

Marjan Sjerps

Nederlands Forensisch Instituut

Volmerlaan 17

2288 GD Rijswijk

m.sjerps@nfi.minjus.nl

Onderzoek

Forensische statistiek

Forensische wetenschappen krijgen de laatste tijd veel aandacht in de media. Populaire programma's als 'Crime scene investigation' laten de mogelijkheden zien van forensisch technisch onderzoek bij opsporing en bewijsvoering in het strafrecht. Dit onderzoek bestrijkt een heel breed gebied en kan variëren van DNA-onderzoek tot schoensporenonderzoek, en van toxicologie tot handschriftkunde. Ook allerlei andere wetenschappers laten regelmatig van zich horen in de rechtszaal. Statistiek en kansrekening liggen vaak ten grondslag aan de verklaringen van deskundigen en spelen daarom een belangrijke rol in het strafrecht, hoewel ze zelden in de schijnwerpers staan als vakgebieden op zich. De 'forensische statistiek' is een heel breed en dynamisch vakgebied, waarin nog erg veel werk te verzetten is. Voor de wiskundige is het een fascinerend toepassingsgebied met spannende voorbeelden. Marjan Sjerps is als statisticus werkzaam bij het Nederlands Forensisch Instituut.

Volgens kenners vervullen deskundigen een toenemende rol in het strafrecht (Nijboer, 2002). Omdat statistiek een belangrijk onderdeel vormt van veel deskundigheidsgebieden is ook hiervoor een toenemende aandacht te bespeuren. In het buitenland zijn er een aantal publicaties en websites gewijd aan de

toepassing van statistiek in het (straf)recht (onder andere Finkelstein en Levin 1990, Aitken en Stoney 1991, Gastwirth 2000, Good 2001, websites van het Federal Judicial Center en van Charles Brenner). In het Nederlands is er eigenlijk heel weinig beschikbaar (er wordt beperkt aandacht aan geschonken in van Koppen et al. 2002, Broeders 2003) en er wordt momenteel door verschillende auteurs gewerkt aan een dergelijk boek (Coster van Voorhout en Sjerps; in voorbereiding). Omdat het zeer uiteenlopende deskundigheidsgebieden betreft, zoals bijvoorbeeld pathologie, biologie, fysica, chemie en psychologie, zijn de statistische technieken waarop men zich baseert ook zeer divers.

Toch zijn er ook een aantal aspecten die deze gebieden bij toepassing in het strafrecht met elkaar gemeen hebben. Zo is er meestal een verdachte in beeld en wordt er een bepaald kenmerk dat geassocieerd is met deze verdachte vergeleken met een spoor dat geassocieerd is met het delict. Het DNA-profiel van de verdachte wordt bijvoorbeeld vergeleken met het DNA-profiel van een spoor in de onderkleding van het slachtoffer van een zedendelict (figuur 1), of een geweer dat de verdachte in huis heeft wordt vergeleken met een kogel die gevonden is op de plek van de gewapende overval. Het kan ook zijn dat een hond de geur van een verdachte vergelijkt

met het geurspoor dat vermoedelijk aanwezig is op een voorwerp dat de dader heeft aangeraakt, en waarbij de deskundige een uitspraak moet doen over de bewijskracht als de hond een geurovereenkomst aangeeft. In dit soort onderzoek gaat het om de bewijskracht van waargenomen overeenkomsten en verschillen.

Tot ongeveer twintig jaar geleden was er geen algemene en bruikbare definitie van het begrip *bewijskracht* en had ieder deskundigheidsgebied zijn eigen methode om de bewijskracht van onderzoeksresultaten weer te geven. Met de opkomst van de zogenaamde Bayesiaanse stroming in de forensische statistiek (ook wel 'logische' of 'likelihood ratio' (LR) stroming genoemd), werd deze lacune gevuld. Er is een model ontwikkeld gebaseerd op de bekende regel van Bayes uit de kansrekening. Het geeft een numerieke uitdrukking voor de bewijskracht van allerlei bewijsmateriaal, maar kan ook als leidraad dienen voor deskundigen die hun conclusie verbaal weergeven. De standaardwerken op dit gebied zijn Aitken (1995) en Robertson en Vignaux (1995). Dit model wordt op steeds meer deelgebieden toegepast, waarbij soms interessante vragen worden gegenereerd (zie kaders). In dit artikel zal het Bayesiaanse model in het kort worden uitgelegd. Ook zullen er een drietal valkuilen aan bod komen waarin de gemiddelde jurist

Formuleren van hypothesen

Een nog onopgelost probleem dat de grondslagen van de statistiek betreft is het volgende. Stel, op de plek van een misdrijf gepleegd door twee daders worden twee bloedvlekken A en B aangetroffen, mogelijk afkomstig van de twee daders. De bloedvlekken zijn echter minuscuul en van bijzonder slechte kwaliteit, waardoor slechts een deel van het DNA-profiel zichtbaar gemaakt kan worden. Hieruit wordt geconcludeerd dat de twee bloedvlekken door verschillende personen zijn gemaakt. De 'match-kans' is één op de miljoen voor vlek A, en één op de honderd voor vlek B. De politie pakt één verdachte op, waarvan het DNA-profiel overeenkomt met dat van vlek A. De bewijswaarde, gemeten met de *likelihood ratio* (LR), hangt nu af van de manier waarop de hypothesen geformuleerd worden. Voor het hypothesenpaar

(1a) De verdachte is donor van één van de twee vlekken

(1b) De verdachte is niet donor van één van de twee vlekken

is de LR een half miljoen; voor het hypothesenpaar

(2a) De verdachte is donor van vlek A

(2b) De verdachte is niet donor van vlek A is de LR één miljoen.

Een verschil tussen de twee hypotheseparen is dat het eerste paar kan worden opgesteld zonder dat de DNA-profielen van de vlekken en de verdachte bekend zijn, terwijl de tweede juist is gebaseerd op deze informatie. De berekening van de LR is in beide gevallen gebaseerd op de aanname dat de vlekken A en B afkomstig zijn van verschillende personen, maar dit is pas bekend nadat de DNA profielen bekend zijn. De discussie gaat nu over het volgende. Mag de deskundige zijn hypothesen formuleren gebruikmakend van het bewijs dat voorhanden is, om vervolgens de LR te rapporteren voor deze hypothesen? Of is het illegaal om gebruik te maken van het voorhandenliggende bewijs, omdat hypothesen onafhankelijk van het bewijs geformuleerd zouden moeten worden (zie ook Meester en Sjerps 2004 en de discussie met Dawid).

gemakkelijk valt bij het interpreteren van het bewijs, maar waarin ook menig wiskundige zal trappen. Voor een vrij compleet overzicht

van dit soort valkuilen kan de lezer terecht in Broeders (2003).

Een heel andere tak van de forensische statistiek concentreert zich op de manier waarop gegevens verzameld worden. Hierbij kan het gaan om milieu-onderzoek, waarbij de vraag bijvoorbeeld is op welke manier er uit een berg afvalstoffen monsters genomen moeten worden, maar ook om partijen verdovende middelen. Het opzetten van experimenten ten behoeve van onderzoek behoort ook tot deze tak van sport. Ik zal hier kort op ingaan.

Om aan te geven waarom forensische statistiek meer is dan alleen de toepassing van gewone statistiek op een ongewoon terrein zal ik verder kort ingaan op de specifieke forensische aspecten. Tenslotte zal ik over de muren van mijn eigen werkterrein, de statistische aspecten van het technisch forensisch onderzoek van het Nederlands Forensisch Instituut, heen kijken en enkele andere terreinen aanstippen waarin statistiek wordt gebruikt ten behoeve van het strafrecht.

Hoe sterk is bewijsmateriaal: Het Bayesiaanse model

Om de kracht van bewijsmateriaal uit te drukken gebruikt de Bayesiaanse stroming in de forensische statistiek een algemeen model dat is gebaseerd op de regel van Bayes:

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]}$$

Hierin is $P[A]$ de kans dat A waar is en $P[A|B]$ de voorwaardelijke kans dat A waar is gegeven dat B waar is. Als we deze regel toepassen op twee hypothesen H_p en H_d en bewijsmateriaal E (de E staat voor 'Evidence') kunnen we afleiden (als we een subjectieve definitie van kans hanteren) dat

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

De interpretatie van deze formule is dat de kansverhouding van de twee hypothesen verandert door de introductie van het bewijsmateriaal. De nieuwe kansverhouding (het linker quotiënt) ontstaat daarbij uit de oude (het middelste quotiënt) door te vermenigvuldigen met de *likelihood ratio* (LR, het rechter quotiënt). In woorden luidt de formule:

De a posteriori kansverhouding is de a priori kansverhouding maal de likelihood ratio.

In het rechtssysteem kan deze regel als volgt worden toegepast: de aanklager heeft een bepaalde hypothese H_p (bijvoorbeeld:



Figuur 1 Het DNA-profiel van een verdachte wordt vergeleken met het DNA-profiel van een spoor in de onderkleding van het slachtoffer van een zedendelict.

de verdachte heeft deze dreigbrief geschreven), en de verdediging heeft een alternatieve hypothese H_d (bijvoorbeeld: iemand anders heeft die brief geschreven). De rechter zal de verdachte veroordelen als de hypothese van de aanklager, gezien al het bewijs, veel waarschijnlijker is dan die van de verdediging, dus als de posteriori kansverhouding tussen deze twee hypothesen groot is. Als het proces begint heeft de rechter a priori een bepaalde inschatting van de kansverhouding. Tijdens het proces worden de bewijsmiddelen aangevoerd: sommige daarvan zijn in het voordeel van de verdachte, andere in het nadeel. Telkens wanneer een bepaald bewijsmiddel wordt aangevoerd zal de rechter zijn inschatting van de kansverhouding aanpassen volgens de regel van Bayes. Hij vermenigvuldigt dus met een LR groter dan één als het gaat om bewijs in het nadeel van de verdachte, en met een LR kleiner dan één als het gaat om bewijs in het voordeel van de verdachte. Als al het bewijs in beschouwing is genomen neemt de rechter vervolgens een beslissing op basis van de uiteindelijk geschatte kansverhouding tussen de hypothese van de aanklager en die van de verdediging.

Hoewel dit een simpele toepassing lijkt, heeft deze denkwijze grote consequenties voor de wijze waarop deskundigen moeten rapporteren. De deskundige kan geen uitspraak kan doen over de a posteriori kansverhouding van de hypothesen van de aanklager en die van de verdediging zonder iets aan te nemen over de a priori kansverhouding. De handschriftdeskundige bijvoorbeeld kan niets zeggen over de kans dat de verdachte de dreigbrief geschreven heeft zonder aannamen te doen over deze kans vóórdat hij de brief onderzocht heeft. Dergelijke aannamen vormen bij uitstek het terrein van de rechter

De waarde van een DNA-databank match

Een bank wordt overvallen door een gemaskerde persoon. Het masker wordt later aangetroffen in de in brand gestoken vluchtauto, en de speekselvlekken die in het half verbrande masker worden aangetroffen geven een DNA-profiel waarvan de kans dat een willekeurig persoon hetzelfde profiel bezit geschat wordt op één op tienduizend (door de slechte kwaliteit van de monsters is slechts een partieel DNA-profiel verkregen). Het profiel wordt vergeleken met de DNA-databank van 3000 verdachten, en één persoon blijkt hetzelfde DNA-profiel te hebben. Het spoor is nu dus vergeleken met 3000 personen, in plaats van met één. Vraag: is de bewijskracht van de DNA-match groter of kleiner door de vergelijking met de databank?

Antwoord 1: Door het maken van meerdere vergelijkingen neemt de kans op een toevalstreffer toe. Als er tienduizend mensen in de databank zouden zitten dan zou je alleen al op grond van toeval een 'match' verwachten. De bewijskracht neemt dus af.

Antwoord 2: De informatie dat de 2999 'niet-matchende' personen in de databank uitgesloten kunnen worden als dader maakt het bewijs alleen maar sterker. Als de hele wereldbevolking in de databank zou zitten en er was maar één 'match', dan zou je met zekerheid weten dat deze persoon de dader was. De bewijskracht neemt dus toe.

Zie voor het antwoord het kader aan het eind van dit artikel.

en niet van de deskundige. De deskundige kan op basis van zijn expertise daarom alleen een uitspraak doen over de LR van het bewijsmateriaal voor de hypothesen H_p versus H_d . Volgens het Bayesiaanse model is dit dan ook de taak van de deskundige: het rapporteren van deze LR, zodat de rechter vervolgens hiermee zijn geschatte kansverhouding van H_p versus H_d kan aanpassen.

Een voorbeeld uit de praktijk

Bij een inbraak snijdt de dader zich aan het glas van de ingeslagen ruit. Van het bloed op dit glas maakt de DNA-deskundige een DNA-profiel dat wordt vergeleken met dat van de verdachte. In de rechtszaal beweert de officier van justitie dat het bloed afkomstig is van de verdachte (H_p), terwijl de advocaat zegt dat zijn cliënt niets met de zaak te maken heeft

en dat het bloed van iemand anders is (H_d). Stel nu dat de twee profielen exact met elkaar overeenkomen, wat is dan de LR? De teller is de kans op deze overeenkomst, als het bloed van de verdachte is. Als er geen fouten worden gemaakt dan is deze kans één. De noemer is de kans op een overeenkomst als het bloed van een willekeurige onbekende afkomstig is. De deskundige weet op basis van referentiebestanden iets over deze 'match-kans': deze is minder dan één op de miljard. De LR is dus 1 miljard, en hiermee wordt de kansverhouding van H_p versus H_d dus met een factor miljard vergroot. Maar hoe groot deze verhouding uiteindelijk is kan de deskundige niet zeggen, immers, dat hangt af van de a priori kansverhouding.

De Bayesiaanse stroming levert zodoende een definitie voor de taak van de deskundige, maar ook een numerieke definitie voor het begrip 'bewijskracht'. Uit het model volgt namelijk dat de grootte van de LR precies weergeeft hoeveel waarschijnlijker de hypothese van de aanklager wordt ten opzichte van die van de verdediging door het toevoegen van het bewijsmiddel. De LR geeft dus precies de bewijskracht weer van het bewijsmiddel. In bovenstaand voorbeeld is de bewijskracht tenminste een miljard. Als het bloedmonster van slechte kwaliteit geweest was waardoor slechts een deel van het DNA-profiel zou zijn verkregen, zodat de 'match-kans' bijvoorbeeld slechts één op de miljoen was, dan zou de LR een miljoen geweest zijn.

Verder kan het Bayesiaanse model gebruikt worden bij het combineren van verschillende bewijsmiddelen. Als de hypothesen hetzelfde zijn, dan is het eenvoudig af te leiden dat de LR van twee onafhankelijke bewijsmiddelen gelijk is aan het product van de afzonderlijke bewijsmiddelen. Zo kan in principe bij haaronderzoek morfologisch onderzoek worden gecombineerd met DNA-onderzoek. In minder eenvoudige zaken wordt de analyse al gauw zeer complex. Momenteel wordt onderzoek gedaan naar het gebruik van 'Bayesian belief networks' om bewijsmiddelen met elkaar te combineren.

De LR benadering is met succes toegepast in gebieden waar getallen voorhanden zijn. Een voorbeeld is DNA-onderzoek (zie onder andere Evett en Weir 1998, Sjerps en Kloosterman 2003), waar de LR niet alleen gebruikt wordt voor simpele vergelijkingen van een spoor met een dader, maar ook in meer complexe zaken waarbij de teller van de LR niet één is. DNA-verwantschapsanalyse is zo'n gebied, waarbij bijvoorbeeld een onbekend slachtoffer vergeleken wordt met de

vermoedelijke ouders, of een foetus met de moeder en haar vermoedelijke verkrachter. Minder bekende voorbeelden zijn automatische sprekerherkenning, waarbij op basis van een geautomatiseerd systeem de stem van de persoon op een telefoontap wordt vergeleken met de stem van een verdachte (zie Broeders 2003 en referenties daarin), en vergelijking van glas (Curran et al. 2000). Ook voor geuridentificatieproeven met honden, het herkennen van gezichten door getuigen en voor leugendetectors kan er een LR worden berekend (zie van Koppen et al. 2002 en referenties hierin, in deze literatuur wordt de LR meestal aangeduid met 'diagnostische waarde'). Hoewel er in de gebieden waar de LR daadwerkelijk berekend kan worden discussie is over het formuleren van de hypothesen, en de manier waarop een LR gerapporteerd moet worden, zijn de meeste onderzoekers het er tegenwoordig wel over eens dat het hun taak is een LR te rapporteren.

In gebieden waar getallen in veel mindere mate voorhanden zijn heeft het Bayesiaanse model veel stof doen opwaaien, en wordt ook zeker niet algemeen geaccepteerd dat een deskundige een LR moet rapporteren. Van oudsher doen deze gebieden verbale uitspraken over de waarschijnlijkheid van de hypothese van de aanklager, en er zijn slechts weinig onderzoekers geneigd om dit te veranderen. De wapendeskundige kan bijvoorbeeld concluderen dat 'de kogel zeer waarschijnlijk werd afgeschoten met het geweer van de verdachte'. Volgens het Bayesiaanse model kan hij op basis van zijn expertise hierover helemaal geen uitspraak doen, omdat deze waarschijnlijkheid afhangt van de a priori kansverhouding. In Bayesiaanse termen zou de deskundige zijn conclusie op een andere manier moeten formuleren, zoals bijvoorbeeld: 'het onderzoek levert een zeer sterke aanwijzing dat de kogel uit het geweer van de verdachte kwam'. Weinig juristen zullen echter het verschil tussen deze twee formuleringen herkennen. De discussie over de rapportage is daarom in dit soort gebieden nog in volle gang.

Denkfouten bij het waarderen van bewijs

Behalve het definiëren van de rol van de deskundige en de kracht van bewijsmateriaal heeft de Bayesiaanse denkwijze ook een aantal denkfouten aan het licht gebracht. Deze berusten op algemeen menselijke denkfouten en waren binnen de psychologie al langer bekend (Kahneman et al. 1982), maar binnen de forensische literatuur is er pas de laatste tien jaar serieuze aandacht voor (Evett 1995, Broeders 2003). De bekendste is de *prose-*

cutor's fallacy. Stel, in een moordzaak wordt bloed van de dader aangetroffen. Het (partitiële) DNA-profiel wordt vergeleken met dat van verdachte J. Jansen en deze blijken overeen te komen. De kans dat een willekeurige persoon dit profiel zal hebben is, zeg, één op de miljoen. De aanklager kan dan de volgende redeneerstappen te berde brengen:

1. De kans dat het profiel overeenkomt terwijl het bloed toch niet van J. Jansen is bedraagt één op de miljoen.
2. Het profiel komt overeen, dus de kans dat het bloed niet van J. Jansen is bedraagt één op de miljoen.
3. De kans dat het bloed wel van J. Jansen is bedraagt dus 99,9999%.

Vanzelfsprekend gaat de aanklager hier de mist in als hij (2) concludeert uit (1). Immers, de kans op overeenkomst gegeven dat het bloed niet van J. Jansen is, is niet gelijk aan de kans dat het bloed niet van J. Jansen is gegeven de overeenkomst. Toch blijkt uit de literatuur dat deze fout op grote schaal in de praktijk gemaakt wordt. Ook uit mijn eigen zeer bescheiden ervaring en die van mijn collega's blijkt dat Nederlandse juristen deze denkfout wel eens maken. De consequenties kunnen groot zijn in gevallen waarin de 'match-kans' relatief groot is en er vrijwel geen overig bewijs is.

De verdediging kan in dezelfde zaak als volgt redeneren:

1. Er zijn in Nederland nog zo'n 16 miljoen andere personen dan J. Jansen.
2. Naar verwachting zijn er dus 17 personen, inclusief J. Jansen, met hetzelfde DNA-profiel.
3. De kans dat het bloed van J. Jansen is bedraagt dus circa 1 op 17, ofwel circa 6%.

Deze redenering is op zich correct, maar er liggen een aantal verzwegen aannamen aan ten grondslag. Bijvoorbeeld dat de dader een Nederlander is, en dat alle Nederlanders (inclusief alle vrouwen, baby's, bejaarden, gehandicapten, ...) a priori met even grote waarschijnlijkheid de dader zijn. Ook deze redenering wordt wel eens in Nederlandse rechtszalen gebruikt. In de literatuur wordt dit de *defense fallacy* genoemd.

Tenslotte is er nog de *base rate fallacy*. Hierbij wordt de LR niet goed gewogen met de a priori kansverhouding. Stel dat er bijvoorbeeld een gemaskerde overval plaatsvindt waarbij speeksel van de dader wordt aangetroffen in een bivakmuts in de vluchtauto. Als er nog geen verdachten in beeld zijn kan de deskundige de frequentie van het DNA-profiel van het speeksel in verschillende bevolkingsgroepen met elkaar vergelijken. Stel dat in be-

waar het om gaat bijzonder groot en hetero-



Figuur 2 Milieu-onderzoek is een gebied waar de monstername van groot belang is voor de uiteindelijke conclusies. Tot op heden hebben laboratoria altijd veel tijd en geld besteed aan het optimaliseren van hun analysemethoden, terwijl er weinig aandacht bestond voor het nemen van de monsters. Tegenwoordig wordt onderkend dat de grootste foutenbron niet in het laboratorium, maar in het veld ligt.

volkingsgroep A de frequentie één op één miljoen is, en in bevolkingsgroep B de frequentie één op één miljard. De LR voor de hypothese dat de dader uit bevolkingsgroep A komt en niet uit B is dan duizend. De base rate fallacy is nu dat men geneigd is te denken dat dit betekent dat de kans dat de dader uit bevolkingsgroep A komt veel groter is dan dat hij uit B komt. Echter, de kans dat de dader uit bevolkingsgroep A komt hangt natuurlijk af van een groot aantal andere factoren, waarvan de grootte van de bevolkingsgroep de meest voor de hand liggende is.

Het bovenstaande maakt duidelijk dat het voor de forensisch deskundige niet alleen zaak is om te bedenken waarover hij precies een uitspraak moet doen, maar ook hoe hij bij de formulering van zijn conclusies denkfouten en misverstanden kan voorkomen. De ultieme oplossing voor het voorkomen van denkfouten en misverstanden is echter nog steeds niet gevonden, en psychologisch onderzoek is daarbij van toenemend belang.

Verzamelen van gegevens

Het bovenstaande heeft betrekking op situaties waarin conclusies moeten worden getrokken uit de resultaten van uitgevoerd onderzoek. Een groot deel van de vragen die mijn collega's en ik in de afgelopen tijd hebben gekregen hebben echter betrekking op situaties waarin de waarnemingen nog gedaan moeten worden. Milieu-onderzoek, bijvoorbeeld, is een gebied waar de monstername van groot belang is voor de uiteindelijke conclusies. Vaak is de vraag of de concentratie van bepaalde stoffen de wettelijke limiet overschrijft. Monsterneming en analyse van de monsters is echter erg duur, terwijl de partijen

geen van aard kunnen zijn (zie figuur 2). In bemonsteringsprotocollen die op wettelijke basis zijn voorgeschreven wordt daarom op basis van slechts enkele analyseresultaten een grote partij goed- of afgekeurd. Bij het handhavingprotocol bouwstoffen (uitvoeringsregeling 1998) bijvoorbeeld wordt op basis van drie analyseresultaten, elk verkregen door vier monsters met elkaar te vermengen, een partij van 2000 ton goed- of afgekeurd. In de praktijk blijkt dat de partijen niet altijd aan de aannames voldoen, bijvoorbeeld omdat er illegaal stoffen worden gemengd door de partij, waardoor de partij veel heterogener is dan aangenomen. Verder kunnen 'hot-spots' in de partij en moeilijkheden bij de analyse uitschieters veroorzaken. Het effect daarvan en de juiste statistische behandeling zijn nog onduidelijk. Voor de statisticus vormt dit gebied een enorme uitdaging (zie onder andere Keith 1996, Gilbert 1987, Patil en Rao 1994, Stelling en Sjerps 1999).

De monstername is vaak nog een ondergeschoven kindje. Laboratoria besteden veel tijd en geld aan het optimaliseren van hun analysemethoden, maar hebben opvallend weinig aandacht voor de monsters die ze binnen krijgen. In de praktijk ligt de grootste foutenbron echter niet in het laboratorium, maar in het veld. De verschillen tussen de monsters zijn meestal veel groter dan de verschillen tussen herhaalde analyses van hetzelfde monster. Ook de wetgever heeft weinig oog voor dit gegeven: de verdachte heeft recht op een contra-analyse in een onafhankelijk laboratorium, maar hiervoor worden monsters gebruikt die genomen zijn direct naast de monsters die al eerder geanalyseerd waren.

In het opzetten van een forensisch onder-



Figuur 3 In de rechtspraak is het dikwijls van belang om de kans te weten dat bijvoorbeeld een schoen van een verdachte een schoenspoor heeft veroorzaakt. Vele forensische statistici zijn van mening dat het Bayesiaanse of subjectieve kansbegrip beter op dit type vraagstukken van toepassing is dan het mathematische kansbegrip, gebaseerd op de axioma's van Kolmogorov.

zoek komt men steeds weer voor de vraag te staan hoe de waarnemingen verzameld moeten worden. Een voorbeeld hiervan is een zaak waarin de verdachte beweerde dat zijn pistool per ongeluk afging toen hij een bepaalde handeling verrichtte. De wapendeskundigen van het NFI hebben toen de waarschijnlijkheid van deze gebeurtenis bepaald door middel van een experiment waarin ze deze handeling met het betreffende wapen circa 400 keer hebben uitgevoerd. Vele dagen en blaren later concludeerden ze dat het wapen geen enkele keer was afgegaan. Meestal worden proefopzetten echter gemaakt in het kader van een onderzoek, bijvoorbeeld om een nieuwe methode te introduceren. Een derde situatie waarin het verzamelen van gegevens belangrijk is betreft de monsternamen uit een partij met discrete eenheden, zoals een partij met zakjes verdovende middelen of pillen, een partij vaten met onbekende inhoud die ergens gedumpt zijn in een verlaten gebied, of een groot aantal aangetroffen vezels op de kleding van een slachtoffer waarvan slechts een beperkt aantal nader geanalyseerd kan worden. Er zijn verschillende methoden om in zo'n situatie monsters te nemen. Momenteel wordt in Europees verband gewerkt aan een richtlijn voor het nemen van monsters uit partijen verdovende middelen, waarin de voor- en nadelen van deze methoden worden besproken (ENFSI drugs WG 2003).

Specifieke forensische aspecten

De forensische statistiek wordt gekenmerkt door een aantal bijzondere aspecten. Ten eer-

ste natuurlijk het veelzijdige gebied waarop het wordt toegepast, hetgeen de inzet van een breed scala aan statistische technieken vereist. Om u een indruk te geven noem ik: proefopzetten, classificatie, betrouwbaarheidsintervallen, Bayesiaanse methoden, simulatietechnieken, niet-parametrische methoden, controlekaarten, regressie en populatiegenetica. Verder zijn de onderzoeksobjecten van niet-alledaagse aard. Bij DNA-onderzoek kan het bijvoorbeeld gaan om bloed-, speeksel- en spermastellen, die soms in zeer kleine hoeveelheden aanwezig, half vergaan of met elkaar vermengd zijn. Bij fysisch onderzoek kan het bijvoorbeeld gaan om vuilniszakken, schroevendraaiers, schoenen (zie figuur 3), kogels, handschrift, stem of geluid; bij chemisch onderzoek om explosieven, schotresten, glasjeeltjes, tape, inkt en autolakken. Ook aan het vergelijken van oren wordt aardig wat statistiek verricht.

Het vergelijkend onderzoek, waarbij een spoor vergeleken wordt met een bepaald kenmerk of voorwerp dat met de verdachte geassocieerd wordt, heeft als specifiek aspect dat de interesse gericht is op de relatie van het spoor met één bepaald voorwerp of individu. De jurist wil bijvoorbeeld weten wat de kans is dat die ene schoen van de verdachte Pieterzen het schoenspoor veroorzaakte. Als gevolg hiervan sluit het Bayesiaanse of subjectieve kansbegrip veel beter aan bij de vraag dan het frequentistische of mathematische kansbegrip gebaseerd op de axioma's van Kolmogorov. Het is erg gekunsteld om in de forensische context een kans te zien als de fractie

van het aantal keren dat de schoen van Pieterzen het spoor maakte als er honderdduizend schoensporen worden gemaakt, of als één of andere abstracte kansmaat.

Vaak vindt de politie een aantal technische bewijsmiddelen in één zaak. Bij een inbraak kunnen bijvoorbeeld vingersporen, schoensporen, DNA-sporen, en werktuigsporen worden gevonden. Een specifieke forensische vraag is dan niet alleen wat de bewijskracht is van de afzonderlijke bewijsmiddelen, maar ook wat de bewijskracht is van de combinatie van deze middelen. Het Bayesiaanse denkraam biedt hier een richtlijn, zoals boven beschreven.

De juridische context waarin het onderzoek steeds wordt uitgevoerd zorgt ook voor een aantal kenmerkende eigenschappen. Zo is aan de deskundige meestal wel bekend wat de hypothese is die de aanklager zal aanvoeren tijdens het proces, maar meestal niet precies wat de hypothese is die de verdediging zal aanvoeren. De hypothesen kunnen verder ieder moment wijzigen als er nieuwe informatie ter beschikking komt, of als de verdediging onverwacht met een alternatief scenario op de proppen komt. De deskundige kan natuurlijk ook zelf alternatieve hypothesen genereren. Zijn de hypothesen eenmaal gesteld, dan is het inschatten van hun a priori kansverhouding voorbehouden aan de rechter. Waar de meeste wetenschappers uitspraken willen doen over de a posteriori kansverhoudingen van hun onderzoekshypothesen (bijvoorbeeld: dit is de kans dat de patiënt een bepaalde ziekte heeft, gegeven de symptomen), zal de forensische deskundige zich in zijn uitspraken volgens de boven beschreven Bayesiaanse theorie moeten beperken tot het rapporteren van de bewijskracht.

De juridische context veroorzaakt ook dat de forensisch deskundige bij het toetsen van hypothesen op de klassieke manier in een ander type fout geïnteresseerd is dan gebruikelijk. Bij het vergelijken van de brekingsindex van glasjeeltjes gevonden in de kleding van de verdachte met de brekingsindex van monsters van de ingeslagen ruit bij een inbraak is de gebruikelijke nulhypothese dat de twee populaties niet van elkaar verschillen (Rudin en Inman 2003). Normaliter is de statisticus dan geïnteresseerd in de type I fout, de kans om de nulhypothese te verwerpen terwijl deze waar is, en is de procedure erop gericht om deze fout onder controle te houden, bijvoorbeeld maximaal 5% of 1%. Deze fout is echter geassocieerd met de kans dat de verdachte ten onrechte wordt vrijgesproken. Immers, dit is de kans dat de deskundige concludeert dat

de glasdeeltjes niet van de ruit afkomstig zijn terwijl dat wel zo is. Voor de deskundige is dit echter niet de belangrijkste fout. Van groter belang is het om de type II fout te controleren, de conclusie dat de glasdeeltjes van de ruit afkomstig zijn terwijl dat niet zo is, wat tot een onterechte veroordeling kan leiden.

Overig onderzoek in de forensische statistiek

Statistiek wordt veel vaker gebruikt in juridische toepassingen dan veel mensen zich realiseren. Soms vormt ze de basis voor de uitspraak van een deskundige, bijvoorbeeld in de strafrechtspsychologie die zich onder andere richt op de bewijswaarde van herkenningen door getuigen, of voor de uitslag van een leugendetector (van Koppen et al. 2002). Ook de bewijswaarde van geuridentificatieproeven met speciaal getrainde honden is onder andere gebaseerd op statistische overwegingen (Schoon en van Koppen 2002). Een heel ander gebied waar statistiek

een rol speelt is bij het inschatten van recidive kansen van personen die vallen onder de TBS maatregel (de Ruiters 2002, Brand en Diks 2001). Hierbij kan gebruik worden gemaakt van een groot scala aan testlijsten, die gecombineerd met de klinische blik van de psychiater of psycholoog leiden tot een schatting van deze kans. Dit wordt vervolgens door de jurist gebruikt voor zijn oordeel of de TBS al dan niet verlengd moet worden.

Ook de kansrekening draagt haar steentje bij aan het strafrecht, bijvoorbeeld bij het beoordelen of een spel gezien moet worden als een behendigheids spel of als een kansspel (Van der Genugten et al. 2001). Andere voorbeelden zijn de kansanalyses die zijn uitgevoerd in zaken als die van Bianca K., de kinderverzorgster die ervan werd beschuldigd een speciale vorm van benauwdheid te hebben veroorzaakt bij een aantal kinderen dat zij in het dagverblijf onder haar hoede had. Een soortgelijke bekende zaak is die van Lucia de

B., de verpleegster die ervan werd verdacht de hand te hebben gehad in een aantal raadselachtige overlijdensgevallen op haar afdeling. Bij dit type zaken speelt de vraag, hoe groot de kans op bepaalde waarnemingen is als er alleen toeval in het spel is, een grote rol. (Elffers 2003).

Oplissing: antwoord 2 is correct. De bewijskracht is sterker, maar dit hoeft niet te betekenen dat de kans dat de verdachte de dader is groter is. Immers, als de verdachte in beeld komt door een databank zoekactie, dan hoeft er geen overig bewijs tegen de verdachte te zijn. Als de verdachte in beeld komt op grond van overig bewijs, dan zal dit de a priori kans dat de verdachte de dader is flink verhogen, en daarmee de posterior kans. Zie Meester en Sjerps (2003) en de daarin opgenomen referenties.

Referenties

- Aitken C.G.G., *Statistics and the evaluation of evidence for forensic scientists*, Wiley, Chichester, 1995.
- Aitken C.G.G. and Stoney D.A. (eds) *The use of statistics in forensic science*, Ellis Horwood Limited, Chichester, 1991.
- Brand E.J.F.M. en Diks G.J.M. 'Richtlijnen voor risicotaxatie in de forensische diagnostiek: theorie en praktijk'. *Tijdschrift voor psychiatrie* (43) 10 (2001), p. 693-704.
- Broeders, A.P.A. *Op zoek naar de bron — over de grondslagen van de criminalistiek en de waardering van het forensisch bewijs*, (proefschrift) Kluwer, Deventer, 2003.
- Coster van Voorhout J.A. en Sjerps M. (eds) *Werktitel: Statistiek voor de strafrechtjurist*, Kluwer, te verschijnen.
- Curran J.M., Hicks T.N., Buckleton J.S. *Forensic interpretation of glass evidence*, CRC Press, Boca Raton, 2000.
- Elffers H. 'Bij toeval veroordeeld? — statistische analyse van dienstroosterdata in het strafproces', *Nederlands Juristen Blad* (34), (2003), p. 1812-1814.
- ENFSI (European Network of Forensic Science Institutes) *Drugs Working Group, Guidelines on representative drug sampling*, 2003.
- Evelt I.W., 'Avoiding the transposed conditional', *Science & Justice* (35), (1995), p. 127-131.
- Evelt I.W. en Weir B.S., *Interpreting DNA-evidence*, Sinauer associates, Sunderland MA, 1998.
- Finkelstein M.O. and Levin B. *Statistics for lawyers*, Springer-Verlag, New York, 1990.
- Gastwirth J.L.(ed.) *Statistical science in the courtroom*, Springer-Verlag, New York, 2000.
- Genugten B.B. van der, Das M. en Borm P.E.M., *Behendig gokken in en rond het casino* (+cd-rom), Academic Service, 2001.
- Gilbert R.O. *Statistical methods for environmental pollution monitoring*, Van Nostrand Reinhold, New York, 1987.
- Good P.I. *Applying Statistics in the Courtroom: A New Approach for Attorneys*, CRC Press, Boca Raton, 2001.
- Kahneman D., Slovic P. en Tversky A., *Judgment under uncertainty: heuristics and biases*, Cambridge University Press, Cambridge, 1982.
- Keith, L.H. (ed.) *Principles of environmental sampling, 2nd ed.*, American Chemical Society, USA, 1996.
- Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002.
- Meester R en Sjerps M. 'The evidential value in the DNA-database search controversy and the two-stain problem', *Biometrics* (59), (2003), p. 727-732.
- Meester R. en Sjerps M. 'Why the effect of prior odds should accompany the likelihood ratio when reporting DNA-evidence', *Law, Probability and Risk* (te verschijnen in 2004).
- Nijboer J.F. *Het gekooid denken — over de rol van discipline binnen feitenonderzoek en bewijs in het recht*, Inaugurele rede, Universiteit Leiden, 18-10-2002.
- Patil G.P. en Rao C.R. (eds.), *Handbook of statistics 12: Environmental statistics*, North Holland, Amsterdam, 1994.
- Robertson B. and Vignaux G.A. *Interpreting evidence*, Chichester, Wiley, 1995.
- Rudin N. en Inman K. 'Articulating Hypotheses-The Null Hypothesis and Beyond', *News of the California association of criminalists*, 1st Quarter, (2003), p. 7-8.
- Ruiters C. de, 'De terbeschikkingstelling', in: Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002, p. 991-1001.
- Schoon A. en van Koppen P.J. 'Identificatie door honden', in: Koppen P.J. van, Hessing D.J., Merckelbach H.L.G.J., Crombag H.F.M., *Het recht van binnen — psychologie van het recht*, Kluwer, Deventer, 2002, p. 597-622.
- Sjerps M. en Kloosterman A.D. 'Statistical aspects of interpreting DNA-profiling in legal cases', *Statistica Neerlandica* (57) 3, (2003), p. 368-389.
- Stelling, M. en Sjerps M. 'A sampling and hypothesis testing method for the legal control of solid hazardous waste', *Environmetrics* (10), (1999), p. 247-259.
- Uitvoeringsregeling bouwstoffenbesluit bijlage F; Handhavingprotocol schone grond, Staatscourant 30 januari 1998.
- Websites: Federal Judicial Center: www.fjc.gov Charles Brenner: www.DNA-view.com.