Jeroen Spandaw
Faculteit Elektrotechniek, Wiskunde en Informatica
TU Delft,
Postbus 5031,
2600 GA Delft
j.g.spandaw@tudelft.nl

Vakantiecursus

Kansrijke symmetrie

Hoe groot is de kans op een stompe driehoek? Hoe groot is de kans dat een koorde 'lang' is? Sinds 1946 organiseert het CWI aan het eind van de zomervakantie een vakantiecursus voor Nederlandse en Vlaamse wiskundeleraren. In 2011 is het thema 'Symmetrie'. Jeroen Spandaw houdt een lezing over kansrekening.

In de geest van de vakantiecursus wilde ik met mijn voordracht over kansrekening vooral mooie en hopelijk verrassende wiskunde laten zien. Een tweede doel was de docenten ervan te doordringen dat kansrekening uitgaat van kansmodellen en dat die modellen keuzes inhouden. Soms kunnen verschillende keuzes gemaakt worden die tot verschillende resultaten leiden.

Om dat punt duidelijk te maken heb ik een deel van het publiek geshockeerd met enkele *ill-posed* vragen, iets waar sommige wiskundig denkende mensen slecht tegen kunnen. Als wiskundedocenten altijd wiskundig onberispelijke vragen stellen, dan ontnemen ze hun leerlingen en studenten de gelegenheid om te leren vage vragen om te zetten in wiskundige vragen. Ik vind het ook niet de taak van de docent om zijn studenten zorgvuldig om iedere valkuil heen te leiden. Integendeel, ik geloof in de heilzame werking van struikelen en opstaan. Achteraf natuurlijk wel goed lessen proberen te trekken uit de struikelpartij. Wees dus op uw hoede, geachte lezer!

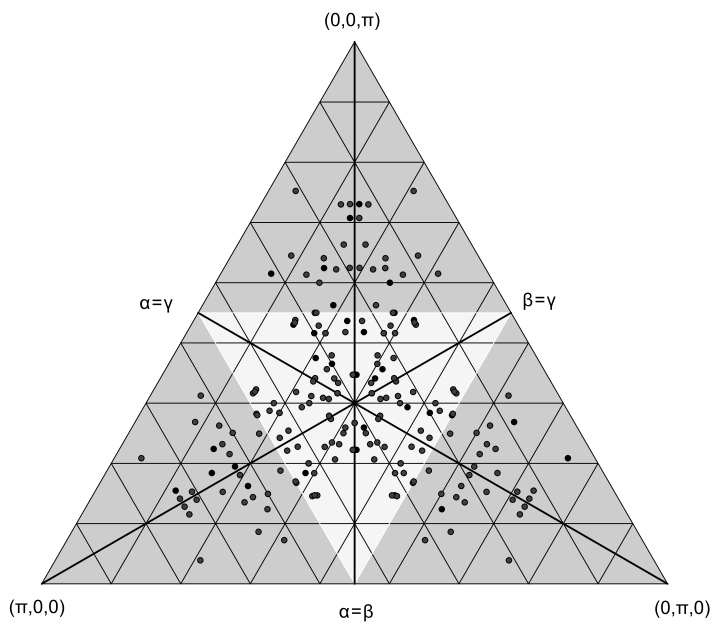
De kans op een stompe driehoek

Een paar jaar geleden vroeg ik mijn studenten in de lerarenopleiding om een 'willekeurige driehoek' te tekenen. Veertien getekende driehoeken waren stomp en zeventien waren scherp. De vakantiecursisten tekenden eveneens vaker een scherpe dan een stompe driehoek. Maar hoe groot is eigenlijk de kans op een stompe driehoek?

Om die vraag te beantwoorden bekijken we een uitkomstenruimte $\mathcal{D}$ die alle driehoeken parametriseert. Dat kan op verschillende manieren gedaan worden, maar een mooie symmetrische manier is

$$\mathcal{D} := \{(\alpha, \beta, \gamma) \in [0, \pi]^3 \,|\, \alpha + \beta + \gamma = \pi\}.$$

Een driehoek is in deze opvatting dus een geordend drietal van hoeken. (Voor onze vraag mogen we gelijkvormige driehoeken identificeren.) We verspillen natuurlijk geen tijd aan de vraag of de hoeken 0 en π wel of niet mee mogen doen. Figuur 1 bevat de resultaten van mijn studenten. De dikke punten (α, β, γ) representeren de getekende driehoeken. Omdat de driehoeken van de studenten onge-



Figuur 1

ordend waren, heb ik alle permutaties van een drietal (α, β, γ) ook genoteerd. In de driehoek $\mathcal{D}$ correspondeert dit met spiegelen in de lijnen $\alpha = \beta$, $\beta = \gamma$ en $\gamma = \alpha$.

De donkergrijze driehoeken parametriseren de driehoeken met een stompe hoek; het lichtgrijze gedeelte in het midden parametriseert de scherpe driehoeken. Dit plaatje suggereert dat de kans op een stompe driehoek maar liefst 75% is! Zouden leraren dus vaker stompe driehoeken moeten tekenen?

U begrijpt het al: de vraag naar de kans op een stompe driehoek is *ill-posed*. Die kans is pas gedefinieerd als we een uitkomstenruimte en een kansmaat hebben vastgelegd. Als we de bovenstaande uitkomstenruimte $\mathcal{D}$ kiezen en als we de kansmaat evenredig aan de oppervlaktemaat op $\mathcal{D}$ nemen, dan is de kans op een stompe driehoek inderdaad 75%. Verschillende studenten hadden verschillende uitkomstenruimten geconstrueerd met verschillende kansmaten die tot verschillende antwoorden leiden. Natuurlijk ontstond een discussie over wat nu 'het goede antwoord' was. Dat was er dus niet. Een heilzaam cognitief conflict, hoop ik.

Na het college dacht ik na over andere, meer natuurlijke kansmaten op de ruimte van alle driehoeken. Ik neem aan dat de zes coördinaten onafhankelijk en standaardnormaal verdeeld zijn. Wat is nu de kans op een stompe driehoek? Leo Breebaart schreef een programmaatje in Python waarmee ik een miljard driehoeken simuleerde. 74.9995% van de driehoeken was stomp! Rogier Brussee en ik hebben vervolgens geverifieerd dat de kans exact $\frac{3}{4}$ is door de zesvoudige integraal

$$\frac{1}{(2\pi)^3}\iiint_{\mathcal{A}} \exp(-\tfrac{1}{2}(A^2+B^2+C^2))\,dA\,dB\,dC$$

over $\mathcal{A} = \{(A,B,C) \in \mathbb{R}^2\times\mathbb{R}^2\times\mathbb{R}^2 \mid \angle BAC \text{ is stomp}\}$ te berekenen. (Substitueer $B = A + b$, $C = A + c$, splits een kwadraat af, integreer A weg, voer poolcoördinaten voor b en c in, ontwikkel de factor $\exp(|b||c|\cos(\angle A))$ in een machtreeks en verwissel sommatievolgorde.) Een veel mooier bewijs is te vinden in [1].

De paradox van Bertrand

Neem een cirkel met een ingeschreven gelijkzijdige driehoek. Een koorde (een lijnstuk waarvan de eindpunten op de cirkel liggen) noemen we *lang* als ze langer is dan de zijde van de ingeschreven gelijkzijdige driehoek. Hoe groot is de kans dat een 'willekeurige koorde' lang is? De volgende drie redeneringen (van Wikipedia) leiden tot verschillende antwoorden.

1. Door de koorde te roteren mogen we aannemen dat het eindpunt P van de koorde PQ samenvalt met het hoekpunt A van de ingeschreven gelijkzijdige driehoek ΔABC. De koorde is lang dan en slechts dan als het andere eindpunt Q van de koorde tussen B en C ligt. De kans is dus $\frac{1}{3}$.
2. De koorde PQ wordt bepaald door haar middelpunt N. (Dit is niet waar als N samenvalt met het middelpunt M van de cirkel, maar de kans daarop is nul. We mogen deze subtiliteit dus negeren.) De koorde is lang dan en slechts dan als N binnen de cirkel met middelpunt M en halve straal ligt. De kans hierop is dus $\frac{1}{4}$.
3. Door te roteren mogen we aannemen dat de koorde horizontaal ligt. Het punt N ligt dus op de verticale diameter. De koorde is lang dan en slechts dan als N hoogstens een halve straal van M ligt, dus de kans op een lange koorde is $\frac{1}{2}$.

Een vierde argument grijpt terug naar de schooldefinitie van Laplace: de kans is het aantal gunstige gebeurtenissen gedeeld door het totaal aantal gebeurtenissen. Dat argument levert $\frac{1}{2}$ of $\frac{1}{3}$ of iets anders. Dit voorbeeld illustreert de beperkingen van Laplace. Niet alleen faalt Laplace bij oneindige uitkomstenruimten, zelfs in eindige gevallen zoals het gooien met twee dobbelstenen moeten we keuzes maken: representeren we $(5,6)$ en $(6,5)$ als verschillende uitkomsten of niet? Je kunt niet met wiskunde alleen bepalen welk kansmodel de werkelijkheid het best beschrijft. Laplace is geen recept.

De belangrijkste boodschap was echter: je kunt pas over de kans op een lange koorde spreken als je een kansmaat op de verzameling koorden hebt vastgelegd. Verschillende kansmaten geven verschillende kansen en verschillende invullingen aan het begrip 'willekeurige koorde'. De vraag was dus *ill-posed*, zelfs als stilzwijgend (onbewust?) wordt aangenomen dat de kansverdeling invariant onder rotaties is. De eerste drie argumenten laten zien dat die symmetrie alleen onvoldoende is om een kansmaat vast te leggen.

Simulatie geeft ook geen uitsluitsel, want voor een simulatie moet je vastleggen hoe je een 'willekeurige' koorde genereert. Daarmee kies je in feite een kansverdeling. In de eerste redenering worden de eindpunten van de koorde onafhankelijk van elkaar getrokken uit de uniforme verdeling op de cirkel. In de tweede redenering wordt het middelpunt N van de koorde getrokken uit de uniforme verdeling op de cirkelschijf. (De kans op $N = M$ is nul, dus de kans is 1 dat N precies één koorde definieert.) In de derde redenering wordt N getrokken uit de uniforme verdeling op de verticale diameter. Vervolgens wordt de koorde geroteerd over een hoek die wordt getrokken uit de uniforme verdeling op $[0, 2\pi]$. Ik heb overigens met opzet in de drie argumenten deze kansverdeling niet uitgespeld. Die zijn immers niet nodig om in te zien dat er verschillende rotatie-invariante oplossingen zijn die verschillende kansen opleveren. De bovenstaande argumenten kunnen besproken worden met leerlingen, die niet weten wat een σ-algebra is.

Rad van Fortuin in drie dimensies

Na de anarchie van Bertrand, waarin verschillende antwoorden mogelijk zijn, is het tijd voor eenduidiger wiskunde. We gaan kijken naar rotaties in drie dimensies. Ik stel me een kandidaat in een spelshow voor die aan een drijvende bol draait. Zo'n rotatie wordt bepaald door een rotatieas en een rotatiehoek. We vragen ons af of alle rotatieassen en -hoeken even waarschijnlijk zijn. Van Bertrand hebben we geleerd dat we eerst een kansmaat op de verzameling rotaties moeten vastleggen. De theorie van Liegroepen leert ons dat er precies één kansmaat is, de *Haarmaat*, die invariant is onder alle rotaties. Deze kansmaat nemen we, dus we spreken nu over welgedefinieerde kansen. (Ik weet niet in hoeverre deze kansmaat een goede weerspiegeling is van zo'n spelshow.)

Door de rotatiesymmetrie zijn alle rotatieassen even waarschijnlijk. Preciezer: de kansdichtheid op de verzameling rotatieassen is uniform. (Ik zou het nog preciezer kunnen formuleren, maar daar wordt niemand wijzer van.) Hoe ziet de kansverdeling op de rotatiehoeken eruit? Allereerst merken we op dat we alleen hoeken tussen 0 en 180° nodig hebben. Is de kansverdeling uniform op het interval $[0, 180^\circ]$? Bas Edixhoven leerde mij dat dat niet het geval is. De kansdichtheid $p(\varphi)$ is evenredig met $\sin^2(\frac{1}{2}\varphi)$. Grote hoeken zijn dus waarschijnlijker dan kleine hoeken! Na de eerste schok vond ik dat wel plausibel: er zijn immers veel meer rotaties over 180° dan over 0°.

Voor de mooie afleiding van de kansdichtheid $p(\varphi)$ met behulp van quaternionen verwijs ik u naar de syllabus [2], die gratis te downloaden is.

Maar Bas had nog een tweede verrassing in petto: we gaan de rotatie verdubbelen. We houden dus dezelfde rotatieas, maar we verdubbelen de rotatiehoek. Dan wordt de kansverdeling op de rotatiehoeken wél homogeen! Ook dit is na de eerste verbazing wel te begrijpen: als we bijvoorbeeld een rotatiehoek van 170° verdubbelen, krijgen we een rotatiehoek van 340° en dat is equivalent met een rotatiehoek van 20°. (Verander de oriëntatie van de rotatieas.) De rotatiehoeken van 10° en 170° worden door verdubbeling dus op een hoop gegooid. Bij verdubbeling moeten we dus de kansdichtheid bij supplementaire hoeken φ en $\pi - \varphi$ optellen. We vinden dat de kansdichtheid na verdubbeling evenredig is met de constante $\sin^2(\frac{1}{2}\varphi)+\sin^2(\frac{1}{2}(\pi-\varphi))$.

Benford en Weyl

Symmetrie is niet beperkt tot meetkundige contexten als stompe driehoeken, lange koorden of 3-dimensionale rotaties. Neem een lange lijst van getallen, bijvoorbeeld de inwoneraantallen van alle landen ter wereld of de afstanden tot andere sterrenstelsels uitgedrukt in lichtjaren of de afstanden tot andere sterrenstelsels uitgedrukt in Rijnlandse roeden. Bekijk nu van al die getallen het eerste cijfer. Zijn alle begincijfers 1 tot en met 9 even waarschijnlijk? Volgens de Wet van Benford is het antwoord ontkennend: we verwachten dat ongeveer 30% van de getallen met een 1 begint, terwijl minder dan 5% met een 9 zal beginnen. Voor begincijfer n geeft Benford de kans

$$p(n) = \log_{10}\left(1 + \frac{1}{n}\right).$$

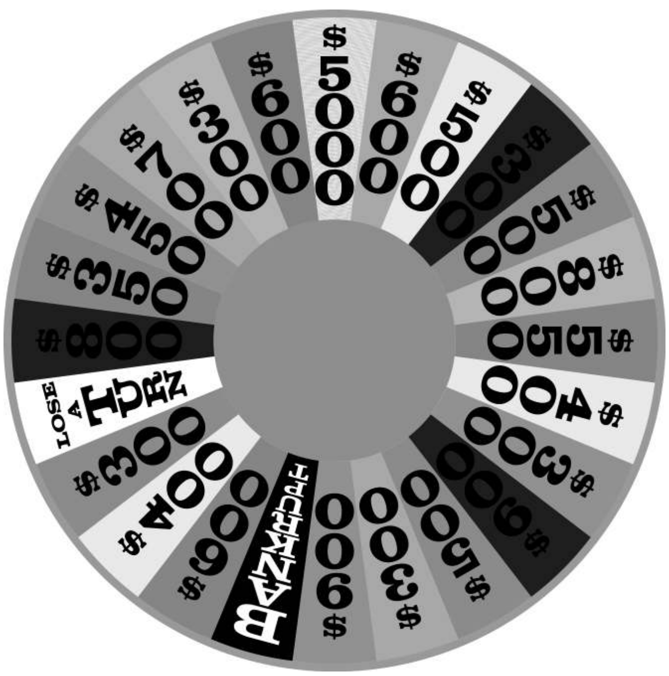
Er bestaan wiskundig exacte formuleringen van de Wet van Benford, maar ik houd het bij deze empirische versie.

Wat heeft dit met symmetrie te maken? Wel, *als* er een 'universele' kansverdeling op de begincijfers is, dan geldt deze ook na verandering van eenheid, zoals in het voorbeeld van de sterrenstelsels. Alle afstanden x in de lijst worden dus met een vaste omrekeningsfactor c vermenigvuldigd, de begincijfers zullen in het algemeen veranderen, maar de relatieve frequenties van de begincijfers moet idealiter gelijk blijven. De kansverdeling is dus invariant onder $x \mapsto cx$. Voor de uitkomstenruimte Ω starten we met $\mathbb{R}^+ = (0, \infty)$. We identificeren x met $10x$, want die hebben hetzelfde begincijfer. De uitkomstenruimte is dus de multiplicatieve groep $\Omega := \mathbb{R}^+/\langle 10\rangle$. Topologisch is Ω een cirkel. Net als in het geval van de groep van 3-dimensionale rotaties is er een unieke maat, de Haar-maat, die invariant is onder alle herschalingen $x \mapsto cx$. De bijbehorende kansdichtheid is evenredig met dx/x, want $d(cx)/cx = dx/x$. Nu kunnen we Benford begrijpen:

$$p(n) \sim \int_n^{n+1} \frac{dx}{x} = \log\left(\frac{n+1}{n}\right).$$

Normeren van de totale kans verandert het grondtal van de logaritme in 10.

Om u niet met een katterig gevoel over een empirische wet achter te laten, sluit ik af met een echte stelling, die ik heb geleerd van Ben de Pagter. Neem een willekeurige macht 7^k van 7. Hoe groot is de kans dat het begincijfer een 4 is? Volgens de gelijkverdelingsstelling van Hermann Weyl luidt het antwoord $\log_{10}(5/4)$, net als in de wet van Benford. Met 'kans' wordt het volgende bedoeld: Neem de machten 7^k voor $k = 1, \ldots, N$ en bekijk de relatieve frequentie van de machten die met het cijfer n beginnen. Volgens Weyl convergeert deze relatieve frequentie naar $\log_{10}(1 + 1/n)$ als $N \to \infty$.



Referenties

1 Stephen Portnoy, A Lewis Carroll Pillow Problem: Probability of an Obtuse Triangle, *Statistical Science*, 1994, **9**(2), 279–284.

2 Symmetrie, Vakantiecursus 2011, CWI syllabus 61, oai.cwi.nl/oai/asset/18770/18770D.pdf.