STAtOR

periodiek van de VVS jaargang 5 nummer 2 juni 2004

THEMA
Statistiek in de rechtszaal

‘Geacht hof, het was geen toeval. De rest is aan u.’

Voorzichtig met statistiek in de rechtszaal!

Statistiek als bewijsmiddel in het strafrecht

Niets nieuws onder de zon

Lucia de B. krijgt in hoger beroep levenslang

STAtOR
Jaargang 5, nummer 2, juni 2004

STAtOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VVS). *STAtOR* wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 4 keer per jaar.

Kopij en reacties richten aan
Prof. dr. ir. D. den Hertog (hoofdredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 466 2122, <D.denHertog@kub.nl>.

Leden- en abonnementenadministratie van de VVS
VVS, Postbus 2095, 2990 DB Barendrecht, telefoon 0180 - 623796, fax 0180 - 623670, e-mail <admin@vvs-or.nl>. Raadpleeg onze website over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op *STAtOR* of op een van de andere periodieken.

VVS-website
http://www.vvs-or.nl

Advertenties
Uiterlijk vier weken voor verschijnen te zenden aan Pharos, Moeflonstraat 5, 6531 JS Nijmegen, telefoon 024 - 3559214, e-mail <hootegem@xs4all.nl>. *STAtOR* verschijnt in maart, juni, september en december.

Ontwerp en opmaak
Pharos / M. van Hootegem, Nijmegen

Uitgever

ISSN 1567-3383

Inhoud

Statistiek in de rechtszaal

In de Angelsaksische landen is toepassing van statistiek in de rechtszaal veel gebruikelijker dan in Nederland. Dat heeft wellicht te maken met het feit dat het vak daar al veel eerder en veel sterker heeft weten door te dringen in de samenleving. Niettemin mochten statistisch deskundigen onlangs ook voor de Nederlandse rechter optreden in twee zogeheten dienstroosterzaken. De meest recente, de strafzaak tegen Lucia de B. (LdB), heeft tot veel discussie en opwinding geleid, vooral onder statistici zelf. De felle discussie is behalve in het *Nederlands Juristenblad* (2003-34, 2004-13) breed uitgemeten in de landelijke pers met artikelen in onder meer de *Haagsche Courant* (30 januari 2004 en 10 maart 2004), *Trouw* (13 maart 2004) en *NRC Handelsblad* (13 maart 2004) en gevolgd door een stroom van ingezonden brieven. Het emotionele gehalte wordt door koppen als ‘Bij toeval veroordeeld?’ en ‘Door statistici veroordeeld?’ aardig geïllustreerd. De VVS organiseerde op 2 april 2004 een lezingen- en discussiemiddag onder de titel ‘Bij toeval veroordeeld?’.

Mede door onbekendheid van het terrein bij journalisten, heeft de discussie tot nu toe op uiterst verwarrende wijze plaatsgevonden. De publicaties hebben veel slordigheden en onjuistheden opgeleverd en het onbegrip bij juristen en in de samenleving over de rol van de statistiek eerder vergroot dan verkleind. Meer dan ooit lijkt Disraeli’s gevleugelde uitspraak: ‘Lies, damned lies, and statistics’ een adequate weergave van het imago van de statistiek.

De zaak tegen LdB is voor de redactie van *STAtOR* aanleiding geweest het voorliggende speciale nummer uit te brengen over de toepassing van statistiek in de rechtsgang. Het doel is niet alleen om te laten zien wat de rechter aan statistiek kan hebben, maar eerst en vooral wat de rol van statistiek en statisticus in de rechtsgang is of zou moeten zijn. De inhoudelijke meningsverschillen tussen statistici sluiten nauw aan bij de verschillende opvattingen over hoe zij hun rol moeten definiëren. Door de hoofdrolspelers in de zaak LdB en andere deskundigen aan het woord te laten, hoopt de redactie een dienst te bewijzen aan allen die met de toepassing van statistiek in de rechtszaal te maken krijgen.

In de formule van *STAtOR* staan duidelijkheid en leesbaarheid voorop. Het betekent dat dit themanummer zich niet alleen tot vakgenoten statistici richt maar evenzeer tot juristen en anderen die interesse hebben in het onderwerp. De auteurs zijn in de gelegenheid gesteld misvattingen en onjuistheden recht te zetten die eerder in de pers tot verwarring aanleiding hebben gegeven.

Auteurs zijn allereerst de statistisch deskundigen opgeroepen in de zaak LdB: Henk Elffers op voordracht van de aanklager en Ronald Meester op voordracht van de verdediging. Vervolgens vergelijkt Marjan Sjerps, als statisticus verbonden aan het Nederlands Forensisch Instituut, de klassieke met de Bayesiaanse benadering en gaat zij in op andere toepassingen van statistiek in het strafrecht. De redactie is ten slotte ingenomen met het feit dat Willem van Zwet, emeritus hoogleraar mathematische statistiek in Leiden en erelid van de VVS, bereid is gebleken de discussie in perspectief te plaatsen en conclusies te formuleren.

Han Oud. redacteur STAtOR

'Geacht hof, het was geen toeval. De rest is aan u.'

Over de bescheiden rol van de statisticus als deskundige in het Nederlandse strafproces, geïllustreerd aan de hand van een dienstroosterzaak.

Henk Elffers

Een strafrechter[1] moet in het Nederlandse strafproces zich beraden of hij datgene wat door het openbaar ministerie (OM) bij monde van de officier van justitie (OvJ, 'de aanklager') aan de verdachte is ten laste gelegd *wettig* en *overtuigend* bewezen acht[2]. Daartoe beschouwt hij de op de openbare strafzitting aangevoerde *wettige bewijsmiddelen*. De criteria voor *wettigheid* van een bewijsmiddel zijn geformuleerd in het wetboek van strafvordering, het wetboek dat de hele gang van zaken in een strafzaak regelt. Die criteria zijn trouwens ruim: wettige bewijsmiddelen zijn *eigen waarneming van de rechter, verklaring van de verdachte, verklaring van een getuige, verklaring van een deskundige, schriftelijke bescheiden*. Of de rechter de aangedragen wettige bewijsmiddelen ook *overtuigend* vindt, staat in principe aan hem en aan hem alleen ter beoordeling: daarvoor zijn geen precieze wettelijke voorschriften vastgelegd. Het is de taak van de rechter om zich daarover te beraden. Hoe de rechter die overtuiging bekomt of niet, dat wordt geheel en al aan hemzelf overgelaten.

In veruit de meeste strafzaken is het bewijs niet moeilijk: de verdachte is wellicht op heterdaad betrapt, hij bekent, er zijn voldoende getuigen, etc. In zulke gevallen worden er in de rechtszaal betrekkelijk weinig woorden vuilgemaakt aan het bewijs, en concentreert de behandeling zich vaak op wat de officier van justitie en de raadsman van de verdachte rond de gepaste straf naar voren brengen. Maar er zijn ook zaken waar het niet zonder meer zonneklaar is dat de verdachte het ten laste gelegde heeft gedaan. Dan concentreert de behandeling zich juist wel op bewijsmiddelen, en ziet men vaak door OvJ, raadsman of rechter de hulp ingeroepen van een deskundige. De rol van de deskundige is wederom in het wetboek van strafvordering geregeld. Art 343 SV stelt: *Onder verklaring van een deskundige wordt verstaan zijn bij het onderzoek op de terechtzitting medegedeeld gevoelen betreffende hetgeen zijne wetenschap hem leert omtrent datgene wat aan zijn oordeel onderworpen is.* Over deskundigen zegt Art 227 SV dat dat personen zijn [...] *die tot taak hebben hem [de rechter] bij te staan, alsmede, zonodig, een onderzoek in te stellen en daarover een schriftelijk verslag uit te brengen. Bij de benoeming worden vermeld de opdracht die moet worden vervuld* [...]

Een deskundige neemt niet de taak van de rechter over. Het is niet van belang of de deskundige overtuigd raakt van de schuld of onschuld van de verdachte, noch ook of hij vindt dat de rechter daar, na zijn verklaring, van overtuigd hoort te zijn. De taak van de deskundige is om antwoord te geven op de vragen die hem worden voorgelegd. Zijn verklaringen vormen bewijsmiddelen, en het is dan weer aan de rechter om te beoordelen op grond van deze verklaringen en andere bewijsmiddelen of hij zich overtuigd weet. OvJ en verdediging zullen niet nalaten onder des rechters aandacht te brengen wat zij vinden dat de verklaring kan bijdragen tot, dan wel afdoen aan, dat overtuigend bewijs. Maar de rechter beslist.

Tot de vaak geraadpleegde deskundigen behoren mensen van de dienst Nationale Recherche Informatie (NRI) en onderzoekers van het Nederlands Forensisch Instituut (NFI). Zij rapporteren bijvoorbeeld over vingerafdrukken, schoensporen, DNA-profielen, reconstructie van hoe hard een verongelukte auto moet hebben gereden uit de botsingsenergie, chemische analyses, spraakherkenning. Ook worden vaak forensisch psychiaters en psychologen geraadpleegd, maar hun rol is vaak niet zozeer om bewijsmiddelen aan te dragen, maar meer hun opinie te geven of verdachte toerekeningsvatbaar is, behandelbaar is, een gevaar voor zichzelf of de maatschappij is.

Dienstroosterdata

Een enkele keer ziet men ook wel eens een statisticus als deskundige in de Nederlandse strafrechtszaal. Recent is in Nederland een tweetal zogeheten dienstroosterzaken behandeld waarbij statistici als deskundigen zijn opgetreden[3]. Dienstroosterzaken betreffen gevallen van mensen die het veroorzaken van allerlei incidenten ten laste wordt gelegd, nadat zij in de ogen van de OvJ wel opvallend vaak dienst hadden wanneer zich een incident voordeed. Wanneer er geen doorslaggevend direct bewijs is dat de verdachte de hand heeft gehad in die incidenten probeert het OM dan aan te voeren dat het 'te toevallig' zou zijn dat alle of de meeste incidenten nou net plaats grijpen als verdachte dienst heeft. Overigens kan dat argument alleen nooit voldoende zijn voor een bewezenverklaring, want in het Nederlandse strafproces geldt, grof gezegd, 'één bewijs is geen bewijs': de rechter dient zich te baseren op minstens twee bewijsmiddelen. In 2001 sprak de rechtbank Haarlem een Hoofddorpse crècheleidster vrij, die door het OM mede op grond van dienstroosterdata poging tot moord op een zestal kinderen ten laste was gelegd, kinderen die in ademhalingsmoeilijkheden waren gekomen als verdachte de zorg over hen had. In hoger beroep werd dit vonnis door het Amsterdamse hof bevestigd[4]. Mede op grond van dienstroosterdata veroordeelde de rechtbank Den Haag in maart 2003 de Haagse verpleegster LdB voor vier moorden en drie pogingen tot moord (tot levenslange gevangenisstraf), terwijl de rechtbank haar vrijsprak van 11 andere ten laste gelegde (pogingen tot) moord[5]. Belangrijkste bewijsmiddel was de verklaring van enkele toxicologen, terwijl statistische argumentatie als ondersteunend bewijs werd aangevoerd. De zaak tegen LdB is dit voorjaar in hoger beroep door het Haagse hof behandeld, en heeft tot grote opwinding geleid rond de rol van statistische overwegingen die door ondergetekende als deskundige naar voren zijn gebracht. Ik zal in het onderstaande schetsen hoe mijn argumentatie in het proces tegen LdB luidde, en welke rol mijn bevindingen mijns inziens in het proces behoren te spelen[6]. Ik zal daarbij voorbijgaan aan concrete details van de strafzaak, die uiteraard een echte zaak altijd compliceren, bijvoorbeeld of alle ten laste gelegde feiten wel juridisch correct zijn omschreven qua tijd en plaats, en of de gegevens waarop de analyse is gebaseerd wel onomstotelijk vast staan. Dat is voor de concrete strafzaak uiteraard van eminent belang, voor het exposeren van de gang van het betoog echter niet. Ik presenteer een geabstraheerde vorm van het argument[7].

De geabstraheerde casus

In een bepaalde periode van 343 dagen hebben op een kinderafdeling van een bepaald ziekenhuis 8 *incidenten* plaatsgevonden, dat wil zeggen dat een kind plotseling in acuut levensgevaar verkeert en een reanimatiepoging, al of niet met succes, wordt ondernomen. Dat is erg opvallend, omdat het een afdeling betreft waar normaal gesproken geen patientjes liggen die in levensgevaar verkeren. In alle gevallen blijkt een bepaalde verpleegster, mevrouw V, dienst te hebben als het incident plaatsvindt, en zij is ook meestal degene die alarm slaat. In alle overige diensten vinden in de betreffende periode geen incidenten plaats. De ziekenhuisleiding heeft gaande de reeks incidenten al verdenking voelen opkomen dat V de hand in de incidenten heeft. Na het achtste incident is de maat vol, en licht het ziekenhuis de politie in, zeggende: 'Dat is te toevallig'. Mevrouw V verdedigt zich door te zeggen dat het wèl alleen maar een ongelukkig toeval is dat uitgerekend zij zo vaak reanimatiegevallen meemaakt. De politie doet onderzoek, achterhaalt nog meer incidenten tijdens eerdere aanstellingen van verdachte bij andere ziekenhuizen, de stoffelijke resten van de overleden kinderen worden toxicologisch onderzocht, en de OvJ besluit V te vervolgen. Eén van diverse bewijsmiddelen die het OM wil aanvoeren is dat het geen toeval kan zijn dat V zo vaak bij een incident betrokken kan zijn. Het OM vraagt daarom een statisticus: 'Kan dat toeval zijn?'

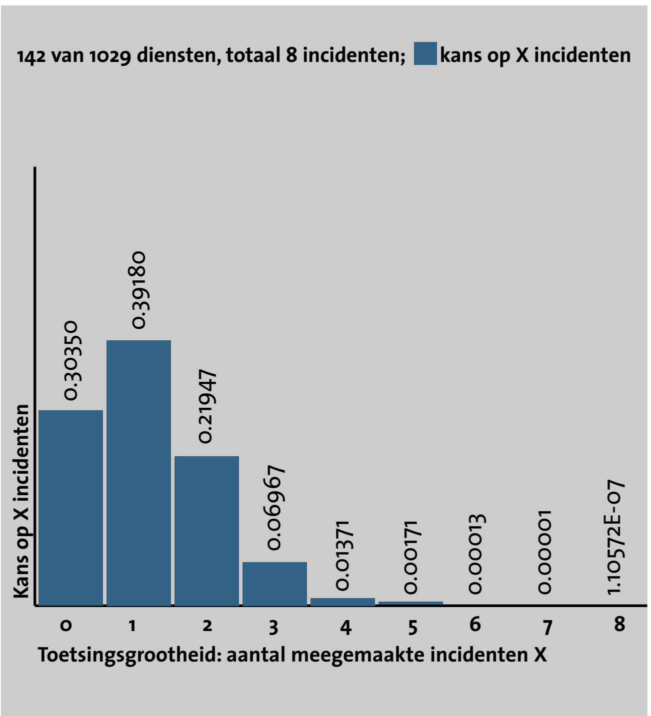
De reconstructie van de vraag 'Kan dat toeval zijn?'

Laten we eerst even stilstaan bij de aan de deskundige gestelde vraag. Die luidt niet: 'Is mevrouw V schuldig?' Dat zou ook niet juist zijn, want die kwestie is aan de rechter voorbehouden. Aan de deskundige wordt gevraagd zijn licht te laten schijnen over één stapje in het argument, dus om één bewijsmiddel aan te leveren. De eerste stap in het werk van de deskundige is natuurlijk zo'n vraag zo te formuleren dat hij binnen zijn vakgebied past en kan worden opgelost. De concrete formulering van de vraag komt in het algemeen tot stand in overleg tussen de vragensteller en de deskundige. Dat leidt uiteraard wel eens tot terminologische discussie. Niet-statistici drukken zich uiteraard niet in door ons vakgebied geheiligde termen uit, maar gebruiken formuleringen als: 'Kunt u uitrekenen wat de kans is dat het toeval was?' Mijns inziens past het niet dat wiskundigen daarover lacherig doen. Dat binnen ons vakgebied een bepaalde formulering ingang heeft gevonden, verplicht immers anderen geenszins deze formuleringen in hun dagelijks spraakgebruik over te nemen. En zo helder is ons vakgebied trouwens ook niet, zoals blijkt uit de controverse tussen Bayesianen en klassieken hoe een probleem als dit aan te pakken. Ik verwerp de herformulering van de vraag als: 'Wat is de kans dat de verdachte het heeft gedaan?' Voor mij is dat een eventualiteit waaraan geen kans kan worden toegekend. V heeft het gedaan of niet, daar zit geen kansruimte op. Anders dan voor aanhangers van de Bayesiaanse benadering is voor mij de kanstaal niet geschikt om evidentiewaarde mee te indiceren.

Mijn benadering is me af te vragen wat het eigenlijk zou betekenen als geldt: 'Het is toeval dat mevrouw V zoveel incidenten meemaakt.' Ik pak dat aan als volgt: formuleer een kansmodel dat 'toevallig voorkomen van incidenten' beschrijft en ga na of de voorliggende data rijmbaar zijn met dat kansmodel, oftewel toets de nulhypothese dat dat kansmodel geldt. Zoals wij weten is deze vraag zò gesteld binnen de klassieke statistiek niet acceptabel geformuleerd: we kunnen een nulhypothese alleen toetsen tegen een bepaalde alternatieve hypothese, zodat we ook daaraan aandacht moeten besteden.

Wat is een goed model voor 'toevallig voorkomen van incidenten'?

Ik stel allereerst voor om te *conditioneren* op de gegeven data, dat wil zeggen ik beschouw de gegeven periode van 343 dagen (dat zijn 1029 diensten), waarin 8 incidenten zijn voorgekomen. Ik formuleer toevallig voorkomen van incidenten als de toepasselijkheid van een urnmodel: of een incident voorkomt in een bepaalde dienst vergelijk ik met het trekken van een bal uit een urn met 1021 witte en 8 zwarte ballen. Een zwarte bal staat voor een incident. De nulhypothese is dat dit model geldt, en dat het voorkomen van incidenten in de diensten waarin V aanwezig was vergeleken kan worden met het trekken (zonder teruglegging) van ballen uit deze vaas, zodat er geen associatie is tussen het dienst hebben van V en het voorkomen van een incident. Ik stel de generale alternatieve hypothese dat er wel sprake is van positieve associatie tussen het dienst hebben van V en het voorkomen van incidenten, of, wat onder de gegeven condities identiek is, dat diensten waarin V aanwezig is met grotere kans dan uit dit urnmodel volgt een zwarte bal trekken. Bij dit model is onder de nulhypothese het aantal zwarte ballen in de 142 diensten die mevrouw V draaide hypergeometrisch verdeeld. De toets voor deze nulhypothese tegen dit alternatief is Fishers exacte toets, met als toetsingsgrootheid het aantal incidenten dat in diensten van V voorkomt, in ons geval 8. De toets verwerpt als de toetsingsgrootheid groter of gelijk is aan een kritieke waarde F_{kr}. De verdeling van de toetsingsgrootheid onder H_0 staat in figuur 1 afgebeeld.



Figuur 1: verdeling toetsingsgrootheid urnmodel onder de nulhypothese.

De kritieke waarde van deze toets bij een eenzijdige onbetrouwbaarheidsdrempel $\alpha = .001$ is bij de gegeven marginalen (1029 diensten, 142 diensten waar V in meedraait, totaal 8 incidenten) $F_{kr}=6$. (De overschrijdingskans van de kritieke waarde is 0.0001384, de overschrijdingskans van de waarde 8 is 0.000000110572). We verwerpen daarom de nulhypothese (V en incidenten zijn ongerelateerd) tegen het alternatief dat er wel sprake is van positieve associatie.

De vertaling naar gewone-mensen-taal van deze statistische operatie is: *het was geen toeval dat mevrouw V zoveel incidenten meemaakte*. De vertaalslag luidt immers:

Het is toeval dat V zoveel incidenten meemaakt
(gewone taal)

is equivalent aan

H_0 geldt (statistische modeltaal)

Nu blijkt:

We verwerpen H_0, dus H_0 geldt niet
(statistische modeltaal)

hetgeen equivalent is aan

Het is geen toeval dat V zoveel incidenten meemaakt
(gewone taal)

Sommige mensen hebben moeite met deze vertaalslag en vinden hem niet terecht. Binnen het statistisch model geldt immers dat het voorkomen van 8 of meer incidenten ook onder H_0 wel mogelijk is (met zeer kleine kans). In die zin tonen we juist aan dat het voorkomen van 8 incidenten juist wel toevallig kan optreden (met kans 0.000000110572). Hier is sprake van verwarring van de gewone taal en de statistische taal. De interpretatie van een zo kleine (overschrijdings-) kans (statistische taal) is immers dat we het model verwerpen (statistische taal), wat betekent dat we niet langer geloven dat mevrouw V toevallig zoveel incidenten meemaakte (gewone taal).

Het is goed op te merken dat 'Het is geen toeval dat V zoveel incidenten meemaakt' in genen dele equivalent is aan de stelling 'Mevrouw V veroorzaakt de incidenten'. Er zijn nog allerlei andere verklaringen denkbaar voor een associatie van V en incidenten dan dat zij er de hand in heeft gehad. Daarop kom ik straks terug. We moeten eerst even ingaan op het probleem van het dubbele gebruik van de data.

Verdenking vooraf

Er zit hier namelijk een addertje onder het gras. We hebben net berekend hoe onwaarschijnlijk het onder het gepresenteerde toevalsmodel is dat een verpleegster die 142 uit 1029 diensten draait 8 incidenten meemaakt. Maar de verdachte was niet 'een' verpleegster, zij was dìe verpleegster die in het oog was gelopen juist òmdat zij zoveel incidenten meemaakte. We gebruiken de data dus eigenlijk tweemaal: eerst om de verdenking op haar te laten vallen, daarna om aan te tonen dat het aantal incidenten dat zij meemaakt niet met een toevalsmodel te rijmen is. Dat is niet correct. We zouden eigenlijk moeten kijken naar de toetsingsgrootheid 'aantal incidenten dat iemand meemaakt, gegeven dat zij opvallend veel incidenten meemaakt', en achterhalen wat daarvan de verdeling onder H_0 zou zijn.

Om dat goed te doen moet je dat 'opvallen' dus kunnen modeleren. Dat is niet zo makkelijk, omdat het opvallen van de verdachte niet goed is gedocumenteerd, en zeker niet alleen is gebaseerd op het aantal incidenten dat zij meemaakte, maar ook op allerlei evidentie ten aanzien van haar gedrag rond de incidenten. Als we zouden weten dat er verdenking tegen haar rees na Y gevallen, en dat dat enkel en alleen op dat aantal was gebaseerd, zouden we natuurlijk de conditionele kansverdeling onder H_0 van

P(V maakt minstens X gevallen mee | zij maakt minstens Y gevallen mee)

kunnen berekenen. Het is evenwel uit het dossier niet duidelijk wat dan de waarde van Y zou zijn, en in ieder geval is de verdenking niet bloot op het aantal gebaseerd geweest. In het extreme geval dat iemand pas opvalt *omdat* zij Y=9 incidenten meemaakt, geldt uiteraard dat die conditionele kans 1 is, en we H_0 zeker niet zouden verwerpen.

Omdat de benodigde gegevens over het rijzen van verdenking niet beschikbaar zijn, stel ik voor om dit probleem op een Bonferroni-achtige manier aan te pakken. De afdeling waar mevrouw V werkt heeft een staf van 27 verpleegsters om te draaien. Als er inderdaad een toevalsproces zoals boven gemodelleerd de incidenten veroorzaakt, is er een kans p dat een bepaalde verpleegster 8 of meer incidenten meemaakt. Wat is dan de kans dat één of meer van 27 verpleegsters dat onder toeval zou overkomen, als die 27 verpleegsters allemaal evenveel diensten als V zouden draaien? Ik schat die kans af met een Bonferroni-methode[8]: de kans dat tenminste iemand dat meemaakt onder H_0 is gelijk aan $1-(1-p)^{27}$ of (conservatief) $27p$, hetgeen in ons geval 0.0000029854 is. Ik noem dat de post-hoc-correctie, en zie het als een benaderende manier om de overschrijdingskans van de toets voor H_0 te corrigeren voor het vooraf rijzen van verdenking. We zien dat deze post-hoc-correctie voor de invloed van gerezen verdenking in dit geval de conclusie niet verandert: we verwerpen H_0 onverminderd.

In de concrete strafzaak tegen de verdachte is bovenstaande analyse aangevuld met een analyse op het voorkomen van incidenten terwijl verdachte dienst had in twee andere ziekenhuizen. In de feitelijke analyse zijn de overschrijdingskansen onder H_0 van de combinatie van die drie datasets gecombineerd, hetgeen resulteerde in een kans van 0.00000000292, in pers steeds aangehaald als '1 op de 342 miljoen'. Inderdaad werd dat in de krant dan wel eens verwoord als: 'de kans dat het toeval was, is 1 op de 342 miljoen'. Ik acht dat niet zo'n gelukkige formulering, maar materieel komt het op hetzelfde neer: we sluiten uit dat er sprake is van toeval.

Bij de behandeling in hoger beroep werd door Ronald Meester, eveneens optredend als deskundige, gesteld dat er geen enkele reden is om de post-hoc-correctie te beperken tot de betreffende afdeling van het ziekenhuis. Hij suggereerde dat je dan ook wel kunt corrigeren voor alle verpleegsters in het ziekenhuizen, alle ziekenhuizen, of voor alle mensen in het land, of wat niet al. Deze tegenwerping komt mij onbegrijpelijk voor. Aangezien de hele analyse *conditioneel* op de gebeurtenissen op de afdeling in een bepaalde periode werkt (*gegeven* die diensten, *gegeven* die incidenten, *gegeven* die bestaffing) is het enige risico waarvoor we willen corrigeren met deze methodiek het risico van de tevoren gerezen verdenking dat van toepassing is op de mensen die op die afdeling werken. Mensen buiten die afdeling hebben immers geen risico gelopen bij toeval tegen juist deze incidenten op te lopen?

Geen toeval, dus ...

Ik heb nu laten zien dat geen sprake kan zijn van toeval, in de zin dat het door mij voorgestelde toevalsmodel niet rijmbaar is met de data. Voorkomen van incidenten en dienstdoen van mevrouw V zijn geassocieerd. Dat impliceert op zich zelf helemaal niet dat mevrouw V de incidenten zou hebben *veroorzaakt*. Andere associaties tussen V's aanwezigheid en het voorkomen van overlijdens zijn uiteraard ook denkbaar. Ik noem, voor de vuist weg, zes hypothesen die men zou kunnen opperen om de aangetoonde associatie van V met het voorkomen van incidenten aannemelijk te maken.

- De aanname dat alle diensten gelijke kans op een incident lopen is onjuist. Bijvoorbeeld: overlijdensrisico is hoger in de vroege dagdienst; mocht het zo zijn dat mevrouw V dan vaker dienst zou hebben gedaan, dan kan dat een associatie opleveren.
- V werkt bij voorkeur samen met W, en W veroorzaakt de incidenten.
- V heeft vaak nachtdienst, en dan is de kans dat een levensbedreigend incident tijdig wordt gesignaleerd kleiner, omdat verpleegsters dan meer patiënten in hun eentje onder zich hebben.
- V is een slechte verpleegster, waardoor ze niet tijdig kritieke situaties onderkent.
- V neemt altijd de moeilijkste gevallen – met groter overlijdensrisico – op zich.
- Iemand heeft de pik op V en probeert haar in diskrediet te brengen.

Of men deze hypotheses als geloofwaardige verklaring voor het meer dan toevallig betrokken zijn van mevrouw V bij incidenten kan kenschetsen, zal met andere middelen dan de hierboven geanalyseerde cijfers moeten worden nagegaan. Sommige hypotheses kunnen ook door nauwkeurige analyses van dienstroostergegevens worden onderzocht, andere zijn niet op basis van dienstroostergegevens op hun merites te beoordelen. Deze dienen op grond van andere overwegingen nader te worden beschouwd. Uiteraard zijn er in het strafdossier allerlei aanwijzingen te vinden die het voor de rechter mogelijk maken zich een oordeel te vormen over deze of andere suggesties: of de verdachte al of niet een goed verpleegster is, of het mogelijk is risicovolle gevallen tevoren te herkennen, of andere verpleegsters vaak samenwerken met de verdachte, daarover kan de rechter veel vinden in het dossier. Uiteraard zou hij desgewenst nadere vragen kunnen stellen aan de recherche, of aan statistische of andere deskundigen.

Van belang is in te zien dat het zich een oordeel vormen over de plausibiliteit van deze verklaringen dan wel de verklaring dat mevrouw V de hand in de incidenten heeft gehad aan de rechter is. Als statisticus heb ik de vraag 'Kan dat toeval zijn?' beantwoord met een ongeclausuleerd 'Nee'. Nu is het aan de rechter om de waarde van dit middel te bepalen bij zijn al of niet komen tot de overtuiging dat de verdachte het ten laste gelegde heeft begaan. Dat statistisch bewijsmiddel is daarbij van beperkte betekenis: een bepaalde mogelijkheid, toeval, is uitgesloten. Het woord is verder aan de rechter. Er is in deze eigenlijk geen fundamenteel verschil met hoe een rechter andere dan statistische bewijsmiddelen beoordeelt. Ook als, bijvoorbeeld, een getuige verklaart een verdachte op een bepaalde tijd en plaats te hebben gezien, overweegt de rechter of alternatieve verklaringen (waarnemingsfouten, herinneringsproblemen, bewust valse verklaringen, etc.) onhoudbaar zijn. Dat de door een deskundige aangedragen argumentatie 'relatief' is, is in genen dele verrassend of bijzonder. Dat geldt voor deskundigenverklaringen precies zo als voor andere elementen in de bewijsvoering van het OM en in de verdediging van de raadsman. Dat is precies de reden waarom de oordeelsvorming aan de rechter is opgedragen, en niet aan een automaat.

Noten

1. Ik spreek steeds over 'de' strafrechter, alhoewel het vaak gaat om een college van drie rechters, en over 'hij', hoewel er zeer veel vrouwelijke rechters zijn. Voor het gemak denken we aan een proces 'in eerste aanleg' voor de rechtbank. Ik ga ook voorbij aan allerlei terminologische details, zoals dat rechters in het hof (waar hoger beroep wordt behandeld) raadsheren heten, etc.
2. Ik ga hier nu niet in op andere overwegingen die de rechter moet maken, onder andere nagaan of het ten laste gelegde, indien bewezen, strafbaar is, en de straftoemeting.
3. Een overzicht van dienstroosterzaken in Angelsaksische landen geven Lucy, D., & Aitken, C. (2002). A review of roster data and evidence of attendance in cases of suspected deaths in a medical context. *Law, Probability and Risk*, *1*, 141-160.
4. Uitspraak te raadplegen op www.rechtspraak.nl, LJN-nummer: AD7198 (rechtbank) en AL7736 (hof)
5. LJN-nummer: AF6172.
6. Ten tijde van het schrijven van dit artikel heeft het hof nog geen uitspraak gedaan.
7. Eerder beschreef ik een nog meer gefictionaliseerde vorm in Elffers, H. (2003), Bij toeval veroordeeld? Statistische analyse van dienstroosterdata in het strafproces *Nederlands JuristenBlad* 78/34,1812-1814.
8. Als we de feitelijke aantallen diensten van alle verpleegsters in rekening zouden nemen, inclusief wie met wie gezamenlijk dienstdeed, zou dit nauwkeuriger kunnen, maar gezien de extreem kleine kansen waar we mee rekenen zal dat weinig verschil maken. Lucy en Aitken hebben getracht dit probleem aan te pakken door de verdeling van de grootste order statistic van de relatieve incidentendichtheid van alle stafleden op een afdeling te berekenen (D. Lucy & C. Aitken. *The evidential value of roster and attendance data in cases where fraud and medical malpractice may be suspected*, submitted).

HENK ELFFERS studeerde wiskundige statistiek aan de UvA en is thans senior-onderzoeker 'Mobiliteit en Spreiding van Criminaliteit' aan het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving NSCR, Leiden, en hoogleraar Rechtspsychologie aan de Universiteit Antwerpen. E-mail: <elffers@nscr.nl>.

Voorzichtig met statistiek in de rechtszaal!

Aan de hand van de recente casus LdB bespreek ik een aantal problemen die samenhangen met het gebruik van statistiek in de rechtszaal. Ik zal uitleggen wat er naar mijn mening mis is gegaan in de genoemde casus. De genoemde problemen spelen niet alleen bij deze casus, maar zijn altijd belangrijk wanneer het gaat om het gebruik van statistiek bij rechtszaken.

Ronald Meester

In de rechtszaken tegen de van meervoudige moord verdachte verpleegkundige LdB speelde de statistiek een belangrijke rol. De verdachte had tijdens haar diensten als verpleegkundige te maken gekregen met een opvallend groot aantal incidenten[1], en de vraag leek gerechtvaardigd of een dergelijk groot aantal incidenten eigenlijk wel in alle redelijkheid 'toevallig' kan plaatsvinden tijdens de diensten van één enkele verpleegkundige. Bij het proces in eerste aanleg en het daarop volgende hoger beroep hebben de rapporten van prof. dr. Henk Elffers een belangrijke rol gespeeld. In zijn rapportage kwam Elffers tot de conclusie dat wanneer we aannemen dat er geen samenhang bestaat tussen incidenten en verantwoordelijke verpleegkundigen, dat dan de kans dat de verdachte minstens het aantal incidenten meemaakt als feitelijk gebeurd is, kleiner is dan 1 op de 342 miljoen. Deze kans is zo extreem klein, aldus Elffers, dat we toeval met een gerust hart kunnen uitsluiten.

Ik denk echter dat dit getal weinig betekenis heeft en zal uitleggen waarom. De problemen die ik signaleer hebben niet alleen betrekking op deze zaak en zijn derhalve van belang in een veel ruimere context.

De methode van Elffers

Ik beperk me tot één van de ziekenhuizen waar de verdachte werkte, het Juliana kinderziekenhuis (JKZ). Binnen het tijdsinterval van de statistische analyse, werden er op de betreffende afdeling van het JKZ in totaal 1029 diensten gedraaid, waarvan de verdachte er 142 voor haar rekening nam. Tijdens die 1029 diensten vonden er 8 zogenoemde incidenten plaats, die alle 8 tijdens een dienst van de verdachte plaatsvonden. Op de afdeling werkten er in die tijd 27 verpleegkundigen.

Elffers redeneert nu als volgt: we weten dat er tijdens de 1029 diensten 8 incidenten zijn geweest. Als de kans op een incident voor iedere verpleegkundige altijd hetzelfde is, en incidenten in verschillende diensten niets met elkaar te maken hebben, dan kunnen we die 8 incidenten beschouwen als een willekeurige greep van 8 diensten uit het totaal van 1029 diensten. Wanneer we 8 diensten toevallig trekken uit het totaal van 1029 diensten, dan kunnen we eenvoudig uitrekenen wat de kans is dat alle 8 getrokken diensten een dienst van de verdachte waren. Deze kans blijkt ongeveer 1 op de 9 miljoen te zijn.

Er zit hier echter een addertje onder het gras, aldus Elffers. Wat hier in feite uitgerekend wordt, is de kans dat een *van tevoren aangewezen* verpleegkundige (namelijk de verdachte) die 142 van de 1029 diensten voor haar rekening neemt, alle 8 incidenten meemaakt. Maar dat is niet helemaal fair; ze is alleen maar aan deze berekening onderworpen omdat ze was opgevallen vanwege het frequent voorkomen van incidenten tijdens haar diensten. Elffers corrigeert hiervoor met de zogenaamde *post hoc correctie*, en komt zo tot een kans van 1 op de 300.000.[2]

Het getal van 1 op de 342 miljoen wordt verkregen door een zelfde analyse te hanteren op de andere afdelingen waar verdachte werkte, en de uitkomsten met elkaar te vermenigvuldigen.

Er is naar mijn mening heel wat mis met deze aanpak. Hieronder bespreek ik een aantal punten van kritiek.

Het schaalprobleem

Stel er wordt een loterij uitgeschreven met loten genummerd 1 tot en met 1.000.000. De hoofdprijs blijkt op nummer 223.478 te vallen, en dat lot blijkt gekocht door een inwoner van de DaCostastraat in Leiden. Gegeven deze feiten kunnen we nu gaan kijken naar de kans dat *iemand* in deze straat de hoofdprijs wint. We kunnen een simpel en oncontroversieel model maken waaruit blijkt dat de kans dat iemand in de DaCostastraat in Leiden de hoofdprijs wint buitengewoon klein is. Moeten we hieruit concluderen dat de loterij niet eerlijk was? Immers, er is iets gebeurd dat bij een eerlijke loterij een heel kleine kans had. Nee, natuurlijk kunnen we dat niet concluderen. Punt is dat de aandacht op de DaCostastraat in Leiden pas gevestigd werd, *nadat* de trekking geweest was. We hebben pas een model voor de DaCostastraat in Leiden gemaakt, nadat we gezien hadden dat zich daar iets opvallends afgespeeld had.

Dit fenomeen is berucht in de statistiek. Als men iets opvallends waarneemt dan is men geneigd om het model, en bijbehorende hypotheses, aan de data aan te passen; in zulke gevallen wordt de hypothese dus opgesteld *nadat* we de data hebben gezien. Het probleem van deze *data-afhankelijke hypothesen* is dat men eerst op grond van de data een idee krijgt voor een hypothese, die men dan vervolgens door diezelfde data wil verifiëren. Het is intuïtief duidelijk dat hier iets mis gaat, maar binnen de statistiek bestaat er geen volledig bevredigende oplossing voor dit probleem. Kern van de zaak is dat je er rekening mee moet houden dat het van tevoren uitermate onwaarschijnlijk was dat de DaCostastraat in Leiden überhaupt in beeld zou komen. Als we de kans gaan uitrekenen dat iemand in Leiden, of iemand in Zuid-Holland de hoofdprijs wint, dan komen daar geheel verschillende getallen uit, en het is belangrijk je te realiseren dat er niet een uniek of voor de hand liggend niveau is waarop je het model moet bekijken. Elk getal is dus relatief aan de keuze voor de *schaal* van het model.

De (enigszins wrange) analogie met de zaak LdB is hopelijk duidelijk: Elffers beperkt zijn model tot de afdeling waar zich iets opvallends heeft afgespeeld, maar hij vergeet dan dat het feit dat het juist deze afdeling betreft, van tevoren uitermate onwaarschijnlijk was. Aan het voorbeeld van de loterij kunnen we zien dat dit helemaal niet bijzonder is; gebeurtenissen met hele kleine kans vinden altijd en overal plaats.

Wat verwerpt Elffers eigenlijk?

De oorspronkelijke vraag aan Elffers was de vraag of het waargenomen aantal incidenten toeval zou kunnen zijn. Nu is het begrip 'toeval' niet eenduidig; het kan op vele manieren geïnterpreteerd worden[3]. Om er als wiskundige naar te kunnen kijken heeft Elffers de vraag of het toeval kan zijn op zijn eigen manier *vertaald* naar een wiskundige context. Hij kan ook niet anders, maar het is belangrijk om te begrijpen dat die vertaling allerlei individuele keuzes en subjectieve afwegingen met zich meebrengt. Laten we er een noemen. Elffers vertaalt toeval met de wiskundige veronderstelling dat elke verpleegkundige tijdens een dienst altijd dezelfde kans op een incident heeft. Maar dat is helemaal niet vanzelfsprekend. Men zou ook kunnen poneren dat toeval betekent dat de kans op een incident gegeven dat verdachte aanwezig is, in alle omstandigheden gelijk is aan de kans gegeven dat ze niet aanwezig is. Déze wiskundige formulering van toeval laat bijvoorbeeld de mogelijkheid open dat er verschil is tussen dag- en nachtdiensten, of de mogelijkheid dat verdachte zelf de ernstig zieke patiënten op zou zoeken. In het laatste geval is de aanname van Elffers onjuist, terwijl er gewoon nog sprake van toeval kan zijn, want ernstig zieke patiënten overlijden *altijd* vaker, wellicht ongeacht de aanwezigheid van de verdachte.

Wanneer Elffers dit model beschouwt en tot de conclusie komt dat er geen sprake kan zijn van toeval, dan verwerpt hij niet alleen de hypothese dat incidenten altijd met dezelfde kans plaatsvinden, onafhankelijk van de aan- of afwezigheid van de verdachte. Nee, hij zou daarmee tot de conclusie moeten komen dat het *totale model* de realiteit blijkbaar niet goed beschrijft. Maar als men nu op grond van het optreden van een onwaarschijnlijke gebeurtenis het model verwerpt, dan verwerpt men het *gehele* model. Echter, welk aspect van het model nu precies onjuist is, blijft dan in het midden. In zijn NJB artikel doet Elffers na alle berekeningen de volgende uitspraak: 'Toeval als verklaring valt niet langer serieus te nemen[4].'

Dat nu is nogal voorbarig. Elffers kan zijn model als geheel verwerpen, maar toeval als verklaring wordt daar niet door uitgesloten: zoals we gezien hebben kan toeval op een hele andere manier wiskundig vertaald worden, en over die (veel ruimere) vertaling zwijgt de methode van Elffers. Bovendien is het zeer discutabel of het optreden van een gebeurtenis met een hele kleine kans toeval uit kan sluiten; denk aan het voorbeeld van de loterij.

Alternatieve hypotheses

Een verwant principieel probleem met de modelkeuze is de onmogelijkheid om daarbinnen alternatieve hypotheses te formuleren. Wanneer we een alternatieve hypothese zouden willen formuleren, dan willen we bijvoorbeeld de kans op een incident voor de verdachte (zeg p) anders inschatten dan voor de overige verpleegkundigen (zeg q). Maar als we uitgaan van de hypothese dat die twee kansen verschillend en onbekend zijn, dan kunnen we de hele methode van Elffers niet meer gebruiken, omdat de kansen die hij wil berekenen niet meer uit te rekenen zijn zonder extra kennis over de kansen p en q. Er is een invloedrijke school in de statistiek die deze procedure volstrekt onacceptabel vindt. Wanneer men de hypothese van toeval wil verwerpen ten faveure van de hypothese dat de kans op een incident tijdens een dienst van de verdachte *groter* is dan tijdens een dienst van de andere verpleegkundigen, dan moeten volgens deze school beide hypotheses in het statistische model geformuleerd en onderzocht kunnen worden. Kan dat niet, dan valt er uit het verwerpen van de hypothese van toeval *geen enkele* conclusie te trekken[5].

Conclusies

In de casus LdB zijn de gebruikte statistische berekeningen en methodes ondeugdelijk, evenals de door Elffers gegeven interpretatie. Hij onderkent het schaalprobleem niet, zijn methode laat geen alternatieve hypotheses toe in het model, en hij interpreteert de getallen onjuist. Reden genoeg om aan het getal van 1 op 342 miljoen weinig waarde toe te kennen.

Noten

1. De precieze definitie van 'incident' is hier niet van belang; het volstaat dit begrip te definiëren als 'noodzaak van reanimatie', waarbij in het midden gelaten wordt of deze succesvol geweest is.
2. Ik heb ook ernstige bezwaren tegen de manier waarop Elffers deze post hoc correctie uitvoert. Bij zijn correctie is het belangrijk dat er juist 27 verpleegsters op de afdeling werkten, maar het is evident dat het *aantal* verpleegsters waarover de resterende diensten verdeeld worden irrelevant is; slechts het aantal diensten doet er toe. Hierdoor is duidelijk dat Elffers' correctie onjuist is. Zie ook M. Van Lambalgen en R. Meester: *Wat zeggen al die getallen eigenlijk?*, te verschijnen.
3. Zie bijvoorbeeld Ronald Meester, *Het pseudoniem van God*, Ten Have 2003.
4. NJB 34, 26 september 2003, blz. 1814.
5. Er bestaat een uitgebreide literatuur over deze kwestie, de zogenoemde *significance test controversy*. (De *significance test* is de statistische toets die alleen de hypothese van toeval verwerpt, zonder een alternatief in het model op te nemen.) Zie bijvoorbeeld V. Barnett, *Comparative Statistical Inference*, Wiley 1999.

*Ronald Meester is hoogleraar Waarschijnlijkheidsrekening aan de Vrije Universiteit Amsterdam.
E-mail: <rmeester@cs.vu.nl>.*

Statistiek als bewijsmiddel in het strafrecht

‘Denkt u dat statistiek eigenlijk wel een plaats heeft in de rechtszaal?’ was één van de vragen die gesteld werd door de voorzitter van het gerechtshof in Den Haag aan de hoogleraren Ronald Meester en Michiel van Lambalgen tijdens het proces in hoger beroep tegen LdB. Het antwoord was dat statistiek en kansrekening veel kunnen bieden, maar dat men er wel heel voorzichtig mee om moet gaan, zoals besproken wordt in het artikel van Ronald Meester in dit nummer van *STATOR*.

In dit artikel zal ik betogen dat statistiek en kansrekening een essentieel onderdeel zijn van vele soorten bewijsmiddelen en hun plaats in de rechtszaal allang veroverd hebben, zij het misschien niet zo opvallend als in de zaken tegen de verpleegster LdB en de crècheleidster Bianca K. Ik zal ook ingaan op het statistische bewijs in de zaak LdB.

Marjan Sjerps

In de rechtszaak tegen LdB stond de statistiek veel in de belangstelling. Ik ben niet bij deze zaak betrokken geweest, maar heb hem wel zoveel mogelijk gevolgd, voornamelijk via de media. Daaruit kwam het beeld naar voren van wetenschappers die het totaal met elkaar oneens waren over de waarde van het statistisch bewijs. Terwijl prof. dr. Henk Elffers beweerde 'het was geen toeval dat patiënten stierven precies op de momenten dat de verdachte dienst had' (*Haagsche Courant*, 30 januari 2004), bleek dat Meester en van Lambalgen 'grote moeite hadden met de methode die hun collega had gevolgd om tot zijn uitspraken te komen' (*Haagsche Courant*, 10 maart 2004). Daarna verscheen er een paginagroot stuk in de kranten van dr. Aart de Vos, waarin werd voorgerekend 'als de statistiek het enige bewijs is, is de kans 80 procent dat ze onschuldig is' (*Trouw*, zaterdag 13 maart 2004).

Interpretatie van bewijs: de Bayesiaanse benadering

De oorzaak van de schijnbare tegenstelling tussen de uitkomsten van de klassieke methode van Elffers (2003) en de Bayesiaanse methode van de Vos (2004) is van fundamentele aard en heeft te maken met het kansbegrip dat deze stromingen gebruiken. De klassieke statistiek is gebaseerd op een frequentistische definitie van kans, of op de axioma's van Kolmogorov. Hierbij zijn kansen op bepaalde gebeurtenissen gedefinieerd, maar niet op hypothesen. De Bayesiaanse statistiek hanteert een wat ruimere en subjectieve definitie van kans, namelijk de mate van geloof, die ook op hypothesen kan worden toegepast. Waar een Bayesiaan dus uitspraken kan doen over de kans dat de verdachte de dader is, daar zal een klassieke statisticus beweren dat de verdachte de dader is of niet, maar dat een kansuitspraak hierover niet gedefinieerd is.

In het rechtssysteem, waar de vraag die de rechter moet beantwoorden uiteindelijk is of de verdachte schuldig dan wel onschuldig is, past een Bayesiaanse benadering mijns inziens beter dan een klassieke benadering (Sjerps, 2000; 2004). Een invloedrijke stroming in de zogenaamde forensische statistiek is gebaseerd op Bayesiaanse ideeën (Robertson en Vignaux, 1995; Aitken en Taroni, 2004). Hierin staat de regel van Bayes centraal, maar dan in de vorm waarbij je kan zien hoe de kansverhouding (Engels: odds) tussen de hypothese van de aanklager (H_p) en die van de verdediging (H_d) verandert door de introductie van een bewijsmiddel (E):

$$\frac{P[H_p \mid E]}{P[H_d \mid E]} = \frac{P[H_p]}{P[H_d]} \cdot \frac{P[E \mid H_p]}{P[E \mid H_d]}$$

of in woorden: *a posteriori kansverhouding* = *a priori kansverhouding · LR*.

De laatste term is de afkorting van Likelihood Ratio (ook wel bekend als aannemelijkheidsquotiënt). Volgens de genoemde stroming is het de taak van de deskundige om deze LR te rapporteren. Het inschatten van de a priori kansverhouding moet de deskundige echter overlaten aan de rechter; immers, de deskundige heeft alleen speciale kennis over het ene bewijsmiddel E. Als gevolg hiervan zou de deskundige dus ook geen uitspraken moeten doen over de a posteriori kansverhouding.

Deze opvatting over de rol van de deskundige past goed binnen de wettelijke bepalingen over deze rol. In 344 Sv lid 1 sub 4 staat dat deskundigen verslag moeten uitbrengen 'behelzende hun gevoelen betreffende hetgeen hunne wetenschap hen leert omtrent datgene wat aan hun oordeel onderworpen is', terwijl de Hoge Raad beslist heeft dat de deskundige niet vooruit mag lopen op het aan de rechter voorbehouden oordeel en de rechter ook geen beslissingen mag opdragen. Omdat de wetenschap van de deskundige hem veelal alleen iets leert over de LR, zou hij niet moeten speculeren over het aan de rechter voorbehouden oordeel over de kansverhouding van de hypothesen.

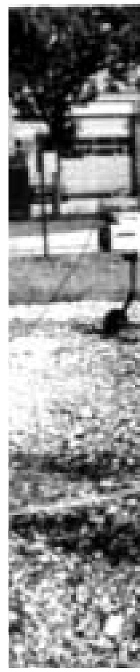
De zaak LdB

Met de bovengenoemde LR theorie in het achterhoofd beschouw ik de analyse van Elffers in de zaak LdB als het uitrekenen van de noemer van de LR. Heel losjes gezegd zijn de hypothesen die hij beschouwt: H_d: er is *geen* verband tussen de momenten dat de verdachte dienst had en het aantal incidenten, en H_p: er is *een* verband tussen de momenten dat de verdachte dienst had en het aantal incidenten. Zijn conclusie dat H_d verworpen moet worden baseert hij op het inschatten van de kans dat de verdachte bij een groot aantal incidenten betrokken is, als er geen verband is, ofwel $P[E \mid H_d]$, de noemer van de LR. De Bayesiaanse theorie maakt echter duidelijk dat dit slechts een deel van het verhaal is: de teller van de LR en de a priori kansverhouding H_p:H_d spelen ook een rol.

De teller van de LR, $P[E \mid H_p]$, is in dit geval de kans dat de verdachte bij een groot aantal incidenten betrokken is, als er wel een verband is. Zonder dit verband nader te specificeren is deze kans niet goed uit te rekenen, maar het moge duidelijk zijn dat de kans om betrokken te raken bij een groot aantal incidenten veel groter is als er wel een verband is dan als er geen verband is. De Vos (2004) schat zijn teller grofweg op 1/2.

Naast de teller van de LR speelt met name de a priori kansverhouding H_p:H_d een beslissende rol in het bepalen van de uiteindelijke kans dat H_d waar is. De Vos (2004) illustreert dat laatste heel goed met rekenvoorbeelden die laten zien dat bij een extreem kleine a priori kansverhouding H_p:H_d ook de a posteriori kansverhouding H_p:H_d klein zal zijn. Zonder dergelijke rekenvoorbeelden zullen veel mensen dit effect van a priori kansen niet beseffen bij het vernemen van de "klassieke" conclusie dat de hypothese H_d moet worden verworpen met een bepaalde p-waarde. Wanneer deze conclusie dan ook nog 'vertaald' wordt in de rechtszaal als 'het was geen toeval' (een stap die ook een klassiek statisticus niet kan maken: de statistische prietpraat is er tenslotte niet voor niks) dan wordt het voor een leek wel heel moeilijk om de rol van de a priori kansen op waarde te schatten. De Vos gaat echter op zijn beurt weer te ver naar mijn smaak door de suggestie te wekken dat het aan de deskundige is om deze a priori kansen in te schatten. Hij gaat hiermee pontificaal op de stoel van de rechter zitten, zoals boven uitgelegd.

Mijn bovengenoemde definitie van het bewijsmateriaal E is eigenlijk wat te slordig. Bij het klassieke hypothese toetsen wordt immers de p-waarde berekend: in dit geval de kans dat LdB betrokken was bij *tenminste* een x-tal incidenten. Bovendien moet er nog een post-hoc correctie worden toegepast voor het selectie-effect: LdB staat in de belangstelling juist omdat zij betrokken is geraakt bij een x-tal incidenten. Beide manoeuvres zijn enigszins arbitrair, zoals al eerder door de Vos (p-waarde; 2004) en van Lambalgen en Meester (post-hoc correctie; 2004) is bekritiseerd. Een mooi aspect van de



Monstername bij milieuonderzoek in het kader van een rechtszaak: statistiek speelt hierin een grote rol.

Bayesiaanse benadering is dat dergelijke manoeuvres overbodig zijn. Het bewijsmateriaal E is dat LdB betrokken was bij een x-tal incidenten, en de LR kan gewoon worden uitgerekend zonder dit te vervangen door *tenminste* een x-tal. Wanneer wij voor het selectie-effect zouden corrigeren door de hypothese H_p te vervangen door bijvoorbeeld H'_p: er is een verband tussen de momenten dat *één van de 27 verpleegsters op de afdeling* dienst had en het aantal incidenten, dan zou de teller van de LR ongeveer een factor 27 kleiner worden, maar de a priori kansverhouding H_p:H_d ongeveer een factor 27 groter. De lagere bewijswaarde als we H'_p kiezen in plaats van H_p wordt dus weer gecompenseerd door een grotere a priori kansverhouding, zodat de a posteriori kansverhouding hetzelfde blijft. Ook wanneer we niet op de afdeling willen focussen maar op een nog grotere groep zal er een soortgelijk compensatie-effect optreden. De a priori kansverhouding corrigeert dus op een heel natuurlijke wijze voor het selectie-effect.

Dat er grote verschillen zijn tussen de 'klassieke' benadering en de 'Bayesiaanse' zal inmiddels wel duidelijk zijn. Toch is hun boodschap uiteindelijk niet zo heel erg verschillend: waar de klassieke stroming H_d verwerpt ten gunste van H_p, komt de Bayesiaanse stroming uit op een grote LR. Beiden concluderen dus min of meer dat het bewijsmateriaal veel waarschijnlijker is als H_p waar is dan als H_d waar is, en in die zin een sterke aanwijzing vormt voor H_p ten opzichte van H_d. Deze sterke aanwijzing moet nu door de rechter gecombineerd worden met de a priori kansverhouding tussen H_p en H_d.

Een parallel met DNA

Een parallel met DNA bewijs is misschien op dit punt verhelderend. Stel dat bij een misdrijf een bloedvlek wordt gevonden die van de dader afkomstig is, en dat hiervan een DNA profiel wordt gemaakt. De analyse lukt slechts gedeeltelijk vanwege de slechte kwaliteit van de bloedvlek, en het resultaat is een zogeheten partieel DNA profiel. Stel verder dat de kans dat een willekeurige persoon dit profiel heeft 1 op 100.000 is, en dat de verdachte hetzelfde DNA profiel bezit. Deze laatste waarneming is, net als bij LdB, extreem onwaarschijnlijk onder de nulhypothese dat de bloedvlek niet van de verdachte is. Het zal echter ook onmiddellijk duidelijk zijn dat dit niet vertaald kan worden als 'de nulhypothese is onjuist'. De kans dat de bloedvlek al dan niet afkomstig is van de verdachte hangt namelijk af van de a priori kansverhouding. Wanneer je hier aannamen over doet, zoals de Vos deed in de zaak LdB, bijvoorbeeld dat er behalve de verdachte nog een half miljoen potentiële daders zijn die allen even waarschijnlijk als de verdachte de bloedvlek achterlieten, dan kun je berekenen wat de kans is dat de bloedvlek van de verdachte afkomstig is. De DNA deskundige zal zich hier echter niet aan wagen, en dat is maar goed ook, want het doen van dergelijke aannamen is niet zijn expertise. Hij zal zich beperken tot het rapporteren van de bovengenoemde match kans 1 op 100.000, een getal dat rechtstreeks is gerelateerd aan de LR (de LR van het bewijsmateriaal 'bloedvlek en verdachte hebben DNA profiel X' voor de hypothesen 'de bloedvlek is afkomstig van de verdachte' versus 'de bloedvlek is afkomstig van een willekeurige niet-verwante persoon' is gelijk aan 1/match-kans van profiel X, als we labfouten e.d. negeren).

Andere toepassingen van statistiek in het strafrecht

In de zaken Bianca K. en LdB stond de statistiek in de schijnwerpers. Maar ook in andere zaken speelt de statistiek een belangrijke rol, zij het minder opvallend. Gaat het om technisch bewijs, dan zijn statistiek en kansrekening vaak gebruikt om de conclusie te onderbouwen. Het DNA profiel ontleent zijn grote bewijskracht aan de extreem lage match-kans, het milieu-onderzoek drijft volledig op representatieve monstername en adequate analyse van de data. In feite zijn er voorbeelden te over waarbij statistiek en kansrekening een rol spelen in de vergaring of waardering van het strafrechtelijk bewijsmateriaal: steekproeven in drugszaken, automatische sprekerherkenningstechnieken, snelheidsbepalingen van auto's, de Oslo confrontatie (de *line-up* van verdachten), de geuridentificatieproef door honden, kansspelen, epidemiologische aspecten van de toxicologie. Statistiek en kansrekening zijn daarmee mijns inziens al heel lang geaccepteerd in de rechtszaal. Heel stiekem denk ik dat juristen eigenlijk ook deel uitmaken van onze beroepsgroep: het inschatten en combineren van waarschijnlijkheden op basis van vaak wankele informatie om tot een beslissing te kunnen komen vereist beslist enig gevoel voor ons vak.

Samenvatting

Statistiek en kansrekening zijn al heel lang een geaccepteerd en een essentieel onderdeel van verschillende soorten bewijsmiddelen in het strafrecht. De zaak LdB leverde veel publiciteit, waarbij het leek of de deskundigen tegenstrijdige conclusies trokken. Mijn mening is dat de verschillende stromingen in essentie de boodschap brengen dat het grote aantal incidenten waarbij LdB betrokken was een sterke aanwijzing is dat er een verband is tussen deze incidenten en de momenten waarop zij dienst had. Deze aanwijzing moet door de rechter echter gecombineerd worden met de a priori kans op een dergelijk verband, die volgt uit het overige bewijsmateriaal.

Literatuur

Aitken, C.G.G., and Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*, 2nd ed. Chichester: Wiley.

Elffers, H. (2003). Bij toeval veroordeeld? *Nederlands JuristenBlad* 34, 1812-1814.

Lambalgen, M. van, en Meester, R. (te verschijnen in 2004). Wat zeggen al die getallen eigenlijk?- de statistiek rond het proces tegen Lucy de B., *Nederlands JuristenBlad*.

Robertson, B., and Vignaux, G.A. (1995). *Interpreting evidence*. Chichester: Wiley.

Sjerps, M. (2000). Pros and cons of Bayesian reasoning in forensic science. In J.F. Nijboer and W.J.J. M. Sprangers (Eds), *Harmonization in Forensic Expertise, Series Criminal Sciences*. Amsterdam: Thela Thesis, pp. 557-585.

Sjerps, M. (2004). Forensische statistiek. *Nieuw Archief voor Wiskunde*, 5 (5), 4-9.

Vos, A. F. de (2004). Door statistici veroordeeld? *Nederlands JuristenBlad*, 13, 686-688.

Marjan Sjerps studeerde wiskunde aan de Katholieke Universiteit Nijmegen en promoveerde op een onderwerp in de evolutionaire ecologie aan de Universiteit Leiden. Sinds 1993 werkt zij als statisticus op het Nederlands Forensisch Instituut.
E-mail: <m.sjerps@nfi.minjus.nl>.

Niets nieuws onder de zon

Willem van Zwet

Vooraf

De redactie van *STAtOR* heeft mij gevraagd conclusies te trekken uit de discussie rond het strafproces tegen LdB. Ik zou willen beginnen deze casus in een wat bredere historische context te plaatsen. De discussie is namelijk klassiek in die zin dat er geen enkel element naar voren is gebracht dat niet al een halve eeuw geleden tamelijk afdoende is besproken. Er is op dit gebied weinig nieuws onder de zon en het lijkt mij nuttig dat wij ons dat realiseren.

We kunnen teruggaan naar de periode direct na de tweede wereldoorlog, toen David van Dantzig net hoogleraar in de kansrekening en de mathematische statistiek was geworden aan de Universiteit van Amsterdam en samen met enkele anderen het Mathematisch Centrum (MC) had opgericht. Van Dantzig had een sterke belangstelling voor de grondslagen van zijn nieuwe vakgebied. Op college sprak hij over de principiële onmogelijkheid om met behulp van de kansrekening iets over de wereld om ons heen te zeggen. Wij trachten dat toch te doen door een wiskundig model aan te nemen en daarbinnen wiskunde te bedrijven. Het probleem zit echter in het in- en uitschakelen van het formalisme: hoe kies ik mijn model en hoe kom ik er weer van af, dat wil zeggen, hoe vertaal ik mijn wiskundige resultaten terug naar praktische conclusies?

Voor Van Dantzig was dit geen theoretische beschouwing, maar een uiterst concreet probleem. Het MC was immers opgericht om toepassingen van de wiskunde in de maatschappij te propageren en van meet af aan stroomden er onderzoekers zijn statistische afdeling binnen met een vraag om advies. Aanvankelijk luidde het antwoord vaak dat het experiment van de onderzoeker geheel verkeerd was opgezet of uitgevoerd, zodat er geen valide conclusies uit getrokken konden worden. Het MC verwierf zich zo al snel een naam als het kerkhof van de medische onderzoekingen. Het werd al spoedig duidelijk dat het MC op die manier weinig toekomst zou hebben en nadat de aanzienlijk praktischer Jan Hemelrijk de leiding van de consultatie had overgenomen ging het beter. Men bleef strikte maatstaven hanteren, maar legde zich er ook op toe om te proberen het uiterste uit een niet ideaal uitgevoerd experiment te halen. Daarnaast probeerde men de 'klanten' uit te leggen hoe een experiment het best kon worden opgezet. Dit laatste was onderdeel van de algemene stelregel dat een samenspraak tussen klant en statisticus in een zo vroeg mogelijk stadium essentieel is. Daarbij bleek dan vrijwel altijd dat de onderzoeker geen precieze voorstelling heeft van wat hij zou willen bereiken en dat de statisticus de vakinhoudelijke problemen onderschat. Een andere stelregel was dat het MC een vetorecht behield op de formulering van de statistische conclusies door de klant. Dat was geen overbodige luxe. De klant realiseert zich soms onvoldoende dat statistische conclusies slechts betekenis hebben binnen een bepaald begrippenkader en heeft daardoor vaak een aanzienlijk rooskleuriger beeld van wat er statistisch is aangetoond.

De discussie

Wat hebben we in de voorafgaande bijdragen en elders over deze zaak vernomen? Aan statisticus Henk Elffers van het NSCR is door een rechter een ogenschijnlijk simpele vraag gesteld. Daarop geeft hij een simpel antwoord. Vervolgens komen er fundamentele bezwaren los van Ronald Meester die ook het model in twijfel trekt. Ook is er kritiek uit de Bayesiaanse hoek van Aart de Vos die het helemaal anders zou doen, zoals verwoord in een interview met Dirk van Delft in *NRC-Handelsblad* van 13 maart 2004. Meester en De Vos hebben hun weinig vleiende gedachten over de analyse van Elffers ook aan het Nederlands Juristenblad toevertrouwd. Wat er vermoedelijk bij de lezers van dit blad zal blijven hangen, is dat de statistici het weer eens totaal oneens zijn en er dus wel weer sprake zal zijn van 'lies, damned lies and statistics'.

De rol van de statisticus en die van de rechter

Als we over een statistische bijdrage tot de bewijsvoering in het strafproces tegen LdB spreken, moeten we goed in de gaten houden wat die bijdrage dan wel zou moeten zijn. Zo is het niet mogelijk om op grond van de ter beschikking staande gegevens over de aanwezigheid van LdB uit te maken of de sterf- en reanimatiegevallen een natuurlijke oorzaak hebben of aan (poging) tot) moord of doodslag te wijten zijn. De statisticus kan hoogstens besluiten dat LdB significant te vaak aanwezig was als er patiënten om niet verklaarde redenen gereanimeerd werden of overleden, maar daarvoor zijn ook andere verklaringen dan moord of doodslag te geven.

Elffers is zich hiervan terdege bewust en beschrijft omstandig de rol die hij als statisticus speelt. Alleen al de titel van zijn bijdrage: 'Geacht hof, het was geen toeval. De rest is aan u' laat weinig aan duidelijkheid te wensen over. Toch zou ik voor mijn verdere betoog wat preciezer met dit feit willen omgaan dan zijn mededeling aan het hof: 'Zoek het verder zelf maar uit.' Of er sprake is van moord – en daarmee bedoel ik voortaan (poging tot) moord of doodslag - dient inderdaad door de rechter op andere gronden te worden beslist. Hij beschikt daarvoor over getuigen, experts, laboratoriumresultaten, secties, etc. Op grond hiervan dient de rechter te besluiten of het wettig en overtuigend bewezen is dat in een aantal gevallen van moord sprake is. Als hij besluit dat dit niet het geval is, zijn we klaar en kan de verdachte worden vrijgesproken. Geen moord, geen moordenaar! Als hij besluit dat er wel moorden zijn gepleegd, komt de vraag aan de orde wie de schuldige is en kan de statisticus aan het werk. De moorden zijn dan dus een juridisch vaststaand feit en het doet er dan niets meer toe, wat de statisticus hiervan vindt. De jurist en de statisticus hebben ieder hun eigen terrein van deskundigheid. We nemen bij onze analyse dus aan dat bij een aantal met name genoemde patiënten sprake is van moord. Dit is van essentieel belang voor het statistische onderzoek.

Een tweede vraag die de rechter op onafhankelijke gronden (getuigen, de situatie ter plaatse) dient te beantwoorden, is wie in deze zaak redelijkerwijs tot de verdachten (mogelijke daders) behoren. Ook hierover kan de statisticus onmogelijk iets zeggen. Als we ons verder beperken tot de 8 moorden die Elffers bespreekt, dan is de rechtbank overduidelijk van oordeel dat de groep verdachten bestaat uit de 27 verpleegsters op de betreffende afdeling. Dit wordt niet expliciet opgemerkt maar is uit het vonnis duidelijk op te maken. Dit klinkt ook uiterst aannemelijk. Het is nauwelijks voorstelbaar dat derden op een afdeling in een ziekenhuis ongestoord medische ingrepen kunnen verrichten en artsen treden zelden onbegeleid op.

We moeten dus aannemen dat 8 moorden zijn gepleegd en dat de groep verdachten uit N=27 verpleegsters bestaat, waaronder LdB. Volgens het dienstrooster was LdB tijdens 142 van de 1029 diensten aanwezig en bij alle 8 moorden.

Bij lezing van het vonnis van de rechtbank valt op dat de zaken in de praktijk weer eens wat anders zijn verlopen dan eigenlijk zou moeten. Men concludeert inderdaad dat er 8 moorden zijn geschied, maar tegelijkertijd dat deze moorden door LdB zijn gepleegd. De argumenten zijn een samenvoeging van de aanwijzingen voor moord en de statistische conclusie van Elffers omtrent de aanwezigheid van LdB. Dit is een staaltje van wanordelijk denken en de statisticus behoort daarop te wijzen. Dit is wat Elffers bedoelt met zijn mededeling 'De rest is aan u', maar het zou duidelijker gezegd kunnen worden. De taakverdeling tussen rechter en statisticus houdt ook in dat de rechter moet weten dat hij bij het vaststellen van de moorden en de groep verdachten niet kan steunen op statistische overwegingen. Het is interessant om te zien dat het hof in hoger beroep de kwestie of er moorden zijn gepleegd opnieuw lijkt te overwegen.

Dit zijn essentiële zaken voor het statistisch onderzoek en ik kom hier later nog op terug. Men zou hier nog aan toe kunnen voegen dat de rechtbank ook overtuigd dient te zijn dat de moorden door één en dezelfde persoon zijn verricht. Zo niet, dan ontvalt eveneens de basis aan de statistische beschouwing.

Om ook verder enige precisie aan te houden definiëren we:

M: het feit dat 8 moorden zijn gepleegd;

R: het roostergegeven dat de 8 moorden tijdens de diensten van LdB plaatsvonden;

BO: de hypothese dat LdB onschuldig is;

BS: de hypothese dat LdB schuldig is;

N: aantal verdachten.

Aangezien het feit M vaststaat dient iedere statistische analyse te geschieden op basis van voorwaardelijke kansen gegeven dat M is opgetreden.

De analyse van Elffers

Wat Elffers uitrekent is de voorwaardelijke kans dat alle moorden plaatsvonden tijdens de LdB's diensten, gegeven dat er k=8 moorden plaatsvonden en dat LdB onschuldig is, dat wil zeggen hij berekent P(R|M, BO). Als er sprake is van D=1029 diensten en LdB tijdens d=142 diensten aanwezig was dan berekent hij met behulp van de hypergeometrische verdeling

$$P(R|M,BO) = \frac{\binom{d}{k}\binom{D-d}{o}}{\binom{D}{k}} = \frac{d!(D-k)!}{(d-k)!D!} \approx 1.1\times 10^{-7}$$

Deze berekening is correct als mag worden aangenomen dat ieder achttal diensten even waarschijnlijk is voor de moorden. Als dit juist is, dan geeft iedere zinnige statistische toets – bij voorbeeld de likelihood ratio test – als conclusie dat de hypothese dat LdB onschuldig is bij iedere redelijke onbetrouwbaarheiddrempel moet worden verworpen, en dat LdB dus schuldig is.

Er zit natuurlijk nog wel een addertje onder het gras, want de hier getoetste hypothese BO dat LdB onschuldig is, is uitsluitend onderzocht omdat de gegevens daartoe aanleiding leken te geven en dat is natuurlijk niet eerlijk. De standaard methode om onder dit verwijt uit te komen is om net te doen alsof we de hypothese van onschuld niet alleen voor LdB hebben onderzocht maar ook voor alle andere verdachten. Dan is de kans dat bij één of meer van deze verdachten iets zo extreems optreedt als bij LdB het geval is, ruwweg gelijk aan het aantal verdachten maal de berekende kans van 1.1×10^{-7}. Vandaar het belang van het afbakenen van de lijst met verdachten die we op N=27 verpleegsters hebben gesteld. We komen zo tot een kans van ongeveer

$$N\times P(R|M,BO) \approx N\times 1.1\times 10^{-7} \approx 3\times 10^{-6}\,.$$

Tenslotte zijn er nog soortgelijke gegevens uit twee andere ziekenhuizen waardoor deze overschrijdingskans met een factor 1000 wordt gereduceerd tot 3×10^{-9}. Deze kans is klein genoeg om de hypothese BO op goede gronden te verwerpen. Dit is de standaard statistische toets om zoiets aan te pakken.

De kritiek van Meester

Ronald Meester komt deels met praktische opmerkingen over het model, maar deels ook met fundamentele kritiek. De laatste betreft het voorbeeld van de loterij waarvan de hoofdprijs in de Leidse Wasstraat valt. Als we tevoren hadden afgesproken de loterij oneerlijk te verklaren als de prijs in die straat zou vallen en dit gebeurt vervolgens inderdaad, dan is er alle reden de loterij niet langer te vertrouwen. Als we echter pas na afloop constateren dat iemand in die straat gewonnen heeft, dan is er geen enkele reden tot wantrouwen. De prijs moet tenslotte ergens terecht komen.

De analogie met ons probleem is duidelijk: de aandacht werd pas op LdB gevestigd door het frappante verschijnsel dat zij steeds op het verkeerde moment aanwezig was en dan mogen we dezelfde gegevens niet zonder meer ook gebruiken om haar te veroordelen.

De analogie tussen beide zaken is echter niet volledig, dus laat ik het fictieve voorbeeld even afmaken zodat het meer op het probleem van mevrouw LdB lijkt zoals ik dat hierboven heb geformuleerd. De prijs is gevallen op Wasstraat 25. Zoals gezegd is dit geen reden tot verdenking, maar stel dat we al dan niet op grond hiervan besluiten de loterij eens tegen het licht te houden. Het blijkt nu dat er met de generator van de aselecte getallen voor de loterij is geknoeid en dat de dader moet behoren tot een groep van vijf hoogleraren in de kansrekening die de generator destijds heeft geleverd. Bij navraag blijkt dat één van deze hoogleraren op nummer 25 in de Wasstraat woont. Het lijkt mij duidelijk dat dit alles voldoende reden is om deze hoogleraar – een zekere Meester – maar eens op te pakken. We weten immers zeker dat er een misdrijf is gepleegd en dat weten wij op andere gronden dan de uitslag van de loterij. Ook de gevolgde methode bij het vervolgens uitgevoerde onderzoek (nagaan of de verdachten een prijs hebben gewonnen) is juist als in het geval van LdB (nagaan wie van de verdachten er bij de moorden in de buurt waren) niet of nauwelijks afhankelijk van de al dan niet opgevatte verdenking. Het is moeilijk in te denken dat we dit niet zouden nagaan! Tenslotte is in de terminologie van Meester – ditmaal bedoel ik de criticus en niet zijn schurkachtige alter ego – de schaal van het model ook duidelijk: de verdachten zijn niet de inwoners van de Wasstraat, of van Leiden, of van Zuid-Holland, maar de vijf kansrekenaars. Het loterijvoorbeeld correspondeert nu beter met het probleem van LdB in mijn formulering. Ik wees er immers op dat de statistische analyse van de dienstroostergegevens alleen dan zin heeft als op andere gronden duidelijk is dat er moorden zijn gepleegd en wie de groep der verdachten vormen. Het komt er dus op neer dat Meester en ik het geheel eens zijn dat er zonder deze beide feiten niet veel zinnigs valt op te merken. Ik zou echter zeker niet zover willen gaan als hij door te stellen dat de getallen van Elffers weinig betekenis hebben. Het tegendeel is het geval. Er is niets mis met de berekening van Elffers. Met zijn uitspraak 'De rest is aan u' heeft hij alleen verzuimd de rechter duidelijk te maken dat die berekening alleen iets betekent als de rechter op andere gronden overtuigd is dat er gemoord is en weet wie de verdachten zijn.

Zoals gezegd maakt Meester ook een aantal kritische opmerkingen over het model dat Elffers hanteert. Nu is het zo dat men bij ieder model de stelling kan verdedigen dat het de situatie niet nauwkeurig genoeg beschrijft. Wil dit echter als serieuze kritiek kunnen gelden, dan zal men toch ook aannemelijk moeten maken dat de conclusies door de keuze van een beter model drastisch zouden kunnen veranderen. Meester doet hiertoe geen poging en dit maakt zijn kritiek op deze punten minder overtuigend. Bij een niet zo heel erg onredelijk model dat uiteindelijk een overschrijdingskans van 3×10^{-9} oplevert zal het ook niet eenvoudig zijn het model zodanig te wijzigen dat deze kleine overschrijdingskans wordt weggepoetst. Laat ik toch in het kort even op de door Meester aangeroerde punten ingaan.

1. Elffers vermenigvuldigt de oorspronkelijke overschrijdingskans van 1.10572×10^{-7} met het aantal N=27 verdachte verpleegsters. Meester meldt ernstige bezwaren te hebben tegen deze post hoc correctie maar zijn argumentatie ontgaat mij. Wellicht heeft hij een andere groep verdachten op het oog, maar ik zie niet in welke dat zou moeten zijn.
2. Het zou bij voorbeeld heel goed kunnen zijn dat er verschil is tussen dag- en nachtdiensten, in die zin dat meer patiënten 's nachts dan overdag sterven en dat LdB juist veel nachtdiensten draaide. Of LdB zou juist ernstig zieke patiënten onder haar hoede hebben. Maar het ging hier helemaal niet om terminale patiënten, maar om patiënten die het redelijk goed ging en die zijn vermoord!
3. Meester mist in het betoog van Elffers een alternatieve hypothese. In mijn formulering is dit de hypothese BS dat LdB schuldig is. Dan is de kans dat zij steeds aanwezig was gelijk aan 1, of iets van dien aard.

Het Bayesiaanse pleidooi van De Vos

In zijn interview in *NRC Handelsblad* geeft Aart de Vos de voorkeur aan de Bayesiaanse aanpak. Dat is zijn goed recht, maar brengt natuurlijk wel het bekende probleem mee dat niet iedereen het over de gepostuleerde a priori kansen (of kansverdelingen) eens zal zijn. Dat is niet erg als het om een persoonlijke beslissing van De Vos gaat, waar anderen niet mee te maken hebben. Bij voorbeeld: Wat voor fiets moet hij kopen afhankelijk van zijn a priori veronderstellingen over het gebruiksgemak, de kans op reparaties, de kosten daarvan, de levensduur etc. van diverse merken. Zoals zo vaak gaat het hier echter om een uitspraak die liefst ook door anderen als zinvol moet worden beschouwd en dan is het noodzakelijk dat er een redelijke mate van overeenstemming bestaat over de toegekende a priori kansen. Een veroordeling wegens moord is niet iets dat op al te subjectieve gronden moet gebeuren.

Dit probleem is Bayesianen goed bekend. Om te voorkomen dat a priori kennis al te veel gewicht krijgt is het niet ongebruikelijk om zogenaamde niet informatieve a priori verdelingen te kiezen. Dit zijn verdelingen die geen sterke voorkeur voor de ene boven de andere hypothese uitdrukken. Een veel voorkomend voorbeeld is de uniforme a priori verdeling die alle mogelijkheden voor even lief neemt. Een ander idee dat door een vooraanstaand Bayesiaan als Jim Berger is uitgewerkt, is om met verschillende a priori verdelingen te werken en te zien wat de invloed daarvan op de conclusies is. Suzie Bayarri vertelde mij vorige zomer dat de nieuwste mode is om niet 'obese' maar O-Bayes te zijn, of wel 'objective Bayes'. Kortom: er wordt in het algemeen naar gestreefd om de invloed van de a priori veronderstellingen binnen de perken te houden en de Bayesiaanse en frequentistische standpunten naderen elkaar na decennia van opgewonden betogen. Dit is De Vos kennelijk ontgaan en hij ziet in Nederland nog steeds een vervolging van Bayesianen door domme wiskundigen die het licht niet hebben gezien. Hij gaat dan ook tegen de heersende trend in en maakt a priori veronderstellingen die de invloed van de waargenomen feiten neutraliseren, zo niet doodslaan. Dan moet je vrezen dat anderen – en met name rechters – daar niet in mee willen gaan.

Een Bayesiaanse analyse

Laat ik eerst mijn eigen Bayesiaanse analyse van het probleem geven en vervolgens proberen de analyse van De Vos in mijn eigen woorden na te vertellen. De Vos introduceert een nieuwe verpleegster, Lucie Klomp, die grote gelijkenis vertoont met mevrouw LdB, maar om verwarring te voorkomen zal ik ook zijn redenering weergeven aan de hand van de gegevens die Elffers in zijn betoog hanteert. Eerst dus maar hoe ik denk dat het zou moeten. Elffers berekende

(1) $P(R|M,BO) \approx 1.1\times 10^{-7}$.

Met de bekende rekenregels voor voorwaardelijke kansen vinden wij voor de kans dat zowel R als M als BO optreden,

$$P(R,BO,M) = P(R|M,BO)\times P(BO|M)\times P(M) \approx 1.1\times 10^{-7}\times P(BO|M)\times P(M)$$

en evenzo

$$P(R,BS,M) = P(R|M,BS)\times P(BS|M)\times P(M) \approx P(BS|M)\times P(M).$$

De laatste stap volgt omdat $P(R|M,BS) \approx 1$, aangezien de moorden en de schuld van LdB inhouden dat de moorden in haar aanwezigheid hebben plaatsgevonden. Nu vinden wij dus

$$P(BO|R,M) = P(R,BO,M) \;/\; P(R,M) = P(R,BO,M) \;/\; [P(R,BO,M) + P(R,BS,M)]$$

$$\approx 1.1\times 10^{-7}\times P(BO|M) \;/\; [1.1\times 10^{-7}\times P(BO|M) + P(BS|M)]\,.$$

Als alleen is gegeven dat de moorden hebben plaatsgevonden, dus zonder de informatie dat deze zich tijdens de diensten van LdB afspeelden, lijkt het redelijk om te stellen dat de kans dat één van de N verdachten de moorden heeft gepleegd, voor ieder van hen gelijk is aan 1/N. Dan geldt dus dat $P(BS|M) = 1/N$ en $P(BO|M) = (N-1)/N$. Omdat $1/N$ veel groter is dan 1.1×10^{-7} vinden wij

(2) $P(BO|R,M) \approx 1.1\times10^{-7}\times P(BO|M) / P(BS|M) = (N-1)\times1.1\times10^{-7}$.

De conclusie moet zijn dat LdB volgens deze Bayesiaanse analyse schuldig is, aangezien de kans dat zij onschuldig is in het licht van de feiten verwaarloosbaar klein is. Het is interessant om te zien dat de toetsingsprocedure van Elffers na post hoc correctie een kans opleverde van

$$N\times P(R|M,BO) \approx N\times1.1\times10^{-7}.$$

Beide analyses leiden dus tot vrijwel dezelfde zeer kleine kans.

Bayes volgens De Vos

Bij De Vos gaat de Bayesiaanse analyse echter anders. Ook hij gaat uit van (1), maar maakt geen onderscheid tussen de eventualiteiten M en R en interpreteert dit zuiver combinatorische resultaat met betrekking tot R gegeven M en BO, als de ‘kans op de doden wanneer LdB onschuldig is’. Zijn uitgangspunt is dus $P(M|BO) \approx 1.1\times10^{-7}$. Vervolgens kiest hij

$$P(BS) \approx (1/4)\times10^{-5} \qquad \text{en} \qquad P(M|BS) = 1/2$$

voor de a priori kans dat een willekeurige verpleegster zo’n serie moorden pleegt en de kans dat de moorden worden gepleegd als LdB schuldig is (deze laatste kans stelden wij hierboven op 1, maar dat maakt voor de orde van grootte van het eindresultaat niet uit). Aangezien $P(M|BO)\times P(BO) \approx 1.1\times10^{-7}[1-(1/4)\times10^{-5}]$ verwaarloosbaar klein is ten opzichte van $P(M|BS)\times P(BS) = (1/8)\times10^{-5}$ leidt dit alles tot

$$P(M) = P(M|BO)\times P(BO) + P(M|BS)\times P(BS) \approx P(M|BS)\times P(BS)$$

en vervolgens tot

(3) $P(BO|M) = P(M|BO)\times P(BO) / P(M) \approx 2.2\times10^{-7}\times[P(BO) / P(BS)] \approx 0.088$.

Nu is de kans dat LdB onschuldig is in het licht van de feiten dus nog steeds niet groot, doch aanzienlijk groter dan de kans van $(N-1)\times1.1\times10^{-7}$ in (2). Vervolgens voegt De Vos nog wat andere elementen toe waardoor de kans op onschuld nog verder stijgt. Anderzijds moeten we natuurlijk ook nog de gegevens van de beide andere ziekenhuizen in rekening brengen, wat de kans weer zou reduceren.

Waar komt het opvallende verschil tussen de resultaten (2) en (3) nu wel vandaan? Zoals gezegd maakt De Vos geen verschil tussen M en R en interpreteert bij $P(R|M,BO)$ als $P(M|BO)$. Waar bij De Vos de keuze van de a priori kans $P(BS) = (1/4)\times10^{-5}$ essentieel is, komt deze grootheid in mijn berekening niet voor. Vergelijking van (2) en (3) toont het verschil heel duidelijk. Afgezien van de onbelangrijke factor 2 in (3) is het verschil dat bij De Vos in (3) het quotiënt $P(BO)/P(BS)$ optreedt waar bij mij in (2) het quotiënt $P(BO|M)/P(BS|M)$ verschijnt. Het laatste ligt ook voor de hand. Als we weten dat er in het ziekenhuis patiënten zijn vermoord en dus één van de N verdachten moorden heeft gepleegd, dan is het niet meer interessant om te weten dat de Nederlandse verpleegster met kans $(1/4)\times10^{-5}$ aan het moorden slaat. We weten namelijk dat er in het ziekenhuis zo’n moordenares rondloopt en zonder nadere informatie heet die met kans $P(BS|M) = 1/N$ mevrouw LdB. Het is interessant dat precies ditzelfde punt aan de orde is geweest in het proces van O.J. Simpson in de VS.[1] Daar werd door de verdediging betoogd dat de kans dat iemand die zijn vriendin mishandelt haar ook vermoordt bijzonder klein is. Dit gaf de fameuze subjectivist I.J. Good aanleiding om op te merken: ‘But somebody did get killed, didn’t she?’

Sjerps over de taak van de deskundige

Marjan Sjerps wijst erop dat de getuige-deskundige vanuit zijn wetenschap antwoord moet geven op de gestelde vragen maar niet op de stoel van de rechter mag gaan zitten. Zij geeft de voorkeur aan een Bayesiaanse benadering, wat haar goed recht is, maar meent dat De Vos te ver gaat wanneer hij de a priori kans P(BO) dat LdB onschuldig is zelf van een getalswaarde voorziet. Zij vindt dat dit aan de rechter is voorbehouden. Zoals uit het voorafgaande blijkt, ben ik het met haar eens dat rechter en statisticus strikt gescheiden taken hebben. Ik meen echter ook dat het noch voor de statisticus, noch voor de rechter een kansrijke onderneming is om een getal P(BO) te bepalen op een manier die door derden als realistisch zal worden aanvaard. Zoals ik heb betoogd is dit ook niet nodig, omdat een statistische analyse alleen dan zin heeft als de rechter van mening is dat op andere gronden wettig en overtuigend bewezen is dat er moorden zijn gepleegd en wie de groep der verdachten vormen. Als dit inderdaad het geval is, is kennis van P(BO) niet meer relevant en zijn we slechts geïnteresseerd in de voorwaardelijke kans P(BO|M) dat LdB onschuldig is gegeven dat er sprake is van moorden. Bij N verdachten en zonder verdere informatie zal er weinig discussie zijn dat het verstandig is deze kans op 1/N te stellen.

Wat dit betreft ligt de opvatting van Sjerps heel dicht bij die van Elffers. Elffers merkt op dat de rechter het zelf verder maar moet uitzoeken, terwijl Sjerps de arme rechter een vrijwel onmogelijke opdracht geeft. Zoals ik hoop te hebben duidelijk gemaakt verdient het echter de voorkeur om meer concreet te zijn en aan te bevelen om over twee zaken – moorden en verdachten – een uitspraak te doen. Dat is iets dat de rechter vroeg of laat in ieder geval moet doen en waarvoor hij uitstekend is toegerust. Daar is geen getal voor vereist.

Conclusies

Zoals ik in deze bijdrage voortdurend heb getracht duidelijk te maken, is het van essentieel belang dat de rechtbank uitzoekt of, en zo ja, hoeveel moorden er zijn gepleegd en op wie, en wie de groep van mogelijke daders vormen. Dit dient met de daartoe geëigende juridische middelen te geschieden, waarbij de statistische analyse van de dienstrooster gegevens geen rol behoort te spelen. Als niet juridisch bewezen is dat er moorden zijn gepleegd, is de vraag naar de schuld van LdB niet aan de orde. Wie Agatha Christie’s *Murder on the Orient Express* heeft gelezen, zal trouwens ook graag de mogelijkheid van multipele daders door de rechtbank zien uitgesloten.

Indien aan deze voorwaarden is voldaan, is de analyse van Elffers overtuigend. Het model is goed te verdedigen en de wiskundige aanpak volgt dan vanzelf en is ook voor leken eenvoudig te begrijpen.

Zonder een uitspraak van de rechter over deze zaken is er reden te twijfelen aan de validiteit van de door Elffers gehanteerde methode. Ook kan men zich in dit geval moeilijk een Baysiaanse analyse voorstellen zonder discutabele a priori veronderstellingen. Die zullen een rechter niet overtuigen.

Mijn voornaamste probleem met de gang van zaken bij de rechtbank is dat Elffers zich tegenover de rechter misschien wat al te vrijblijvend heeft opgesteld en dat dit tot misverstand bij de rechter heeft geleid. In Elffers’ visie is de rechter de baas. De getuige geeft advies en dan moet de rechter maar zien wat hij daarmee doet. Zoals ik in mijn inleiding over de goede oude tijd getracht heb duidelijk te maken, heeft men al 50 jaar geleden ontdekt dat dit geen goed model voor statistisch advieswerk is. Bij een statistisch advies hoort altijd een aanduiding van de context waarbinnen dit advies geldig is en van de straf die erop staat als men het advies buiten deze context plaatst. Een duidelijke aanduiding van wat de rechter geacht wordt te doen wil het advies relevant zijn, had misschien kunnen voorkomen dat in het vonnis de argumenten voor moord met die voor de schuld van LdB werden vermengd. Hoewel er zeker sterke argumenten bestaan om tot het bewezen zijn van moorden te besluiten is nu uit het vonnis niet duidelijk of de rechter dit ook gedaan heeft.

Hetzelfde bezwaar, maar dan in een Bayesiaanse context, geldt voor de suggestie van Sjerps om de rechter op te knappen met iets dat ook de vakman niet bevredigend kan doen.

De zaak dient nog in hoger beroep voor het gerechtshof en het lijkt dus niet gewenst hier veel over te zeggen. Ik vermoed echter dat Meester het hof een beter handvat zou hebben aangereikt als hij zou hebben benadrukt dat de rechter een onafhankelijke uitspraak over moord en verdachten moet doen. Ik betwijfel of zijn principiële afwijzing van de analyse van Elffers zonder daar iets voor in de plaats te stellen de rechter op nieuwe ideeën heeft gebracht. Aangezien ironie op papier vaak wordt misverstaan, zeg ik voor de zekerheid maar even dat ik hem niet van loterijfraude beschuldig.

Tot slot nog één opmerking die door geen der betrokkenen wordt gemaakt. Toen men in het ziekenhuis verdenking tegen LdB had opgevat, ging men alle gevallen van onverwacht overlijden of reanimatie na. Bij zo'n onderzoek ligt er het levensgrote gevaar op de loer dat de onderzoekers al overtuigd zijn van wat zij moeten vinden. Als er dus een onverklaarbaar overlijden wordt gevonden waarbij LdB aanwezig was dan wordt dit verder onderzocht en vindt zo'n geval al gauw een plaats onder de moorden. Als daarentegen LdB niet aanwezig was dan zal ongetwijfeld de neiging bestaan om dit geval als een natuurlijke dood te classificeren. Dit is een overbekend verschijnsel in de statistiek en ik hoop maar dat iemand ook echt met de onderzoekers gepraat heeft om uit te vinden of zoiets mogelijkerwijs kan zijn gebeurd.

Noot

1. Mondelinge mededeling van onze vakhistoricus Stephen Stigler.

*Willem van Zwet is emeritus-hoogleraar in de mathematische statistiek aan de Universiteit Leiden.
E-mail: <vanzwet@math.leidenuniv.nl>.*

Lucia de B. krijgt in hoger beroep levenslang.

Den Haag, 19 juni.
Het gerechtshof in Den Haag heeft verpleegster Lucia de B. een levenslange gevangenisstraf met daarop aansluitend tbs met dwangverpleging opgelegd voor de moord op zeven patiënten en drie moordpogingen. [...]

Lucia de B. was vorig jaar maart door de rechtbank al tot levenslang veroordeeld voor de moord op vier patiënten en drie moordpogingen op patiënten. Zij was beschuldigd van dertien moorden en vijf moordpogingen tussen 1997 en 2001, toen zij in vier ziekenhuizen in en rond Den Haag werkte. Zowel de verdediging als het Openbaar Ministerie (OM) waren tegen de uitspraak van de rechtbank in hoger beroep gegaan. [...]

De kritiek van de verdediging op de statistische rapportages van het OM omzeilde het hof door niet één woord aan statistiek te wijden. Het OM had op basis van statistische rapporten nog betoogd dat de aanwezigheid van De B. bij alle incidenten geen toeval kon zijn, en had het hof gevraagd aan deze conclusie 'een groot gewicht' toe te kennen. De conclusie was door de verdediging fel bekritiseerd, en ook statistici hadden het oordeel van hun door OM aangetrokken collega in twijfel getrokken. Wellicht daarom besloot het hof de statistiek, in tegenstelling tot de rechtbank, geheel niet in haar overweging mee te nemen. [...]

In: *NRC Handelsblad*, zaterdag 19 juni 2004

column

Gerijmd en ongerijmd in de statistiek

Fred Steutel

Nederlanders zijn onverbeterlijke rijmelaars. Reclameteksten rijmen heel vaak; voorbeelden: ‘Twee halen, één betalen’, ‘Wees slim, gebruik Glim’. De volgende zinnetjes werden ingestuurd bij een slagzinnenwedstrijd: ‘Gezond gezin: Blue Band erin’ en ‘Bij ’t gebruik van Medinos laten al uw tanden los’. De eerste van deze twee werd bekroond, maar niet gebruikt, van de tweede is nooit meer iets vernomen. Op 5 december verscheen de voorpagina van dagblad *Tubantia* helemaal op rijm.

Ook in de statistiek is rijm populair; niet allemaal even fraai, maar toch. Behalve in rijm hebben statistici zich uitgeput in neologismen, Nederlandse woorden voor nieuwe statistische begrippen. Ook op dit terrein was er wisselend succes.

Rijmen, dichten en andere onzin.

Een belangrijke bron voor statistische rijmelarij, die tot de dag van vandaag doorwoekert, zijn motto’s en titels van voordrachten op Statistische Dagen. Wat dat betreft is het misschien jammer dat op onze jaarlijkse toogdag steeds meer Engels wordt gesproken. Wat Nederlandse voorbeelden; de auteurs laat ik, barmhartigerwijs, ongenoemd:

Meten = weten.
Gissen is missen.
Van gissen naar beslissen.
Van geval tot getal.
Rechtszekerheid bij onzekerheid.
Met kans meer mans.
Fraai model, past het wel?
Verloren illusie of geslaagde fusie?

Tot slot twee Engelse voorbeeldjes; het laatste rijmt niet helemaal, maar bijna.

Certainty about uncertainty
How to poison Poisson.

Dit alles om komende sprekers te ontmoedigen om hun voordrachten op rijm aan te kondigen: alle grappen zijn al gemaakt, en zo leuk waren ze nou ook weer niet. In het Engels zijn er ook versjes; het volgende vind ik aardig:

Negative correlation
means that
as one goes up
the other comes down
for example
when rain comes down
umbrellas go up.

Ook wel aardig vind ik de in het kader van de variantie-analyse door A.J. (Ton) Bosch geïntroduceerde Boschanova.

Nederlandse kansen.

Van het begin af, zeg vanaf 1946 was er discussie over de vertaling van kanstheoretische termen uit het Engels in het Nederlands. Het begon met de meest elementaire begrippen als *‘random variable’, (probability) distribution (function).* Nu ingeburgerd als respectievelijk toevalsgrootheid of stochast, (kans)verdeling en verdelingsfunctie. Toch is er nog steeds enige verwarring. Vooral het woord 'verdeling', dat door wiskundigen wordt gebruikt voor 'kansmaat op de reële as', wordt door sommigen gebruikt voor kansdichtheid en door anderen voor verdelingsfunctie.

Het woord stochast heeft een wonderlijke geschiedenis. Het is afkomstig van Van Dantzig, die geen 'klassieke opleiding' had. Hij vroeg destijds aan mensen op het MC (CWI) of stochast een redelijk zelfstandig naamwoord zou zijn bij het (passieve) werkwoord stochazomai (gissen, mikken). Wij wisten dat natuurlijk ook niet. Ik was wel klassiek opgeleid, maar het enige wat mij te binnen schoot was een parallel met 'ballast', waarvan ik dacht dat het afgeleid was van 'ballein', dat gooien betekent. Uiteindelijk koos Van Dantzig, al of niet op grond van mijn of andermans inbreng, voor stochast.

Freudenthal had zijn eigen ideeën en koos voor 'stochastiek', het woord dat nu de kansrekening, de statistiek en een deel van de OR omvat; soms wordt er speciaal de 'stochastische' OR mee bedoeld. Omdat Freudenthal nogal wat invloed had in kringen van de didactiek van de wiskunde (denk aan het tegenwoordige Freudenthal Instituut), heeft een aantal jaargangen middelbare scholieren het woord 'stochastiek' geleerd voor toevalsgrootheid. Freudenthal ging nog een stapje verder: wat wij tegenwoordig een normaal verdeelde stochast noemen heette bij hem een 'De Moivre stochastiek'. Ik geloof dat we blij moeten zijn dat deze uitdrukking geen gemeen goed is geworden.

Vertalen en onderstrepen.

In de statistiek zijn natuurlijk legio begrippen die vragen om een Nederlands woord. Een zo'n begrip is 'zuiverheid' van schatters, dat in het Engels met '*unbiased'* wordt aangegeven. Het Nederlandse woord is positief en daarom, vind ik, mooier dan het Engelse. Het gebruik van de uitdrukking 'negentig procent zuiver', dat gebruikt werd in de Wetenschapsquiz is ongebruikelijk en op zijn minst verwarrend.

Andere begrippen waarvoor Nederlandse uitdrukkingen goed zijn ingeburgerd, zijn 'voldoende' voor '*sufficient*', 'meest aannemelijke schatter' voor '*maximum likelihood estimator*'. De uitdrukking 'asymptotisch raak' voor '*consistent*' is niet echt populair geworden.

In Nederland en sporadisch in België werden stochastische grootheden onderscheiden van niet-stochastische door onderstreping. Zo schreef men $P(\underline{x} > x)$ in plaats van $P(X > x)$, zoals nu meer gebruikelijk is. Er is natuurlijk geen wezenlijk verschil; je gebruikt in beide gevallen twee alfabetten: onderstreept en niet-onderstreept of kapitaal en niet-kapitaal. Toch was er een klein onderscheid: In de theoretische statistiek zijn stochasten functies op een uitkomstenruimte Ω. Een experiment komt neer op het kiezen van een element ω in Ω. Nu is ω zelf in feite een stochast en we zouden dus eigenlijk $\underline{\omega}$ moeten schrijven. We komen dan uit op $\underline{x} = x(\underline{\omega})$. Het probleem is nu wanneer we ω onderstrepen en wanneer niet; vóór het experiment wel, daarna niet meer. Hoewel het onderstrepen enkele voordelen heeft (levert voor elk alfabet een extra alfabet) is het gebruik vrijwel beperkt gebleven tot Nederland; het is nu vrijwel verdwenen. Het heeft, denk ik, zijn nut gehad, vooral in de begintijd van de Nederlandse statistiek, toen het onderscheid tussen stochasten en hun mogelijke waarden niet altijd duidelijk was. In rapporten van het Mathematisch Centrum stond soms de wat verwarrende voetnoot: *'underlined letters denote random variables*'.

We kunnen ons nauwelijks nog voorstellen dat deze dingen ooit belangrijk werden gevonden. De Nederlandse stochastiek is volwassen geworden.

Fred Steutel is emeritus hoogleraar kansrekening aan de TU Eindhoven; hij is redacteur van STAtOR.
E-mail: <f.w.steutel@tue.nl>.