Record-problemen

R. Jansen/S.L. Kemme
RU Groningen

Samenvatting

De auteurs beschrijven hoe zij, gegrepen door een record-probleem, dit probleem geleidelijk aan opgelost hebben. Hierbij hebben ze een breed instrumentarium aan wiskundige technieken gebruikt. In het bijzonder spelen intuïtie en computersimulatie een belangrijke rol.

Inleiding

Op school leer je sommen maken. Later, in je wiskundestudie, wordt dat niet veel anders. De sommen zijn moeilijker. Je moet er langer over nadenken. Vaak moet je iets bewijzen. Maar het blijven sommen. Pas aan het eind van je studie ga je je een beetje bezig houden met problemen waarvan de oplossing niet allang bekend is bij je docenten. Dan pas kun je je misschien een beetje creatief gaan inleven in het vak waar je al zo lang mee bezig bent. Je ontdekt dat het vak helemaal niet zo systematisch deductief in elkaar zit, dat de boeken een wat groot uitgevallen samenvatting zijn van gedaan onderzoek, dat wiskunde een vak kan zijn vol onverwachte ogenblikken, fascinerende vragen, gevoelens van spanning, vreugde en teleurstelling.
Dan ga je het onderwijs in en houd je je verder weer bezig met sommen. Sommen opgeven, sommen voormaken, sommen uitleggen, sommen bedenken, sommen nakijken. Voor al dat andere heb je weinig tijd meer.

Het was daarom voor ons een hele verrassende ervaring om te ontdekken dat een wiskundig probleem, waarvan we dachten dat het gewoon een grote moeilijke som was, ineens vol allerlei onverwachte vragen, strategieën en oplossingen bleek te zitten en dat we daar zelf creatief wiskundig mee aan de gang konden gaan.

In dit artikel willen we dat oplossingsproces zo getrouw mogelijk weergeven. Het artikel is bedoeld als een uitnodiging aan alle wiskundedocenten om de moed niet te laten zakken en het proefwerk, dat nodig nagekeken moet worden, een keer opzij te leggen en gewoon ouderwets met je gezonde wiskundeverstand een avondje wiskunde te bedrijven.

Het probleem

Arthur Engel heeft eens, we weten niet meer waar, het volgende probleem opgeschreven:

> Je kiest aselect een getal uit de verzameling 1,2,...100.
> Dat eerste getal is meteen je eerste record, want het is het hoogste getal tot nu toe. Vervolgens kies je weer aselect een getal uit dezelfde verzameling. Is dit getal kleiner of gelijk aan het vorige, dan gooi je het weg; is het groter, dan is dit je nieuwe record. Zo ga je door tot je record 100 is geworden. Je telt alleen maar het aantal records (en noemt dit R).
> *Vraag:* Wat is de verwachtingswaarde voor het aantal records?

Zo op het eerste gezicht lijkt dit een doodgewone kansrekeningsom. Je probeert gewoon alle mogelijke gevallen uit te schrijven en daar de kansen bij uit te rekenen.
Met andere woorden: je bepaalt de kansverdeling voor R. Als je daaraan begint verzeil je algauw in een oerwoud van gevallen. Dat we daar niet uitkomen ligt natuurlijk aan ons, we zijn gewoon niet (meer?) behendig genoeg voor dit soort sommen.

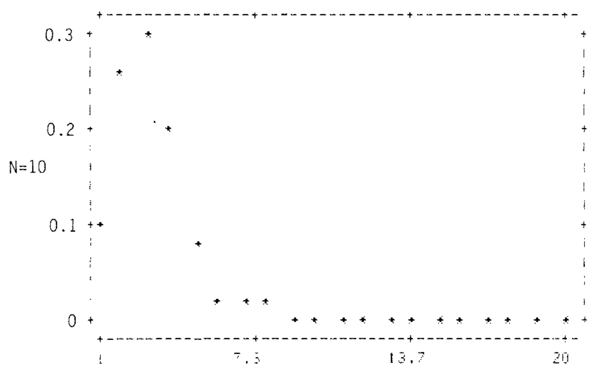
Je kunt het probleem echter wel heel gemakkelijk simuleren met behulp van een computer. We besluiten het probleem op de HEWET-nascholingscursus te presenteren en denken daarmee twee vliegen in één klap te vangen: je kunt mooi zien hoe je de computer kunt gebruiken als 'onderzoeks'-instrument bij theoretische problemen èn wat een kansverdeling eigenlijk is. Aan de bovenstaande vraag worden nog enkele vragen toegevoegd.

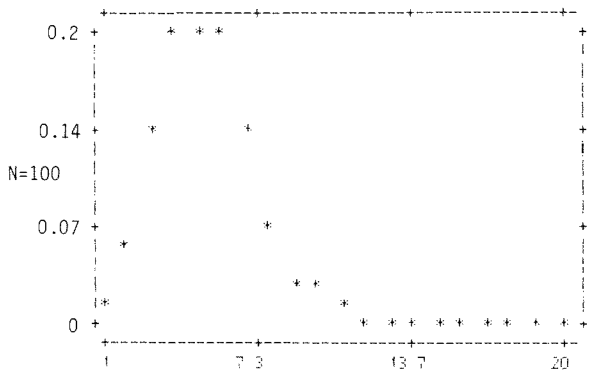
Vraag: wat gebeurt er met die verwachtingswaarde als je de grens 100 vervangt door een andere: 1000, 10, N?
Hoe ziet de kansverdeling er dan uit?

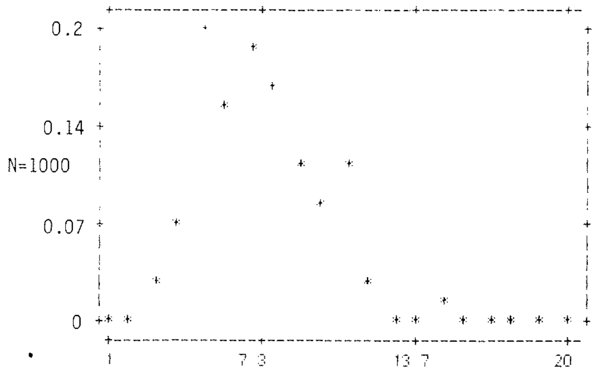
Op de cursussen wordt er ijverig aan gerekend. Velen raken geboeid door het probleem (al weet niemand waarom). Het laat ons ook niet los. Ieder jaar proberen we weer of we het probleem ook gewoon op kunnen lossen. We worden daarbij trouw ontmoedigd door prof. dr. W. Schaafsma die telkens opnieuw langs zijn neus weg verklaarde dat "record-problemen vréselijk ingewikkeld zijn".

Het eerste vermoeden

Op de cursussen vragen we naar de kansverdeling en de verwachtingswaarde voor R als de grenzen 10, 100 respectievelijk 1000 zijn. Met behulp van structuurdiagrammen worden BASIC-programma's geschreven en ingevoerd in de computer. Voor 10 en 100 zijn de resultaten al na een kwartier bekend. Voor het geval 1000 duurt het wat langer.







De gevonden kansverdelingen zien er erg scheef uit. Voor R = 1 kun je de kans zo uitrekenen. Dan moet het getrokken getal namelijk meteen de eindgrens zijn en de kans daarop is natuurlijk 0.1, 0.01 of 0.001. Dat is algauw erg klein.
De gevonden verwachtingswaarden zijn:

N	E(R)
10	2.91
100	5.15
1000	7.32

Als N 10 keer zo groot wordt, neemt E(R) met een constante waarde (2.2) toe. Natuurlijk zijn drie waarnemingen veel te weinig om een verantwoorde uitspraak te kunnen doen. Toch ontstaat er vanwege de onverwachte langzame toename van E(R) een vermoeden.

De verwachtingswaarde is een logaritmische functie van de eindgrens.

Verder dan dit vermoeden komen we voorlopig niet. Het probleem verdwijnt ieder jaar weer in de koelkast totdat we een nieuwe HEWET-cursus gaan organiseren.

Regressie (de missing link) en het tweede vermoeden

Het Praediniusgymnasium in Groningen heeft de hand weten te leggen op een pracht pakket van computerprogramma's voor wiskunde A. Het is door een Engelse groep leraren ontwikkeld voor de BBC computer en nu, helaas, alleen nog maar in het Engels beschikbaar. [1] De mogelijkheden variëren van matrixrekening, statistiek tot regressieanalyse. Vooral dat laatste kunnen we goed gebruiken om het gebruik van logaritmisch papier te demonstreren. Door de assen te transformeren tot logaritmische schalen en daarna een rechte lijn door de punten te tekenen, spoort de computer een verband op tussen twee variabelen. We besluiten het programma los te laten op ons record probleem. Eerst bepalen we een tiental waarden voor E(R) als N met stapgrootte 10 van 10 naar 100 loopt. De gevonden waarden van E(R) worden het regressieprogramma ingevoerd waarbij we voor de horizontale as een logaritmische schaal kiezen vanwege ons logaritmische vermoeden.
De computer vindt:

$$E(R) = 0.99 * \ln N + 0.69$$

Van die 0.99 maken we natuurlijk meteen 1. Met 0.69 zitten we een beetje in onze maag. Natuurlijk zijn dit geschatte waarden die aan toeval onderhevig zijn. We weten alleen niet hoe groot die toevalsschommelingen kunnen zijn.

Als N = 1 dan moet E(R) = 1. Dat geeft het vermoeden dat:

$$E(R) = \ln N + 1$$

Een mooie eenvoudige formule die strookt met alles wat we tot nu hebben gevonden. Er kan gewoon niets meer misgaan. We moeten alleen nog wat theoretische argumenten zien te vinden die onze formule bewijzen.

Een slimmerik, die we de formule hebben voorgelegd, beweert dat je dat gewoon even met genererende functies kunt opschrijven. Hij heeft vast gelijk. Maar als hij dat kan, dan kunnen wij dat ook.

Een doorbraak

Tot zover was alles alleen nog maar empirie. Op basis van een stel waarnemingen hebben we een vermoeden uitgesproken (het eerste). Dit hebben we getoetst aan meer waarnemingen waardoor het vermoeden werd aangescherpt tot een mooie formule. We besluiten die formule (het tweede vermoeden) ook weer te toetsen aan verdere waarnemingen. Daarvoor kiezen we kleine waarden van N. De verandering van 0.69 tot 1 werd immers ingegeven door voor N de waarde 1 te kiezen.

Zo vinden we de volgende tabel:

N	E(R)	ln N + 1
1	1	1
2	1.46	1.69
3	1.87	2.09
4	2.04	2.38
5	2.34	2.60
6	2.33	2.79
7	2.56	2.94
8	2.74	3.08
9	2.79	3.20
10	2.75	3.30

Hoewel de waarden van E(R) nogal grillig schommelen, is er toch een duidelijk toenemende trend te bespeuren. De verschillen met de derde kolom zijn echter steeds groot. Zelfs zó groot dat we vinden dat die formule gewoon niet klopt. Wat nu? Voor die kleine waarden van N kun je misschien toch nog wel iets met de hand uitrekenen. Als N = 1, is de zaak heel eenvoudig: er is maar één getal en dat is meteen het record, dus E(R) = 1.

Stel N = 2. Dan zijn de getallen 1 en 2. De kans op R = 1 is 0.5, als je meteen het getal 2 kiest. Hoe kan R = 2 optreden? Je moet dan altijd eerst een 1 trekken. Vervolgens kun je nog een hele rij enen trekken, het doet er niet toe hoeveel, als je maar met een 2 eindigt.

Dus deze rijtjes (de kans staat erachter):

1 2	1/4
1 1 2	1/8
1 1 1 2	1/16
.	
$\underbrace{1\ 1\ 1\ 1 \ldots\ldots 1\ 2}_{n}$	$\frac{1}{2^n}$

De totale kans is dus:

$$\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots = \sum_{n=2}^{\infty}\frac{1}{2^n}=\frac{1}{4}\sum_{n=0}^{\infty}\frac{1}{2^n}=\frac{1}{4}\cdot 2=\frac{1}{2}$$

En $E(R) = 0.5 \cdot 1 + 0.5 \cdot 2 = 1.5$.
Dat ligt mooi dichtbij de gevonden 1.46. Veel mooier dan de 1.69 van ln 2 + 1.
Voor N = 3 lukt het ook nog wel om alles door te rekenen, maar dat kost al heel wat meer moeite. Voor E(R) vinden we 1.83. Ook weer mooi dichtbij de experimenteel gevonden 1.87.
Het geval N = 4 laten we maar zitten, dat wordt echt te ingewikkeld.

Om wat overzicht te krijgen kun je natuurlijk ook alleen maar de records opschrijven. We bekijken het geval N = 3.
Dan krijg je de volgende combinaties:

	record(s)
R = 1	3
R = 2	1 3
	2 3
R = 3	1 2 3

Alle gekozen getallen tussen de records zijn dus, voor het overzicht, weggelaten. Maar kun je niet hieruit ook meteen de kansen uitrekenen?
Even proberen.

	record(s)	kans	verklaring
R = 1	3	$\frac{1}{3}$	want keuze uit 1, 2 en 3
R = 2	1 3	$\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{6}$	want eerst keuze uit 1, 2 en 3 en daarna alleen uit 2 en 3
	2 3	$\frac{1}{3}\cdot 1$	want voor het record 3 heb je geen keuze meer
R = 3	1 2 3	$\frac{1}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{6}$	zie boven

Dat klopt precies met onze eerste rekenpartij voor N = 3! Dan is:

$$E(R) = \frac{1}{3}\cdot 1 + \left(\frac{1}{6}+\frac{1}{3}\right)\cdot 2 + \frac{1}{6}\cdot 3 = \frac{1}{3}+1+\frac{1}{2} = 1\frac{5}{6} = 1.83$$

Dat geeft moed. In deze afdeling zit echter een gewaagde veronderstelling. We baseren alles op een uitkomstenruimte van records {1.2.3} met gelijke kans, waaruit zonder teruglegging wordt getrokken.

Dat niet-terugleggen is zelfs een soort wegpakken. Als je bijvoorbeeld meteen het record 2 trekt, verdwijnt ook het record 1 uit de uitkomstenruimte. Dus trekking van een record heeft tot gevolg dat ook alle kleinere records uit de uitkomstenruimte worden genomen.

Deze aanname wordt nog eens ondersteund door het computerprogramma. We geven alleen het structuurschema.

```
N=0
Herhaal
    S=0 : N=N+1
    Herhaal 100 maal
        R=0 : T=0 : X=0
        Herhaal zolang R< N
            Herhaal zolang X<=R
                X=Random [1 .. N]
            R=X
            T=T+1
        S=S+T
    Print S/100
Totdat N=10
```

In de binnenste herhaling kies je een X waar je alleen maar iets mee doet als X>R. Je kunt dus net zo goed X random kiezen uit [R+1N]. Die binnenste herhaling kun je er eigenlijk gewoon uitknippen en vervangen door een aselecte keuze uit $\{R+1, R+2, \ldots, N\}$. De binnenste herhaling wordt dan:

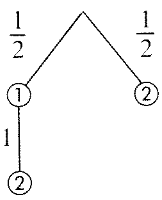
```
Herhaal zolang R<N
    R=Random [R+1 .. N]
    T=T+1
```

Daar wordt het programma meteen een stuk sneller van. Bovendien vinden we hier weer onze vreemde vorm van teruglegging terug.
Al met al komen we tot de volgende aanname:

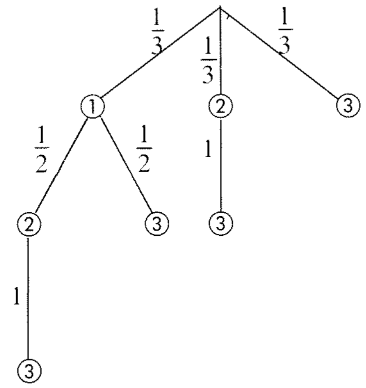
> Het record probleem laat zich beschrijven door een trekking zonder teruglegging en met wegname van lagere uitkomsten uit de verzameling records.

Een formeel bewijs voor deze aanname hebben we niet. Maar dat interesseert ons op dit ogenblik ook niet zo. Eerst maar eens kijken of we hiermee verder kunnen komen.
Misschien geven boomdiagrammen nog wat meer overzicht.

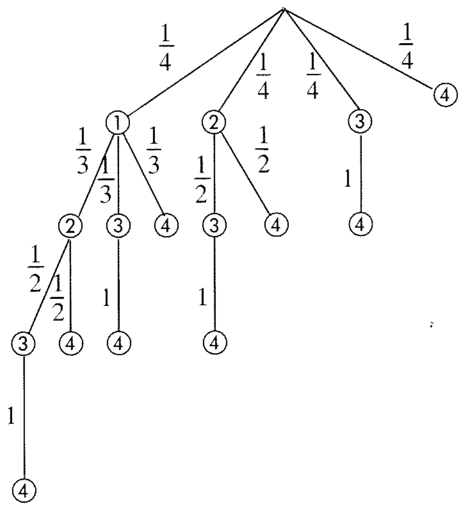
N = 2



N = 3



N = 4



Dat gaat mooi. Even E(R) uitrekenen:

$$E(R) = \frac{1}{4}\cdot 1 + (\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{3})\cdot 2 + $$
$$+ (\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{1}{3}\cdot\frac{1}{2})\cdot 3 +$$
$$+ \frac{1}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot 4 = 2\frac{1}{12} = 2.08$$

We vonden 2.04. Het blijft kloppen! We hebben het geval N=4 nu ook gekraakt. De boom voor N=5 wordt ons te ingewikkeld. Bovendien lokt dat gereken met breuken ook niet erg. Even kijken of er één of andere regelmaat valt te ontdekken:

N	E(R)	verschil
1	1	
		$\frac{1}{2}$
2	$1\frac{1}{2}=\frac{3}{2}$	
		$\frac{1}{3}$
3	$1\frac{5}{6}=\frac{11}{6}$	
		$\frac{1}{4}$
4	$2\frac{1}{12}=\frac{25}{12}$	

Dat is haast te mooi om waar te zijn.
We gokken: voor N=5 geldt:

$$E(R) = \frac{25}{12} + \frac{1}{5} = \frac{137}{60} = 2.28$$

We vonden 2.34. Dat is nogal wat hoger. Maar de volgende, bij N=6, was 2.33, dat kan een aanwijzing zijn dat die 2.34 door het toeval zo hoog is uitgevallen. Zo vinden we een heel andere formule voor E(R), namelijk:

$$E(R) = \sum_{k=1}^{N} \frac{1}{k}$$

Meteen springt er een lamp op groen. Heel vroeger leerden we bij het vak functie-theorie, dat deze reeks zich voor grote waarden van N gedraagt als:

$$\ln(N) + j$$

Hierin is j het zogenaamde getal van Euler: $j = 0.58$. Nu kunnen we niet meer stuk. Dit klopt prima met de eerder gevonden benadering:

$$E(R) = \ln(N) + 0.69$$

Zo komen we tot de notatie:

$$E_N(R) = \sum_{k=1}^{N} \frac{1}{k}$$

Het ligt voor de hand om dit vermoeden opnieuw te confronteren met de harde werkelijkheid. Weer is de computer hierbij een onmisbaar instrument. We nemen nu kleine èn grote waarden van N:

N	E(R)	
1	1	1
2	1.46	1.5
3	1.87	1.83
4	2.04	2.08
5	2.34	2.28
6	2.33	2.45
7	2.56	2.59
8	2.74	2.71
9	2.79	2.82
10	2.75	2.92
20	3.57	3.60
30	3.91	3.99
40	4.36	4.28
50	4.35	4.50
60	4.24	4.68
70	5.10	4.83
80	5.00	4.97
90	4.97	5.08
100	5.23	5.19

Afgezien van lokale toevallige schommelingen blijft er toch een redelijke overeenstemming tussen de twee tabellen. Als het verschil groot is, zijn er duidelijke aanwijzingen dat het getal in de middenkolom te hoog of te laag is als we naar de vorige en de volgende getallen kijken. Dit versterkt ons vermoeden tot zekerheid. Het is nu een stelling geworden waarvan we het bewijs moeten zien te vinden.

Bewijzen

Bewijzen is blijkbaar een bezigheid die in alle eenzaamheid dient te geschieden. Tot op dit moment was er een voordurende wisselwerking van suggesties van verschillende kanten.

Hier hebben we ieder voor zich naar een bewijs zitten zoeken. Het leuke is achteraf dat dit twee volledig andere verhalen oplevert.

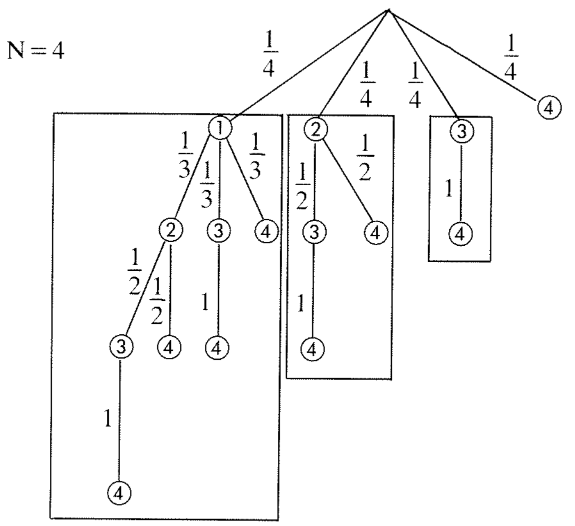
Het ene verhaal
Het vermoeden is dat:

$$E_N(R) = \sum_{k=1}^{N} \frac{1}{k}$$

Dat betekent dat:

$$E_{N+1} = E_N + \frac{1}{N+1}$$

$\frac{1}{N+1}$ is de kans op één record bij bovengrens N+1. Herleiding van het geval N+1 tot het geval N. In ons achterhoofd hebben we dan een bewijs met volledige inductie. Eerst nog maar weer even naar de boom kijken of we ook een regelmaat kunnen ontdekken:



In de boom vinden we niet terug:

$$E_{N+1}(R) = E_N(R) + \frac{1}{N+1}$$

Het geval N=4 blijkt een combinatie van de bomen voor N=1, 2 en 3. Voor de bomen bij N=2 en 3 vinden we een soortgelijke regelmaat. Met het startpunt meegerekend, hebben deze bomen allemaal één weg meer. Voor de verwachtingswaarde betekent dit:

$$E_4 = \frac{1}{4} + \frac{1}{4}(E_1 + 1) + \frac{1}{4}(E_2 + 1) + \frac{1}{4}(E_3 + 1) =$$

$$= 1 + \frac{1}{4}(E_1 + E_2 + E_3)$$

Of algemeen:

$$E_{N+1} = 1 + \frac{1}{N+1}(E_1 + E_2 + \ldots + E_N) \qquad (*)$$

Stel nu (dit is de inductie-aanname) dat:

$$E_N = E_{N-1} + \frac{1}{N}$$

Omdat ook:

$$E_N = 1 + \frac{1}{N}(E_1 + E_2 + \ldots + E_{N-1})$$

dus:

$$E_1 + E_2 + \ldots + E_{N-1} = N(E_N - 1)$$

kunnen we (*) omzetten in:

$$\begin{aligned} E_{N+1} &= 1 + \frac{1}{N+1}(N(E_N - 1) + E_N) = \\ &= 1 + E_N - \frac{N}{N+1} = \\ &= E_N + \frac{1}{N+1} \end{aligned}$$

Mooi, dat is gelukt. Zwak punt is het bewijs van (*). Dit is gewoon afgeleid uit de regelmaat in de boomdiagrammen voor $N=2$, 3 en 4. Voor een volledig bewijs zullen we dieper in de theorie van de verwachtingswaarden moeten duiken. Het patroon in de boomdiagrammen is natuurlijk wel weer de zoveelste aanwijzing dat we op de goede weg zitten.

Het andere verhaal
Het vermoeden is dat:

$$E_N(R) = \sum_{k=1}^{N} \frac{1}{k}$$

Aangezien we ook wat over de kansverdeling van R willen weten, eerst maar weer de situatie voor $N=3$, 4 en 5.

N=3	N=4	N=5
3	4	5
13 23	14 24 34	15 25 35 45
123	124 134 234	125 135 145 235 245 345
	1234	1235 1245 1345
		12345
4 uitkomsten	8 uitkomsten	16 uitkomsten

Het aantal mogelijke uitkomsten verdubbelt telkens. Hierdoor ontstaat het vermoeden dat een uitkomst bij $N=3$ twee andere uitkomsten bij $N=4$ voortbrengt. Na wat proberen komt de volgende mogelijkheid tot stand:

$3 \to 4$, $3 \to 34$; $13 \to 14$, $13 \to 134$; $23 \to 24$, $23 \to 234$; $123 \to 124$, $123 \to 1234$

In het algemeen brengt een uitkomst $a_1, \ldots a_j$, N de uitkomsten:

$a_1, \ldots a_j, N+1$ en $a_1, \ldots a_j, N, N+1$ voort.

Dit levert nog niet zoveel op. Eigenlijk wil ik ook de andere kant op, want in mijn achterhoofd denk ik aan een of andere recurrente betrekking.

$N=4$ en $R=3$: $\quad 123 \to 124 \qquad 13 \to 134 \quad 23 \to 234$

$\underbrace{123 \to 124}_{N=3 \text{ en } R=3} \qquad \underbrace{13 \to 134 \quad 23 \to 234}_{N=3 \text{ en } R=2}$

Hier ontstaat het volgende vermoeden:

$$P_4(R=3) = f(P_3(R=3), P_3(R=2))$$

Of algemener:

$$P_N(R=k) = f(P_{N-1}(R=k), P_{N-1}(R=k-1)).$$

Eerst maar weer eens proberen bij $N=3$ en $N=4$.

$$P_3(R=1) = \frac{1}{3}; \quad P_3(R=2) = \frac{1}{3}(\frac{1}{2} + 1);$$

$$P_3(R=3) = \frac{1}{3} \cdot \frac{1}{2} \cdot 1$$

$$P_4(R=1) = \frac{1}{4}; \quad P_4(R=2) = \frac{1}{4}(\frac{1}{3} + \frac{1}{2} + 1);$$

$$P_4(R=3) = \frac{1}{4}(\frac{1}{6} + \frac{1}{3} + \frac{1}{2})$$

$$P_4(R=4) = \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 1$$

Nog niet op het eerste gezicht duidelijk wat het moet worden. Maar ik weet nu waar ik naar zoek:

$$P_4(R=3) = \frac{1}{4}(\underbrace{\frac{1}{6} + \frac{1}{3}}_{P_3(R=2)} + \underbrace{\frac{1}{2}}_{3 \cdot P_3(R=3)})$$

Dus:

$$P_4(R=3) = \frac{1}{4} P_3(R=2) + \frac{3}{4} P_3(R=3)$$

Wat proberen bij $N=5$ levert op:

$$P_5(R=3) = \frac{1}{5} P_4(R=2) + \frac{4}{5} P_4(R=3)$$

Dit leidt tot het vermoeden:

$$P_N(R=k) = \frac{1}{N} P_{N-1}(R=k-1) + \frac{N-1}{N} P_{N-1}(R=k)$$

Het bewijs door inductie lijkt mij nog een hele klus, eerst maar eens kijken wat er bij de 'randen' gebeurt:

$$\begin{aligned} P_4(R=1) &= \frac{1}{4} P_3(R=0) + \frac{3}{4} P_3(R=1) \\ &= \frac{1}{4} P_3(R=0) + \frac{3}{4} \cdot \frac{1}{3} \end{aligned}$$

Als we afspreken dat $P_3(R=0) = 0$ wat voor de hand ligt, dan is de uitkomst 1/4.
Op deze manier vinden we een recurrente betrekking voor de kansverdeling. Ook nu zouden we met de computer hiermee de kansen kunnen bepalen. Zijn de kansen opgeteld wel 1?

$$\begin{aligned} \sum_{k=1}^{N} P_N(R=k) &= \frac{1}{N} \sum_{k=1}^{N} P_{N-1}(R=k-1) + \frac{N-1}{N} \sum_{k=1}^{N} P_{N-1}(R=k) = \\ &= \frac{1}{N} \sum_{k=2}^{N} P_{N-1}(R=k-1) + \frac{N-1}{N} \sum_{k=1}^{N} P_{N-1}(R=k) = \\ &= \frac{1}{N} \sum_{k=1}^{N-1} P_{N-1}(R=k) + \frac{N-1}{N} \cdot 1 = \\ &= \frac{1}{N} \cdot 1 + \frac{N-1}{N} = 1 \end{aligned}$$

Nu nog even de verwachtingwaarde E_N uitrekenen:

$$E_N = \sum_{k=1}^{N} k \cdot P_N(R=k) =$$

$$= \frac{1}{N} \sum_{k=1}^{N} k \cdot P_{N-1}(R=k-1) + \frac{N-1}{N} \sum_{k=1}^{N} k \cdot P_{N-1}(R=k) =$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} (k+1) \cdot P_{N-1}(R=k) + \frac{N-1}{N} \sum_{k=1}^{N-1} k \cdot P_{N-1}(R=k) =$$

$$= \frac{1}{N} \sum_{k=1}^{N-1} (k+1) \cdot P_{N-1}(R=k) + \frac{N-1}{N} \sum_{k=1}^{N-1} k \cdot P_{N-1}(R=k) =$$

$$= \frac{1}{N} \cdot E_{N-1} + \frac{1}{N} \sum_{k=1}^{N-1} P_{N-1}(R=k) + \frac{N-1}{N} \cdot E_{N-1} =$$

$$= \frac{1}{N} E_{N-1} + \frac{1}{N} \cdot 1 + \frac{N-1}{N} \cdot E_{N-1} = E_{N-1} + \frac{1}{N}$$

Tevreden besluit ik de volgende dag mijn verhaal te laten zien. Er blijft nog genoeg over: het bewijs van de recurrente betrekking en het berekenen van de kansverdeling voor bepaalde waarden van N met behulp van de computer. Zou die kansverdeling misschien gewoon de Poisson-verdeling blijken te zijn? Als je naar de vorm kijkt dan lijkt dat er wel erg op. Uit de recurrente betrekking zou je eigenlijk ook nog een echte formule voor de kansverdeling moeten kunnen destilleren. Voorlopig besluit ik nog even te kijken wat ik morgen in klas 3 moet doen.

Tot slot

De bewijzen blijven een zwak punt. Eigenlijk is er nog niets strikt wiskundig bewezen. We hebben alleen zo veel mogelijk ondersteunend materiaal gevonden. Toch laten we het hierbij. We weten nu hoe het verder gaat, we hebben vat op het probleem gekregen en daarmee het gevoel dat we de zaak hebben opgelost, hoe moeizaam het formele vervolg ook nog mag zijn.

Inmiddels heeft Gerard Kroon, één van onze HEWET-cursisten die door het probleem gegrepen werd, een aantal formele bewijzen geleverd èn een expliciete formule voor de kansverdeling waaruit blijkt dat het niet de Poisson-verdeling is.

Het is interessant om achteraf nog eens na te gaan wat er allemaal een rol heeft gespeeld bij het oplossingsproces.
We noemen:

- Een onrustig gevoel over het probleem. Zo van: dat moeten we toch de baas kunnen!
- Uitgedaagd worden door het probleem: het ziet er zo eenvoudig uit, tegelijkertijd weet je dat dat niet zo is.
- Een vorm van rivaliteit: als hij (de slimmerik) beweert dat hij het zo maar even opschrijft, dan moeten wij dat toch ook kunnen?
- De durf om gewoon maar eens wat aan te rommelen met vermoedens, computers, aannames, vage theoretische overwegingen, enz.
- De systematiek van het voortdurend heen en weer gaan tussen vermoeden en testen, bijstellen van het vermoeden en weer testen.
- Het brede karakter van de oplossing. Er komt van alles bij te pas: kansrekening, programmeren, boomdiagrammen, functietheorie, het gebruik van standaardprogramma's, regressie-analyse, interpreteren van gegevens, getalpatronen, de waarde van een uitspraak (vermoeden, bewijs, aanname, ...) kunnen vaststellen en daarna kunnen handelen.
- De manier waarop je elkaar stimuleert. De één begint net een vermoeden, de ander rekent dat wat verder door en weet dat om te zetten in een formule, waarop in samenspraak de formule op juistheid wordt onderzocht, enz. Bij een dergelijk 'exploratief' onderzoek weten ook bij wiskunde twee meer dan één.

Literatuur

[1] M.E.I., 41a West Street, Oundle, Peterborough PE8 4EJ.