goede voorstelling maken van echte toevalsrijen. Databestanden die random tot stand gekomen zijn, hebben vaak onverwachte eigenschappen die tegen de intuïtie indruisen. Deze eigenschappen zouden gebruikt kunnen worden om te testen of aan databestanden geknoeid is wanneer het vermoeden daartoe bestaat. Beschouw eens de volgende twee lijsten van 1'en en 0'en met 1 is kop en 0 is munt die de resultaten zouden moeten zijn van 100 worpen met een zuivere munt. Welke lijst is naar uw mening niet door toeval tot stand gekomen maar is gefabriceerd?

1 1 0 0 1 0 0 1 0 1
 1 0 0 1 0 0 0 1 1 0
 1 0 1 0 0 1 1 0 1 0
 0 1 0 1 0 1 1 0 1 1
 0 0 1 1 0 1 1 1 0 1
 0 0 1 0 0 1 1 0 1 0
 0 1 1 0 1 0 0 1 1 0
 1 0 1 1 0 0 1 1 1 0
 0 1 0 1 0 1 0 0 0 1
 0 1 0 1 0 1 0 1 0 1

of

1 1 1 0 0 0 1 1 1 0
 1 0 1 1 1 1 1 1 0 1
 0 0 0 1 1 0 0 1 1 0
 1 0 1 0 0 0 1 1 0 1
 0 0 1 1 1 0 1 0 0 0
 0 1 0 1 1 1 0 1 1 0
 0 1 1 1 0 1 1 0 0 1
 1 1 1 1 1 0 1 1 0 1
 0 1 1 1 0 0 0 0 0 0
 0 0 1 1 0 1 1 1 0 1

De Amerikaanse wiskundige Ted Hill had de gewoonte om bij zijn eerste college kansrekening de studenten te vragen thuis een experiment uit te voeren en de resultaten mee te nemen naar het volgende college. De studenten werd gevraagd de 200 opeenvolgende uitkomsten van een daadwerkelijk opgeworpen zuivere munt te noteren of net te doen alsof de munt 200 keer opgegooid was en 200 opeenvolgende uitkomsten op te schrijven van de zogenaamd uitgevoerde muntworpen. Tot grote verbazing van de studenten haalde Ted Hill vrijwel altijd de gefingeerde resultaten er feilloos uit. Een snelle blik volstond om de inzendingen van de studenten

Runs van alleen kop of alleen munt bij worpen met een munt

n/r	3	4	5	6	7	8	9	10
10	0,826	0,465	0,217	0,094	0,039	0,016	0,006	0,002
25	0,993	0,848	0,550	0,300	0,151	0,073	0,035	0,017
50	1,000	0,981	0,821	0,544	0,309	0,162	0,082	0,041
75	1,000	0,998	0,929	0,703	0,438	0,242	0,126	0,064
100	1,000	1,000	0,972	0,807	0,542	0,315	0,169	0,087
150	1,000	1,000	0,996	0,918	0,697	0,440	0,247	0,131
200	1,000	1,000	0,999	0,965	0,799	0,542	0,318	0,172

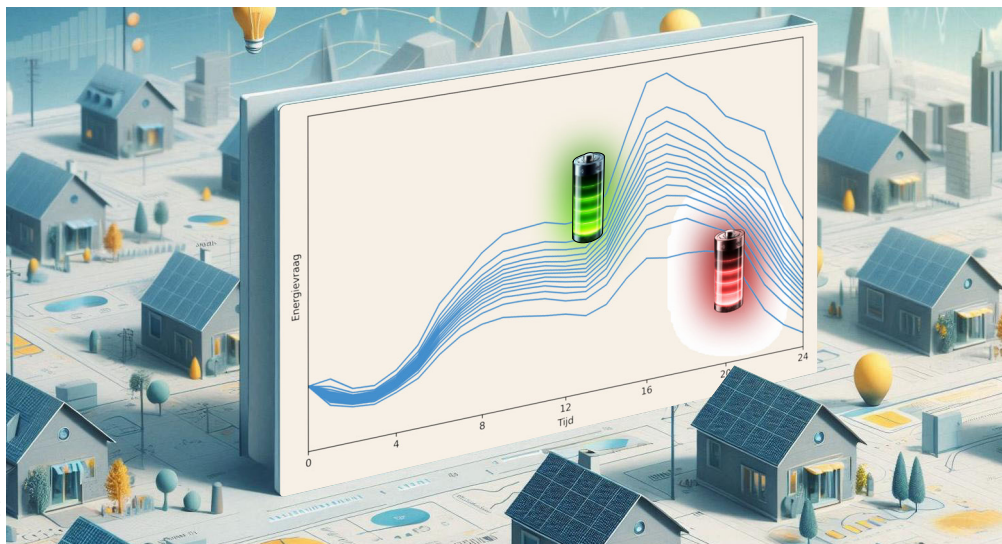
op waarde te schatten. Vanzelfsprekend was het doel van Ted Hill de studenten aan den lijve te laten ervaren dat verreweg de meeste mensen zich geen goede voorstelling kunnen maken van echte toevalsrijen. In een interview met de New York Times legde Ted Hill uit dat het geheim van zijn aanpak heel simpel was: elk resultaat dat niet een ononderbroken reeks van tenminste zes keer kop of tenminste zes keer munt bevatte, werd door hem als gefingeerd bestempeld. De kans op een dergelijke run heeft de verrassend hoge waarde van 0,965. De meeste mensen onderschatten de lengte van de langste run in een reeks van worpen met een munt. Als vuistregel is $\log_2(\frac{1}{2}n) + 1$ een redelijke schatting voor de lengte van de langste run van alleen kop of alleen munt in n worpen als n voldoende groot is (de vuistregel is $\log_2(\frac{1}{2}n)$ voor runs met alleen kop). Deze vuistregel is afkomstig van Mark Schilling, "The surprising predictability of long runs", Mathematics Magazine, Vol. 85 (2012), 141-149. De kans op een run van tenminste r keer kop of tenminste r keer munt in n worpen met een zuivere munt kan worden berekend met behulp van een absorberende Markov keten die als toestand heeft het aantal keren dat achtereenvolgende muntworpen tot dezelfde uitkomst geleid hebben. In de tabel geven we deze kans voor een aantal waarden van r en n . Een approximatie voor de kans op een run van tenminste r keer kop of tenminste r keer munt in n worpen met een zuivere munt is $1 - e^{-2n(1-p)^r}$ voor n voldoende groot, waarbij $p = 0,5$. Dit is een praktisch bruikbare approximatie zoals blijkt uit vergelijking met de exacte resultaten in de tabel.

Ook in roulette kunnen lange runs met alleen de uitkomst rood of alleen de uitkomst zwart optreden. In dit verband is een aardige anecdote te geven. Aan het einde van de 19de eeuw publiceerde de krant Le Monaco regelmatig de reeksen van getallen die gevallen waren in uitvoeringen van roulette in het

casino van Monte Carlo. De beroemde statisticus Karl Pearson (1857-1936) bestudeerde deze gegevens. Hij vond dat rood en zwart een vergelijkbaar aantal keren voorkwamen, maar ook vond hij dat de lengtes van runs met alleen rood of alleen zwart veel korter waren dan hij had verwacht. Wat was de oorzaak? Nou, het bleek dat de journalisten de roulette getallen gewoon verzonden hadden aan de bar van het casino. Ze dachten dat niemand het zou merken.

In het casino van Monte Carlo viel op 18 augustus 1913 het roulette balletje 26 keer achter elkaar op zwart. Hoe uitzonderlijk is een dergelijke gebeurtenis over een langere periode bezien? Een vuistregel is dat een redelijke waarde voor de lengte van de langste run van alleen rood of alleen zwart in n draaiingen van het roulette wiel gegeven wordt door $\log_{1/p}(n(1-p)) + 1$ voor n groot, waarbij $p = \frac{18}{37}$ in Europees roulette. Dit houdt in dat in $n = 100$ miljoen draaiingen van het roulette wiel de kansverdeling van de lengte van de langste run geconcentreerd is rond de waarde 26, precies de waarde die optrad in 1913 in het Monte Carlo casino. Een goede benadering voor de kans dat in n draaiingen van het roulette wiel een run met lengte r of meer optreedt is $1 - e^{-2n(1-p)^r}$ voor n groot. Voor $r = 26$ heeft deze benadering de waarden 0,528 en 0,950 voor $n = 10^8$ en $n = 4 \times 10^8$. In het licht van deze berekeningen en het feit dat er talloze casino's op de wereld zijn, is het dus bepaald niet onwaarschijnlijk als ergens een keer een run van lengte 26 of meer optreedt. Het record behoort een Amerikaans casino toe waar in 1943 het roulette balletje 32 keer achter elkaar op rood viel.

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Dit najaar verscheen bij Epsilon Uitgaven zijn boek "Kansrekening in Vogelvlucht", dat geschreven is voor studenten aan hogeschool en universiteit. E-mail: h.c.tijms@xs4all.nl



Kansverdelingsvoorspellingen: De manier om je te wapenen tegen onzekere parameters bij optimalisatie?

Pieter Knops

In dit artikel bespreken we hoe je, met behulp van clustering, de energievraag kan voorspellen. Deze voorspellingen gebruiken we om met batterijen pieken in het elektriciteitsnetwerk plat te slaan. In plaats van een deterministische voorspelling, gebruiken we een tijdreeks aan kansverdelingsvoorspellingen. We introduceren het optimalisatieprobleem van piekverschuivingen en beschrijven strategieën om een optimaal batterijoplaadplan te berekenen. De strategie gebaseerd op de kansverdelingsvoorspellingen lijkt het meest robuust te zijn tegen afwijkingen van de voorspelling.

Mismatch tussen opwek en gebruik

Ons energienetwerk heeft het lastig met de groeiende energievraag en de lokale energieproductie, zoals zonne-energie. De bulk van de hernieuwbare energieproductie is doorgaans niet wanneer het gebruik het hoogst is (Nottrott, Kleissl en Washom, 2013), zeker niet in woonwijken, terwijl opwek en gebruik altijd in balans moeten zijn. Als dit niet het geval is, raakt het elektriciteitsnetwerk overbelast met black-outs als gevolg. Het energiegebruik is lastig te voorspellen, omdat het weer en het gedrag van consumenten daar een sterk effect op hebben. Daarbij is ons energienetwerk ontworpen voor centrale, planbare productie, maar hernieuwbare energie maakt nu beide kanten van de balans lastig te voorspellen.

Het gat in tijd tussen productie en consumptie zou gedeeltelijk opgelost kunnen worden met batterijen. Ons doel is om met batterijen de piek zo laag mogelijk te houden door op te laden wanneer het energiegebruik het laagst is. Dit heet bandbreedte minimalisatie. Er zijn simpele planningsmodellen, zoals de ONOFF strategie, waarin de batterij gebaseerd op de tijd op volle snelheid gaat opladen of ontladen. Een uitgebreidere versie, genaamd Real-Time, laadt af gebaseerd op het huidige verbruik, maar dit heeft als risico dat even later het verbruik hoger is dan verwacht, en de batterij te vroeg leeg is. Het plannen van de batterij gebaseerd op een voorspelling met continue bijsturing op basis van het huidige verbruik, een zogenaamde “rolling horizon” strategie werkt het beste (Nottrott, Kleissl en Washom, 2013).

Kansverdelingsvoorspellingen

Plannen begint dus met het hebben van een voorspelling. Gerelateerd onderzoek wees er op dat een hogere nauwkeurigheid behaald kan worden wanneer individuele voorspellingen opgeteld worden om een voorspelling te krijgen per wijk of stad (Bañales, Dormido en Duro, 2021). Om de nauwkeurigheid verder te verhogen, is het mogelijk om huishoudens te clusteren op vergelijkbaar energiegebruik en één machine learning model per cluster te trainen (Bañales, Dormido en Duro, 2021). Het kost namelijk veel rekenkracht om voor ieder huishouden een eigen model te trainen. Echter, het gebruiken van slechts één model voor alle huishoudens maakt het trainen weer lastig, omdat dit ene model dan van alle huishoudens het gedrag moet leren.

In tegenstelling tot de genoemde onderzoeken voorspellen we niet een deterministische tijdreeks, met één waarde per uur, maar een tijdreeks aan kansverdelingen, per uur en 24 uur vooruit. Er zijn enkele onderzoeken waarin een onderliggende kansverdeling gesimuleerd wordt door meerdere losse voorspellingen te doen. Bijvoorbeeld door met verschillende temperatuurscenario's (Xie en Hong, 2016) verschillende voorspellingen te maken, of door het voorspellen van kwartielen waarin waarden voorspeld worden waar je met een 25% waarschijnlijkheid onder of boven zit (Xu, Hu en Fan, 2022). Een andere mogelijkheid is om deterministische voorspelling te gebruiken samen met een analyse van de nauwkeurigheid daarvan om betrouwbaarheidsintervallen voor het toekomstig energiegebruik te berekenen. Ander onderzoek (Kodaira, Jung en Han, 2019) koppelt dit aan het plannen van batterijen.

Ze benoemen dat het direct voorspellen van een betrouwbaarheidsinterval lastig is, omdat het energiegebruik geen normaalverdeling volgt.

Om dit probleem op te lossen, en toch een directe kansverdelingsvoorspelling te kunnen maken, gebruiken we een “Exponentially Modified Gaussian” (EMG) kansverdeling, wat neerkomt op een normaalverdeling plus een exponentiële verdeling. De gedachte is dat het piekloze gebruik met een normaalverdeling voorspeld kan worden, en pieken in gebruik (die bij huishoudens veel voorkomen, bijvoorbeeld door het gebruik van een waterkoker) beter voorspeld kunnen worden met een exponentiële verdeling. De EMG-kansverdeling heeft als parameters μ , σ en λ en heeft als verwachtingswaarde $\mu + \frac{1}{\lambda}$. Het voorspellen van een EMG-kansverdeling komt neer op het berekenen van de juiste waarden voor de drie parameters, die dus de output zullen zijn van ons model.

De gebruikte data

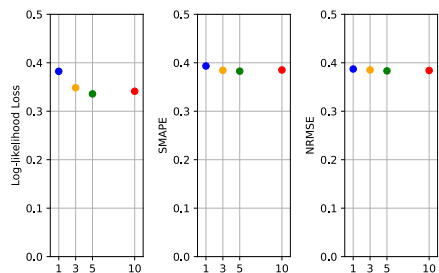
We gebruiken van 1772 Nederlandse huishoudens bijna twee jaar aan gebruiksdata, met waardes per uur. Dit waren echter niet de meest ‘schone’ data; bij alle tijdreeksen ontbraken datapunten, met soms gaten van meer dan een week. Na opschonen hebben we 500 willekeurig tijdreeksen kunnen selecteren.

Voor weerdata hebben we historische voorspellingen verkregen van het KNMI tot 48 uur vooruit. Het KNMI werkt met meerdere voorspellingen voor dezelfde tijd, een zogenaamd “ensemble” van voorspellingen. Hiervan gebruikten we het gemiddelde en standaardafwijking als input voor ons model, zodat dit weet hoe zeker de weersvoorspellingen zijn.

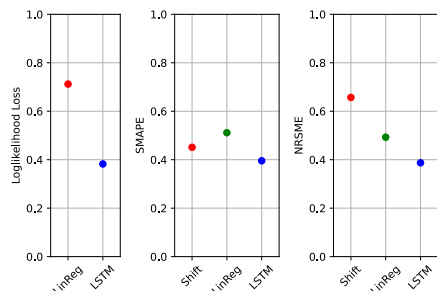
Het voorspellen van de energievraag met kansverdelingen

Als eerste stap hebben we consumenten ingedeeld in groepen op basis van gebruik. Hiervoor hebben we gekeken naar hoe het energiegebruik verschilt over het jaar, over de week, en hebben we gebruik gemaakt van twee “representatieve dagen”. Deze twee dagen beschrijven alle andere dagen van deze consument het beste. Op deze manier gaan we van 8760 datapunten (1 jaar aan data) naar 140 datapunten, waarmee we ‘de vloek van dimensionaliteit’ ook uit de weg gaan.

Hierna hebben we meerdere “Long-Short Term Memory” (LSTM) neurale netwerken gemaakt, één voor elk cluster, en is het trainen gestart. Om de



Figuur 1: Nauwkeurigheid van de LSTM modellen voor verschillende cluster hoeveelheden.



Figuur 2: Nauwkeurigheid van de basismodellen en het LSTM model gebaseerd op vijf clusters.

resultaten te kunnen vergelijken, hebben we twee basismodellen gemaakt. De eerste, TimeShift, kopieert het gebruik van de vorige week als voorspelling voor de volgende week, een veelgebruikt basismodel. Als tweede basismodel gebruiken we Lineaire Regressie (LR) die ook een kansverdeling voorspelt.

Om de voorspellingen te beoordelen, gebruiken we drie metrieken. De Symmetric Mean Average Percentage Error (SMAPE) geeft weer hoeveel procent de voorspelling gemiddeld genomen af zit van het echte gebruik. Voor de Normalized Root Mean Squared Error (NRMSE) score is het goed voorspellen van pieken belangrijker. De Loglikelihood Loss score geeft weer hoe goed de kansverdeling de echte waarden weerspiegelt. Aangezien TimeShift geen kansverdeling produceert, kan de Loglikelihood Loss hiervoor niet worden uitgerekend.

Het nemen van vijf clusters bleek het beste voor alle verschillende scores (zie Figuur 1). TimeShift komt gemiddeld genomen dicht bij het echte gebruik dan LR, te zien aan de SMAPE scores. Echter heeft TimeShift meer moeite met pieken, die mogelijk beter voorspeld worden door de kansverdeling van LR, te zien aan de NRMSE scores. Onze LSTM bleek in alle gevallen het beste model (zie Figuur 2).

Pieken platslaan met een batterij

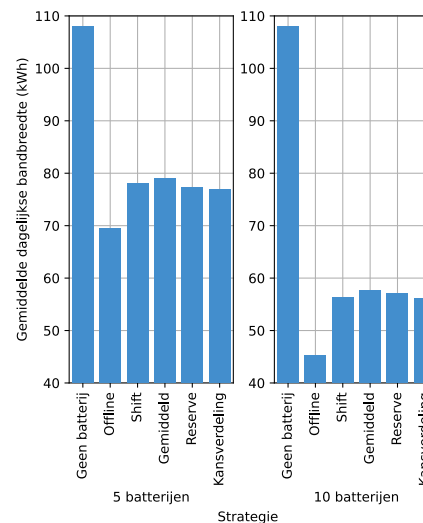
Voor de bandbreedte minimalisatie is het belangrijk om rekening te houden met de efficiëntie, capaciteit en maximum op- en ontladingsnelheid van de batterij. Deze kunnen namelijk niet ongelimiteerd snel laden, en er gaat energie verloren bij het opladen en afladen. Het doel is om te onderzoeken hoe je het

beste kan omgaan met het onbekend toekomstig gebruik.

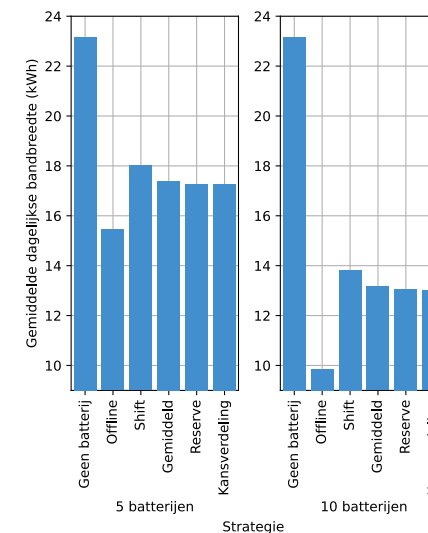
Wanneer het verbruik vooraf bekend is, kan een optimaal laadschema voor de batterij gemaakt worden. Het plannen met deze kennis heet "offline" plannen. Voor wanneer we het toekomstig gebruik niet weten (de "online" variant) hebben we verschillende Rolling Horizon strategieën.

- Shift: We nemen het gebruik van zeven dagen geleden en berekenen daar een optimaal laadschema voor, en gedurende de dag sturen we bij op verschillen.
- Gemiddelde: We gebruiken alleen de verwachtingswaarde van de voorspelde kansverdelingen als deterministisch toekomstig gebruik, en sturen bij.
- Reserve: We houden een paar procent van de capaciteit als reserve, zowel voor het opladen als voor het afladen. Zo kunnen we voorkomen dat de batterij te vaak te vroeg leeg raakt bij een hoger gebruik dan verwacht.
- Kansverdeling: Ook hier hanteren we een reserve, maar dan op basis van de kansverdeling. Is de kansverdeling breder (meer onzekerheid), dan is de reserve ook groter.

Deze strategieën zijn toegepast in testscenario's met vijf of tien Tesla Powerwall thuisbatterijen van 13.5kWh en een efficiëntie van 90%. Deze batterijen hebben we verdeeld over vijf groepen van 100 huishoudens, en over 500 huishoudens tegelijkertijd, om ook over verschillende schalen te testen.



Figuur 3: Gemiddelde dagelijkse bandbreedte voor de zes verschillende strategieën voor vijf en tien batterijen voor alle 500 huishoudens.



Figuur 4: Gemiddelde dagelijkse bandbreedte voor de zes verschillende strategieën voor vijf en tien batterijen verdeeld over vijf losse groepen van 100 huishoudens.

Enerzijds om wat te kunnen zeggen over hoe goed de modellen werken op verschillende schalen, maar ook om te zien hoe de batterijen het beste ingezet kunnen worden.

In alle gevallen is de strategie gebaseerd op de kansverdeling het beste, met een wisselende tweede plek voor de Shift en de Reserve strategie (zie Figuur 3 en 4).

Discussie

Op basis van MAPE, NRMSE en Loglikelihood loss scores zien we dat het LSTM model met vijf clusters de LR en TimeShift modellen verslaat. Ons voorspelmodel is dus beter dan de benchmark. Verder hebben we laten zien dat, voor het inplannen van batterijen, het nuttig is om gebruik te maken van voorspellingen. Het weten van de zekerheid van de voorspellingen door middel van kansvoorspellingen verhoogt de resultaten nog verder. Het zou goed kunnen dat dit geldt voor alle planningsproblemen waar een mate van robuustheid nodig is, zeker als in een oplossingsalgoritme voor deze planningsproblemen een robuustheidsparameter verwerkt kan worden. Zijn de voorspellingen onzeker, maar

is haalbaarheid van de oplossing belangrijk? Dan kan er daar rekening mee gehouden worden door extra middelen achter de hand te houden. Is het model zekerder in zijn voorspelling? Dan hoeven we minder achter de hand te houden en kunnen we een uitdagendere planning maken.

Echter, voordat ons planningsalgoritme ingezet kan worden, zou het goed zijn te testen op een bredere dataset. Zo hebben we nu data van één regio in Nederland gebruikt, en worden de batterijen op slechts de zomer getest (de rest van de data was als training of validatiedata gebruikt). Het zou goed zijn meerdere jaren aan data te hebben van verschillende landen om echte conclusies te vormen. Ook zouden schonere data de resultaten kunnen verbeteren.

We hebben laten zien dat het toekomstig energiegebruik voorspeld kan worden met de EMG-kansverdeling, maar mogelijk zijn andere kansverdelingen beter. Verder zou er een vergelijking gemaakt kunnen worden tussen het direct voorspellen van het gebruik en de zekerheid, en twee losse modellen: één die een verbruik voorspelt, en een ander die de zekerheid voorspelt of berekent.

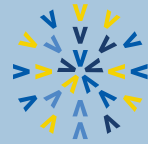
Literatuur

- S. Bañales, R. Dormido en N. Duro. „Smart meters time series clustering for demand response applications in the context of high penetration of renewable energy resources“. In: *Energies* 14.12 (2021), p. 3458.
- D. Kodaira, W. Jung en S. Han. „Optimal energy storage system operation for peak reduction in a distribution network using a prediction interval“. In: *IEEE Transactions on Smart Grid* 11.3 (2019), p. 2208–2217.
- A. Nottrott, J. Kleissl en B. Washom. „Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems“. In: *Renewable Energy* 55 (2013), p. 230–240.
- J. Xie en T. Hong. „Temperature scenario generation for probabilistic load forecasting“. In: *IEEE Transactions on Smart Grid* 9.3 (2016), p. 1680–1687.
- L. Xu, M. Hu en C. Fan. „Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks“. In: *Journal of Building Engineering* 46 (2022), p. 103853.

Dankwoord

Dit artikel is een samenvatting van mijn masterthesis, gebaseerd op het onderzoek onder leiding van Dr. ir. Marjan van den Akker vanuit Computing Science, Dr. Sjoerd Dirksen vanuit Mathematical Sciences en Rinse Veltman vanuit het bedrijf Energyworx. Ik wil dit drietal hartelijk bedanken voor de prettige samenwerking en de begeleiding tijdens dit onderzoek. De volledige versie van de thesis is te lezen op pdknops.nl/probability_forecasting_thesi.html

Pieter Knops heeft in 2023 aan de Universiteit Utrecht een dubbele master in Mathematical Sciences en in Computing Sciences voltooid, met een focus op Scientific Computing en Operations Research. Zijn passie ligt in het oplossen van praktische, maatschappelijk relevante optimalisatie-vraagstukken. Pieter heeft, met de thesis waarop dit artikel gebaseerd is, de prijs van Best Master Thesis in Applied Mathematics 2023 gewonnen, uitgereikt door het Centrum voor Wiskunde en Informatica. Ter ontspanning doet Pieter aan hardlopen en wielrennen, en is hij imker. linkedin.com/in/pieter-knops, pdknops.nl.



Young statisticians

The Young Statisticians recently hosted their annual YS Pub Quiz to kick off the new academic year at Café Einstein in Leiden. It was wonderful to see many familiar faces alongside a number of new MSc students. The teams were highly motivated and competed fiercely in rounds covering R, statistics, and funny graphs. Congratulations to the winning team for earning their well-deserved YS mugs!



Next event

Our next event will be a Statistics Café on Sports Statistics, organized by our new board. If you'd like to learn more, sign up for our newsletter, follow us on LinkedIn or Instagram, or keep an eye on our website: vvsor.nl/young-statisticians.

Kleine oorzaken ...

Kleine oorzaken kunnen grote gevolgen hebben. Daar kennen we allemaal voorbeelden van, zowel privé als in ons werkzame leven.

Heb u zich ooit afgevraagd waarom eind jaren tachtig SPSS ineens zo'n massale verspreiding kreeg? Daar was een plotselinge ingeving die ik kreeg deels verantwoordelijk voor.

Ik kwam in 1982 in dienst bij het Rekencentrum van de toenmalige Landbouwhogeschool Wageningen. Alle rekenwerk ging toen nog op grote centrale mainframes, bij de LH was dat de DEC-10 van Digital Equipment Corporation. Later werden dat VAX-computers van dezelfde fabrikant. Vanuit het Wageningse Rekencentrum werden de HBO-opleidingen in de landbouwsector ondersteund, en vanzelfsprekend was ook daar DEC de huisleverancier van alle apparatuur.

Voorzichtig kwamen de eerste PC's op de markt, en ook DEC kwam met een eigen ontwerp. Deze PC's hadden twee processoren, waardoor men zonder emulatie met twee verschillende operating systemen kon werken, als ik goed onthouden heb een 8-bits en een 16-bits. Heel belangrijk was ook dat ze konden werken als de fameuze VT100 terminals. Daardoor had je maar een enkel apparaat op je bureau waarmee je direct verbinding had met de mainframes maar dat ook een zelfstandige PC was.

Er was maar één minpuntje dat uiteindelijk dit ontwerp de das om heeft gedaan: ze waren niet 100% compatibel met de IBM-achtige PC's. Zoals het zo vaak gaat was dit ontwerp in technische zin feitelijk superieur maar faalde de marketing.

Allerlei softwareleveranciers begonnen hun producten om te zetten voor gebruik op een PC, uiteraard koos men daarbij voor het DOS-systeem. Ook SPSS deed dat. De prijzen waren aan de forse kant: enkele duizenden guldens voor een enkele licentie. Maar ja, de PC's van toen hadden ook prijzen die boven de f 5000 uitkwamen dus we waren wel gewend aan grote bedragen. Voor grotere organisaties wa-

ren er wel voordeliger massa-licenties beschikbaar, vanaf minimaal tien gebruikers.

Als coördinator van de statistische software in het Wageningse kreeg ik al snel het verzoek die PC-versie van SPSS aan te schaffen. Maar door de enorme dominantie van de iets afwijkende DEC-apparatuur waren er binnen de hele universiteit maar enkele 'standaard' PC's en had ik niet genoeg afnemers voor die massa-licentie.

En ja, toen kreeg ik die ingeving die grote gevolgen zou hebben. Ik had uitstekende contacten met SPSS, ik was onder andere mede-oprichter en eerste voorzitter van de onafhankelijke Europese SPSS-gebruikersclub. Ik heb voorgesteld om de universiteit én de HBO-opleidingen als één gezamenlijke klant te beschouwen. Samen hadden we dan wel genoeg PC's. Het Wageningse Rekencentrum zou als contactpunt fungeren en zorgen voor de landelijke distributie. Dat idee hebben we uitgewerkt en zelfs verder uitgebreid voor SPSS op alle VAX en Mini-VAX mainframes en alle PC's. Dat liep voortreffelijk.

Er was in die tijd een regelmatig overleg van directeurs van de universitaire rekencentra, de CVDUR. Later is dat overgegaan in SURF, dat in eerste instantie uit diezelfde groep directeurs bestond. In zo'n overleg vertelde Henk le Grand, de Wageningse directeur, over die succesvolle massalicensie en zijn collega's waren erg geïnteresseerd. Kort en goed: CVDUR/SURF opende besprekingen met SPSS voor een landelijke licentie naar Wagenings model. Dat verliep voorspoedig, iedereen zag grote voordelen. Op 30 juni 1989, dus 35 jaar geleden, werd met enig ceremonieel de SURF-SPSS licentieovereenkomst voor heel academisch Nederland gesloten. Hoewel ik er niets meer mee te maken had, ik was inmiddels bij het CWI werkzaam, kreeg ik toch een mooie pen met inscriptie, als dank voor die uit nood geboren ingeving die zulke grote gevolgen had.

Gerrit Stemerding is eindredacteur van STATOR.
E-mail: gjstemerding@hotmail.com



3D-Beleggen: Portefeuilleconstructie voor beleggers met duurzaamheidsdoelen

Rogier Swierstra

Iedere belegger staat voor het probleem van portefeuilleconstructie: waarin te beleggen om in een onzekere wereld je doelen zo goed mogelijk te behalen? Pensioenfondsen hebben onder beleggers uitzonderlijk lange horizons en kunnen zich daarom richten op langetermijndoelen. Dan spelen er meer factoren mee dan alleen het rendement en risico dat de basis vormt van beleggingstheorie, zoals duurzaamheid. Het meenemen van deze nieuwe dimensie in beleggingsafwegingen is 3D beleggen gaan heten (Gasser, Rammerstorfer en Weinmayer, 2017). In dit verhaal zullen we eerst kort ingaan op de klassieke portefeuilleconstructietheorie van Markowitz, om begrippen te introduceren. Daarna breiden we het model uit naar meerdere doelen. Ten slotte doen we voorstellen voor het communiceren van deze begrippen aan de bestuurders van pensioenfondsen, om hen te ondersteunen bij het maken van de juiste beleggingsafwegingen.

Moderne portefeuilletheorie (MPT)

Het standaardprobleem van een belegger is om een portefeuille samen te stellen uit N risicovolle beleggingscategorieën (denk aan aandelen, obligaties, vastgoed, enzovoort) met voor hem de beste verhouding tussen rendement en risico – uitgedrukt als de variantie van de waarde. Schrijf μ voor de vector van verwachte rendementen van de categorieën (op een vastgestelde horizon) en Σ voor de covariantiematrix; deze matrix geeft alle onderlinge covarianties van de beleggingscategorieën. We zoeken als een portefeuille een genormaliseerde vector van gewichten w (sommarend tot 1) met hoog rendement en laag risico:

$$\max_w w^T \mu - \frac{\gamma}{2} w^T \Sigma w,$$

waar de parameter γ uitdrukking geeft aan de risicoaversie van de belegger. Dit probleem laat zich analytisch oplossen met behulp van Lagrange multiplicatoren. De optimale portefeuille, gegeven door w_μ , is een lineaire combinatie van de raakportefeuille $w_\mu = \Sigma^{-1} \mu$ met het hoogste rendement-per-eenheid-risico, en de minimum-variantie portefeuille $w_e = \Sigma^{-1} e$ (waarbij we e schrijven voor een vector van 1-en) om het risico eventueel terug te brengen.

De oplossing verandert wanneer een risicovrije belegging wordt verondersteld, zoals kasgeld in een beleggingscontext. Beleggers kunnen dit ook lenen om meer te beleggen, als ze rente μ_0 betalen over het geleende bedrag. Nu is het excesrendement van categorieën ten opzichte van risicovrij rendement $(\mu - \mu_0 e)$ van belang voor portefeuilleconstructie; we schrijven $w_\mu^* = \Sigma^{-1} \cdot (\mu - \mu_0 e)$ voor de raakportefeuille en w_0 voor de risicovrije belegging. Het inzicht van Markowitz (Markowitz, 1952) was dat de optimale portefeuille

$$w_\gamma^* = \lambda_\gamma w_\mu^* + (1 - \lambda_\gamma) w_0$$

voor alle beleggers bestaat uit een combinatie van dezelfde markt-portefeuille w_μ^* en een van hun risicohouding afhankelijke kaspositie. Hier is λ_γ een normalisatiefactor afhankelijk van de risicoaversie (die ook groter dan 1 kan worden). Zie voor een illustratie de Figuur 1, waarbij risico is uitgedrukt in standaarddeviatie, dat is de wortel van de variantie.

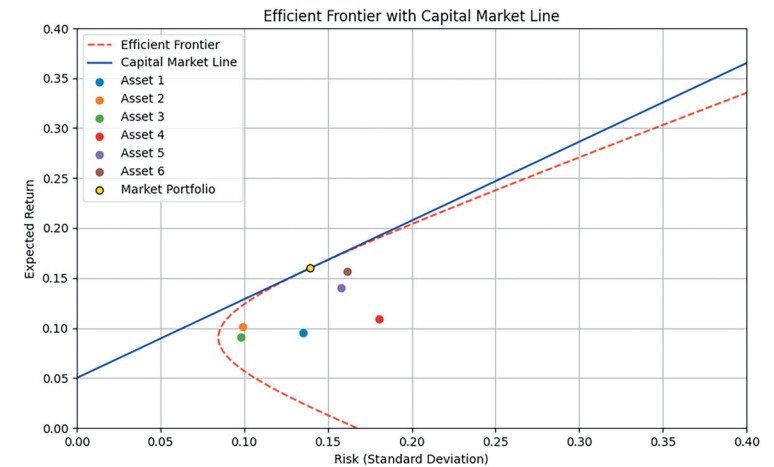
Deze alinea's hebben de nadruk gelegd op het analytische probleem. In de praktijk zijn er al snel complicerende randvoorwaarden zoals dat de componenten w_i beperkt zijn (niet negatief mogen wor-

den, bijvoorbeeld) en moeten oplossingen numeriek worden gezocht. Daarnaast zijn er nog belangrijke kwalitatieve overwegingen zoals de keuze van de horizon en de bepaling van de beleggingscategorieën, en de schatting van hun karakteristieken μ en Σ . Tenslotte, om tot een oplossing te komen is het nodig om een uitspraak te doen over de risico-aversie γ waar een bestuur met verantwoordelijkheid voor andermans spaargeld het lastig over eens wordt.

Uitbreiding naar meerdere doelen

Nu is de vraag hoe de portefeuilleconstructie uit te breiden om nieuwe doelen mee te wegen. Zoals gezegd betreft dit voor pensioenfondsen de bestuurlijke wens om duurzaam te beleggen. Dit kan intuïtief op meerdere manieren, onder andere:

1. Restricties opleggen; zoals het uitsluiten van categorieën of sectoren.
2. De opties uitbreiden; met nieuwe categorieën met betere duurzaamheidskarakteristieken.
3. De parameters (μ, Σ) herberekenen; om werkelijke kosten en risico's in een veranderende wereld juist te modelleren.
4. De doelfunctie uitbreiden.



Figuur 1: Schematische weergave van categorieën en portefeuilles in een risico-rendementsruimte, met punten voor verschillende categorieën, een efficiënte grenslijn (rood) en de efficiënte lijn in aanwezigheid van een risicovrije belegging (blauw). De marktportefeuille w_μ^* (geel) ligt op hun raakpunt, de risicovrije op de verticale as.

We richten ons hier op de laatste optie (Blitz e.a., 2024), omdat deze de afweging tussen de doelen expliciet maakt. Daarom veronderstellen we dat we een vector van duurzaamheidskarakteristieken θ (bijvoorbeeld CO2-uitstoot of bijdrage aan de ontwikkelingsdoelstellingen [SDGs] van de VN) van de bestaande categorieën kunnen bepalen, en dat de portefeuille-score $w^T\theta$ een gewogen som van deze scores is. Het portefeuille-probleem schrijven we nu als

$$\max_w \phi w^T(\mu - \mu_0 e) + \psi w^T(\theta - \theta_0 e) - \frac{1}{2} w^T \Sigma w$$

(nu zonder de eis $|w| = 1$ dus aan te vullen met kas met parameters ϕ en ψ , die de voorkeuren weergeven voor rendement respectievelijk duurzaamheid ten opzichte van risico.

Een oplossingsrichting zou zijn om beide doelen samen te voegen in een voor duurzaamheid gecorrigeerd rendement $(\mu + \frac{\psi}{\phi}\theta)$, en de standaardoplossing te volgen met deze waarde. Dit vereist echter dat de parameters bekend en betrouwbaar zijn, en daarmee dat de uitruil tussen rendement en duurzaamheid expliciet gemaakt kan worden.

Als alternatief suggereert dezelfde argumenta-

tie als hiervoor het introduceren van een risico-optimale duurzaamheidsportefeuille

$$w_\theta^* = \Sigma^{-1}(\theta - \theta_0 e)$$

voor het geval $\phi = 0$. Hierbij hoort opnieuw een raaklijn samen met de risicovrije belegging, net als in de vorige paragraaf. Samen met de rendementsraaklijn wordt dus een vlak gespannen door $\{w_0, w_\mu^*, w_\theta^*\}$ dat efficiënte oplossingen geeft van het probleem, voor verschillende keuzes van de parameters (ϕ, ψ) .

Feitelijk zijn de te kiezen parameters dus coördinaten op een vlak. De bijbehorende maximale opbrengsten leiden tot het efficiënte grens-oppervlak in 3D, een uitbreiding ten opzichte van Figuur 1. Omdat er in het algemeen interactie is tussen het risico van w_μ^* en w_θ^* , zogeheten diversificatie, is dit oppervlak gekromd. In Figuur 2 is dit grensoppervlak gevisualiseerd voor de situatie zonder de aanwezigheid van een risicovrije belegging.

Eenzelfde logica geldt voor ieder additioneel doel van deze lineaire vorm, bijvoorbeeld een additioneel duurzaamheids-eigenschap θ' . Maar er is nog een gangbaar doel dat in deze aanpak past. Vaak willen beleggers de afwijking van een benchmarkportefeuille b beheersen, voor pensioenfondsen is dit hun toegezegde uitkeringen. Dit geeft in het probleem een strafterm $(w - b)^T \Sigma (w - b)$. Het uitwerken van dit kwadraat geeft ons termen $b^T \Sigma b$ (constant dus irrelevant voor optimalisatie), $w^T \Sigma w$ (de bekende term voor risico) en de kruissterm $2w^T \Sigma b$: deze is weer lineair in w . We kunnen dus weer een bijbehorende portefeuille $w_{\Sigma b}^*$ introduceren, en merken op dat $w_{\Sigma b}^* = \Sigma^{-1}(\Sigma b) = b$. Om de afwijking met de benchmark-portefeuille b te minimaliseren moet je natuurlijk in b zelf beleggen.

Dit verklaart waarom Markowitz' conclusie over de marktportefeuille in de praktijk niet opgaat. Beleggers hebben verschillende referentiewaarden voor b (naast andere doelen); de samenstelling van hun portefeuilles is dus van meer afhankelijk dan alleen hun risico-aversie.

Communicatie met bestuurders

Het bepalen van de risico-aversie γ , nodig voor de afweging tussen risico en rendement, is psychologisch gezien een taai vraagstuk. Zeker voor een groep en al helemaal wanneer ze optreden als fiduciair met

verantwoordelijkheid voor de belangen van anderen. Het probleem wordt exponentieel moeilijker wanneer er meer dimensies zijn en meerdere parameters (ϕ, ψ, \dots) in samenhang moeten worden bepaald.

Daarom is de lineariteit van onze aanpak zo nuttig. De bestuurlijke afweging hangt dan niet af van de positie op de niet-lineaire grenscurve. Een bestuur kan een aantal illustratieve opties gepresenteerd krijgen, bestaand uit verschillende wegen van optimale portefeuilles w_μ^* voor rendement, w_θ^* voor duurzaamheid, $w_{\Sigma b}^* = b$ voor relatief risico en w_0 voor absoluut risico, en onderling schuiven tussen deze portefeuilles totdat een combinatie gevonden wordt die aansluit bij hun gemeenschappelijke voorkeur. De adviseurs kunnen de impliciete voorkeursparameters afleiden, en gebruiken voor numerieke oplossingen en bij gedelegeerd portefeuillebeheer.

Dankwoord

Dit artikel kwam tot stand op basis van een project van studenten van de Universiteit Utrecht: Gijs Bartholomeus, Erwin Dijkstra, Maarten de Groot, Ezra Hulleman, Lay de Jong, Tomas Koekoek, Arco de Kort, Jochem Lange, en Edward Tollington.

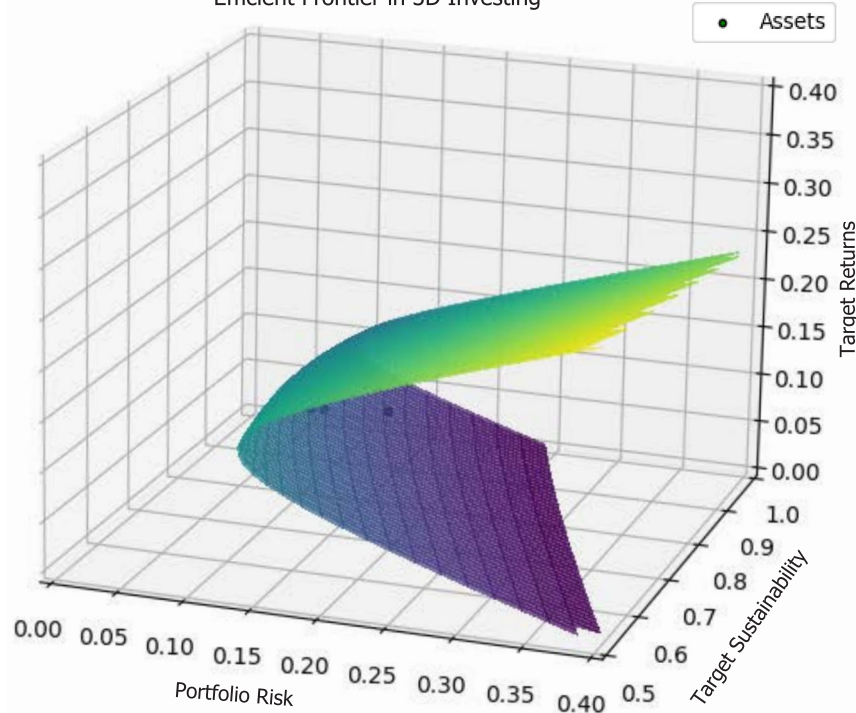
Naast het formaliseren van de wiskunde van het probleem met meerdere doelen, hebben zij onderzocht hoe een beleggingscommissie hun opdrachten kon delegeren zodat afzonderlijke uitvoerders optimaal kunnen samenwerken.

Literatuur

- D. Blitz e.a. „3D Investing: Jointly optimizing return, risk, and sustainability“. In: *Financial Analysts Journal* 80.3 (2024), p. 1–17. DOI: 10.1080/0015198X.2024.2335142.
- S. Gasser, M. Rammerstorfer en K. Weinmayer. „Markowitz revisited: Social portfolio engineering“. In: *European Journal of Operational Research* 258 (2017), p. 1181–1190. DOI: 10.1016/j.ejor.2016.10.043.
- H. Markowitz. „Portfolio Selection“. In: *Journal of Finance* 7 (1952), p. 77–91. DOI: 10.2307/2975974.

Rogier Swierstra is portefeuillemanager bij PGGM en verbonden aan het Mathematisch Instituut van de Universiteit Utrecht. E-mail: rogier.swierstra@pggm.nl.

Efficient Frontier in 3D Investing



Figuur 2: Schematische weergave van de uitbreiding van het klassieke model naar een derde dimensie, zonder risicovrije categorie. Optimale portefeuilles vormen het efficiënte oppervlak, waarbij enkele assets met stippen zijn weergegeven. De kleuren benadrukken de hoogtelijnen; hoe lichter, hoe beter – in dit geval betekent dit een hoger rendement.

Statistiek, astronomie en de constructie van data

Richard Starmans

Het jaarcongres van de VVSOR was dit jaar geheel gewijd aan het thema "Seeing the invisible". Dat klinkt zowel ambitieus als paradoxaal; het impliceert immers niets minder dan dat het onzichtbare zichtbaar moet worden gemaakt. Toch klinkt het thema ook wel weer vertrouwd. De gedachte dat achter de wereld van de verschijnselen een diepere, vooralsnog onzichtbare werkelijkheid schuilgaat, de echte wereld, die de verklaring vormt van de zintuiglijk waarneembare werkelijkheid (onze soms bedrieglijke percepties en impressies) en die kenbaar is, maar ontsluit moet worden, is immers zo oud als de wetenschap zelf.

Theorie en Data

Rond 600 voor Chr. begon zich een naturalistische wending te voltrekken, toen Ionische en Dorische natuurfilosofen niet langer de verschijnselen wilden duiden als acties van intentioneel handelende homeïrische goden, maar deze trachtten te verklaren vanuit onderliggende abstracte principes, causale werkingsmechanismen en algemene theorieën. Geen machten, maar krachten, werkzaam in de natuur en dat alles beschreven in een veel meer betogende en argumenterende dan narratieve stijl. Reductie-nisme was daarbij essentieel: het terugbrengen van de verschijnselen tot (abstracte) eerste principes of begrippen, van de kwalitatieve deeltjes van Empedocles (aarde, water, lucht, vuur), via Anaximanders "apeiron" of "het onbepaalde" tot de ondeelbare, eeuwige atomen van Demokritos en Leukippos.

Zo ontstond het spanningsveld, de wisselwerking tussen data en theorie en daarmee het project van de wetenschap. Enerzijds moeten de observaties en impressies worden gecodeerd in data; deze zijn concreet, particulier en "eindig" en vormen de empirische "basis". Anderzijds is er de theorie met haar al dan niet causale verklaringen: abstract, algemeen en "oneindig". Deze relatie speelt tot op heden filosofen en wetenschappers parten. Hoe ziet die empirische basis eruit? Zijn het "sense data", percepties of impressies en kunnen daaruit objectieve feiten worden afgeleid? En hoe komen we vervolgens bij een theorie? En omgekeerd: moet die theorie dan alle bekende observaties verklaren en nieuwe voorspellen? Licht het primaat bij de data of bij de theorie? Hoe beïnvloeden ze elkaar en in welke mate behoren zij afhankelijk dan wel onafhankelijk te zijn? Vragen als deze vormden een opmaat tot klassieke filosofische thema's, zoals het inductieprobleem (Hume), de theoriegeladenheid van de waarneming (Hanson, Popper), de onderdeterminatie van de theorie door data (Quine, Kuhn) et cetera. Wetenschappers daarentegen worstelen met de vraag hoe abstracte begrippen moeten worden geoperationaliseerd; mensurandum versus mensuratum, grootheid versus eenheid, theoretische constructies versus observationele begrippen. En dat liefst op een betrouwbare en valide wijze. Hoe dan ook, bij het "zichtbaar maken van het onzichtbare" en de daarbij onvermijdelijke wisselwerking tussen theorie en data, speelt statistiek dikwijls op verschillende manieren een rol. We beperken ons hier tot één aspect van de problematiek. De grens van het waarneembare schuift steeds verder op; door betere optica, instrumenten en designs, maar ook door de evolutie van onze notie van het waarneembare; van ongeholpen observatie en opsporend experiment, via detectie en simulatie naar actieve constructie van data. Een tweetal voorbeelden passeert hier kort de revue. Allereerst de astronomie en het zichtbaar maken van zwarte gaten; vervolgens de ontwikkeling van de deeltjesfysica in de 20e eeuw en het zichtbaar maken van de subatomaire wereld.

De Event Horizon telescoop

Op 10 april 2019 werd door de *Event Horizon Telescope Collaboration (EHT)* voor het eerst een foto van (de schaduw van) een zwart gat gepubliceerd. Dat was wereldnieuws en de prestatie werd door het tijdschrift *Science* prompt uitgeroepen tot de doorbraak van het jaar. De gebeurtenis vond bovendien plaats precies honderd jaar nadat Sir Arthur Edding-

ton (1882-1944) in 1919 zijn expeditie naar het eiland Principe in de Golf van Guinee had georganiseerd, om daar een zonsverduistering te kunnen waarnemen. Licht dat langs de verduisterde zon trok, bleek te worden afgebogen door de zwaartekracht van de zon en zo kon Eddington een eerste empirische bevestiging van Einsteins algemene relativiteitstheorie geven. Beide "waarnemingen" spraken enorm tot de verbeelding van het grote publiek en werden dan ook breed uitgemeten in de media als historische en heroïsche pogingen het "onzichtbare zichtbaar" te maken. Het onderzoeksproject *EHT*, dat in 2009 relatief kleinschalig van start was gegaan met een drietal telescopen in de VS was tien jaar later uitgegroeid tot een mondiaal netwerk van telescopen met 347 onderzoekers, werkzaam aan 60 instituten uit twintig landen. In 2022 kreeg het een officieel vervolg met *The Next Generation Event Horizon Telescope Collaboration (ngETC)*, een samenwerkingsverband waarin vele honderden onderzoekers wereldwijd samenwerken, niet alleen natuurkundigen en astronomen, maar opmerkelijk genoeg ook historici, filosofen en sociale wetenschappers. De Amerikaanse natuurkundige, wetenschapshistoricus en wetenschapsfilosoof Peter Galison (1955) speelt hierbij een belangrijke rol, onder meer als initiatiefnemer en coördinator van *The History, Philosophy, and Culture Working Group*, één van de acht binnen de ngETC ingestelde officiële werkgroepen (Galison, 2023). In het oprichtingsdocument bouwt hij voort op vroeger werk (Galison, 1987, 1997), waarin hij de wereld van "big science" projecten analyseert; internationale onderzoeksgroepen, politieke spanningsvelden, mondiale subsidieprogramma's en peperdure deeltjesversnellers. Maar ook de protocollen en conventies, de meetings tussen experts, inclusief onderhandelingen, debatten, conflicten, compromissen en andere consensus- en besluitvormingsmechanismen. Die bepalen zeker ten dele wanneer een experiment wordt beëindigd, wat telt als een resultaat en ook wat een waarneming is, wat moet worden getoond en hoe dit geschiedt.

Zo stelt Galison dat astronomische beelden het resultaat zijn van een "array" van keuzes en beslissingen "and they also participate in their larger historical and cultural contexts". De creatie en beoogde interpretatie ervan hangen af van "pre-existing visual traditions that establish the norms, expectations and methods by which a scientific image is given meaning". Hij vervolgt: "Another concern for the EHT was the legibility of the images for a wide audience...The color palette - a ring on orange-red hues against a black background - was chosen with this in mind; orange was believed to

more signify heat than blue (even though blue has shorter wavelengths and is therefore “hotter” than orange. Because the EHT Collaboration wanted to share one image with audiences of varying degrees of specialization... a single averaged image was created from multiple images based on different imaging methods”. Het constructieve karakter van deze “gemiddelde” waarneming die in 2019 aan de wereld werd getoond, is onmiskenbaar. Om de consensusvorming bij dit soort kwesties te duiden grijpt Galison terug op de gegeneraliseerde “Condorcet Jury Theorem”, vernoemd naar Marquis de Condorcet (1743-1794), de 18e eeuwse statisticus en Verlichtingsdenker, die geldt als een van de grondleggers van de sociale keuzetheorie, waarin het aggregeren van individuele overtuigingen, preferenties en belangen tot een gemeenschappelijke “keuze” of “volonté generale” centraal staat. Iets dat, enigszins gechargeerd en anachronistisch uitgedrukt, zogenaamde “stackings-algorithmen” in “ensemble learning” heden ten dage eveneens trachten te bewerkstelligen. Dit alles is saillant, maar nauwelijks verrassend in het licht van Galisons jarenlange preoccupatie met de verhouding tussen natuurkunde, cultuur en maatschappij. In een recente bijdrage gaat hij nog een stap verder en pleit voor betrokkenheid van de humaniora en sociale wetenschappen in alle “high-tech-big-science” projecten (Marconi, 2023).

Deeltjesfysica en statistiek

Galison vormt ook de verbindende schakel met het tweede voorbeeld, de evolutie van de deeltjesfysica in de 20e eeuw, waarnaar hij al ruim veertig jaar onderzoek doet. Zijn historisch-filosofische benadering maakt hem één van de vernieuwers in de hedendaagse wetenschapsfilosofie. In *How Experiments End* (1987) schetst hij de teloorgang van een roemruchte traditie, waarbij fysici op eigen houtje of in zeer kleine teams opereren, zelfstandig hun experimenten uitvoeren en desnoods de benodigde instrumenten zelf bouwen. De auteur verricht enkele historische casestudies, gewijd aan een drietal experimentele perioden uit de fysica, die zich kenmerken door een toenemende graad van complexiteit en organisatie. Allereerst gaat Galison in op de kleinschalige, min of meer traditionele, macroscopische studies van onder meer Einstein en De Haas naar gyromagnetische effecten. Vervolgens besteedt hij aandacht aan experimenten met kosmische straling, die in de jaren 30 zowel door Amerikaanse als Europese onderzoeksgroepen werden uitgevoerd en die zouden leiden tot de ontdekking

van het mu-meson (muon). Tot slot analyseert Galison in ruim honderd pagina's de ontdekking van zogenaamde weak electric current als resultaat van mega-onderzoeksprojecten met langdurige experimenten, waarbij honderden onderzoekers waren betrokken. Belangrijk is dat hij met deze casestudies laat zien dat experimenten “a life of their own” bezitten en daarmee een rehabilitatie van het experiment bewerkstelligde in de wetenschapsleer. Hij deed dit samen met de Canadese wetenschapsfilosoof en historicus van de statistiek Ian Hacking (1936-2023). Laatstgenoemde had in zijn *Representing and Intervening* (1983) vooral langs filosofische weg dezelfde positie verdedigd. Daartoe was ook wel enige aanleiding. Experimenten hadden uiteraard lange tijd een cruciale rol gespeeld als “scherprechter” bij het toetsen van hypothesen of theorieën. Men denke aan het “experimentum crucis” van Robert Hooke in de 17e eeuw, aan de opvattingen van de logisch-positivisten (verificatie, confirmatie) of aan het werk van Karl Popper (falsificatie). Toch waren ook daar experimenten in zekere zin ondergeschikt en vooral instrumenteel aan de theorie, waar alles om draaide, hetgeen volgens sommigen tot een monomane preoccupatie met theorievorming leidde. Door de befaamde “Quine-Duhem-these” over de principiële onmogelijkheid van het testen van geïsoleerde hypothesen, maar vooral door de populaire kuhniaanse opvatting over de “incommensurabiliteit van paradigma's” werd de rol van het experiment als scherprechter steeds twijfelachtiger. Hoe dan ook, Galison en Hacking zouden beide boegbeelden worden van *The New Experimentalism*, een stroming die eind vorige eeuw opkwam, de eenzijdige focus op theorie ten koste van het experiment kritiseert en nog steeds relevant is, zeker in tijden van data-science en “data-driven everything”. Belangrijker nog, Galison betoogde met verve dat vooruitgang in de natuurkunde naast een theoretische laag een zelfstandige, grotendeels autonome experimentele laag behelst.



Peter L. Galison

Maar Galison ging nog een stap verder in zijn bijkans klassieke boek *Image and Logic; a material culture of microphysics* uit 1997. Vreemd genoeg spreekt hij over “logic”, maar al spoedig wordt zonneklaar dat de auteur feitelijk statistiek en probabilistisch redeneren bedoelt. In deze studie analyseert de auteur twee experimentele tradities in de ontwikkeling van de 20e eeuwse deeltjesfysica met eigen instrumenten, detectiemethoden en “visualiserings”. Beide dragen op eigen wijze bij aan het zichtbaar maken van het onzichtbare en het opschuiven van de grens van het waarneembare. Allereerst de beeldtraditie, die begon met de nevelkamers of wolkenkamers van Wilson in 1910, daarna een vervolg kreeg met de nucleaire emulsietechnieken en die uiteindelijk zou leiden tot de populaire bellenvaten in de jaren vijftig. Vele nieuwe deeltjes zijn op deze wijze “gevonden”. Ook hier speelde statistiek een substantiële rol, onder meer bij het exploreren van de vele data, anomaliedetectie en het ontwaren van patronen. Maar ruwweg gaat het toch om het fotografisch vastleggen van een “Golden Event”, een soms vluchtig moment dat wordt gefixeerd in de tijd. De eerste wolkenkamers betroffen het fotograferen van geïsoleerde deeltjes, die via condensatie een spoor achterlieten in een nevel. Bij de nucleaire emulsies daarentegen ging het vooral om het beeld dat de deeltjes direct achterlieten op de fotografische plaat. Maar, “*there stood an abiding faith in the power of the individual image*”, aldus Galison. Dat wordt allemaal anders bij de tweede traditie met haar elektronische, statistische aanpak. Deze traditie kwam voort uit de geiger-müllertellers van de jaren twintig en zou onder meer leiden tot vonkenkamers en later draadkamers.

Anders dan bij de beeldtraditie was visualisatie, sterk gebaseerd op statistische en logische afleidingen, waarbij “ruimtelijke” informatie wordt gegeven door configuraties van (geiger)tellers te creëren. De auteur stelt onder meer: “*For all these reasons, counters considered individually were statistical devices. Then, over and above their individual behavior, coincidence counting provided a second, irreducibly statistical feature of the logic tradition. When four, five, even six counters fired within a short time, there was always the possibility that they had been triggered, not by a single penetrating ray, but by the separate particles of a widespread avalanche—or even by coincidental spurious discharge. Any argument for the penetration of a single particle (as opposed to a shower) or for a shower (as opposed to a single particle) had to be of the form: one hypothesis is more probable than another. Consequently, a single event, whether it was the*

click of a Geiger counter or the pulse from a complex array of counters, was meaningless. Data in the logic tradition became persuasive only in their statistical aggregation.”

Kortom, de statistiek in de rol van Plato's demiurg, die als een superingenieur de brokstukken assembleert tot één geheel; een probabilistisch universum, gemodelleerd naar de eeuwige platoonse Ideeën. Beide tradities bestonden jarenlang onafhankelijk naast elkaar, gaven het landschap van de experimentele fysica daarmee een enigszins verzuilde aanblik en zouden pas in de jaren zeventig van de vorige eeuw convergeren.

Galison analyseert aan de hand van de verschillende deeltjesdetectoren de “status” van instrumenten en constateert dat deze op hun beurt relatief onafhankelijk zijn van experiment en theorie. Met andere woorden, ook instrumenten hebben “*a life of their own*” en de natuurkundige praktijk krijgt daarmee een gelaagde structuur. Theorie, experiment en instrument bestaan naast -en deels onafhankelijk van elkaar, veranderen doorgaans niet simultaan en vormen een soort “tradingzones”, waarin de natuurkundige praktijk gestalte krijgt. De auteur introduceert daarbij de “metafoor van de muur”, waarbij de verschillende lagen geschakeld zijn samengevoegd; intern consistent, maar veranderingen, hetzij groot en abrupt (revolutie) of klein en langzaam (evolutie) verlopen niet gelijktijdig. Communicatie is mogelijk via het oude antropologische concept van pidgins, eenvoudige mengtalen die ontstaan in zones waarin verschillende culturele groepen elkaar ontmoeten en handeldrijven. Dus geen incommensurabiliteit! Vooruitgang in met name “big science” komt tot stand in deze uitwisselingsgebieden en daar ligt ook de rationaliteit van de wetenschap. Revoluties vinden niet globaal, maar veeleer lokaal plaats in één van de lagen, waarmee zowel continuïteit (van de natuurkundige praktijk) als discontinuïteit (verandering) wordt verdisconteerd en recht wordt gedaan aan de dynamiek van het wetenschappelijk bedrijf, inclusief “theory-change”, samenwerking, vooruitgang en rationaliteit.

Epiloog

Twee kanttekeningen tot slot. We hebben in dit korte essay uiteraard slechts enkele aspecten van het construeren van data kunnen bespreken. Onderzoek behelst al lang niet meer de noeste arbeid, creatieve invallen of diepzinnige overpeinzingen van individuele geesten of kleine onderzoeksgroepen, die in hun eigen vertrouwde laboratorium zelfstandig hypothe-

sen bedenken, eensgezind experimenten opzetten, deze tussentijds naar eigen inzicht of bevindingen aanpassen en die uiteindelijk zelf bepalen wat als resultaat geldt en welke conclusies daaraan verbonden kunnen worden. Galison lijkt zich bewust van het feit dat hij de aloude en eerbiedwaardige traditie van de astronomie, bij uitstek een objectieve en exacte wetenschap, een wel heel forse sociale, culturele inbedding verschaft. Hij haast zich dan ook om afstand te nemen van modieuze, filosofische bewegingen, zoals het sociaal-constructivisme, dat stelt dat de gehele werkelijkheid een constructie is en kennis volledig cultuurbepaald is. Deze stroming zou in een zeer genuanceerde versie onder meer gestalte krijgen in het werk van de natuurkundige en socioloog Andrew Pickering (1948), die op imposante wijze de sociale achtergrond van de constructie van quarks analyseert (Pickering, 1984). In andere domeinen zou het sociaal-constructivisme echter ontspreken, tot scepticisme, relativisme en “science denial” leiden en ten grondslag liggen aan de “science wars” in de VS, waarbij elke zelfverklaarde gemarginaliseerde groep zijn eigen concepties van “waarheid, werkelijkheid en weten” claimt (Starmans, 2022). In dit licht is het voorbehoud van Galison begrijpelijk, maar de kloof met Pickering lijkt soms overbrugbaar. Feit is dat van het aloude, naïeve epistemische ideaal van een objectieve, waardenvrije empirische basis weinig overblijft en dat “waarnemen” onvermijdelijk een conceptuele evolutie doorloopt. Een evidente schaduwzijde van het construeren van de data ligt op de loer. Immers als een project vele stakeholders kent, het maatschappelijk belang groot is en de politieke druk toeneemt, kunnen constructie en manipulatie een dunne scheidslijn krijgen en geraakt de onafhankelijk van de queeste naar kennis in het geding; zeker in tijden van polarisatie, pandemieën en klimaatcrises. Dat in deze “constructies” statistiek soms een methodologische en ethische waarborg kan bieden, lijkt een geruststellende gedachte.

De tweede kanttekening betreft Galisons participatie met de rol van historici bij big-science projecten. In 2016 formuleerde de bekende historicus van de statistiek Steven Stigler (1941) zijn Seven Pillars of Statistical Wisdom, te weten: “aggregation”, “information measurement”, “likelihood”, “comparison”, “regression”, “design” and “residuals” (Stigler, 2016). De auteur, zelf gastspreker op het VVSOR-jaarcongres in 2014, geeft een voor sommigen wat onorthodoxe taxonomie van concepten, maar deze zijn historisch diep verankerd en hij schetst hoe ze ontstonden, waarom en hoe ze vaak van de weeromstuit of tegen wil en dank werden toegepast en

in de praktijk verder werden ontwikkeld. De evolutie van de probabilistische deeltjesdetectoren zoals door Galison beschreven vormt niet alleen een doorwrochte historische casestudie, waaraan een theorie / taxonomie als die van Stigler kan worden getoetst, maar evenzeer een prachtige proeftuin voor grondslagen van statistiek en data science in ruimere zin. In een vorige bijdrage aan STATOR benadrukten we de unieke samenwerking tussen wetenschapshistorica Deborah Mayo en statisticus Andrew Gelman (Starmans, 2023). Op zijn eigen manier past ook Stigler in deze traditie en onderschrijft daarmee de visie van denkers als Gelman, Mayo en Galison, dat de historisch-filosofische verankering van statistische begrippen en inzichten uit de huidige onderzoekspraktijk niet alleen zinvol, maar dikwijls ook nodig kan zijn.

Referenties

- Galison, P. (1987) “How experiments end”, Chicago Press, Chicago.
- Galison, P. (1997) “Image and Logic; a material culture of microphysics”, Chicago.
- Galison, P. et al. (2023) The Next Generation Event Horizon Telescope Collaboration: History, Philosophy, and Culture. *Galaxies* 2023, 11, 32. <https://doi.org/10.3390/galaxies11010032>
- Hacking I. (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.
- Marcoci, A., Thresher, A.C., Martens, N.C.M., Galison, P. et al. (2023) Big STEM collaborations should include humanities and social science. *Nat Hum Behav* 7, 2023.
- Pickering, A. (1984) *Constructing Quarks: A Sociological History of Particle Physics*. Chicago.
- Starmans, R.J.C.M. (2022) De Sokal-affaire in retrospectief; over parodieën, crises en wantrouwen in wetenschap In: *Filosofie-Tijdschrift*. 32, 1.
- Starmans, R.J.C.M. (2023) Clio's stiefkind en de geschiedenis van het probabilistische denken, in: *STATOR*, 24, 2.
- Stigler, S. (2016) “The seven pillars of statistical wisdom”, Harvard University Press.
- Richard Starmans is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht en aan Tilburg University. Hij doet onderzoek op het snijvlak van filosofie, statistiek, informatica en AI.
E-mail: starmans@cs.uu.nl



Samenwerking VVSOR met Instats

De VVSOR gaat samenwerken met Instats (instats.org), een online platform met cursusmateriaal op het gebied van statistiek en aanverwante onderzoekgebieden. Het onderwijsmateriaal bij Instats varieert van inleidende tutorials tot gedetailleerde meerdaagse cursussen.

Onze samenwerking levert een aantal voordelen voor VVSOR-leden:

1. Op het Instats-platform kan elke gebruiker cursussen aanmaken die, naar keuze, gratis of tegen betaling zijn. Deze cursussen kunnen elke lengte en opzet hebben die gewenst is: van een kort webinar tot een meerdaagse workshop. Het platform is geïntegreerd met Zoom en ChatGPT wat het aanmaken van cursussen ondersteunt. Seminars die via Instats live gestreamd worden, worden opgenomen en bewerkt zodat ze gebruikt kunnen worden voor een on-demand streamingversie van de cursus. Al het intellectueel eigendom van de cursussen blijft bij de cursusleider, in dit geval het VVSOR-lid.
2. Eenieder die cursussen aanbiedt via Instats heeft de keuze om deze gratis of tegen een vergoeding aan te bieden. Bij cursussen die tegen betaling aangeboden worden, wordt een klein percentage afgedragen aan Instats. Voor het resterende bedrag heeft de aanbieder de keuze of dit geld gaat naar de aanbieder zelf, de werkgever van de aanbieder, de VVSOR, of een combinatie van deze groepen. Tevens kan de aanbieder groepen selecteren – bijvoorbeeld alle VVSOR-leden of alle studenten van de eigen universiteit – die gratis deel kunnen nemen aan de cursus.
3. Als VVSOR-lid krijg je 30% korting op alle (betaalde) content van het Instats platform. Om hiervoor in aanmerking te komen, moet het account aangemaakt worden via de speciale link die je kan vinden nadat je op <https://vvsor-site.e-captain.nl/mijn-captain/login> ingelogd hebt bij jouw VVSOR-account.



65th ISI World Statistics Congress

Het International Statistical Institute brengt hierbij het laatste nieuws over het 65e World Statistics Congress dat 5-9 oktober 2025 plaats zal vinden.

Het programma voor de Invited Paper Sessions is nu gereed en online beschikbaar op onze website. Het omvat een divers aanbod aan onderwerpen die de huidige staat en toekomstige ontwikkelingen van statistiek en de toepassingen daarvan weerspiegelen. U vindt sessies over onderwerpen zoals officiële statistieken, enquête-statistieken, milieustatistieken, bedrijfs- en industriële statistieken, statistische geletertheid en nog veel meer.

Ook zijn wij verheugd aan te kondigen dat professor Ionica Smeets een keynote spreker zal zijn op het congres. Als voorzitter van de vakgroep Science Communication and Society aan de Universiteit Leiden is professor Smeets bekend om haar vermogen de kloof in communicatie tussen wetenschappers en het publiek te overbruggen. Haar onderzoek, media-aanwezigheid en deelname aan het Nationaal Expertisecentrum Wetenschap & Samenleving hebben haar gemaakt tot een invloedrijk figuur in haar vakgebied. Haar unieke vaardigheid om complexe ideeën op een toegankelijke manier te presenteren met humor en herkenbare voorbeelden, belooft frisse inzichten te brengen op het congres. U kunt meer informatie vinden over de Keynote speakers op de website.

Het Short Courses Committee, met professor Peter van der Heijden als voorzitter, zal een programma voor korte cursussen samenstellen dat aansluit bij de behoeften van de deelnemers aan het World Statistics Congress. Wij nodigen u uit om bij te dragen aan dit programma, specifiek gericht op jonge statistici, om samen de inhoud vorm te geven.

Ook zijn we blij om onze samenwerking met het Centraal Bureau voor de Statistiek te kunnen aankondigen om van het congres in Den Haag een onvergetelijke gebeurtenis te maken. Van 5-9 oktober 2025 zullen meer dan 2.000 statistici en datawetenschappers van over de hele wereld samenkomen. Het programma omvat meer dan 200 sessies, inspirerende keynote sprekers, netwerk mogelijkheden en sociale activiteiten.

Meld u aan vóór 12 mei 2025 om gebruik te maken van de early-bird tarieven. Bezoek onze website voor meer informatie: <https://isi-next.org>



Bron: The Daily Echo, 10 december 2019

PEILINGPRAKTIJEN

The UK Polling Disaster

Op 22 mei 2024 kondigde de Britse premier Rishi Sunak aan dat de Britten op 4 juli 2024 naar de stembus zouden gaan voor de parlementsverkiezingen. Dat was een half jaar eerder dan veel deskundigen dachten. Waarschijnlijk hoopte de premier zo het verlies van zijn partij (de Conservatieven) binnen de perken te houden.

De peilingen in 2024

In de aanloop naar deze verkiezingen werd er weer heel veel gepeild. Vrijwel al die peilingen voorspelden een grote overwinning voor Labour. Je kunt je afvragen wat die peilingen waard zijn als ze al ver voor de verkiezingen worden gehouden. Er kan immers van alles misgaan. Het is daarom nuttig om achteraf de peilingen nog eens te vergelijken met de werkelijke uitslag van de verkiezingen in 2024.

Tabel 1 geeft een opsomming van de peilingen in

de laatste week voor de verkiezingen, in de periode na de verkiezingsdebatten tussen de partijleiders Rishi Sunak (Conservatieven) en Keir Starmer (Labour). De peilingen zijn geordend naar tijdstip van afname, van oud (onderaan) naar recent (bovenaan). De gegevens zijn afkomstig van de verkiezingswebsite van de BBC.

Alle peilingen hebben een behoorlijk grote steekproefomvang. Die ligt zo tussen de 1.000 en 20.000 personen. Die steekproefomvang zou voldoende moeten zijn om redelijk nauwkeurig schattingen te kunnen maken van de werkelijke uitslag.

De kolom "Lab" bevat de prognose (in procenten) voor Labour. Die prognoses liggen allemaal tussen de 36 en 41 procent. De kolom "Con" bevat de prognoses (ook in procenten) voor de Conservatieven. Al deze prognoses liggen tussen de 16 en 24 procent. De laatste kolom in de tabel bevat het verschil tussen de prognose voor Labour en die voor de Conservatieven. Die prognoses liggen vrijwel allemaal tussen de 13 en 20 procentpunten.

Tabel 1: Peilingen in de laatste week voor de verkiezingen van 4 juli 2024

Peiler	Datum	Steekproef	Lab	Con	Vershil
JL Partners	2-3 jul	2.005	38	23	15
Savanta	2-3 jul	2.110	39	20	19
WeThink	2-3 jul	1.325	41	23	18
Norstat	1-3 jul	3.134	37	24	13
Opinium	1-3 jul	3.010	41	21	20
Deltapol	29 jun-3 jul	1.737	39	22	17
PeoplePolling	2 jul	1.260	36	16	20
Whitestone Insight	1-2 jul	2.008	38	21	17
BMG	30 jun-2 jul	1.854	39	22	17
Techne	28 jun-2 jul	5.503	40	21	19
Redfield & Wilton	28 jun-2 jul	20.000	41	22	19
Survation	26 jun-2 jul	1.744	38	18	20
Verian	28 jun-1 jul	2.135	36	21	15
JL Partners	28 jun-1 jul	2.028	39	24	15

Uit een eerste blik op de tabel kunnen we wel de conclusie trekken dat er maar weinig verschillen zijn tussen al deze peilingen. Ze voorspellen allemaal een grote overwinning voor Labour. De voorsprong van Labour op de Conservatieven bedraagt minimaal 13 procentpunten en kan oplopen tot 20 procentpunten. Het gemiddelde over de 14 peilingen is 17,4 procentpunt.

Je kunt je afvragen of er misschien sprake is van significante verschillen tussen de peilingen, of dat er alleen maar sprake is van statistische ruis die is veroorzaakt door het loten van de steekproef. Om dit te onderzoeken, hebben we een dotplot gemaakt, zie Figuur 1. Hierbij veronderstellen we dat de steekproef van elke peiler lijkt op een aselechte steekproef.

De horizontale stippellijnen corresponderen met de verschillende peilingen. De blauwe bolletjes geven de prognoses van de peilers aan. De horizontale blauwe lijnsegmenten corresponderen met de onzekerheidsmarges. Vrijwel alle onzekerheidsmarges overlappen elkaar. Is dus geen sprake van significante verschillen tussen de peilingen.

De grafiek in Figuur 1 bevat ook de werkelijke verkiezingsuitslag. Dat is de verticale rood gekleurde lijn. Die geeft aan dat er in de officiële uitslag een verschil van 10 procentpunten is tussen Labour en de Conservatieven. Uit de grafiek blijkt duidelijk dat er sprake is van een ernstig probleem bij de peilingen. Ze voorspellen immers allemaal een veel te groot verschil. Al die verschillen zijn significant te groot. Het gemiddelde verschil is 17,4 procentpunten en dat is 7,4 procentpunten meer dan de 10 procent van de officiële uitslag.

Uit contacten met prof. John Curtice van de University of Strathclyde in Schotland (peiling-expert van de BBC) blijkt dat het elke keer weer fout gaat bij het schatten van het verschil tussen Labour en Conservatieven. Ook in 1997 en 2015 was er sprake van een overschatting van 7 procentpunten. En 1992 was er zelfs een overschatting van 9 procentpunten.

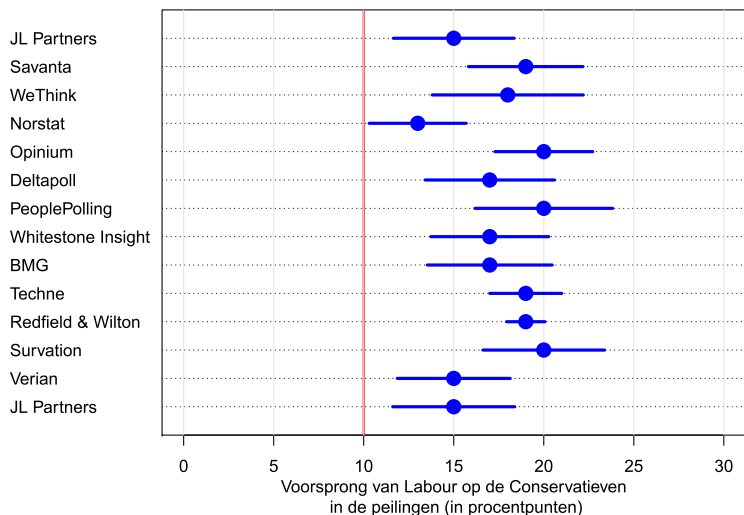
De peilingen in 2015

De peilers zaten er dus in 2024 behoorlijk naast bij de verkiezingen. Dat was niet de eerste keer. Een bijna iconisch, voorbeeld is *The UK Polling Disaster* van 2015. Die beschrijft wat er mis ging tijdens de peilingen voor de Britse parlementsverkiezingen van 2015. We beschrijven hieronder wat er toen gebeurde.

Die parlementsverkiezingen vonden in het Verenigd Koninkrijk plaats op 7 mei 2015. Er waren heel wat peilingen tijdens de verkiezingscampagne. Tot verbazing van velen, bleken de laatste prognoses van de peilers er allemaal naast te zitten. Alle peilers voorspelden een nek-aan-nekrace tussen de Conservatieven en Labour. Figuur 2 toont een dergelijke foute prognose, gemaakt op 4 april 2015. Hij is afkomstig van de peiler Opinium en werd gemaakt in opdracht van de krant The Observer. De Conservatieven en Labour stonden toen beide op 33 procent.

Een dergelijke uitslag is nogal ongebruikelijk in het Verenigd Koninkrijk. Hij zou immers coalitieonderhandelingen noodzakelijk maken, en dat gebeurt bijna nooit in het VK. Maar het liep anders. Nadat de stemmen waren geteld, bleek dat de Conservatieven

Peilingen voor de Parlementsverkiezingen in het VK in 2024



Figuur 1: Peilingen voor de parlementsverkiezingen in 2024 in het Verenigd Koninkrijk.

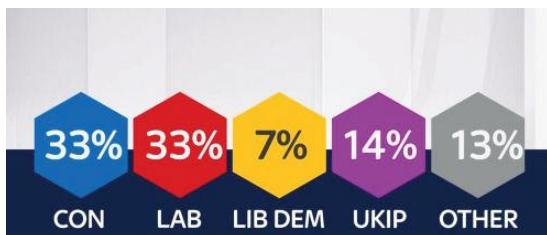
de verkiezingen hadden gewonnen met een ruime voorsprong van 6,5 procentpunten op Labour. Alle peilers zaten er naast en ze zaten er allemaal op dezelfde manier naast. Vandaar dat de deskundigen en de media spraken van een UK Polling Disaster.

De prognoses in 2015

Tabel 2 bevat voor elf peilers de prognose van het verschil tussen de Conservatieven en Labour (in procentpunten) in 2015. Het gaat hier om peilingen in

de laatste twee dagen voor de verkiezingen. Een verschil van 0 procent betekent dat er geen verschil is tussen de prognoses voor beide partijen. En een verschil van 1 procent betekent dat de prognoses heel dicht bij elkaar liggen. De tabel wijst inderdaad op een nek-aan-nekrace. Vrijwel alle verschillen zijn immers 0 of 1 procentpunt.

Merk op dat de omvang van de steekproef van de peilingen ligt tussen de (ongeveer) 1.000 en 10.000 personen. De uitkomst voor de peiler met de grootste steekproef (YouGov) is niet anders dan die voor



Figuur 2: Een peiling van de peiler Opinium voor The Observer. De prognoses voor de Conservatieven en Labour zijn hetzelfde (33%).

de peilers met kleine steekproeven. Dit is een bekend verschijnsel bij peilingen: een systematische afwijking kun je niet wegwerken door een grotere steekproef te trekken.

Onzekerheidsmarges

Figuur 3 brengt de prognoses en de uitslag in beeld in de vorm van een dotplot. De grafiek bevat voor de elf peilers de prognose van het verschil tussen de Conservatieven en Labour. De rode en blauwe bolletjes geven het verschil aan tussen de Conservatieven en Labour (in procentpunten). Alle bolletjes, en dus ook alle verschillen tussen beide partijen, liggen dicht bij 0. Ook de dotplot geeft dus aan dat er sprake is van een nek-aan-nekrace. De horizontale gekleurde lijnstukken duiden de onzekerheidsmarges aan. De marge van YouGov is klein omdat de steekproef groot was (ongeveer 10.000 personen). De onzekerheidsmarge is een stuk groter bij, bijvoorbeeld, BMG (met een steekproefomvang van ongeveer 1.000).

De verkiezingsuitslag is ook weergegeven in de grafiek, namelijk de verticale zwarte lijn ter hoogte van de waarde 6,5. Deze uitslag valt buiten alle onzekerheidsmarges. We moeten daarom concluderen dat de prognoses significant afwijken van de werkelijke uitslag. Alle peilers zitten systematisch te laag.

In Tabel 2 is ook aangegeven op welke manier de gegevens zijn verzameld. De punten/lijnen van de online peilingen zijn blauw en die van de telefonische peilingen zijn rood. Er lijkt weinig verschil te zijn tussen de uitkomsten van de online peilingen

en de telefonische peilingen. Ze zitten er allemaal evenveel naast.

Onderzoek van de afwijkingen

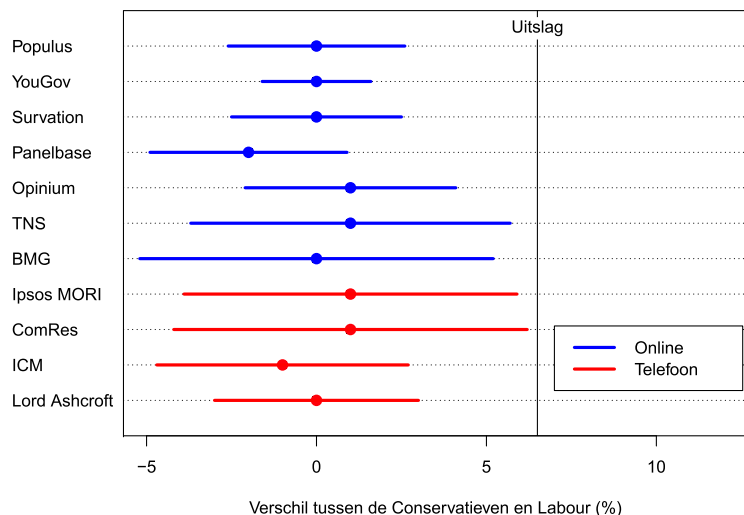
De uitkomsten van de peilingen leidden tot veel kritiek op de Britse peilers. Dat was aanleiding voor de British Polling Council (BPC) een onafhankelijk onderzoek in te stellen. De BPC is een organisatie van opinieonderzoekers die regelmatig uitkomsten van peilingen publiceren. De onderzoekscommissie concludeerde dat de belangrijkste oorzaak van de problemen het gebrek aan representativiteit van de steekproeven was. De online peilers werkten met steekproeven uit online panels die niet representatief waren. En de telefonisch peilers werkten met Random Digit Dialing (RDD, het willekeurig genereren van telefoonnummers), waarbij meestal heel veel non-respons optreedt

De onderzoekscommissie keek ook nog naar andere mogelijke oorzaken van de problemen. Eén mogelijk probleem was de Shy Tory Factor. Dat is het verschijnsel dat vooral stemmers op de Conservatieven dat niet willen toegeven in peilingen. Ze zeggen dan maar dat ze niet gaan stemmen. Dat zou leiden tot een ondervertegenwoordiging van de Conservatieven in de peilingen. De onderzoekscommissie kwam tot de conclusie dat er geen grote Shy Tory Factor was geweest.

De onderzoekscommissie onderzocht ook een mogelijke *Late Swing*. Dat is het verschijnsel dat stemmers op het laatste moment, na de laatste peilingen, toch nog van mening veranderingen. Ze stemmen dan dus op een andere partij dan ze hebben ge-

Tabel 2: De prognoses van het verschil tussen Conservatieven en Labour van elf Britse peilers in 2015. Ze voerspelden allemaal een nek-aan-nekrace.

Peiler	Modus	Steekproef	Verskil
Populus	Online	3.917	0
YouGov	Online	10.307	0
Survation	Online	4.088	0
PanelBase	Online	3.019	-2
Opinium	Online	2.916	1
TNS	Online	1.185	1
BMG	Online	1.009	0
Ipsos MORI	Telefoon	1.186	1
ComRes	Telefoon	1.007	1
ICM	Telefoon	2.023	-1
Lord Ashcroft	Telefoon	3.028	0



Figuur 3: Een dotplot met de prognoses van elf Britse peilers in 2015 inclusief de bijbehorende onzekerheidsmarges.

noemd in de laatste peiling. De onderzoekscommissie kon geen Late Swing aantonen.

Het was opmerkelijk dat alle peilers op dezelfde manier in de fout gingen. Dit zou een aanwijzing kunnen zijn voor *herding*. Dit is het verschijnsel dat peilers bij het opzetten en uitvoeren van hun peiling bewust of onbewust beslissingen nemen die er voor zorgen dat hun prognoses niet teveel afwijken van de prognoses van andere peilers. Dat verkleint het risico voor een peiler dat hij als enige met een foute voorspelling komt. De onderzoekscommissie kon niet volledig uitsluiten dat er sprake was geweest van enige vorm van *herding*.

In Tabel 2 en Figuur 3 is te zien dat er twee soorten peilingen waren: online peilingen en telefonische peilingen. De eerste zeven peilers deden online peilingen. Ze hadden allemaal de beschikking over online panels waarvoor de leden waren gerekruteerd via zelfselectie. Ze liepen dus een groot risico dat hun peilingen niet representatief waren. En daarom waren de steekproeven uit deze panels vermoedelijk ook niet representatief. De peilers probeerden het gebrek aan representativiteit te repareren door het uitvoeren van een wegingsprocedure. Het was de conclusie van de onderzoekscommissie dat die wegingen onvoldoende effectief waren. Ze waren

niet in staat de vertekeningen weg te werken. Vier peilers gebruikten telefonische peilingen. Omdat telefoonboeken ernstige dekkingsproblemen hebben (niet alle vaste nummers staan er in en mobiele nummers ontbreken geheel), hebben ze gebruik gemaakt van RDD. Zo kun je zonder dekkingsproblemen zowel vaste als mobiele nummers genereren. Probleem van deze aanpak is de lage respons. De respons ligt vaak niet hoger dan 20 procent en is soms zelfs lager dan 10 procent. Dit betekent dat de vertekening ten gevolge van non-respons aanzienlijk kan zijn. Wegen hielp ook hier niet om het gebrek aan representativiteit te repareren.

Merk op dat de telefonische peilingen zo ongeveer verdwenen waren in 2024. Vrijwel alle peilingen waren gebaseerd op steekproeven uit online panels. Slechts enkele peilingen bestonden uit een mix van online data en telefonische data. Kennelijk zijn de peilers er niet in geslaagd om de peilingen behoorlijk te verbeteren. Dat zal vooral moeten gebeuren door het verbeteren van de procedures die corrigeren voor het gebrek aan representativiteit van de online peilingen.

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



80 Years of
Statistics and Operations Research
in the Netherlands

Past, Present and Beyond

VVSOR

Annual meeting
March 20, 2025

Oproep voor het insturen van abstracts voor de Annual Meeting

Het thema van de annual meeting op 20 maart 2025 is terugkijken en vooruitblikken op onderwerpen en prestaties van Statistiek en Operations Research. In het bijzonder omdat dit een jubileumjaar is: de VVSOR bestaat 80 jaar. Wij nodigen u uit om uw voorstellen voor een presentatie in te dienen vóór **16 december 2024**.

Ingezonden abstracts kunnen gekozen worden als korte presentaties in een parallelsessie of voor een posterpresentatie. De gekozen abstracts worden gepubliceerd op de VVSOR-website en in STATOR van maart 2025. Het voorstel moet in ieder geval het volgende bevatten:

- Voorkeursformaat (presentatie/poster)
- Titel
- Een abstract van maximaal 250 woorden, waarin de inhoud en doelstellingen van de presentatie of poster wordt beschreven
- Een korte biografie van uzelf
- Kernwoorden om het hoofdthema te beschrijven

Gebruik de link of de QR-code om een abstract in te sturen. Inzenden via e-mail is ook mogelijk.

Bericht van acceptatie ontvangt zult u uiterlijk **15 januari 2025** ontvangen.

Wij kijken uit naar uw voorstellen en deelname. Heeft u een vraag of wilt u meer informatie, neem dan gerust contact met ons op via annualmeeting@vvsor.nl.



<https://shorturl.at/1XHLE>



VVSOR

Uitmuntende master's of PhD thesis begeleid?

Oproep om kandidaten te nomineren voor de Jan Hemelrijk en Willem R. van Zwet Awards 2024

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs/hoger beroepsonderwijs looft de VVSOR al vanaf 1989 een scriptieprijs uit. In 2014 kreeg deze de naam Jan Hemelrijk Award. Afgelopen jaar was er geen winnaar. Sinds 2012 is er ook een prijs voor dissertaties: de Willem R. van Zwet Award. Deze werd vorig jaar gewonnen door Richard Post.

De afgelopen jaren kwam het overgrote deel van de nominaties uit de hoek van de operations research en de mathematische statistiek. Zonder iets af te willen doen aan de goede kwaliteit van deze nominaties, is de jury ervan overtuigd dat ook in de andere secties uitstekend promotie- en afstudeeronderzoek gedaan wordt. Ook statistisch onderzoek met een meer toegepast karakter kan in aanmerking komen voor deze prijs.

Wij willen dan ook supervisors (begeleiders) uit alle vakgebieden van harte uitnodigen om een uitmuntende afstudeerscriptie (Master) of dissertatie (Ph.D.) te nomineren. De indiening van een nominatie dient vergezeld te gaan van een aanbevelingsbrief van de supervisor van de genomineerde. De precieze procedure voor beide prijzen, alsmede de reglementen en het nominatieformulier zijn te downloaden via de website van de VVSOR, www.vvsor.nl. De nominatie dient uiterlijk **26 januari 2024** binnen te zijn.

Namens de VVSOR,

Dr. Ad Ridder, juryvoorzitter Jan Hemelrijk Award
Prof. dr. Jelle Goeman, juryvoorzitter Willem R. van Zwet Award
Dr. Sander Scholtus, Secretaris der beide jury's
Prof. dr. Casper Albers, voorzitter VVSOR