Voortgezette Kansrekening

voor het 2e jaar wiskunde

docent: Hans Maassen

November 2005

Mathematisch Instituut
Radboud Universiteit Nijmegen

Inhoudsopgave

1 Het vaasmodel **4**

2 Het algemene model **14**

3 Onafhankelijkheid **26**

4 Verwachtingswaarde **36**

5 Spreiding **49**

6 Continue verdelingen **58**

7 De normale verdeling **72**

8 Rekenen met voorwaardelijke kansen **83**

9 Kansverdelingen in $\mathbb{R}^2$ **92**

10 De Poisson-verdeling **101**

11 Simulatie **113**

12 Kansgenererende functies **130**

13 Markovketens **137**

A Tabel voor de normale verdeling **169**

B Tentamenopgaven **170**

Hoofdstuk 1

Het vaasmodel

1.1 Dobbelstenen zijn een belangrijke inspiratiebron geweest voor de kansrekening. In de 17^e eeuw verschenen de eerste wiskundige verhandelingen over kansspelen, o.a. van Pascal en Huijgens. Kanstheorie gebaseerd op axioma's werd ingevoerd door Kolmogorov in 1933.
We geven enige historische probleempjes, waarvan het oplossen voor je nu wellicht nog wat lastig is; het benodigde gereedschap zal in het komende hoofdstuk aangedragen worden.

Een vraag die aan Galileï (17^e eeuw) gesteld werd: Als je 3 dobbelstenen gooit, dan kan het totaal 10 op zes manieren verkregen worden, nl. 631, 622, 541, 532, 442, 433; het totaal 9 kan eveneens op zes manieren verkregen worden (controleer!). Toch blijkt in de praktijk het totaal 10 vaker voor te komen dan 9. Verklaring?

Pascal en De Méré (17^e eeuw): Is het even "moeilijk" om minstens één zes te gooien in 4 worpen met één dobbelsteen als om minstens één dubbelzes te gooien in 24 worpen met twee dobbelstenen? (Redenering: het is zes maal zo "moeilijk" om dubbelzes met twee dobbelstenen te gooien, als om zes te gooien met één dobbelsteen; vandaar dat er zesmaal zo veel worpen gedaan mogen worden).

1.2 Bekijk een vaas met n ballen (n is een of ander natuurlijk getal), die gemerkt zijn, bv. door nummering of kleuring. We willen één bal uit de vaas nemen, maar de vaas heeft een nauwe hals, zodat we niet kunnen zien welke bal we pakken. Als we aannemen dat de ballen goed geschud zijn, dan ligt het voor de hand aan alle ballen dezelfde kans toe te kennen om gepakt te worden, nl. $\frac{1}{n}$.
Als we de ballen $\omega_1, \ldots, \omega_n$ noemen, dan kunnen we (vooruitlopend op de formele definitie) spreken van een *eindige uitkomstenruimte*

$$\Omega = \{\omega_1, \ldots, \omega_n\}$$

met *gelijkwaardige uitkomsten* (namelijk voor elke $\omega \in \Omega$ is de kans op die ω: $\frac{1}{n}$). Voorlopig noemen we dit model gemakshalve: het *vaasmodel.*

1.3 Kansen op bepaalde resultaten zijn in het vaasmodel i.h.a. vrij eenvoudig te berekenen. Stel bijvoorbeeld dat de vaas 15 ballen bevat, waarvan er 5 rood zijn en 10 wit. Neem één bal uit de vaas. De kans dat deze rood is, is $\frac{5}{15}$, immers: het totaal aantal uitkomsten is 15, het aantal voor "rood" gunstige uitkomsten is 5.

1.4 Er zijn natuurlijk talloze voorbeelden die je in het vaasmodel kunt vertalen. Ga zelf na dat hieronder ook vallen:

1. Het werpen van een dobbelsteen. Hoe groot is de kans op een aantal ogen van minstens 5?

2. Het werpen van twee dobbelstenen. Bereken de kans op hetzelfde aantal ogen bij beide stenen.

3. Het werpen van m munten, waarbij je kijkt welke zijden boven komen. Hoe groot is n, het aantal mogelijke uitkomsten?

4. Het trekken van een kaart uit een kaartspel. Wat is de kans op $\heartsuit$?

5. Het trekken van achtereenvolgens twee kaarten uit een kaartspel (de eerste kaart wordt niet teruggestoken). Wat is de kans op twee $\heartsuit$?

6. Idem, als de eerste kaart wèl wordt teruggelegd.

7. Het nemen van vijf ballen uit een vaas die tien rode en twintig witte ballen bevat. De kans dat er hiervan precies drie rood zijn, zullen we in paragraaf 1.8 berekenen.

Om aan te geven dat inderdaad alle uitkomsten voor het toeval gelijkwaardig zijn, gebruikt men vaak uitdrukkingen als: de dobbelsteen is eerlijk (of zuiver), de kaarten zijn goed geschud, men kiest blindelings, enz..

1.5 Gebeurtenissen. We voeren enige namen en notaties in. In een vaasmodel, dus in een eindige uitkomstenruimte Ω met n gelijkwaardige uitkomsten $\omega_1, \ldots, \omega_n$, is een *gebeurtenis* een uitspraak over de uitkomst, en dus te identificeren met een deelverzameling van Ω: namelijk met de verzameling van die uitkomsten die de betreffende eigenschap hebben.

Zo is de gebeurtenis in 1.4.2 precies de verzameling

$$\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

een deelverzameling van

$$\begin{aligned}\Omega &= \{1,\ldots,6\}\times\{1,\ldots,6\}\\ &= \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}.\end{aligned}$$

Gebeurtenissen geven we bij voorkeur aan met hoofdletters uit het begin van het alfabet: A, B, C, ..., of, als er veel nodig zijn, A_1, A_2, A_3, ...

Notatie. Wanneer we bijvoorbeeld met 2 dobbelstenen gooien, zullen we de gebeurtenis "er wordt minstens 9 gegooid" (de verzameling $\{(x,y)\in\Omega | x+y\geq 9\}$) ook wel noteren als [minstens 9 ogen].

1.6 Het aantal elementen van een verzameling (gebeurtenis) A geven we aan met $\#A$.
De *kans* dat A optreedt noteren we met $\mathbb{P}(A)$.
We hebben in het vaasmodel de volgende formule (**kansdefinitie van Laplace**):

$$\boxed{\mathbb{P}(A) = \frac{\#A}{\#\Omega}}$$

(Laten we maar aannemen dat $\Omega \neq \emptyset$).

1.7 Vaak is flink wat rekenwerk nodig om $\#A$ vast te kunnen stellen. Probeer maar eens de in 1.4.7 gevraagde kans te bepalen.
Om dergelijk rekenwerk te vergemakkelijken, hebben we een aantal *combinatorische rekenregels* nodig:

(a) Vermenigvuldigingsregel. Als er m mogelijkheden zijn voor de variabele x en voor elke mogelijkheid van x er n mogelijkheden zijn voor de variabele y, dan zijn er $m \times n$ mogelijkheden voor het paar (x, y).
Voorbeeld: bij het werpen van twee dobbelstenen hebben we 6×6 mogelijkheden.

(b) Rangschikking (Permutaties). Een toepassing van (a): n elementen kunnen op $n\times(n-1)\times(n-2)\times\ldots\times 3\times 2\times 1$ manieren gerangschikt worden. Verkorte schrijfwijze:

$$n! = n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1.$$

Afspraak: $0! = 1$.

Voorbeeld. Het aantal manieren waarop we 5 mensen op 5 stoelen kunnen zetten (ieder op 1 stoel) is $5! = 120$.

Intermezzo. In het boek "Our Man in Havanna" van Graham Greene (blz. 114), is de geheime agent in Havanna het sleutelgetal van de safe, waarin zijn geheime zender bewaard wordt, kwijt, net nu hij dringend een bericht moet verzenden. Hij praat dan met zijn assistente:

– I'll remember it in a moment, it's something like 77539.

– We could try all the combinations of 77539. Do you know how many there are?

– Somewhere around six hundred I'd guess.

– I hope your cable is not urgent.

Hoe goed is de gok van de geheime agent?

(c) Geordende grepen. Nog een toepassing van (a). Uit n elementen worden er achtereenvolgens k gekozen ($k \leq n$), waarbij we afspreken dat we op de volgorde letten. Zo'n geordende k-greep kan genoteerd worden als $(x_1, x_2, \ldots, x_k)$. Het aantal mogelijkheden is

$$n \cdot (n-1) \cdot (n-2) \cdots (n-(k-1)) = \frac{n!}{(n-k)!}$$

(dit wordt wel afgekort tot $(n)_k$ in sommige boeken).

Voorbeelden. Een woord van drie verschillende letters kan op $26 \times 25 \times 24$ manieren gevormd worden.

Uit een vereniging van 100 mensen kan een bestuur van voorzitter, secretaris, penningmeester en vice-voorzitter gekozen worden op $\frac{100!}{96!}$ manieren (dus $100 \times 99 \times 98 \times 97$).

(d) Ongeordende grepen. Als (c), maar nu letten we niet op de volgorde. Zo'n "ongeordende k-greep" kan genoteerd worden als $\{x_1, x_2, \ldots, x_k\}$. Het aantal mogelijkheden is nu $k!$ keer zo klein als in (c) (ga na!), dus gelijk aan

$$\frac{n!}{k!(n-k)!} \text{ ofwel } \binom{n}{k}$$

(spreek uit: n-boven-k). Uitgeschreven:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

Deze laatste formule gebruiken we als definitie van $\binom{n}{k}$; met het oog op latere toepassingen definiëren we $\binom{n}{k}$ aldus voor alle $n \in \mathbb{R}$ en alle $k \in \mathbb{N}$.

Voorbeeld. Het aantal manieren waarop we een groepje van drie verschillende letters uit het alfabet kunnen kiezen is $\binom{26}{3} = \frac{26\times25\times24}{3\times2\times1}$.
Uit onze vereniging in (c) kan een afvaardiging van 4 mensen gekozen worden op $\binom{100}{4} = \frac{100\times99\times98\times97}{4\times3\times2\times1}$ manieren.

(e) In vakjes verdelen. Het getal $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ heeft nog een tweede interpretatie, nl. het aantal manieren om n elementen over 2 vakjes te verdelen, waarbij er k in vakje 1 en $n-k$ in vakje 2 geplaatst moeten worden.
In het verenigingsvoorbeeld: k $(= 4)$ mensen die wèl afgevaardigd worden, $n-k$ $(= 96)$ mensen die niet afgevaardigd worden.
Dit kunnen we veralgemenen tot het verdelen van n elementen over r vakjes, waarbij er k_1 in vakje 1, k_2 in vakje 2 , ..., en k_r in vakje r geplaatst dienen te worden (met $k_1 + k_2 + \ldots + k_r = n$). Het aantal mogelijkheden hiervoor is

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{k_r}{k_r}.$$

Reken zelf na dat dit hetzelfde is als

$$\frac{n!}{k_1!k_2!\cdots k_r!}.$$

Voorbeeld. In onze vereniging moet een bestuur gekozen worden, bestaande uit voorzitter, secretaris, penningmeester en drie overige bestuursleden. Dit kan op $\frac{100!}{1!1!1!3!94!}$ manieren.

Trekken met en zonder terugleggen

1.8 Trekken zonder terugleggen

We keren terug naar voorbeeld 1.4.7 en beschouwen nu, algemener, een vaas met N ballen, waarvan er r rood en w wit zijn ($r, w \in \mathbb{N}$; $r + w = N$). We nemen n ballen ($n \leq N$) uit de vaas (zonder teruglegging). Er zijn, wegens 1.7(d), $\binom{N}{n}$ manieren om dit te doen; we nemen aan dat deze alle gelijkwaardig zijn ("de vaas is goed geschud"). We herleiden aldus ons model tot een elementairder "supervaasmodel", waarbij we nu als uitkomsten opvatten alle ongeordende n-tallen ballen: $\omega = \{\omega_1, \ldots, \omega_n\}$ met $\omega_i \neq \omega_j$ als $i \neq j$.
Zij X het aantal rode onder de n uit de vaas gehaalde ballen. We willen de kans bepalen dat $X = k$ (Ga na dat dit alleen zin heeft als $k \in \mathbb{N}$, $k \leq n$, $k \leq r$ en $n - k \leq w$). Er zijn $\binom{r}{k}$ manieren om k rode uit het geheel van r rode te halen. Er moeten ook $n-k$ witte ballen gepakt worden; dit kan op $\binom{w}{n-k}$ manieren. Zo leiden

we, m.b.v. 1.7(a), (d) en de definitie van Laplace, af:

$$\mathbb{P}[X{=}k] = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

(we schrijven, voor de overzichtelijkheid, $\mathbb{P}[\ldots]$ in plaats van $\mathbb{P}([\ldots])$). Men noemt X *hypergeometrisch verdeeld*, met parameters N, r en n.

Voorbeeld: $N = 5$, $r = 2$, $n = 2$. Nummer de ballen 1 t/m 5; de ballen 1 en 2 zijn rood en de overige ballen zijn wit. Ω bestaat uit ongeordende paren getallen uit $\{1, \ldots, 5\}$, zoals $\{3, 5\}$, $\{2, 4\}$ enzovoorts. Dan is $\#\Omega = \binom{5}{2} = 10$.
Ga na dat $[X{=}2] = \{\{1, 2\}\}$ en $[X{=}0] = \{\{3, 4\}, \{3, 5\}, \{4, 5\}\}$. Bepaal zelf de elementen van $[X{=}1]$. Controleer nu de antwoorden die bovenstaande formule oplevert.

1.9 Toepassing.

Een bedrijf schaft 10 nieuwe machines aan, 4 van merk A en 6 van merk B. Na 2 jaar blijken 1 A-machine en 4 B-machines versleten. Alle machines zijn even intensief gebruikt. Kunnen we concluderen dat A-machines gemiddeld langer meegaan dan B-machines?

	Versleten	Intact	Totaal
A	1	3	4
B	4	2	6
Totaal	5	5	10

We vertalen dit probleem in een 2×2-tabel, dat is het volgende vaasmodel: Een vaas bevat 10 ballen waarvan 4 rood (A) en 6 wit (B). We poneren nu een *hypothese*: *stel* dat de A- en de B-machines even goed zijn.
Als die hypothese juist is, dan is de "greep" van 5 (versleten) machines op te vatten als een willekeurige greep van 5 ballen uit deze vaas. De hypothese dat alle machines dezelfde kans hebben om na 2 jaar versleten te zijn, komt overeen met de hypothese dat alle ballen dezelfde kans hebben om in de steekproef van 5 voor te komen.
De vraag is of de hypothese juist is: wijst het resultaat van de steekproef (1 rode bal en 4 witte ballen getrokken) niet op het tegendeel, namelijk dat de A-machines langer meegaan? Of is het toeval, dat het resultaat zo scheef is?
Een beslissing hierover wordt genomen op grond van de kans dat er in een steekproef van 5 uit onze vaas slechts 1 rode bal voorkomt (of zelfs nog minder).
Deze kans blijkt vrij groot te zijn, nl.:

$$\frac{\binom{4}{1}\binom{6}{4}}{\binom{10}{5}} + \frac{\binom{4}{0}\binom{6}{5}}{\binom{10}{5}} \approx 0{,}26$$

daarom is er geen reden om de hypothese te verwerpen, m.a.w. het is best mogelijk dat inderdaad alle machines gelijkwaardig zijn.

1.10 Trekken met terugleggen.

We nemen dezelfde vaas als in paragraaf 1.8. We nemen er weer n ballen uit, maar leggen telkens de gepakte bal terug. We nummeren de ballen weer van 1 tot en met N, en zo dat de ballen 1 t/m r rood zijn en de ballen $r+1$ t/m N wit.

Het is nu mogelijk dat meerdere keren dezelfde bal wordt getrokken, en we zullen dus op een andere manier moeten "boekhouden". Daarom letten we nu (eventjes) wèl op de volgorde. De uitkomstenruimte Ω bestaat dan uit rijtjes $\omega = (\omega_1, \ldots, \omega_n)$, waarbij elke ω_i een element is van $\{1, \ldots, N\}$. Ofwel: $\Omega = \{1, \ldots, N\}^n$. Ω bevat N^n gelijkwaardige uitkomsten (weer: een "supervaas").

Zij X het aantal keren dat een rode bal wordt getrokken; we bepalen weer $\mathbb{P}[X{=}k]$, voor $k \in \mathbb{N}$ met $k \leq n$. We kunnen bijvoorbeeld eerst k rode en dan $n-k$ witte pakken. Dit kan op $r^k w^{n-k}$ manieren. Maar ook elke andere volgorde van pakken, die k rode en $n-k$ witte oplevert, kan op $r^k w^{n-k}$ manieren gerealiseerd worden. En aangezien er $\binom{n}{k}$ mogelijke volgorden zijn die tot k rode en $n-k$ witte ballen leiden, vinden we als antwoord:

$$\mathbb{P}[X{=}k] = \frac{\binom{n}{k} r^k w^{n-k}}{N^n} = \binom{n}{k} p^k q^{n-k}$$

waarbij $p = \frac{r}{N}$ de fractie rode en $q = \frac{w}{N}$ de fractie witte ballen is.

Dit is de beroemde *binomiale verdeling* met parameters n en p. In hoofdstuk 3 komen we erop terug.

Voorbeeld. Dezelfde vaas als in het voorbeeld uit paragraaf 1.8. Nu is $\#\Omega = 25$, Ω bevat elementen als $(3,5)$, $(2,2)$ enzovoorts. Schrijf zelf de elementen op van $[X{=}0]$, $[X{=}1]$ en $[X{=}2]$, en controleer de bijbehorende kansen.

Opgaven

1. Men werpt 3 dobbelstenen. Bepaal de kans op:

   (a) 3 gelijke;

   (b) 3 verschillende aantallen ogen;

   (c) een totaal van minstens 16.

2. Hoe groot is de kans dat in een groep van 20 personen er minstens 2 op dezelfde dag jarig zijn? (Laat schrikkeljaren buiten beschouwing)

3. Uit een goed geschud pak van 52 kaarten krijgt men er 13. Hoe groot is de kans:

   (a) dat er precies 6 $\spadesuit$ bij zijn;

   (b) dat er minstens één aas bij is;

   (c) dat het 3 $\heartsuit$, 3 $\clubsuit$, 3 $\diamondsuit$ en 4 $\spadesuit$ zijn;

   (d) dat het een (3,3,3,4)-hand is (van één kleur 4, van iedere andere 3)?

4. n personen ($n \geq 3$), waaronder Jo en Ien, worden op willekeurige wijze aan een tafel gezet. Hoe groot is de kans:

   (a) dat Jo en Ien naast elkaar zitten, als alle personen aan één kant van de tafel gaan zitten;

   (b) dat Jo en Ien naast elkaar zitten, als de tafel rond is;

   (c) dat Jo en Ien naast elkaar zitten, als n even is, en er aan beide lange zijden van de tafel $\frac{1}{2}n$ personen plaatsnemen;

   (d) dat Jo en Ien in de situatie van (c) tegenover elkaar zitten?

5. Los het probleem van de geheim agent in 1.7 op.

6. Uit een zak met a witte en b zwarte knikkers worden één voor één knikkers genomen totdat er nog één over is. Bereken de kans dat deze overblijvende knikker wit is.

7. Voor 7 mensen zijn 5 bioscoopkaartjes beschikbaar. Op hoeveel manieren kunnen deze verdeeld worden:

   (a) als de 5 kaartjes allemaal voor dezelfde voorstelling zijn;

(b) als de 5 kaartjes allemaal voor verschillende voorstellingen zijn;

(c) als er één kaartje is voor "Bambi" en er vier kaartjes zijn voor "De kleine zeemeermin"?

8. Los het probleem van Galileï in paragraaf 1.1 op.

9. Los het probleem van Pascal en De Méré in paragraaf 1.1 op.

10. Drie personen gooien elk een dobbelsteen; wie het hoogste aantal ogen heeft, wint. Als meerdere personen het hoogste aantal ogen gooien, beginnen ze opnieuw. Wat is de kans dat de beslissing meteen valt?

11. In een kast liggen $2n$ schoenen (n paren) willekeurig door elkaar. Men grijpt blindelings k schoenen ($k \leq n$).

    (a) Hoe groot is de kans dat daar minstens één paar bij is?

    (b) Hoe groot is de kans dat er precies één paar bij is?

12. Lotto. Ik kies zes van de getallen $1, 2, \ldots, 41$. Zes van deze getallen worden door het lot aangewezen. Hoe groot is de kans:

    (a) dat dat juist mijn zes getallen zijn?

    (b) dat ik vijf goede getallen heb?

13. In het pokerspel werpt men 5 dobbelstenen. Bereken de kans op elk van de volgende pokercombinaties:

    (a) poker (5 gelijke);

    (b) carré (4 gelijke, 1 anders);

    (c) vol huis (een drietal en een paar);

    (d) grote straat (2,3,4,5,6);

    (e) kleine straat (1,2,3,4,5);

    (f) drie van één soort (een drietal en twee losse);

    (g) twee paar (twee paren en een losse);

    (h) één paar (een paar en drie losse);

    (i) de rest.

14. Op een kermis beweert iemand telekinetische gaven te hebben. Het publiek gooit munten op, en hij zal proberen om de munten met kop omhoog te laten vallen. Hij slaagt 5 van de 6 keren. Welke kans zou hij hebben op zo'n goed resultaat (of nog beter), wanneer hij geen bovennatuurlijke gaven bezat?

15. Bij het Risk-spel spelen telkens twee spelers tegen elkaar: de 'aanvaller' en de 'verdediger'.
We beschouwen nu één enkel duel. Daarbij werpt de aanvaller met drie dobbelstenen, de verdediger met één of met twee.
We noemen de uitslagen van de aanvaller: X_1, X_2 en X_3, en nummeren deze van hoog naar laag; dus $6 \geq X_1 \geq X_2 \geq X_3 \geq 1$.

    (a) Stel eerst dat de verdediger met één dobbelsteen werpt. (Uitslag: $Y \in \{1,2,3,4,5,6\}$). We zeggen dat hij *één slag wint* wanneer $Y \geq X_1$. Bereken de kans hierop.

    (b) Stel nu dat de verdediger met twee dobbelstenen werpt. We noemen de uitslagen: Y_1 en Y_2 ($6 \geq Y_1 \geq Y_2 \geq 1$).
    We zeggen dat de verdediger *de eerste slag wint* als $Y_1 \geq X_1$, en *de tweede slag wint* als $Y_2 \geq X_2$.
    Zij V_i het aantal door de verdediger gewonnen slagen minus het aantal door hem verloren slagen, bij verdediging met i dobbelstenen ($i = 1$ of 2).
    De verdediger mag, nadat hij de uitslag van de aanvaller gezien heeft, kiezen of hij met één of met twee dobbelstenen verdedigt.
    Stel dat de uitslag van de aanvaller is: $X_1 = 5$, $X_2 = 4$ en $X_3 = 2$.
    Bereken de kansen $\mathbb{P}(V_i = j)$ voor $i = 1, 2$; $j = -2, 0, 2$.

    (c) Als je in de situatie van onderdeel (c) verdediger was, met hoeveel dobbelstenen zou je dan verdedigen (bewijs is niet nodig, alleen een argumentatie is voldoende)?

    (d) Overigens: heb je wel eens Risk gespeeld? (Het antwoord op deze vraag beïnvloedt je practicumcijfer niet.)

Hoofdstuk 2

Het algemene model

2.1 In hoofdstuk 1 hebben we al een flink instrumentarium opgebouwd waarmee we vele problemen kunnen oplossen. Toch blijven er nog genoeg vraagstukken over die eenvoudig te formuleren zijn, maar niet opgelost kunnen worden met de methoden uit hoofdstuk 1. Enige voorbeelden:

1. In casino's speelt men een gokspel, roulette, waarbij alle mogelijke uitkomsten: $0, 1, 2, \ldots, 36$ dezelfde kans ($\frac{1}{37}$) zouden moeten hebben. De man die het balletje werpt, kan echter de vaardigheid ontwikkelen om het balletje met grote kans in een bepaald deel van de getallencirkel terecht te laten komen, zodat de gelijkwaardigheid der uitkomsten verloren gaat.

2. Werp een (eerlijke) dobbelsteen. Hoe lang duurt het vòòr de eerste 6 verschijnt? Dit kan willekeurig lang duren: de uitkomstenruimte is (aftelbaar) oneindig geworden.

3. Werp een pijltje op een schijf met een straal van 20 cm. Hoe groot is de kans dat het pijltje hoogstens 3 cm van het middelpunt terecht komt?

Uitkomstenruimten

2.2 Blijkbaar moeten we twee essentiële vooronderstellingen uit hoofdstuk 1 laten vallen: de uitkomsten hoeven niet allen gelijkwaardig te zijn (2.1.1), en de uitkomstenruimte hoeft niet eindig te zijn (2.1.2 en 2.1.3).
De uitkomstenruimte Ω van een experiment kan een eindige of oneindige, aftelbare[1] of overaftelbare verzameling zijn.
Voorbeelden: $\{1, 2, 3, 4, 5, 6\}$, $\{\text{kop}, \text{munt}\}$, $\{\text{januari}, \text{februari}, \ldots, \text{december}\}$, $\mathbb{N}$, $\mathbb{Z}$, $[0, 1]$, $[0, 1] \times [0, 1]$.

[1]Een oneindige verzameling V heet *aftelbaar* als de elementen van V genummerd kunnen worden.

2.3 Wanneer we binnen een uitkomstenruimte gebeurtenissen $A, B, A_1, A_2, \ldots$ hebben, dan kunnen we daaruit allerlei nieuwe gebeurtenissen maken door verzamelingsoperaties toe te passen.

Voer eenmaal het betreffende experiment uit en noem de uitkomst ω (Je kunt ook zeggen: laat het toeval een uitkomst ω aanwijzen).

Dan correspondeert de uitspraak "A en B treden beide op" met "$\omega \in A \cap B$". We zetten een aantal van dergelijke verbanden op een rijtje.

Vooraf enige afspraken: de verzameling $\{\omega : \omega \in \Omega, \omega \notin A\}$ heet het *complement* van A en wordt genoteerd als A^c. Het *verschil* van A en B definiëren we als $A \cap B^c$; we schrijven ervoor $A \setminus B$.

(a)	A treedt op	$\omega \in A$
(b)	A en B treden beide op	$\omega \in A \cap B$
(c)	$A_1, A_2, \ldots, A_n$ treden alle op	$\omega \in \bigcap_{k=1}^{n} A_k$
(d)	A of B treedt op (minstens één van de twee)	$\omega \in A \cup B$
(e)	Minstens één van de gebeurtenissen $A_1, \ldots, A_n$ treedt op	$\omega \in \bigcup_{k=1}^{n} A_k$
(f)	Of A òf B treedt op (precies één van de twee)	$\omega \in (A \cup B) \setminus (A \cap B)$. Deze verzameling heet ook wel het *symmetrisch verschil* van A en B, en wordt genoteerd als $A \triangle B$.
(g)	A treedt niet op	$\omega \in A^c$

Verder kunnen er verbanden bestaan tussen het optreden van A en B, bv.:

(h)	A en B sluiten elkaar uit, kunnen niet tegelijk optreden	$A \cap B = \emptyset$ (A en B zijn *disjunct*)

(i)	Uit A volgt B, d.w.z. als A optreedt, dan zeker ook B	$A \subset B$

2.4 * We willen dadelijk aan iedere gebeurtenis een kans toekennen. Maar de uitkomstenruimte Ω kan zó veel deelverzamelingen hebben (anders gezegd: $\mathcal{P}(\Omega)$ kan zó groot zijn), dat het niet mogelijk is om aan alle deelverzamelingen van Ω een kans toe te kennen. Daarom beperken we ons veelal tot een deelklasse $\mathcal{G}$ van $\mathcal{P}(\Omega)$, en slechts elementen van die deelklasse $\mathcal{G}$ zullen we gebeurtenissen noemen.
We vooronderstellen dat $\mathcal{G}$ de volgende eigenschappen heeft:

(a) $\Omega \in \mathcal{G}$;
(b) Als $A \in \mathcal{G}$, dan $A^c \in \mathcal{G}$;
(c) Als $A_1, A_2, \ldots \in \mathcal{G}$, dan $\bigcap_{n=1}^{\infty} A_n \in \mathcal{G}$.
(Men noemt zo'n $\mathcal{G}$ wel een σ-algebra).
In woorden: combinaties van gebeurtenissen, zoals ingevoerd in paragraaf 2.3, zijn zelf óók gebeurtenissen.
Als Ω (eindig of) aftelbaar is, kunnen we altijd $\mathcal{G} = \mathcal{P}(\Omega)$ nemen; bij overaftelbare Ω zal $\mathcal{G}$ i.h.a. echt kleiner zijn dan $\mathcal{P}(\Omega)$.
We zullen verder aan deze beperkingen geen aandacht schenken (dat gebeurt wèl in het college "Maat en Integraal"). We nemen steeds aan dat alle voorkomende deelverzamelingen van Ω gebeurtenissen zijn. In de praktijk, dat betekent voor ons: bij het oplossen van vraagstukken, is dit altijd het geval.

Kansruimten

2.5 Ons doel is, zoals aangekondigd, aan elke gebeurtenis A een kans toe te kennen, die we zullen noteren met $\mathbb{P}(A)$. In feite is $\mathbb{P}$ dus een functie op (een deel van) $\mathcal{P}(\Omega)$.

De formele definitie van kansruimte:
Een *kansruimte* bestaat uit een uitkomstenruimte Ω, een collectie gebeurtenissen $\mathcal{G}$ ($\mathcal{G} \subset \mathcal{P}(\Omega)$) en een functie $\mathbb{P} : \mathcal{G} \to [0, 1]$ die aan de volgende voorwaarden voldoet:

(a) $\mathbb{P}(\Omega) = 1$
(b) Als A en B elkaar uitsluitende gebeurtenissen zijn, dan is $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (men zegt wel: "$\mathbb{P}$ is additief").

Voor oneindige kansruimten stellen we bovendien de volgende (continuïteits)eis:

(c) Als $A_1, A_2, \ldots$ gebeurtenissen zijn met $A_1 \subset A_2 \subset A_3 \subset \ldots$ (men zegt wel: de rij $A_1, A_2, \ldots$ is monotoon stijgend), dan is $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

$\mathbb{P}$ heet de kansfunctie, of kortweg *kans*.
Met volledige inductie bewijs je uit eis (b) eenvoudig dat voor elkaar uitsluitende gebeurtenissen $A_1, \ldots, A_n$ geldt:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Maar eis (c) volgt *niet* uit de eisen (a) en (b) (Bewijs: zie het college "Maat en Integraal").
Ten overvloede zij opgemerkt, dat $\mathbb{P}(A)$ altijd een getal in $[0,1]$ is.

2.6 De in de vorige paragraaf genoemde voorwaarden lijken misschien wat mager, maar het blijkt dat er veel formules uit af te leiden zijn (zie opgave 2). We schrijven er een aantal op (met A, B, $C \in \mathcal{G}$):

(a) $\mathbb{P}(\emptyset) = 0$;
(b) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
(c) $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$;
(d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cap B^c)$;
(e) $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$;
(f) $$\begin{aligned}\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ + \mathbb{P}(A \cap B \cap C);\end{aligned}$$
(g) Als $A \subset B$, dan $\mathbb{P}(A) \leq \mathbb{P}(B)$.

2.7 Met behulp van de eigenschappen van de kansfunctie $\mathbb{P}$ in paragraaf 2.5 en 2.6 kan men het volgende bewijzen:

(a) Als $A_1, A_2, \ldots$ gebeurtenissen zijn met $A_1 \supset A_2 \supset A_3 \supset \ldots$ (men zegt wel: de rij $A_1, A_2, \ldots$ is monotoon dalend), dan is

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(b) Als $A_1, A_2, \ldots$ elkaar uitsluitende gebeurtenissen zijn, dan is

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

2.8 Als Ω aftelbaar is, en we dus $\mathcal{G} = \mathcal{P}(\Omega)$ kunnen nemen, dan is $\mathbb{P}$ gemakkelijk vast te leggen door elke uitkomst $\omega \in \Omega$ op te vatten als gebeurtenis $A = \{\omega\}$, en de kans dáárop vast te leggen. Als volgt: schrijf $p_\omega = \mathbb{P}(\{\omega\})$ voor $\omega \in \Omega$. Wegens 2.7(b) moet nu voor elke $A \in \mathcal{G}$ gelden:

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega.$$

Merk op dat deze som altijd bestaat en eindig is. De sommatievolgorde is niet van belang, want de termen zijn allemaal positief.
Controleer zelf de voorwaarden in paragraaf 2.5.

2.9 Opmerking. Uit deze constructie volgt dat in een aftelbaar oneindige uitkomstenruimte nooit alle uitkomsten gelijkwaardig kunnen zijn. Stel namelijk maar dat er een constante $c \in [0,1]$ bestaat met $\mathbb{P}(\{\omega\}) = c$ voor elke $\omega \in \Omega$. Als $c > 0$, dan volgt: $\mathbb{P}(\Omega) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega} c = \infty$, in tegenspraak met de axioma's in paragraaf 2.5. En als $c = 0$, dan volgt op dezelfde wijze: $\mathbb{P}(\Omega) = 0$, hetgeen evenzeer onjuist is.

Stochasten

2.10 Vaak zijn we niet zozeer geïnteresseerd in de uitkomst ω zelf van het experiment, maar in een functie van ω. De uitkomst ω van het experiment en daarmee de waarde van deze functie wordt bepaald door het toeval; daarom heet deze functie toevalsgrootheid, of stochastische variabele, of kortweg stochast[2]. We gebruiken voor stochasten meestal hoofdletters uit het eind van het alfabet: X, Y, Z, X_1, $X_2, \ldots$
Een *stochast* X is dus een functie $X : \Omega \to W$, waarbij W de waardenverzameling van X is. Als $W \subset \mathbb{R}$ heet X een *stochastische grootheid* (random variable); als $W \subset \mathbb{R}^n$ voor een of andere n, dan heet X een *stochastische vector*. Als W aftelbaar of zelfs eindig gekozen kan worden, noemen we X een *discrete stochast*.

2.11 Voorbeelden van stochasten.

[2]Grieks: $\sigma\tau o\chi\alpha\zeta o\mu\alpha\iota$ = gissen of mikken; $\sigma\tau o\chi\alpha\sigma\tau o$ = datgene waarmee men gist; werpspeer.

1. We doen 2 worpen met een dobbelsteen, en zijn slechts geïnteresseerd in de som X van de beide worpen. Een uitkomst $\omega = (\omega_1, \omega_2)$ van het experiment levert voor X de waarde $X(\omega) = \omega_1 + \omega_2$. De stochast X heeft waardenverzameling $\{2, 3, 4, \ldots, 12\}$ en is derhalve discreet.

2. We herhalen het experiment in 1, maar kijken nu slechts naar de hoogste, Y, van de twee worpen, Dan is $Y(\omega) = \max(\omega_1, \omega_2)$, een stochast met waarden in $\{1, 2, 3, 4, 5, 6\}$, en dus discreet.

3. Bij het werpen op de schijf noteren we de uitkomsten door middel van coördinaten $\omega = (x, y)$. De roos is $(0, 0)$. We zijn geïnteresseerd in de horizontale afwijking X t.o.v. de roos, de verticale afwijking Y t.o.v. de roos, en de afstand R tot de roos.
   Dan zijn $X(x, y) = x$, $Y(x, y) = y$, terwijl $R(x, y) = \sqrt{x^2 + y^2}$. Merk op dat $R = \sqrt{X^2 + Y^2}$.
   Elke worp wordt met een aantal punten Z gewaardeerd, als volgt: $Z(x, y) = n$ als $\frac{1}{n+1} < R(x, y) \leq \frac{1}{n}$ (dus: $Z = [\frac{1}{R}]$). Ga na dat X, Y en R niet discreet zijn, en Z wel.

2.12 Stochasten zijn we ongemerkt al tegengekomen in paragraaf 1.8. Ze maken het makkelijk om kort en duidelijk gebeurtenissen te omschrijven. Zo wordt de gebeurtenis "geen vijven of zessen in twee worpen met een dobbelsteen" in 2.11.2: $\{\omega : Y(\omega) \leq 4\}$, in de notatie van 1.5: $[Y \leq 4]$.
Voorbeeld. Als we drie dobbelstenen gooien met aantallen ogen X_1, X_2, resp. X_3, dan is de gebeurtenis "de laatste worp is de hoogste" te schrijven als

$$[X_1 \leq X_3 \text{ en } X_2 \leq X_3]$$

hetgeen weer de afkorting is van

$$\{\omega \in \Omega : X_1(\omega) \leq X_3(\omega) \text{ en } X_2(\omega) \leq X_3(\omega)\}.$$

2.13 Beschouw een discrete stochast X met waardenverzameling W. We zijn geïnteresseerd in de *kansverdeling* van X, d.w.z. de waarde van $\mathbb{P}[X{=}x]$ voor $x \in W$. Als we deze kansen eenmaal kennen, liggen alle verdere kansen op gebeurtenissen die betrekking hebben op X, vast door 2.7(b). Zij nl. $V \subset W$, dan is $\mathbb{P}[X{\in}V] = \sum\limits_{x \in V} \mathbb{P}[X{=}x]$.
Voorbeeld. Neem de stochast Y uit 2.11.2. Reken na dat $\mathbb{P}[Y{=}k] = \frac{2k-1}{36}$ voor $k = 1, 2, 3, 4, 5, 6$. Bereken zelf de kansverdeling van de stochast X uit 2.11.1.

2.14 Bij elke gebeurtenis A kunnen we een stochast $\mathbf{1}_A$ maken door te definiëren:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{als } \omega \in A \\ 0 & \text{als } \omega \in A^c. \end{cases}$$

$\mathbf{1}_A$ heet de *indicator* (aanwijzer) van A. Van deze stochast kunnen we de kansverdeling eenvoudig bepalen:

$$\begin{aligned} \mathbb{P}[\mathbf{1}_A{=}1] &= \mathbb{P}(A); \\ \mathbb{P}[\mathbf{1}_A{=}0] &= \mathbb{P}(A^c) = 1 - \mathbb{P}(A). \end{aligned}$$

Indicatoren worden vaak gebruikt om berekeningen waarin een discrete stochast X voorkomt, te vereenvoudigen. Men kan namelijk X "ontbinden" m.b.v. indicatoren:

$$X = \sum_{x \in W} x \cdot \mathbf{1}_{[X=x]}$$

waar W de waardenverzameling van X is. Controleer deze identiteit. Later komen we hierop terug.

2.15 Een stochast X die slechts de waarden 0 en 1 aanneemt heet *alternatief verdeeld*. Als $\mathbb{P}[X{=}1] = p$ (en dus $\mathbb{P}[X{=}0] = 1-p$), dan zeggen we dat X alternatief verdeeld is met parameter p. We schrijven wel: $X \sim \mathrm{Alt}(p)$.
Uit paragraaf 2.14 volgt dat voor gebeurtenissen A geldt:

$$\mathbf{1}_A \sim \mathrm{Alt}(\mathbb{P}(A)).$$

2.16 Stel we hebben een experiment met uitkomstenruimte Ω, en daarbij twee discrete stochasten X en Y met waardenverzameling V resp. W. Dan kunnen we (X, Y) opvatten als een discrete stochast met waarden in (een deelverzameling van) $V \times W$. Ook deze nieuwe stochast is discreet en heeft een kansverdeling, nl.:

$$\mathbb{P}[(X, Y) = (x, y)] = \mathbb{P}[X{=}x \text{ en } Y{=}y]$$

(voluit: $\mathbb{P}(\{\omega : X(\omega) = x \text{ en } Y(\omega) = y\})$). Deze kansverdeling heet de *simultane kansverdeling* van de stochasten X en Y. De kansverdeling van de stochast X alleen noemt men de *marginale kansverdeling* van X.
Voorbeeld. Werp twee dobbelstenen. Uitkomst: $\omega = (\omega_1, \omega_2)$. Zij X_1 het aantal ogen van de eerste worp: $X_1(\omega) = \omega_1$, en evenzo $X_2(\omega) = \omega_2$, en zij X de som der beide worpen: $X(\omega) = \omega_1 + \omega_2$. Dan is de simultane kansverdeling van X_1 en X_2:

$$\begin{aligned} \mathbb{P}[(X_1, X_2) = (x_1, x_2)] &= \\ \mathbb{P}[X_1{=}x_1 \text{ en } X_2{=}x_2] &= \begin{cases} \frac{1}{36} & \text{als } x_1, x_2 \in \{1, \ldots, 6\} \\ 0 & \text{anders;} \end{cases} \end{aligned}$$

de simultane kansverdeling van X_1 en X is:

$$\begin{aligned}
\mathbb{P}[(X_1, X) = (x, y)] &= \\
\mathbb{P}[X_1{=}x \text{ en } X_2{=}y-x] &= \begin{cases} \frac{1}{36} & \text{als } x, y{-}x \in \{1, \ldots, 6\} \\ 0 & \text{anders.} \end{cases}
\end{aligned}$$

De marginale kansverdelingen van X_1 en X_2 kennen we natuurlijk allang:

$$\mathbb{P}[X_1{=}x] = \tfrac{1}{6}$$

$$\mathbb{P}[X_2{=}x] = \tfrac{1}{6}$$

als $x = 1, \ldots, 6$. De marginale kansverdeling van X heb je in paragraaf 2.13 uitgerekend.

2.17 Discreet of niet discreet.

In een aftelbare uitkomstenruimte krijg je de kans op een gebeurtenis A door de kansen p_ω van de afzonderlijke uitkomsten waarbij A gerealiseerd wordt bij elkaar op te tellen (zie paragraaf 2.8). Dit recept geldt alleen voor de zogenaamde *discrete situatie*, waarbij de totale “kansmassa” (1 dus) opgedeeld is in discrete porties p_ω (“puntmassa’s”) geconcentreerd in afzonderlijke punten ω van Ω. Deze punten, het zijn er hoogstens aftelbaar veel, vormen een verzameling D (de zogenaamde drager van de kansmaat) die alle kans opslokt: $\mathbb{P}(D) = \sum_{\omega \in D} p_\omega = 1$. Dus $\mathbb{P}(\Omega \setminus D) = 0$. Het kan zijn dat $D = \Omega$, maar dat hoeft niet, het is best mogelijk dat $D = \mathbb{N}$ en $\Omega = \mathbb{R}$.

Naast deze discrete situatie zullen we ons verderop (hoofdstuk 8 en verder) uitgebreid bezighouden met situaties waarbij de kansmassa continu (men zegt ook wel diffuus) is uitgespreid over de uitkomstenruimte, die dan noodzakelijk overaftelbaar is. Karakteristiek daarbij is dat iedere afzonderlijke uitkomst kans nul krijgt: $\mathbb{P}(\{\omega\}) = 0$ voor iedere $\omega \in \Omega$.

Een eenvoudig, maar belangrijk voorbeeld is het volgende:

2.18 Uniforme toevalsgetallen (Uniform Random Numbers).

De uitkomstenruimte bestaat nu uit het interval $[0, 1]$ in de reële getallen, $\Omega = [0, 1]$. De kansmaat $\mathbb{P}$ is zodanig dat voor ieder interval $I = (a, b)$ (en ook voor elk interval $[a, b]$) geldt: $\mathbb{P}(I) = b - a$, de lengte van I.

Dit is wat gewoonlijk bedoeld wordt wanneer men het heeft over het kiezen (eventueel door het toeval) van een willekeurig reëel getal tussen 0 en 1.

Definieer je bij deze uitkomstenruimte de stochast U door $U(\omega) = \omega$, dan geldt:

$\mathbb{P}[a < U < b] = b - a$ als $0 \le a \le b \le 1$ en $\mathbb{P}[U \le x] = \mathbb{P}[U < x] = x$ voor $x \in [0, 1]$.

Zo'n stochast U noemt men uniform verdeeld over $[0, 1]$; notatie: $U \sim \mathrm{Un}(0, 1)$.

Toepassing. Treinen in de richting Juistweg vertrekken van station Netgemist met tussenpozen van precies een uur. Zonder het spoorboekje te raadplegen ga je op goed geluk naar het station. Je hoeft hoogstens een uur te wachten op de eerstvolgende trein naar Juistweg. Het kan zijn dat je minder dan 20 minuten moet wachten, maar net zo goed kan je wachttijd meer dan 40 minuten zijn. Het lijkt redelijk om aan te nemen dat de wachttijd U (in uren) uniform$(0, 1)$ verdeeld is.

Opmerking. Er is hier sprake van idealisering. In de werkelijkheid zul je de wachttijd hoogstens tot op één minuut nauwkeurig willen weten. De wachttijd (in uren) wordt dan beschreven door de discrete stochast U' met $\mathbb{P}[U' = \frac{k}{60}] = \frac{1}{60}$ voor $k = 0, 1, \ldots, 59$.

Iemand anders wil de wachttijd op de seconde nauwkeurig hebben, dat geeft een andere beschrijving. Het zal duidelijk zijn dat de geïdealiseerde continue beschrijving hier het voordeel van de eenvoud heeft.

2.19 Simulatie.

Computers (en zelfs zakrekenmachientjes met een knopje RND of RAN) kunnen op aanroep een getal tussen 0 en 1 produceren, dat dienst kan doen als uniform toevalsgetal. Bij benadering tenminste, want het geproduceerde getal heeft een beperkt aantal decimalen, 8 bijvoorbeeld, zodat we in feite een discrete stochast hebben met 10^8 mogelijke waarden, waarvan we hopen dat ze allemaal dezelfde kans hebben.

Heb je een stochast $U \sim \mathrm{Un}(0, 1)$ dan kun je daarmee stochasten maken met iedere gewenste verdeling, zowel discreet als continu. Het werpen van een dobbelsteen kan als volgt nagebootst (gesimuleerd) worden:

Laat U het door de computer geproduceerde toevalsgetal zijn. Stel $X = [6U] + 1$. Ga na dat $\mathbb{P}[X = k] = \mathbb{P}[\frac{k-1}{6} \le U < \frac{k}{6}] = \frac{1}{6}$ als $U \sim \mathrm{Un}(0, 1)$.

Bedenk zelf een recept voor een stochast $Y \sim \mathrm{Alt}(p)$ (wat denk je van $Y = [U + p]$?).

2.20 Intuïtieve achtergrond van het kansbegrip.

Als je 600 keer een eerlijke dobbelsteen gooit, hoe vaak verwacht je dan vijf ogen te gooien? Een redelijk antwoord lijkt: ongeveer 100 keer, waarbij je met "ongeveer" zou kunnen bedoelen dat het wel gek moet gaan als je minder dan 70 of meer dan 130 keer vijf ogen krijgt. Experimenteel blijkt dat dergelijke relatief grote afwijkingen inderdaad weinig voorkomen. Dit ervaringsfeit staat bekend als de *Experimentele wet van de grote aantallen.*

Als je 6 keer een dobbelsteen gooit, hoe vaak verwacht je dan vijf ogen te gooien? "Gemiddeld één keer" lijkt een redelijk antwoord, maar relatief grote afwijkingen als nul keer of twee keer zijn zeer wel mogelijk (kans ongeveer 0,33 resp. 0,20).
Het begrip kans zegt dus i.h.a. weinig over de uitkomst van één enkel experiment, maar krijgt pas betekenis als het experiment heel vaak herhaald wordt.
Zij n het aantal experimenten, en n_A het aantal experimenten daaronder die gebeurtenis A opleveren. Dan zal het frequentiequotiënt $\frac{n_A}{n}$ met grote kans dicht bij $\mathbb{P}(A)$ liggen als n groot is.
De precieze formulering van deze uitspraak staat bekend als de *Theoretische (zwakke) wet van de grote aantallen*; deze stelling zullen we in hoofdstuk 6 bewijzen.

Voor gebeurtenissen met zeer kleine of zeer grote kans, kan men wèl een zinnige voorspelling doen over de uitkomst van één enkel experiment. Als de kans op een gebeurtenis A bijvoorbeeld kleiner is dan 0,001, dan zegt men wel: we mogen er op vertrouwen, dat A bij eenmalige uitvoering van het experiment niet optreedt, dat risico kunnen we wel nemen.
Verzin zelf een interpretatie voor het geval dat $\mathbb{P}(A) > 0,999$.

Opgaven

1. Laat A, B en C willekeurige gebeurtenissen zijn. Druk de volgende gebeurtenissen in A, B en C uit:

   (a) Geen van de drie (treedt op);

   (b) Precies één van de drie;

   (c) Minstens twee van de drie.

2. Bewijs de formules in paragraaf 2.6.

3. Bewijs:

   (a) $\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mathbb{P}(A_k)$;

   (b) $\sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j\leq n} \mathbb{P}\left(A_i \cap A_j\right) \leq \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mathbb{P}(A_k)$.

4. Bewijs:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{k\leq n} \mathbb{P}(A_k) - \sum_{k_1<k_2\leq n} \mathbb{P}(A_{k_1} \cap A_{k_2}) \\ &\quad + \sum_{k_1<k_2<k_3\leq n} \mathbb{P}(A_{k_1} \cap A_{k_2} \cap A_{k_3}) - \dots \\ &\quad - (-1)^n \mathbb{P}(A_1 \cap \dots \cap A_n).\end{aligned}$$

   Deze generalisatie van 2.6(d) en (f) noemt men wel de "regel van inclusie en exclusie".

5. Lootjes trekken. n personen besluiten lootjes te trekken voor Sinterklaas. Wat is de kans dat minstens één van de personen het lootje met zijn eigen naam trekt (en er dus opnieuw getrokken moet worden)?
   Aanwijzing: Neem $A_i = [i^{\text{e}}$ persoon trekt zijn eigen lootje]. Gebruik de regel van inclusie en exclusie. Antwoord:

$$1 - \frac{1}{2!} + \frac{1}{3!} - \dots - (-1)^n \cdot \frac{1}{n!}$$

   (voor grote n is dit ongeveer $1 - \frac{1}{e}$).

6. We werpen twee dobbelstenen en noteren de uitkomst als $\omega = (\omega_1, \omega_2)$. X is het verschil tussen de beide worpen: $X(\omega) = |\omega_1 - \omega_2|$. Bepaal de kansverdeling van X.

7. Bewijs 2.7(a) en (b). Aanwijzing bij (b): Zij $B_1 = A_1$, $B_n = A_n \cup B_{n-1} (n \geq 2)$, dan is de rij $B_1, B_2, \ldots$ monotoon stijgend en $\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k$ en $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.

8. Bewijs:
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

9. Jan en Piet werpen om de beurt een dobbelsteen. Jan begint. Degene die als eerste een zes gooit, wint.

   (a) Bepaal de kans dat Jan wint. Aanwijzing: neem $\Omega = \{1, 2, \ldots, \infty\}$. Uitkomst $k \in \mathbb{N}$ betekent: eerste zes bij de k^{e} worp. De kans hierop is $\frac{5^{k-1}}{6^k}$. Uitkomst ∞ betekent: er wordt nooit een zes gegooid.

   (b) Bewijs zo ook dat de kans dat Jan of Piet wint, 1 is.

10. A en B komen (onafhankelijk van elkaar) tussen 0 en 1 uur binnen in café "Houdoe". Neem voor Ω het vierkant van alle punten (x, y), $0 \leq x \leq 1$, $0 \leq y \leq 1$, waarin x en y de aankomsttijden van A en B voorstellen. Veronderstel dat de kans dat het punt (x, y) in zeker deel van Ω valt, evenredig is met de oppervlakte van dat deel. A en B blijven ieder een kwartier in het café. Bepaal de kans dat ze elkaar daar aantreffen.

11. n rode en n blauwe kralen worden in een kring gelegd ($n \in \mathbb{N}$, $n \geq 2$). Zij X het aantal der rode kralen die minstens één rode buurkraal hebben. Bepaal $\mathbb{P}[X{=}1]$ en $\mathbb{P}[X{=}0]$.

12. In een dichtgevouwen hand bevinden zich tien touwtjes die elk met één uiteinde boven de hand en met het andere uiteinde onder de hand uitsteken. De tien bovenuiteinden worden twee aan twee aan elkaar gebonden; idem voor de tien onderuiteinden. Wat is de kans dat de tien touwtjes, na opening van de hand, samen één gesloten ring vormen?

13. Er worden 12 dobbelstenen geworpen. Er zijn nu twee mogelijkheden: Alle getallen van 1 t/m 6 komen minstens één keer in de worp voor, of van de getallen 1 t/m 6 komen er één of meer niet voor in de worp.
    Welke gebeurtenis heeft de grootste kans?
    Aanwijzing. Stel $A_i = [i \text{ ogen komt niet voor}]$, en gebruik de ongelijkheid van opgave 3(b).

Hoofdstuk 3

Onafhankelijkheid

Voorwaardelijke kansen

3.1 Voorbeeld: We kunnen de nederlandse bevolking enerzijds indelen naar mannen en vrouwen; anderzijds naar personen onder de 65 jaar en personen van 65 jaar en ouder. De volgende tabel[1] geeft een overzicht van deze gegevens voor 1990:

%	< 65	≥ 65	Totaal
Man	44,3	5,1	49,4
Vrouw	42,9	7,7	50,6
Totaal	87,2	12,8	100,0

Als we een willekeurige Nederlander aanwijzen, hoe groot is dan de kans dat die 65 jaar of ouder is?
Als er niets bekend zou zijn, dan is de kans op $B =$ [65 of ouder] gelijk aan 12,8%. Weten we echter dat deze Nederlander een vrouw is, dus dat $A =$ [vrouw] optreedt, dan is het veel waarschijnlijker geworden, dat deze persoon 65 jaar of ouder is. Immers, voor 15,2% van de vrouwen uit A geldt, dat ze 65 jaar of ouder zijn. Het ligt dus voor de hand om te zeggen: de kans dat de persoon 65 jaar of ouder is als *gegeven* is dat de persoon een vrouw is, is 15,2%.
Het woordje "als" geeft hier aan dat we te doen hebben met een zogenaamde voorwaardelijke kans: de voorwaardelijke kans op B, gegeven A. Die kans vinden we dus door binnen A (de voorwaarde) te kijken welke uitkomsten bovendien tot B, dus tot $A \cap B$ behoren.

3.2 Definitie: Zij A een gebeurtenis met $\mathbb{P}(A) \neq 0$. We definiëren dan de *voorwaardelijke kans op B gegeven A*, notatie $\mathbb{P}(B|A)$, als volgt:

[1]Gegevens ontleend aan: Statistisch Jaarboek 1991, Centraal Bureau voor de Statistiek.

$$\boxed{\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.}$$

In feite verklein je zo de uitkomstenruimte Ω tot de uitkomstenruimte A.
Het is gebruikelijk om in het geval dat $\mathbb{P}(A) = 0$ te zeggen dat $\mathbb{P}(B|A)$ niet gedefinieerd is. Voor zulke A is de definitie immers zinloos geworden.
Maar de formule $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$ die je uit de definitie kunt afleiden, is wèl geldig voor A met $\mathbb{P}(A) = 0$, immers, dan is ook $\mathbb{P}(A \cap B) = 0$.
Voorbeeld. Laat A en B als in paragraaf 3.1. Dan is $\mathbb{P}(B|A) = 15{,}2\%$, en $\mathbb{P}(A|B) = 60{,}2\%$. Merk op dat $\mathbb{P}(B|A) \neq \mathbb{P}(A|B)$!

	♠	♡	♢	♣
Aas				
Heer				
Vrouw				
Boer				♣B
10				
9				
8				
7				
6				
5				
4				
3				
2				

3.3 De 52 kaarten van het kaartspel zijn op twee manieren in te delen: naar soort (♠, ♡, ♢, ♣) en naar rang (aas, heer,. . . ,2). Iemand neemt blindelings een kaart uit de stapel en zegt dat het ♣ is. Geeft deze mededeling ons enige informatie over de rang van de getrokken kaart, bijvoorbeeld, of deze kaart een boer is? Nee: binnen iedere kleur is de verdeling over de rangen precies dezelfde. In meer wiskundige termen: $\mathbb{P}(\text{Boer}|♣) = \mathbb{P}[\text{Boer}]$, en dus: $\mathbb{P}[♣\ \text{Boer}] = \mathbb{P}[♣] \cdot \mathbb{P}[\text{Boer}]$.
Blijkbaar beïnvloeden de gebeurtenissen [Boer] en [♣] elkaar niet. We zullen zulke "dwars door elkaar heenlopende" gebeurtenissen onafhankelijk noemen.

Merk op dat we eigenlijk $\mathbb{P}([\text{Boer}]|[♣])$ zouden moeten schrijven. We zullen echter, om een woud van haakjes en haken te vermijden, de notatie $\mathbb{P}(\text{Boer}|♣)$ aanhouden.

Onafhankelijkheidsdefinities

3.4 Definitie. Stel A en B zijn gebeurtenissen. We noemen A en B *onafhankelijk*, notatie $A \perp\!\!\!\perp B$, als

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Als de gebeurtenissen A en B niet onafhankelijk zijn, dan noemen we ze *afhankelijk*.

Merk op dat voor gebeurtenissen A en B geldt:

$$A \perp\!\!\!\perp B \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$$

(als tenminste $\mathbb{P}(A) \neq 0$) en ook:

$$A \perp\!\!\!\perp B \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$$

(als tenminste $\mathbb{P}(B) \neq 0$).
De mededeling $A \perp\!\!\!\perp B$ wil dus zoiets zeggen als: A en B beïnvloeden elkaar niet, de kans dat B optreedt hangt niet af van het optreden van A, en andersom.
Het begrip onafhankelijkheid is de kern van de kansrekening.
Voorbeelden. De gebeurtenissen A en B uit paragraaf 3.1 zijn niet onafhankelijk; de gebeurtenissen [boer] en [♣] uit paragraaf 3.3 zijn het wel.

3.5 Het is intuïtief duidelijk, dat als het optreden van A het optreden van B niet beïnvloedt, het optreden van A ook geen invloed heeft op het niet-optreden van B, en andersom. Dit valt wiskundig ook precies te maken:
Stellinkje. Stel A en B zijn gebeurtenissen zo dat $A \perp\!\!\!\perp B$. Dan geldt ook $A \perp\!\!\!\perp B^c$, $A^c \perp\!\!\!\perp B$ en $A^c \perp\!\!\!\perp B^c$.
Je mag dit zelf bewijzen, in opgave 1.

3.6 Partities.
We komen nog even terug op het kaartenvoorbeeld uit paragraaf 3.3. We trokken reeds de conclusie, dat de gebeurtenis [Boer] geen invloed heeft op de gebeurtenis [♣].
Maar dat geldt natuurlijk ook voor elke andere combinatie van rang en kleur. We kunnen, meer algemeen, stellen dat de indeling van het kaartspel in rangen niets te maken heeft met de indeling van het kaartspel in kleuren.
We zullen dat nu wiskundig precies maken. Daarvoor de volgende
Definitie. Een *partitie (indeling, opsplitsing)* α van Ω is een stel disjuncte gebeurtenissen $\alpha = \{A_1, A_2, \ldots\}$ (aftelbaar of eindig veel) zó dat $\bigcup_k A_k = \Omega$.

3.7 Definitie. We noemen twee partities $\alpha = \{A_1, A_2, \ldots\}$ en $\beta = \{B_1, B_2, \ldots\}$ *onafhankelijk* als voor elke gebeurtenis $A \in \alpha$ en elke gebeurtenis $B \in \beta$ geldt dat $A \perp\!\!\!\perp B$. Notatie: $\alpha \perp\!\!\!\perp \beta$.
Voorbeeld. De partities van het kaartspel naar kleur en rang zijn onafhankelijk, want bv. $\mathbb{P}[\heartsuit \text{ Aas}] = \mathbb{P}[\heartsuit] \cdot \mathbb{P}[\text{Aas}]$ $(\frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13})$. Binnen iedere kleur komen evenveel

azen voor, of ook: binnen iedere rang komen evenveel $\heartsuit$ voor.
Ook het Amerikaanse leger is in te delen naar kleur en rang. Zouden dit ook onafhankelijke partities zijn?

Gevolg. Als $\alpha = \{A_1, A_2, \ldots\}$ en $\beta = \{B_1, B_2, \ldots\}$ onafhankelijke partities zijn, en A is vereniging van een stel elementen van α, en B is vereniging van een stel elementen van β, dan is $A \perp\!\!\!\perp B$.
Bewijs dit zelf (opgave 12).

Toepassing. Werp twee dobbelstenen, met aantal ogen X_1 resp. X_2. In de uitkomstenruimte Ω van de 36 punten (i, j) (met $i, j \in \{1, 2, 3, 4, 5, 6\}$) hebben we de onafhankelijke partities $\alpha = \{A_1, \ldots, A_6\}$ en $\beta = \{B_1, \ldots, B_6\}$ met $A_k = [X_1{=}k]$ en $B_k = [X_2{=}k]$. Dan geldt bijvoorbeeld: $\mathbb{P}[X_1 \text{ is even en } X_2 \leq 3] = \mathbb{P}[X_1 \text{ is even}] \cdot \mathbb{P}[X_2 \leq 3]$.
Hier is $[X_1 \text{ is even}] = A_2 \cup A_4 \cup A_6$ de vereniging van een stel elementen van α; evenzo is $[X_2 \leq 3] = B_1 \cup B_2 \cup B_3$.

Opmerking. In paragraaf 3.5 hebben we gezien:
De gebeurtenissen A en B zijn onafhankelijk $\Leftrightarrow$ de partities $\{A, A^c\}$ en $\{B, B^c\}$ zijn onafhankelijk.

3.8 Bij iedere discrete stochast hoort een partitie van Ω. Zij nl. X een discrete stochast met waardenverzameling $\{x_1, x_2, \ldots\}$. Dan vormen de gebeurtenissen $A_k = [X{=}x_k]$ een partitie $\alpha(X) = \{A_1, A_2, \ldots\}$ van Ω. $\alpha(X)$ heet de *partitie voortgebracht door* X.
X is te schrijven als $X = \sum_k x_k \mathbf{1}_{A_k}$ (zie paragraaf 2.14).

Voorbeeld. Neem X_1 en X_2 als in paragraaf 3.7, en zij $Y = \max\{X_1, X_2\}$. Teken Ω als een 6×6 rooster, en teken daarin de elementen van $\alpha(X_1)$ en $\alpha(Y)$.

3.9 Onafhankelijke stochasten.

Definitie. We noemen de discrete stochasten X en Y *onafhankelijk* als de partities $\alpha(X)$ en $\alpha(Y)$ onafhankelijk zijn. Notatie: $X \perp\!\!\!\perp Y$.
Als V de waardenverzameling van X is en W de waardenverzameling van Y, dan is onafhankelijkheid van X en Y, volgens de definities, equivalent met

$$\mathbb{P}[X{=}x \text{ en } Y{=}y] = \mathbb{P}[X{=}x] \cdot \mathbb{P}[Y{=}y]$$

voor alle $x \in V$ en alle $y \in W$.

Voorbeeld: X_1, X_2 en Y zijn als in de paragrafen 3.7 en 3.8. De stochasten X_1 en X_2 zijn onafhankelijk, want voor elke keuze van $x, y \in \{1, \ldots, 6\}$ is

$$\mathbb{P}[X_1{=}x \text{ en } X_2{=}y] = \tfrac{1}{36}$$

en

$$\mathbb{P}[X_1{=}x] = \mathbb{P}[X_2{=}y] = \tfrac{1}{6}.$$

De stochasten X_1 en Y zijn niet onafhankelijk, want bv. $\mathbb{P}[X_1{=}6, Y{=}4] = 0$ terwijl $\mathbb{P}[X_1{=}6] \cdot \mathbb{P}[Y{=}4] \neq 0$.
Laat zelf zien dat ook Y en $Z = |X_1 - X_2|$ afhankelijk zijn.
Opmerking. Als X en Y onafhankelijke stochasten zijn, dan zijn ook $f(X)$ en $g(Y)$ onafhankelijke stochasten (voor alle functies f en g).
Opmerking. Als A en B gebeurtenissen zijn, dan komt onafhankelijkheid van de stochasten $\mathbf{1}_A$ en $\mathbf{1}_B$ precies neer op onafhankelijkheid van de gebeurtenissen A en B (vergelijk het stellinkje in 3.5).

3.10 Meerdere onafhankelijke gebeurtenissen.

Definitie. De gebeurtenissen $A_1, \ldots, A_n$ noemen we *onafhankelijk* als voor elke greep $i_1, \ldots, i_k$ uit $\{1, \ldots, n\}$ geldt:

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Voorbeeld: A, B en C zijn, per definitie, onafhankelijke gebeurtenissen als $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$ en $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.
Het is dus in het algemeen *niet* voldoende als alleen aan $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$ is voldaan, hetgeen je misschien vooronderstelde; opgave 2 laat hiervan een voorbeeld zien.
Als de vermenigvuldigingsregel in de definitie wèl geldt voor paren, maar eventueel niet voor grotere grepen, dan noemt men de gebeurtenissen $A_1, \ldots, A_n$ *paarsgewijs onafhankelijk*.
Dat dit echt zwakker is dan "gewoon" onafhankelijk, kun je ervaren in opgave 3.

3.11 Meerdere onafhankelijke partities en stochasten.

Voorbeeld. We doen n worpen met een dobbelsteen. De uitkomstenruimte Ω bestaat uit 6^n rijtjes $\omega = (\omega_1, \ldots, \omega_n)$. X_k is het aantal ogen van de k^{e} worp, en dus gedefinieerd door $X_k(\omega) = \omega_k$.
Dan is $\alpha(X_k) = \{A_1^{(k)}, \ldots, A_6^{(k)}\}$, waarin bv. $A_2^{(k)} = [X_k{=}2]$ de verzameling is van alle 6^{n-1} rijtjes $(\omega_1, \ldots, \omega_n)$ met $\omega_k = 2$.
We hebben dan n partities, die op grond van de volgende definitie onafhankelijk zijn.

Definitie. De partities $\alpha^{(1)}, \dots, \alpha^{(n)}$ zijn *onafhankelijk* als voor alle gebeurtenissen $A_1, \dots, A_n$ met $A_1 \in \alpha^{(1)}$, $A_2 \in \alpha^{(2)}, \dots, A_n \in \alpha^{(n)}$ geldt:

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n).$$

Ga na dat de definitie van onafhankelijkheid voor twee partities in 3.7 een speciaal geval is van de definitie hierboven.
Op dezelfde manier als in 3.9 definiëren we nu onafhankelijkheid van meerdere stochasten.
Definitie. De discrete stochasten $X_1, \dots, X_n$ zijn *onafhankelijk* als de partities $\alpha(X_1), \dots, \alpha(X_n)$ onafhankelijk zijn.
Is W_1 de waardenverzameling van X_1, W_2 de waardenverzameling van $X_2, \dots$, en W_n de waardenverzameling van X_n, dan is onafhankelijkheid van $X_1, \dots, X_n$ weer equivalent met:

$$\mathbb{P}[X_1{=}w_1 \text{ en } \dots \text{ en } X_n{=}w_n] = \mathbb{P}[X_1{=}w_1] \cdots \mathbb{P}[X_n{=}w_n]$$

voor alle $w_1 \in W_1$, alle $w_2 \in W_2, \dots$, en alle $w_n \in W_n$.
Ook de definitie van onafhankelijkheid van meerdere stochasten is weer een uitbreiding van de definitie van onafhankelijkheid van twee stochasten.
Opmerking. Er geldt:

De stochasten $X_1, \dots, X_n$ zijn onafhankelijk

$$\Leftrightarrow$$

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

voor iedere keuze van gebeurtenissen $A_1, \dots, A_n$ zó, dat A_1 alleen afhangt van X_1 (bv. $A_1 = [X_1 \le -5]$), A_2 alleen van X_2, enzovoorts.

3.12 Soms kijken we naar nóg grotere collecties gebeurtenissen, partities en stochasten.
Aftelbaar veel gebeurtenissen $A_1, A_2, \dots$ noemen we *onafhankelijk* als elk eindig stel gebeurtenissen eruit onafhankelijk is.
Net zo voor partities en stochasten.

3.13 Voor de volledigheid vermelden we nog enige eigenschappen van onafhankelijkheid.

(a) Een deelcollectie van een stel onafhankelijke stochasten (of partities, of gebeurtenissen) is ook weer onafhankelijk.

(b) Als $X_1, \ldots, X_n$ onafhankelijke stochasten zijn, en $Y = f(X_1, \ldots, X_k), Z = g(X_{k+1}, \ldots, X_n)$, dan $Y \perp\!\!\!\perp Z$.

De binomiale verdeling

3.14 Rijen onafhankelijke experimenten. We hebben nu voldoende gereedschap m.b.t. onafhankelijkheid verzameld om een wiskundig model af te kunnen leiden voor rijen onafhankelijke experimenten.
Wanneer men bij een aantal toevalsexperimenten, die na of naast elkaar gedaan worden, mag aannemen dat er geen verband bestaat tussen de uitkomsten van verschillende experimenten, en men heeft een stel stochasten $X_1, X_2, \ldots$ zó, dat X_k slechts bepaald wordt door de uitkomst van het k^{e} experiment, dan mag men aannemen dat $X_1, X_2, \ldots$ onafhankelijke stochasten zijn. We spreken dan ook van onafhankelijke experimenten.
Als Ω_k de uitkomstenruimte van het k^{e} experiment is, dan kunnen we de hele (aftelbare) rij experimenten opvatten als één groot experiment met uitkomstenruimte $\Omega_1 \times \Omega_2 \times \ldots$, waarvan de elementen rijen $\omega = (\omega_1, \omega_2, \ldots)$ zijn. Hierin is ω_k de uitkomst van het k^{e} experiment. Zie het voorbeeld in paragraaf 3.11, waar $\Omega_1 = \Omega_2 = \ldots = \Omega_n = \{1, 2, 3, 4, 5, 6\}$.

3.15 De binomiale verdeling.
We hebben n onafhankelijke experimenten. Bij ieder experiment kan een zeker resultaat optreden, dat we een succes zullen noemen, en wel steeds met dezelfde kans p. Met kans q $(= 1-p)$ treedt géén succes ("mislukking") op. We willen de kansverdeling bepalen van X, het aantal successen. X neemt waarden aan in $\{0, 1, \ldots, n\}$.
Voorbeelden:

1. X is het aantal zessen in n worpen met een dobbelsteen.

2. X is het aantal kop in n worpen met een munt.

3. De deelnemers aan de voetbaltoto voorspellen van 12 wedstrijden, club A tegen club B, de uitslag: A wint, B wint, of gelijkspel. Vult men het formulier in m.b.v een dobbelsteen (1 of 2 ogen: A wint; 3 of 4 ogen: B wint; 5 of 6 ogen: gelijkspel), dan is voor iedere wedstrijd de kans op een juiste voorspelling $\frac{1}{3}$. Het gaat nu om X, het aantal juiste voorspellingen. $X = 12$ betekent de hoofdprijs.

Laat X_k de stochast zijn die de waarde 1 of 0 aanneemt al naargelang het k^e experiment een succes oplevert of niet. X_k is dus de indicator van de gebeurtenis $A_k =$[het k^e experiment levert een succes op]. Dan is $X_1 + \ldots + X_n$ het aantal successen. De stochasten $X_1, \ldots, X_n$ hebben de volgende eigenschappen:
(a) $X_1, \ldots, X_n$ zijn onafhankelijk;
(b) $X_k \sim \mathrm{Alt}(p)$ voor $k = 1, \ldots, n$.
We leiden nu af dat $X = X_1 + \ldots + X_n$ binomiaal verdeeld is met parameters n en p (notatie: $X \sim \mathrm{Bin}(n, p)$):

$$\mathbb{P}[X{=}k] = \binom{n}{k} p^k q^{n-k}$$

voor $k = 0, 1, \ldots, n$. Het bewijs is in wezen hetzelfde als in paragraaf 1.10.
$[X{=}k]$ treedt op bv. als de eerste k experimenten successen opleveren en de overige $n - k$ mislukkingen. De kans hierop is wegens de onafhankelijkheid $p^k q^{n-k}$. Maar ook een willekeurige andere k-greep uit de n experimenten kan voor de k successen zorgen; en er zijn $\binom{n}{k}$ van die grepen.
We hebben zo de volgende stelling bewezen.
Stelling. Als $X_1, \ldots, X_n$ onafhankelijk zijn, $X_k \sim \mathrm{Alt}(p)$ voor elke k, en $X = X_1 + \ldots + X_n$, dan is $X \sim \mathrm{Bin}(n, p)$.

Opgaven

1. Bewijs het stellinkje in paragraaf 3.5.

2. Voor onafhankelijkheid van de gebeurtenissen A, B en C is in het algemeen niet voldoende dat $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$. Laat dit zien aan de hand van drie muntworpen, waarin $A =$[eerste worp geeft kop], $B =$[meer keren kop dan munt], $C =$[de laatste twee worpen geven dezelfde uitkomst].

3. Het licht op de overloop op de eerste verdieping kan in veel huizen op twee plaatsen bediend worden: zowel beneden als op de overloop zelf is een lichtschakelaar aangebracht. Beide lichtschakelaars kunnen in twee standen staan: "naar boven" ("0") en "naar beneden" ("1") gericht. Doordat er twee schakelaars zijn, valt uit de stand van één schakelaar niet af te leiden of het licht op de overloop brandt of niet.
   Laat X_1 de stand van de schakelaar beneden zijn, en X_2 de stand van de schakelaar op de overloop zelf. We nemen aan dat het licht op de overloop brandt, precies wanneer beide schakelaars in dezelfde stand staan.
   Zij $A = [X_1{=}1]$, $B = [X_2{=}1]$, $C =$ [Het licht op de overloop brandt]. Ga na dat A, B en C wel paarsgewijs onafhankelijk zijn, maar niet "echt" onafhankelijk.

4. Een test bestaat uit 10 ja-nee vragen. Iemand die van toeten noch blazen weet, besluit de vragen op goed geluk te beantwoorden. We geven hem bij iedere vraag een kans $\frac{1}{2}$ op een goed antwoord. Bepaal de kans op 8 of meer goede antwoorden.

5. Men werpt n munten op. Hoe groot is de kans dat het aantal kop even is? (Het antwoord is een eenvoudig rationaal getal)

6. Bepaal de kans op 11 of 12 goed bij de voetbaltoto.

7. Rijexamen. Veronderstel dat bij elk examen de kans om te slagen p is $(0 < p < 1)$ en dat de resultaten van opvolgende examens onderling onafhankelijk zijn. Laat Y het aantal examens zijn dat iemand nodig heeft om te slagen. Bepaal de kansverdeling van Y, en $\mathbb{P}[Y < \infty]$. Y noemt men *geometrisch verdeeld.*

8. In een rij onafhankelijke proeven met kans p op succes zij Y_n het aantal proeven dat men moet doen om n successen te behalen. Bereken $\mathbb{P}[Y_n{=}k]$.

9. Laat in de vorige opgave X_1 het aantal proeven tot en met het eerste succes zijn, X_2 het aantal proeven dat na het eerste succes nodig is om het tweede succes te bereiken, enz., zodat $Y_n = X_1 + \ldots + X_n$. Bewijs dat X_1, X_2 en X_3 onafhankelijk zijn en alle drie dezelfde kansverdeling hebben.

10. Laat $X_0, X_1, \ldots, X_n$ onafhankelijke stochasten zijn, met $X_k \sim \text{Alt}(p)$ voor elke $k \in \{0, 1, \ldots, n\}$. Hierin is $p \in [0, 1]$. Definieer $Y_k = |X_k - X_{k-1}|$ voor $k \in \{1, \ldots, n\}$. Voor welke waarden van p zijn $Y_1, \ldots, Y_n$ onafhankelijk?

11. Bewijs: Als $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ en X en Y onafhankelijk, dan $X + Y \sim \text{Bin}(n + m, p)$.

12. Bewijs het "gevolg" in paragraaf 3.7

Hoofdstuk 4

Verwachtingswaarde

4.1 Huygens begint zijn verhandeling met het volgende principe:
1^{e} Voorstel: Als ick gelijcke kans hebbe om a of b te hebben, dit is my so veel weerdt als $\frac{a+b}{2}$.
By exempel, so yemandt sonder mijn weeten in d'eene handt 3 schellingen verbergt, en in d'andere 7 schellingen, ende my te kiesen geeft welck van beyde ick begeere te hebben, ick segge dit my even soo veel weerdt te zijn als of ick 5 schellingen seecker hadde.

4.2 Voorbeelden.

(a) Bij een zeker kaartspel wint men f 1,– met kans $\frac{1}{2}$, f 2,50 met kans $\frac{1}{5}$, f 100,– met kans $\frac{1}{100}$. Men moet x gulden betalen om mee te mogen doen.
Bij welke waarden van x is het voordelig om mee te doen? Bij welke waarden van x wordt er (op den duur althans) winst gemaakt door degene die het spel organiseert?

(b) In n worpen met een scheve dobbelsteen hebben we n_i keer het resultaat i ogen gekregen (i =1,2,3,4,5,6). Ons gemiddelde is dan: het totaal gedeeld door n, dat is

$$\frac{1}{n}(n_1 \times 1 + n_2 \times 2 + \ldots + n_6 \times 6) =$$
$$1 \cdot \frac{n_1}{n} + 2 \cdot \frac{n_2}{n} + \ldots + 6 \cdot \frac{n_6}{n}.$$

Als n groot is, dan verwachten we op grond van de Experimentele wet van de grote aantallen (zie paragraaf 2.20) dat het frequentie-quotiënt $\frac{n_i}{n}$ dicht bij p_i, de kans op i ogen, zal liggen. Dan zal het gemiddelde ook dicht bij $\sum_{i=1}^{6} ip_i$ liggen. Deze laatste uitdrukking is een soort ideaal gemiddelde: de verwachtingswaarde. Gewone gemiddelden, bepaald met eindig veel worpen, zijn benaderingen (die nog van het toeval afhangen) voor dit ideale gemiddelde.

Definitie en eigenschappen

4.3 Definitie. Laat X een discrete stochast zijn met waardenverzameling W. De *verwachtingswaarde* $\mathbb{E}(X)$ van X definiëren we als

$$\boxed{\mathbb{E}(X) = \sum_{w \in W} w \mathbb{P}[X{=}w].}$$

Voorbeeld. Als X het aantal ogen bij een worp met een eerlijke dobbelsteen is, dan is $\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3\frac{1}{2}$.
Merk op: blijkbaar hoeft $\mathbb{E}(X)$ niet tot de waardenverzameling van X te behoren.

4.4 Niet voor elke X bestaat $\mathbb{E}(X)$.
Als de waardenverzameling van X eindig is, levert de definitie in paragraaf 4.3 geen problemen op. Bij oneindige waardenverzamelingen staat er echter een oneindige som van reële getallen, en je hebt geleerd voorzichtig te zijn met oneindige sommen. De oneindige sommen die in voorgaande hoofdstukken voorkomen, leverden nog geen problemen op, omdat het daar telkens gaat om het optellen van een aantal van de positieve getallen $p_1, p_2, \ldots$, waarvan we op voorhand weten dat de totale som 1 is. Nu echter kunnen er wèl moeilijkheden ontstaan: de totale som van de termen in de definitie van $\mathbb{E}(X)$ kan wel eens ∞ zijn; ook kunnen er negatieve waarden in voorkomen.

(a) Neem eerst het geval dat $X \geq 0$ (dus alle waarden $x_1, x_2, \ldots$ zijn ≥ 0). Dan laten we toe dat $\mathbb{E}(X) = \infty$.
Voorbeeld. Werp zolang een munt op tot voor het eerst kop verschijnt. Als de eerste kop bij de k^{e} worp verschijnt, dan is de uitbetaling X gelijk aan 2^k gulden, en het spel is afgelopen.
Daar $\mathbb{P}[X{=}2^k] = (\frac{1}{2})^k$ (ga na), volgt: $\mathbb{E}(X) = 1 + 1 + 1 + \ldots = \infty$.

(b) Analoog aan (a). Als $X \leq 0$, dan laten we toe dat $\mathbb{E}(X) = -\infty$.

(c) * Moeilijkheden kunnen ontstaan als sommige van de waarden positief zijn, en andere negatief. Noem de mogelijke positieve waarden van X even $x_1, x_2, \ldots,$ en de mogelijke negatieve waarden van X: $y_1, y_2, \ldots$. Bereken nu eerst $a = \sum_k x_k \mathbb{P}[X{=}x_k]$ en $b = \sum_l y_l \mathbb{P}[X{=}y_l]$. Dan is $a > 0$ en $b < 0$. We onderscheiden nu vier gevallen:

1. Als $a < \infty$ en $b > -\infty$, is er geen probleem: dan is $\mathbb{E}(X) = a + b$.

2. Als $a = \infty$ en $b > -\infty$, dan vinden we volgens de definitie: $\mathbb{E}(X) = \infty$.
3. Als $a < \infty$ en $b = -\infty$, dan is analogiter: $\mathbb{E}(X) = -\infty$.
4. Maar als $a = \infty$ en $b = -\infty$, dan is $\mathbb{E}(X)$ niet eenduidig uit te rekenen. We zeggen dan dat $\mathbb{E}(X)$ niet gedefinieerd is.

   Voorbeeld. een dobbelsteen heeft twee zijden met 1 oog, twee zijden met 2 ogen en twee zijden met 3 ogen. We werpen tot voor het eerst een 2 of een 3 verschijnt. Als dat gebeurt bij de k^{e} worp, dan ontvangen we ingeval we 3 ogen gegooid hebben, 3^k gulden; maar als we 2 ogen gegooid hebben, moeten we 3^k gulden betalen. Ga na dat de verwachtingswaarde van onze winst X niet gedefinieerd is.

(d) Tot slot: in verreweg de meeste gevallen die wij zullen tegenkomen, zullen er geen problemen ontstaan, omdat vaak de waardenverzameling eindig is en/of de betreffende stochast positief. We spreken af: als we in dit hoofdstuk over $\mathbb{E}(X)$ spreken, dan nemen we daarbij stilzwijgend aan dat $\mathbb{E}(X)$ bestaat (dus dat (a) of (b) of (c)1 of (c)2 of (c)3 geldt).

4.5 Eenvoudige eigenschappen van $\mathbb{E}$.

(a) Als X constant is, $X = c$ (d.w.z. $X(\omega) = c$ voor alle $\omega \in \Omega$), dan is $\mathbb{E}(X) = c$;

(b) $\mathbb{E}(aX) = a \cdot \mathbb{E}(X)$ (voor alle $a \in \mathbb{R}$);

(c) Als $X \geq 0$, dan $\mathbb{E}(X) \geq 0$;

(d) Als $X \geq Y$ (d.w.z. $X(\omega) \geq Y(\omega)$ voor alle $\omega \in \Omega$), dan $\mathbb{E}(X) \geq \mathbb{E}(Y)$;

(e) $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

Bewijs deze eigenschappen zelf (opgave 15).

4.6 Laat $\alpha = \{A_1, A_2, A_3\}$ een partitie zijn en laat X op A_1 constant 5 zijn, op A_2 constant 7 en op A_3 weer constant 5; dus $X = 5 \cdot \mathbf{1}_{A_1} + 7 \cdot \mathbf{1}_{A_2} + 5 \cdot \mathbf{1}_{A_3}$. Dan is volgens de definitie:

$$\begin{aligned}
\mathbb{E}(X) &= 5\mathbb{P}[X{=}5] + 7\mathbb{P}[X{=}7] = \\
&= 5\mathbb{P}(A_1 \cup A_3) + 7\mathbb{P}(A_2) = \\
&= 5\mathbb{P}(A_1) + 7\mathbb{P}(A_2) + 5\mathbb{P}(A_3).
\end{aligned}$$

Dit geldt algemeen: als $\alpha = \{A_1, A_2, \ldots\}$ een partitie is, en $X = \sum_i a_i \mathbf{1}_{A_i}$, dan is uiteraard

$$\mathbb{E}(X) = \sum_i a_i \cdot \mathbb{P}(A_i).$$

In het bijzondere geval dat de getallen a_i allemaal verschillend zijn, is bovenstaande formule slechts een herformulering van de definitie van $\mathbb{E}(X)$.

4.7 Als $f : \mathbb{R} \to \mathbb{R}$ een willekeurige functie is, $\alpha = \{A_1, A_2, \ldots\}$ een partitie is, en $X = \sum_i a_i \cdot \mathbf{1}_{A_i}$, dan is $f(X) = \sum_i f(a_i) \cdot \mathbf{1}_{A_i}$, en dus volgens paragraaf 4.6: $\mathbb{E}(f(X)) = \sum_i f(a_i) \cdot \mathbb{P}(A_i)$.
In het bijzonder, met $\alpha = \alpha(X)$:

$$\boxed{\begin{aligned} \mathbb{E}(f(X)) &= \sum_i f(x_i) \cdot \mathbb{P}[X = x_i] \\ &= \sum_{w \in W} f(w) \mathbb{P}[X = w] \end{aligned}}$$

waarin W de waardenverzameling van X is.
Voorbeelden:
$\mathbb{E}(X^2) = \sum_{w \in W} w^2 \cdot \mathbb{P}[X{=}w]$
$\mathbb{E}(|X|) = \sum_{w \in W} |w| \cdot \mathbb{P}[X{=}w]$
Als X het aantal ogen van een zuivere dobbelsteen is, dan is $\mathbb{E}(X^2) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = 15\frac{1}{6}$. Merk op dat $\mathbb{E}(X^2) \neq (\mathbb{E}(X))^2$. Algemener: $\mathbb{E}(f(X))$ en $f(\mathbb{E}(X))$ hoeven niet hetzelfde te zijn. In hoofdstuk 6 zullen we zien dat wèl altijd geldt: $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$.

4.8 Laat Z het totale aantal ogen zijn als 2 dobbelstenen gegooid worden. Met behulp van de kansverdeling van Z: $\mathbb{P}[Z{=}2] = \frac{1}{36}$, $\mathbb{P}[Z{=}3] = \frac{2}{36}$,…,$\mathbb{P}[Z{=}12] = \frac{1}{36}$, kunnen we $\mathbb{E}(Z) = \sum_{k=2}^{12} k\mathbb{P}[Z{=}k]$ berekenen. Er komt uit $\mathbb{E}(Z) = 7$, dus twee keer zoveel als bij één dobbelsteen. Hier geldt dus: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, als X en Y de uitkomsten van de afzonderlijke dobbelstenen zijn. Dit blijkt algemeen te gelden:

$$\boxed{\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)}$$

Bewijs. Zij V de waardenverzameling van X, en W de waardenverzameling van Y. Schrijf $X = \sum_{v \in V} v \cdot \mathbf{1}_{[X=v]}$, $Y = \sum_{w \in W} w \cdot \mathbf{1}_{[Y=w]}$.

Dan is

$$X + Y = \sum_{(v,w)\in V\times W} (v{+}w)\mathbf{1}_{[X=v,Y=w]}.$$

Wegens paragraaf 4.6 is dan (ga na!)

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{(v,w)\in V\times W} (v{+}w)\mathbb{P}[X{=}v, Y{=}w] = \\
&= \sum_{(v,w)\in V\times W} v\mathbb{P}[X{=}v, Y{=}w] + \sum_{(v,w)\in V\times W} w\mathbb{P}[X{=}v, Y{=}w] = \\
&= \sum_{v\in V} v\mathbb{P}[X{=}v] + \sum_{w\in W} w\mathbb{P}[Y{=}w] = \\
&= \mathbb{E}(X) + \mathbb{E}(Y). \quad \square
\end{aligned}$$

Gevolg:

$$\mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n).$$

Bewijs. Met volledige inductie. $\square$

4.9 Toepassingen.

(a) Als $X \sim \text{Alt}(p)$, dan is $\mathbb{E}(X) = 0 \cdot \mathbb{P}[X{=}0] + 1 \cdot \mathbb{P}[X{=}1] = p$.

(b) Als je honderd vragen moet beantwoorden, die je elk met kans $\frac{3}{4}$ goed hebt, dan verwacht je 75 goede antwoorden. Algemeen: Als $X \sim \text{Bin}(n,p)$, dan $\mathbb{E}(X) = np$. Bewijs: zie (c). $\square$

(c) Ook de hypergeometrische verdeling heeft verwachtingswaarde np.

Bewijs van (b) en (c): de binomiale en de hypergeometrische verdeling zijn beide afkomstig uit het vaasmodel (paragraaf 1.10 resp. paragraaf 1.8. Beschouw dus een vaas met N ballen; r rode en w witte ($r + w = N$), laat $p = \frac{r}{N}$, en trek n ballen uit deze vaas (met, resp. zonder terugleggen).

Laat $X_i = \mathbf{1}_{\text{bal } i \text{ is rood}}$, en $X = X_1 + \ldots + X_n$. X is het aantal getrokken rode ballen. X is binomiaal (in het geval "met terugleggen") resp. hypergeometrisch ("zonder terugleggen") verdeeld. Wegens het gevolg in paragraaf 4.8 is in beide gevallen $\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) \stackrel{\text{a}}{=} np$. $\square$

Opmerking. Realiseer je wel dat de X_i's in het hypergeometrische geval *niet* onafhankelijk zijn. Toch mogen de verwachtingen bij elkaar opgeteld worden.

(d) Merk op dat we in (b) en (c) de verwachting van een stochast X hebben uitgerekend zonder daarbij de kansverdeling van X te gebruiken. De gebruikte methode is vaak toepasbaar: als X op te vatten is als het aantal successen in een zeker aantal (n) proeven, schrijf dan $X = X_1 + \ldots + X_n$, waarbij $X_k = \mathbf{1}_{[\text{succes in } k^{\text{e}} \text{ proef}]}$. Wegens paragraaf 4.8 is dan $\mathbb{E}(X) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = \mathbb{P}[X_1{=}1] + \ldots + \mathbb{P}[X_n{=}1]$. Deze laatste kansen zijn dikwijls eenvoudig te berekenen, terwijl X zelf een vreselijk ingewikkelde kansverdeling kan hebben.

Voorbeeld. Lootjes trekken voor Sinterklaas (hoofdstuk 2, opgave 5). Wat is de verwachting van het aantal personen dat zijn/haar eigen lootje trekt? Met $X_i = \mathbf{1}_{A_i}$ is $X_i \sim \mathrm{Alt}(\frac{1}{n})$, dus $\mathbb{E}(X) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = n \cdot \frac{1}{n} = 1$.

Loterijen, verzekeringen en wat dies meer zij: Een moraliserend intermezzo

4.10 De Petersburg-paradox.

Lees nog eens het voorbeeld in paragraaf 4.4(a) door. Hoeveel is het je waard, aan dat spel mee te mogen doen?

Iemand zegt dat het hem 10 gulden waard is. Wij zullen laten zien dat het meer waard is.

Vergelijk het spel in paragraaf 4.4(a) met de volgende variant: na 10 worpen breken we het spel af. Heeft de speler dan nog geen kop gegooid, dan krijgt hij niets.

De verwachte winst van de speler in dit spel is 10 gulden, zoals je gemakkelijk kunt nagaan. Dus als de speler vooraf 10 gulden zou moeten betalen, dan is de verwachte winst 0 gulden. Maar dit spel is voor hem ongunstiger dan het oorspronkelijke spel, immers het scheelt hem een aantal winstmogelijkheden. Dus als hij bij het oorspronkelijke spel 10 gulden zou moeten betalen als inleg, dan is de verwachte uitbetaling nog steeds positief.

Nu hadden we in plaats van 10 gulden ook wel 100.000 gulden kunnen nemen in bovenstaand verhaal. We hadden dan gevonden, dat meedoen met het spel meer dan 100.000 gulden waard is.

Blijkbaar is elk (eindig) bedrag, dat men van de speler als inleg vraagt, acceptabel voor de speler: de verwachte winst blijft voor hem positief. Anderzijds heb je – intuïtief – het gevoel, dat je het spel niet moet spelen als er een heel hoge inzet wordt verlangd. Je loopt dan tenslotte een groot risico om veel geld te verliezen, terwijl je slechts een heel kleine kans hebt om (heel erg) veel geld te winnen.

Deze tegenstrijdigheid staat bekend als de *Petersburg-paradox*.

4.11 Zó win je altijd.

We spelen een wat reëler spel. Elke ronde bestaat uit een muntworp. Vòòr elke

worp bepaalt de speler zijn inzet, zeg, k gulden. Als de worp kop oplevert, ontvangt hij $2k$ gulden van de spelleider (dus wint k gulden); als munt verschijnt, is hij zijn k gulden kwijt. Zo'n spel noemt men eerlijk, omdat in elke ronde de verwachte winst 0 gulden is. Toch is er een strategie om als speler met zekerheid 1 gulden te winnen. Zet voor de eerste worp 1 gulden in. Als kop bovenkomt, stop dan en ga naar huis. Als munt bovenkomt, zet dan voor de tweede worp 2 gulden in. Als bij de tweede worp kop bovenkomt, dan is je winst per saldo 1 gulden, stop dus. Als munt bovenkomt, zet dan voor de derde worp 4 gulden in. Enzovoorts (verdubbel steeds de inzet zolang je verliest).
Als na k ronden de eerste kop verschijnt, dan is de spelerswinst:

$$(-1) + (-2) + (-4) + \ldots + (-2^{k-2}) + 2^{k-1} = 1 \text{ gulden}$$

(ga na).
Dus wànneer ook de kop verschijnt, je winst is altijd 1 gulden. En met zekerheid verschijnt ooit de eerste kop. Het beroerde is alleen, dat de strategie alleen uitvoerbaar is met oneindig veel geld op zak. Is je kapitaal eindig, dan is er een kans dat je alles kwijtraakt.
We maken het nog bonter. We werpen nu niet met een munt, maar met een dobbelsteen; de speler wint alleen als een worp met de dobbelsteen zes ogen oplevert. Met precies dezelfde strategie bereik je dan precies hetzelfde resultaat; alleen duurt het waarschijnlijk wel wat langer voordat je uitgespeeld bent.
Deze strategie is blijkbaar toepasbaar voor elk spel waarin je als speler in elke ronde een kans $p > 0$ hebt om te winnen. Hiermee wapen je je dus tegen geslepen munten en dito tegenstanders.
Maar waarom dan niet meteen naar de Waalkade gesneld en mezelf rijk gespeeld? Wel, er zijn praktische bezwaren. Vaak is er een maximum-inleg in casino's. En verder ontbreekt je als speler het geld om lang genoeg door te gaan als het in het begin flink tegenzit.

4.12 De praktijk.

Verwachtingswaarden kun je eenvoudig toepassen bij zaken als verzekeringen en loterijen. Eigenlijk speel je in beide gevallen een spel: je betaalt premie (resp. inzet), en krijgt met (kleine) kans geld terug. Zowel de verzekeringsmaatschapij als de organisatie van de loterij willen winst maken, dus beide spelen hebben een negatieve verwachtingswaarde voor de deelnemers.
Toch zit er (psychologisch) verschil tussen een verzekering aangaan en meedoen aan een loterij: je tegen bepaalde risico's verzekeren wordt algemeen als zeer verstandig

aangemerkt; het meedoen aan een loterij als geldverspilling (of: geld schenken aan "het goede doel").
Bij een verzekering betaal je namelijk een (relatief) klein bedrag, om je in te dekken tegen een risico, en bij een loterij alleen om een (kleine) kans te maken op een grote winst.
Blijkbaar hangt het niet alleen van de verwachtingswaarde af, of het verstandig is om aan een bepaald spel mee te doen of niet!

Onafhankelijkheid en verwachting

4.13 We behandelen nu enkele eigenschappen van de verwachting die in verband staan met onafhankelijkheid. X en Y zijn stochasten.

(a) $\boxed{\text{Als } X \perp\!\!\!\perp Y, \text{ dan is } \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).}$

Bewijs. Laten V, W de waardenverzamelingen van X resp. Y zijn, en schrijf X en Y weer als

$$X = \sum_{v \in V} v \mathbf{1}_{[X=v]}$$

en

$$Y = \sum_{w \in W} w \mathbf{1}_{[Y=w]}.$$

Dan is

$$XY = \sum_{(v,w) \in V \times W} v \cdot w \mathbf{1}_{[X=v, Y=w]}.$$

Wegens paragraaf 4.6 is dan weer:

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{(v,w) \in V \times W} v \cdot w \mathbb{P}[X=v, Y=w] \\ &= \sum_{(v,w) \in V \times W} v \cdot w \mathbb{P}[X=v] \mathbb{P}[Y=w] \\ &= \left(\sum_{v \in V} v \mathbb{P}[X=v] \right) \left(\sum_{w \in W} w \mathbb{P}[Y=w] \right) \\ &= \mathbb{E}(X) \cdot \mathbb{E}(Y). \quad \square \end{aligned}$$

(b) Als $X \perp\!\!\!\perp Y$, dan geldt voor elk tweetal functies f en g:

$$\mathbb{E}(f(X) \cdot g(Y)) = \mathbb{E}(f(X)) \cdot \mathbb{E}(g(Y)).$$

(c) Als $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, dan volgt daaruit nog niet $X \perp\!\!\!\perp Y$.

Voorbeeld. Gooi 2 keer een munt op, zij $X_i = \mathbf{1}_{[i^e \text{ keer kop}]}$ $(i = 1, 2)$. Dan zijn $X = X_1 + X_2$ en $Y = X_1 - X_2$ niet onafhankelijk, maar wel geldt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Als $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, dan heten X en Y wel *ongecorreleerd*; notatie: $X \perp Y$. Onafhankelijke stochasten (met bestaande verwachting) zijn dus ongecorreleerd, maar ongecorreleerde stochasten hoeven niet onafhankelijk te zijn.

4.14 Voorwaardelijke verwachting.

Voorbeeld. In een worp met een dobbelsteen is X het aantal ogen. Zoals bekend is $\mathbb{E}(X) = 3\frac{1}{2}$. Wat is nu echter de verwachting van X als gegeven is dat X even is? Er zijn dan nog drie mogelijke waarden van X, namelijk 2, 4 en 6, die alle drie optreden met kans $\frac{1}{3}$. Intuïtief zeg je dus: de gevraagde verwachting is $\frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 4 + \frac{1}{3} \cdot 6 = 4$. We noteren dit als volgt: $\mathbb{E}(X|[X \text{ is even}]) = 4$, of, om een teveel aan haken te vermijden: $\mathbb{E}(X|X \text{ is even}) = 4$.

Definitie. Zij A een gebeurtenis met $\mathbb{P}(A) > 0$. Dan definiëren we de *voorwaardelijke verwachting van X gegeven A* door

$$\boxed{\mathbb{E}(X|A) = \sum_{w \in W} w\mathbb{P}([X{=}w]|A)}$$

waar W de waardenverzameling van X is.

Het verschil met de "gewone" verwachting is, dat hier de "gewone" kansen vervangen zijn door voorwaardelijke kansen.

In bovenstaand voorbeeld is $\mathbb{E}(X|X \text{ is even}) = 1 \cdot 0 + 2 \cdot \frac{1}{3} + 3 \cdot 0 + 4 \cdot \frac{1}{3} + 5 \cdot 0 + 6 \cdot \frac{1}{3} = 4$.

Onmiddelijk uit de definitie volgt:

Eigenschap: $\mathbb{E}(\mathbf{1}_B|A) = \mathbb{P}(B|A)$.

In paragraaf 8.2 zagen we dat het bij het berekenen van kansen wel eens handig is, de waarde van een bepaalde stochast te kennen. Dat geldt ook bij het berekenen van verwachtingen. Als X een stochast is met waardenverzameling W, en $\mathbb{E}(Y)$ bestaat, dan is

$$\mathbb{E}(Y) = \sum_{w \in W} \mathbb{P}[X{=}w] \cdot \mathbb{E}(Y|X{=}w).$$

Deze formule is een speciaal geval $(\alpha = \alpha(X))$ van de volgende, waarin α een partitie is:

$$\boxed{\mathbb{E}(Y) = \sum_{A \in \alpha} \mathbb{P}(A)\mathbb{E}(Y|A).}$$

BEWIJS van deze formule:

$$
\begin{aligned}
\sum_{A\in\alpha} \mathbb{P}(A)\mathbb{E}(Y|A) &= \sum_{A\in\alpha} \mathbb{P}(A) \sum_{w\in W} w\cdot \mathbb{P}([Y{=}w]|A) \\
&= \sum_{w\in W} \left(w \cdot \sum_{A\in\alpha} \mathbb{P}(A)\mathbb{P}([Y{=}w]|A) \right) \\
&= \sum_{w\in W} w\mathbb{P}[Y{=}w] \\
&= \mathbb{E}(Y). \quad \square
\end{aligned}
$$

4.15 Stelling. Als X alleen *gehele waarden* ≥ 0 aanneemt, dan is

$$
\boxed{\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}[X{>}k].}
$$

BEWIJS: zij $p_k = \mathbb{P}[X{=}k]$, dan is

$$
\begin{aligned}
\mathbb{E}(X) &= p_1 + 2p_2 + 3p_3 + \ldots \\
&= (p_1 + p_2 + p_3 + \ldots) + (p_2 + p_3 + \ldots) + (p_3 + \ldots) + \ldots \\
&= \mathbb{P}[X{>}0] + \mathbb{P}[X{>}1] + \mathbb{P}[X{>}2] + \ldots. \quad \square
\end{aligned}
$$

Opmerking. In dit bewijs wordt de sommatievolgorde van twee oneindige sommen verwisseld. Dat levert hier geen problemen op, maar dat zullen we niet bewijzen. Vaak is het rekenwerk met deze formule wat eenvoudiger dan wanneer men de definitie van $\mathbb{E}(X)$ gebruikt.

Voorbeeld. Als Y het aantal proeven is dat nodig is om één succes te behalen (onafhankelijke proeven, elke proef kans p op succes) dan is

$$
\mathbb{P}[Y{>}k] = \mathbb{P}[\text{eerste t/m } k^{\text{e}} \text{ proef mislukt}] = q^k
$$

(met $q = 1{-}p$), zodat $\mathbb{E}(Y) = 1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q} = \frac{1}{p}$.

Overigens kan $\mathbb{E}(Y)$ ook berekend worden m.b.v. de formule $\displaystyle\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$ (voor $|x| < 1$).

Opgaven

1. Men gooit 12 dobbelstenen. Wat is de verwachting van X, het aantal zessen, en Y, het totale aantal ogen?

2. Als de machine weigert zijn er twee mogelijkheden: onderdeel A is kapot (kans 0,3), of onderdeel B is kapot (kans 0,7). De kans dat beide kapot zijn, stellen we 0. Het nakijken en eventueel vervangen van A duurt 10 minuten; voor B is dit 20 minuten.

   (a) Men begint te kijken bij A. Is A in orde, dan gaat men verder met B. Is A kapot, dan hoeft men niet naar B te kijken. Bepaal de verwachte tijdsduur van de werkzaamheden.

   (b) Zelfde vraag, als men bij B begint.

3. In een serie onafhankelijke spelen hebben A en B bij ieder spel kans $\frac{1}{2}$ om te winnen. Iedere winstpartij levert één punt op. Bepaal de verwachting van het aantal spelen als:

   (a) er wordt gespeeld tot A of B drie punten behaald heeft;

   (b) er wordt gespeeld tot iemand een voorsprong van 2 punten op zijn tegenstander heeft.

4. Men neemt 13 kaarten uit een goed geschud kaartspel. Laat X het aantal azen, Y het aantal ♠, en Z het aantal kwartetten zijn (een kwartet is een viertal kaarten van gelijke rang).

   (a) Bepaal $\mathbb{E}(X)$ en $\mathbb{E}(Y)$.

   (b) Bepaal $\mathbb{E}(Z)$. Aanwijzing. Schrijf $Z = Z_1 + \ldots + Z_{13}$, waar bv. $Z_6 = \mathbf{1}_{[\text{zessen vormen een kwartet}]}$.

5. Vervolg op hoofdstuk 3, opgave 10. Zij $Y = Y_1 + \ldots + Y_n$. Bereken $\mathbb{E}(Y)$.

6. Men krijgt 13 kaarten uit een goed geschud spel. Nadat genoteerd is, welke kaarten dit waren, worden de kaarten teruggegeven, geschud en opnieuw krijgt men 13 kaarten. Bij deze kaarten blijken er X te zijn die ook in de eerste hand voorkwamen.

   Ook deze kaarten worden genoteerd en teruggegeven. Na het schudden krijgt men voor de derde keer een hand van 13 kaarten. In deze hand komen Y kaarten voor die ook in de beide voorgaande handen reeds voorkwamen (dus zeker: $Y \leq X$). Bepaal:

(a) $\mathbb{P}[X \geq 1]$;

(b) $\mathbb{E}(X)$;

(c) $\mathbb{E}(Y)$.

7. Werp zesmaal een dobbelsteen. X is het aantal kanten dat nooit bovenkomt. Bereken $\mathbb{E}(X)$.

8. Bereken de verwachting van het aantal gepakte paren schoenen in opgave 11 van hoofdstuk 1.

9. Bereken $\mathbb{E}(X)$ in opgave 11 van hoofdstuk 2.

10. Y_n is het aantal proeven dat men moet doen om n successen te behalen (we hebben te maken met onafhankelijke proeven met vaste succeskans p). Bepaal $\mathbb{E}(Y_n)$. Aanwijzing: Hoofdstuk 3, opgave 9.

11. Het recordprobleem.
Er wordt een onbeperkte rij proeven gedaan, die resultaten $X_1, X_2, \ldots$ geven. Als er al n proeven gedaan zijn, dan kunnen we de resultaten naar grootte ordenen. We nemen aan dat daarbij alle mogelijke rangschikkingen even waarschijnlijk zijn, en dat precies gelijke resultaten met kans 0 optreden.
We noemen de n^{e} proef een record als X_n al zijn voorgangers overtreft. Deze gebeurtenis noemen we A_n.
Laat R_n het aantal records zijn onder de eerste n proeven. Laat T_2 het nummer van de proef zijn waarbij het 2^{e} record optreedt. Toon aan:

    (a) $\mathbb{E}(R_n) = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$;

    (b) $\mathbb{P}[T_2 > k] = \frac{1}{k}$ voor $k \geq 1$, en $\mathbb{E}(T_2) = \infty$.

12. X is het aantal worpen nodig om alle kanten van de dobbelsteen minstens één keer boven te krijgen. Bepaal $\mathbb{E}(X)$ (Antwoord: $\mathbb{E}(X) = 14{,}7$).

13. Werp zolang een dobbelsteen tot voor het eerst twee keer achter elkaar hetzelfde aantal ogen optreedt. Dit gebeurt bij de X^{e} worp. Bepaal $\mathbb{E}(X)$.

14. In een reeks onafhankelijke proeven, elk met succeskans p $(0 < p < 1)$ zij Y het aantal proeven nodig om één succes te behalen. In paragraaf 4.15 hebben we $\mathbb{E}(Y)$ berekend; een tweede manier daarvoor is de volgende. Zij $A = [1^{\mathrm{e}}$ proef succes$]$.

    (a) Noem even $\mathbb{E}(Y)$: μ. Beredeneer (géén bewijs) dat $\mathbb{E}(Y|A^c) = 1 + \mu$.

(b) Bereken nu $\mathbb{E}(Y)$ m.b.v. de laatste formule uit paragraaf 4.14.

15. Bewijs de eigenschappen van de verwachtingswaarde in paragraaf 4.5.

Hoofdstuk 5

Spreiding

5.1

(a) We gooien met pijltjes naar een prikbord dat in smalle verticale stroken verdeeld is die genummerd zijn: $\ldots, -2, -1, 0, 1, 2, \ldots$ De bedoeling is de 0-strook te raken. De persoon die aan het gooien is, heeft kansen

$$\ldots, p_{-1}, p_0, p_1, \ldots$$

om de verschillende stroken te raken (verschillende personen hebben in principe verschillende kansverdelingen). Zij X het nummer van de geraakte strook. Laten we aannemen dat afwijkingen naar links ongeveer even vaak voorkomen als overeenkomstige afwijkingen naar rechts, dus dat $\mathbb{E}(X) = \sum_i ip_i = 0$ is. Bij iemand die het goed kan, zal de spreiding klein zijn: de stroken dicht bij 0 nemen bijna alle kans op. Bij een slordige gooier zal ook nog flink wat kans verder weg liggen. Mogelijke maten voor die spreiding zijn bv. $\sum_i |i|p_i$ en $\sum_i i^2 p_1$.

(b) We gooien n munten en krijgen daarbij X_n keer kop. Het verwachte aantal kop $\mathbb{E}(X_n)$ is $\frac{1}{2}n$. We doen dit experiment enkele keren. Is n klein, laten we zeggen 10, dan zullen afwijkingen van het verwachte aantal kop van 20 % of meer ($|X_{10} - 5| \geq 2$) geregeld voorkomen (de kans hierop is ongeveer 0,34). Nemen we n wat groter, bv. 40, dan worden afwijkingen van 20 % of meer ($|X_{40} - 20| \geq 8$) tamelijk zeldzaam (de kans hierop is $\approx$ 0,016). Met toenemende n wordt de spreiding van $\frac{X_n}{n}$ kennelijk kleiner.

Variantie

5.2 Definitie. Als maat voor de spreiding van een stochast X, die verwachting $\mathbb{E}(X)$ heeft, voeren we in de *variantie*, afgekort $\mathbb{V}\mathrm{ar}$:

$$\boxed{\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).}$$

Vaak noteren we $\mathbb{E}(X)$ als μ, zodat dan $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}((X-\mu)^2)$.
Als X een *discrete* stochast is met waardenverzameling W en verwachting μ, dan wordt de variantie, m.b.v. paragraaf 4.7:

$$\boxed{\mathbb{V}\mathrm{ar}(X) = \sum_{w \in W} (w-\mu)^2 \cdot \mathbb{P}[X{=}w].}$$

Voorbeelden:

1. Voor het aantal ogen X bij een worp met een dobbelsteen vinden we:

   $$\mathbb{V}\mathrm{ar}(X) = \sum_{i=1}^{6} (i - 3\tfrac{1}{2})^2 \cdot \tfrac{1}{6} = \tfrac{35}{12}.$$

2. Laat X de waarden $+a$ en $-a$ ieder met kans $\frac{1}{2}$ aannemen. Dan is $\mathbb{V}\mathrm{ar}(X) = a^2$.

3. Laat X de waarden $+a$ en $-a$ ieder met kans $\frac{1}{4}$ aannemen, en de waarde 0 met kans $\frac{1}{2}$. Dan is $\mathbb{V}\mathrm{ar}(X) = \frac{1}{2}a^2$.

5.3 Bestaat $\mathbb{V}\mathrm{ar}(X)$ altijd?

Nee: als $\mathbb{E}(X)$ niet bestaat, bestaat $\mathbb{V}\mathrm{ar}(X)$ natuurlijk ook niet.
Als $\mathbb{E}(X) = -\infty$ of $\mathbb{E}(X) = \infty$, is het weinig zinvol om naar de verwachting van de "stochast" $(X - \mathbb{E}(X))^2$ te kijken. Ook voor zulke X definiëren we $\mathbb{V}\mathrm{ar}(X)$ niet.
Als echter $|\mathbb{E}(X)| < \infty$, dan bestaat $\mathbb{V}\mathrm{ar}(X)$ en kan zowel eindig als oneindig zijn.
We maken een soortgelijke afspraak als in hoofdstuk 5: als we in het vervolg ergens $\mathbb{V}\mathrm{ar}(X)$ (of $\mathbb{E}(X)$) schrijven, dan nemen we daarbij stilzwijgend aan dat $\mathbb{V}\mathrm{ar}(X)$ (resp $\mathbb{E}(X)$) bestaat en eindig is.

5.4 Eigenschappen van $\mathbb{V}\mathrm{ar}$.

(a)	$\mathbb{V}\mathrm{ar}(X) \geq 0$;
(b)	Als X constant is, $X = c$, dan $\mathbb{V}\mathrm{ar}(X) = 0$, en omgekeerd;
(c)	$\mathbb{V}\mathrm{ar}(aX) = a^2\mathbb{V}\mathrm{ar}(X)$, voor elke $a \in \mathbb{R}$;
(d)	$\mathbb{V}\mathrm{ar}(X + a) = \mathbb{V}\mathrm{ar}(X)$, voor elke $a \in \mathbb{R}$;
(e)	$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$;
(f)	Als $X \perp\!\!\!\perp Y$, dan $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$;
(g)	Als $X_1, \ldots, X_n$ onafhankelijk zijn, dan $\mathbb{V}\mathrm{ar}(X_1 + \ldots + X_n) = \mathbb{V}\mathrm{ar}(X_1) + \ldots + \mathbb{V}\mathrm{ar}(X_n)$.

Bewijs. (a), (b), (c) en (d) volgen direct uit de definitie door gebruik te maken van bekende eigenschappen van $\mathbb{E}$.
(e). Noem $\mathbb{E}(X) = \mu$, dan is

$$\begin{aligned} \mathbb{V}\mathrm{ar}(X) &= \mathbb{E}\left((X-\mu)^2\right) \\ &= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\ &= \mathbb{E}(X^2) - 2\mu \cdot \mu + \mu^2 \\ &= \mathbb{E}(X^2) - \mu^2. \end{aligned}$$

(f). Noem $\mathbb{E}(X) = \mu$, $\mathbb{E}(Y) = \nu$, dan is $\mathbb{E}(X + Y) = \mu + \nu$, en

$$\begin{aligned} \mathbb{V}\mathrm{ar}(X+Y) &= \mathbb{E}\left(((X+Y) - (\mu+\nu))^2\right) \\ &= \mathbb{E}\left(((X-\mu) + (Y-\nu))^2\right) \\ &= \mathbb{E}\left((X-\mu)^2 + (Y-\nu)^2 + 2(X-\mu)(Y-\nu)\right) \\ &= \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\mathbb{E}\left((X-\mu)(Y-\nu)\right). \end{aligned}$$

Daar X en Y onafhankelijk zijn, is de laatste term gelijk aan $2\mathbb{E}(X-\mu)\mathbb{E}(Y-\nu) = 0$.
(g). Met volledige inductie uit (f). □

5.5 Standaardafwijking.

Definitie. De *standaardafwijking* of *standaarddeviatie* σ_X van een stochast X is

$$\boxed{\sigma_X = \sqrt{\mathbb{V}\mathrm{ar}(X)}.}$$

σ_X is in dezelfde eenheden (bv. kg, cm) als de stochast X (de waarnemingen van X); $\mathbb{V}\mathrm{ar}(X)$ niet.
Uit 5.4(f) volgt direct dat standaardafwijkingen van onafhankelijke stochasten "optellen volgens Pythagoras":

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y.$$

5.6 De 2σ-regel.

Zij X een stochast met verwachting $\mathbb{E}(X) = \mu$ en variantie $\mathbb{V}\mathrm{ar}(X) = \sigma^2$. Een vuistregel (de 2σ-regel) zegt dat meestal meer dan 95 % der waarnemingen van X in het interval

$$(\mu{-}2\sigma, \mu{+}2\sigma)$$

ligt. Anders gezegd: afwijkingen van 2σ (of meer) van de verwachtingswaarde μ komen weinig voor; afwijkingen van 1σ daarentegen zijn heel gewoon (vandaar de uitdrukking "standaard"afwijking). σ is dus een maat voor de breedte van de kansverdeling.

5.7 Variantie van enige bekende kansverdelingen

(a) Als $X \sim \mathrm{Alt}(p)$, dan is $\mathbb{V}\mathrm{ar}(X) = pq$ (waar $q = 1{-}p$). Immers, $X = X^2$, dus $\mathbb{E}(X^2) = p$, zodat $\mathbb{V}\mathrm{ar}(X) = p - p^2 = pq$.

(b) Als $X \sim \mathrm{Bin}(n,p)$, dan is $\mathbb{V}\mathrm{ar}(X) = npq$. We schrijven namelijk $X = X_1 + \ldots + X_n$, met $X_k \sim \mathrm{Alt}(p)$ (elke k), en $X_1, \ldots, X_n$ onafhankelijk.
Op grond van eigenschap 5.4(g) vinden we m.b.v. (a) dat

$$\mathbb{V}\mathrm{ar}(X) = \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(X_k) = npq.$$

(c) Zij X het aantal proeven nodig voor het eerste succes in een reeks onafhankelijke proeven, elk met succeskans p.
We weten al dat $\mathbb{P}[X{=}k] = q^{k-1}p$ ($q = 1{-}p$), en $\mathbb{E}(X) = \frac{1}{p}$. Rechtstreekse berekening van $\mathbb{V}\mathrm{ar}(X)$ uit de definitie wordt bemoeilijkt door analytische problemen (verwisselen van sommatievolgorden e.d.). Het lukt echter wel met voorwaardelijke verwachting (vergelijk opgave 14 in hoofdstuk 5).
Zij weer $A = $ [succes bij 1$^{\mathrm{e}}$ proef]. Dan is $\mathbb{E}(X^2|A^c) = \mathbb{E}\left((1+X)^2\right)$ (intuïtief duidelijk, exact bewijs niet nodig), zodat

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}(X^2|A)\cdot\mathbb{P}(A) + \mathbb{E}(X^2|A^c)\cdot\mathbb{P}(A^c)\\
&= 1\cdot p + \mathbb{E}\Big((1+X)^2\Big)\cdot q\\
&= 1 + 2q\mathbb{E}(X) + q\mathbb{E}(X^2)
\end{aligned}$$

waaruit volgt dat $\mathbb{E}(X^2) = \frac{1+q}{p^2}$ en $\mathbb{V}\mathrm{ar}(X) = \frac{q}{p^2}$.
Zie voor een andere methode opgave 6.

De ongelijkheid van Chebyshev

5.8 De spreiding van een stochast hangt nauw samen met de verdeling van de kansdichtheid rond het gemiddelde. Dit blijkt nog eens uit de volgende ongelijkheid:

> **Ongelijkheid van Chebyshev:**
>
> Voor elke stochast X en elk getal $a > 0$ is
> $$\mathbb{P}[|X - \mathbb{E}(X)| \geq a] \leq \frac{1}{a^2}\mathbb{V}\mathrm{ar}(X).$$

Bewijs. Schrijf $\mu = \mathbb{E}(X)$. De $\omega \in \Omega$ met $|X(\omega) - \mu| \geq a$ dragen aan de variantie van X minstens $a^2 \cdot \mathbb{P}[|X(\omega) - \mu| \geq a]$ bij. Dus:

$$\mathbb{V}\mathrm{ar}(X) \geq a^2 \cdot \mathbb{P}[|X - \mu| \geq a]. \quad \square$$

Opmerking. In veel gevallen is de ongelijkheid van Chebyshev slechts een ruwe afschatting voor de grootte van de kans $\mathbb{P}[|X - \mathbb{E}(X)| \geq a]$. Zo levert de keuze $a = 2\sigma_X$ de afschatting $\frac{1}{4}$ op; maar in paragraaf 5.6 vermeldden we al dat deze kans vaak ongeveer 0,05 is. Er zijn echter ook voorbeelden te bedenken waarin de ongelijkheid van Chebyshev wèl scherp is:

$$X = \begin{cases} -a & \text{met kans } p; \\ 0 & \text{met kans } 1 - 2p; \\ a & \text{met kans } p. \end{cases}$$

Reken zelf na dat in dit voorbeeld geldt

$$\mathbb{P}[|X - \mathbb{E}(X)| \geq a] = \frac{1}{a^2}\mathbb{V}\mathrm{ar}(X).$$

Zonder extra informatie over de verdeling van X kan de ongelijkheid van Chebyshev dus niet verbeterd worden.

5.9 Een toepassing van paragraaf 5.8 is de Zwakke Wet van de grote aantallen, al genoemd in paragraaf 2.20. Hier volgt een speciaal geval (in opgave 5 staat de algemene formulering):

> Zij S_n het aantal successen in n onafhankelijke proeven, elk met succeskans p. Zij P_n het frequentiequotiënt: $P_n = \frac{1}{n}S_n$.
> Dan geldt voor elk getal $\varepsilon > 0$:
> $$\lim_{n\to\infty} \mathbb{P}[|P_n - p| < \varepsilon] = 1.$$

BEWIJS. $S_n \sim \text{Bin}(n, p)$, dus $\mathbb{E}(S_n) = np$ en $\mathbb{V}\text{ar}(S_n) = npq$ (waar $q = 1-p$). We vinden: $\mathbb{E}(P_n) = p$, $\mathbb{V}\text{ar}(P_n) = \frac{1}{n}pq$ en met behulp van de ongelijkheid van Chebyshev:

$$\mathbb{P}[|P_n - p| < \varepsilon] > 1 - \frac{pq}{n\varepsilon^2}$$

en dit gaat naar 1 als $n \to \infty$ $\square$.

Commentaar. Als we n groot genoeg nemen, kunnen we de kans dat P_n meer dan ε van p afwijkt, zo klein krijgen als we willen, bv. $< 0{,}001$. En aangezien we in paragraaf 2.20 aangenomen hebben dat we er op mogen vertrouwen dat gebeurtenissen met zeer kleine kansen bij eenmalige uitvoering van het experiment niet optreden, mogen we dan óók vertrouwen hebben dat in een lange serie van onafhankelijke proeven het frequentiequotiënt niet veel van p zal afwijken. De zojuist bewezen Zwakke Wet van de grote aantallen voorspelt dus dat in lange reeksen onafhankelijke proeven, frequentiequotiënten zich gaan stabiliseren.

Covariantie

5.10 Definitie.

De onafhankelijkheid in paragraaf 5.4(f) is heel wezenlijk. Bekijk maar eens de volgende twee voorbeelden van "zwaar afhankelijke" stochasten:

(a) Stel $X = Y$. X en Y zijn dan zgn. positief afhankelijk: omdat ze steeds dezelfde waarde aannemen, versterken hun afwijkingen (van $\mathbb{E}(X)$) elkaar. En inderdaad: $\mathbb{V}\text{ar}(X+Y) = \mathbb{V}\text{ar}(2X) = 4\mathbb{V}\text{ar}(X)$, terwijl $\mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y) = 2\mathbb{V}\text{ar}(X)$.

(b) Als $X = -Y$, hebben we negatieve afhankelijkheid: de afwijkingen van X en Y werken elkaar lijnrecht tegen. We zien: $\mathbb{V}\text{ar}(X+Y) = 0$, terwijl $\mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y) = 2\mathbb{V}\text{ar}(X)$.

Kunnen we dan i.h.a. niets zeggen over het verband tussen $\mathbb{V}\text{ar}(X+Y)$ enerzijds en $\mathbb{V}\text{ar}(X)$ en $\mathbb{V}\text{ar}(Y)$ anderzijds, als X en Y afhankelijke stochasten zijn?
Gelukkig wel. We voeren daarvoor een nieuwe grootheid in.
Definitie. De *covariantie* van de stochasten X en Y wordt genoteerd met $\mathbb{C}\text{ov}(X, Y)$ en gedefinieerd als

$$\boxed{\mathbb{C}\text{ov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\right].}$$

5.11 Eigenschappen.

(a) $\mathbb{C}\mathrm{ov}(X, X) = \mathbb{V}\mathrm{ar}(X)$;
(b) $\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
(c) Als X en Y onafhankelijk zijn, dan is $\mathbb{C}\mathrm{ov}(X, Y) = 0$ (andersom hoeft niet);
(d) X en Y ongecorreleerd $\Leftrightarrow \mathbb{C}\mathrm{ov}(X, Y) = 0$ (voor de definitie van ongecorreleerdheid, zie paragraaf 4.13);
(e) $\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{C}\mathrm{ov}(Y, X)$;
(f) $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\mathbb{C}\mathrm{ov}(X, Y)$;
(g) $\mathbb{V}\mathrm{ar}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i) + 2\sum_{i<j} \mathbb{C}\mathrm{ov}(X_i, X_j)$;
(h) $\mathbb{C}\mathrm{ov}(X, aY) = a\mathbb{C}\mathrm{ov}(X, Y)$ voor alle $a \in \mathbb{R}$.

De meeste van deze eigenschappen zijn heel gemakkelijk te bewijzen; we stellen dit uit tot opgave 2.

Voorbeeld. Het lootjesprobleem (zie opgave 5 in hoofdstuk 2 en paragraaf 4.9(d)). $X_1, \ldots, X_n$ zijn niet onafhankelijk, dus we gebruiken bovenstaande eigenschap (f): voor elke i is $X_i \sim \mathrm{Alt}(\frac{1}{n})$, zodat $\mathbb{V}\mathrm{ar}(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$. Voor $i \neq j$ is $\mathbb{E}(X_iX_j) = \mathbb{P}[X_i = X_j = 1] = \frac{1}{n(n-1)}$, zodat

$$\begin{aligned} \mathbb{C}\mathrm{ov}(X_i, X_j) &= \mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &= \frac{1}{n(n-1)} - \frac{1}{n^2}. \end{aligned}$$

Enig rekenwerk levert dan: $\mathbb{V}\mathrm{ar}(X) = 1$.

5.12 De covariantie kan zowel positief als negatief zijn. Een positieve covariantie duidt op een positieve afhankelijkheid, een negatieve covariantie duidt op een negatieve afhankelijkheid. Nu zegt de covariantie van X en Y niet veel over de mate van afhankelijkheid tussen X en Y: als we X en Y beide verdubbelen, wordt $\mathbb{C}\mathrm{ov}(X, Y)$ $4 \times$ zo groot, terwijl de mate van afhankelijkheid natuurlijk dezelfde blijft.

Daarom bestaat er ook nog een “gestandaardiseerde covariantie”, de zgn. *correlatie-coëfficiënt* $\rho_{X,Y}$:

$$\boxed{\rho_{X,Y} = \frac{\mathbb{C}\mathrm{ov}(X, Y)}{\sigma_X \sigma_Y}.}$$

Altijd geldt: $-1 \leq \rho_{X,Y} \leq 1$. Dit volgt direct uit:

Ongelijkheid van Cauchy-Schwarz:

$$|\mathbb{C}\text{ov}(X,Y)| \leq \sigma_X \sigma_Y.$$

Gelijkheid treedt op dan en slechts dan als $Y = c \cdot X$ voor zekere c.

Bewijs. Omdat $\mathbb{V}\text{ar}(X + \lambda Y) \geq 0$ voor alle $\lambda \in \mathbb{R}$, moet de discriminant van de kwadratische functie

$$\lambda \mapsto \mathbb{V}\text{ar}(X) + 2\lambda \mathbb{C}\text{ov}(X,Y) + \lambda^2 \mathbb{V}\text{ar}(Y)$$

≤ 0 zijn. Dus moet

$$(\mathbb{C}\text{ov}(X,Y))^2 \leq \sigma_X^2 \sigma_Y^2.$$

□

Ga zelf na dat de uiterste waarden -1 en 1 van de covariantie juist bereikt worden in de twee voorbeelden in paragraaf 5.10. Verder is het duidelijk dat $\rho_{X,Y} = 0$ zodra X en Y ongecorreleerd zijn, iets wat de naamgeving al suggereert.

Opgaven

1. Men werpt n dobbelstenen. X is het totaal aantal ogen, $Y = \frac{X}{n}$ het gemiddelde per worp. Bepaal $\mathbb{V}\mathrm{ar}(X)$ en $\mathbb{V}\mathrm{ar}(Y)$.

2. Bewijs de eigenschappen van $\mathbb{C}\mathrm{ov}$ in paragraaf 5.11.

3. **Hypergeometrische verdeling.** Notaties als in paragraaf 8.3(b) en paragraaf 4.9(c). Omdat $X_1, \ldots, X_n$ niet onafhankelijk zijn, is nu i.h.a.
$$\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{n} X_i\right) \neq \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i).$$
Gebruik formule 5.11(f) om $\mathbb{V}\mathrm{ar}(X)$ te berekenen.
(Antwoord: $\mathbb{V}\mathrm{ar}(X) = npq \cdot \frac{N-n}{N-1}$).

4. We gooien n keer een munt op. $P_n = \frac{1}{n} S_n$ is daarbij het frequentiequotiënt voor het optreden van kop. Hoe groot zou volgens de afschatting in het bewijs van de Wet van de grote aantallen (paragraaf 5.9) n moeten zijn, opdat de kans op een afwijking $|P_n - \frac{1}{2}| \geq 0,01$ kleiner dan 0,02 zij?

5. **Zwakke wet van de grote aantallen voor gemiddelden.**
Laat $X_1, X_2, \ldots$ onafhankelijke stochasten zijn met $\mathbb{E}(X_i) = \mu$ en $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2$ (elke i).
Zij $\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. Bewijs dat voor elke $\varepsilon > 0$: $\lim_{n\to\infty} \mathbb{P}[|\overline{X}_n - \mu| < \varepsilon] = 1$.

6. De stochast X neemt de waarden 0,1,2,... aan met kansen $p_0, p_1, p_2, \ldots$. Zij $Q_k = \mathbb{P}[X > k] = p_{k+1} + p_{k+2} + \ldots$
In paragraaf 4.15 is aangetoond dat $\mathbb{E}(X) = \sum_{k=0}^{\infty} Q_k$. We gaan met deze methode een stap verder:
   (a) Zij $R_k = Q_{k+1} + Q_{k+2} + \ldots$. Toon aan dat $\sum_{k=0}^{\infty} R_k = \mathbb{E}\left(\frac{1}{2} X(X-1)\right)$.
   (b) Laat Y het aantal proeven zijn nodig om één succes te behalen (onafhankelijke proeven met succeskans p). Gebruik de methode in (a) om $\mathbb{E}\left(Y(Y-1)\right)$ en daarmee $\mathbb{V}\mathrm{ar}(Y)$ te bepalen.
   (c) Laat Y_n het aantal proeven zijn nodig om n successen te behalen. Bepaal $\mathbb{V}\mathrm{ar}(Y_n)$. Aanwijzing: gebruik opgave 9 in hoofdstuk 3.

7. De schoenenkast (opgave 11 in hoofdstuk 1). Om het rekenwerk wat te beperken nemen we $n = 3$. Bereken de variantie van het aantal gepakte paren schoenen. Zie ook opgave 8 in hoofdstuk 5.

Hoofdstuk 6

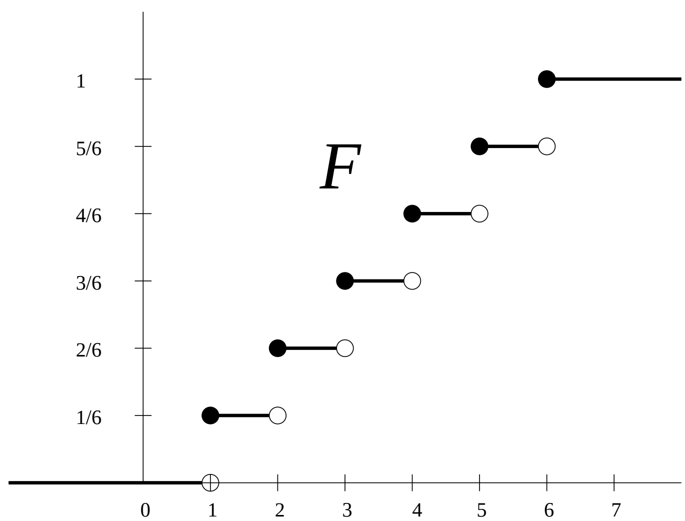
Continue verdelingen

Verdelingsfuncties

6.1 Tot nog toe hebben we ons vrijwel uitsluitend beziggehouden met discrete stochasten. In dit hoofdstuk introduceren we een ander soort stochasten, nl. continu verdeelde stochasten. Vòòr het zover is (in paragraaf 6.3), zullen we eerst het begrip verdelingsfunctie invoeren, en wel voor àlle reëelwaardige stochasten.
Definitie. Voor een stochast X wordt de *verdelingsfunctie* F als volgt gedefinieerd:

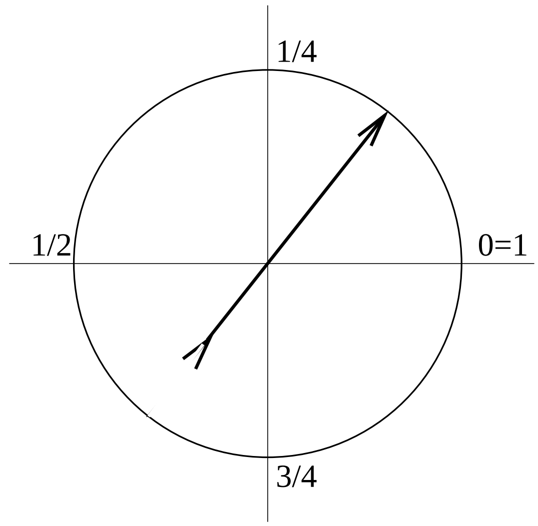
$$\boxed{F(x) = \mathbb{P}[X \leq x] \text{ voor elke } x \in \mathbb{R}.}$$

Voorbeeld. Zij X het aantal ogen van een worp met een dobbelsteen. Dan heeft F de volgende grafiek:



Voorbeeld. Rad van avontuur. In het middelpunt van een schijf is een vrij draaiende wijzer aangebracht. Op de omtrek van de schijf is een schaalverdeling van $[0,1)$ geplaatst. Het punt X waar de wijzer tot stilstand komt is een zeker reëel getal uit $[0,1)$.

Het heeft geen zin om te gaan kijken naar kansen $\mathbb{P}[X = x]$ voor punten $x \in [0,1)$; die kansen zijn allemaal 0.
We zullen in plaats daarvan kansen $\mathbb{P}[a < X < b]$ beschouwen, m.a.w. de kans dat X in het interval (a,b) terechtkomt.
Is er geen enkele voorkeur voor bepaalde richtingen, dan hangt deze kans alleen van de lengte van het interval af: $\mathbb{P}[a < X < b] = b - a$ voor ieder interval $(a,b) \subset [0,1)$. Dit betekent dat de kans met een constante dikte is uitgesmeerd over $[0,1)$.



De verdelingsfunctie van X is dus

$$F_X(x) = \begin{cases} 0 & \text{als } x < 0 \\ x & \text{als } 0 \leq x \leq 1 \\ 1 & \text{als } x > 1 \end{cases}$$

6.2 Algemene eigenschappen van verdelingsfuncties. Voor elke verdelingsfunctie gelden de volgende eigenschappen:

(a) F is stijgend, d.w.z. als $a < b$, dan $F(a) \leq F(b)$.

BEWIJS. $[X \leq a] \subset [X \leq b]$, dus $F(a) = \mathbb{P}[X \leq a] \leq \mathbb{P}[X \leq b] = F(b)$. $\square$

(b)

$$\boxed{\lim_{x\to\infty} F(x) = 1.}$$

BEWIJS. Omdat we al weten dat F stijgend is, is het voldoende om te bewijzen dat $\lim_{n\to\infty} F(n) = 1$. Wel, met 2.5(c) is

$$\begin{aligned} \lim_{n\to\infty} F(n) &= \lim_{n\to\infty} \mathbb{P}[X \leq n] \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} [X \leq n]\right) \\ &= \mathbb{P}[X < \infty] = 1. \quad \square \end{aligned}$$

(c)
$$\boxed{\lim_{x\to-\infty} F(x) = 0.}$$

Bewijs. Net als dat van (b). □

(d)
$$\boxed{\lim_{x\downarrow a} F(x) = F(a) \quad \text{en} \quad \lim_{x\uparrow a} F(x) = \mathbb{P}[X < a].}$$

Bewijs.

$$\begin{aligned}
\lim_{x\downarrow a} F(x) &= \lim_{n\to\infty} F\left(a + \frac{1}{n}\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left[X \le a + \frac{1}{n}\right] \\
&\overset{2.7(a)}{=} \mathbb{P}\left(\bigcap_{n=1}^{\infty} \left[X \le a + \frac{1}{n}\right]\right) \\
&= \mathbb{P}[X \le a] = F(a)
\end{aligned}$$

en

$$\begin{aligned}
\lim_{x\uparrow a} F(x) &= \lim_{n\to\infty} F\left(a - \frac{1}{n}\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left[X \le a - \frac{1}{n}\right] \\
&\overset{2.5(c)}{=} \mathbb{P}\left(\bigcup_{n=1}^{\infty} \left[X \le a - \frac{1}{n}\right]\right) \\
&= \mathbb{P}[X < a].
\end{aligned}$$

□

We zien dat de grootte van de sprong die F maakt in $x = a$ gelijk is aan

$$\lim_{x\downarrow a} F(x) - \lim_{x\uparrow a} F(x) = \mathbb{P}[X \le a] - \mathbb{P}[X < a] = \mathbb{P}[X = a].$$

Dichtheidsfuncties en continu verdeelde stochasten

6.3 Definitie. De stochast X is *continu verdeeld* als de verdelingsfunctie F van X continu is.

Opmerking. Uit 6.2(d) volgt: als X continu verdeeld is, dan is $\mathbb{P}[X = a] = 0$ voor alle $a \in \mathbb{R}$.

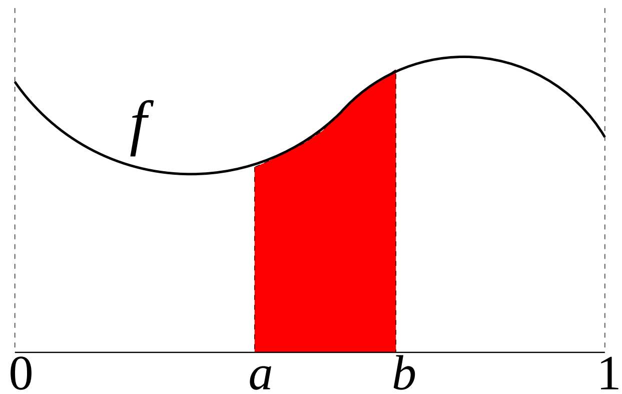
Voorbeeld. Het rad van avontuur uit 6.1. We nemen aan, dat de wijzer geen enkele voorkeur heeft voor bepaalde richtingen. Merk op dat onder deze aanname

$$F(x) = \begin{cases} 0 & \text{als } x \leq 0; \\ x & \text{als } 0 < x < 1; \\ 1 & \text{als } 1 \leq x \end{cases}$$

zodat X een continu verdeelde stochast is. Zou er wèl voorkeur zijn voor bepaalde richtingen, dan geldt dat

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx$$

voor een of andere functie f, die er bijvoorbeeld als in het plaatje kan uitzien. Deze functie f zullen we de *dichtheid* van X noemen; de grafiek van f geeft aanschouwelijk weer hoe dik de kansmassa plaatselijk uitgesmeerd is.



6.4 Definitie. De stochast X is *continu verdeeld met dichtheidsfunctie*[1] f als voor ieder interval (a, b) geldt:

$$\boxed{\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx.}$$

Hierbij zijn de keuzen $a = -\infty$ en $b = \infty$ toegestaan.
Wanneer je $a = -\infty$ invult, vind je: $\int_{-\infty}^b f(x)dx = \mathbb{P}[X \leq b] = F(b)$, dus

$$\boxed{F(x) = \int_{-\infty}^x f(u)du \text{ voor elke } x \in \mathbb{R}.}$$

Een stochast die continu verdeeld is met dichtheidsfunctie f (volgens definitie 6.4) is dus ook inderdaad continu verdeeld volgens definitie 6.3.
Opmerking. De functie f ligt nooit helemaal vast. Als je nl. de waarde van f in één punt verandert, dan verandert de integraal van f daardoor niet. We hebben daardoor altijd een heel klein beetje keuzevrijheid voor de functie f. Wel zullen we

[1] In de literatuur heet "continu met dichtheidsfunctie f" ook wel: "absoluut continu".
Men kan laten zien dat er stochasten zijn die continu verdeeld zijn, maar geen dichtheidsfunctie hebben, maar zulke stochasten zullen wij hier niet tegenkomen.

f altijd zo kiezen dat $f(x) \geq 0$ voor alle $x \in \mathbb{R}$.
Natuurlijk is altijd

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

6.5 Verband tussen f en F.

(a) Zij X een continu verdeelde stochast met onbekende dichtheid f, waarvan we wèl de verdelingsfunctie F kennen. Dan vinden we f door F te differentiëren (bewijs: De hoofdstelling van de Analyse):

$$\boxed{f(x) = F'(x).}$$

F hoeft niet in élk punt x differentieerbaar te zijn. F is echter wèl differentieerbaar in die punten waar we f continu kunnen maken. In de resterende punten, en dat zullen er in dit dictaat meestal één of twee zijn, is de waarde van f niet van belang.

(b) Met de hoofdstelling van de Analyse vinden we ook:

$$\boxed{\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx = F(b) - F(a).}$$

Voorbeelden van continu verdeelde stochasten

Het wordt tijd voor een aantal voorbeelden van continu verdeelde stochasten. In de opsomming hieronder zal het belangrijkste voorbeeld, de zogenaamde "normale verdeling" ontbreken. Dit hiaat zal ruimschoots worden opgevuld door hoofdstuk 10, dat in zijn geheel is gewijd aan deze verdeling.

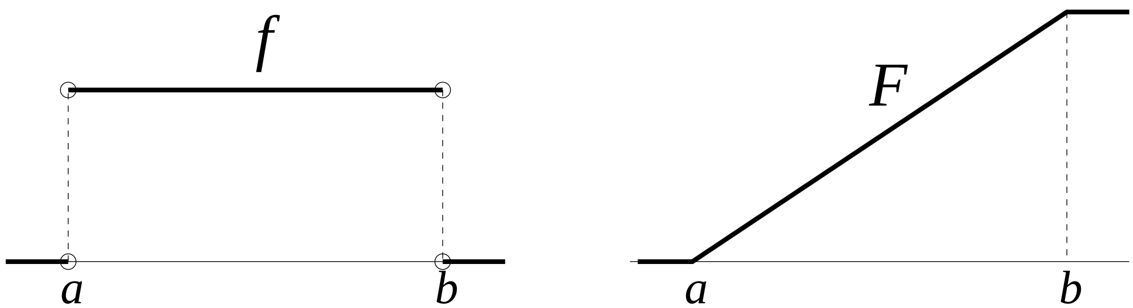
6.6 De uniforme (of homogene) verdeling. Een stochast X is *uniform verdeeld* over (a, b) als X zijn waarden aanneemt tussen a en b, en zonder voorkeur voor bepaalde delen van (a, b). Notatie: $X \sim \mathrm{Un}(a, b)$. De dichtheidsfunctie f van X is dan constant op (a, b); omdat $\int_{-\infty}^{\infty} f(x)dx = 1$ moet zijn, zal die constante waarde van f gelijk zijn aan $\frac{1}{b-a}$. Dus:

$$\boxed{f(x) = \begin{cases} 0 & \text{als } x < a; \\ \frac{1}{b-a} & \text{als } a < x < b; \\ 0 & \text{als } x > b. \end{cases}}$$

$f(a)$ en $f(b)$ schrijven we niet voor.
De verdelingsfunctie F berekenen we nu m.b.v. 6.5(b) (of 6.4):

$$\boxed{F(x) = \begin{cases} 0 & \text{als } x \leq a; \\ \frac{x-a}{b-a} & \text{als } a < x < b; \\ 1 & \text{als } x \geq b. \end{cases}}$$

De grafieken:



Voorbeeld 1. In het rad van avontuur is $X \sim \text{Un}(0, 1)$. Dus:

$$f(x) = \begin{cases} 0 & \text{als } x < 0; \\ 1 & \text{als } 0 < x < 1; \\ 0 & \text{als } x > 1; \end{cases} \qquad F(x) = \begin{cases} 0 & \text{als } x \leq 0; \\ x & \text{als } 0 < x < 1; \\ 1 & \text{als } x \geq 1. \end{cases}$$

Voorbeeld 2. Een draad van lengte 1, langs de x-as gespannen tussen de punten 0 en 1, breekt in een toevallig punt. De kans dat het breekpunt in een interval (a, b) ligt, stellen we gelijk aan $b - a$, en dit voor ieder interval $(a, b) \subset [0, 1]$. Als dan X de lengte is van het linkerstuk, dan is $X \sim \text{Un}(0, 1)$. Welke verdeling heeft de lengte Y van het rechterstuk?

Voorbeeld 3. Afrondingsfouten. Een tabel geeft de logaritmen van de getallen 1 tot en met 10.000 in 4 decimalen. De vierde decimaal is verkregen door afronding. Is Y een willekeurig getal, dan kun je $\log Y$ schrijven als

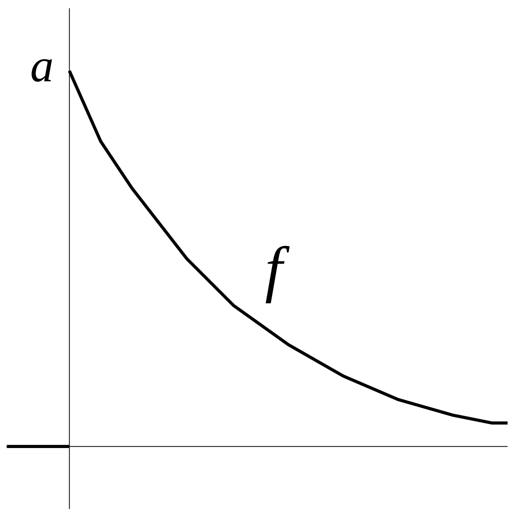
$$\log Y = \texttt{a,bcde} + 10^{-4}X,$$

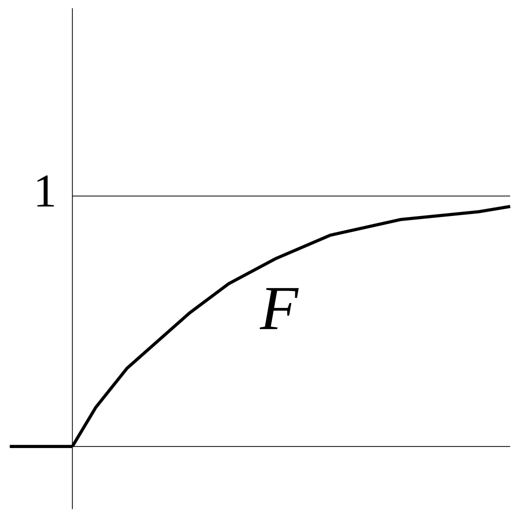
waarin `a,bcde` het getal is dat we in de tabel vinden en $10^{-4}X$ de (niet vermelde) afrondingsfout. Als de afronding correct is, dan geldt: $-\frac{1}{2} \leq X < \frac{1}{2}$.
We mogen aannemen dat $X \sim \text{Un}(-\frac{1}{2}, \frac{1}{2})$. Door er een grotere logaritmentafel bij te halen, met bv. 8 decimalen, kunnen we voor een aantal willekeurige waarden van Y de bijbehorende afrondingsfouten in de kleine tabel bepalen (in slechts vier decimalen nauwkeurig weliswaar). Is bv. $\log Y = 0{,}36277281$ volgens de grote tabel, dan is in de kleine tabel $\log Y = 0{,}3628$, zodat $X = -0{,}2719$.

6.7 De exponentiële verdeling. Zij a een positief reëel getal. Een stochast X is *exponentieel verdeeld met parameter* a (kortweg $X \sim \mathrm{Exp}(a)$) als $\mathbb{P}[X > x] = e^{-ax}$ voor alle $x \geq 0$. We vinden dan eenvoudig de dichtheids- en verdelingsfunctie van X:

$$F(x) = \begin{cases} 0 & \text{als } x \leq 0; \\ 1 - e^{-ax} & \text{als } x > 0; \end{cases} \qquad f(x) = \begin{cases} 0 & \text{als } x < 0; \\ ae^{-ax} & \text{als } x > 0. \end{cases}$$

De grafieken:





Voorbeeld 1. De levensduur T van sommige voorwerpen (borden, kopjes) heeft een exponentiële verdeling. Immers, het aantal borden dat er op een dag sneuvelt is evenredig met het aantal borden dat er nog aanwezig is, dus

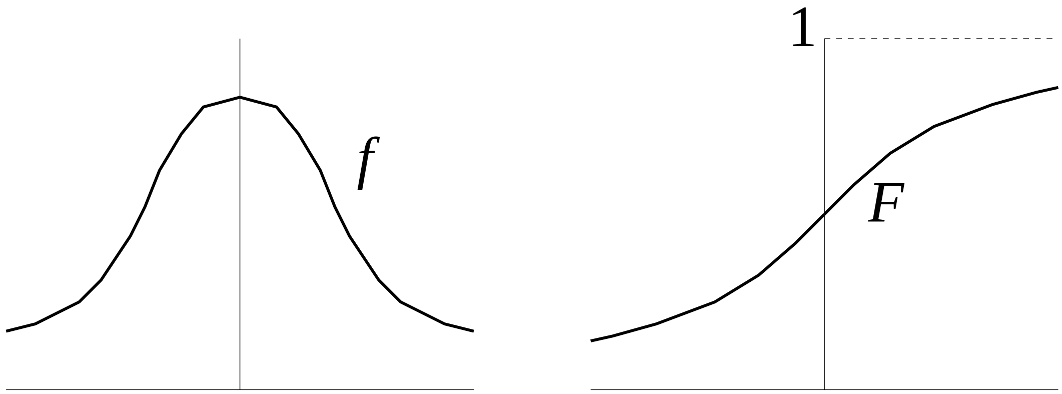
$$\frac{d}{dt}\mathbb{P}[T > t] = -a\mathbb{P}[T > t]$$

voor zekere $a > 0$.
Oplossen van deze differentiaalvergelijking levert $\mathbb{P}[T > t] = e^{-at}$ (voor $t > 0$), zodat T inderdaad exponentieel verdeeld is.
Voorbeeld 2. Wachten op de eerste klant in het Poissonproces. In hoofdstuk 7, opgave 6, is de volgende berekening gemaakt: $\mathbb{P}[T > t] = \mathbb{P}[K_t = 0] = e^{-\lambda t}$, zodat $\mathbb{P}[T \leq t] = 1 - e^{-\lambda t}$.
Conclusie: $T \sim \mathrm{Exp}(\lambda)$.

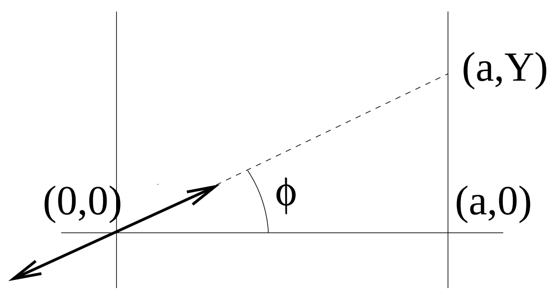
6.8 De Cauchy- (of arctan-)verdeling. Zij a een positief getal. Een stochast X is *Cauchy verdeeld met parameter* a als X de volgende dichtheids- en verdelingsfunctie heeft:

$$f(x) = \frac{a}{\pi(a^2 + x^2)} \quad (x \in \mathbb{R}); \qquad F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x}{a}\right) \quad (x \in \mathbb{R}).$$

De grafieken:



Voorbeeld. Een tweezijdige wijzer draait om het punt (0,0) in het x, y-vlak en komt in een willekeurige richting tot stilstand.
De hoek die de wijzer maakt met de positieve x-as noemen we φ. We veronderstellen dat φ uniform verdeeld is over $\left(-\frac{1}{2}\pi, \frac{1}{2}\pi\right)$. Op de lijn $x = a$ wordt door de wijzer een punt (a, Y) aangewezen. We laten het als een opgave (5) om te bewijzen dat deze stochast Y Cauchy verdeeld is met parameter a.



6.9 Laat X continu verdeeld zijn met dichtheid f. Als Y een functie is van X, bv. $Y = X^2$, hoe ziet de kansverdeling van Y er dan uit? Dit soort vragen is meestal het eenvoudigst te beantwoorden door de verdelingsfunctie van Y te berekenen.
Voorbeeld. Stel de omtrek X van een vierkant is uniform verdeeld over $(10, 20)$. Welke kansverdeling heeft de oppervlakte Y van dat vierkant?
Oplossing: Daar $Y = \left(\frac{1}{4}X\right)^2$, is voor $y > 0$: $F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}\left[\frac{1}{4}X \leq \sqrt{y}\right] = \mathbb{P}[X \leq 4\sqrt{y}] = F_X(4\sqrt{y})$, dus

$$F_Y(y) = \begin{cases} 0 & \text{als } y \leq 6\frac{1}{4}; \\ \frac{2}{5}\sqrt{y} - 1 & \text{als } 6\frac{1}{4} < y < 25; \\ 1 & \text{als } y \geq 25 \end{cases}$$

en

$$f_Y(y) = \frac{1}{5\sqrt{y}} \cdot \mathbf{1}_{(6\frac{1}{4}, 25)}(y).$$

N.B. De onderindex "Y" of "X" bij dichtheids- en verdelingsfuncties gebruiken we als we het nodig vinden om duidelijk aan te geven over welke stochast we het hebben.

Onafhankelijkheid, verwachtingswaarde, variantie.

6.10 Onafhankelijkheid De definitie in paragraaf 3.9 is niet langer bruikbaar, omdat $\mathbb{P}[X{=}x] = 0$ voor continu verdeelde stochasten X; De volgende definitie heeft zin voor álle stochasten, en stemt voor discrete stochasten overeen met die van paragraaf 3.9.
Definitie. X en Y heten *onafhankelijk* als voor alle x en y geldt:

$$\mathbb{P}[X{\leq}x, Y{\leq}y] = \mathbb{P}[X{\leq}x] \cdot \mathbb{P}[Y{\leq}y].$$

6.11 Verwachting van continu verdeelde stochasten. De verwachting van een stochast X die continu verdeeld is met dichtheid f berekenen we met de volgende formule:

$$\boxed{\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx.}$$

Voorbeeld 1. Als $X \sim \mathrm{Un}(a,b)$, dan is

$$\mathbb{E}(X) = \int_a^b x \cdot \frac{1}{b-a}dx = \frac{1}{2}(b^2 - a^2) \cdot \frac{1}{b-a} = \frac{1}{2}(a+b),$$

een antwoord dat ook intuïtief geheel duidelijk is.
Voorbeeld 2. Als $X \sim \mathrm{Exp}(a)$, dan volgt uit de definitie:

$$\mathbb{E}(X) = \int_0^{\infty} x \cdot ae^{-ax}dx.$$

Als deze integraal bestaat, moet hij gelijk zijn aan

$$\lim_{n\to\infty} \int_0^n x \cdot ae^{-ax}dx.$$

Wel, met behulp van partiële integratie:

$$\begin{aligned} \int_0^n x \cdot ae^{-ax}dx &= \Big[-xe^{-ax}\Big]_{x=0}^{x=n} + \int_0^n e^{-ax}dx \\ &= -ne^{-an} + \frac{1}{a}(1 - e^{-an}) \end{aligned}$$

en

$$\lim_{n\to\infty} \left(-ne^{-an} + \frac{1}{a}\left(1 - e^{-an}\right)\right) = \frac{1}{a}$$

dus $\mathbb{E}(X) = \frac{1}{a}$.
Voorbeeld 3. Als $X \sim \text{Cauchy}(a)$, bestaat $\mathbb{E}(X)$ niet.
Immers, $\int_{-\infty}^{\infty} x \frac{a}{\pi(a^2+x^2)} dx$ bestaat slechts als $\int_{-\infty}^{0} x \frac{a}{\pi(a^2+x^2)} dx$ en $\int_{0}^{\infty} x \frac{a}{\pi(a^2+x^2)} dx$ bestaan.
Maar

$$\int_0^n \frac{a}{\pi} \cdot \frac{x}{a^2+x^2} dx = \left[\frac{a}{2\pi} \log(a^2+x^2) \right]_{x=0}^{x=n} = \frac{a}{2\pi} \log\left(\frac{a^2+n^2}{a^2} \right);$$

deze uitdrukking heeft geen limiet als $n \to \infty$, dus $\mathbb{E}(X)$ bestaat niet.

Voorwaarde. Voorbeeld 3 maakt duidelijk dat we een voorwaarde moeten opleggen aan de dichtheidsfunctie f van een continu verdeelde stochast X, wil $\mathbb{E}(X)$ bestaan. Deze voorwaarde luidt:

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty.$$

6.12 * Verband tussen de twee definities van de verwachting.

indexverwachting v. discrete en continue stochasten Voor een discrete stochast Y met waardenverzameling W hebben we gedefinieerd: $\mathbb{E}(Y) = \sum_{w \in W} w \cdot \mathbb{P}[Y = w]$. Voor een continu verdeelde stochast X met dichtheidsfunctie f hebben we gedefinieerd: $\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$.
Er bestaat wel degelijk een verband tussen deze twee definities. In het nu volgende verhaal doen we alsof we alleen de eerste definitie kennen; we leiden de tweede dan af.
Zij X een continu verdeelde stochast met dichtheidsfunctie f. Kies een klein positief getal h (bijvoorbeeld $\frac{1}{10}$).
We gaan X afronden (naar beneden) op gehele veelvouden van h. Het resultaat is een stochast X_h die de waarde kh aanneemt ($k \in \mathbb{Z}$) als X een waarde aanneemt in $[kh, (k+1)h)$. De afgeronde stochast X_h is dus een discrete stochast met mogelijke waarden

$$\ldots, -3h, -2h, -h, 0, h, 2h, 3h, \ldots$$

en kansverdeling

$$\mathbb{P}[X_h = kh] = \mathbb{P}[kh \leq X < (k+1)h] = \int_{kh}^{(k+1)h} f(x) dx.$$

X_h ligt steeds heel dicht bij X, want $X_h \leq X < X_h + h$.
De verwachting van X zal dus, wanneer we haar zinnig willen definiëren, ook dicht bij $\mathbb{E}(X_h)$ moeten liggen. Hoe kleiner h, des te dichter zal $\mathbb{E}(X)$ bij $\mathbb{E}(X_h)$ moeten liggen. Dit leidt ons ertoe te stellen:

$$\mathbb{E}(X) = \lim_{h \downarrow 0} \mathbb{E}(X_h).$$

Omdat X_h een discrete stochast is, kunnen we $\mathbb{E}(X_h)$ makkelijk uitrekenen:

$$\mathbb{E}(X_h) = \sum_{k\in\mathbb{Z}} kh \cdot \mathbb{P}[X_h{=}kh] = \sum_{k\in\mathbb{Z}} \int_{kh}^{(k+1)h} kh \cdot f(x)dx.$$

Als $x \in [kh, (k{+}1)h)$, dan is $kh \leq x < (k{+}1)h$, zodat

$$\int_{kh}^{(k+1)h} kh \cdot f(x)dx \leq \int_{kh}^{(k+1)h} x \cdot f(x)dx \leq \int_{kh}^{(k+1)h} (k{+}1)h \cdot f(x)dx$$

en

$$\int_{kh}^{(k+1)h} (k{+}1)h \cdot f(x)dx = \int_{kh}^{(k+1)h} kh \cdot f(x)dx + h \int_{kh}^{(k+1)h} f(x)dx.$$

Door sommeren over k volgt hieruit:

$$\mathbb{E}(X_h) \leq \int_{-\infty}^{\infty} x \cdot f(x)dx \leq \mathbb{E}(X_h) + h.$$

Dus:

$$\lim_{h\downarrow 0} \mathbb{E}(X_h) = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

6.13 Eigenschappen van de verwachting. Alle eigenschappen van de verwachtingswaarde $\mathbb{E}$ die we in hoofdstuk 5 zijn tegengekomen en die niet specifiek handelen over discrete stochasten, gelden ook voor continu verdeelde stochasten. Een opsomming:

(a) $\mathbb{E}(aX) = a\mathbb{E}(X)$ voor alle $a \in \mathbb{R}$.
(b) als $X \geq 0$, dan $\mathbb{E}(X) \geq 0$.
(c) als $X \geq Y$, dan $\mathbb{E}(X) \geq \mathbb{E}(Y)$.
(d) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
(e) $\mathbb{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{E}(X_k)$.
(f) als $X \perp\!\!\!\perp Y$, dan $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

Bewijzen geven we niet. Ze zijn in de stijl van paragraaf 6.12.

Als X continu verdeeld is, met dichtheid f, en g is een enigszins nette functie, dan is $g(X)$ weer een stochast, waarvan we de verwachtingswaarde als volgt berekenen:

$$\boxed{\mathbb{E}\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)f(x)dx}$$

(geen bewijs). In het bijzonder levert de keuze $g(x) = (x - \mu)^2$ (waar $\mu = \mathbb{E}(X)$):

$$\boxed{\mathbb{V}\mathrm{ar}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.}$$

Ook voor de variantie geldt dat alle niet specifiek discrete eigenschappen geldig blijven:

(g) $\mathbb{V}\mathrm{ar}(X) \geq 0$.

(h) $\mathbb{V}\mathrm{ar}(aX) = a^2\mathbb{V}\mathrm{ar}(X)$ voor alle $a \in \mathbb{R}$.

(i) $\mathbb{V}\mathrm{ar}(X+a) = \mathbb{V}\mathrm{ar}(X)$ voor alle $a \in \mathbb{R}$.

(j) $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

(k) Als $X \perp\!\!\!\perp Y$ dan is $\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$.

(l) Als $X_1, \ldots, X_n$ onafhankelijk zijn, dan is $\mathbb{V}\mathrm{ar}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(X_k)$.

(m) $\mathbb{P}[|X - \mathbb{E}(X)| \geq a] \leq \frac{1}{a^2}\mathbb{V}\mathrm{ar}(X)$ (Chebyshev).

Ook σ_X en $\mathbb{C}\mathrm{ov}(X)$ kunnen weer ingevoerd worden. We noemen slechts twee eigenschappen:

(n) $\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\mathbb{C}\mathrm{ov}(X,Y)$.

(o) $\mathbb{V}\mathrm{ar}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(X_k) + 2\sum_{i<j} \mathbb{C}\mathrm{ov}(X_i, X_j)$.

Opgaven

1. Laat in voorbeeld 2 van paragraaf 6.6 het langste van de twee stukken de lengte Z hebben, dus $Z = \max\{X, 1-X\}$. Bepaal verdelingsfunctie, dichtheidsfunctie en verwachting van Z.

2. Bepaal de variantie van de

   (a) uniforme verdeling;

   (b) exponentiële verdeling.

3. Wachten op de trein. Treinen naar Duistervoort vertrekken met tussenpozen van precies 1 uur. Als we op een willekeurig tijdstip op het station komen, moeten we een tijd T wachten tot de volgende trein vertrekt. Bepaal verwachting en variantie van T (T is, in minuten, uniform verdeeld over $(0, 60)$.).

4. Bewijs dat de stochast Y in het voorbeeld van paragraaf 6.8 inderdaad Cauchy-verdeeld is.

5. Laat X een Cauchy-verdeling hebben met parameter $a = 1$. Toon aan dat $Y = \frac{1}{X}$ dezelfde kansverdeling heeft als X.
   Aanwijzing: $\arctan x + \arctan \frac{1}{x} = \begin{cases} \frac{1}{2}\pi & \text{als } x > 0; \\ -\frac{1}{2}\pi & \text{als } x < 0. \end{cases}$

6. Laat X en Y onafhankelijke stochasten zijn, beide uniform verdeeld over $(0, 1)$. Bepaal verdelingsfunctie, dichtheidsfunctie en verwachting van $Z = \max\{X, Y\}$.

7. Stel de levensduur T van een voorwerp is exponentieel verdeeld met parameter a. Als gegeven is dat het voorwerp al minstens 100 dagen oud is, wat is dan de kans dat het ook de 101$^{\mathrm{e}}$ dag ongeschonden doorkomt? M.a.w. bereken de kans $\mathbb{P}(T > 101 | T > 100)$.
   Moraal: Slijtage speelt geen rol. De exponentiële verdeling is de enige verdeling die deze eigenschap heeft.

8. Laat X continu verdeeld zijn met verdelingsfunctie F en dichtheidsfunctie f.

   (a) Toon aan dat $Y = X^2$ een verdelingsfunctie G en een dichtheidsfunctie g heeft die gegeven worden door

$$\begin{aligned} G(y) &= F(\sqrt{y}) - F(-\sqrt{y}) & (y > 0); \\ g(y) &= \tfrac{1}{2\sqrt{y}} \Big(f(\sqrt{y}) + f(-\sqrt{y}) \Big) & (y > 0). \end{aligned}$$

(b) Idem voor $Z = aX + b$ ($a > 0$):

$$\begin{aligned} H(z) &= F\left(\frac{z-b}{a}\right); \\ h(z) &= \frac{1}{a} f\left(\frac{z-b}{a}\right). \end{aligned}$$

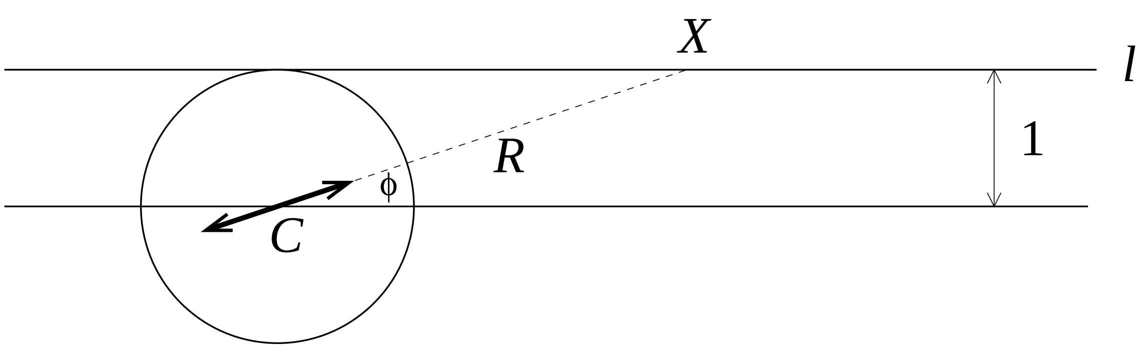
9. Zij $X_1, \ldots, X_n$ een rij onafhankelijke stochasten met dezelfde verdelingsfunctie F. Zij $Y_n = \min\{X_1, \ldots, X_n\}$.

   (a) Bepaal de verdelingsfunctie van Y_n.
   Aanwijzing: $[Y_n > x] = \bigcap_{k=1}^{n} [X_k > x]$.

   (b) Toepassing: tien lampen, allen met een levensduur die exponentieel verdeeld is met parameter $\frac{1}{20}$, worden tegelijk ingeschakeld. Zij T het tijdstip waarop de eerste lamp kapot gaat. Bereken verdelingsfunctie, dichtheidsfunctie en verwachting van T.

10. X_1 en X_2 zijn onafhankelijke stochasten. $\mathbb{E}(X_1) = \mu_1$, $\mathbb{E}(X_2) = \mu_2$, $\mathbb{V}\mathrm{ar}(X_1) = \sigma_1^2$, $\mathbb{V}\mathrm{ar}(X_2) = \sigma_2^2$. Bepaal de verwachtingswaarde en de variantie van het product $Y = X_1 \cdot X_2$.

11. Een tweezijdige pijl draait rond op een pen op C, en komt zodanig tot stilstand dat de hoek φ met een rechte ℓ op afstand 1 uniform over $(0, \pi)$ verdeeld is. De tot stilstand gekomen pijl wijst een punt X aan op ℓ. Zij R de afstand tussen C en X, en zij $Y = \frac{1}{R^2}$. Bepaal verdelings- en dichtheidsfunctie van Y.



12. Stel, X en Y zijn onafhankelijke stochasten; X neemt de waarden $0, \ldots, n-1$ aan elk met kans $\frac{1}{n}$, en $Y \sim \mathrm{Un}(0, 1)$.
    Bewijs: $X + Y \sim \mathrm{Un}(0, n)$.

Hoofdstuk 7

De normale verdeling

7.1 In paragraaf 9.5 zagen we dat herhaalde convoluties van de uniforme dichtheid een steeds mooiere vorm krijgen. Dit blijkt voor veel meer verdelingen te gelden. Door dit voor een aantal verdelingen uit te rekenen, ontdekte men al in de 18$^{\text{e}}$ en 19$^{\text{e}}$ eeuw speciale gevallen en beperkte versies van wat later de centrale-limietstelling zou gaan heten.
De formulering van deze stelling, die in paragraaf 7.2 staat, is bewezen door Lyapunov omstreeks 1900; later vond men allerlei veralgemeningen.

Theorie.

7.2 De centrale-limietstelling. Zij $X_1, X_2, \ldots$ een rij *onafhankelijke* stochasten zó, dat alle X_k dezelfde verdelingsfunctie hebben, en dus ook dezelfde verwachting μ ($|\mu| < \infty$) en dezelfde variantie σ^2 ($0 < \sigma < \infty$).
We zijn geïnteresseerd in de zgn. partiële sommen $S_n = X_1 + \ldots + X_n$.
Nu zal de rij $S_1, S_2, \ldots$ zelf i.h.a. niet interessant zijn, omdat de verdeling van S_n "wegloopt" als $n \to \infty$, immers $\mathbb{E}(S_n) = n\mu$. Daarom bekijken we liever de stochasten $S_1 - \mu, S_2 - 2\mu, S_3 - 3\mu, \ldots$; deze hebben alle verwachting 0.
Maar een nadeel blijft dat de verdeling van $S_n - n\mu$ zich steeds verder "uitspreidt" als n stijgt, immers $\mathbb{V}\text{ar}(S_n - n\mu) = \mathbb{V}\text{ar}(S_n) = n\sigma^2$. Daarom bekijken we de volgende stochasten:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

anders geschreven:

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}\text{ar}(S_n)}}.$$

Merk op dat $\mathbb{E}(Z_n) = 0$, $\mathbb{V}\text{ar}(Z_n) = 1$ voor elke n.
We noemen Z_n daarom *gestandaardiseerde stochasten* .

Voor de achtereenvolgende verdelingsfuncties van $Z_1, Z_2, Z_3, \ldots$ kan men het volgende bewijzen:

Centrale-limietstelling:

Zij $X_1, X_2, \ldots$ een rij onafhankelijke stochasten, alle met dezelfde verdelingsfunctie, verwachting μ en variantie σ^2.
Definieer

$$S_n = \sum_{k=1}^{n} X_k \text{ en } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Dan geldt voor elke $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}[Z_n \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du.$$

We bewijzen deze stelling niet.

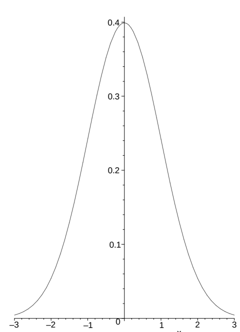
7.3 Standaard-normale verdeling.

De zojuist gevonden functie is de verdelingsfunctie van de *standaard-normale verdeling* . Deze is zo belangrijk, dat men er een speciale letter voor heeft gereserveerd, Φ, in plaats van de gebruikelijke F.
Dus:

$$\boxed{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du.}$$

De bijbehorende dichtheidsfunctie krijgen we nu cadeau; hiervoor wordt de letter φ gebruikt:

$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.}$$

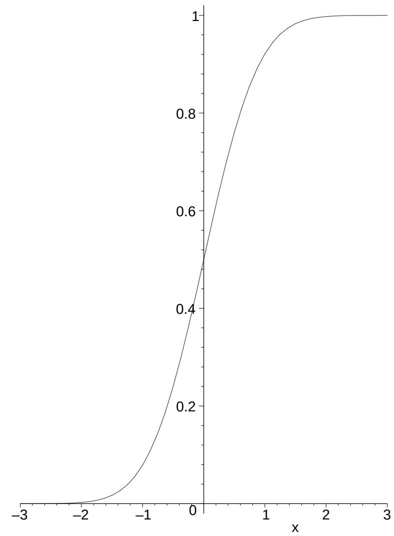


Het lijkt wellicht wat vreemd dat er in de definitie van een verdelingsfunctie een integraalteken staat. De reden hiervoor is dat we de primitieve functie van φ (Φ dus) niet op een elementaire manier in bekende functies kunnen uitdrukken (veeltermen, sin, arctan, log, exp, ...).

Om dezelfde reden is het evenmin mogelijk om rechtstreeks te bewijzen dat

$$\int_{-\infty}^{\infty} \varphi(x)dx = 1,$$

iets wat we wel degelijk moeten weten, willen we van een dichtheid kunnen spreken. Met een omweg lukt het echter wel: er is een bewijs mogelijk m.b.v. poolcoördinaten, zie opgave (13).



7.4 Verwachting en variantie.

Zij Z standaard-normaal verdeeld. Dan is

$$\boxed{\mathbb{E}(Z) = 0 \text{ en } \mathbb{V}\mathrm{ar}(Z) = 1.}$$

Dit ligt voor de hand, want in paragraaf 7.2 hebben we gezien dat

$$F_Z(z) = \Phi(z) = \lim_{n\to\infty} F_{Z_n}(z)$$

voor zekere stochasten $Z_1, Z_2, \ldots$ met $\mathbb{E}(Z_n) = 0$ en $\mathbb{V}\mathrm{ar}(Z_n) = 1$.

Bewijs. Om

$$\mathbb{E}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}x^2} dx$$

te bepalen, berekenen we

$$\int_0^n xe^{-\frac{1}{2}x^2} dx = \left[-e^{-\frac{1}{2}x^2}\right]_{x=0}^{x=n} = 1 - e^{-\frac{1}{2}n^2},$$

zodat

$$\int_0^\infty xe^{-\frac{1}{2}x^2} dx = 1.$$

Evenzo volgt $\int_{-\infty}^0 xe^{-\frac{1}{2}x^2} dx = -1$, zodat $\mathbb{E}(Z) = 0$.

$\mathbb{V}\mathrm{ar}(Z)$ berekenen we met partiële integratie:

$$\begin{aligned}
\int_0^n x^2 e^{-\frac{1}{2}x^2} dx &= \int_0^n x \cdot xe^{-\frac{1}{2}x^2} dx \\
&= \left[-xe^{-\frac{1}{2}x^2}\right]_{x=0}^{x=n} + \int_0^n e^{-\frac{1}{2}x^2} dx \\
&= -ne^{-\frac{1}{2}n^2} + \int_0^n e^{-\frac{1}{2}x^2} dx,
\end{aligned}$$

zodat

$$\int_0^\infty x^2 e^{-\frac{1}{2}x^2} dx = \int_0^\infty e^{-\frac{1}{2}x^2} dx.$$

Evenzo is

$$\int_{-\infty}^{0} x^2 e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{0} e^{-\frac{1}{2}x^2} dx,$$

zodat

$$\mathbb{V}\mathrm{ar}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = 1. \quad \square$$

7.5 Normale verdeling. We zeggen dat de stochast Y *normaal verdeeld* is met (twee!) parameters μ en σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$) als de gestandaardiseerde stochast $Z = \frac{Y-\mu}{\sigma}$ standaardnormaal verdeeld is.
Merk op: $\mu = \mathbb{E}(Y)$, $\sigma^2 = \mathbb{V}\mathrm{ar}(Y)$.
Notatie: $Y \sim \mathrm{N}(\mu, \sigma^2)$ (en dus $Z \sim \mathrm{N}(0,1)$).
De verdelingsfunctie van Y is:

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}\left[Z \leq \tfrac{y-\mu}{\sigma}\right] = \Phi\left(\tfrac{y-\mu}{\sigma}\right),$$

en de dichtheid (kettingregel!):

$$f_Y(y) = F_Y'(y) = \tfrac{1}{\sigma}\varphi\left(\tfrac{y-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}.$$

Bij normaal verdeelde stochasten kunnen we, als de parameters μ en σ bekend zijn, alle kansen berekenen m.b.v. de functie Φ. In elk statistiekboek is wel een tabel van Φ te vinden. In dit dictaat hebben we 'm in Bijlage A gezet.

Eigenschappen van Φ. Om snel te kunnen rekenen met de tabel is het handig enkele eigenschappen van de normale verdeling te kennen.

(a) $$\boxed{\Phi(0) = \tfrac{1}{2}.}$$

Bewijs. $\Phi(0) = \int_{-\infty}^{0} \varphi(x)\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{\infty} \varphi(x)\mathrm{d}x = \frac{1}{2}$ wegens de symmetrie van φ. $\square$

(b) $$\boxed{\Phi(-x) = 1 - \Phi(x).}$$

Bewijs. Ook dit is onmiddelijk duidelijk uit de symmetrie van φ:

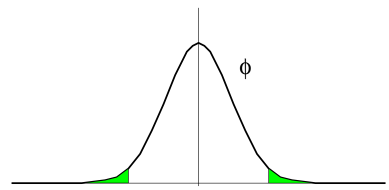
$$\begin{aligned} \Phi(-x) &= \mathbb{P}[Z \leq -x] = \mathbb{P}[Z \geq x] \\ &= 1 - \Phi(x). \quad \square \end{aligned}$$

(c) Als $Y \sim \mathrm{N}(\mu, \sigma^2)$ en $a > 0$ dan $\mathbb{P}[|Y-\mu| > a\sigma] = 2(1-\Phi(a))$.
Immers (bij ($\star$) maken we gebruik van de symmetrie van φ):

$$\begin{aligned}
\mathbb{P}[|Y-\mu| > a\sigma] &= \mathbb{P}[Y-\mu > a\sigma \text{ of } Y-\mu < -a\sigma] \\
&\overset{\star}{=} 2\mathbb{P}[Y-\mu > a\sigma] \\
&= 2\left(1-\mathbb{P}[Y-\mu \leq a\sigma]\right) \\
&= 2\left(1-\mathbb{P}\left[\frac{Y-\mu}{\sigma} \leq a\right]\right) \\
&= 2\left(1-\Phi(a)\right). \quad \square
\end{aligned}$$

(d) Een bijzonder gevolg van (c) krijgen we door $a = 2$ te kiezen:

$$\mathbb{P}[|Y-\mu| > 2\sigma] = 2(1-\Phi(2)) \approx 0{,}046 < 5\%).$$



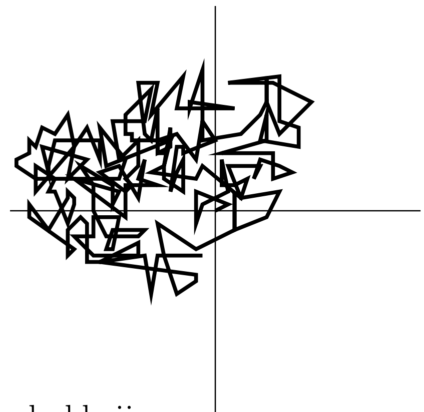
We zien: elke normaal verdeelde stochast voldoet aan de 2σ-regel.
Ga zelf na dat voor normaal verdeelde stochasten afwijkingen van $\geq 1\sigma$ heel "normaal" zijn (kans $\approx$0,318), en dat afwijkingen van $\geq 3\sigma$ uiterst weinig voorkomen (kans $\approx$ 0,0026).

Praktijk.

7.6 * Brownse beweging in $\mathbb{R}^2$.

Microscopisch kleine deeltjes zwevend in een vloeistof zijn in een rusteloze chaotische beweging. Zoals we nu weten (de botanicus Brown dacht dat al in 1828) is de beweging van zo'n deeltje het gevolg van de stoten die het krijgt van de omringende vloeistofmoleculen die er onophoudelijk van alle kanten tegenaan botsen.

Een deeltje dat op tijdstip 0 in de oorsprong is, bevindt zich na 1 tijdseenheid in een willekeurig punt (X, Y). Als men nu de volgende vooronderstellingen maakt:



(a) X en Y zijn onafhankelijk;

(b) (X, Y) is continu verdeeld met een dichtheid f die constant is op iedere cirkel om de oorsprong;

dan kan men bewijzen dat X en Y beide normaal verdeeld zijn.

7.7 Toepassing van de Centrale Limietstelling.

Door de stelling in paragraaf 7.2 wat anders te formuleren vinden we:

> Als een stochast X te schrijven is als som van een groot aantal onafhankelijke stochasten die allen dezelfde verdeling hebben, dan is X bij benadering normaal verdeeld, dus (met $\mu = \mathbb{E}(X)$ en $\sigma^2 = \mathbb{V}\mathrm{ar}(X)$):
>
> $$\mathbb{P}[X \leq a] \approx \Phi\left(\tfrac{a-\mu}{\sigma}\right)$$

Deze uitdrukking is de zogenaamde *normale benadering* voor $\mathbb{P}[X \leq a]$. In veel gevallen blijkt de benadering al bij de som van een klein aantal termen vrij goed.

Continuïteitscorrectie. Als X alleen gehele waarden aanneemt, an zou $\mathbb{P}[X = k]$ in de normale benadering 0 worden, voor elke k. We redden de situatie door $\mathbb{P}[X = k]$ te schrijven als

$$\mathbb{P}[k - \tfrac{1}{2} < X \leq k + \tfrac{1}{2}];$$

deze kans heeft als normale benadering

$$\Phi\left(\frac{k + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{k - \frac{1}{2} - \mu}{\sigma}\right)$$

(als $\mathbb{E}(X) = \mu$, $\mathbb{V}\mathrm{ar}(X) = \sigma^2$). Dientengevolge vervangen we ook

$$\begin{aligned}
&\mathbb{P}[X < k] \quad \text{door} \quad \mathbb{P}[X \leq k - \tfrac{1}{2}];\\
&\mathbb{P}[X \leq k] \quad \text{door} \quad \mathbb{P}[X \leq k + \tfrac{1}{2}];\\
&\mathbb{P}[X > k] \quad \text{door} \quad \mathbb{P}[X \geq k + \tfrac{1}{2}];\\
&\mathbb{P}[X \geq k] \quad \text{door} \quad \mathbb{P}[X \geq k - \tfrac{1}{2}].
\end{aligned}$$

Als X alleen waarden in $\{\ldots, -\frac{2}{7}, -\frac{1}{7}, 0, \frac{1}{7}, \frac{2}{7}, \ldots\}$ aanneemt, moet je natuurlijk ook een continuïteitscorrectie gebruiken. Nu niet met $\frac{1}{2}$, maar met $\frac{1}{14}$.

7.8 Voorbeelden bij 7.7.

(a) Zij X het totaal aantal ogen in n worpen met een dobbelsteen. Dan is

$$\mathbb{P}[X \leq k] \approx \Phi\left(\frac{k + \frac{1}{2} - 3\frac{1}{2}n}{\sqrt{\frac{35}{12}n}}\right);$$

al bij kleine n blijk je mooie resultaten te krijgen.

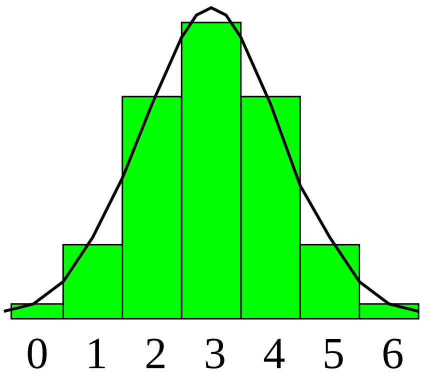
(b) Als $X \sim \mathrm{Bin}(n, p)$, dan kunnen we X schrijven als $X = X_1 + \ldots + X_n$, waarin de X_k's onafhankelijk zijn en dezelfde (alternatieve) verdeling hebben. 7.7 geeft dus voor "grote" n:

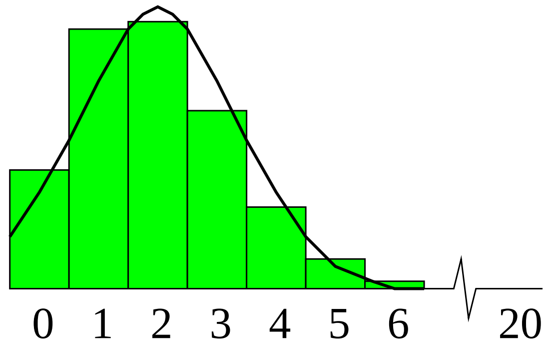
$$\mathbb{P}[X \leq k] \approx \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Hóé groot n moet zijn hangt ook van p af. Als p dicht bij $\frac{1}{2}$ ligt, is X zelf ook al nagenoeg symmetrisch verdeeld; X zal dan bij toenemende n veel eerder de "klok"vorm gaan benaderen dan wanneer p dicht bij 0 of 1 ligt: dan is de verdeling van X veel schever. Ter illustratie tekenen we de exacte $\mathrm{Bin}(n, p)$ kansen en de benaderde $\mathrm{N}(np, np(1-p))$ dichtheid voor $n = 6$, $p = \frac{1}{2}$ en voor $n = 20$, $p = \frac{1}{10}$. Oordeel zelf.

7.9 Een uitgewerkt voorbeeld.

Een stadion heeft 65.000 plaatsen. Omdat gemiddeld 2% van de mensen die een kaartje gekocht hebben, niet komt opdagen, wordt besloten in totaal n kaartjes te





verkopen ($n > 65.000$). Men wil echter de kans dat er meer dan 65.000 mensen komen, kleiner dan 0,05 houden. Hoe groot mag n maximaal zijn?

Oplossing. Zij X het aantal kaartjeskopers dat daadwerkelijk komt. Dan is $X \sim \mathrm{Bin}(n; 0{,}98)$. Er moet gelden:

$$\mathbb{P}[X > 65.000] < 0{,}05.$$

Wel,

$$\begin{aligned}\mathbb{P}[X > 65.000] &\approx 1 - \Phi\left(\frac{65.000 + \frac{1}{2} - 0{,}98n}{\sqrt{0{,}02 \cdot 0{,}98 \cdot n}}\right) \\ &= 1 - \Phi\left(\frac{65.000 + \frac{1}{2} - 0{,}98n}{0{,}14\sqrt{n}}\right) < 0{,}05;\end{aligned}$$

uit de tabel halen we dat moet gelden

$$\frac{65.000 + \frac{1}{2} - 0{,}98n}{0{,}14\sqrt{n}} > 1{,}65$$

waaruit volgt

$$n \leq 66.266.$$

Opgaven

1. Zij X het aantal kop in vier muntworpen.

   (a) Bepaal de exacte waarde van $\mathbb{P}[X < 3]$.

   (b) Bepaal de normale benadering van $\mathbb{P}[X < 3]$.

2. In 12 worpen met een dobbelsteen zij X het aantal zessen en Y het totaal aantal ogen. Benader de normale benaderingen voor de kansen

   (a) $\mathbb{P}[X \geq 5]$;

   (b) $\mathbb{P}[|Y - 42| \geq 10]$.

3. Een test bestaat uit 30 vragen. Bij elke vraag moet men kiezen uit drie antwoorden, waarvan er één het goede antwoord is. Stel dat de proefpersoon zuiver op de gok bij iedere vraag een antwoord kiest. X is het aantal goede antwoorden.

   (a) Benader $\mathbb{P}[X > 15]$;

   (b) Bepaal het kleinste natuurlijke getal m waarvoor $\mathbb{P}[X > m] < 0,01$ (bij m goede antwoorden is men geslaagd).

4. $X_1, \ldots, X_{10}$ zijn onafhankelijke, normaal verdeelde stochasten, bv. 10 herhalingen van een bepaalde meting.
   Stel $\mathbb{E}(X_i) = \mu = 3$ en $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 = 1$. Bepaal de volgende kansen (n.b. $\Phi(3{,}16) \approx 0{,}9992$):

   (a) $\mathbb{P}[2{,}1 < X_1 < 4{,}0]$;

   (b) $\mathbb{P}[2{,}1 < \overline{X} < 4{,}0]$ (waar $\overline{X} := \frac{1}{10}(X_1 + \ldots + X_{10})$).

5. Verkeersongelukken.

   (a) Laat het aantal verkeersongelukken X in 1 jaar te N. een Poissonverdeling hebben met $\lambda = 400$. Hoe groot is $\mathbb{P}[X > 440]$?

   (b) Laat X_1 en X_2 de aantallen in twee opeenvolgende jaren zijn; stel X_1, X_2 beide Poisson(400) en onafhankelijk. Hoe groot is $\mathbb{P}[|X_1 - X_2| \geq 50]$?

6. Laat de lengte X van de Nederlandse man normaal verdeeld zijn met gemiddelde $\mu = 172$ cm en standaarddeviatie $\sigma = 6$ cm. Hoeveel procent van de nederlandse mannen is

(a) langer dan 180 cm? (antwoord 9,2%)

(b) korter dan 160 cm? (antwoord 2,3%)

7. In n worpen met een munt krijgen we S_n keer kop. Zij $Z_n = \frac{S_n}{n}$. In hoofdstuk 6, opgave 4, vonden we dat de ongelijkheid van Chebyshev garandeert dat voor $n \geq 125.000$ geldt dat $\mathbb{P}[|Z_n - \frac{1}{2}| \geq 0,01] \leq 0,02$. Laat zien dat dit met de normale benadering al bij veel kleinere n bereikt wordt.

8. Stel dat in een land de lengte X van mannen normaal verdeeld is, met $\mu_X =$ 180 cm, $\sigma_X = 6$ cm, terwijl de lengte Y van vrouwen normaal verdeeld is met $\mu_Y = 173$ cm, $\sigma_Y = 5$ cm. Hoe groot is de kans dat een willekeurig gekozen man langer is dan een willekeurig gekozen vrouw? (Het antwoord op deze vraag geeft aan bij hoeveel % van de echtparen men kan verwachten dat de man langer is dan de vrouw, als tenminste de lengte bij partnerkeuze geen rol zou spelen).

9. Laat X en Y onafhankelijk zijn, beide standaardnormaal verdeeld. Bepaal de verdelingsfunctie van $U = \frac{X}{Y}$.

10. Twee theaters, A en B, strijden om de gunst van 1600 bezoekers. Veronderstel dat iedere bezoeker, totaal willekeurig en onafhankelijk van alle andere, één der theaters kiest, elk met kans $\frac{1}{2}$. Hoeveel zitplaatsen moet theater A minimaal hebben opdat de kans dat in dit theater minstens één bezoeker geweigerd zal moeten worden wegens plaatsgebrek, minder is dan 1% ?

11. Een verzekeringsmaatschappij betaalt per schadegeval een vast bedrag van 1000 gulden. Het aantal verzekerden is 10.000. Voor iedere verzekerde is de kans op zo'n schadegeval in 1 jaar 0,005. De kans op meer dan éé́n schadegeval in 1 jaar voor één verzekerde is verwaarloosbaar. Hoe hoog moet de jaarpremie zijn opdat de kans dat de maatschappij in een jaar verlies zou lijden, kleiner zij dan 0,001? (Φ^{-1}(0,999) = 3,08).

12. Twee artikelen, I en II, van verschillende merken, hebben levensduur T_1 resp. T_2. T_1 en T_2 zijn onafhankelijk en exponentieel verdeeld met parameter λ_1 resp. λ_2. Stel $\lambda_1 > \lambda_2 > 0$.

    (a) De verwachte levensduur van een artikel van merk I is korter dan de verwachte levensduur van een artikel van merk II. Laat zien dat de kans dat een artikel van merk I tòch langer meegaat, $\frac{\lambda_2}{\lambda_1+\lambda_2}$ is.

(b) Stel $\lambda_1 = 1$. De consumentenbond gaat een test uitvoeren waarbij 100 keer de levensduur van artikel I en die van artikel II worden vergeleken. De fabrikant van artikel II besluit daarom zijn product te verbeteren (λ_2 te veranderen), want hij wil de kans dat zijn artikel vaker dan 20 $\times$ verliest, kleiner maken dan 0,01. Door welke waarden van λ_2 wordt dit gegarandeerd?

13. * Om te bewijzen dat $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}dx = \sqrt{2\pi}$ doen we het volgende. Zij $f(x) = ce^{-\frac{1}{2}x^2}$. Te bepalen c zó dat f een dichtheid wordt, d.w.z. $\int_{-\infty}^{\infty} f(x)dx = 1$. Neem twee onafhankelijke stochasten X en Y, beide met dichtheid f.

(a) Bepaal de simultane dichtheid $f_{X,Y}$.

Zij $Z = \sqrt{X^2 + Y^2}$.

(b) Druk F_Z uit in $f_{X,Y}$.

We schakelen nu over op poolcoördinaten. Zij $C = \{(x, y) : x^2 + y^2 \leq z^2\}$.

(c) Gebruik dat

$$\int\int_C e^{-\frac{1}{2}(x^2+y^2)}dxdy = \int_0^z \int_0^{2\pi} e^{-\frac{1}{2}r^2} \cdot rd\varphi dr$$

om te bewijzen dat

$$F_z(z) = 2\pi c^2(1 - e^{-\frac{1}{2}z^2})$$

(voor $z > 0$).

(d) Bereken nu c. Aanwijzing: Bedenk dat

$$\lim_{n\to\infty} F_Z(z) = 1.$$

Hoofdstuk 8

Rekenen met voorwaardelijke kansen

8.1 Een nuttige eigenschap. We gaan het begrip voorwaardelijke kans, dat we aan het begin van hoofdstuk 3 hebben ingevoerd, aan een nader onderzoek onderwerpen. De fundamentele eigenschap van de voorwaardelijke kans is:

(a) $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$. Zie paragraaf 3.2.

We breiden deze eigenschap nu uit:

(b) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A \cap B)$

$$\text{BEWIJS: } \mathbb{P}(A \cap B \cap C) = \mathbb{P}((A \cap B) \cap C) = \\ = \mathbb{P}(A \cap B) \cdot \mathbb{P}(C|A \cap B) = \\ = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A \cap B). \quad \square$$

Met behulp van volledige inductie bewijs je eenvoudig dat

(c) $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$.

Toepassing. Men neemt blindelings 3 kaarten uit het kaartspel. De kans dat het drie ♠ kaarten zijn, is $\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$. Hierin stelt bijv. de factor $\frac{11}{50}$ de voorwaardelijke kans voor dat de 3e kaart ♠ is als de beide voorgaande ook ♠ kaarten waren.
In hoofdstuk 1 hebben we de kans op 3 ♠ uitgerekend volgens $\frac{\binom{13}{3}}{\binom{52}{3}}$. Dit leidt tot dezelfde uitkomst. In paragraaf 8.3 gaan we hier verder op in.

8.2 Als $\{A_1, A_2, \ldots\}$ een partitie (van Ω) is, dan is

$$\boxed{\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B).}$$

en dus is ook

$$\boxed{\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\mathbb{P}(B|A_i).}$$

Vergeet in de tweede formule de factor $\mathbb{P}(A_i)$ niet! Vooral deze tweede formule is van groot praktisch nut om vraagstukken op te lossen.

Voorbeeld. 0,01 % van de bevolking van een land lijdt aan TBC. Een test geeft in 90 % van de gevallen waarin inderdaad sprake is van TBC, een positieve uitslag. Heeft een persoon geen TBC, dan is er toch nog 1 % kans dat de test positief uitvalt. Wat is de kans op een positieve uitslag?

Voorbeeld. Op de hele wereld zijn er maar twee fabrieken die gorgels maken. 20 % van de gorgels die van fabriek I komen en 5 % van die van fabriek II zijn defect. Fabriek I produceert elke week twee keer zoveel gorgels als fabriek II.

Wat is de kans dat een gorgel, willekeurig gekozen, niet defect is?

Dat hangt natuurlijk af van de fabriek, waarin hij is gemaakt. Als B de gebeurtenis [niet defect] is, en A de gebeurtenis [gemaakt in fabriek I], dan is

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c) \\
&= \tfrac{2}{3} \cdot \tfrac{4}{5} + \tfrac{1}{3} \cdot \tfrac{19}{20} \\
&= \tfrac{51}{60}.
\end{aligned}$$

8.3 Kaarten uit een kaartspel.

Uit een spel kaarten worden 2 kaarten getrokken, zonder teruglegging. We willen de kans op 2 ♡ berekenen. Dat kan op twee manieren:

(a) We gebruiken de kansdefinitie van Laplace ofwel de hypergeometrische verdeling:

$$\begin{aligned}
\mathbb{P}[\heartsuit\heartsuit] &= \frac{\text{Aantal paren } \heartsuit}{\text{Totaal aantal paren kaarten}} \\
&= \frac{\binom{13}{2}}{\binom{52}{2}} \\
&= \tfrac{13 \cdot 12}{2} \cdot \tfrac{2}{52 \cdot 51} \\
&= \tfrac{13}{52} \cdot \tfrac{12}{51}.
\end{aligned}$$

(b) We gebruiken voorwaardelijke kansen:

$$\begin{aligned} \mathbb{P}[\heartsuit\heartsuit] &= \mathbb{P}[\text{1e kaart } \heartsuit] \cdot \mathbb{P}(\text{2e kaart } \heartsuit | \text{1e kaart } \heartsuit) \\ &= \tfrac{13}{52} \cdot \tfrac{12}{51}. \end{aligned}$$

Methode (a) gaat er van uit dat elk paar kaarten dezelfde kans heeft om getrokken te worden. Methode (b) heeft als uitgangspunt, dat alle nog niet getrokken kaarten dezelfde kans hebben om als eerstvolgende getrokken te worden.
De berekening hierboven laat zien, dat (in dit geval) methode (a) en (b) hetzelfde resultaat hebben!
Dit geldt meer algemeen:
Als we uit een vaas met N knikkers, waarvan er r rood en w wit zijn ($r+w=N$), n knikkers nemen ($n \le N$), en de kans op k rode ($k \le n$ en $k \le r$) willen berekenen, dan is die kans gelijk aan

$$\begin{aligned} &\frac{\binom{r}{k} \cdot \binom{w}{n-k}}{\binom{N}{n}} \\ = &\frac{\overbrace{r \cdot (r-1) \cdots (r-k+1)}^{k \text{ factoren}}}{k!} \times \frac{\overbrace{w \cdot (w-1) \cdots (w-n+k+1)}^{n-k \text{ factoren}}}{(n-k)!} \times \frac{n!}{\underbrace{N \cdot (N-1) \cdots (N-n+1)}_{n \text{ factoren}}} \\ = &\binom{n}{k} \times \frac{r}{N} \cdot \frac{r-1}{N-1} \cdots \frac{r-k+1}{N-k+1} \times \frac{w}{N-k} \cdots \frac{w-n+k+1}{N-n+1}. \end{aligned}$$

De aanpak met behulp van voorwaardelijke kansen levert nog een extra gegeven op. Kijk maar in het voorbeeld: [1e kaart $\heartsuit$] en [2e kaart $\heartsuit$] zijn afhankelijk, want $\mathbb{P}$[2e kaart $\heartsuit$|1e kaart $\heartsuit$] $\neq$ $\mathbb{P}$[2e kaart $\heartsuit$| 1e kaart geen $\heartsuit$]. Het resultaat van de eerste kaart beïnvloedt het resultaat van de tweede kaart.

8.4 De formule van Bayes

Stel er doen zich een aantal mogelijkheden voor, beschreven door de gebeurtenissen $A_1, A_2, \ldots$ Dus: $\{A_1, A_2, \ldots\}$ is een partitie van Ω. Neem aan: we kennen de kansen $\mathbb{P}(A_i)$ op elk van deze mogelijkheden; we noemen $\mathbb{P}(A_i)$ de "a priori"-kans op A_i.
Neem aan dat we ook de voorwaardelijke kansen $\mathbb{P}(B|A_i)$ op een zekere gebeurtenis B kennen.
Nu wordt ons verteld, dat B inderdaad optreedt. Door dit gegeven veranderen de kansen op de A_i's in de zogenaamde "a posteriori" kansen $\mathbb{P}(A_i|B)$, die te berekenen zijn met behulp van de **formule van Bayes**:

$$\boxed{\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i) \cdot \mathbb{P}(B|A_i)}{\sum_j \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j)}}$$

De formule van Bayes vertelt je dus, hoe kansen veranderen door het verkrijgen van extra informatie.

Voorbeeld. We keren terug naar het TBC-voorbeeld uit paragraaf 8.2. Stel je voor: Je ondergaat de desbetreffende test. Vóórdat de uitslag bekend is, is de kans op TBC 0,01 %. Nu blijkt de uitslag positief te zijn. Bereken met behulp van de formule van Bayes de (a posteriori) kans, dat je inderdaad drager van TBC bent.

Toepassing. Drie collegezalen zitten boordevol studenten. Student S bevindt zich in één van de drie (a priori: kans $\frac{1}{3}$ voor elke zaal). De voorwaardelijke kans dat we S zien bij het vluchtig doorkijken van zaal i , gegeven dat S in zaal i is, noemen we α_i, voor $i = 1, 2, 3$ ($\alpha_i < 1$ is toegestaan). Tijdens een vluchtig bezoek aan zaal 1 treffen we S niet aan. Wat is nù de kans dat S zich toch in zaal 1 bevindt?

Oplossing. Zij A_i de mogelijkheid [S is in zaal i], $i = 1, 2, 3$, en B het feitelijk gegeven [we zien S niet tijdens een vluchtig bezoek aan zaal 1]. Gevraagd wordt: $\mathbb{P}(A_1|B)$. De formule van Bayes geeft (controleer!)

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1) \cdot \mathbb{P}(B|A_1)}{\sum_{i=1}^{3} \mathbb{P}(A_i) \cdot \mathbb{P}(B|A_i)} = \frac{1-\alpha_1}{3-\alpha_1}.$$

8.5 De eerste zes.

Opgave 9 van hoofdstuk 2 werd door Huygens als volgt opgelost. Laat q de kans zijn dat Jan wint. Dit kan direct bij de eerste worp gebeuren, of later. We definiëren:

$$\begin{aligned} A_i &= [i^{\text{e}} \text{ worp levert een zes op}],\ i = 1, 2, 3; \\ B &= [\text{Jan wint}], \end{aligned}$$

en vinden:

$$\begin{aligned} q &= \mathbb{P}(B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_1^c \cap B) = \mathbb{P}(A_1) + \mathbb{P}(A_1^c \cap A_2^c \cap B) = \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_1^c \cap A_2^c) \cdot \mathbb{P}(B|A_1^c \cap A_2^c) = \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_1^c) \cdot \mathbb{P}(A_2^c) \cdot \mathbb{P}(B). \end{aligned}$$

Toelichting: de voorwaardelijke kans $\mathbb{P}(B|A_1^c \cap A_2^c)$ is gewoon $\mathbb{P}(B)$, want onder de gegeven voorwaarde (1$^{\text{e}}$ en 2$^{\text{e}}$ worp geen zes) zijn we weer terug in de begintoestand. Dus $q = \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot q$. Hieruit volgt: $q = \frac{6}{11}$.

Opmerking. Op dezelfde manier kunnen we aantonen dat de kans dat ooit een zes optreedt in een rij worpen met een dobbelsteen, 1 is. Stel deze kans nl. q. Dan geldt: $q = \frac{1}{6} + \frac{5}{6}q$. Hieruit volgt $q = 1$.

8.6 Het machine-voorbeeld.

In paragraaf 1.9 rekenden we de kans uit dat het gevonden slijtage-resultaat optreedt onder de hypothese dat de kwaliteit van beide merken hetzelfde is (dus dat kwaliteit en merk onafhankelijk zijn). Ook deze kans kunnen we opvatten als een voorwaardelijke kans. Immers zij X het aantal A-machines en Y het aantal B-machines dat na twee jaar versleten is.

Dan is $X \sim \text{Bin}(4, p_1)$ en $Y \sim \text{Bin}(6, p_2)$ voor zekere succeskansen p_1 en p_2. X en Y zijn onafhankelijk: het al of niet kapotgaan van de ene machine staat los van dat van de andere. Het gaat om de voorwaardelijke kans

$$\mathbb{P}(X \leq 1 | X + Y = 5).$$

Nemen we nu aan dat de gestelde hypothese geldt, dus dat $p_2 = p_1$, dan is $X + Y \sim \text{Bin}(10, p_1)$ (opgave 11 van hoofdstuk 3). Nu volgt:

$$\begin{aligned}
\mathbb{P}(X = 1 | X + Y = 5) &= \frac{\mathbb{P}[X = 1, Y = 4]}{\mathbb{P}[X + Y = 5]} = \\
&= \frac{\mathbb{P}[X = 1] \cdot \mathbb{P}[Y = 4]}{\mathbb{P}[X + Y = 5]} = \\
&= \frac{\binom{4}{1} p_1^1 (1 - p_1)^3 \cdot \binom{6}{4} p_1^4 (1 - p_1)^2}{\binom{10}{5} p_1^5 (1 - p_1)^5} = \\
&= \frac{\binom{4}{1}\binom{6}{4}}{\binom{10}{5}} = \frac{5}{21}
\end{aligned}$$

en dat is precies de (in hoofdstuk 1 geponeerde) hypergeometrische kans. Net zo bereken je $\mathbb{P}(X = 0 | X + Y = 5)$.

Algemener geldt: als $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, en X en Y zijn onafhankelijk, dan is

$$\mathbb{P}[X = k | X + Y = j] = \frac{\binom{n}{k} \cdot \binom{m}{j-k}}{\binom{n+m}{j}}.$$

Opgaven

1. Het weer in een weekeinde. Zij $A =$ [mooi weer op zaterdag], $B =$ [mooi weer op zondag]. Stel dat $\mathbb{P}(A) = \mathbb{P}(B) = 0,4$ en $\mathbb{P}(A \cap B) = 0,3$. Bepaal $\mathbb{P}(B|A)$, $\mathbb{P}(B^c|A)$, $\mathbb{P}(B|A^c)$ en $\mathbb{P}(B^c|A^c)$.

2. In opgave 1 blijkt $\mathbb{P}(B|A) > \mathbb{P}(B)$ (er is een positieve afhankelijkheid: als het eenmaal mooi weer is, dan is het waarschijnlijker dat het de volgende dag ook nog zo is). Laat zien dat in het algemeen uit $\mathbb{P}(B|A) > \mathbb{P}(B)$ volgt dat ook $\mathbb{P}(B^c|A^c) > \mathbb{P}(B^c)$. (neem aan: $0 < \mathbb{P}(A) < 1$).

3. In een zak zitten drie kaarten. Eén ervan is aan beide kanten wit. De tweede is aan beide kanten rood en de derde heeft een witte en een rode kant. Men pakt blindelings een kaart en legt die op tafel zó dat we alleen de bovenkant zien. Die blijkt wit te zijn. Hoe groot is de kans dat de onderkant ook wit is? Aanwijzing: het antwoord $\frac{1}{2}$ is niet goed.

4. Nog een keer het probleem van de eerste zes. Jan en Piet gooien om de beurt een dobbelsteen. Jan begint. Winnaar is degene die als eerste een zes gooit. We zeggen dat de beslissing in de eerste ronde valt als bij de eerste of tweede worp een zes verschijnt. De beslissing valt in de tweede ronde als de eerste zes pas bij de derde of vierde worp optreedt, enzovoort.
Zij $A_k =$ [beslissing in k^{e} ronde], $B =$ [Jan wint].
Bepaal eerst $\mathbb{P}(B|A_k)$ en bereken dan $\mathbb{P}(B)$ m.b.v. de tweede formule van paragraaf 8.2. Merk op dat het in dit geval niet nodig is de afzonderlijke kansen $\mathbb{P}(A_k)$ te kennen. Het is genoeg dat we weten dat $\sum_k \mathbb{P}(A_k) = 1$.

5. We doen een rij onafhankelijke proeven waarbij steeds de kans op succes p is. Stel p_n de kans op een oneven aantal successen in de eerste n proeven; $q_n = 1 - p_n$. Bewijs:

   (a) $p_{n+1} = q \cdot p_n + p \cdot q_n$;

   (b) $p_n = \frac{1}{2} - \frac{1}{2}(q - p)^n$.

6. In een rij onafhankelijke spelen kan men bij ieder spel een gulden winnen of een gulden verliezen, beide met kans $\frac{1}{2}$. Laat, voor $n \geq 1$, p_n de kans zijn dat we op den duur al ons geld verliezen als we beginnen met n gulden op zak. Bewijs:

   (a) $p_n = \frac{1}{2}p_{n+1} + \frac{1}{2}p_{n-1}$ $(n \geq 2)$; $p_1 = \frac{1}{2}p_2 + \frac{1}{2}$.

(b) $p_n = 1$ voor iedere n.

7. Hoe moeten we 10 witte en 10 zwarte knikkers over 2 vazen verdelen als we willen dat, wanneer we op goed geluk een vaas pakken (kans $\frac{1}{2}$ voor beide vazen) en uit die vaas blindelings een knikker, de kans dat die knikker wit is, zo groot mogelijk is?

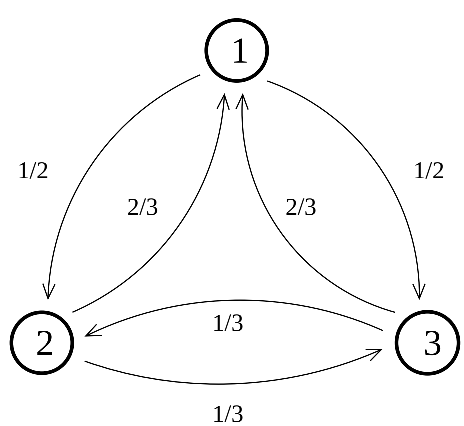
8. Doos 1 bevat 1 witte en 9 zwarte knikkers, doos 2 bevat 3 witte en 3 zwarte knikkers, doos 3 bevat 4 witte en 0 zwarte knikkers.
   We kiezen op goed geluk één van de drie dozen, en uit die doos nemen we (zonder terugleggen) 2 knikkers. Hoe groot zijn de voorwaardelijke kansen dat we doos i gekozen hebben als die twee knikkers beide wit bleken te zijn?

9. De scheidsrechter gooit een dobbelsteen, resultaat Z ogen. De spelers A en B krijgen dan ieder Z munten en gooien die op. A krijgt daarbij X keer kop en B krijgt Y keer kop. A wint als $X > Y$, B wint als $Y > X$.

   (a) Bepaal de kans op gelijkspel.

   (b) Bepaal $\mathbb{P}[X = 6]$ en $\mathbb{P}[X = 6 \text{ en } Y = 6]$. Zijn X en Y onafhankelijk?

10. Een vlo springt heen en weer over de hoekpunten van een driehoek, die we toestanden 1, 2 en 3 zullen noemen. Uit toestand 1 springt hij met kans $\frac{1}{2}$ naar 2 en met kans $\frac{1}{2}$ naar 3; uit 2 met kans $\frac{2}{3}$ naar 1 en met kans $\frac{1}{3}$ naar 3; uit 3 met kans $\frac{2}{3}$ naar 1 en met kans $\frac{1}{3}$ naar 2. Dit zijn de zogenaamde overgangskansen. Laat X_n de toestand na n sprongen zijn. Als de begintoestand $X_0 = 1$, bepaal



dan de voorwaardelijke kansen dat de vlo na 2 sprongen in toestand i is (i=1, 2, 3) (Zoiets noemt men een Markovketen; zie ook hoofdstuk 13).

11. Als het in Nijmegen een bepaalde dag regent, dan is het de volgende dag met kans $\frac{1}{4}$ de gehele dag droog, terwijl het met kans $\frac{3}{4}$ ook de volgende dag regent. Als het een bepaalde dag droog is, dan is het ook de volgende dag droog met kans $\frac{3}{5}$, terwijl het met kans $\frac{2}{5}$ de volgende dag regent. Verder nemen we aan dat het geen enkele dag met zekerheid droog is; evenmin regent het met zekerheid een bepaalde dag.

    (a) Stel dat het vandaag regent. Wat is onder dit gegeven de voorwaardelijke kans dat het overmorgen droog is?

    (b) Zijn de gebeurtenissen [donderdag 16 april 2020 regent het] en [vrijdag 17 april 2020 regent het] onafhankelijk?

    (c) Stel nu dat de kans, dat het een bepaalde dag regent, voor alle dagen hetzelfde is. Bereken die kans.

12. In een stad rijden 70 % gele taxi's en 30 % groene taxi's. Op een donkere avond veroorzaakt een taxi een aanrijding en rijdt door. Een getuige verklaart dat een groene taxi de aanrijding veroorzaakte. In een proces, dat tegen het groene taxibedrijf wordt aangespannen, gelast de rechter na te gaan hoe goed de getuige bij matig licht een groene van een gele taxi kan onderscheiden. Een experiment wijst uit dat de getuige in 80 % van de gevallen de juiste kleur noemt. Bereken de voorwaardelijke kans dat inderdaad een groene taxi de aanrijding heeft veroorzaakt (aangenomen mag worden dat taxi's en chauffeurs van beide bedrijven gemiddeld even betrouwbaar zijn).

13. (Knock-out-systeem) Gegeven 2^n spelers, allen even sterk. Ieder van hen speelt tegen een van de anderen. Vervolgens spelen de 2^{n-1} winnaars twee-aan-twee tegen elkaar, enzovoorts, tot uiteindelijk 2 spelers overblijven en tegen elkaar spelen. Via loting wordt telkens vastgesteld, welk tweetal spelers tegen elkaar speelt. In elke partij hebben beide spelers kans $\frac{1}{2}$ op winst. A en B zijn twee van de deelnemers.

    (a) Bepaal de kans dat A in precies i partijen meespeelt (i =1, 2,. . . ,n).

    (b) Bepaal de kans p_n dat A en B elkaar ooit treffen. Aanwijzing: leid eerst een verband af tussen p_n en p_{n-1}. Bereken dan zonodig p_1, p_2, p_3,. . .

14. Je doet mee aan een TV-quiz en bent er, wonder boven wonder, in geslaagd om de finale te bereiken. De presentator houdt je nu drie dozen voor. In één van deze dozen zit een fantastische hoofdprijs; in de twee andere dozen zit een zak

aardappelen. Helaas kun je niet zien in welke doos de hoofdprijs zit. Je mag nu kiezen welke van de drie dozen je wilt hebben. Nadat je je keuze gemaakt hebt, maakt de presentator (die weet waar de hoofdprijs zit) één van de dozen die je *niet* gekozen hebt open, en haalt er een zak aardappelen uit. Hij vraagt vervolgens, of je nog van keuze wilt veranderen. Wat biedt de meeste kans op de hoofdprijs, bij je aanvankelijke keuze blijven, of wisselen? Of maakt het niet uit?

Hoofdstuk 9

Kansverdelingen in $\mathbb{R}^2$

9.1

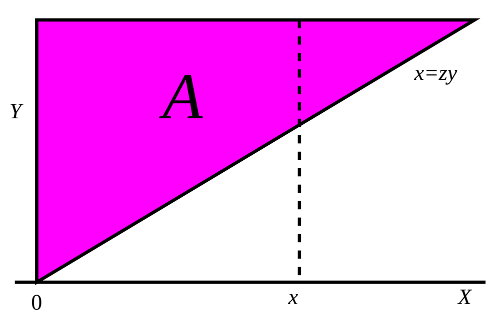
9.2 Voorbeeld met de exponentiële verdeling. Stel X en Y zijn onafhankelijke stochasten, beide exponentieel verdeeld met parameter 1. Bepaal verdelingsfunctie, dichtheidsfunctie en verwachting van $Z = \frac{X}{Y}$.

Oplossing. het is duidelijk dat $F_Z(z) = 0$ als $z \leq 0$. Zij verder $z > 0$. Omdat X en Y onafhankelijk zijn, en $f_X(x) = e^{-x}$ $(x > 0)$ en $f_Y(y) = e^{-y}$ $(y > 0)$, heeft de stochastische vector (X, Y) wegens **??**(b) dichtheid

$$f_{X,Y}(x,y) = e^{-x}e^{-y} \qquad (x, y > 0).$$

Zo volgt, met

$$A = \{(x,y) : x, y > 0 \text{ en } x \leq zy\}:$$



$$\begin{aligned}
F_Z(z) &= \mathbb{P}\left[\frac{X}{Y} \leq z\right] = \mathbb{P}[(X,Y) \in A] \\
&= \iint_A e^{-x}e^{-y}dydx \\
&= \int_0^\infty e^{-x}\left(\int_{\frac{x}{z}}^\infty e^{-y}dy\right)dx \\
&= \int_0^\infty e^{-x}e^{-\frac{x}{z}}dx = \int_0^\infty e^{-\left(1+\frac{1}{z}\right)x}dx = \frac{1}{1+\frac{1}{z}} = \frac{z}{z+1},
\end{aligned}$$

zodat

$$F_Z(z) = \begin{cases} 0 & (z < 0) \\ \frac{z}{z+1} & (z \geq 0) \end{cases} \quad \text{en } f_Z(z) = \begin{cases} 0 & (z < 0) \\ \frac{1}{(z+1)^2} & (z > 0) \end{cases}$$

Om $\mathbb{E}(Z)$ te bepalen, bereken we $\int_0^n z f_Z(z)dz$;
we vinden

$$\int_0^n \frac{z}{(z+1)^2}dz = \int_0^n \left(\frac{1}{z+1} - \frac{1}{(z+1)^2}\right)dz$$

$$
\begin{aligned}
&= \left[\log(z+1) + \frac{1}{z+1}\right]_{z=0}^{z=n} \\
&= \log(n+1) + \frac{1}{n+1} - 1,
\end{aligned}
$$

zodat

$$\mathbb{E}(Z) = \lim_{n\to\infty} \int_0^n \frac{z}{(z+1)^2} dz = \infty.$$

9.3 Sommen van normaal verdeelde stochasten.

$$\boxed{\left.\begin{array}{r} U_1 \sim \mathrm{N}(\mu_1, \sigma_1^2) \\ U_2 \sim \mathrm{N}(\mu_2, \sigma_2^2) \\ U_1 \perp\!\!\!\perp U_2 \end{array}\right\} \Longrightarrow U_1 + U_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$$

Bewijs: We mogen wel aannemen dat $\mu_1 = \mu_2 = 0$ (ga anders over op de stochasten $U_1 - \mu_1$ en $U_2 - \mu_2$).

De meest voor de hand liggende wijze is de convolutie te berekenen van de stochasten U_1 en U_2. Dat blijkt echter erg veel werk te zijn. Het volgende gaat veel mooier. Schrijf

$$S = U_1 + U_2$$

en

$$X = \frac{U_1}{\sigma_1} \qquad Y = \frac{U_2}{\sigma_2}.$$

Merk op dat X en Y standaard-normaal verdeeld zijn.

Wegens de onafhankelijkheid van X en Y is de simultane dichtheid $f_{X,Y}$ (zie **??**(b)):

$$f_{X,Y}(x,y) = \varphi(x)\varphi(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}.$$

We zien dat $f_{X,Y}$ op elke cirkel in $\mathbb{R}^2$ met de oorsprong als middelpunt een constante functie is, m.a.w. $f_{X,Y}$ is invariant voor rotaties om de oorsprong.

Nog anders gezegd: ls het gebied A door een rotatie om de oorsprong overgaat in gebied B, dan is

$$\iint_A f_{X,Y}(x,y) dy dx = \iint_B f_{X,Y}(x,y) dy dx.$$

Zij nu $s \geq 0$. Dan is

$$
\begin{aligned}
\mathbb{P}[S \leq s] &= \mathbb{P}[\sigma_1 X + \sigma_2 Y \leq s] \\
&= \mathbb{P}[(X,Y) \in A] \\
&= \iint_A f_{X,Y}(x,y) dy dx,
\end{aligned}
$$

waarin

$$A = \{(x, y) : \sigma_1 x + \sigma_2 y \leq s\}\,.$$

Door een rotatie over de hoek ϑ (zie plaatje) in de richting van de wijzers van de klok gaat A over in het halfvlak

$$B = \left\{(x, y) : \sqrt{\sigma_1^2 + \sigma_2^2}\,x \leq s\right\}$$

(ga dit na). We vinden zo:

$$\begin{aligned}
\mathbb{P}[(X, Y) \in A] &= \iint_A f_{X,Y}(x, y)dydx \\
&= \iint_B f_{X,Y}(x, y)dydx \\
&= \mathbb{P}\left[(X, Y) \in B\right] \\
&= \mathbb{P}\left[X \leq \frac{s}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right] \\
&= \mathbb{P}\left[\sqrt{\sigma_1^2 + \sigma_2^2}\,X \leq s\right]
\end{aligned}$$

Dus $S \sim \sqrt{\sigma_1^2 + \sigma_2^2}\,X$, d.w.z. $S \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$. $\square$

Opmerking. Het is nogal bijzonder dat de som van twee onafhankelijke stochasten met eenzelfde verdeling, weer eenzelfde verdeling heeft. Zoals opgemerkt in paragraaf 7.1 krijg je door het optellen van onafhankelijke stochasten (convolutie van bijbehorende dichtheidsfuncties) verdelingen met steeds mooieren klokvormige dichtheden. De normale verdeling hééft kennelijk al deze "ideale" vorm; er valt dus niets meer te verfraaien.

Convoluties

9.4 De verdeling van $X+Y$. Vaak moeten in de stochastiek stochasten opgeteld worden: wat is de kansverdeling van $X + Y$ als de kansverdelingen van X en Y bekend zijn en X en Y onafhankelijk verondersteld worden? Het blijkt niet nodig dit telkens opnieuw uit te rekenen: in deze paragraaf leiden we een formule af die voor alle continu verdeelde stochasten gebruikt kan worden.

Concreet. Stel X en Y zijn onafhankelijke stochasten, beide continu verdeeld met dichtheden f en g. Wat is dan de verdeling van $Z = X + Y$?

Antwoord: Z is ook continu verdeeld en wel met een dichtheid h die gegeven wordt door

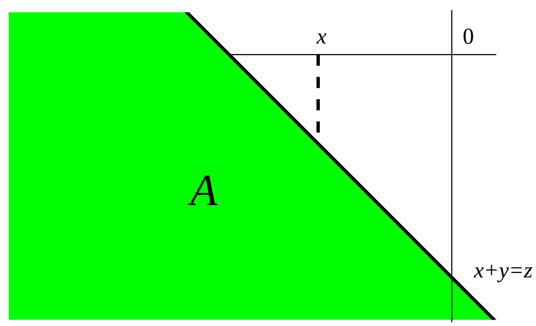
$$\boxed{h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{-\infty}^{\infty} f(z-y)g(y)dy}$$

(beide formules leveren hetzelfde resultaat).
Men noemt h de *convolutie* van f en g, notatie: $h = f * g$.
(Vergelijk dit met het discrete geval: Als X en Y onafhankelijke geheelwaardige stochasten zijn, dan is

$$\mathbb{P}[X+Y=n] = \sum_{k=0}^{n} \mathbb{P}[X{=}k] \cdot \mathbb{P}[Y{=}n{-}k].)$$

Bewijs. De verdelingsfunctie H van $Z = X + Y$ wordt gegeven door

$$H(z) = \mathbb{P}[Z \leq z] = \mathbb{P}[(X,Y) \in A]$$



met $A = \{(x,y) : x + y \leq z\}$. Met **??**(a) zien we:

$$\begin{aligned} H(z) &= \iint_A f_{X,Y}(x,y)dydx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f(x)g(y)dy \right) dx. \end{aligned}$$

We nemen nu even x vast en beschouwen de "binnenste" integraal:

$$\int_{-\infty}^{z-x} f(x)g(y)dy.$$

We vervangen (substitueren) de variabele y door de nieuwe variabele $t = x + y$, en vinden dan:

$$\int_{-\infty}^{z-x} f(x)g(y)dy = \int_{-\infty}^{z} f(x)g(t{-}x)dt.$$

Zo zien we:

$$\begin{aligned} H(z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f(x)g(t{-}x)dt \right) dx \\ &= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f(x)g(t{-}x)dx \right) dt \end{aligned}$$

en blijkt dat we de functie

$$t \mapsto \int_{-\infty}^{\infty} f(x)g(t{-}x)dx$$

kunnen nemen als dichtheid van Z. □

9.5 Sommen van uniform verdeelde stochasten.

Bij numerieke berekeningen komen vaak problemen voor als beschreven in het begin van paragraaf 9.4 (vergelijk paragraaf 6.6, voorbeeld 3). Laat $X_1, \ldots, X_n$ de afrondingsfouten zijn in de afgeronde grootheden; $Y_n = X_1 + \ldots + X_n$ is dan de fout in de som van die grootheden.
We nemen aan dat $X_1, \ldots, X_n$ onafhankelijk zijn en allemaal uniform verdeeld over $(-\frac{1}{2}, \frac{1}{2})$.
Laat f_k de dichtheidsfunctie van Y_k zijn; dan is f_1 dus de dichtheid van zowel X_1, als $X_2, \ldots, X_n$.
Omdat $Y_{k+1} = Y_k + X_{k+1}$ is (wegens paragraaf 9.4) $f_{k+1} = f_k * f_1$, dus

$$f_{k+1}(x) = \int_{-\infty}^{\infty} f_k(t) f_1(x-t) dt = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} f_k(t) dt.$$

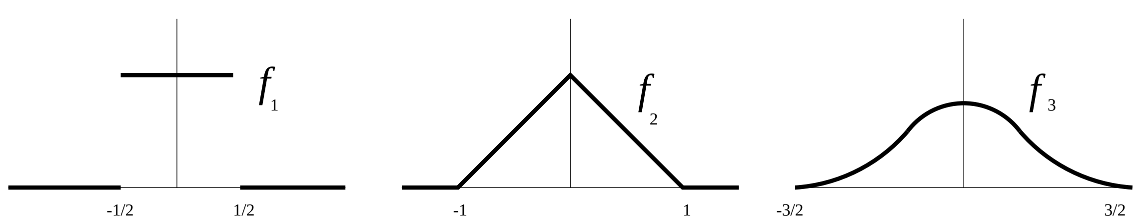
Door $k = 1$ te kiezen vinden we

$$f_2(x) = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \mathbf{1}_{(-\frac{1}{2},\frac{1}{2})}(t) dt.$$

Voor vaste x moeten we dus t's hebben met

$$x - \tfrac{1}{2} < t < x + \tfrac{1}{2} \text{ en } -\tfrac{1}{2} < t < \tfrac{1}{2}.$$

Voor $|x| > 1$ lukt dit niet, dus dan is $f_2(x) = 0$.
Voor $-1 < x < 0$ betekent het bovenstaande: $-\frac{1}{2} < t < x + \frac{1}{2}$, zodat $f_2(x) = \int_{-\frac{1}{2}}^{x+\frac{1}{2}} 1 dt = x + 1$.
Voor $0 < x < 1$ vind je $x - \frac{1}{2} < t < \frac{1}{2}$, dus $f_2(x) = \int_{x-\frac{1}{2}}^{\frac{1}{2}} 1 dt = 1 - x$.
Daarmee is f_2 gevonden.
Zo doorgaande kun je f_3 uitrekenen m.b.v. f_2, f_4 m.b.v. f_3, enzovoorts; het rekenwerk wordt wel steeds omvangrijker. We besluiten met de grafieken van f_1, f_2 en f_3 en merken op dat deze gladder worden naarmate het rangnummer toeneemt:



9.6 Sommen van exponentieel verdeelde stochasten. Aangezien mijn theepotten nogal snel sneuvelen, besluit ik er maar 12 tegelijk te kopen (dat is nog voordeliger ook). Wat is de kans dat ik binnen drie jaar opnieuw naar de winkel moet? Nummer de theepotten; zij X_n de levensduur van de n^{e} theepot (ingaande op het moment van ingebruikneming, dus als de $(n-1)^{\mathrm{e}}$ kapotvalt). We veronderstellen $X_1, \ldots, X_{12}$ onafhankelijk, en alle exponentieel verdeeld met parameter λ; zij $Y_n = X_1 + \ldots + X_n$ de totale levensduur van de eerste n potten. Noem de dichtheid van Y_n: f_n, dan is weer $f_{n+1} = f_n * f_1$ zodat

$$\begin{aligned} f_{n+1}(t) &= \int_{-\infty}^{\infty} f_n(s) f_1(t-s) ds \\ &= \int_0^t f_n(s) \lambda e^{-\lambda(t-s)} ds \\ &= e^{-\lambda t} \int_0^t f_n(s) \lambda e^{\lambda s} ds. \end{aligned}$$

Dit geeft achtereenvolgens:

$$\begin{aligned} f_2(t) &= \lambda^2 t e^{-\lambda t} \\ f_3(t) &= \frac{\lambda^3}{2} t^2 e^{-\lambda t} \end{aligned}$$

en met behulp van volledige inductie:

$$f_n(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}.$$

Dit is een van de zeldzame gevallen waarin convoluties gemakkelijk te berekenen zijn.
Sommen van exponentieel verdeelde stochasten hangen nauw samen met het Poissonproces: zie paragraaf 10.8.

Opgaven.

1. Beantwoord de vraag in de eerste alinea van paragraaf 9.4.

2. Laat (X, Y) continu verdeeld zijn met dichtheid

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{als } x^2 + y^2 < 1; \\ 0 & \text{als } x^2 + y^2 > 1. \end{cases}$$

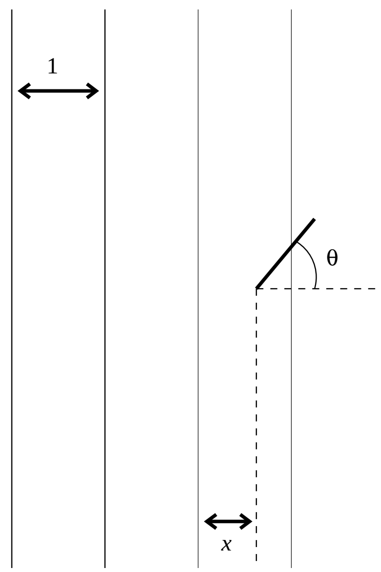
   (a) Bepaal de dichtheidsfuncties van X en Y.

   (b) Bepaal verdelings- en dichtheidsfunctie van $R = \sqrt{X^2 + Y^2}$.

(c) Zijn X en Y onafhankelijk?

3. Zij $X \sim \text{Un}(0,1)$, $Y \sim \text{Exp}(1)$, $Z = X + Y$. Stel X en Y onafhankelijk. Bepaal dichtheids- en verdelingsfunctie van Z.

4. Een munt met diameter d ($d < 1$) wordt op een schaakbord geworpen, met vierkante vakjes van 1×1. Bepaal de kans dat de munt helemaal binnen één vakje terechtkomt.



5. Een vloer bestaat uit planken van 1 meter breed. Een breinaald met lengte ℓ ($\ell < 1$) wordt op de vloer geworpen. We willen de kans bepalen dat de breinaald geheel op één plank valt. We noteren zijn positie met (X, ϑ), waarin X de afstand is van het meest linkse punt van de breinaald tot de naad er links naast (dus $X \in (0,1)$) en ϑ de hoek van de naald met de positieve X-as (dus $\vartheta \in (-\frac{1}{2}\pi, \frac{1}{2}\pi)$).
We nemen aan: $X \sim \text{Un}(0,1)$, $\vartheta \sim \text{Un}(-\frac{1}{2}\pi, \frac{1}{2}\pi)$, X en ϑ onafhankelijk.

(a) Leid af dat moet gelden: $X + \ell \cos\vartheta < 1$, wil de naald geheel op één plank vallen.

(b) Teken een (X, ϑ)-assenstelsel ($0 < X < 1$, $-\frac{1}{2}\pi < \vartheta < \frac{1}{2}\pi$) en daarin het gebied dat voldoet aan (a).

(c) Bereken nu de gevraagde kans.

6. De stochasten X en Y zijn onafhankelijk en uniform verdeeld over $(0,1)$. Verdeel $(0,1)$ in drie delen:

$$(0, \min\{X,Y\}), \quad (\min\{X,Y\}, \max\{X,Y\}), \quad (\max\{X,Y\}, 1).$$

Probeer een driehoek te vormen met zijden, congruent aan deze intervallen. Wat is de kans dat dit lukt?

7. X en Y zijn onafhankelijke, exponentieel(1) verdeelde stochasten.

   (a) Bepaal de dichtheidsfunctie van (X,Y).

   (b) Zij $Z = |X - Y|$. Bewijs dat ook $Z \sim \text{Exp}(1)$.

8. X en Y zijn onafhankelijke stochasten. $X \sim \text{Un}(0,1)$, $Y \sim \text{Un}(0,a)$, met $a > 1$.

   (a) Bereken verdelings- en dichtheidsfunctie van $Y - X$.

   (b) Bereken a als gegeven is dat $\mathbb{E}(Y - X) = \frac{19}{12}$.

9. Een mier (dikte nul) beweegt zich kriskras over een tafelblad met lengte $2n$ en breedte $2m$ $(n > m > 0)$. De plaats waar de mier zich op zeker ogenblik bevindt, is uniform verdeeld over het blad. Zij X de afstand tot de dichtstbijzijnde rand.

   (a) Bereken de verdelingsfunctie van X.

   (b) Bereken de dichtheidsfunctie van X.

   (c) Zij $n = 2m$. Bereken $\mathbb{E}(X)$ en $\mathbb{V}\text{ar}(X)$.

10. De levensduur van grote vliegtuigmotoren is exponentieel (λ) verdeeld, die van kleine exponentieel (2λ). De levensduren van verschillende motoren zijn onafhankelijk.

    (a) In vliegtuigtype A bevinden zich een grote en een kleine motor die slechts samen het vliegtuig in werking houden. Zij T_A de levensduur van het vliegtuig (reparatie of vervanging is niet mogelijk). Bereken verdelingsfunctie en verwachting van T_A.

    (b) In type B bevinden zich een grote en een kleine motor die ook elk apart het vliegtuig in werking houden. Bij de start wordt de grote motor ingeschakeld; als deze uitvalt, start automatisch de kleine motor. Bereken dichtheid en verwachting van de levensduur T_B.

(c) Type C is uitgerust met één grote en twee kleine motoren. Dit vliegtuig blijft in werking tot ofwel de grote motor uitvalt, ofwel beide kleine uitgevallen zijn. Alle drie de motoren werken vanaf de start. Bereken verdelingsfunctie en verwachting van de levensduur T_C.

Hoofdstuk 10

De Poisson-verdeling

10.1 Definitie en berekening.

Laat K het aantal klanten zijn dat in een tijdsinterval I van bijvoorbeeld een uur een winkel binnengaat. We willen de kansverdeling van K achterhalen.
Natuurlijk moeten we daartoe iets over het gedrag van de klanten aannemen. Eerst wat notatie: als $J \subset I$ een deelinterval is, dan verstaan we onder $K(J)$ het aantal klanten dat gedurende J binnenkomt. We veronderstellen:

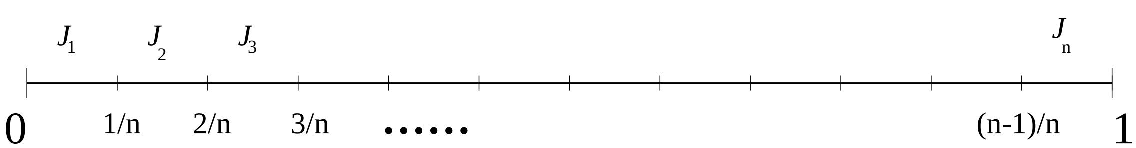
1. Als de deelintervallen J_1 en J_2 van I even lang zijn, dan hebben $K(J_1)$ en $K(J_2)$ dezelfde kansverdeling.

2. Als $J_1 \cap J_2 = \emptyset$, dan zijn $K(J_1)$ en $K(J_2)$ onafhankelijk.

3. Er komen nooit twee of meer klanten tegelijk binnen.

4. De verwachting $\lambda := \mathbb{E}(K)$ bestaat (en is eindig).

Het blijkt dat deze uitgangspunten de kansverdeling van K (en ook van $K(J)$ voor elk deelinterval J) helemaal vastleggen. Dat gaan we nu bewijzen.



Hiertoe verdelen we $I = (0, 1]$ in n gelijke stukken $J_1, \ldots, J_n$, dus $J_k = (\frac{k-1}{n}, \frac{k}{n}]$, $k = 1, \ldots, n$. Met $A_j^{(n)}$ duiden we de volgende gebeurtenis aan:

$$A_j^{(n)} := [\text{er komt minstens één klant binnen gedurende } J_j].$$

Wegens aanname 1 is de kans $\mathbb{P}\left(A_j^{(n)}\right)$ dezelfde voor alle waarden van j; we noemen deze kans: p_n.

Wegens aanname 2 zijn de gebeurtenissen $A_1^{(n)}, A_2^{(n)}, \ldots, A_n^{(n)}$ onafhankelijk. Het aantal K_n van de $A_j^{(n)}$'s die optreden (dus van de tijdsintervallen waarin er iemand binnenkomt) is dus binomiaal verdeeld met parameters n en p_n: voor alle $k \in \mathbb{N}$ is

$$\mathbb{P}[K_n = k] = \binom{n}{k} p_n^k (1 - p_n)^{n-k}. \tag{1}$$

Nu kan het aantal tijdsintervallen waarop er iemand binnenkomt natuurlijk niet groter zijn dan het totale aantal klanten dat binnenkomt: $K_n \leq K$. Maar ook volgt, uit aanname 3, dat, als er maar voldoende deelstreepjes in het interval I worden geplaatst, elke klant zijn eigen tijdsintervalletje krijgt toebedeeld:

$$\forall \omega \in \Omega : \lim_{n\to\infty} K_n(\omega) = K(\omega). \tag{2}$$

Het lijkt daarom een goed idee, in (1) de limiet $n \to \infty$ te nemen. Of nog liever: we stellen $n = 2^l$ en nemen $l \to \infty$. Dit is handig omdat (ga na!)

$$K_1 \leq K_2 \leq K_4 \leq K_8 \leq K_{16} \leq \ldots. \tag{3}$$

We duiden deze "limiet over de machten van 2" aan met $\lim^*_{n\to\infty}$. We zullen aantonen dat:

$$\begin{array}{ll} \text{(a)} & \lim^*_{n\to\infty} \mathbb{P}[K_n = k] = \mathbb{P}[K = k] \text{ voor } k \in \mathbb{N}); \\ \text{(b)} & \lim^*_{n\to\infty} np_n = \lambda. \end{array}$$

(a) bewijzen we als volgt: Voor willekeurige k geldt wegens (3) dat de gebeurtenissen $[K_n{>}k]$ met $n = 2^l$ een stijgende rij vormen: $[K_1{>}k] \subset [K_2{>}k] \subset [K_4{>}k] \subset \ldots$.
De vereniging van deze rij gebeurtenissen is (wegens (2)) gelijk aan de gebeurtenis $[K{>}k]$.
Volgens eigenschap 2.5(c) geldt dus

$$\lim^*_{n\to\infty} \mathbb{P}[K_n{>}k] = \mathbb{P}[K{>}k]. \tag{4}$$

En dus ook

$$\begin{aligned} \lim^*_{n\to\infty} \mathbb{P}[K_n{=}k] &= \lim^*_{n\to\infty} (\mathbb{P}[K_n{>}k{-}1] - \mathbb{P}[K_n{>}k]) \\ &= \mathbb{P}[K{>}k{-}1] - \mathbb{P}[K{>}k] \\ &= \mathbb{P}[K{=}k], \end{aligned}$$

waarmee (a) is aangetoond.

(b) gaat nu als volgt: Uit (3) volgt dat de verwachtingen $\mathbb{E}(K_n)$ met $n = 2^l$ een stijgende rij vormen. Deze rij is begrensd, want $K_n \leq K$, dus $\mathbb{E}(K_n) \leq \mathbb{E}(K)$ voor alle n.
De rij verwachtingen is dus convergent, en

$$\lim_{n\to\infty}^{*} \mathbb{E}(K_n) \leq \mathbb{E}(K). \tag{5}$$

Wegens $\mathbb{E}(K_n) = \sum_{k=0}^{\infty} \mathbb{P}[K_n{>}k] \geq \sum_{k=0}^{N} \mathbb{P}[K_n{>}k]$ krijgen we met (4):

$$\lim_{n\to\infty}^{*} \mathbb{E}(K_n) \geq \lim_{n\to\infty}^{*} \sum_{k=0}^{N} \mathbb{P}[K_n{>}k] = \sum_{k=0}^{N} \mathbb{P}[K{>}k].$$

Daar dit voor iedere N geldt, krijgen we:

$$\lim_{n\to\infty}^{*} \mathbb{E}(K_n) \geq \sum_{k=0}^{\infty} \mathbb{P}[K{>}k] = \mathbb{E}(K).$$

Dit gecombineerd met (5) geeft (b).

Met behulp van (a) en (b) kunnen we nu de limiet $n \to \infty$ over de machten van 2 in (1) uitvoeren:

$$\begin{aligned}
\mathbb{P}[K = k] &= \lim_{n\to\infty}^{*} \mathbb{P}[K_n = k] \\
&= \lim_{n\to\infty}^{*} \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{1}{k!} \lim_{n\to\infty}^{*} (np_n)\,((n{-}1)p_n) \cdots ((n{-}k{+}1)p_n) \cdot \frac{(1-p_n)^n}{(1-p_n)^k} \\
&= \frac{\lambda^k}{k!} e^{-\lambda},
\end{aligned} \tag{6}$$

De verdeling van K wordt de *Poisson-verdeling met parameter* λ genoemd. Als een stochast X deze kansverdeling heeft, schrijven we $X \sim \text{Poisson}(\lambda)$.

10.2 Verwachting en variantie.

Dat $\lambda = \mathbb{E}(K)$ moet achteraf natuurlijk ook uit de kansverdeling van K berekend kunnen worden. Dat gaat zó:

$$\begin{aligned}
\mathbb{E}(K) &= \sum_{k=0}^{\infty} k\mathbb{P}[K = k] \\
&= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda.
\end{aligned}$$

Berekening van $\mathbb{V}\mathrm{ar}(K)$ gaat het gemakkelijkst via dezelfde methode: we berekenen eerst

$$\begin{aligned}
\mathbb{E}(K(K-1)) &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} \\
&= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} \\
&= \lambda^2 e^{-\lambda}e^{\lambda} \\
&= \lambda^2.
\end{aligned}$$

Vervolgens: $\mathbb{V}\mathrm{ar}(K) = \mathbb{E}(K^2) - \mathbb{E}(K)^2 = \mathbb{E}(K(K-1)) + \mathbb{E}(K) - \mathbb{E}(K)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. Dus is ook

$$\boxed{\mathbb{V}\mathrm{ar}(K) = \lambda.}$$

10.3 Verband met de binomiale verdeling.

De laatste drie stappen uit berekening (6) zijn ook geldig voor binomiaal (n, p_n) verdeelde stochasten K_n mits

$$\lim_{n\to\infty} np_n = \lambda \tag{7}$$

(in dat geval hoeft de limiet ook niet meer over de machten van 2 genomen te worden).
Hier wordt triviaal aan voldaan door: $p_n = \frac{\lambda}{n}$.
Dus als de rij stochasten $X_1, X_2, X_3, \ldots$ voldoet aan: $X_n \sim \mathrm{Bin}(n, \frac{\lambda}{n})$, dan is

$$\lim_{n\to\infty} \mathbb{P}[X_n = k] = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{8}$$

Men gebruikt dit resultaat vaak om voor kansen uit een binomiale verdeling met grote n en kleine p en benadering te vinden door niet de binomiale maar de "overeenkomstige" Poisson-kans te berekenen; dat scheelt een parameter en het is vaak wat minder lastig rekenwerk. Zie opgave 2. Het resultaat werd gevonden door Siméon Poisson, al in 1837, en was de reden dat zijn naam aan deze kansverdeling gekoppeld werd. Pas veel later zag men de zelfstandige betekenis van deze verdeling, die we in paragraaf 10.1 hebben gezien.
In (8) staat dat een binomiale verdeling met grote n en kleine p sterk lijkt op een Poisson-verdeling met parameter $\lambda = np$. Hun verwachtingen zijn gelijk, en hun varianties bijna gelijk (npq resp. $\lambda = np$; bedenk dat $q = 1 - p$ bijna 1 is). Omdat de Poisson-verdeling te voorschijn komt als limiet van een binomiale verdeling

met aanzwellende n (het aantal proeven) en krimpende succeskans p, noemt men de Poisson-verdeling wel de verdeling der "zeldzame gebeurtenissen". Deze benaming is eigenlijk misleidend. De succeskans p mag dan klein zijn, het verwachte aantal successen $np = \lambda$ hoeft helemaal niet klein te zijn. Een belangrijk *verschil* tussen de binomiale verdeling en de Poisson-verdeling is dat de uitkomstenruimte bij de binomiale verdeling *eindig* is ($\{0, 1, \ldots, n\}$) en bij de Poisson-verdeling *oneindig* ($\mathbb{N}$).

10.4 Zeldzame gebeurtenissen en paardebloemen.

Er zijn in het dagelijks leven veel voorbeelden van stochasten die kunnen worden opgevat als het aantal successen in een lange reeks onafhankelijke pogingen met allemaal dezelfde, kleine succeskans. Zulke stochasten hebben in benadering een Poisson-verdeling. We noemen ze "zeldzame gebeurtenissen".
Voorbeelden:

(a) Het aantal drukfouten op een bladzijde.

(b) Het aantal telefoonnummers dat in Nijmegen op één dag verkeerd wordt gedraaid.

(c) Het aantal gebarsten straatklinkers in 10 meter straat.

Nog groter is het aantal voorbeelden waarbij, zoals in 10.1, de "pogingen" niet meer goed te onderscheiden zijn, maar elk stukje tijd (of lengte, of oppervlak, of...) als "poging" kan worden opgevat, waarbij de succeskans evenredig is met de duur (of lengte of oppervlakte of...). We noemen ze "paardebloemen". Voorbeelden:

(d) Het aantal paardebloemen in 1 are weiland.

(e) Het aantal klanten dat in één uur een winkel binnenkomt.

(f) Het aantal klanten dat in één uur een winkel binnenkomt en ook iets koopt.

(g) Het aantal borelingen op één dag in Nijmegen.

(h) Het aantal α-deeltjes dat in een bepaald tijdsinterval uitgezonden wordt door een radioactief preparaat.

10.5 Eigenschappenvan de Poisson-verdeling.

(a) Opteleigenschap.

$$\left.\begin{array}{r} K_1 \sim \text{Poisson}(\lambda_1) \\ K_2 \sim \text{Poisson}(\lambda_2) \\ K_1 \perp\!\!\!\perp K_2 \end{array}\right\} \Longrightarrow K_1 + K_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

Dus de som van onafhankelijke Poisson-verdeelde stochasten is weer Poisson-verdeeld.
Dit is heel logisch als je het formuleert in de termen van 10.1:
Wanneer $K_1 = K(J_1)$ en $K_2 = K(J_2)$, terwijl $J_1 \cap J_2 = \emptyset$ (zodat $K_1 \perp\!\!\!\perp K_2$) dan is natuurlijk $K_1 + K_2 = K(J_1 \cup J_2)$, en dit is Poisson met parameters $\mathbb{E}(K(J_1 \cup J_2)) = \mathbb{E}(K_1) + \mathbb{E}(K_2) = \lambda_1 + \lambda_2$.
We geven hier echter een direct bewijs:
BEWIJS:

$$\begin{aligned}
\mathbb{P}[K_1 + K_2 = n] &= \sum_{k=0}^{n} \mathbb{P}[K_1 = k, K_2 = n-k] \\
&= \sum_{k=0}^{n} \mathbb{P}[K_1 = k]\mathbb{P}[K_2 = n-k] \\
&= \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\
&\overset{\text{Bin. v. Newton}}{=} e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1+\lambda_2)^n}{n!}.
\end{aligned}$$

□

(b) Onafhankelijkheid van de individuen

$$\left.\begin{array}{r} K_1 \sim \text{Poisson}(\lambda_1) \\ K_2 \sim \text{Poisson}(\lambda_2) \\ K_1 \perp\!\!\!\perp K_2 \end{array}\right\} \implies \mathbb{P}(K_1 = k | K_1 + K_2 = n) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\text{met } p = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \text{ alle } k, n.$$

In termen van het klantenprobleem: als gegeven is dat in het interval $J_1 \cup J_2$ in totaal n klanten binnenkomen, dan kunnen hun tijdstippen van binnenkomst als n onafhankelijke grootheden worden opgevat (elke klant maakt onafhankelijk van de andere klanten een keuze tussen J_1 en J_2); in het bijzonder zijn de n keuzes tussen J_1 en J_2 die de klanten maken, onafhankelijk van elkaar.
Het aantal k van de klanten die J_1 uitkiezen om binnen te komen is dus $\text{Bin}(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$ verdeeld.
Deze eigenschap van de Poisson-verdeling is veel minder begrijpelijk op grond van de berekening in paragraaf 10.1. Ze kan desgewenst als alternatieve karakterisering van de Poisson-verdeling worden gebruikt. Het bewijs echter is wel eenvoudig:

BEWIJS:

$$\begin{aligned}
\mathbb{P}(K_1 = k | K_1 + K_2 = n) &= \frac{\mathbb{P}[K_1 = k, K_1 + K_2 = n]}{\mathbb{P}[K_1 + K_2 = n]} \\
&= \frac{\mathbb{P}[K_1 = k, K_2 = n - k]}{\mathbb{P}[K_1 + K_2 = n]} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k}.
\end{aligned}$$

□

(c) Splitsings-eigenschap.

$$\left. \begin{array}{r} K \sim \text{Poisson}(\lambda) \\ K = K_1 + K_2 \\ \mathbb{P}(K_1 = k | K = n) = \binom{n}{k} p^k (1-p)^{n-k} \end{array} \right\} \Longrightarrow \left\{ \begin{array}{l} K_1 \sim \text{Poisson}(\lambda p) \\ K_2 \sim \text{Poisson}(\lambda(1-p)) \\ K_1 \perp\!\!\!\perp K_2 \end{array} \right.$$

Als je door loting een Poisson-verdeeld aantal individuen in twee groepen verdeelt, dan zijn de aantallen van deze groepen onafhankelijk en weer Poisson-verdeeld.
Dit is een soort omkering van (b) met gebruikmaking van (a).
In het klantenmodel is deze eigenschap weer goed te verklaren (als we de onafhankelijkheid van de klanten, (b), aannemen): zij $K = K(I)$. Op grond van (b) kunnen we de loting per klant mooi uitvoeren door eenvoudig het interval I in twee stukken te hakken: $J_1 \cup J_2 = I$ en $J_1 \cap J_2 = \emptyset$. Natuurlijk zijn nu $K_1 = K(J_1)$ en $K_2 = K(J_2)$ Poisson-verdeeld en onafhankelijk.
BEWIJS:

$$\begin{aligned}
\mathbb{P}[K_1 = k, K_2 = m] &= \mathbb{P}[K_1 + K_2 = k + m, K_1 = k] \\
&= \mathbb{P}[K_1 + K_2 = k + m] \cdot \mathbb{P}(K_1 = k | K_1 + K_2 = k + m) \\
&= \frac{\lambda^{k+m}}{(k+m)!} e^{-\lambda} \cdot \binom{k+m}{k} \cdot p^k \cdot (1-p)^m \\
&= \left(\frac{(\lambda p)^k}{k!} e^{-\lambda p} \right) \cdot \left(\frac{(\lambda(1-p))^m}{m!} e^{-\lambda(1-p)} \right).
\end{aligned}$$

Hieruit volgt:

$$\begin{aligned}
\mathbb{P}[K_1 = k] &= \sum_{m=0}^{\infty} \mathbb{P}[K_1 = k, K_2 = m] \\
&= \sum_{k=0}^{\infty} \left(\frac{(\lambda p)^k}{k!} e^{-\lambda p} \right) \cdot \left(\frac{(\lambda(1-p))^m}{m!} e^{-\lambda(1-p)} \right) \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda p},
\end{aligned}$$

dus $K_1 \sim$ Poisson(λp). Net zo volgt dat $K_2 \sim$ Poisson($\lambda(1-p)$). Tevens hebben we gevonden dat

$$\mathbb{P}[K_1 = k, K_2 = m] = \mathbb{P}[K_1 = k] \cdot \mathbb{P}[K_2 = m]$$

voor elke k, m, dus dat K_1 en K_2 onafhankelijk zijn. $\square$

Voorbeeld 1. Als het aantal klanten Poisson(λ) verdeeld is, en als iedere klant met kans p iets koopt, dan is het aantal kopers Poisson(λp) verdeeld, en het aantal niet-kopers Poisson($\lambda(1-p)$).

Voorbeeld 2. Als het aantal kinderen dat in een jaar geboren wordt Poisson(λ) verdeeld is, en bij iedere geboorte de kans op een jongen p is, dan is het aantal jongens K_1 dat geboren wordt Poisson(λp) verdeeld. Het aantal meisjes K_2 is Poisson($\lambda(1-p)$) verdeeld, en K_1 en K_2 zijn onafhankelijk.

10.6 Een toepassing.

Het aantal ongelukken K is Poisson(λ) verdeeld. Bij elk ongeluk bedraagt de schade 1, 2 of 3 eenheden, met kans p_1, p_2 resp. p_3 (en $p_1 + p_2 + p_3 = 1$). Zij S de totale schade. Bepaal $\mathbb{E}(S)$.

1e oplossing. $\mathbb{E}(S) = \sum_{k=0}^{\infty} \mathbb{E}(S|K = k)\mathbb{P}[K = k]$. Als $K = k$, kun je S schrijven als $S = Y_1 + \ldots + Y_k$, waarin Y_j de schade bij het j^{e} ongeluk voorstelt. Dan is $\mathbb{E}(Y_j) = p_1 + 2p_2 + 3p_3$ voor elke j, dus $\mathbb{E}(S|K = k) = k(p_1 + 2p_2 + 3p_3)$. We vinden dan

$$\begin{aligned}
\mathbb{E}(S) &= \sum_{k=0}^{\infty} k(p_1 + 2p_2 + 3p_3)e^{-\lambda}\frac{\lambda^k}{k!} \\
&= (p_1 + 2p_2 + 3p_3)e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda(p_1 + 2p_2 + 3p_3).
\end{aligned}$$

2e oplossing. Schrijf $K = K_1 + K_2 + K_3$, waarbij K_i het aantal ongelukken is dat schade i veroorzaakt. Dan is $S = K_1 + 2K_2 + 3K_3$. Wegens 10.5(c) en opgave 11 is elke K_i Poisson(λp_i) verdeeld, zodat $\mathbb{E}(K_i) = \lambda p_i$. Gevolg:

$$\mathbb{E}(S) = \mathbb{E}(K_1) + 2\mathbb{E}(K_2) + 3\mathbb{E}(K_3) = \lambda(p_1 + 2p_2 + 3p_3).$$

10.7 Het Poissonproces.

Over het klantenmodel uit paragraaf 10.1 valt nog meer interessants te zeggen.

Laat I eens het tijdsinterval $(0, t]$ voorstellen. Met een kleine wijziging in de oorspronkelijke opzet kunnen we best de tijd ná het tijdstip 1 dóór laten lopen. Zij

nu $K_t = K([0,t))$, het aantal klanten dat vóór tijdstip t binnenkomt. Natuurlijk is voor elke t de stochast K_t Poisson-verdeeld. Op grond van aannamen 1 en 4 leid je gemakkelijk af dat $\mathbb{E}(K_t) = \lambda t$ (Immers de functie $f : t \mapsto \mathbb{E}(K_t)$ is stijgend en additief, en voldoet aan $f(1) = \lambda$).
Dus $K_t \sim \text{Poisson}(\lambda t)$.
λ is het verwachte aantal klanten per tijdseenheid.
De familie van stochasten $(K_t)_{t\geq 0}$ wordt wel het *Poissonproces* met *intensiteit* λ genoemd.
Opmerking. Als we in het model hadden willen opnemen dat de intensiteit van het klantenbezoek gedurende de dag varieert (om half zes zal het gemiddeld drukker zijn dan om 10 uur), dan kunnen we de constante intensiteit λ vervangen door een intensiteitsfunctie $\lambda : t \mapsto \lambda(t)$. K_t heeft dan Poisson-verdeling met $\mathbb{E}(K_t) = \int_0^t \lambda(s)ds$.
Dit komt erop neer dat we aanname 1 vervangen door:

1'. Als $\int_{J_1} \lambda(t)dt = \int_{J_2} \lambda(t)dt$, dan hebben $K(J_1)$ en $K(J_2)$ dezelfde kansverdeling.

10.8 Tussenpozen. De volgende eigenschap van het Poissonproces is van groot belang in toepassingen en simulaties.

> De tussenpozen $X_1 := T_1$, $X_2 := T_2 - T_1$, $X_3 := T_3 - T_2$, $\cdots$ in het Poissonproces zijn exponentieel (λ) verdeeld en onderling onafhankelijk.

Dat de aankomsttijd T_1 van de eerste klant exponentieel (λ) verdeeld is zul je zelf bewijzen in opgave 6. de exponentiele verdeling van de overige tussenpozen, en hun onafhankelijkheid kan men bewijzen op eenzelfde manier als in het geval van de geometrische verdeling (zie Hoofdstuk 3, opgave 9), maar we laten het bewijs hier achterwege.
De aankomsttijden kunnen dus worden opgevat als sommen van onafhankelijke exponentieel verdeelde stochasten, zoals we bekeken hebben in paragraaf 9.6. En dus is de verdeling van de aankomsttijd T_n van de n^{e} klant ($T_n = X_1 + \ldots + X_n$) precies dezelfde als we daar gevonden hebben . Hun dichtheidsfuncties zijn:

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}.$$

Opmerking. Deze dichtheidsfuncties kan men ook anders afleiden. Als het aantal klanten K_t in $(0,t]$ Poisson (λt) verdeeld is, dan is nl.

$$F_{T_n}(t) = \mathbb{P}[T_n \leq t]$$

$$
\begin{aligned}
&= \mathbb{P}[K_t \geq n] \\
&= 1 - \mathbb{P}[K_t \leq n-1] \\
&= 1 - e^{-\lambda t}\left((1 + \lambda t + \frac{(\lambda t)^2}{2!} + \ldots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right).
\end{aligned}
$$

Differentieer vervolgens: $f_{T_n}(t) = F'_{T_n}(t)$ en we vinden de gezochte dichtheid (opgave 7).

10.9 Het Poissonproces in $\mathbb{R}^n$.

In plaats van de tijdstippen van binnenkomst van klanten en tijdsinterval I hadden we in paragraaf 10.1 ook wel kunnen kiezen: paardebloemen in een weiland W. Verdeling in hokjes $H \subset W$ in plaats van deelintervallen leidt tot dezelfde uitkomst voor de verdeling van $K(W)$, en ook van $K(H)$ met $H \subset W$: een Poisson-verdeling met als parameter de gemiddelde dichtheid $\lambda \times$ het oppervlak van H.
Op overeenkomstige wijze kan men de verdeling van sterren over een stukje heelal beschrijven door een Poissonproces in $\mathbb{R}^3$. Dat geldt ook voor de verdeling van melkwegstelsels over een groter stuk heelal.

10.10 Simulatie van de Poissonverdeling. We kunnen de eigenschap van de onafhankelijke exponentieel verdeelde tussenpozen uit paragraaf 10.8 gebruiken om het Poissonproces, en daarmee de Poissonverdeling, te simuleren. Immers K_t, het aantal aankomsttijdstippen vóór tijdstip t, is Poisson-verdeeld met parameter λt. Kies nu even $\lambda = 1$ en noem t weer: λ. We simuleren de tijdstippen $T_1, T_2, \ldots$ door exponentieel(1) verdeelde stochasten bij elkaar te tellen: $T_1 = -\ln U_1$, $T_2 = T_1 + (-\ln U_2)$, ...
Stop zodra $T_{k+1} \geq \lambda$. Als $T_k < \lambda$ en $T_{k+1} \geq \lambda$, dan geven we X de waarde k.
In het bijzonder: $X = 0$ als $T_1 \geq \lambda$.
We moeten dus steeds controleren of T_k nog kleiner dan λ is.

$$
\begin{aligned}
T_k &= -\ln U_1 - \ln U_2 - \ldots - \ln U_k \\
&= -\ln(U_1 \cdots U_k) < \lambda
\end{aligned}
$$

en dit is precies het geval als

$$U_1 \cdots U_k > e^{-\lambda}.$$

We vermenigvuldigen dus randomgetallen $U_1, U_2, \ldots$ zolang het product groter blijft dan $e^{-\lambda}$. Op den duur duikt het product onder het niveau $e^{-\lambda}$. De laatste k waarvoor $U_1 \cdots U_k > e^{-\lambda}$ is de gesimuleerde van onze Poisson(λ)-stochast.

Opgaven

1. Zij X Poisson(λ) verdeeld. Voor welke $k \in \mathbb{N}$ is $\mathbb{P}[X = k]$ maximaal?

2. Stel dat iedere schroef die door een machine gemaakt wordt, met kans 0,01 een gebrek heeft (onafhankelijk van de andere schroeven). Benader de kans dat er in een doosje met 200 schroeven minder dan 5 defecte schroeven zitten (Aanwijzing: paragraaf 10.3).

3. Bepaal in paragraaf 10.6 de kans $\mathbb{P}[S{=}k]$ voor $k = 0, 1, 2, 3$. Heeft S een Poisson-verdeling?

4. Het aantal geboorten X in Nijmegen op één dag heeft een Poisson(λ) verdeling. Bij iedere geboorte kan het (onafhankelijk van andere geboorten) een eenling (kans $p_1 \in (0, 1)$) of een tweeling (kans p_2) zijn; $p_1 + p_2 = 1$. Dus als X_1 het aantal eenlingen en X_2 het aantal tweelingen is, dan is $X = X_1 + X_2$ het aantal geboorten en $Y = X_1 + 2X_2$ het aantal kinderen.

   (a) Bepaal $\mathbb{P}[Y = 4]$.

   (b) Bepaal $\mathbb{E}(Y)$ en $\mathbb{V}\mathrm{ar}(Y)$.

   (c) Heeft ook Y een Poisson-verdeling?

5. In een zeker gebied treden aardbevingen bij benadering op volgens een Poissonproces met een gemiddelde van 2 aardbevingen per week.

   (a) Bereken de kans dat er de komende twee weken minstens drie aardbevingen optreden.

   (b) Wat is de kans dat de eerstvolgende aardbeving minstens drie weken op zich laat wachten?

6. (a) Klanten komen aan volgens een Poisson(λ) proces. Zij T de aankomsttijd van de 1$^{\mathrm{e}}$ klant. Zij $t > 0$. Bepaal de kans $\mathbb{P}[T > t]$ (Aanwijzing: Druk de gebeurtenis $[T > t]$ uit in een gebeurtenis m.b.v. X_t, het aantal klanten in $(0, t]$).

   (b) Zij in paragraaf 10.9 R de afstand tot de oorsprong van de dichtstbijzijnde paardebloem. Bepaal $\mathbb{P}[R > t]$.

   (c) Zij in paragraaf 10.9 R de afstand van de dichtstbijzijnde ster tot de oorsprong. Bepaal $\mathbb{P}[R > t]$.

7. Bepaal f_{T_n} in paragraaf 10.8 door de aldaar genoemde differentiatie uit te voeren.

8. Hagelstenen vallen op een vlak. Voor ieder deel van het vlak met oppervlakte a is het aantal hagelstenen Poisson verdeeld met verwachting λa (voor elke $a > 0$). (X, Y) zijn de coördinaten van de steen die het dichtst bij $(0, 0)$ valt. Zij $R = \sqrt{X^2 + Y^2}$. Bepaal verdelingsfunctie en dichtheid van R. Is R exponentieel verdeeld?

9. Grassprietjes groeien in een tuin volgens een Poisson(λ) proces[1]. Er staat een paaltje (dikte nul) in de tuin.

   (a) Er staan 10.000 grassprietjes op hoogstens 1 m afstand van het paaltje. Hoe groot is de kans dat er hiervan precies één op hoogstens 1 cm van het paaltje staat?

   (b) Geef een Poissonbenadering voor de in (a) gevonden kans.

10. Tussen 11 en 12 uur komen klanten een winkel binnen volgens een Poisson-proces met intensiteit 3. Elke klant blijft twintig minuten in de winkel. Hoe groot is de kans dat twee klanten elkaar ontmoeten?

11. (Een uitbreiding van de "Splitsings-eigenschap") Bewijs:

$$\left.\begin{array}{r} K \sim \text{Poisson}(\lambda) \\ K = K_1 + K_2 + K_3 \\ \mathbb{P}[K_i = k | K = n] = \binom{n}{k} p_i^k (1 - p_i)^{n-k} \end{array}\right\} \Rightarrow \left\{\begin{array}{l} K_i \sim \text{Poisson}(\lambda p_i) \\ K_1, K_2, K_3 \text{ zijn onafhankelijk.} \end{array}\right.$$

[1] d.w.z. gemiddeld λ sprietjes per cm^2.

Hoofdstuk 11

Simulatie

11.1 De experimentele weg. Toevalsverschijnselen kunnen we analyseren met behulp van de theorie in de voorgaande hoofdstukken.

Voorbeeld. Hoeveel worpen met een dobbelsteen zijn nodig om iedere kant minstens één keer boven te krijgen? Langs theoretische weg (opgave 12 uit hoofdstuk 5) vind je de verwachting: 14,7 $(= 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + 6)$, dus gemiddeld lukt het in ongeveer 15 worpen.

Hoe groot is de kans dat je 20 worpen nodig hebt? Als je geen zin hebt deze kans uit te rekenen (hoe zou je dat trouwens moeten doen?), dan kun je nog altijd proefondervindelijk een indruk krijgen van de grootte van deze kans, en wel op twee manieren:

(a) Doe de proef echt. Gooi 20 dobbelstenen en kijk of je "alles hebt". Herhaal dit (100 keer bijvoorbeeld).

(b) Doe alsof (simuleer). Gebruik de computer. De randomgenerator produceert een getal U uniform verdeeld tussen 0 en 1. Met $[6U] + 1$ krijg je met gelijke kansen uitkomsten 1, 2, 3, 4, 5, en 6 net als bij een echte dobbelsteen. Je kunt de computer dit laten herhalen, bijvoorbeeld tot iedere uitkomst aan de beurt geweest is. Zo krijg je een gesimuleerde waarde van de stochast X: het benodigde aantal worpen. Laat de computer 1000 keer een waarde voor X produceren. Daarbij wordt bijgehouden hoe vaak ieder van de mogelijke waarden $6, 7, 8, \ldots$ optreedt. Het resultaat kan op het scherm gebracht worden in de vorm van een histogram en een tabel.

11.2 Randomgenerator. De ideale randomgenerator produceert een rij toevalsgetallen $U_1, U_2, U_3, \ldots$ die voldoet aan:

(a) Iedere U_i is uniform verdeeld over het interval $(0, 1)$: $\mathbb{P}[U_i \leq u] = u$ voor u tussen 0 en 1.

(b) De stochasten U_1, U_2, $U_3, \ldots$ zijn onafhankelijk.

Aan deze eisen is in de praktijk nooit helemaal exact voldaan. Wegens het gebruik van een beperkt aantal (zeg k) decimalen is de verdeling van een randomgetal eigenlijk niet eens continu, zodat aan de eerste eis niet voldaan is. Dit is voor de meeste praktische doeleinden niet zo erg, als alle 10^k mogelijke uitkomsten maar dezelfde kans hebben, d.w.z. op den duur ongeveer even vaak voorkomen.
Ernstiger zijn meestal de tekortkomingen wat betreft de tweede eis. De onafhankelijkheid van de uniforme stochasten U_n en U_{n+1} impliceert dat de punten (U_n, U_{n+1}) uniform verdeeld zijn over het eenheidsvierkant. Eenvoudige contrôle: Breng een groot aantal van die punten op het scherm. Doe dit een paar keer. Ontstaan er systematisch bepaalde patronen dan is er iets mis met eis (b).

Er zijn allerlei methoden en principes voor het maken van randomgeneratoren. Zie hiervoor bijvoorbeeld: D.E. Knuth, *The Art of Computer Programming*, deel 2, hoofdstuk 3. Daar wordt ook beschreven hoe je kunt testen of de geproduceerde toevalsrijen wel voldoende toevallig, chaotisch, onvoorspelbaar zijn. Wij zullen in het volgende uitgaan van een ideale randomgenerator en enkele dingen bespreken die we ermee kunnen doen.
In de rest van dit hoofdstuk stelt U steeds een uniforme stochast voor, en U_1, $U_2, \ldots$ een rij van onafhankelijke uniforme stochasten zoals geproduceerd door een ideale randomgenerator.

11.3 Enkele discrete verdelingen

(a) Discreet uniform. Zo noemt men een stochast X met $\mathbb{P}[X{=}k] = \frac{1}{N}$ voor $k = 1, 2, \ldots, N$. Deze verdeling krijgen we met $X = [NU] + 1$.

(b) Alternatieve verdeling: $\mathbb{P}[X{=}1] = 1 - \mathbb{P}[X{=}0] = p$.
Recept: $X = [p + U]$.

(c) Binomiale verdeling (met parameters n en p).
Uit de interpretatie van X: het aantal successen in n onafhankelijke proeven met succeskans p, volgt dat we X kunnen simuleren met $X = X_1 + \ldots + X_n$, waarbij $X_i = [p + U_i]$ voor $i = 1, 2, \ldots, n$.

(d) Geometrische verdeling. $\mathbb{P}[X{=}k] = pq^{k-1}$ voor $k = 1, 2, \ldots$.

Interpretatie: X is het aantal proeven dat nodig is om één succes te behalen. Dit geeft ons de *eerste methode* om X te simuleren: Maak $X_1, X_2, \ldots$ met $X_i = [p + U_i]$ voor $i = 1, 2, \ldots$. Stop zodra je een 1 krijgt. Bijvoorbeeld: Als $X_1 = 1$ dan $X = 1$. Als $X_1 = X_2 = \ldots = X_{k-1} = 0$ en $X_k = 1$, dan $X = k$. Het aantal randomgetallen dat nodig is om één waarde van X te produceren is bij deze methode variabel (gemiddeld $\frac{1}{p}$ stuks).

Het is ook mogelijk om het met slechts één randomgetal klaar te spelen. *Tweede methode:* $X := 1 + \left[\dfrac{\ln U}{\ln q}\right]$ (ga na dat $\mathbb{P}[X{>}k] = q^k$ en dus $\mathbb{P}[X{=}k] = \mathbb{P}[X > k{-}1] - \mathbb{P}[X{>}k] = q^{k-1} - q^k = pq^{k-1}$).

11.4 De trekking van de lottogetallen. Uit de bol met ballen, genummerd $1, 2, \ldots, N$, wordt n keer een bal getrokken. Dit gaat zonder teruglegging. Zo krijgen we een rijtje: $X_1, X_2, \ldots, X_n$ waarbij X_i het resultaat van de i^{e} trekking voorstelt. Bij de Nederlandse lotto: $N = 41$, $n = 6$ (of 7 met het reservegetal erbij). In Duitsland: $N = 47$, $n = 6$ (of 7).

Je speelt mee door op een formuliertje de nummers te voorspellen die getrokken gaan worden. Is je voorspelling goed of bijna goed, dan heb je een prijs.

Na de trekking worden de getrokken nummers op volgorde van grootte gelegd. Het *geordende* resultaat geven we aan met $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

Bijvoorbeeld bij een trekking $(X_1, \ldots, X_6) = (23, 16, 34, 2, 15, 20)$ wordt de gerangschikte uitslag: $(X_{(1)}, \ldots, X_{(6)}) = (2, 15, 16, 20, 23, 34)$.

Hoe kunnen we zulke lotto-uitslagen met de computer nabootsen? Eerst een paar algemene opmerkingen:

1. We hebben hier te doen met een steekproef van omvang n uit de verzameling $\{1, \ldots, N\}$. Deze steekproef veronderstellen we *aselect* te zijn. Dat betekent: Alle $\binom{N}{n}$ mogelijkheden voor $(X_{(1)}, \ldots, X_{(n)})$ hebben dezelfde kans, en ook: alle $n!\binom{N}{n}$ mogelijkheden voor $(X_1, \ldots, X_n)$.

2. Ieder element van $\{1, \ldots, N\}$ heeft dezelfde kans om bij de i^{e} trekking te voorschijn te komen: $\mathbb{P}[X_i{=}k] = \frac{1}{N}$, voor $i = 1, \ldots, n$ en $k = 1, \ldots, N$. Dus: iedere X_i is discreet uniform.

3. Ieder element van de verzameling $\{1, \ldots, N\}$ heeft dezelfde kans om in de steekproef terecht te komen. Deze kans is $\frac{n}{N}$.

De beweringen 2 en 3 volgen uit de veronderstellingen in 1. Bewijs dit zelf.

Simulatie van $(X_1, \ldots, X_n)$**.**

Eerste methode: (primitief). De X_i's zijn discreet uniform, dus neem $X_i = [NU_i] + 1$, $i = 1, \ldots, n$ (dus net als bij een steekproef met teruglegging). Controleer nu of alle n getallen verschillend zijn. Zo nodig uitkomsten die eerder zijn voorgekomen weggooien en extra trekkingen doen totdat er n verschillende resultaten zijn.

Je kunt dit op verschillende manieren programmeren. Je kunt bijvoorbeeld bij iedere trekking controleren of het resultaat al eerder is voorgekomen, zo ja, weg ermee en opnieuw trekken. Bij deze methode is het aantal trekkingen (en dus het aantal benodigde randomgetallen) variabel.

De volgende methode heeft altijd precies n randomgetallen nodig.

Tweede methode. Het idee is als volgt. Zet N bakjes op een rij. In ieder bakje een bal. $A(i)$ is de inhoud van bakje i (het nummer van de bal in het i^{e} bakje).

Beginsituatie: $A(i) = i$ voor $i = 1, \ldots, N$ (bal i in bakje i).

Eerste trekking: Kies een bakje, ieder bakje met dezelfde kans. Neem de bal uit dat bakje. Dat is de eerste getrokken bal. Die kan ook niet opnieuw getrokken worden, want het bakje waar hij in zat is nu leeg.

We doen nu de bal uit bakje N in dat lege bakje, zodat bakjes 1 t/m $N-1$ gevuld zijn en bakje N leeg (had je bakje N gekozen, dan hoef je niets te doen). We zijn nu klaar voor de

Tweede trekking: Kies een van de bakjes $1, 2, \ldots, N-1$ met gelijke kansen. De bal uit dat bakje is de tweede getrokken bal.

De bal uit bakje $N-1$ doen we in het bakje dat net geleegd is. Zo gaan we door.

Iets formeler: Beginsituatie: $A(i) = i$ voor $i = 1, \ldots, N$.

1^{e} trekking: Kies i_1 uniform uit $\{1, \ldots, N\}$:

```
i1 := [NU1] + 1
X1 := A(i1)
A(i1) := A(N)
```

(Bakje i_1 gekozen)
(Eerste trekking wordt: bal uit bakje i_1)
(Nieuwe inhoud bakje i_1 wordt gelijk aan inhoud bakje N)

2^{e} trekking: Kies i uniform uit $\{1, \ldots, N-1\}$:

```
i2 := [(N-1)U2] + 1
X2 := A(i2)
A(i2) := A(N-1)
```

Het is duidelijk hoe dit verder gaat.

Door een variabele v in dienst te nemen die bijhoudt hoeveel bakjes nog gevuld zijn, kunnen we de procedure iets netter beschrijven:

Methode A

Beginsituatie: $v = N$; $A(i) = i$ voor $i = 1, \ldots, N$.
Voor $k = 1, \ldots, n$ doe je het volgende:

`i := [vU] + 1`	(keuze bakje i uniform uit $\{1, \ldots, v\}$)
`X`$_k$ `:= A(i)`	(k^e trekking: bal uit bakje i)
`A(i) := A(v)`	(Nieuwe vulling bakje i wordt bal uit meest rechtse bakje)
`v := v - 1`	(Aantal gevulde bakjes wordt één minder)

Het is niet moeilijk in te zien dat deze methode correct is, d.w.z. de steekproef is aselect zoals gedefinieerd op blz.115. Immers, wat gebeurt er precies?
Er worden achtereenvolgens bakjes $i_1, i_2, \ldots, i_n$ gekozen. Het aantal mogelijkheden voor $i_1, i_2, \ldots, i_n$ is

$$N(N-1)\cdots(N-n+1) = \frac{N!}{(N-n)!}.$$

Bij ieder rijtje $(i_1, \ldots, i_n)$ hoort precies één steekproef $(x_1, \ldots, x_n)$ (bijvoorbeeld bij $(i_1, \ldots, i_n) = (1, 1, \ldots, 1)$ krijg je $(x_1, \ldots, x_n) = (1, N, N-1, \ldots, N-n+2)$).
Verander je het rijtje $(i_1, \ldots, i_n)$ dan verandert ook $(x_1, \ldots, x_n)$.
Het aantal mogelijkheden voor $(x_1, \ldots, x_n)$ is $n!\binom{N}{n} = \frac{N!}{(N-n)!}$, evenveel als voor $(i_1, \ldots, i_n)$. Er is dus een 1-1 correspondentie.
De kans op een bepaald rijtje $(i_1, \ldots, i_n)$ is gelijk aan

$$\frac{1}{N} \cdot \frac{1}{N-1} \cdots \frac{1}{N-n+1}$$

(hier gebruik je de onafhankelijkheid van de gebruikte randomgetallen $U_1, \ldots, U_n$).
Dus iedere mogelijkheid heeft dezelfde kans.

Random permutatie. Neem je $n = N$, dan maakt de voorgaande procedure een random rangschikking van de getallen $1, \ldots, N$. Dit vindt allerlei toepassingen: van CD-spelers tot bridgedrives.

Simulatie van $(X_{(1)}, \ldots, X_{(n)})$.
De gerangschikte steekproef $(X_{(1)}, \ldots, X_{(n)})$ kunnen we krijgen door eerst $(X_1, \ldots, X_n)$ te maken zoals we hiervoor besproken hebben, en daar de getrokken nummers in de goede volgorde te zetten (sorting). Er is echter ook een procedure die ons rechtstreeks de gerangschikte steekproef geeft.

Het idee is als volgt: Bij ieder element van $\{1,\ldots,N\}$ kiest het toeval of dit element zal worden opgenomen in de steekproef of niet. Het resultaat van dit kiesproces wordt weergegeven als een rijtje $(Y_1,\ldots,Y_N)$ bestaande uit n enen en $N-n$ nullen, $Y_i = 1$ als element i gekozen wordt, en $Y_i = 0$ als dat niet zo is (voorbeeld: de steekproef $(X_{(1)},\ldots,X_{(4)}) = (3,4,7,9)$ uit $\{1,\ldots,10\}$ wordt weergegeven door $(Y_1,\ldots,Y_{10}) = (0,0,1,1,0,0,1,0,1,0)$). Volgens opmerking 3 op blz. 115 geldt voor iedere i: $\mathbb{P}[Y_i = 1] = \frac{n}{N}$. Maar de stochasten $Y_1,\ldots,Y_N$ zijn niet onafhankelijk. De afhankelijkheid is als volgt: Als van de elementen $1,\ldots,k-1$ er al s gekozen zijn dan is de kans dat k gekozen wordt gelijk aan $\frac{n-s}{N-k+1}$, immers van de overblijvende $N-k+1$ elementen $k,\ldots,N$ moeten er nog $n-s$ gekozen worden om de steekproef vol te maken. Dus

$$\mathbb{P}(Y_k = 1 | Y_1 + \ldots + Y_{k-1} = s) = \frac{n-s}{N-k+1}.$$

We kunnen het ook zo zeggen: De kans op een één (of een nul) bij de volgende keuze is evenredig met het aantal ontbrekende enen (nullen). Dit leidt automatisch tot het juiste aantal enen want zodra je n enen hebt wordt de kans op nog een één nul. Deze intuïtieve gedachtengang leidt tot de volgende

Methode B

Beginsituatie: $s = 0$ (s houdt bij hoeveel énen je al hebt, dus hoeveel elementen er al gekozen zijn).
Voor $k = 1,\ldots,N$ doe je het volgende:

$\mathtt{Y_k} := [\mathtt{U} + \frac{\mathtt{n-s}}{\mathtt{N-k+1}}]$
`s := s + Y`$_\mathtt{k}$
`ALS Y`$_\mathtt{k}$ `= 1 DAN X`$_\mathtt{(s)}$ `:= k`

Op deze manier maken we een rijtje $(Y_1,\ldots,Y_N)$ en tegelijk de bijbehorende gerangschikte steekproef $(X_{(1)},\ldots,X_{(n)})$.
De methode is correct: alle $\binom{N}{n}$ mogelijkheden hebben dezelfde kans. Neem maar zo'n rijtje $(y_1,\ldots,y_n)$ met n enen en $N-n$ nullen.
De kans op dit rijtje is een breuk met noemer $N(N-1)\cdots 2\cdot 1$. De teller bevat alle factoren n, $n-1$,..., 2, 1 en $N-n$, $N-n-1$,... (de ontbrekende aantallen enen en nullen). Dus ieder rijtje $(y_1,\ldots,y_N)$ heeft kans $\frac{n!(N-n)!}{N!}$.

11.5 Steekproef en populatie. Behalve steekproeven uit $\{1,\ldots,N\}$ zoals op zondagavond bij de trekking van de lottogetallen, komt de oplettende burger door

de week nog heel wat steekproeven tegen uit allerlei verzamelingen. Dikwijls staat men er niet bij stil dat de gegevens die we tegenkomen slechts een steekproef zijn uit een groter geheel, een topje van de ijsberg. Van de grote massa onder water hebben we een vaag vermoeden.
Door het nemen van steekproeven kan men iets te weten komen over een grotendeels onbekende verzameling die men wil onderzoeken.

De algemene situatie kan, abstract, zo voorgesteld worden:
Er is een populatie $\mathcal{P}$. De elementen van $\mathcal{P}$ kunnen getallen zijn of vectoren, of functies, of ...
Voor de eenvoud houden wij het bij getallen: $\mathcal{P}$ is een eindig rijtje van getallen $(x_1, \ldots, x_N)$. Het gaat bijvoorbeeld over de studenten in een bepaalde plaats. x_i is het geldbedrag dat student i per maand van zijn/haar ouders krijgt, of het aantal glazen dat i drinkt, of waar men zich ook maar voor interesseert.
Men kiest nu, op goed geluk, aselect, wat elementen uit $\mathcal{P}$. Deze vormen de steekproef.
Hoe kunnen we zo'n steekproef trekken?
Omdat de elementen van $\mathcal{P}$ genummerd zijn, komt dit neer op het trekken van een steekproef uit $\{1, \ldots, N\}$. Je kunt je voorstellen dat $\mathcal{P}$ een kaartsysteem is met kaarten genummerd van 1 tot en met N. Het rangnummer i staat onopvallend in de linkerbovenhoek van kaart i en het getal x_i waar het om gaat staat dik gedrukt midden op die kaart. Dus: trek maar n nummers uit $\{1, \ldots, N\}$. De gegevens op de bijbehorende kaartjes vormen dan de steekproef $(X_1, \ldots, X_n)$.
Voor het *simuleren* van een steekproef (aselect en zonder terugleggen) $(X_1, \ldots, X_n)$ *uit* $\{x_1, \ldots, x_N\}$ kunnen we methode A gebruiken. We hoeven alleen de eerste regel maar te veranderen. Die wordt nu:

Methode A′

Beginsituatie: $v = N$; $A(i) = x_i$ voor $i = 1, \ldots, N$.
De rest is hetzelfde als bij **Methode A**

11.6 De hypergeometrische verdeling. Uit een doos met N ballen, m witte en $N-m$ rode, neemt men een aselecte steekproef (zonder teruglegging) van n ballen. In de steekproef treft men X witte ballen aan. X heeft de bekende hypergeometrische verdeling.
Dit is een bijzonder geval van de algemene situatie hiervoor. De populatie $\mathcal{P}$ is hier

$(x_1, \ldots, x_N)$ met $x_i = 1$ als $1 \leq i \leq m$ en $x_i = 0$ als $m < i \leq N$. Tellen we de n getallen $X_1, \ldots, X_n$ bij elkaar op, dan hebben we X.

Simulatie van X kan dus zoals hierboven beschreven, maar het gaat ook heel goed als volgt:

Beginsituatie: $X = 0$ (X houdt bij hoeveel witte ballen er getrokken zijn)
Voor $k = 1$ tot en met n doen we het volgende:

$$\mathtt{X_k\ :=\ [\frac{m-X}{N-k+1}\ +\ U]}$$
$$\mathtt{X\ :=\ X\ +\ X_k}$$

Er zijn bij deze methode n randomgetallen nodig. Soms kan het met minder. Bijvoorbeeld als $m < n$.
Voorbeeld: Het aantal azen in een hand van 13 kaarten heeft dezelfde verdeling als het aantal $\heartsuit$ in een greep van 4 kaarten. In het eerste geval hebben we een hypergeometrische verdeling met $n = 13$, $m = 4$ ($N = 52$). In het tweede geval: $n = 4$, $m = 13$.
De verwisselbare rol van m en n blikt als we de kans op $X = k$ uitschrijven:

$$\mathbb{P}[X = k] = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{n!m!(N-n)!(N-m)!}{N!k!(n-k)!(m-k)!(N-n-m+k)!}.$$

Ook de uitdrukkingen voor verwachting en variantie vertonen deze symmetrie:

$$\mathbb{E}(X) = \frac{nm}{N} \text{ en } \mathbb{V}\mathrm{ar}(X) = \frac{nm(N-m)(N-n)}{N^2(N-1)}.$$

Je kunt het ook zo bekijken: De populatie is ingedeeld naar kleur: wit (m stuks) en rood ($N - m$). Door het trekken van de steekproef komt er een tweede, toevallige, indeling bij: Een bal kan in de steekproef zitten (n stuks) of niet ($N - n$ stuks). De vier aantallen n, m, $N - n$ en $N - m$ zijn de randgetallen van een 2×2 tabelletje

	wit	rood	
In steekproef	X	$n-X$	n
Niet in steekproef	$m-X$	$N-n-m+X$	$N-n$
	m	$N-m$	

De tweede indeling wordt aangebracht onafhankelijk van de eerste. De aantallen in de 4 hokjes (cellen) hebben alle vier de hypergeometrische verdeling.
Voor simulatie komt het vakje met de kleinste randgetallen het eerst in aanmerking. het aantal randomgetallen is dan gelijk aan het kleinste randgetal.

11.7 Simulatie in de statistiek. *11.7.1 Twee gemiddelden*

Een test meet zoiets als kennis, begrip, woordenschat van het Engels. 8 jongens en 10 meisjes hebben die test afgelegd en daarbij de volgende scores behaald:

Jongens:	48, 58, 62, 72, 74, 80, 88, 98	gemiddeld: $\frac{580}{8} = 72{,}5$
Meisjes:	63, 72, 74, 74, 74, 80, 80, 85, 92, 94	gemiddeld: 78,8

Het gemiddelde van de meisjes ligt 6,3 hoger dan het gemiddelde van de jongens. Hoe beoordeel je zoiets?

Ook als er in het algemeen geen verschil zou bestaan tussen de prestaties van jongens en meisjes op deze test dan zul je toch bijna nooit twee groepen met precies hetzelfde gemiddelde tegenkomen. Toeval speelt hier een rol.

Laten we uitgaan van de veronderstelling dat er geen wezenlijk verschil is. Deze aanname noemt men in de statistiek de "nulhypothese". Deze houdt in dat geconstateerde verschillen op toeval berusten. De invloed van het toeval kunnen we onderzoeken d.m.v. simulatie.

Dat gaat zo. Doe alle uitslagen bij elkaar in één rijtje

$$\mathcal{P} = (x_1, \ldots, x_{18}) = (48, 58, \ldots, 98, 63, \ldots, 94).$$

Trek steekproeven van 8 uit $\mathcal{P}$: $(X_1, \ldots, X_8)$. Vergelijk het gemiddelde

$$\overline{X} = \tfrac{1}{8}(X_1 + \ldots + X_8)$$

(het gemiddelde van de jongens) met het gemiddelde $\overline{Y}$ van de rest (de meisjes).

$8\overline{X} + 10\overline{Y} = 1368$ (het totaal van alle 18 uitslagen), dus $\overline{Y} = 136{,}8 - \frac{8}{10}\overline{X}$.

Het verschil van de gesimuleerde gemiddelden is dus

$$V = \overline{X} - \overline{Y} = 1{,}8\overline{X} - 136{,}8.$$

Door een groot aantal van zulke verschillen te simuleren krijgen we een indruk of het feitelijk waargenomen verschil uitzonderlijk is. Krijg je bijvoorbeeld in 1000 herhalingen maar 20 keer (2%) V-waarde kleiner dan $-6{,}3$ (het verschil tussen gemiddelden zoals oorspronkelijk waargenomen) dan hebben we hier toch wel een sterke aanwijzing dat het waargenomen verschil niet toevallig is (een *significant verschil* heet zoiets in de statistiek).

11.7.2 Correlatie.

n personen hebben ieder aan twee tests meegedaan. Persoon i heeft score x_i behaald op de 1e test (Engels) en score y_i op de 2e test (wiskunde).

Dit geeft n punten (x_i, y_i). Deze "puntenwolk" (scatter diagram) kan al een indruk

geven van een zekere samenhang tussen de prestaties door beide tests gemeten. Een kwantitatieve maat voor die samenhang is de zogenaamde *correlatiecoëfficiënt* r. Deze wordt als volgt berekend (vergelijk ook paragraaf 5.12):
Bepaal eerst de gemiddelden

$$\overline{x} = \tfrac{1}{n} \sum x_i \text{ en } \overline{y} = \tfrac{1}{n} \sum y_i.$$

Dan:

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left(\sum (x_i - \overline{x})^2\right)\left(\sum (y_i - \overline{y})^2\right)}}.$$

Het is niet moeilijk in te zien dat r altijd tussen -1 en $+1$ ligt. $r = 1$ als alle punten op een lijn $y = ax + b$ met $a > 0$ liggen. $r = -1$ als alle punten op een lijn $y = ax + b$ met $a < 0$ liggen.
Ook als er in werkelijkheid geen enkel verband is tussen x en y dan zal r toch, af en toe, een tamelijk hoge waarde kunnen aannemen.
Om te kunnen beoordelen of de correlatiecoëfficiënt die we uit onze waarnemingen berekend hebben echt iets betekent, moeten we de toevalsspreiding van r kennen. Daar kunnen we een goede indruk van krijgen met een simulatie-experiment. We maken de verbinding tussen de x-waarden en de y-waarden in paren (x_i, y_i). We zetten de x-waarden in een vaste volgorde en maken een randompermutatie van de y-waarden: $Y_1, \ldots, Y_n$.
Voor de puntenwolk (x_i, Y_i), $i = 1, \ldots, n$, berekenen we r.
Dat doen we een groot aantal keren, 1000 keer bijvoorbeeld.
Dan kun je zien of de feitelijke waarde van r uitzonderlijk hoog (of laag) is.
Men kan aantonen dat voor de verwachting en spreiding van r (over alle $N!$ gelijkwaarschijnlijke permutaties) geldt:

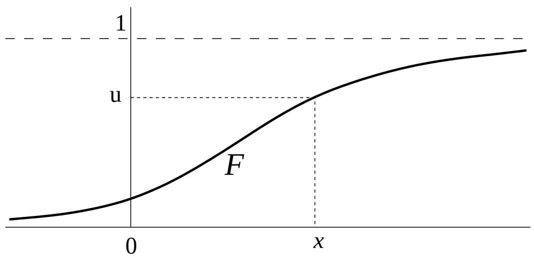
$$\mathbb{E}(r) = 0 \text{ en } \mathbb{V}\mathrm{ar}(r) = \frac{1}{n-1}$$

(Dit geldt ongeacht de waarden van de getallen x_i, y_i. De vorm van de verdeling van r hangt natuurlijk *wel* van deze waarden af).

11.8 De methode met de inverse verdelingsfunctie. Een stochast X met een strikt stijgende continue verdelingsfunctie F, kunnen we simuleren met

$$X = F^{-1}(U).$$

F^{-1} is de inverse functie van F. Het plaatje maakt dit duidelijk: Als $x = F^{-1}(u)$ (dus $u = F(x)$), dan is $X \leq x$ precies dan als $U \leq u$, en dit gebeurt met kans $\mathbb{P}[X \leq x] = \mathbb{P}[U \leq u] = u = F(x)$.



Om de gewenste uitdrukking $X = F^{-1}(U)$ te krijgen moeten we X oplossen uit de vergelijking $F(X) = U$. Je mag ook de vergelijking $F(X) = 1 - U$ nemen, want $1 - U$ is ook uniform-$(0, 1)$-verdeeld.

Voorbeeld 1. De exponentiële verdeling. $F(x) = 1 - e^{-\lambda x}$. Op te lossen: $F(X) = 1 - U$, $e^{-\lambda X} = U$, dus

$$X = -\tfrac{1}{\lambda} \ln U.$$

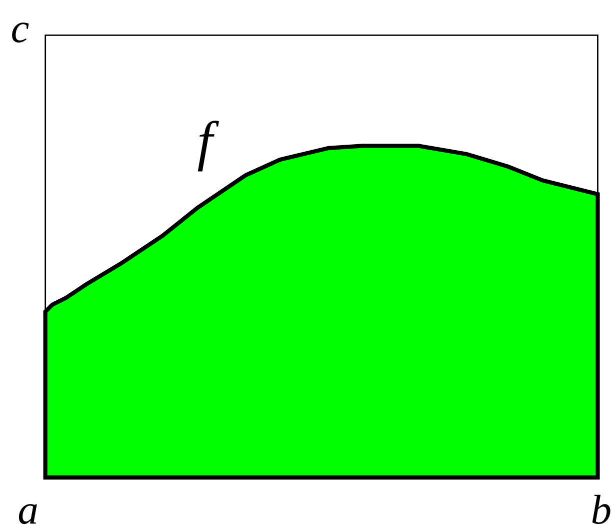
Voorbeeld 2. De Cauchy-verdeling. $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$, dichtheid $f(x) = \frac{1}{\pi(1+x^2)}$. Uit $F(X) = U$ vinden we:

$$X = \tan(\pi(U - \tfrac{1}{2})).$$

11.9 De wegwerpmethode (rejection method). Een continue verdelingsfunctie heeft niet altijd een eenvoudige inverse. De normale verdelingsfunctie, bijvoorbeeld. In zulke gevallen is de methode van paragraaf 11.8 minder praktisch. Er is gelukkig een andere algemene methode waarbij we alleen de dichtheidsfunctie nodig hebben.

We demonstreren het principe voor het geval van een dichtheidsfunctie die begrensd is en nul wordt buiten een interval (a, b).

Begrensd wil zeggen dat er een c is zodat $f(x) \leq c$ voor alle x.
Een stochast Z met dichtheid f laat zich als volgt simuleren:
Met $X = a + (b-a)U_1$ en $Y = cU_2$ maken we een punt (X, Y) uniform verdeeld over de rechthoek $(a, b) \times (0, c)$. Valt dit punt in het gearceerde gebied onder de grafiek van f dan is het een goed punt en accepteren we de x-coördinaat als een gesimuleerde waarde van X. Punten die buiten het gearceerde gebied vallen gooien we weg. De x-coördinaten van de geaccepteerde punten zijn verdeeld volgens de dichtheidsfunctie f.



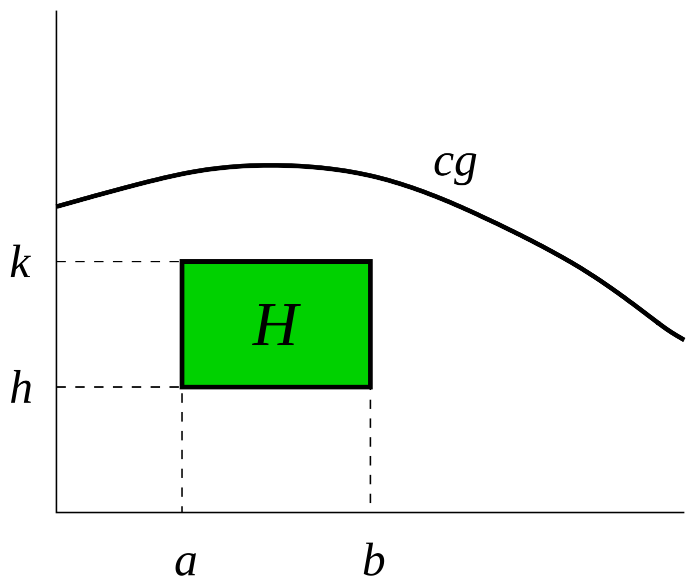
11.10 Wat te doen wanneer je een stochast wilt simuleren (normaal verdeeld, bijvoorbeeld) waarvan de dichtheidsfunctie f niet binnen een begrensde rechthoek te vangen is? Stel je hebt een stochast met een dichtheid g die je gemakkelijk kunt simuleren. Neem verder aan dat voor zekere $c > 0$ geldt: $f(x) \leq cg(x)$ voor alle x. We kunnen dan een punt (X, Y) genereren, uniform over

$$A = \{(x, y) : -\infty < x < \infty, 0 < y < cg(x)\}.$$

Als dat in orde is dan accepteren we het punt (X, Y) wanneer het in

$$B = \{(x, y) : -\infty < x < \infty, 0 < y < f(x)\}$$

valt.
De x-coördinaat van het geaccepteerde punt heeft dan dichtheid f, beweren we.
Hoe maken we een punt (X, Y) uniform over A? Wel, dat gaat zo:
Maak X volgend bekend recept met dichtheid g (bv. $X = tan(\pi(U_1 - \frac{1}{2}))$ heeft dichtheid $g(x) = (\pi(1 + x^2))^{-1}$).
Neem $Y = cg(x)U_2$, dan is (X, Y) uniform over A.



Bewijs: Neem een rechthoekje $H = (a, b) \times (h, k)$ binnen A. Bij gegeven x is de kans dat Y in (h, k) valt gelijk aan $\frac{k-h}{cg(x)}$. Dit geïntegreerd van a tot b geeft:

$$\begin{aligned} \mathbb{P}[(X, Y) \in H] &= \int_a^b \frac{k - h}{cg(x)} g(x) dx \\ &= \frac{(k - h)(b - a)}{c} \\ &= \frac{\text{Opp}(H)}{\text{Opp}(A)}. \end{aligned}$$

Zie opgave 2 voor een toepassing op de normale verdeling. We geven echter eerst twee andere manieren om normaal verdeelde stochasten te simuleren.

11.11 De normale verdeling. De *eerste methode* berust op de centrale limietstelling: Als n groot is dan lijkt de verdeling van $S_n = U_1 + \ldots + U_n$ sterk op een normale verdeling (met $\mathbb{E}(S_n) = \frac{n}{2}$ en $\mathbb{V}\mathrm{ar}(S_n) = \frac{n}{12}$).
Reeds bij $n = 4$ is de benadering fraai. Door standaardiseren:

$$Z = \sqrt{3}(S_4 - 2)$$

krijg je een stochast die moeilijk te onderscheiden is van een standaardnormale.
Wil je het helemaal mooi maken, neem dan $n = 12$.

$$Z = S_{12} - 6$$

doet het fantastisch.

De *tweede methode* (de poolcoördinatenmethode) maakt tegelijk twee onafhankelijke standaardnormale stochasten X en Y zodat (X, Y) een punt in het vlak is met dichtheid

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}.$$

Dit gaat via de poolcoördinaten: straal R en hoek H:

$$X = R\cos H \text{ en } Y = R\sin H.$$

Het is duidelijk dat we H uniform $(0, 2\pi)$ moeten nemen.
Wat R betreft:

$$\mathbb{P}[R \le r] = \iint_C \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dxdy,$$

waarbij

$$C = \{(x, y) : x^2 + y^2 \le r^2\}.$$

Met behulp van poolcoördinaten reken je die integraal uit:

$$\begin{aligned} \mathbb{P}[R \le r] &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^r e^{-\frac{1}{2}u^2} udud\varphi \\ &= 1 - e^{-\frac{1}{2}r^2}. \end{aligned}$$

Dus $\mathbb{P}[R^2 \le x] = \mathbb{P}[R \le \sqrt{x}] = 1 - e^{-\frac{1}{2}x}$. Dus R^2 is exponentieel verdeeld met parameter $\frac{1}{2}$: $R^2 \sim \mathrm{Exp}(\frac{1}{2})$.
Volgens het recept in paragraaf 11.8, voorbeeld 2, maken we R^2 dus met $r^2 = -2\ln U$

en dus $R = \sqrt{-2 \ln U}$.
Samenvattend: Met

$$H = 2\pi U_1 \text{ en } R = \sqrt{-2 \ln U_2}$$

en

$$Z_1 = R \cos H \text{ en } Z_2 = R \sin H$$

maak je tegelijk twee standaardnormale stochasten die nog onafhankelijk zijn ook.
Een voorbeeld. De gemiddelde lengte van jonge mannen in Nederland is 182 cm; standaardafwijking 8 cm. Zulke lengten kunnen we simuleren met $X = 182 + 8Z$, waarbij Z een standaardnormale stochast is. Hoe de verdeling van de lengten in een aselecte steekproef van 1000 personen uit deze verdeling er uit zou kunnen zien kunnen we door simulatie aan het licht brengen (gesimuleerde waarden worden afgerond op hele centimeters en in klassen van 1 cm geteld).

Opgaven

1. X is het aantal worpen met een dobbelsteen dat nodig is om iedere zijde minstens één keer boven te krijgen. Beschrijf een methode om X te simuleren, waarbij slechts 5 randomgetallen $U_1, \ldots, U_5$ gebruikt worden.

2. Een foute methode:
   In het boek *Understanding and Learning Statistics by Computer* van M.C.K. Yang en D.H. Robinson (World scientific 1986) wordt op blz. 142-143 een recept gegeven voor het trekken van een steekproef van 10 uit $\{1, \ldots, 20\}$.
   Het resultaat kan weergegeven worden door een rijtje $(Y_1, \ldots, Y_{20})$ met 10 enen en 10 nullen ($Y_i = 1$ als element i getrokken wordt).
   Yang en Robinson doen het zo: Te beginnen bij $k = 1$, om vast te stellen of $Y_k = 1$ of $Y_k = 0$, gooi je een munt op. Ga ook zo te werk voor $k = 2$, $k = 3$, enzovoorts. Om te zorgen dat je het juiste aantal enen en nullen krijgt houd je bij hoeveel je er al van ieder hebt. Zodra één van die aantallen de waarde 10 bereikt stop je en vul je de rij aan met de ontbrekende nullen of enen.
   Bijvoorbeeld, wanneer bij $k = 17$ de tiende één krijgt, dan zet je op de laatste drie plaatsen drie nullen.

   (a) Toon aan dat bij deze methode niet alle $\binom{20}{10}$ mogelijke steekproeven dezelfde kans krijgen. Bereken bijvoorbeeld

   $$\mathbb{P}[(X_{(1)}, \ldots, X_{(10)}) = (1, 2, \ldots, 10)]$$

   en

   $$\mathbb{P}[(X_{(1)}, \ldots, X_{(10)}) = (1, 3, 5, \ldots, 17, 19)]$$

   (zie voor de notatie paragraaf 11.4 over de lotto).

   (b) Is het wel waar dat ieder element dezelfde kans heeft om in de steekproef voor te komen?

   (c) Als $(Y_1, \ldots, Y_{20})$ weer de rij voorstelt met 10 enen en 10 nullen, bepaald zoals boven beschreven, wat is dan $\mathbb{P}[Y_1 = Y_2]$? En $\mathbb{P}[Y_{19} = Y_{20}]$? Hoe groot zou $\mathbb{P}[Y_i = Y_j]$ moeten zijn bij een aselecte steekproef?

3. Toppen tellen.
   $U_0, U_1, \ldots, U_n, U_{n+1}$ is een rij uniforme randomgetallen. We hebben een top bij k ($k = 1, \ldots, n$) als $U_k > U_{k-1}$ en $U_k > U_{k+1}$. Het aantal toppen in de rij is $T = T_1 + \ldots + T_n$, waarbij $T_k = 1$ als er een top is bij k (en 0 als dat niet zo is).

Een te grote of te kleine waarde van T kan een aanwijzing zijn dat er iets mis is met de randomgenerator. Bij een ideale randomgenerator geldt $\mathbb{E}(T) = \frac{n}{3}$, $\mathbb{V}\text{ar}(T) = \frac{2(n+3)}{45}$ (zie onderdeel c).
Bijvoorbeeld als $n = 99$, dan: $\mathbb{E}(T) = 33$, $\mathbb{V}\text{ar}(T) = 4{,}53$. De standaardafwijking is slechts 2,13.

(a) Wat is, bij gegeven n, de maximale waarde van T?

(b) Toon aan dat $\mathbb{E}(T) = \frac{n}{3}$.

(c) Bepaal $\mathbb{V}\text{ar}(T)$.
Aanwijzing: $\mathbb{V}\text{ar}(T) = \sum \mathbb{V}\text{ar}(T_i) + \sum \mathbb{C}\text{ov}(T_i, T_j) =$
$= n\mathbb{V}\text{ar}(T_1) + 2(n-1)\mathbb{C}\text{ov}(T_1, T_2) + 2(n-2)\mathbb{C}\text{ov}(T_1, T_3) + \ldots$

(d) Schrijf een programma dat het aantal toppen telt in rijen randomgetallen van gegeven lengte. Laat ook de frequentie bijhouden waarmee de verschillende waarden van T optreden.

4. U_1, U_2, U_3 zijn onafhankelijke uniforme randomgetallen. Gerangschikt naar grootte: $U_{(1)}$, $U_{(2)}$, $U_{(3)}$.

(a) Bepaal de verdelingsfunctie en de dichtheidsfunctie van $U_{(k)}$, $k = 1, 2, 3$. Bepaal ook $\mathbb{E}(U_{(k)})$.

(b) * Hoe wordt dit bij $U_{(1)}, \ldots, U_{(n)}$?

5. $g(x) = \frac{1}{2}e^{-|x|}$ is de dichtheidsfunctie van de zogenaamde *Laplace-verdeling*. Er zijn verschillende manieren om deze verdeling te simuleren.

(a) Laat zien dat de methode met de inverse verdelingsfunctie leidt tot:

$$X = \operatorname{sgn}(1 - 2U) \ln(1 - |2U - 1|)$$

($\operatorname{sgn}(x)$, het teken van x, is gedefinieerd door $\operatorname{sgn}(x) = 1$ als $x > 0$, $\operatorname{sgn}(x) = -1$ als $x < 0$, en $\operatorname{sgn}(0) = 0$).

(b) Een eenvoudiger formule krijg je als je twee randomgetallen gebruikt: $X = \operatorname{sgn}(U_1) \ln U_2$. Laat zien dat dit juist is.

(c) Laat zien dat

$$X = \ln \frac{U_2}{U_1}$$

ook deze verdeling genereert.

6. De normale verdeling met de verwerpingsmethode.

(a) Bepaal de kleinste c zo dat

$$e^{-\frac{1}{2}x^2} \leq ce^{-|x|}$$

voor alle x (uitkomst: $c = \sqrt{e}$).

(b) Laat zien dat je met de volgende methode een standaardnormale stochast simuleert:

$$\text{Maak } X = \ln \tfrac{U_1}{U_2} \text{ en } Y = e^{\frac{1}{2}-|x|} \cdot U_3.$$
$$\text{Accepteer } X \text{ als } Y < e^{-\frac{1}{2}x^2} \text{ (dus als } U_3 < e^{|x|-\frac{1}{2}(1+x^2)}\text{).}$$

(c) Bepaal de acceptatiekans p bij de bovenstaande methode (het aantal randomgetallen dat gemiddeld nodig is om een X-waarde te produceren is dan $\frac{3}{p}$).

7. Beschrijf een methode om een punt (X, Y) te genereren, uniform verdeeld over $C = \{(x, y) : x^2 + y^2 \leq a^2\}$

(a) Met het wegwerpprincipe;

(b) Met behulp van poolcoördinaten: $X = R\cos H$ en $Y = R\sin H$.

Hoofdstuk 12

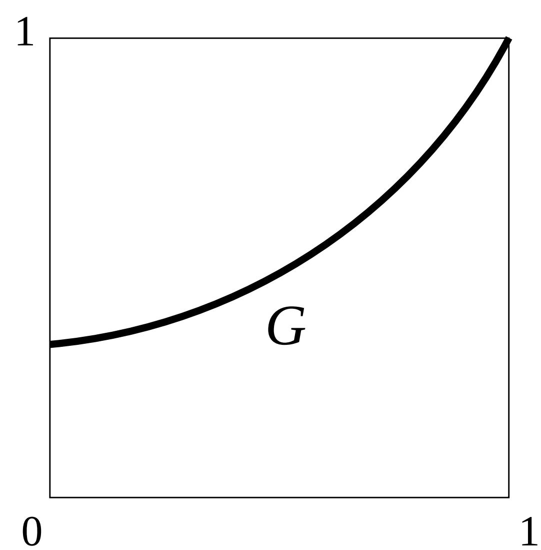
Kansgenererende functies

12.1 Definitie. Als een stochast X alleen natuurlijke getallen als waarden aanneemt, dan kunnen we bij X een functie $G = G_X : [0,1] \to [0,1]$ maken op de volgende manier:

$$G(s) = \mathbb{E}(s^X) = \sum_{n=0}^{\infty} s^n \mathbb{P}[X{=}n].$$

Deze *kansgenererende functie* G_X is een bijzonder handig instrument voor het berekenen en bestuderen van de kansverdeling van X.

12.2 Hieronder geven we enkele **eigenschappen:**



(a) G is stijgend en convex;

(b) $G(0) = \mathbb{P}[X = 0]$ en $G(1) = 1$;

(c) $\mathbb{E}(X) = G'(1)$ (eventueel beide ∞);

(d) $\mathbb{V}\mathrm{ar}(X) = G''(1) + G'(1) - G'(1)^2$;

(e) $X \perp\!\!\!\perp Y \Rightarrow G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$.

Bewijs:

(a) Omdat $s^n \leq t^n$ voor $0 < s < t$ en $n \in \mathbb{N}$, is G stijgend; omdat $s \mapsto s^n$ convex is voor alle $n \in \mathbb{N}$, is G convex.

(b) Dit zie je direct.

(c) Als $\mathbb{E}(X) < \infty$, dan is

$$\begin{aligned}
G'(1) &= \lim_{s\uparrow 1} \frac{G(1)-G(s)}{1-s} \\
&= \lim_{s\uparrow 1} \sum_{n=0}^{\infty} \frac{1-s^n}{1-s} \cdot \mathbb{P}[X{=}n] \\
&= \lim_{s\uparrow 1} \sum_{n=0}^{\infty} \left(1+s+s^2+\ldots+s^{n-1}\right) \mathbb{P}[X{=}n] \\
&= \sum_{n=0}^{\infty} n\mathbb{P}[X{=}n] \\
&= \mathbb{E}(X).
\end{aligned}$$

Ga zelf na dat de limiet $s\uparrow 1$ hier achter het somteken uitgevoerd mag worden[1].

(d)

$$\begin{aligned}
G''(1) &= \left.\frac{d^2}{ds^2} \sum_{n=0}^{\infty} s^n \mathbb{P}[X{=}n]\right|_{s=1} \\
&= \sum_{n=2}^{\infty} n(n-1)\mathbb{P}[X{=}n] \\
&= \mathbb{E}(X^2) - \mathbb{E}(X).
\end{aligned}$$

Dus als $G''(1) < \infty$, dan is ook $G'(1) < \infty$, en

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = G''(1) + \mathbb{E}(X) - \mathbb{E}(X)^2 = G''(1) + G'(1) - G'(1)^2.$$

(e) Als X en Y onafhankelijk zijn, dan is

$$G_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X \cdot s^Y) = \mathbb{E}(s^X)\mathbb{E}(s^Y) = G_X(s)G_Y(s).$$

12.3 Kansgenererende functies van bekende verdelingen. Laten we nu eens de kansgenererende functies uitrekenen van de belangrijkste geheelwaardige stochasten die we tot dusver zijn tegengekomen.

De alternatieve verdeling. Als $X \sim \mathrm{Alt}(p)$, dan is

$$G_X(s) = \mathbb{P}[X{=}0] \cdot s^0 + \mathbb{P}[X{=}1] \cdot s^1 = q + ps.$$

[1] In het vervolg zullen we zulke limieten en oneindige sommen zonder veel omhaal verwisselen. Zolang de coëfficiënten van s^n in de som positief zijn, is dit toegestaan op grond van de stelling van Abel. Een voorbeeld volgt bij (d)

De binomiale verdeling. Als $Y \sim \text{Bin}(n, p)$, dan is

$$G_Y(s) = \sum_{k=0}^{n} \mathbb{P}[Y{=}k]s^k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k = (q + ps)^n.$$

Merk op dat $G_Y(s) = G_X(s)^n$. Dit hadden we direct kunnen zeggen op grond van eigenschap (e) hierboven. Immers de som van n onafhankelijke Alt(p)-verdeelde stochasten is Bin(n, p)-verdeeld.

De Poisson-verdeling. Als $Z \sim \text{Poisson}(\lambda)$, dan is

$$G_Z(s) = \sum_{n=0}^{\infty} \mathbb{P}[Z{=}n]s^n = \sum_{n=0}^{\infty} \left(\frac{\lambda^n}{n!}e^{-\lambda}\right) s^n = e^{\lambda(s-1)}.$$

Ook deze kansgenererende functie hadden we uit de voorgaande af kunnen leiden: Vul in G_Y in: $p = \frac{\lambda}{n}$ en neem de limiet $n \to \infty$:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n = e^{\lambda(s-1)}.$$

12.4 Een toepassing: Veel dobbelstenen. Wat is de kans op m ogen bij een worp met n dobbelstenen?

In het voorgaande hebben we gezien dat, als X_1 kansgenererende functie G_1 heeft, en $X_1, X_2, \ldots, X_n$ onafhankelijk en gelijk verdeeld zijn, $S_n = X_1 + \ldots + X_n$ de kansgenererende functie

$$G_n(s) = G_1(s)^n$$

heeft.

Laten we voor X_1 nu het aantal ogen van een dobbelsteen nemen. Dan is

$$G_1(s) = \tfrac{1}{6}\left(s + s^2 + s^3 + s^4 + s^5 + s^6\right).$$

Het totale aantal ogen bij een worp met n dobbelstenen heeft dus kansgenererende functie $G_n(s) = G_1(s)^n$.

De kans op m ogen is de coëfficiënt van s^m hierin. Op het eerste gezicht is het bepalen van deze coëfficiënt net zo moeilijk als de oorspronkelijke vraag. Nu kunnen we echter iets doen: We kunnen schrijven

$$G_1(s) = \frac{s}{6} \cdot \frac{1 - s^6}{1 - s},$$

dus

$$G_n(s) = \left(\frac{s}{6}\right)^n \cdot \frac{(1 - s^6)^n}{(1 - s)^n}.$$

De factor $(1-s^6)^n$ kunnen we uitschrijven met de binomiaalformule van Newton, en de factor $(1-s)^{-n}$ kunnen we in een machtreeks ontwikkelen door de formule

$$\frac{1}{1-s} = \sum_{k=0}^{\infty} s^k$$

links en rechts $n-1$ maal te differentiëren, en te delen door $(n-1)!$. Er komt

$$\begin{aligned}
\frac{1}{(1-s)^n} &= \frac{1}{(n-1)!}\frac{d^n}{ds^n}\sum_{k=0}^{\infty} s^k \\
&= \frac{1}{(n-1)!}\sum_{k=n-1}^{\infty} k(k-1)\cdots(k-n+2)s^{k-n+1} \\
&= \sum_{k=n-1}^{\infty}\binom{k}{n-1}s^{k-n+1} \\
&= \sum_{l=0}^{\infty}\binom{l+n-1}{n-1}s^l.
\end{aligned}$$

Resultaat:

$$G_n(s) = \frac{s^n}{6^n}\left(\sum_{k=0}^{n}\binom{n}{k}(-s^6)^k\right)\left(\sum_{l=0}^{\infty}\binom{l+n-1}{n-1}s^l\right).$$

De coëfficiënt van s^m hierin vinden we door die termen bijeen te rapen waarvoor $n+6k+l=m$:

$$\mathbb{P}[S_n{=}m] = \frac{1}{6^n}\sum_{k=0}^{\left[\frac{m-n}{6}\right]}(-1)^k\binom{n}{k}\binom{m-6k-1}{n-1}.$$

Zo is bijvoorbeeld de kans om 24 ogen te gooien met 8 dobbelstenen gelijk aan

$$\begin{aligned}
\mathbb{P}[S_8{=}24] &= \frac{1}{6^8}\left(\binom{8}{0}\binom{23}{7} - \binom{8}{1}\binom{17}{7} + \binom{8}{2}\binom{11}{7}\right) \\
&= \frac{98813}{1679616} \sim 0,05883.
\end{aligned}$$

Het valt niet mee, dit resultaat door direct tellen te verkrijgen.

12.5 Een tweede toepassing: Het Sinterklaasprobleem. Een gezelschap van n personen trekt lootjes om te bepalen wie er voor wie een Sinterklaassurprise zal maken. We willen de kansverdeling weten van M, het aantal deelnemers dat hun eigen naam trekt (De kans dat $M=0$ heb je al berekend in opgave 2.5.).

Onze kansruimte Ω_n is de verzameling van de $n!$ permutaties $\omega : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$, waarbij $\omega(j)=i$ staat voor: persoon j trekt de naam van persoon i. Elke permutatie heeft evenveel kans; de kansmaat $A \mapsto \frac{\#(A)}{n!}$ geven we aan met $\mathbb{P}_n$, en middelen over $(\Omega_n, \mathbb{P}_n)$ geven we aan met $\mathbb{E}_n$.

Verder beschouwen we bij elke deelverzameling K van ons gezelschap (zeg $K \subset \{1, 2, \ldots, n\}$) de gebeurtenis A_K dat deze personen allemaal hun eigen naam trekken (geheel afgezien van wat de rest doet):

$$A_K = \{\omega \in \Omega_n | \forall_{j \in K} : \omega(j) = j\}.$$

Merk op dat $\mathbb{P}_n(A_K) = \frac{(n-\#(K))!}{n!}$, omdat er voor de overige $n - \#(K)$ personen $(n-\#(K))!$ mogelijkheden overblijven.
Zij verder $V(\omega)$ de vaste-punten-verzameling van ω:

$$V(\omega) = \{j \in \{1, 2, \ldots, n\} | \omega(j) = j\}.$$

We zijn geïnteresseerd in de stochast

$$M(\omega) = \#(V(\omega)).$$

We berekenen nu de kansgenererende functie G_n van M (bij deelname van n personen) met behulp van de volgend truc. Als een verzameling V m elementen heeft, dan is (ga na!)

$$(1 + x)^m = \sum_{k=0}^{m} \binom{m}{k} x^k = \sum_{K \subset V} x^{\#(K)}.$$

Laten we dit los op de vaste-punten-verzameling $V(\omega)$ van $\omega \in \Omega_n$, dan vinden we

$$(1 + x)^{M(\omega)} = \sum_{K \subset V(\omega)} x^{\#(K)} = \sum_{K \subset \{1,2,\ldots,n\}} \mathbf{1}_{A_K}(\omega) x^{\#(K)}.$$

En dus is

$$\begin{aligned} G_n(1 + x) &= \mathbb{E}\left((1 + x)^M\right) \\ &= \sum_{K \subset \{1,2,\ldots,n\}} \mathbb{P}(A_K) x^{\#(K)} \\ &= \sum_{k=0}^{n} \binom{n}{k} \cdot \frac{(n-k)!}{n!} x^k \\ &= \sum_{k=0}^{n} \frac{x^k}{k!} \end{aligned}$$

Conclusie:

$$\boxed{G_n(s) = \sum_{k=0}^{n} \frac{(s-1)^k}{k!}.}$$

De kans dat niemand zijn eigen naam trekt, vinden we terug als

$$\mathbb{P}_n[M{=}0] = G_n(0) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Ook zien we dat, voor $n \geq 2$,

$$\mathbb{E}_n(M) = G_n'(1) = 1;$$

$$\mathbb{V}\mathrm{ar}_n(M) = G_n''(1) + G_n'(1) - G_n'(1)^2 = 1.$$

Verder blijkt dat $\lim_{n\to\infty} G_n(s) = e^{s-1}$, de kansgenererende functie van de Poisson-verdeling met parameter 1.

Tenslotte kunnen we de hele kansverdeling van M uit onze formule voor G_n aflezen:

$$\begin{aligned}
\mathbb{P}_n[M{=}m] &= \sum_{k=0}^{n} \frac{1}{k!} \left(\text{coëfficiënt van } s^m \text{ in } (s{-}1)^k\right) \\
&= \sum_{k=m}^{n} \frac{1}{k!} \cdot (-1)^{k-m} \binom{k}{m} \\
&= \frac{1}{m!} \sum_{l=0}^{n-m} \frac{(-1)^l}{l!} \\
&= \frac{1}{m!} \left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \ldots \pm \frac{1}{(n{-}m)!}\right).
\end{aligned}$$

Opgaven

1. Zij N de wachttijd op het eerste succes in een rij onafhankelijke pogingen met slaagkans $p \in (0,1)$ en faalkans $q = 1 - p$.

   (a) Bereken de kansgenererende functie G_N van N.

   (b) Bereken hieruit $\mathbb{E}(N)$ en $\mathbb{V}\mathrm{ar}(N)$ (vergelijk met eerdere resultaten!).

   (c) Zij M de wachttijd op het tweede succes.
   Laat zien dat $G_M(s) = G_N(s)^2$.
   Kun je dit resultaat ook op een andere manier verklaren?

2. Een bepaalde soort bacteriën plant zich voort volgens het volgende schema: ieder uur loot ieder individu tussen twee mogelijkheden: zich delen in twee individuen (kans p) of sterven (kans $q = 1 - p$). Op tijdstip 0 is in ons kweekschaaltje één bacterie van deze soort aanwezig. Het aantal individuen na k uur noemen we: N_k.

   (a) Bereken de kansgenererende functies G_0, G_1 en G_2 van respectievelijk N_0, N_1, en N_2.

   (b) Laat zien dat voor alle $k \in \mathbb{N}$ geldt:

   $$G_{k+1}(s) = G_k(G_1(s))$$

   door te conditioneren naar N_k.

   (c) Toon aan dat de kans op uitsterven van de bacteriekolonie in ons kweekschaaltje gelijk is aan

   $$\lim_{k\to\infty} G_k(0) = \lim_{k\to\infty} \underbrace{G_1 \circ G_1 \circ \ldots \circ G_1}_{k \text{ keer}}(0).$$

   (d) Teken de grafiek van G_1. Laat in het plaatje zien wat de uitsterfkans is. Maak onderscheid tussen de gevallen $p > \frac{1}{2}$ en $p \leq \frac{1}{2}$.

Hoofdstuk 13

Markovketens

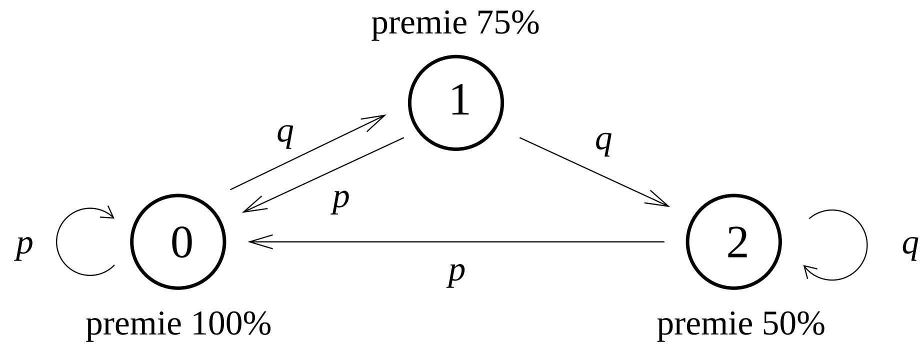
13.1 Inleiding. Een rij van verschijnselen kan meer of minder systematisch verlopen.
Het minst systematisch is een rij onafhankelijke proeven zoals bijvoorbeeld een rij worpen met een dobbelsteen. Om het resultaat van de volgende worp te voorspellen hebben we geen enkel houvast aan hetgeen de voorgaande worpen hebben opgeleverd.
Wat meer systeem vind je in een rij letters van een tekst. Zelfs als de taal waarin die tekst geschreven is je onbekend is, zie je toch dat na bepaalde letters bepaalde andere letters meer of minder waarschijnlijk zijn.
Dergelijke voorbeelden brachten Markov (Andrej Andrejewitsj Markov, 1856-1922) ertoe het volgende model voor te stellen:
Er is een verzameling van mogelijke toestanden (de *toestandsruimte* S, bijvoorbeeld de letters van het alfabet). Als het systeem zich op tijdstip n (nu) in toestand x bevindt, dan zal het zich op tijdstip $n+1$ in toestand y bevinden met kans $P(x,y)$. De overgangskans $P(x,y)$ hangt af van x, de toestand nu, maar niet van het verleden (toestanden op tijdstippen vóór n).
Deze aanname (de Markov-aanname) geldt niet zo goed voor de letters van een tekst (Als we onder de toestand niet de laatste letter maar de laatste drie letters verstaan, dan klopt het al wat beter). Er zijn echter veel voorbeelden waarbij de Markov-aanname een redelijke veronderstelling is. Er is een uitgebreide literatuur over de theorie en de toepassingen van het Markov-model. In dit hoofdstuk geven we een korte, enigszins intuïtieve inleiding.

13.2 Wat voorbeelden.

(a) No-claim-korting. Als je je auto verzekert bij General Accident N.V., dan betaal je voor het eerste jaar de volle premie (toestand 0). Heb je in een jaar geen schade gemeld dan betaal je voor het volgende jaar slechts 75% (25%

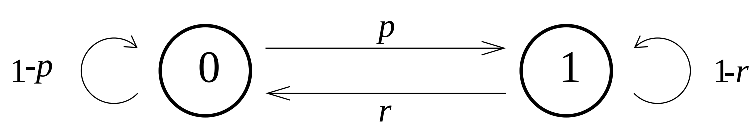
korting). Je bent dan in toestand 1. Heb je het jaar daarop weer geen schade te melden dan krijg je de maximale korting van 50% (toestand 2).

We nemen aan dat er ieder jaar een kans p is op meldenswaardige schade en dat je na een schademelding weer de volle premie moet betalen (je komt dan weer in toestand 0). Het volgend schema geeft de situatie weer ($q = 1 - p$):



We hebben hier een systeem met 3 toestanden. Toestandsruimte $S = \{0, 1, 2\}$. De pijlen geven de mogelijke overgangen aan. De bijbehorende overgangskansen staan er bij.

(b) Twee toestanden. Bij dit eenvoudige model met $S = \{0, 1\}$ hoort het volgende schema:

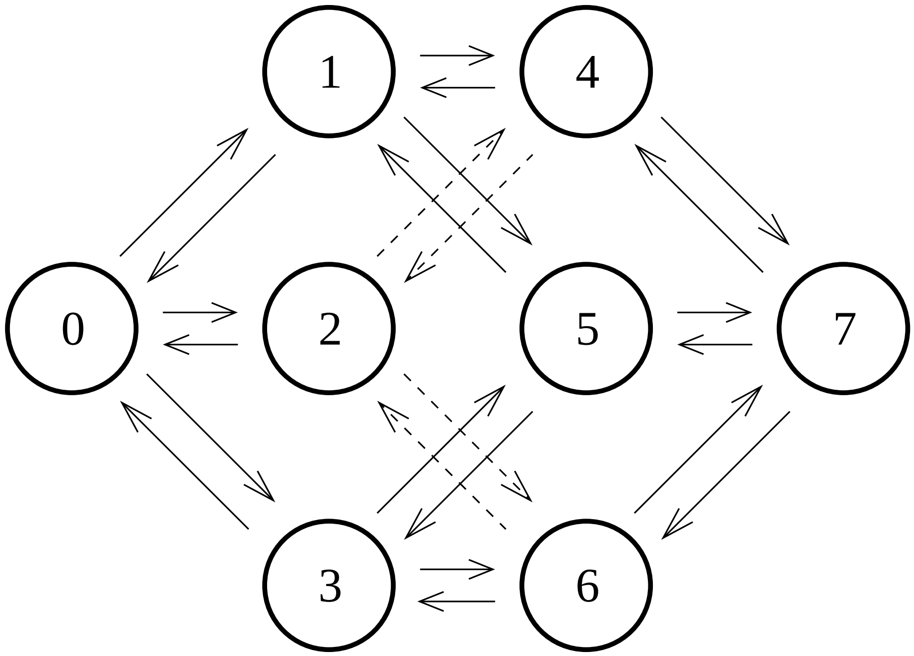


Er zijn allerlei interpretaties mogelijk:

1. Een apparaat (machine, voertuig,...) kan in goede conditie zijn: toestand 0, of niet: toestand 1. Er is iedere dag een kans p dat er een mankement zal optreden. Als dat gebeurt gaan we er direct mee naar de reparatieafdeling en dan is er iedere dag een kans r dat we het weer goed terugkrijgen.
2. Het weer. Toestand 0: droog. Toestand 1: regen. Zie hoofdstuk 4, opgave 11, pagina 90.

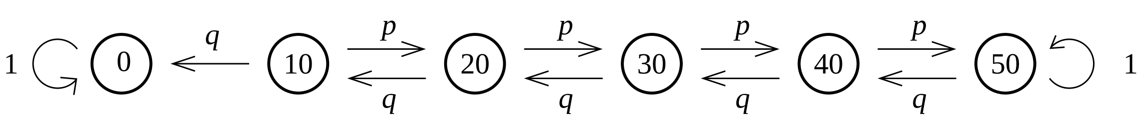
(c) Toevalswandeling op de hoekpunten van een kubus.
Maak via de ribben van een kubus een wandeling langs hoekpunten. Iedere keer wandel je van een hoekpunt naar een willekeurig aangrenzend hoekpunt (elk

aangrenzend hoekpunt heeft evenveel kans). $S = \{0, 1, \ldots, 7\}$. Langs iedere ribbe van de kubus lopen pijlen die de overgangskansen (steeds $\frac{1}{3}$) aangeven.



(d1) Gokken.

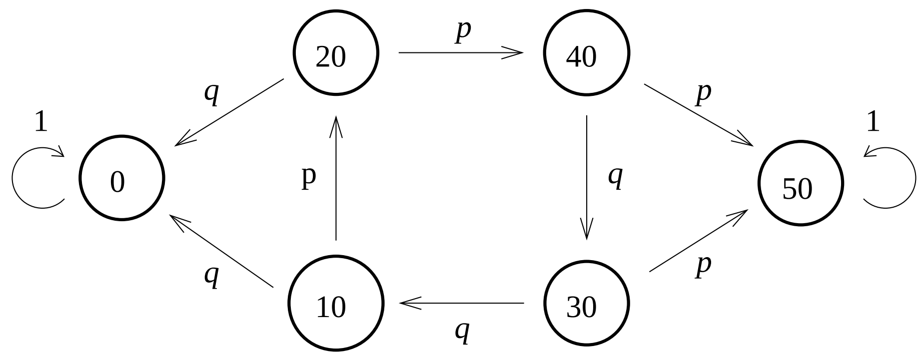
Je gaat het casino binnen met 20 gulden op zak. Je bent van plan te stoppen met spelen wanneer je 50 gulden hebt, of niets. Je zet telkens een tientje in op rood. Met kans p win je er een tientje bij; met kans $q = 1 - p$ ben je dat tientje kwijt.



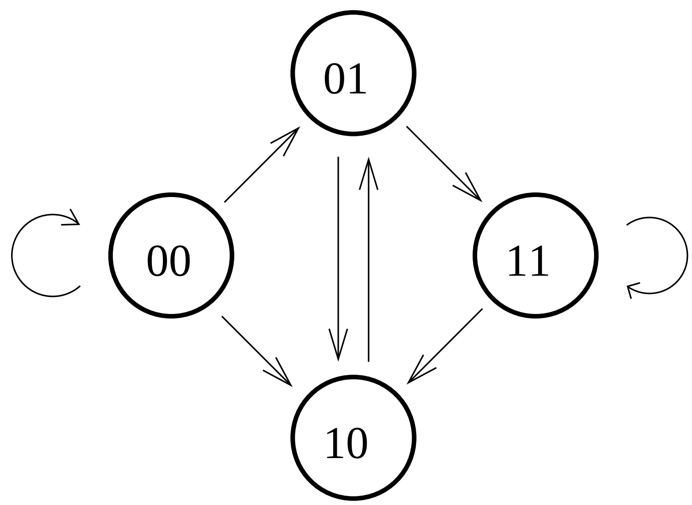
Hoe groot is de kans dat je de 50 gulden haalt?

(d2) Bold Play.

Volgens Dubins en Savage (How to Gamble if You Must 1965) heb je meer kans de 50 te halen wanneer je dat zo snel mogelijk probeert te bereiken (Hoe langer je speelt hoe meer het casino aan je verdient). Dus je zet meteen je twee tientjes in en vervolgens (zolang je geld hebt) zet je net zoveel in als nodig is om 50 te bereiken, of al je geld, als dat niet in één keer kan. Schema:



(e) In een rij worpen met een munt letten we steeds op de laatste twee uitkomsten. Schrijf 0 voor munt, 1 voor kop. $S = \{00, 01, 10, 11\}$ Schema:



Bij iedere pijl hoort een overgangskans van $\frac{1}{2}$.
Teken zelf een dergelijk schema voor het geval dat de drie laatste uitkomsten de toestand bepalen ($S = \{000, 001, \ldots, 111\}$).

13.3 Grondbeginselen, terminologie.

De toestandsruimte S. We beperken ons hier tot een eindige of hoogstens aftelbaar oneindige verzameling toestanden. Nummeren we de toestanden $0, 1, 2\ldots$ dan krijgen we $S = \{0, 1, \ldots, M\}$ of $S = \{0, 1, 2, \ldots\}$.

Discrete tijd. Verandering van toestand kan optreden tussen de tijdstippen 0, 1, 2,.... De toestand op tijdstip n geven we aan met X_n. De begintoestand wordt aangegeven door X_0. De rij stochasten $X_0, X_1, \ldots$ beschrijft de evolutie van het systeem (zoiets heet: stochastisch proces).

Overgangskansen. Vanuit toestand x kan het systeem overgaan naar toestand y met kans $P(x,y)$:

$$P(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x).$$

De overgangskansen vormen een matrix P, de *overgangsmatrix*. In voorbeeld (a) hebben we bijvoorbeeld

$$P = \begin{pmatrix} P(0,0) & P(0,1) & P(0,2) \\ P(1,0) & P(1,1) & P(1,2) \\ P(2,0) & P(2,1) & P(2,2) \end{pmatrix} = \begin{pmatrix} p & q & 0 \\ p & 0 & q \\ p & 0 & q \end{pmatrix}$$

Merk op dat voor iedere $x \in S$ geldt:

$$\sum_{y \in S} P(x,y) = 1.$$

Beginverdeling. De kansverdeling van X_0 geven we aan met π_0 (dit is een rijvector met elementen $\pi_0(x)$, $x \in S$). $\pi_0(x) = \mathbb{P}[X_0{=}x]$.
Beginverdeling en overgangsmatrix leggen de Markovketen vast. De kansverdeling van X_1 (aan te duiden met π_1) krijgen we als volgt:

$$\pi_1(x) = \mathbb{P}[X_1{=}x] = \sum_{u \in S} \pi_0(u) P(u,x).$$

In matrixnotatie:

$$\pi_1 = \pi_0 P.$$

Net zo krijgen we de verdeling π_2 van X_2 uit die van X_1:

$$\mathbb{P}[X_2{=}x] = \sum_{u \in S} \pi_1(u) P(u,x)$$

en dus

$$\pi_2 = \pi_1 P = (\pi_0 P) P = \pi_0 P^2.$$

Meerstapsovergangskansen. Het element $P^2(x,y)$ van de matrix P^2 geeft de kans op in twee stappen van x naar y te komen:

$$P^2(x,y) = \sum_{z \in S} P(x,z) P(z,y).$$

De regels van de matrixvermenigvuldiging en die van de kansrekening zijn hier helemaal met elkaar in overeenstemming.
Fundamenteel voor de theorie van de Markovketens is de volgende formule:

$$\mathbb{P}[X_0{=}x_0, X_1{=}x_1, \ldots, X_n{=}x_n] = \pi_0(x) P(x_0,x_1) \cdots P(x_{n-1},x_n).$$

Deze formule berust op de *Markovaanname*, die we zo kunnen formuleren: Voor ieder rijtje toestanden $x_0, \ldots, x_n$ geldt:

$$\mathbb{P}(X_n{=}x_1|X_0{=}x_0, \ldots, X_{n-1}{=}x_{n-1}) = \mathbb{P}(X_n{=}x_n|X_{n-1}{=}x_{n-1}) = P(x_{n-1}, x_n).$$

Neem bijvoorbeeld $n = 2$, dan geldt altijd (zie de rekenregels voor voorwaardelijke kansen)

$$\mathbb{P}[X_0{=}x_0, X_1{=}x_1, X_2{=}x_2] =$$

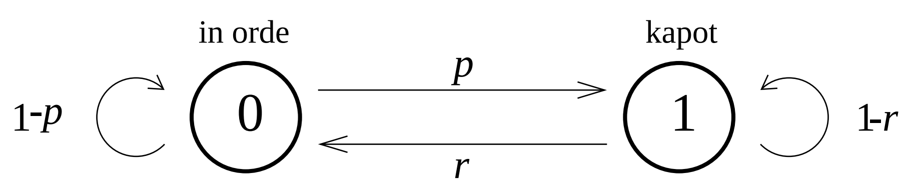
$$\mathbb{P}[X_0{=}x_0]\mathbb{P}(X_1{=}x_1|X_0{=}x_0)\mathbb{P}(X_2{=}x_2|X_0{=}x_0, X_1{=}x_1).$$

Dankzij de Markovaanname kunnen we de laatste factor

$$\mathbb{P}(X_2{=}x_2|X_0{=}x_0, X_1{=}x_1)$$

vervangen door $\mathbb{P}(X_2{=}x_2|X_1{=}x_1) = P(x_1, x_2)$.

Kansen bij een gegeven begintoestand. In het bijzondere geval dat de beginverdeling π_0 geconcentreerd is in één enkele toestand x zullen we de kans op iets, zeg A, aangeven met $\mathbb{P}_x(A)$, dus: $\mathbb{P}_x(A) = \mathbb{P}(A|X_0{=}x)$. Zo krijg je bijvoorbeeld $\mathbb{P}_x(X_1{=}y) = P(x, y)$ en $\mathbb{P}_x(X_n{=}y) = P^n(x, y)$.



13.4 De keten met twee toestanden.(zie interpretatie 2 bij voorbeeld b uit paragraaf 13.2)

In dit geval kunnen we de kans $\mathbb{P}[X_n{=}0]$ dat de machine na n dagen nog (of weer) heel is expliciet berekenen, gegeven de kans $\pi_0(0) = \mathbb{P}[X_0{=}0]$. Dat gaat zo:

$$\begin{aligned}
\mathbb{P}[X_{n+1}{=}0] &= \mathbb{P}[X_n{=}0]P(0,0) + \mathbb{P}[X_n{=}1]P(1,0) \\
&= \mathbb{P}[X_n{=}0](1-p) + (1 - \mathbb{P}[X_n{=}0])r \\
&= (1-p-r)\mathbb{P}[X_n{=}0] + r.
\end{aligned}$$

Voor het gemak schrijven w: $a_n = \mathbb{P}[X_n{=}0]$; $c = 1 - p - r$. Dan staat er:

$$a_{n+1} = ca_n + r.$$

Deze recursie heeft als vaste waarde $\frac{r}{1-c}$ (als $a_n = \frac{r}{1-c}$ dan is ook $a_{n+1} = \frac{r}{1-c}$). Voor de afwijking van a_n t.o.v. deze vaste waarde geldt:

$$a_{n+1} - \frac{r}{c-1} = c\left(a_n - \frac{r}{1-c}\right)$$

waaruit volgt dat

$$a_n = \frac{r}{1-c} + c^n\left(a_0 - \frac{r}{1-c}\right).$$

Met $1-c = p+r$, $a_n = \mathbb{P}[X_n{=}0]$ en $a_0 = \pi_0(0)$ wordt dit:

$$\mathbb{P}[X_n{=}0] = \frac{r}{p+r} + c^n\left(\pi_0(0) - \frac{r}{p+r}\right) \tag{1}$$

Op dezelfde manier vinden we voor $\mathbb{P}[X_n{=}1]$ (verwissel r en p, en ook 1 en 0):

$$\mathbb{P}[X_n{=}1] = \frac{p}{p+r} + c^n\left(\pi_0(1) - \frac{p}{p+r}\right) \tag{2}$$

Schrijven we $\pi = \left(\frac{r}{p+r}, \frac{p}{p+r}\right)$ en, zoals afgesproken, π_n voor de verdeling van X_n dan kunnen de formules (1) en (2) in vectornotatie gecombineerd worden tot

$$\pi_n = \pi + c^n(\pi_0 - \pi) \tag{3}$$

Hieruit blijkt het volgende:

(a) Bij beginverdeling $\pi_0 = \pi$ krijgen we: $\pi_n = \pi$ voor iedere n. De verdeling π noemt men de *evenwichtsverdeling* (ook wel *stationaire verdeling*). Deze verdeling voldoet aan

$$\pi P = \pi.$$

In lineaire-algebra-taal: de rijvector π is een links-eigenvector met eigenwaarde 1 voor de matrix P (Een andere eigenvector is $(1,-1)$. Deze heeft eigenwaarde $c = 1-p-r$. Ga na dat $\pi_0 - \pi$ een veelvoud is van $(1,-1)$).

(b) Convergentie naar de evenwichtsverdeling. Omdat $|c| = |1-p-r| < 1$ (tenzij $p = r = 0$ of $p = r = 1$) geldt dat $\pi_n \to \pi$ als $n \to \infty$. Dus, onafhankelijk van de beginverdeling, zullen we het systeem op den duur in toestand x aantreffen met kans $\pi(x)$ in overeenstemming met de evenwichtsverdeling.

(c*) Voor de n-staps overgangsmatrix P^n krijgen we:

$$P^n = \frac{1}{p+r}\left[\begin{pmatrix} r & p \\ r & p \end{pmatrix} + c^n \begin{pmatrix} p & -p \\ -r & r \end{pmatrix}\right].$$

Afleiding: De bovenste rij van P^n:

$$(P^n(0,0), P^n(0,1)) = (\mathbb{P}_0(X_n{=}0), \mathbb{P}_0(X_n{=}1))$$

krijgen we door in formule (3) in te vullen: $\pi_0 = (1,0)$. De onderste rij van P^n wordt gegeven door formule (3) met $\pi_0 = (0,1)$.
We zien dat P^n convergeert naar een matrix met rijen die gelijk zijn aan de evenwichtsverdeling.

13.5 Praktische betekenis van de evenwichtsverdeling. Stel dat in een fabriek N machines in gebruik zijn die zich ieder gedragen volgens het schema van voorbeeld (b) uit paragraaf 13.2 met kans $p = 0{,}1$ om kapot te gaan en $r = 0{,}5$ om, kapotzijnd, weer na één dag gerepareerd te worden: Van de machines in de reparatieafdeling wordt gemiddeld per dag de helft weer in goede staat terug gebracht.
Na zekere tijd zijn $n(0)$ machines in bedrijf (toestand 0) en $n(1)$ zijn er kapot (toestand 1). We kunnen van evenwicht spreken als de in- en uitgaande stromen van de fabriekshal naar de reparatieafdeling en terug even groot zijn. Dus het aantal machines dat op een dag kapot gaat (gemiddeld $n(0)p$) is gelijk aan het aantal machines dat op een dag gerepareerd terugkomt (gemiddeld $n(1)r$).
De verhouding $n(0) : n(1) = r : p$ is dan in overeenstemming met de evenwichtsverdeling. Als $p = 0{,}1$ en $r = 0{,}5$ dan zal gemiddeld een fractie $\frac{p}{r+p} = \frac{1}{6}$ van de machines in reparatie zijn.
In voorbeeld (a) (verzekering met no-claim-korting) zegt de evenwichtsverdeling ons iets over welk deel van de verzekerden op den duur hoeveel procent korting krijgt.

Opgave 1. Bepaal de evenwichtsverdeling bij voorbeeld (a) als $p = 0{,}2$.

Opgave 2. Voorbeeld (e). Schrijf de overgangsmatrix P op. Bepaal P^n voor $n = 2, 3, \ldots$, en de evenwichtsverdeling.

13.6 Evenwicht en convergentie naar evenwicht. In deze paragraaf bespreken we een paar fundamentele stellingen over evenwichtsverdelingen bij eindige Markovketens (Markovketens met eindige toestandsruimte). De uitspraken van deze stellingen zijn eenvoudig. De bewijzen zijn minder eenvoudig. Daarvoor verwijzen we naar de literatuur.

Stelling 1. Een eindige Markovketen heeft minstens één evenwichtsverdeling.

In lineaire-algebra-taal (laat $S = \{1, 2, \ldots, n\}$):

Bij een willekeurige $n \times n$ overgangsmatrix P is er minstens één $1 \times n$ rijvector $\pi = (\pi(1), \ldots, \pi(n))$ met $\sum_{k=1}^{n} \pi(k) = 1$, $\pi(k) \geq 0$ voor $k = 1, 2, \ldots, n$, zó dat

$$\pi = \pi P.$$

Voorbeeld 1. $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\pi = (\frac{1}{2}, \frac{1}{2})$ is de enige evenwichtsverdeling.

Voorbeeld 2. $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Voor iedere p, $0 \leq p \leq 1$, voldoet $\pi = (p, 1-p)$ aan $\pi = \pi P$.

De volgende voorwaarde zorgt ervoor dat er hoogstens één evenwichtsverdeling is.
Definitie: Een Markovketen heet *irreducibel* als bij ieder tweetal toestanden x en y er een n is zó dat $P^n(x, y) \neq 0$.
Dit betekent dat iedere toestand vanuit iedere andere toestand bereikbaar moet zijn.

Stelling 2. Een eindige Markovketen die irreducibel is heeft precies één evenwichtsverdeling.

13.7 Convergentie. In paragraaf 13.4 hebben we gezien dat bij de Markovketen met 2 toestanden de verdeling van X_n convergeert naar de evenwichtsverdeling: $\pi_n(x) = \mathbb{P}[X_n{=}x] \to \pi(x)$ als $n \to \infty$, ongeacht de beginverdeling.
In het bijzonder: $P^n(z, x) = \mathbb{P}_z(X_n{=}x) \to \pi(x)$ voor iedere $z \in S$. De n-staps overgangsmatrix P^n convergeert dus naar de matrix waarvan alle rijen overeenstemmen met de evenwichtsverdeling.

Zoiets geldt niet altijd. Bijvoorbeeld als $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ dan is $\pi = (\frac{1}{2}, \frac{1}{2})$ maar P^n blijft altijd heen en weer flippen tussen $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ en $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Nog een voorbeeld: $P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $S = \{1, 2, 3, 4\}$.

Teken het pijlenschema. Je ziet: deze Markovketen is irreducibel. De evenwichtsverdeling heeft de vorm $\pi = (\frac{1}{3}, \cdot, \cdot, \cdot)$. P^n convergeert niet. Bijvoorbeeld: $P^n(1, 1) = 1$ als n een 3-voud is, en $P^n(1, 1) = 0$ als n geen drievoud is.
Dit voorbeeld lijdt aan *periodiciteit*.

Definitie: Een Markovketen met toestandsruimte S en overgangsmatrix P noemen we *periodiek* als S is op te splitsen in $d \geq 2$ disjuncte delen $S_1, S_2, \ldots, S_d$, zó dat

$P(x, y) \neq 0$ alleen kan voorkomen als $x \in S_i$ en $y \in S_{i+1}$ ($i = 1, 2, \ldots, d-1$), of $x \in S_d$ en $y \in S_1$.
De grootste d waarbij zo'n opsplitsing mogelijk is noemen we de *periode* van de Markovketen.

De toestand maakt steeds een rondreis door $S_1, S_2, \ldots, S_d, S_1, \ldots$. Dit impliceert dat terugkeer in dezelfde toestand alleen mogelijk is in een aantal stappen dat een veelvoud is van d.
Noem k een *terugkeertijd* van toestand x als $P^k(x, x) \neq 0$. Dan zijn alle terugkeertijden (van alle toestanden) veelvouden van d. Anders gezegd: De periode d van een Markovketen is de grootste gemene deler van de terugkeertijden.
Een Markovketen heet *aperiodiek* als er voor geen enkele $d > 1$ een opsplitsing zoals zojuist beschreven mogelijk is. De grootste gemene deler van de terugkeertijden is dan 1.

Stelling 3. In een eindige Markovketen die irreducibel en aperiodiek is, geldt: $\pi_n \to \pi$ als $n \to \infty$, d.w.z. de verdeling van X_n convergeert naar de evenwichtsverdeling en P^n convergeert naar de matrix met rijen die allemaal gelijk zijn aan de evenwichtsverdeling.

Opgave 3. Voorbeeld (c) van paragraaf 13.2. Is deze Markovketen aperiodiek?

13.8 Hoe vinden we de evenwichtsverdeling? Laat P de $n \times n$ overgangsmatrix zijn van een irreducibele Markovketen. De eerste manier om π te vinden ligt voor de hand: Los op: $\pi P = \pi$, n vergelijkingen met n onbekenden.
De tweede manier is soms handig als het pijlenschema van de toestandsruimte niet te ingewikkeld is. Denk aan de toestandsruimte als een verzameling plaatsen (honderuggen bijvoorbeeld). Op plaats x zitten $n(x)$ vlooien. Het gaat in totaal over $N = \sum_{x \in S} n(x)$ vlooien, behoorlijk veel.
Op de tijdstippen $0, 1, 2, \ldots$ (of liever: net na ieder van die tijdstippen) springen al die vlooien op. Een vlo, opgesprongen uit x landt in y met kans $P(x, y)$.
Verdeel S in twee stukken S_0 en S_1 (disjunct). Als er evenwicht is, dan zullen bij zo'n springpartij gemiddeld evenveel vlooien uit S_0 vertrekken naar S_1 als er aankomen uit S_1: *Instroom = Uitstroom.*
Van S_0 vertrekken gemiddeld

$$\sum_{x \in S_0, y \in S_1} n(x) P(x, y)$$

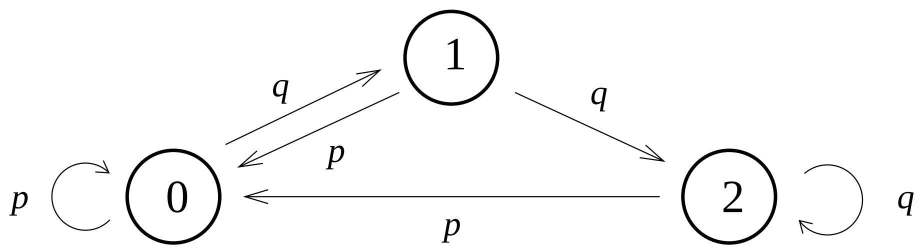
vlooien naar S_1, en van S_1 vertrekken gemiddeld

$$\sum_{x \in S_1, y \in S_0} n(x)P(x,y)$$

vlooien naar S_0. Dit levert de relatie

$$\sum_{x \in S_0, y \in S_1} n(x)P(x,y) = \sum_{x \in S_1, y \in S_0} n(x)P(x,y).$$

Door een aantal van zulke opsplitsingen geschikt te kiezen krijg je een aantal relaties waaruit de aantallen $n(x)$ te bepalen zijn, en daaruit, door normeren, de evenwichtskansen.
Voorbeeld: Voorbeeld (a) uit paragraaf 13.2. $S = \{0, 1, 2\}$, schema:



Met $S_0 = \{1\}$ krijg je: $n(1)p + n(1)q = n(0)q$, dus:

$$n(1) = n(0)q.$$

Met $S_0 = \{0, 1\}$:

$$n(1)q = n(2)p.$$

Uit deze twee relaties volgt dat

$$n(0) : n(1) : n(2) = p : pq : q^2.$$

13.9 Gemiddelde terugkeertijd. In een irreducibele Markovketen kun je vanuit iedere toestand naar iedere andere toestand. De stochast T_x die het aantal stappen telt waarin x voor het eerst bereikt wordt, heet de *aankomsttijd* in x (arrival time, hitting time). Algemener: Voor een deelverzameling A van S is de aankomsttijd in A:

$$T_A = \min\{n : n \geq 1, X_n \in A\}.$$

Wordt A nooit bereikt ($X_n \notin A$ voor alle $n \geq 1$) dan $T_A = \infty$.
Bestaat A uit een enkele toestand, $A = \{x\}$, dan schrijven we T_x in plaats van $T_{\{x\}}$.

Bij begintoestand x noemen we T_x de *terugkeertijd* in x.

Voor een irreducibele eindige Markovketen geldt:

Stelling 4. Voor iedere x en y in S geldt $\mathbb{P}_x(T_y{<}\infty) = 1$.
In het bijzonder: $\mathbb{P}_x(T_x,\infty) = 1$ en voor de verwachting van T_x bij begintoestand $X_0 = x$ geldt:

$$m(x) := \mathbb{E}_x(T_x) = \sum_{k=1}^{\infty} k\mathbb{P}_x(T_x{=}k) = \sum_{k=0}^{\infty} \mathbb{P}_x(T_x{>}k)$$

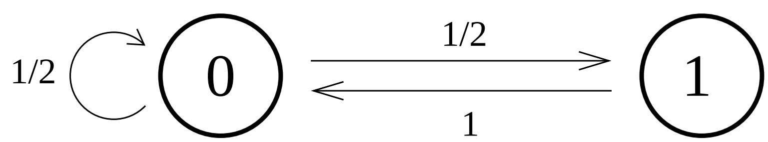
en:

$$m(x) = \frac{1}{\pi(x)},$$

waarbij $\pi(x)$ de kans op toestand x bij de evenwichtsverdeling is.

Dit is een prachtige eigenschap van de eindige irreducibele Markovketen: Iedere toestand komt steeds weer terug. De gemiddelde terugkeertijd is korter of langer naarmate de evenwichtskans van die toestand groter of kleiner is.

Voorbeeld: Twee toestanden, schema:

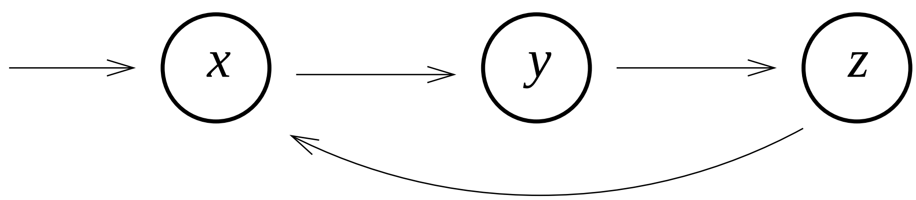


Terugkeer in toestand 0 kan alleen in 1 of 2 stappen. De gemiddelde terugkeertijd van 0 is dus $m(0) = \frac{3}{2}$. Dus: $\pi(0) = \frac{2}{3}$. Dan blijft er $\pi(1) = \frac{1}{3}$ over voor toestand 1. Dit zou betekenen dat toestand 1 een gemiddelde terugkeertijd van $m(1) = 3$ heeft. Ga na dat dit klopt.

We merken nog op dat de uitspraken van stelling 4 niet altijd waar zijn bij een irreducibele Markovketen met (aftelbaar) oneindige toestandsruimte. Daar gelden ze alleen als er een evenwichtsverdeling is. Die hoeft er echter niet te zijn. Het kan dan gebeuren dat terugkeer niet zeker is: $\mathbb{P}_x(T_x{<}\infty) < 1$. Dan heet x *voorbijgaand* (transient). Als terugkeer zeker is, $\mathbb{P}_x(T_x{<}\infty) = 1$, toestand x is *terugkerend* (recurrent), dan hoeft $m(x)$ nog niet eindig te zijn.
Voorbeeld (zonder bewijs): Een stochastische wandeling in $\mathbb{Z}$. $P(x, x{+}1) = p$, $P(x, x{-}1) = q$, $p + q = 1$. Als $p \neq q$ dan is iedere toestand voorbijgaand. Als $p = \frac{1}{2}$ dan is iedere toestand terugkerend met oneindig gemiddelde.

13.10 Fuiken en vangstkansen. Als je je geld verspeeld hebt (in het casino bijvoorbeeld) kun je niet verder spelen, want je hebt niets meer om in te zetten. Vanuit toestand 0 (in de voorbeelden (d1) en (d2) van paragraaf 13.2) kun je nergens anders heen: $P(0, x) = 0$ voor $x \neq 0$ en $P(0, 0) = 1$. Zo'n toestand heet *absorberend*. x is een absorberende toestand betekent dus: $P(x, x) = 1$.
In het diagram kun je voor het gemak de pijl (lus) van x naar x weglaten, dus in plaats van

Een groepje toestanden $F \subset S$ wordt een *fuik* genoemd (Engels: closed set) als $P(x, y) = 0$ voor iedere $x \in F$, $y \notin F$.
Een *priemfuik* (irreducible closed set) is een fuik die geen kleinere fuik bevat. Een fuik die maar één toestand bevat (een absorberende toestand) is vanzelf een priemfuik. Een priemfuik die uit meerdere toestanden bestaat kan er bijvoorbeeld zó uitzien:



Bij een irreducibele Markovketen is de hele toestandsruimte S de enige priemfuik. In een eindige Markovketen die niet irreducibel is zijn er altijd een of meer priemfuiken. Het systeem komt dan vroeg of laat in een van die priemfuiken terecht.
In de voorbeelden (d1) en (d2) van paragraaf 13.2 komt de speler op den duur in één van de priemfuiken $\{0\}$ of $\{50\}$.

Het is van belang de bijbehorende kansen te weten. De kans om, startend in x, uiteindelijk in priemfuik F terecht te komen noemen we de *vangstkans* van F, notatie $v(x, F)$, of korter $v(x)$ als we weten over welke F we het hebben.

De functie v voldoet aan:

(a) $v(x) = 1$ als $x \in F$, $v(x) = 0$ als x in een andere priemfuik. Dit zijn de "randvoorwaarden".

(b) $v(x) = \sum_y P(x,y)v(y)$ (matrixnotatie: $v = Pv$).

Bij een eindige Markovketen is v door (a) en (b) eenduidig vastgelegd. We passen dit toe op het probleem van de gokker uit paragraaf 13.2:

13.11 Ruïneringsprobleem (Gamblers Ruin).

Neem het schema van voorbeeld (d1).

De kans om te winnen, de succeskans, is (bij begintoestand x) de vangstkans $v(x)$ van toestand $\{50\}$.

Er geldt:

$$\begin{cases} \text{(a)} \quad v(0) = 0,\ v(5) = 1 \\ \text{(b)} \quad v(x) = qv(x{-}10 + pv(x{+}10) \end{cases}$$

Als $p\frac{1}{2}$ dan volgt uit $v(x) = \frac{1}{2}v(x{-}10) + \frac{1}{2}v(x{+}10)$ dat de grafiek van de functie v rechtlijnig is en dus $v(x) = \frac{x}{50}$. In het bijzonder: $v(20) = \frac{20}{50} = 0{,}4$.

Dit is eerlijk spel. De verwachting aan het eind is gelijk aan het kapitaal van de speler in het begin.

Voor $p \neq \frac{1}{2}$ krijgen we

$$lv(x) = (p{+}q)v(x) = pv(x{+}1) + qv(x{-}1)$$

dus

$$v(x{+}1) - v(x) = \frac{q}{p}(v(x) - v(x{-}1))$$

Schrijf even $a(x) = v(x) - v(x{-}1)$ en $r = \frac{q}{p}$, dan staat hier:

$$a(x) = ra(x{-}1).$$

Dit betekent dat

$$\begin{aligned} v(x) &= v(0) + a(1) + a(2) + \ldots + a(x) \\ &= 0 + a(1) + ra(1) + \ldots + r^{x-1}a(1) \\ &= a(1) \cdot \frac{r^x - 1}{r - 1}. \end{aligned}$$

Uit de randvoorwaarde $v(5) = 1$ volgt: $a(1) = \frac{r-1}{r^5-1}$ en dus:

$$v(x) = \frac{r^x - 1}{r^5 - 1} = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^5 - 1}.$$

Bij $p = \frac{18}{37}$, $q = \frac{19}{37}$ (roulette: 18 rode en 18 zwarte nummers waar je op kunt inzetten en één nul, waarbij alle inzetten naar het casino gaan) krijgen we $v(2) = \frac{(19/18)^2-1}{(19/18)^5-1} =$

0,368.
Bij de tweede manier van spelen (schema (d2)) krijgen we

$$\begin{cases} v(2) = pv(4) \\ v(4) = p + qv(3) \\ v(3) = p + qv(1) \\ v(1) = pv(2) \end{cases} \text{randvoorwaarden:} v(0) = 0, v(5) = 1$$

Voor $v(2)$ krijgen we dan $v(2) = \frac{p^2+p^2q}{1-p^2q^2}$.
Met $p = \frac{18}{37}$ en $q = \frac{19}{37}$: $v(2) = 0{,}382$.
Volgens Dubins en Savage is dit de maximale kans om te winnen (Bold play is optimal in subfair casinos).
Het verschil tussen de speelwijzen van (d1) en (d2) wordt duidelijker wanneer het om grotere bedragen gaat (zie opgave 5).

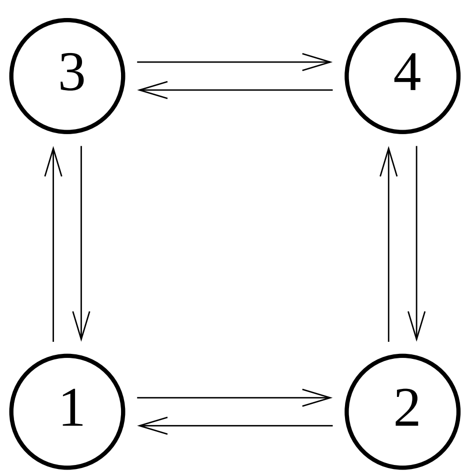
13.12 Hoe lang gaat het spel duren? Beginnend in toestand x zal het systeem na T stappen in één van de absorberende priemfuiken terechtkomen.
Algemener: Laat $T := T_A$ de aankomsttijd zijn in een verzameling $A \subset S$ (bij het gokprobleem bijvoorbeeld: $A = \{0, 50\}$).
Voor de verwachting $u(x) := \mathbb{E}_x(T)$ geldt:

(a) $u(x) = 0$ als $x \in A$;

(b) $u(x) = 1 + \sum_{y \in S} P(x, y)u(y)$ als $x \notin A$.

Uit deze relatie kan $u(x)$ berekend worden. In voorbeeld (d1) krijgen we als $p = \frac{1}{2}$: $u(x) = \frac{x(50-x)}{100}$.

Ander voorbeeld: Toevalswandeling op het vierkant.



Hoe lang duurt het tot je in 4 aankomt, als je begint in 1 (de overgangskansen zijn allemaal $\frac{1}{2}$)?
Uit

$$\begin{cases} u(1) = 1 + \frac{1}{2}u(2) + \frac{1}{2}u(3) \\ u(2) = 1 + \frac{1}{2}u(1) \\ u(3) = 1 + \frac{1}{2}u(1) \end{cases}$$

volgt: $u(1) = 4$; $u(2) = u(3) = 3$.

Met de volgende truc, bedacht door David van Danzig (Amsterdam (1948)) kunnen we ook de voortbrengende functie $G(s)$ van T afleiden.
Het idee is $G(s) := \mathbb{E}_x(s^T) = \sum_{k=1}^{\infty} s^k \mathbb{P}_x(T{=}k)$ te interpreteren als een kans. Hierbij is s de kans dat een stap veilig verloopt (s = safe). Bij iedere stap van het proces kan er een ongeluk gebeuren. Er ontploft een bom, of iemand krijgt een woedeaanval en slaat alles kort en klein.
$G(s)$ is dan de kans dat het goed afloopt. Immers, wanneer er $T = k$ stappen nodig zijn om A te bereiken, dan is er kans s^k dat die alle k veilig gedaan worden.
Met deze interpretatie krijgen we voor $a_x := \mathbb{E}_x(s^T)$:

(a) $a_x = 1$ als $x \in A$;

(b) $a_x = \sum_{y \in S} P(x,y) s a_y$ als $x \notin A$.

Voor de toevalswandeling op het vierkant krijgen we voor de aankomsttijd T in 4, beginnende in x:

$$\begin{cases} a_1 = \frac{1}{2}sa_2 + \frac{1}{2}sa_3 \\ a_2 = \frac{1}{2}sa_1 + \frac{1}{2}s \\ a_3 = \frac{1}{2}sa_1 + \frac{1}{2}s \end{cases}$$

Hieruit volgt $a_1 = G(s) = \frac{\frac{1}{2}s^2}{1-\frac{1}{2}s^2}$ waarmee we de hele kansverdeling van T te pakken hebben.

Opgaven

4. $S = \{1, 2, 3\}$, $P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$ Is de Markovketen met deze overgangsmatrix irreducibel?

   Toch is er maar één evenwichtsverdeling. Bepaal die.

5. Rood en zwart.

   Het verschil tussen de beide manieren van spelen (voorbeelden (d1) en (d2)) wordt groter naarmate het om meer geld gaat. Bepaal de succeskansen als je begint met 200 gulden, stoppen bij 500 of 0

   (a) als je iedere keer een tientje inzet;

   (b) met "bold play".

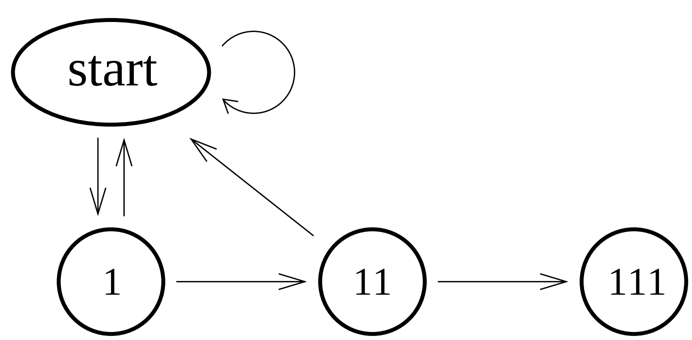
6. Voorbeeld (c) uit paragraaf 13.2. Bij iedere sprong verplaatst het beestje zich van een hoekpunt van de kubus naar een van de drie aangrenzende hoekpunten (met gelijke kansen). De reis begint in hoekpunt 1. Na T sprongen wordt voor het eerst het tegenoverliggende hoekpunt bezocht.

   (a) Bepaal $\mathbb{E}(T)$.

   (b) Bepaal de voortbrengende functie van T.

7. Een reeks worpen met een munt.

   Na T worpen heb ik voor het eerst een rijtje van 3 keer achter elkaar kop. Bepaal $\mathbb{E}(T)$. Schema (munt = 0, kop = 1, bij iedere pijl hoort kans $\frac{1}{2}$):



   Bepaal ook de verwachting van het aantal worpen dat nodig is om voor het eerst een rijtje 010 te krijgen.

8. Speler A wint als in de rij worpen met de munt het rijtje 111 eerder komt dan het rijtje 010 van speler B. Bepaal de kans dat A wint.

Appendix A

Tabel voor de normale verdeling

Voor de cijfers moet “0,” worden gezet. 637 betekent dus: 0,637.

x	0	1	2	3	4	5	6	7	8	9
0,0	500	504	508	512	516	520	524	528	532	536
0,1	540	544	548	552	556	560	564	567	571	575
0,2	579	583	587	591	595	599	603	606	610	614
0,3	618	622	626	629	633	637	641	644	648	688
0,4	655	659	663	666	670	674	677	681	684	688
0,5	691	695	698	702	705	709	712	716	719	722
0,6	726	729	732	736	739	742	745	749	752	755
0,7	758	761	764	767	770	773	776	779	782	785
0,8	788	791	794	797	800	802	805	808	811	813
0,9	816	819	821	824	826	829	831	834	836	839
1,0	841	844	846	848	851	853	855	858	860	862
1,1	864	867	869	871	873	875	877	879	881	883
1,2	885	887	889	891	893	894	896	898	900	901
1,3	903	905	907	908	910	911	913	915	916	918
1,4	919	921	922	924	925	926	928	929	931	932
1,5	933	934	936	937	938	939	941	942	943	944
1,6	945	946	947	948	949	951	952	953	954	954
1,7	955	956	957	958	959	960	961	962	962	963
1,8	964	965	966	966	967	968	969	969	970	971
1,9	971	972	973	973	974	974	975	976	976	977
2,0	977	978	978	979	979	980	980	981	981	982
2,1	982	983	983	983	984	984	985	985	985	986
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9952	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986

Appendix B

Tentamenopgaven

1. Een machine schroeft doppen op flessen met een gemiddeld draaimoment van 24 eenheden en een standaardafwijking van 3 eenheden. De flessen zijn gemiddeld bestand tegen een draaimoment van 35 eenheden met een standaardafwijking van 4 eenheden. Bepaal het percentage breuk, aannemende dat de verschillende grootheden onafhankelijk en normaal verdeeld zijn.

2. We beschouwen twee willekeurige personen X en Y, beide geboren ná het jaar 1900.

   (a) Bereken $\mathbb{P}[X \text{ jarig op 29 februari}]$ en $\mathbb{P}[X \text{ jarig op 1 april}]$.

   (b) Bereken $\mathbb{P}[X \text{ en } Y \text{ op verschillende dagen jarig}]$.

   Houd hierbij rekening met schrikkeljaren.

3. Zij $\Omega = \{1, 2, 3, 4, 5, 6\}$ en $\mathbb{P}\{j\} = \frac{1}{6}$ voor elk punt j van Ω. Laten A, B en C drie gebeurtenissen (deelverzamelingen van Ω) zijn, alle drie verschillend van $\emptyset$ en Ω. Bewijs:

   (a) als A en B onafhankelijk zijn, dan heeft één van beide kans $\frac{1}{2}$, en de ander $\frac{1}{3}$ of $\frac{2}{3}$.

   (b) A, B en C zijn niet onafhankelijk.

4. De verdeling F van een stochast Y wordt gegeven door

$$F(y) = \begin{cases} 0, & \text{als } y < 0, \\ \frac{2}{\pi} \arcsin \sqrt{y}, & \text{als } 0 \leq y \leq 1, \\ 1, & \text{als } y > 1, \end{cases}$$

   waarbij arcsin de inverse is van de functie sin, beperkt tot $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

   (a) Bereken $\mathbb{E}(Y)$.

(b) Bereken $\mathbb{P}[\frac{1}{4} \leq Y \leq \frac{3}{4}]$.

Achtergrond (géén noodzakelijke informatie voor het maken van de opgave): Als Peter en Paul $2n$ onafhankelijke gokspelen met gelijke kans op winst spelen om inzet 1, en Y_n de fractie van de tijd tot $2n$ is dat Peter vóór staat, dan is de hierboven gegeven verdeling van Y de limietverdeling van Y_n voor $n \to \infty$.

5. (a) In een vaas zitten n rode en n groene ballen ($n \in \mathbb{N}$). Telkens wordt blindelings een bal uit de vaas getrokken totdat van één der twee kleuren alle ballen zijn getrokken. Hoe groot is de kans dat de vaas dan nog k ballen bevat ($k \in \{1, 2, \ldots, n\}$)?

   (b) Er zijn twee doosjes met elk n lucifers ($n \in \mathbb{N}$). Telkens wordt blindelings één van beide doosjes genomen en dan hieruit een lucifer, net zolang tot één der doosjes leeg is. Hoe groot is de kans dat het andere doosje dan k lucifers bevat ($k \in \{1, 2, \ldots, n\}$)?

6. In een vaas bevinden zich n ballen, genummerd $1, 2, 3, \ldots, n$. We trekken willekeurig en zonder terugleggen k ballen uit de vaas. Het kleinste nummer dat zo getrokken wordt noemen we X. We willen $\mathbb{E}(X)$ berekenen.

   (a) Trek nog een $(k{+}1)^{\mathrm{e}}$ "hulpbal" uit de vaas. Het nummer dat daarop staat noemen we: Y. Laat zien dat voor $x = 1, 2, \ldots, n{-}k{+}1$ geldt:

$$\mathbb{P}(Y < X | X = x) = \frac{x-1}{n-k}.$$

   (b) Leid uit (a) af dat

$$\mathbb{P}[Y < X] = \frac{1}{n-k}(\mathbb{E}(X) - 1).$$

   (c) Laat zien dat ook moet gelden:

$$\mathbb{P}[Y < X] = \frac{1}{k+1}.$$

   (d) Bereken $\mathbb{E}(X)$ uit (b) en (c).

7. Bij ieder van de 18 vragen van een multiple choice-test moet gekozen worden uit drie mogelijke antwoorden, waarvan er één goed is. De score X die een leerling krijgt, is het aantal keren dat hij of zij goed gekozen heeft. We zeggen dat iemand de stof beheerst met *graad* p als hij/zij elke vraag, onafhankelijk van de andere vragen, met kans p goed beantwoordt.

(a) Benader $\mathbb{P}[X \geq 10]$ voor iemand met beheersingsgraad $\frac{1}{3}$.

(b) In een klas met 16 jongens en 9 meisjes halen de jongens een gemiddelde $\overline{X}$ en de meisjes een gemiddelde $\overline{Y}$. Bepaal verwachting en variantie van $V := \overline{X} - \overline{Y}$, als alle leerlingen de stof evengoed beheersen (zeg met graad p).
Veronderstel waar nodig dat de prestaties van de leerlingen onafhankelijk zijn, en geef aan waar deze veronderstelling nodig is.

(c) Bereken $\mathbb{P}[|V| \geq 1]$ (benaderd) in de situatie van (b) met $p = \frac{1}{3}$. (Zonder continuïteitscorrectie.)

8. Een mier (met dikte nul) beweegt zich kriskras over een tafelblad met lengte $2N$ en breedte $2M$ $(0 < M < N)$. De plaats waar de mier zich op zeker ogenblik bevindt, is homogeen verdeeld over het blad. Zij X de afstand tot de dichtstbijzijnde rand.

   (a) Bereken de verdelingsfunctie van X.

   (b) Zij $N = 2M$. Bereken $\mathbb{E}(X)$ en $\mathbb{V}\mathrm{ar}(X)$.

9. De stochasten X en Y zijn onafhankelijk; $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathbb{E}(X^3) = 1$ en $\mathbb{E}(Y^3) = 3$. Bereken $\mathbb{E}((X+Y)^3)$ en $\mathbb{E}((X+Y)^2(X-Y))$.

10. We werpen een dobbelsteen totdat er voor het eerst tweemaal achtereen hetzelfde aantal ogen boven komt. Zij X het aantal worpen dat daartoe nodig is.

   (a) Bereken $\mathbb{P}[X > 4]$.

   (b) Bepaal $\mathbb{E}(X)$.

11. Een deelnemer aan een televisiequiz (genaamd X), mag als beloning voor zijn goede prestaties een prijs uitkiezen. Er zijn drie prijzen beschikbaar: A, B en C. X weet niet wat dit voor prijzen zijn, maar we nemen aan dat hij, als hij er twee van ziet, een duidelijke voorkeur zal hebben voor één van de twee. Hij krijgt de prijzen één voor één aangeboden, en mag telkens alleen "ja" of "nee" zeggen, en nooit meer op een keuze terugkomen. Heeft hij eenmaal "ja" gezegd, dan krijgt hij geen verdere prijzen meer te zien.
Er dienen zich de volgende strategieën aan:

   (I) Kies A.
   (II) Als B aantrekkelijker is dan A, kies dan B. Kies anders C.

Bepaal voor beide strategieën de kans dat X de voor hem aantrekkelijkste prijs kiest uit A, B en C, en de kans dat hij de minst aantrekkelijke kiest.

12. Men wijst aselect (dat wil zeggen volgens een homogene verdeling) een punt in (0,1) aan. Nu is (0,1) in twee deelintervallen verdeeld. In het grootste van deze twee wijst men weer aselect een punt aan. Nu is (0,1) in drie intervallen verdeeld. Bepaal de verdelingsfunctie, dichtheid en verwachting van de lengte van het middelste interval.

13. In 8 worpen met een dobbelsteen krijgt men X keer een zes en Y keer een vijf. Gevraagd:

    (a) $\mathbb{P}[X = 3]$;

    (b) $\mathbb{P}[X + Y = 4]$;

    (c) $\mathbb{P}(X = 4|Y = 4)$;

    (d) Zijn X en Y onafhankelijk?

14. Laten X en Y onafhankelijke stochasten zijn; laat X uniform verdeeld zijn over $[0, 1]$ terwijl
$$\mathbb{P}[Y = 1] = 1 - \mathbb{P}[Y = 0] = \frac{1}{2}.$$
Bovendien definiëren we stochastische variabelen S, T en U als volgt:
$$\begin{aligned} S &= \sqrt{X}; \\ T &= 1 - \sqrt{1 - X}; \\ U &= YS + (1 - Y)T. \end{aligned}$$
Bereken de verwachting en de variantie van S en de verdelingsfunctie van U.

15. Men doet twee worpen met een dobbelsteen. X_1 is het aantal ogen bij de eerste worp, X_2 bij de tweede. $U := \min(X_1, X_2)$ is het kleinste van de twee en $V := \max(X_1, X_2)$ is het grootste.

    (a) Zijn U en V onafhankelijk?

    (b) Bepaal de kansverdeling van U.

    (c) Hoe groot is $\mathbb{E}(U)$ en $\mathbb{E}(U + V)$?

    (d) Bepaal $\mathbb{P}(X_1 = 3|U = 3)$.

16. De kansdichtheid van de stochast X, de levensduur van een bepaald elektronisch apparaat (uitgedrukt in uren), wordt gegeven door

$$f(x) = \begin{cases} \frac{c}{x^2} & \text{voor } x > 10 \\ 0 & \text{voor } x \leq 10 \end{cases}$$

   (a) Bepaal c zó, dat f inderdaad een kansdichtheid is.

   (b) Bereken $\mathbb{P}[X > 20]$.

   (c) Bereken de verdelingsfunctie van X.

   (d) Bereken de kans dat van 6 van dergelijke apparaten er 3 tenminste 20 uur functioneren.

17. Zij $X_1, X_2, X_3, \ldots$ een rij onafhankelijke stochasten die alle uniform over het interval $[0,1]$ verdeeld zijn.

   (a) Bereken $\mathbb{E}(X_1 \cdot X_2 \cdots X_n)$ voor $n = 1, 2, 3, \ldots$.

   (b) Bepaal de verdelingsfunctie en de kansdichtheidsfunctie van $-\log X_1$.

   (c) Bewijs dat voor $n = 1, 2, 3, \ldots$ de kansdichtheidsfunctie f_n van $X_1 \cdot X_2 \cdots X_n$ gegeven wordt door

$$f_n(x) = \frac{1}{(n-1)!}(-\log x)^{n-1}.$$

   Aanwijzing: Gebruik (b).

18. We besluiten, ongehinderd door enige kennis van voetbal of voetbalclubs, deel te nemen aan de voetbaltoto, wat neerkomt op het voorspellen van de uitslagen van 13 wedstrijden door middel van het aankruisen van een 1, 2 of 3 op het formulier bij ieder van de 13 wedstrijden (1 = thuisclub wint, 2 = thuisclub verliest, 3 = gelijkspel). Uit een groot aantal voorafgaande wedstrijden hebben we de frequenties bepaald waarmee de cijfers 1, 2 en 3 voorkomen: zeg p_1, p_2 en p_3, waarbij $p_1 + p_2 + p_3 = 1$. We nemen aan dat dit de kansen zijn waarmee elke uitslag zich, onafhankelijk van de andere, realiseert. We merken op dat p_1 groter is dan p_2 en p_3. We overwegen de volgende strategieën:

   Strategie 1: We vullen 13 keer 1 in.

   Strategie 2: Bij elke wedstrijd loten we onafhankelijk wat we invullen; met kans p_i vullen we i in ($i = 1, 2, 3$).

   Zij U het aantal goed voorspelde uitslagen.

   (a) Bepaal $\mathbb{E}(U)$ voor beide strategieën.

(b) Voor welke van de beide strategieën is $\mathbb{E}(U)$ het grootst?
Aanwijzing: $p_1 = p_1(p_1 + p_2 + p_3)$.

19. We werpen net zolang met een zuivere dobbelsteen totdat voor het eerst een 5 of een 6 bovenkomt. Dit gebeurt bij de Y^{e} worp. Vervolgens werpen we Y eerlijke munten op en krijgen daarbij X keer kop.

    (a) Bepaal de kansverdeling en de verwachting van Y.

    (b) Bewijs dat $\mathbb{E}(X) = \frac{1}{2}\mathbb{E}(Y)$.

20. De geheeltallige stochast S_n is de som van n onafhankelijke stochasten

$$X_1, X_2, \ldots, X_n,$$

die alle de volgende kansmassafunctie f hebben:

$$f(0) = \tfrac{1}{4} \quad f(1) = \tfrac{1}{2} \quad f(2) = \tfrac{1}{4} \quad f(x) = 0 \text{ als } x \notin \{0, 1, 2\}.$$

    (a) Bepaal de genererende functies van X_1 en van S_n.

    (b) Bereken hieruit de kansmassafunctie, zeg h_n, van S_n.

    (c) Wat wordt de functie h_3 als we f als volgt kiezen:

$$f(0) = f(1) = f(2) = \tfrac{1}{3} \quad f(x) = 0 \text{ elders?}$$

21. Een bepaald type lamp heeft een levensduur die exponentieel verdeeld is met een gemiddelde van drie weken. In een kamer wordt dit type lamp gebruikt op twee verschillende plaatsen, in fitting A en in fitting B. Op zeker tijdstip worden in beide fittingen nieuw lampen ingedraaid. De lampen branden daarna voortdurend, en als er één kapot gaat, wordt hij direct vervangen door een lamp van hetzelfde type. Zij N het aantal lampen dat in de eerste week kapot gaat.
De levensduren van de lampen mogen onderling onafhankelijk worden verondersteld.

    (a) Bereken de kans dat $N = 0$.

    (b) Bereken de kans dat $N = 1$.

    (c) Bereken de kans dat $N = 2$.

    (d) Kun je bedenken wat de kansverdeling van N is? (Een bewijs is niet nodig.)

22. Bij het Risk-spel spelen telkens twee spelers tegen elkaar: de 'aanvaller' en de 'verdediger'.

We beschouwen nu één enkel duel. Daarbij werpt de aanvaller met drie dobbelstenen, de verdediger met één of met twee.

We noemen de uitslagen van de aanvaller: X_1, X_2 en X_3, en nummeren deze van hoog naar laag; dus $6 \geq X_1 \geq X_2 \geq X_3 \geq 1$.

(a) Stel eerst dat de verdediger met één dobbelsteen werpt. (Uitslag: $Y \in \{1,2,3,4,5,6\}$). We zeggen dat hij *één slag wint* wanneer $Y \geq X_1$.
Bereken de kans hierop.

(b) Stel nu dat de verdediger met twee dobbelstenen werpt. We noemen de uitslagen: Y_1 en Y_2 ($6 \geq Y_1 \geq Y_2 \geq 1$).
We zeggen dat de verdediger *de eerste slag wint* als $Y_1 \geq X_1$, en *de tweede slag wint* als $Y_2 \geq X_2$.
Zij V_i het aantal door de verdediger gewonnen slagen minus het aantal door hem verloren slagen, bij verdediging met i dobbelstenen ($i = 1$ of 2).
De verdediger mag, nadat hij de uitslag van de aanvaller gezien heeft, kiezen of hij met één of met twee dobbelstenen verdedigt.
Stel dat de uitslag van de aanvaller is: $X_1 = 5$, $X_2 = 4$ en $X_3 = 2$. Met hoeveel dobbelstenen moet de verdediger dan spelen? Anders gezegd: voor welke i is $\mathbb{E}(V_i | X_1 = 5, X_2 = 4, X_3 = 2)$ maximaal?

(c) Overigens: heb je wel eens Risk gespeeld? (Het antwoord op deze vraag beïnvloedt je cijfer niet.)

23. Zij X het aantal zessen in één worp met een eerlijke dobbelsteen. (Dus $X = 0$ of $X = 1$.)

(a) Bepaal de kansgenererende functie G_X van X.

(b) Zij N Poisson-verdeeld met parameter λ.
Bepaal de kansgenererende functie G_N van N.

(c) Nu wordt de dobbelsteen N keer geworpen. Zij Z het aantal keren dat een zes wordt gegooid.
Toon aan dat $G_Z(s) = G_N(G_X(s))$, en bereken G_Z. Bereken ook de kansgenererende functie G_{N-Z} van $N - Z$.

(d) Toon nu aan dat de gezamenlijke kansgenererende functie $G(s,t) :=$

$\mathbb{E}(s^Z t^{N-Z})$ van Z en $N - Z$ gegeven wordt door

$$G(s,t) = G_N\left(tG_X\left(\frac{s}{t}\right)\right).$$

(e) Bewijs met behulp van deze gezamenlijke kansgenererende functie dat Z en $N - Z$ onafhankelijk van elkaar zijn.