STAtOR

periodiek van de VVSOR jaargang 19, nummer 3, september 2018

STAtOR

Jaargang 19, nummer 3, september 2018

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

Redactie

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

Kopij en reacties richten aan

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

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl
Secretaris a.i.: dr. Laurence Frank, db@vvsor.nl
Penningmeester: dr. Ad Ridder, db@vvsor.nl
Overige bestuursleden: dr. Eric Cator (SMS), prof. dr. Ernst Wit (BMS), Maarten Kampert MSc., prof. dr. Albert Wagelmans (NGB), dr. Michel van de Velden (ECS), prof. dr. Jelte Wicherts (SWS), Elian Griffioen (Young Statisticians).

Leden- en abonnementenadministratie van de VVSOR

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

Advertentieacquisitie

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

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

ISSN 1567-3383

WARM

Wat was het warm deze zomer. Maar het werk aan *STAtOR* ging gewoon door. Plasjes zweet op het toetsenbord en automatisch uitgeschakelde computers waren niet ongewoon. Maar omdat iedereen *STAtOR* een warm hart toedraagt is het allemaal in orde gekomen. Het resultaat ligt voor u.

Onze taal kent veel uitdrukkingen en zegswijzen die met warmte te maken hebben, zoals 'het gaat er warm aan toe' en 'een warm onthaal', om er maar een paar te noemen. Best vreemd voor een land waar het in het verleden nooit zó extreem warm is geweest.

Zoals altijd is dit een afwisselend nummer geworden. Han Hoogeveen beschrijft hoe men op een optimale wijze stekjes van moederplanten kan oogsten, Sander Scholtus heeft het over het automatisch opsporen van fouten in de officiële statistiek en Stef Kleinluchtenbeld en Gerhard Post vergelijken verschillende methoden voor het toewijzen van voorkeursactiviteiten.

Verder een artikel over het veilig en efficiënt inspecteren van het spoor. Mocht er na deze warme zomer een stormachtige herfst voor blaadjes op de rails en vierkante wielen zorgen dan weet u dat vanuit ons vakgebied een belangrijke bijdrage is geleverd om tóch over betrouwbare rails te reizen.

In ons dichtbevolkte land zijn we gewend dat ambulances altijd zeer snel ter plaatse kunnen zijn, mede op basis van onderzoek naar de optimale lokalisering van de posten. In het dunbevolkte Noorwegen lost men dit op met helikopters en met een Nederlandse inbreng heeft men ook hier de optimale verdeling daarvan over het land bepaald. Jo Røislien schrijft er een boeiend artikel over.

In Groningen ging het er in het verleden ook warm aan toe is? Wiskundigen gebruikten de methoden van Newton en Leibniz en sommeerden oneindig veel oneindig kleine rechthoeken tot een oppervlakte. De theologen vonden dat maar niets: hier werd iets uit niets geschapen en dat is alleen de Schepper gegeven! Heftige polemieken met Johann Bernoulli in het middelpunt, Gerard Sierksma vertelt er smakelijk over en John Poppelaars wijdt zijn column aan de hype rond het verzamelen van gigantische hoeveelheden data.

Tijdens de Annual Meeting van dit jaar werd Richard Gill benoemd tot erelid van de VVSOR, Maarten Kampert vertelt waarom in een ode aan deze warme en kleurrijke statisticus.

Ten slotte kunnen we melden dat de redactie onderzoekt of en hoe er kan worden samengewerkt met het blog Wetenschap.nu. Als eerste stap hebben we een artikel van Sanne Willems op dit blog overgenomen.

Wij hopen dat u een warm gevoel krijgt bij dit nummer, en wensen u veel leesplezier.

INHOUD

Tegenwoordig verzamelt ieder zichzelf respecterend bedrijf data van al zijn processen; we zijn in het tijdperk van de Big Data beland, en 'machine learning' is het toverwoord. Op het NGB-seminar van 18 januari 2018 was iedereen het erover eens dat big data grote kansen biedt voor de Operations Research, maar het is nog onduidelijk hoe we die kansen kunnen benutten. In dit artikel laten we zien hoe een basale vorm van datamining kan worden gecombineerd met lineaire programmering bij het oplossen van een belangrijk probleem bij de teelt van bloemen en planten.

STEKJES OOGSTEN

een combinatie van optimalisering en datamining

Han Hoogeveen

Calibrachoa Aloha Kona Dark Red van Dummen Orange

Maak kennis met persoon X: een enthousiaste amateur wielrenner. Hij wil geld ophalen voor een goed doel door de Alpe d'Huez op te fietsen. Omdat het voor hem geen uitdaging is om boven te komen (dat hij heeft hij al diverse malen gedaan) wil hij zich voor ieder van de 21 bochten laten sponsoren: voor iedere bocht is een deadline afgesproken, en hij verdient een behoorlijk bedrag door die deadline te halen en nog een extra bonus voor iedere minuut dat hij er is voordat de deadline is verstreken. De betreffende dag is aangebroken, en X voelt zich in topvorm. Soepel rondt hij de eerste paar bochten en voelt dat hij nog veel over heeft: tijd om te versnellen om de extra bonus voor de volgende bochten te incasseren. Enthousiast pakt hij een flinke bonus in de volgende bochten, maar vanaf bocht 10 gaat het mis en haalt hij de deadlines niet meer: hij heeft teveel gegeven in het begin.

Voldoende stekjes

Een soortgelijk probleem speelt ook bij het oogsten van stekjes bij Dümmen Orange, een internationaal bedrijf dat zich onder andere bezig houdt met het kweken en veredelen van bloemen en planten. De stekjes worden geproduceerd door moederplanten te laten groeien waar iedere week één of meer stekjes van worden afgesneden. Vooraf zijn contracten afgesloten die aangeven hoeveel stekjes per week geleverd moeten worden (vergelijkbaar met de deadlines per bocht) en een eventueel overschot wordt op de markt verkocht (vergelijkbaar met de extra bonus per bocht). Het aantal benodigde moederplanten wordt bepaald op basis van de bekende leververplichtingen met een buffer van 10%. Wanneer het bestaan van deze buffer wordt onthuld, gaat de afdeling Verkoop enthousiast aan de slag, maar wanneer er teveel extra worden verkocht, kan later niet aan de verplichtingen worden voldaan (de wielrenner mist de deadlines van de latere bochten). Dümmen Orange heeft dit probleem aangemeld bij Wiskunde voor de Industrie 2016[1]; dit artikel is gebaseerd op de toen voor dit probleem gevonden oplosmethode.

Lineaire programmering

Simpel gesteld komt het probleem van Dümmen Orange erop neer dat we moeten bepalen hoeveel moederplanten we nodig hebben en hoe vaak we welke snijpatronen moeten gebruiken, waarbij een snijpatroon aangeeft hoeveel stekjes we gemiddeld afsnijden per moederplant per week; noem dit aantal a_{jt} voor snijpatroon A_j in week t. Deze a_{jt} waarden kunnen fractioneel zijn, omdat het gaat om een gemiddelde opbrengst over alle moederplanten. Uiteraard mogen we alleen snijpatronen gebruiken die niet te veel vragen van een moederplant; zulke snijpatronen noemen we *geldig*. In de context van de fietser komt een snijpatroon overeen met een schema voor een beklimming, waarbij alle doorkomsttijden zijn gespecificeerd; we bekijken alleen schema's die X vol kan houden. Als alle mogelijke geldige snijpatronen bekend zijn, dan kunnen we het minimale aantal benodigde moederplanten bepalen door het volgende lineair programmeringsprobleem (LP) op te lossen:

$$\begin{aligned} &\min M = \sum_{j=1}^{n} x_j \\ &\text{onder de voorwaarden} \\ &\sum_{j=1}^{n} a_{jt} x_j \geq b_t \qquad \forall t = 1, \ldots, T \\ &x_j \geq 0 \qquad \forall j = 1, \ldots, n \end{aligned}$$

Hierin geeft x_j het aantal moederplanten aan waarop we snijpatroon A_j toepassen, en b_t is het aantal stekjes dat nodig is in week t. Aangezien het om grote aantallen gaat is het niet nodig om te eisen dat x_j geheeltallig is; een toegelaten oplossing kan worden gevonden door naar boven af te ronden. Deze oplossing blijkt ook optimaal te zijn, omdat er in een optimale oplossing precies één snijpatroon wordt gebruikt, zoals we later zullen zien.

Datamining

Helaas is er geen lijst beschikbaar met snijpatronen die gebruikt kunnen worden. Geen nood, een impliciete lijst in de vorm van een verzameling beperkingen waar een geldig snijpatroon aan moet voldoen is ook goed: we kunnen dan met behulp van *kolomgeneratie* (Ford & D Fulkerson, 1958) het bovenstaande LP oplossen, net zoals bij het *cutting stock* probleem (Gilmore & Gomory, 1961; 1963). Echter, die aanpak gaat voor Dümmen Orange ook niet op: een belangrijke onderzoeksvraag is juist het vinden van zo'n lijst van eisen die noodzakelijk en voldoende zijn voor de geldigheid van een snijpatroon. Wat wel beschikbaar is zijn de oogstdata van de afgelopen tien jaar; in de context van de fietser zijn dit de doorkomsttijden van tien eerdere beklimmingen. Uit deze data zijn voor de hand liggende vuistregels af te leiden: aangezien er in geen enkele week gemiddeld meer dan 2,0 stekjes zijn geoogst, mogen we eisen dat $a_{jt} \leq 2{,}0$. Uiteraard kunnen er ook andere redenen aan ten grondslag liggen dat a_{jt} altijd ten hoogste 2,0 is, maar volgens de domeinexpert

was $a_{jt} \leq 2.0$ een redelijke eis. Gesterkt door dit succes hebben we vervolgens door middel van een basale vorm van *data mining* een hele lijst met beperkingen afgeleid waar ieder snijpatroon $(a_1, a_2, \ldots, a_T)$ aan moet voldoen, waarbij a_t de gemiddelde oogst per moederplant in week t is. Hiertoe bepalen we voor ieder k-tupel $(t_1, \ldots, t_k)$ met $t_1 < t_2 < \ldots < t_k$, waarbij $k \leq 6$ en $t_k - t_1 \leq 10$ de minimale waarde Q zodanig dat voor ieder jaar van de oogstdata geldt

$$a_{t_1} + a_{t_2} + \ldots + a_{t_k} \leq Q.$$

Dit leidt tot een hele lijst met beperkingen die de geldigheid van een snijpatroon beschrijven; een kleine bloemlezing hieruit (deze moeten gelden voor alle t):

a_t	$\leq 2{,}0$
$a_t + a_{t+1}$	$\leq 3{,}90$
$a_t + a_{t+2}$	$\leq 3{,}85$
$a_t + a_{t+1} + a_{t+2}$	$\leq 5{,}75$
$a_t + a_{t+2} + a_{t+4}$	$\leq 5{,}71$
$a_t + a_{t+1} + a_{t+2} + a_{t+3}$	$\leq 7{,}60$
$a_t + a_{t+1} + a_{t+3} + a_{t+4}$	$\leq 7{,}61$
$a_t + a_{t+1} + a_{t+2} + a_{t+3} + a_{t+4}$	$\leq 9{,}46$
$a_t + a_{t+1} + a_{t+3} + a_{t+4} + a_{t+5}$	$\leq 9{,}21$
$a_t + a_{t+1} + a_{t+2} + a_{t+3} + a_{t+4} + a_{t+5}$	$\leq 11{,}15$
$a_t + a_{t+1} + a_{t+3} + a_{t+4} + a_{t+5} + a_{t+6}$	$\leq 10{,}90$

Hoeveel moederplanten zijn er minimaal nodig?

Met behulp van de volledige lijst met beperkingen is het mogelijk om met behulp van kolomgeneratie het gegeven LP op te lossen, maar het blijkt dat er een elegantere en veel simpelere oplossing beschikbaar is. De gevonden beperkingen zijn allemaal lineair (want dit zijn de enige beperkingen waar we naar hebben gezocht) en dit impliceert dat iedere *convexcombinatie* $\lambda_1 A_1 + \ldots + \lambda_k A_k$ van geldige snijpatronen $A_1, \ldots, A_k$ ook een geldig snijpatroon is. Als $(x_1^*, \ldots, x_n^*)$ nu een optimale oplossing van het LP is met uitkomstwaarde $M = x_1^* + x_2^* + \ldots + x_n^*$, dan kunnen we hetzelfde resultaat bereiken door M maal het snijpatroon te gebruiken dat gelijk is aan

$$\sum_{j=1}^{n} \frac{x_j^*}{M} A_j.$$

Hieruit volgt dat we kunnen volstaan met precies één snijpatroon, en aangezien we in periode t minstens b_t stekjes nodig hebben, is het optimaal om gemiddeld $a_t = b_t/M$ stekjes te oogsten in periode t. Nu moeten we nog de waarde van M bepalen zodanig dat het snijpatroon $(b_1/M, \ldots, b_T/M)$ aan de beperkingen voldoet.

Beschouw een willekeurige eis, bijvoorbeeld $a_t + a_{t+1} \leq 3{,}9$. Als we hierin $a_t = b_t/M$ invullen en uitwerken, dan vinden we dat $b_t + b_{t+1} \leq 3{,}9\, M$, waaruit volgt

$$M \geq \frac{b_t + b_{t+1}}{3{,}9} \qquad \text{voor alle } t = 1, \ldots, T-1.$$

Op deze manier kan iedere beperking worden vertaald naar een ondergrens op M, waarna we M gelijk kunnen kiezen aan de maximale ondergrens.

Maximaliseren extra opbrengst

Uiteraard kan er worden besloten om meer moederplanten te planten. Indien bekend is hoeveel stekjes extra kunnen worden verkocht in week t (noem dit aantal D_t) tegen prijs p_t, dan rijst natuurlijk de vraag: hoeveel moederplanten moeten we planten en hoeveel stekjes moeten we per week oogsten? Ook nu kunnen we bewijzen dat alle moederplanten volgens hetzelfde snijpatroon moeten worden gesneden. We gebruiken M als de variabele die aangeeft hoeveel moederplanten moeten worden geplant; dit brengt kosten $c(M)$ met zich mee, waarvan we eisen dat deze functie lineair is in M. Definieer z_t als het aantal stekjes dat wordt geoogst in week t. Aangezien we maar één snijpatroon gebruiken, worden er gemiddeld $a_t = z_t/M$ stekjes per moederplant geoogst in week t; deze waarden a_t moeten aan de geldigheidsbeperkingen voldoen. Dit is niet lineair, aangezien M een variabele is, maar als we iedere beperking met M vermenigvuldigen, dan houden we lineaire beperkingen over: de beperking $a_t + a_{t+1} \leq 3{,}9$ gaat dan over in $z_t + z_{t+1} \leq 3{,}9\, M$.

Van deze z_t stekjes hebben we er b_t nodig voor de vaste contracten, en de rest kan worden verkocht. De opbrengst moet uiteraard worden gemaximaliseerd, en dit leidt tot het volgende LP:

$$\max \sum_t p_t (z_t - b_t) - c(M)$$
onder de voorwaarden
$$b_t \leq z_t \leq b_t + D_t \quad \forall t$$
'$z_1, \ldots, z_T, M$ voldoen aan de geldigheidsbeperkingen'

In het bovenstaande LP wordt er van uit gegaan dat alle extra stekjes voor dezelfde prijs kunnen worden verkocht. Dit kan uiteraard worden veralgemeniseerd door een stuksgewijs-lineaire opbrengstfunctie te gebruiken. Evenzo kunnen we voor $c(M)$ een stuksgewijs-lineaire kostenfunctie gebruiken.

Conclusie

In het bovenstaande hebben we aangegeven hoe datamining kan worden gebruikt binnen de Operations Research; voor zover wij weten is dit de eerste toepassing van datamining in OR in plaats van omgekeerd. Onze basale vorm van datamining levert simpele, lineaire beperkingen op die eenvoudig te combineren zijn met LP, wat een groot voordeel is ten opzichte van het gebruik van geavanceerde technieken uit de machine learning.

Dit artikel is gebaseerd op het rapport van Han Hoogeveen met co-auteurs Jakub Tomczyk en Tom C. van der Zanden. Verder wil ik nogmaals de organisatoren van Wiskunde voor de Industrie 2016 en Dümmen Orange bedanken.

Noot

1. Studiegroep Wiskunde met de Industrie. http://www.ru.nl/math/research/vmconferences/swi-2016/

Literatuur

Ford, J.L.R., & Fulkerson, D.R. (1958). A suggested computation for maximal multicommodity network flows. *Management Science*, 5, 97–101. http://dx.doi.org/10.1287/mnsc.5.1.97.

Gilmore, P.C., & Gomory, R.E. (1961). A linear programming approach to the cutting-stock problem. *Operations Research*, 9, 849–859. http://dx.doi.org/10.1287/opre.9.6.849.

Gilmore, P.C., & Gomory, R.E. (1963). A linear programming approach to the cutting stock problem: Part ii. *Operations Research*, 11, 863–888. http://dx.doi.org/10.1287/opre.11.6.863.

Hoogeveen, J.A., Tomczyk, J., & Van der Zanden, T.C. (2018). *Flower power: Finding optimal plant cutting strategies through a combination of optimization and data mining*. Technical Report Utrecht University, UU-CS-2018-003. http://www.cs.uu.nl/research/techreps/UU-CS-2018-003.html

Han Hoogeveen studeerde econometrie in Groningen en deed daarna bij het CWI onderzoek op het gebied van multicriteria scheduling resulterend in een promotie aan de Universiteit Eindhoven bij Jan Karel Lenstra. Hij is als universitair docent verbonden aan de afdeling Information and Computing Sciences van de Universiteit van Utrecht. In zijn onderzoek richt hij zich vooral op het oplossen van op de praktijk geïnspireerde problemen uit de transport en logistiek met behulp van technieken uit de combinatorische optimalisering.
E-mail: j.a.Hoogeveen@uu.nl

NGB YOUNG OR DAY

We have the great pleasure of inviting you to the first NGB Young OR Day to be held at Schiphol Airport on September 28, 2018. The Young OR Day is an initiative to bring together young* professionals and academics working in the field of Operations Research.

In this first edition, we will look at applications of OR in Aviation. On this day, there will be talks by both experts from KLM and researchers from academia. There will also be a guided tour, and we will end the day with a social event. Note that all presentations will be in English.

This event is free for NGB members; 25 euros for non members. Sign up by sending your name, date of birth, company/institute, and job title to YoungOR@vvsor.nl

* There is no age limit. If you feel young, you qualify. If you don't feel young, perhaps you have young colleagues that could be interested in this event. Please feel free to forward this information to them.

We hope to see you on September 28!

The organizing committee,
Renée Buijs, Ortec
Martijn van Ee, Netherlands Defence Academy
Caroline Jagtenberg, University of Auckland
Bernard Zweers, CWI & Free University Amsterdam

VEILIGE EN EFFICIËNTE INSPECTIE VAN HET SPOOR

Met het project 'Veilige en efficiënte inspectie van het spoor' wonnen de bedrijven Inspectation[1] en CQM[2] de Hendrik Lorentz Prijs[3]. In dit project wordt onder meer deep-learningtechniek intelligent ingezet in combinatie met visualisatietechnieken waardoor spoorweginspecteurs veel sneller en effectiever gebreken in het spoor kunnen identificeren en controleren.*

Johan van Rooij & Huub van den Broek

Veilige semiautomatische inspectie van het spoor

Nederland heeft het drukst bereden spoornet van Europa. Gemiddeld maken reizigers elke dag 1,1 miljoen treinreizen – in totaal 17 miljard reizigerskilometers per jaar. Het is dan ook van groot belang zorgvuldig en efficiënt met het spoor om te gaan. Door monitoring en inspectie van spoorstaven en wissels is het mogelijk beginnende defecten aan het spoor eerder te zien en vroegtijdig in te grijpen. Dit verhoogt niet alleen de veiligheid, maar ook de beschikbaarheid van het drukbezette spoor.

Inspectation is een onderneming gespecialiseerd in advisering en engineering van beheer, onderhoud en conditie-monitoring van infrastructuur.

Sherloc, de meettrein van Inspectation, levert een helder, accuraat en actueel beeld van de kwaliteit van wissels en sporen. Sherloc is voorzien van verschillende lasersystemen en HD-fotocamera's. Deze treinen maken drie à vier keer per jaar van elk stukje spoor een foto. Dat levert een enorme hoeveelheid beeldmateriaal op dat door inspecteurs van Inspectation wordt bekeken. Dit inspectieproces is een veel veiliger en efficiënter alternatief ten opzichte van de vroegere situatie waarbij inspecteurs langs het spoor liepen.

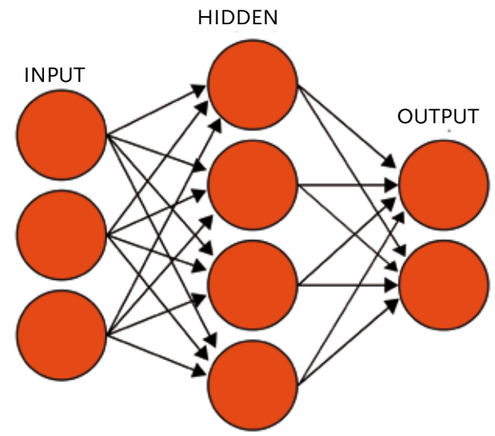
Met behulp van deep learning hebben wij, samen met Inspectation, een systeem ontwikkeld voor het automatisch detecteren van defecten in spoorstaven gebaseerd op het beeldmateriaal geproduceerd door Sherloc. In de nieuwe opzet bekijkt de inspecteur slechts die stukken spoor die door het algoritme als verdacht worden bestempeld. Dit betekent dat hij 80% van de beelden niet meer hoeft te bekijken. De overige foto's bekijkt hij nog wel, maar dan om te controleren of er daadwerkelijk van een defect sprake is en vervolgens te beoordelen welke acties nodig zijn om het eventuele defect te repareren. Men ziet beginnende defecten eerder en kan vroegtijdig ingrijpen als dat nodig is. Dit verhoogt niet alleen de veiligheid, maar vermindert ook de reizigershinder door onderhoud.

Hoe werkt deep learning?

Deep learning is een vorm van machine learning gebaseerd op neurale netwerken. Een neuraal netwerk, ondanks de op biologie geïnspireerde naam, is in zijn meest eenvoudige vorm gewoon een functie f_θ: $\mathbb{R}^n \to \mathbb{R}^k$ gerepresenteerd door een gerichte graaf (zie figuur 1). In deze graaf heeft iedere gerichte kant a een gewicht θ_a. Het uitrekenen van deze functie (het evalueren van het neurale netwerk) werkt als volgt: de input knopen (de knopen zonder binnenkomende kanten) krijgen een waarde overeenkomstig met de n verschillende coördinaten van de input. Vervolgens worden de waarden van de overige knopen bepaald door, voor iedere knoop v, $g_v\,(\Sigma_a \theta_a\, x_a)$ uit te rekenen. Hier wordt gesommeerd over alle in v binnenkomende kanten a, is x_a de waarde van de knoop waar kant a vandaan komt, en zijn de g_v vooraf gedefinieerde niet-lineaire functies. Het resultaat van de functie f_θ is uiteindelijk af te lezen als de waarden van de output knopen (knopen zonder uitgaande kanten).

Neurale netwerken zijn er in vele verschijningsvormen. Zo kun je eindeloos variëren met de topologie van het netwerk, dus ook met het aantal parameters θ_a, en met de deelfuncties g_v. In het voorbeeldplaatje zie je een gelaagd netwerk met één input laag, één interne (hidden) laag en één output laag. Wij gebruiken een topologie die ook wel een convolutioneel neuraal netwerk genoemd wordt. De term deep learning verwijst naar neurale netwerken die diep zijn, waarmee bedoeld wordt dat ze meer dan alleen een input en een output laag hebben. Het voorbeeld is dus een heel eenvoudig deep learning model.

Een neuraal netwerk wordt pas nuttig als de functie die men ermee uitrekent daadwerkelijk iets betekent. In theorie kunnen neurale netwerken iedere functie willekeurig goed benaderen, mits het netwerk groot genoeg mag zijn. Door middel van een set trainingsdata, data waarvan de gewenste uitkomst van het netwerk bekend is, wordt gezocht naar waarden voor de parameters zodanig dat het model deze data benadert, èn daarnaast op nieuwe ongeziene data ook gewenste uitkomsten geeft. Dit is niet heel anders dan voor statistische modellen,



Figuur 1. Voorbeeld van een neuraal netwerk

echter met een aantal belangrijke verschillen:

- Ten eerste is het vaak niet mogelijk en praktisch niet relevant om de best fittende parameters te vinden, zoals bij maximum likelihood, omdat het bijbehorende optimalisatie probleem niet convex is en daarnaast ook heel hoog dimensionaal. Sterker nog, om overfitting te voorkomen, wil je het optimum niet eens vinden.
- Ten tweede is het aantal parameters vaak groter (soms wel ordes groter) dan het aantal punten in de trainingsdata. Hierdoor is het optimale netwerk dus zelfs niet uniek bepaald.
- Ten slotte kijken we meer naar het neurale netwerk als *black box* waarbij we tevreden zijn als de uitkomsten op zowel de trainingsdata als de niet eerder geziene data goed genoeg zijn, zonder precies te willen begrijpen waarom juist deze model parameters werken.

Door implementaties op krachtige grafische kaarten die veel meer rekenkracht hebben dan moderne processoren is het mogelijk neurale netwerken van significante omvang op niet al te dure hardware te trainen. Wij gebruiken hier een Titan X grafische kaart van Nvidia voor en het TensorFlow framework van Google.

Tot zover de theorie, nu de toepassing.

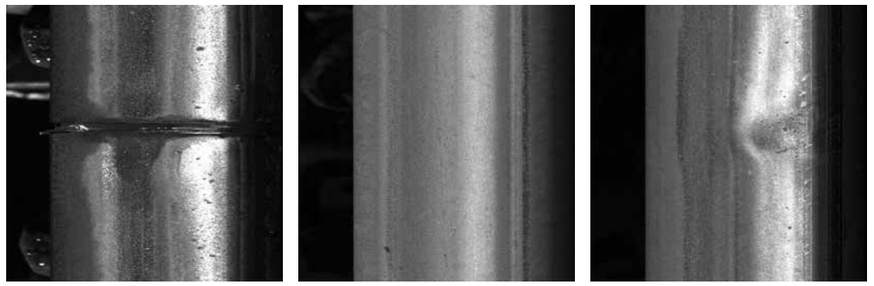
Sherloc levert een film van het spoor. Als je de beelden van deze film achter elkaar legt ontstaat een hele grote langwerpige foto van het spoor. Om deze beelden geschikt te maken voor classificatie met een neuraal netwerk gaan we de film eerst omzetten in losse plaatjes van een kleiner, vast formaat: de inputs x van het neurale netwerk. Zie de resulterende plaatjes in figuur 2. Vervolgens zorgen we ervoor dat het netwerk een getal tussen 0 en 1 als output geeft dat te interpreteren is als de kans dat dit specifieke plaatje een defect bevat.

Omdat de film ongeveer 3 keer zo breed is als de spoorstaaf, en deze spoorstaaf niet altijd op dezelfde plek in de film ligt, zoeken we eerst de spoorstaaf op. Dit doen we door per frame, per kolom van pixels in het frame, een histogram van de grijswaarden te maken. Uit deze histogrammen (en de locatie van de spoorstaaf in het vorige frame) is op te maken waar de spoorstaaf ligt. Hierna, kunnen we een venster (*sliding window*) over de spoorstaaf schuiven, en om de zoveel pixels een plaatje maken.

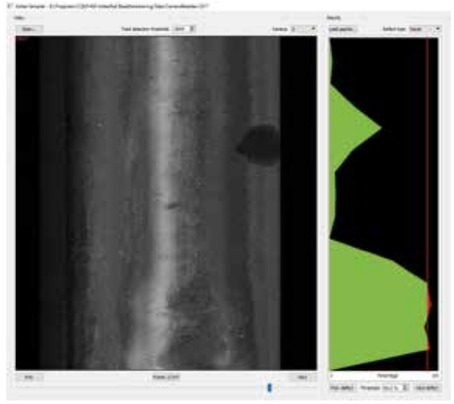
Om een verzameling trainingsdata te maken, worden films handmatig door inspecteurs geannoteerd met waar de defecten zich bevinden. Dit is te vertalen naar of een plaatje binnen het sliding window een defect bevat of niet. Als het netwerk uiteindelijk gebruikt wordt om defecten in het spoor te detecteren dan levert dit per verschuiving van het venster een kans op een defect op. Dit visualiseren we naast het spoor als een grafiek met op de verticale as de afstand over het spoor en op de horizontale as de kans op een defect, zie figuur 3.

De gebruiker kan vervolgens kiezen bij welke kans hij een defect wil zien en zo de *trade-off* tussen de hoeveelheid werk en de kans dat er defecten gemist worden beinvloeden. Hierbij is het van groot belang dat het aantal *false negatives* (wel een defect, niet gedetecteerd) klein is, omdat de inspecteur deze dus niet meer ziet, terwijl *false positives* (geen defect, wel als dusdanig gedetecteerd) niet zo erg zijn. Gelukkig voor de staat van ons spoornet speelt hierbij de extra complicatie dat we juist heel veel beelden van schoon spoor hebben terwijl het aantal beelden van defecten vele ordes van grootte kleiner is. Hiermee houden we expliciet rekening met het samenstellen van de trainingsdata en tijdens het trainen straffen we fouten op het classificeren van defecten veel harder af dan classificatiefouten op schoon spoor.

Een van de grootste uitdagingen van dit project was het zodanig samenstellen van de trainingsdata dat hierin zo veel mogelijk van alle natuurlijke variatie die in de praktijk voorkomt terugkomt. Zo moet het netwerk leren dat *squats* wel defecten zijn, maar tegelijkertijd dat regendruppels en vetvlekken dit niet zijn. Daarnaast kan het



Figuur 2. Links een stuk spoor met een ES-las (geen defect), in het midden schoon spoor, en rechts een squat categorie B (een defect)



Figuur 3. Visualisatie van de kans op een defect

spoor er in verschillende omstandigheden, zoals bij laagstaande zon, ineens heel anders uit zien.

Ten slotte, hebben we om het vertrouwen in het netwerk te vergroten geëxperimenteerd met verschillende technieken om zichtbaar te maken op welke onderdelen van een input plaatje het neurale netwerk zijn beslissing baseert, zie figuur 4. Dit is essentieel om inzicht te krijgen in de werking van de black box en om vertrouwen in het systeem te kweken bij inspecteurs.

Conclusie

Automatische spoorinspectie met behulp van deep-learningtechnieken is een voorbeeld van hoe data science in de praktijk kan werken. CQM ontwikkelde samen met Inspectation een zelflerend systeem voor het automatisch detecteren van defecten op het spoor. In de toekomst biedt de huidige oplossing mogelijkheden om de meest bereden kritische sporen nog vaker te inspecteren en de kwaliteit nog verder verhogen. Verder kan de oplossing ook worden ingezet op het buitenlandse spoornetwerk of metro- en tramnetwerken. Daarnaast is de onderliggende techniek geschikt om toe te passen voor het inspecteren van wegen, bruggen en tunnels.

Noten

1. Inspectation, voortgekomen uit onderhoudsaannemer VolkerRail, is een onderneming gespecialiseerd in advisering en engineering van beheer, onderhoud en conditie monitoring van infrastructuur.
2. CQM (Consultants in Quantitative Methods) is een adviesbureau op het gebied van logistieke optimalisatie, industriële statistiek en data science gevestigd in Eindhoven.
3. De Hendrik Lorentz Prijs is bedoeld voor een organisatie binnen het bedrijfsleven of de overheid die op een onderscheidende en innovatieve manier data science toepast. https://nederlandsedatascienceprijzen.nl

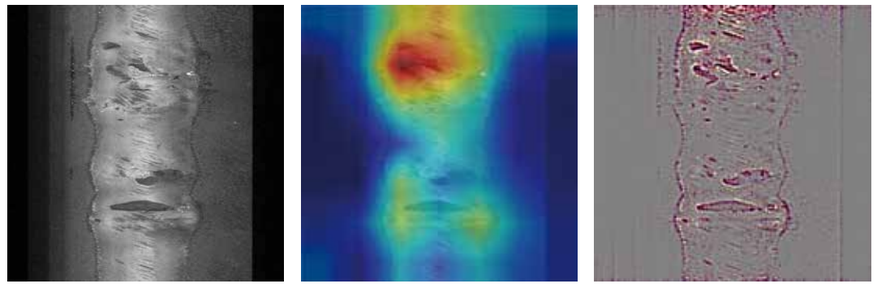
Literatuur

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge MA: MIT Press [een gedegen inleiding in het vakgebied].

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Gradient-weighted Class Activation Mapping*. ICCV

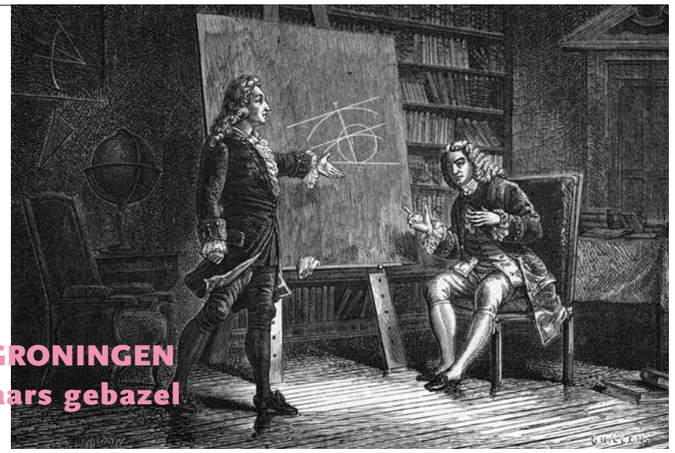
Johan van Rooij is als Senior Consultant werkzaam bij CQM. Daarnaast heeft hij een aanstelling als universitair docent aan het Departement Informatica aan de Universiteit Utrecht.
E-mail: Johan.vanRooij@cqm.nl

Huub van den Broek is als Senior Consultant werkzaam bij CQM en is projectleider van dit project.
E-mail: Huub.vandenBroek@cqm.nl



Figuur 4. Links een stuk spoor met overduidelijke defecten; in het midden een heatmap van regio's in de afbeelding waarop het neurale netwerk zijn beslissing baseert; rechts een meer gedetailleerde weergave van de belangrijkste pixels. (Naar Selvaraju et al., 2017)

Gerard Sierksma



Johann en Jacob Bernoulli,

EEN BUITENLANDER IN GRONINGEN
over integreren en barbaars gebazel

Waarheid en wetenschap staan onder druk en hebben dat in het verleden vaker gedaan. Soms was die druk tiranniek, maar vaak zat er ook een hilarisch kantje aan. Zo hebben wiskundigen en statistici bij de begrippen 'integreren' en 'differentiëren' veelal een volstrekt andere eerste connotatie dan niet-vakgenoten, zoals we verderop zullen zien. Niet-wiskundigen denken bij integreren veelal aan de integratie van nieuwe landgenoten hun samenleving.

We beleven barre tijden vol nepnieuws en met politieke leiders zonder beschavingsideaal. Wetenschap of het (on)gezonde verstand, waarheid of nepnieuws, Obama of Trump, Galileo of de paus, om een paar tegenpolen te noemen uit het verleden en het heden. En wie had twintig, of zelfs tien jaren geleden, gedacht dat onze fel Verlichte samenleving overschaduwd zou worden door populistische barbarij? Op het verzetsmonument bij de Weteringschans in Amsterdam staan de volgende drie regels van H.M. van Randwijk die betrekking hebben op de donkere jaren van de Tweede Wereldoorlog:

> een volk dat voor tirannen zwicht
> zal meer dan lijf en goed verliezen
> dan dooft het licht ...

De drie puntjes na 'licht' suggereren een naargeestig vervolg. Met de stijgende invloed van de Erdogans en de Trumps wordt waarheid vervangen door nep en daalt het gezag van de wetenschap, inclusief dat van de wiskunde en de statistiek.

Wetenschap, politiek en religie hebben helaas veel met elkaar te maken. Gelukkig zijn er ook 'wetenschap versus godsdienst'-ruzies met een hilarisch randje. Het volgende verhaal is daar een mooi voorbeeld van. Het laat zien dat dergelijke controversen ook een positieve bijwerking kunnen hebben.

Johann Bernoulli

In 1695 stapt de jonge 28-jarige wiskundige Johann Bernoulli met vrouw en baby in Bazel aan boord van een vrachtschip. Vandaaruit gaat het over de Rijn naar Nijmegen en daarna per koets naar Utrecht, om vervolgens via Amsterdam over de *Suydersee* naar *Wonnen* (Zwolle) te gaan. Vanuit Wonnen komt het gezin uiteindelijk via een aantal omwegen op 22 oktober 1695 aan in Groningen. Daar vindt op 28 november van datzelfde jaar de feestelijke inauguratie plaats van de nieuwe hoogleraar wiskunde in Groningen. Zijn inaugurale rede *In Laudem Matheseos* oogst veel bijval. Maar met de waardering groeit ook de tegenstand. Binnen een jaar wordt Bernoulli, de buitenlander, door predikanten en theologen uitgekreten voor een 'verderver der jeugd'.

Zo breken in 1695 voor het gezin Bernoulli tien turbulente jaren aan. De theologische twisten in Groningen gaan gepaard met geruzie met zijn geleerde broer Jacob Bernoulli, die de leerstoel Wiskunde in Bazel bezet. De ruzie tussen de beide broers wordt ongetwijfeld ook gevoed door Johanns afgunst ten aanzien van Jacob. In december 1695 verschijnt in *Acta Eruditorium* een artikel van Jacob met een heftige aanval op zijn broertje, waarin hij hem op sarcastische toon verwijt zich ontdekkingen toe te eigenen die hem, Jacob, toekomen. Johann probeert zijn gram te halen door zijn broer vervolgens allerlei lastige wiskundevraagstukken voor te leggen in de hoop dat hij zich erin zal verslikken. Jacob lost ze echter allemaal op en antwoordt zelf met een nieuw probleem, onder de toevoeging dat Johann bij een goede oplossing 50 Thaler ontvangt. Johann beweert in enkele uren – later maakt hij er drie minuten van – de oplossing gevonden te hebben. Deze wordt evenwel door Jacob voor foutief verklaard. Johann op zijn beurt houdt vol dat zijn oplossing correct is en beschuldigt zijn broer ervan 'hem te willen chicaneren, en de armen, voor wie hij de prijs heeft bestemd, te kort te doen'. Johann stelt voor om Leibniz als scheidsrechter te laten optreden. Jakob accepteert het voorstel mits ook Newton, de grote tegenpool van Leibniz, en De l'Hôpital scheidsrechter zijn. Van een arbitrage met deze drie heren is nooit iets gekomen en de 50 Thaler zijn Johanns armen misgelopen.

De gebroederlijke animositeit is overigens uitermate vruchtbaar geweest voor de ontwikkeling van de differentiaal- en integraalrekening, waarmee ook de hierboven genoemde problemen werden opgelost. Zo zijn de beide Bernoulli's belangrijke pleitbezorgers geworden van Leibniz's *Nova Methodus Pro Maximus et Minimus* die gaat over de beginselen van de differentiaal- en integraalrekening.

De theologische twisten in Groningen zijn mede een gevolg van Bernoulli's polemische karakter en draaien om een prikkelende uitlating van hem tijdens een college. Daar vraagt Johann de studenten naar de mogelijke theologische consequenties van 'de voortdurende stofwisseling van het menselijk lichaam'. De Groningse theoloog Paulus Hulsius ontketent daarop een ware ketterjacht tegen Bernoulli en beschuldigt hem ervan 'de opstanding des vleses' te loochenen. De ophef over de vermeende ketterijen is van dien mate 'dat de stad er vol van is'. Een student, met de naam Petrus Venhuysen, doet er nog een schep bovenop en beschuldigt Bernoulli ervan 'het Gereformeerde geloof te weerstreven en de troost voor de gelovigen uit het lijden van Christus weg te nemen'. Bernoulli antwoordt: 'Laat die ongehoorde uitlegger toch inzien dat hij hier zelf zeer vreemde gedachten heeft, die via de buik en het toilet verwijderd moeten worden'. Pittige taal, die de hedendaagse lachspieren niet onberoerd laat.

Johanns vrouw, Dorothea Falkner, is het Groningse getwist na tien jaar meer dan beu en heeft heimwee naar haar familie. In 1705 keren ze terug naar Bazel, teleurgesteld en dan nog niet wetend dat daar een fel begeerd hoogleraarschap wacht. Nog maar net vertrokken ontvangt Johann op 22 augustus in Amsterdam een brief met het overlijdensbericht van zijn broer Jacob.

Als het gezin in 1705 vertrekt, woeden ook buiten Groningen de theologische twisten onverminderd verder. Ongekend fel is de toon en toorn van calvinistische anti-cartesiaanse theologen tegen de *Nova Methodus*. 'Wiskundigen in hun arrogantie nemen plaats op de troon van de Schepper', aldus de anti-cartesianen: 'zij (de wiskundigen dus) menen iets-uit-niets te kunnen creëren...'.

'Iets' uit 'niets' maken. Hoe deden ze dat dan wel? De redenering ging ongeveer als volgt. In de ogen van de toenmalige wiskundigen is het principe van de integraalrekening het kunnen opdelen van een gegeven stuk materie in oneindig veel oneindig kleine delen. Leibniz noemde die oneindig kleine entiteiten *monaden*. De oppervlakte van het gebied tussen een deel van een gegeven kromme en de horizontale as in het platte vlak is, in het algemene geval, niet de som van een eindig aantal rechthoeken door die kromme bepaald, maar de S*umma* (vandaar het bekende S-symbool voor integraal), van een oneindig aantal 'rechthoeken' met oneindig kleine breedte. Daarmee is, zo zei men, dat oppervlak dus de S*umma* van oneindig veel oneindig kleine grootheden.

'Iets' dus gemaakt uit 'niets': de wiskundige op Gods troon. Voltaire schrijft bij Johann Bernoulli's overlijden het volgende gedichtje:

> zijn Geest zag Waarheid
> zijn Hart kende Gerechtigheid
> Zwitserland en de Gehele Mensheid was hij tot Eer.

We zouden daaraan kunnen toevoegen:

> waarheid en wetenschap
> overwinnen de barbaren
> en het licht dooft nooit.

Noot

Een van mijn wiskundige fietsvrienden wees me na lezing van een vorige versie van dit verhaal op een belangrijke omissie, namelijk het niet noemen van Johann Bernoulli's oplossing van het in 1696 door Galileo geformuleerde brachystochrone probleem, waarbij wordt gevraagd naar een wiskundige formule van de vorm van de glijbaan waarin een knikker het snelst van boven naar beneden rolt. 'De oplossing is het neergaande deel van de baan van het ventiel van je voortrollende voorwiel', vertelde mijn fietsvriend. Nooit eerder had ik op deze wijze naar dat ventiel gekeken. Ik pomp er altijd acht bar doorheen.

Literatuur

Sierksma, G. (1989), Johann Bernoulli (1667–1748); een Zwitsers Wiskundige bekneld tussen Stad en Ommelanden. In G.A. van Gemert, J. Schuller tot Peursum-Meijer, & A.J. Vanderjagt (eds.), *Om niet aan Onwetendheid en Barbarij te Bezwijken, Groningse Geleerden 1614–1989* (p. 65–82). Hilversum: Verloren.

Sierksma, G. (1992). Johan Bernoulli (1667–1748); His Ten Turbulent Years in Groningen, *Mathematical Intelligencer, 14(4)*, 22-31.

Maanen, J.A. van, (1995), *Een Complexe Grootheid; Leven en Werk van Johann Bernoulli 1667–1748*. Amsterdam: Epsilon Uitgaven.

Sierksma, G., & Sierksma, W. (1999). The Great Leap to the Infinitely Small. Johann Bernoulli: Mathematician and Philosopher, *Annals of Science, 56*(4), 433-449.

Gerard Sierksma is emeritus hoogleraar Kwantitatieve Logistiek en Sportstatistiek aan de Rijksuniversiteit Groningen.
E-mail: g.sierksma@rug.nl

Automatisch opsporen van fouten in data voor officiële statistiek

Sander Scholtus

Meetfouten komen voor in praktisch alle gegevens die gebruikt worden voor statistisch onderzoek. In de officiële statistiek wordt veel aandacht besteed aan het opsporen en verbeteren van meetfouten, om te voorkomen dat deze de kwaliteit van te publiceren statistieken aantasten. Traditioneel gebeurde dit handmatig. Er bestaan ook methoden om fouten automatisch op te sporen, wat in potentie veel tijd en kosten zou kunnen besparen. Echter, de kwaliteit van automatische foutlocalisatie valt tot nu toe vaak tegen, zodat in de praktijk veel handwerk nodig blijft. In dit artikel ga ik in op een bestaande methode voor automatische foutlocalisatie, en op manieren om deze methode te verbeteren.

Gegevens die zijn verzameld voor statistisch onderzoek bevatten in de praktijk altijd meetfouten. Bij vragenlijstonderzoek kunnen fouten bijvoorbeeld optreden wanneer een respondent een vraag verkeerd begrijpt, of het antwoord niet precies weet, of zelfs bewust een verkeerd antwoord geeft. Gegevens die afkomstig zijn uit een register kunnen ook meetfouten bevatten. In dit geval is een extra bron van fouten dat personen of bedrijven er belang bij kunnen hebben om (ten onrechte) op een bepaalde manier geregistreerd te staan.

Meetfouten werken door in statistieken die uit de data worden afgeleid. In het gunstigste geval zorgen meetfouten alleen voor extra onzekerheid (variantie), maar ze kunnen ook leiden tot vertekening. Het is daarom belangrijk om bij statistisch onderzoek alert te zijn op de invloed van meetfouten en deze eventueel te corrigeren.

In de officiële statistiek is van oudsher veel aandacht besteed aan gestandaardiseerde methoden om gegevens te controleren en waar nodig te corrigeren voor meetfouten. Traditioneel werden alle data 'met de hand' gecontroleerd door inhoudelijk deskundigen. Dit was tijdrovend en kostbaar; naar schatting besteedden statistische bureaus tot 40% van hun budget aan dit controlewerk.

Vanaf de jaren tachtig van de vorige eeuw groeide het besef dat deze aanpak verbeterd kon worden. Veel individuele meetfouten zijn nauwelijks zichtbaar in een geaggregeerde statistiek. Door het handmatige controlewerk toe te spitsen op verdachte waarden die naar verwachting de grootste invloed hebben op te publiceren statistieken, kan veel tijd en geld worden bespaard terwijl de nauwkeurigheid van de statistieken nauwelijks verandert. De kwaliteit van de statistieken kan zelfs worden verhoogd door een deel van het vrijgekomen budget te besteden aan andere verbeteringen, zoals het bestrijden van non-respons. De afgelopen decennia is een standaard-methodologie ontwikkeld voor het op een effectieve manier selecteren van invloedrijke verdachte waarden (De Waal et al., 2011).

Automatische foutlocalisatie

Tegelijk met de ontwikkeling van selectiemethoden om het handmatige controlewerk terug te dringen, is er ook onderzoek gedaan naar methoden om meetfouten automatisch te verbeteren. Het idee is dat men dergelijke methoden toepast op het deel van de data dat niet voor handmatige controle wordt geselecteerd. Als meetfouten redelijk betrouwbaar worden gecorrigeerd door een automatische methode, kan de selectie voor handmatige controle verder worden ingeperkt, zodat statistieken nog efficiënter en sneller worden geproduceerd. Een ander voordeel van automatische methoden is dat ze transparanter en beter reproduceerbaar zijn dan handmatige correctie.

Om fouten in data automatisch te herkennen kan men controleregels opstellen waaraan foutloze data zouden moeten voldoen. Deze regels bevatten inhoudelijke kennis die ook (maar soms impliciet) wordt gebruikt bij handmatige controles. Met name bij bedrijfsstatistieken kunnen vaak sterke controleregels worden opgesteld, op basis van boekhoudkundige definities. Stel bijvoorbeeld dat bij een bedrijf de drie variabelen *omzet*, *kosten* en *resultaat (winst/verlies)* worden uitgevraagd. In een bepaalde bedrijfstak moeten deze variabelen voldoen aan de volgende regels:

omzet – kosten = resultaat,
omzet ≥ 0,
kosten ≥ 0,
kosten ≥ 60% × omzet.

Binnengekomen data die niet voldoen aan alle controleregels bevatten blijkbaar minimaal één fout. De eerste rij van tabel 1 toont een voorbeeld van een ingevulde vragenlijst die de eerste controleregel schendt.

In het algemeen is niet onmiddellijk duidelijk *welke* waarden fout zijn, omdat controleregels meerdere variabelen kunnen bevatten. Om fouten automatisch aan te wijzen is naast de controleregels een selectiecriterium nodig. Fellegi en Holt (1976) gaven de eerste logisch sluitende aanpak voor automatische foutlocalisatie. Hun voorstel is uitgegroeid tot de standaardaanpak in de officiële statistiek. Volgens het principe van Fellegi en Holt moet de kleinst mogelijke deelverzameling van de variabelen worden aangewezen als fout waarbij het mogelijk is om (alleen) deze variabelen aan te passen zodat voldaan wordt aan alle controleregels.

In het voorbeeld uit tabel 1 kan aan alle controleregels

	OMZET	KOSTEN	RESULTAAT
oorspronkelijke data	200	125	755
data na handmatige correctie	200	125	75
data na automatische correctie (Fellegi-Holt)	200	125	75

Tabel 1. Eerste voorbeeld van foutlocalisatie in een dataset met drie variabelen

	OMZET	KOSTEN	RESULTAAT
oorspronkelijke data	100	50	50
data na handmatige correctie	100	50	50
data na automatische correctie (Fellegi-Holt; alle controleregels meegenomen)	100	60	40

Tabel 2. Tweede voorbeeld van foutlocalisatie in een dataset met drie variabelen

worden voldaan door alleen de waarde van *resultaat* aan te passen, maar niet door alleen de waarde van *omzet* of *kosten* aan te passen. Volgens het principe van Fellegi en Holt moet daarom de oorspronkelijke waarde van *resultaat* als fout worden aangewezen (derde rij). In dit voorbeeld is dit ook de oplossing die een inhoudelijk deskundige zou kiezen (tweede rij).

Uitgaande van het principe van Fellegi en Holt komt automatische foutlocalisatie neer op een minimalisatieprobleem onder restricties. Voor een grote klasse van controleregels die in de praktijk voorkomen kan dit worden geschreven als een gemengd geheeltallig lineair programmeringsprobleem (Engelse afkorting: MILP) – een optimalisatieprobleem waarvan de restricties en de doelfunctie lineair zijn en waarbij sommige beslisvariabelen geheeltallig zijn (De Jonge & Van der Loo, 2014). In de praktijk moeten foutlocalisatieproblemen worden opgelost die maximaal enkele honderden variabelen en enkele honderden controleregels bevatten; dit lukt meestal goed met standaard-softwarepakketten voor MILP-problemen. Daarnaast zijn ook specifieke algoritmen ontwikkeld om dit foutlocalisatieprobleem op te lossen (De Waal et al., 2011).

In het voorbeeld uit tabel 1 leidde het principe van Fellegi en Holt tot dezelfde oplossing die een inhoudelijk deskundige zou kiezen. In toepassingen blijkt echter vaak dat data die automatisch zijn gecorrigeerd volgens dit principe systematisch verschillen van data die handmatig zijn gecorrigeerd. Dit beperkt de toepasbaarheid van automatische controle en correctie in de praktijk. In mijn promotieonderzoek (Scholtus, 2018) heb ik gewerkt aan twee uitbreidingen van het principe van Fellegi en Holt die deze aanpak flexibeler maken, zodat de uitkomsten beter aansluiten bij die van inhoudelijk deskundigen.

	OMZET	KOSTEN	RESULTAAT
oorspronkelijke data	100	60.000	40.000
data na handmatige correctie	100	60	40
data na automatische correctie (Fellegi-Holt; alleen harde controleregels meegenomen)	100	60.000	–59.000

Tabel 3. Derde voorbeeld van foutlocalisatie in een dataset met drie variabelen

	OMZET	KOSTEN	RESULTAAT
oorspronkelijke data	90	130	40
data na handmatige correctie	130	90	40
data na automatische correctie (Fellegi-Holt)	170	130	40

Tabel 4. Vierde voorbeeld van foutlocalisatie in een dataset met drie variabelen

Zachte controleregels

In het bovenstaande voorbeeld komen de eerste drie controleregels direct voort uit de definities van de variabelen. Dit zijn 'harde' restricties: elke vragenlijst die een van deze regels schendt bevat gegarandeerd een fout. De vierde regel is 'zacht': een vragenlijst waarin de opgegeven kosten minder dan 60% van de opgegeven omzet bedragen is verdacht, maar bevat niet per se een fout. De grens van 60% is gekozen omdat verwacht wordt dat deze restrictie geldt voor de meeste bedrijven in deze branche, maar er zijn uitzonderingen.

Zachte regels worden veel gebruikt tijdens handmatige controles. Echter, het principe van Fellegi en Holt ziet alle regels als hard. Bij automatische foutlocalisatie moeten zachte regels dus ofwel worden opgevat als harde regels, of worden weggelaten. Beide opties zijn niet ideaal, zoals blijkt uit de voorbeelden in tabel 2 en 3. De vragenlijst uit tabel 2 schendt de zachte controleregel. Bij handmatige controle is geconcludeerd dat deze (kleine) schending voor dit bedrijf terecht is. Bij automatische foutlocalisatie worden de data wel aangepast. In dit voorbeeld zou het beter zijn geweest om de zachte regel weg te laten. Het weglaten van de zachte regel kan echter ook leiden tot ongewenste aanpassingen, zoals blijkt uit tabel 3.

Een betere oplossing is mogelijk door onderscheid te maken tussen harde en zachte controleregels. In de besliskunde is een bekende aanpak om zachte restricties mee te nemen in een minimalisatieprobleem, een extra term aan de doelfunctie toe te voegen die kosten verbindt aan schendingen van deze restricties. Deze aanpak is direct toepasbaar op de MILP-formulering van het principe van Fellegi en Holt (Scholtus, 2013). De oplossing van het foutlocalisatieprobleem hoeft dan niet altijd aan alle zachte controleregels te voldoen; een eventuele schending van een zachte regel wordt afgewogen tegen het aanwijzen van een extra fout om deze schending op te heffen.

Algemenere aanpassingen

Een tweede belangrijke bron van systematische verschillen tussen handmatige en automatische foutlocalisatie is dat het principe van Fellegi en Holt ervan uitgaat dat elke variabele onafhankelijk wel of geen fout bevat. Inhoudelijk deskundigen zien echter dat respondenten ook fouten maken waarbij meerdere variabelen tegelijk betrokken zijn. Ze passen dan correcties toe die vanuit het oogpunt van Fellegi en Holt suboptimaal zijn. Bekijk bijvoorbeeld tabel 4. De inhoudelijke deskundige concludeert hier dat de respondent de bedragen bij *omzet* en *kosten* heeft verwisseld; hij/zij ziet dit als één fout. De automatische methode ziet dit als twee aanpassingen en kiest daarom een andere oplossing waarbij maar één waarde wordt aangepast.

In Scholtus (2016) is een uitbreiding van het principe van Fellegi en Holt voorgesteld die ruimte biedt aan dit soort correcties. Definieer per toepassing een verzameling toegelaten aanpassingen. Inhoudelijk komen deze aanpassingen overeen met correcties voor fouttypen waarvan men verwacht dat ze regelmatig voorkomen in de data. Dit kunnen fouten in individuele waarden zijn, zoals in de oorspronkelijke formulering van Fellegi en Holt, maar ook ingewikkeldere correcties zoals omwisselingen en overhevelingen van bedragen. Technisch is elke aanpassing een affiene transformatie van de data, waarbij eventueel vrije parameters worden ingevoerd. (In het oorspronkelijke probleem van Fellegi en Holt is de nieuwe waarde voor een als fout aangewezen variabele zo'n vrije parameter.) Het criterium voor foutlocalisatie is nu om te zoeken naar de combinatie van het kleinste aantal toegelaten aanpassingen die ervoor zorgt dat vol-

daan wordt aan alle controleregels. Het oorspronkelijke principe van Fellegi en Holt is een speciaal geval van deze formulering, voor een specifieke verzameling toegelaten aanpassingen.

Recent is gebleken dat ook dit uitgebreide foutlocalisatieprobleem is te schrijven als MILP-probleem (Daalmans en Scholtus, 2018). Het kan daarom op dezelfde manier worden opgelost als het oorspronkelijke probleem van Fellegi en Holt. Resultaten op data met gesimuleerde fouten suggereren dat hiermee een verbetering in de kwaliteit van automatisch gecorrigeerde data kan worden bereikt, mits men in staat is om relevante toegelaten aanpassingen te vinden.

Slot

De twee besproken uitbreidingen maken de automatische foutlocalisatie flexibeler, maar ze introduceren ook extra keuzes die per toepassing moeten worden gemaakt, zoals de precieze kostenterm voor geschonden zachte regels en de keuze van de toegelaten aanpassingen. Vervolgonderzoek zal zich richten op het uitwerken en testen van deze methoden voor concrete toepassingen bij het Centraal Bureau voor de Statistiek.

Literatuur

Daalmans, J., & Scholtus, S. (2018). *A MIP Approach for a Generalised Data Editing Problem*. Discussion paper, Den Haag: Centraal Bureau voor de Statistiek.

Fellegi, I.P., & Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*, *71*, 17–35.

Jonge, E. de, & Loo, M. van der (2014). *Error Localization as a Mixed Integer Problem with the Editrules Package*. Discussion paper 2014-07. Den Haag: Centraal Bureau voor de Statistiek.

Scholtus, S. (2013). Automatic Editing with Hard and Soft Edits. *Survey Methodology*, *39*, 59–89.

Scholtus, S. (2016). A Generalized Fellegi-Holt Paradigm for Automatic Error Localization. *Survey Methodology*, *42*, 1–18.

Scholtus, S. (2018). *Editing and Estimation of Measurement Errors in Administrative and Survey Data*. Proefschrift, Amsterdam: Vrije Universiteit. http://dare.ubvu.vu.nl/handle/1871/55568.

Waal, T. de, Pannekoek, J., & Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. Hoboken, NJ: John Wiley & Sons.

Sander Scholtus is methodoloog bij het Centraal Bureau voor de Statistiek in Den Haag. In maart 2018 is hij cum laude gepromoveerd aan de Vrije Universiteit op een proefschrift dat onder andere gaat over het onderwerp van dit artikel.
E-mail: s.scholtus@cbs.nl.

ONBEKEND MAAKT ONBEMIND

Het is inmiddels alweer een jaar geleden dat aan onze verliefdheid op 'het algoritme' een eind kwam.[1] We hielden het niet langer vol. Het algoritme zorgde voor filterbubbels, beïnvloedde verkiezingen, propageerde vooroordelen en elimineerde bovenal onze privacy. Cathy O'Neil heeft ons de ogen geopend, het 'algoritme' is manipulatief en heeft niet het beste met ons voor.[2] Nee, het kon niet langer zo, het moest stoppen. Nu we een jaar verder zijn, komt het besef dat de breuk wellicht wat impulsief was. Zijn we te generalistisch geweest door alle algoritmes in de ban te doen? Maar hoe kunnen we elkaar dan weer vertrouwen?

In mijn werk als *analytics consultant* kom ik, jammer genoeg, steeds meer organisaties tegen die het vertrouwen in het gebruik van data en algoritmes aan het verliezen zijn. In de afgelopen jaren hebben ze, veelal op aangeven van strategieconsultants en technologieleveranciers, geïnvesteerd in opslagcapaciteit en analysesoftware en zijn ze als een dolle data gaan verzamelen en analyseren. Dit heeft geleid tot een overspannen vraag naar statistici en econometristen terwijl al deze inspanningen tot niet meer dan een open-deurenshow[3] van inzichten hebben geleid. Er zijn ook maar weinig echt impactvolle toepassingen van data science en kunstmatige intelligentie in de praktijk te vinden. Andrew Ng geeft aan dat bijna alle economische waarde die gegenereerd wordt met machine-leren op het conto komt van *supervised learning*[5] toegepast op online *display ad*.[4] Een kleine verbetering van de *click trough rates* kan in deze industrie veel geld opleveren.

Ondanks dat deze toepassing van machine-leren kennelijk veel oplevert, moet ik bekennen dat ik nog steeds niet erg onder de indruk ben van de relevantie van de advertenties die ik online te zien krijg. Ik blijf me verbazen dat er in de media zoveel aandacht geschonken wordt aan de beloftes van data science en kunstmatige intelligentie, terwijl tastbare resultaten schaars zijn. De voorbeelden zijn beperkt en hebben niet de potentie ook in de praktijk veel impact te hebben. Zeg nu zelf, zou je een algoritme[6] dat GO kan spelen en winnen de opdracht geven je auto te besturen of je agenda te beheren? Vragen naar bewijs voor de gouden bergen die worden beloofd[7] worden bij voorkeur genegeerd, het lijkt wel een religie.

Dat de inzet van statistiek en operations research juist veel impact heeft, komt minder vaak over het voetlicht. Onze overheid bespaart miljarden euro's doordat we met statistiek en operations research kunnen vaststellen welke dijkhoogte[8] optimaal is en de hoge benutting van ons spoornet kan alleen bereikt worden omdat de NS operations research gebruikt om treinen te *schedulen*. Begin juli was ik als jurylid van de EURO Excellence in Practice Award9 (EEPA) op de EURO-conferentie in Valencia. Ik vond het fantastisch om te zien hoeveel praktische toepassingen van operations research gepresenteerd werden en te horen welke impact die hebben gehad. Toepassingen als het opstellen van een eerlijk spelschema voor de kwalificatie van het WK-voetbal, het bepalen van de beste manier om UNESCO-erfgoed te conserveren, het optimaal aansluiten van offshore windmolens op het energienet, het dynamisch plannen van melkcollectieritten, het verbeteren van het kinderbeschermingsbeleid en het bepalen van een marktmechanisme voor de verkoop van visserijrechten. Stuk voor stuk praktische toepassingen met veel impact, zowel maatschappelijk als economisch. Ik heb er alleen niets over gelezen in de media.

Ondanks dat ik al lang in ons vakgebied werkzaam ben blijf ik het lastig vinden aan iemand van buiten ons vakgebied uit te leggen wat operations research nu precies is. Voor hen is er geen verschil tussen operations research, data science, analytics, machine learning of kunstmatige intelligentie. In elk van deze disciplines gebruik je vanuit hun optiek data en wiskunde om inzichten te genereren die besluitvorming ondersteunen. Als gevolg projecteren ze hun teleurstellende ervaringen op allemaal en dat is niet terecht. Om hun vertrouwen te herwinnen is het belangrijk ze te laten zien wat de impact van operations research werkelijk is. Dat kan in mijn optiek het beste aan de hand van voorbeelden zoals de inzendingen voor de EEPA, de Franz Edelman Award[10] of natuurlijk je eigen ervaringen. Het is belangrijk dat duidelijk wordt wat operations research te bieden heeft. Zeker nu de wereld steeds complexer wordt en veranderingen zich sneller aandienen is de behoefte aan inzicht en op feiten gebaseerde, gestructureerde besluitvorming groter dan ooit. Wij moeten de wereld laten weten dat operations research daarbij een instrumentele rol kan spelen. Het verleden heeft dat al vele malen bewezen en – je kunt me vertrouwen – in dit geval zijn resultaten uit het verleden een garantie voor de toekomst.

Noten

1. Https://www.wired.com/story/2017-was-the-year-we-fell-out-of-love-with-algorithms/
2. https://www.wired.com/2016/10/big-data-algorithms-manipulating-us/
3. Een van mijn klanten vatte de resultaten van het data science initiatief binnen zijn organisatie zo samen
4. https://www.mckinsey.com/featured-insights/artificial-intelligence/how-artificial-intelligence-and-data-add-value-to-businesses
5. https://en.wikipedia.org/wiki/Supervised_learning
6. https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/
7. https://www.gartner.com/newsroom/id/3872933
8. https://www.cpb.nl/persbericht/3213331/franz-edelman-award-voor-project-optimale-dijkhoogtes-nederland
9. https://www.euro-online.org/web/pages/209/excellence-in-practice-award-eepa
10. https://www.informs.org/Recognizing-Excellence/INFORMS-Prizes/Franz-Edelman-Award

John Poppelaars is Practice leader of the Advanced Analytics and Business Intelligence unit of BearingPoint.
E-mail: john.poppelaars@bearingpoint.com

HET TOEWIJZEN VAN VOORKEURSACTIVITEITEN

Stef Kleinluchtenbeld & Gerhard Post

Een groep deelnemers moet ingedeeld worden bij een aantal activiteiten, waarbij op elke activiteit het aantal toe te wijzen deelnemers gelimiteerd wordt door een minimum en een maximum aantal. In eerste instantie gaan we ervan uit dat elk groepslid aan precies één activiteit moet deelnemen. We kunnen bij deze toewijzing op verschillende manieren rekening houden met de voorkeuren van de deelnemers.

Voorkeuren

De eerste mogelijkheid is om elke deelnemer drie activiteiten te laten aanvinken; bij de toewijzing moet dan één van de aangevinkte activiteiten genomen worden. Deze rangschikking noteren we als '111'; we zeggen dat elke deelnemer drie 'eerste' voorkeuren opgeeft. Natuurlijk is het mogelijk dat de voorkeuren zo eenzijdig zijn dat er geen geldige toewijzing mogelijk is; we zullen steeds aannemen dat dit niet optreedt.

Iets subtieler is dat elke deelnemer zijn eerste, tweede en derde voorkeur opgeeft, een voorbeeld van *ordinal ranking* (Wikipedia, 2018), die we noteren als '123'. Nu zijn er al verschillende manieren om de toewijzing tot stand te brengen. Weer aannemend dat iedereen toe te wijzen is, kunnen we ervoor kiezen om eerst het aantal toegekende derde voorkeuren te minimaliseren, en daarbinnen de toewijzing van tweede voorkeuren. In termen van een lineair programma kunnen we bij n deelnemers een toewijzing van een derde voorkeur kosten n^2 geven, en een toewijzing van een tweede voorkeur kosten n.

Een andere methode is dat elke deelnemer 100 punten kan verdelen. Hierbij stellen we als eis dat aan tenminste drie activiteiten 1 of meer punten gegeven moet worden. Dit leidt tot allerlei interessante vragen voor de deelnemers en voor de planner. Hoe moeten we een score 98-1-1 interpreteren, en hoe zetten we die tegenover 50-25-25 of 49-49-2 of 30-30-20-20? Als we een lineair programma zouden maken en de scores op een lineaire manier zouden verwerken, dan zou de '98' in 98-1-1 eerder toegekend worden dan de '50' in 50-25-25, of de '49's in 49-49-2, of de '30's in 30-30-20-20. Zo kan een deelnemer met meerdere eerste voorkeuren gekoppeld worden aan zijn derde of vierde voorkeur, omdat andere deelnemers maar één 'hoge' eerste voorkeur opgeven. Dit maakt strategisch gedrag aantrekkelijk: door alle punten te concentreren, wordt de kans groter dat je een eerste voorkeur krijgt. Vanuit de planner gezien is een scorelijst met equivalente eerste voorkeuren veel plezieriger en is, van de genoemde scores, 30-30-20-20 de meest flexibele optie.

Standard competition ranking

Het idee van 100 punten verdelen klinkt mooi maar leidt tot de vraag of de verschillen tussen de scores betekenis hebben. Grote verschillen in score interpreteren als grote voorkeursverschillen kan, naast oprechte sterke voorkeur voor één activiteit, strategisch gedrag bevorderen. Omgekeerd zou je als planner het aangeven van meerdere eerste keuzes willen aanmoedigen, als de deelnemer eigenlijk geen of weinig voorkeur heeft. Het gevolg daarvan is dat (meestal) meer deelnemers hun eerste voorkeur kunnen krijgen. Deze kwesties kunnen worden aangepakt door de toegekende scores om te zetten in een rangorde volgens de *standard competition ranking* (Wikipedia, 2018). De score 98-1-1 wordt dan de ranking 122, net als de score 50-25-25. De score 49-49-2 wordt ranking 113 en de score 30-30-20-20 wordt ranking 1133. We kunnen nu net als bij de ordinal ranking hierboven, het probleem met *goal programming* oplossen, waarbij we weer de toewijzing aan hoge rangnummers minimaliseren.

In plaats van punten uitdelen, kunnen we de deelnemers ook rechtstreeks hun ranking laten opgeven, waarbij dus meerdere 1e, 2e en 3e keuzes mogelijk zijn. Deze keuzes zullen we wel naar de standard competition ranking omzetten. Merk op dat in dit systeem een deelnemer geen direct voordeel heeft om een 113 ranking in plaats van een 123 ranking op te geven. Het zijn juist deelnemers met een 123 ranking die kunnen profiteren; hun tweede voorkeur kan vermeden worden door de flexibiliteit van anderen. Om de deelnemers aan te moedigen toch meerdere eerste voorkeuren op te geven, zullen we bij het toewijzen aan een derde voorkeur de deelnemers die twee eerste voorkeuren opgaven vermijden, ten kosten van hen die dat niet deden.

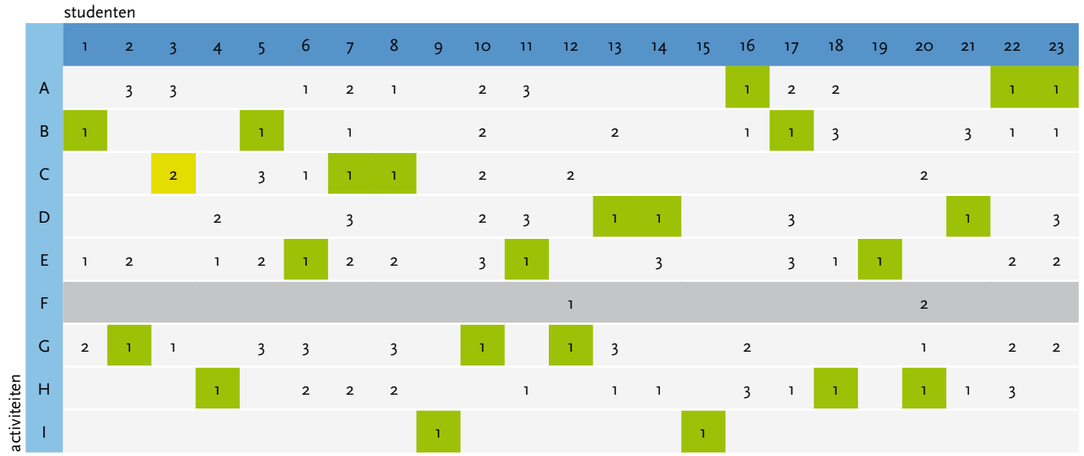
Voorbeeld 1. Toewijzing Bacheloropdrachten

In het voorjaar van 2014 waren er bij de opleiding Toegepaste Wiskunde van de Universiteit Twente 23 studenten die ingedeeld dienden te worden in groepen om gezamenlijk één van de negen bacheloropdrachten (met namen 'A' tot 'I') te doen. De ideale groepsgrootte was drie studenten, maar groepen van twee studenten was ook mogelijk. Bij voorbaat lag al vast dat de studenten 9 en 15 samen opdracht I zouden doen. Een ander speciaal geval was student 19, die opdracht E zelf ingebracht had; hij moest dan ook toegewezen worden aan deze opdracht. In eerdere jaren gaven de studenten een 1e, 2e en 3e voorkeur op, wat (uiteraard) leidde tot veel studenten die niet op hun 1e voorkeur ingepland konden worden.

Om deze reden kregen de studenten in 2014 de mogelijkheid om zoveel 1e, 2e en 3e voorkeuren op te geven als ze wilden, met de belofte dat meer sterke (bijvoorbeeld 1e) voorkeuren aangeven, zou leiden tot een betere kans op toewijzing aan een sterke voorkeur. Van de 20 studenten die hun keuze nog moesten opgeven, gaven maar liefst 16 studenten tenminste twee 1e voorkeuren op, zie tabel 1. Er konden hiermee zeven groepen van drie studenten geformeerd worden, met toewijzing van twintig 1e voorkeuren, en één 2e voorkeur. Opdracht F verviel. Dat een *matching* met alleen 1e voorkeuren onmogelijk is, valt te zien aan de keuzes voor de opdrachten F en G.

Voorbeeld 2. Activiteiten op een middelbare school

Het Staring College te Borculo organiseert één keer per jaar een activiteitendag waarop zo'n 300 leerlingen in drie tijdsblokken deelnemen aan één van 21 activiteiten. In 2017 konden de leerlingen vijf verschillende voorkeuren opgeven, een 1e tot en met 5e voorkeur. Aangezien elke leerling in elk tijdsblok een andere activiteit doet (afgezien van enkele activiteiten, die twee tijdsblokken



Tabel 1. De keuzes van de 23 studenten en de gemaakte groepsindeling (■1e voorkeur; ■2e voorkeur)

beslaan) worden idealiter de eerste drie (of in een enkel geval twee) voorkeuren toegewezen. Deze voorkeuren noemen we de *sterke* voorkeuren, de andere de *zwakke* voorkeuren.

De opgegeven sterke voorkeuren blijken niet erg evenwichtig verdeeld te zijn. Als we proberen activiteiten toe te wijzen met uitsluitend sterke voorkeuren, dan kunnen 166 activiteitsblokken van leerlingen niet toegewezen worden. Het is daarmee een goede case om het effect van verschillende score- en rankingsmethodes te onderzoeken (Kleinluchtenbeld, 2018). De zaak is eigenlijk nog wat gecompliceerder, doordat leerlingen ook drie medeleerlingen kunnen opgeven waarmee ze gedurende de dag graag samen worden ingepland. De kosten op het niet samen inplannen bij een activiteit zullen we gelijkstellen aan het verschil in kosten tussen de laagste sterke voorkeur en de hoogste zwakke voorkeur.

Bij het plannen van meerdere activiteiten per deelnemer lijkt goal programming minder geschikt. In plaats daarvan bekijken we twee doelfuncties:

- **Lineair**: de kosten voor het inplannen van voorkeur c levert in de doelfunctie kosten $c - 1$.
- **Stap**: als hierboven, maar met kosten 30 extra voor de zwakke voorkeuren.

Tabel 2 geeft een overzicht van de resultaten. We zien precies het te verwachten effect: in het lineaire model worden 286 van de 299 eerste voorkeuren gehonoreerd. Daartegenover staat dat dan 15 meer zwakke voorkeuren zijn toegekend. In het stap-model zijn slechts 171 zwakke voorkeuren toegekend, waarvan er 166 onvermijdbaar zijn.

Om te onderzoeken wat het effect van het toestaan van meerdere 1e, 2e, 3e, 4e of 5e voorkeuren is, hebben we voor een deel van leerlingen (oplopend van 0% t/m 100%) de voorkeuren random gegenereerd. Bij deze leerlingen is wel de eerste voorkeur blijven staan, maar op de tweede tot en met de vijfde plaats is uniform verdeeld een voorkeur tussen 1 en 5 getrokken. De zo ontstane getallen zijn weer met de standard competition ranking genormaliseerd. We kunnen de beide modellen ('Lineair' en 'Stap') weer toepassen, maar interessanter is wat er gebeurt met het aantal activiteiten dat niet toegewezen kan worden aan een sterke voorkeur.

Tabel 3 geeft een overzicht, gebaseerd op 10 random

	STERKE VOORKEUREN				ZWAKKE VOORKEUREN				NIET
	1	2	3	(som)	3	4	5	(som)	samen
LINEAIR	286	252	173	711	19	128	39	186	30
STAP	270	261	195	726	18	106	47	171	32

Tabel 2. Aantal toegekende voorkeuren bij ordinal ranking

0%	20%	40%	60%	80%	100%
166	155	143	132	119	106

Tabel 3. Aantal onvermijdbare zwakke voorkeuren als het aangegeven percentage random tot stand komt

trekkingen. We zien het aantal onvermijdbare zwakke voorkeuren in de toewijzing dalen als meer leerlingen meerdere sterke voorkeuren kunnen opgeven, zonder dat de capaciteit van de activiteiten te vergroot wordt. Als de leerlingen inderdaad niet altijd uitgesproken voorkeuren hebben, zoals voorbeeld 1 suggereert, kunnen we door het toelaten van meerdere 1e, 2e en 3e voorkeuren de tevredenheid verhogen.

Conclusie

Deze bijdrage wil de aandacht vestigen op een iets minder gebruikelijke methode om voorkeuren vast te leggen. We geloven dat deze methode in een aantal gevallen leidt tot toewijzingen die de deelnemers als beter waarderen. De methode met ordinal ranking dwingt de deelnemers te onderscheiden wat men als gelijkwaardig beschouwt. De puntenmethode moet in onze ogen op z'n minst gecombineerd worden met een rankingmethode, om strategisch gedrag te ontmoedigen. Maar ook bij de hier gepropageerde methode kan de deelnemer strategisch kiezen, namelijk door voor '3' een populaire activiteit te kiezen. Hopelijk is dat dan niet de eigen eerste voorkeur...

Onze dank gaat uit naar Marc Uetz en Johann Hurink voor gesprekken over dit onderwerp.

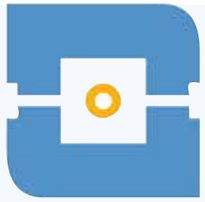
Literatuur

Wikipedia, *Ranking*, https://en.wikipedia.org/wiki/Ranking, geraadpleegd op 13 juli 2018.

Kleinluchtenbeld, S. (2018). *Het indelen van leerlingen voor een activiteitendag* (opdracht voor Onderzoek van Wiskunde, master Educatie en Communicatie in de Bètawetenschappen. Enschede: Universiteit Twente.

Stef Kleinluchtenbeld voltooide de bachelor in Applied Mathematics en deed bovenstaand onderzoek als onderdeel van de master Educatie en Communicatie in de Bètawetenschappen aan de Universiteit Twente.
E-mail: s.kleinluchtenbeld@alumnus.utwente.nl

Gerhard Post is OR expert bij ORTEC te Zoetermeer en universitair docent aan de Universiteit Twente. Hij publiceerde diverse wetenschappelijke artikelen over roosterproblemen.
E-mail: g.f.post@utwente.nl



After a warm summer with record-breaking temperatures, the Young Statisticians are back on track.

There is a young statistician in everyone

Last academic year we organized company visits to Marktplaats/eBay, SciSports, Alliander, the Dutch Court of Audit (Algemene Rekenkamer), to learn about their ways of applying statistics. We also organized symposia about sports & statistics and causality and contributed to the VVSOR Annual Meeting by hosting a lunch and a pub quiz. We enjoyed these events: bringing so much young statistical enthusiasts together gives us a lot of energy and motivation to continue this valuable board work. During fall, we will organize new company visits and a statistics café (symposium) and prepare for the events in 2019.

If you are interested in our events, please visit our website https://www.vvsor.nl/young-statisticians/ (where you can sign up for the newsletter) to sign up for our events. We hope to see many new statistics lovers, because the fun increases exponentially with the number of people joining clearly illustrated by the graph below. We believe that there is a young statistician in everyone!

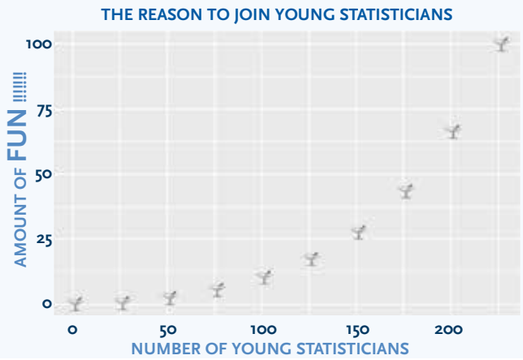


Foto: Airbus

FROM POLITICS TO MATHEMATICS

Exploring optimal air ambulance base locations in Norway

In Norway the nationwide physician-manned air ambulance service is considered essential in order to achieve the desired equality in healthcare. Yet, the location of the air ambulance bases has rarely been approached by the scientific community. A collaboration between mathematicians in the Netherlands and medical researchers in Norway is exploring the optimal location of air ambulance bases using mathematical modelling. The research has produced numerous novel insights, demonstrating that population coverage could be increased by moving existing bases, and that openly available population data might not be a reasonable proxy for actual incident data.

Jo Røislien

On May 12, 2014, Norwegian national broadcaster NRK published a news story that created a stir: "166,000 people do not receive help from a medical helicopter within 45 minutes." Placing a pair of compasses at each of the existing 12 air ambulance bases in Norway and drawing circles, combined with openly available population data from Statistics Norway, the journalists could simply count how many Norwegians could be reached by a medical helicopter within 45 minutes – and how many could not. The areas outside the circles were termed 'the black holes' on the map. More news stories followed. "Cardiac arrest in the black hole" and "Mayors demand better helicopter coverage". Interviews, radio, television.

Unfair?

The news story sparked a debate, but did the health service really perform poorly? A Norwegian Government white paper from 2001 states that "90 percent of the country's population should be reached by a physician manned ambulance within 45 minutes". As the news stories surfaced in 2014 the Norwegian population was about 5.1 million. The 166,000 living in 'the black holes' thus make for only 3.3% of the population. The medical helicopter service was not only performing due to government regulations: they were performing significantly better, covering almost 97% of the population within the target time.

While NRK was quick to point this out in editorial revisions, the debate was rooted in something bigger. Equality is a strong ideal in Norway. Equal opportunities. Equal pay. Equality for the law. Equality in health care. A famous photo in Norway shows King Olav taking the tram in 1973 during the petroleum shortage (Figure 1). So when there is indication that health services are unevenly distributed, that some people don't get something that others do, there's bound to be reactions.

Delft calling

Born in Norway professor Karen Aardal at Delft University of Technology (TU Delft) still follows the news from her home country regularly. And the news stories about the air ambulance caught her eye. Aardal is a professor of applied mathematics specializing in operations research and logistics, and she and her team have been working on various base location problems for emergency vehicles in The Netherlands, including ground ambulance allocation, and the location of fire trucks in Amsterdam. Aardal sent an e-mail to the Norwegian Air Ambulance Foundation (SNLA), a non-profit charity organization working to improve pre-hospital critical care. The e-mail initiated a collaboration between Aardal and a team of mathematicians at TU Delft and professor of medical statistics Jo Røislien and a team of medical researchers in Norway, aimed at exploring the optimal base locations for the Norwegian helicopter emergency service (HEMS).

This is Norway

Norway is a long-stretched country to the far North of Europe, covering 323,000 km². With a population

Figure 1. The Norwegian King Olav (right) riding the tram in 1973 during the petroleum crisis

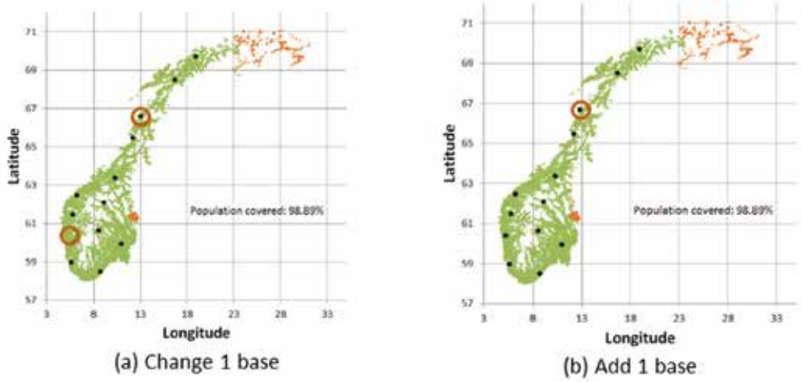


Figure 2. Optimizing air ambulance base locations conditioned on the existing base structure: Moving or adding one base (upper row), and moving or adding two bases (lower row)

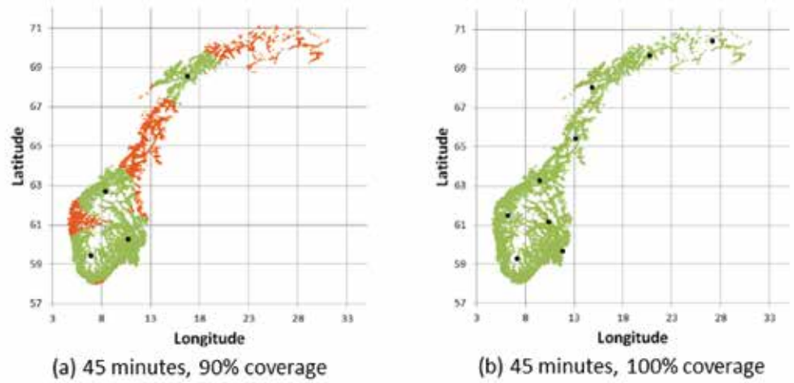


Figure 3. Optimal air ambulance locations in greenfield analyses for 90% and 100% population coverage: 45 minute target time (upper row), and 30 minute target time (lower row)

of about 5.1 million, the average is only 16 inh/km². For comparison the Netherlands has 411 inh/km², the mountainous Switzerland has 203 inh/km², and Europe's largest country Ukraine has 73.6 inh/km². The Norwegian population is also unevenly distributed, with large urban-rural differences. In the capitol Oslo the population density is 1132 inh/km², while in Northern Norway it is 1.5 inh/km²: less than in Mongolia. Yet, potential inequalities in health care services make headline news.

With large geographical distances and large uninhabited areas the nationwide physician manned HEMS is considered essential in order to achieve the desired equality in health care in Norway. Yet, the question of how to localize HEMS bases has rarely been approached by the scientific community. The 12 helicopter ambulance bases in Norway providing HEMS are established through historical local engagement from the late 1970s.

From news to maths

The journalists' approach gave a good estimate of the current population coverage. The main shortcoming of their approach was the lack of a mathematical model, and the added flexibility provided by that. A mathematical model is like a laboratory: It allows for experimentation. It allows for looking beyond the current situation, with its 'black holes', and whether – and how – it might be improved. Also, pre-hospital critical care time is an important factor, and it is of significant interest to explore the consequences of lowering the target threshold of 45 minutes. It was time to let mathematics get to work.

We collected official population statistics for Norway on a fine grid with cells of dimension 1km x 1km from Statistics Norway. For possible base locations we used a coarser 10km x 10km grid. Helicopter ground speed depends on wind direction and strength, and varies according different helicopter types and speeds used during missions. As an overall average number we used 220 km/h. The empirical average pre-flight preparation time of 5.5 minutes for air ambulances in Norway was also entered into the calculations.

Several mathematical models can be used to optimize the location of emergency vehicles in order to maximize coverage, one being the Maximal Covering Location Problem (MCLP). The MCLP maximizes the weighted number of demand locations covered within a desired service distance, or time, from a facility by allocating a fixed number of facilities. Conversely, the model can be used to determine the least number of bases needed in order to guarantee a certain pre-specified coverage.

Population coverage

Using a 45 minute threshold the existing 12 bases cover an estimated 97.84% of the population. By moving the base in Bergen, Norway's second largest city, further north (Figure 2a) or adding a new base at this northern location (Figure 2b) coverage increases to 98.89%. Moving or adding two bases further north (Figure 2c and 2d) increases coverage further to 99.88%. With minor adjustments to the existing base structure almost full population coverage is within reach.

Performing greenfield analyses demonstrates that with a 45 minute threshold 90% of the population can be covered with only four bases (Figure 3a), and the whole population with nine (Figure 3b). Reducing the target time to 30 minutes increases the number of bases needed markedly. It takes eight bases to cover 90% of the population (Figure 3c) and 21 bases to cover the whole population (Figure 3d). The greenfield numbers are shockingly low. Chances they might ever be implemented are slim to none. In Norway the air ambulance is a symbol of safety, and removing an existing base will not be taken lightly.

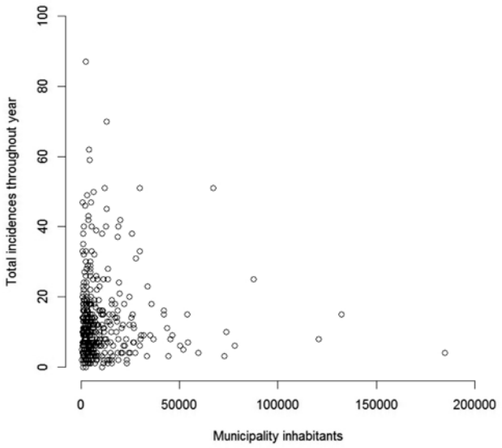


Figure 4. Municipality inhabitants and yearly total of incidences

It's all about the data

Government regulations are with regards to population coverage. But the main concern of the air ambulance is not where people live. After all, it is not a postal service. The main interest is the incidents for which HEMS is needed. Basing the analyses on population data is only meaningful if population density data is a reasonable proxy for actual incidents.

Norway is a country of large distances and changing weather. In winter Norwegians flock to the snow covered mountains, and in summer they spend time along the coast. The Norwegian air ambulance service already has seasonal bases, for example one in Hovden in southern Norway during Easter. "The location is essential with regards to the mountain tourism", says regional head Arild Sørensen.

The variation in the number of municipality incidents per 1000 inhabitants is substantial, with the higher numbers in Northern Norway, where the population density is the lowest. Indeed, the correlation between municipality population and yearly number of incidents is only -0.0027 (Figure 4). It's a common saying that 'people get injured where they live', but for incidents for which HEMS is needed this does not seem to be the case. We decided to redo the analysis, comparing results using both population density and incident data.

Where we live and where we are

Using population density data and a 45 minute threshold coverage can be increased by moving the Bergen base further north (Figure 5). Using incident data the Bergen base is still the least contributive, but in order to increase coverage the base should be moved significantly further north (Figure 5).

Reducing the target time to 30 minutes, the difference between the two data types remain. Using population data a base in northern Norway should be moved to the larger Oslo region to increase coverage (Figure 5). Using incident data it's more beneficial to move one of the bases along the southern coast further up the mountain – closer to the location of the Hovden seasonal base (Figure 5).

Where we live and where we spend our time is not the same thing.

It's all about the model

In our work on the location of HEMS bases in Norway (Røislien et al., 2017 and 2018) we have used the MCLP model. While regarded as a robust model for estimating optimal base locations for emergency vehicles, it has a major flaw: it assumes that there is always a vehicle available at a base whenever needed. This is rarely the case in practice. The model is thus suitable for exploring the optimal location of bases, but it cannot answer how to best distribute the available vehicles across these bases. For this a more advanced mathematical model is needed. Several such models exist that take the busy fraction into account, the fraction of the time emergency vehicles are busy with concurrent tasks, but there is no simple, well-established such model that can model Norway's uneven distribution of inhabitants and incidents. Newer theoretical results exist, but these are not readily available. We are currently working on testing and comparing various models.

When we published our first article, 153 of the thousands of articles in the journal contained the search term 'HEMS'. Ours was the first to contain the term 'MCLP'. Searching for the term 'EMS' (emergency medical service) in the extensive library of medical research literature PubMed returns about 15,000 search results. Searching for 'EMS AND busy fraction': Zero.

Mathematics is a cornerstone in modern medical research, and in 2000 the *New England Journal of Medicine* highlighted statistical analysis as one of the 11 most important medical developments of the last 1000 years. The

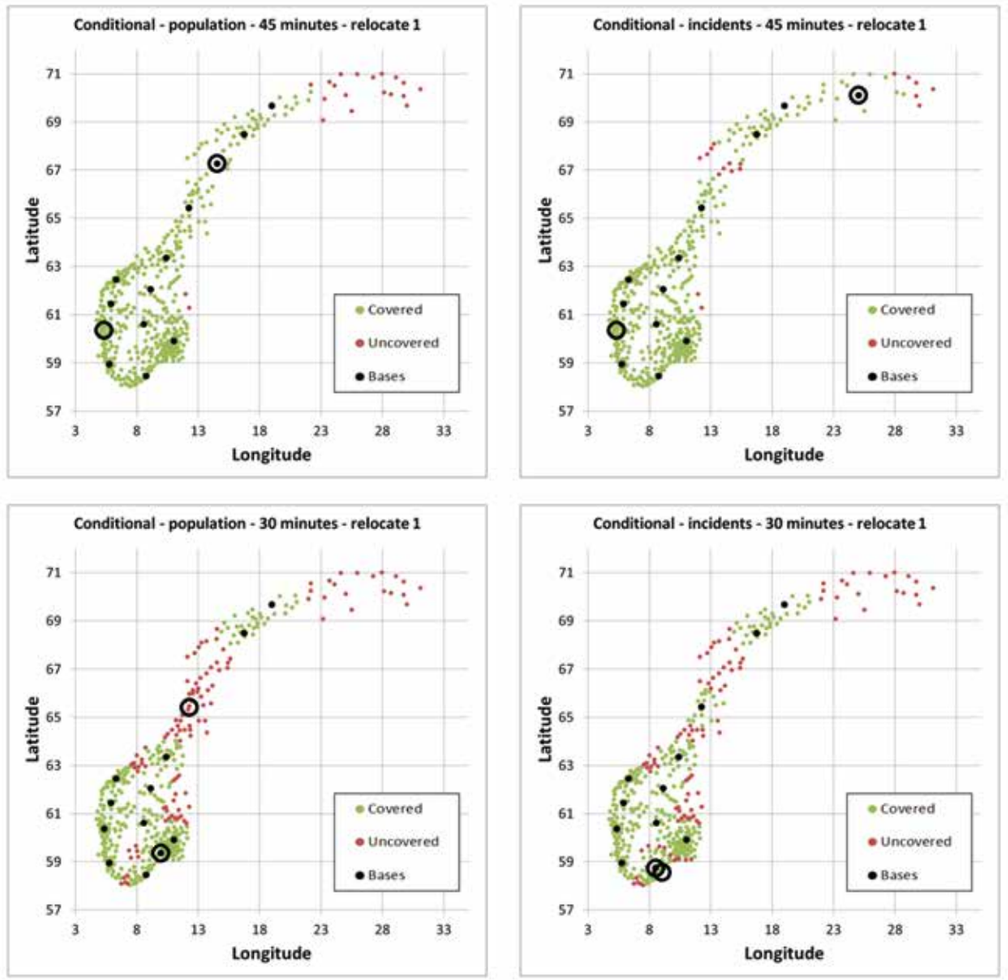


Figure 5. Optimal air ambulance locations in Norway comparing population and incident data for 45 minute threshold (upper row), and 30 minute threshold (lower row)

mathematical literature holds numerous results potentially of great value for medical researchers, and more advanced mathematical modelling is slowly finding its way into medical research. But if not applied to real world problems – problems that healthcare personnel actually care about – this mathematics will remain hidden in plain sight, read only by the few. This would be unfortunate. Healthcare is too important to be left to health personnel alone.

References

Røislien, J., Van den Berg, P.L., Lindner, T., Zakariassen, E., Aardal, K., & Van Essen, J.T. (2017). Exploring optimal air ambulance base locations in Norway using advanced mathematical modelling. *Injury Prevention,* 23(1), 10–15. doi: 10.1136/injuryprev-2016-041973.

Røislien, J., Van den Berg, P.L., Lindner, T., Zakariassen, E., Uleberg, O., Aardal, K., & Van Essen, J.T. (2018). Comparing population and incident data for optimal air ambulance base locations in Norway. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *26*(1), 1–11. doi: 10.1186/s13049-018-0511-4.

Jo Røislien is professor of medical statistics at the Faculty of Health Sciences, University of Stavanger, senior scientist with the Norwegian Air Ambulance Foundation, and adjunct associate professor at the Department of Mathematical Sciences, NTNU. He is a renowned science communicator and TV host in Norway, with numerous appearances on radio, TV and in the written press. www.joroislien.no. E-mail: jo@joroislien.no

ONZEKERHEID HOORT BIJ ONDERZOEK

Je komt tegenwoordig overal het begrip ‘kansen’ tegen: in het nieuws, in voorlichtingsfolders, noem maar op. Schattingen lopen vaak ook uiteen. Is de kans op een bepaalde ziekte nou 10 procent, 12 procent of 30 procent? Dat brengt lezers in de war, waardoor ze het resultaat wantrouwen. Door eerlijk te zijn over de (on)nauwkeurigheid van schattingen kan men het wantrouwen wegnemen.

Wantrouw onnauwkeurige resultaten daarom niet, maar gebruik ze bij je keuzes

Sanne JW Willems

Onderzoekers proberen de onnauwkeurigheid van hun resultaat te laten zien door de kans aan te geven met een interval. Bijvoorbeeld een kans van 10 tot 14 procent. Om dit uit te leggen, moeten we eerst iets weten over puntschatters en betrouwbaarheidsintervallen. Laten we hiervoor kijken naar een simpel voorbeeld.

Betrouwbaarheidsinterval

Stel dat we geïnteresseerd zijn in de gemiddelde lengte van volwassen Nederlandse vrouwen. Om hiervan een beeld te krijgen, kun je een steekproef nemen uit alle Nederlandse vrouwen en van ieder van hen de lengte meten. Hieruit blijkt dat de gemiddelde lengte van de vrouwen in de steekproef 1,65m is. Dan is je beste inschatting van het populatiegemiddelde, de gemiddelde lengte van alle volwassen Nederlandse vrouwen, precies dit gemiddelde van 1,65m. Deze schatting wordt een *puntschatter* genoemd. Een puntschatter is één getal, gebaseerd op één streekproef. Het nadeel van een puntschatter is dat het niks zegt over de nauwkeurigheid van die schatting. Ligt het populatiegemiddelde wel dicht bij het gemiddelde van de steekproef?

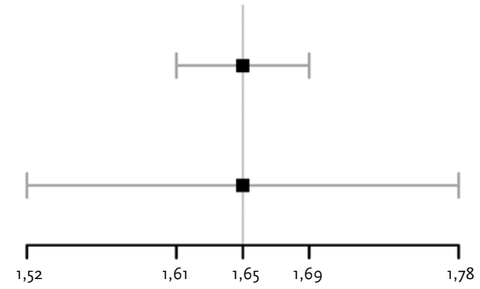
Precies om die reden geven onderzoekers vaak een *betrouwbaarheidsinterval*. Dit is een interval om de puntschatter heen en geeft aan hoe nauwkeurig de puntschatter is. Vaak is dat een 95 procent-betrouwbaarheidsinterval. Lezers van het onderzoek denken vaak dat dit betekent dat de onderzoeker met 95 procent zekerheid kan zeggen dat de schatting klopt. Ofwel, dat het 95 procent zeker is dat de gemiddelde lengte van Nederlandse vrouwen 1,65 meter is. Maar eigenlijk betekent het iets anders. Namelijk dat, als we heel veel steekproeven zouden nemen, in 95 procent van die steekproeven het populatiegemiddelde binnen het betrouwbaarheidsinterval ligt.

Volg je het nog? Mogelijk niet, en dat is ook niet gek. De precieze interpretatie is erg ingewikkeld en wordt helaas ook vaak door onderzoekers fout gedaan. Laten we daarom niet te veel focussen op de precieze interpretatie, maar meer op wat een betrouwbaarheidsinterval je kan vertellen over een puntschatter.

Nauwkeurigheid

Eigenlijk is vooral de breedte van het betrouwbaarheidsinterval belangrijk. Deze geeft namelijk de nauwkeurigheid van je puntschatter aan. Kijken we weer naar het voorbeeld over de lengte van vrouwen, waarbij de puntschatter 1,65m was, dan zou een betrouwbaarheidsinterval van 1,61m – 1,69m aangeven dat de puntschatter heel nauwkeurig is. Er is maar weinig speling. Bij een interval van 1,52m – 1,78m is de onnauwkeurigheid veel groter. De figuur hiernaast geeft je een idee van het verschil.

Onderzoekers hopen dus uit te komen op een smal betrouwbaarheidsinterval. Dat geeft immers aan dat je resultaat erg nauwkeurig is.



Figuur 1. Twee betrouwbaarheidsintervallen

De grootte van de steekproef en de variatie

Maar waar hangt de breedte van het interval dan van-af? Twee factoren hebben veel invloed: de grootte van je steekproef en de variatie binnen de populatie.

Hoe groter je steekproef, dus van hoe meer vrouwen je de lengte meet, des te smaller je betrouwbaarheidsinterval. Want hoe meer informatie je verzamelt, hoe meer je over de populatie in het algemeen weet. Denk maar eens in: de gemiddelde lengte over 10.000 vrouwen zegt natuurlijk veel meer over de lengte van vrouwen in het algemeen dan de gemiddelde lengte over maar 100 vrouwen. Hoe meer verzamelde data, hoe nauwkeuriger je puntschatter, en dus hoe smaller het betrouwbaarheidsinterval.

Verder hangt het betrouwbaarheidsinterval ook af van de variatie in de populatie. Als er bijvoorbeeld weinig verschil zit in de lengte van volwassen Nederlandse vrouwen, dan maakt het niet zo veel uit welke vrouwen je selecteert voor de steekproef. De gemiddelden van alle steekproeven liggen dan erg dicht bij elkaar liggen, omdat de lengten van de vrouwen in elke steekproef op elkaar lijken. Maar hoe groter de verschillen in die populatie, hoe onzekerder de schatting uit een steekproef.[1]

Waar komt dat wantrouwen vandaan?

Nu we wat meer weten over puntschatter en betrouwbaarheidsintervallen, kunnen we terug naar het wantrouwen dat veroorzaakt wordt door betrouwbaarheidsintervallen.

Als een onderzoeksresultaat wordt gegeven door middel van een betrouwbaarheidsinterval, dan lijkt het alsof een onderzoeker niet zo zeker is van zijn resultaat. Als uit onderzoek blijkt dat ‘de gemiddeld lengte van Nederlandse vrouwen 1,61m – 1,69m is’, dan lijkt dit op een vrij grove schatting. Terwijl de uitspraak ‘de gemiddeld lengte van Nederlandse vrouwen is 1,65m’ heel precies oogt. Je zou dus kunnen denken dat het eerste onderzoeksresultaat aangeeft dat er veel onzekerheid is in het onderzoek. Maar eigenlijk geeft de puntschatter in de tweede uitspraak eerder een vals gevoel van precisie. We kunnen zo’n uitspraak eigenlijk alleen doen als we alle Nederlandse vrouwen hebben opgemeten. Want alleen dan weten we echt het exacte gemiddelde.

Eerlijk zijn over (on)nauwkeurigheid

Ten slotte is het goed om te weten dat niet alle waardes in een betrouwbaarheidsinterval even waarschijnlijk zijn. De puntschatter ligt namelijk waarschijnlijk het dichtst bij het populatiegemiddelde. En hoe meer je naar de uiteindes van een betrouwbaarheidsinterval gaat, hoe kleiner de kans dat het echte gemiddelde die waarde heeft. Kijk nog maar een keer naar de grafiek hierboven. De puntschatter is dus het belangrijkste, maar het betrouwbaarheidsinterval is nodig om een idee te geven van de nauwkeurigheid.

Onzekerheid is helaas een onderdeel van onderzoek. We kunnen daarom maar beter eerlijk zijn over de (on)nauwkeurigheid. Onnauwkeurigheden in een onderzoek benoemen, zou daarom moeten worden gestimuleerd. Ze moeten in ieder geval zeker geen reden zijn om een onderzoeksresultaat weg te wuiven.

1. In een filmpje op YouTube worden beide factoren uitgelegd aan de hand van een voorbeeld over het gewicht van appels. https://www.youtube.com/watch?v=tFWsuO9f740

Sanne JW Willems doet promotieonderzoek aan de Universiteit van Leiden en richt zich op optimal-scalingmethoden voor *generalized linear models*. Ze is actief geweest in de Young Statisticians en is momenteel druk bezig met het oprichten van een VVSOR-sectie over Statistiek Communicatie. Daarnaast schrijft zij bijdragen voor het voor het wetenschapsblog Wetenschap.nu. Dit artikel is eerder op dat blog gepubliceerd.
E-mail: s.j.w.willems@math.leidenuniv.nl
www.SanneJWWillems.nl.

EEN SENIOR ONDER DE YOUNG STATISTICIANS

Een ode aan prof. dr. Richard D. Gill die benoemd is tot erelid van de VVSOR

Maarten Kampert


Foto: Marc de Haan

Dat Richard Gill erelid zou worden van de vereniging was voor veel van de actieve leden binnen de VVSOR vanzelfsprekend. De vraag was eerder *wanneer* het zou gebeuren. Al vrij snel nadat Jacqueline Meulman als eerste vrouwelijke voorzitter van de VVSOR het stokje van Richard overnam, kwam deze wanneer-vraag dan ook aan de orde Het duurde niet lang meer voordat Richard met emeritaat zou gaan; de daaropvolgende algemene ledenvergadering leek ons hét geschikte moment om hem extra in het zonnetje te zetten. Tijdens de Annual Meeting van 28 maart 2018 is het voorstel gedaan Richard tot erelid te benoemen; dit voorstel werd onder luid applaus aangenomen.

Het bestuur hoefde dit keer niet eens een toelichting te geven op het voorstel iemand tot erelid te benoemen. Het stond immers buiten kijf dat Richard Gill op nationaal en internationaal gebied bijzondere resultaten tot stand heeft gebracht in de ontwikkeling van de statistiek en OR. Niet alleen in de wetenschap zelf, maar ook in haar toepassing in de samenleving. In de afgelopen jaren lag voor Richard de nadruk op het laatste. Zo heeft hij als statisticus meegewerkt aan een onderzoek van de Verenigde Naties naar de 'aanslag' op de eerste minister in Libanon. Hij heeft buiten de wereld van de statistiek en wiskunde vooral bekendheid verworven door zijn tomeloze inzet als activist en pro-deo statisticus om de rechtszaak voor Lucia de Berk te heropenen. Dit bracht hem op de Nederlandse televisie, bijvoorbeeld als gast bij *Pauw & Witteman* en *De Wereld Draait Door*, maar ook in het tv-programma *Per Seconde Wijzer* als het antwoord behorende bij de omschrijving: 'Een beroemd statisticus in Nederland'.

In de bladen van de VVSOR en ook daarbuiten duikt zijn naam met enige regelmaat op. Een selectie van vijf bronnen – die verre van uitputtend noch representatief zijn – waaruit kan worden opgemaakt wat Richard wel niet allemaal betekend heeft voor de vereniging, de wetenschap en de samenleving is:

'The Van Dantzig Award' door Paul van der Laan in *Statistica Neerlandica*, 44(1990)4, p. 185-193.

'Thank you Richard Gill' door Hans van Houwelingen in *Statistica Neerlandica*, 50(1996)3, p. 337-338.

Interview met Richard Gill door Han Oud en Gerrit Stemerdink in *STAtOR*, *4*(2003)3, p. 4-9.

Interview met Richard Gill door Klaas Landsman en Hans Maassen in *NAW,* (2012)3, p. 5-13.

'Openheid is Noodzakelijk' door Richard Gill in *STAtOR, 13*(2012)3-4, p. 4-6.

Voor wie de hierboven genoemde interviews heeft gelezen moge het gelijk duidelijk zijn waarom Richard zo'n icoon is binnen de vereniging en de statistiek. Echter, het is niet alleen vanwege zijn prestaties. Vooral ook zijn open en prettig excentrieke persoonlijkheid heeft hier wel degelijk aan bijgedragen. Richard is een persoon die bij vele mensen een onvergetelijke indruk achterlaat.

Ook zonder iets van hem af te weten is hij snel herkenbaar. Een goed voorbeeld is misschien de Annual Meeting van 2012. Tijdens het laatste praatje van prof. dr. Sam Savage, waar iedereen hutje-mutje op elkaar gepropt zat in een volle en steeds warmer wordende zaal, kon men Richard – tot verbazing van de mensen die hem nog niet kenden – op de grond terugvinden. Hij volgde het praatje 'liggend' in het zijpad, waarschijnlijk een stuk prettiger en koeler. Het zijn dit soort anekdotes die ervoor zorgen dat Richard als authentiek persoon in het hart van velen gemakkelijk een plek verovert.

Een andere anekdote komt van Jacobus (Kobus) Oosterhoff, zijn promotor aan de Vrije Universiteit Amsterdam. Kobus vertelde dat Richard hem veel over *Survival Analysis* – het onderwerp van Richards proefschrift – heeft bijgeleerd, maar dat hijzelf Richard Gill heeft moeten aanleren correct Engels te schrijven, en dat terwijl Richard een geboren en getogen Engelsman is, afgestudeerd aan de Cambridge University in Wiskunde en Statistiek.

Wat ook wel degelijk heeft meegespeeld bij de voordracht voor het erelidmaatschap, is Richards grote betrokkenheid bij de studenten van de masterstudie Statistical Science in Leiden en zijn bijdragen aan de beginjaren van de Young Statisticians.

Al dan niet ietwat chaotisch is het voor de studenten altijd een genot geweest om Richards lessen (en levenslessen) te mogen bijwonen. Voor het eerste cohort van de in 2009 net opgerichte master Statistical Science betrof de eerste kennismaking met Richard een college over Maximum Likelihood Estimation, daar waar volgens het programma het toch echt de beurt was aan het onderwerp Sufficient Statistics. Om de slides te kunnen projecteren moest Richard eerst zijn beamer aan de gang zien te krijgen. Wij, de studenten, mochten dat alles meemaken en ook zien wat er zoal op het bureaublad van zijn computer stond. Dat bureaublad stond vol met allerlei bestanden met daarbij een folder 'Desktop'. In deze folder 'Desktop' zaten ook weer allerlei bestanden met een folder 'Desktop'. Het Droste-effect was al snel zichtbaar. Een fenomeen waar nu onder oud-studenten gekscherend naar gerefereerd wordt als 'Gill's Desktop Recursion Theorem'. De legende gaat dat om tot de originele 'Desktop'-folder van Richard Gill te kunnen komen er tot in de oneindigheid dient te worden doorgeklikt. Tijdens het zoeken naar de slides voor het college kon Richard het natuurlijk niet laten om trots de foto's van zijn kleinkinderen te laten zien. Iets dat in zo'n klein groepje van studenten positief en 'ontwapenend' verwelkomd werd.

Richard was aan ons geïntroduceerd als een enorme autoriteit in de statistiek, de Van Dantzig-prijswinnaar 1990, de Lorentz NIAS-fellow 2010, de voorzitter van de Nederlandse Vereniging voor de Statistiek (2007 – 2011), en als iemand die een voor ons onbegrijpelijk probleem in de survival analyse kon oplossen, en zich bezighield met statistiek in de quantummechanica. Diezelfde man was dus eigenlijk hartstikke toegankelijk en bereid om ook zijn tijd in ons, de studenten, te investeren. Hij lunchte samen met zijn studenten en zat vol verhalen. Wanneer wij erom vroegen, wilde hij ons ook graag van advies voorzien in de statistiek als een informeel studieadviseur en *early adopter* binnen het Statistical Science-programma van 2009–2013. Dat ging tot in de meeste concrete details, bijvoorbeeld: 'Wanneer je dan toch de programmeertaal R leert, raak dan ook bekend met R Markdown.'

Ook binnen de Young Statisticians was Richards enthousiasme voor 'R' makkelijk op te merken. Veel Young Statisticians leerden hoe je ook zelf R kon installeren op je iPhone of Android. Als (ex)-voorzitter van de VVSOR was Richard nauw betrokken bij de beginjaren van de Young Statisticians in meerdere rollen: als een van de medesponsors (uit zijn persoonlijke budget!), als medeorganisator van een activiteit, maar vooral ook als bezoeker en mede-Young Statistician. Met zijn 'jonge' en open geest inspireert Richard nog steeds jongeren in de statistiek in Nederland. Hij is eindelijk erelid van onze vereniging!

Maarten Kampert is docent en promovendus (onderwerp Clustering objects on subsets of attributes) binnen de groep Statistical Science aan het Mathematisch Instituut van de Universiteit van Leiden. Hij is één van de mede-oprichters van de Young Statisticians in 2010. Van 2011 tot 1 augustus 2018 was hij lid van het dagelijks bestuur van de VVSOR.
E-mail: mkampert@math.leidenuniv.nl

E.T. JAYNES

Harm Jan Boonstra & Leon Willenborg

Edwin Thompson Jaynes (1922–1998) was een natuurkundige die een groot deel van zijn leven heeft besteed aan het ophelderen van de rol van kansen binnen de natuurkunde, en aan de fundamenten van de kansrekening/statistiek zelf. Hij begon zijn promotieonderzoek bij J. Robert Oppenheimer ('de vader van de atoombom'), maar trok zich terug omdat zijn ideeën over de (interpretatie) van de kwantummechanica te veel verschilden van die van Oppenheimer. Jaynes promoveerde uiteindelijk bij Eugene Wigner. Net als Albert Einstein en Erwin Schrödinger heeft Jaynes de kwantummechanica nooit als een complete theorie geaccepteerd. Vooral de rol van en betekenis van kansen in de Kopenhaagse interpretatie van de kwantummechanica – die voornamelijk door Niels Bohr en Werner Heisenberg is ontwikkeld – was hem een doorn in het oog.

Volgens Jaynes zijn kansen geen fundamentele eigenschappen van fysieke objecten of systemen, maar representeren ze 'slechts' kennis hierover bij een persoon ('de waarnemer'). In filosofische termen gesteld: kansen bestaan op het epistemologische en niet op het ontologische niveau. Dit idee is een rode draad in zijn werk. Het toeschrijven van kansen aan de natuur, of dat nu aan elementaire deeltjes is of aan een dobbelsteen, noemt Jaynes een *Mind Projection Fallacy*.

MaxEnt

Jaynes' pogingen de kwantummechanica te demystificeren lijken grotendeels gefaald te hebben. Desondanks heeft hij belangrijke bijdragen aan onder andere de kwantumoptica geleverd. In de statistische mechanica – de natuurkunde van veel-deeltjessystemen – heeft Jaynes naam gemaakt met zijn maximum entropie raamwerk (afgekort MaxEnt), voortbouwend op het werk van onder andere J. Willard Gibbs. Hij zag statistische mechanica als een vorm van statistische inferentie en legde met MaxEnt de verbinding met de informatietheorie van Claude Shannon. MaxEnt is een manier om een kansverdeling af te leiden door de entropie (bij discrete verdeling gelijk aan $-\sum_i p_i \ln p_i$ en bij continue verdeling gelijk aan $-\int f(x)\ln f(x)dx$) te maximaliseren onder randvoorwaarden, vaak in de vorm van verwachtingswaarden.

In de statistische mechanica is het onmogelijk om de posities en snelheden van alle deeltjes in, zeg, een gas te kennen. Vaak zijn er wel metingen van macroscopische variabelen van het systeem zoals de temperatuur of de druk, die geïnterpreteerd kunnen worden als verwachtingswaarden van de kansverdeling over de microscopische vrijheidsgraden van het systeem. De kansverdeling met maximale entropie gegeven deze verwachtingswaarden geeft dan de beste beschrijving van het systeem, in die zin dat deze verdeling geen andere informatie veronderstelt dan de gegeven verwachtingswaarden. De zo verkregen kansverdeling kan dan weer gebruikt worden om andere eigenschappen van het systeem te voorspellen. Zo leidt bijvoorbeeld deze aanpak bij een systeem met discrete energiewaarden $E_i > 0$ tot de Boltzmann-verdeling, waarbij de kans om het systeem in toestand E_i aan te treffen evenredig is aan $e^{-\beta E_i}$ voor een zekere $\beta > 0$.

Probability Theory; The Logic of Science

In statistiekkringen is Jaynes vooral bekend vanwege zijn postuum uitgegeven boek *Probability Theory; The Logic of Science*. Dit is een boek over de fundamenten van de kansrekening. Kansrekening is volgens Jaynes een uitbreiding van de logica naar de situatie van onzekerheid. Deze opvatting deelt hij met de econoom Maynard Keynes, de logicus en filosoof Frank Ramsey en de filosoof Rudolf Carnap. Kritiek op de aanpak van Keynes is geleverd door onder andere de Franse wiskundige Émile Borel.

Het boek van Jaynes begint met de afleiding van de regels van de kansrekening volgens Richard T. Cox. Jaynes zegt hierover: '*The basic rules, from which all else follows, are nothing but the standard product and sum rules of probability theory; but now they are derived uniquely as principles of logic, from very elementary qualitative requirements of consistency and rationality.*' Kansen worden toegekend op grond van (onvolledige) informatie of kennis. Een uitdaging is daarbij om de informatie op een objectieve manier te vertalen naar een kansverdeling. MaxEnt is één manier, als er informatie in de vorm van verwachtingswaarden is.

Jaynes heeft, in navolging van de Britse wiskundige, statisticus, geofysicus en astronoom Harold Jeffreys, die hij zeer bewonderde, ook bijgedragen aan de theorie van transformatiegroepen voor het toekennen van prior verdelingen als er sprake is van bepaalde symmetrieën. Als er data beschikbaar zijn en een kansmodel voor de data gegeven de onbekenden, dan vindt binnen de *Probability Theory as Extended Logic*-aanpak inferentie plaats via de regel van Bayes, zoals de productregel in deze context wordt genoemd. De inferentie hangt ook af van een prior verdeling, die bijvoorbeeld met MaxEnt of de theorie van transformatiegroepen is toegekend. De uitkomst van de regel van Bayes is de posterior verdeling die de informatie in de data en de prior informatie combineert.

Optimale statistische procedures zijn volgens Jaynes altijd gebaseerd op de regels van de kansrekening. Zo laat hij in zijn boek zien hoe sommige statistische procedures uit de kansrekening volgen, en laat hij de beperkingen zien van statistische praktijken die de regels schenden. Jaynes besteedt in zijn boek ook aandacht aan modelselectie, en merkt hierover op: '*Instead of fearing wrong predictions, we look eagerly for them; it is only when predictions based on our present knowledge fail that probability theory leads us to fundamental new knowledge.*' Dit geldt misschien wel vooral, maar zeker niet uitsluitend voor de fundamentele natuurwetenschappen, en geeft in ieder geval het belang aan van evaluatie van de onzekere elementen en onvolledigheid van de veronderstelde prior informatie en modelkeuze.

Jaynes staat net als Jeffreys bekend als aanhanger van de objectieve Bayesiaanse statistiek. Weliswaar zijn kansen volgens Jaynes onvermijdelijk subjectief in de zin dat ze iemands informatie over een systeem of situatie representeren, maar objectief in de zin dat twee personen met dezelfde informatie tot dezelfde toekenning van kansen zouden moeten komen (indien ze het eens zijn over de hierbij te volgen procedure). Het nemen van beslissingen op basis van een statistische inferentie, waaronder bijvoorbeeld het samenvatten van een posterior verdeling in een puntschatting, bevat dan weer een duidelijk subjectief element, namelijk de keuze van een verliesfunctie.

Jaynes' boek en artikelen zijn zeer helder geschreven en bieden vaak verrassende inzichten. Door de vele historische uitweidingen geven ze ook een interessant beeld van de geschiedenis van de statistiek.

De geïnteresseerde lezer verwijzen we graag naar de site http://bayes.wustl.edu, waar een grote verzameling van Jaynes' geschriften te vinden is, waaronder ongepubliceerde en gepubliceerde artikelen op uiteenlopende gebieden binnen de natuurkunde en de statistiek, een interessant boek over muziek, en een vroege versie van zijn 'Probability Theory' boek. Ook Wikipedia bevat informatie over Jaynes: http://en.wikipedia.org/wiki/Edwin_Thompson_Jaynes.

In de literatuurlijst staan enkele geschriften vermeld waarin de MaxEnt aanpak wordt toegepast, van de hand van econometristen Henri Theil en Denzil Fiebig. Een bekend artikel waarin de informatietheorie als uitgangspunt is genomen voor een statistische aanpak is van de cryptografen en statistici Solomon Kullback en Richard Leibler.

Literatuur

Borel, E. (1964). Apropos of a Treatise on Probability. In H.E. Kyburg, & H.E. Smokler (eds.), *Studies in Subjective Probability*. Hoboken: Wiley.

Cox, R.T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics*, *14*, 1–13.

Cox, R.T. (1961). *The Algebra of Probable Inference*. Baltimore, MD: Johns Hopkins University Press.

Cox, R.T. (1979). Of inference and inquiry—An essay in inductive logic. In R.D. Levine, & M. Tribus (eds.), *The Maximum Entropy Formalism*. Cambridge, MA: MIT Press.

Jaynes, E.T. (1957). Information theory and statistical mechanics, *The Physical Review*, 106, 620–630.

Jaynes, E.T. (1986). Bayesian methods: general background. In J. H. Justice (ed.), *Maximum-Entropy and Bayesian Methods in Applied Statistics*. New York: Cambridge University Press.

Jaynes, E.T. (1989). Clearing up mysteries; The original goal. In J. Skilling (ed.), *Maximum-Entropy and Bayesian Methods*. Dordrecht: Kluwer.

Jaynes, E.T. (1993). A backward look to the future. In W.T. Grandy, Jr., & P.W. Milonni (eds.), *Physics and Probability: Essays in Honor of Edwin T. Jaynes*. Cambridge: Cambridge University Press.

Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*. New York: Cambridge University Press.

Jeffreys, H. (1931). *Scientific Inference*. New York: Cambridge University Press.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). New York: Oxford University Press.

Keynes, J.M. (1921). *Treatise on Probability*. London: Macmillan.

Kullback, S., & Leibler, R.A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, 22, 79–86.

Ramsey, F.P. (1964). Truth and Probability. In H.E. Kyburg, & H.E. Smokler (eds.), *Studies in Subjective Probability*. Hoboken: Wiley.

Theil, H. (1972). *Statistical Decomposition Analysis*. Amsterdam: North-Holland Publishing.

Theil, H. & Fiebig, D.Z. (1984). *Exploiting Continuity; Maximum Entropy Estimation of Continuous Distributions*. Pensacola, FL: Ballinger.

Harm Jan Boonstra en Leon Willenborg werken bij de afdeling Methodologie van het CBS, in Heerlen en Den Haag. E-mails: hjh.boonstra@cbs.nl en lcrj.willenborg@cbs.nl.

Are you up for the VeRoLog Solver Challenge?!

VeRoLog organizes the VeRoLog Solver Challenge (VSC) aimed at promoting the development of effective algorithms for real-world applications in Vehicle Routing. The problem subject of the challenge is defined by the VeRoLog board in collaboration with a company and is accompanied by representative instances of the problem. The challenge is open to everybody, regardless of affiliation. Registration is free.

An exciting routing problem

The present challenge revolves around an exciting routing problem which received little attention in the literature so far. This problem combines distribution and subsequent installation of equipment, such as vending machines. The organization provides a comprehensive description of the problem, instances and some resources such as a solution validator on the challenge website https://verolog2019.ortec.com/.

The challenge

The challenge consists of two parts that run in parallel. Participants may compete – and win money prizes – in one part or in both parts of the challenge.

During the first part of the challenge, solutions submitted for a set of public instances are ranked weekly. For each instance, the team that leads on that instance at the end of the week earns money. Additional prizes are awarded per instance to the team with the all-time best solution for that instance. Solutions submitted during the first part can be obtained by any means available.

The second part of the challenge is restricted to the teams that submit implementations either in the form of a Windows executable or a Python script (see https://verolog2019.ortec.com/ for specifications) and that are selected by the organization based on their solution quality. In this part, the organization applies the selected implementations on a set of not previously disclosed instances. IBM, Gurobi and FICO supply licenses of their solvers (CPLEX, Gurobi and Xpress, respectively) which ORTEC may use during the second part. This means that the algorithms submitted may call on these solvers.

Selected participants will have the possibility to present during the upcoming VeRoLog conference in Seville, 3-5 June 2019, and to publish in a special issue of the journal *Networks*. The winning team receives € 2019, the second best € 500 and the third € 250. The prizes are sponsored by ORTEC and will be awarded during the aforementioned VeRoLog conference.

The first edition of the VSC was held in 2014. PTV organized the first and second editions of the VSC, while the third was organized by ORTEC. The present (fourth) edition of the challenge, once again organized by ORTEC, was officially launched in September 2018. **Visit https://verolog2019.ortec.com/ for all the details, registration and participation**.

The organizers hope that many VSSOR members will participate in this exciting challenge. May the best team win!