Jurjen Bos heeft een eigenzinnige manier van het opvouwen van zijn sokken: random er een uit de wasmand nemen, kijken of er een bijpassende sok op schoot ligt, zo niet: sok erbij, zowel: vouwen maar.
Een werkwijze als deze levert mooie vragen op met verbazingwekkende antwoorden.

Losse sokken

Inleiding

Wiskundigen zijn net mensen. Net als biologen overal interessante beestjes en plantjes tegenkomen, zien wiskundigen overal interessante wiskundeproblemen. Als wiskundige vallen mij de interessante beestjes en plantjes niet zoveel op, maar af en toe kom ik wel een wiskundeprobleem tegen dat zowel relevant als lastig genoeg is om interessant te zijn.
Dit artikel gaat over een probleem dat aanvankelijk veel eenvoudiger leek dan het was. Ik kwam het probleem tegen tijdens het opvouwen van mijn schone sokken. Nadat ik er eens goed over had nagedacht, heb ik een aantal vragen die ik mezelf stelde kunnen oplossen, maar er is ook nog een aantal vragen overgebleven.
Uiteindelijk besloot ik (na veel uitstel) om het eens netjes op te schrijven; dat is het artikel geworden dat u nu voor u ziet. De meeste wiskundige artikelen zijn geschreven als een keurig uitgeschreven betoog dat laat zien hoe de oplossing in elkaar zit, zonder de dwalingen van de ontdekker te tonen. Dit artikel gaat meer over de zoektocht zelf; ik wil laten zien hoe wispelturig de zoektocht kan zijn om een ogenschijnlijk simpel probleem op te lossen. Ik hoop hiermee te laten zien waarom wiskunde zo leuk is. Maar dat wist u natuurlijk al, anders las u dit niet.

Hoe ik sokken opvouw

Om te beginnen een korte introductie in het opvouwen van sokken, zoals ik het doe. Ik geef toe dat het makkelijker is om de sokken op kleur bij elkaar te zoeken, maar gelukkig ben ik niet zo gestructureerd, anders was dit artikel er nooit gekomen.
Ik pak steeds zonder in de wasmand te kijken één sok. Als deze sok niet past bij het rijtje sokken dat ik op schoot heb liggen, leg ik hem erbij. Het rijtje op mijn schoot bevat zo steeds de sokken van een paar waarvan de andere sok nog in de wasmand zit. Als ik een sok heb gepakt die wel in het rijtje voorkomt, pak ik de bijbehorende sok erbij, vouw ik ze samen op, en leg ze weg.
Het rijtje sokken op mijn schoot wordt dus steeds één sok langer of korter. Tijdens het sokkenvouwen wordt het rijtje langzamerhand steeds langer, en na verloop van tijd wordt het weer korter. Als het goed is, is het rijtje leeg aan het eind van het proces, anders ben ik een sok kwijt (dat komt maar al te vaak voor, maar daar hebben we het later over).
Als ik teveel sokken op mijn schoot heb liggen, wordt het vouwen lastig. Op een dag vroeg ik me af hoe groot het rijtje eigenlijk wordt, gegeven hoe vol de wasmand is. Die vraag bleek gemakkelijker om te stellen dan om te beantwoorden, zoals dat meestal gaat.

Variabelen

Om hier een 'echt' wiskundeprobleem van te maken, moeten we eerst wat variabelen invoeren. We splitsen net als bij echte wiskunde de tijd in discrete eenheden: één stap voor elke sok die opgevouwen moet worden.

- We bekijken steeds de situatie na het pakken van t sokken. t loopt dus van 0 (aan het begin) naar $2n$ (alle sokken opgevouwen).
- De wasmand is aanvankelijk gevuld met n paar sokken, dus $2n$ sokken. Na t stappen heb ik dus nog $2n - t$ sokken in de wasmand; losse sokken behandelen we later.
- Het pakken van de sokken uit de mand is een aselecte trekking. Het aantal paren dat ik heb gevouwen en weggelegd, is dan een stochastische functie: die noem ik $p(t)$. (Het aantal gevouwen *sokken* op tijdstip t is dan $2p(t)$.)
  Voorlopig ga ik er bij de formules van uit dat ik volkomen willekeurig sokken uit de mand haal; later ga ik alternatieven verkennen.
- De hoeveelheid sokken in het rijtje op mijn schoot noem ik $s(t)$.

Gebruikmakend van bovenstaande notatie kan de 'wet van behoud van sokken' worden geschreven als $t = s(t) + 2p(t)$.

Kansverdeling

Met bovenstaande notatie kunnen we aan de slag.
De kans dat bij het pakken van sok nummer $t + 1$ een nieuw paar wordt gevormd, wordt bepaald door de situatie na t sokken. Op dat moment moet ik een sok kiezen uit

een wasmand met $2n - t$ sokken; $s(t)$ van de sokken in de mand hebben een bijbehorende sok op mijn schoot.
Stel dat ik inderdaad een passende sok op mijn schoot heb liggen. De formule wordt dan voor $t > 0$:

$$P_{paar}(t+1) = \frac{s(t)}{2n-t}$$

In dit geval is $s(t + 1) = s(t) - 1$ en $p(t + 1) = p(t) + 1$. De kans dat er geen nieuw paar wordt gevormd bij het pakken van sok $t + 1$ is:

$$P_{los}(t) = 1 - P_{paar}(t+1)$$

Het is duidelijk dat in dit geval geldt dat $s(t + 1) = s(t) + 1$ en $p(t + 1) = p(t)$.
Gebruikmakend van het bovenstaande kan de functie $s(t)$ recursief worden gedefinieerd als:

$$s(0) = 0$$

en

$$s(t+1) = \begin{cases} s(t) - 1 \text{ met kans } P_{paar}(t) = \dfrac{s(t)}{2n-t} \\ s(t) + 1 \text{ met kans } P_{los}(t) = 1 - P_{paar}(t+1) \end{cases}$$

Tijdens het was opvouwen is het gemakkelijk in te zien dat $s(2n) = 0$ als er geen losse sokken in de mand zitten; aan de formules is dit niet zo duidelijk te zien. Later gaan we dit bewijzen.

Voorbeeld
In figuur 1 staat een voorbeeld van het verloop van het aantal sokken op mijn schoot in een werkelijke sokkenvouwsessie van zeven paar sokken.

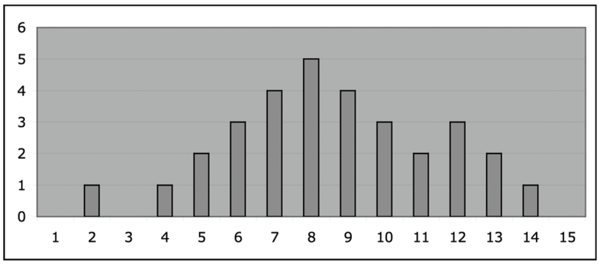

fig. 1 Het opvouwen van zeven paar sokken

Uitwerking

Nu we de verschillende variabelen hebben geïntroduceerd, kunnen we gaan rekenen.
De oorspronkelijke vraag is hoeveel ruimte de sokken maximaal innemen op mijn schoot. In formules is dat:

$$\max_{0 \le t \le 2n} s(t)$$

Dit is een stochast, waarvan we het liefst de kansverdeling willen weten. Deze stochast is niet zo eenvoudig als de stochasten die we op school tegenkwamen. De functie s, met zijn $2n + 1$ waarden voor $t = 0, 1, ..., 2n$, is als geheel als een stochast te beschouwen. s is te zien als een trekking uit de verzameling van alle mogelijke functies S.

De verwachtingswaarde van het aantal sokken op mijn schoot is bijvoorbeeld:

$$\underset{s \in S}{E}\left(\max_{0 \le t \le 2n} s(t)\right)$$

Hier komen we later op terug.

Continue vorm
Het eerste wat ik deed was proberen een idee te krijgen van de uitkomst. Daarom maakte ik $s(t)$ eerst continu, in de hoop dat het me een idee gaf waar de uitkomst naartoe gaat. Ik noem de continue versie van de functie $\bar{s}$. Voor $\bar{s}(t)$ kunnen we een differentiaalvergelijking maken door voor zijn afgeleide $\bar{s}'(t)$ de verwachtingswaarde te kiezen van $s(t + 1) - s(t)$:

$$\begin{aligned} \bar{s}'(t) &= E[s(t+1)] - E[s(t)] = E[s(t+1) - s(t)] = \\ &= -1 \cdot P_{paar}(t) + 1 \cdot P_{los}(t) \\ &= 1 - 2 \cdot P_{paar}(t) \\ &= 1 - 2\frac{\bar{s}(t)}{2n-t} \end{aligned}$$

De oplossing van deze differentiaalvergelijking is een eenvoudige kwadratische functie, die ik voor de duidelijkheid op drie verschillende manieren schrijf:

$$\bar{s}(t) = \frac{t}{2n}(2n-t) = -\frac{1}{2n}t^2 + t = \frac{n^2 - (t-n)^2}{2n}$$

Dat deze oplossing aan de voorwaarden voldoet, is gemakkelijk te zien door substitutie:

$$1 - 2\frac{\bar{s}(t)}{2n-t} = 1 - 2\frac{\frac{t}{2n}(2n-t)}{2n-t} = 1 - t/n = \bar{s}'(t)$$

Ik geef eerlijk toe dat ik die oplossing van de differentiaalvergelijking door een computer heb laten bepalen. De lol van wiskunde zit niet in het rekenwerk, maar in het inzicht in de materie.
Toen ik dit resultaat had, ging ik natuurlijk meteen kijken of het aan de verwachtingen voldeed. De functie voldoet keurig aan $\bar{s}(0) = \bar{s}(2n) = 0$, het bereik van $\bar{s}$ is $\left[0, \frac{n}{2}\right]$, en het domein is $[0, 2n]$: dat zou kunnen kloppen met mijn ervaringen uit de praktijk. Het grafiekje hieronder ziet er ook plausibel uit. Dit geeft me het gevoel dat ik op de goede weg ben.

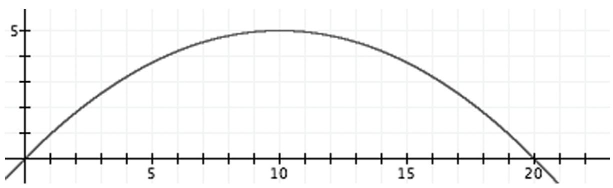

fig. 2 De continue vorm van s(t) voor n = 10

Eerste analyse
Met deze continue vorm, die in zekere zin de ‘gemiddelde’ situatie aangeeft, kunnen we al zien wat voor oplossing we kunnen verwachten. Het aantal sokken op m’n schoot zal

eerst snel, dan steeds langzamer stijgen. Op het moment dat ik de helft van de sokken uit de mand heb gehaald is het aantal sokken op mijn schoot ongeveer de helft van het aantal paren. Daarna daalt het aantal sokken langzaam, maar steeds sneller tot de laatste sok is opgevouwen.
Het aantal sokken in de wasmand $p(t)$ stijgt kwadratisch met de tijd:

$$\bar{p}(t) = \frac{t - \bar{s}(t)}{2} = \frac{t^2}{4n}$$

zodat aan het eind de mand het snelst vol gaat.
De oplossing is overduidelijk symmetrisch. Op de vraag of dit door de benadering komt, of doordat het probleem inherent symmetrisch is, kom ik later terug.
Deze continue vorm geeft ons wel een idee van de oplossing van het oorspronkelijke probleem, maar nog niet de uitkomst. We waren oorspronkelijk geïnteresseerd in het maximum van $s(t)$. Het is duidelijk dat het maximum van $\bar{s}(t)$ gelijk is aan $\frac{n}{2}$. Het echte aantal sokken $s(t)$ fluctueert waarschijnlijk om de continue oplossing $\bar{s}(t)$ heen, zodat het verwachte aantal sokken dat ik maximaal op m'n schoot leg tijdens het opvouwen niet precies op $t = n$ zal plaatsvinden en groter zal zijn dan $\frac{n}{2}$. Hoeveel groter kan ik hier niet zien.

Kansverdeling van s(t)

Voor de hoofdvraag van de hoeveelheid ruimte op mijn schoot is het nodig om de kansverdeling van de functies $s \in S$ te bepalen.
We bepalen eerst hoeveel sokken er op een gegeven moment op mijn schoot liggen. In wiskundetaal: de kansverdeling van $s(t)$ voor een gegeven waarde van t.
Hiervoor moeten we om te beginnen de kans berekenen dat op een gegeven moment een bepaald aantal sokken op mijn schoot ligt. Die kans hangt af van wat er in de wasmand lag voordat ik een sok pakte, en natuurlijk ook van wat er toen op mijn schoot lag. Laten we dit even een echt wiskundige naam geven:

- De kansfunctie $k_t(i)$ geeft de kans op i sokken op mijn schoot op tijdstip t:
  $$k_t(i) = P\,[s(t) = i\,]$$

Dat we echt iets hebben aan deze definitie, kun je zien aan het feit dat we gemakkelijk k_{t+1} kunnen berekenen uit k_t met de recursieformule voor s. Kijk maar:

$$\begin{aligned} k_{t+i}(i) &= k_t(i+1) \cdot P_{paar}(t) + k_t(i-1) \cdot P_{los}(t) = \\ &= k_t(i+1) \cdot \frac{i+1}{2n-t} + k_t(i-1) \cdot \left(1 - \frac{i-1}{2n-t}\right) \end{aligned}$$

en we moeten aannemen dat $k_t(i) = 0$ voor $i < 0$. Door systematisch de waarden van i en t op te hogen, kunnen we alle waarden van $k_t(i)$ bepalen.

Numerieke benadering

Bovenstaande berekening van $k_t(i)$ leent zich voor een computerprogramma. Daar ben ik toevallig goed in.
We kunnen de waarden van k_t voor een bepaalde waarde van i gemakkelijk in een *array* opslaan, en telkens voor elke volgende waarde van i alle nieuwe waarden berekenen. Dat levert snel numerieke waarden op van de kansverdeling van $s(t)$, zodat we een idee kunnen krijgen van het verloop van de stapel sokken op m'n schoot. Je kunt de waarden van $k_t(i)$ zelfs met een spreadsheet berekenen: zo heb ik onderstaande tabellen vastgesteld.
De waarden voor $100 \cdot k_t(i)$ zijn te vinden in onderstaande tabel. In deze tabel staan de kansen 'in procenten' dat i sokken op mijn schoot liggen (verticale as) na t stappen (horizontale as). Waarden die 0 zijn, zijn weggelaten.

Tabel 1: Kansverdeling van s(t) in procenten voor n = 6

$i = 6$							7						
5						24		24					
4					48		52		48				
3				73		61		61		73			
2			91		48		39		48		91		
1		100		27		15		15		27		100	
0	100		9		3		2		3		9		100
$t =$	0	1	2	3	4	5	6	7	8	9	10	11	12

Ik heb veel geleerd door deze tabel te bestuderen. Uiteraard zijn de kolomsommen gelijk aan 100 (op afronding na), omdat iedere kolom een kansverdeling is. Dit is niet zonder meer aan de formule voor $k_t(i)$ te zien. Ik heb de kolomsommen gebruikt om fouten uit de formule in het spreadsheet te halen.
Wat me het meest opviel van s is dat $s(t)$ even is als n even is, en oneven als n oneven. Bij nader inzien is dit eenvoudig aan de recursieformule te zien, en het is ook eenvoudig te beredeneren. Ik had me dit niet van tevoren gerealiseerd; pas toen ik deze tabel had uitgerekend, werd me duidelijk dat dit zo is. Zo zie je maar dat inzicht niet zomaar komt; meestal moet je eerst gewoon hard werken.
Zo'n numerieke benadering kan helpen om een probleem beter te begrijpen. Zelfs grote wiskundigen als Leibniz gebruikten numerieke benaderingen om inzicht te krijgen in hun formules. Dit is vooral bijzonder als u bedenkt dat Leibniz die numerieke benaderingen allemaal met de hand heeft moeten doen, bij gebrek aan rekenmachines.

Symmetrie

Het tweede dat mij opvalt aan de numerieke gegevens is dat de verdeling symmetrisch is om $t = n$, net als de continue benadering. Deze symmetrie is geen gevolg van de benadering op twee cijfers: nader onderzoek (met Excel) leert dat de waarden zo precies zijn als van de precisie van de berekening verwacht kan worden. Het is zeer onwaarschijnlijk dat het dan niet exact is. Toen ik dit ontdekt had, moest ik het natuurlijk ook bewijzen.
De gemakkelijkste manier om het in te zien, is dat het proces symmetrisch is in de tijd, zoals ik nu zal laten zien. Hiermee bewijzen we en passant dat $s(2n) = 0$, zoals hierboven beloofd. Het bewijs is in het Nederlands; er staan al genoeg formules in dit verhaal.
We bekijken alleen het aantal sokken op schoot, en laten de wasmand even weg. Van een gegeven paar sokken wordt eerst één sok gepakt, en later de andere. In de tijd tussen deze twee gebeurtenissen ligt er één sok op mijn

schoot. Het moment waarop een sok (eerste of tweede) gepakt wordt, is ook een stochast. De eerste en tweede sok van een paar hebben dezelfde kansverdeling, zodat het niet uitmaakt als je de tijd omkeert, want dan veranderen alleen de namen ('eerste' en 'tweede') van de sokken. De kansverdeling van de sokken op schoot is dus symmetrisch in de tijd.

De vergissing

Met bovenstaande kansverdeling kunnen we gemakkelijk het verwachte aantal sokken op mijn schoot op een gegeven moment in de tijd berekenen. Voor het voorbeeld hierboven van zes paar sokken krijgen we de verwachtingswaarden uit de volgende tabel. Hier is af te lezen dat de maximum verwachtingswaarde gelijk is aan 3,3; die wordt geheel naar verwachting in het midden aangenomen.

Tabel 2: verwachtingswaarde van s(t) voor n = 6

$t =$	0	1	2	3	4	5	6	7	8	9	10	11	12
E =	0	1,0	1,8	2,5	2,9	3,2	3,3	3,2	2,9	2,5	1,8	1,0	1,0

We hebben nu wel het maximum uitgerekend van de verwachtingswaarde, maar niet de verwachtingswaarde van het maximum.

We zijn gestuit op een subtiliteit waarover menig wiskundige het hoofd heeft gebroken: je mag niet klakkeloos operaties verwisselen, hoe onschuldig dat ook lijkt. We willen namelijk de verwachtingswaarde (over alle mogelijke functies s) van het maximum (over t) van $s(t)$ uitrekenen:

$$\underset{s \in S}{E}\left(\max_{0 \le t \le 2n} s(t)\right)$$

en dat is niet hetzelfde als het maximum (over t) van de verwachtingswaarde (over alle functies s):

$$\max_{0 \le t \le 2n}\left(\underset{s \in S}{E} s(t)\right)$$

We hebben dus het verkeerde probleem opgelost.

U weet dat nu al; ik deed er een paar weken over om daarachter te komen. Maar ja, dat is nou eenmaal de manier waarop onderzoek gaat. Andrew Wiles, die beroemd is geworden door na meer dan tien (!) jaar werk de laatste stelling van Fermat op te lossen, zei:

> 'Het oplossen van dit probleem was als het zoeken van je weg in een vreemd huis in het donker. Eerst stommel je een tijdje door een kamer, je steeds stotend aan het meubilair, totdat je het lichtknopje hebt gevonden. Dan kun je je weg vinden in de kamer, en zie je de deuren naar de volgende kamers. Pas als je alle kamers van het huis op deze manier hebt verkend, is het probleem opgelost.'

Goed, het gaat om het subtiele verschil tussen de verwachtingswaarde van het maximum aantal sokken en het maximum van de verwachtingswaarde van het aantal sokken. Bij slechts drie paar sokken maakt dit al uit: ik zal dit precies laten zien aan de hand van tabel 3.

Er zijn vijftien verschillende volgordes waarop ik drie paar sokken kan pakken, zoals in tabel 3 is uitgeschreven. Ik ga er voor het gemak van uit dat de eerste sok van elk paar respectievelijk (r)ood, (g)roen of (b)lauw is. Bij het invullen van de tabel kom je vanzelf uit op de rechteronderhoek. Dan sta je voor de keuze wat we moeten invullen:

- Het maximum van de verwachtingswaarden $\frac{9}{5} = 1{,}8$. Dit is de waarde die we net hebben uitgerekend in tabel 2.
- De verwachtingswaarde van het maximum $\frac{7}{3} \approx 2{,}33$. Dit is de waarde die ik wil weten: de te verwachten hoeveelheid ruimte die ik op schoot nodig heb.

Tabel 3: Alle mogelijkheden voor drie paar sokken

Volgorde						Op schoot					Max.
b	b	g	g	r	r	1	0	1	0	1	1
b	b	g	r	g	r	1	0	1	2	1	2
b	b	g	r	r	g	1	0	1	2	1	2
b	g	b	g	r	r	1	2	1	0	1	2
b	g	b	r	g	r	1	2	1	2	1	2
b	g	b	r	r	g	1	2	1	2	1	2
b	g	g	b	r	r	1	2	1	0	1	2
b	g	g	r	b	r	1	2	1	2	1	2
b	g	g	r	r	b	1	2	1	2	1	2
b	g	r	b	g	r	1	2	3	2	1	3
b	g	r	b	r	g	1	2	3	2	1	3
b	g	r	g	b	r	1	2	3	2	1	3
b	g	r	g	r	b	1	2	3	2	1	3
b	g	r	r	b	g	1	2	3	2	1	3
b	g	r	r	g	b	1	2	3	2	1	3
Verwachtingswaarde						1	$\frac{8}{5}$	$\frac{9}{5}$	$\frac{8}{5}$	1	$\frac{9}{5}$ of $\frac{7}{3}$

Oplossing

Dan wordt het nu tijd om het probleem echt op te lossen, zoals ik heb beloofd. Om de verwachte waarde van het aantal sokken:

$$\underset{s \in S}{E}\left(\max_{0 \le t \le 2n} s(t)\right)$$

te bepalen, moeten we eerst $\max_{0 \le t \le 2n} s(t)$ bepalen voor verschillende waarden van s.

Het vervelende is dat s een functie is, zodat het lastig is om de kansruimte S te beschrijven. Ik heb een tijdje zitten worstelen om iets over $\max_{0 \le t \le 2n} s(t)$ uit te rekenen, maar ik merkte dat dat niets werd. Het is alsof de formule helemaal niet zo benaderd wil worden. In de taal van Andrew Wiles: ik bleef me maar stoten aan het meubilair. Na lang denken realiseerde ik met dat ik niets eens wist hoe ik mij S moest voorstellen; daar moest ik dus eerst maar eens mee beginnen.

In woorden is de meest voor de hand liggende definitie van S: 'de kans dat alle punten gelijk zijn aan een gegeven functie $S(t)$, gegeven de voorgaande punten'. In formule-vorm is dit nauwkeuriger op te schrijven:

$$P[s = S] = \prod_{0 \le t \le n} P[s(t) = S(t) | \forall_{0 \le k \le t}\ s(k) = S(k)]$$

Omdat de definitie van *s* recursief is, zit er niks anders op dan de beschrijving van *S* ook recursief te maken in *t*, hoe onlogisch dat ook lijkt. Dit inzicht was het lichtknopje dat ik nodig had; toen kon ik eindelijk weer verder. Dus, we gaan het hebben over een 'voorstuk' van *s*, tot aan *m*:

$$P_{0,\ldots,m}[s(t) = S(t)] = \prod_{0 \le t \le m} P[s(t) = S(t) | \forall_{0 \le k < t}\ s(k) = S(k)]$$

waarbij we *m* laten lopen van 0 tot *n*; als $m = n$, hebben we de uitkomst.

Kans op een gegeven verdeling

Dan kunnen we nu eindelijk aan de slag. Uit de formule hiervoor kunnen we zien dat het gaat om een eenvoudige vermenigvuldiging:

$$P_{0,\ldots,m}[s(t) = S(t)] = P_{0,\ldots,m-1}[s(t) = S(t)] \cdot \\ \cdot P[s(m) = S(m) | \forall_{0 \le k < m}\ s(k) = S(k)]$$

De factor waar we mee vermenigvuldigen hangt in principe af van alle voorgaande waarden $s(0)$ tot en met $s(m-1)$, maar hierboven hebben we al gezien dat de enige waarde die van belang is de voorgaande is. Eerder staat dat de kans afhangt of er een paar wordt gevormd of niet:

$$P[s(t+1) = S(t+1)] = \begin{cases} \frac{S(t)}{2n-t} \text{ als } S(t+1) = S(t) - 1 \\ 1 - \frac{S(t)}{2n-t} \text{ als } S(t+1) = S(t) + 1 \\ 0 \qquad \text{anders} \end{cases}$$

Wiskundigen hebben de ziekelijke neiging om zo'n formule met keuzes als een onbegrijpelijke gesloten uitdrukking op te schrijven. Ik kan natuurlijk niet achterblijven:

$$P[s(t+1) = S(t+1)] = \frac{1}{2} + [S(t+1) - S(t)] \cdot \left(\frac{S(t)}{2n-t} - \frac{1}{2}\right)$$

Nu kunnen we dus de kans op een gegeven verdeling precies uitschrijven door $m = t + 1$ te stellen:

$$P[s = S] = \prod_{0 \le t+1 \le n} P[s(t+1) = S(t+1)] =$$

$$= 1 \cdot \prod_{0 \le t < n} P[s(t+1) = S(t+1)] =$$

$$= \prod_{0 \le t < n} \frac{1}{2} + [S(t+1) - S(t)] \cdot \left(\frac{S(t)}{2n-t} - \frac{1}{2}\right)$$

We hebben niets aan deze formule, maar we kunnen in elk geval nu de kans op een gegeven functie uitrekenen. Hiermee kunnen we de uiteindelijke uitkomst verkrijgen:

$$\underset{s \in S}{E}\left(\max_{0 \le t \le 2n} s(t)\right) = \sum_{s \in S} \max_{0 \le t \le 2n} s(t) \cdot P(s) = S$$

Maar, zoals u ziet, echt handig is de formule niet: het is voor kleine waarden van *n* een hele berg rekenwerk, en we kunnen er ook geen schatting mee maken. In feite is dat wat we in tabel 3 hebben gedaan.

Tabel 4: Numerieke benadering voor drie paar sokken

k	*s(k)*	$\max_{0 \le t \le k} s(k)$	kans
0	0	0	1,000
1	1	1	1,000
2	0	1	0,200
2	2	2	0,800
3	1	1	0,200
3	1	2	0,400
3	3	3	0,400
4	0	1	0,067
4	0	2	0.133
4	2	2	0,400
4	2	3	0,400
5	1	1	0,067
5	1	2	0,533
5	1	3	0,400
6	0	1	0,067
6	0	2	0,533
6	0	3	0,400

Numerieke benadering

Als het gaat om een numerieke benadering, kunnen we een techniek gebruiken die op de vorige paragraaf is geïnspireerd: gewoon van $t = 0$ naar $t = 2n$ werken. We berekenen

$$\underset{s \in S}{E}\left(\max_{0 \le t \le k} s(t)\right)$$

waarbij we k laten lopen van 0 tot $2n$.

In iedere stap van deze bewerking moeten we de gegevens bewaren die nodig zijn in de volgende stappen. Het grappige is dat we niet alle kansen van alle $s \in S$ hoeven te berekenen, het gaat om de functies die een gegeven maximum hebben.

In woorden: we houden steeds bij, voor een voorstuk van het traject van 0 tot $2n$, wat de mogelijke maxima zijn, en welke kansverdeling daar bijhoort. Voor de volgende stap kunnen we dan gemakkelijk uitrekenen of het maximum verandert, en wat de nieuwe kansverdeling is. Aan het eind van de berekening kunnen we zo de hele kansverdeling van het maximum zien. Voor het verlengen van het traject met één sok hoeven we dan alleen de kansverdelingen door te rekenen, en te kijken wat er met het maximum gebeurt. Iets wiskundiger geformuleerd houden we de kansverdeling bij van de paren

$$\left(s(t), \max_{0 \le t \le k} s(t)\right)$$

De kans op zo'n paar is gemakkelijk te bepalen met de formule van het begin in combinatie met het het triviale feit

$$\max_{0 \le t \le k} s(t) \;=\; \max\left(s(k+1), \max_{0 \le t \le k} s(t)\right)$$

Een en ander is gedemonstreerd in tabel 4. U ziet dat de kansverdeling van het maximum aan het eind van de tabel staat; de verwachtingswaarde (of wat je maar wil) kan zo worden uitgerekend. Dit alles is in een computer te krijgen. Het is wat werk om te programmeren, maar we kunnen nu eindelijk tabel 2 uitbreiden met de echte verwachtingswaarden van het maximum, zoals te zien in tabel 5. U ziet het maximum oplopen, en ook steeds iets hoger dan de verwachtingswaarde. Dit klopt met mijn intuïtie dat de kans dat er in het verleden een groot aantal sokken was het gemiddelde ietsje hoger maakt.

Tabel 5: De juiste berekening van max s(t) voor n=6

$t =$	0	1	2	3	4	5	6	7	8	9	10	11	12
E =	0	1,0	1,8	2,5	2,9	3,2	3,3	3,2	2,9	2,5	1,8	1,0	1,0
max=	0	1,0	1,9	2,6	3,2	3,6	3,9	4,1	4,2	4,2	4,2	4,2	4,2

Even stug doorrekenen levert ons tabel 6 op met de uitkomst tot 30 paar sokken.

Tabel 6: Verwachtingswaarde van het maximum voor $1 \le n \le 30$

n	E	n	E	n	E
1	1,000	11	7,055	21	12,570
2	1,667	12	7,618	22	13,111
3	2,333	13	8,177	23	13,652
4	2,962	14	8,733	24	14,191
5	3,574	15	9,287	25	14,730
6	4,172	16	9,839	26	15,267
7	4,762	17	10,388	27	15,804
8	5,343	18	10,936	28	16,339
9	5,919	19	11,482	29	16,874
10	6,489	20	12,026	30	17,408

Nadelen van numerieke benaderingen

Nu hebben we een prachtige tabel, maar die zegt niets over hoe de verwachting verloopt voor zeer grote waarden van *n*. Dat is voor mijn originele probleem niet zo interessant, maar het is wel leuk vanuit een theoretisch perspectief. Ik heb me daar verder niet in verdiept, ook door gebrek aan wiskundige kennis, moet ik toegeven. De lijn is in elk geval net niet recht, zoals u kunt zien in figuur 3.

Uitbreidingen

We hebben nu het wiskundige probleem van het sokken vouwen redelijk onder de knie. Bij wiskundige benaderingen van huishoudproblemen is het meestal zo dat als gevolg van het modelleren een aantal praktische omstandigheden worden verwaarloosd. We zullen een aantal van deze omstandigheden in de formule proberen op te nemen.

Niet-blind grijpen

In de praktijk zal ik niet precies blindelings naar de volgende sok grijpen, maar toch met iets meer kans de bijpassende sok grijpen. Dit kan ontstaan doordat de sokken dicht bij elkaar in de wasmand liggen, of doordat ik expres ga zoeken als mijn schoot te vol is.

Het zou mooi zijn als ik kan kwantificeren hoeveel nut het heeft om eventjes rond te kijken op zoek naar de andere sok voordat ik een sok pak. Het lijkt me duidelijk dat het verwachte aantal plaatsen op schoot dan minder wordt, maar hoeveel? Om te beginnen onderzoeken we het effect als ik steeds *a* sokken bekijk, en als daar een sok tussen zit waarvan de bijpassende sok op mijn schoot ligt, dan pak ik die. De kans P_{los} is dan de kans dat alle sokken die ik pak niet op mijn schoot liggen; dit berekenen we met de 'loterijformule'

$$P_{los}(t) = \frac{\binom{2n-t-s(t)}{a}}{\binom{2n-t}{a}}$$

en natuurlijk

$$P_{paar}(t) \;=\; 1 - P_{los}(t)$$

(Opgave voor de lezer: controleer dat dit het geval $a = 1$ klopt met de oorspronkelijke formules.) Als we ons voorbeeld voor $n = 6$ opnieuw doorrekenen, krijgen we tabel 7. De verschillen zijn schokkend: om te beginnen is de kans dat van alle paren een sok op mijn schoot terecht komt bijna 0 geworden. Maar nog schokkender is dat de symmetrie is verdwenen! Het moment dat het verwachte aantal sokken maximaal is ligt nu duidelijk bij $t = 5$, terwijl dat oorspronkelijk in het midden bij $t = 6$ was.

Tabel 7: s(t) voor n=6; steeds twee sokken bekijken

i													
i=6							0						
5						5		2					
4					21		19		6				
3				51		53		45		18			
2			82		68		68		73		45		
1		100		49		42		53		82		100	
0	100		18		11		12		21		55		100
$t =$	0	1	2	3	4	5	6	7	8	9	10	11	12

Na wat verder rekenen heb ik de resultaten uit de grafiek in figuur 3 kunnen berekenen. U kunt goed zien dat het inderdaad nut heeft om meerdere sokken te bekijken.
Bedenk wel dat de berekening ervan uitgaat dat als je meerdere sokken pakt, je wel steeds opnieuw aselect de

groepjes pakt. Als je dit niet doet, werkt het minder goed, want de sok die de vorige keer niet bij een sok op mijn schoot paste, past vermoedelijk nog steeds niet.

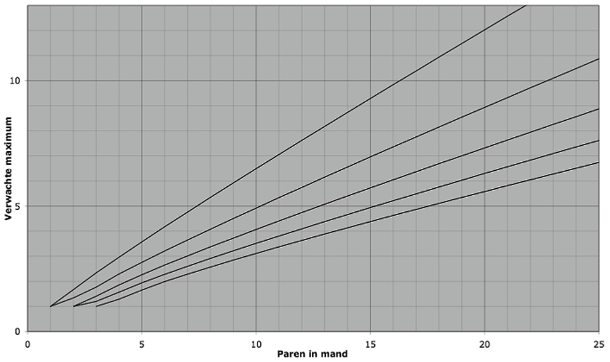

fig. 3 Meerdere sokken tegelijk pakken

Een andere ontdekking die ik deed, is dat het verschil tussen het verwachte maximum en het maximum van de verwachting steeds groter wordt als je meerdere sokken tegelijk bekijkt. Dit komt omdat de periode waarin m'n schoot bijna vol is, langer wordt. Hierdoor is het maximum meer afhankelijk van toevallige uitschieters in het aantal sokken. Dit zie je niet als je het maximum neemt van de verwachtingswaarde, maar wel als je de verwachtingswaarde van het maximum neemt. Dit was het punt waarop ik bij de eerste versie van het artikel ondekte dat er iets mis was. Toen kon ik weer helemaal opnieuw beginnen...

Losse sokken

Afhankelijk van de organisatiegraad van uw huishouden blijven er soms een paar losse sokken in de mand zitten, die geen bijpassende sok hebben. (Zo'n ding heet bij ons een 'oneven sok', maak ik weet niet of dit in Van Dale staat.) We gaan er even van uit dat er één sok ontbreekt in de mand. Je kunt het zien als een sokkenvouwproces waarbij één stap voor het einde wordt gestopt. De uitkomst voor een mand met een ontbrekende sok is dus hetzelfde als wanneer de sok er wel was.

Ideeën om ruimte op mijn schoot te sparen

De vraag is of er nog andere slimme dingen zijn te bedenken die zorgen dat het geheel op mijn schoot blijft passen. Een idee is om de sokken die niet meer op mijn schoot passen gewoon in de mand terug te gooien. Het opvouwen duurt dan wel langer natuurlijk, omdat ik vaker sokken moet pakken; het kan zelfs willekeurig lang duren! Tabel 8 laat zien hoeveel extra tijd het kost om sokken terug te gooien in de mand. In de kolommen 'extra' is te zien hoe vaak je een sok moet teruggooien. Hieruit blijkt het volgende:

- Teruggooien is geen slecht idee als je schoot bijna groot genoeg is.
- Je moet niet te gauw sokken teruggooien, anders blijf je bezig.
- Het bekijken van meerdere sokken tegelijk bespaart tijd, als je beperkte schootruimte hebt.

Tabel 8: sokken teruggooien bij 16 paar in de mand

	1 bekijken		2 bekijken		3 bekijken		4 bekijken	
Schoot	E max	extra	E max	extra	E max	extra	E max	extra
1	1,000	225,0	1,000	105,0	1,000	65,3	1,000	45,5
2	2,000	102,5	2,000	47,9	2,000	29,6	2,000	20,4
3	3,000	61,3	3,000	28,2	3,000	16,9	3,000	11,1
4	4,000	40,3	4,000	17,7	3,994	9,8	3,962	5,6
5	5,000	27,5	4,990	10,9	4,905	5,0	4,695	2,2
6	5,999	18,7	5,909	6,1	5,560	2,0	5,065	0,6
7	6,985	12,2	6,631	2,8	5,896	0,6	5,186	0,1
8	7,918	7,3	7,071	1,0	6,015	0,1	5,212	0,0
9	8,714	3,8	7,271	0,3	6,044	0,0	5,215	0,0
10	9,289	1,7	7,338	0,1	6,049	0,0	5,216	0,0
∞	9,839	0,0	7,357	0,0	6,049	0,0	5,216	0,0

Een ander idee is om een aparte stapel te maken voor sokken die niet meer op mijn schoot passen. Als er ruimte op mijn schoot ontstaat, vul ik die op met sokken van de stapel. Dit kost natuurlijk een extra stapeltje naast de wasmand, maar het levert wel wat op. En zo kun je doorgaan met dingen uitrekenen ...

Conclusie

Ik hoop dat u na dit artikel geïnspireerd bent om ook om u heen te kijken en de wiskunde te zien die ons omringt. Verder sta ik open voor ideeën om het sokkenprobleem verder uit te werken: is er nog een manier te bedenken om het aantal sokken op mijn schoot te reduceren? Kun je zo'n oplossing ook nog doorrekenen? Ik ben benieuwd.

Jurjen Bos
Interpay Nederland B.V., Utrecht