

Kansloos: van Willem Ruis tot Lucia de B.

Peter Grünwald

Centrum voor Wiskunde en Informatica

Kruislaan 413, 1098 XJ Amsterdam

homepages.cwi.nl/~pdg

Uitspraken van de vorm “deze gebeurtenis heeft X procent kans” zijn in de praktijk vaak betekenisloos. In veel alledaagse situaties kan men eigenlijk niet spreken van “kansen”, hoewel de meeste mensen (inclusief wiskundigen!) dit vaak wel doen. Dit is het tweede deel van een artikel gaat over dit soort “kansloze situaties,” die ik bespreek aan de hand van drie voorbeelden. In het eerste deel ging het over:

-Het 3-Gevangenen Probleem. Een wiskundige puzzel die laat zien dat een eenduidige “kans” soms niet bestaat (Deel 1, Sectie 2).

-Het 3-Deuren Probleem. Een, veel bekendere, wiskundige puzzel die laat zien dat onze intuïtie hierover vaak verkeerd is (Deel 1, Sectie 3).

Hieronder, in Sectie 4, gaan we verder met:

-Het 1-Gevangene Probleem. Dit is het waargebeurde verhaal van een recente rechtzaak waarin de veroordeling mede op basis van kansberekening is gebeurd. Het laat zien dat onze pogingen om over “kansen” te praten als die er niet zijn, desastreuze gevolgen kunnen hebben!

Vervolgens, in Sectie 5, gaan we meer in het algemeen kijken naar het boeiende gebied van statistiek in de rechtzaal: zijn er überhaupt situaties in de rechtzaal waarin het gebruik van kansberekening te rechtvaardigen valt? En wat denken statistici hierover? Zoals we zullen zien, zijn er hierbij nogal wat controverses. Zo is er onenigheid tussen statistici over de vraag of het in bepaalde situaties gerechtvaardigd of zelfs wenselijk is bepaalde data *weg te gooien*. Ook is er onenigheid over de vraag hoe met DNA evidentie om te gaan, wanneer een “verdachte” gevonden is door te zoeken in een grote database met DNA profielen. De algemene conclusie zal zijn dat, ondanks alle onopgeloste moeilijkheden, het gebruik van statistiek in de rechtzaal soms toch wenselijk en in feite onvermijdelijk is.

4. Het ÉÉN Gevangene Probleem

In 2004 werd verpleegkundige **Lucia de B.** in hoger beroep veroordeeld tot levenslang voor *7 moorden en 3 pogingen tot moord*. Zij heeft nooit schuld bekend. Ton Derksen, auteur van het boek *Lucia de B., Reconstructie van een gerechtelijke dwaling* heeft de zaak aangekaart bij de Commissie Posthumus-II (evaluatie afgesloten strafzaken). Deze commissie heeft drie wijze mannen aangesteld, die in oktober 2007 adviseerden dat de zaak inderdaad diende te worden heropend. Vervolgens heeft de hoge raad zelf ook onderzoek laten doen. Op 2 april 2008 adviseerde de Procureur-Generaal van de Hoge Raad de zaak te heropenen, en werd Lucia de B. (na meer dan zes jaar gevangenis) voorlopig vrijgelaten. Op 7 oktober 2008 besloot de Hoge Raad dat de zaak inderdaad heropend dient te worden. De zaak werd verwezen naar het Gerechtshof in Arnhem, waar de zaak in 2009 opnieuw gevoerd zal worden.



In deze zaak heeft statistiek een cruciale rol gespeeld.¹ Er zijn in de statistische analyse een aantal fouten gemaakt. Zoals we zullen zien is een van de belangrijkste fouten dat er van een ‘kans’ wordt gesproken in een situatie waarin dat eigenlijk niet kan – de kans is dus betekenisloos, net als in het 3-gevangenen probleem.

Lucia werkte op de Medium Care Unit van het Juliana Kinderziekenhuis (JKZ) in Den Haag. Er waren veel meer “incidenten” (plotselinge sterfgevallen en reanimaties) wanneer Lucia wel dienst had, dan wanneer Lucia geen dienst had. Op deze manier is de eerste verdenking gerezen: dat kon toch geen toeval zijn!?!?

Het hof vroeg de statisticus E. om de gegevens nader te analyseren. Deze statisticus berekende dat de kans dat een verpleegkundige bij toeval een dergelijk incidenten-patroon zou meemaken, kleiner is dan 1 op 342.000.000. Hij trok hieruit de conclusie dat het geen toeval kon zijn. Ik moet er meteen bijzeggen dat hij benadrukte dat dit *niet* betekent dat Lucia een moordenaar is. Ter illustratie geeft de statisticus een aantal mogelijke alternatieve verklaringen, in de trant van ‘het zou bijvoorbeeld kunnen dat Lucia vaker nachtdiensten draait, en dat ’s nachts meer patiënten sterven’. Het hof schrijft echter in zijn arrest (11.13):

¹ De statistiek was cruciaal voor de veroordeling door de rechtbank, in eerste instantie. Maar sommige mensen beweren nog steeds dat statistiek geen rol heeft gespeeld in het hoger beroep. Maar wie het arrest en de getuigenverklaringen bekijkt ziet wel degelijk een aantal (foutieve) statistische redeneringen. Dat blijkt al uit de citaten uit het arrest op de volgende pagina. Het boek van Derksen bevat nog andere voorbeelden van de impliciete rol van de statistiek in het hoger beroep. Ik geef er hier slechts een: bij een van de “vermoorde” patiënten is aan zes medische experts gevraagd of het om een natuurlijke dood ging. Vijf van de zes dachten van wel. De *enige* expert die dacht dat het niet om een natuurlijke dood ging, is dezelfde arts die oorspronkelijk een natuurlijke doodverklaring had afgegeven. Maar aan die niet-natuurlijke dood dacht hij pas vier jaar later, zoals hij zelf verklaart, nadat “in de media aandacht werd besteed aan onverklaarbare sterfgevallen in de diverse Haagse ziekenhuizen”. Het hof volgt deze laatste expert, die zich duidelijk heeft laten leiden door de statistische redenering dat zoveel onverklaarbare sterfgevallen ‘geen toeval kunnen zijn.’

Er is geen enkele aannemelijke verklaring gevonden voor het feit dat de verdachte in die korte periode bij zoveel overlijdensgevallen en levensbedreigende incidenten betrokken was.

Verderop in het arrest lezen we dat dit als belastend feit voor de verdachte wordt gezien. Dit speelt een belangrijke rol in de bewijsvoering. Dit wordt nog eens bevestigend in de inhoudsindicatie op pagina 1 (!) van het arrest. Daar lezen we: “Er is geen statistisch bewijs *in de vorm van toevalsberekeningen* gebruikt. *Wel* is voor het bewijs van belang geacht dat de in het JKZ gepleegde delicten een betrekkelijk korte periode bestrijken ... en hiervoor geen verklaring is gevonden.” (cursivering door mij toegevoegd). Het hof gebruikt dus wel degelijk statistiek – hoewel er geen getal genoemd wordt, wordt de statistische conclusie ‘het kan geen toeval zijn’ wel degelijk overgenomen (als het wel gewoon toeval kan zijn, is het niet vreemd dat er geen verklaring wordt gevonden voor de aanwezigheid van de verdachte bij al die incidenten, en kan de aanwezigheid op zich zeker niet als belastend worden gezien). Helaas blijft er bij nadere analyse niets over van de conclusie “het kan geen toeval zijn”.

De statisticus deed een “nulhypothese toets” met significantieniveau 1/10000. Dit is een standaard statistische methode. In grote lijnen werkt het als volgt: we formuleren eerst een zogenoemde ‘nulhypothese’ en een ‘alternatieve’ hypothese. In dit geval was de nulhypothese ‘Lucia heeft dezelfde kans om een incident mee te maken als andere verpleegkundigen’. De alternatieve hypothese is ‘Lucia heeft een hogere kans om een incident mee te maken als andere verpleegkundigen’. We kijken nu wat de kans² is op de daadwerkelijk geobserveerde gegevens als de nulhypothese waar zou zijn. Als die zgn. *overschrijdingskans* kleiner is dan het gekozen significantieniveau (in dit geval, 1 op 10000), dan verwerpen we de nulhypothese. Stel bijv. dat er 10 incidenten waren in de tijd dat Lucia op de afdeling werkte, en dat Lucia bij 8 incidenten aanwezig was. Dan berekenen we de kans dat Lucia 8 *of meer* van die 10 incidenten meemaakt onder de aanname dat Lucia een even grote kans heeft op een incident als andere verpleegkundigen. De statisticus berekende dat de kans dat Lucia evenveel of meer incidenten meemaakte, dan zij daadwerkelijk meemaakte op haar afdeling, kleiner was dan 1 op 342 miljoen.³ Dat is veel kleiner dan 1 op 10000. Daarom verwerpt hij de hypothese “Lucia heeft dezelfde kans op incidenten als andere verpleegkundigen,” en hij concludeert hieruit “wat er gebeurd is, is geen toeval.”

² We moeten hierbij heel voorzichtig zijn. We mogen niet zomaar de kans op de gegevens berekenen, want *elke* verzameling gegevens heeft uiteindelijk een hele kleine kans. Als we 10 keer met een eerlijke dobbelsteen gooien, dan heeft de uiteindelijke reeks die we gooien een kans van (1/6) tot de macht 10, vele malen kleiner dan 1 op 10000. Dit geldt altijd, welke reeks we ook gooien. We mogen hieruit natuurlijk niet concluderen dat de dobbelsteen vals is! Bij een nulhypothese toets bepalen we daarom niet de kans op de gegevens zelf, maar een zgn. *overschrijdingskans*. Dit is de kans op een speciaal gekozen *eigenschap* van de gegevens, waarbij die eigenschap aan bepaalde voorwaarden moet voldoen en *gekozen moet zijn vóórdat we de gegevens observeren*. Bij de dobbelsteen kunnen we bijv. kijken naar het gemiddeld aantal ogen. Als we waarnemen dat dat 4.5 in plaats van de verwachte 3.5 is, en de kans op een aantal ogen van 4.5 *of hoger* is kleiner dan 1 op 10000, dan kunnen we wel degelijk concluderen dat de dobbelsteen vermoedelijk niet eerlijk is. In het geval Lucia kijken we naar de (overschrijdings-) kans op *evenveel of meer incidenten* dan Lucia heeft meegemaakt.

³ In feite kwam de statisticus uit op een kans van 1 op 243 miljoen dat Lucia haar specifieke incidentenpatroon mee zou maken, onder de aanname dat haar kans een incident mee te maken even groot is als die van de andere verpleegkundigen. Met de verderop beschreven correctie (vermenigvuldigen met het aantal verpleegkundigen op de afdeling, 27) kwam hij op 1 op 9 miljoen: de kans dat iemand op de afdeling zo iets mee zou maken. Dit werd dan weer 1 op 342 miljoen door de cijfers te combineren met de cijfers van het Rode Kruis Ziekenhuis, waar Lucia eerder had gewerkt. Voor het gemak gaan we er verder vanuit dat alle gegevens van één enkel ziekenhuis afkomstig zijn.

4.1 “De” kans bestaat niet

Hoe werkt nulhypothese toetsen nou precies? De methode zit zo in elkaar, dat, als een statisticus hem herhaaldelijk (en correct) zou toepassen, dan zou gelden dat de statisticus gemiddeld maximaal 1 op 10000 keer zegt “dat kan geen toeval zijn” terwijl het wèl toeval is. Hij doet zo’n verkeerde uitspraak dus gemiddeld maximaal 1 op de 10000 keer dat hij de toets toepast. We kunnen bij een goed uitgevoerde hypothese toets trouwens *niet* zeggen (a) dat maximaal 1 op de 10000 keer dat er sprake is van toeval, er geconcludeerd wordt dat het geen toeval is; we kunnen alleen zeggen dat (b) maximaal 1 op de 10000 keer van alle keren dat de toets wordt toegepast, we in de situatie zitten dat het wel toeval is, maar dat we zeggen van niet. Het verschil tussen (a) en (b) is dat we, bij de definitie van “keer”, in (a) kijken naar alleen die toetsen waarbij het in werkelijkheid toeval is, en bij (b), naar alle toetsen, of het nou in werkelijkheid wel of geen toeval is.

Helaas is de nulhypothese toets in het geval Lucia niet correct toegepast, en *kan* hij ook helemaal niet correct toegepast worden. Dat zien we meteen als we ons gaan afvragen *wat “herhaaldelijk toepassen” hier zou moeten betekenen.*

Herhalen we de berekening, en doen we de uitspraak ‘wel/geen toeval’

1. ieder jaar, voor elke verpleegkundige op *de medium care unit van het Juliana Kinderziekenhuis?*
2. ieder jaar, voor elke verpleegkundige *in Den Haag/Nederland/in Europa/op de wereld?*
3. elke keer als een verpleegkundige *in het Juliana Kinderziekenhuis* zoveel sterfgevallen meemaakt dat het nader onderzocht dient te worden?
4. elke keer als een verpleegkundige *ergens in Nederland/in Europa/op de wereld* zoveel sterfgevallen meemaakt dat het onderzocht dient te worden?
5. of telkens als er een *rechtzaak* is waarin het Openbaar Ministerie van een nulhypothese toets gebruik maakt?

Dit is volstrekt onduidelijk. En als we de berekening proberen aan te passen aan de vijf gevallen hierboven, komen we steeds op iets anders uit (dit wordt verderop nader uitgelegd). Met andere woorden: zonder een precieze *context* aan te geven, is de uitspraak “het kan geen toeval zijn want de kans is 1 op 342 miljoen” simpelweg *betekenisloos*. Net als in het 3-gevangenen probleem is dit een ‘kansloze situatie’ waarin we niet, of in ieder geval niet zonder meer, van kansen kunnen spreken. In het hoger beroep is de rechter hierop expliciet geattendeerd door de hoogleraren M. Van Lambalgen (logica) en R. Meester (kansrekening), die optraden als deskundigen van de verdediging. Maar de rechter wilde hier niet aan, en bleef maar vragen ‘als u het niet met de statisticus E. eens bent, wat is volgens u de kans dan wèl?’

Normaalgesproken worden nulhypothese toetsen toegepast in situaties waarbij de nul- en alternatieve hypothese van te voren geformuleerd worden, en getest worden op nieuwe, onafhankelijk verkregen gegevens. Er wordt bijvoorbeeld een speciaal experiment opgezet om die gegevens te verkrijgen. Als men dit zorgvuldig doet, dan kan men garanderen dat *gemiddeld van alle keren dat iemand, in wat voor verdere context dan ook, een nulhypothese toets correct uitvoert, de nulhypothese maar 1 op de 10000 keer onterecht verworpen zal worden.* De verschillende toetsen hoeven niet over hetzelfde fenomeen te gaan: sommige van die 10000 toetsen kunnen bijvoorbeeld gaan over een nieuw

geneesmiddel, andere over de levensduur van gloeilampen of wat dan ook; als we alle toetsen bij elkaar nemen, dan kan de 1 op 10000 garantie toch gegeven worden. Maar zo een domein-onafhankelijke garantie kan alleen gegeven worden als de toetsen steeds op *nieuwe* gegevens worden toegepast. In het geval van Lucia wordt de nulhypothese echter getoetst aan dezelfde data waardoor hij gesuggereerd is. Dan kan de 1 op 10000 garantie alleen gegeven worden als precies bekend is in wat voor context de toets uitgevoerd wordt, en de berekening daaraan aangepast wordt. Als die context niet bekend is, is de uitkomst van de toets feitelijk betekenisloos.

De statisticus E. realiseerde zich wel dat er een probleem was, en pastte daarom een 'post-hoc correctie' op zijn hypothesetoets toe – hij vermenigvuldigde de overschrijdingskans met het aantal verpleegkundigen op Lucia's afdeling (27, in dit geval). Maar met deze correctie kan nog steeds niet gegarandeerd worden dat maar 1 op de 10000 gevallen onterecht gezegd wordt 'het is geen toeval', omdat nog steeds onduidelijk is wat de context is: 1 op *welke* 10000 gevallen? E.'s correctie zou de berekening min of meer correct maken voor de eerste interpretatie in het lijstje hierboven. Inderdaad: als we E.'s procédé elk jaar opnieuw uitvoeren op precies één afdeling van één ziekenhuis in Nederland, dan is zijn correctiemethode gerechtvaardigd, en dan zullen we maximaal 1 op 10000 keer dat we zijn procédé uitvoeren, in de situatie zitten dat we concluderen "het is geen toeval" terwijl het dat wel was. Als we echter, beducht geworden op gevaarlijke verpleegkundigen, besluiten het procédé elk jaar automatisch per computer uit te laten voeren voor elke verpleegkundige op elke afdeling in Nederland (context 2B in het lijstje hierboven), dan zouden we, voor iedere verpleegkundige, de gevonden kans moeten vermenigvuldigen met (ruwweg) het product van het aantal mensen op zijn/haar afdeling, en het totaal aantal ziekenhuisafdelingen in Nederland; we zouden dan alleen zeggen "geen toeval" voor een verpleegkundige waarvoor de resulterende – veel grotere – kans, nog steeds kleiner is dan 1 op 10000. Met deze correctie wordt de procedure enigszins verdedigbaar voor context 2B; "enigszins", want over complicaties zoals het feit dat sommige verpleegkundigen op verschillende afdelingen werken heb ik het nog niet eens. Een andere optie is om alleen een statisticus in te huren en hem/haar een uitspraak "wel/geen toeval" te laten doen in gevallen waarin "er al iets vreemds gebeurd is" ergens in een ziekenhuis in Nederland (context 4A hierboven). Dan moeten we weer anders corrigeren; om een ook maar enigszins betrouwbaar resultaat te krijgen, zouden we precies moeten weten wat bedoeld wordt met "er is iets vreemds gebeurd". Wordt er bijv. een statisticus ingehuurd zodra er een incidentenpatroon is opgetreden met een kans kleiner dan 1 op 100? Dan moeten we de gevonden kans al vermenigvuldigen met (ruwweg) 100. In de praktijk zullen we niet weten wat de criteria zijn om een statisticus ernaar te laten kijken, en dan weten we dus ook niet hoe we moeten corrigeren om de 1 op 10000 garantie te verkrijgen. Dus, zelfs als we iets over de "echte" context durven te zeggen (in het Lucia geval lijkt context 4A misschien nog het beste), kunnen we vaak nog geen correcte kansuitspraak doen omdat we de benodigde correctiefactor niet kunnen achterhalen. Het blijft hoe dan ook natte vinger werk – een "kansloze" situatie!

Het voorgaande suggereert dat de statistische analyse, hoewel die een zeer grote impact heeft gehad, eigenlijk niet zoveel zegt. Wanneer we andere relevante gegevens (beschikbaar ten tijde van de rechtszaak maar genegeerd door het hof) bekijken, dan krijgen we de indruk dat 'de statistiek' in feite heel weinig zegt, en in ieder geval niet in de richting van een seriemoordenaar wijst. Het blijkt nl. dat in de drie jaar dat Lucia op de medium care unit van het JKZ werkte, er daar **zes** sterfgevallen waren. *In de drie voordat ze er werkte, waren er*

zeven. Voor een nadere analyse van wat dit betekent, verwijs ik naar Derksen's boek. Derksen maakt ook aannemelijk dat de gegevens waarop de statisticus zijn analyse baseerde niet betrouwbaar zijn. Verder werd er ook nog een soort rekenfout gemaakt (vermenigvuldigen van overschrijdingskansen). Hiermee blijft er niets, maar dan ook niets van de oorspronkelijke statistiek over.

5. Wat kunnen we dan wèl zeggen?

Had de statisticus een andere methode kunnen gebruiken, waarmee hij wel een tot op zekere hoogte correcte ("geijkte") kansuitspraak had kunnen doen? Er zijn twee mogelijkheden. Beide zijn ze niet geheel bevredigend, maar wel te prefereren boven wat er daadwerkelijk gebeurd is:

5.1. Mogelijkheid 1: Data weggooien

Lucia had in een eerdere periode in het eveneens in Den Haag gelegen Rode Kruis Ziekenhuis (RKZ) gewerkt. Tijdens haar werkzaamheden daar was er geen verdenking gerezen. Als we ons tot deze gegevens beperken, zijn we feitelijk in dezelfde situatie als wanneer we onze hypothese formuleren *voordat* de gegevens geobserveerd worden: de gegevens waren weliswaar al eerder gerealiseerd, maar ze hebben de hypothese niet gesuggereerd, en dat is waarom het gaat: we hebben al aangegeven dat in dat geval de 1 op 10000 garantie gegeven kan worden zonder de exacte context van herhaling te kennen. We zouden dus de JKZ gegevens kunnen weggooien, en een hypothese toets op de RKZ gegevens kunnen baseren, zoals gesuggereerd door Meester, Collins, Gill en Van Lambalgen (2007). We komen dan uit op een overschrijdingskans van rond de 1 op 50, en bij de gekozen belofte "maximaal 1 op 10000 keer een foute uitspraak" kunnen we dus *niet* concluderen dat het geen toeval was. Het nadeel van deze methode is natuurlijk dat we op zich relevante data zomaar weggooien. Dit lijkt verkeerd: we moeten voor het dubbel gebruik van deze data corrigeren, maar weten niet hoe. Om de data dan gewoon maar weg te gooien is wel een heel drastische stap.

5.2 Mogelijkheid 2: Bayesiaanse methode

In deze methode veranderen we de vraagstelling. In plaats van de vraag of het allemaal wel of geen toeval was, proberen we *direct* te kijken naar de *kans dat Lucia schuldig is, gegeven de data*. Volgens de beroemde *Stelling van Bayes* geldt (waarbij we data afkorten tot *D*) :

$$P(\text{Lucia is Schuldig}|D)/P(\text{Lucia Onschuldig}|D) = \\ P(D|\text{Lucia Schuldig}) / P(D|\text{Lucia Onschuldig}) * P(\text{Lucia Schuldig})/P(\text{Lucia Onschuldig})$$

Hierbij is $P(\text{Lucia is Schuldig}|D)$ de conditionele kans dat Lucia schuldig is, gegeven de data *D*. Als we nog enkele verdere aannames doen, is de linkerbreek uit te rekenen. Deze zal veel

groter dan 1 zijn, omdat de kans op de gegevens onder aanname van Lucia's schuld veel groter is dan onder de aanname van haar onschuld. Echter, de rechterbreuk zal veel kleiner dan 1 zijn, omdat de a priori kans (dus voordat we verdere evidentie hebben verzameld) dat iemand een seriemoordenaar is, nu eenmaal heel klein is. Met een soortgelijke analyse, gebaseerd op de originele JKZ gegevens en een denkbeeldige a priori kans dat Lucia schuldig is van 1 op 100000, komt A. de Vos uit op een "a posteriori" kans (dus *gegeven* de data) van 90% dat Lucia schuldig is. Het grote voordeel van deze methode is dat er niet gecorrigeerd hoeft te worden voor de context waarin de uitspraak wordt gedaan: *als* de a priori kansen kloppen, dan is de uitkomst betrouwbaar, onafhankelijk van de context. Het nadeel is dat de a priori kansen onkenbaar zijn: we zouden naar statistieken over seriemoordenaars kunnen kijken om te kijken wat de a priori kans is dat een *volstrekt willekeurig gekozen* iemand een seriemoordenaar is (die zal nog ruim kleiner zijn dan 1 op 100000), maar Lucia is natuurlijk niet willekeurig gekozen. Om deze en dergelijke redenen wordt de aanpak met a priori kansen door sommige statistici a priori verworpen. Echter, men zou kunnen stellen dat juist in een rechtzaak, tenzij het tegendeel is bewezen, iedere verdachte als "normaal" beschouwd zou moeten worden. Zo is het bijvoorbeeld expliciet verboden om in een rechtszaak bij de schuldbeoordeling rekening te houden met de ethniciteit van de verdachte, ook al is het bekend dat mensen van sommige ethniciteiten nu eenmaal vaker dan gemiddeld bij bepaalde misdrijven betrokken zijn. Op dezelfde manier is het te rechtvaardigen om voor Lucia gewoon de a priori kans voor een willekeurig element van de bevolking te nemen. Nu doet zich natuurlijk weer de vraag voor: *welke* bevolking? Die van Noord-Holland, Nederland, Europa, de Wereld? In feite zitten we hier toch weer in een "kansloze" situatie, net als in het begin van dit artikel: de kansen zijn niet in te schatten, tenzij we specifiekere aannames doen. Toch is de situatie wat minder kansloos: het is duidelijk dat we een grote bevolkingsomvang moeten nemen, want we dienen Lucia als een "normaal" iemand te beschouwen. En of we nu Noord-Holland, Nederland, Europa of de Wereld nemen, het percentage seriemoordenaars in de bevolking zal hoe dan ook heel klein zijn. We kunnen dus misschien wel een enigszins betrouwbare *grens* aangeven: hoe we hem ook bepalen, de a priori kans dat iemand een seriemoordenaar is, is in ieder geval kleiner dan, zeg, 1 op 100000. Dan is de a posteriori kans in ieder geval kleiner dan 9 op 10, en dat is – op zichzelf – bij lange na niet overtuigend.

5.3. Statistiek in de Rechtzaal – moeilijk, hachelijk, maar onvermijdelijk

Beide methodes hebben dus voor- en nadelen. Ze worden hier natuurlijk toegepast op onbetrouwbaar gebleken data, en dus moeten alle uitkomsten met een grote korrel zout genomen worden. Maart 2007 besprak ik beide aanpakken tijdens een lezing in het *evidence seminar* aan University College London. Het publiek bestond uit statistici, juristen, psychologen en filosofen die gezamenlijk onderzoek doen naar nieuwe (betere) manieren om met "evidence" (bewijsmateriaal) om te gaan. Hoewel men het unaniem eens was dat de oorspronkelijke aanpak in de Lucia-zaak niet deugde, ontstond er een buitengewoon felle discussie tussen aanhangers van Bayesiaanse en van de data-weggooi aanpak. Het is duidelijk dat de deskundigen er zelf nog niet uit zijn wat de beste aanpak is. Mede gezien deze onenigheid onder de deskundigen zelf, zou men misschien moeten concluderen dat statistiek gewoon helemaal niet in de rechtzaal thuishoort. Een ander, vaak gehoord argument voor deze conclusie is dat je, als je een statistische uitspraak doet, je altijd een bepaalde kans hebt op een fout. Je hebt dus nooit zekerheid, en dus hoort het niet in de rechtzaal thuis. Dit is echter ook weer wat te kort door de bocht. Bij de *meeste* vormen van

bewijsmateriaal heb je namelijk geen zekerheid: het is bekend dat een getuige vaak meent in een persoon de dader te herkennen, terwijl die persoon het later niet geweest blijkt te zijn. Dit gebeurt ook bij de meest geavanceerde methoden van daderherkenning. Ook een uitspraak van zo'n getuige is niet met zekerheid waar. Hier kan statistiek juist helpen om een idee van de onzekerheid te krijgen: het is nuttig voor een rechter om te weten of getuigen zich nou 1 op de 10 of 1 op de 100 keer vergissen als ze denken de dader te herkennen. Zo kunnen er nog tal van andere voorbeelden worden gegeven waarbij onzekerheid in de rechtzaal een onvermijdbaar gegeven is, en statistiek juist gebruikt kan worden om een idee van die onzekerheid te krijgen.

Een onderdeel van de bewijsvoering waarbij statistiek hoe dan ook onvermijdelijk is, is *DNA matching*. Stel, er wordt een DNA-spoor gevonden op de plek van een misdrijf, en er wordt een persoon gevonden met hetzelfde DNA profiel. Als het DNA spoor van goede kwaliteit is, zal er maar één iemand op aarde zijn met hetzelfde profiel (de mogelijkheid van een-eiige tweelingen laten we even buiten beschouwing). Als er een persoon met dat profiel gevonden wordt, moet het DNA wel van hem/haar zijn – wat overigens nog niet betekent dat die persoon de dader is, het DNA kan natuurlijk om andere redenen op het plaats van delict terecht gekomen zijn. In de praktijk is het gevonden DNA spoor vaak echter maar een klein fragment, en van slechte kwaliteit bovendien. Dan is het bijvoorbeeld mogelijk dat de kans dat het DNA matcht met een volstrekt willekeurig gekozen persoon, ongeveer 1 op 1 miljoen is. Dit lijkt nog steeds heel klein, maar zoals we zometeen zullen zien is het vaak onvoldoende om ook maar enige conclusie te trekken. Het is hoe dan ook belangrijk dat, wanneer het gevonden DNA matcht met dat van een verdachte, deze zgn. *random match probability* door de getuige-deskundige wordt vermeld. Het maakt natuurlijk nogal een verschil of het 1 op 10, 1 op 1 miljoen of 1 op 1 biljoen is. Maar hoe moet je deze random match probability nu interpreteren? Dat gaan we in de volgende sectie bekijken.

5.4. De DNA-database controverse

Vaak is het zo dat er al een verdachte is, en dat vervolgens blijkt dat het DNA van de verdachte matcht met het DNA op de plaats van delict. Deze match geeft dan extra bewijsmateriaal. In sommige gevallen echter, wordt een potentiële verdachte *gevonden* door te zoeken in een database van DNA profielen. Dit gebeurde bijvoorbeeld in de zaak Diana Sylvester. Deze verpleegkundige werd meer dan 30 jaar geleden in San Francisco verkracht en vermoord. Een dader werd nooit gevonden, totdat, in 2004, gezocht werd in de DNA database van veroordeelde criminelen in Californië. Er werd precies een match gevonden met het DNA van het sperma van de verkrachter. Toen bleek dat de gevonden persoon, ene John Puckett, weleens veroordeeld was geweest wegens verkrachting en ten tijde van de moord op Sylvester in Californie woonde, is men tot vervolging overgegaan. John Puckett, inmiddels 70 en aan een rolstoel gekluisterd, heeft geen schuld bekend, maar is uiteindelijk wel veroordeeld.

Omdat de kwaliteit van het DNA in het sperma inmiddels achteruit was gegaan, was de kans op een match met een volstrekt willekeurig persoon ongeveer 1 op 1.1 miljoen. De database was echter zo groot, dat, ook al was niemand in de database de dader, de kans dat het DNA zomaar met één iemand in de database zou matchen, ongeveer 1 op 3 was. Men zou dus kunnen stellen dat het feit dat zo iemand gevonden is, nauwelijks iets zegt. Het is in eerste

instantie dan ook verbazend om te horen dat de jury in deze rechtzaak *niet* is verteld dat de verdachte gevonden is door te zoeken in een database, en dus ook niet hoe groot de database was, en dus ook niet dat er een redelijk grote kans was op een toevallige match. Nog verbazingwekkender wordt het als blijkt wat de reden is waarom dit niet verteld is: juist om de verdachte te *beschermen*, en de jury niet te bevooroordeelen, wilde de rechter niet laten vermelden dat de verdachte opgenomen was in een database vol met veroordeelde criminelen.

Er zijn statistici die deze gang van zaken volstrekt verkeerd vinden, en in eerste instantie lijkt het dat ook. Er zijn echter ook veel statistici die denken dat de grootte van de database er eigenlijk niet toe doet. Volgens deze, Bayesiaanse, redenering beginnen we weer met de *a priori* kans dat het DNA van Puckett is, als we verder nog niets over zijn eventuele betrokkenheid weten. We zouden deze bijvoorbeeld kunnen stellen op 1 op het aantal inwoners van Californie ten tijde van de moord, zeg 20 miljoen. De *a posteriori* kans dat het van Puckett is, gegeven dat hij gevonden is in de database, kan dan weer berekend worden met de regel van Bayes, op dezelfde manier als hierboven "de kans dat Lucia schuldig is, gegeven haar incidentenpatroon" is berekend. De *a posteriori* kans dat het DNA van Puckett is, wordt inderdaad ongeveer 1.1 miljoen keer zo groot, dus, zeg, 1 op 20, of de database nou klein of groot is. Dus: ook al is de kans op een toevallige match in de database erg groot, de *relevante* kans, nl. de *a posteriori* kans dat Pucket de dader is, is toch veel groter geworden nu blijkt dat hij als enige matcht. *Echter*, deze kans is dan natuurlijk nog steeds maar 1 op 20, dus bij lange na niet groot in absolute termen. Persoonlijk denk ik dat deze redenering in grote lijnen correct is, en dat dus, vanuit wiskundig oogpunt, de jury niet verteld hoeft te worden hoe groot de database is. Maar, omdat rechters en juries geen wiskundigen zijn, moet, denk ik, het volgende wèl *altijd* verteld moet worden: het is heel goed mogelijk dat tegelijkertijd geldt:

De kans dat het DNA matcht met X gegeven dat het DNA niet echt van X is, is, bijvoorbeeld, 1 op 1 miljoen

De kans dat het DNA niet echt van X is, gegeven dat het DNA matcht met X, is, bijvoorbeeld, 19 op de 20.

$P(\text{data} \mid \text{hypothese})$ en $P(\text{hypothese} \mid \text{data})$ zijn niet hetzelfde: het verband wordt gegeven door de Stelling van Bayes. Helaas hebben juries, maar ook rechters en officiers van justitie nogal eens de neiging om deze twee kansen met elkaar te verwarren. Deze redeneerfout staat in de literatuur zelfs bekend als *The Prosecutor's Fallacy*. Ik weet natuurlijk niet precies wat de jury in het geval Puckett is verteld, maar het zou me niet verbazen als de prosecutor's fallacy een rol heeft gespeeld in de veroordeling...

Samenvattend: sommige statistici vinden dat de grootte van de database absoluut relevant is, anderen denken dat hij irrelevant is. Dit probleem staat bekend als de *DNA database search controversy*. Het lijkt er echter op dat toch de meeste statistici met ervaring in rechtzaken, voor de Bayesiaanse aanpak kiezen waarbij de grootte van de database niet belangrijk is. Echter, deze statistici hameren er telkens weer op in de rechtzaal dat zij, als statistici, slechts iets kunnen zeggen over de kans op de data gegeven de verschillende hypothesen, bijvoorbeeld de kans op een match gegeven de hypothese "DNA is van Puckett", en de kans op een match gegeven "DNA is niet van Puckett". De *a priori* kans P van de hypothese, bijv. $P(\text{DNA is van Pucket})$, of zelfs $P(\text{Puckett is de dader})$, is enkel en alleen aan de rechter om te bepalen.

Tot Slot

Zoals we gezien hebben is het gebruik van statistiek in de rechtzaal vaak een hachelijke, maar toch in het algemeen onvermijdelijke onderneming. Er is veel onenigheid over wat de beste aanpak is, en ik vermoed dat we nog maar aan het begin staan van wat eigenlijk een nieuw onderzoeksgebied is: het op de juiste manier omgaan met kwantificeerbare "evidentie" in de context van rechtzaken. Ik hoop dat door voortschrijdende inzicht de huidige controverses langzaam maar zeker zullen worden vervangen door een consensus onder statistici en juristen, maar zover is het nog lang niet...

Colofon

Dit artikel is een uitbreiding van mijn bijdrage voor de Zomercursus Wiskunde 2008, georganiseerd door het Centrum voor Wiskunde en Informatica (CWI) te Amsterdam. Deze bijdrage was zelf weer gebaseerd op een lezing gehouden 12-12 2006, ter ere van het 60-jarig bestaan van te Amsterdam.

Literatuur

P.D. Grünwald and [J. Halpern](#). [When ignorance is bliss](#). *Proceedings of the Twentieth Annual Conference on Uncertainty in Artificial Intelligence (UAI 2004)*, Banff, Canada, July 2004 (beschikbaar via mijn homepage)

P.D. Grünwald and [J. Halpern](#). [Updating probabilities](#). *Journal of Artificial Intelligence Research (JAIR)* 19, pages 243-278, 2003 (beschikbaar via mijn homepage; bevat verdere verwijzingen naar literatuur over het quizmaster probleem)

T. Derksen. Lucia de B. Reconstructie van een gerechtelijke dwaling. Veen Magazines, Diemen, 2006.

R. Meester, M. Collins, R. Gill, M. van Lambalgen, . [On the \(ab\)use of statistics in the legal case against the nurse Lucia de B \(with discussion by David Lucy\)](#) . *Law, Probability and Risk*, 2007.