

Kansrekenen: Beliefs & Bayes

L. Schomaker, juni 2001

- Bereik van kansen

$$0 \leq P(A) \leq 1 \quad (1)$$

- Kansen op valide en onvervulbare proposities

$$P(Waar) = 1, P(Onwaar) = 0 \quad (2)$$

- Somregel

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) \quad (3)$$

De kans op A of B, disjunctief, is gelijk aan de kans op A plus de kans op B, verminderd met de kans op A en B, conjunctief.

- Productregel

$$P(A \wedge B) = P(A|B) * P(B) \quad (4)$$

$$= P(B|A) * P(A) \quad (5)$$

De kans op A en B, conjunctief, is gelijk aan de kans op A, gegeven B, vermenigvuldigd met de kans op B, en tevens gelijk aan de kans op B, gegeven A, vermenigvuldigd met de kans op A.

- Kettingregel

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \quad (6)$$

Dit is de productregel, herschreven. We kunnen deze gebruiken om de kans op conjuncties te berekenen. Voorbeeld: gegeven $P(A) = 0.5$, $P(B|A) = 0.6$, $P(C|A \wedge B) = 0.8$, wat is dan $P(A \wedge B \wedge C)$? Antwoord:

$$\begin{aligned} P(A \wedge B) &= P(B|A) * P(A), \\ P(A \wedge B \wedge C) &= P(C|A \wedge B) * P(A \wedge B) \Rightarrow \\ P(A \wedge B \wedge C) &= P(C|A \wedge B) * P(B|A) * P(A) = 0.8 * 0.6 * 0.5 \end{aligned}$$

- Kanstotaal

Gegeven n exclusieve gebeurtenissen A_i met $\sum_{i=1}^n P(A_i) = 1$, dan geldt, voor het geval ik de A_i beschouw als gegevens voor het observeren van een gebeurtenis B :

$$P(B) = \sum_{i=1}^n P(B|A_i) * P(A_i) \tag{7}$$

Voorbeeld: de kans dat de zon schijnt als Piet lacht is 0.3, de kans dat de zon schijnt als Mia lacht is 0.4. De kans dat Piet zomaar lacht is 0.1. De kans dat Mia zomaar lacht is 0.2. Ze lachen nooit tegelijk. De kans dat de zon schijnt is: $0.3 * 0.1 + 0.4 * 0.2 = 0.03 + 0.08 = 0.11$. Zoek uit: wat zijn in dit voorbeeld de proposities B , A_1 en A_2 , $B|A_1$ en $B|A_2$?

De volgende figuur (Fig. 1) geeft een Venn-diagram van $P(B)$ berekening uit de conditionele kansen.

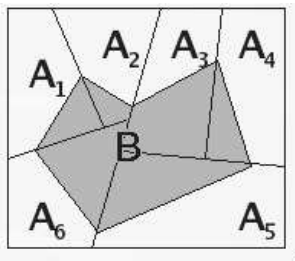


Figure 1: De totaalkans $P(B)$ is de som van de samenstellende conditionele kansen op de bewijzen A_i , gewogen met de kans op elk bewijs.

- Bayes regel

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)} \tag{8}$$

in woorden:

$$posterior = \frac{prior * likelihood}{evidence}$$

De a posteriori kans (dwz, de kans op een feit A , gegeven een bewijs B) is dus een weging van de a priori kans op feit A met het quotient van de voorwaardelijke kans op het bewijs en de a priori kans op het bewijs. Merk op dat de teller van de Bayes formule (Verg. 8) gelijk is aan de formule voor $P(A \wedge B) = P(A) * P(B|A)$ (Verg. 4). We kunnen dus de Bayes formule herschrijven net als Verg. 6:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

die nog eens duidelijk maakt dat de Bayes kans in feite de genormaliseerde kans op een conjunctie $A \wedge B$ is, dwz. genormaliseerd voor het feit dat één van de twee gegevens (B) al bekend is ('evidence').

Een medisch-diagnostisch voorbeeld

Voorbeeld. S = stijve nek, M =Meningitis. Wat is de kans dat ik meningitis heb wanneer ik last heb van een stijve nek? Hiervoor moeten we weten: de kans op een stijve nek (algemeen), $P(S) = 1/20$, de kans op meningitis (algemeen), $P(M) = 1/5000$, en de kans dat ik een stijve nek heb wanneer ik meningitis heb: $P(S|M) = 0.5$.

Antwoord:

$$\begin{aligned}P(M|S) &= P(M) * \frac{P(S|M)}{P(S)} \\ &= 0.0002 * 0.5/0.05 = 0.002 \\ &= 1/500\end{aligned}$$

Dit illustreert de essentie van Bayes: al is het vrij waarschijnlijk dat ik een stijve nek voel wanneer ik meningitis heb, dan kan het omgekeerde, de kans dat ik meningitis heb wanneer ik een stijve nek heb veel minder waarschijnlijk zijn. Zelfs in het geval $P(S|M) = 1$ (altijd een stijve nek bij meningitis) geldt dat de kans op meningitis bij een stijve nek zeer laag is:

$$\begin{aligned}P(M|S) &= \frac{P(M) * 1}{P(S)} \\ &= 0.0002/0.05 \\ &= 0.004 \\ &= 1/250\end{aligned}$$

NB: $P(S|M) \leq \text{Min}(1, P(S)/P(M))$

Nog een medisch-diagnostisch voorbeeld

De kans op een niersteen (N) is 0.01. De kans dat de urinetest (U) positief is als je een niersteen hebt is 0.80. De kans dat de urinetest vals alarm slaat is 0.10. Wat is de kans dat je een niersteen hebt? Het verschil met het vorige voorbeeld is dat we $P(U)$, de kans dat de urinetest 'ja' zegt niet weten, we weten alleen de kans op vals alarm $P(U|\neg N)$.

$$\begin{aligned}P(N|U) &= P(N) * \frac{P(U|N)}{P(U)} \\ &= \frac{P(N) * P(U|N)}{P(N) * P(U|N) + P(\neg N) * P(U|\neg N)} \\ &= 0.01 * 0.80 / (0.01 * 0.80 + 0.99 * 0.10) = 0.075\end{aligned}$$

(Opdracht: bewijs dat $P(U) = P(N) * P(U|N) + P(\neg N) * P(U|\neg N)$)

Hetzelfde maar dan anders ...

Vaak wordt het gemakkelijker gevonden om met frequenties (geturfde aantallen) te werken. Gemiddeld tien van de 1000 mensen hebben een niersteen. Als 8 van die 10 mensen met een niersteen een positieve test opleveren, terwijl er van de overgebleven 990 toch nog 99 mensen zijn waarop de urinetest vals alarm geeft, dan reduceert bovenstaande probabilistische notatie tot de volgende eenvoudige frequentistische notatie voor de kans dat ik een niersteen heb gegeven een positieve urinetest:

$$\begin{aligned}P(N|U) &= \frac{f(\text{Urinetest : Ja : OK})}{f(\text{Urinetest : Ja : OK}) + f(\text{Urinetest : Ja : ValsAlarm})} \\ &= \frac{8}{8 + 99} \\ &= 0.075\end{aligned}$$

Dit lijkt voor mensen beter te begrijpen, maar het is niet altijd mogelijk of gemakkelijk om tellingen in plaats van kansen te gebruiken. In dit telvoorbeeld gebruiken we een enkele set van gegevens over 1000 mensen waarop de urinetest was uitgevoerd en waarvan op een andere (onbesproken) wijze vastgesteld was of ze de niersteen ook echt hadden:

$$\text{Allen} = 1000 \left\{ \begin{array}{l} \text{Ziek} = 10 \left\{ \begin{array}{l} \text{Urinetest : Ja : OK} = 8 \\ \text{Urinetest : Nee : Misser} = 2 \end{array} \right. \\ \text{Gezond} = 990 \left\{ \begin{array}{l} \text{Urinetest : Ja : ValsAlarm} = 99 \\ \text{Urinetest : Nee : OK} = 891 \end{array} \right. \end{array} \right.$$

Hoe gebruiken we Bayes om onze 'Beliefs' mbt de wereld aan te passen als er nieuwe feiten worden waargenomen?

Voorbeeld: tandproblemen. Heb je een gaatje ja of nee?

- Eerste observatie: tandpijn.

Gebruikmakend van Bayes (T='Tandpijn', G='Gaatje'):

$$P(G|T) = P(G) * \frac{P(T|G)}{P(T)} \quad (9)$$

- Tweede observatie: het haakje van de tandarts blijft ergens steken (H).

Een nieuwe observatie wordt als volgt meegenomen. Uitgaande van *Verg. 8* gebruiken we wat we al weten, nl. $P(G|T)$ als '*prior*' voor een tweede toepassing van de vergelijking, die uitgaat van een voorwaarde die een conjunctie is: $G|T \wedge H$. Het nieuwe bewijs 'H=haakje haakt' is (vooralsnog) op te vatten als voorwaardelijk afhankelijk van de geconstateerde tandpijn, dus we hebben als *evidence* in de noemer nodig: $P(H|T)$. Voor de *likelihood* moeten we ons realiseren dat we zoeken naar de kans op haken, gegeven de gezochte conditie (G) en wat we al weten (T), nl. $P(H|T \wedge G)$. Dus wordt de gecombineerde vergelijking:

$$P(G|T \wedge H) = P(G|T) * \frac{P(H|T \wedge G)}{P(H|T)} \quad (10)$$

$$= P(G) * \frac{P(T|G)}{P(T)} * \frac{P(H|T \wedge G)}{P(H|T)} \quad (11)$$

Dit levert op zich niet veel op omdat het in het algemeen moeilijk zal zijn om $P(H|T \wedge G)$ te kennen, d.i. de kans dat het haakje blijft steken terwijl ik tandpijn heb en er daadwerkelijk een gat in mijn tand zit. In andere woorden: elk nieuw bewijs is voorwaardelijk gerelateerd aan het reeds bekende bewijs. Dat wordt nogal omslachtig.

Als echter de bewijzen (T=Tandpijn en H='haakje haakt') voorwaardelijk **onafhankelijk** van elkaar zijn: $P(H|T \wedge G) = P(H|G)$ (of het haakje ergens blijft steken hangt niet af van hoe de patient zich voelt) en $P(T|H \wedge G) = P(T|G)$ (of ik tandpijn voel hangt niet af van de kans dat het haakje van de tandarts ergens blijft steken), dan kunnen we dit vereenvoudigen tot:

$$P(G|T \wedge H) = P(G) * \frac{P(T|G)}{P(T)} * \frac{P(H|G)}{P(H|T)} \quad (12)$$

We kunnen nog een stap verder gaan (het boek slaat dit over) en stellen dat als $P(H|T \wedge G) = P(H|G)$ (haken hangt alleen af van de aanwezigheid van een gaatje) dat het dan redelijk is om ook te stellen $P(H|T) = P(H)$.

$$P(G|T \wedge H) = P(G) * \frac{P(T|G)}{P(T)} * \frac{P(H|G)}{P(H)} \quad (13)$$

Hiermee komen we tot een meer algemene vergelijking (z.o.z)

Meer algemeen, de Bayes vergelijking voor het combineren van meervoudig bewijs:

$$P(Z|A, B, C, \dots) = \frac{P(Z) * P(A|Z) * P(B|Z) * P(C|Z) * \dots}{1 * P(A) * P(B) * P(C) * \dots} \quad (14)$$

waarin Z staat voor een conclusie (of 'Belief') en A, B, C, \dots voor de geconstateerde bewijzen. Vergeet niet dat de eis van onafhankelijkheid geldt, die in de realiteit lastig te garanderen valt.

De noemer in *Verg. 14* bestaat uit normalisatietermen, die kunnen worden ingevoerd omdat het gegevene (bijv. A) reeds is geconstateerd. De normalisatie zorgt ervoor dat de Bayes' of **a posteriori kans** als maximum de waarde 1 kan bereiken. We kunnen een normalisatiefactor α nemen om een vereenvoudigde versie van de (incrementele) Bayes regel op te stellen:

$$P(Z|A, B, C, \dots) = \alpha * P(Z) * P(A|Z) * P(B|Z) * P(C|Z) * \dots \quad (15)$$

Dit is met name praktisch als ik op basis van een aantal observaties A, B, C, \dots de meest waarschijnlijke consequentie Z_i uit N mogelijke consequenties wil trekken. Voor α nemen we:

$$\alpha = \frac{1}{\sum_{i=1}^N P(Z_i|A, B, C, \dots)} \quad (16)$$

zodat de som van de kansen op het trekken van de conclusies gelijk wordt aan 1. Het is namelijk zeker dat ik de conclusie Z_i trek die de hoogste Bayes kans op basis van de observaties heeft. We mogen deze normalisatie uitvoeren zonder gevolgen voor de te trekken conclusie omdat de a priori kansen op A, B etc. voor elk van de vergelijkingen dezelfde zullen zijn (als de wereld ondertussen niet verandert) voor elke te trekken conclusie Z_i . Na een dergelijke normalisatie is het echter niet mogelijk om de resulterende Bayes kans te relateren aan de fysieke wereld, of de sterkte van een 'Belief' te relateren aan geheel andere conclusies of 'Beliefs' die op andere feiten zijn gestoeld.

Als laatste is het nuttig om op te merken dat we in de Bayes vergelijking totnutoe een belangrijk aspect zijn vergeten, nl. dat er altijd een wereld of context bestaat. Daarom kan het belangrijk zijn om de propositie c voor 'context' als voorwaarde mee te nemen:

$$P(A|B, c) = \frac{P(A|c) * P(B|A, c)}{P(B|c)} \quad (17)$$

Voorbeeld

Onderstaand voorbeeld is illustratief omdat het contra-intuïtief is (het is bekend dat mensen zich over het algemeen niet volgens de optimale Bayes-strategie gedragen).

Stel, je doet mee aan een quiz. Er zijn drie deuren. Van tevoren is de prijs willekeurig achter één van de drie deuren geplaatst. Je kiest een deur door er voor te gaan staan. De quizmaster echter wil het spannend maken, gaat voor een van de andere deuren staan, en zegt: "deze is het NIET". Je mag 1x van deur wisselen. Vraag: wat is de optimale strategie, blijven staan of de overgebleven vrije deur kiezen?

Denk hierover goed na. Wat zou je doen?

Oplossing voor het quizmaster probleem

Het probleem is een typisch voorbeeld van een situatie waarin we Bayes goed kunnen gebruiken. Er is sprake van een a priori kans op een prijs: $P(\text{Prijs}) = 1/3$. Er komt echter nieuwe informatie binnen die je visie op de wereld verandert. Hoe integreer je de informatie die de quizmaster introduceert met wat je al weet?

Er zijn meerdere manieren om de situatie te interpreteren, maar volgens Bayes gaat het als volgt. We noemen de deur die jij hebt gekozen '1'. De deur die de quizmaster aanwijst noemen we '3'. De deur die je nog zou kunnen kiezen noemen we '2'. Realiseer je dat, nadat jij deur '1' hebt gekozen, de quizmaster nog uit twee deuren kan kiezen.

De kans dat de prijs achter *deur1* ligt, gegeven de kennis dat er achter *deur3* niets ligt:

$$P(\text{Prijsin1}|\text{ToontLeeg3}) = \frac{P(\text{Prijsin1}) * P(\text{ToontLeeg3}|\text{Prijsin1})}{P(\text{ToontLeeg3})} \quad (18)$$

De kans dat de prijs achter *deur2* ligt, gegeven de kennis dat er achter *deur3* niets ligt:

$$P(\text{Prijsin2}|\text{ToontLeeg3}) = \frac{P(\text{Prijsin2}) * P(\text{ToontLeeg3}|\text{Prijsin2})}{P(\text{ToontLeeg3})} \quad (19)$$

(probeer de twee laatste vergelijkingen in woorden uit te drukken!)

We weten a priori: $P(\text{Prijsin1}) = P(\text{Prijsin2}) = P(\text{Prijsin3}) = 1/3$. Verder weet jij dat de quizmaster na jouw keuze voor *deur1*, een keuze moet maken uit twee deuren. Dus: $P(\text{ToontLeeg3}) = 1/2 (= P(\text{ToontLeeg2}))$. Dit vullen we vast in:

$$P(\text{Prijsin1}|\text{ToontLeeg3}) = \frac{1/3 * P(\text{ToontLeeg3}|\text{Prijsin1})}{1/2} \quad (20)$$

$$P(\text{Prijsin2}|\text{ToontLeeg3}) = \frac{1/3 * P(\text{ToontLeeg3}|\text{Prijsin2})}{1/2} \quad (21)$$

We hebben nu alle a priori kansen ingevoerd op de juiste plek. Nu komt het lastigste: het bepalen van de voorwaardelijke kansen. Je moet denken vanuit het standpunt van de quizmaster.

Zou jij gekozen hebben voor *deur1* met de prijs erachter, dan heeft de quizmaster een vrije keuze uit de twee overgebleven deuren:

$P(\text{ToontLeeg3}|\text{Prijsin1}) = P(\text{ToontLeeg2}|\text{Prijsin1}) = 1/2$. Dus:

$$P(\text{Prijsin1}|\text{ToontLeeg3}) = \frac{1/3 * 1/2}{1/2} = 1/3 \quad (22)$$

Blijven staan levert je dus eenzelfde kans op, als bij het begin van het spel ($p=1/3$).

Heb jij echter niet gekozen voor de deur met de prijs erachter, dan moet de quizmaster zijn keuze laten bepalen door kennis: hij weet precies waar de prijs staat en kiest de overgebleven deur.

Dat wil zeggen: $P(\text{ToontLeeg3}|\text{Prijsin2}) = 1$.

Conclusie:

$$P(\text{Prijsin2}|\text{ToontLeeg3}) = \frac{1/3 * 1}{1/2} = 2/3 \quad (23)$$

Je kunt dus maar beter wisselen naar de overgebleven deur!