Als er iets de gemoederen hoog kan doen oplopen, dan zijn het wel conditionele kansen. Hoe moet je je keuzes en beslissingen bijstellen naarmate je meer informatie verkrijgt? **Ronald Meester** bespreekt enkele problemen.

Verwarring door onvolledige informatie: conditionele kansen

Een voorbeeld

Drie gevangenen, A, B en C, weten dat twee van hen veroordeeld zijn en één niet. Zonder extra informatie is voor A de kans dat hij veroordeeld is gelijk aan $\frac{2}{3}$, er even van uit gaande dat ze allemaal met dezelfde kans veroordeeld zijn.

Stel nu dat A de cipier (die weet wie er veroordeeld zijn) tegenkomt en hem vraagt: 'Kunt u mij de naam geven van een van de andere gevangenen die veroordeeld is?' De cipier is de rotste niet en vertelt A dat B veroordeeld is. A knikt blij, bedenkt dat van de overgebleven mogelijkheden A en C er slechts één veroordeeld is (A heeft nu immers extra informatie dat B veroordeeld is) en hij concludeert dat hijzelf en C allebei met kans $\frac{1}{2}$ veroordeeld zijn. Dit is minder dan $\frac{2}{3}$, de oorspronkelijke kans dat A veroordeeld was. De extra informatie heeft dus voor A (denkt hij) de kans op veroordeling kleiner gemaakt. Heeft A gelijk of niet en waarom? Straks meer hierover.

Een klein beetje theorie

De meeste mensen hebben wel wat intuïtie voor kansen en verwachtingen. Kansen zijn in principe gebaseerd op frequenties waarmee gebeurtenissen optreden, en een verwachting is zoiets als een 'gemiddelde uitkomst'. Als in een experiment de kans op een gebeurtenis bijvoorbeeld $\frac{1}{3}$ is, dan zal bij vele herhalingen van dat experiment in ongeveer $\frac{1}{3}$ van de gevallen deze gebeurtenis optreden. Dus als ik een dobbelsteen vele malen, zeg n keer, achter elkaar gooi, dan zal in ongeveer $\frac{n}{3}$ van die worpen een 1 of een 2 bovenkomen. Buiten een wiskundige context wordt dit de *empirische wet van de grote aantallen* genoemd. De wiskundige formulering van dit fenomeen is subtiel en bepaald niet eenvoudig, maar daar wil ik het hier niet over hebben. De verwachting van een worp met een dobbelsteen is natuurlijk $3\frac{1}{2}$, het gemiddelde aantal ogen.

Heel anders wordt het wanneer we over *conditionele* kansen en verwachtingen praten. Een conditionele kans (verwachting) is niets anders dan een gewone kans (verwachting) waarbij we extra informatie hebben over de uitkomst van het experiment. Als ik een dobbelsteen gooi dan is de kans op een 1 of een 2 natuurlijk $\frac{1}{3}$, en de verwachting is $3\frac{1}{2}$. Maar als ik bijvoorbeeld weet dat het aantal ogen ten hoogste 3 is, dan verandert de zaak. In die situatie zijn alle mogelijke uitkomsten 1, 2 en 3 en er is geen reden om aan te nemen dat een bepaalde uitkomst meer kans heeft dan een andere. We concluderen dan ook dat de conditionele kans op een 1 of een 2, gegeven dat het aantal ogen ten hoogste 3 is, gelijk is aan $\frac{2}{3}$. De conditionele verwachting wordt nu 2.

Misschien dat je je afvraagt *hoe* ik weet dat het aantal ogen ten hoogste 3 is. Dit is een cruciale vraag. Ik kan bijvoorbeeld iemand naar de uitkomst laten kijken, maar deze persoon heeft veel keuze in welke informatie hij mij zal geven. Als een 1 bovenligt, kan hij zeggen: 'Het aantal ogen is niet meer dan 3'. Maar hij kan ook zeggen: 'Het aantal ogen is niet meer dan 5'. In beide gevallen heeft hij gelijk en geeft hij relevante informatie weg, maar de conditionele kans op de gebeurtenis dat er een 1 of een 2 ligt, is in het eerste geval $\frac{2}{3}$ en in het tweede geval $\frac{2}{5}$. Om in dit geval een conditionele kans goed te kunnen definiëren, moeten we *van tevoren* met elkaar afspreken welke informatie weggegeven wordt. Zo kunnen we bijvoorbeeld afspreken dat de 'kijkpersoon' mij zal vertellen of de uitkomst al dan niet hoger is dan 3. Als de uitkomst niet hoger is dan 3, wordt de conditionele kans op een 1 of een 2 gelijk aan $\frac{2}{3}$, als de uitkomst wel hoger is dan 3, dan is die conditionele kans 0. De conditionele verwachting is respectievelijk 2 en 5.

Alle kansen zijn conditioneel

Als je er even over nadenkt, realiseer je je dat bijna alle kansen en verwachtingen die we in het dagelijks leven tegenkomen conditionele kansen en verwachtingen zijn. Neem bijvoorbeeld het (bepaald niet geruststellende) voorbeeld van de slagingskans van een bepaalde operatie. De dokter kan zeggen: 'Eens kijken, u bent 34 jaar oud, verder gezond, rookt niet en drinkt zeer matig, dus het slagingspercentage is in uw geval 80%'. Echt geruststellend is dit niet (even afgezien van de vraag of 80% überhaupt geruststellend kan zijn).

Het percentage dat de dokter noemt, is gebaseerd op gedeeltelijke informatie. Bij meer (of minder) informatie kan de slagingskans heel anders zijn. Net als in het geval van de dobbelsteen, waar het antwoord van de informatie afhangt, kan het hier best zijn dat als de dokter zou weten dat ik getrouwd ben, de kans op slagen toe zou nemen. Dus afgezien van het feit dat je je al af kunt vragen wat de betekenis van een kans is in een experiment dat je niet kunt herhalen, moet je zeer voorzichtig zijn met interpretaties van conditionele kansen.

Gevoel voor conditionele kansen

Hoe kan ik studenten gevoel voor conditionele kansen en verwachtingen bijbrengen? Ik ervaar elke keer weer dat mijn studenten (in dit geval tweedejaars wiskundestudenten) bijzonder makkelijk in verwarring te brengen zijn. Er is iets met conditionele kansen en verwachtingen dat ze voor veel studenten moeilijk maken. Laten we nog eens naar het voorbeeld van de gevangenen kijken. Veel studenten denken dat A gelijk heeft. De volgende redenering laat echter zien dat A geen gelijk *kan* hebben en geeft direct een antwoord op de vraag aan het begin van deze alinea.

Als gedachtenexperiment 'herhalen' we de gehele situatie (inclusief de uitspraak van de rechter) van de drie gevangenen vele malen. Het is zonder meer duidelijk dat A in ongeveer $\frac{2}{3}$ van de gevallen veroordeeld zal worden, ongeacht of hij de cipier vraagt de naam van een andere veroordeelde te geven. Het vragen naar de naam van een veroordeelde is volstrekt nutteloos, aangezien er altijd ten minste een andere veroordeelde is wiens naam door de cipier genoemd kan worden. In feite kunnen we hier dus ook de wet van de grote aantallen gebruiken: bij het herhalen van het experiment zien we direct of de extra informatie ook daadwerkelijk invloed heeft op kansen en verwachtingen.

Deze redenering geeft het goede antwoord, maar vertelt ons niet wat er nu precies verkeerd is aan de redenering van A. A heeft immers gelijk dat er maar twee gevangenen over zijn waarvan er één veroordeeld is. Blijkbaar hebben ze niet dezelfde (conditionele) kans. De volgende keer zal ik stilstaan bij de vraag hoe dat wiskundig in elkaar steekt.

Vragen

Er volgen nu een aantal vragen over conditionele kansen en verwachtingen. Mijn ervaring is dat deze vragen tot bijzonder levendige discussies kunnen leiden (dit is een eufemisme). Het zou daarom jammer zijn als ik direct zou vertellen hoe de vork in de steel zit. In de volgende *Nieuwe Wiskrant* kom ik op de vraagstukken terug. Bij de eerste drie vragen zal een goede discussie uiteindelijk op 'ja' of 'nee' uitkomen. De laatste vraag biedt wat ruimte voor creativiteit, vooral in het bedenken van een goede beslissingsprocedure. Veel plezier ermee!

Vraag 1
Stel dat in het bovenstaande voorbeeld A de cipier had gevraagd: 'Is B veroordeeld?' en dat de cipier 'Ja' had geantwoord. Zou dit voor A de conditionele kans dat hij veroordeeld is wel veranderen? Waarom?

Vraag 2
We zagen dat het antwoord van de cipier voor A geen relevante informatie opleverde. Levert het antwoord van de cipier wel relevante informatie op voor C? Met andere woorden, is de conditionele kans (van A uit gezien) dat C veroordeeld wordt ook gewoon $\frac{2}{3}$ of is hier toch iets meer aan de hand?

Vraag 3
Stel ik heb twee enveloppen met verschillende hoeveelheden geld erin, waarvan ik jou er één laat trekken. Je maakt hem *niet* open. Je hebt geen idee hoeveel geld erin zit en de kans dat je de envelop met het hoogste bedrag hebt getrokken, is uiteraard $\frac{1}{2}$. Nu geef ik je extra informatie: ik vertel je namelijk dat de ene envelop (welke weet ik niet) tweemaal zoveel geld bevat als de andere. Vervolgens krijg je van mij de gelegenheid om de enveloppen om te ruilen. Is het verstandig om dit te doen?

Hier is een interessante redenering: we weten natuurlijk niet hoeveel geld er in de verschillende enveloppen zit, maar stel dat jij x gulden in jouw envelop hebt. Op grond van mijn extra informatie weet je dan dat de andere envelop $2x$ of $\frac{1}{2}x$ gulden bevat en dat deze mogelijkheden gelijke kans hebben. De gemiddelde opbrengst van de andere envelop is dan dus $\frac{1}{2} \cdot 2x + \frac{1}{2} \cdot \frac{1}{2}x = \frac{5}{4}x$. Dit is meer dan ik nu heb (namelijk x) dus het is zeker verstandig om te ruilen, aangezien de conditionele verwachte opbrengst op grond van deze redenering hoger wordt na ruilen. Is dit correct, en waarom (niet)?

(Om te beslissen of je inderdaad moet ruilen, is het niet nodig om de hele redenering al dan niet te weerleggen.)

Vraag 4
Ik heb weer twee enveloppen met geld waarvan jij er één trekt. In dit geval maak je hem *wel* open en je ziet een bepaald bedrag, zeg 30 gulden. In vergelijking met het vorige vraagstuk heb je nu dus andere informatie: je weet precies hoeveel geld jouw envelop bevat, maar je weet niets meer over de verhouding van de hoeveelheden geld. Is na deze extra informatie de conditionele kans dat jouw envelop het hoogste bedrag bevat nog steeds $\frac{1}{2}$? Is er nu een of andere beslissingsprocedure die vertelt of je al dan niet moet ruilen en die ook winstgevend is?

Merk wel op dat in sommige gevallen het antwoord op deze vraag flauw is. Als je bijvoorbeeld weet dat de enveloppen dertig en vijftig gulden bevatten, dan weet je alles na het openen van jouw envelop. Je moet er dus vanuit gaan dat je van tevoren niet weet welke bedragen ik in de envelop heb gedaan.

Ronald Meester, Mathematisch Instituut, Universiteit Utrecht