beeld verkrijgen van de leefomstandigheden van de patiënten.

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

Conclusie

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

Literatuur

Dutch Tech Center For Life Sciences (2017). *Manifesto of the Personal Health Train consortium.* <https://www.dtls.nl/wp-content/uploads/2017/12/PHT_Manifesto.pdf>.

Schram, M. T., Sep, S. J. S., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., et al. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European Journal of Epidemiology, 29*, 439–451.

Sun, C., Ippel, L., Wouters, B., van Soest, J., Malic, A., Adekunle, O., et al. (2018). *Analyzing Partitioned FAIR Health Data Responsibly.* <http://arxiv.org/abs/1812.00991>.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018.

Lianne Ippel is postdoc bij het Institute of Data Science aan de Universiteit Maastricht.
E-mail: lianne.ippel@maastrichtuniversity.nl

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

Henk Tijms

Sinterklaas en de laatste lootjes

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

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

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

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

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

waarbij

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

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

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

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

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