

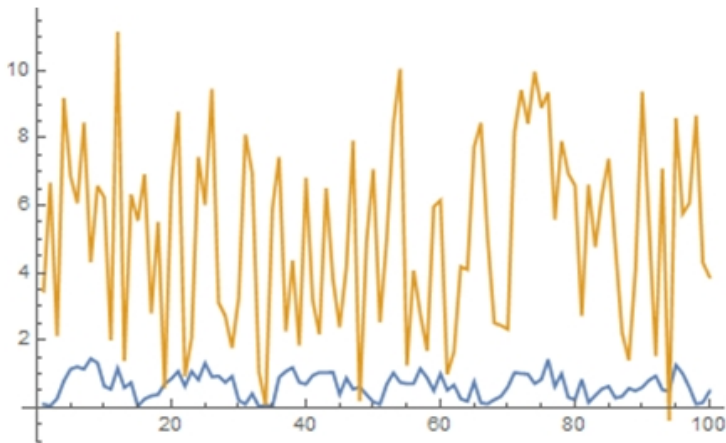
Hoofdstuk 1

Ruis verklapt richting

“De vraag naar causaliteit is een hele lastige. Je komt in bijna filosofische discussies terecht”, aldus wiskundige Sándor Kolumbán. Toch was dat de vraag die het Centraal Bureau voor de Statistiek aan de wiskundigen van de Studiegroep Wiskunde met de Industrie stelde - wanneer kan je concluderen dat overheidsingrijpen daadwerkelijk effect heeft gehad.

Het Centraal Bureau voor de Statistiek (CBS) krijgt vanuit ministeries geregeld de vraag om uit te zoeken of een wijziging in het beleid merkbaar effect heeft gehad. “Vaak zullen wij dan wat voorzichtig antwoorden met ‘we zien dit in de data’ en ‘dat zou kunnen komen door dit, maar misschien ook door dat’. Bij het CBS praten we veel over correlaties, maar wagen we ons niet zo snel aan harde uitspraken over causaliteit”, vertelt CBS-methodoloog Frank Pijpers. “Maar we willen onze klanten natuurlijk toch graag verder helpen.” Om beter zicht te krijgen wanneer het wel - of juist niet - mogelijk is om over causaliteit te spreken, wendde hij zich tot de Studiegroep Wiskunde met de Industrie. “Ook een negatieve uitspraak, dat het in bepaalde gevallen helemaal niet mogelijk is om causaliteit aan te tonen, zou voor mij al waarde hebben. Dan kan ik namelijk beter onderbouwen waarom het CBS de vraag niet kan beantwoorden.”

“Je kan betogen dat het uitgesloten is om met zekerheid te kunnen concluderen dat overheidsingrijpen effect heeft”, zegt wiskundige Sándor Kolumbán in zijn kantoor aan de Technische Universiteit Eindhoven. “De wereld is altijd in beweging. Er verandert dus altijd van alles, terwijl jij je maatregelen neemt. Je weet achteraf nooit zeker of een effect het gevolg is van die veranderende wereld, of van jouw ingrijpen.” Zo filosofisch pakt de wiskundigen de vraag van Pijpers echter niet op. Pijpers had zijn vraag vanuit wiskundig oogpunt namelijk glashelder geformuleerd: hij wilde graag weten of op basis van twee datareeksen te concluderen valt of één van de twee de drijvende kracht is van de ander. Met nog een addertje onder het gras - de dataserie bestonden maar uit een paar honderd datapunten, wat het probleem extra lastig maakt.



Figuur 1.1: Twee reeksen. Drijft de ene reeks de ander?

Pijpers presenteerde zijn vraag aan de hand van een voorbeeld: de aardbevingen in Groningen. De overheid bouwt de gaswinning daar af. “De komende jaren wordt er minder gas gewonnen dan voorheen. De overheid vraagt het CBS of het effect van die reductie aan te tonen is”, legt Pijpers uit. Data zijn er: van de gaswinning, van de bevingen. Kolumbán: “Wat het lastig maakt is dat er op meerdere plekken gas wordt gewonnen, en het onduidelijk is hoe de ondergrond daarop reageert. En dus ook op welke tijdschaal winning effect heeft.” Daardoor is het moeilijk te zeggen of en hoeveel de bevingen verminderen als de winning wordt verlaagd.

Daar komt bij dat de tijdreeks die er is, relatief kort is. Er wordt sinds 1963 gas gewonnen, en vanaf begin jaren 90 is er een toenemend aantal aardbevingen. Volgens Pijpers is ‘Groningen’ daarmee een goed voorbeeld voor de vragen die het CBS vanuit ministeries krijgt. Pijpers: “Veel van onze data komen maandelijks binnen. Beleid kan om de vier jaar veranderen. Het CBS werkt meestal met datareeksen van minder dan 100 datapunten.”

De wiskundigen begonnen tijdens de week niet met de Groningse data. Ze kregen van Pijpers drie oefenreeksen om mee te werken, in oplopende moeilijkheidsgraad. Pijpers wist welk verband er al dan niet tussen de data was, de wiskundigen niet. Aan hen de opgave om uit te vogelen, wat er in de data verstopt zat. Drijft de ene reeks de ander, of andersom, of is er geen drijver en geen volger? Of is het onmogelijk hier iets over te zeggen?

“Doordat het probleem zo helder geformuleerd was, konden we meteen aan de slag. Het fundament voor onze aanpak hebben we op de eerste dag van de week al gelegd”, aldus Kolumbán. Hij is tevreden met het resultaat: de groep is zelfs van plan het na de studiegroep verder uit te werken tot een wetenschappelijke publicatie.

De wiskundigen gingen als volgt te werk. “We maakten een hypothese over de relatie tussen reeks 1 en 2. Hier is achtergrondkennis voor nodig, van een expert, of bijvoorbeeld kennis uit soortgelijke datasets uit een andere landen. Je neemt dan een veronderstelde functie die beschrijft hoe die twee samenhangen.” Bovenop dat signaal zit ruis. Kolumbán benadrukt dat die ruis uit allerlei soorten verstoringen bestaat: “Fouten in de database, verstoringen door andere bronnen, andere toevalligheden - dat noemen we in dit geval allemaal ruis.” De wiskundigen namen aan dat die ruis onafhankelijk is van de drijvende functie. Die aanname is essentieel. “Want alleen als dat zo is, kun je iets zinnigs over causaliteit zeggen”, vertelt Kolumbán. Hij geeft een voorbeeld. “Stel dat je twee dingen bestudeert die steeds exact op het zelfde moment gebeuren, en er is geen ruis. Er gaat bijvoorbeeld steeds een lichtje aan en er klinkt een piep. Dan kun je zonder verdere achtergrondkennis onmogelijk zeggen of het lampje de piep veroorzaakt, of de piep zorgt dat het lampje aangaat, of dat er nog iets anders is dat beide veroorzaakt.” Ruis zorgt ervoor dat dit wel kan. Als de ruis uit reeks 1 invloed blijkt te hebben op het gedrag van reeks 2, en omgekeerd niet, alleen dan kan je concluderen dat reeks 1 reeks 2 drijft. Pijpers: “Dit vind ik echt een prachtige gedachte. Normaal gesproken vinden we verstoringen in de data alleen maar vervelend, maar nu blijken ze essentieel te zijn om onderscheid te kunnen maken tussen oorzaak en gevolg.”

De testmethode die de wiskundigen ontwikkelden was een stappenplan. “We begonnen met een hypothese over de ruis en over de relatie tussen de twee reeksen. We gebruikten een standaardmethode, de Wald-Wolfowitztest, om te testen of onze aanname over de ruis aannemelijk was of niet. Als dat zo was, dan konden we door.” Het uiteindelijke resultaat van de testmethode is een waarschijnlijkheidswaarde voor de functie, die is gekozen als hypothese voor de relatie tussen de reeksen. “En voor de aannamen over de ruis”, benadrukt Kolumbán.

Uiteindelijk aan de gebruiker van de testmethode om te beslissen bij welke uitkomst hij vindt dat causaliteit aannemelijk is - absolute zekerheid geeft deze test immers niet. Pijpers: “Dat vind ik prettig aan deze methode, dat je kan zeggen ‘ik kan bij deze reeks met zoveel procent waarschijnlijkheid zeggen dat X inderdaad de veroorzaker is van Y ’.”

De wiskundigen testten hun methode op de oefenreeksen die Pijpers gesimuleerd had. Bij twee van de drie klopten de resultaten niet. “Dat kwam iedere keer doordat de reeksen niet voldeden aan de eisen die wij bij deze methode hebben. Zo gaan wij uit ervan uit dat de reactietijd van de drijver op de volger niet variabel is, terwijl dat in een van de reeksen wel zo was. Dan werkt het niet”, verklaart Kolumbán.

De testmethode voor causaliteit die de wiskundigen ontwikkelden, moet aan drie voorwaarden voldoen. Zo mag het verband tussen de drijver en de volger niet lineair zijn, moet er een constante tijdsvertraging tussen die twee zitten en moet de ruis van meetpunt tot meetpunt onafhankelijk zijn. Volgens Pijpers valt met alle drie eisen te leven voor het CBS. “Het lastigst is eigenlijk nog die eis van niet-lineariteit; als wij de relatie tussen twee reeksen nog niet kennen, is onze eerste aanname gewoonlijk dat

het verband lineair is. Maar dat is niet onoverkomelijk. Ik denk dat het klopt dat de meeste relaties in werkelijkheid niet-lineair zijn, dus wanneer we deze toets inzetten zullen we een andere inschatting kunnen doen voor het verband tussen de reeksen.”

De week was te kort om daarna nog diep in de Groningse data te duiken. Kolumbán: “We hebben wel een eerste poging gedaan, maar die resultaten waren onvoldoende. De hypothesefunctie die we hadden opgesteld voor de relatie tussen winning en aantal bevingen was vermoedelijk niet goed genoeg.” Waarmee hij niet wil zeggen dat causaliteit aantonen niet mogelijk is. “Wij hebben het in deze korte tijd niet kunnen doen. Je zou meer tijd moeten stoppen in het ontwikkelen van een goede hypothesefunctie.”

Pijpers is niet teleurgesteld dat dit niet is gelukt: “Dat zou ook echt teveel gevraagd zijn.” Met het voorbeeld van de gaswinning wilde hij vooral illustreren wat voor soort vragen aan het CBS wordt gesteld. Dat de wiskundigen hier niet een-twee-drie een conclusie over konden trekken, is volgens Pijpers volstrekt logisch. “Het laat zien hoe lastig de vraag naar de relatie tussen winning en bevingen is. Dat begint al met de vraag wat je als maat neemt voor de aardbevingen? Het aantal aardbevingen? De energie die vrijkomt bij die bevingen? Er is meer achtergrondkennis nodig om daar goed mee aan de slag te kunnen.”

In de slotpresentatie aan het eind van de week concludeerden de wiskundigen dat hun methode de medewerkers van de ministeries niet altijd blij zal maken. ‘Zomaar’ uit twee datareeksen causaliteit afleiden gaat niet; er moet altijd een onderbouwde hypothese aan te pas komen over hoe de twee samenhangen. Bovendien zijn er eisen waaraan voldaan moet zijn om deze methode te kunnen gebruiken. In wiskundig opzicht waren ze tevreden; de groep had immers een nieuwe toets ontwikkeld voor causaliteit bij korte tijdreeksen, en dat resultaat wil ze graag in een wetenschappelijk publicatie uitwerken.

Opdrachtgever Pijpers was uitermate tevreden over de resultaten. “Dit laat zien dat je korte tijdreeksen niet als een black box kan benaderen. Je hebt inhoudelijke kennis nodig voordat je conclusies kan trekken over causaliteit. Ik moet in mijn werk vaak aangeven waar de grenzen liggen van wat je op basis van de data kan concluderen, en wat niet. Ik heb nu een heldere definitie over wanneer je wel, maar vooral ook wanneer je géén conclusies over oorzaak en gevolg kan trekken.” Hij voelt hiermee zich beter beslagen ten ijs komen bij overleg met klanten van het CBS. “Ik had gedacht dat er met korte tijdreeksen niet zoveel mogelijk zou zijn. Dit is toch een mooie toets om in strak gedefinieerde omstandigheden soms iets over causaliteit te kunnen zeggen. Dat is bijzonder waardevol.” Hij vindt dat de studiegroep zijn verwachtingen overtroffen heeft: “Dit is duidelijk meer dan waar ik vooraf op had durven rekenen.”

Team CBS

Sándor Kolumbán (Technische Universiteit Eindhoven)¹, Richard Post¹, Rui Castro¹, Robert Fitzner¹, Arthur Vromans¹, Giovanni Eugenio Comi (Scuola Normale Superiore, Pisa), Frank Pijpers (CBS).