

DE REVOLUTIE DER PROBABILISTEN

Jan-Willem Romeyn

Wouter Meijs

Subjective and Objective Bayesian Statistics: Principles, Models, and Applications, James Press, John Wiley, 2003.

Scientific Inference: the Bayesian Approach, Colin Howson en Peter Urbach, Open Court Publishers, 2006.

Bayesian Nets and Causality: Philosophical and Computational Foundations, Jon Williamson, Oxford University Press, 2005.

Bayesian Epistemology, Luc Bovens en Stephan Hartmann, Oxford University Press, 2003.

Against Coherence, Erik Olsson, Oxford University Press, 2005.

Inleiding

In 1763 publiceerde de Royal Society in haar 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". Door de aanvankelijke stilte rondom het essay is maar moeilijk voorstelbaar dat daarmee een werkelijk revolutionaire stap in de statistiek werd gezet. En het klinkt mogelijk nog merkwaardiger dat de afgelopen vijftien jaar een groeiend aantal publicaties in de statistiek en de epistemologie de eerbiedwaardige dominee in de titel vermelden. Van deze recente opleving zijn de recente boeken van Press, Williamson, Howson en Urbach, en Bovens en Hartmann uitstekende voorbeelden. Hoog tijd dus om het gedachtegoed van Bayes bij een groter publiek onder de aandacht te brengen.

Van casino naar academie

Als 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 deze verhandeling gaat uitsluitend nog over kans, dat wil zeggen, over dobbelstenen, muntworpen en roulette.

Eigenlijk wordt vijf jaar later, wanneer Pascal en zijn Jansenistische broeders de zogenaamde Port Royal logica presenteren, een geheel ander onderwerp aangesneden. Hierin wordt waarschijnlijkheid in verband gebracht met onzekere overtuigingen en het verwachte nut van beslissingen. En weer vijf jaar later stelt Leibniz voor om de kracht van juridische argumentatie te meten met zogenaamde *probabilitas*, weer te geven op een schaal van nul tot één. In deze verhandelingen gaat het niet om kansen op uitkomsten van herhaalbare worpen van een dobbelsteen, of kortweg om kansen op gebeurtenissen, maar in eerste instantie om de waarschijnlijkheid van overtuigingen. Het is precies dit duale karakter van het probabilisme, tot uitdrukking gebracht in het onderscheid tussen de termen kans en waarschijnlijkheid, dat de inzet van Bayes' revolutionaire bijdrage vormt.

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 steevast om, gegeven een bepaald kansspel, dobbelen bijvoorbeeld, een uitdrukking te vinden voor de kans op een bepaalde uitkomst of gebeurtenis, bijvoorbeeld 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 kunnen berekenen wat de waarschijnlijkheid is dat een dobbelsteen eerlijk of verzaaid is, gegeven dat we drie zessen hebben gegooid.

Natura magister

Het ligt voor de hand te denken dat de dominee Bayes met deze redeneringen een agenda had. Ongeveer een kwart eeuw eerder had 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*, waarin net als in de bijbel Gods bedoelingen geopenbaard worden, is weliswaar leesbaar, maar het gaat te ver om er wetteksten in te zien.

Met de methode van Bayes wordt dit probleem op het eerste gezicht opgelost. Op basis van geobserveerde verschijnselen kunnen we volgens Bayes wel degelijk conclusies trekken over de wetten die achter de verschijnselen steken. Maar is al die kennis dan helemaal gratis? Merk op dat Bayes ons niets vertelt over het soort kansspel waarmee de natuur ons confronteert. Om de methode van Bayes op natuurlijke verschijnselen toe te passen moeten we eerst een verzameling van mogelijke kansspelen formuleren volgens welke de natuur ons de uitkomsten van observaties en experimenten voorschotelt. We worden, met andere woorden, 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 dat 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.

Fishers thee drinkende dame

De Gentse hoogleraar Quetelet, inderdaad die van de Quetelet-index, 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.

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 uitsluitend nog van kansen op uitkomsten wordt gesproken, 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, waaronder Ramsey, Jeffreys, De Finetti, Savage, en veel recenter Howson en Urbach, blijven in al die jaren hameren op de tekortkomingen in de klassieke visie op kansen. En die concentreren zich, hoe kan het ook anders, weer 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.

Harvard Medical School

Een van de controverses mag daarbij zeker 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 zij slechts in één op de duizend gevallen de noodklok luidt, en dus positief uitvalt, onder mensen die helemaal niet ziek zijn, terwijl de test altijd positief uitvalt onder 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 nu als we ook nog weten dat de bewuste ziekte slechts bij een op de miljoen mensen optreedt? Als we bijvoorbeeld een miljoen mensen testen, dan zijn er daarvan hooguit een paar mensen echt ziek, terwijl er rond de duizend mensen 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 een 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 metingen 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 dat niet wordt gemeten, een kans nul geeft!

Het belang van bovenstaande controverses werd kort geleden weer eens duidelijk. In de rechtszaak over de betrokkenheid van de verpleegkundige Lucy de B. bij enkele onverwachtse 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 controverses tonen dat, als we statistische technieken opvatten als redeneringen, alleen de Bayesiaanse statistiek garantie geeft op geldigheid.

Causaliteit en de computer

Er zijn, kortom, conceptuele redenen om gebruik te maken van de Bayesiaanse statistiek. Maar is die statistiek 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 makkelijker om ook de veel uitgebreidere Bayesiaanse 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 voor de Bayesiaanse statistiek. Zoals in het boek van Williamson overzichtelijk wordt uitgelegd, zijn in de hoek van de artificiële intelligentie over de laatste jaren ook een aantal 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 een oogopslag afleesbaar wordt. De 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.

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

Probabilistische epistemologie

Terwijl haar invloed op de empirische wetenschappen dus nog niet tot volle bloei is gekomen, heeft het Bayesianisme al wel een volwaardige opdrachtgever gevonden in de wetenschapsfilosofie. En niet alleen zijn er successen behaald bij het analyseren van waarschijnlijkheidsuitspraken die in wetenschapsfilosofische argumenten gebruikt worden, 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 bijvoorbeeld geneigd om te geloven dat een theorie die op meer gevarieerde evidentie berust een grotere waarschijnlijkheid heeft dan een theorie waarvoor dat niet het geval is. Soortgelijke “epistemische deugden” zijn eenvoud, samenhang en mathematische elegantie. Echter, in de discussies blijft vaak onduidelijk wat precies eenvoud, samenhang of verscheidenheid van evidentie is en, ten tweede, hoe deze deugden precies van invloed kunnen zijn op de waarschijnlijkheid van een theorie.

De laatste jaren zijn er een aantal pogingen ondernomen om Bayesiaanse modellen te bouwen waarmee dit soort methodologische vraagstukken kunnen worden aangevat. Modelleren is natuurlijk niets nieuws: de evolutie, een ontploffende supernova, het maagdarkanaal 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, maar bovendien 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 geboekt kan worden. Zo kunnen we met behulp van een geïdealiseerd model van een wetenschappelijke theorie, bestaand 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 de theorie ondersteund wordt door wederom geïdealiseerde evidentie (de zogenaamde confirmatiematen). Met behulp van de laatste soort maten kan dan aangetoond worden 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 de slaap gehouden heeft: is er een verband tussen waarheid en coherentie? Zogenaamde coherentisten zijn ervan overtuigd dat onze overtuigingen meer of minder gerechtvaardigd zijn naar mate 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.

Verscheidenheid aan evidentie

Het Bayesiaanse instrumentarium kan daarnaast gebruikt worden om uitspraken te doen over de validiteit van methodologische principes zoals eenvoud en verscheidenheid. Als 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 maar bij tijd en wijle zeer technische boek laten zij een aantal methodologische vraagstukken de revue passeren en proberen 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 onverwachts resultaat een intuïtieve verklaring te geven. De reden voor het feit 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 toenemen. 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 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. En dat komt uiteindelijk de inzetbaarheid van hun resultaten in de wetenschappelijke praktijk ten goede.

Gevaren van probabilistisch modelleren

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 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. Echter, haar grootste zwakte is tevens haar kracht: precies omdat het zich niet toespitst op een specifieke vorm van wetenschappelijk onderzoek, staat het ons toe om de algemene conclusie te trekken dat er een verband is 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 wetenschappelijk praktijk betreft, moeten wij ervoor waken om te sterke conclusies te verbinden aan de gevonden 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 Olsson's argument in *Against Coherence*, waarin hij een model voor het vergaren van informatie presenteert waarin er niet altijd een relatie is tussen coherentie en waarheid. Op basis van dit soort modellen trekt hij een aantal pessimistische conclusies over het coherentisme. Maar voordat men overgaat 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.

Personalia

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 juni 2005 op een proefschrift over probabilistische epistemologie, getiteld *Probabilistic Measures of Coherence*.