

De orde van de waarschijnlijkheid

Dominee Thomas Bayes veroorzaakte een revolutie in de statistiek

De Academische Boekengids 55, maart 2006, pp. 15-17.

Het probabilisme van Thomas Bayes veroorzaakte niet alleen in de achttiende eeuw een revolutie in de wiskunde. Dit waarschijnlijkheidsprincipe wordt nu, mede door de onstuitbare opmars van de computer, opnieuw ingezet om actuele vragen te beantwoorden.

In 1763 publiceerde de Londense Royal Society in haar tijdschrift *Philosophical Transactions* een essay van de twee jaar eerder overleden dominee **Thomas Bayes**, onder de titel 'An Essay Towards Solving a Problem in the Doctrine of Chances'. Gezien de aanvankelijke stilte rond dit artikel is het moeilijk voorstelbaar dat Bayes hiermee een werkelijk revolutionaire stap in de statistiek zette. Het klinkt mogelijk nog merkwaardiger dat de afgelopen vijftien jaar een groeiend aantal publicaties in de statistiek en de epistemologie de eerbiedwaardige dominee in hun titel vermelden. Uitstekende voorbeelden van deze opleving zijn de recent verschenen boeken van Press, Williamson, Howson en Urbach, en Bovens en Hartmann. Hoog tijd dus om het gedachtegoed van Bayes bij een groter publiek onder de aandacht te brengen.

Wanneer Bayes zijn essay schrijft, zijn de begrippen 'kans' en 'waarschijnlijkheid' al ongeveer een eeuw onderwerp van discussie. Rond 1654 discussiëren Fermat en Pascal over de manier waarop, bij een voortijdig afgebroken kansspel, de inzet moet worden verdeeld; een probleem dat dan al een lange, en ongetwijfeld ook gewelddadige, geschiedenis heeft. Huygens schrijft in 1657 de eerste gedrukte verhandeling over kansspelen en maakt het gokken daarmee tot een tak van wiskunde. Maar bij hem gaat het nog uitsluitend over kans. Dat wil zeggen: over dobbelstenen, muntworpen en roulette.

Vijf jaar later snijden Pascal en zijn jansenistische broeders een geheel ander onderwerp aan met hun presentatie van de zogenaamde Port Royal logica. Deze brengt waarschijnlijkheid in verband met onzekere overtuigingen en het verwachte nut van beslissingen. Weer vijf jaar later stelt Leibniz voor om de kracht van een juridische argumentatie te meten met zogenaamde *probabilitas*, weer te geven op een schaal van nul tot één. Bij Pascal en Leibniz gaat het niet om kansen op uitkomsten van herhaalbare worpen van een dobbelsteen, of kortweg om kansen op gebeurtenissen, maar vooral om de waarschijnlijkheid van overtuigingen. Precies dit duale karakter van het probabilisme, tot uitdrukking gebracht in het onderscheid tussen de termen 'kans' en 'waarschijnlijkheid', vormt de inzet van Bayes' revolutionaire bijdrage.

'BAYES IS DE EERSTE DIE DE OMGEKEERDE BEWEGING MAAKT: ALS DE OPZET VAN EEN KANSSPEL ONS VERTELT WAT DE KANS OP EEN BEPAALDE UITKOMST IS, DAN MOET EEN BEPAALDE UITKOMST ONS OOK IETS KUNNEN VERTELLEN OVER DE OPZET VAN HET KANSSPEL DAT DIE UITKOMSTEN LEVERT.'

Hoewel in Bayes' tijd de noties kans en waarschijnlijkheid beide al bekend zijn, beperken toepassingen van wiskundige berekeningen zich bijna altijd tot kans. De vraag is dan stevast om, gegeven een bepaald kansspel - dobbelen bijvoorbeeld - een uitdrukking te vinden voor de kans op een bepaalde uitkomst of gebeurtenis, zoals de kans op drie zessen in de eerste drie worpen. Bayes is de eerste die de omgekeerde beweging maakt, op basis van een even geniale als eenvoudige overweging: als de opzet van een kansspel ons vertelt wat de kans op een bepaalde uitkomst is, dan moet een bepaalde uitkomst ons ook iets kunnen vertellen over de opzet van het kansspel dat die uitkomsten levert. De dobbelsteen in het kansspel kan eerlijk zijn, maar het is ook mogelijk dat de zijde waarop de één staat, verzaaid is, zodat er vaker zes wordt gegooid. Het idee van Bayes is dat we de waarschijnlijkheid kunnen berekenen dat een dobbelsteen eerlijk of verzaaid is, gegeven dat we drie zessen hebben gegooid.

Het ligt voor de hand te denken dat dominee Bayes met deze redeneringen een agenda had. Ongeveer een kwarteeuw eerder heeft de bijtende kritiek van David Hume de vernislaag van de religieus gemotiveerde natuurwetenschap ernstig aangetast. Hume argumenteert dat elke conclusie over natuurlijke verschijnselen die de onmiddellijke ervaringsinhoud van die verschijnselen overstijgt, gebaseerd moet zijn op instinct, gewoonte of loze speculatie. Het boek der natuur, dat net als de Bijbel Gods bedoelingen openbaart, is weliswaar leesbaar, maar het gaat te ver om er wetteksten in te zien.

De methode van Bayes lost dit probleem op het eerste gezicht op. Op basis van geobserveerde verschijnselen kunnen we volgens Bayes wel degelijk conclusies trekken over de wetten die daarachter steken. Maar al die kennis is niet helemaal gratis: Bayes vertelt ons niets over het soort kansspel waarmee de natuur ons confronteert. Om de methode van Bayes toe te passen op natuurlijke verschijnselen moeten we eerst een verzameling van mogelijke kansspelen formuleren volgens welke de natuur ons de uitkomsten van observaties en experimenten voorschotelt. Met andere woorden: we worden gedwongen om hypothesen te formuleren.

In de gebruiksaanwijzing van deze dure conceptuele aankoop staan bovendien nog wat kleine lettertjes. Om Bayes' methode toe te passen, moeten we over de gekozen verzameling hypothesen beginwaarschijnlijkheden aannemen. We moeten dus vooraf al een oordeel hebben over de hypothesen, en dat terwijl we dit oordeel juist op basis van de ervaring wilden vormen. Het is precies deze ogenschijnlijke circulariteit waarop Bayes door de generatie statistici van begin twintigste eeuw wordt afgerekend.

De Gentse hoogleraar Quetelet, inderdaad die van de Quetelet Index (over de relatie tussen lengte en gewicht), doet in de jaren dertig van de negentiende eeuw grootscheeps demografisch en sociaal-wetenschappelijk onderzoek. Hij merkt natuurlijk dat er overal variaties tussen individuen zijn, maar raakt vooral onder de indruk van de regelmaat die zichtbaar wordt door naar een artificiële 'gemiddelde mens' te kijken. Dit thema wordt herhaald in de kinetische gastheorie van de fysicus Maxwell: hoewel het gedrag van individuele gasdeeltjes grillig en onvoorspelbaar is, zijn via de statistiek allerlei interessante conclusies te trekken over het gedrag van de gaswolk als geheel.

'DE MASCOTTE VAN DE KLASSIEKE STATISTIEK IS DE DAME DIE BEWEERT AAN HAAR THEE TE KUNNEN PROEVEN IN WELKE VOLGORDE THEE EN MELK IN HAAR KOPJE ZIJN GESCHONKEN.'

Het kernpunt in de alternatieve visie op statistiek, die vanaf Quetelet en Maxwell tot ontwikkeling komt, is dat kansen onderdeel uitmaken van de natuurlijke orde, en dat voor het bepalen daarvan helemaal geen beginwaarschijnlijkheden nodig zijn. Gesteund door deze overtuiging trekken Neyman, Fisher en vader en zoon Pearson de zogenaamde klassieke statistiek op, waarin zij uitsluitend nog van kansen op uitkomsten spreken, en met nadruk niet van waarschijnlijkheden over hypothesen. De mascotte van de klassieke statistiek is de dame die beweert aan haar thee te kunnen proeven in welke volgorde thee en melk in haar kopje zijn geschonken. Fisher laat zien dat wij deze theekransopschepperij op waarheid kunnen toetsen door enkel te testen op uitkomsten, zonder dat we ons vooraf al een mening hoeven te vormen. Wanneer de dame in een door ons ingeschonken serie kopjes thee met melk de volgorde te vaak verkeerd raadt, mogen we haar bewering verwerpen.

De statistiek van Fisher en Neyman-Pearson slaat goed aan in de wetenschappen, maar tegelijk is de tegenbeweging al ingezet. Veel wiskundigen en filosofen, onder wie Ramsey, Jeffreys, De Finetti, Savage, en veel recenter Howson en Urbach, blijven in al die jaren de tekortkomingen in de klassieke visie op kansen onderstrepen. En zij concentreren zich, hoe kan het ook anders, opnieuw op de kwestie van de beginwaarschijnlijkheden over hypothesen. Voor de deur van de Faculteit Wiskunde, een discipline die toch het toonbeeld van redelijkheid moet zijn, gaan de statistici rollebollend over straat.

Een van de controversen mag daarbij niet onvermeld blijven: de zogenaamde Harvard Medical School test. Het blijft een van de eenvoudigste illustraties van het belang van beginwaarschijnlijkheden. Stel dat wij beschikken over een test waarvan bekend is dat deze slechts in één op de duizend gevallen de noodklok luidt, en dus positief uitvalt, bij iemand die helemaal niet ziek is, terwijl de test altijd positief uitvalt bij mensen die wel ziek zijn. Als een zekere Joep positief test, moeten we dan besluiten dat Joep waarschijnlijk ziek is, en dus daarnaar handelen? Volgens de officiële lezing van de klassieke statistiek wel. Maar wat als we ook nog weten dat de bewuste ziekte slechts bij één op de miljoen mensen optreedt? Als we bijvoorbeeld één miljoen mensen testen, dan zijn er van hen hooguit een paar echt ziek, terwijl er rond de duizend zijn die ten onrechte positief testen. De waarschijnlijkheid is dus veel groter dat onze Joep deel uitmaakt van de groep die ten onrechte positief test.

Een ander voorbeeld. Twee onderzoekers meten aan dezelfde opstelling. De één meet tot zij vijftien meetresultaten heeft, de ander tot zij zin krijgt in de lunch. Maar stel nu dat de honger bij de tweede onderzoeker precies begint te knagen als zij vijftien metingen heeft verricht, en dat deze bovendien precies dezelfde zijn als die van de eerste onderzoeker. Dan nog mogen beide onderzoekers volgens de klassieke statistiek niet dezelfde conclusies trekken. De tweede onderzoeker had een ander aantal metingen kunnen verrichten, en moet daarom volgens de klassieke statistiek een andere analyse doen. De invloed van deze niet-gerealiseerde metingen is zelfs zo groot dat, in de befaamde woorden van de statisticus Harold Jeffreys, een ware hypothese kan worden verworpen omdat zij iets wat niet wordt gemeten, een kans nul geeft!

'ALS WE STATISTISCHE TECHNIEKEN OPVATTEN ALS REDENERINGEN GEEFT ALLEEN DE BAYESIAANSE STATISTIEK GARANTIE OP GELDIGHEID.'

Het belang van bovenstaande controversen werd kort geleden weer eens duidelijk. In de rechtszaak over de betrokkenheid van de verpleegkundige Lucy de B. bij enkele onverwachte sterfgevallen maakten statistische analyses expliciet deel uit van de juridische argumentatie. In zulke gevallen is het van levensbelang dat de argumentatie geldig is. De bovenstaande controversen tonen dat, als we statistische technieken opvatten als redeneringen, alleen de bayesiaanse statistiek garantie geeft op geldigheid.

Kortom, er zijn conceptuele redenen om gebruik te maken van de bayesiaanse statistiek. Maar is zij wel praktisch toepasbaar? Een deel van de populariteit van de klassieke statistiek is gelegen in de rekentechnisch spaarzame technieken die zij aandraagt, terwijl de bayesiaanse statistiek juist om arbeidsintensieve berekeningen vraagt. Gelukkig wordt het sinds een jaar of tien steeds gemakkelijker

om ook deze berekeningen uit te voeren, dankzij de onstuitbare opmars van de microcomputer en de ontwikkeling van allerhande simulatietechnieken. Een belangrijk praktisch bezwaar tegen de bayesiaanse statistiek is daarmee van de baan.

De computer biedt overigens niet alleen de noodzakelijke rekenkracht. Zoals Williamson overzichtelijk uitlegt, zijn de laatste jaren in de hoek van de artificiële intelligentie ook verscheidene interessante toepassingen van bayesiaanse ideeën ontwikkeld. Het blijkt dat statistische modellen elegant te representeren zijn in termen van netwerken, waarin modelvariabelen met pijltjes worden verbonden. Met enige welwillendheid kunnen deze bayesiaanse netwerken bovendien voorzien worden van een interpretatie in termen van oorzaak en gevolg, zodat de causale structuur van een statistisch model in één oogopslag afleesbaar wordt. Deze bayesiaanse netwerken, waarin causale hypothesen van waarschijnlijkheden voorzien worden en waarin waarschijnlijkheden snel en overzichtelijk kunnen worden bijgesteld, zijn bij uitstek bruikbaar voor toepassing in het wetenschappelijk onderzoek. Ze zijn een aanwinst in de gereedschapskist van elke empirische onderzoeker.

‘DEZE BAYESIAANSE NETWERKEN ZIJN EEN AANWINST IN DE GEREEDSCHAPSKIST
VAN ELKE EMPIRISCHE ONDERZOEKER.’

Hoe mooi al dit bayesiaanse gereedschap ook is, het is de vraag of het snel zijn weg naar de wetenschappelijke consument zal vinden. Statistiek staat nu eenmaal niet in het centrum van de onderzoeks aandacht. Bovendien bekijken de gezaghebbende tijdschriften methodologische vernieuwingen vaak met argwaan. De trage dissipatie van nieuwe statistische technieken in de wetenschappen is het logische gevolg. Hopelijk brengen overzichtswerken van de bayesiaanse technieken, zoals die van James Press, daarin verandering.

Terwijl de invloed van het bayesianisme op de empirische wetenschappen nog niet tot volle bloei is gekomen, heeft het al wel een volwaardige opdrachtgever gevonden in de wetenschapsfilosofie. Niet alleen zijn er successen behaald bij het analyseren van waarschijnlijkheidsuitspraken in wetenschapsfilosofische argumenten, ook is er vooruitgang geboekt in het modelleren van epistemologische vraagstukken. Wetenschappers passen een groot aantal methodologische principes toe bij het bepalen van hun voorkeur voor de ene of de andere wetenschappelijke theorie. Zo zijn we geneigd te geloven dat een theorie die op meer gevarieerde evidentie berust een grotere waarschijnlijkheid heeft dan eentje waarvoor dat niet geldt. Soortgelijke ‘epistemische deugden’ zijn eenvoud, samenhang en mathematische elegantie. Maar in de discussies blijft ten eerste vaak onduidelijk wat precies verscheidenheid van evidentie, eenvoud, samenhang en mathematische elegantie zijn en, ten tweede, hoe deze deugden precies van invloed kunnen zijn op de waarschijnlijkheid van een theorie.

De laatste jaren is een aantal pogingen ondernomen om bayesiaanse modellen te bouwen waarmee dit soort methodologische vraagstukken kunnen worden aangepakt. Modelleren is natuurlijk niets nieuws: de evolutie, een ontploffende supernova, het maagdarmkanaal van de mens en de vertragingen op het spoorwegennet – je kunt het zo gek niet bedenken of er is wel een betrekkelijk goede simulatie van gemaakt. De voordelen daarvan zijn evident: niet alleen kun je zeldzame gebeurtenissen bij herhaling nabootsen, ook kun je met een krachtige computer allerlei alternatieve mogelijkheden narekenen. In de afgelopen tien jaar zijn epistemologen zich gaan realiseren dat met zulke simulaties grote winst te boeken is. Zo kunnen we met behulp van een geïdealiseerd model van een wetenschappelijke theorie, bestaande uit een aantal hypothesen waarvan we de precieze waarschijnlijkheden kennen, formules definiëren die aangeven hoe goed de theorie intern samenhangt (de zogenaamde coherentiematen) of hoe sterk zij ondersteund wordt door opnieuw geïdealiseerde evidentie (de zogenaamde confirmatiematen). Met behulp van de laatste soort maten kunnen we dan aantonen dat, om maar een oud voorbeeld van stal te halen, de stelling ‘alle raven zijn zwart’ minder sterk ondersteund wordt door een witte schoen dan door een zwarte raaf.

In dit verhaal nemen de coherentiematen een speciale plaats in. Zij stellen ons in staat nieuw licht te werpen op een vraag die al generaties filosofen uit hun slaap gehouden heeft: bestaat er een verband tussen waarheid en coherentie? Zogenaamde coherentisten menen dat onze overtuigingen meer of minder gerechtvaardigd zijn naarmate zij meer of minder met elkaar samenhangen. Om deze opvatting van rechtvaardiging opgeld te laten doen, moeten zij aantonen dat de betere coherentie van een theorie leidt tot een hogere waarschijnlijkheid voor die theorie. Lange tijd bleven veel van de discussies rond dit onderwerp steken in theoretisch gesteggel over de definitie van coherentie en de manier waarop deze eigenschap waarheidsbevorderend zou kunnen werken. Maar met behulp van een maat van coherentie is de vraag naar de relatie tussen coherentie en waarheid vrij eenvoudig te beantwoorden: kijk naar de gevallen waarin de coherentie van een theorie toeneemt en bereken of in al die gevallen de waarschijnlijkheid van de theorie ook toeneemt.

Het bayesiaanse instrumentarium is daarnaast bruikbaar voor het doen van uitspraken over de validiteit van methodologische principes als eenvoud en verscheidenheid. Wanneer we een model maken waarin we als invoer een aantal bronnen nemen die informatie rapporteren, en als uitvoer de waarschijnlijkheid bekijken van de totale gerapporteerde informatie, dan kunnen we evalueren in hoeverre een grotere verscheidenheid van bronnen, gemodelleerd als een grotere onafhankelijkheid tussen deze bronnen, leidt tot een hogere waarschijnlijkheid van het gerapporteerde. Dat deze vorm van probabilistisch modelleren succesvol kan zijn, laten Bovens en Hartmann zien in hun *Bayesian Epistemology*. In dit relatief korte en

bij tijd en wijle zeer technische boek laten zij een aantal methodologische vraagstukken de revue passeren en proberen zij met een sterk gesimplificeerd probabilistisch model uitspraken te doen over de geldigheid van deze principes. Dat dit in sommige gevallen tot verbazingwekkende conclusies kan leiden, blijkt al uit hun stelling dat de verscheidenheid van evidentie alleen een pre is in gevallen waarin we ervan overtuigd zijn dat de bronnen die de informatie rapporteren redelijk betrouwbaar zijn. Zodra we ervan uitgaan dat de bronnen eigenlijk onbetrouwbaar zijn, is het juist beter om twee overeenkomende resultaten uit soortgelijke bronnen te ontvangen.

De gemiddelde filosoof zal zich misschien weinig gelegen laten liggen aan een theoretisch model van de verscheidenheid van bronnen, waaruit met een computer bepaalde uitkomsten berekend zijn. Daarom is het voor het probabilistisch modelleren van zulke epistemologische vraagstukken van het grootste belang om bij elke verbazingwekkende conclusie ook uit te leggen hoe en waarom deze afwijkt van onze verwachtingen. Bovens en Hartmann zijn zich hiervan terdege bewust en proberen voor ieder onverwacht resultaat een intuïtieve verklaring te geven. De reden dat we bijvoorbeeld in het geval van ogenschijnlijk zeer onbetrouwbare bronnen liever twee identieke resultaten hebben van soortgelijke bronnen, is dat deze resultaten ons vertrouwen in de bronnen sterk zullen doen stijgen. De waarschijnlijkheid dat een mogelijk erg onbetrouwbare bron twee keer met hetzelfde verkeerde resultaat komt, is immers klein. Als twee soortgelijke bronnen dan toch met dezelfde informatie komen, heeft dat zijn weerslag op de betrouwbaarheid ervan. Met een dergelijke uitleg houden Bovens en Hartmann mooi het midden tussen traditionele filosofie en computationeel ingenieurswerk. Terwijl de modellen resulteren in precieze kwantitatieve voorspellingen, leiden de beschouwingen die erop volgen tot een wijsgerig begrip van de modellen. Dat komt uiteindelijk de inzetbaarheid van hun resultaten in de wetenschappelijke praktijk ten goede.

‘JUUST OMDAT HET PROBABILISTISCH MODELLEREN ZICH NIET TOESPITST OP EEN SPECIFIEKE VORM VAN WETENSCHAPPELIJK ONDERZOEK, STAAT HET ONS TOE DE ALGEMENE CONCLUSIE TE TREKKEN DAT ER EEN VERBAND BESTAAT TUSSEN DE WAARDE VAN GEVARIËERDE EVIDENTIE EN DE BETROUWBAARHEID VAN DE BRONNEN. DIT IS, VANUIT EEN FILOSOFISCH PERSPECTIEF, EEN ZEER INTERESSANT RESULTAAT.’

Aan het probabilistisch modelleren van epistemologische en wetenschapsfilosofische vragen kleeft een even groot als evident nadeel. Want net zoals modellen van de vertragingen op het spoor geen rekening kunnen houden met tal van oorzaken van de vertragingen, blijkt ook de wetenschappelijke praktijk een stuk weerbarstiger dan de formele idealisatie. Niemand zal bijvoorbeeld willen beweren dat het bovengenoemde model voor de verscheidenheid van evidentie een exacte tegenhanger heeft in een daadwerkelijk uitgevoerd wetenschappelijk onderzoek. Maar de grootste zwakte vormt tegelijk de kracht: precies omdat het probabilistisch modelleren zich niet toespitst op een specifieke vorm van wetenschappelijk onderzoek, staat het ons toe de algemene conclusie te trekken dat er een verband bestaat tussen de waarde van gevarieerde evidentie en de betrouwbaarheid van de bronnen. En dit is, vanuit een filosofisch perspectief, een zeer interessant resultaat.

Juist omdat het een idealisering van de wetenschappelijke praktijk betreft, moeten we ervoor waken te sterke conclusies te verbinden aan de resultaten. Dat we bijvoorbeeld geen harde relatie vinden tussen coherentie en waarheid in ons model betekent niet noodzakelijkerwijs dat zo'n relatie er in de praktijk ook niet is, laat staan dat zij er niet kan zijn. Een goed voorbeeld van dit type valkuil is Olssons argument in *Against Coherence*, waarbij hij een model presenteert voor het vergaren van informatie waarin niet altijd een relatie bestaat tussen coherentie en waarheid. Op basis van dit soort modellen trekt hij een aantal pessimistische conclusies over het coherentisme. Maar voordat we overgaan tot zulke sterke conclusies moet natuurlijk eerst overtuigend aangetoond zijn dat de waargenomen verschijnselen niet slechts een artefact zijn van het model, maar ons daadwerkelijk iets kunnen leren over de praktijk. En helaas kunnen formalismen en computers ons hierbij slechts zeer beperkt terzijde staan.

Jan-Willem Romeijn is als wetenschapsfilosoof verbonden aan de Faculteit der Maatschappij- en Gedragwetenschappen van de Universiteit van Amsterdam. Hij promoveerde in juni 2005 op een proefschrift over de filosofische grondslagen van de statistiek, getiteld *Bayesian Inductive Logic*.

Wouter Meijs is als wetenschapsfilosoof verbonden aan de Faculteit der Wijsbegeerte van de Erasmus Universiteit Rotterdam. Hij promoveerde in mei 2005 op een proefschrift over probabilistische epistemologie, getiteld *Probabilistic Measures of Coherence*.

Besproken boeken:

Against Coherence. Truth, Probability and Justification
door **Erik Olsson**
Oxford University Press. Oxford 2005.
246 pag., € 57,95

Bayesian Epistemology
door **Luc Bovens en Stephan Hartmann**
Oxford University Press. Oxford 2003.
180 pag., € 28,15

Bayesian Nets and Causality. Philosophical and Computational Foundations
door **Jon Williamson**
Oxford University Press. Oxford 2005.
239 pag., € 67,00

Scientific Reasoning. The Bayesian Approach
door **Colin Howson en Peter Urbach**
Open Court Publishers. Illinois 2006.
272 pag., € 34,34

Subjective and Objective Bayesian Statistics. Principles, Models, and Applications
door **James Press**
John Wiley & Sons. New York 2003.
600 pag., € 100,00