

Statisticus Aad van der Vaart (1959), hoogleraar Stochastiek aan de Universiteit Leiden, ontving in juni de Spinozapremie van NWO voor zijn werk aan het ontwikkelen van modellen om statistiek te bedrijven op grote databestanden. Hij is een van de topwetenschappers in zijn vakgebied, zowel in Nederland als internationaal. Eerder werd zijn werk bekroond met de Van Dantzigprijs en ontving hij een Advanced Grant van de EU voor grensverleggend onderzoek. Door Edith van Gameren

bespreking

Waar richt uw werk zich op?

'Ik doe onderzoek op het gebied van statistiek en kansrekening, en daar dan weer een klein onderdeel van, de niet-parametrische, bayesiaanse statistiek, en dan in het bijzonder onzekerheidskwantificatie.

Ik ben met theorie bezig, maar ook met toepassingen. Bijvoorbeeld op het vlak van genomics, maar ook PET-imaging, signaalanalyse en historische demografie.'

Wat fascineert u daar zo in?

'Statistiek is een hulpwetenschap, die in veel andere wetenschappen gebruikt wordt. Het is leuk daar iets van mee te krijgen en connecties uit andere vakgebieden op te doen. Tegelijkertijd heeft statistiek als vak een geschiedenis van eeuwen en ontwikkelt het zich steeds verder. Ook op zichzelf is het interessant, als tak van wiskunde, met stellingen, bewijzen en theorievorming. Multidisciplinaire projecten zijn interessant, maar ik vind het ook wel eens lekker hier de deur dicht te doen en me helemaal op statistiek te focussen.'

Op welke manier is de statistiek de afgelopen jaren veranderd?

'De eerste grote verandering vond plaats in de jaren '90, de opkomst van de computer was daarin de drijvende kracht. De mogelijkheid ontstond om berekeningen te maken die eerst niet te doen waren. Ik heb specifiek gewerkt aan modellen met veel parameters. Grote datasets zijn daar handig bij: hoe meer datapunten, des te interessanter de berekening. De tweede grote verandering is big data. Nu is dat een beetje een hype-woord van de laatste jaren, maar data is



Statisticus Aad van der Vaart, hoogleraar Stochastiek aan de Universiteit Leiden (Foto: Ivar Pel)

al lange tijd omvangrijker aan het worden. Bijvoorbeeld in de genetica is die groei er het afgelopen decennium heel sterk geweest. Bij het automatisch verzamelen van allerlei data door bedrijven als Google gaat het nog sneller.'

Welke rol speelt statistiek in die big data?

'Onze taak is om in al die data het signaal te ontdekken dat niet toevallig is, de toevalsvariantie. Bijvoorbeeld: je hebt 30.000 genen, waarvan je de expressie wilt meten – of ze actief zijn of niet. Je doet onderzoek bij enkele honderden of wellicht enkele duizenden mensen. Die populatie is nog altijd te klein om de ruis uit de metingen te halen. De statistiek moet helpen het signaal te vinden en een goede foutmarge hiervan aan te geven. Dat kun je doen met allerlei modellen. De onzekerheidskwantificatie is een van de belangrijkste uitdagingen. Hoe ga je om met die multidimensionaliteit die je krijgt als je zo veel factoren tegelijk onderzoekt? Toen het genenonderzoek net begon dacht men dat gepersonaliseerde geneeskunde binnen handbereik lag, maar het blijkt toch wel heel erg lastig om de netwerken te ontrafelen van alle genen en de manieren waarop ze samenwerken.'

Ligt hier een rol voor informatica-onderzoek?

'Er zijn talloze computationele vragen, waarvoor nieuwe algoritmes nodig zijn, bij uitstek het terrein van informatici uiteraard. Er is ook zeker veel uitwisseling tussen die vakgebieden. Dat geldt vooral op het gebied van machine learning, waar de statistiek en de ICT samenkomen. Overigens houdt slechts een klein groepje zich daarmee bezig in Nederland, de omgeving hier kan wat dat betreft beter.'

Hoe gaat u de Spinozapremie inzetten?

'Ik ga wat kleine dingen doen, zoals congressen bezoeken, maar de hoofdmoot van het geld gaat uiteraard naar salarissen: mensen aanstellen. Op dit moment ben ik veel bezig in de niche van bayesiaanse statistiek, maar ik hoef daar niet per se in uit te breiden. Ik zoek juist mensen die hun eigen weg vinden in de wetenschap, hun eigen specialisme hebben. Mijn doel is statistiek nog beter neer te zetten in Leiden. Liefst zou ik een structurele uitbreiding zien, in de VS of Engeland bijvoorbeeld is statistiek als vakgebied meer ontwikkeld.'