

Statistische paradoxen in de rechtszaal - opdrachten

Charlotte Vlek

www.charlottevlek.nl

c.s.vlek@rug.nl

1 februari, 2014

Opdracht 1: de regel van Bayes

De regel van Bayes is

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Leid de regel van Bayes af uit de definitie van voorwaardelijke kansen $P(A|B) = \frac{P(A,B)}{P(B)}$ (tip: $P(A, B) = P(B, A)$).

Opdracht 2: groene en blauwe taxi's

Een voetganger werd aangereden door een taxi. De taxi reed vervolgens door. In het stadje waar dit plaatsvond zijn slechts twee taxicentrales: een waar alle taxi's blauw zijn, en een waar alle taxi's groen zijn. De verdeling is als volgt:

- 85 taxi's zijn groen
- 15 taxi's zijn blauw

1. Hoe groot is de a priori kans $P(H)$ dat een blauwe taxi verantwoordelijk is voor de aanrijding?

Er heeft zich inmiddels een getuige gemeld. Deze getuige zegt dat hij een blauwe taxi weg zag rijden. De politie deed een aantal tests met deze getuige onder dezelfde omstandigheden, om te bepalen hoe betrouwbaar dit bewijs was. Daaruit bleek dat de getuige in 80% van de gevallen een blauwe taxi correct identificeerde als blauw, en een groene correct als groen. In 20% van de gevallen ging het mis. We hebben dus de volgende tabel (met H : de taxi verantwoordelijk voor de aanrijding was blauw, $\neg H$: de taxi verantwoordelijk voor de aanrijding was groen en E : de getuige zegt een blauwe taxi te hebben gezien).

	H	$\neg H$
E	0,8	0,2
$\neg E$	0,2	0,8

2. De kans dat de hypothese H en het bewijs E zich tegelijk voordoen, wordt gegeven door $P(E, H)$. Dit is dus geen voorwaardelijke kans, maar het kan wel berekend worden op basis van de voorwaardelijke kans $P(E|H)$. Gebruik de definitie van voorwaardelijke kansen om $P(E, H)$ uit te drukken in de voorwaardelijke kans $P(E|H)$ en de a priori kans $P(H)$. Bereken $P(E, H)$.
3. Bereken ook $P(E, \neg H)$.
4. De kans $P(E)$ is gelijk aan de kans dat E en H tegelijk voorkomen, plus de kans dat E en $\neg H$ tegelijk voorkomen. Oftewel: $P(E) = P(E, H) + P(E, \neg H)$. Dit geldt omdat H en $\neg H$ samen alle gevallen beschrijven. Bereken de kans $P(E)$.

5. Bereken de kans $P(H|E)$ dat een blauwe taxi verantwoordelijk is voor de aanrijding gegeven de getuigenis. Gebruik de regel van Bayes:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}.$$

6. Bereken ook de kans $P(\neg H|E)$. Welk taxibedrijf moet een boete krijgen?

Opdracht 3: de Adams-zaak

In 1991 vond in Hemel Hempstead een verkrachting plaats. In 1993 werd Denis John Adams gearresteerd voor een ander misdrijf en bij een routine-check bleek zijn DNA te matchen met DNA in sperma gevonden bij het slachtoffer van de verkrachting in 1991. Deze zaak lijkt op de Vaatstra-zaak in de zin dat de verdachte voorafgaand aan de DNA-test geen bijzonder geval was. We definiëren H en E als volgt:

- H is de hypothese dat Adams schuldig is
- E is het DNA-bewijs

We hebben de volgende gegevens:

- De kans dat het DNA-bewijs gevonden zou zijn terwijl Adams onschuldig is (de random match probability), is 1 op 2 miljoen.
- De kans dat het DNA-bewijs er is op voorwaarde dat Adams schuldig is, is 1.
- Er wonen 300.000 mannelijke volwassenen in Hemel Hempstead, en er zouden nog eens 100.000 andere volwassen mannen op het moment van de verkrachting in de omgeving kunnen zijn geweest. Adams is dus een van de 400.000 potentiële verkrachters.

Beantwoord de volgende vragen:

1. Een rechter concludeert: de kans dat Adams onschuldig is, is 1 op 2 miljoen. Wat gaat hier mis?
2. Wat is de a priori kans $P(H)$, dat Adams schuldig is?
3. Neem onderstaande tabel over en vul die in met de juiste kansen.

	H	$\neg H$
E	...	...
$\neg E$	...	...

4. Bereken de kans $P(E)$ op het vinden van bewijsstuk E . Gebruik $P(E) = P(E, H) + P(E, \neg H)$.
5. Bereken de a posteriori kans $P(H|E)$ dat Adams schuldig is gegeven het DNA-bewijs. Gebruik de regel van Bayes.
6. Is het DNA-bewijs voldoende om overtuigend te concluderen dat Adams schuldig is?

Opdracht 4: de Adams-zaak in een Bayesiaans netwerk

We kijken nog eens naar de Adams-zaak, ditmaal met meerdere variabelen. De (on)afhankelijkheden zijn gemodelleerd in Figuur 1. De tabellen in 1 geven de bijbehorende conditionele kansen.

1. Bekijk de conditionele kanstabellen van het Bayesiaanse netwerk. Leg uit welke kans de 0 in de onderste tabel (voor de knoop Getuigenis) voorstelt.
2. Waarom is deze kans op 0 geschat?

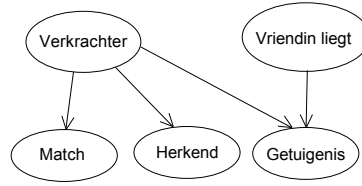


Figure 1: Een Bayesiaans netwerk voor de Adams-zaak

Verkrachter = waar	0,0000025	Vriendin liegt = waar	0,4
Verkrachter = onwaar	0,9999975	Vriendin liegt = onwaar	0,6
	Verkrachter = waar	Verkrachter = onwaar	
Match = waar	1	0,0000005	
Match = onwaar	0	0,9999995	

	Verkrachter = waar	Verkrachter = onwaar
Herkend = waar	0,9	0,1
Herkend = onwaar	0,1	0,9

	Verkrachter = waar		Verkrachter = onwaar	
	Vr l = waar	Vr l = onwaar	Vr l = waar	Vr l = onwaar
Getuigenis = waar	0,99	0	0,1	0,9
Getuigenis = onwaar	0,01	1	0,9	0,1

Table 1: Tabellen bij het netwerk voor Adams

3. Hoe groot is de kans op de getuigenis (de vriendin getuigt: Adams was bij haar die avond) als Adams de verkrachter was, en de vriendin liegt?

Berekeningen binnen een Bayesiaans netwerk kunnen gebruik maken van onafhankelijkheden in het netwerk. Daartoe kijken we naar de collectie ouders van een knoop X . De ouders (genoteerd als $\text{pa}(X)$) zijn alle knopen Y die met een pijl verbonden zijn met X in de richting $Y \rightarrow X$. Kortom: $\text{pa}(X) = \{Y | Y \rightarrow X\}$. Om de kans op X te berekenen hoeven we alleen te kijken naar alle voorouders van X : dit zijn de ouders van ouders van ouders (etc) van X . Voor een netwerk met knopen $A_1, \dots, A_n$ (over het domein $U = \{A_1, \dots, A_n\}$) vinden we de gezamenlijke kansverdeling als volgt:

$$P(U) = \prod_{i=1}^n P(A_i | \text{pa}(A_i))$$

Voor de kans $P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{waar}, \text{Vriendin liegt} = \text{waar})$ vermenigvuldigen we de (conditionele) kansen van deze knopen, maar hoeven we de knopen *Herkend* en *Match* niet mee te nemen, omdat het geen ouders zijn van *Getuigenis*, *Verkrachter* of *Vriendin liegt*:

$$\begin{aligned}
 &P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{waar}, \text{Vriendin liegt} = \text{waar}) \\
 &= P(\text{Getuigenis} = \text{waar} | \text{Verkrachter} = \text{waar}, \text{Vriendin liegt} = \text{waar}) \\
 &\cdot P(\text{Verkrachter} = \text{waar}) \cdot P(\text{Vriendin liegt} = \text{waar})
 \end{aligned}$$

Als we niets weten over de waarde van Verkrachter en Vriendin liegt, dan tellen we de kansen voor alle mogelijke waarden van die twee variabelen bij elkaar op:

$$\begin{aligned}
 &P(\text{Getuigenis} = \text{waar}) \\
 &= P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{waar}, \text{Vriendin liegt} = \text{waar}) \\
 &+ P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{waar}, \text{Vriendin liegt} = \text{onwaar}) \\
 &+ P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{onwaar}, \text{Vriendin liegt} = \text{waar}) \\
 &+ P(\text{Getuigenis} = \text{waar}, \text{Verkrachter} = \text{onwaar}, \text{Vriendin liegt} = \text{onwaar})
 \end{aligned}$$

4. Hoe groot is de kans op de getuigenis als Adams de verkrachter was (zonder verdere informatie over de vriendin)?
5. Uiteindelijk zijn we natuurlijk geïnteresseerd in de kans dat Adams de verkrachter was, gegeven het bewijs. In opdracht 3 hebben we de kans $P(\text{Verkrachter} = \text{waar} | \text{Match} = \text{waar}) \approx 0,84$ berekend. Wat gebeurt er met deze kans nu we een Bayesiaans netwerk hebben met meer variabelen?
6. Behalve de DNA match zijn er nog twee andere bewijsstukken: Adams is niet herkend door het slachtoffer (Herkend = onwaar) en de vriendin van Adams getuigde dat hij bij haar was die avond (Getuigenis = waar). De regel van Bayes kunnen we als volgt generaliseren voor meerdere bewijsstukken:

$$P(H|E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3|H) \cdot P(H)}{P(E_1, E_2, E_3)}$$

Schrijf uit hoe we hiermee de kans kunnen berekenen dat Adams de verkrachter was gegeven de drie bewijsstukken. Optioneel (voor liefhebbers met een rekenmachine bij de hand): bereken deze kans.

Opdracht 5: De verklaring van een verdachte

In de Anjumer moordzaak werden twee lijken gevonden op het terrein van een pension in het Friese Anjum. Eigenaresse van het pension, Marjan van der E. is veroordeeld voor beide moorden. Het politieonderzoek begon nadat Beekman, de buurman van Marjan, een verklaring aflegde dat Marjan tegen hem had gezegd dat ze iemand had vermoord: Leo de Jager, om precies te zijn. In deze opgave houden we ons alleen bezig met deze ene moord op Leo.

Gedurende het politieonderzoek kwam er allerlei informatie boven water waaruit vrij duidelijk bleek dat Leo was gedood in het pension, doordat hij met een hamer op zijn hoofd was geslagen. Er was echter weinig bewijs dat duidelijk een verdachte aanwees. De verklaring van Beekman was een van de weinige bewijsstukken dat direct Marjan aanwees als verdachte. Er was echter nog een andere aanwijzing: het lijk van Leo lag buiten, terwijl er allerlei bloedsporen werden gevonden in de gang. Als Marjan Leo had gedood in de gang en helemaal alleen naar buiten had gemanoeuvreerd, dan zou er vrijwel zeker een bloederig sleepspoor in de gang te vinden zijn. Als zij niet degene was die Leo gedood had, dan was er misschien geen sleepspoor geweest (bijvoorbeeld omdat de aanvaller sterk genoeg was om Leo te tillen in plaats van te slepen).

Deze informatie kunnen we modelleren in een Bayesiaans netwerk met knopen voor Marjan als aanvaller ('Marjan'), het sleepspoor en de verklaring van Beekman, zie Figuur 2a. Echter, wanneer het Beekman zelf was die Leo heeft gedood, zou dit ook kunnen verklaren waarom hij Marjan beschuldigt: hij wil de aandacht op haar vestigen zodat hij hopelijk zelf niet verdacht wordt. Daarom nemen we de mogelijkheid dat Beekman het gedaan heeft mee in het netwerk (een extra knoop 'Beekman'), zie Figuur 2b. In deze vraag zullen we zien dat het sleepspoor een cruciale rol speelt.

De kanstabellen bij het netwerk in Figuur 2b staan in Tabel ??.

1. Het politieonderzoek begon met de verklaring van Beekman dat Marjan hem had verteld dat ze Leo had vermoord. Hoe groot is de kans dat Marjan het gedaan heeft, gegeven deze verklaring van Beekman?
2. Hoe groot is de kans dat Beekman het gedaan heeft, gegeven zijn eigen verklaring?

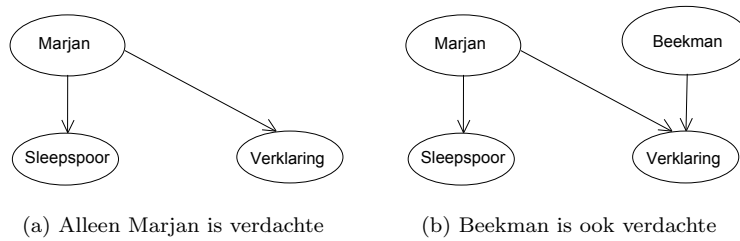


Figure 2: Netwerken voor de Anjumer moordzaak

Marjan = waar	0,0001	Beekman = waar	0,0001
Marjan = onwaar	0,9999	Beekman = onwaar	0,9999
	Marjan = waar	Marjan = onwaar	
Sleepspoor = waar	0,99	0,5	
Sleepspoor = onwaar	0,01	0,5	

	Beekman = waar		Beekman = onwaar	
	Marjan = waar	Marjan = onwaar	Marjan = waar	Marjan = onwaar
Verklaring = waar	0,1	0,6	0,99	0,001
Verklaring = onwaar	0,9	0,4	0,01	0,999

Table 2: Kanstabellen bij de Anjumer moordzaak

- Met alleen de verklaring van Beekman, wie is dus de meest voor de hand liggende verdachte?
- Er wordt geen sleepspoor gevonden in het pension. Overige sporen zijn overduidelijk, dus we kunnen concluderen dat het sleepspoor er echt niet is geweest. Hoe groot is nu de kans dat Marjan het heeft gedaan? En Beekman?