

# Berekening en toepassing van forensische bewijswaarde: frequentistisch of Bayesiaans?

Klaas Slooten

Nederlands Forensisch Instituut, en VU Universiteit Amsterdam

*k.slooten@vu.nl*

Nationale Wiskundedagen, 3 februari 2018

# Overview

Forensisch bewijs

**Klaas Slooten**

Inleiding

Inleiding

Forensisch  
DNA-onderzoek

Forensisch DNA-onderzoek

Afweging  
methodes

Afweging methodes

Conclusies

Conclusies

De evaluatie van forensisch bewijs gaat vaak over het koppelen van een spoor aan een bekende bron, b.v.

- ▶ Een DNA-spoor aan een persoon
- ▶ Een vingerafdruk aan een persoon
- ▶ Een handschrift aan een persoon
- ▶ Een XTC-pil aan een drugslab
- ▶ Een document aan een printer
- ▶ Een bandenspoor aan een auto
- ▶ Een DNA-spoor aan een broer van de mogelijke dader,...

of van sporen aan elkaar.

# Uitsluiting, versus geen uitsluiting

Als de mogelijke bron niet de echte bron is, is dat vaak (maar niet altijd) met zekerheid vast te stellen:

- ▶ Dit DNA-spoor kan niet afkomstig zijn van verdachte
- ▶ Deze vingerafdruk komt niet van verdachte
- ▶ Dit is niet het handschrift van verdachte
- ▶ Deze XTC-pil komt niet uit dit drugslab
- ▶ Dit document komt niet van de bewuste printer
- ▶ Dit bandenspoor kan niet door deze auto gemaakt zijn

# Koppeling bron-spoor

Als een spoor wel van een bron kan komen, wil een rechtbank graag weten hoe sterk die link is. In veel forensische disciplines was het gebruikelijk om te rapporteren in de trant van

- ▶ een vingerafdruk is met absolute zekerheid afkomstig van verdachte
- ▶ een werktuigspoor is met aan zekerheid grenzende waarschijnlijkheid afkomstig van een bepaald werktuig

In alle gevallen wordt een kans gegeven: de kans dat het spoor van de bron afkomstig is, is

- ▶ 0 (uitsluiting)
- ▶ 1 (zekerheid)
- ▶ niet precies omschreven maar “aan zekerheid grenzend”

Mooi! Of..misschien te mooi om waar te zijn?

- ▶ Waar is de context eigenlijk gebleven? Deze uitspraken zijn puur op basis van spoor en bron
- ▶ Als die niet nodig is, is het dus kennelijk zo dat een even grote overeenkomst tussen spoor en bron steeds dezelfde kans geeft dat het spoor van die bron is.
- ▶ Maar dat kan nooit in zijn algemeenheid waar zijn! Stel b.v. dat een schoenspoor overeenkomt met de schoen van verdachte. Is het dan zo dat de kans dat verdachte het schoenspoor achter heeft gelaten, even groot is als voor ieder ander met een identiek paar?
- ▶ Als dader bloedgroep  $AB$  heeft, is dan voor iedereen met bloedgroep  $AB$  de kans even groot dat hij het bloedspoor heeft achtergelaten? Natuurlijk niet!

Stel we vinden na een delict een spoor van een dader waaruit een DNA-profiel kan worden vastgesteld. Laten we zeggen, dat 1 op de miljoen mensen dat profiel heeft. Dan maakt het natuurlijk uit of

- ▶ Er een match wordt geconstateerd bij een verdachte die al in beeld was
- ▶ Er een match wordt geconstateerd in een databank van 1 miljoen DNA-profielen van Nieuw-Zeelandse zuigelingen.

Zolang je niet kan uitsluiten dat er ook andere bronnen kunnen bestaan die het spoor kunnen hebben achtergelaten, kan je niet zeggen wat de kans is dat het de bewuste onderzochte mogelijke bron is.

# Alles gaat zonder meer goed als...

- ▶ De bron kan worden uitgesloten
- ▶ Alle mogelijke bronnen onderzocht kunnen worden, en het er maar eentje kan zijn.

Helaas voor de forensiche (maar gelukkig voor de wiskundige) praktijk is dit geen uitputtende opsomming van de mogelijkheden.

Uiteindelijk wil een rechtbank een kans kennen van de vorm

$$P(H_p | E),$$

waarbij  $H_p$  staat voor de kans dat het door de aanklager beschreven scenario klopt, en  $E$  voor het bewijs dat is ingebracht.

Een eenvoudig voorbeeld is  $H_p$ : verdachte heeft een spoor achtergelaten, en  $E$ : DNA-match.

We kunnen dit op twee manieren proberen te benaderen:

- ▶ Frequentistisch: Vanuit een *nulhypothese* van onschuld, die we verwerpen als het bewijs sterk genoeg is
- ▶ Bayesiaans: Door de kans  $P(H_p | E)$  (zo goed mogelijk) te bepalen

# Frequentistisch

Hypothesetoets:

- ▶ Nulhypothese  $H_0$  waaronder je kan voorspellen met welke kans je welke data krijgt
- ▶ Experiment dat data oplevert, volgens  $H_0$  uit een of andere kansverdeling
- ▶ Als de data “te onwaarschijnlijk” zijn (b.v. tot de minst verwachte 5% behoren), verlies je je geloof in  $H_0$
- ▶ Experiment leidt dus tot al of niet *verwerpen* van  $H_0$ .

De beslissing hierover hangt af van de  $p$ -waarde: de kans dat de data die je had kunnen krijgen, even veel of meer afwijkt van wat je had verwacht als wat je daadwerkelijk hebt gezien. Dus:

- ▶ verwerpen of niet hangt ook af van de data die je *had kunnen krijgen*
- ▶ maar niet van iets anders dan de kansverdeling van de data onder  $H_0$

- ▶ De data zijn hoe dan ook ontstaan. Was het niet door  $H_p$ , dan moet er een andere verklaring zijn
- ▶ Noem de (een?) alternatieve verklaring  $H_d$
- ▶ Nu zegt elementaire kansrekening dat

$$\frac{P(H_p | E)}{P(H_d | E)} = \frac{P(E | H_p)}{P(E | H_d)} \times \frac{P(H_p)}{P(H_d)}.$$

- ▶ “posterior odds=likelihood ratio  $\times$  prior odds”
- ▶ Idee: LR door forensisch laboratorium;  
kans(verhouding)en op hypotheses door beslisser  
(rechtbank).

# Sterkte bewijs versus sterke zaak

De *Likelihood Ratio* wordt door het forensische laboratorium berekend en doorgegeven. De voordelen zijn:

- ▶ De kansen om het bewijs  $E$  te zien worden berekend *uitgaande van* een verklaring  $H_p$  of  $H_d$ ; dus het lab hoeft niet te weten hoe waarschijnlijk deze zijn
- ▶ Het lab kan zich beperken tot het evalueren van welke verklaring beter is ( $H_p$  of  $H_d$ ) en hoeveel beter
- ▶ Scheiding tussen evaluatie bewijs an sich (LR; lab) en duiding in de context van de zaak (rechter; kansverhouding hypothesen)
- ▶ De LR geeft aan hoe sterk het bewijs is; de posterior odds hoe sterk de zaak is.

- ▶ Voor de berekening van de LR kan vaak een statistisch model worden gebruikt (b.v. voor de analyse van DNA-sporen); dus goed verankerd in klassieke opvatting van kansen
- ▶ De kansen op de hypothesen  $H_p$  en  $H_d$  drukken eerder de onzekerheid van een waarnemer uit die bepaalde informatie tot zich heeft genomen
- ▶ Dit valt onder de subjectieve interpretatie van kansen; lastig begrip al voor wiskundigen, laat staan voor juristen

Stel  $LR = 10.000$  voor  $H_p$ : verdachte liet spoor achter,  
versus  $H_d$ : nietes.

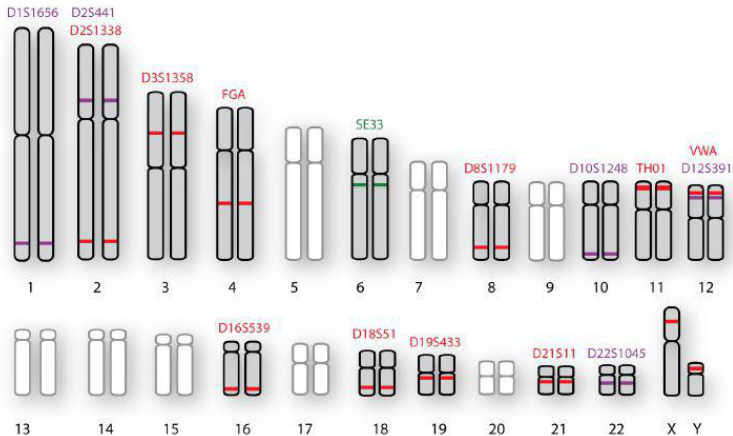
## Prosecutor's fallacy

- ▶ Begrepen als: de kans(verhouding) is 1 op 10.000 dat de verdachte het spoor niet achter liet.
- ▶ Fout: herkomstuitspraak kan niet zonder context (= prior) dus er is kennelijk impliciet een prior gebruikt
- ▶ Die is 50%: dan is de uitspraak correct.

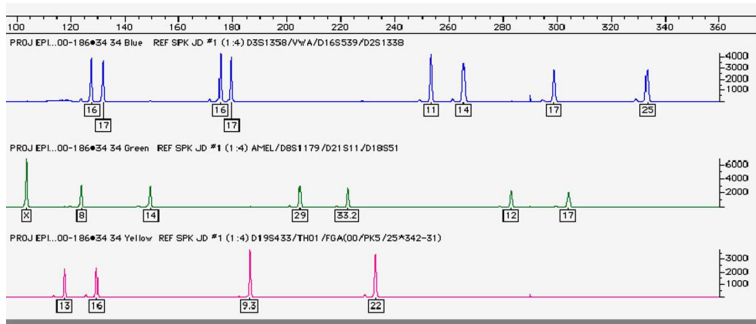
# Forensische DNA profielen

- ▶ Op sommige plekken (*loci*) op het niet-coderende DNA, wordt een 'woord' (bv. ATAG) een variabel aantal keren herhaald
- ▶ Forensische laboratoria meten hoeveel herhalingen iemand op zijn of haar chromosomen heeft, voor een aantal van dat soort loci.
- ▶ Per locus dus twee varianten, want elk chromosoom komt in tweevoud voor (gekregen van vader en moeder)
- ▶ Genotype (12, 15) betekent dat op één van de chromosomen 12 herhalingen zitten en op de andere 15 (geen ordening; je ziet niet welke op het paternale of maternale chromosoom zat)
- ▶ Het DNA profiel is de collectie genotypes ( $a, b$ )
- ▶ Standaard: 15 loci (+ bepaling geslacht X/X of X/Y)
- ▶ Komt eraan in 2018: 23 autosomaal + geslacht + 3 Y-chromosomale loci

# Veelgebruikte autosomale loci



# Voorbeeld van een DNA profiel



AMEL	D3	vWA	D16	D2	D8	D21	D18	D19	TH01	FGA
X,X	16, 17	16, 17	11, 14	17, 25	8, 14	29, 33.2	12, 17	13, 16	9.3, 9.3	22, 22

## Sporen

Sporen bevatten vaak DNA van meerdere personen.

- ▶ Soms allen onbekend
- ▶ Soms slachtoffer + dader(s)

en de LR die je dan voor een verdachte  $S$  berekent is voor de hypotheses

$H_p$ : verdachte  $S$  is een van de DNA-donoren

$H_d$ : geen van de DNA-donoren is (verwant aan) verdachte.

dus voor mengsel  $M$

$$LR(M, S) = \frac{P(M | H_p)}{P(M | H_d)}.$$

# p-waarden voor mengsels

Stel  $LR(M, S) = t$ , b.v.  $t = 1000$ .

## Bayes

Dit betekent dat

- ▶ De kans om  $M$  te krijgen  $t$  keer groter is als  $S$  heeft bijgedragen, dan als dat niet zo is
- ▶ De kansverhouding tussen de kansen dat  $S$  wel, c.q. niet, heeft bijgedragen,  $t$  keer zo groot is geworden als voorafgaand aan de DNA-typering

## p-waarde

Je kan ook uitrekenen wat de kans  $P(LR(M, S) \geq t \mid H_d)$  is: dit is de kans dat tegen een onschuldige verdachte bewijs zal worden gevonden dat minstens even sterk is als in de onderhavige zaak.

# Wat te rapporteren? LR's..

## ISFG

- ▶ ISFG: International Society for Forensic Genetics (belangrijkste vakorganisatie)
- ▶ Over de interpretatie van DNA-mengsels:  
*Recommendation 1: The likelihood ratio is the preferred approach to mixture interpretation.*
- ▶ *Recommendation 2: Even if the legal system does not implicitly appear to support the use of the likelihood ratio, it is recommended that the scientist is trained in the methodology and routinely uses it in case notes, advising the court in the preferred method before reporting the evidence in line with the court requirements. The scientific community has a responsibility to support improvement of standards of scientific reasoning in the court-room*

## Oppositie

- ▶ “Likelihood ratio moeilijk uit te leggen en te begrijpen ...”
- ▶ Peter Gill and Haned (2012): “... there is no reason why the performance test itself could not be used *instead* of the LR statistic. But this debate is reserved for future work.” (italics in original)
- ▶ Mitchell et al.(2014) volgen deze suggestie en stellen voor de sterkte van DNA-bewijs aan rechters and jury's voor te leggen in de vorm van een  $p$ -waarde.
- ▶ (Perlin 2017) “While the LR summarizes evidence, the RMP (=  $p$ -waarde) estimates error. Both statistical measures assist a trier of fact in understanding DNA evidence.”

# Relatie tussen $LR$ en $p$ -waarde

## Ongelijkheid

Het is niet zo moeilijk om in te zien dat

$$p - \text{waarde} \leq 1/LR,$$

in alle situaties. Dus de  $p$ -waarde zal altijd een - op het eerste gezicht - overtuigender bewijswaarde leveren.

## Reden

- ▶ Er geldt dat  $P(LR = x \mid H_p) = xP(LR = x \mid H_d)$
- ▶ Dus de kans dat een onschuldige verdachte toch een LR van  $t$  krijgt in het voordeel van DNA-donor zijn, is *precies*  $t$  keer zo klein als voor een echte DNA-donor
- ▶ De kans dat je een LR van  $x > t$  hebt als onschuldige verdachte, is eveneens  $x$  keer kleiner dan voor een echte DNA-donor, dus minder dan  $t$  keer kleiner.

## Interpretatie: makkelijk?

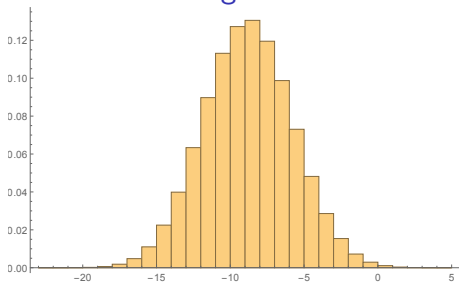
- ▶ Stel, dat de kans op een LR van 1000 of meer, voor een willekeurige persoon gelijk is aan 1 op 100.000
- ▶ Dat betekent dat als ik een willekeurige persoon neem, en ik bereken de LR, ik een kans heb van 1 op 100.000 dat ik minstens 1000 krijg
- ▶ Maar dat is niet wat we willen weten!!
- ▶ De vraag is: wat zegt een LR van 1000?
- ▶ De conclusie dat de kans 1 op 100.000 is, dat het om een onschuldige verdachte gaat, is een soort prosecutor's fallacy
- ▶ Je kan eenvoudigweg *niets* zeggen over de kans dat de verdachte het spoor heeft achtergelaten, ook al is de  $p$ -waarde nog zo klein
- ▶ LR's worden inderdaad vaak verkeerd begrepen. Maar de prosecutor's fallacy waar ze toe kunnen leiden, wordt alleen maar erger met  $p$ -waardes.

# Verwerpen, maar dan?

- ▶ Zelfs als we mee zouden gaan in het verwerpen van de hypothese  $H_d$  waaronder de verdachte niet heeft bijgedragen, wat geloven we dan wel?
- ▶ De precieze formulering van  $H_d$  is dat aan het mengsel geen aan verdachte verwante persoon heeft bijgedragen
- ▶ De verwerping hiervan houdt geenszins in dat we zouden moeten besluiten dat verdachte dan zelf moet hebben bijgedragen

# $p$ -waarden kunnen bewijs suggereren dat er niet is

Verdeling van LR van willekeurige personen voor een fictief DNA-mengsel



## $p$ -waarde versus LR

- ▶ Bij een  $LR = 1$  hoort een  $p$ -waarde van 0.002
- ▶ Bij een  $LR = 1/10$  hoort een  $p$ -waarde van 0.004
- ▶ Bij een  $LR = 1/100$  hoort een  $p$ -waarde van 0.01
- ▶ Bij een  $LR = 1/1000$  hoort een  $p$ -waarde van 0.03

Locus	$G_1$		$G_2$	
D2S1338	22	23	17	23
D3S1358	17	18	16	18
FGA	22.2	24	22	23
D8S1179	14	17	12	14
TH01	8	8	8	9
VWA	18	18	17	18
D16S539	12	12	12	13
D18S51	13	16	13	16
D19S433	13.2	16.2	14	14
D21S11	30.2	35.2	29	31.2

Table: Genotypes  $G_1$  and  $G_2$  van personen 1 and 2.

$H_1$  : persoon 1 en persoon 2 zijn broers,

$H_2$  : persoon 1 en persoon 2 zijn niet verwant.

In de LR kan je conditioneren op 1 van de personen, als je wilt:

$$LR_0 = \frac{P(G_1, G_2|H_1)}{P(G_1, G_2|H_2)}, \quad LR_1 = \frac{P(G_2|G_1, H_1)}{P(G_2|G_1, H_2)}, \quad LR_2 = \frac{P(G_1|G_2, H_1)}{P(G_1|G_2, H_2)},$$

numeriek identiek, en (in dit geval) gelijk aan bijna precies 1.

Maar...

$$p_0 := P(LR_0(G_1, G_2) \geq 1 | H_2) = 0.035,$$

$$p_1 := P(LR_1(G_2) \geq 1 | H_2) = 0.008,$$

$$p_2 := P(LR_2(G_1) \geq 1 | H_2) = 0.050.$$

Het verschil komt doordat persoon 1 relatief zeldzame DNA-kenmerken heeft en persoon 2 juist vrij algemene. De profielen ondersteunen geen hypothese meer dan de andere, aangezien de LR ongeveer 1 is. Toch zijn de  $p$ -waarden klein, hetgeen tot de (onjuiste) gevolgtrekking kan leiden dat het DNA de hypothese ondersteunt dat het hier om broers gaat.

# De $p$ -waarde hangt van meer af dan wat er gezien is

## Voorbeeld

Stel dat persoon 1 genotype  $(a, b)$  en persoon 2  $(a, c)$  heeft. De LR voor het zijn van broers is hangt alleen af van de frequentie van het gemeenschappelijke allel  $a$ . Stel dat  $p_a = 1/7$  en dit gebeurt op 15 loci, dan is de LR gelijk aan ongeveer 6.

## $p$ -waardes

Hangen wel van de frequenties af!

- ▶ Als er 4 allelen zijn met frequentie  $(1/7, 1/7, 1/7, 4/7)$  is de  $p$ -waarde  $P(LR > 6 \mid H_d)$  gelijk aan 0.02
- ▶ Als er 7 allelen zijn elk met frequentie  $1/7$ , is de  $p$ -waarde  $P(LR > 6 \mid H_d)$  gelijk aan 0.004
- ▶ Als er 103 allelen zijn met frequenties  $(1/7, 1/7, 1/7, 4/700, \dots, 4/700)$  is de  $p$ -waarde gelijk aan 0.00007.

1. p-values zijn niet eenvoudig te begrijpen en te interpreteren
2. Hoe combineer je p-waardes van verschillende bewijsstukken ?

## LR

De berekening van de LR is het minst problematische deel van de evaluatie van forensisch bewijs. Vaak kan dit aan de hand van betrouwbare modellen, vooral bij DNA-bewijs.

## Bewijswaarde

- ▶ De LR geeft aan hoe de kansen op de hypothesen moeten worden bijgesteld en is daarom de relevante grootte voor een besliser.
- ▶ De LR dwingt je om het bewijs hoe dan ook te verklaren en dat is een vereiste, omdat het zich heeft voorgedaan. Dat het onwaarschijnlijk is onder een bepaald scenario zegt niks. Misschien is het wel onwaarschijnlijk onder elk scenario

## Toepassing Bayes

De toepassing van de regel van Bayes in de rechtspraak is echter niet eenvoudig omdat

- ▶ A priori kansen bijna niet zinvol vast te stellen zijn
- ▶ Juristen er niet voldoende aan gewend zijn en interpretatiefouten op de loer liggen
- ▶ Combineren van verschillende LR's zonder extra informatie over onafhankelijkheid niet kan.

## p-waardes

p-waardes werken zonder prior en lossen dus 1 probleem op.  
Echter,

- ▶ Ze zijn misschien nog wel lastiger te begrijpen dan LR's
- ▶ Ze werken onder 1 hypothese, en zijn daarmee voorspellend en niet verklarend
- ▶ Ze zijn vaak op meerdere manieren te definiëren en daarmee ambigu
- ▶ Ze stellen het bewijs sterker voor dan het is
- ▶ Ze hangen af van data die we niet gezien hebben

en daarom zijn p-waardes niet geschikt voor het overbrengen van de sterkte van forensisch bewijs.

## Gedeeltelijk gebaseerd op:

Maarten Kruijver, Ronald Meester, Klaas Slooten, *p-Values should not be used for evaluating the strength of DNA evidence*, Forensic Science International: Genetics 16 (2015), 226–231.

## Vragen?

k.slooten@vu.nl