Kansrekening in het kort

Vincent Icke
Sterrewacht Leiden

Beschrijven van kansen

Het is de bedoeling te zien hoe we kansen in getallen kunnen uitdrukken, en ermee rekenen. Als je zeker weet dat iets niet gebeurt, noemen we de kans op die gebeurtenis 0. Als je zeker weet dat iets wèl gebeurt, noemen we de kans op die gebeurtenis 1. Dat getal mag van alles wezen, bijvoorbeeld 8 of 100, als je je maar aan de eenmaal gemaakte keuze houdt. Vaak gebruikt men 100, en noemt de kans dan een 'percentage'. Soms gebruikt men 1.000.000, en noemt de kans een ppm (*parts-per-million*). Maar de waarde 1 rekent het handigst. Bij deze keuze kunnen we de kans of waarschijnlijkheid P (voor *probabilitas*) uitrekenen, door het aantal malen N_A dat een bepaalde gebeurtenis A voorvalt te delen door het totale aantal gebeurtenissen N van welk type dan ook. Zo is de kans dat een willekeurig iets gebeurt, een getal van 0 tot en met 1:

$$P_A = \frac{N_A}{N} \tag{1}$$

Hierbij veronderstellen we wel dat we te maken hebben met gebeurtenissen die elkaar uitsluiten, en die gezamenlijk alle mogelijkheden overdekken; bijvoorbeeld of een persoon man of vrouw is, of dat een van de zes zijden van een dobbelsteen bovenkomt. Dan is de som van die kansen gelijk aan 1, want er gebeurt altijd *iets* (alle mogelijkheden zijn overdekt) en er gebeuren nooit twee dingen tegelijk (de mogelijkheden sluiten elkaar uit). Als $\diamondsuit$ staat voor de gebeurtenis 'ruitenaas trekken uit een compleet spel kaarten', dan is

$$P_{\diamondsuit} = \frac{1}{52} = 0,01923 = 1,92\,\% \tag{2}$$

Als ♣ staat voor de gebeurtenis ‘een klaverkaart trekken’ dan is

$$P_{\clubsuit} = \frac{13}{52} = 0,25 \tag{3}$$

Zo kunnen we talloze kansberekeningen baseren op eenvoudige uitgangspunten. Merk op dat we steeds veronderstellen dat de gebeurtenissen *onderling onafhankelijk* zijn en dat zij *elkaar uitsluiten.* Deze schijnbaar onschuldige voorwaarde heeft verregaande gevolgen. Stel bijvoorbeeld dat ik wil weten wat de kans is op de gebeurtenis ‘een aas trekken of een klaverkaart’. Er zijn 13 klaveren en 4 azen, maar... één van die klaveren is klaveraas! Dus is de kans daarop niet $(13+4)/52$ maar $(13+3)/52$, een verschil van 3,7% waar een gladde gokker zijn voordeel mee kan doen. In praktische gevallen zijn zulke afhankelijkheden (de technische term is *voorwaardelijke waarschijnlijkheid*) aan de orde van de dag, en aan het eind van dit stuk komen we daarop terug. Je kunt er heel goed mee rekenen, maar het is een kwestie van voortdurend oppassen.

Als twee gebeurtenissen A en B onafhankelijk zijn en elkaar uitsluiten, dan is de kans op A *of* B gegeven door

$$P_{A|B} = \frac{N_A + N_B}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P_A + P_B \tag{4}$$

Dit noemen we de *optelregel voor onafhankelijke kansen.* De kans op A *èn* B ligt iets ingewikkelder, want het is een kwestie van mogelijke combinaties tellen. Vanwege dat verraderlijk korte ‘èn’ gaat het om *paren* van gebeurtenissen. Stel dat van een totaal N gebeurtenissen, N_A maal geval A voorkomt. En stel dat van de M gebeurtenissen, M_B maal B gebeurt. Als ik die combineer, heb ik $N \times M$ mogelijkheden voor paren, waarvan er $N_A \times M_B$ paren A *èn* B zijn. Dus is de kans

$$P_{A\&B} = \frac{N_A \times M_B}{N \times M} = \frac{N_A}{N} \times \frac{M_B}{M} = P_A \times P_B \tag{5}$$

Dit is de *productregel voor samengestelde kansen.* Deze regel is de basis voor de ‘vertakkingen’ die we dadelijk zullen tegenkomen.

Doet-ie het of doet-ie het niet?

In het hier beschreven voorbeeld gebruiken we een ‘ding’ dat in twee toestanden kan verkeren: het werkt of het werkt niet. Het ding bestaat uit onderdelen; een onderdeel is heel of het is stuk. Het ding werkt alleen als alle onderdelen heel zijn. We gaan berekenen wat de kans is dat een ding werkt, als we weten wat de kans is dat een onderdeel stuk is.

Stel dat het ding uit één onderdeel bestaat. Laat K de kans zijn dat het onderdeel stuk is. Dan is de kans dat het ding niet werkt ook K. De kans dat het ding wel werkt is $1-K$, want de som van de kansen moet 1 zijn.

Stel nu dat het ding uit *twee* onderdelen bestaat. Voorlopig nemen we aan dat de kans dat een onderdeel stuk is, altijd K is. Dat is niet erg realistisch, want bijvoorbeeld de accu van een auto of de banden gaan veel sneller stuk dan de cylinderwanden, maar het rekent makkelijker. Aan het eind van dit verhaal komen we hierop terug.

Vertakkingen en kansen

We rekenen* nu alsvolgt. In een fractie K van alle gevallen is onderdeel 1 stuk; de kans is $1-K$ dat het nog heel is. Hetzelfde geldt voor onderdeel 2. Hoe groot is nu de kans dat het ding werkt? We kunnen dit zien als een boom, die zich vertakt met takken ter dikte K en $1-K$, of een rivier, of zoiets. Elke tak vertakt zich weer, *in dezelfde verhouding*: als een tak een dikte D heeft, vertakt hij zich in een twijg met dikte $D \times K$ en een andere met dikte $D \times (1-K)$. Het aantal lagen van vertakking is gelijk aan het aantal onderdelen in het ding.

De tak met dikte K splitst zich in de verhouding $(K) : (1-K)$, dus de takken krijgen na twee splitsingen een dikte

* Het komt hier en daar voor dat een berekening of bewijs wat lang of lastig is. In dat geval wordt verwezen naar aanhangsels aan het eind van dit stuk, gemerkt met ♠.

$K \times K = K^2$ en $K \times (1-K) = K - K^2$. De tak $(1-K)$ splitst zich in $(1-K) \times K = K - K^2$ en $(1-K) \times (1-K) = (1-K)^2 = 1 - 2K + K^2$. Tellen we alles op, dan komt er

$$K^2 + K - K^2 + K - K^2 + 1 - 2K + K^2 = 1 \tag{6}$$

dus we hebben een totale kans 1, dat klopt: alle boomtakken samen zijn net zo dik als de stam, ofwel er stroomt net zoveel water door de zijrivieren als uiteindelijk in zee uitkomt.

Alleen de tak waarlangs geen enkel onderdeel stuk is, levert een ding op dat werkt. Dus de kans dat een ding met twee onderdelen werkt, is $1-K$ tot de macht 2,

$$P_0 = (1-K)^2 = 1 - 2K + K^2 \tag{7}$$

We zullen deze kans P_0 noemen, de kans-op-nul-fouten. De kans dat het ding niet werkt noemen we P_f, en omdat de som van alle mogelijke kansen gelijk aan 1 moet zijn, is

$$P_f = 1 - P_0 \tag{8}$$

Een voorbeeld. Als de kans dat één onderdeel stuk is, gelijk is aan 0,1 (tien procent), dan is

$$P_0 = (1-0,1)^2 = 1 - 0,2 + 0,01 = 0,81 \tag{9}$$
$$P_f = 1 - P_0 = 0,19 \tag{10}$$

Benadering: Het is vaak handig en snel om gebruik te maken van het feit dat K in practische gevallen meestal klein is (je wilt immers met betrouwbare onderdelen werken). Als dat het geval is, dan zien we direct dat 1 veel groter is dan K, zodat K veel groter is dan K^2, die weer veel groter is dan K^3, enzovoorts. We gebruiken het symbool "$\simeq$" als we bedoelen 'is bij benadering gelijk aan'. Dan volgt uit het bovenstaande, voor *kleine* waarde van K (zeg b.v. kleiner dan 0,1; hoe kleiner K, hoe nauwkeuriger de schatting)

$$P_0 \simeq 1 - 2K \tag{11}$$
$$P_f \simeq 2K \tag{12}$$

Algemene vertakkingen

Stel nu dat het ding uit N onderdelen bestaat. Dan kun je een boom of rivier tekenen die zich $N-1$ maal vertakt. Je ziet direct dat de kans dat het ding werkt wordt gevonden langs slechts één tak: namelijk die waarlangs elk onderdeel heel is. Dus

$$P_0 = (1-K)\times(1-K)\times(1-K)\times\cdots\times(1-K) = (1-K)^N \quad (13)$$

Je moet dus het getal $1-K$ steeds met zichzelf vermenigvuldigen totdat je N factoren bij elkaar hebt.

Laten we eens zien wat dat oplevert. Boven hadden we al $N = 2$, dat wil zeggen $P_0 = (1-K)^2$, en voor hogere N komt er

$$(1-K)^3 = 1 - 3K + 3K^2 - K^3 \quad (14)$$
$$(1-K)^4 = 1 - 4K + 6K^2 - 4K^3 + K^4 \quad (15)$$
$$(1-K)^5 = 1 - 5K + 10K^2 - 10K^3 + 5K^4 - K^5 \quad (16)$$

Kijk even naar de cijfers vóór de machten van K. Die vormen een speciale figuur, als we de gevallen $N = 0, 1, 2, 3\cdots$ van boven naar beneden opstapelen:

$$\begin{array}{lccccccccccccc}
N=0 & & & & & & & 1 & & & & & & \\
N=1 & & & & & & 1 & & 1 & & & & & \\
N=2 & & & & & 1 & & 2 & & 1 & & & & \\
N=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
N=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
N=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
N=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & \cdots & \cdots & \cdots & & \cdots & & \cdots & & \cdots & \cdots & \cdots & &
\end{array} \quad (17)$$

Dat heet de *driehoek van Pascal*. We zullen hem nog tegenkomen (zie ♠**1**). Je kunt elke volgende rij eenvoudig berekenen door een 1 aan weerszijden te zetten (op de stippeltjes uiterst links en rechts), en het getal in een tussenruimte te vinden door de twee getallen erboven op te tellen. Zie bijvoorbeeld de rij voor $N = 5$: die staat zo omdat $1 + 4 = 5$ en $4 + 6 = 10$.

Gebruiken we nu de benadering dat K klein is, dan zien we meteen aan de driehoek dat

$$P_0 \simeq 1 - NK \tag{18}$$

$$P_f \simeq NK \tag{19}$$

Voorbeeld: als $K = 0,01$ (een procent), en $N = 20$, dan is

$$P_0 \simeq 1 - 20 \times 0,01 = 0,8 \tag{20}$$

Dus zelfs als de onderdelen redelijk betrouwbaar zijn (99%) gaat de kans dat het ding werkt snel omlaag als het uit veel onderdelen bestaat (80% bij 20 onderdelen).

Een belangrijke conclusie uit (18,19) is, dat de kans dat een ding niet werkt kan worden verkleind door *de betrouwbaarheid van de onderdelen te vergroten* (K kleiner maken), of door *het aantal onderdelen te verkleinen* (N kleiner maken), en dat *beide methoden evenveel aan het resultaat bijdragen* omdat P_f gelijk is aan het *product* van K en N. In de praktijk is het soms makkelijker om het aantal onderdelen te verkleinen dan om de betrouwbaarheid per onderdeel te vergroten.

Omgekeerd kunnen we de kans dat een onderdeel stuk is, uitrekenen als we weten wat de kans is dat een ding niet werkt, mits we ook weten uit hoeveel onderdelen het bestaat. We bepalen P_f uit waarnemingen, bijvoorbeeld door te kijken hoeveel dingen uit een partij van een miljoen niet werken. Zo berekenen we

$$K \simeq \frac{P_f}{N} \tag{21}$$

Als bijvoorbeeld een ding in 10 procent van de gevallen niet blijkt te werken, is $P_f = 0,1$. Weten we ook N, zeg $N = 50$, dan is

$$K \simeq \frac{0,1}{50} = 0,002 \tag{22}$$

dus in dit voorbeeld is de kans dat een onderdeel stuk is, 2 promille.

Een willekeurig ding stuk

Tot dusver zijn we ervan uitgegaan 'dat een ketting zo sterk is als de zwakste schakel'. Maar in de praktijk zijn kettingproblemen eigenlijk zeldzaam. Als je in een muur een enkele baksteen stukslaat, blijft de muur vrijwel altijd staan. Laten we dus eens kijken naar 'aftakeling' inplaats van 'niet werken'. We bezien een ding dat altijd min of meer werkt, maar des te slechter naarmate er meer onderdelen stuk zijn.

In het bovenstaande berekenden we de kans dat er *geen enkel* onderdeel stuk is. In (8, 13) zagen we dat die kans gegeven wordt door

$$P_0 = (1 - K)^N \simeq 1 - NK \tag{23}$$

Op soortgelijke manier zien we, dat de kans dat *alle* onderdelen stuk zijn gelijk is aan

$$P_N = K^N \tag{24}$$

Wat is nu de kans P_1 dat precies één onderdeel stuk is? laten we eerst het geval $N = 2$ bekijken. Uit de boomstructuur zien we direct dat er twee mogelijkheden zijn, namelijk

$$\begin{array}{ll} (1-K)K & \text{onderdeel 1 werkt, 2 is stuk} \\ K(1-K) & \text{onderdeel 2 werkt, 1 is stuk} \end{array} \tag{25}$$

Samen hebben we dus

$$P_1 = 2K(1 - K) = 2K - 2K^2 \tag{26}$$

want elk van deze twee boomtakken levert dezelfde bijdrage.

In het algemene geval, dus met N onderdelen, is de kans langs één enkele boomtak

$$P_1 = K(1 - K)^{N-1} \tag{27}$$

dat wil zeggen, een factor K voor de kans dat een onderdeel stuk is, en een factor $1 - K$ voor elk van de $N - 1$ onderdelen die nog heel zijn. Omdat elk van de N onderdelen stuk kan gaan, zijn er N takken die het resultaat (27) geven, en dus

$$P_1 = NK(1-K)^{N-1} \tag{28}$$

In benadering voor kleine K komt er

$$\begin{aligned} P_1 &= NK(1-K)^{N-1} \simeq NK\,(1-(N-1)K) \\ &= NK - N(N-1)K^2 \end{aligned} \tag{29}$$

Voor kleine K is dit dus vrijwel gelijk aan de eerder gevonden P_f. Dat is niet zo vreemd: als er maar een heel kleine kans is dat een enkel onderdeel stuk is, is de kans dat er twee stuk zijn dienovereenkomstig kleiner. Het exacte verschil is

$$\begin{aligned} P_f - P_1 &= 1-(1-K)^N - NK(1-K)^{N-1} \\ &= 1-(1-K)^{N-1}(1-K+NK) \end{aligned} \tag{30}$$

Hoeveel wegen leiden naar Rome?

Nog steeds uitgaande van de veronderstelling dat alle onderdelen hetzelfde zijn, of althans dezelfde kans hebben om stuk te zijn, vragen we ons af: wat is de kans dat niet slechts één, maar twee of meer onderdelen stuk zijn? In het algemeen komt dat neer op het beantwoorden van de vraag: op hoeveel manieren kunnen M van de N onderdelen stuk zijn? Als bijvoorbeeld $N=2$, heb je de volgende mogelijkheden. Ofwel alles is stuk; dat kan maar op een enkele manier. Ofwel alle onderdelen zijn nog heel; ook dat kan maar op één manier. Ofwel er is precies 1 onderdeel stuk; dat kan het eerste onderdeel zijn of het tweede, dus 2 manieren. Kortom:

$$\begin{array}{cccccc} N=0 & & & 1 & & \\ N=1 & & 1 & & 1 & \\ N=2 & 1 & & 2 & & 1 \\ & \text{beide} & & \text{een} & & \text{geen} \\ M= & 2 & & 1 & & 0 \end{array} \tag{31}$$

Doen we hetzelfde spelletje met $N = 3$, dan komt er

$$\begin{array}{lcccccccc} N=0 & & & & 1 & & & \\ N=1 & & & 1 & & 1 & & \\ N=2 & & 1 & & 2 & & 1 & \\ N=3 & 1 & & 3 & & 3 & & 1 \\ & \text{alle} & & & & & & \text{geen} \\ M= & 3 & & 2 & & 1 & & 0 \end{array} \tag{32}$$

We zien nu direct in, dat *het aantal manieren* Δ_M^N *waarop* M *van de* N *onderdelen stuk kunnen zijn, wordt gegeven door de driehoek van Pascal.* Voorwaarde is steeds dat we ons er niet om bekommeren welke onderdelen dat precies zijn; elke combinatie van M niet-werkende onderdelen wordt geacht hetzelfde resultaat te hebben.

Hoe komt het dat we Δ hier weer terugvinden? De reden is de bijzondere eigenschap van de driehoek van Pascal, namelijk dat elk getal wordt gevormd door de som van de twee getallen erboven. Zien we de driehoek als een boom of rivier, met de stam (of de monding) bij de eerste 1, dan herkennen we weer het verhaal uit bovengenoemde voorbeelden. Omlaag werkend in de driehoek hebben we op elke 'vork van de boomtakken' de keuze: doet-ie het, of doet-ie het niet? Zo ja, dan nemen we de vertakking naar (bijvoorbeeld) links, zo nee, dan nemen we de weg naar rechts. De kans op 'onderdeel is stuk', overeenkomend met 'vork naar rechts', is K; de kans op 'onderdeel is niet stuk', ofwel 'vork naar links', is $1 - K$. Zodoende is de kans dat M willekeurige onderdelen stuk zijn

$$P_M^N = \Delta_M^N \, K^M (1 - K)^{N-M} \tag{33}$$

Dit is de *kansverdeling van Bernoulli.* Hierin is Δ_M^N het aantal manieren waarop M van de N onderdelen stuk kunnen zijn, dus de Pascal-factor. Voor $M = 0$ krijgen we formule (8) weer terug voor de kans dat geen enkel onderdeel stuk is. Als $M = N$ vinden we de uitdrukking (24) voor de kans dat *alle* onderdelen stuk zijn. Dat hoort ook zo, want de uiterste getallen van de driehoek zijn altijd gelijk aan 1: $\Delta_0^N = \Delta_N^N = 1$.

De kansverdeling van Bernoulli

Er zit echter nog een adder onder het gras. We hadden immers in het begin afgesproken dat *de som van alle kansen is 1.* Voor de driehoek van Pascal is de som van de getallen op een rij absoluut niet gelijk aan 1; integendeel, een beetje werk toont aan dat die som gegeven wordt door $1, 2, 4, 8, 16, \cdots$, dat wil zeggen

$$\text{som van rij } N \text{ is } 2^N \tag{34}$$

Bovendien zien we, dat de grootste waarde die op een rij bereikt wordt, de pan uitrijst wanneer N steeds groter wordt. We kunnen dat proberen te beteugelen door te delen door die factor 2^N:

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1{,}0 & & & & & & & \\
 & & & & & & 0{,}5 & & 0{,}5 & & & & & & \\
 & & & & & 0{,}25 & & 0{,}5 & & 0{,}25 & & & & & \\
 & & & & 0{,}13 & & 0{,}38 & & 0{,}38 & & 0{,}13 & & & & \\
 & & & 0{,}06 & & 0{,}25 & & 0{,}38 & & 0{,}25 & & 0{,}06 & & & \\
 & & 0{,}03 & & 0{,}16 & & 0{,}31 & & 0{,}31 & & 0{,}16 & & 0{,}03 & & \\
 & 0{,}02 & & 0{,}09 & & 0{,}23 & & 0{,}31 & & 0{,}23 & & 0{,}09 & & 0{,}02 & \\
0{,}01 & & 0{,}05 & & 0{,}16 & & 0{,}27 & & 0{,}27 & & 0{,}16 & & 0{,}05 & & 0{,}01
\end{array} \tag{35}$$

Deze driehoek voldoet natuurlijk *niet* meer aan de regel dat een getal gevonden wordt door de twee bovenliggende bij elkaar op te tellen!

Maar in het geval van de boven gevonden kansverdeling van Bernoulli,

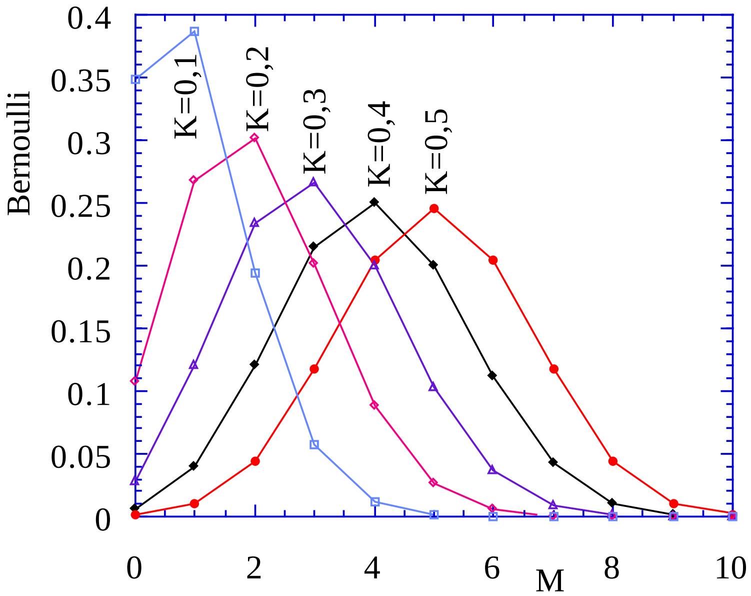
$$P_M^N = \Delta_M^N \, K^M (1 - K)^{N-M} \tag{36}$$

hebben we geluk. Het blijkt namelijk dat hiervoor *wel* geldt dat de som van de kansen 1 is (zie ♠**1**). Een grafiek van deze verdeling voor $N = 10$, en voor K lopend van 0,5 tot 0,1 is gegeven in onderstaande figuur.

Laten we bijvoorbeeld nemen $K = 0,5$ ('kruis of munt'), dan vinden we bij $N = 7$ de rij voor $M = 0$ tot en met $M = 7$,

$$P_M^N = \Delta_M^N \left(\frac{1}{2}\right)^N = 2^{-N} \Delta_M^N \tag{37}$$

0,0078	0,055	0,164	0,273	0,273	0,164	0,055	0,0078	(38)
M=7	M=6	M=5	M=4	M=3	M=2	M=1	M=0	



Alles gaat mis, zeg je wel eens, maar je weet dat dat eigenlijk overdreven is, en (38) laat zien waarom. Zelfs als er slechts 7 dingen mis kunnen gaan, en we nemen aan dat de kans dat iets mis gaat 0,5 bedraagt, is de kans kleiner dan 8 promille dat werkelijk *alles* in de soep loopt. Daar staat tegenover dat de kans net zo klein dat er *niets* aan de hand is, dus ‘Met mij gaat alles goed’ is waarschijnlijk evenmin waar.

Wedden dat...

Wanneer je een lot koopt waarop fl. 25 wordt uitgekeerd als het wordt getrokken, en de kans op zo’n trekking is 10%, verwacht je gemiddeld fl. 2,50 te krijgen. Betaal je meer dan een knaak voor zo’n lot, dan word je beetgenomen. Dat is uiteraard bij *alle* loterijen het geval. Twee gulden zou een voordelige prijs zijn voor dat lot, want per trekking houd je er *gemiddeld* twee kwartjes aan over. Het is verlakkerij om te zeggen ‘ik heb zo-en-zoveel gulden gewonnen’, tenzij je daarvan eerst je totale inleg aftrekt. Niemand doet dat, om niet als ezel te boek te staan, want gemiddeld moet je dan bij alle bekende loterijen steeds van verlies spreken.

Hieruit zien we dat de 'waarde' van een kans gegeven wordt door *het product van de uitkomst en de kans daarop*:

$$W = M \times P_M \tag{39}$$

Wat is nu de te verwachten waarde als we alle mogelijke trekkingen in aanmerking nemen? Deze grootheid V, de *verwachtingswaarde* van de kansverdeling P, vinden we door (39) te sommeren over alle toegelaten waarden van M:

$$V = \sum_M M\, P_M \tag{40}$$

(zie ♠**2** voor de betekenis van het somteken Σ).

Laten we de verwachtingswaarde eens uitrekenen in een zeer eenvoudig geval: het werpen met een dobbelsteen. In dat geval is de kans op een uitkomst van 1 tot en met 6 altijd gelijk, namelijk 1/6 tenzij er met de dobbelsteen geknoeid is. Dus vinden we

$$V = \sum_{i=1}^{6} i \times P_i = \sum_{i=1}^{6} \frac{i}{6} = \frac{1+2+3+4+5+6}{6} = 3,5 \tag{41}$$

Stel dat je een spel speelt waarbij je evenveel guldens krijgt uitgekeerd als er ogen bij een worp bovenkomen, dan moet je per spel hoogstens f. 3,50 inzetten, anders ga je de boot in.

In het geval van de Bernoulli-verdeling (36) kunnen we de verwachtingswaarde uitrekenen volgens bovenstaand recept:

$$\begin{aligned} P_M^N &= \Delta_M^N\, K^M (1-K)^{N-M} \\ &= \frac{N(N-1)\cdots(N-M+1)}{1\times 2\times 3\times\cdots\times M} K^M (1-K)^{N-M} \end{aligned} \tag{42}$$

$$V = \sum_{M=0}^{N} M\, P_M^N = NK \tag{43}$$

(zie voor het bewijs ♠**3**). We vinden dus dat het verwachte aantal onderdelen dat stuk is, wordt gegeven door het product van het aantal N der onderdelen en de kans K dat een onderdeel stuk is. De bijzondere rol van het product NK kwamen we al veel eerder tegen (bijvoorbeeld in (18) en (19)). We zien nu dat dat geen toeval is, maar berust op een exacte eigenschap van de Bernoulli-kansverdeling. Ook hier concluderen we dus weer: je kunt de verwachtingswaarde verlagen (en dus de prestaties van het 'ding' verhogen) door het aantal onderdelen N waaruit het ding bestaat kleiner te maken, of de kans K dat een onderdeel stuk is te verkleinen, of beide. Wegens de productregel (43) werkt elk van deze alternatieven even goed. In de praktijk hangt het er natuurlijk van af wat het makkelijkst te bereiken is; meestal is het verkleinen van N de aangewezen weg.

Als je een ding koopt dat uit N onderdelen bestaat, en er is een kans K dat een willekeurig onderdeel stuk is, verwacht je dat gemiddeld NK onderdelen stuk zullen zijn. Voorbeeld: een huis bestaat uit minstens een half miljoen onderdelen: spijkers, buizen, stukken betonijzer, planken, loodslabben enzo. Laat de kans dat een willekeurig onderdeel stuk is, zeer klein zijn, zeg 1 op een miljoen. Dan verwachten we dat een opgeleverd huis $500.000/1.000.000 = 0,5$ kapotte onderdelen bevat. De *vereniging eigen huis* beweert dat elk nieuw opgeleverd huis 38 gebreken vertoont. Zo te zien doet de bouw-industrie het dus geweldig goed, namelijk 76 maal beter dan verwacht! Natuurlijk is deze berekening onzin: een kromme spijker ergens in een vloer heeft niet dezelfde gevolgen als een gescheurde loodslab. Ook is de kans op een kromme spijker niet dezelfde als de kans op een scheur in het lood. Op dit soort problemen komen we aan het eind van het verhaal terug.

Strooiing

Het getal V is een *verwachtings*waarde: er is geen garantie dat elke trekking uit een kansverdeling steeds de waarde V heeft.

Sterker nog, dat kan soms helemaal niet: bij het gooien van een dobbelsteen is de verwachtingswaarde 3,5 – een getal dat je bij het dobbelen nooit boven zult zien komen.

Er is dus bij elke trekking een zeker verschil tussen de getrokken waarde en de verwachtingswaarde. De grootte van dat verschil heet *strooiing*, soms ook wel *spreiding* genoemd. We zullen deze strooiing aan het eind van dit verhaal tegenkomen in de vorm van het getal 'sigma', σ. Om de spreiding S te berekenen vinden we eerst de verwachtingswaarde van de variabele M, en vergelijken die met de verwachtingswaarde van *het kwadraat van* M. Dat doen we omdat een afwijking van V zowel positief kan zijn (als we dobbelen en gooien 6, is de afwijking $6 - 3,5 = 2,5$) als negatief (als we 1, 2 of 3 gooien). Door te kwadrateren tellen beide afwijkingen even sterk mee. Dus we nemen

$$S^2 = \sum_i i^2 P_i - V^2 = \sum_i i^2 P_i - \left(\sum_i i\, P_i \right)^2 \tag{44}$$

In het geval van de dobbelsteen vinden we

$$\begin{aligned} S^2 &= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) - 3,5^2 = 2,9166\ldots \\ S &= \sqrt{2,91667} = 1,708 \end{aligned}$$

Je moet dus niet raar opkijken als je bij het dobbelen een paar punten van de waarde 3,5 vandaan zit. Merk overigens op dat de strooiing S *niet* gelijk is aan het verschil tussen de verwachtingswaarde 3,5 en de uiterste waarde (6 of 1) die je kunt gooien! Het verschil is *kleiner*, een feit waar een beroepsgokker zijn voordeel mee kan doen.

De strooiing voor de Bernoulli-verdeling berekenen we uit

$$\sum_{M=0}^{N} M^2\, P_M^N = NK(1 - K) + N^2K^2 \tag{45}$$

(voor het bewijs, zie ♠**4**). Gebruiken we hierbij de verwachtingswaarde (43), dan vinden we

$$S^2 = NK(1-K) + N^2K^2 - V^2 = NK(1-K) \tag{46}$$
$$S = \sqrt{NK(1-K)} \tag{47}$$

Hieruit zien we iets zeer belangrijks: *de strooiing neemt slechts toe met de wortel uit het aantal onderdelen.* Als we naïef te werk gaan, zouden we dat niet hebben verwacht: de verwachtingswaarde V is recht evenredig met N, dus verdubbelen van het aantal onderdelen verdubbelt de verwachtingswaarde van het aantal falende delen. Doch de strooiing is *niet* recht evenredig met N, maar slechts met $\sqrt{N}$. Als we dus viermaal zoveel onderdelen gebruiken, neemt de strooiing slechts toe met een factor 2. Deze regel geldt veel algemener dan hier beschreven, en heet de *wet van de grote getallen.*

De Wet van De Moivre en het getal σ (sigma)

We zagen al eerder dat de kans dat M van de N onderdelen stuk zijn, wanneer alle onderdelen dezelfde kans K hebben om te falen, gegeven wordt door de Bernoulli-verdeling

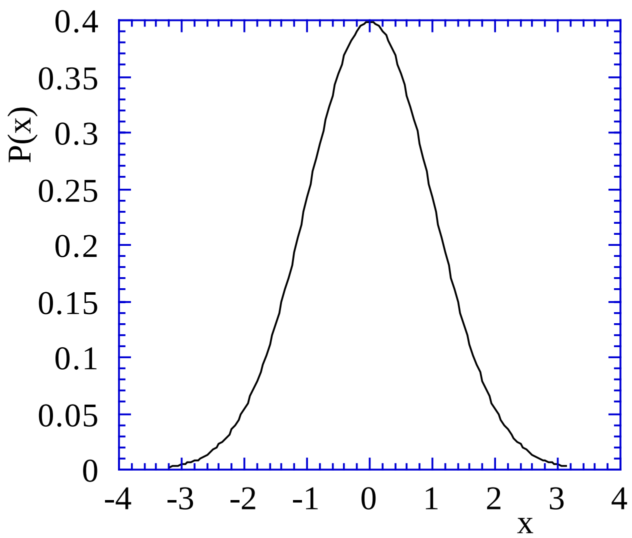
$$P_M^N = \frac{N(N-1)\cdots(N-M+1)}{1 \times 2 \times 3 \times \cdots \times M}\, K^M (1-K)^{N-M} \tag{48}$$

Wanneer we P_M^N tekenen, door bij een gegeven waarde van N te nemen $M = 0, 1, 2, 3, \cdots, N$, dan zien we dat de zo gevormde grafiek steeds minder van vorm verandert naarmate N groter wordt. Door de wiskundige De Moivre is bewezen dat de vorm die zo uiteindelijk ontstaat, in het geval $K = 0,5$, gegeven wordt door de *exponentiële functie* (zie ♠**5**)

$$P(x) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-x^2/2\sigma^2} \tag{49}$$

Hierin is nu x een continue variabele, die de rol van de discrete waarden van M heeft overgenomen.

Later werd dit opnieuw gedaan door Gauss, en naar hem wordt dit vaak de 'Gauss-functie' genoemd ('eer wie ere toekomt' gaat in de wetenschap niet altijd op). De De Moivre-functie heet in de Angelsaksische literatuur *Gauss curve*, *normal distribution*, of ook wel *bell curve* (vanwege de klokvorm):



Door de constructie die we boven hebben gevolgd, kan deze kromme alsvolgt gelezen worden. Aan het ene uiterste ($x = -\infty$) staat de kans dat van een oneindig aantal mogelijkheden, *alles* fout gaat. Die kans is nul, want de exponent van een negatief oneindig getal is nul. Aan het andere uiterste ($x = +\infty$) staat de kans dat, van een oneindig aantal mogelijkheden, alles *goed* gaat. Die kans is eveneens nul. In het midden (bij $x = 0$) is de kansverdeling maximaal. Door vergelijken met de redeneringen die we tot dusver hebben gevolgd, zien we meteen dat hierbij de ene helft van alle dingen goed gaat, de andere helft fout. Het maximum bij $x = 0$ in de De Moivre-functie is de wiskundige uitdrukking van het gevoelen *'t kan vriezen, 't kan dooien.*

Uit de vorm van de functie (36) zien we direct dat de kans op een bepaalde afwijking van het gemiddelde (bij $x = 0$) bepaald wordt door de verhouding x/σ. Het getal σ heet de *strooiing* of ook wel *standaarddeviatie*. In de wandeling zegt men wel 'een kans van twee sigma' als men de waarde van $P(x)$ bedoelt op het punt dat $x = 2\sigma$.

Vanwege het kwadraat in de exponent, neemt de functie $P(x)$ razendsnel af naarmate x/σ verder van nul af ligt. De kans dat een bepaald resultaat meer dan 2σ van het gemiddelde afwijkt is slechts 2%. Het 'aantal sigma' wordt wel kortweg gebruikt om de kwaliteit van een bepaald product of resultaat aan te duiden.

Wanneer iemand dus beweert dat de nagestreefde kwaliteitsnorm *six sigma* is, grenst dat aan opschepperij, want het betekent dat de kans op een fout van de orde van een op 10 miljoen is. Alleen in het geval van uiterst uitgebreide foutcorrectieprocedures (zoals gebruikt in computer- en communicatie-software) kan dat ten naaste bij bereikt worden. Een voorbeeld van wat zoiets betekent: over de *Space Shuttle* werd oorspronkelijk beweerd dat de kans op een fatale fout tijdens de lancering ongeveer *five sigma* was, iets van de orde van 1 fout op 100.000 lanceringen. Dat lijkt een begrijpelijk getal, maar – zoals Feynman opmerkte – als je het nagaat, betekent dit dat je slechts één ongeluk verwacht als je *een ruimteveer per dag gedurende driehonderd jaar achtereen* zou lanceren. Zo uitgedrukt zie je meteen dat die foutkans grootspraak is, zoals op tragische wijze bleek bij de ontploffing van het ruimteveer *Challenger*. Dat was ruwweg de vijftigste lancering, overeenkomend met een kans van $2\,\sigma$.

Slotopmerking

Tenslotte moet nog worden opgemerkt dat het hele sigma-verhaal uitgaat van de sterk geïdealiseerde situatie zoals boven geschetst. In werkelijkheid zijn natuurlijk lang niet alle foutkansen gelijk aan dat ene getal K, en – erger nog – de kans op een bepaalde fout is meestal niet onafhankelijk van de andere mogelijke foutbronnen. Dan krijgen we te maken met *voorwaardelijke waarschijnlijkheid*, en gaat het sigma-sprookje niet door.

—◇—

Aanhangsel met meer wiskundige uitleg

♠1. Pascal-getallen

De driehoek van Pascal wordt gevormd door aan een rij getallen steeds een nieuwe toe te voegen volgens een bepaald recept. We beginnen met de rij genummerd $N = 0$; die rij bestaat uit het getal 1. Elke volgende rij heeft één getal meer dan de vorige, en elke rij begint en eindigt met 1. Daarmee ligt de rij voor $N = 1$ vast. De rij voor $N = 2$ is de eerste die een 'open plek' in het midden heeft. Elke open plek van een rij wordt gevuld met het getal dat gelijk is aan de som van de twee getallen in de rij erboven die ter linker- en ter rechterzijde staan. Zodoende komt er

$$\begin{array}{lccccccccccccc}
N=0 & & & & & & & 1 & & & & & & \\
N=1 & & & & & & 1 & & 1 & & & & & \\
N=2 & & & & & 1 & & 2 & & 1 & & & & \\
N=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
N=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
N=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
N=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & & & & &
\end{array}$$

We willen nu uiteraard weten wat de algemene formule is voor de N-de regel van de driehoek, want het is teveel gedoe om het steeds helemaal uit te rekenen. Als we het elke keer weer van voren af aan willen doen, moeten we steeds met het Pascal-recept een hele pyramide bouwen van de vorm

$$\begin{array}{cccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 1+6 & & 6+15 & & 15+20 & & 20+15 & & 15+6 & & 6+1 & & 1 \\
1 & & =7 & & =21 & & =35 & & =35 & & =21 & & =7 & & 1
\end{array}$$

Op de N-de rij staan de Pascal-getallen Δ_M^N voor $M = 0$ tot en met $M = N$. Na enig prutsen vermoeden we dat

$$\Delta_M^N = \frac{N(N-1)(N-2)\cdots(N-M+1)}{1 \times 2 \times 3 \times \cdots \times M}$$

Proberen we het eens voor $N = 4$ en $M = 3$, dan zien we dat het klopt:

$$\Delta_3^4 = \frac{4 \times 3 \times 2}{1 \times 2 \times 3} = 4$$

Het algemene bewijs berust op het feit dat we uit regel N altijd die voor $N+1$ kunnen uitrekenen. Tellen we, volgens het Pascal-voorschrift, het $(M-1)$-de getal van regel N op bij het M-de getal van diezelfde regel, dan vinden we

$$\begin{aligned}\mathcal{P}_M^{N+1} =& \frac{N(N-1)(N-2)\cdots(N-(M-1)+1)}{1 \times 2 \times 3 \times \cdots \times (M-1)} \\ &+ \frac{N(N-1)(N-2)\cdots(N-M+1)}{1 \times 2 \times 3 \times \cdots \times M}\end{aligned}$$

Nu brengen we dit geheel op één noemer door in de eerste term de teller en de noemer met M te vermenigvuldigen. Brengen we de gemeenschappelijke factor buiten haakjes, dan kunnen we schrijven

$$\mathcal{P}_M^{N+1} = \frac{N(N-1)\cdots(N+1-M+1)}{1 \times 2 \times 3 \times \cdots \times M} \times (M + (N - M + 1))$$

We zien direct dat dit precies hetzelfde is als de formule voor Δ_M^N, maar nu met $N+1$ inplaats van N, want in de laatste factor tussen haakjes valt M weg tegen $-M$ zodat we $N+1$ overhouden, en deze factor zetten we links:

$$\mathcal{P}_M^{N+1} = \frac{(N+1)N(N-1)\cdots((N+1)-M+1)}{1 \times 2 \times 3 \times \cdots \times M}$$

Hiermee is bewezen dat het M-de Pascal-getal op de N-de regel voor alle M en N gegeven is door

$$\Delta_M^N = \frac{N(N-1)\cdots(N-M+1)}{1 \times 2 \times \cdots \times M}$$

De Pascal-getallen Δ zijn van groot belang omdat zij in doorlopende vermenigvuldigingen ('machten') voorkomen. De algemene vorm van (8) is, voor een product van n factoren,

$$(a+b) \times (a+b) \times (a+b) \cdots \times (a+b) = (a+b)^n$$

Uitschrijven levert het volgende resultaat:

$$(a+b)^n = \Delta_0^n a^0 b^n + \Delta_1^n a^1 b^{n-1} + \Delta_2^n a^2 b^{n-2} + \cdots + \Delta_n^n a^n b^0$$

waarin Δ_m^n het m-de Pascal-getal op rij n voorstelt. Zo'n oneindige som wordt afgekort met de notatie (zie ♠**2**)

$$(a+b)^n = \sum_{m=0}^{n} \Delta_m^n a^m b^{n-m}$$

(merk op dat $a^0 = 1$ en $a^1 = a$). Door deze uitdrukking te vergelijken met de formule voor de kansverdeling van Bernoulli, zien we dat deze sommeert tot 1:

$$\sum_{M=0}^{N} P_M^N = \sum_{M=0}^{N} \Delta_M^N K^M (1-K)^{N-M} = (K + (1-K))^N = 1 \tag{50}$$

♠2. Sommering

De hoofdletter sigma, Σ, wordt gebruikt als afkorting om de som van een rij getallen aan te geven. Zo wordt de som van n getallen

$$a_1 + a_2 + a_3 + \cdots + a_n$$

afgekort weergegeven door

$$\sum_{i=1}^{n} a_i$$

hetgeen te lezen is als *de som van alle getallen a_i waarin i de waarden van 1 tot en met n aanneemt.* Zo schrijven we bijvoorbeeld

$$\sum_{i=0}^{n} a_i = a_0 + a_1 + a_2 + a_3 + \cdots + a_n$$

en, voor een oneindige som,

$$\sum_{i=0}^{\infty} a_i = a_0 + a_1 + a_2 + a_3 + \cdots$$

♠3. Verwachtingswaarde van de Bernoulli-kansverdeling

De verwachtingswaarde V van een discrete kansverdeling P is gedefinieerd als

$$V = \sum_i i\, P_i$$

In het geval van de Bernoulli-verdeling moeten we dus berekenen

$$\begin{aligned} V &= \sum_{M=0}^{N} \Delta_M^N\, M\, K^M (1-K)^{N-M} \\ &= \sum_{M=1}^{N} K^M (1-K)^{N-M} \frac{N(N-1)\cdots(N-M+1)}{1 \times 2 \times 3 \cdots \times (M-1)} \end{aligned}$$

We halen nu een factor NK links van het somteken en schrijven

$$\begin{aligned} V &= NK \sum_{M=1}^{N-1} K^{M-1} (1-K)^{N-M} \times \\ &\qquad \frac{(N-1)(N-2)\cdots(N-M+1)}{1 \times 2 \times 3 \cdots \times (M-1)} \\ &= NK \sum_{M=1}^{N-1} K^{M-1} (1-K)^{(N-1)-(M-1)} \times \\ &\qquad \frac{(N-1)(N-2)\cdots((N-1)-(M-1)+1)}{1 \times 2 \times 3 \cdots \times (M-1)} \end{aligned}$$

Nu zien we iets aardigs: de som over M is *precies* van de vorm gegeven in ♠**1**, met dien verstande dat $m = M-1$ en $n = N-1$. We concluderen dus dat

$$V = NK\,(K + (1-K))^{N-1} = NK$$

hetgeen we wilden bewijzen. We gebruikten deze truc ook al in ♠**1** . De Bernoulli-verdeling is dus, dankzij de bijzondere eigenschappen van de Pascal-getallen, een zeer nette functie.

♠4. Strooiing van de Bernoulli-kansverdeling

Ook als we de strooiing van de Bernoulli-vorm uitrekenen kunnen we gebruik maken van de kunstgreep uit ♠**3**. We moeten berekenen

$$\sum_{M=0}^{N} K^M (1-K)^{N-M}\, M^2 \frac{N(N-1)\cdots(N-M+1)}{1\times 2\times 3\cdots\times M}$$

Meteen een factor M van het kwadraat in de teller wegstrepen tegen de M in de noemer, heeft weinig zin omdat we dan nog met een M in de teller blijven zitten. We gebruiken de truc

$$M^2 = M(M-1+1) = M + M(M-1)$$

De eerste term M behandelen we zoals in ♠**3**; dat levert een waarde NK op, zoals we al zagen. In de tweede term $M(M-1)$ kunnen we zowel een factor M als een factor $M-1$ wegstrepen tegen $M\times(M-1)$ in de noemer. Dan houden we precies dezelfde vorm over als in ♠**3**, behalve dan dat we inplaats van NK een factor $NK\times(N-1)K$ links van het somteken halen. Zo vinden we

$$\sum_{M=0}^{N} K^M (1-K)^{N-M}\, M^2 \ldots =$$

$$NK + NK(N-1)K = NK(1-K) + N^2K^2$$

We wisten al dat de verwachtingswaarde NK is, en dus

$$S^2 = NK(1-K) + N^2K^2 - (NK)^2 = NK(1-K)$$
$$S = \sqrt{NK(1-K)}$$

De strooiing neemt dus slechts met de *wortel* uit N toe. Hoe groter het aantal onderdelen N, hoe zekerder je bent dat het aantal te verwachten fouten NK is.

♠5. Exponentiële functie

De *exponentiële functie* is gedefinieerd door de oneindige som

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{1 \times 2 \times 3 \times \cdots \times n}$$

Nemen we hierin $x = 1$, dan blijkt dat het *exponentiële grondgetal* e de waarde heeft van

$$\mathrm{e}^1 = \mathrm{e} = \sum_{n=0}^{\infty} \frac{1}{1 \times 2 \times 3 \times \cdots \times n} = 2,71828182846\ldots$$

De formule van De Moivre geeft zo'n exponentiële functie, maar dan niet met de exponent x maar met $-x^2/2\sigma^2$. Deze vorm wordt bereikt door van de variabele M over te gaan op

$$x = (M - V)\frac{\sigma}{\sqrt{NK(1-K)}}$$

dat wil zeggen een variabele die de verwachtingswaarde nul heeft (omdat we NK van M hebben afgetrokken), en een strooiing σ (omdat we door $\sigma\sqrt{NK(1-K)}$ hebben gedeeld).