Meer weten, minder kansen

Jean Paul Van Bendegem

Aanleiding

In dit kort stukje wil ik een probleem aankaarten in verband met waarschijnlijkheden en kansen. We weten allemaal, dankzij de ondertussen ontelbare studies en onderzoeken (ik bespaar de lezer de nodige referenties), dat wij mensen niet echt geschikt zijn voor het inschatten van waarschijnlijkheden. Het gaat vaak samen met de nodige en interessante verklaringen vanuit onze evolutionaire voorgeschiedenis. De voorbeelden die hierbij ter sprake komen zijn ons allen zeer wel bekend, om niet te zeggen té bekend. Keer op keer krijgen we de Lotto voorgeschoteld, het verjaardagsexperiment of het Monty Hall probleem, ook bekend als het drie deuren probleem. Ik mag de lezer verwijzen naar mijn bijdrage . “Waarlijk, wonder en is gheen wonder” in dit tijdschrift, 2de jaargang, nummer 3, pp. 7-10 in 2002. Het zijn en blijven uiteraard zeer geschikte voorbeelden, maar door steeds maar deze voorbeelden te hernemen loopt men het risico de indruk te wekken dat er geen andere zouden te vinden zijn. Impliciet geeft men daarbij aan dat deze gevallen eerder uitzonderingen zijn dan de regel en in die zin naast ons mogen worden neergelegd. Anders gezegd, het blijft zeer belangrijk om nieuwe of toch minder gekende voorbeelden te vinden, wat de bescheiden bedoeling van dit stukje is. Hierbij zo’n geval. Ik moet eerlijk bekennen dat ik nog nooit zo’n contra-intuïtief voorbeeld heb meegemaakt; dit gaat werkelijk voor de volle honderd procent in tegen mijn wereldbeeld. In die mate zelfs dat ik neig naar de gedachte dat onze intuïties best wel in orde zijn, maar dat simpelweg de wereld een heel eigenaardige plek is voor nuchtere wezens zoals wij. Maar dat is een andere discussie. Hier gaat het om het voorbeeld dat ik probeer zo exact mogelijk te presenteren, anders is er geen lezer die mij gelooft.

De situatie is de volgende: er zijn drie mensen A, B en C, die in running zijn voor een prijs die door een jury zal toegekend worden. Wat geweten is, is de kans dat ze de prijs binnenhalen: voor A is dat 1/4, voor B 1/2 en voor C 1/4. De jury heeft beraadslaagd, maar nog niets bekend gemaakt. A vraagt aan een jurylid dat hij kent of hij kan zeggen wie van B of C het *niet* heeft gehaald. Daarmee is niets gezegd over A zelf en uiteindelijk weet je dat één van die twee het toch niet haalt, dus echt verrassend kan dat nieuws niet zijn. Het jurylid is het hiermee eens en deelt mee dat C het niet heeft gehaald. A haalt opgelucht adem, want hij meent dat daardoor zijn kansen zijn verbeterd, vermits alleen hij en B nog in competitie zijn. Maar heeft A gelijk om zo optimistisch te zijn? Het antwoord is neen: door dit bericht is de kans van A om de prijs te winnen gezakt van 1/4 naar een 1/5. Jawel, *gezakt*!

Een informele redenering

Neem het standpunt van A in. Wat betekent deze informatie voor hem? Dat er maar twee scenario’s overblijven: ofwel heeft A de prijs gewonnen ofwel heeft B de prijs gewonnen.

Scenario 1: *A* heeft de prijs gewonnen met een kans van 1/4, zoals gegeven. Dat wil zeggen dat het jurylid de keuze heeft om *B* of *C* aan te duiden, vermits beiden de prijs niet hebben gekregen. Dus dit lid maakt een willekeurige keuze uit *B* of *C*, wat betekent dat hij met kans 1/2 meedeelt dat het *C* is. Dus de totale kans van dit scenario is 1/4 maal 1/2, zijnde 1/8.

Scenario 2: *B* heeft de prijs gewonnen. Dit heeft kans 1/2, maar nu kan het jurylid niemand anders dan *C* aanduiden, dus dat is noodzakelijk zo, wat wil zeggen met een kans gelijk aan 1, dus dit scenario heeft een kans 1/2 maal 1, zijnde 1/2 of 4/8.

De som van de waarschijnlijkheden van deze twee scenario's is 1/8 + 4/8 = 5/8, dus is de kans voor het eerste scenario 1/8 gedeeld door 5/8, met andere woorden 1/5, en voor het tweede scenario 4/5.

Zou men zich afvragen waar de resterende 3/8 naartoe zijn, dan is het antwoord dat dit de waarschijnlijkheid is van de som van de scenario's indien het jurylid *B* had aangeduid als de persoon die het niet had gehaald. Zoals bekend, zijn informele redeneringen altijd delicaat, dus toch ook snel even de zaak formeel weergeven.

Een formele benadering (met een bescheiden geloofsact)

Willen we de zaak formeel benaderen, dan is de Bayesiaanse methode aangewezen. We hebben drie mogelijkheden: *A* of *B* of *C* krijgt de prijs en de waarschijnlijkheden zijn bekend:

$P(A) = 1/4$,
$P(B) = 1/2$,
$P(C) = 1/4$,

waarbij *A*, *B* en *C* afkortingen zijn voor "*A* krijgt de prijs", "*B* krijgt de prijs" en "*C* krijgt de prijs". Het bewijsstuk dat wordt geleverd is het jurylid dat *C* aanduidt, dus

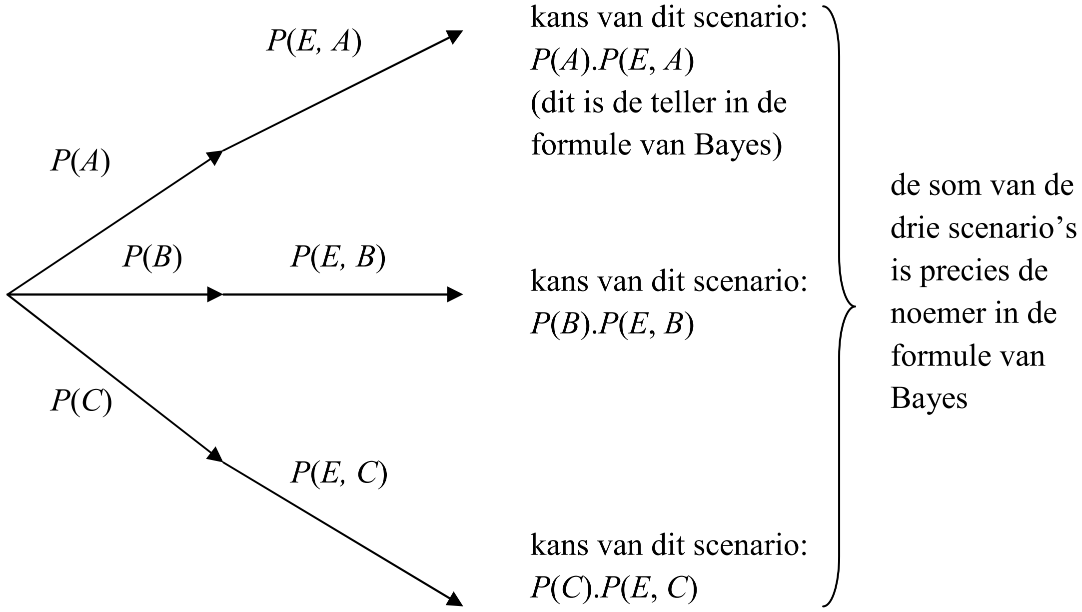
E : het jurylid deelt mee dat *C* het niet heeft gehaald.

Met behulp van deze gegevens laat de Bayesiaanse methode toe om $P(A, E)$ te berekenen, waarbij deze uitdrukking de waarschijnlijkheid uitdrukt nadat *E* gebeurd is. Dit is precies wat we willen weten, namelijk wat de kansen zijn na *E*.

De formule van Bayes zegt hierover het volgende:

$$P(A,E) = \frac{P(A).P(E,A)}{P(A).P(E,A) + P(B).P(E,B) + P(C).P(E,C)}$$

Ik bespaar de lezer de exacte afleiding van deze uitspraak – dit is de geloofsact waarover sprake in de hoofding van deze sectie –, maar ik durf hopen dat de uitspraak dat één beeld meer is dan duizend woorden (of mathematische symbolen) ook in deze context van toepassing is:



Om deze uitdrukking te kunnen berekenen moeten we nog juist weten wat $P(E, A)$, $P(E, B)$ en $P(E, C)$ zijn, met andere woorden, we worden gevraagd de kans te berekenen dat het jurylid zegt dat C niet tot de gelukkigen behoort, gegeven dat A, B, resp. C, de prijs heeft gewonnen.

$P(E, A) = 1/2$, omdat het jurylid zowel B als C kan aanduiden als niet-winnaar,
$P(E, B) = 1$, want het jurylid heeft geen keuze en
$P(E, C) = 0$, want dan zou het jurylid zichzelf tegenspreken.

(Bemerk de overeenkomst hier met de informele redenering, waar ook de getallen 1/2 en 1 opduiken).

Nu moeten we maar deze waarden invullen in de formule van Bayes en onvermijdelijk verschijnt er:

$$P(A,E) = \frac{(1/4).(1/2)}{(1/4).(1/2)+(1/2).1+(1/4).0} = \frac{1}{5}$$

En op juist dezelfde wijze komt er:

$$P(B, E) = \frac{4}{5}$$

en, voor alle duidelijkheid, zoals het hoort:

$P(C, E) = 0.$

Hopelijk door zowel een informele als een formele benadering voor te stellen, beschikt iedereen over de mogelijkheid om zich te overtuigen van de correctheid van dit antwoord. Hoe contra-intuïtief ook.

Een veralgemening

Misschien zijn er wel lezers die menen dat dit probleem verdacht goed lijkt op het Monty Hall probleem of zijn meest bekende variant, namelijk het drie gevangenen probleem. Is er een verband? Het antwoord is wel degelijk ja. Het prijsprobleem is een veralgemening van het Monty Hall probleem. In het laatste geval starten we met waarschijnlijkheden $P(A) = P(B) = P(C) = 1/3$, want de prijs kan achter elke deur zitten met evenveel kans. Als dan de mededeling E komt dat achter deur C geen prijs zit, dan is het duidelijk dat, volkomen gelijkaardig als in het prijsprobleem, $P(E, A) = 1/2$, $P(E, B) = 1$ en $P(E, C) = 0$, zodat de formule van Bayes ons nu zegt dat $P(A, E) = 1/3$, $P(B, E) = 2/3$ en $P(C, E) = 0$.

Zoals bekend blijft in het Monty Hall probleem de kans van A ongewijzigd, maar, als de oorspronkelijke kansen niet gelijk zijn, zoals in het prijsprobleem, dan hoeft de kans niet alleen niet ongewijzigd te blijven, maar het kan zowel naar boven als naar beneden. Wat op zich ook al een interessant gegeven is, want bij Monty Hall insisteert men altijd dat het toch niet anders kan dan dat de kansen van A ongewijzigd blijven. Niet dus in het algemene geval.

Het algemene geval vertrekt dan ook van willekeurige waarschijnlijkheden voor A, B en C, stel, resp., p, q en r, dus $P(A) = p$, $P(B) = q$ en $P(C) = r$, zodanig dat $p + q + r = 1$, met andere woorden, er zijn geen andere mogelijkheden. Als we dan mogen aannemen dat $P(E, A) = 1/2$, $P(E, B) = 1$ en $P(E, C) = 0$, dan zegt de formule van Bayes dat de kansen na E gehoord te hebben de volgende moeten zijn:

$$P(A, E) = \frac{p.(1/2)}{p.(1/2) + q} = \frac{p}{p + 2q},$$

$$P(B, E) = \frac{q}{p.(1/2) + q} = \frac{2q}{p + 2q}, \text{ en, onvermijdelijk}$$

$$P(C, E) = 0.$$

Even checken (voor alle zekerheid, we zijn en blijven sceptici):

* Monty Hall probleem: $p = q = r = 1/3$, dus $P(A, E) = 1/3$ en $P(B, E) = 2/3$, wat klopt,
* het prijsprobleem: $p = r = 1/4$, $q = 1/2$, dus $P(A, E) = 1/5$ en $P(B, E) = 4/5$, wat ook klopt.

Stel nu even dat $p = 1/2$, $q = 1/8$ en $r = 3/8$. Het verrassende resultaat (hoewel, kan hier nog iets verrassend zijn?) is dat nu $P(A, E) = 2/3$ en $P(B, E) = 1/3$, met andere woorden, in dit geval zijn de kansen van *A* effectief gestegen. Nog niet vreemd genoeg? De kansen van *A* stijgen nog steeds, ook als p een zeer kleine waarde heeft, zoals bijvoorbeeld $p = 4/100$, $q = 1/100$ en $r = 95/100$. Even narekenen en $P(A, E) = 2/3$, wat je terecht een dramatische stijging mag noemen, en $P(B, E) = 1/3$, ook een drastische stijging, maar niet goed genoeg.

Er is nog meer. Bij het Monty Hall probleem maken we ons vaak vrolijk over het feit dat mensen denken dat, indien maar twee deuren *A* en *B* overblijven, de kansen gelijk verdeeld moeten zijn, dus 1/2, waardoor het niet uitmaakt of je van deur verandert of niet. Het is beter om voorzichtig te zijn, want een variant op Monty Hall, waarbij $P(A) = 4/7$, $P(B) = 2/7$ en $P(C) = 1/7$, laat een verrassend resultaat zien. Nadat *C* wordt geopend, blijkt bij toepassing van Bayes dat, jawel, $P(A, E) = 1/2$ en $P(B, E) = 1/2$, zodat in dit geval wisselen effectief niets uitmaakt.

Skeptische moraal van dit mathematisch verhaal: altijd twee keer nadenken voor je besluit dat iemand er niets van begrepen heeft. De wereld zit vreemder in elkaar dan we vermoeden. En niet alleen omdat wij er deel van uitmaken.

Met dank aan

Voor wat het tot stand komen van deze bijdrage betreft, moet ik dr. Helen de Cruz, als onderzoekster verbonden aan mijn onderzoekscentrum aan de VUB en haar zus, lic. Lesley de Cruz, die een doctoraat aan de Universiteit Gent voorbereidt, bedanken om mij, één, dit voorbeeld aan de hand te hebben gedaan en, twee, mij overtuigd te hebben van de juistheid van de analyse. Ook dank aan alle andere leden van het CLWF (zie www.vub.ac.be/CLWF/) die dit denkproces in gang hebben gezet toen ik ze heb lastig gevallen met de vraag of ze de volgende vaststelling contra-intuïtief vonden of niet. In een kamer bevinden zich drie vrouwen en één man. Volkomen willekeurig worden twee mensen gekozen en die komen als een koppel naar buiten. Dan is de kans even groot dat dat koppel bestaat uit een man en een vrouw als uit twee vrouwen. De uitleg hierbij is weer een ander verhaal.