

Wiskunde 1 voor kunstmatige intelligentie (BKI 116)

Bernd Souvignier

voorjaar 2005

Inhoud

I	Lineaire Algebra	4
Les 1	Stelsels lineaire vergelijkingen	5
1.1	Gauss-eliminatie	6
1.2	Oplosbaarheid	9
1.3	Homogeen/inhomogeen systeem	10
1.4	Toepassingen	12
Les 2	Lineaire afbeeldingen	15
2.1	Vectorruimten	15
2.2	Lineaire onafhankelijkheid, basis	16
2.3	Lineariteit	17
2.4	Matrix van een lineaire afbeelding	19
2.5	Kern en beeld	21
2.6	Stelsels lineaire vergelijkingen als lineaire afbeeldingen . .	24
Les 3	Matrix product	27
3.1	Vermenigvuldiging van matrices	27
3.2	Samenstelling van lineaire afbeeldingen	30
3.3	Inverse van een matrix	31
3.4	Basistransformatie	33
Les 4	Eigenwaarden en eigenvectoren	36
4.1	Eigenvectoren	36
4.2	Determinanten	37
4.3	Het vinden van eigenwaarden en eigenvectoren	39
4.4	Limieten van overgangsmatrices	41
4.5	Hoofdcomponenten analyse	42
Les 5	Inproducten	47
5.1	Het standaardinproduct	48
5.2	Inproducten	49
5.3	Orthogonale projecties	51
5.4	Beste approximatie	54
5.5	Zoekmachines	55
	Opgaven voor Lineaire Algebra	58

II Calculus	61
Les 6 Differentiatie van functies	62
6.1 Functies	62
6.2 Continue functies	65
6.3 De afgeleide van een functie	67
6.4 Regels voor differentiatie	70
Les 7 Speciale functies	74
7.1 Exponentiële functie en natuurlijke logaritme	74
7.2 Trigonometrische functies	77
7.3 Hyperbolische functies	83
Les 8 Minima en maxima van functies	88
8.1 Minima en maxima van gewone functies	88
8.2 Functies van meerdere variabelen en de partiële afgeleide	91
8.3 Minima en maxima van functies van meerdere variabelen	94
Les 9 Primitieve en integraal	98
9.1 De oppervlakte onder een grafiek	99
9.2 De primitieve en de integraal	100
9.3 Partieel integreren	103
Les 10 Substitutie	106
10.1 Rekenen met differentiaal	106
10.2 De substitutieregels	109
10.3 Toepassingen van de integraal	113
Opgaven voor Calculus	116
III Kansrekening	118
Les 11 Combinatoriek	119
11.1 Tellen van uitkomsten	119
11.2 Geordende grepen	120
11.3 Ongeordende grepen	121
Les 12 Kansverdelingen	128
12.1 Discrete kansverdelingen	128
12.2 Continue kansverdelingen	134
Les 13 Verwachtingswaarde en spreiding	141
13.1 Stochasten	141
13.2 Verwachtingswaarde	142
13.3 Spreiding	146
13.4 Covariantie en correlatie	150
Les 14 Voorwaardelijke kansen en de regel van Bayes	153
14.1 Voorwaardelijke kansen	154
14.2 Regel van Bayes	156
14.3 Onafhankelijkheid	159
14.4 Bernoulli-model	161
Les 15 Schatten en simuleren	164
15.1 Maximum likelihood schatting	164
15.2 Simulatie	168

15.3 Simulatie van speciale verdelingen	171
Opgaven voor Kansrekening	176

Deel I

Lineaire Algebra

Aanbevolen achtergrondliteratuur met veel opgaven (en oplossingen):

- Seymour Lipschutz, Marc L. Lipson: (Schaum's Outline of Theory and Problems of) Linear Algebra. McGraw-Hill Companies, 2000, 368 p., ISBN: 0071362002.
- Frank Ayres: (Schaum's Outline of Theory and Problems of) Matrices. McGraw-Hill Companies, 1974, 256 p., ISBN: 0070843791.

Les 1 Stelsels lineaire vergelijkingen

Om te beginnen is hier een puzzeltje:

- Vandaag is Annie twee jaar jonger dan Ben en Cees samen.
- Over vijf jaar is Annie twee keer zo oud als Ben.
- Twee jaar geleden was Ben half zo oud als Cees.

Hoe oud zijn Annie, Ben en Cees?

Dit kunnen we makkelijk met een stelsel lineaire vergelijkingen beschrijven, namelijk:

$$\begin{aligned} A &= B + C - 2 \\ A + 5 &= 2(B + 5) \\ B - 2 &= \frac{1}{2}(C - 2). \end{aligned}$$

Dit kunnen we op verschillende manieren oplossen, bijvoorbeeld door vervangen van één van de onbekenden met behulp van één van de vergelijkingen:

Uit de eerste vergelijking zien we meteen dat $A = B + C - 2$ en als we dit in de tweede vergelijking inzetten, wordt deze

$$B + C - 2 + 5 = 2(B + 5) \text{ of te wel } C + 3 = B + 10$$

Uit de derde vergelijking zien we dat $B = \frac{1}{2}C + 1$ en als we B nu in de nieuwe tweede vergelijking vervangen krijgen we

$$C + 3 = \frac{1}{2}C + 1 + 10 \text{ of } \frac{1}{2}C = 8$$

Dus vinden we als oplossing $C = 16$ en door achteruit door de stappen te gaan volgt dat $B = 9$ en $A = 23$.

Een iets dramatischer voorbeeld: Twee vliegtuigen vliegen in dezelfde hoogte op koersen die een snijpunt hebben. Vliegtuig A heeft een afstand van 45 mijl van het snijpunt en vliegt met een snelheid van 350 mijl per uur, vliegtuig B heeft een afstand van 75 mijl en een snelheid van 600 mijl per uur. Is er gevaar van een crash?

Een botsing zal gebeuren als de twee vliegtuigen het snijpunt op hetzelfde tijdstip bereiken, dus als er een x is met

$$350x = 45 \text{ en } 600x = 75.$$

Maar uit $350x = 45$ volgt $x = 9/70$ en $600 \cdot 9/70 = 540/7 \approx 77.14$ dus is er geen gevaar van een crash. Aan de andere kant kunnen we ook berekenen dat het wel mis kan gaan als vliegtuig B iets te langzaam is, namelijk als hij met snelheid

$75 \cdot 70/9 = 583.\bar{3}$ vliegt. In dit geval is inderdaad $x = 9/70$ een oplossing voor beide vergelijkingen.

Dit zijn natuurlijk eenvoudige voorbeelden van stelsels lineaire vergelijkingen, maar omdat we dit soort problemen heel vaak tegen komen zou het nuttig zijn hier een algemene oplossingsmethode voor te hebben.

1.1 Gauss-eliminatie

Lineaire vergelijkingen heten *lineair* omdat er veelvouden en sommen van onbekenden voorkomen, maar geen producten of machten van onbekenden (zo als A^3 of AB). Dus vinden we alle informatie van een vergelijking in de *coëfficiënten* van de onbekenden en de rechterzijde terug (als we alle onbekenden naar de linkerzijde brengen).

Daarom kunnen we een stelsel lineaire vergelijkingen in een schema schrijven dat we een *matrix* noemen. Dit is een rechthoekig schema van getallen waarbij de rijen met vergelijkingen corresponderen en de kolommen met de onbekenden. Een extra kolom wordt eraan toegevoegd om de rechterzijde van een vergelijking op te slaan.

In het puzzeltje over de leeftijd hebben we het stelsel vergelijkingen

$$\begin{aligned} A - B - C &= -2 \\ A - 2B &= 5 \\ B - \frac{1}{2}C &= 1. \end{aligned}$$

De hierbij horende matrix (het hierbij horende schema) is

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & -2 \\ 1 & -2 & 0 & 5 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array} \right).$$

Wat hebben we hier nu aan? We kunnen op zo'n matrix een aantal operaties toepassen, de zogeheten *elementaire operaties*, waaronder de oplossingen van het stelsel vergelijkingen ongedeerd blijven. Deze operaties zijn:

- het verwisselen van twee rijen;
- het vermenigvuldigen van een rij met een getal $\neq 0$;
- het optellen van een veelvoud van een rij bij een andere rij.

Het is duidelijk dat een oplossing van het oorspronkelijke stelsel ook een oplossing van het zo veranderde stelsel is. Het belangrijke punt is nu, dat we er op deze manier ook geen (kunstmatige) oplossingen bij krijgen, want we kunnen alle deze operaties weer ongedaan maken (omkeren) en daarmee zien we dat een oplossing van het veranderde stelsel ook een oplossing van het oorspronkelijke stelsel is.

De strategie is nu, de matrix van het stelsel door elementaire operaties in een vorm te brengen waarvan we de oplossingen makkelijk af kunnen lezen. Zo'n vorm is de *rijtrapvorm*.

I.1 Definitie Een matrix heeft *rijtrapvorm* als elke rij met meer nullen begint dan de voorafgaande rijen.

Als we niet-nul elementen met \bullet en willekeurige elementen met $*$ noteren, zien typische rijtrapvormen er als volgt uit

$$\begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \\ 0 & 0 & \bullet \end{pmatrix}, \quad \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \end{pmatrix}, \quad \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & * \end{pmatrix}, \quad \begin{pmatrix} \bullet & * & * & * & * \\ 0 & \bullet & * & * & * \\ 0 & 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix},$$

terwijl de volgende matrices niet in rijtrapvorm zijn

$$\begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \\ 0 & \bullet & * \end{pmatrix}, \quad \begin{pmatrix} 0 & \bullet & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & * \end{pmatrix} \text{ of } \begin{pmatrix} \bullet & * & * & * & * \\ 0 & \bullet & * & * & * \\ 0 & 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet & * \end{pmatrix}.$$

Om het op rijtrapvorm brengen (vaak ook *vegen* geheten) beter te kunnen bestuderen (en controleren) is het handig om afkortingen voor de elementaire operaties af te spreken. Het verwisselen van de i -de en de j -de rij noemen we $W_{i,j}$ het vermenigvuldigen van de i -de rij met het getal c noemen we $V_i(c)$ en het optellen van het c -voud van de i -de rij bij de j -de rij noemen we $O_{i,j}(c)$.

In het voorbeeld van het puzzeltje proberen we nu in de eerste kolom overal 0en te produceren (behalve in de eerste rij). De derde rij is al klaar en van de tweede rij moeten we de eerste aftrekken, dus passen we de operatie $O_{1,2}(-1)$ toe. Dit levert de volgende nieuwe matrix (met dezelfde oplossingen) op:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & -2 \\ 0 & -1 & 1 & 7 \\ 0 & 1 & -\frac{1}{2} & 1 \end{array} \right)$$

Nu moeten we alleen maar de tweede bij de derde rij optellen, dus $O_{2,3}(1)$ toepassen, om de rijtrapvorm te bereiken. Dit geeft de matrix:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & -2 \\ 0 & -1 & 1 & 7 \\ 0 & 0 & \frac{1}{2} & 8 \end{array} \right)$$

Dit stelsel kunnen we nu eenvoudig van onder naar boven oplossen.

We kunnen de oplossingen ook door verdere toepassingen van elementaire operaties vinden. Hiervoor berekenen we de *gereduceerde rijtrapvorm*. Deze heeft de eigenschap dat de eerste niet-nul in elke rij 1 is en dat verder alle elementen in de kolom van zo'n 1 allemaal 0 zijn.

In het voorbeeld passen we dus eerst de operaties $V_2(-1)$ en $V_3(2)$ toe, dit geeft:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & -2 \\ 0 & 1 & -1 & -7 \\ 0 & 0 & 1 & 16 \end{array} \right)$$

De tweede kolom is klaar als we $O_{2,1}(1)$ uitvoeren:

$$\left(\begin{array}{ccc|c} 1 & 0 & -2 & -9 \\ 0 & 1 & -1 & -7 \\ 0 & 0 & 1 & 16 \end{array} \right)$$

Tenslotte moeten we nog de derde kolom *vegen* door $O_{3,1}(2)$ en $O_{3,2}(1)$ toe te passen en we eindigen dus met de matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 23 \\ 0 & 1 & 0 & 9 \\ 0 & 0 & 1 & 16 \end{array} \right)$$

die de oplossing expliciet aan geeft.

Wat we in dit voorbeeld hebben gedaan kunnen we ook in een algemeen algoritme beschrijven die we de *Gauss-eliminatie* noemen:

I.2 Gauss-eliminatie (Vegen)

- Begin met de eerste kolom die niet alleen maar 0en bevat, zeg maar de j -de kolom;
- zorg (als nodig) door verwisseling van rijen ervoor dat in de eerste rij van deze kolom een getal c staat dat niet 0 is, dit getal noemen we een *pivot*;
- pas voor de rijen i van 2 t/m n (het aantal van rijen) de elementaire operatie $O_{1,i}(-d/c)$ toe, waarbij d het getal in de j -de kolom van rij i is, op die manier wordt de j -de kolom geveegd;
- herhaal deze procedure op de submatrix die wordt verkregen door de eerste rij te schrappen.

Als we de gereduceerde rijtrapvorm willen bereiken hebben we nog de volgende stappen nodig:

- Pas voor i van 1 t/m n de operatie $V_i(c_i^{-1})$ toe, waarbij c_i het eerste element $\neq 0$ in de i -de rij is;
- veeg de kolom j waarin de laatste rij zijn eerste 1 heeft, door $O_{n,i}(-d_i)$ voor i van 1 t/m $n - 1$ toe te passen, waarbij d_i het element in de j -de kolom van rij i is;
- herhaal deze procedure op de submatrix die wordt verkregen door de laatste rij te schrappen.

1.2 Oplosbaarheid

In het voorbeeld van het puzzel met de leeftijden hebben we een eenduidige oplossing gevonden. In het probleem van de vliegtuigen hadden we de matrix

$$\left(\begin{array}{c|c} 350 & 45 \\ \hline 600 & 75 \end{array} \right)$$

die we door toepassen van $O_{1,2}(-\frac{12}{7})$ naar

$$\left(\begin{array}{c|c} 350 & 45 \\ \hline 0 & 75 - \frac{12}{7}45 \end{array} \right)$$

brengen. Maar $75 - \frac{12}{7}45 = \frac{-15}{7}$, dus staat de laatste rij voor de vergelijking $0 \cdot x = \frac{-15}{7}$ en deze heeft natuurlijk geen oplossing. We kunnen nagaan dat dit al het algemeen geval van een stelsel zonder oplossing is, er geldt namelijk de volgende stelling:

I.3 Stelling *Een stelsel lineaire vergelijkingen is dan en slechts dan niet oplosbaar, als er in de rijtrapvorm een rij met louter nullen als coëfficiënten voor de onbekenden maar een getal $\neq 0$ voor de rechterzijde staat.*

Als we geen zo'n rij vinden, weten we dus dat het systeem oplosbaar is. Maar is de oplossing eenduidig?

Het is duidelijk dat de oplossing niet eenduidig is als we een kolom met alleen maar nullen vinden, want dan mag de bijhorende onbekende elke willekeurige waarde hebben. Maar hetzelfde geldt ook als we trappen van meer dan een kolom breedte vinden. Voor de kolommen die geen *pivot* (het eerste niet-nul element van een rij) bevatten, mogen we ook willekeurige waarden kiezen, want bij het van onder naar boven oplossen zijn alleen maar de pivots van belang.

Dit zal het volgende voorbeeld illustreren. Stel we vinden voor een stelsel vergelijkingen in x, y, z de rijtrapvorm

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ \hline 0 & 1 & -3 & 4 \end{array} \right)$$

dan kunnen we het stelsel voor een willekeurig gekozen waarde van z oplossen door $y = 4 + 3z$ en $x = 9 - 2z - y = 5 - 5z$. We noemen z dan ook een *vrije parameter*.

I.4 Stelling *Een oplosbaar stelsel lineaire vergelijkingen heeft zoveel vrije parameters als het verschil tussen het aantal onbekenden en het aantal pivots in de rijtrapvorm aangeeft.*

In het bijzonder is een stelsel lineaire vergelijkingen eenduidig oplosbaar dan en slechts dan als er in de rijtrapvorm even veel pivots als onbekenden zijn.

Als voorbeeld kijken we eens naar een iets groter stelsel vergelijkingen dat we al op rijtrapvorm hebben gebracht. We nemen aan dat de rechterzijden alle 0 zijn:

$$\left(\begin{array}{ccccc|c} \boxed{1} & 1 & 1 & 1 & 1 & 0 \\ \hline 0 & 0 & \boxed{1} & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 \end{array} \right).$$

Dit stelsel heeft 5 onbekenden, 3 pivots (door de hokjes gemarkeerd) en dus 2 vrije parameters, die we in de tweede en vierde kolom kunnen kiezen. Als we de onbekenden x, y, z, v, w noemen, krijgen we bij het van beneden naar boven oplossen:

$$\begin{aligned} w &= 0 \\ z + 2v &= 0 \Rightarrow z = -2v \\ x + y - 2v + v &= 0 \Rightarrow x = -y + v \end{aligned}$$

De algemene oplossing is dus van de vorm $(-y + v, y, -2v, v, 0)$. Merk op dat we hier y en v als vrije parameters hebben *gekozen*. We hadden voor de vrije parameters ook een andere keuze kunnen maken, maar voor de manier hoe we uit de rijtrapvorm de oplossingen door achteruit oplossen bepalen is onze keuze de meest voor de hand liggende.

1.3 Homogeen/inhomogeen systeem

Bij het van onder naar boven oplossen van een lineair stelsel vergelijkingen met vrije parameters is het soms lastig om met de parameters te rekenen. Echter kunnen we dit ook op een andere manier doen. Hiervoor is de volgende eenvoudige opmerking belangrijk:

Merk op: *Als we twee oplossingen van een stelsel lineaire vergelijkingen van elkaar aftrekken is het resultaat een oplossing van hetzelfde stelsel, waarbij we de rechterzijden alle door 0 vervangen.*

Dit is in een voorbeeld makkelijk in te zien: Stel dat (X, Y, Z) en (X', Y', Z') oplossingen zijn van de vergelijking $ax + by + cz = d$, dan is $aX + bY + cZ = d = aX' + bY' + cZ'$ en dus is $a(X - X') + b(Y - Y') + c(Z - Z') = 0$.

Een stelsel vergelijkingen waarbij alle rechterzijden 0 zijn, noemen we een *homogeen stelsel*, een algemeen stelsel heet *inhomogeen*. Het stelsel vergelijkingen dat we uit een inhomogeen stelsel verkrijgen door de rechterzijden door 0 te vervangen heet het *bijhorende homogene stelsel*. We kunnen dus het volgende zeggen:

I.5 Stelling *De oplossingen van een stelsel lineaire vergelijkingen zijn van de vorm $P + H$, waarbij P een vaste particuliere oplossing van het inhomogene stelsel is en H een oplossing van het bijhorende homogene stelsel.*

Als namelijk P en P' oplossingen van het stelsel vergelijkingen zijn, dan is $P' = P + H$ met $H = P' - P$ en H is een oplossing van het bijhorende stelsel lineaire vergelijkingen.

Het aardige aan een homogeen stelsel lineaire vergelijkingen is, dat we oplossingen van een homogeen stelsel bij elkaar kunnen optellen en zo weer een oplossing van hetzelfde stelsel krijgen. Als namelijk $aX + bY + cZ = 0$ en $aX' + bY' + cZ' = 0$, dan is ook $a(X + X') + b(Y + Y') + c(Z + Z') = 0$. Net zo goed kunnen we een oplossing ook met een factor vermenigvuldigen, dan blijft het ook een oplossing.

We vinden *alle* oplossingen van een homogeen stelsel vergelijkingen als lineaire combinaties van een aantal *basis-oplossingen*, namelijk van zo veel basis-oplossingen als er vrije parameters zijn. Hierbij noemen we een oplossing een basis-oplossing als één van de vrije parameters de waarde 1 en alle andere vrije parameters de waarde 0 hebben. Omdat we voor elke keuze van de vrije parameters een oplossing van het homogene stelsel krijgen, geldt dit natuurlijk ook voor onze speciale keuze van één parameter 1 en de anderen 0.

Als we bijvoorbeeld naar de vergelijking

$$x + y + z = 0$$

kijken, zijn y en z vrije parameters en zijn $(x, y, z) = (-1, 1, 0)$ en $(x, y, z) = (-1, 0, 1)$ de basis-oplossingen.

We kunnen nu de basis-oplossingen met een willekeurige factor vermenigvuldigen en bij elkaar optellen. In het voorbeeld krijgen we zo bijvoorbeeld de oplossing $(x, y, z) = (-7, 3, 4)$ door de eerste basis-oplossing met 3 en de tweede met 4 de vermenigvuldigen en deze twee bij elkaar op te tellen.

Het is duidelijk dat lineaire combinaties van de basis-oplossingen verschillend zijn als de factoren van de basis-oplossingen verschillend zijn. Andersom kunnen we een willekeurige oplossing van het homogene stelsel steeds schrijven als een combinatie van de basis-oplossingen door de waarden van de vrije parameters als factoren te kiezen.

Tenslotte kunnen we ook voor het vinden van een particuliere oplossing van het inhomogene stelsel nog een trucje toepassen om het makkelijker te maken. Bij een oplossing P van het inhomogene stelsel hebben de vrije parameters zekere waarden en als we de oplossing H van het homogene stelsel waarvoor de vrije parameters dezelfde waarden hebben van P aftrekken, krijgen we een nieuwe oplossing $P' = P - H$ van het inhomogene stelsel. Maar P' is zo gemaakt dat alle vrije parameters de waarde 0 hebben. We vinden dus een particuliere oplossing van het inhomogene stelsel door alle vrije parameters 0 te zetten en het stelsel voor de resterende onbekenden (die bij de pivots horen) op te lossen.

Bij elkaar genomen hebben we dus de volgende stelling ingezien:

I.6 Stelling *Laat P een particuliere oplossing van een stelsel lineaire vergelijkingen zijn en noem de basis-oplossingen van het bijhorende homogene stelsel h_1, \dots, h_s , dan zijn alle oplossingen van het stelsel eenduidig te schrijven als $P + c_1 h_1 + c_2 h_2 + \dots + c_s h_s$, waarbij c_i willekeurige getallen zijn.*

Let op: Als je de oplossingen van een inhomogeen stelsel bepaalt moet je *eerst* het originele stelsel *met* rechte zijden op rijtrapvorm brengen en pas vervolgens de rechte zijden door 0en vervangen. De rijtrapvorm heeft namelijk in het algemeen andere rechterzijden dan het oorspronkelijke stelsel.

In het voorbeeld van boven met rijtrapvorm

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -3 & 4 \end{array} \right)$$

hebben we één vrije parameter, namelijk de derde onbekende. Als we $z = 0$ kiezen vinden we $y = 4$ en $x = 5$ als particuliere oplossing van het inhomogene stelsel. Voor het homogene stelsel kiezen we nu $z = 1$, dit geeft $y = 3$ en $x = -5$ en dus vinden we de algemene oplossing als $(x, y, z) = (5, 4, 0) + t \cdot (-5, 3, 1)$ en dit is precies wat we ook eerder als oplossing hadden gevonden (met z in plaats van t).

In het grotere voorbeeld met 5 onbekenden veranderen we even de rechterzijden en bekijken het stelsel in de onbekenden x, y, z, v, w met rijtrapvorm

$$\left(\begin{array}{ccccc|c} \boxed{1} & 1 & 1 & 1 & 1 & 3 \\ 0 & 0 & \boxed{1} & 2 & 3 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{array} \right).$$

De vrije parameters horen bij de tweede en vierde kolom, dus bij de onbekenden y en v . Als we $y = 0$ en $v = 0$ kiezen, vinden we de particuliere oplossing $(10, 0, -11, 0, 4)$.

Voor de basis-oplossingen van het bijhorende homogene stelsel kiezen we eerst $y = 0$ en $v = 1$, dan krijgen we als eerste basis-oplossing $(1, 0, -2, 1, 0)$. Vervolgens kiezen we $y = 1$ en $v = 0$, dit geeft de tweede basis-oplossing $(-1, 1, 0, 0, 0)$. We vinden nu alle oplossingen van het homogene stelsel door de twee basis-oplossingen met willekeurige factoren te vermenigvuldigen en bij elkaar op te tellen (dit geeft precies de oplossingen die we boven als hebben bepaald). Om de oplossingen van het inhomogene stelsel te krijgen, moeten we hierbij nog de particuliere oplossing optellen, en we krijgen zo als oplossingen:

$$\begin{aligned} (x, y, z, v, w) &= (10, 0, -11, 0, 4) + r \cdot (1, 0, -2, 1, 0) + s \cdot (-1, 1, 0, 0, 0) \\ &= (10 + r - s, s, -11 - 2r, r, 4). \end{aligned}$$

1.4 Toepassingen

Een toepassing van de stelsels lineaire vergelijkingen is het vinden van een grafiek door voorgeschreven punten, bijvoorbeeld door de punten van een meting. Hierbij maken we gebruik van de volgende stelling die we hier niet gaan bewijzen.

I.7 Stelling *Laten x_1, \dots, x_n verschillende (reële) getallen zijn en y_1, \dots, y_n willekeurig gekozen waarden. Dan is er precies één veelterm $a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$ van graad $\leq n - 1$ zodat de grafiek van $y = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$ door de punten $(x_1, y_1), \dots, (x_n, y_n)$ gaat.*

De toepassing van deze stelling ziet er als volgt uit: Neem aan we hebben bij een meting voor de x -waarden 1, 2 en 3 de resultaten 6, 3 en 2 verkregen. We zijn dus op zoek naar een grafiek door de punten $(1, 6)$, $(2, 3)$ en $(3, 2)$. Volgens de stelling kunnen we hiervoor een kwadratische veelterm vinden. We schrijven zo'n veelterm in algemene vorm neer, dan zijn de coëfficiënten de onbekenden voor een stelsel lineaire vergelijkingen. In ons geval hebben we dus de veelterm

$a_0 + a_1x + a_2x^2$ en door invullen van de gewenste punten vinden we het stelsel vergelijkingen:

$$\begin{aligned} a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 &= 6 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 &= 3 \\ a_0 + a_1 \cdot 3 + a_2 \cdot 3^2 &= 2. \end{aligned}$$

De matrix van dit stelsel is

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 1 & 2 & 4 & 3 \\ 1 & 3 & 9 & 2 \end{array} \right)$$

en de rijtrapvorm hiervan wordt

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 3 & -3 \\ 0 & 0 & 2 & 2 \end{array} \right).$$

Door van onder naar boven op te lossen vinden we $a_2 = 1$, $a_1 = -6$ en $a_0 = 11$, dus is de gezochte grafiek gegeven door $y = x^2 - 6x + 11$.

Een voorbeeld uit de scheikunde laat zien dat soms ook de oplossingen van een homogeen stelsel interessant kunnen zijn. Bij een chemische reactie zijn de uitgangs- en de eindproducten bekend, maar om de hoeveelheden van de stoffen te bepalen moet een chemische vergelijking in evenwicht gebracht worden. In de reactie $Ca + H_3PO_4 \rightarrow Ca_3P_2O_8 + H_2$ hebben we dus coëfficiënten a_1, a_2 en b_1, b_2 nodig, zodat in de vergelijking $a_1Ca + a_2H_3PO_4 \rightarrow b_1Ca_3P_2O_8 + b_2H_2$ op beide zijden hetzelfde aantal van alle soorten van atomen staat.

Dit levert de vergelijkingen $a_1 = 3b_1$, $3a_2 = 2b_2$, $a_2 = 2b_1$ en $4a_2 = 8b_1$ op. Het bijhorende stelsel lineaire vergelijkingen is dus

$$\left(\begin{array}{cccc|c} 1 & 0 & -3 & 0 & 0 \\ 0 & 3 & 0 & -2 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 4 & -8 & 0 & 0 \end{array} \right)$$

Door toepassen van $O_{3,4}(-4)$, $O_{3,2}(-3)$ en $W_{2,3}$ vinden we de rijtrapvorm

$$\left(\begin{array}{cccc|c} 1 & 0 & -3 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 6 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

We zien dat we b_2 als vrije parameter kunnen kiezen, als we die t noemen wordt de oplossing dus $b_2 = t$, $b_1 = \frac{1}{3}t$, $a_2 = \frac{2}{3}t$ en $a_1 = t$. Omdat het om een chemische reactie gaat, willen we graag de kleinste oplossing in de natuurlijke getallen vinden. Hiervoor laten we t door $1, 2, \dots$ lopen en zien dat we voor $t = 3$ inderdaad zo'n oplossing vinden.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- stelsel lineaire vergelijkingen
- matrix van een stelsel vergelijkingen
- elementaire operaties
- (gereduceerde) rijtrapvorm, pivot
- Gauss-eliminatie
- vrije parameter
- homogeen/inhomogeen stelsel
- basis-oplossing, particuliere oplossing

OPGAVEN

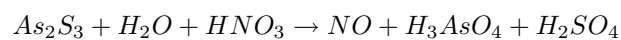
1. Vind alle oplossingen van het lineaire stelsel vergelijkingen

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + y + z &= 8. \end{aligned}$$

2. Ga na dat een lineair stelsel vergelijkingen met meer onbekenden dan vergelijkingen nooit een eenduidige oplossing heeft. Bestaat er in dit geval altijd een oplossing?
3. Zij $c_1, c_2, c_3, b \in \mathbb{R}$ en $c_1 \neq 0$. Vind alle oplossingen van $c_1x_1 + c_2x_2 + c_3x_3 = b$. Hoeveel vrije parameters heeft dit 'stelsel'?
4. Vind een veelterm $a_3x^3 + a_2x^2 + a_1x + a_0$ van graad 3, waarvan de grafiek door de punten $(-1, 10)$, $(0, 4)$, $(1, 2)$ en $(2, -2)$ gaat.
5. In een experiment zijn de resultaten afhankelijk van een parameter t . Gemeten worden de waarden $y = 0$ voor $x = 0$, $y = t$ voor $x = 1$ en $y = 3t$ voor $x = 2$. Geef een veelterm $a_2x^2 + a_1x + a_0$ van graad 2 aan die door deze drie punten gaat (de coëfficiënten zijn natuurlijk van t afhankelijk).
6. Een stelsel vergelijkingen heeft de rijtrapvorm

$$\left(\begin{array}{cccccc|c} 0 & 1 & 2 & 0 & 1 & 1 & 8 \\ 0 & 0 & 2 & 1 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array} \right)$$

- (i) Bepaal alle oplossingen van het bijhorende homogene stelsel.
- (ii) Vind een oplossing van het inhomogene stelsel.
- (iii) Beschrijf alle oplossingen van het inhomogene stelsel.
7. Breng de chemische vergelijking



in evenwicht.

Les 2 Lineaire afbeeldingen

Als een robot bij de *robocup* (het voetbaltoernooi voor robots) een doelpunt wil maken moet hij eerst in de goede positie komen, d.w.z. geschikt achter de bal staan, voordat hij trapt. Hiervoor moet hij naar de positie van de bal lopen en hij moet zo om zijn as draaien, dat hij in de richting van het doel mikt. Als we dit in wiskundige woorden beschrijven, is eerst een translatie van zijn momentele positie naar de positie van de bal nodig, en vervolgens een rotatie om zijn as.

In de berekeningen die voor zo'n probleem nodig zijn, komen we de volgende vraag tegen: Hoe kunnen we de nieuwe coördinaten van een punt (x, y) berekenen, als we om een as in het punt (x_0, y_0) om een hoek van φ (bijvoorbeeld van 90 graden) draaien?

Voordat we met dit soort vragen aan de slag kunnen, hebben we een paar basisbegrippen nodig.

2.1 Vectorruimten

Een \mathbb{R} -vectorruimte is een verzameling V van vectoren die afgesloten is onder optelling en onder schaling met factoren uit de reële getallen \mathbb{R} . Deze factoren worden vaak *scalaires* genoemd. Soms wordt \mathbb{R} ook door de complexe getallen \mathbb{C} vervangen, dan spreekt men van een \mathbb{C} -vectorruimte. In deze cursus zullen we ons tot \mathbb{R} -vectorruimten beperken die we van nu af gewoon *vectorruimten* noemen. In het kort is een verzameling V een vectorruimte als geldt:

- (i) voor alle $v \in V$ en $\lambda \in \mathbb{R}$ is $\lambda v \in V$, d.w.z. met een vector v bevat V ook alle *schalingen* van v ,
- (ii) voor alle $v, w \in V$ is $v + w \in V$, d.w.z. met twee vectoren bevat V ook hun som.

Men kan dit ook zo uitdrukken, dat een vectorruimte een verzameling van vectoren is, die afgesloten is onder het vormen van *lineaire combinaties*, d.w.z. eindige sommen van scalaire veelvouden van de vectoren.

Men ziet onmiddellijk in dat een (niet lege) vectorruimte steeds de *nulvector* $0 = 0 \cdot v$ bevat en voor elke vector $v \in V$ een tegengestelde vector $-v = (-1) \cdot v$ met $v + (-v) = 0$.

De belangrijkste vectorruimten zijn de 2-dimensionale vlakke

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x, y \in \mathbb{R} \right\}$$

en de 3-dimensionale ruimte

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}.$$

Vaak is het echter ook handig met hoger-dimensionale vectorruimten te werken, bijvoorbeeld als bij een meting n onafhankelijke parameters een rol spelen. De

n -dimensionale vectorruimte \mathbb{R}^n bestaat dan gewoon uit de kolommen met n componenten.

In veel situaties is het interessant naar een deelverzameling van een vectorruimte te kijken die zelfs ook weer een vectorruimte is. Dit noemen we een *deelvectorruimte* of *lineaire deelruimte*, maar meestal gewoon kort een *deelruimte*. Een voorbeeld is een lijn door de oorsprong in \mathbb{R}^2 , bijvoorbeeld de lijn $L = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x = 2y \right\}$. Alle elementen van L zijn van de vorm $\lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ en men gaat eenvoudig na dat dit inderdaad een vectorruimte is.

2.2 Lineaire onafhankelijkheid, basis

Een stelsel (v_1, \dots, v_n) van vectoren in een vectorruimte V heet *lineair onafhankelijk* of kort *onafhankelijk* als we geen van de v_i als lineaire combinatie van de andere vectoren v_j kunnen schrijven. Een equivalente beschrijving is dat (v_1, \dots, v_n) onafhankelijk zijn als de enige mogelijkheid voor een vergelijking $\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = 0$ gegeven is door alle $\lambda_i = 0$ te kiezen, dus hebben we de definitie:

I.8 Definitie Een stelsel (v_1, \dots, v_n) van vectoren V heet *onafhankelijk* als

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = 0 \Rightarrow \lambda_1 = \lambda_2 = \dots = \lambda_n = 0.$$

Hieruit laat zich snel een handig criterium voor lineaire onafhankelijkheid afleiden: Een stelsel (v_1, \dots, v_n) van vectoren is dan en slechts dan onafhankelijk is als

- (i) $v_1 \neq 0$,
- (ii) voor $2 \leq i \leq n$ is de vector v_i geen lineaire combinatie van v_1, \dots, v_{i-1} , d.w.z. $v_i \neq \lambda_1 v_1 + \dots + \lambda_{i-1} v_{i-1}$ voor alle keuzes van $\lambda_1, \dots, \lambda_{i-1}$.

Bijvoorbeeld is $\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$ onafhankelijk, want $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is geen veelvoud van $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (kijk naar de tweede component).

Aan de andere kant is $\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$ wel afhankelijk, omdat $\begin{pmatrix} 1 \\ 2 \end{pmatrix} = -1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ geldt.

I.9 Definitie Zij V een vectorruimte:

- (i) Een onafhankelijk stelsel $B = (v_1, \dots, v_n)$ heet een *basis* van V als zich elke vector $v \in V$ als lineaire combinatie $v = \lambda_1 v_1 + \dots + \lambda_n v_n$ van de vectoren in B laat schrijven.
- (ii) Het aantal van vectoren in een basis voor een vectorruimte V is voor alle bases hetzelfde en heet de *dimensie* van V .

Merk op dat achter het feit dat elke basis van een vectorruimte even veel vectoren bevat een serieuze stelling zit.

We noemen de λ_i de *coördinaten* van v met betrekking tot B . Uit de onafhankelijkheid van B volgt dat de coördinaten eenduidig zijn.

Voorbeeld: Het stelsel $B = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$ is een basis van \mathbb{R}^2 , want elke vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ laat zich schrijven als $\begin{pmatrix} x \\ y \end{pmatrix} = (x - y) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ en \mathbb{R}^2 is dus 2-dimensionaal (wat geen verrassing is).

We zullen later in deze les zien hoe we met behulp van een homogeen stelsel lineaire vergelijkingen makkelijk na kunnen gaan of een stelsel vectoren onafhankelijk en een basis is.

Door voor de vectorruimte \mathbb{R}^n de *standaardbasis*

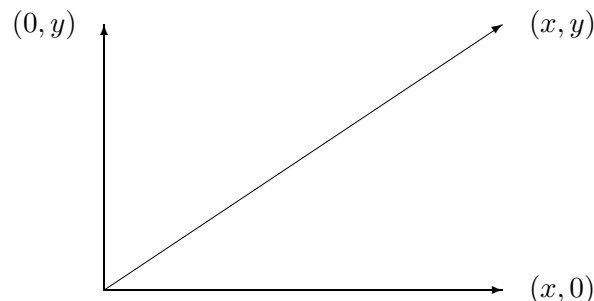
$$B := \left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right)$$

te kiezen (waarbij de i -de vector een 1 in de i -de component heeft, en 0 elders) zien we dat \mathbb{R}^n dimensie n heeft. De standaardbasis heeft de mooie eigenschap dat de coördinaten van een vector precies de componenten van de kolomvector zijn.

Let op: Niet elke vectorruimte die uit vectoren met n componenten bestaat (dus een deelverzameling van \mathbb{R}^n is) heeft dimensie n , hij kan ook een lagere dimensie hebben. Het boven genoemde voorbeeld van de lijn L in \mathbb{R}^2 is een 1-dimensionale vectorruimte.

2.3 Lineariteit

We gaan een eenvoudig voorbeeld van een draaiing in het vlak bekijken. Stel, de as waar we om draaien is in de oorsprong, het punt $(0,0)$. Als we om 90 graden (linksom) draaien, komt het punt $(1,0)$ op $(0,1)$ terecht en het punt $(0,1)$ op $(-1,0)$. Het is duidelijk, dat een schaling (rekken of krimpen) geen probleem is, daarom gaat het punt $(x,0)$ naar $(0,x)$ en het punt $(0,y)$ naar $(-y,0)$. Een algemeen punt (x,y) krijgen we als som van $(x,0)$ en $(0,y)$, zo als in het volgende plaatje te zien is:

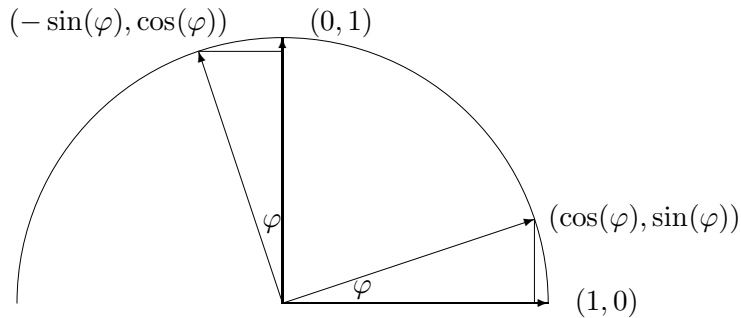


Omdat bij een rotatie de hele rechthoek gedraaid wordt, is het beeld van de som $(x, 0) + (0, y)$ hetzelfde als de som van de beelden $(0, x)$ en $(-y, 0)$ en dus gelijk aan $(-y, x)$.

We kunnen op deze manier gemakkelijk berekenen dat het punt $(47, 11)$ bij een rotatie om 90 graden op $(-11, 47)$ terecht komt.

Als we nu in plaats van 90 graden om 60 graden draaien, kunnen we met dezelfde methode weer het beeld van een algemeen punt bepalen, als we weten waar de punten $(1, 0)$ en $(0, 1)$ naar toe gaan. Met een beetje meetkunde uit de middelbare school zien we in dat $(1, 0)$ naar $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ en $(0, 1)$ naar $(-\frac{\sqrt{3}}{2}, \frac{1}{2})$ gaat. Een punt (x, y) gaat dus naar het beeldpunt $(\frac{1}{2}x - \frac{\sqrt{3}}{2}y, \frac{\sqrt{3}}{2}x + \frac{1}{2}y)$.

In het algemeen kunnen we een draaiing om een hoek φ in het vlak met behulp van de cosinus en de sinus beschrijven, namelijk het punt $(1, 0)$ gaat naar $(\cos(\varphi), \sin(\varphi))$, het punt $(0, 1)$ naar $(-\sin(\varphi), \cos(\varphi))$ en dus gaat het algemene punt (x, y) naar $(x \cos(\varphi) - y \sin(\varphi), x \sin(\varphi) + y \cos(\varphi))$.



We hebben in dit voorbeeld gebruik gemaakt van een belangrijk principe, namelijk dat een draaiing een *lineaire afbeelding* is. Het enige wat we inderdaad gebruikt hebben is, dat we een schaling of een optelling voor en na het afbeelden kunnen uitvoeren zonder het resultaat te veranderen. Dit geeft de volgende definitie:

I.10 Definitie Een afbeelding f op een vectorruimte V met de eigenschappen

- (i) $f(\lambda v) = \lambda f(v)$ voor alle vectoren $v \in V$ en alle $\lambda \in \mathbb{R}$
- (ii) $f(v + w) = f(v) + f(w)$ voor alle vectoren $v, w \in V$

heet een *lineaire afbeelding*.

We hebben boven al gezien dat een rotatie een lineaire afbeelding is, we hoeven namelijk alleen maar de punten (x, y) door vectoren $\begin{pmatrix} x \\ y \end{pmatrix}$ te vervangen.

Maar: Niet elke eenvoudige afbeelding is lineair. Bijvoorbeeld is een translatie geen lineaire afbeelding, want bij een translatie f over de vector t geldt voor twee vectoren v en w dat $f(v + w) = v + w + t$, maar $f(v) + f(w) = v + t + w + t = v + w + 2t$ en dit is alleen maar hetzelfde als $t = 0$ is.

Een noodzakelijke voorwaarde dat een afbeelding f lineair is, bestaat erin dat $f(0) = 0$ moet gelden, want $f(0) = f(v-v) = f(v) + f(-v) = f(v) - f(v) = 0$. Hiermee kunnen we al zien dat translaties niet lineair zijn.

Omdat translaties zo belangrijk en eenvoudig zijn is er wel een eigen begrip voor de combinaties van lineaire afbeeldingen en translaties: *affiene afbeeldingen*. Een affiene afbeelding φ op een vectorruimte V is een afbeelding van de vorm $\varphi(v) = f(v) + t$ waarbij f een lineaire afbeelding op V is en $t \in V$ de translatie vector.

We kunnen omgekeerd heel eenvoudig checken of een gegeven afbeelding φ affien is. Omdat elke lineaire afbeelding de nulvector vast laat moet $\varphi(0)$ de translatie vector zijn. Dus is φ een affiene afbeelding dan en slechts dan als de afbeelding $f(v) := \varphi(v) - \varphi(0)$ een lineaire afbeelding is.

2.4 Matrix van een lineaire afbeelding

We kunnen nu de manier hoe we in het voorbeeld de beeldvectoren van een rotatie bepaald hebben nader toelichten. Het idee is, voor een basis van de vectorruimte de beelden te berekenen en vervolgens de lineariteit te gebruiken. Immers, als $B = (v_1, \dots, v_n)$ een basis van V is, dan vinden we voor $v = \lambda_1 v_1 + \dots + \lambda_n v_n$ het beeld $f(v)$ als $f(v) = \lambda_1 f(v_1) + \dots + \lambda_n f(v_n)$.

Om een lineaire afbeelding op te schrijven, gebruiken we weer een matrix. Als de afbeelding f van een n -dimensionale vectorruimte V naar een m -dimensionale vectorruimte W gaat, dan hebben we een $m \times n$ -matrix A nodig. We kiezen een basis (v_1, \dots, v_n) voor V en een basis (w_1, \dots, w_m) voor W .

In de eerste kolom van de matrix A schrijven we nu het beeld $f(v_1)$ van de eerste basisvector, in de tweede het beeld $f(v_2)$ van de tweede basisvector, enzovoorts. Hierbij zetten we voor de beelden alleen maar de coördinaten met betrekking tot de basis (w_1, \dots, w_m) neer, dus:

$$f(v_i) = a_{1i}w_1 + a_{2i}w_2 + \dots + a_{mi}w_m \Rightarrow i\text{-de kolom van } A \text{ is } \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix}.$$

Merk op: Als V en W gewoon de vectorruimten \mathbb{R}^n en \mathbb{R}^m zijn, en we hier-

voor de standaardbases $\left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right)$ van n - resp. m -dimensionale

vectoren kiezen, bevat de matrix van f precies de kolomvectoren $f(v_i)$.

Maar in sommige situaties is het handig een andere basis dan de standaardbasis te kiezen, en dan hebben we de coördinaten met betrekking tot zo'n basis nodig.

In het voorbeeld van boven schrijven we een rotatie om de hoek φ (met betrekking tot de standaardbases) als $\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$.

Het beeld van een vector bepalen we nu, door de matrix van de afbeelding met de (coördinaten van de) vector te vermenigvuldigen. Deze vermenigvuldiging moet voor een vector $v = \lambda_1 v_1 + \dots + \lambda_n v_n$ de lineaire combinatie $f(v) = \lambda_1 f(v_1) + \dots + \lambda_n f(v_n)$ weerspiegelen. Dit bereiken we met de definitie:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \ddots & \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} := \begin{pmatrix} a_{11}\lambda_1 + \dots + a_{1n}\lambda_n \\ \vdots \\ a_{m1}\lambda_1 + \dots + a_{mn}\lambda_n \end{pmatrix}$$

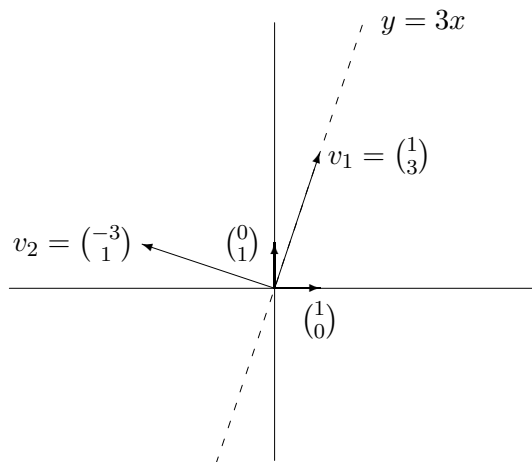
Dus, de i -de coördinaat van de beeldvector vinden we door tegelijkertijd over de i -de rij van de matrix van de afbeelding en over de λ -vector (de vector met de coördinaten) te lopen en de producten van de paren waar we overheen lopen op te tellen.

Voorbeelden voor lineaire afbeeldingen in de vlakte zijn naast rotaties ook spiegelingen. De spiegeling in de diagonaal $x = y$ in de vlakte beschrijven we met betrekking tot de standaardbasis door de matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ en het beeld van een algemene vector is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$.

Als we in plaats van de standaardbasis de basis $\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$ kiezen, dan wordt de matrix van de spiegeling $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Iets ingewikkelder wordt het voorbeeld, als we de spiegeling in de lijn $y = 3x$ bekijken. Hierbij is het handig om een basis zo te kiezen dat één vector op de spiegellingsas ligt, en de andere er loodrecht op staat. Zo'n basis is bijvoorbeeld

$$B := \left(v_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, v_2 = \begin{pmatrix} -3 \\ 1 \end{pmatrix} \right).$$



De matrix van de spiegeling met betrekking tot deze basis is (net als boven) $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, want v_1 wordt afgebeeld op $1 \cdot v_1 + 0 \cdot v_2$ en v_2 gaat naar $0 \cdot v_1 + (-1) \cdot v_2$.

Om beelden van de vectoren in de standaardbasis te vinden, moeten we de vectoren $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ en $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ als lineaire combinaties van de gekozen basis schrijven. Er geldt

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \frac{3}{10} \begin{pmatrix} -3 \\ 1 \end{pmatrix} \text{ en } \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{3}{10} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} -3 \\ 1 \end{pmatrix}.$$

De coördinaten van het beeld van $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ zijn dus $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{10} \\ \frac{-3}{10} \end{pmatrix} = \begin{pmatrix} \frac{1}{10} \\ \frac{3}{10} \end{pmatrix}$ en dus is het beeld van $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ gelijk aan

$$\frac{1}{10}v_1 + \frac{3}{10}v_2 = \begin{pmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{pmatrix}.$$

Op dezelfde manier volgt dat het beeld van $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ gelijk aan $\begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \end{pmatrix}$ is. Met betrekking tot de standaardbasis heeft de spiegeling in de as $y = 3x$ dus de matrix $A = \frac{1}{5} \begin{pmatrix} -4 & 3 \\ 3 & 4 \end{pmatrix}$.

Ook projecties zijn lineaire afbeeldingen. Als we bijvoorbeeld de loodrechte projectie van de 3-dimensionale ruimte in de $x - y$ -vlakte bekijken, dan wordt $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ op $\begin{pmatrix} x \\ y \end{pmatrix}$ afgebeeld. De matrix die hier bij hoort is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

2.5 Kern en beeld

Om een lineaire afbeelding goed te kunnen beschrijven zijn twee begrippen nuttig: de *kern* en het *beeld* van een lineaire afbeelding.

De *kern* van een lineaire afbeelding $f : V \rightarrow W$ bestaat uit de vectoren, die door f naar 0 worden afgebeeld, dus

$$\ker(f) := \{v \in V \mid f(v) = 0\}.$$

Merk op, dat met $v \in \ker(f)$ ook alle scalaire veelvouden $\lambda v \in \ker(f)$ zijn, en dat uit $v, w \in \ker(f)$ volgt, dat $v + w \in \ker(f)$ is. De kern is dus afgesloten onder het vormen van lineaire combinaties en is dus een deelruimte van V .

Het *beeld* van een lineaire afbeelding is de verzameling van vectoren, waarop een vector afgebeeld wordt, dus

$$\text{im}(f) := \{f(v) \in W \mid v \in V\}.$$

Ook hier geldt, dat lineaire combinaties van elementen in het beeld ook in het beeld liggen, dus is het beeld een deelruimte van W .

Het bepalen van de kern van een lineaire afbeelding komt erop neer, het homogene stelsel lineaire vergelijkingen op te lossen, dat door de matrix van de afbeelding gegeven is. Voor de kern zijn we namelijk op zoek naar coördinaten $\lambda_1, \dots, \lambda_n$ zo dat $f(\lambda_1 v_1 + \dots + \lambda_n v_n) = \lambda_1 f(v_1) + \dots + \lambda_n f(v_n) = 0$. Met andere woorden zoeken we een vector $\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$ met $A \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = 0$. Als we dit uitschrijven is het niets anders dan het stelsel lineaire vergelijkingen met matrix A en de λ_i als onbekende.

Met hetzelfde argument zien we nu ook hoe we kunnen testen of een stelsel vectoren onafhankelijk of een basis is: We schrijven de vectoren v_i als kolommen in een matrix A en lossen het bijhorende homogene stelsel lineaire vergelijkingen op. Een vector in de kern van de lineaire afbeelding met matrix A levert dan een combinatie $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$ en als niet alle $\lambda_i = 0$ zijn is het stelsel afhankelijk. Maar omdat we het hier over een homogeen stelsel hebben is dit het geval dan en slechts dan als de rijtrapvorm van A een vrije parameter laat zien (minder pivots dan kolommen heeft). Het stelsel vectoren is dus onafhankelijk als de kern van de matrix A alleen maar uit de nulvector bestaat.

Omdat het aantal elementen in elke basis hetzelfde is, is een onafhankelijk stelsel vectoren automatisch een basis van een vectorruimte V van dimensie n als het n elementen bevat. Let op dat je niet de dimensie van V met het aantal componenten van de vectoren uit V verwarrt, dit kan namelijk echt groter dan de dimensie zijn.

Het beeld van een lineaire afbeelding vinden we door de afbeelding op een basis los te laten. Dit levert vectoren op die het beeld voortbrengen maar deze zijn niet noodzakelijk onafhankelijk. Als we een basis (v_1, v_2, \dots, v_n) afbeelden, dan is $f(v_i)$ voor het voortbrengen van het beeld alleen maar nodig als $f(v_i)$ geen lineaire combinatie van $f(v_1), \dots, f(v_{i-1})$ is.

Er bestaat een belangrijke samenhang tussen de dimensies van de kern en het beeld van een lineaire afbeelding. Stel dat f een lineaire afbeelding van een n -dimensionale vectorruimte V naar een m -dimensionale vectorruimte W is, dan is de matrix van f een $m \times n$ -matrix. Om de kern van de afbeelding te bepalen, brengen we de matrix op (gereduceerde) rijtrapvorm. Omdat we voor de kern het homogene stelsel oplossen, is de dimensie van de kern het aantal vrije parameters en dit is het verschil van het aantal n van kolommen en het aantal pivot-kolommen.

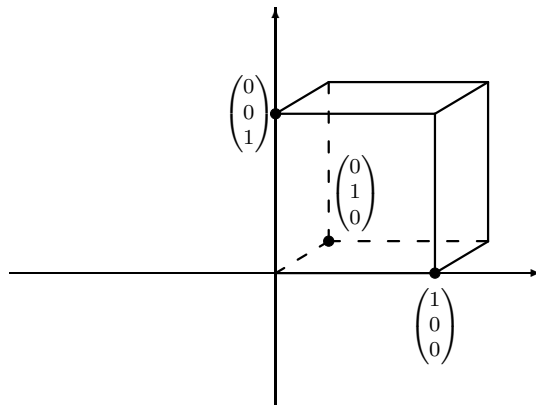
Aan de andere kant kunnen we van de gereduceerde rijtrapvorm heel makkelijk de dimensie van het beeld aflezen, dat is namelijk precies het aantal pivot-kolommen. Verder is het niet moeilijk om in te zien, dat de elementaire operaties waarmee we naar de rijtrapvorm zijn gekomen de dimensie van het beeld niet hebben veranderd (terwijl het beeld zelf wel is veranderd). Dit ligt eraan dat we de elementaire operaties terug kunnen draaien en dat een lineaire relatie tussen kolommen ook na het toepassen van een elementaire operatie nog bestaat. We hebben dus de volgende stelling ingezien:

I.11 Stelling *Als f een lineaire afbeelding van een n -dimensionale vectorruimte is, dan is de som van de dimensies van $\ker(f)$ en $\text{im}(f)$ gelijk aan n .*

Het voordeel van deze stelling ligt voor de hand: we hoeven alleen maar één van de dimensies van $\ker(f)$ of $\text{im}(f)$ te bepalen en kunnen de andere daaruit afleiden.

Als we bijvoorbeeld een lineaire afbeelding met matrix $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & 2 & 3 \end{pmatrix}$ hebben, dan berekenen we dat de rijtrapvorm hiervan $\begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$ is en concluderen dus dat de kern 1-dimensionaal en het beeld 2-dimensionaal is. We zien dat de eerste twee kolommen $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ en $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ in de (originele) matrix van de afbeelding onafhankelijk zijn, dus zijn deze vectoren een basis voor het beeld en daarom zijn de vectoren in het beeld van de vorm $\begin{pmatrix} x + y \\ x \\ x + 2y \end{pmatrix}$.

We bekijken de begrippen kern en beeld van een lineaire afbeelding nog eens aan de hand van een projectie die een kubus op de volgende manier in het 2-dimensionaal vlak afbeeldt:



We kiezen de kubus zo dat de drie aangegeven hoekpunten met de vectoren uit de standaardbasis in \mathbb{R}^3 overeen komen. Bij de projectie beelden we nu de vector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ af op de vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in het 2-dimensionale vlak en de vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ op $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in het vlak. Dit heeft het effect dat de voorkant van de kubus een niet vertekend vierkant wordt. Door de keuze van het beeld van de derde basisvector bepalen we nu of we meer van voren of van boven op de kubus kijken. In het plaatje is het beeld van $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ de vector $\begin{pmatrix} 1/5 \\ 1/5 \end{pmatrix}$.

De matrix van deze projectie (met betrekking tot de standaardbases) is $\begin{pmatrix} 1 & \frac{1}{3} & 0 \\ 0 & \frac{1}{5} & 1 \end{pmatrix}$ want de kolommen van deze matrix zijn de beelden van de basisvectoren in de standaardbasis van \mathbb{R}^3 .

Het beeld van een algemeen vector berekenen we nu door

$$\begin{pmatrix} 1 & \frac{1}{3} & 0 \\ 0 & \frac{1}{5} & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + \frac{1}{3}y \\ \frac{1}{5}y + z \end{pmatrix}.$$

De kern van deze projectie is precies de lijn langs welke we projecteren want dit zijn de punten die op het nulpunt in de vlakke afgebeeld worden. Merk op dat alle lijnen die parallel met deze projectielijn zijn ook op één punt in het 2-dimensionale vlak afgebeeld worden. We berekenen de kern door het homogene stelsel vergelijkingen met als matrix de matrix van de afbeelding op te lossen. In ons geval heeft deze matrix al rijtrapvorm, we kunnen de z -coördinaat als

vrije parameter kiezen en krijgen zo als kern de lijn $\left\{ \begin{pmatrix} 5t \\ -15t \\ 3t \end{pmatrix} \in \mathbb{R}^3 \mid t \in \mathbb{R} \right\}$.

Omdat de vectoren $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ en $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in het beeld van de projectie liggen is het duidelijk dat het beeld van deze projectie het hele vlak \mathbb{R}^2 is. Dit hadden we ook uit het feit kunnen concluderen dat de dimensie van de kern 1 is, want dan moet die dimensie van het beeld volgens stelling I.11 gelijk aan $3 - 1 = 2$ zijn.

2.6 Stelsels lineaire vergelijkingen als lineaire afbeeldingen

We hebben in de vorige les stelsels lineaire vergelijkingen in een schema geschreven en vervolgens met behulp van eenvoudige operaties op dit schema de oplossingen van het schema bepaald. We kunnen nu met de nieuwe terminologie van lineaire afbeeldingen de resultaten herformuleren.

Laat A de $m \times n$ -matrix van een stelsel lineaire vergelijkingen met m vergelijkingen en n onbekenden zijn. Verder zij b de rechterzijde van het stelsel. De lineaire afbeelding met matrix A (met betrekking tot de standaardbases) noemen we f . Dan geldt:

- Het stelsel is oplosbaar dan en slechts dan als b in $\text{im}(f)$ ligt.
- Het aantal vrije parameters voor de oplossingen is de dimensie van $\ker(f)$.
- De oplossingen zijn van de vorm $P+h$, waarbij P een particuliere oplossing is en $h \in \ker(f)$.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- vectorruimte, deelruimte
- onafhankelijke vectoren, basis, dimensie

- lineaire afbeelding, affiene afbeelding
- matrix van een lineaire afbeelding
- kern en beeld van een lineaire afbeelding

OPGAVEN

8. Een (stomme) robot bij de robocup mikt altijd in de richting waarin hij het laatst gelopen is, behalve als hij expres nog een extra rotatie uitvoert. Onze robot is van het punt $(-10, 10)$ naar de bal toe gelopen die op het punt $(20, -10)$ ligt. Nu draait hij nog om 30 graden (linksom) en haalt vervolgens uit. Het doel ligt tussen de punten $(100, -5)$ en $(100, 5)$. Wordt dit een doelpunt (als we van tegenstanders afzien)? Hoe zit het met een rotatie om 45 graden in plaats van 30? Kun je een hoek gokken, waar het een doelpunt wordt?
9. De veeltermen van graad $\leq n$ vormen een $(n + 1)$ -dimensionale vectorruimte met basis $B = (1, x, x^2, \dots, x^n)$. De afgeleide $f'(x)$ van een veelterm $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ is gedefinieerd als $f'(x) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \dots + a_1$.
- Ga na dat de afbeelding $f(x) \rightarrow f'(x)$ een lineaire afbeelding is, d.w.z. toon aan dat $(f + g)'(x) = f'(x) + g'(x)$ en $(\lambda f(x))' = \lambda f'(x)$.
 - Bepaal voor $n = 3$ de matrix A van deze lineaire afbeelding (m.b.t. de basis B).
 - Bepaal kern en beeld van A .
10. Ook de afbeelding die een veelterm $f(x)$ naar $g(x) := f(x+1)$ afbeeldt is een lineaire afbeelding (bijvoorbeeld gaat x^2 naar $(x+1)^2 = x^2 + 2x + 1$).
- Bepaal voor $n = 4$ de matrix A van deze lineaire afbeelding (m.b.t. de basis B uit de vorige opgave).
 - Bepaal (met behulp van de matrix A) het beeld van de veelterm $f(x) = x^4 - 3x^3 + 5x^2 - 7x + 9$.
11. Bepaal alle veeltermen van graad ≤ 3 die een oplossing voor de differentiaalvergelijking $x f'(x) - f(x+1) = 0$ zijn.
(Aanwijzing: Ga na dat de afbeelding $f(x) \rightarrow x f'(x) - f(x+1)$ lineair is. De kern van deze lineaire afbeelding is dan de verzameling van oplossingen van de differentiaalvergelijking.)
12. Bij een vaak gebruikte projectie van \mathbb{R}^3 naar \mathbb{R}^2 gaat de vector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ naar $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, de vector $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ naar $\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ en de vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ naar $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
- Bepaal de matrix van deze projectie (m.b.t. de standaardbases).
 - Bepaal het beeld van een kubus met hoekpunten $(\pm 1, \pm 1, \pm 1)$ bij deze projectie. Bereken de beelden van de hoekpunten en maak een plaatje.
 - Bereken de kern van deze projectie. Dit zijn de punten die onder de projectie naar de oorsprong in de vlakke gaan.

13. De kubus met hoekpunten $(\pm 1, \pm 1, \pm 1)$ wordt zo om de as door $(-1, -1, -1)$ en $(1, 1, 1)$ gedraaid dat hij weer hetzelfde uitziet. Hiervoor zijn er twee mogelijkheden, waarvan we degene kiezen waarbij het punt $(1, 1, -1)$ naar $(1, -1, 1)$ gaat.
- (i) Toon aan dat de vectoren naar de punten $(1, 1, -1)$, $(1, -1, 1)$ en $(-1, 1, 1)$ onafhankelijk en dus een basis voor \mathbb{R}^3 zijn. Dit is een geschikte basis omdat deze drie punten de punten zijn die in de kubus met $(1, 1, 1)$ verbonden zijn en daarom door de rotatie onder elkaar verwisseld worden.
 - (ii) Bepaal de matrix van deze rotatie met betrekking tot de in (i) gekozen basis.
 - (iii) Bepaal het beeld van de punten $(1, 0, 0)$ (middenpunt van een zijvlak) en $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4})$ (punt ergens binnen de kubus).

Les 3 Matrix product

We hebben gezien hoe we matrices kunnen gebruiken om lineaire afbeeldingen te beschrijven. Om het beeld van een vector onder een afbeelding te bepalen hebben we al een soort product van een matrix met een vector gedefinieerd. In deze les zullen we nu naar producten van matrices kijken en zien wat we daarmee kunnen.

3.1 Vermenigvuldiging van matrices

Om de manier, hoe we matrices vermenigvuldigen, te motiveren, kijken we eerst naar een toepassing. Begin 2002 zijn de Euro-munten geïntroduceerd en we kunnen ons afvragen hoe het met de verspreiding van de munten uit de verschillende landen zit. (Er zijn zelfs onderzoeksgroepen die regelmatig *Euro-scouts* bericht laten geven hoeveel van de verschillende munten ze op een gegeven moment in hun portemonnee hebben om zo de verdere verspreiding te voorspellen.) In een heel eenvoudig model kunnen we *Euroland* splitsen in Nederland en niet-Nederland. Om het proces van de verspreiding van de munten te beschrijven, moeten we alleen maar afschatten, hoeveel munten in een jaar van Nederland naar het buitenland gaan en hoeveel munten van het buitenland naar Nederland binnen komen.

Stel dat in een jaar een tiende van de Nederlandse munten naar het buitenland gaan. Verder zullen misschien nog 2% van de munten verdwijnen, dus blijven 88% van de munten in eigen land. Omgekeerd nemen we aan dat van de munten in het buitenland elk jaar 1% naar Nederland komen en 5% verdwijnen. Wat verwachten we na 1, 2, 3 jaar voor een mix van munten in Nederland?

We kunnen het proces van de verspreiding van de munten door een *overgangsmatrix* beschrijven, waar we in de eerste kolom schrijven waar de munten uit Nederland naar toe gaan, in de tweede, waar de munten uit het buitenland naar toe gaan en in de derde, waar de verdwenen munten terecht komen (deze blijven natuurlijk verdwenen). Voor de boven aangegeven schattingen is deze matrix dus

$$A = \begin{pmatrix} 0.88 & 0.01 & 0 \\ 0.10 & 0.94 & 0 \\ 0.02 & 0.05 & 1 \end{pmatrix}$$

Als we nu willen weten hoeveel Nederlandse en buitenlandse munten na 1 jaar in Nederland terecht komen, hoeven we alleen maar A met de vector $v = \begin{pmatrix} N \\ B \\ 0 \end{pmatrix}$ te vermenigvuldigen, waarbij N het aantal geslagen Nederlandse munten en B het aantal geslagen buitenlandse munten zijn. Het resultaat is de vector

$$\begin{pmatrix} 0.88 \cdot N + 0.01 \cdot B \\ 0.10 \cdot N + 0.94 \cdot B \\ 0.02 \cdot N + 0.05 \cdot B \end{pmatrix}$$

die precies weer geeft, dat na een jaar in Nederland 88% Nederlandse en 1% buitenlandse munten zijn. De tweede component geeft de mix voor het buitenland en de derde de mix van verdwenen munten.

Hoe zit het nu na het tweede en na het derde jaar? In principe doen we hetzelfde opnieuw, we vermenigvuldigen de overgangsmatrix met de vector waarin de hoeveelheden munten staan, dus met het resultaat na het eerste jaar. Dit geeft dus

$$\begin{aligned}
 A &= \begin{pmatrix} 0.88 & 0.01 & 0 \\ 0.10 & 0.94 & 0 \\ 0.02 & 0.05 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.88N + 0.01B \\ 0.10N + 0.94B \\ 0.02N + 0.05B \end{pmatrix} \\
 &= \begin{pmatrix} 0.88 \cdot (0.88N + 0.01B) + 0.01 \cdot (0.10N + 0.94B) \\ 0.10 \cdot (0.88N + 0.01B) + 0.94 \cdot (0.10N + 0.94B) \\ 0.02 \cdot (0.88N + 0.01B) + 0.05 \cdot (0.10N + 0.94B) + 1 \cdot (0.02N + 0.05B) \end{pmatrix} \\
 &= \begin{pmatrix} 0.7754N + 0.0182B \\ 0.182N + 0.8846B \\ 0.0426N + 0.0972B \end{pmatrix}
 \end{aligned}$$

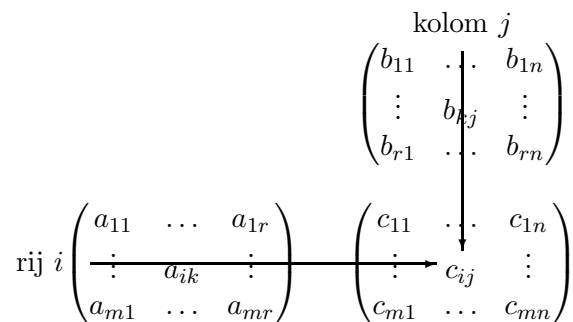
Dit zegt dat na twee jaar 18.2% van de Nederlandse munten in het buitenland zitten en nog 77.54% in Nederland.

Om de mix na drie jaar te bepalen kunnen we nu in principe weer de matrix A met de resultaatvector na twee jaar vermenigvuldigen. Maar we kunnen dit ook op een andere manier beschrijven.

We kunnen namelijk niet alleen een matrix met een vector vermenigvuldigen, maar ook een matrix met een matrix als het aantal kolommen van de eerste matrix hetzelfde is als het aantal rijen van de tweede matrix. Om het product $A \cdot B$ voor een $m \times r$ -matrix A en een $r \times n$ -matrix B te bepalen, beschouwen we B als een collectie van n kolommen. Hoe we een matrix met een kolom vermenigvuldigen hebben we al eerder gezien. Als we A dus met de n kolommen van B vermenigvuldigen, geeft dit n kolommen met telkens m componenten, en deze schrijven we (in de goede volgorde) weer als een $m \times n$ -matrix.

In formules wordt dit zo geschreven: Als A de $m \times r$ -matrix (a_{ij}) is met a_{ij} het element in de i -de rij en j -de kolom, en B de $r \times n$ -matrix (b_{ij}) , dan is $C = A \cdot B$ de $m \times n$ -matrix (c_{ij}) met $c_{ij} = \sum_{k=1}^r a_{ik}b_{kj}$.

Dit ziet er erger uit als het is. Een goede manier om de vermenigvuldiging van matrices te onthouden is het volgende schema:



Om het element c_{ij} in de i -de rij en j -de kolom van C te bepalen lopen we gelijktijdig over de i -de rij van A en de j -de kolom van B . De elementen van A

en B waar we tegelijkertijd op staan worden vermenigvuldigd en deze producten worden bij het doorlopen opgeteld.

In het voorbeeld van de verspreiding van de Euro-munten hebben we de mix na één jaar bepaald door $A \cdot v$ te berekenen, de mix na twee jaar was gegeven door de vector $(A \cdot (A \cdot v))$ en dus wordt de mix na drie jaar de vector $(A \cdot (A \cdot (A \cdot v)))$. Maar dit kunnen we ook berekenen als $(A \cdot A \cdot A) \cdot v$ waar we ook $A^3 \cdot v$ voor schrijven. Met de boven aangegeven methode vinden we dat

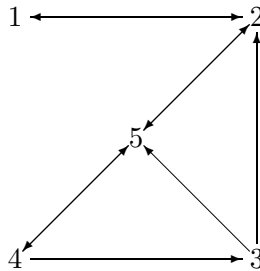
$$A^2 = A \cdot A = \begin{pmatrix} 0.7754 & 0.0182 & 0 \\ 0.1820 & 0.8846 & 0 \\ 0.0426 & 0.0972 & 1 \end{pmatrix}$$

en

$$A^3 = A^2 \cdot A = \begin{pmatrix} 0.684172 & 0.024862 & 0 \\ 0.248620 & 0.833344 & 0 \\ 0.067208 & 0.141794 & 1 \end{pmatrix}$$

De mix na afloop van drie jaar vinden we door de matrix A^3 met de vector $v = \begin{pmatrix} N \\ B \\ 0 \end{pmatrix}$ te vermenigvuldigen, en hieruit kunnen we zien dat er na drie jaar van de in Nederland geslagen munten 68.4% nog (of weer) in Nederland, 24.9% in het buitenland en 6.7% verdwenen zijn.

Een verdere toepassing van het machtsverheffen van matrices is de analyse van netwerken. Neem aan we hebben het volgende netwerk van (geörienteerde) paden:



Als we de vraag willen beantwoorden hoe lang het kortste pad van 1 naar 3 is, kunnen we dit aan de hand van de machten van een geschikte matrix aflezen. Voor een netwerk met n punten gebruiken we een $n \times n$ -matrix $A = (a_{ij})$, waarbij $a_{ij} = 1$ is als er een directe verbinding van punt i naar punt j bestaat en $a_{ij} = 0$ anders. Voor het voorbeeld is dus

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

We kunnen deze matrix nu ook anders interpreteren: Het element a_{ij} zegt, hoeveel paden er van het punt i naar het punt j zijn. Als we nu $C = A^2$

berekenen, dan is het element c_{ij} hiervan het aantal van paden van lengte 2 die van i naar j gaan. Immers, elk pad van lengte 2 van i naar j is van de vorm $i \rightarrow k \rightarrow j$ en het aantal van deze paden is $a_{ik}a_{kj}$. Als we deze aantallen voor alle k optellen, hebben we dus het aantal paden van lengte 2 van i naar j .

Als we dit herhalen, vinden we met hetzelfde argument dat het element c_{ij} in de matrix $C = A^m$ het aantal van paden van lengte m is die van punt i naar punt j gaan.

In het voorbeeld is

$$A^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 2 \end{pmatrix}, A^3 = \begin{pmatrix} 0 & 2 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 3 \\ 1 & 2 & 1 & 1 & 2 \\ 2 & 1 & 1 & 1 & 3 \\ 0 & 4 & 0 & 2 & 1 \end{pmatrix}, A^4 = \begin{pmatrix} 2 & 0 & 1 & 0 & 3 \\ 0 & 6 & 0 & 3 & 1 \\ 2 & 4 & 1 & 2 & 4 \\ 1 & 6 & 1 & 3 & 3 \\ 4 & 1 & 2 & 1 & 6 \end{pmatrix}$$

dus zien we dat het kortste pad van punt 1 naar punt 3 lengte 4 heeft en dat alle andere verbindingen hooguit lengte 3 nodig hebben.

We kunnen ook eenvoudig het aantal paden van lengte $\leq m$ bepalen, dit is namelijk het element c_{ij} in $C = A + A^2 + \dots + A^m$.

3.2 Samenstelling van lineaire afbeeldingen

In de vorige les hebben we gezien hoe we bijvoorbeeld rotaties en spiegelingen in het 2-dimensionale vlak of in de 3-dimensionale ruimte kunnen beschrijven. Nu we weten hoe we matrices vermenigvuldigen, kunnen we ook zeggen hoe we de samenstelling van twee (of meer) lineaire afbeeldingen kunnen beschrijven. Hierbij betekent samenstelling dat we eerst bijvoorbeeld een rotatie en dan een spiegeling op een vector toepassen.

Stel dat A de matrix van een lineaire afbeelding f en B de matrix van een lineaire afbeelding g is. We willen op een (coördinaten-)vector v eerst f en op het resultaat dan g toepassen. Hiervoor zeggen we ook dat we de *samengestelde* of *gecomponeerde afbeelding* $g \circ f$ toepassen. Het beeld onder f vinden we als $u = A \cdot v$ en het beeld hiervan onder g dan als $w = B \cdot u = B \cdot (A \cdot v) = (B \cdot A) \cdot v$. Dus, in plaats van de twee afbeeldingen achter elkaar toe te passen kunnen we ook hun matrices (in de goede volgorde) vermenigvuldigen en vervolgens het product met de vector vermenigvuldigen.

Als we bijvoorbeeld eerst om een hoek φ willen draaien en vervolgens in de as $y = x$ willen spiegelen is de matrix van de samenstelling van deze afbeeldingen het product

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} = \begin{pmatrix} \sin(\varphi) & \cos(\varphi) \\ \cos(\varphi) & -\sin(\varphi) \end{pmatrix}$$

Waarschuwing: Let op dat de vermenigvuldiging van matrices bijna nooit commutatief is, d.w.z. in het algemeen is $A \cdot B \neq B \cdot A$. Bijvoorbeeld heeft de spiegeling in de x -as de matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ en de rotatie om 90 graden de matrix $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Als we eerst de rotatie en dan de spiegeling toepassen,

wordt dit door de matrix $A \cdot B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ beschreven. Het is eenvoudig om na te gaan dat dit de spiegeling in de as $y = x$ is. Als we omgekeerd eerst de spiegeling en dan de rotatie toepassen hoort hierbij de matrix $B \cdot A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ en dit is de spiegeling in de as $y = -x$.

3.3 Inverse van een matrix

Bij de stelsels lineaire vergelijkingen hebben we als argument dat we door een elementaire operatie geen oplossingen verliezen of erbij krijgen gebruikt, dat we de operatie kunnen omkeren. In de taal van lineaire afbeeldingen heet dit dat we bij een afbeelding f van een vectorruimte naar zichzelf een afbeelding g zoeken zo dat de samenstelling van f en g de *identieke afbeelding* is, die elke vector op zich zelf afbeeldt. Omdat de identieke afbeelding in het bijzonder elke basis vector op zich zelf afbeeldt, is de matrix van de identieke afbeelding de matrix

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

met 1 op de diagonaal en 0 elders, die we de *identiteitsmatrix* noemen en met I_n noteren.

Als A de $n \times n$ -matrix van de afbeelding f is, hebben we dus een $n \times n$ -matrix B nodig zo dat $B \cdot A = I_n$ is. De matrix B noemen we dan de *inverse* van A en noteren deze met A^{-1} (zo als we dat bij getallen gewend zijn).

We kunnen ook omgekeerd naar een matrix C zoeken zo dat $A \cdot C = I_n$ is, maar als we zo'n matrix gevonden hebben, kunnen we deze vergelijking van links met A^{-1} vermenigvuldigen, dit geeft $A^{-1} \cdot A \cdot C = A^{-1}$ en dus $C = A^{-1}$ omdat $A^{-1} \cdot A = I_n$. De samenstelling van een afbeelding en zijn inverse geeft dus onafhankelijk van de volgorde de identieke afbeelding.

Hoe kunnen we de inverse matrix A^{-1} nu bepalen?

De eigenschap die we eisen is, dat $A^{-1} \cdot A = I_n$ of, wat equivalent is, dat $A \cdot A^{-1} = I_n$ is. Maar het laatste kunnen we als stelsel lineaire vergelijkingen opvatten, waarbij A de matrix van het stelsel vergelijkingen is en de kolommen van A^{-1} de vectoren met de onbekenden zijn. Voor de i -de kolom v_i van A^{-1}

geldt, dat hij voldoet aan $A \cdot v_i = e_i$ is, waarbij e_i de i -de standaardvector $\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$

(met 1 in de i -de component en 0 elders) is. Het inverteren van een $n \times n$ -matrix komt dus neer op het oplossen van een stelsel lineaire vergelijkingen met n verschillende rechterzijden. Hiervoor brengen we het stelsel op rijtrapvorm, waarbij we de verschillende rechterzijden simultaan behandelen.

We zien op deze manier eenvoudig in, dat een $n \times n$ -matrix A dan en slechts dan een inverse heeft (we zeggen ook: *inverteerbaar* is), als de rijtrapvorm geen 0-rij heeft, of, met andere woorden, als de rijtrapvorm van A geen vrije parameters maar n pivots heeft. In dit geval is namelijk het stelsel vergelijkingen met elke rechterzijde eenduidig oplosbaar. Omgekeerd hebben we ook nodig, dat het stelsel met elke rechterzijde oplosbaar is, omdat de eenheidsvectoren (e_1, \dots, e_n) een basis van \mathbb{R}^n , de vectorruimte van alle rechterzijden, vormt.

I.12 Stelling *Een $n \times n$ -matrix A is dan en slechts dan inverteerbaar als het bij A horende stelsel lineaire vergelijkingen met elke rechterzijde oplosbaar is. Dit is dan en slechts dan het geval als de rijtrapvorm van A in elke kolom een pivot heeft.*

Hier is een voorbeeld: We willen de inverse van de matrix $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ bepalen. Hiervoor brengen we het volgende stelsel lineaire vergelijkingen met drie rechterzijden op gereduceerde rijtrapvorm:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right)$$

De stappen hiervoor zijn:

$$\begin{aligned} &\xrightarrow{\substack{o_{1,2}(-2) \\ o_{1,3}(-1)}} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array} \right) \xrightarrow{w_{2,3}} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & -3 & 1 & 1 \end{array} \right) \\ &\xrightarrow{v_3(-\frac{1}{2})} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array} \right) \xrightarrow{o_{3,1}(-1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{-1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array} \right) \end{aligned}$$

Dus is $A^{-1} = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 \\ -2 & 0 & 2 \\ 3 & -1 & -1 \end{pmatrix}$ de inverse van A .

Voor 2×2 -matrices kunnen we zelfs in het algemeen bepalen, of een matrix inverteerbaar is en de inverse dan ook aangeven. Voor een matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ bepalen we net als boven de rijtrapvorm van het bijhorende stelsel lineaire vergelijkingen. We nemen eerst aan, dat $a \neq 0$ is. De stappen zijn dan:

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array} \right) \xrightarrow{o_{1,2}(-\frac{d}{a})} \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d - \frac{cb}{a} & -\frac{c}{a} & 1 \end{array} \right) \xrightarrow{v_2(a)} \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array} \right).$$

De matrix A is dus inverteerbaar dan en slechts dan als $ad - bc \neq 0$, en in dit geval kunnen we verder doorgaan naar de gereduceerde rijtrapvorm:

$$\xrightarrow{o_{2,1}(-\frac{b}{ad-bc})} \left(\begin{array}{cc|cc} a & 0 & \frac{ad}{ad-bc} & -\frac{ba}{ad-bc} \\ 0 & ad - bc & -c & a \end{array} \right) \rightarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array} \right).$$

Dus bestaat de inverse als $ad - bc \neq 0$ is en is voor het geval $a \neq 0$ gelijk aan

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Voor het geval $a = 0$ vinden we als rijtrapvorm de matrix $\left(\begin{array}{cc|cc} c & 0 & -\frac{d}{b} & 1 \\ 0 & b & 1 & 0 \end{array} \right)$ als $b \neq 0$ is. Als $b = 0$ is, is het duidelijk dat de matrix niet inverteerbaar is, evenzo moet $c \neq 0$ zijn. We komen dan verder naar de gereduceerde rijtrapvorm $\left(\begin{array}{cc|cc} 1 & 0 & -\frac{d}{bc} & \frac{1}{c} \\ 0 & 1 & \frac{1}{b} & 0 \end{array} \right)$. Dit is inderdaad hetzelfde als de boven gevonden inverse A^{-1} , als we daar $a = 0$ invullen, dus is A^{-1} de algemene vorm van de inverse van een 2×2 -matrix.

3.4 Basistransformatie

In de vorige les hebben we gezien dat het soms handig is om de matrix van een lineaire afbeelding met betrekking tot een andere basis dan de standaardbasis te schrijven. De overgang van de matrix van een lineaire afbeelding met betrekking tot één basis naar de matrix met betrekking tot een andere basis noemen we een *basistransformatie*. Met behulp van de inverse matrix kunnen we een basistransformatie nu heel simpel beschrijven.

Het enige wat hiervoor nodig is, is de matrix van de lineaire afbeelding die een basis naar de andere vertaalt. Zij A de matrix van de afbeelding f met betrekking tot de basis $B = (v_1, \dots, v_n)$, en laat T de matrix zijn die de basisvectoren van een nieuwe basis $C = (w_1, \dots, w_n)$ met betrekking tot de basis B uitdrukt. Hiermee wordt bedoeld dat de i -de kolom van T de coördinaten van de (nieuwe) basisvector w_i met betrekking tot de (oude) basis B bevat. Als B speciaal de standaardbasis is, dan zijn de kolommen van T gewoon de vectoren w_1, \dots, w_n . De matrix van de lineaire afbeelding f met betrekking tot de basis C is dan

$$A' = T^{-1} \cdot A \cdot T.$$

Waarom is dit zo?

De matrix T drukt de vectoren uit de basis C in de basis B uit. Omdat T^{-1} de inverse matrix is, drukt deze omgekeerd de vectoren uit de basis B in de nieuwe basis C uit. Als we de samenstelling $A' = T^{-1} \cdot A \cdot T$ nu van rechts naar links lezen en op de i -de basisvector w_i toe passen, dan geeft $T \cdot w_i$ de coördinaten van de vector w_i met betrekking tot de basis B en $A \cdot T \cdot w_i$ is dus de coördinaten van het beeld $f(w_i)$ met betrekking tot de basis B . Door nu nog met T^{-1} te vermenigvuldigen drukken we dit beeld weer met betrekking tot de basis C uit en vinden dus zo de gewenste coördinaten van het beeld $f(w_i)$ met betrekking tot (w_1, \dots, w_n) .

In een voorbeeld in de vorige les hebben we naar de spiegeling in de lijn $y = 3x$ gekeken. We waren toen (met iets moeite) erachter gekomen dat deze spiegeling met betrekking tot de standaardbasis $B = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$ de matrix

$A = \frac{1}{5} \begin{pmatrix} -4 & 3 \\ 3 & 4 \end{pmatrix}$ heeft.

Als we als nieuwe basis de basis $C = \left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \end{pmatrix} \right)$ met één vector in de spiegellingslijn en één vector loodrecht erop kiezen, kunnen we de transformatiematrix T meteen opschrijven, die is namelijk $T = \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}$. Met de formule voor de inverse van een 2×2 -matrix berekenen we T^{-1} en krijgen: $T^{-1} = \frac{1}{10} \begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}$.

De matrix van de spiegeling in de lijn $y = 3x$ heeft dus met betrekking tot de nieuwe basis de matrix

$$A' = \frac{1}{10} \begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix} \cdot \frac{1}{5} \begin{pmatrix} -4 & 3 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We hadden overigens net zo goed met de eenvoudige matrix A' kunnen beginnen en hieruit de matrix met betrekking tot de standaardbasis kunnen afleiden. Uit $A' = T^{-1} \cdot A \cdot T$ volgt namelijk (door van links met T en van rechts met T^{-1} te vermenigvuldigen) dat $A = T \cdot A' \cdot T^{-1}$ en dit is in principe de berekening die we in de afgelopen les stapsgewijs hebben uitgevoerd.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- vermenigvuldiging en machtsverheffing van matrices
- overgangsmatrix
- matrix van een netwerk
- samenstelling van lineaire afbeeldingen
- inverse van een matrix
- basistransformatie

OPGAVEN

14. Bij een zeldzame soort van vogels geven observaties de volgende informatie: In het eerste jaar zijn de vogels in hun jeugd, in het tweede zijn ze pubers en in het derde zijn ze volwassen. Van de volwassen vrouwtjes hebben elk jaar 33% een vrouwelijk kuiken. Slechts 18% van de vrouwelijke kuikens overleven het eerste jaar, 71% van de vrouwelijke pubers overleven het tweede jaar en van de volwassen vrouwtjes overlijden per jaar 6%.
- (i) Beschrijf de ontwikkeling van deze populatie door een (3×3) overgangsmatrix, waarbij het element in de i -de kolom en j -de rij van de matrix voor de overgang van het i -de naar het j -de jaar staat.

- (ii) Bij een telling van de vrouwelijke vogels worden op een gegeven moment 300 jeugdige vogels, 50 pubers en 1000 volwassen vogels geteld. Hoeveel vogels van de verschillende leeftijden verwachten we na afloop van 1, 2, 3 jaar.
- (iii) Heb je een verwachting over de ontwikkeling van deze populatie? Kun je er redenen voor aangeven?
15. In het 2-dimensionale vlak zij f de draaiing om een hoek van 45 graden en g de spiegeling in de lijn $y = x$. Bepaal de matrices van de samengestelde afbeeldingen

$$(i) f \circ g \quad (ii) g \circ f \quad (iii) g \circ f \circ g.$$

16. Ga na of de volgende matrices inverteerbaar zijn en bepaal, zo ja, de inverse matrix:

$$(i) \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} \quad (ii) \begin{pmatrix} 2 & 1 & 0 \\ -1 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} \quad (iii) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 1 & 2 & 2 & 2 \\ 1 & 3 & 2 & 4 \end{pmatrix}.$$

17. Van een lineaire afbeelding is bekend, dat de vectoren op de lijn $y = x$ onveranderd blijven en dat de vectoren op de lijn $y = -x$ om een factor 2 verlengd worden. Geef de matrix van deze lineaire afbeelding met betrekking tot de standaardbasis aan.
18. Een netwerk met 5 punten wordt beschreven door de matrix

$$A = (a_{ij}) = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

waarbij $a_{ij} = 1$ betekent dat de punten i en j door een ribbe verbonden zijn en $a_{ij} = 0$ dat ze niet (direct) verbonden zijn.

- (i) Teken een plaatje van het netwerk.
- (ii) Welke paren van punten zijn door een eenduidig pad van lengte hoogstens 2 verbonden?
- (iii) Bepaal het aantal paden van lengte (precies) 3 die van punt 2 naar punt 3 lopen.

Les 4 Eigenwaarden en eigenvectoren

In het voorbeeld van de verspreiding van de Euro-munten hebben we gezien hoe we de mix van munten na afloop van n jaar uit de n -de macht A^n van de overgangsmatrix af kunnen lezen. Een voor de hand liggende vraag is, of er uiteindelijk een stabiel evenwicht wordt bereikt dat niet meer verandert.

Nou zijn we in het voorbeeld ervan uit gegaan dat munten verdwijnen en deze munten ook verdwenen blijven. Dan is duidelijk dat uiteindelijk alle munten verdwijnen, en de toestand waarin in elk land helemaal geen munten meer zijn is natuurlijk stabiel, maar ook flauw. Een betere aanpak is dat we de verdwenen munten door nieuw geslagen munten weer opvullen.

In het voorbeeld krijgen we zo de nieuwe overgangsmatrix

$$A = \begin{pmatrix} 0.90 & 0.01 \\ 0.10 & 0.99 \end{pmatrix}.$$

Het totale aantal munten blijft altijd hetzelfde omdat alle kolommen de som 1 hebben. Als er een stabiele evenwichtstoestand bereikt wordt, dan voldoet de vector v met de hoeveelheden munten in Nederland en het buitenland aan $A \cdot v = v$. Anders gezegd is $A \cdot v - v = 0$ ofwel v ligt in de kern van $A - I$ (waarbij I de identiteitsmatrix is). Dit is een makkelijk probleem, want we hoeven alleen maar een stelsel lineaire vergelijkingen op te lossen. Dit stelsel heeft de matrix

$$A - I = \begin{pmatrix} -0.10 & 0.01 \\ 0.10 & -0.01 \end{pmatrix}$$

en de kern hiervan is $t \cdot \begin{pmatrix} 1 \\ 10 \end{pmatrix}$. Het evenwicht ziet er dus zo uit, dat tien keer zo veel munten in het buitenland dan in Nederland zijn. Verder kunnen we ook bepalen, hoe de mix er uitziet, die zal namelijk in Nederland en in het buitenland hetzelfde zijn. Waarom dat zo is zullen we later in deze les zien. Als we hogere en hogere machten van A berekenen zullen die uiteindelijk naar

$$A_\infty = \begin{pmatrix} 1/11 & 1/11 \\ 10/11 & 10/11 \end{pmatrix}$$

convergeren. Merk op dat de kolommen van A_∞ veelvoudigen van de evenwichtsvector zijn die zo gescaled zijn dat de som van de componenten 1 is.

4.1 Eigenvectoren

Om een evenwichtstoestand te vinden hebben we een vector v gezocht waarvoor geldt dat $A \cdot v = v$ is. Iets algemener is het een belangrijke taak om vectoren te vinden die door een lineaire afbeelding alleen maar met een factor vermenigvuldigd worden, maar niet van richting veranderen. Dit zijn dus vectoren die aan de vergelijking

$$A \cdot v = \lambda v$$

voldoen, waarbij λ een scalaire factor is. Een vector v die aan de vergelijking $A \cdot v = \lambda v$ voldoet heet een *eigenvector* voor de *eigenwaarde* λ (soms ook *eigenvector met eigenwaarde* λ).

De evenwichtsvector van boven is dus een eigenvector voor de eigenwaarde 1. We hebben al eerder met eigenvectoren te maken gehad, zonder ze zo te noemen. Bijvoorbeeld zijn bij een spiegeling in het vlak de vectoren op de spiegelingas eigenvectoren voor de eigenwaarde 1 en de vectoren op een lijn loodrecht op de spiegelingas zijn eigenvectoren voor de eigenwaarde -1 . Vectoren die in de kern van een lineaire afbeelding liggen zijn natuurlijk eigenvectoren voor de eigenwaarde 0.

Merk op: De 0-vector noemen we geen eigenvector, want hij is een eigenvector voor elke eigenwaarde, en we willen graag dat een eigenvector bij een unieke eigenwaarde hoort.

Het is duidelijk dat voor een eigenvector v met eigenwaarde λ ook alle veelvoud cv (met $c \neq 0$) eigenvectoren met eigenwaarde λ zijn, want uit $A \cdot v = \lambda v$ volgt wegens de lineariteit van A dat $A \cdot (cv) = c(A \cdot v) = \lambda(cv)$.

De vraag is nu natuurlijk hoe we eigenwaarden en eigenvectoren van een lineaire afbeelding kunnen vinden. Het belangrijke punt is, dat een eigenvector v met eigenwaarde λ voldoet aan $A \cdot v = \lambda v$, ofwel $(A - \lambda \cdot I) \cdot v = 0$. De vector v ligt dus in de kern van de matrix $A - \lambda \cdot I$ en dit kan alleen maar voor een vector $v \neq 0$ gebeuren als het stelsel lineaire vergelijkingen met matrix $A - \lambda \cdot I$ een vrije parameter heeft. Omdat A even veel rijen en kolommen heeft, kan dit alleen maar gebeuren als we in de rijtrapvorm van de matrix een 0-rij vinden. We hebben dus gezien:

Een eigenwaarde van A is een getal λ zo dat de rijtrapvorm van $A - \lambda \cdot I$ een vrije parameter en dus een 0-rij heeft.

4.2 Determinanten

Sommigen hebben zeker eens van de determinant van een matrix gehoord. Dat is een getal, die we voor een $n \times n$ -matrix kunnen berekenen. De voor ons belangrijkste eigenschap van de determinant is:

Een $n \times n$ -matrix is dan en slechts dan inverteerbaar als zijn determinant niet 0 is.

Verder is er ook een meetkundige interpretatie: De determinant is het volume van het parallellepipedum dat door de kolommen van de matrix opgespannen wordt.

De samenhang van inverteerbaarheid en de determinant van een matrix volgt meteen uit de manier hoe we de determinant kunnen berekenen. Hier is een methode, die misschien iets van achter door de borst lijkt, maar die in feite de manier is, hoe in software pakketten als MAPLE of MATHEMATICA de determinant wordt berekend. Het belangrijkste ingrediënt in deze methode is weer de rijtrapvorm.

I.13 Karakterisatie van de determinant

- Voor een matrix in rijtrapvorm is de determinant het product van de elementen op de diagonaal.

- Als we twee rijen (of kolommen) van een matrix verwisselen, wordt de determinant met -1 vermenigvuldigd.
- Als we een rij (of kolom) met een getal c vermenigvuldigen, wordt ook de determinant met c vermenigvuldigd.
- Als we een veelvoud van een rij (of kolom) bij een andere optellen, verandert de determinant niet.

Er is een stelling die zegt, dat de determinant door deze karakterisatie (tot op een schaling na) eenduidig bepaald is, maar dit gaan we hier gewoon als een feit accepteren. De manier, hoe we de determinant van een matrix berekenen is dus heel eenvoudig: We brengen de matrix op rijtrapvorm, waarbij we de vermenigvuldigingen met factoren en de verwisselingen van rijen in een product c verwerken, dan is de determinant het product van de diagonaalelementen gedeeld door c .

Het zal nu duidelijk zijn dat een $n \times n$ -matrix dan en slechts dan inverteerbaar is als de determinant niet 0 is: De matrix is inverteerbaar als de rijtrapvorm n pivots heeft die niet 0 zijn, maar dan is ook de determinant niet nul. Omgekeerd is een matrix niet inverteerbaar als er in de rijtrapvorm een 0-rij is, en dan is ook de determinant 0.

Voor 2×2 -matrices kunnen we de determinant voor een algemene matrix berekenen: De matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

heeft voor $a \neq 0$ de rijtrapvorm

$$\begin{pmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{pmatrix}$$

en omdat we geen rijen hebben verwisseld en ook geen rij met een factor hebben vermenigvuldigd is de determinant

$$\det(A) = a \cdot \left(d - \frac{c}{a}b\right) = ad - bc.$$

Dit klopt ook voor $a = 0$, want dan verwisselen we de rijen en hebben als rijtrapvorm $\begin{pmatrix} c & d \\ 0 & b \end{pmatrix}$ en omdat we een keer rijen hebben verwisseld is de determinant $-bc$.

De determinant van een 2×2 -matrix is dus het product van de diagonaalelementen min het product van de dwarsdiagonaal-elementen.

Ook voor 3×3 -matrices kunnen we nog een algemene formule uit de rijtrapvorm berekenen. We nemen weer een algemene matrix

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

en negeren voor het moment de gevallen waar een noemer 0 kan zijn. De stappen naar de rijtrapvorm zijn nu:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \rightarrow \begin{pmatrix} a & b & c \\ 0 & e - \frac{d}{a}b & f - \frac{d}{a}c \\ 0 & h - \frac{g}{a}b & i - \frac{g}{a}c \end{pmatrix} = \begin{pmatrix} a & b & c \\ 0 & \frac{1}{a}(ae - bd) & \frac{1}{a}(af - cd) \\ 0 & \frac{1}{a}(ah - bg) & \frac{1}{a}(ai - cg) \end{pmatrix}$$

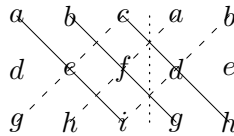
$$\rightarrow \begin{pmatrix} a & b & c \\ 0 & \frac{1}{a}(ae - bd) & \frac{1}{a}(af - cd) \\ 0 & 0 & \frac{1}{a}(ai - cg) - \frac{ah - gb}{ae - bd} \frac{af - cd}{a} \end{pmatrix}.$$

Als product van de diagonaalelementen vinden we

$$\begin{aligned} \det(A) &= \frac{1}{a}(ae - bd)(ai - cg) - \frac{1}{a}(ah - gb)(af - cd) \\ &= \frac{1}{a}(a^2ei - aceg - abdi + bcdg - a^2fh + acdh + abfg - bcdg) \\ &= aei + bfg + cdh - ceg - afh - bdi. \end{aligned}$$

Ook in dit geval kunnen we nagaan dat deze formule ook voor de exceptionele gevallen $a = 0$ en $ae - bd = 0$ geldt.

Een goede manier om het berekenen van de determinant van een 3×3 -matrix te onthouden is het volgende plaatje:



Schrijf de eerste twee kolommen nog eens rechts naast de matrix en teken dan de drie diagonalen en de drie dwarsdiagonalen. De determinant is dan de som van de producten op de diagonalen min de producten op de dwarsdiagonalen.

Waarschuwing: Het schema van de 3×3 matrices geldt niet voor grotere matrices. Vanaf grootte 4×4 is het inderdaad nodig de matrix op rijtrapvorm te brengen.

4.3 Het vinden van eigenwaarden en eigenvectoren

Om eigenwaarden en eigenvectoren te vinden hoeven we nu alleen nog de resultaten van boven te combineren. Een eigenvector voor de eigenwaarde λ is een vector in de kern van $A - \lambda \cdot I$ en de kern bevat dan en slechts dan andere elementen dan de 0-vector als de determinant $\det(A - \lambda \cdot I) = 0$ is. Hierbij interpreteren we λ als een onbekende, dan wordt de determinant van $A - \lambda \cdot I$ een veelterm in de onbekende λ . Deze veelterm heet het *karakteristieke polynoom* van de matrix A en de eigenwaarden van A zijn juist de nulpunten van het karakteristieke polynoom.

Als voorbeeld bepalen we de eigenwaarden en eigenvectoren van de matrix

$$A := \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}.$$

We hebben

$$A - \lambda \cdot I = \begin{pmatrix} 1 - \lambda & 2 \\ -1 & 4 - \lambda \end{pmatrix} \text{ en}$$

$$\det(A - \lambda \cdot I) = (1 - \lambda)(4 - \lambda) - (-2) = \lambda^2 - 5\lambda + 4 + 2 = \lambda^2 - 5\lambda + 6.$$

Om de eigenwaarden van A te vinden moeten we dus de nulpunten van $\lambda^2 - 5\lambda + 6$ bepalen. Dit kan altijd met de abc -formule, maar in dit geval zien we makkelijk dat $\lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3)$ en we hebben dus

$$\lambda^2 - 5\lambda + 6 = 0 \Leftrightarrow \lambda = 2 \text{ of } \lambda = 3.$$

De eigenwaarden van A zijn dus 2 en 3.

Het bepalen van de eigenvectoren is nu eenvoudig. We weten dat de matrix $A - \lambda \cdot I$ voor de gevonden eigenwaarden λ een vector $\neq 0$ in de kern heeft, en elke van deze vectoren is een eigenvector. Merk op dat met een eigenvector v ook elke veelvoud $c \cdot v$ een eigenvector voor dezelfde eigenwaarde is. We vinden dus de eigenvectoren van A als volgt:

$\lambda = 2$: $A - \lambda \cdot I = \begin{pmatrix} -1 & 2 \\ -1 & 2 \end{pmatrix}$ heeft kern $t \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, dus is $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ een eigenvector van A voor de eigenwaarde 2. Men gaat eenvoudig na dat inderdaad $A \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$.

$\lambda = 3$: $A - \lambda \cdot I = \begin{pmatrix} -2 & 2 \\ -1 & 1 \end{pmatrix}$ heeft kern $t \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, dus is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ een eigenvector van A voor de eigenwaarde 3. Men gaat weer na dat $A \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$.

Voor een algemene 2×2 -matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is de determinant van $A - \lambda \cdot I$ de determinant van $\begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$ en dus gelijk aan $(a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$. Dit is een kwadratische veelterm waarvan we met de abc -formule de nulpunten kunnen bepalen.

De matrix in het voorbeeld van de Euro-munten was $A = \begin{pmatrix} 0.90 & 0.01 \\ 0.10 & 0.99 \end{pmatrix}$ en hiervoor hebben we

$$\det(A - \lambda \cdot I) = \lambda^2 - 1.89\lambda + 0.89 = (\lambda - 1)(\lambda - 0.89).$$

De eigenwaarden zijn dus 1 en 0.89 en we kunnen voor deze eigenwaarden nu ook de eigenvectoren bepalen:

De eigenvectoren voor de eigenwaarde 1 zijn de vectoren in de kern van $A - 1 \cdot I = \begin{pmatrix} -0.10 & 0.01 \\ 0.10 & -0.01 \end{pmatrix}$ en die is gelijk aan $t \cdot \begin{pmatrix} 1 \\ 10 \end{pmatrix}$ (wat we al wisten).

De eigenvectoren voor de eigenwaarde 0.89 zijn de vectoren in de kern van $A - 0.89 \cdot I = \begin{pmatrix} 0.01 & 0.01 \\ 0.10 & 0.10 \end{pmatrix}$ en die is gelijk aan $t \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

4.4 Limieten van overgangsmatrices

Bij processen die we door een overgangsmatrix A beschrijven zijn we vaak geïnteresseerd in de ontwikkeling op langere termijn. Hiervoor hebben we de machten A^m voor grotere waarden van m nodig en soms zelfs een limiet voor $m \rightarrow \infty$. Dit is natuurlijk een lastig probleem, maar als we de eigenwaarden van A kennen en een basis uit eigenvectoren kunnen vinden, valt het inderdaad mee.

Stel dat we voor een $n \times n$ -matrix A de eigenwaarden hebben bepaald en n lineair onafhankelijke basisvectoren hebben gevonden. Als (v_1, \dots, v_n) de eigenvectoren zijn en $(\lambda_1, \dots, \lambda_n)$ de bijhorende eigenwaarden, dan zij T de matrix met v_i als i -de kolom. Door kolomsgewijs te vergelijken zien we dat $A \cdot T = T \cdot D$, waarbij D een diagonaalmatrix met de λ_i op de diagonaal is:

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Hieruit volgt dat $A = T \cdot D \cdot T^{-1}$ en $D = T^{-1} \cdot A \cdot T$.

Het aardige is dat we de machten A^m van A nu veel makkelijker kunnen berekenen, want

$$(T \cdot D \cdot T^{-1})^m = T \cdot D^m \cdot T^{-1}$$

omdat

$$\begin{aligned} & (T \cdot D \cdot T^{-1}) \cdot (T \cdot D \cdot T^{-1}) \cdot \dots \cdot (T \cdot D \cdot T^{-1}) \\ &= T \cdot D \cdot (T^{-1}T) \cdot D \cdot \dots \cdot (T^{-1}T)DT^{-1} = T \cdot D \cdot D \cdot \dots \cdot D \cdot T^{-1}. \end{aligned}$$

Maar de machten van D zijn gewoon

$$D^m = \begin{pmatrix} \lambda_1^m & 0 & \dots & 0 \\ 0 & \lambda_2^m & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n^m \end{pmatrix}$$

en hiervoor kunnen we voor $m \rightarrow \infty$ makkelijk zeggen wat er gaat gebeuren: Als $|\lambda_i| < 1$ dan gaat λ_i^m naar 0, als $|\lambda_i| > 1$ dan gaat λ_i^m naar ∞ , en als $|\lambda_i| = 1$ hangt het ervan af of $\lambda_i = 1$, $\lambda_i = -1$ of λ_i een complex getal is.

Bij veel processen heeft de overgangsmatrix A de eigenschap dat de som van alle kolommen gelijk is aan 1. In dit geval laat zich aantonen dat 1 altijd een eigenwaarde is, dat de kern van $A - I$ dimensie 1 heeft (dus dat er op scalaire na een unieke eigenvector met eigenwaarde 1 bestaat) en dat alle andere eigenwaarden van absolute waarde < 1 zijn. Als we voor zo'n proces een basis

uit eigenvectoren zo kiezen dat de eerste basisvector de eigenvector $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$

met eigenwaarde 1 is, volgt dat D^m voor grote m naar de matrix gaat waarbij het $(1, 1)$ -element 1 is en alle andere elementen 0 zijn. Maar dan gaat A^m naar

$$A_\infty = T \cdot \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1}$$

$$= \begin{pmatrix} v_1 & 0 & \dots & 0 \\ v_2 & 0 & \dots & 0 \\ & & \ddots & \\ v_n & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1} = \begin{pmatrix} s_1 v_1 & s_2 v_1 & \dots & s_n v_1 \\ s_1 v_2 & s_2 v_2 & \dots & s_n v_2 \\ & & \ddots & \\ s_1 v_n & s_2 v_n & \dots & s_n v_n \end{pmatrix}$$

waarbij (s_1, \dots, s_n) de elementen in de eerste rij van T^{-1} zijn. In dit geval is dus A_∞ een matrix waarin elke kolom een veelvoud van de eigenvector v met eigenwaarde 1 is.

In het voorbeeld van de Euro-munten vinden we dat de eerste rij van T^{-1} gelijk is aan $(\frac{1}{11}, \frac{1}{11})$, dus gaan de machten A^m voor de matrix $A = \begin{pmatrix} 0.90 & 0.01 \\ 0.10 & 0.99 \end{pmatrix}$ naar $A_\infty = \begin{pmatrix} \frac{1}{10} & \frac{1}{10} \\ \frac{1}{11} & \frac{1}{11} \end{pmatrix}$. Dit zegt dat er uiteindelijk zowel in Nederland als ook in het buitenland relatief hetzelfde aandeel van Nederlandse en van buitenlandse munten terecht komt, maar dat er in het buitenland tien keer zo veel munten zullen zijn dan in Nederland (onafhankelijk van de hoeveelheden munten in het begin).

In het voorbeeld van opgave 14 is de overgangsmatrix

$$A = \begin{pmatrix} 0 & 0 & 0.33 \\ 0.18 & 0 & 0 \\ 0 & 0.71 & 0.94 \end{pmatrix}.$$

De eigenwaarden hiervan zijn de wortels van de veelterm $\lambda^2(\lambda - 0.94) - 0.042174$. Deze zijn niet zo makkelijk met de hand te berekenen, maar bijvoorbeeld met een computeralgebra pakket als MAPLE vinden we als numerieke benaderingen: $\lambda_1 = 0.9835927398$, $\lambda_2 = -0.02179636988 + 0.2059184805i$, $\lambda_3 = -0.02179636988 - 0.2059184805i$. De eigenwaarden λ_2 en λ_3 zijn complexe getallen met absolute waarde 0.2070688348, dus zijn alle eigenwaarden van absolute waarde < 1 , en dus gaan de machten van A tegen de nulmatrix. De vogels sterven dus uit, als er niets verandert!

4.5 Hoofdcomponenten analyse

Bij de statistische evaluatie kijken we vaak naar het gemiddelde van de metingen en naar de afwijkingen van de metingen van het gemiddelde. Als de metingen x_1, \dots, x_n zijn, dan is de gemiddelde waarde (verwachtingswaarde) $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ en de kwadratische afwijkingen (variantie) zijn dan $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$.

Als we nu een experiment hebben met verschillende parameters, dan zijn de metingen vectoren van een aantal componenten. Het gemiddelde vinden we weer door de metingen op te tellen en door het aantal metingen te delen,

dit geeft voor drie parameters bijvoorbeeld een vector $\mu = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}$. Het is nu

niet meer vanzelfsprekend dat de grootste spreiding van de metingen in één van de richtingen van de parameters ligt, want misschien is er een combinatie van parameters, die het grootste effect op het experiment heeft. We willen dus graag de richting bepalen waarin de spreiding maximaal wordt. De afstand in

de richting van de vector $v = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ die een vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ van de gemiddelde

vector μ heeft, bepalen we als $(a(x - \mu_x) + b(y - \mu_y) + c(z - \mu_z))^2$ (dit zullen we in de volgende les nog verder toelichten) en voor de spreiding in deze richting krijgen we dus (analoog als boven):

$$\sigma_v^2 = \sum_{i=1}^n (a(x_i - \mu_x) + b(y_i - \mu_y) + c(z_i - \mu_z))^2.$$

Dit kunnen we (een beetje kunstmatig) ook als een product van matrices schrijven, dan wordt het

$$\begin{aligned} \sigma_v^2 &= \sum_{i=1}^n (a \quad b \quad c) \cdot \begin{pmatrix} x_i - \mu_x \\ y_i - \mu_y \\ z_i - \mu_z \end{pmatrix} \cdot (x_i - \mu_x \quad y_i - \mu_y \quad z_i - \mu_z) \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\ &= (a \quad b \quad c) \cdot \left(\sum_{i=1}^n \begin{pmatrix} x_i - \mu_x \\ y_i - \mu_y \\ z_i - \mu_z \end{pmatrix} \cdot (x_i - \mu_x \quad y_i - \mu_y \quad z_i - \mu_z) \right) \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \end{aligned}$$

De matrix $V = \sum_{i=1}^n \begin{pmatrix} x_i - \mu_x \\ y_i - \mu_y \\ z_i - \mu_z \end{pmatrix} \cdot (x_i - \mu_x \quad y_i - \mu_y \quad z_i - \mu_z)$ heet de *covariantiematrix*.

We zijn geïnteresseerd in de richting zo dat de spreiding σ_v maximaal wordt. Maar dat gebeurt precies als we voor v de eigenvector voor de grootste eigenwaarde van V kiezen.

In het algemeen wordt deze methode gebruikt om uit een groot aantal parameters de belangrijkste combinaties eruit te vissen. Deze combinaties heten de *hoofdcomponenten*. Men neemt dan bijvoorbeeld in een experiment met honderd parameters de eigenvectoren die bij de grootste tien eigenwaarden horen en heeft hiermee vaak de belangrijkste eigenschappen voor het experiment beschreven.

Een voorbeeld is de automatische spraakherkenning, waarbij de signalen door een Fourier-analyse met betrekking tot een groot aantal frequenties worden beschreven. Om nu de verschillende klinkers en medeklinkers te kunnen

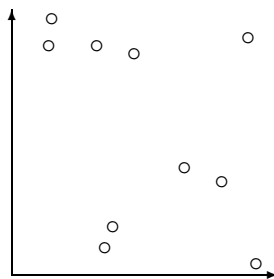
onderscheiden, wordt gekeken welke combinaties van frequenties een grote spreiding hebben en op die manier wordt het signaal met veel minder parameters beschreven. Een voordeel hierbij is ook nog, dat de gereduceerde parameters robuuster tegen ruis en andere storingen zijn.

Een andere toepassing zit in het comprimeren van informatie. Bij een kleurplaatje wordt vaak de kleur in het RGB-systeem aangegeven, dat wil zeggen er wordt een waarde voor de intensiteiten van de kleuren rood, groen en blauw aangegeven. Meestal zijn de kleuren niet onafhankelijk, dan laat het plaatje zich ook met de intensiteiten van twee combinaties van rood, groen en blauw beschrijven, de spreiding in de richting van de derde eigenvector is dan heel klein.

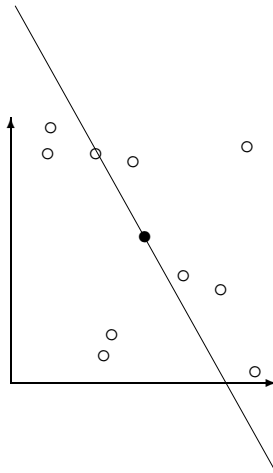
We gaan een 2-dimensionaal voorbeeld bekijken, om te zien hoe de hoofdcomponenten analyse werkt. In het volgende plaatje zijn er tien punten met de coördinaten:

$$\begin{aligned}(x_1, y_1) &= (14, 86) \\(x_2, y_2) &= (15, 96) \\(x_3, y_3) &= (92, 4) \\(x_4, y_4) &= (65, 40) \\(x_5, y_5) &= (35, 10) \\(x_6, y_6) &= (89, 89) \\(x_7, y_7) &= (79, 35) \\(x_8, y_8) &= (32, 86) \\(x_9, y_9) &= (38, 18) \\(x_{10}, y_{10}) &= (46, 83).\end{aligned}$$

(De x -waarden zijn de eerste 10 paren van cijfers in π , de y -waarden de eerste 10 paren van cijfers in π^2 .)



Het middelpunt heeft de coördinaten $(50.5, 54.7)$ en voor de covariantiematrix vinden we $C = \begin{pmatrix} 7578.5 & -3721.5 \\ -3721.5 & 12162.1 \end{pmatrix}$. De eigenwaarden van C zijn $\lambda_1 = 14240.87313$ en $\lambda_2 = 5499.726869$ en de eigenvector voor de eigenwaarde λ_1 is $\begin{pmatrix} -0.4876624923 \\ 0.8730322409 \end{pmatrix}$. Dit is dus de richting van de grootste spreiding, die in het volgende plaatje als lijn door het middelpunt te zien is.



BELANGRIJKE BEGRIPPEN IN DEZE LES

- eigenwaarden, eigenvectoren
- determinant
- karakteristieke polynoom
- hoofdcomponenten analyse

OPGAVEN

19. Bereken de determinanten van de volgende matrices:

$$A := \begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}.$$

20. Bereken de determinant van de matrix

$$A := \begin{pmatrix} 1+x & 1 & 1 \\ 1 & 1+x & 1 \\ 1 & 1 & 1+x \end{pmatrix}.$$

21. Bepaal voor de matrix $A := \begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}$ de eigenwaarden en eigenvectoren. Beschrijf de lineaire afbeelding met matrix A meetkundig.

22. In de 3-dimensionale ruimte \mathbb{R}^3 heeft een rotatie om een hoek φ rond de z -as de matrix $\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Veronderstel dat de matrix $A := \begin{pmatrix} -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} \\ -\frac{1}{2} & 1 & \frac{1}{2} \\ \frac{13}{6} & -\frac{2}{3} & \frac{1}{6} \end{pmatrix}$ de matrix van een rotatie is, maar niet rond de z -as. Wat is de rotatie-as van de rotatie A ?

23. Zij A de matrix $A := \begin{pmatrix} -4 & 1 \\ -6 & 1 \end{pmatrix}$.

- (i) Bereken het karakteristieke polynoom $\det(A - \lambda \cdot I)$, de eigenwaarden en de eigenvectoren van A .
- (ii) Bepaal voor willekeurige $m \in \mathbb{N}$ de matrix A^m .

24. Bepaal voor de matrix

$$A = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{pmatrix}$$

- (i) het karakteristieke polynoom $\det(A - \lambda \cdot I)$,
- (ii) de eigenwaarden van A ,
- (iii) de eigenvectoren van A ,
- (iv) de matrix A_∞ waar A^m voor grote m naar toe gaat.

(Hint: $\lambda = 1$ is een eigenwaarde.)

Les 5 Inproducten

Als we het in de meetkunde (of elders) over afstanden en hoeken hebben, dan hebben we daar intuïtief wel een idee van. Maar wat is eigenlijk de afstand van twee punten in het 2-dimensionale vlak? Als we twee punten (x_1, y_1) en (x_2, y_2) hebben dan is hun afstand de *lengte* van de vector tussen de twee punten. We kunnen dus het meten van afstanden terug brengen op het meten van de lengte van vectoren. Maar de lengte van een vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is natuurlijk $\sqrt{x^2 + y^2}$. Waarom is dit zo? We passen hier stiekem de stelling van Pythagoras toe, namelijk op de driehoek met hoekpunten $(0, 0)$, $(x, 0)$ en (x, y) . De schuine zijde is dan de vector waarvan we de lengte l willen weten en de rechthoekzijden hebben lengtes x en y , dus geldt $l^2 = x^2 + y^2$.

Ook in de 3-dimensionale ruimte berekenen we de lengte van een vector op een soortgelijke manier, de lengte van $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is $\sqrt{x^2 + y^2 + z^2}$. Ook hier wordt de stelling van Pythagoras toegepast, in dit geval twee keer: eerst in de $x - y$ -vlakke en vervolgens in de vlakke loodrecht op de vector $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$.

Het is nu voor de hand liggend, dat we in een n -dimensionale vectorruimte de lengte van een vector $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ aangeven met $\sqrt{x_1^2 + \dots + x_n^2}$. Dit komt er op neer, de stelling van Pythagoras herhaald $n - 1$ keer toe te passen.

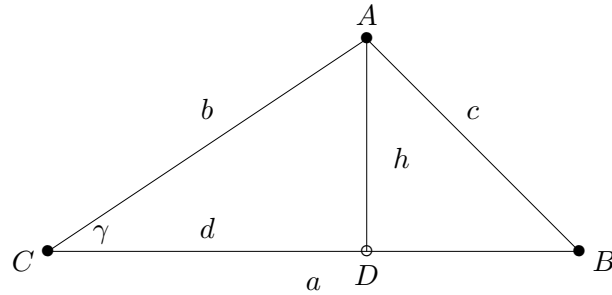
En hoe zit het met hoeken? Als we weten wat een hoek in het 2-dimensionale vlak is zijn we klaar, want ook in een n -dimensionale vectorruimte liggen twee vectoren altijd in een 2-dimensionale deelruimte en we definiëren de hoek tussen de vectoren gewoon als de hoek in dit vlak. Het aardige is nu, dat we een hoek alleen maar uit afstanden kunnen berekenen, hiervoor hebben we trouwens wel de cosinusregel nodig.

I.14 Cosinusregel *In een driehoek met zijden a , b en c en met de hoek γ tussen de zijden a en b geldt:*

$$c^2 = a^2 + b^2 - 2ab \cos(\gamma).$$

In het geval van een rechthoekige driehoek vinden we de stelling van Pythagoras terug, want dan is γ een hoek van 90 graden en dus $\cos(\gamma) = 0$. Als γ een scherpe hoek is (kleiner dan 90 graden) is $\cos(\gamma) > 0$, dus wordt de tegenover γ liggende zijde c in dit geval korter dan in het rechthoekige driehoek, voor een stompe hoek γ wordt c groter.

Het bewijs van de cosinusregel kunnen we eenvoudig uit het volgende plaatje aflezen:



De driehoeken CDA en ADB zijn rechthoekig, dus geldt $b^2 = h^2 + d^2$, $c^2 = h^2 + (a - d)^2$ en $d = b \cos(\gamma)$. Als we nu deze vergelijkingen van elkaar aftrekken krijgen we $b^2 - c^2 = d^2 - a^2 + 2ad - d^2 = -a^2 + 2ad$ en dus $c^2 = a^2 + b^2 - 2ad$. Als we nu $b \cos(\gamma)$ voor d invullen, hebben we precies de bewering van de cosinusregel. Voor een stompe hoek γ werkt het bewijs op een analoge manier.

5.1 Het standaardinproduct

We vertalen het berekenen van een hoek in een driehoek nu naar vectoren. Als we de hoek tussen twee vectoren v en w willen weten, verbinden we de punten van de vectoren, en de vector die daar bij hoort is $v - w$. De lengte van een vector v schrijven we als $\|v\|$ en noemen dit vaak ook de *norm* van de vector v . De cosinusregel zegt dat voor de hoek γ tussen de vectoren v en w geldt dat

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2 \|v\| \|w\| \cos(\gamma).$$

Als we deze vergelijking nu voor de twee vectoren $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ en $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ uitschrijven, krijgen we

$$(v_1 - w_1)^2 + (v_2 - w_2)^2 = v_1^2 + v_2^2 + w_1^2 + w_2^2 - 2 \|v\| \cdot \|w\| \cos(\gamma)$$

dus $-2v_1w_1 - 2v_2w_2 = -2 \|v\| \cdot \|w\| \cos(\gamma)$. Voor de hoek γ tussen v en w geldt dus:

$$\cos(\gamma) = \frac{v_1w_1 + v_2w_2}{\|v\| \|w\|}.$$

De teller van deze formule kunnen we beschouwen als een functie van de vectoren v en w , namelijk $\Phi(v, w) = v_1w_1 + v_2w_2$ als v en w zo als boven gegeven zijn.

Deze functie noemen we een *inproduct* en gebruiken meestal de notatie $\Phi(v, w) = \langle v, w \rangle$. Met deze notatie kunnen we nu de hoek tussen de vectoren v en w definiëren door

$$\cos(\gamma) = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

De berekening waarmee we geconcludeerd hebben dat $\|v\| \|w\| \cos(\gamma) = v_1w_1 + v_2w_2$ gaat precies hetzelfde door als v en w niet meer 2-dimensionaal

maar n -dimensionale vectoren zijn. Voor $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ en $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$ krijgen we dan

$$\|v\|\|w\| \cos(\gamma) = v_1w_1 + v_2w_2 + \dots + v_nw_n.$$

Merk op: Ook in een n -dimensionale ruimte liggen twee vectoren steeds in een 2-dimensionale deelruimte, en de hoek tussen de twee vectoren definiëren we als de gewone hoek in deze 2-dimensionale ruimte.

Het *standaardinproduct* in de n -dimensionale ruimte definiëren we nu door

$$\left\langle \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle := v_1w_1 + v_2w_2 + \dots + v_nw_n = \sum_{i=1}^n v_iw_i.$$

Dit kunnen we ook als een matrix product opvatten, hiervoor schrijven we de eerste vector als een matrix met 1 rij en n kolommen en de tweede als een matrix

met n rijen en 1 kolom, dan krijgen we $\langle v, w \rangle = (v_1 \ \dots \ v_n) \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v^{tr} \cdot w$.

Hierbij is v^{tr} de *getransponeerde* vector van v , d.w.z. de vector v geschreven als een rij in plaats van een kolom.

5.2 Inproducten

Als we het standaardinproduct als matrix product interpreteren, zien we meteen in dat het zekere eigenschappen heeft:

$$\begin{aligned} \langle v_1 + v_2, w \rangle &= \langle v_1, w \rangle + \langle v_2, w \rangle & \text{en} & \quad \langle \lambda v, w \rangle = \lambda \langle v, w \rangle \\ \langle v, w_1 + w_2 \rangle &= \langle v, w_1 \rangle + \langle v, w_2 \rangle & \text{en} & \quad \langle v, \lambda w \rangle = \lambda \langle v, w \rangle \end{aligned}$$

Als we de tweede vector in het inproduct vast laten, hebben we dus een lineaire afbeelding van het eerste argument, net zo krijgen we een lineaire afbeelding van het tweede argument als we de eerste vector vast laten. Dit geeft aanleiding om de inproducten in een algemener context in te bedden. Op een soortgelijke manier kunnen we namelijk een heel algemene klasse van afbeeldingen beschrijven, de *bilineaire afbeeldingen*.

Voor een vectorruimte V noemen we een afbeelding Φ , die aan een paar (v, w) van vectoren een getal $\Phi(v, w) \in \mathbb{R}$ toewijst een *bilineaire afbeelding* als geldt:

- (i) $\Phi(v_1 + v_2, w) = \Phi(v_1, w) + \Phi(v_2, w)$ voor alle $v_1, v_2, w \in V$
- (ii) $\Phi(\lambda v, w) = \lambda \Phi(v, w)$ voor alle $v, w \in V, \lambda \in \mathbb{R}$
- (iii) $\Phi(v, w_1 + w_2) = \Phi(v, w_1) + \Phi(v, w_2)$ voor alle $v, w_1, w_2 \in V$
- (iv) $\Phi(v, \lambda w) = \lambda \Phi(v, w)$ voor alle $v, w \in V, \lambda \in \mathbb{R}$.

Als we voor een bilineaire afbeelding één van de argumenten vast laten, krijgen we dus een lineaire afbeelding in het andere argument, daarom de naam.

We kunnen nu ook eens naar een voorbeeld van een bilineaire afbeelding van een heel andere soort kijken: Neem als vectorruimte de veeltermfuncties van graad ≤ 2 , dan kunnen we een bilineaire afbeelding definiëren door

$$\Psi(f(x), g(x)) := \int_0^1 f(x)g(x)dx.$$

Dat dit inderdaad een bilineaire afbeelding is volgt uit de rekenregels voor integralen. We kiezen nu $(1, x, x^2)$ als basis, dan geldt $\Psi(1, 1) = 1$, $\Psi(1, x) = \Psi(x, 1) = \frac{1}{2}$, $\Psi(1, x^2) = \Psi(x, x) = \Psi(x^2, 1) = \frac{1}{3}$, $\Psi(x, x^2) = \Psi(x^2, x) = \frac{1}{4}$ en $\Psi(x^2, x^2) = \frac{1}{5}$.

We hebben gezien dat het standaardinproduct een speciaal geval van een bilineaire afbeelding is. Maar niet alle bilineaire afbeeldingen zijn geschikt als we het over afstanden, lengtes en hoeken willen hebben.

In het geval van het standaardinproduct zien we dat de lengte $\|v\|$ van een vector v gelijk is aan $\sqrt{\langle v, v \rangle}$. Als we nu op een abstracte vectorruimte een bilineaire afbeelding Φ hebben waarmee we een lengte willen definiëren, dan zullen we precies deze samenhang gebruiken. Omdat het schrijven van wortels soms onhandig wordt, wordt meestal het kwadraat van de lengte gedefinieerd. Maar hier zit al een belangrijk punt in: als $\Phi(v, v) < 0$ is, is het kwadraat van de lengte negatief, maar dan is de lengte een complex en geen reëel getal meer. Ook vectoren van lengte 0 zijn onhandig, we willen dat alleen maar de 0-vector lengte 0 heeft. Hieruit volgen de volgende eisen aan een 'nuttige' bilineaire afbeelding Φ :

- (i) $\Phi(v, v) \geq 0$ voor alle $v \in V$
- (ii) $\Phi(v, v) = 0 \Rightarrow v = 0$.

Een bilineaire afbeelding die aan de eisen (i) en (ii) voldoet noemen we *positief definitief*.

Verder hadden we voor het standaardinproduct gezien dat we de hoek tussen twee vectoren kunnen berekenen uit de formule $\cos(\gamma) = \frac{\Phi(v, w)}{\|v\|\|w\|}$. Maar de hoek is natuurlijk onafhankelijk van de volgorde van de vectoren, dus willen we nog graag hebben dat Φ *symmetrisch* is, dus

- (iii) $\Phi(v, w) = \Phi(w, v)$ voor alle $v, w \in V$.

Een bilineaire afbeelding met de eigenschappen (i), (ii) en (iii), dus een positief definitief, symmetrische bilineaire afbeelding, noemen we een *inproduct* op V .

We zien nu makkelijk in dat het standaardinproduct $\langle v, w \rangle = \sum_{i=1}^n v_i w_i$ een inproduct op \mathbb{R}^n is, de naam is dus terecht gekozen.

Dat de bilineaire afbeelding $\Psi(f(x), g(x)) = \int_0^1 f(x)g(x)dx$ symmetrisch is, is duidelijk en Ψ voldoet ook aan eigenschap (i), want $f(x)^2 \geq 0$ en een integraal over waarden ≥ 0 is zelfs ook ≥ 0 . Er zit wel een klein probleempje met eigenschap (ii). De functie $f(x)$ die alleen maar in het punt 1 de waarde

1 heeft, maar elders 0 is, heeft inderdaad integraal 0 zonder zelf de 0-vector te zijn. Maar dit is een functie met een sprong, dus een niet-continue functie, en als we ons tot continue functies beperken, zien we dat Ψ dan wel positief definitief is: Als $f(x_0)^2 > 0$ dan vinden we een klein stukje om x_0 heen, waar ook $f(x)^2 > 0$ geldt, omdat $f(x)$ geen sprongen heeft. Maar dan hebben we een positieve vlakte onder de grafiek van $f(x)^2$ gevonden en is de integraal dus > 0 .

Waarschuwing: In de toepassingen zijn niet alleen maar inproducten belangrijk. Het meest prominente voorbeeld van een bilineaire afbeelding die wel symmetrisch maar niet positief definitief is, is de afstand in de 4-dimensionale ruimtetijd van de relativiteitstheorie. Hier wordt de zogeheten *Minkowski-afstand* gebruikt die voor een punt met ruimte-coördinaten (x, y, z) en tijd t de afstand $x^2 + y^2 + z^2 - ct^2$ geeft. Hierbij is c de snelheid van het licht. Het idee is dat twee punten afstand nul hebben als hun posities in de ruimtetijd door een lichtsignaal kunnen worden verbonden.

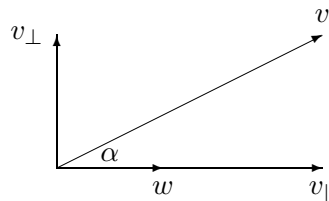
5.3 Orthogonale projecties

In de 2-dimensionale vlakte en de 3-dimensionale ruimte is het duidelijk dat twee vectoren *loodrecht* op elkaar staan als hun standaardinproduct 0 is. We gaan daarom ook in het algemeen zeggen, dat twee vectoren v, w ten opzichte van een inproduct loodrecht op elkaar staan, als $\langle v, w \rangle = 0$ is. We zeggen in dit geval, dat v en w *orthogonaal* zijn en schrijven dit ook als $v \perp w$. Een stelsel (v_1, \dots, v_n) van vectoren (bijvoorbeeld een basis) heet orthogonaal, als $\langle v_i, v_j \rangle = 0$ voor alle $i \neq j$ is. Verder is het nog bijzonder handig als de vectoren van een basis lengte 1 hebben, dus als $\|v_i\|^2 = \langle v_i, v_i \rangle = 1$ voor alle i is. Zo'n basis noemen we *orthonormaal*.

Een van de belangrijkste eigenschappen van het inproduct is, dat het de orthogonale projectie van een vector op de lijn door een andere vector beschrijft. Er geldt namelijk, dat de lineaire afbeelding

$$\pi_w(v) := \langle v, w \rangle \frac{w}{\|w\|^2} = \frac{\langle v, w \rangle}{\langle w, w \rangle} w$$

de orthogonale projectie op de lijn door w weergeeft. Dit kunnen we goed aan het volgende plaatje aflezen:



We schrijven v als $v = v_{\parallel} + v_{\perp}$, waarbij v_{\parallel} parallel met w is en v_{\perp} loodrecht op w staat. Omdat v_{\perp} loodrecht op w staat, geldt $\langle v_{\perp}, w \rangle = 0$ en dus $\langle v - v_{\parallel}, w \rangle = 0$. Maar v_{\parallel} ligt in de richting van w en is dus een veelvoud λw van w . Er geldt dus $\langle v - \lambda w, w \rangle = \langle v, w \rangle - \langle \lambda w, w \rangle = 0$ en dus $\langle v, w \rangle = \lambda \langle w, w \rangle$. We hebben dus aangetoond dat $v_{\parallel} = \lambda w = \frac{\langle v, w \rangle}{\langle w, w \rangle} w$.

Maar hoe kunnen we nu de orthogonale projectie op een 2-dimensionaal vlak of zelfs een hoger-dimensionale deelruimte berekenen? We kunnen voor een basis natuurlijk de orthogonale projecties op de lijnen door de basisvectoren berekenen. Maar let op: Als de basisvectoren niet loodrecht op elkaar staan is er een probleem. De projectie in de richting van de tweede basisvector heeft dan misschien ook een component in de richting van de eerste basisvector, en dan is de orthogonale projectie in het vlak niet de som van de twee projecties op de lijnen.

Waarschuwend voorbeeld: We willen de vector $v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ in het $x - y$ -

vlak projecteren. Het is duidelijk dat de projectie $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ is, want we hoeven alleen maar de z -componente te vergeten. Maar stel we nemen als basis voor het $x - y$ -vlak de vectoren $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ en $v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. Dan is de projectie van v in de richting van v_1 gegeven door $\frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = 1 \cdot v_1$ en de projectie van v in de richting van v_2 is $\frac{\langle v, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 = \frac{3}{2} \cdot v_2$. De som van deze twee projecties is $\begin{pmatrix} \frac{5}{2} \\ \frac{3}{2} \\ 0 \end{pmatrix}$ en dit is niet de juiste projectie in het $x - y$ -vlak.

Dit voorbeeld is een illustratie van de volgende stelling:

I.15 Stelling *De projectie in een deelruimte met basis (v_1, v_2, \dots, v_r) is de som van de projecties in de richtingen van de basisvectoren v_i als de v_i een orthogonale basis vormen. Dit geldt niet als de v_i niet orthogonaal zijn.*

Voordat we met een projectie aan de slag kunnen moeten we dus eerst ervoor zorgen dat de ruimte waar we op willen projecteren een orthogonale basis heeft. Maar dit kunnen we bereiken, door stapsgewijs van elke basisvector de projecties op de eerder behandelde basisvectoren af te trekken. Dit geeft de volgende procedure:

I.16 Gram-Schmidt orthogonalisatie

We gaan uit van een basis (v_1, \dots, v_n) en willen een nieuwe basis (v'_1, \dots, v'_n) vinden, die orthogonaal is. We construeren de orthogonale basis vector voor vector.

De eerste vector v_1 kunnen we ongedeerd laten, we nemen dus $v'_1 := v_1$.

Nu hebben we nodig dat v'_2 loodrecht op v'_1 staat. Hiervoor trekken we de orthogonale projectie van v_2 op de lijn door v'_1 van v_2 af, de rest staat dan loodrecht op v'_1 . We krijgen dus $v'_2 := v_2 - \frac{\langle v_2, v'_1 \rangle}{\langle v'_1, v'_1 \rangle} v'_1$.

Op deze manier gaan we nu door. Van de vector v_i trekken we de projecties op alle eerder bepaalde basisvectoren in de orthogonale basis af, dus krijgen we

algemeen:

$$v'_i := v_i - \left(\sum_{j=1}^{i-1} \frac{\langle v_i, v'_j \rangle}{\langle v'_j, v'_j \rangle} v'_j \right) = v_i - \frac{\langle v_i, v'_1 \rangle}{\langle v'_1, v'_1 \rangle} v'_1 - \dots - \frac{\langle v_i, v'_{i-1} \rangle}{\langle v'_{i-1}, v'_{i-1} \rangle} v'_{i-1}.$$

Als we de basis nog mooier willen maken, kunnen we elk van de orthogonale vectoren nog op lengte 1 brengen door v'_i door $\|v'_i\|$ te delen. Dit geeft dan een orthonormale basis.

We gaan nu een orthogonale basis voor het voorbeeld van de veeltermfuncties met inproduct $\Psi(f(x), g(x)) = \int_0^1 f(x)g(x)dx$ bepalen. We gaan uit van de basis $(v_1, v_2, v_3) = (1, x, x^2)$.

We beginnen met $v'_1 = v_1 = 1$ en hebben $v'_2 = v_2 - \frac{\Psi(v_2, v'_1)}{\Psi(v'_1, v'_1)} v'_1 = x - \frac{1}{1} \cdot 1 = x - \frac{1}{2}$. Tenslotte is $v'_3 = v_3 - \frac{\Psi(v_3, v'_1)}{\Psi(v'_1, v'_1)} v'_1 - \frac{\Psi(v_3, v'_2)}{\Psi(v'_2, v'_2)} v'_2 = x^2 - \frac{1}{1} \cdot 1 - \frac{1}{\frac{1}{12}} (x - \frac{1}{2}) = x^2 - \frac{1}{3} - (x - \frac{1}{2}) = x^2 - x + \frac{1}{6}$.

Het principe van een orthogonale projectie in een deelruimte wordt ook in de Fourieranalyse gebruikt die bijvoorbeeld in de analyse en verwerking van geluidsignalen een rol speelt. Hierbij worden periodieke functies (trillingen) beschreven door hun projectie in de deelruimte die door de functies $\sin(kx)$ en $\cos(kx)$ wordt opgespannen. De functies $\sin(kx)$ en $\cos(kx)$ vormen namelijk een orthogonaal stelsel en worden door een geschikte scaling een orthonormale basis.

Bij de hoofdcomponenten analyse hebben we de covariantiematrix $A = \sum_{i=1}^n v_i \cdot v_i^{tr}$ berekend, waarbij elke vector v_i het verschil tussen een gemeten vectoren en het gemiddelde van de gemeten vectoren is. De eigenvector van de matrix A voor de grootste eigenwaarde geeft de richting van de grootste spreiding van de resultaten. Het aardige is nu, dat de eigenvectoren voor verschillende eigenwaarden met betrekking tot het standaardinproduct automatisch orthogonaal zijn omdat A een symmetrische matrix is, d.w.z. omdat $A^{tr} = A$. Neem aan dat v een eigenvector met eigenwaarde λ en w een eigenvector met eigenwaarde μ is en dat $\lambda \neq \mu$. Er geldt:

$$\begin{aligned} \lambda \langle v, w \rangle &= \langle \lambda v, w \rangle = \langle Av, w \rangle = (Av)^{tr} w = (v^{tr} A^{tr}) w = v^{tr} (A^{tr} w) \\ &= v^{tr} (Aw) = \langle v, Aw \rangle = \langle v, \mu w \rangle = \mu \langle v, w \rangle \end{aligned}$$

dus is $(\lambda - \mu) \langle v, w \rangle = 0$, en dus $\langle v, w \rangle = 0$ omdat $\lambda \neq \mu$.

We hebben dus de volgende belangrijke stelling gevonden:

I.17 Stelling *Voor een symmetrische matrix $A = A^{tr}$ zijn eigenvectoren met verschillende eigenwaarden orthogonaal (met betrekking tot het standaardinproduct).*

In de hoofdcomponenten analyse hoeven we dus alleen maar de eigenvectoren van de m grootste eigenwaarden te kiezen en deze op lengte 1 te normeren, dan hebben we een orthonormaalbasis voor de m -dimensionale deelruimte met de grootste spreiding van metingen en we kunnen ook meteen de orthogonale projectie van de hoger-dimensionale metingen in deze deelruimte aangeven.

5.4 Beste approximatie

We hebben gezien dat we de orthogonale projectie v_0 van een vector v in een deelruimte U met orthogonale basis (v_1, \dots, v_r) kunnen berekenen door

$$v_0 := \frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \dots + \frac{\langle v, v_r \rangle}{\langle v_r, v_r \rangle} v_r.$$

Het aardige aan deze formule is dat we op deze manier de *beste approximatie* van v in de deelruimte U vinden. Dat wil zeggen: v_0 is de vector in U waarvoor het verschil $v - v_0$ de minimale lengte heeft. Om dit in te zien schrijven we $v = v_0 + w$, dan is $\langle w, u \rangle = 0$ voor alle $u \in U$ omdat v_0 de orthogonale projectie van v in U is. Maar als we nu de lengte van het verschil van v met een willekeurige vector $u \in U$ berekenen, dan is

$$\begin{aligned} \|v - u\|^2 &= \|(v - v_0) + (v_0 - u)\|^2 = \langle (v - v_0) + (v_0 - u), (v - v_0) + (v_0 - u) \rangle \\ &= \langle v - v_0, v - v_0 \rangle + \langle v_0 - u, v_0 - u \rangle + 2\langle v - v_0, v_0 - u \rangle \\ &= \|v - v_0\|^2 + \|v_0 - u\|^2 + 2\langle w, v_0 - u \rangle = \|v - v_0\|^2 + \|v_0 - u\|^2. \end{aligned}$$

Maar dit is minimaal als $\|v_0 - u\| = 0$ is, dus als $u = v_0$ is.

Op deze manier kunnen we bijvoorbeeld projecties op een willekeurig 2-dimensionaal vlak in de 3-dimensionale ruimte uitrekenen. Stel we willen de projectie op het vlak U waarin de vectoren $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ en $v_2 = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix}$ liggen.

Eerst hebben we dus een orthogonale basis voor U nodig. Dit wordt (met behulp van de Gram-Schmidt orthogonalisatie) $v'_1 = v_1$ en $v'_2 = v_2 - \frac{1}{3}v'_1 = \frac{1}{3} \begin{pmatrix} -1 \\ 5 \\ -4 \end{pmatrix}$.

De factor $\frac{1}{3}$ kunnen we ook weglaten, daarom kiezen we als orthogonale basis van U de basis $\left(w_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, w_2 = \begin{pmatrix} -1 \\ 5 \\ -4 \end{pmatrix} \right)$. De kwadraten van de lengtes van deze vectoren zijn $\|w_1\|^2 = \langle w_1, w_1 \rangle = 3$ en $\|w_2\|^2 = \langle w_2, w_2 \rangle = 42$.

Voor de inproducten van deze basis met de eenheidsvectoren $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ geldt $\langle e_1, w_1 \rangle = 1$, $\langle e_2, w_1 \rangle = 1$, $\langle e_3, w_1 \rangle = 1$, $\langle e_1, w_2 \rangle = -1$, $\langle e_2, w_2 \rangle = 5$, $\langle e_3, w_2 \rangle = -4$. De projecties van de eenheidsvectoren in het vlak dat door v_1 en v_2 opgespannen is, zijn dus:

$$e_1 \mapsto \frac{1}{3}w_1 + \frac{-1}{42}w_2 = \begin{pmatrix} \frac{1}{3} + \frac{1}{42} \\ \frac{1}{3} - \frac{1}{42} \\ \frac{1}{3} + \frac{1}{42} \end{pmatrix} = \begin{pmatrix} \frac{5}{14} \\ \frac{1}{3} \\ \frac{6}{14} \end{pmatrix},$$

$$e_2 \mapsto \frac{1}{3}w_1 + \frac{5}{42}w_2 = \begin{pmatrix} \frac{1}{3} - \frac{5}{42} \\ \frac{1}{3} + \frac{5}{42} \\ \frac{1}{3} - \frac{5}{42} \end{pmatrix} = \begin{pmatrix} \frac{3}{14} \\ \frac{13}{14} \\ \frac{-2}{14} \end{pmatrix},$$

$$e_3 \mapsto \frac{1}{3}w_1 + \frac{-4}{42}w_2 = \begin{pmatrix} \frac{1}{3} + \frac{4}{42} \\ \frac{1}{3} - \frac{4}{42} \\ \frac{1}{3} + \frac{4}{42} \end{pmatrix} = \begin{pmatrix} \frac{6}{14} \\ \frac{14}{14} \\ \frac{10}{14} \end{pmatrix}.$$

Maar meestal zijn we inderdaad niet geïnteresseerd in de projecties zelf maar in hun coëfficiënten met betrekking tot een orthonormaalbasis van de deelruimte, bijvoorbeeld als we de projecties in een vlak willen tekenen. Als orthonormaalbasis van U krijgen we ($w'_1 = \frac{1}{\sqrt{3}}w_1, w'_2 = \frac{1}{\sqrt{42}}w_2$) en beschouwen deze als de standaardbasis van U . Dit levert de projecties

$$e_1 \mapsto \frac{1}{\sqrt{3}}w'_1 + \frac{-1}{\sqrt{42}}w'_2 \rightarrow \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{42}} \end{pmatrix},$$

$$e_2 \mapsto \frac{1}{\sqrt{3}}w'_1 + \frac{5}{\sqrt{42}}w'_2 \rightarrow \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{5}{\sqrt{42}} \end{pmatrix},$$

$$e_3 \mapsto \frac{1}{\sqrt{3}}w'_1 + \frac{-4}{\sqrt{42}}w'_2 \rightarrow \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{-4}{\sqrt{42}} \end{pmatrix}.$$

5.5 Zoekmachines

Een eenvoudige maar wel redelijk efficiënte manier om een zoekmachine in elkaar te zetten is gebaseerd op een variatie van het standaardinproduct. We nemen aan dat we een aantal documenten D_1, D_2, \dots, D_n hebben en dat deze de woorden t_1, t_2, \dots, t_m (t voor *term*) bevatten. We beschrijven een document of een aanvraag Q (Q voor *query*) nu door de vector $f(Q) = (f_1, f_2, \dots, f_m)$ waarbij f_i aangeeft hoe vaak het woord t_i in Q voorkomt. We zeggen nu dat een aanvraag Q sterk op een document D_i lijkt als de hoek tussen de vectoren $f(Q)$ en $f(D_i)$ klein is, dus als het inproduct $\langle f(Q), f(D_i) \rangle$ tussen de vectoren groot is. Nu komen er nog gewichten bij, omdat sommige woorden belangrijker zijn dan andere: Een woord dat in alle documenten voorkomt geeft geen informatie over de overeenkomst met een aanvraag, terwijl een woord dat alleen in één document voorkomt een heel sterke indicator is. Daarom wordt het inproduct $\langle f(Q), f(D_i) \rangle$ niet met het standaardinproduct berekend maar door

$$\langle f(Q), f(D_i) \rangle := \sum_{j=1}^m w_j f(Q)_j f(D_i)_j.$$

De gewichten w_j worden meestal via de *inverse document frequency* bepaald, namelijk als $w_j := \log(n/n_j)$ waarbij n_j het aantal documenten in de collectie is, dat het woord t_j bevat. Het maat, waarmee een aanvraag Q op een document D_i lijkt is dus

$$l(Q, D_i) := \frac{\langle f(Q), f(D_i) \rangle}{\|f(Q)\| \|f(D_i)\|}$$

waarbij de lengte $\|f(Q)\|$ weer als $\sqrt{\langle f(Q), f(Q) \rangle}$ gedefinieerd is. De documenten worden nu zo gerangschikt dat de waarde $l(Q, D_i)$ afneemt.

In de praktijk zijn er natuurlijk een aantal trucjes bij de implementatie van zo'n zoekmachine. Een van de belangrijkste is de *inverse index*, die voor een woord t_i zegt in welke documenten hij überhaupt voorkomt. Dan hoeft bij een aanvraag alleen maar naar de documenten te worden gekeken, waarin een van de woorden in de aanvraag voorkomt en dat zijn er meestal niet zo erg veel omdat zinvolle aanvragen alleen maar informatie dragende woorden bevatten. Een lijst van heel onbelangrijke woorden (zogenoemde *stopwords*) wordt meestal expliciet uit de lijst van woorden uitgesloten.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- standaardinproduct
- bilineaire afbeelding
- inproduct
- orthogonaliteit, Gram-Schmidt orthogonalisatie
- orthogonale projectie
- beste approximatie

OPGAVEN

25. Bepaal in een gewone kubus de volgende hoeken tussen vectoren die van het middelpunt van de kubus uit gaan:
- (i) Tussen de vectoren naar twee hoekpunten die door een ribbe verbonden zijn.
 - (ii) Tussen de vectoren naar een hoekpunt en het middelpunt van een zijvlak die het hoekpunt bevat.
 - (iii) Tussen de vectoren naar de middelpunten van twee ribben die zich op een zijvlak tegenover liggen.
26. Op de 2×2 -matrices definiëren we een afbeelding $\Phi(A, B) := sp(A \cdot B^{tr})$. Hierbij betekent $sp(X) := X_{11} + X_{22}$ de *spoor* van de matrix X , dus de som van de elementen op de diagonaal.
- (i) Laat zien dat Φ een bilineaire afbeelding is.
 - (ii) Ga na dat Φ een inproduct is.
 - (iii) Bepaal (met behulp van de Gram-Schmidt orthogonalisatie) een orthonormaalbasis van $\mathbb{R}^{2 \times 2}$ met betrekking tot Φ .
27. Bereken de orthogonale projectie (met betrekking tot het standaardinproduct) van de vector $v = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$ in het vlak dat de vectoren $v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ en $v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ bevat.

28. Zij $V = \mathbb{R}^4$ met standaardinproduct. De deelruimte $U \subseteq V$ zij opgespannen door de vectoren $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ en $v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$. Bereken de beste approximatie van $v = \begin{pmatrix} 2 \\ 2 \\ 3 \\ 1 \end{pmatrix}$ in U , d.w.z. de vector $v_0 \in U$ zo dat $\|v - v_0\|$ minimaal is. (Let op dat (v_1, v_2, v_3) nog geen orthogonaal stelsel is.)

Opgaven voor Lineaire Algebra

Opgave 1.

Breng de volgende stelsels lineaire vergelijkingen op rijtrapvorm, concludeer of ze oplosbaar zijn en geef, zo ja, alle oplossingen aan.

$$\begin{array}{ll}
 \begin{array}{l}
 2x_1 - x_2 + 3x_3 = 2 \\
 x_1 + 2x_2 + x_3 = 1 \\
 3x_1 - 4x_2 + 5x_3 = 3
 \end{array} & \text{(i)} \\
 \begin{array}{l}
 x_1 + 2x_2 - x_3 = 0 \\
 2x_1 + 5x_2 + 2x_3 = 0 \\
 x_1 + 4x_2 + 7x_3 = 0 \\
 x_1 + 3x_2 + 3x_3 = 0
 \end{array} & \text{(iii)} \\
 \begin{array}{l}
 x_1 + 2x_2 - 3x_3 + 2x_4 = 2 \\
 2x_1 + 5x_2 - 8x_3 + 6x_4 = 5 \\
 3x_1 + 4x_2 - 5x_3 + 2x_4 = 4
 \end{array} & \text{(v)}
 \end{array}
 \quad
 \begin{array}{ll}
 \begin{array}{l}
 x_1 - x_2 + 2x_3 - 3x_4 = 7 \\
 4x_1 + 3x_3 + x_4 = 9 \\
 2x_1 - 5x_2 + x_3 = -2 \\
 3x_1 - x_2 - x_3 + 2x_4 = -2
 \end{array} & \text{(ii)} \\
 \begin{array}{l}
 x_1 + 2x_2 + 3x_3 = 1 \\
 4x_1 + 5x_2 + 6x_3 = 2 \\
 7x_1 + 8x_2 + 9x_3 = 3 \\
 5x_1 + 7x_2 + 9x_3 = 4
 \end{array} & \text{(iv)} \\
 \begin{array}{l}
 x_1 + 2x_2 + x_3 + x_4 + x_5 = 0 \\
 -x_1 - 2x_2 - 2x_3 + 2x_4 + x_5 = 0 \\
 2x_1 + 4x_2 + 3x_3 - x_4 = 0 \\
 x_1 + 2x_2 + 2x_3 - 2x_4 - x_5 = 0
 \end{array} & \text{(vi)}
 \end{array}$$

Opgave 2.

Bepaal de waarden van de parameter a zo dat het stelsel lineaire vergelijkingen: (a) een eenduidige oplossing heeft, (b) meerdere oplossingen heeft, (c) niet oplosbaar is:

$$\begin{array}{ll}
 \begin{array}{l}
 x_1 + x_2 - x_3 = 1 \\
 2x_1 + 3x_2 + ax_3 = 3 \\
 x_1 + ax_2 + 3x_3 = 2
 \end{array} & \text{(i)} \\
 \begin{array}{l}
 x_1 + x_2 + ax_3 = 2 \\
 3x_1 + 4x_2 + 2x_3 = a \\
 2x_1 + 3x_2 - x_3 = 1
 \end{array} & \text{(ii)}
 \end{array}$$

Opgave 3.

Aan welke voorwaarden moeten a , b en c voldoen opdat het volgende stelsel lineaire vergelijkingen een oplossing heeft?

$$\begin{array}{ll}
 \begin{array}{l}
 x_1 + 2x_2 - 3x_3 = a \\
 2x_1 + 6x_2 - 11x_3 = b \\
 x_1 - 2x_2 + 7x_3 = c
 \end{array} & \text{(i)} \\
 \begin{array}{l}
 x_1 + 2x_2 - 3x_3 = a \\
 3x_1 - x_2 + 2x_3 = b \\
 x_1 - 5x_2 + 8x_3 = c
 \end{array} & \text{(ii)}
 \end{array}$$

Opgave 4.

Ga na of de volgende stelsels van vectoren lineair afhankelijk of onafhankelijk zijn:

$$\begin{array}{ll}
 \text{(i)} \left(\begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right) & \text{(ii)} \left(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \right) \\
 \text{(iii)} \left(\begin{pmatrix} 1 \\ -2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 6 \\ -5 \\ -4 \end{pmatrix} \right) & \text{(iv)} \left(\begin{pmatrix} 0 \\ 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \right)
 \end{array}$$

Opgave 5.

Ga na of de volgende afbeeldingen lineair zijn en geef, zo ja, de matrix van de afbeelding met betrekking tot de standaardbases aan:

- (i) $f : \mathbb{R}^2 \rightarrow \mathbb{R} : f(x, y) := xy$
- (ii) $f : \mathbb{R} \rightarrow \mathbb{R}^2 : f(x) := (2x, 3x)$
- (iii) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 : f(x, y) := (2x - y, x)$
- (iv) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 : f(x, y) := (x + 1, 2y, x + y)$
- (v) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 : f(x, y) := (x, 2y, x + y)$
- (vi) $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 : f(x, y) := (x^2, 2y, x + y)$
- (vii) $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2 : f(x, y, z) := (z, x + y)$
- (viii) $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2 : f(x, y, z) := (x + 1, y + z)$

Opgave 6.

Bereken de volgende determinanten:

$$\det \begin{pmatrix} t-5 & 7 \\ -1 & t+3 \end{pmatrix}, \quad \det \begin{pmatrix} 2 & 0 & -1 \\ 3 & 0 & 2 \\ 4 & -3 & 7 \end{pmatrix}, \quad \det \begin{pmatrix} 3 & 2 & -4 \\ 1 & 0 & -2 \\ -2 & 3 & 3 \end{pmatrix},$$

$$\det \begin{pmatrix} 7 & 6 & 5 \\ 1 & 2 & 1 \\ 3 & -2 & 1 \end{pmatrix}, \quad \det \begin{pmatrix} t+3 & -1 & 1 \\ 5 & t-3 & 1 \\ 6 & -6 & t+4 \end{pmatrix}, \quad \det \begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix},$$

$$\det \begin{pmatrix} 2 & 5 & -3 & -2 \\ -2 & -3 & 2 & -5 \\ 1 & 3 & -2 & 2 \\ -1 & -6 & 4 & 3 \end{pmatrix}, \quad \det \begin{pmatrix} 3 & -2 & -5 & 4 \\ -5 & 2 & 8 & -5 \\ -2 & 4 & 7 & -3 \\ 2 & -3 & -5 & 8 \end{pmatrix}, \quad \det \begin{pmatrix} 1 & 2 & 2 & 3 \\ 1 & 0 & -2 & 0 \\ 3 & -1 & 1 & -2 \\ 4 & -3 & 0 & 2 \end{pmatrix}$$

Opgave 7.

Bepaal de eigenwaarden en eigenvectoren van de volgende matrices. Geef in elk geval aan of er een basis bestaat die alleen maar uit eigenvectoren bestaat.

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}, \quad D = \begin{pmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{pmatrix},$$

$$E = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 7 & -6 \\ -1 & 4 & 0 \\ 0 & 2 & -2 \end{pmatrix}, \quad G = \begin{pmatrix} 2 & -2 & 3 \\ 10 & -4 & 5 \\ 5 & -4 & 6 \end{pmatrix}$$

Opgave 8.

Bepaal een basis van de deelruimten die (met betrekking tot het standaardinproduct) orthogonaal op de volgende vectoren staan:

$$(i) \begin{pmatrix} 1 \\ -2 \\ 3 \\ 4 \end{pmatrix} \text{ en } \begin{pmatrix} 3 \\ -5 \\ 7 \\ 8 \end{pmatrix} \in \mathbb{R}^4 \quad (ii) \begin{pmatrix} 1 \\ 2 \\ 3 \\ -1 \\ 2 \end{pmatrix} \text{ en } \begin{pmatrix} 2 \\ 4 \\ 7 \\ 2 \\ -1 \end{pmatrix} \in \mathbb{R}^5$$

Opgave 9.

Laat zien dat de afbeelding $f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) := x_1x_2 - x_1y_2 - y_1x_2 + 3y_1y_2$ een inproduct op \mathbb{R}^2 definieert (d.w.z. laat zien dat f een positief definitieve, symmetrische bilineaire afbeelding is).

Opgave 10.

Laten $v_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ en $v_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ vectoren in \mathbb{R}^2 zijn.

(i) Voor welke $a \in \mathbb{R}$ definieert de afbeelding

$$f(v_1, v_2) := x_1x_2 - 3x_1y_2 - 3y_1x_2 + ay_1y_2$$

een inproduct op \mathbb{R}^2 ?

(ii) Voor welke $a, b, c, d \in \mathbb{R}$ definieert de afbeelding

$$f(v_1, v_2) := ax_1x_2 + bx_1y_2 + cy_1x_2 + dy_1y_2$$

een inproduct op \mathbb{R}^2 ?

Opgave 11.

Bepaal een orthonormaalbasis (met betrekking tot het standaardinproduct)

voor de deelvectorruimte $U \subseteq \mathbb{R}^4$ die de vectoren $\begin{pmatrix} -3 \\ -3 \\ 3 \\ 3 \end{pmatrix}$, $\begin{pmatrix} -5 \\ -5 \\ 7 \\ 7 \end{pmatrix}$ en $\begin{pmatrix} 4 \\ -2 \\ 0 \\ 6 \end{pmatrix}$

bevat.

Opgave 12.

Geef de orthogonale projectie (met betrekking tot het standaardinproduct) van

$v = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ in de richting van $w = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ aan.

Deel II

Calculus

Aanbevolen achtergrondliteratuur met veel opgaven (en oplossingen):

- Frank Ayres: (Schaum's Outline of Theory and Problems of) Calculus. McGraw-Hill Companies, 1999, 578 p., ISBN: 0070419736.
- Michael Spivak: Calculus. Addison Wesley World Student Series, 1967, 588 p., ISBN: 0805390235.

Les 6 Differentiatie van functies

Waarschijnlijk heeft iedereen wel een idee ervan wat een functie is, maar voor de duidelijkheid herhalen we voor de meest belangrijke begrippen de definities. We zullen ons beperken tot reële functies, dus functies die op (een deel van) de reële getallen gedefinieerd zijn.

6.1 Functies

Een *reële functie* is een voorschrift dat aan ieder element $x \in D$ van een deelverzameling $D \subseteq \mathbb{R}$ een *functiewaarde* $f(x) \in \mathbb{R}$ toewijst. We noemen D het *domein* van de functie f en de verzameling $\{f(x) \mid x \in D\}$ van alle functiewaarden het *bereik* van de functie f . Merk op dat we hier nog niets over de structuur van het domein D hebben geëist, dit kan inderdaad een willekeurige deelverzameling van \mathbb{R} zijn. Belangrijke voorbeelden zijn:

- (i) alle reële getallen, dus $D = \mathbb{R}$,
- (ii) alle reële getallen behalve één, bijvoorbeeld $\mathbb{R} \setminus \{0\}$,
- (iii) intervallen, hierbij onderscheiden we:
 - (a) afgesloten intervallen zoals $D = [0, 1] := \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$,
 - (b) open intervallen zoals $D = (0, 1) := \{x \in \mathbb{R} \mid 0 < x < 1\}$,
 - (c) halfopen intervallen zoals $D = (0, 1] := \{x \in \mathbb{R} \mid 0 < x \leq 1\}$,
- (iv) de vereniging van (afgesloten, open of halfopen) intervallen.

Een speciaal type van halfopen intervallen zijn intervallen die in een richting onbegrensd zijn. Een voorbeeld hiervoor zijn de niet-negatieve reële getallen $D = \{x \in \mathbb{R} \mid x \geq 0\}$. Deze worden vaak met $D = [0, \infty)$ genoteerd.

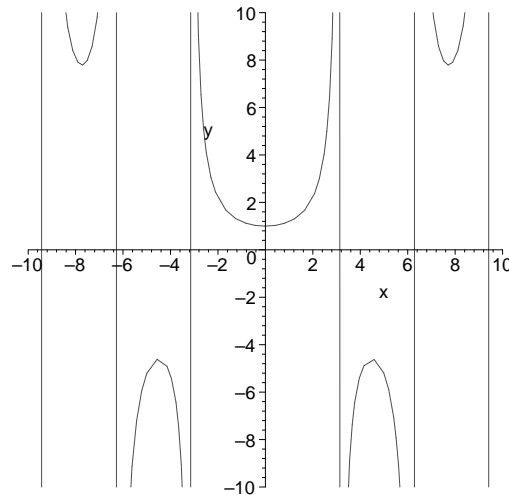
Er zijn echter ook rare domeinen mogelijk, bijvoorbeeld $D = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Dit is (net als de rationale getallen \mathbb{Q}) een deelverzameling van \mathbb{R} waar tussen elk paar van punten uit D overaftelbaar veel punten uit \mathbb{R} liggen die niet in D bevat zijn. Aan de andere kant liggen de punten van \mathbb{Q} willekeurig dicht bij elkaar, dus geldt dit ook voor D . Dit soort rare domeinen komen we in de praktijk zelden tegen, maar ze zijn in de wiskunde belangrijk om de fundamentele begrippen zuiver te definiëren.

Een functie f wordt beschreven door zijn domein D en een afbeeldingsvoorschrift, dat zegt hoe je aan een $x \in D$ zijn functiewaarde $f(x)$ toewijst. Dit wordt vaak geschreven als $f : D \rightarrow \mathbb{R}, x \mapsto y$ als y de waarde van f in het punt x is, dus als $y = f(x)$ geldt.

Hier zijn een paar voorbeelden van functies, die laten zien dat het soms om redelijk gekke dingen gaat:

- (i) $D = \mathbb{R}, f : x \mapsto x^2$,

- (ii) $D = \mathbb{R}$, $f(x) := \begin{cases} -1 & \text{als } x < 0 \\ 0 & \text{als } x = 0 \\ 1 & \text{als } x > 0 \end{cases}$ (dit heet soms ook de *signum*-functie),
- (iii) $D = \mathbb{R}$, $f(x) := \begin{cases} x & \text{als } x \in \mathbb{Q} \\ 0 & \text{als } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$
- (iv) $D = \mathbb{R} \setminus \{-1, 1\}$, $f : x \mapsto \frac{x^2 - 3x + 2}{x^2 - 1}$ (merk op dat voor $x \in D$ de noemer nooit nul wordt),
- (v) $D = \mathbb{R} \setminus \{n \cdot \pi \mid n \in \mathbb{Z}\}$, $f : x \mapsto \frac{x}{\sin(x)}$,



Figuur II.1: $f(x) := \frac{x}{\sin(x)}$

- (vi) $D = \mathbb{R}$, $f(x) := n$ als n het aantal van 7en in de decimale ontwikkeling van x is en $f(x) := \pi$ als er oneindig veel 7en in de decimale ontwikkeling zitten.

Om ingewikkeldere functies te bouwen, worden vaak eenvoudigere functies gecombineerd. Hierbij wordt behalve van het optellen, aftrekken, vermenigvuldigen en delen van functies ook de *samenstelling* van functies gebruikt. Voor een functie f met domein D_f en bereik B_f en een functie g met domein D_g waarvoor $D_g \supseteq B_f$ geldt, definiëren we de *samengestelde functie* $g \circ f$ met domein D_f door

$$g \circ f : D_f \rightarrow \mathbb{R}, x \mapsto g(f(x)).$$

Bijvoorbeeld kunnen we de functie $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{1+x^2}$ beschrijven als $f = h_1 \circ g_1$ met $h_1 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{x}$ en $g_1 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto 1+x^2$, maar ook door $f = h_2 \circ g_2$ met $h_2 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{1+x}$ en $g_2 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2$ en zelfs door $f = h_1 \circ g_3 \circ g_2$ met $g_3 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto 1+x$.

Merk op dat $h_2 = h_1 \circ g_3$ en $g_1 = g_3 \circ g_2$, dus is $(h_1 \circ g_3) \circ g_2 = h_1 \circ (g_3 \circ g_2)$. Omdat ook in het algemeen geldt dat de samenstelling van functies associatief

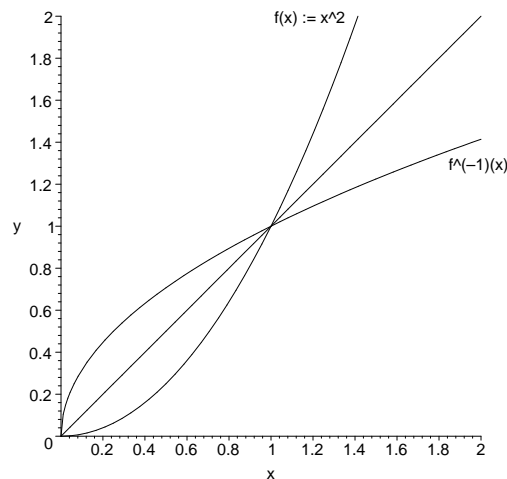
is, d.w.z. dat $(h \circ g) \circ f = h \circ (g \circ f)$, hoeven we in de samenstelling van drie of meer functies niet op haakjes te letten.

Om een vergelijking $f(x) = y$ naar x op te lossen, moeten we bij een functie f een punt x vinden die een gegeven waarde y oplevert. Hiervoor hebben we een soort van omkering van f nodig. Zo'n omkering kunnen we voor willekeurige y vinden, als er voor de functie f een functie g bestaat, zo dat $g \circ f(x) = x$ voor alle x in het domein van f . Dan geldt namelijk $x = g(f(x)) = g(y)$. Als zo'n functie g bestaat noemen we g de *inverse functie* van f en noteren deze als f^{-1} . Merk op dat het domein van f^{-1} het bereik van f is.

Merk op: De inverse functie f^{-1} kan alleen maar bestaan als f aan verschillende punten $x_1 \neq x_2$ ook verschillende functiewaarden $f(x_1) \neq f(x_2)$ toewijst. Voor $f(x_1) = f(x_2)$ is namelijk $x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2$, dus zou voor $f(x_1) = f(x_2)$ met $x_1 \neq x_2$ de inverse functie twee verschillende waarden aan $f(x_1)$ moeten toewijzen, en dat mag niet omdat een functie een *eenduidige* waarde aan een punt toewijst.

Een functie met $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$ heet een *injectieve functie*. We hebben dus gezien dat alleen maar injectieve functies een inverse functie hebben.

We kunnen de grafiek van de inverse functie gemakkelijk uit de grafiek van f afleiden: De grafiek van f bestaat uit de punten $(x, f(x))$ in het $x - y$ -vlak, en omdat $f^{-1}(f(x)) = x$, bestaat de grafiek van f^{-1} uit de punten $(f(x), x)$. Dit is dus de gespiegelde van de grafiek van f in de diagonaal $x = y$. We zien nu ook meteen in dat f injectief moet zijn, want anders is er een horizontale lijn $x = a$ die twee (of meer) snijpunten met de grafiek van f heeft, en dan heeft de verticale lijn $y = a$ twee snijpunten met de grafiek van f^{-1} en dit kan niet voor een functie.

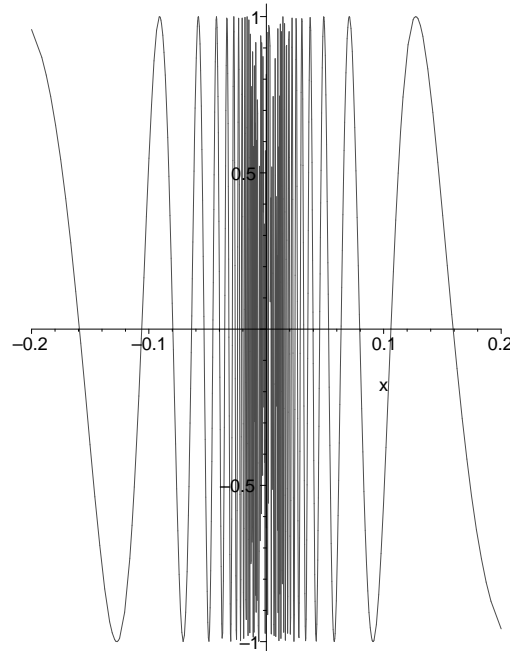


Figuur II.2: $f(x) := x^2$ heeft voor $x \geq 0$ de inverse functie $f^{-1}(x) := \sqrt{x}$

6.2 Continue functies

Een belangrijke klasse van functies zijn functies die we intuïtief *glad* zouden noemen, omdat we hun grafiek kunnen tekenen zonder de pen af te zetten. Dit wil zeggen, dat er geen *sprongen* in de grafiek zijn. Omdat wiskundigen soms heel rare functies verzinnen waarvoor we niet eens weten hoe we de grafiek zouden kunnen tekenen, hebben we een iets algemenere definitie van gladheid nodig.

Bijvoorbeeld schommelt de functie $f(x) := \sin(\frac{1}{x})$ in de buurt van $x = 0$ sneller en sneller tussen -1 en 1 en we kunnen niet zeggen wat voor een waarde we aan 0 moeten toewijzen.



Figuur II.3: $f(x) := \sin(\frac{1}{x})$ in de buurt van $x = 0$

We zeggen dat een functie $f : D \rightarrow \mathbb{R}$ *continu* in het punt $a \in D$ is, als de functiewaarden in een klein stukje om a heen allemaal dicht bij $f(a)$ liggen. Dit wordt door de beroemde $\varepsilon - \delta$ -definitie uitgedrukt:

II.1 Definitie Een functie $f : D \rightarrow \mathbb{R}$ heet continu in het punt $a \in D$ als er voor alle $\varepsilon > 0$ een $\delta > 0$ bestaat, zo dat uit $|x - a| < \delta$ volgt, dat $|f(x) - f(a)| < \varepsilon$ is.

Deze definitie betekent, dat er voor een in a continue functie f voor een willekeurig klein interval $(f(a) - \varepsilon, f(a) + \varepsilon)$ om de functiewaarde $f(a)$ een interval $I = (a - \delta, a + \delta)$ om a heen bestaat, zo dat alle functiewaarden $f(x)$ voor $x \in I$ in het gekozen ε -interval om $f(a)$ liggen.

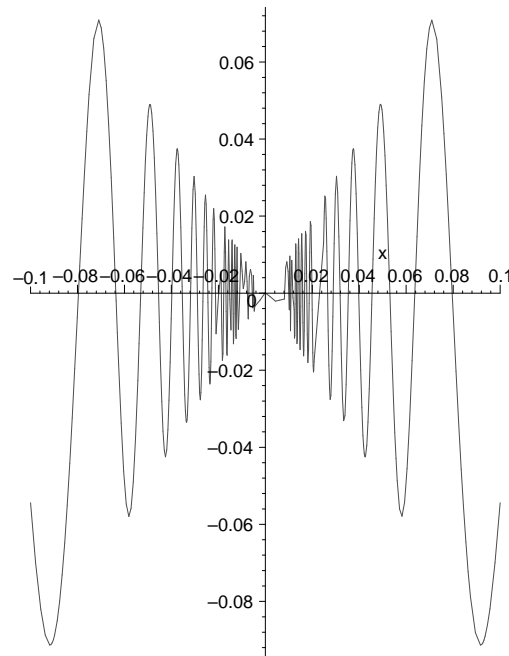
In het voorbeeld van de functie $f(x) := \sin(\frac{1}{x})$ neemt de functie tussen $-\frac{1}{n}$ en $\frac{1}{n}$ elke waarde tussen -1 en 1 aan. Voor elke mogelijke waarde die we voor $f(0)$ zouden kiezen, vinden we dus al voor $\varepsilon = \frac{1}{2}$ geen $\delta > 0$ zo dat alle

functiewaarden in het interval $(f(0) - \frac{1}{2}, f(0) + \frac{1}{2})$ liggen, want δ zou kleiner dan $\frac{1}{n}$ voor elke n moeten zijn. Dit betekent dat $\sin(\frac{1}{x})$ niet tot een in 0 continue functie voortgezet kan worden.

Aan de andere kant is de functie

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \begin{cases} x \cdot \sin(\frac{1}{x}) & \text{als } x \neq 0 \\ 0 & \text{als } x = 0 \end{cases} .$$

wel continu, want $|\sin(x)| \leq 1$, en dus is $|f(x) - f(0)| = |x \sin(\frac{1}{x})| \leq |x|$, dus kunnen we altijd $\delta = \varepsilon$ kiezen.



Figuur II.4: $f(x) := x \cdot \sin(\frac{1}{x})$

Omgekeerd zien we dat een functie in een punt, waar een sprong is, niet continu is. We kijken als voorbeeld naar de signum-functie

$$\text{sign} : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \begin{cases} -1 & \text{als } x < 0 \\ 0 & \text{als } x = 0 \\ 1 & \text{als } x > 0 \end{cases} .$$

Als we $\varepsilon = \frac{1}{2}$ kiezen, zien we dat $\text{sign}(x)$ niet continu in 0 is, want voor elke $\delta > 0$ is $\text{sign}(\frac{\delta}{2}) = 1$ en $\text{sign}(-\frac{\delta}{2}) = -1$, dus liggen de functiewaarden niet in het ε -interval om $f(0) = 0$.

We hebben nu gezien wat het betekent dat een functie continu in een punt is. We noemen een functie een *continue functie* als hij continu in ieder punt van zijn domein is.

In toepassingen hebben we het meestal met continue functies te maken, maar er zijn ook situaties waar we functies met sprongen nodig hebben. Een

voorbeeld hiervoor is het volgende experiment: we willen de intensiteit van het licht op een lijn beschrijven, waar in $x = 1$ een gloeilamp staat en in $x = 0$ een niet-transparante muur. De intensiteit van het licht in een afstand r van de lamp is $1/r^2$. Dus wordt de intensiteit beschreven door de functie

$$I : \mathbb{R} \setminus \{1\} \rightarrow \mathbb{R}, x \mapsto \begin{cases} 0 & \text{als } x \leq 0 \\ \frac{1}{(x-1)^2} & \text{als } x > 0 \end{cases} .$$

Voor de beschrijving van dit soort functies is de *Heaviside-functie* H handig: $H : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \begin{cases} 0 & \text{als } x \leq 0 \\ 1 & \text{als } x > 0 \end{cases} .$

We kunnen de functie I dan beschrijven als $I(x) = \frac{H(x)}{(x-1)^2}$ voor $x \neq 1$.

6.3 De afgeleide van een functie

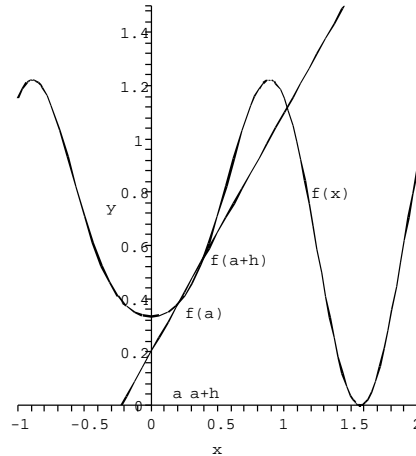
Het meest belangrijke bij een functie zijn natuurlijk de waarden, maar soms zijn we ook nog in andere dingen geïnteresseerd, bijvoorbeeld of een functie rond een gegeven punt afneemt of toeneemt. Bijvoorbeeld willen we weten of de temperatuur stijgend of dalend is en vragen we ons af of het heelal eeuwig expandeert of uiteindelijk weer in elkaar stort. (Eigenlijk willen we zelfs weten of de snelheid van de expansie afneemt of toeneemt.) Als we naar de grafiek van een functie kijken, kunnen we natuurlijk in de meeste gevallen meteen zien, wat er met stijgen en dalen aan de hand is. Maar soms is het voorschrift van een functie te ingewikkeld om er een grafiek van te maken, en soms zijn zelfs de grafieken zo complex, dat we niet kunnen zeggen of een functie stijgt of daalt. Daarom hebben we ook hier (net als bij de continue functies) een precieze definitie nodig.

Het idee is, dat we een functie in een punt a door een lijn benaderen, die de grafiek van de functie in het punt a raakt. Als deze raaklijn een positieve richtingscoëfficiënt heeft, noemen we de functie stijgend, als hij negatief is noemen we de functie dalend. Maar hoe vinden we de richtingscoëfficiënt van de raaklijn in een punt? Hiervoor kijken we naar de functiewaarden van f in de 'buurt' van a , dus we kijken naar $f(a+h)$ voor kleine waarden van h . Als we nu een koorde door de punten $(a, f(a))$ en $(a+h, f(a+h))$ trekken, heeft deze lijn de richtingscoëfficiënt $\frac{f(a+h)-f(a)}{h}$. Als we h kleiner laten worden, lijkt de koorde door $(a, f(a))$ en $(a+h, f(a+h))$ steeds meer op een raaklijn in het punt a , dus vinden we de richtingscoëfficiënt als de 'limiet' van $\frac{f(a+h)-f(a)}{h}$ als h naar 0 gaat.

We moeten nu eerst definiëren, wat het betekent dat een functie een *limiet* voor x gaat naar a heeft. Dit lijkt erg op de definitie van continue functies.

II.2 Definitie Een functie $f : D \rightarrow \mathbb{R}$ heeft voor $x \rightarrow a$ de limiet b , als er voor alle $\varepsilon > 0$ een $\delta > 0$ bestaat, zo dat uit $|x - a| < \delta$ volgt, dat $|f(x) - b| < \varepsilon$ is. We noteren dit als $\lim_{x \rightarrow a} f(x) = b$.

Merk op dat we niet eisen dat a in het domein van f ligt. Als de limiet $\lim_{x \rightarrow a} f(x)$ bestaat en de waarde b heeft, betekent dit, dat we door $f(a) := b$ een functie definiëren die continu in a is.



Figuur II.5: Benadering van de raaklijn in een punt door een koorde

We kunnen het nu over limieten van functies hebben, en we gebruiken deze nieuwe definitie meteen voor de definitie van de afgeleide.

II.3 Definitie Een functie $f : D \rightarrow \mathbb{R}$ heet in een punt $a \in D$ *differentieerbaar*, als de limiet

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

bestaat. In dit geval noteren we de limiet met $f'(a)$ en noemen dit de *afgeleide* van f in het punt a .

Met andere woorden heeft een functie f de afgeleide $f'(a)$ in het punt a , als de functie

$$\Delta : D \rightarrow \mathbb{R}, h \mapsto \begin{cases} \frac{f(a+h) - f(a)}{h} & \text{als } h \neq 0 \\ f'(a) & \text{als } h = 0 \end{cases}$$

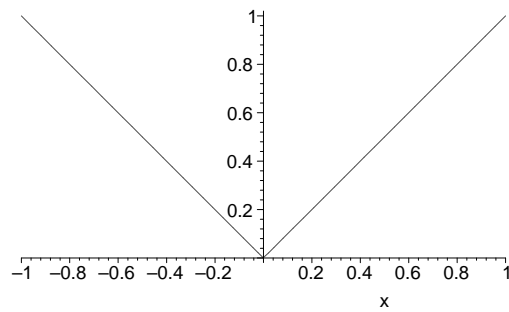
in het punt $x = 0$ continu is.

We noemen een functie f een *differentieerbare functie* als de afgeleide $f'(a)$ voor elke $a \in D$ bestaat. In dit geval heet de functie

$$f' : D \rightarrow \mathbb{R}, x \mapsto f'(x) \text{ de afgeleide functie van } f.$$

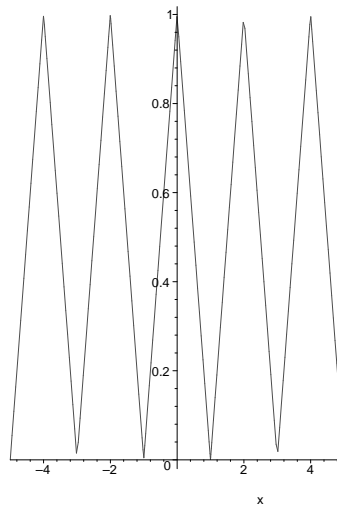
We hebben gezien dat continue functies geen sprongen hebben. Dit is niet voldoende, om een differentieerbare functie te hebben. Bijvoorbeeld is de functie $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto |x|$ continu, maar in het punt $x = 0$ niet differentieerbaar. Er geldt namelijk $\frac{f(a+h) - f(a)}{h} = 1$ voor $h > 0$ en $\frac{f(a+h) - f(a)}{h} = -1$ voor $h < 0$, dus bestaat $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ in dit geval niet. Het probleem ligt in het feit dat de absoluutwaarde-functie in het punt $x = 0$ een knik heeft. Differentieerbare functies zijn dus functies die geen knikken hebben.

Men zou misschien denken, dat continue functies, alleen maar een paar knikken kunnen hebben, maar in de meeste punten wel differentieerbaar



Figuur II.6: $f(x) := |x|$ is voor $x = 0$ niet differentieerbaar

zijn. Helaas is dit niet zo. Het is inderdaad mogelijk een functie aan te geven, die in elk punt continu, maar in geen punt differentieerbaar is. Zo'n functie bestaat dus alleen maar uit knikken.



Figuur II.7: Een zaag-functie

Het idee voor zo'n functie begint met een zaag-functie zo als in Figuur II.7. Vervolgens wordt elk lijnsegment tussen twee knikken in drie even grote delen onderverdeeld. Voor een lijnsegment met richtingscoëfficiënt c krijgt het eerste stuk de oude richtingscoëfficiënt c , het tweede de richtingscoëfficiënt $-c$ en het derde de richtingscoëfficiënt $3c$. Op deze manier komen er twee nieuwe knikken in elk recht lijnsegment. Als we deze constructie oneindig vaak herhalen, leidt dit tot een 'limiet-functie' die inderdaad continu maar in geen punt differentieerbaar is.

We komen nu terug op de vraag, of een functie stijgend of dalend is. In principe kunnen we dit ook voor functies definiëren, die niet continu zijn. Zij $f : [a, b] \rightarrow \mathbb{R}$ een functie:

- (i) f heet stijgend, als $x_2 > x_1 \Rightarrow f(x_2) \geq f(x_1)$.
- (ii) f heet strikt stijgend, als $x_2 > x_1 \Rightarrow f(x_2) > f(x_1)$.

- (iii) f heet dalend, als $x_2 > x_1 \Rightarrow f(x_2) \leq f(x_1)$.
- (iv) f heet strikt dalend, als $x_2 > x_1 \Rightarrow f(x_2) < f(x_1)$.

Voor differentieerbare functies kunnen we dit nu vertalen naar een criterium voor de afgeleide. Zij f op een interval $[a, b]$ differentieerbaar:

- (i) Is $f'(x) \geq 0$ voor alle $x \in [a, b]$ dan is f stijgend op $[a, b]$.
- (ii) Is $f'(x) > 0$ voor alle $x \in [a, b]$ dan is f strikt stijgend op $[a, b]$.
- (iii) Is $f'(x) \leq 0$ voor alle $x \in [a, b]$ dan is f dalend op $[a, b]$.
- (iv) Is $f'(x) < 0$ voor alle $x \in [a, b]$ dan is f strikt dalend op $[a, b]$.

6.4 Regels voor differentiatie

Als we een functie $f : D \rightarrow \mathbb{R}$ hebben dan is f vaak verkregen uit eenvoudigere functies, die we door optellen, aftrekken, vermenigvuldigen, delen en samenstellen combineren. Daarom zou het handig zijn om regels voor de afgeleide van dit soort combinaties te hebben.

We gaan er nu van uit, dat de functies die we bekijken in de aangegeven punten ook inderdaad differentieerbaar zijn.

- (0) Constante functies $f(x) = c$ hebben afgeleide 0, de functie $f(x) = x$ heeft de afgeleide 1.
- (1) Optellen:

$$(f + g)'(a) = f'(a) + g'(a).$$

Dit geldt omdat $(f + g)(a + h) - (f + g)(a) = (f(a + h) - f(a)) + (g(a + h) - g(a))$ is. Hierbij hebben we nog nodig, dat $\lim_{x \rightarrow a} (f + g)(x) = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x)$ als de twee limieten op de rechterzijde bestaan.

- (2) Vermenigvuldigen met een factor $c \in \mathbb{R}$:

$$(cf)'(a) = cf'(a).$$

Hier gebruiken we dat $(cf)(a + h) - (cf)(a) = c(f(a + h) - f(a))$ is.

Uit (1) en (2) volgt, dat de afbeelding $f \rightarrow f'$ een lineaire afbeelding op de vectorruimte van differentieerbare functies is.

- (3) Vermenigvuldigen (**productregel**):

$$(f \cdot g)'(a) = f'(a)g(a) + f(a)g'(a).$$

We hebben $f(a + h)g(a + h) - f(a)g(a) = f(a + h)g(a + h) - f(a)g(a + h) + f(a)g(a + h) - f(a)g(a) = (f(a + h) - f(a))g(a + h) + f(a)(g(a + h) - g(a))$, dus is $\frac{f(a+h)g(a+h) - f(a)g(a)}{h} = \frac{f(a+h) - f(a)}{h}g(a+h) + f(a)\frac{g(a+h) - g(a)}{h}$, en dus $\lim_{h \rightarrow 0} \frac{f(a+h)g(a+h) - f(a)g(a)}{h} = f'(a)g(a) + f(a)g'(a)$.

(4) Delen (*quotiëntenregel*):

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}.$$

Hiervoor laten we eerst zien dat $(\frac{1}{g})'(a) = -\frac{g'(a)}{g^2(a)}$. Uit (3) volgt dat $0 = (g\frac{1}{g})'(a) = g'(a)\frac{1}{g(a)} + g(a)(\frac{1}{g})'(a)$, dus is $g(a)(\frac{1}{g})'(a) = -\frac{g'(a)}{g(a)}$. We passen nu (3) op $f \cdot \frac{1}{g}$ toe, dan geldt: $(\frac{f}{g})'(a) = \frac{f'(a)}{g(a)} - \frac{f(a)g'(a)}{g^2(a)}$, en dit geeft de bewering.

(5) Samenstelling (*kettingregel*):

$$(g \circ f)'(a) = (g' \circ f)(a)f'(a) = g'(f(a))f'(a).$$

In plaats van een voetzame 0 die we bij de productregel gebruikt hebben, gebruiken we nu een voetzame 1: $\frac{(g \circ f)(a+h) - (g \circ f)(a)}{h} = \frac{g(f(a+h)) - g(f(a))}{h} = \frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} \cdot \frac{f(a+h) - f(a)}{h} = \frac{g(f(a) + (f(a+h) - f(a))) - g(f(a))}{f(a+h) - f(a)} \cdot \frac{f(a+h) - f(a)}{h}$. Als we $k := k(h) = f(a+h) - f(a)$ definiëren, gaat wegens de continuïteit van f voor $h \rightarrow 0$ ook $k \rightarrow 0$. Dus is $\lim_{h \rightarrow 0} \frac{(g \circ f)(a+h) - (g \circ f)(a)}{h} = \lim_{h \rightarrow 0} \frac{g(f(a)+k) - g(f(a))}{k} \cdot \frac{f(a+h) - f(a)}{h} = g'(f(a)) \cdot f'(a)$.

We hebben hier op een plek gesjoemeld, want $f(a+h) - f(a)$ kan ook voor $h \neq 0$ gelijk aan 0 zijn en dan mogen we hier niet door delen. Maar dit kunnen we herstellen, door in het geval dat $f(a+h) - f(a) = 0$ de quotiënt $\frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)}$ te vervangen door $g'(f(a))$. Het laat zich aantonen dat de zo gedefinieerde functie continu is en het argument gaat door.

Een slimme toepassing van de kettingregel geeft de afgeleide van de inverse functie $f^{-1}(x)$: Er geldt $(f \circ f^{-1})(x) = x$, dus volgt met de kettingregel dat $f'(f^{-1}(x)) \cdot f^{-1}'(x) = 1$ en dus geldt

$$f^{-1}'(x) = \frac{1}{f'(f^{-1}(x))}.$$

We laten nog een paar belangrijke voorbeelden van afgeleiden zien, die we eenvoudig kunnen bepalen.

Voor $n \in \mathbb{N}$ is de afgeleide van $f_n(x) = x^n$ de functie $f'_n(x) = nx^{n-1}$. Dit is duidelijk voor $n = 1$ en als we het voor een n hebben bewezen dan geldt $f'_{n+1}(x) = (f_n f_1)'(x) = f'_n(x)f_1(x) + f_n(x)f'_1(x) = nx^{n-1}x + x^n = (n+1)x^n$. Dus klopt het ook voor $n+1$ en dus per volledige inductie voor alle $n \in \mathbb{N}$.

Maar dezelfde formule geldt ook voor $f_{-n}(x) = x^{-n} = \frac{1}{x^n}$ voor $n \in \mathbb{N}$. We hebben $f'_{-n}(x) = (\frac{1}{f_n})'(x) = -\frac{nx^{n-1}}{x^{2n}} = (-n)x^{-n-1}$.

En dezelfde formule klopt zelfs voor $n = \frac{1}{2}$. Er geldt namelijk $\frac{\sqrt{a+h} - \sqrt{a}}{h} = \frac{1}{\sqrt{a+h} + \sqrt{a}} \frac{\sqrt{a+h}^2 - \sqrt{a}^2}{h} = \frac{1}{\sqrt{a+h} + \sqrt{a}}$, en dus is $\lim_{h \rightarrow 0} \frac{\sqrt{a+h} - \sqrt{a}}{h} = \frac{1}{2\sqrt{a}}$. Dus zien

we dat ook hier geldt, dat $f'_{\frac{1}{2}}(x) = \frac{1}{2}x^{\frac{1}{2}-1}$. Voor de wortelfunctie $f(x) := \sqrt{x}$ geldt dus:

$$f(x) = \sqrt{x} \Rightarrow f'(x) = \frac{1}{2\sqrt{x}}.$$

Het vermoeden ligt nu voor de hand, dat de formule voor de afgeleide van x^n ook voor algemene machtsfuncties $f(x) := x^c$ met $c \in \mathbb{R}$ klopt. Dit is inderdaad het geval, er geldt:

$$f(x) = x^c \Rightarrow f'(x) = cx^{c-1}.$$

Wij gaan dit hier alleen maar voor rationale machten $c = \frac{m}{n}$ aantonen. Hiervoor gebruiken we de formule voor de afgeleide van de inverse functie. We weten dat $f_n^{-1}(x) = x^{\frac{1}{n}} = \sqrt[n]{x}$ de inverse functie van $f_n(x) = x^n$ is, daarom geldt $f_n^{-1}'(x) = \frac{1}{n(x^{\frac{1}{n}})^{n-1}} = \frac{1}{n}x^{-\frac{n-1}{n}} = \frac{1}{n}x^{\frac{1}{n}-1}$.

We kunnen nu met de kettingregel de afgeleide van $f(x) = x^{\frac{m}{n}} = (x^m)^{\frac{1}{n}}$ berekenen als $f'(x) = \frac{1}{n}(x^m)^{\frac{1}{n}-1}mx^{m-1} = \frac{m}{n}x^{\frac{m}{n}-m}x^{m-1} = \frac{m}{n}x^{\frac{m}{n}-1}$.

Om van rationale machten $c = \frac{m}{n}$ naar algemene reële machten te komen, gebruikt men het feit dat een reëel getal willekeurig goed door een rationaal getal benaderd kan worden.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- functies, domein, bereik, inverse functie
- continuïteit
- afgeleide van een functie
- stijgen en dalen van functies
- productregel, quotiëntenregel, kettingregel
- afgeleide van de inverse functie

OPGAVEN

29. Bepaal de limiet

$$\lim_{y \rightarrow x} \frac{x^n - y^n}{x - y}$$

op twee manieren: rechtstreeks (met behulp van een staartdeling) en via afleiden.

30. Een functie f heet *even* als $f(x) = f(-x)$ voor alle x en *oneven* als $f(x) = -f(-x)$ voor alle x in het domein van f . De grafiek van een even functie is symmetrisch ten opzichte van een spiegeling in de y -as, de grafiek van een oneven functie is symmetrisch ten opzichte van een puntspiegeling in de oorsprong.

- (i) Laat zien dat voor een differentieerbare functie f , die even is, geldt dat $f'(x) = -f'(-x)$ (dus de afgeleide van een even functie is oneven).

(ii) Laat zien dat voor een differentieerbare functie f , die oneven is, geldt dat $f'(x) = f'(-x)$ (dus de afgeleide van een oneven functie is even).

31. Bepaal voor de volgende functies de intervallen waar de functies stijgen en waar ze dalen:

(i) $D = \mathbb{R}, f(x) = -x^2,$

(ii) $D = \mathbb{R}, f(x) = (x - 3)^2,$

(iii) $D = [-1, 1], f(x) = \sqrt{1 - x^2},$

(iv) $D = [4, \infty), f(x) = \sqrt{x - 4},$

(v) $D = \mathbb{R}, f(x) = x^5 + 20x - 6.$

32. Zij $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto 7 + \sqrt[5]{3x - 2}$. Bepaal de inverse functie van f .

33. Bepaal de afgeleiden van:

(i) $f_1(x) := \sqrt{1 + \sqrt{x}},$ (ii) $f_2(x) := 3x^{\frac{1}{2}} - x^{\frac{3}{2}} + 2x^{-\frac{1}{2}},$ (iii) $f_3(x) := \sqrt[3]{x^5 + 7x},$

(iv) $f_4(x) := 2x^2 \sqrt{x^2 - 2x + 2},$ (v) $f_5(x) := \left(\frac{x^3 - 1}{2x^3 + 1} \right)^4.$

Les 7 Speciale functies

We hebben in de vorige les een aantal elementaire functies bekeken en hiervoor gezien hoe we deze functies kunnen afleiden. In wezen waren alle deze functies samengesteld uit machtsfuncties $f(x) := x^c$ met $c \in \mathbb{R}$. In deze les hebben we het over verschillende andere elementaire functies die een belangrijke rol in alle soorten van toepassingen spelen.

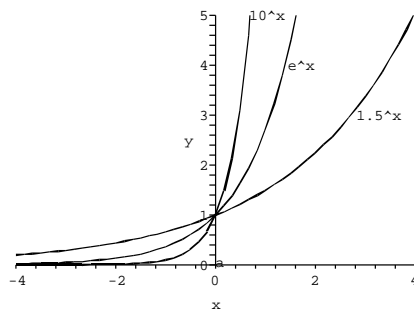
7.1 Exponentiële functie en natuurlijke logaritme

Als we de ontwikkeling van een populatie beschrijven, hebben we het vaak met de volgende situatie te maken: Er is een beginpopulatie van C konijnen en elk jaar verdubbelt het aantal konijnen. Dan zijn er na afloop van één jaar $2C$ konijnen, na twee jaar $4C$, na drie jaar $8C$ enzovoorts. Na afloop van x jaar zijn er dan $2^x C$ konijnen.

Het zou dus handig zijn, iets meer over functies als $f(x) := a^x$ voor $a \in \mathbb{R}$ te weten.

Om te beginnen moeten we iets erover zeggen hoe we de functiewaarden van zo'n functie berekenen. Voor breuken $x = \frac{m}{n}$ kunnen we a^x wel berekenen, dit is gewoon $\sqrt[n]{a^m}$. Hier zien we dat $a > 0$ moet zijn, anders zouden we (voor oneven m) uit negatieve getallen de wortel moeten trekken.

Omdat we graag willen dat $f(x) := a^x$ een continue functie wordt, hebben we nu geen keuze meer bij de berekening van a^x voor een willkeurige $x \in \mathbb{R}$. Als we x namelijk door breuken $\frac{m}{n}$ steeds beter benaderen, moet $\sqrt[n]{a^m}$ een steeds betere benadering van de functiewaarde a^x zijn (dat is juist de definitie van continuïteit).



Figuur II.8: De functies $f(x) := a^x$ voor $a = 1.5, e, 10$

Zoals we dat uit de grafieken in Figuur II.8 zouden verwachten, laat zich aantonen dat de functie $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto a^x$ voor $a > 0$ in het punt $x = 0$ differentieerbaar is. Er laat zich ook algemeen bewijzen dat de afgeleide $f'(0)$

groter is naarmate a groter is. Maar als we de afgeleide van a^x in 0 kennen, kunnen we de afgeleide van a^x in elk punt berekenen, want

$$\frac{a^{x+h} - a^x}{h} = a^x \cdot \frac{a^h - 1}{h} \text{ en dus } f'(x) = a^x \cdot f'(0).$$

Als we nu voor verschillende waarden van a de afgeleide van $f(x) := a^x$ in het punt $x = 0$ berekenen, kunnen we door een benaderingsproces een waarde voor a vinden, zo dat $f'(0) = 1$ is. Op die manier vinden we het *Euler-getal* $e \approx 2.71828$ met de eigenschap dat de functie $f(x) := e^x$ in 0 de afgeleide 1 heeft.

Zo als boven opgemerkt volgt uit het feit dat de afgeleide van $f(x) := e^x$ voor $x = 0$ gelijk aan 1 is, dat $f'(x) = e^x \cdot f'(0) = e^x \cdot 1 = f(x)$. Dit betekent dat $f(x) := e^x$ een functie is die aan de vergelijking $f(x) = f'(x)$ voldoet. De functie e^x heet de *exponentiële functie* en wordt vaak ook met $f(x) := \exp(x)$ genoteerd. Samenvattend hebben we dus:

$$\exp'(x) = \exp(x) \quad \text{en} \quad \exp(0) = 1.$$

Er laat zich zelf aantonen dat de exponentiële functie door de eigenschappen $f'(x) = f(x)$ en $f(0) = 1$ eenduidig bepaald is:

Neem aan dat $f(x)$ een functie is met $f'(x) = f(x)$ en $f(0) = 1$, dan bepalen we de afgeleide van de functie $g(x) := \frac{f(x)}{\exp(x)}$. Hiervoor geldt

$$g'(x) = \frac{f'(x)\exp(x) - f(x)\exp'(x)}{\exp(x)^2} = \frac{f(x)\exp(x) - f(x)\exp(x)}{\exp(x)^2} = 0$$

omdat $\exp'(x) = \exp(x)$ en $f'(x) = f(x)$. Maar hieruit volgt dat $g(x)$ een constante functie is, dus is $f(x) = c \cdot \exp(x)$ en uit $f(0) = \exp(0) = 1$ volgt $c = 1$, dus $f(x) = \exp(x)$.

De exponentiële functie speelt in veel toepassingen een rol, bijvoorbeeld (zo als eerder al gezegd) bij de ontwikkeling van populaties of in de beschrijving van radioactief verval. Maar ook bij het remmen van een auto of bij het verloop van de temperatuur tussen twee kamers met verschillende temperaturen is de functie $\exp(x)$ van toepassing.

We weten (uit ervaring) dat we met evenveel remkracht niet zo snel van 100 naar 80 km per uur kunnen afremmen als van 50 naar 30. De verandering van de snelheid bij het remmen is dus afhankelijk van de snelheid zelfs. Ook bij de temperatuur zien we een soortgelijk effect: als we een kamer van 0° naast een kamer van 50° hebben, zullen de temperaturen sneller veranderen dan bij kamers van 20° en 30° . Bij veel processen vinden we dus een afhankelijkheid tussen de snelheid van de verandering van de functie en de waarde van de functie, d.w.z. een afhankelijkheid van de vorm

$$f'(x) = C \cdot f(x),$$

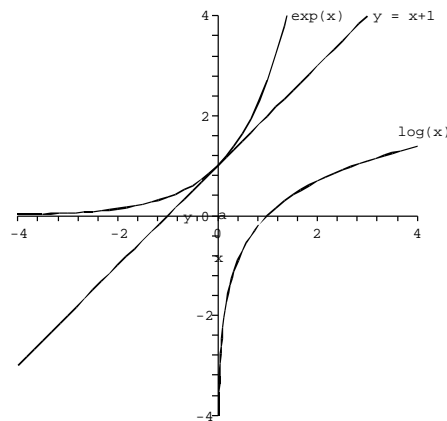
waarbij C een constante is. Er laat zich aantonen dat alle functies die aan deze vergelijking voldoen van de vorm

$$f(x) := x_0 \cdot \exp(Cx)$$

zijn, waarbij x_0 door de randvoorwaarde $x_0 = f(0)$ bepaald is (bijvoorbeeld de temperatuur of positie op het tijdstip $x = 0$).

Algemeen noemt men vergelijkingen tussen een functie $f(x)$ en hun afgeleiden $f'(x)$, $f''(x)$ enz. een *differentiaalvergelijking*.

Uit het feit dat $e > 1$ volgt dat $\exp(x) > 0$ voor alle x en $\exp(x) > 1$ voor alle $x > 0$, daarom is $\exp(y) - \exp(x) = (\exp(y - x) - 1)\exp(x) > 0$ voor $y > x$. Dit toont aan dat $\exp(x)$ een op \mathbb{R} strikt stijgende functie is. Het bereik is $(0, \infty)$, dus kunnen we op het open interval $(0, \infty)$ de inverse functie van $\exp(x)$ definiëren. Deze noemen we de (*natuurlijke*) *logaritme* en noteren deze met $\log(x)$.



Figuur II.9: Exponentiële functie en natuurlijke logaritme

Merk op: De omkeersfunctie van de algemene functie $f(x) := a^x$ heet de *logaritme met basis a* en wordt met ${}^a \log(x)$ genoteerd. Soms (bijvoorbeeld op de middelbare school of bij ingenieurs) wordt met $\log(x)$ de logaritme met basis 10 bedoeld, de natuurlijke logaritme wordt dan met $\ln(x)$ aangegeven. In deze cursus gaan we echter $\log(x)$ steeds voor de logaritme met basis e gebruiken, een andere basis wordt altijd expliciet aangegeven (bijvoorbeeld ${}^{10} \log(x)$ en ${}^2 \log(x)$ voor de logaritmes met basis 10 en 2).

Ook zakrekenmachines kunnen tot verwarring leiden: Vaak is **ln** de toets voor de natuurlijke logaritme terwijl de toets **log** voor de logaritme met basis 10 staat.

We kunnen logaritmes voor verschillende bases makkelijk omrekenen, want er geldt:

$${}^a \log(x) = \frac{\log(x)}{\log(a)}.$$

We hebben namelijk $e^{\log(x)} = x = a^{{}^a \log(x)} = (e^{\log(a)})^{{}^a \log(x)} = e^{\log(a) \cdot {}^a \log(x)}$ en dus $\log(x) = \log(a) \cdot {}^a \log(x)$.

Uit onze formule voor de afgeleide van de inverse functie kunnen we de afgeleide van $\log(x)$ makkelijk berekenen, er geldt

$$\log'(x) = \frac{1}{\exp'(\log(x))} = \frac{1}{\exp(\log(x))} = \frac{1}{x}.$$

We hebben hiermee een belangrijk gat gevuld: We hadden in de vorige les gezien dat we voor een geheel getal $n \in \mathbb{Z}$ de afgeleide van $f(x) := x^n$ vinden als $f'(x) = n \cdot x^{n-1}$. In het bijzonder vinden we elke van de functies x^n als afgeleide van een andere machtsfunctie, namelijk als afgeleide van $\frac{1}{n+1}x^{n+1}$. De enige uitzondering hierbij is het geval $n = -1$, want de afgeleide van x^0 is natuurlijk 0. Maar nu hebben we een functie gevonden, die $x^{-1} = \frac{1}{x}$ als afgeleide heeft, namelijk de natuurlijke logaritme $\log(x)$.

Om de algemene exponentiële functie $f(x) := a^x$ af te leiden is het handig om de relatie $a = e^{\log(a)}$ en dus $a^x = e^{\log(a) \cdot x} = \exp(\log(a) \cdot x)$ te gebruiken. Met de kettingregel volgt dan namelijk dat

$$(a^x)' = \exp(\log(a)x) \cdot \log(a) = \log(a) \cdot a^x.$$

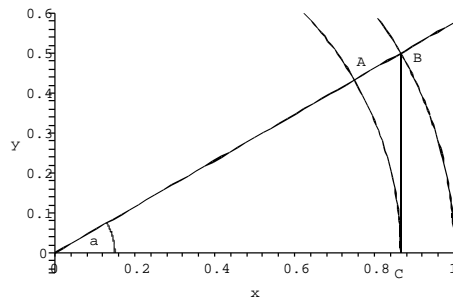
Tenslotte nog twee belangrijke relaties voor het optellen en vermenigvuldigen bij exp en log:

$$\exp(x + y) = \exp(x) \exp(y) \quad \text{en} \quad \log(xy) = \log(x) + \log(y).$$

7.2 Trigonometrische functies

De trigonometrische (of goniometrische) functies zijn gebaseerd op de meetkunde van rechthoekige driehoeken.

Als in een rechthoekige driehoek de schuine zijde lengte 1 heeft, en a één van de niet-rechte hoeken is, dan noemen we de lengte van de zijde tegenover a de *sinus van a* , genoteerd met $\sin(a)$ en de lengte van de andere rechthoekszijde de *cosinus van a* , genoteerd met $\cos(a)$.

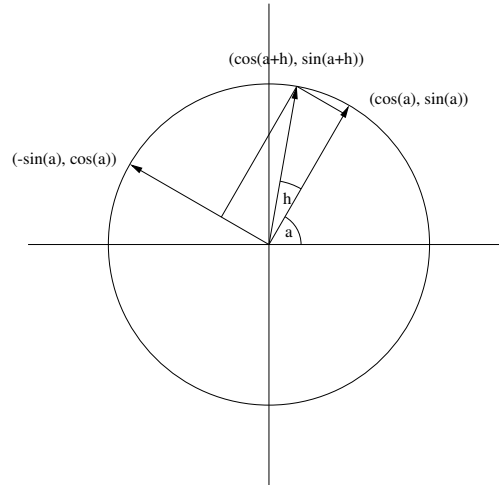


Figuur II.10: $\sin(a) = |BC|$, $\cos(a) = |OC|$

In het plaatje van Figuur II.10 is OB de schuine zijde in de driehoek $OB C$ en we hebben $\sin(a) = |BC|$ en $\cos(a) = |OC|$. Een van de belangrijkste relaties voor sinus en cosinus volgt meteen uit de stelling van Pythagoras, namelijk

$$\sin^2(x) + \cos^2(x) = 1.$$

Om de afgeleide van $\sin(x)$ te bepalen moeten we iets over de quotiënt $\frac{\sin(a+h)-\sin(a)}{h}$ zeggen. Maar hoe kunnen we de sinus van een som van twee hoeken bepalen? Hiervoor geeft het volgende plaatje een aanleiding.



Figuur II.11: De sinus van de som van twee hoeken

We weten (uit Lineaire Algebra) dat we de vector $w = \begin{pmatrix} \cos(a+h) \\ \sin(a+h) \end{pmatrix}$ kunnen schrijven als de som van zijn orthogonale projecties op de orthogonale basisvectoren $v_1 = \begin{pmatrix} \cos(a) \\ \sin(a) \end{pmatrix}$ en $v_2 = \begin{pmatrix} -\sin(a) \\ \cos(a) \end{pmatrix}$. Maar de lengte van de projectie van w in de richting van v_1 is $\cos(h)$ en de lengte van de projectie in de richting van v_2 is $\sin(h)$. Dus geldt:

$$\begin{pmatrix} \cos(a+h) \\ \sin(a+h) \end{pmatrix} = \cos(h) v_1 + \sin(h) v_2 = \begin{pmatrix} \cos(a) \cos(h) - \sin(a) \sin(h) \\ \sin(a) \cos(h) + \cos(a) \sin(h) \end{pmatrix}.$$

Dit geeft de twee belangrijke opteltheorema's:

$$\cos(a+h) = \cos(a) \cos(h) - \sin(a) \sin(h),$$

$$\sin(a+h) = \sin(a) \cos(h) + \cos(a) \sin(h).$$

We hebben dus $\sin(a+h) - \sin(a) = \sin(a) \cos(h) + \cos(a) \sin(h) - \sin(a) = \sin(a)(\cos(h) - 1) + \cos(a) \sin(h)$ en hieruit volgt dat

$$\frac{\sin(a+h) - \sin(a)}{h} = \sin(a) \frac{\cos(h) - 1}{h} + \cos(a) \frac{\sin(h)}{h}.$$

We weten dat $\lim_{h \rightarrow 0} \sin(h) = 0$ en $\lim_{h \rightarrow 0} \cos(h) = 1$, maar dit is nog niet voldoende om de limiet van $\frac{\sin(a+h)-\sin(a)}{h}$ te berekenen.

Merk op: Vaak worden hoeken niet in graden maar in *radialen* aangegeven. Het idee hierbij is, een hoek door de lengte van de bijhorende cirkelboog in een cirkel van straal 1 te beschrijven. Een hoek van 360° heeft een volle cirkel als boog en die heeft lengte 2π . Omgekeerd hoort een boog van π bij

een hoek van 180° . Dus komen we van graden naar radialen door de hoek in graden met $\frac{\pi}{180}$ te vermenigvuldigen en van radialen naar graden door met $\frac{180}{\pi}$ te vermenigvuldigen. We zullen hoeken meestal in radialen aangeven.

We kunnen nu ook de lengte van een algemene cirkelboog aangeven, dat is namelijk $r \cdot a$, als r de straal van de cirkel is en a de bij de boog horende hoek in radialen. In Figuur II.10 heeft dus de boog van B naar 1 lengte a en de boog van A naar C lengte $a \cos(a)$. Omdat de boog AC korter is dan de lijn BC geldt $a \cos(a) < \sin(a)$ en omdat de lijn BC korter is dan de boog $B1$ geldt $\sin(a) < a$. Hieruit volgt (voor hoeken a met $0 \leq a \leq \frac{\pi}{2}$) dat

$$\cos(a) < \frac{\sin(a)}{a} < 1.$$

Omdat $\lim_{h \rightarrow 0} \cos(h) = 1$, volgt hieruit rechtstreeks dat

$$\lim_{h \rightarrow 0} \frac{\sin(h)}{h} = 1.$$

Verder is

$$\begin{aligned} \frac{\cos(h) - 1}{h} &= \frac{(\cos(h) - 1)(\cos(h) + 1)}{h(\cos(h) + 1)} = \frac{\cos^2(h) - 1}{h(\cos(h) + 1)} = \frac{-\sin^2(h)}{h(\cos(h) + 1)} \\ &= \frac{\sin(h)}{h} \cdot \frac{-\sin(h)}{\cos(h) + 1} \end{aligned}$$

en omdat $\frac{-\sin(h)}{\cos(h)+1}$ voor $h \rightarrow 0$ naar $\frac{0}{2} = 0$ gaat, volgt hieruit

$$\lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} = 1 \cdot 0 = 0.$$

Als we alles bij elkaar nemen volgt dus

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\sin(a+h) - \sin(a)}{h} &= \lim_{h \rightarrow 0} \sin(a) \frac{\cos(h) - 1}{h} + \lim_{h \rightarrow 0} \cos(a) \frac{\sin(h)}{h} \\ &= \sin(a) \cdot 0 + \cos(a) \cdot 1 = \cos(a). \end{aligned}$$

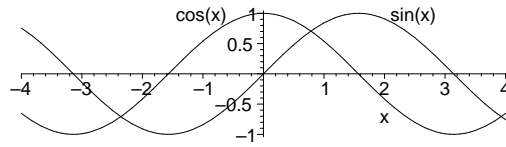
Kort en goed: de afgeleide van de sinus is de cosinus, ofwel

$$\sin'(x) = \cos(x).$$

We kunnen de afgeleide van de cosinus nu op dezelfde manier bepalen, maar met een klein trucje gaat het sneller. We weten dat $\sin^2(x) + \cos^2(x) = 1$ is, dus geldt $0 = (\sin^2(x) + \cos^2(x))' = 2\cos(x)\cos'(x) + 2\sin(x)\sin'(x) = 2\cos(x)(\cos'(x) + \sin(x))$. Hieruit volgt meteen:

$$\cos'(x) = -\sin(x).$$

Net zo als we de exponentiële functie $\exp(x)$ door de differentiaalvergelijking $f'(x) = f(x)$ hebben gekarakteriseerd, kunnen we ook sinus en cosinus door een differentiaalvergelijking karakteriseren. Het is duidelijk dat voor de



Figuur II.12: Sinus- en cosinus-functie

tweede afgeleiden geldt dat $\sin''(x) = -\sin(x)$ en $\cos''(x) = -\cos(x)$. Differentiaalvergelijkingen van de vorm $f''(x) = C \cdot f(x)$ spelen bijvoorbeeld bij de beschrijving van trillingen een belangrijke rol.

De bewering is nu, dat een functie $f(x)$ met $f''(x) + f(x) = 0$ een lineaire combinatie van $\sin(x)$ en $\cos(x)$ is, preciezer gezegd:

$$f''(x) + f(x) = 0 \Rightarrow f(x) = a \cdot \sin(x) + b \cdot \cos(x) \text{ met } a = f'(0), b = f(0).$$

Neem eerst aan we hebben een functie $f(x)$ met $f''(x) + f(x) = 0$, $f(0) = 0$ en $f'(0) = 0$. Dan is $0 = f'(x)(f''(x) + f(x)) = \frac{1}{2}(f'(x)^2 + f(x)^2)'$, dus is $f'(x)^2 + f(x)^2$ een constante functie. Maar omdat $f'(0) = f(0) = 0$, is $f'(x)^2 + f(x)^2 = 0$ voor alle x . Maar een som van kwadraten is alleen maar 0 als alle kwadraten 0 zijn, dus volgt hieruit dat $f(x) = 0$ voor alle x , dus is $f(x)$ de constante 0-functie.

Neem nu aan dat $f''(x) + f(x) = 0$, $f'(0) = a$ en $f(0) = b$. Dan geldt voor $g(x) := f(x) - a \sin(x) - b \cos(x)$ dat $g''(x) + g(x) = 0$, $g'(0) = f'(0) - a = 0$ en $g(0) = f(0) - b = 0$. Dus is $g(x) = 0$ en dus $f(x) = a \sin(x) + b \cos(x)$.

Uit de functies $\sin(x)$ en $\cos(x)$ wordt een aantal verdere functies afgeleid, de belangrijkste hiervan is de *tangens* die gedefinieerd is door

$$\tan(x) := \frac{\sin(x)}{\cos(x)}.$$

Het domein van de tangens zijn de punten $x \in \mathbb{R}$ met $\cos(x) \neq 0$, dus $x \neq \frac{\pi}{2} + n\pi$ met $n \in \mathbb{Z}$.

Voor $\tan(x)$ geldt de relatie

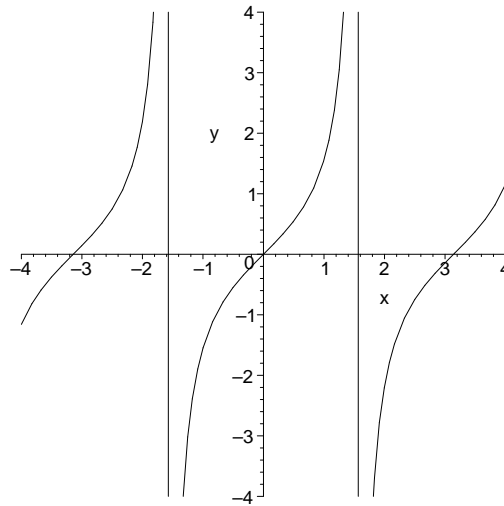
$$1 + \tan^2(x) = \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}.$$

Toevallig geeft dit juist ook de afgeleide van de tangens, want

$$\tan'(x) = \left(\frac{\sin(x)}{\cos(x)}\right)' = \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)} = \frac{1}{\cos^2(x)}.$$

We hebben dus:

$$\tan'(x) = 1 + \tan^2(x) = \frac{1}{\cos^2(x)}.$$



Figuur II.13: Tangens-functie

Inverse functies van de trigonometrische functies

De inverse functies van de trigonometrische functies heten *arcus*-functies en worden als $\arcsin(x) := \sin^{-1}(x)$, $\arccos(x) := \cos^{-1}(x)$ en $\arctan(x) := \tan^{-1}(x)$ genoteerd. De afgeleiden van deze functies kunnen we makkelijk met de formule

$$f^{-1}'(x) = \frac{1}{f'(f^{-1}(x))}$$

bepalen.

Het bereik van $\sin(x)$ is het interval $[-1, 1]$ dus heeft $\arcsin(x)$ dit interval als domein. Met behulp van het trucje $\cos(x) = \sqrt{1 - \sin^2(x)}$ vinden we:

$$\arcsin'(x) = \frac{1}{\cos(\arcsin(x))} = \frac{1}{\sqrt{1 - \sin^2(\arcsin(x))}} = \frac{1}{\sqrt{1 - x^2}}.$$

Het domein voor $\arccos(x)$ is ook $[-1, 1]$ en met behulp van $\sin(x) = \sqrt{1 - \cos^2(x)}$ tonen we aan dat

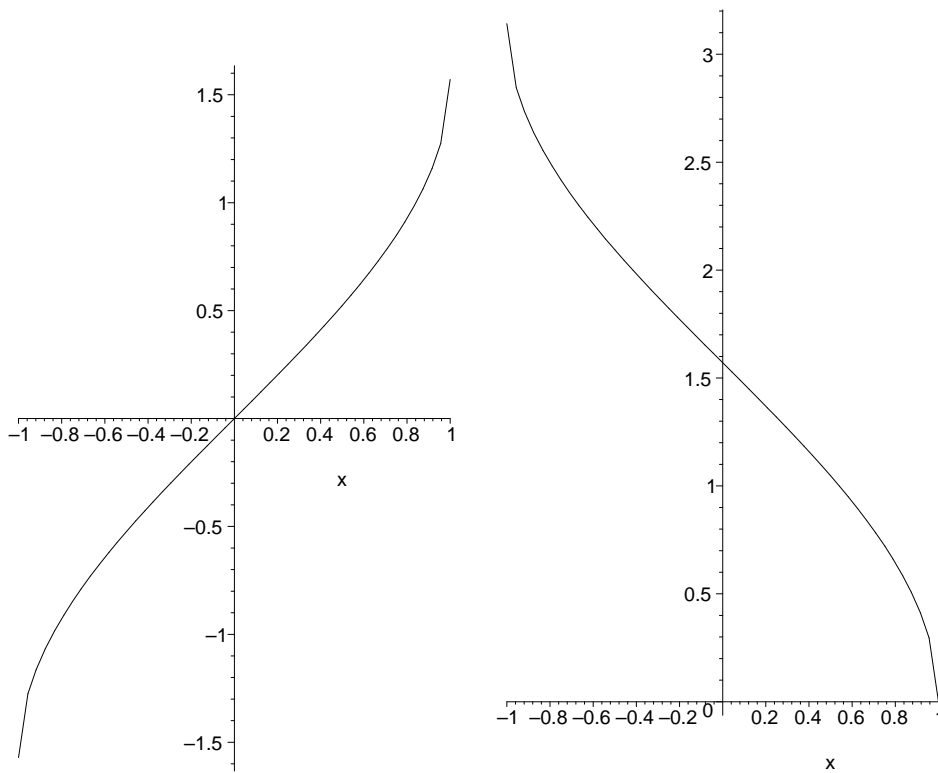
$$\arccos'(x) = \frac{1}{-\sin(\arccos(x))} = \frac{1}{-\sqrt{1 - \cos^2(\arccos(x))}} = \frac{1}{-\sqrt{1 - x^2}}.$$

We hebben dus:

$$\arcsin'(x) = \frac{1}{\sqrt{1 - x^2}}, \quad \arccos'(x) = \frac{1}{-\sqrt{1 - x^2}}.$$

De meest belangrijke toepassing van de arcussinus en de arcuscosinus ligt in de integratie van functies. We zullen zien dat de integratie de omkering van de differentiatie is, dus hebben we de functie $\arcsin(x)$ nodig om integralen over functies zo als $f(x) := \frac{1}{\sqrt{1-x^2}}$ te berekenen.

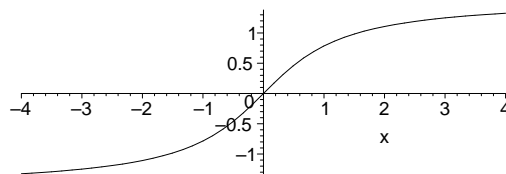
Het bereik van $\tan(x)$ is \mathbb{R} , maar de functie is alleen maar injectief op een interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ (of een verschuiving hiervan om $n\pi$). De arcustangens-functie



Figuur II.14: Arcussinus- en arcuscosinus-functie

is dus op \mathbb{R} gedefinieerd en heeft waarden tussen $-\frac{\pi}{2}$ en $\frac{\pi}{2}$. Voor de afgeleide vinden we met de formule voor de afgeleide van de inverse functie en de relatie $\cos^2(x) = \frac{1}{1+\tan^2(x)}$:

$$\begin{aligned} \arctan'(x) &= \frac{1}{\left(\frac{1}{\cos^2(\arctan(x))}\right)} = \cos^2(\arctan(x)) = \frac{1}{1 + \tan^2(\arctan(x))} \\ &= \frac{1}{1 + x^2}. \end{aligned}$$



Figuur II.15: Arcustangens-functie

De arcustangens-functie wordt (naast zogeheten *sigmoid-functies*) vaak gebruikt om experimentele waarden naar een genormeerd interval af te beelden. Bijvoorbeeld wil men als waarden, die een zoekmachine voor de kwaliteit van een zoekresultaat aangeeft, meestal waarden tussen 0 en 1 (of tussen 0% en 100%). Maar de intern in een zoekmachine gebruikte methode levert vaak

waarden die niet eens naar beneden of boven begrensd zijn. Dan is het handig om deze waarden af te beelden met de functie

$$f : \mathbb{R} \rightarrow [0, 1], x \mapsto \frac{1}{\pi}(\arctan(x) + \frac{\pi}{2})$$

die strikt stijgend is en als bereik het interval $[0, 1]$ heeft.

7.3 Hyperbolische functies

Een verdere klasse van belangrijke functies zijn de hyperbolische functies. Deze zijn afgeleid van de exponentiële functie, maar hebben eigenschappen die op eigenschappen van $\sin(x)$ en $\cos(x)$ lijken. We definiëren de *sinushyperbolicus* en *cosinushyperbolicus* door

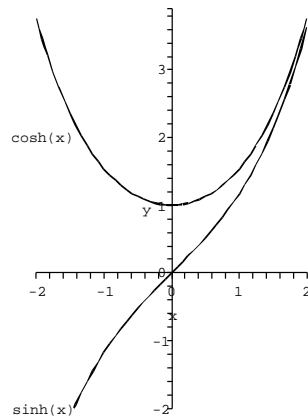
$$\sinh(x) := \frac{1}{2}(\exp(x) - \exp(-x)), \quad \cosh(x) := \frac{1}{2}(\exp(x) + \exp(-x)).$$

Met behulp van $\exp'(x) = \exp(x)$ gaat men eenvoudig na dat

$$\sinh'(x) = \cosh(x) \text{ en } \cosh'(x) = \sinh(x).$$

Verder vinden we dat

$$\cosh^2(x) - \sinh^2(x) = 1.$$



Figuur II.16: Sinushyperbolicus en cosinushyperbolicus

De naam van de hyperbolische functies heeft betrekking tot de *hyperbolische meetkunde*. Terwijl we in de Euclidische meetkunde afstanden in het vlak door $\sqrt{x^2 + y^2}$ berekenen, wordt dit in de hyperbolische meetkunde met $\sqrt{x^2 - y^2}$ gedaan. In de Euclidische meetkunde liggen de punten met afstand r van het nulpunt op een cirkel die we met $r(\cos(t), \sin(t))$, $0 \leq t \leq 2\pi$ kunnen aangeven. Een analoge constructie levert in de hyperbolische meetkunde de punten $r(\cosh(t), \sinh(t))$,

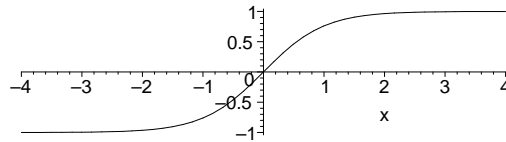
die op een hyperbool liggen (dus de naam). Een van de belangrijkste toepassingen van de hyperbolische meetkunde is de ruimtetijd uit de speciale relativiteitstheorie.

Analoog met de tangens-functie wordt ook een *tangenshyperbolicus* gedefinieerd:

$$\tanh(x) := \frac{\sinh(x)}{\cosh(x)}.$$

We hebben $1 - \tanh^2(x) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} = \frac{1}{\cosh^2(x)}$ en voor de afgeleide geldt $\tanh'(x) = \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} = \frac{1}{\cosh^2(x)}$, dus

$$\tanh'(x) = 1 - \tanh^2(x) = \frac{1}{\cosh^2(x)}.$$



Figuur II.17: Tangenshyperbolicus

Merk op dat ook de functie $\tanh(x)$ net als $\arctan(x)$ voor het normaliseren van experimentele waarden gebruikt kan worden.

Inverse functies van de hyperbolische functies

Ook de hyperbolische functies hebben inverse functies, deze heten de *area*-functies en worden met $\operatorname{arsinh}(x) := \sinh^{-1}(x)$, $\operatorname{arcosh}(x) := \cosh^{-1}(x)$ en $\operatorname{artanh}(x) := \tanh^{-1}(x)$ genoteerd.

We kunnen deze inverse functies expliciet bepalen, want uit $y = \sinh(x) = \frac{1}{2}(\exp(x) - \exp(-x))$ volgt door vermenigvuldiging met $\exp(x)$ dat

$$\exp(x)^2 - 2y \exp(x) - 1 = 0.$$

Dit geeft de oplossingen $\exp(x) = y \pm \sqrt{y^2 + 1}$, maar wegens $\exp(x) > 0$ is alleen maar het plusteken mogelijk. Het domein van $\operatorname{arsinh}(x)$ is \mathbb{R} omdat dit het bereik van $\sinh(x)$ is. Dus geldt voor $x \in \mathbb{R}$:

$$\operatorname{arsinh}(x) = \log(x + \sqrt{x^2 + 1}).$$

Voor de afgeleide vinden we met behulp van $\cosh(x) = \sqrt{1 + \sinh^2(x)}$

$$\operatorname{arsinh}'(x) = \frac{1}{\cosh(\operatorname{arsinh}(x))} = \frac{1}{\sqrt{1 + \sinh^2(\operatorname{arsinh}(x))}} = \frac{1}{\sqrt{1 + x^2}}.$$

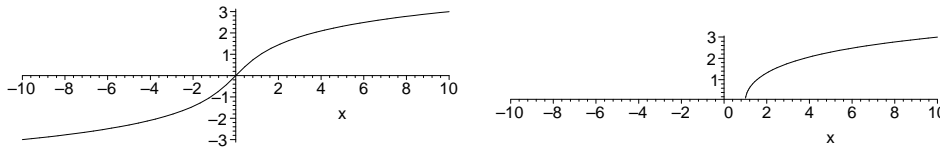
Het trucje van $\sinh(x)$ toegepast op $\cosh(x)$ geeft $\exp(x)^2 - 2y \exp(x) + 1 = 0$, dus $\exp(x) = y \pm \sqrt{y^2 - 1}$. In dit geval moeten we erop letten, dat $\cosh(x)$ niet

injectief is, we kunnen dus of een inverse functie voor $x > 0$ of voor $x < 0$ aangeven. Voor de inverse functie van $\cosh(x)$ met $x > 0$ geldt het plusteken, dus is

$$\operatorname{arcosh}(x) = \log(x + \sqrt{x^2 - 1}).$$

De afgeleide van $\operatorname{arcosh}(x)$ vinden we net als voor $\operatorname{arsinh}(x)$, maar deze keer gebruiken we de relatie $\sinh(x) = \sqrt{\cosh^2(x) - 1}$:

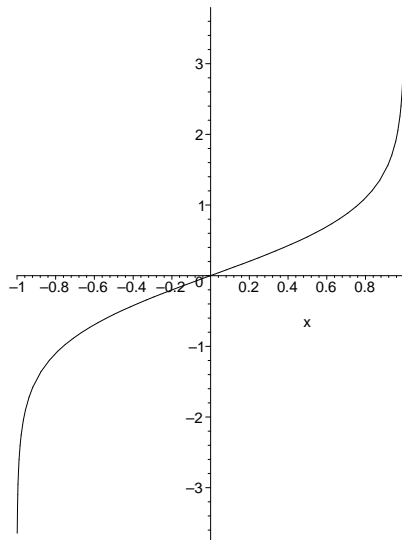
$$\operatorname{arcosh}'(x) = \frac{1}{\sinh(\operatorname{arcosh}(x))} = \frac{1}{\sqrt{\cosh^2(\operatorname{arcosh}(x)) - 1}} = \frac{1}{\sqrt{x^2 - 1}}.$$



Figuur II.18: Areasinushyperbolicus en areacosinushyperbolicus

Tenslotte kijken we naar de inverse functie van de *tangenshyperbolicus*, de *areatangenshyperbolicus* $\operatorname{artanh}(x)$. Uit $y = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$ volgt $1 + y = \frac{2 \exp(x)}{\exp(x) + \exp(-x)}$ en $1 - y = \frac{2 \exp(-x)}{\exp(x) + \exp(-x)}$, dus geldt $1 + y = \exp(2x)(1 - y) = \exp(x)^2(1 - y)$ en dus $\exp(x) = \sqrt{\frac{1+y}{1-y}}$. Hieruit volgt

$$\operatorname{artanh}(x) = \log\left(\sqrt{\frac{1+x}{1-x}}\right) = \frac{1}{2} \log\left(\frac{1+x}{1-x}\right).$$



Figuur II.19: Areatangenshyperbolicus

De afgeleide van $\operatorname{artanh}(x)$ vinden we met behulp van $\cosh^2(x) = \frac{1}{1-\tanh^2(x)}$ door $\operatorname{artanh}'(x) = \frac{1}{\tanh'(\operatorname{artanh}(x))} = \cosh^2(\operatorname{artanh}(x)) = \frac{1}{1-\tanh^2(\operatorname{artanh}(x))} = \frac{1}{1-x^2}$, dus is

$$\operatorname{artanh}'(x) = \frac{1}{1-x^2}.$$

Ook in dit geval is het belangrijkste argument om de functie $\operatorname{artanh}(x)$ te behandelen, dat we hiermee de integraal over functies zo als $\frac{1}{1-x^2}$ kunnen oplossen.

Deze les wordt samengevat door een tabel met de behandelde functies en hun afgeleiden.

$f(x)$	$f'(x)$
$\exp(x)$	$\exp(x)$
$\log(x)$	$\frac{1}{x}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$\frac{1}{\cos^2(x)}$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$-\frac{1}{\sqrt{1-x^2}}$
$\arctan(x)$	$\frac{1}{1+x^2}$
$\sinh(x)$	$\cosh(x)$
$\cosh(x)$	$\sinh(x)$
$\tanh(x)$	$\frac{1}{\cosh^2(x)}$
$\operatorname{arsinh}(x)$	$\frac{1}{\sqrt{1+x^2}}$
$\operatorname{arcosh}(x)$	$\frac{1}{\sqrt{x^2-1}}$
$\operatorname{artanh}(x)$	$\frac{1}{1-x^2}$

BELANGRIJKE BEGRIPPEN IN DEZE LES

- exponentiële functie, logaritme
- trigonometrische functies
- inverse trigonometrische functies
- hyperbolische functies
- inverse hyperbolische functies

OPGAVEN

34. Laten $f, g : \mathbb{R} \rightarrow \mathbb{R}$ de functies zijn met $f(x) := \log(x^2 + 1)$ en $g(x) := \exp(3x)$. Bereken de samengestelde functies $f \circ g$ en $g \circ f$ en de afgeleiden $f'(x)$, $g'(x)$, $(f \circ g)'(x)$ en $(g \circ f)'(x)$.
35. Toon aan dat voor alle $x \in (0, \infty)$ geldt dat $\log(x) \leq 2\sqrt{x} - 2$.
36. Laat zien dat $\sin x + \tan x > 2x$ voor alle $x \in (0, \pi/2)$.
(Hint: Differentiëren.)
37. Definieer $f : \mathbb{R} \rightarrow \mathbb{R}$ door $f(x) := x + \sin x + \arctan(3x)$.
Toon aan dat f een inverse functie met domein \mathbb{R} bezit. Daarvoor moet je bewijzen dat f strikt stijgend of dalend is en het geheel van \mathbb{R} als bereik heeft.
38. Bepaal de afgeleiden van:
- (i) $f_1(x) = x^x$, (ii) $f_2(x) = x^{x \sin(x)}$, (iii) $f_3(x) = \log(\cosh(x) + \sinh(x))$,
- (iv) $f_4(x) = \sin\left(\frac{x^3}{\cos(x^3)}\right)$, (v) $f_5(x) = \exp(-x^2)$, (vi) $f_6(x) = x \exp(\arctan(x))$,
- (vii) $f_7(x) = 5^{\cos(x)}$, (viii) $f_8(x) = \log\left(\sqrt{\frac{1-x^2}{1+x^2}}\right)$, (ix) $f_9(x) = \arcsin\left(\frac{1-x}{1+x}\right)$.
39. Bereken voor $f(x) := \frac{1}{1+x}$ de functies $g(x) := f(f'(x))$ en $h(x) := f'(f(x))$.

Les 8 Minima en maxima van functies

Een reden waarom we de afgeleide van een functie bekijken is dat we iets over het stijgen of dalen van de functie willen weten. Als we met een differentieerbare functie te maken hebben, is de functie stijgend als de afgeleide positief is en dalend als de afgeleide negatief is. Maar soms zijn we ook geïnteresseerd in de verandering van het stijgen en dalen van een functie. Bijvoorbeeld gaat het er in de economie vaak niet om of een bedrijf een groei in de omzet heeft, maar alleen maar of de groei toeneemt of afneemt. De groei wordt beschreven door de afgeleide van de omzet-functie, de verandering van de groei door de afgeleide van de groei-functie, dus door de tweede afgeleide van de omzet-functie.

We kijken dus voor een functie $f(x)$ niet alleen maar naar de eerste afgeleide $f'(x)$ maar ook naar de tweede afgeleide $f''(x) := (f'(x))'$ en ook naar hogere afgeleiden $f'''(x)$ enz. Omdat het onhandig wordt, meer en meer streepjes te schrijven, gebruiken we een nieuwe notatie: We schrijven $f^{(3)}(x)$ voor $f'''(x)$ en in het algemeen $f^{(n)}(x)$ voor de n -de afgeleide.

Merk op: Hogere afgeleiden hoeven niet altijd te bestaan, zelfs als de functie $f(x)$ differentieerbaar is. Bijvoorbeeld heeft de functie

$$f(x) := \begin{cases} -x^2 & \text{als } x < 0 \\ x^2 & \text{als } x \geq 0 \end{cases}$$

de afgeleide $f'(x) = 2|x|$ en is dus een differentieerbare functie, maar de tweede afgeleide bestaat voor $x = 0$ niet.

Men zegt soms dat een functie bijvoorbeeld *drievoudig differentieerbaar* is, om aan te duiden dat de derde afgeleide $f'''(x)$ bestaat. De 'beste' functies (d.w.z. de meest gladde functies) zijn functies die willekeurig vaak differentieerbaar zijn. Hierbij is het toegestaan dat de n -de afgeleiden $f^{(n)}(x)$ vanaf een zekere n alle 0 zijn, bijvoorbeeld bij veeltermfuncties van graad $n - 1$.

Een belangrijke toepassing van de hogere afgeleiden is de karakterisering van minima en maxima van functies. Dit zullen we in deze les uitvoerig bekijken.

8.1 Minima en maxima van gewone functies

In veel toepassingen komt een probleem er op neer een waarde x te bepalen zo dat een functie $f(x)$ een minimale (of maximale) waarde aanneemt. Bijvoorbeeld wordt er in de productie van blikken naar gekeken, een zo klein mogelijk hoeveelheid materiaal voor een gegeven volume te gebruiken. In de economie hangt de vraag naar een object natuurlijk ook van de prijs af. Die wordt dus zo gekozen dat het product van de prijs en het (verwachte) aantal verkochte objecten maximaal wordt.

De definities van (absolute) minima en maxima van functies in hun domein is heel voor de hand liggend. Een functie $f : D \rightarrow \mathbb{R}$ heeft in a een *absoluut minimum* als $f(a) \leq f(x)$ voor alle $x \in D$. Evenzo heeft een functie een *absoluut maximum* als $f(a) \geq f(x)$ voor alle $x \in D$.

Soms is het ook interessant om naar lokale minima en maxima te kijken. Dat zijn punten $a \in D$ zo dat f voor een klein interval om a heen een absoluut minimum/maximum in a heeft. Preciezer zeggen we: Een functie f heeft in $a \in D$ een *lokaal* (of relatief) minimum/maximum als er een $\delta > 0$ bestaat zo dat $f(a) \leq f(x)$ ($f(a) \geq f(x)$) voor alle $x \in (a - \delta, a + \delta)$.

Voor willekeurige functies f kunnen we niet veel verder dan deze definities, maar als f een differentieerbare functie is, zegt de afgeleide inderdaad iets over minima en maxima.

Stel dat f in a een lokaal minimum heeft en differentieerbaar in a is. Dan geldt voor kleine $h < 0$ dat $f(a + h) \geq f(a)$ en dus is $\frac{f(a+h)-f(a)}{h} \leq 0$. Maar ook voor kleine $h > 0$ is $f(a + h) \geq f(a)$ en dus $\frac{f(a+h)-f(a)}{h} \geq 0$. Omdat de afgeleide $f'(a)$ alleen maar bestaat als de limiet $\lim_{h \rightarrow 0} \frac{f(a+h)-f(a)}{h}$ voor $h < 0$ en $h > 0$ bestaat en dezelfde waarde heeft, is noodzakelijk $f'(a) = 0$.

Dezelfde redenering geeft ook voor een lokaal maximum dat $f'(a) = 0$ is. We hebben dus gezien:

II.4 Stelling Voor een functie $f : D \rightarrow \mathbb{R}$ die in $a \in D$ differentieerbaar is en een lokaal minimum/maximum in a heeft, geldt $f'(a) = 0$.

Helaas geldt de omkering van deze stelling niet: Bijvoorbeeld is voor $f(x) = x^3$ de afgeleide $f'(x) = 3x^2$ en dus is $f'(0) = 0$, maar 0 is geen lokaal minimum of maximum, omdat $f(x) < 0$ voor $x < 0$ en $f(x) > 0$ voor $x > 0$.

Maar tenminste kunnen we met behulp van deze stelling vaak een lijstje van kandidaten x maken, waar een functie mogelijk een minimum/maximum heeft, namelijk de punten x waarvoor $f'(x) = 0$ is. Dit noemen we ook de *kritieke punten* van $f(x)$. Als er punten zijn waar $f(x)$ niet differentieerbaar is, tellen we deze ook bij de kritieke punten. Bijvoorbeeld is $f(x) = |x|$ in $x = 0$ niet differentieerbaar, maar zijn (absoluut) minimum zit in $x = 0$. Als een functie op een interval gedefinieerd is horen ook nog de randwaarden van het interval bij de kritieke punten.

Hoe kunnen we nu aflezen of een functie in een punt nu echt een lokaal minimum of maximum heeft? Voor een lokaal minimum weten we al dat $f'(a) = 0$ moet zijn. Verder weten we dat $f(x)$ dalend is als $f'(x) \leq 0$ en stijgend als $f'(x) \geq 0$. Als we dus zien dat $f'(x) \leq 0$ voor $x < a$ en $f'(x) \geq 0$ voor $x > a$, heeft $f(x)$ inderdaad in a een lokaal minimum. Omgekeerd heeft $f(x)$ een lokaal maximum in a als $f'(a) = 0$, $f'(x) \geq 0$ voor $x < a$ en $f'(x) \leq 0$ voor $x > a$. Kort gezegd heeft een functie $f(x)$ een lokaal minimum of maximum in het punt a als het teken van de afgeleide $f'(x)$ in a verandert. Merk op dat we de ongelijkheden $f'(x) \geq 0$ en $f'(x) \leq 0$ slechts in een klein interval rond a hoeven te bekijken.

Om een lokaal minimum te identificeren, kunnen we soms ook de tweede afgeleide gebruiken: Als $f'(a) = 0$ en $f''(x) > 0$ voor $x \in (a - \delta, a + \delta)$, dan is $f'(x)$ een op dit interval strikt stijgende functie en dus is $f'(x) < 0$ voor $x < a$ en $f'(x) > 0$ voor $x > a$. Net zo vinden we een lokaal maximum door $f'(a) = 0$ en $f''(x) < 0$ op $(a - \delta, a + \delta)$. Dit houden we in de volgende stelling vast:

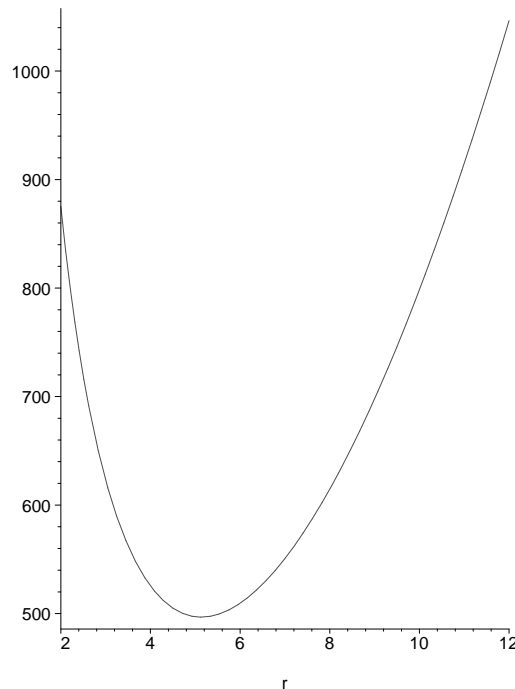
II.5 Stelling Een differentieerbare functie $f(x)$ heeft een lokaal minimum in a als er een $\delta > 0$ bestaat zo dat $f'(a) = 0$, $f'(x) \leq 0$ voor $x \in (a - \delta, a)$ en $f'(x) \geq 0$ voor $x \in (a, a + \delta)$. Dit geldt in het bijzonder als $f'(a) = 0$ en $f''(x) > 0$ voor $x \in (a - \delta, a + \delta)$.

Een differentieerbare functie $f(x)$ heeft een lokaal maximum in a als er een $\delta > 0$ bestaat zo dat $f'(a) = 0$, $f'(x) \geq 0$ voor $x \in (a - \delta, a)$ en $f'(x) \leq 0$ voor $x \in (a, a + \delta)$. Dit geldt in het bijzonder als $f'(a) = 0$ en $f''(x) < 0$ voor $x \in (a - \delta, a + \delta)$.

Voorbeeld 1

We gaan na hoe we de vorm van een blik moeten kiezen zo dat het gebruikte materiaal (de oppervlakte) minimaal wordt. Een blik nemen we aan als een cilinder van hoogte h met als grondvlak een cirkel van straal r . Dan is het volume V van de cilinder gegeven door $V = \pi r^2 h$ en de oppervlakte O door $O = 2\pi r^2 + 2\pi r h = 2\pi r(r + h)$. Bij een gegeven volume willen we nu de minimale oppervlakte vinden. Uit de vergelijking voor het volume vinden we $h = \frac{V}{\pi} r^{-2}$, dus kunnen we O schrijven als een functie van r door

$$O(r) = 2\pi r(r + \frac{V}{\pi} r^{-2}) = 2\pi r^2 + 2Vr^{-1}.$$



Figuur II.20: Oppervlakte van een blik van 850ml afhankelijk van de straal van het grondvlak

Voor de afgeleide van $O(r)$ geldt $O'(r) = 4\pi r - 2Vr^{-2}$ en we hebben $O'(r) = 0$ dan en slechts dan als $4\pi r = 2Vr^{-2}$, dus voor $2\pi r^3 = V = \pi r^2 h$. Hieruit volgt $2r = h$, dus de straal van de cirkel is half zo groot als de hoogte van de blik.

Voor een blik van 850ml vinden we dus $r = \sqrt[3]{\frac{V}{2\pi}} \approx 5.13\text{cm}$ en $h \approx 10.27\text{cm}$. Dit zijn inderdaad ongeveer de afmetingen van een standaardblik van dit volume.

Voorbeeld 2

De kosten van een auto zijn bepaald door de kosten van de benzine en de vaste kosten die alleen maar afhankelijk zijn van de tijd die de auto rijdt. Neem aan dat de kosten voor de benzine met het kwadraat van de snelheid stijgen. Wat is dan de optimale snelheid om een gegeven afstand zo goedkoop mogelijk te rijden? Hiervoor moeten we de kosten per gereden km bepalen. Als in de tijd t de afstand s met snelheid v gereden wordt, dan geldt $v = \frac{s}{t}$. De kosten voor de benzine op een afstand s zijn dus $k_b = cv^2t = cv^2\frac{s}{v} = csv$ (waarbij c de benzineprijs aangeeft) en de vaste kosten voor dezelfde afstand zijn $k_v = dt = d\frac{s}{v}$ (voor een constante d). De totale kosten afhankelijk van de snelheid v zijn dus $k(v) = csv + dsv^{-1}$ en de afgeleide hiervan is $k'(v) = cs - dsv^{-2}$. We hebben $k'(v) = 0$ voor $v^2 = \frac{d}{c}$, dus is de meest economische snelheid $v = \sqrt{\frac{d}{c}}$. Dit kunnen we ook kwalitatief bevestigen, want als de vaste kosten relatief hoger worden, is het goedkoper om sneller te rijden.

Voorbeeld 3

We hebben n punten a_1, \dots, a_n op de x -as gegeven en willen een punt x bepalen zo dat de som $f(x) := (x - a_1)^2 + \dots + (x - a_n)^2$ van de kwadratische afstanden van x van de gegeven punten minimaal wordt. Voor de afgeleide $f'(x)$ geldt $f'(x) = 2(x - a_1) + \dots + 2(x - a_n) = 2nx - 2(a_1 + \dots + a_n)$. We hebben dus $f'(x) = 0$ als $x = \frac{1}{n}(a_1 + \dots + a_n)$, dus als x het rekenkundig gemiddelde van de a_i is.

8.2 Functies van meerdere variabelen en de partiële afgeleide

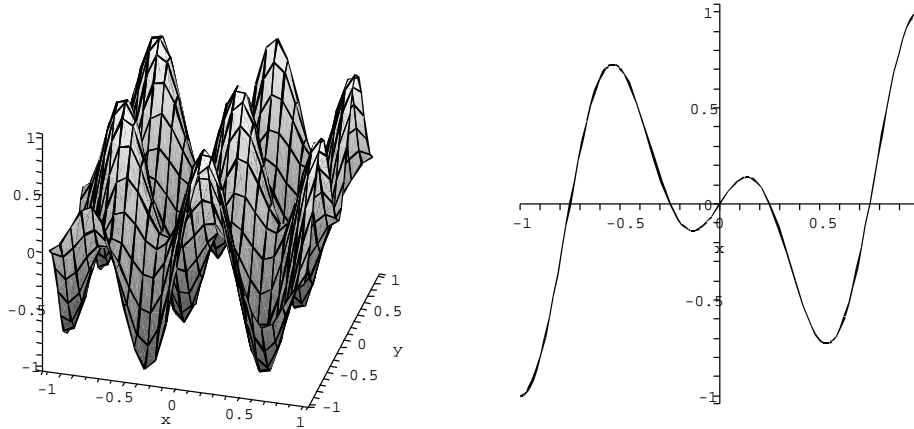
Tot nu toe hebben we alleen maar naar functies gekeken die van één variabele afhangen. Maar in de praktijk komen we vaak functies tegen die van een aantal parameters afhangen. Bijvoorbeeld is de afstand van de oorsprong van een punt (x, y, z) in de 3-dimensionale ruimte gegeven door de functie $f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ die drie variabelen heeft.

Het is nu niet meer onmiddellijk duidelijk, hoe we een afgeleide van zo'n functie moeten definiëren. Het zou handig zijn als de afgeleide weer de richtingscoëfficiënt van de raaklijn aan de grafiek van de functie is, maar het probleem is dat er in één punt raaklijnen voor elke mogelijke richting zijn. We moeten dus ook de richting aangeven in die we de raaklijn aan de grafiek willen leggen.

Het belangrijkste zijn de afgeleiden in de richtingen van de coördinatenassen. Deze vinden we door alle variabelen tot op één na als constant op te vatten. Dan is de functie slechts nog een functie van de overgebleven variabele en die kunnen we met de gewone regels afleiden.

Voor een functie van twee variabelen kunnen we dit ook grafisch interpreteren: De grafiek van zo'n functie kunnen we zien als de verzameling van punten $(x, y, f(x, y))$ in de 3-dimensionale ruimte, net zo als we de grafiek van een gewone functie als de verzameling van punten $(x, f(x))$ in het 2-dimensionale

vlak bekijken. Als we nu y tot een constante y_0 verklaren, dan kijken we naar de doorsnede van de grafiek $(x, y, f(x, y))$ met het vlak dat bepaald is door de vergelijking $y = y_0$, dus de punten (x, y_0, z) . Dit is een gewone grafiek van een functie in één variabeel zo als in de plaatjes in Figuur II.21 te zien.



Figuur II.21: Functie $f(x, y)$ en snede door deze functie voor $y = 0.5$

Maar voor zo'n doorsnede kunnen we natuurlijk zeggen wat de afgeleide zou zijn, namelijk de richtingscoëfficiënt van de raaklijn aan deze gewone grafiek. Deze afgeleide noemen we de *partiële afgeleide* naar x . Als we nu terug gaan naar de definitie van de gewone afgeleide, zien we dat de partiële afgeleide naar x in dit geval gegeven is door de limiet

$$\lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h}.$$

Omdat we voor de partiële afgeleide naar x alleen maar naar de verandering van $f(x, y)$ in de richting van x kijken, noemen we dit ook de *richtingsafgeleide* in de richting van x .

Analoog definiëren we nu de partiële afgeleide van een algemene functie $f(x_1, \dots, x_n)$ van n variabelen. Voor de partiële afgeleide naar x_i behandelen we de variabelen $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ als constanten (d.w.z. net zo als de constanten 5 of π) en interpreteren f zodanig als een functie van één variabeel (namelijk x_i). Als we voor de zo geïnterpreteerde functie de definitie van de gewone afgeleide toepassen krijgen we voor de partiële afgeleide naar x_i :

$$\frac{\partial f}{\partial x_i} := \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}$$

als deze limiet bestaat. Vaak wordt de partiële afgeleide $\frac{\partial f}{\partial x_i}$ ook kort als f_{x_i} geschreven.

Natuurlijk gebruiken we nooit deze limiet-definitie om een partiële afgeleide te berekenen, maar passen de gewone regels voor functies van één variabeel toe (waarbij we gewoon een paar constanten meer in de functie hebben).

Voorbeeld: Zij $f(x, y, z)$ de functie van drie variabelen, gegeven door $f(x, y, z) := x \log(yz)$, dan geldt:

$$\frac{\partial f}{\partial x} = \log(yz), \quad \frac{\partial f}{\partial y} = x(yz)^{-1}z = \frac{x}{y}, \quad \frac{\partial f}{\partial z} = x(yz)^{-1}y = \frac{x}{z}.$$

Merk op dat we ook partiële afgeleide kunnen itereren, dus we kunnen $\frac{\partial f}{\partial x} = \log(yz)$ weer partieel afleiden. Als we dit bijvoorbeeld partieel naar y afleiden, schrijven we dit als

$$\frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial y \partial x} = (yz)^{-1}z = y^{-1}.$$

Omgekeerd kunnen we ook $\frac{\partial f}{\partial y}$ partieel naar x afleiden, dit geeft $\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial x \partial y} = y^{-1}$. We zien dus dat het geen verschil maakt of we eerst partieel naar x en dan naar y afleiden, of andersom.

Merk op: Het is helemaal niet vanzelfsprekend dat $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$ geldt, maar dit is inderdaad altijd (behalve in kunstmatig geconstrueerde gevallen) zo en heet de *Stelling van Schwarz*. We hoeven dus bij het achter elkaar uitvoeren van meerdere partiële afgeleiden *niet* op de volgorde te letten.

Notatie: We hebben al gezien dat we een tweede partiële afgeleide $\frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)$ schrijven als $\frac{\partial^2 f}{\partial y \partial x}$. Als we twee keer in dezelfde richting afleiden, wordt dit afgekort met $\frac{\partial^2 f}{\partial x^2} := \frac{\partial^2 f}{\partial x \partial x}$.

In de vorige les hebben we gezien, dat sommige belangrijke functies door relaties met hun afgeleiden gekarakteriseerd kunnen worden. Zo is de exponentiële functie een oplossing van de differentiaalvergelijking $f'(x) = f(x)$ en de sinus en cosinus zijn oplossingen van de vergelijking $f(x) + f''(x) = 0$.

Dit soort vergelijkingen bestaat ook voor de partiële afgeleiden, we noemen zo'n vergelijking een *partiële differentiaalvergelijking*. Als voorbeeld kijken we naar de vergelijking:

$$x^2 \frac{\partial^2 f}{\partial x^2} + 2xy \frac{\partial^2 f}{\partial x \partial y} + y^2 \frac{\partial^2 f}{\partial y^2} = 0.$$

We laten zien dat de functie $f(x, y) := \frac{xy}{x-y}$ een oplossing hiervan is: We hebben $\frac{\partial f}{\partial x} = \frac{y(x-y) - xy}{(x-y)^2} = -\frac{y^2}{(x-y)^2}$ en $\frac{\partial f}{\partial y} = \frac{x(x-y) + xy}{(x-y)^2} = \frac{x^2}{(x-y)^2}$. Hieruit volgt $\frac{\partial^2 f}{\partial x^2} = \frac{2y^2}{(x-y)^3}$, $\frac{\partial^2 f}{\partial y^2} = \frac{2x^2}{(x-y)^3}$ en $\frac{\partial^2 f}{\partial x \partial y} = \frac{2x(x-y)^2 - x^2 \cdot 2(x-y)}{(x-y)^4} = \frac{-2xy}{(x-y)^3}$ en we zien dat $f(x, y)$ inderdaad aan de vergelijking voldoet.

In het algemeen is het oplossen van partiële differentiaalvergelijkingen een ingewikkeld probleem, meestal laten zich oplossingen alleen maar door numerieke benaderingsmethoden vinden.

Soms spelen bij de beschrijving van problemen verschillende soorten van variabelen een rol, bijvoorbeeld ruimtelijke coördinaten en tijd of temperatuur

en druk. Als voorbeeld hiervoor bekijken we de functie

$$f(x, y, t) := \exp(-t)(\sin(x) + \cos(y)).$$

Hiervoor geldt: $\frac{\partial f}{\partial x} = \exp(-t) \cos(x)$, $\frac{\partial f}{\partial y} = -\exp(-t) \sin(y)$, dus hebben we $\frac{\partial^2 f}{\partial x^2} = -\exp(-t) \sin(x)$ en $\frac{\partial^2 f}{\partial y^2} = -\exp(-t) \cos(y)$. Aan de andere kant geldt $\frac{\partial f}{\partial t} = -\exp(-t)(\sin(x) + \cos(y))$, dus voldoet de functie aan de vergelijking

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial f}{\partial t}.$$

Deze vergelijking beschrijft de uitbreiding van de hitte (de parameter t) op een gegeven oppervlak (met coördinaten x en y) en heet de *hittevergelijking*.

8.3 Minima en maxima van functies van meerdere variabelen

Ook bij functies van meerdere variabelen kunnen we ons natuurlijk afvragen hoe we minima en maxima kunnen vinden. Door de interpretatie van de partiële afgeleide als richtingscoëfficiënt van de raaklijn parallel met een van de coördinatenassen zien we, dat in een lokaal minimum alle partiële afgeleiden gelijk aan 0 moeten zijn. Net als bij de functies van één variabele is dit alleen maar een noodzakelijke voorwaarde, maar geeft toch kandidaten voor minima en maxima.

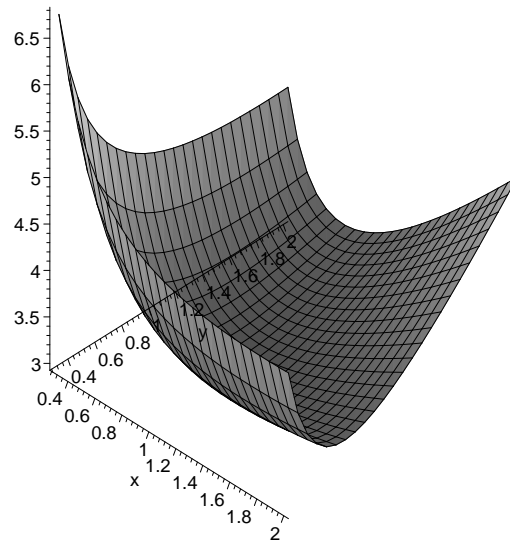
We kunnen bijvoorbeeld na gaan dat een kubus de (gesloten) doos met minimale oppervlakte voor gegeven volume is. Het volume van een doos met zijden x, y, z is geven als $V = xyz$ de oppervlakte als $O(x, y, z) = 2(xy + yz + zx)$. We kunnen de derde coördinaat z uitdrukken door $z = \frac{V}{xy}$, dan wordt

$$O = O(x, y) = 2\left(xy + \frac{V}{x} + \frac{V}{y}\right).$$

We hebben $\frac{\partial O}{\partial x} = 2\left(y - \frac{V}{x^2}\right)$ en $\frac{\partial O}{\partial y} = 2\left(x - \frac{V}{y^2}\right)$. Uit $\frac{\partial O}{\partial x} = \frac{\partial O}{\partial y} = 0$ volgt dus $x^2y = V$ en $xy^2 = V$. Als we deze twee vergelijkingen van elkaar aftrekken vinden we $xy(x - y) = 0$, dus $x = y$. Dan geldt $V = x^2y = x^3$, dus is ook $z = x$.

Hoe zou dit (niet onverwachte) resultaat veranderen als we naar een open doos zonder deksel kijken? We kunnen deze vraag ook zo interpreteren, dat het materiaal voor bodem en deksel van een gesloten doos slechts half zo duur is als het materiaal voor de zijvlakken. Dan zouden we verwachten dat het goedkoper wordt, als we bodem en deksel iets groter en de zijvlakken iets kleiner maken.

Het volume van een open doos is nog steeds $V = xyz$, maar de oppervlakte wordt nu $O(x, y, z) = xy + 2yz + 2zx$ (als de z -as de verticale as is). Door weer z door $\frac{V}{xy}$ te vervangen krijgen we $O(x, y) = xy + \frac{2V}{x} + \frac{2V}{y}$. Voor de partiële afgeleiden geldt $\frac{\partial O}{\partial x} = y - \frac{2V}{x^2}$ en $\frac{\partial O}{\partial y} = x - \frac{2V}{y^2}$, dus krijgen we uit $\frac{\partial O}{\partial x} = 0$ en $\frac{\partial O}{\partial y} = 0$ de noodzakelijke voorwaarden $x^2y = 2V$ en $xy^2 = 2V$. Deze vergelijkingen trekken we weer van elkaar af, dan volgt $xy(x - y) = 0$ en dus $x = y$ en dus $x^3 = 2V$. We hebben dus $x = y = \sqrt[3]{2V}$ en hieruit volgt $z = \sqrt[3]{\frac{1}{4}V}$.



Figuur II.22: Oppervlakte van een (gesloten) doos van volume 1 afhankelijk van de lengten van de x - en y -zijden

Tegenover de gesloten doos met $x = y = z = \sqrt[3]{V}$ zijn dus x en y een factor $\sqrt[3]{2} \approx 1.26$ groter en de hoogte z is een factor $\sqrt[3]{\frac{1}{4}} \approx 0.63$ kleiner.

Een verdere toepassing voor het vinden van minima van een functie van meerdere variabelen ligt in het vinden van een regressielijn door een aantal gegeven punten. De regressielijn door de punten is bepaald door de eigenschap dat de som van de kwadraten van de verticale afstanden tussen de punten en de lijn minimaal wordt. Stel we hebben punten $(x_1, y_1), \dots, (x_n, y_n)$ gegeven. Als we door deze punten de lijn $y = ax + b$ leggen, wordt de som van de kwadratische afstanden gelijk aan

$$f(a, b) := (ax_1 + b - y_1)^2 + \dots + (ax_n + b - y_n)^2.$$

De partiële afgeleide van f naar a is gegeven door $\frac{\partial f}{\partial a} = 2(ax_1 + b - y_1)x_1 + \dots + 2(ax_n + b - y_n)x_n = \sum_{i=1}^n 2(ax_i + b - y_i)x_i$ en de partiële afgeleide naar b is $\frac{\partial f}{\partial b} = 2(ax_1 + b - y_1) + \dots + 2(ax_n + b - y_n) = \sum_{i=1}^n 2(ax_i + b - y_i)$. De voorwaarden $\frac{\partial f}{\partial a} = 0$ en $\frac{\partial f}{\partial b} = 0$ kunnen we nu schrijven als een stelsel lineaire vergelijkingen voor a en b . Hiervoor is het handig een notatie van Gauss te gebruiken, namelijk een som $\sum_{i=1}^n z_i$ af te korten als $[z]$. We hebben dan:

$$\begin{pmatrix} [x^2] & [x] \\ [x] & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} [xy] \\ [y] \end{pmatrix}.$$

Deze matrix kunnen we expliciet invertieren, de determinant is $n[x^2] - [x]^2$ en dus geldt

$$(n[x^2] - [x]^2) \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} n & -[x] \\ -[x] & [x^2] \end{pmatrix} \begin{pmatrix} [xy] \\ [y] \end{pmatrix} = \begin{pmatrix} n[xy] - [x][y] \\ [x^2][y] - [x][xy] \end{pmatrix}$$

en dus vinden we de coëfficiënten van de regressielijn door

$$a = \frac{n[xy] - [x][y]}{n[x^2] - [x]^2} = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2},$$

$$b = \frac{[x^2][y] - [x][xy]}{n[x^2] - [x]^2} = \frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n x_i y_i)}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2}.$$

BELANGRIJKE BEGRIPPEN IN DEZE LES

- hogere afgeleide
- lokaal (absoluut) minimum/maximum
- kritieke waarden
- functies van meerdere variabelen
- partiële afgeleide
- regressielijn

OPGAVEN

40. Bepaal voor ieder van de volgende functies de lokale minima en maxima op het gegeven domein (let ook op de randpunten).

- (i) $f : [-\frac{1}{2}, \frac{1}{2}] \rightarrow \mathbb{R}, x \mapsto 3x^4 - 8x^3 + 6x^2,$
- (ii) $f : [-\frac{1}{2}, 1] \rightarrow \mathbb{R}, x \mapsto \frac{1}{x^5 + x + 1},$
- (iii) $f : [-1, \frac{1}{2}] \rightarrow \mathbb{R}, x \mapsto \frac{x+1}{x^2+1},$
- (iv) $f : [0, 5] \setminus \{1\} \rightarrow \mathbb{R}, x \mapsto \frac{x}{x^2-1}.$

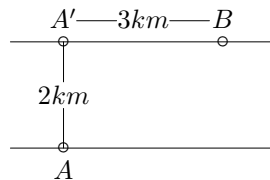
41. Bepaal voor de functie $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2 \exp(x)$ de lokale minima en maxima.

42. Zij $a > 0$ en $f : \mathbb{R} \rightarrow \mathbb{R}$ gegeven door

$$f(x) := \frac{1}{1 + |x|} + \frac{1}{1 + |x - a|}.$$

Bepaal het absolute maximum van f . (Hint: Bepaal de afgeleide apart op de deelintervallen $(-\infty, 0), (0, a)$ en (a, ∞)).

43. Iemand wil van een punt A aan de oever van een $2km$ breed kanaal naar een punt B aan de andere oever van het kanaal. Het punt A' rechtstreeks tegenover A aan het andere oever heeft een afstand van $3km$ van het punt B . Op het kanaal kan hij met een snelheid van $3km/h$ roeien, aan land loopt hij met een snelheid van $6km/h$. Wat is de snelste weg om van A naar B te komen?



44. Bepaal de hoogte en het volume van de grootste cilinder (qua volume) die in een kogel van straal r past.
45. Baron Münchhausen wordt op zijn kogel met een hoek van α tegen de grond en een snelheid van v afgevuurd. Zijn traject wordt beschreven door $(x, y) = (v \cos(\alpha)t, v \sin(\alpha)t - \frac{g}{2}t^2)$, waarbij g de acceleratie door de aantrekking van de aarde is (dus ongeveer $9.81 \frac{m}{s^2}$).
- (i) Na welke tijd t komt de Baron weer naar de grond?
 - (ii) Bepaal de hoek α zo dat de Baron zo ver als mogelijk op zijn kogel kan rijden.
 - (iii) Als de Baron van een hogere punt (bijvoorbeeld zijn dakterras) wordt afgevuurd, moet de optimale hoek α dan groter of kleiner gekozen worden?
46. Bepaal voor elk van de volgende functies de partiële afgeleiden $\frac{\partial f}{\partial x}$ en $\frac{\partial f}{\partial y}$ naar x en naar y :
- (i) $f(x, y) := \frac{x^2}{y} + \frac{y^2}{x}$,
 - (ii) $f(x, y) := \sin(2x + 3y)$,
 - (iii) $f(x, y) := \arctan(x^2y) + \arctan(xy^2)$,
 - (iv) $f(x, y) := \exp(x^2 + xy)$.
47. Een ellipsoïd is gegeven door de vergelijking $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ (voor $a = b = c = r$ geeft dit een kogel van straal r). Bepaal het maximale volume van een doos die in het ellipsoïd past.

Les 9 Primitieve en integraal

Een motivatie om naar de afgeleide $f'(x)$ van een functie $f(x)$ te kijken is het bepalen van de richtingscoëfficiënt van de raaklijn aan de grafiek van $f(x)$. Uiteindelijk hebben we de afgeleide gedefinieerd als de limiet

$$f'(x) := \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

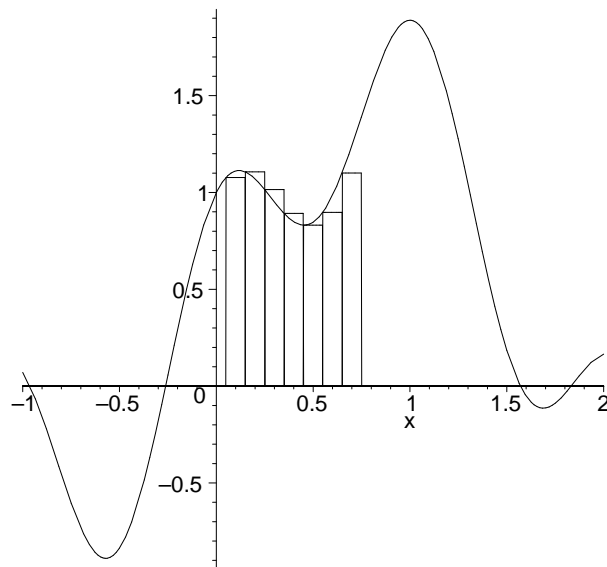
als deze bestaat. De afgeleide $f'(x)$ geeft informatie over het stijgen en dalen van $f(x)$ en is daarom ook een belangrijk hulpmiddel bij het opsporen van minima en maxima van de functie $f(x)$.

Het is nu een (min of meer) voor de hand liggende vraag, of we ook aan de omgekeerde overgang van $f'(x)$ naar $f(x)$ iets hebben. Met andere woorden: Stel, $F(x)$ is een functie zo dat $F'(x) = f(x)$, wat voor informatie geeft dan de functie $F(x)$? Er geldt in dit geval

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} \text{ en dus } f(x) \cdot h \approx F(x+h) - F(x)$$

en dus benadert het verschil $F(x+h) - F(x)$ de oppervlakte van de rechthoek met hoogte $f(x)$ en breedte h . We kunnen dus verwachten, dat de functie $F(x)$ iets met de oppervlakte onder de grafiek van $f(x)$ te maken heeft. Preciezer:

Om de oppervlakte onder de grafiek van $f(x)$ tussen a en b te bepalen, delen we het interval $[a, b]$ in N even grote delen, deze hebben dus breedte $h = \frac{b-a}{N}$. De oppervlakte O onder de grafiek wordt dan benaderd door de som van de oppervlakten van de N rechthoeken van hoogte $f(a + jh)$ en breedte h voor $j = 0, 1, \dots, N - 1$.



Figuur II.23: Benadering van de oppervlakte onder een grafiek door rechthoeken

We hebben dus $O \approx f(a) \cdot h + f(a+h) \cdot h + \dots + f(a+(N-1)h) \cdot h$. Maar aan de andere kant hebben we boven gezien dat $f(x) \cdot h \approx F(x+h) - F(x)$, en

hieruit volgt dat

$$\begin{aligned} O &\approx (F(a+h) - F(a)) + (F(a+2h) - F(a+h)) + \dots \\ &\dots + (F(a+(N-1)h) - F(a+(N-2)h)) + (F(b) - F(a+(N-1)h)) \\ &= F(b) - F(a). \end{aligned}$$

Het lijkt dus, dat we een functie $F(x)$ met $F'(x) = f(x)$ kunnen gebruiken om de oppervlakte onder de grafiek van $f(x)$ te bepalen.

9.1 De oppervlakte onder een grafiek

Boven hebben we gezien dat een functie $F(x)$ met $F'(x) = f(x)$ iets met de oppervlakte onder de grafiek van $f(x)$ te maken heeft.

We gaan nu omgekeerd aantonen dat we uit twee voor de hand liggende eisen aan een functie, die de oppervlakte onder de grafiek van $f(x)$ aangeeft, kunnen concluderen dat deze functie noodzakelijk de afgeleide $f(x)$ heeft.

Voor een functie $f(x)$ zij $O_f(a, b)$ de oppervlakte onder de grafiek van $f(x)$ in het interval $[a, b]$. Onze twee eisen zijn als volgt:

- (i) $O_f(a, b) + O_f(b, c) = O_f(a, c)$, dus het opsplitsen van een interval in twee deelintervallen verandert de oppervlakte niet.
- (ii) Als $m \leq f(x) \leq M$ voor alle $x \in [a, b]$ dan is $m(b-a) \leq O_f(a, b)$ en $O_f(a, b) \leq M(b-a)$, dus de oppervlakte ligt tussen de oppervlakten van de rechthoeken met hoogte het minimum en het maximum van alle functiewaarden.

Omdat een lijn geen oppervlakte heeft, geldt $O_f(a, a) = 0$ en met eis (i) volgt hieruit dat $O_f(a, b) = -O_f(b, a)$. Dit betekent dat een oppervlakte die we van rechts naar links aangeven de negatieve waarde van dezelfde oppervlakte met de gewone oriëntatie (van links naar rechts) heeft.

Ook voor een functie met $f(x) < 0$, dus met een grafiek onder de x -as, krijgen we een negatieve oppervlakte, want in eis (ii) kunnen we $M = 0$ kiezen. Het is afhankelijk van de toepassing of we inderdaad de oppervlakten onder de x -as negatief of positief willen tellen, in het laatste geval moeten we dan naar de functie $g(x) := |f(x)|$ in plaats van $f(x)$ kijken.

Als we het interval $[a, b]$ onderverdelen in N deelintervallen $[x_1, x_2], [x_2, x_3], \dots, [x_N, x_{N+1}]$ met $x_1 = a$ en $x_{N+1} = b$, dan geldt volgens eis (i):

$$O_f(a, b) = O_f(x_1, x_2) + O_f(x_2, x_3) + \dots + O_f(x_N, x_{N+1}).$$

Als we nu aannemen dat $f(x)$ op $[a, b]$ continu is, weten we dat in een interval $(x_i - \delta, x_i + \delta)$ om x_i de functiewaarden $f(x)$ in het interval $(f(x_i) - \varepsilon, f(x_i) + \varepsilon)$ liggen. Als we een $\varepsilon > 0$ kiezen, kunnen we een onderverdeling aannemen zo dat $|f(x) - f(x_i)| < \varepsilon$ voor alle $x \in [x_i, x_{i+1}]$. Dan geldt volgens eis (ii) dat $(f(x_i) - \varepsilon)(x_{i+1} - x_i) \leq O_f(x_i, x_{i+1}) \leq (f(x_i) + \varepsilon)(x_{i+1} - x_i)$ en dus

$$(f(x_i) - \varepsilon) \leq \frac{O_f(x_i, x_{i+1})}{x_{i+1} - x_i} \leq (f(x_i) + \varepsilon).$$

Door de limiet $\varepsilon \rightarrow 0$ te nemen zien we dat de limiet

$$\lim_{h \rightarrow 0} \frac{O_f(x_i, x_i + h)}{h}$$

bestaat en de waarde $f(x_i)$ heeft. In het bijzonder kunnen we nu voor een vast gekozen punt x_0 op het interval $[a, b]$ een functie F definiëren door

$$F(x) := O_f(x_0, x)$$

Dan geldt $F(x+h) - F(x) = O_f(x_0, x+h) - O_f(x_0, x) = O_f(x, x+h)$ en dus is $F(x)$ een differentieerbare functie met afgeleide $f(x)$. We hebben dus gezien dat de oppervlakte $O_f(a, b)$ gegeven is door

$$O_f(a, b) = F(b) - F(a) \text{ voor een functie } F(x) \text{ met } F'(x) = f(x).$$

Stel nu we hebben een tweede functie $G(x)$ met $G'(x) = f(x)$, dan geldt $(F - G)'(x) = F'(x) - G'(x) = f(x) - f(x) = 0$, dus is $(F - G)(x)$ een constante functie en dus $G(x) = F(x) + C$ voor een constante $C \in \mathbb{R}$. Maar dan is $G(b) - G(a) = F(b) + C - F(a) - C = F(b) - F(a)$ en dus kunnen we de oppervlakte $O_f(a, b)$ ook met behulp van de functie $G(x)$ aangeven. We hebben dus aangetoond:

II.6 Stelling *De oppervlakte onder de grafiek van een continue functie $f(x)$ in het interval $[a, b]$ is $F(b) - F(a)$ voor een functie $F(x)$ met $F'(x) = f(x)$. Dit is onafhankelijk van de keuze van de functie $F(x)$.*

9.2 De primitieve en de integraal

We noemen een functie $F(x)$ met $F'(x) = f(x)$ een *primitieve* van $f(x)$. Als $F(x)$ een primitieve van $f(x)$ dan is ook $F(x) + C$ voor een constante $C \in \mathbb{R}$ een primitieve van $f(x)$, dus is de primitieve niet eenduidig bepaald. Aan de andere kant verschillen twee primitieve functies van $f(x)$ alleen maar een constante, daarom wordt de primitieve van een functie vaak met $F(x) + C$ aangegeven, waarbij C een niet verder bepaalde constante is.

We hebben boven gezien dat een continue functie altijd een primitieve heeft en dat deze een differentieerbare functie is. Als een functie $f : [a, b] \rightarrow \mathbb{R}$ continu is behalve in een punt $c \in [a, b]$, kunnen we een differentieerbare primitieve $F_1(x)$ op het interval $[a, c]$ en een differentieerbare primitieve $F_2(x)$ op het interval $[c, b]$ vinden. Dan is de functie

$$F(x) := \begin{cases} F_1(x) & \text{als } x \leq c \\ F_2(x) - F_2(c) + F_1(c) & \text{als } x > c \end{cases}$$

een continue functie op $[a, b]$ die voor $x \neq c$ differentieerbaar is met $F'(x) = f(x)$.

Op deze manier kunnen we continue primitieven voor alle functies vinden, die alleen maar in geïsoleerde punten sprongen hebben.

Voor functies $f(x)$ die in willekeurig kleine intervallen in oneindig veel punten niet continu zijn is een iets ingewikkeldere definitie van een primitieve nodig, maar dit soort exotische functies gaan we hier niet verder bekijken.

De gebruikelijke notatie voor de primitieve functie $F(x)$ van $f(x)$ is de *integraal*

$$F(x) = \int_{x_0}^x f(t) dt \quad \text{of} \quad F(x) = \int f(x) dx.$$

De eerste vorm (met grenzen) noemen we ook de *bepaalde integraal* de tweede (zonder grenzen) de *onbepaalde integraal*. Bij de onbepaalde integraal identificeren we primitieven die een constante verschillen.

De notatie van de integraal is ingevoerd door Gottfried Wilhelm Leibniz (1646-1716) die parallel met Isaac Newton (1643-1727) de cruciale principes van de calculus ontwikkeld heeft. De notatie is afgeleid van de betekenis van de primitieve voor de oppervlakte onder de grafiek van een functie:

$$\begin{aligned} \int_a^b f(t) dt &= O_f(a, b) = \sum_{i=1}^N O_f(x_i, x_{i+1}) \approx \sum_{i=1}^n f(x_i)(x_{i+1} - x_i) \\ &= \sum_{i=1}^n f(x_i)\Delta x_i. \end{aligned}$$

De Σ is de Griekse letter S en staat voor *som* (of *Summe*), het teken \int voor de integraal lijkt op een uitgerekte S . Om aan te duiden dat er een limiet voor $\Delta x_i \rightarrow 0$ plaats vindt, wordt het symbool dx geschreven. Dit is net als $\frac{\partial}{\partial x_i}$ in de partiële afgeleide een formeel symbool dat aangeeft welke variabele gevarieerd wordt.

Het lijkt een flauwe opmerking dat een functie $f(x)$ een primitieve voor zijn afgeleide $f'(x)$ is. Maar dit betekent, dat het integreren de omkering van het afleiden is, want we hebben

$$\int f'(x) dx = f(x).$$

Deze formule staat bekend als de *Hoofdstelling van de Calculus*.

Uit de definitie van de integraal en de eigenschappen van de afgeleide volgen meteen twee belangrijke rekenregels voor de integraal:

$$\begin{aligned} \int_a^b f(x) dx + \int_a^b g(x) dx &= \int_a^b (f + g)(x) dx \\ c \cdot \int_a^b f(x) dx &= \int_a^b (cf)(x) dx \quad \text{voor } c \in \mathbb{R} \end{aligned}$$

We zien dus dat de integraal (net zo als de afgeleide) een lineaire afbeelding op de vectorruimte van continue functies is.

Waarschuwing: Bij het afleiden van functies hebben we gezien dat er regels bestaan zo dat we de afgeleide $f'(x)$ van een functie $f(x)$ die uit elementaire functies (veelterm-functies, exponentiële functie, logaritme, trigonometrische functies en hun inverse functies) opgebouwd is, weer als combinatie van elementaire functies kunnen schrijven. Voor de integraal geldt dit niet! Er zijn functies $f(x)$ zo dat we de integraal $\int f(x) dx$ niet als combinatie van elementaire functies kunnen schrijven. Dit ligt niet eraan dat we te stom zijn om zo'n functie te vinden maar het is mogelijk te bewijzen dat zo'n functie niet bestaat.

Een voorbeeld hiervoor is de functie $f(x) := \exp(-x^2) = e^{-x^2}$ die een belangrijke rol in de statistiek speelt (bijvoorbeeld bij de normale verdeling). Deze functie is continu (zelfs differentieerbaar), maar de enige manier om de primitieve $F(x)$ van deze functie te schrijven is de integraal $F(x) = \int \exp(-x^2) dx$.

Een ingewikkelder voorbeeld is de Γ -functie $\Gamma(x) := \int_0^\infty \exp(-t)t^{x-1} dt$. Ook deze functie is niet zonder integraal te schrijven. Er laat zich wel aantonen dat $\Gamma(n) = (n - 1)!$ voor $n \in \mathbb{N}$, dus is de Γ -functie een soort interpolatie voor de faculteit van natuurlijke getallen.

Een aantal integralen kunnen we al berekenen, omdat we bij het differentiëren gezien hebben, dat zekere functies 'eenvoudige' afgeleiden hebben. Een paar voorbeelden zijn:

- (i) $\int x^c dx = \frac{1}{c+1}x^{c+1}$ voor $c \neq -1, c \in \mathbb{R}$,
- (ii) $\int \exp(x) dx = \exp(x)$ en $\int \frac{1}{x} dx = \log(x)$,
- (iii) $\int \sin(x) dx = -\cos(x)$ en $\int \cos(x) dx = \sin(x)$,
- (iv) $\int \frac{1}{1+x^2} dx = \arctan(x)$ en $\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin(x)$.

Soms verschilt een functies maar een beetje (bijvoorbeeld door een constante) van een functie waar je de integraal al van kent. Dan laat zich vaak ook de integraal van de veranderde functie makkelijk berekenen. Bijvoorbeeld heeft $\int \cos(ax) dx$ voor een constante a zeker iets met $\sin(ax)$ te maken, maar $\sin(ax)' = a \cos(ax)$, dus moet je nog met $\frac{1}{a}$ vermenigvuldigen en krijgt zo

$$\int \cos(ax) dx = \frac{1}{a} \sin(ax).$$

Een iets slimmer trucje is bij de integraal $\int \frac{1}{\sqrt{a^2-x^2}} dx$ nodig. Bij de primitieve speelt zeker de *arcussinus* een rol, en als je $\arcsin(bx)$ afleidt, krijg je (met behulp van de kettingregel): $\arcsin(bx)' = \frac{b}{\sqrt{1-(bx)^2}} = \frac{b}{b\sqrt{\frac{1}{b^2}-x^2}} = \frac{1}{\sqrt{\frac{1}{b^2}-x^2}}$. We moeten dus $b = \frac{1}{a}$ kiezen, dit geeft

$$\int \frac{1}{\sqrt{a^2-x^2}} dx = \arcsin\left(\frac{x}{a}\right).$$

We kunnen met behulp van de integraal nu bijvoorbeeld de oppervlakte van een cirkel met straal 1 bepalen. De punten (x, y) op de cirkel voldoen aan $x^2 + y^2 = 1$, dus is de bovenhelft van de cirkel de grafiek van $f(x) := \sqrt{1 - x^2}$. De oppervlakte van de cirkel vinden we dus als $4 \cdot \int_0^1 \sqrt{1 - x^2} dx$. We zullen later nog een andere manier zien hoe we deze integraal kunnen berekenen, maar door een goede gok te doen, kunnen we hem ook meteen oplossen.

Er geldt $(x \cdot \sqrt{1 - x^2})' = \sqrt{1 - x^2} + x \cdot \frac{-2x}{2\sqrt{1 - x^2}} = \sqrt{1 - x^2} + \frac{1 - x^2}{\sqrt{1 - x^2}} = 2\sqrt{1 - x^2} - \frac{1}{\sqrt{1 - x^2}}$. We brengen nu $2\sqrt{1 - x^2}$ naar een zijde en delen door 2, dit geeft $\sqrt{1 - x^2} = \frac{1}{2}((x \cdot \sqrt{1 - x^2})' + \frac{1}{\sqrt{1 - x^2}})$ en volgens de Hoofdstelling van de Calculus geeft dit:

$$\int \sqrt{1 - x^2} dx = \frac{1}{2}(x \cdot \sqrt{1 - x^2} + \int \frac{1}{\sqrt{1 - x^2}} dx) = \frac{1}{2}(x \cdot \sqrt{1 - x^2} + \arcsin(x)).$$

Hieruit volgt voor de bepaalde integraal

$$\int_0^1 \sqrt{1 - x^2} dx = \frac{1}{2}(1 \cdot 0 + \arcsin(1)) - \frac{1}{2}(0 \cdot 1 + \arcsin(0)) = \frac{1}{2} \cdot \frac{\pi}{2} = \frac{\pi}{4}$$

en we vinden (niet onverwacht) π als oppervlakte van een cirkel met straal 1.

Een handige notatie om integralen kort door hun primitieven te beschrijven is de volgende: Als $F(x)$ een primitieve van $f(x)$ is, dus $F'(x) = f(x)$, dan schrijven we

$$\int_a^b f(x) dx = F(b) - F(a) = F(x)|_a^b.$$

9.3 Partieel integreren

Omdat de integraal de omkering van de afgeleide is, kunnen we verwachten dat ook uit de productregel iets nuttigs voor het primitiveren volgt. De productregel voor het afleiden zegt dat

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

en door deze vergelijking te primitiveren (dat wil zeggen op beide zijden de integraal te nemen), krijgen we

$$f(x)g(x)|_a^b = \int_a^b f'(x)g(x) dx + \int_a^b f(x)g'(x) dx.$$

Door één van de integralen naar de andere kant te brengen volgt hieruit de regel voor de *partiële integratie*:

$$\begin{aligned} \int_a^b f'(x)g(x) dx &= f(x)g(x)|_a^b - \int_a^b f(x)g'(x) dx, \\ \int_a^b f(x)g'(x) dx &= f(x)g(x) - \int_a^b f'(x)g(x) dx. \end{aligned}$$

De grap bij deze formule is dat we sommige functies eenvoudig kunnen primitiveren en dat een integraal misschien eenvoudiger wordt als we een deel primitiveren en de rest afleiden.

Helaas zijn er geen regels hoe we een functie als product van twee functies moeten schrijven en welke van de twee delen we als $f'(x)$ en welke we als $g(x)$ moeten kiezen. Dit is een kwestie van oefening en ervaring en soms is het verstandig de verschillende mogelijkheden te proberen. Meestal zijn er namelijk niet zo veel mogelijkheden om een functie als product van twee functies te schrijven en we moeten dus alleen maar kiezen of we een factor als $f'(x)$ of als $g(x)$ willen gebruiken.

Een paar typische toepassingen zijn de beste manier om te zien hoe de partiële integratie werkt:

$$\int \underbrace{x}_{g(x)} \underbrace{\exp(x)}_{f'(x)} dx = x \exp(x) - \int \exp(x) dx = (1-x) \exp(x).$$

In het volgende voorbeeld moet men eerst iets kunstmatig een product van twee functies produceren:

$$\int \log(x) dx = \int \underbrace{1}_{f'(x)} \cdot \underbrace{\log(x)}_{g(x)} dx = x \log(x) - \int x \frac{1}{x} dx = x \log(x) - x.$$

Soms lukt het bij de partiële integratie, de oorspronkelijke integraal op de rechterzijde terug te vinden, maar met een andere coëfficiënt. Dit is in het bijzonder bij functies met $\sin(x)$ of $\cos(x)$ vaak het geval, ook al is het hier soms nodig meer dan een keer partieel te integreren omdat $\sin''(x) = -\sin(x)$. Twee typische voorbeelden hiervoor zijn:

$$\begin{aligned} \int \sin^2(x) dx &= \int \underbrace{\sin(x)}_{f'(x)} \underbrace{\sin(x)}_{g(x)} dx = -\cos(x) \sin(x) - \int -\cos(x) \cos(x) dx \\ &= -\cos(x) \sin(x) + \int (1 - \sin^2(x)) dx \\ &= -\cos(x) \sin(x) + \int 1 dx - \int \sin^2(x) dx \\ &= -\cos(x) \sin(x) + x - \int \sin^2(x) dx \\ \Rightarrow 2 \int \sin^2(x) dx &= -\cos(x) \sin(x) + x \\ \Rightarrow \int \sin^2(x) dx &= \frac{1}{2}(-\cos(x) \sin(x) + x) \end{aligned}$$

en (met twee keer partieel integreren)

$$\begin{aligned} \int \underbrace{\exp(x)}_{f'(x)} \underbrace{\sin(x)}_{g(x)} dx &= \exp(x) \sin(x) - \int \underbrace{\exp(x)}_{f'(x)} \underbrace{\cos(x)}_{g(x)} dx \\ &= \exp(x) \sin(x) - \exp(x) \cos(x) - \int \exp(x) \sin(x) dx \\ \Rightarrow \int \exp(x) \sin(x) dx &= \frac{1}{2}(\exp(x) \sin(x) - \exp(x) \cos(x)). \end{aligned}$$

BELANGRIJKE BEGRIPPEN IN DEZE LES

- oppervlakte onder een grafiek
- primitieve functie
- integraal van een functie
- Hoofdstelling van de calculus
- partieel integreren

OPGAVEN

48. Bepaal primitieven voor de volgende functies:

$$(i) f_1(x) := a^x \text{ met } a > 0, a \neq 1, \quad (ii) f_2(x) := \frac{1}{1+x}, \quad (iii) f_3(x) := \frac{x}{1+x},$$

$$(iv) f_4(x) := \frac{1}{a^2 + x^2} \text{ met } a \neq 0, \quad (v) f_5(x) := \frac{1}{\sqrt{x-1} + \sqrt{x+1}}.$$

49. Bereken de volgende integralen:

$$\int_0^1 (1-x)^n dx \text{ voor } n \in \mathbb{N} \quad \text{en} \quad \int_0^\pi \sin(mx) dx \text{ voor } m \in \mathbb{Z}.$$

50. Bepaal de oppervlakte van het gebied dat door de grafieken van $f(x) := x^2$ en $g(x) := \frac{x^2}{2} + 2$ wordt ingesloten (dus het gebied tussen de grafieken op het interval tussen hun snijpunten).

51. Bepaal de volgende integralen door partiële integratie:

$$(i) \int x \sin(x) dx, \quad (ii) \int x^2 \exp(x) dx, \quad (iii) \int \sqrt{x} \log(x) dx, \quad (iv) \int \log^2(x) dx,$$

$$(v) \int \log^3(x) dx, \quad (vi) \int \cos(\log(x)) dx, \quad (vii) \int x \arctan(x) dx.$$

Les 10 Substitutie

We hebben gezien dat de productregel voor het differentiëren aanleiding geeft tot een manier, om voor zekere functies een primitieve te vinden, namelijk door partiële integratie. Een regel voor het differentiëren die we bij het integreren nog niet hebben gebruikt is de kettingregel en het zou geen verrassing zijn dat we ook hier iets voor het primitiveren aan hebben.

Als we bijvoorbeeld naar de functie $F(x) := \exp(x^2)$ kijken, dan geldt $F'(x) = 2x \cdot \exp(x^2)$, dus kunnen we de integraal $\int x \exp(x^2) dx$ bepalen door

$$\int_a^b x \exp(x^2) dx = \frac{1}{2} \exp(x^2) \Big|_a^b = \frac{1}{2} (\exp(b^2) - \exp(a^2)).$$

Algemeen zien we dat we een integraal van de vorm $\int g'(f(x)) \cdot f'(x) dx$ kunnen oplossen, want volgens de kettingregel is $g'(f(x)) \cdot f'(x) = g(f(x))'$ en hieruit volgt met de hoofdstelling van de calculus dat

$$\int g'(f(x)) \cdot f'(x) dx = \int g(f(x))' dx = g(f(x)).$$

Dit geeft dus een manier hoe we voor zekere functies met behulp van de kettingregel een primitieve kunnen vinden. Maar we zullen zien dat we dit trucje nog veel slimmer kunnen toepassen, en niet alleen maar als we rechtstreeks een functie van de vorm $g'(f(x)) \cdot f'(x)$ in een integraal vinden.

Om dit op een handige manier te kunnen beschrijven, gaan we eerst nog eens de notatie van Leibniz voor de afgeleide nader bekijken.

10.1 Rekenen met differentiaal

De afgeleide hebben we gedefinieerd als limiet van de quotiënt $\frac{f(x+h)-f(x)}{h}$. Als we met Δ een (klein) verschil van waarden noteren, kunnen we dit ook schrijven als $\frac{\Delta f(x)}{\Delta x}$. Om nu de limiet $\Delta x \rightarrow 0$ duidelijk te maken, vervangen we de Δ door een d , dit geeft dus

$$f'(x) = \frac{df(x)}{dx} = \frac{df}{dx}.$$

Tot hier toe hebben we alleen maar de notatie veranderd en zien $\frac{df}{dx}$ als een formeel symbool, dat we als *df naar dx* lezen. Inderdaad hebben we deze notatie al bij de partiële afgeleide $\frac{\partial f}{\partial x}$ gebruikt. Maar we kunnen $\frac{df}{dx}$ ook als breuk zien en met teller en noemer net zo rekenen als bij breuken. We noemen de uitdrukking dx dan een *differentiaal*.

Het belangrijke punt is nu dat in de integraal $\int f(x) dx$ de uitdrukking dx inderdaad een differentiaal is. De samenhang is als volgt:

Laat $F(x)$ een primitieve van de functie $f(x)$ zijn, dan is $F'(x) = \frac{dF}{dx} = f(x)$ en dus $dF(x) = f(x) dx$. Als we nu $dF(x)$ weer als limiet van $\Delta F(x)$ zien, dan is de integraal $\int dF(x)$ de limiet van de som $\sum \Delta F(x)$ en dus gelijk aan $F(x)$.

Op deze manier vinden we de ons al bekende formule voor de (onbepaalde) integraal terug:

$$\int f(x) dx = F(x) = \int dF(x).$$

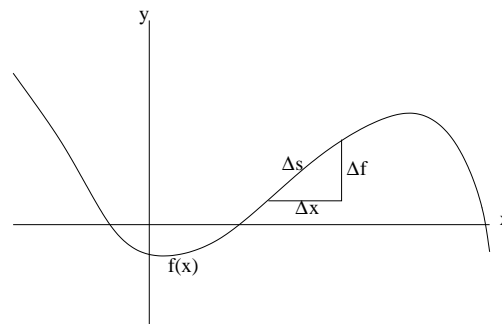
Merk op: Dat we met differentiaal net zo mogen rekenen als met breuken is helemaal niet vanzelfsprekend, omdat we het impliciet altijd met limieten te maken hebben. Voor continue functies laat zich aantonen, dat het inderdaad mag, terwijl er gevallen van niet-continue functies zijn, waar het mis gaat. In de wiskunde wordt er veel aandacht aan besteed, precies te beschrijven wanneer het rekenen met differentiaal wel goed gaat en wanneer niet.

Als we de productregel en kettingregel voor het afleiden nu nog eens met behulp van differentiaal opschrijven, zien ze er zo uit:

$$\begin{aligned} \frac{d(fg)}{dx} &= (fg)'(x) = f'(x)g(x) + f(x)g'(x) = \frac{df}{dx}g(x) + f(x)\frac{dg}{dx} \\ \frac{d(g \circ f)}{dx} &= (g \circ f)'(x) = g'(f(x)) \cdot f'(x) = \frac{d(g \circ f)}{df} \cdot \frac{df}{dx} \end{aligned}$$

In de kettingregel lijkt het als of we df in de teller en noemer tegen elkaar kunnen schrappen. Dit weerspiegelt hoe we de kettingregel hebben 'bewezen': Toen hebben we met een voedzame 1 de quotiënt $\frac{(g \circ f)(x+h) - (g \circ f)(x)}{h}$ geschreven als $\frac{g(f(x+h)) - g(f(x))}{f(x+h) - f(x)} \cdot \frac{f(x+h) - f(x)}{h}$ en de limiet hiervan is precies wat we net met differentiaal hebben opgeschreven.

Om aan de notatie met differentiaal te wennen gaan we na een toepassing kijken, namelijk de lengte van een kromme. Stel we hebben een functie $f(x)$ en we willen weten hoe lang de grafiek van deze functie tussen de punten a en b is. Hiervoor delen we het interval in kleine deelintervallen Δx die zo klein zijn dat de grafiek op zo'n interval goed door een rechte lijn benaderd wordt. De lengte van de kromme vinden we nu (zo als in Figuur II.24 geschetst) als de som van de lengten Δs van de lijnstukken op de deelintervallen Δx .



Figuur II.24: Berekening van de lengte van een kromme

Volgens de stelling van Pythagoras is $(\Delta s)^2 = (\Delta f)^2 + (\Delta x)^2$, dus

$$\Delta s = \sqrt{(\Delta f)^2 + (\Delta x)^2} = \Delta x \cdot \sqrt{\left(\frac{\Delta f}{\Delta x}\right)^2 + 1}.$$

Als we nu de lengte van de deelintervallen Δx naar 0 laten gaan, moeten we de Δ 's door differentiaalvoren vervangen, dit geeft

$$ds = dx \cdot \sqrt{\left(\frac{df}{dx}\right)^2 + 1} = \sqrt{f'(x)^2 + 1} dx.$$

Omdat de lengte van de kromme de som over de lijnstukken Δs en dus (in de limiet) de integraal over de differentiaalvoren ds is, volgt voor de lengte s van de grafiek van $f(x)$ tussen a en b :

$$s = \int_a^b ds = \int_a^b \sqrt{f'(x)^2 + 1} dx.$$

Helaas is deze integraal (van wege de wortel) slechts voor weinig functies $f(x)$ makkelijk op te lossen. Een voorbeeld waar het wel lukt is de cirkel, die door de functie $f(x) := \sqrt{1 - x^2}$ beschreven wordt. Als we voor deze functie op het interval $[0, 1]$ de lengte van de grafiek bepalen, zouden we als resultaat de lengte van een kwartcirkel, dus $\frac{\pi}{2}$ verwachten. Er geldt $f'(x) = \frac{1}{2\sqrt{1-x^2}} \cdot 2x = \frac{x}{\sqrt{1-x^2}}$, dus

$$\sqrt{f'(x)^2 + 1} = \sqrt{\frac{x^2}{1-x^2} + 1} = \sqrt{\frac{x^2 + (1-x^2)}{1-x^2}} = \sqrt{\frac{1}{1-x^2}} = \frac{1}{\sqrt{1-x^2}}$$

en hieruit volgt

$$s = \int_0^1 \sqrt{f'(x)^2 + 1} dx = \int_0^1 \frac{1}{\sqrt{1-x^2}} dx = \arcsin(x) \Big|_0^1 = \frac{\pi}{2} - 0 = \frac{\pi}{2}.$$

Maar al voor de gewone parabool $f(x) = x^2$ met $f'(x) = 2x$ wordt het berekenen van de lengte van de grafiek een hele klus waar we al een hele rij van kleine trucjes moeten toepassen. Er geldt:

$$\begin{aligned} \int \sqrt{f'(x)^2 + 1} dx &= \int \sqrt{4x^2 + 1} dx = x\sqrt{4x^2 + 1} - \int x \frac{8x}{2\sqrt{4x^2 + 1}} dx \\ &= x\sqrt{4x^2 + 1} - \int \frac{(4x^2 + 1) - 1}{\sqrt{4x^2 + 1}} dx \\ &= x\sqrt{4x^2 + 1} - \int \sqrt{4x^2 + 1} dx + \int \frac{1}{\sqrt{4x^2 + 1}} dx \\ &= \frac{1}{2} \left(x\sqrt{4x^2 + 1} + \int \frac{1}{\sqrt{4x^2 + 1}} dx \right) \\ &= \frac{1}{2} x\sqrt{4x^2 + 1} + \frac{1}{4} \operatorname{arsinh} 2x = \frac{1}{2} x\sqrt{4x^2 + 1} + \frac{1}{4} \log(2x + \sqrt{4x^2 + 1}) \end{aligned}$$

dus heeft de standaardparabool $f(x) = x^2$ op het interval $[0, x]$ de lengte

$$s = \frac{1}{2} x\sqrt{4x^2 + 1} + \frac{1}{4} \log(2x + \sqrt{4x^2 + 1}).$$

10.2 De substitutieregels

Er zijn twee manieren hoe we van de kettingregel gebruik kunnen maken om de primitieve van een functie te vinden. Bij de eerste (meer directe) manier vervangen we een functie $f(x)$ door een nieuwe variabele u , als we ook $f'(x)$ in de integraal kunnen vinden. Bij de tweede manier gaat het andersom, we vervangen x door een functie $f(u)$ en moeten dan ook $f'(u)$ toevoegen, waardoor soms een eenvoudigere integraal ontstaat.

Type A: Vervangen van $f(x)$ door u

Stel dat $g(x)$ een functie is waarvan we een primitieve $G(x)$ kennen, dus $G'(x) = g(x)$. Als we nu een integraal van de vorm $\int g(f(x)) \cdot f'(x) dx$ vinden, dan kunnen we dit oplossen door naar de functie $F(x) := G(f(x))$ te kijken, want dan geldt

$$F'(x) = G'(f(x)) \cdot f'(x) = g(f(x)) \cdot f'(x), \text{ dus } \int g(f(x)) \cdot f'(x) dx = F(x)$$

d.w.z. $F(x) = G(f(x))$ is een primitieve van $g(f(x)) \cdot f'(x)$.

Met behulp van differentiaalrekenen kunnen we hetzelfde zo schrijven:

$$\int g(f(x)) \cdot f'(x) dx = \int \frac{dG}{df} \cdot \frac{df}{dx} dx = G(f(x)) = F(x).$$

Het lijkt weer als of we de differentiaalrekenen df en dx kunnen schrappen, maar we moeten wel opletten dat we G in $f(x)$ evalueren omdat we G naar f afleiden.

Wat hier eigenlijk gebeurt is het volgende: We definiëren $u := f(x)$, dan is u natuurlijk een functie van x en er geldt $\frac{du}{dx} = u'(x) = f'(x)$ en door vermenigvuldigen met dx vinden we $du = f'(x) dx$. Voor de integraal geldt dus

$$\int g(f(x)) f'(x) dx = \int g(u) du = G(u) = G(f(x)) = F(x).$$

Omdat we $f(x)$ door u hebben vervangen, noemen we dit een *substitutie*. Het recept voor dit type van substitutie is dus:

- (1) Vervang een functie $f(x)$ door u en $f'(x) dx$ door du . De integraal mag nu geen x meer bevatten.
- (2) Vind een primitieve $G(u)$ voor $g(u)$.
- (3) Vervang u in $G(u)$ door $f(x)$ om zo de primitieve $F(x) = G(f(x))$ te vinden.

Tot nu toe hebben we ons met een onbepaalde integraal (zonder grenzen) bemoeid. Wat gebeurt er nu bij de substitutie voor een bepaalde integraal met grenzen a en b ? Als x van a tot b loopt, dan loopt $f(x)$ van $f(a)$ tot $f(b)$, dus loopt u van $f(a)$ tot $f(b)$. We hebben dus

$$\int_a^b g(f(x)) f'(x) dx = \int_{f(a)}^{f(b)} g(u) du = G(u) \Big|_{f(a)}^{f(b)} = (G \circ f)(x) \Big|_a^b = F(x) \Big|_a^b$$

d.w.z. nadat we $g(x)$ voor u terug hebben gesubstitueerd zijn de grenzen weer hetzelfde.

De beste manier, om de substitutie te begrijpen is naar een aantal voorbeelden te kijken. Daarbij komen ook verschillende standaard-trucjes naar voren.

(i) $\int \sin^5(x) \cos(x) dx$:

We kiezen $u = f(x) = \sin(x)$, dan is $du = \cos(x) dx$. Er geldt dus:
 $\int \sin^5(x) \cos(x) dx = \int u^5 du = \frac{1}{6}u^6 = \frac{1}{6} \sin^6(x)$.

(ii) $\int \frac{x}{1+x^2} dx$:

We zien dat de teller (tot op een factor na) de afgeleide van de noemer is, dus kiezen we $u = f(x) = 1 + x^2$, dan is $du = 2x dx$. Dan geldt:
 $\int \frac{x}{1+x^2} dx = \frac{1}{2} \int \frac{2x}{1+x^2} dx = \frac{1}{2} \int \frac{1}{u} du = \frac{1}{2} \log(u) = \frac{1}{2} \log(1 + x^2)$.

(iii) $\int \tan(x) dx$:

Dit is niet helemaal voor de hand liggend, maar omdat $\tan(x) = \frac{\sin(x)}{\cos(x)}$ is, kunnen we $u = f(x) = \cos(x)$ kiezen, dan is $du = -\sin(x) dx$. We hebben dus:
 $\int \tan(x) dx = \int \frac{\sin(x)}{\cos(x)} dx = \int \frac{-1}{u} du = -\log(u) = -\log(\cos(x))$.

(iv) $\int \frac{1}{x \log(x)} dx$:

Ook hier is er niet zo veel keuze, omdat we voor $f(x)$ een functie moeten kiezen, waarvan we ook de afgeleide terug vinden. In dit geval is $u = f(x) = \log(x)$ de voor de hand liggende substitutie, dan is $du = \frac{1}{x} dx$. Er geldt dus:
 $\int \frac{1}{x \log(x)} dx = \int \frac{1}{u} du = \log(u) = \log(\log(x))$.

Een eenvoudige maar belangrijke type van substitutie is het vervangen van een lineaire functie $ax + b$ door u , in het bijzonder het vervangen van een veelvoud ax door u . Voor $u = ax + b$ is $du = a dx$, dus is $\int f(ax + b) dx = \frac{1}{a} \int f(u) du$.

Een voorbeeld hiervoor is de integraal $\int \frac{1}{x^2+c^2} dx$. We weten dat $\int \frac{1}{x^2+1} dx = \arctan(x)$, maar om dit te kunnen gebruiken moeten we eerst nog de c^2 kwijt raken. Er geldt $x^2 + c^2 = c^2((\frac{x}{c})^2 + 1)$, dus is $\int \frac{1}{x^2+c^2} dx = \frac{1}{c^2} \int \frac{1}{(\frac{x}{c})^2+1} dx = \frac{1}{c} \int \frac{\frac{1}{c}}{(\frac{x}{c})^2+1} dx$. Voor de substitutie $u = \frac{x}{c}$ is $du = \frac{1}{c} dx$, dus geldt

$$\int \frac{1}{x^2 + c^2} dx = \frac{1}{c} \int \frac{\frac{1}{c}}{(\frac{x}{c})^2 + 1} dx = \frac{1}{c} \int \frac{1}{u^2 + 1} du = \frac{1}{c} \arctan(u) = \frac{1}{c} \arctan(\frac{x}{c}).$$

Type B: Vervangen van x door $f(u)$

Voor het type A van substitutie hadden we nodig dat we naast een functie $f(x)$ (die we door u vervangen) ook de afgeleide $f'(x)$ in de integraal konden vinden. Dit is natuurlijk een bijzonder mooie situatie, maar helaas vaak niet het geval (of tenminste niet makkelijk te zien).

Bij veel integralen is dus een andere aanpak nodig: In plaats van $f(x)$ door u te substitueren, vervangen we x door een (geschikte) functie $f(u)$. In dit geval

hebben we $x = f(u)$ en $u = f^{-1}(x)$, waarbij f^{-1} de inverse functie van f is. Er geldt $\frac{dx}{du} = x'(u) = f'(u)$ en dus $dx = f'(u) du$. Voor de integraal $\int g(x) dx$ volgt hieruit:

$$\int g(x) dx = \int g(f(u)) \cdot f'(u) du.$$

We gebruiken dus in principe de kettingregel in de omgekeerde richting als bij type A.

Stel nu dat $G(u)$ een primitieve is van $g(f(u)) \cdot f'(u)$, d.w.z. er geldt $G'(u) = g(f(u)) \cdot f'(u)$. Omdat $f(u) = x$, is $u = f^{-1}(x)$ en we zien door differentiëren dat $G(u) = G(f^{-1}(x))$ een primitieve van $g(x)$ is:

$$\begin{aligned} (G(f^{-1}(x)))' &= G'(f^{-1}(x)) \cdot (f^{-1}(x))' = G'(f^{-1}(x)) \cdot \frac{1}{f'(f^{-1}(x))} \\ &= g(f(f^{-1}(x))) f'(f^{-1}(x)) \cdot \frac{1}{f'(f^{-1}(x))} = g(x). \end{aligned}$$

Het recept voor dit type van substitutie is dus:

- (1) Vervang x (overal) door een functie $f(u)$ en dx door $f'(u) du$.
- (2) Vind een primitieve $G(u)$ voor de nieuwe functie in de integraal.
- (3) Vervang u in $G(u)$ door $f^{-1}(x)$ om zo naar de oorspronkelijke integraal terug te komen.

Bij een bepaalde integraal $\int_a^b g(x) dx$ loopt x van a tot b , dus loopt ook $f(u)$ van a tot b en dus moet u van $f^{-1}(a)$ tot $f^{-1}(b)$ lopen. Hier zien we ook dat er een beperking voor de functie $f(u)$ bestaat: deze moet op het interval $[f^{-1}(a), f^{-1}(b)]$ een inverse functie hebben, dus $f(u)$ moet of op $[f^{-1}(a), f^{-1}(b)]$ strikt stijgend of dalend zijn. Voor de bepaalde integraal geldt dus

$$\int_a^b g(x) dx = \int_{f^{-1}(a)}^{f^{-1}(b)} g(f(u)) \cdot f'(u) du.$$

Het zal duidelijk zijn dat we door een minder slimme keuze van $f(u)$ de integraal alleen maar moeilijker maken. De goede keuze van $f(u)$ is een vraag van oefenen, ervaring en soms ook een beetje gokken. Aan de andere kant is er ook vaak niet zo heel veel keuze voor een substitutie, want op een of andere manier moet de afgeleide $f'(u)$ iets in $g(f(u))$ vereenvoudigen.

Voor verschillende typen van integralen zijn er standaard substituties, een paar voorbeelden zullen we nu bekijken.

- (i) $\int \sqrt{1-x^2} dx$:

Bij integralen met een wortel en een kwadraat zijn vaak de 'trigonometrische substituties' van toepassing. Wat hierbij gebruikt wordt, is de relatie $\sin^2(x) + \cos^2(x) = 1$. Hierdoor kunnen we vaak de wortel kwijt raken. We kiezen $x = \sin(u)$, dan is $dx = \cos(u) du$. Omdat $1 - \sin^2(u) = \cos^2(u)$ wordt de integraal nu $\int \sqrt{1-x^2} dx = \int \cos(u) \cos(u) du$ en dit kunnen we

met partiële integratie oplossen. Er geldt $\int \cos^2(u) du = \cos(u) \sin(u) + \int \sin^2(u) du = \cos(u) \sin(u) + \int (1 - \cos^2(u)) du = \frac{1}{2}(\cos(u) \sin(u) + u)$. Uit $x = \sin(u)$ volgt $u = \arcsin(x)$ en met $\cos(u) = \sqrt{1 - \sin^2(u)}$ volgt dus $\int \sqrt{1 - x^2} dx = \frac{1}{2}(x\sqrt{1 - x^2} + \arcsin(x))$.

(ii) $\int \frac{\sqrt{x}}{x+1} dx$:

Hier hebben we niet zo veel keuze, om de wortel kwijt te raken moeten we $x = u^2$ kiezen, dan is $dx = 2u du$. Dit geeft $\int \frac{\sqrt{x}}{x+1} dx = \int \frac{u \cdot 2u}{u^2+1} du = 2 \int \left(\frac{u^2+1-1}{u^2+1}\right) du = 2 \int \left(1 - \frac{1}{u^2+1}\right) du = 2u - 2 \arctan(u)$. Omdat $u = \sqrt{x}$ is dus $\int \frac{\sqrt{x}}{x+1} dx = 2\sqrt{x} - 2 \arctan(\sqrt{x})$.

(iii) $\int \frac{e^{2x}}{\sqrt{e^x+1}} dx$:

Dit is een voorbeeld van een substitutie waar het niet onmiddellijk duidelijk is dat de substitutie een vereenvoudiging geeft. We willen graag weer de wortel kwijt, daarom proberen we $u = \sqrt{e^x + 1}$, dan is omgekeerd $x = \log(u^2 - 1)$ en dus $dx = \frac{2u}{u^2-1} du$. Verder zien we nog dat $e^{2x} = (u^2 - 1)^2$ geldt. We krijgen dus $\int \frac{e^{2x}}{\sqrt{e^x+1}} dx = \int \frac{(u^2-1)^2}{u} \frac{2u}{u^2-1} du = \int 2(u^2 - 1) du = \frac{2}{3}u^3 - 2u = \frac{2}{3}\sqrt{e^x + 1}^3 - 2\sqrt{e^x + 1}$.

We kunnen deze integraal trouwens ook met een substitutie van type A oplossen: Kies $u = f(x) = e^x$, dan is $du = e^x dx$. Er geldt $\int \frac{e^{2x}}{\sqrt{e^x+1}} dx = \int \frac{e^x}{\sqrt{e^x+1}} e^x dx = \int \frac{u}{\sqrt{u+1}} du = \int \frac{(u+1)-1}{\sqrt{u+1}} du = \int \sqrt{u+1} du - \int \frac{1}{\sqrt{u+1}} du = \frac{2}{3}\sqrt{u+1}^3 - 2\sqrt{u+1} = \frac{2}{3}\sqrt{e^x+1}^3 - 2\sqrt{e^x+1}$.

Een toepassing van de substitutie is het bepalen van de integraal van de inverse functie. We zullen de integraal $\int f^{-1}(x) dx$ op twee manieren bepalen, een keer beginnen we met een substitutie en de andere keer met partiële integratie.

Omdat we over de functie $f(x)$ niets weten, is de enige voor de hand liggende substitutie $x = f(u)$, want dan is $f^{-1}(f(u)) = u$. Voor $x = f(u)$ hebben we $dx = f'(u) du$, dus is $\int f^{-1}(x) dx = \int u f'(u) du$. Dit schreeuwt nu naar partiële integratie, namelijk $\int u f'(u) du = u f(u) - \int f(u) du = x f^{-1}(x) - \int f(u) du$. Voor een bepaalde integraal geldt dus:

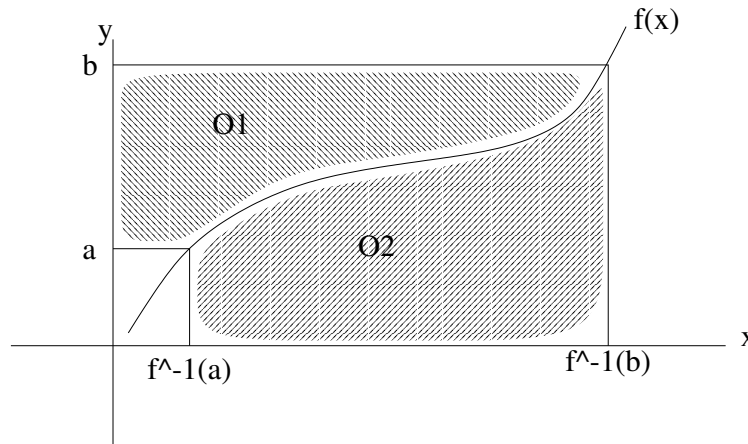
$$\int_a^b f^{-1}(x) dx = x f^{-1}(x) \Big|_a^b - \int_{f^{-1}(a)}^{f^{-1}(b)} f(u) du.$$

We kunnen ook eerst met partiële integratie beginnen. De afgeleide van $f^{-1}(x)$ is $\frac{1}{f'(f^{-1}(x))}$, dus hebben we $\int 1 \cdot f^{-1}(x) dx = x f^{-1}(x) - \int \frac{x}{f'(f^{-1}(x))} dx$. Als we nu $x = f(u)$ substitueren, is weer $dx = f'(u) du$ en dus $\int \frac{x}{f'(f^{-1}(x))} dx = \int \frac{f(u)}{f'(u)} f'(u) du = \int f(u) du$. Dus vinden we dezelfde formule als boven.

Dat dit resultaat inderdaad klopt kunnen we aan de hand van de grafiek in Figuur II.22 aflezen. Door het gebied $O1$ aan de diagonaal te spiegelen zien we dat $O1$ de integraal $\int_a^b f^{-1}(x) dx$ is. Aan de andere kant is het gebied $O2$ de

integraal $\int_{f^{-1}(a)}^{f^{-1}(b)} f(x) dx$. De grote rechthoek heeft oppervlakte $bf^{-1}(b)$ en de kleine rechthoek links onder heeft oppervlakte $af^{-1}(a)$, dus geldt

$$bf^{-1}(b) = af^{-1}(a) + \int_{f^{-1}(a)}^{f^{-1}(b)} f(x) dx + \int_a^b f^{-1}(x) dx.$$



Figuur II.25: Integraal van de inverse functie

10.3 Toepassingen van de integraal

In de vorige les hebben we gezien dat we met behulp van de integraal de oppervlakte onder een grafiek kunnen berekenen en in deze les dat we ook de lengte van een grafiek kunnen vinden.

Een verdere toepassing is het volume van een rotatielichaam. Het idee hierbij is heel eenvoudig: Als we de grafiek van een functie $f(x)$ om de x -as laten roteren, geeft dit een *rotatielichaam*. Net zo als we de oppervlakte onder een grafiek als som van dunne rechthoeken onder de grafiek hebben benaderd, kunnen we het volume van een rotatielichaam door de som van dunne cirkelschijven benaderen. Een cirkelschijf tussen x en $x + \delta$ met straal $f(x)$ heeft het volume $\pi f(x)^2 \cdot \delta$, dus krijgen we het volume van een rotatielichaam als

$$V = \int_a^b \pi f(x)^2 dx.$$

Een eenvoudig voorbeeld is een kogel met straal r . De kogel is het rotatielichaam voor de grafiek van $f(x) = \sqrt{r^2 - x^2}$ tussen $-r$ en r . Het volume van de kogel is dus $\int_{-r}^r \pi \sqrt{r^2 - x^2}^2 dx = \pi \int_{-r}^r (r^2 - x^2) dx = \pi(r^2x - \frac{1}{3}x^3)|_{-r}^r = 2\pi(r^3 - \frac{1}{3}r^3) = \frac{4}{3}\pi r^3$.

Een andere toepassing van de integraal is het oplossen van eenvoudige differentiaalvergelijkingen. Hier zijn twee voorbeelden.

Voorbeeld 1: Bij een zekere chemische reactie is de reactiesnelheid proportioneel met de hoeveelheid van een van de uitgangproducten dat we P noemen. Stel dat na afloop van 3 minuten de helft van P is omgezet. Hoe lang duurt het tot dat 90% van P zijn omgezet?

De aanvankelijke hoeveelheid van P noemen we A , en de hoeveelheid die na x minuten is omgezet noemen we $f = f(x)$. We weten dat op het moment x de reactiesnelheid $f'(x)$ proportioneel met de hoeveelheid $A - f$ van het product P is. Dit geeft $f'(x) = \frac{df}{dx} = k(A - f)$ voor een zekere constante k . Hieruit volgt $dx = \frac{1}{k(A-f)}df$ en door integreren krijgen we

$$x = \int \frac{1}{k(A-f)} df = -\frac{1}{k} \log(A-f) + C.$$

Dit lossen we nu naar $f = f(x)$ op:

$$-k(x - C) = \log(A - f) \Rightarrow A - f = e^{-k(x-C)} \Rightarrow f(x) = A - e^{kC} \cdot e^{-kx}.$$

Omdat $f(0) = 0$ is, geldt $A = e^{kC}$, hieruit volgt

$$f(x) = A(1 - e^{-kx}).$$

Er geldt $f(3) = \frac{1}{2}A$, dus is $\frac{1}{2}A = A(1 - e^{-3k})$ en dus $e^{-3k} = \frac{1}{2}$. Dit geeft $k = \frac{1}{3} \log(2)$. We willen nu x bepalen zo dat $f(x) = \frac{9}{10}$ geldt. Hiervoor moeten we hebben dat $1 - e^{-kx} = \frac{9}{10}$, dus $e^{-kx} = e^{-\frac{1}{3} \log(2)x} = \frac{1}{10}$. Hieruit volgt $\frac{1}{3} \log(2)x = \log(10)$ en dus $x = 3 \frac{\log(10)}{\log(2)} \approx 9.97$.

Voorbeeld 2: Een auto wordt met constante remkracht tot stilstand gebracht. Hoe ver komt de auto nog als zijn beginsnelheid v_0 is?

De snelheid $v = v(t)$ van de auto verandert volgens $v'(t) = \frac{dv}{dt} = -k$ voor de remconstante k . Er geldt dus $dv = -k dt$ en dus $v = \int -k dt = -kt + C$. Omdat $v(0) = v_0$ is, moet $C = v_0$ zijn. We hebben dus $v(t) = v_0 - kt$ en de auto komt tot stilstand voor $v(t) = 0$ dus voor $t_1 = \frac{v_0}{k}$. Om de afstand te bepalen die de auto nog aflegt, moeten we nog eens integreren, om uit de snelheid het traject $s(t)$ te bepalen. Er geldt $v = s'(t) = \frac{ds}{dt}$, dus is $ds = v dt = (v_0 - kt) dt$ en dus $s = \int_0^{t_1} (v_0 - kt) dt = v_0 t - \frac{k}{2} t^2 \Big|_0^{t_1} = v_0 \frac{v_0}{k} - \frac{k}{2} \frac{v_0^2}{k^2} = \frac{v_0^2}{2k}$. Als de beginsnelheid twee keer zo groot is, komt de auto dus nog vier keer zo ver.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- differentialen
- lengte van een kromme
- substitutie (type A en type B)
- toepassingen van de integraal

OPGAVEN

52. Bepaal de volgende integralen door substitutie:

$$(i) \int e^x \sin(e^x) dx, \quad (ii) \int x e^{-x^2} dx, \quad (iii) \int \frac{x}{\sqrt{1-x^4}} dx, \quad (iv) \int \frac{e^{\sqrt{x}}}{\sqrt{x}} dx,$$

$$(v) \int \frac{\log(x)}{x} dx, \quad (vi) \int \log(\cos(x)) \tan(x) dx, \quad (vii) \int \frac{\log(\log(x))}{x \log(x)} dx.$$

53. Bereken de volgende bepaalde integralen:

$$(i) \int_4^8 \frac{x}{\sqrt{x^2-15}} dx, \quad (ii) \int_0^1 \log(x^2+1) dx, \quad (iii) \int_0^{2\pi} \sin\left(\frac{1}{2}x\right) dx.$$

54. Bepaal de lengte van de grafieken voor de volgende functies op de aangegeven intervallen:

(i) $f(x) = \frac{a}{2}(\exp(\frac{x}{a}) - \exp(-\frac{x}{a})) = a \cdot \cosh(\frac{x}{a})$ op het interval $[-a, a]$ (deze kromme heet de *kettinglijn*);

(ii) $f(x) = \log(x)$ op het interval $[1, 3]$ (hint: substitueer $u = \sqrt{x^2+1}$);

(iii) $f(x) = x^2$ op het interval $[0, 2]$ (hint: integreer partieel met 1).

Opgaven voor Calculus

Opgave 1.

Bepaal de afgeleiden van de volgende functies:

- (i) $f(x) := \sin(x+x^2)$, (ii) $f(x) := \sin(x)+\sin(x^2)$, (iii) $f(x) := \sin(\cos(x))$,
 (iv) $f(x) := \sin(\sin(x))$, (v) $f(x) := \sin\left(\frac{\cos(x)}{x}\right)$, (vi) $f(x) := \frac{\sin(\cos(x))}{x}$,
 (vii) $f(x) := \sin(x + \sin(x))$, (viii) $f(x) := \sin(\cos(\sin(x)))$.

Opgave 2.

Als je gewone afgeleiden vervelend vindt, zou je het misschien interessanter vinden om van de volgende functies de afgeleide $f'(x)$ te berekenen:

- (i) $f(x) := \sin((x+1)^2(x+2))$, (ii) $f(x) := \sin^3(x^2 + \sin(x))$,
 (iii) $f(x) := \sin^2((x + \sin(x))^2)$, (iv) $f(x) := \sin\left(\frac{x^3}{\cos(x^3)}\right)$,
 (v) $f(x) := \sin(x \sin(x)) + \sin(\sin(x^2))$, (vi) $f(x) := \sin^2(x) \sin(x^2) \sin^2(x^2)$,
 (vii) $f(x) := (x + \sin^5(x))^6$, (viii) $f(x) := \sin(\sin(\sin(\sin(\sin(x))))))$,
 (ix) $f(x) := \sin((\sin^7(x^7) + 1)^7)$, (x) $f(x) := (((x^2 + x)^3 + x)^4 + x)^5$,
 (xi) $f(x) := \sin(x^2 + \sin(x^2 + \sin(x^2)))$, (xii) $f(x) := \sin(6 \cos(6 \sin(6 \cos(6x))))$,
 (xiii) $f(x) := \frac{\sin(x^2) \sin^2(x)}{1 + \sin(x)}$, (xiv) $f(x) := \sin\left(\frac{x^3}{\sin\left(\frac{x^3}{\sin(x)}\right)}\right)$.

Opgave 3.

Vind lokale en globale minima en maxima voor de volgende functies (die op \mathbb{R} met uitzondering van eventuele nulpunten van de noemer gedefinieerd zijn):

- (i) $f(x) := x^3 - x$, (ii) $f(x) := x^4 - 2x^2$, (iii) $f(x) := \frac{x^2 - 2x + 2}{x - 1}$,
 (iv) $f(x) := x^3 - x^2 - 8x + 1$, (v) $f(x) := x^5 + x + 1$, (vi) $f(x) := 2 + x^{\frac{2}{3}}$,
 (vii) $f(x) := x^{\frac{4}{3}}(1 - x)^{\frac{1}{3}}$, (viii) $f(x) := x^3 + \frac{48}{x}$,

Opgave 4.

Een ellips wordt beschreven door de vergelijking $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. De parameters a en b geven de lengten van de twee halfassen van de ellips. Voor $a = b = 1$ is de ellips een cirkel. We bekijken rechthoeken met zijden evenwijdig aan de x - en y -assen, die in de ellips liggen. Bepaal de afmetingen van de rechthoeken: (a) met maximale oppervlakte, (b) met maximale omvang, die in de ellips passen.

Opgave 5.

Een bal wordt uit een hoogte h_0 met snelheid v_0 verticaal omhoog of omlaag gegooid (als $v_0 > 0$ wordt hij omhoog gegooid, als $v_0 < 0$ omlaag). De hoogte van de bal wordt (afhankelijk van de tijd) door een functie $h(t)$ beschreven, voor de snelheid $v(t)$ geldt $v(t) = h'(t)$ en de acceleratie $a(t) = v'(t) = -g$ is constant (we nemen de acceleratie met een minteken, omdat deze omlaag gericht is, de waarde van g is ongeveer $9.81m/s^2$).

- (i) Toon aan dat $h(t) = -\frac{1}{2}gt^2 + v_0t + h_0$.
- (ii) Wat is de maximale hoogte die de bal bereikt?
- (iii) Wat is de snelheid waarmee de bal de grond raakt?

Opgave 6.

Bepaal primitieven voor de volgende functies:

(i) $f(x) := \frac{a^x}{b^x}$, (ii) $f(x) := \frac{1}{1 + \sin(x)}$, (iii) $f(x) := \frac{1}{\sqrt{a^2 - x^2}}$.

Opgave 7.

Bepaal de volgende integralen (bijvoorbeeld door partiële integratie):

(i) $\int \cos^3(x) dx$, (ii) $\int x^3 \exp(x^2) dx$, (iii) $\int x^2 \sin(x) dx$, (iv) $\int x \log^2(x) dx$.

Opgave 8.

Bepaal de volgende integralen (bijvoorbeeld door substitutie):

(i) $\int \frac{\log(\log(x))}{x} dx$, (ii) $\int \frac{\exp(x)}{\exp(2x) + 2 \exp(x) + 1} dx$,
 (iii) $\int \exp(\exp(x)) \exp(x) dx$, (iv) $\int x \sqrt{1 - x^2} dx$.

Opgave 9.

Bereken de volgende bepaalde integralen:

(i) $\int_0^2 (2 + x) dx$, (ii) $\int_0^2 (2 - x)^2 dx$, (iii) $\int_0^3 (3 - 2x + x^2) dx$,
 (iv) $\int_{-1}^2 x(1 - x^2) dx$, (v) $\int_1^4 \sqrt{x}(1 - x) dx$, (vi) $\int_1^8 \sqrt{1 + 3x} dx$,
 (vii) $\int_0^2 x^2(x^3 + 1) dx$, (viii) $\int_0^3 \frac{1}{\sqrt{1+x}} dx$, (ix) $\int_0^1 x(1 - \sqrt{x})^2 dx$,
 (x) $\int_0^a \sqrt{a^2 - x^2} dx$, (xi) $\int_3^4 \frac{1}{25 - x^2} dx$, (xii) $\int_0^{\frac{\pi}{3}} x^2 \sin(3x) dx$.

Opgave 10.

Bereken de volgende bepaalde integralen:

(i) $\int_3^{11} \sqrt{2x + 3} dx$, (ii) $\int_4^9 \frac{1 - \sqrt{x}}{1 + \sqrt{x}} dx$, (iii) $\int_0^{\sqrt{2}} x^3 \exp(x^2) dx$.

Deel III

Kansrekening

Aanbevolen achtergrondliteratuur met veel opgaven (en oplossingen):

- Murray R. Spiegel, John J. Schiller, R. A. Srinivasan: (Schaum's Outline of Theory and Problems of) Probability and Statistics. McGraw-Hill Companies, 2000, 408 p., ISBN: 0071350047.

Les 11 Combinatoriek

Als we het over de *kans* hebben dat iets gebeurt, hebben we daar wel intuïtief een idee over, wat we hiermee bedoelen. Bijvoorbeeld zeggen we, dat bij het werpen van een munt de kans $\frac{1}{2}$ is, dat de zijde met cijfer (munt) boven te liggen komt, evenzo als de kans voor de koningin (kop) $\frac{1}{2}$ is. Op een soortgelijke manier behandelen we het werpen van een dobbelsteen: de kans voor elke van de getallen 1, 2, 3, 4, 5, 6 is $\frac{1}{6}$, maar we kunnen ook iets over de kans zeggen, dat we een even getal werpen, die is namelijk de som van de kansen voor 2, 4 en 6, dus $\frac{1}{2}$.

Het algemeen principe dat hier achter zit, is dat er een aantal mogelijke uitkomsten is, en we een deel hiervan als *gunstige* uitkomsten aanzien. De relatieve frequentie van gunstige uitkomsten interpreteren we dan als kans voor een gunstige uitkomst.

Principe van de relatieve frequentie: *De kans op een gunstige uitkomst berekenen we als het aantal gunstige uitkomsten gedeeld door het totaal aantal mogelijke uitkomsten.*

Het Simpson paradox

Soms kan zelfs het bepalen van kansen met behulp van relatieve frequenties tot verrassingen leiden. Stel we hebben een fruithandelaar die sinaasappels van minstens 100g per stuk wil verkopen. Hij heeft twee leveranciers, *A* en *B*, van sinaasappels.

In een eerste levering krijgt hij van *A* 110 sinaasappels waarvan er 50 te licht zijn en van *B* 70 sinaasappels waarvan 30 te licht zijn. Op dit moment zou hij ervan uit gaan dat *B* de betere leverancier is, omdat $\frac{5}{11} > \frac{3}{7}$ is.

Een week later krijgt hij van *A* een levering van 90 sinaasappels waarvan 60 te licht zijn en van *B* 140 sinaasappels waarvan 90 te licht zijn. Ook in dit geval is *B* de betere leverancier, want $\frac{6}{9} > \frac{9}{14}$.

Maar als we nu de twee leveringen bij elkaar nemen, waren bij *A* 110 van 200 sinaasappels te licht, terwijl bij *B* 120 van 210 sinaasappels te licht waren. Er geldt $\frac{11}{20} < \frac{12}{21}$, dus is over de twee weken gezien *A* de betere leverancier!

Het probleem is, dat we uit de twee leveringen apart kunnen concluderen dat $\frac{5}{11} + \frac{6}{9} > \frac{3}{7} + \frac{9}{14}$. Maar als we de leveringen gezamenlijk vergelijken, moeten we $\frac{5+6}{11+9}$ met $\frac{3+9}{7+14}$ vergelijken, en dat is niet de som van de breuken.

11.1 Tellen van uitkomsten

Om goed over kansen en kansverdelingen te kunnen praten, moeten we kijken, hoe we bij iets ingewikkeldere problemen dan het werpen van een dobbelsteen gunstige uitkomsten kunnen tellen. De kunst van het tellen van uitkomsten heet *combinatoriek*.

Bij het dobbelen met drie dobbelstenen kunnen we ons afvragen of de kans groter is dat de som van de ogen 11 of 12 is. Hiervoor moeten we 11 of 12 schrijven als sommen van drie getallen uit de verzameling $\{1, 2, 3, 4, 5, 6\}$. We

hebben

$$11 = 6 + 4 + 1 = 6 + 3 + 2 = 5 + 5 + 1 = 5 + 4 + 2 = 5 + 3 + 3 = 4 + 4 + 3$$

$$12 = 6 + 5 + 1 = 6 + 4 + 2 = 6 + 3 + 3 = 5 + 5 + 2 = 5 + 4 + 3 = 4 + 4 + 4$$

dus zijn er in elk geval 6 mogelijkheden en de kans lijkt even groot te zijn. Maar als we dit in een experiment na gaan (bijvoorbeeld met een computersimulatie), zien we dat de kans voor de som 11 ongeveer $P(11) = 0.125$ is en de kans voor de som 12 ongeveer $P(12) = 0.116$, dus kleiner dan die voor de som 11. Wat is hier mis gegaan?

Bij het tellen van de mogelijkheden hebben we alleen maar afstijgende sommen opgeschreven, maar als we even aannemen dat de drie dobbelstenen rood, blauw en groen zijn, is het duidelijk dat er verschillende manieren zijn, hoe we $6 + 4 + 1$ kunnen krijgen. De 6 kan namelijk op elke van de drie dobbelstenen verschijnen en in elk van deze drie gevallen hebben we nog twee mogelijkheden om 4 en 1 op de andere twee dobbelstenen te verdelen. We moeten dus de som $6 + 4 + 1$ zes keer tellen, omdat er zes verschillende manieren zijn hoe we deze som kunnen krijgen. Bij een som met twee verschillende getallen (zo als $5 + 5 + 1$) hebben we drie mogelijkheden en bij drie dezelfde getallen alleen maar eentje. Als we de mogelijkheden voor de som 11 zo bepalen vinden we $3 \cdot 6 + 3 \cdot 3 = 27$ mogelijkheden en voor de som 12 krijgen we $3 \cdot 6 + 2 \cdot 3 + 1 = 25$. Omdat er $6^3 = 216$ mogelijke uitkomsten met drie dobbelstenen zijn, is de kans voor de som 11 dus $\frac{27}{216} = \frac{1}{8}$ en die voor som 12 is $\frac{25}{216}$, en dit is wat we ook experimenteel zouden vinden.

Het belangrijke punt bij dit voorbeeld is, dat we de dobbelstenen kunnen onderscheiden en dat we daarom op de volgorde van de resultaten moeten letten. Het is afhankelijk van het experiment of we inderdaad op de volgorde willen letten of niet. Bijvoorbeeld zijn we bij een kwaliteitscontrole alleen maar geïnteresseerd hoeveel slechte stukken we in een steekproef hebben, maar niet of de eerste of de laatste in de steekproef slecht is.

11.2 Geordende grepen

We gaan eerst na hoe we het aantal uitkomsten berekenen als de volgorde een rol speelt, dus als we het resultaat van de eerste greep en het resultaat van de tweede greep willen onderscheiden. Dit is bijvoorbeeld het geval voor het dobbelen met meerdere dobbelstenen, maar ook voor het toewijzen van nummers aan de spelers van een voetbalploeg.

Hier is een voorbeeld: Stel een exclusieve restaurant biedt een keuze van 4 voorgerechten, 3 hoofdgerechten en 3 desserts. Je mag elke combinatie van de drie gangen kiezen, hoeveel mogelijke menu's kun je dan bestellen? Het is duidelijk dat je $4 \cdot 3 \cdot 3$ mogelijkheden hebt. Algemeen geldt:

Principe van de vermenigvuldiging van uitkomsten: *Het aantal uitkomsten voor een geordende greep is $\prod_{i=1}^r n_i$ als we r keer trekken en er voor de i -de greep n_i mogelijkheden zijn.*

Van dit principe zijn er twee heel belangrijke speciale gevallen, het trekken *met* en het trekken *zonder* terugleggen.

Trekken met terugleggen

Uit een verzameling van n objecten kiezen we r keer een element, waarbij we het getrokken element weer terugleggen. Dan hebben we voor elke keuze n mogelijkheden en het aantal uitkomsten is dus

$$\underbrace{n \cdot n \cdot \dots \cdot n}_r = n^r.$$

Dit is het aantal rijen (a_1, \dots, a_r) met $a_i \in \{1, \dots, n\}$.

Trekken zonder terugleggen

Uit een verzameling van n objecten kiezen we r keer een element, maar een getrokken element wordt niet terug gelegd, dus is er na elke greep een element minder in de verzameling. Voor de eerste greep hebben we dus n mogelijkheden, voor de tweede $n - 1$, voor de derde $n - 2$ enzovoorts. Het aantal uitkomsten is dus

$$n \cdot (n - 1) \cdot \dots \cdot (n - r + 1) = \frac{n!}{(n - r)!}.$$

Dit is het aantal rijen (a_1, \dots, a_r) met $a_i \in \{1, \dots, n\}$ waarbij alle a_i verschillend zijn. In het bijzonder geldt:

Permutaties van n elementen: *Het aantal manieren hoe we de getallen $\{1, \dots, n\}$ kunnen ordenen is gelijk aan $n!$.*

11.3 Ongeordende grepen

Bij veel toepassingen speelt de volgorde geen rol, bijvoorbeeld als we alleen maar geïnteresseerd zijn hoeveel objecten met een bepaalde eigenschap in een steekproef zitten. Als de volgorde geen rol speelt, kunnen we de elementen in de rij van getrokken elementen omordenen en zo ervoor zorgen dat ze in een zekere volgorde zitten. Op die manier zijn de uitkomsten van een ongeordende greep alleen maar de rijen (a_1, \dots, a_r) met $a_i \leq a_{i+1}$.

Merk op: Hier ligt een bron van mogelijke verwarring : Bij een *ongeordende greep* mogen we de elementen *omordenen* en krijgen dan een *geordende* rij.

Ook voor de ongeordende grepen zijn er weer twee mogelijkheden: We kunnen met of zonder terugleggen trekken. Omdat het geval zonder terugleggen eenvoudiger is, gaan we dit eerst bekijken.

Trekken zonder terugleggen

Het misschien meest bekende voorbeeld van een ongeordende greep zonder terugleggen is het trekken van de lottogetallen. Hierbij worden de ballen met de nummers weliswaar achter elkaar getrokken en we kunnen de ballen ook onderscheiden, maar op het eind worden de nummers in opstijgende volgorde gesorteerd, daarom speelt het geen rol in welke volgorde de nummers getrokken werden en de greep is dus ongeordend.

We hebben gezien, dat er $\frac{n!}{(n-r)!}$ mogelijke uitkomsten van een geordende greep zonder terugleggen zijn. Maar van zo'n greep zijn er precies $r!$ permutaties en alleen maar één van deze permutaties heeft de eigenschap dat de elementen opstijgend geordend zijn. Dus is het aantal uitkomsten voor ongeordende grepen zonder terugleggen

$$\frac{1}{r!} \cdot \frac{n!}{(n-r)!} = \frac{n!}{r!(n-r)!} =: \binom{n}{r}.$$

We noemen $\binom{n}{r}$ een *binomiaalcoëfficiënt* en spreken dit 'n over r'. De binomiaalcoëfficiënt $\binom{n}{r}$ geeft aan op hoeveel manieren we een deelverzameling van r elementen uit een verzameling van n elementen kunnen kiezen. Dit is hetzelfde als het aantal rijen (a_1, \dots, a_r) met $a_i \in \{1, \dots, n\}$ en $a_i < a_{i+1}$. Merk op dat de binomiaalcoëfficiënt $\binom{n}{r} = 0$ voor $r > n$, omdat we geen r elementen uit $n < r$ kunnen kiezen.

In het geval van de lottogetallen is $n = 49$ en $r = 6$ (we negeren even extra- en supergetallen), dus is het aantal mogelijke uitkomsten van de lotto $\binom{49}{6} = 13983816$, dus bijna 14 miljoen.

Een andere samenhang waar we de binomiaalcoëfficiënt tegen komen (en waar ook de naam vandaan komt), is bij veeltermen: De (algemene) binomische formule is

$$(a+b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} = a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a b^{n-1} + b^n$$

dus bijvoorbeeld $(a+b)^4 = b^4 + 4ab^3 + 6a^2b^2 + 4a^3b + a^4$. Het is geen toeval dat de binomiaalcoëfficiënt hier naar voren komt: Als we het product $(a+b)^n$ uitschrijven als $(a+b) \cdot (a+b) \cdot \dots \cdot (a+b)$ en dan uitvoerig vermenigvuldigen krijgen we een term $a^r b^{n-r}$ als we in r van de factoren a kiezen en in de $n-r$ andere factoren b . Maar het aantal manieren om de r factoren met a uit de n factoren te kiezen is $\binom{n}{r}$, daarom wordt dit de coëfficiënt van $a^r b^{n-r}$.

We kunnen makkelijk een paar belangrijke eigenschappen van de binomiaalcoëfficiënten afleiden:

$$(i) \quad \binom{n}{r} = \binom{n}{n-r}$$

Dit volgt meteen uit de definitie, omdat we alleen maar de factoren in de noemer omruilen. Maar we kunnen het ook anders inzien: Als we r uit de n elementen van een verzameling hebben gekozen, dan hebben we $n-r$ elementen niet gekozen, dus hoort bij elke deelverzameling van r elementen een eenduidige deelverzameling van $n-r$ elementen, dus is het aantal deelverzamelingen met r elementen gelijk aan het aantal deelverzamelingen met $n-r$ elementen. We noemen dit ook de *symmetrie* van de binomiaalcoëfficiënten.

$$(ii) \quad \sum_{r=0}^n \binom{n}{r} = 2^n$$

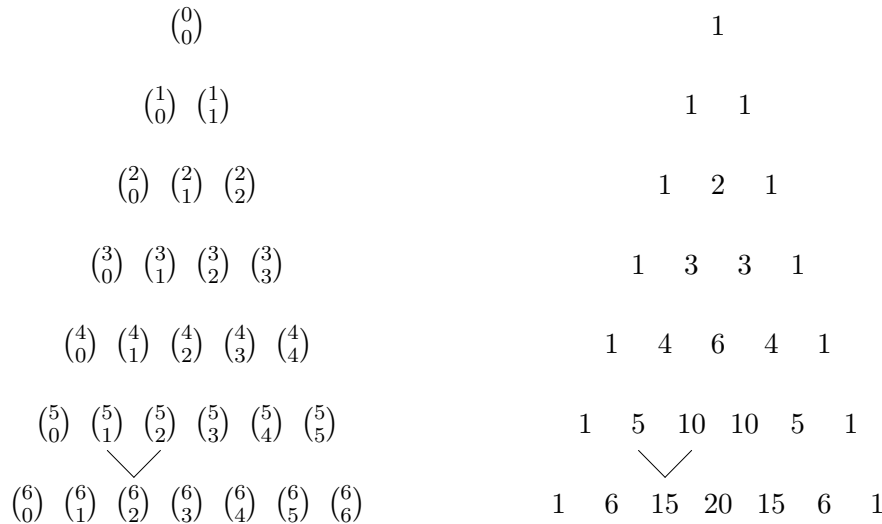
Dit volgt uit de binomische formule als we $a = b = 1$ invullen. Maar we

kunnen dit ook uit het aftellen van deelverzamelingen zien: Een verzameling Ω van n elementen heeft $\binom{n}{r}$ deelverzamelingen met r elementen, dus is de som over de binomiaalcoëfficiënten het aantal van alle deelverzamelingen van Ω . Maar elk element $a \in \Omega$ is of in een deelverzameling $A \subseteq \Omega$ bevat of is er niet in bevat. Dit geeft 2 mogelijkheden voor elk element en dus 2^n mogelijkheden om de uitkomsten $a \in A$ of $a \notin A$ op de n elementen van Ω te verdelen en dus zijn er 2^n deelverzamelingen van Ω .

(iii) $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$

Hiervoor tellen we de $\binom{n+1}{r}$ deelverzamelingen $A \subseteq \{1, \dots, n+1\}$ met r elementen op de volgende manier: Of het element $n+1$ ligt in een deelverzameling A , dan bevat A nog $r-1$ elementen uit de resterende n elementen en er zijn dus $\binom{n}{r-1}$ mogelijkheden voor A . Of het element $n+1$ zit niet in de deelverzameling A , dan zijn de r elementen van A uit de resterende n elementen gekozen en hiervoor zijn er $\binom{n}{r}$ mogelijkheden.

Een handige manier om de binomiaalcoëfficiënten op te schrijven (en uit te rekenen) is de *driehoek van Pascal* die in Figuur III.1 afgebeeld is. In de driehoek van Pascal heeft de eerste rij één element, de tweede heeft twee elementen enz., de n -de rij heeft dus n elementen. Als r -de element in de n -de rij schrijven we de binomiaalcoëfficiënt $\binom{n-1}{r-1}$. Merk op dat $\binom{0}{0} = 1$ omdat $0! = 1$ is. De formule $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$ zegt nu dat we een element op een zekere plek in de driehoek van Pascal krijgen door de twee direct links en rechts boven dit element staande binomiaalcoëfficiënten op te tellen zo als in Figuur III.1 voor het element $\binom{6}{2}$ aangetoond.



Figuur III.1: Driehoek van Pascal

Trekken met terugleggen

Als we na een greep het getrokken element weer terugleggen maar niet op de volgorde letten, willen we het aantal rijen (a_1, \dots, a_r) bepalen met $a_i \in$

$\{1, \dots, n\}$ en $a_i \leq a_{i+1}$. Merk op dat we het aantal van dit soort rijen niet zo makkelijk uit het aantal van geordende rijen kunnen bepalen, omdat het aantal permutaties van een rij met herhalingen ervan afhangt hoeveel elementen hetzelfde zijn.

Maar hier komen we met een trucje en het resultaat voor het trekken zonder terugleggen verder: Stel we hebben een rij (a_1, \dots, a_r) met $a_i \leq a_{i+1}$, dan kunnen we hieruit een rij zonder herhalingen maken door $(i-1)$ bij het element a_i op te tellen. Dit geeft de rij (b_1, \dots, b_r) waarbij $b_i = a_i + i - 1 < a_{i+1} + i = b_{i+1}$. Voor de elementen b_i geldt $1 \leq b_i \leq n + r - 1$, dus hoort deze rij bij een ongeordende greep zonder terugleggen uit $n + r - 1$ elementen.

Omgekeerd kunnen we uit elke rij (b_1, \dots, b_r) met $b_i < b_{i+1}$ door aftrekken van $(i-1)$ van het element b_i een rij (a_1, \dots, a_r) maken met $a_i \leq a_{i+1}$. We zien dus dat er even veel rijen (a_1, \dots, a_r) zijn met $1 \leq a_i \leq n$ en $a_i \leq a_{i+1}$ als er rijen (b_1, \dots, b_r) zijn met $1 \leq b_i \leq n + r - 1$ en $b_i < b_{i+1}$. Maar we hebben gezien dat het aantal van het laatste soort rijen gelijk is aan

$$\binom{n+r-1}{r}$$

dus is dit ook het aantal van ongeordende r -grepen met terugleggen.

We hebben nu vier soorten van grepen gezien, namelijk geordende en ongeordende grepen die we telkens met of zonder terugleggen kunnen bekijken. Dit kunnen we overzichtelijk in een 2×2 -schema beschrijven:

	geordend	ongeordend
met terugleggen	I	III
zonder terugleggen	II	IV

Deze vier gevallen kunnen we als volgt karakteriseren:

- I: Noteer de uitslag van elke greep en leg terug $\Rightarrow n^r$ uitkomsten.
- II: Noteer de uitslag van elke greep en leg niet terug $\Rightarrow \frac{n!}{(n-r)!} = \binom{n}{r} r!$ uitkomsten.
- III: Noteer voor elke $a \in \Omega$ alleen maar het aantal grepen die a opleveren en leg terug $\Rightarrow \binom{n+r-1}{r}$ uitkomsten.
- IV: Noteer voor elke $a \in \Omega$ alleen maar het aantal grepen die a opleveren en leg niet terug $\Rightarrow \binom{n}{r}$ uitkomsten.

Het Verjaardagsparadox

We willen de kans berekenen, dat er in en groep van r mensen twee mensen op dezelfde dag jarig zijn. Als verzameling nemen we de verzameling van verjaardagen, dus $|\Omega| = 365$ (we nemen aan dat niemand op 29 februari jarig is). Voor het aantal mogelijke uitkomsten zijn we in geval I, omdat we de mensen kunnen onderscheiden, dus het aantal is 365^r . Nu gebruiken we een klein trucje: We bepalen de kans van het complement van de gewenste uitkomst, dus

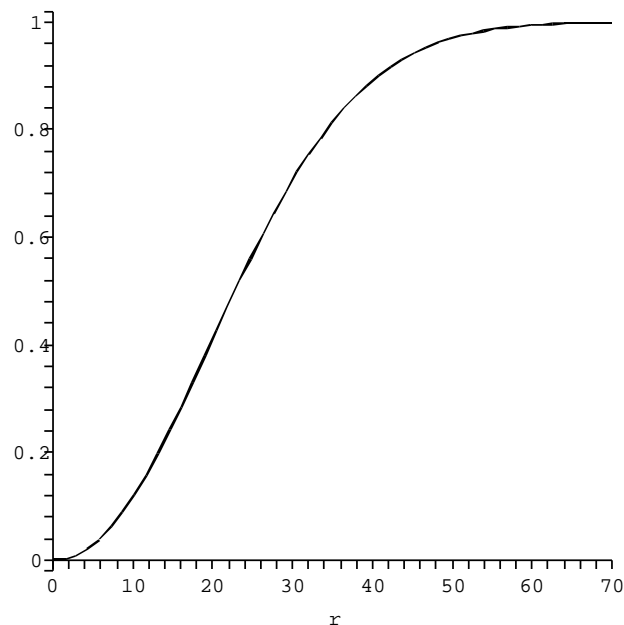
we bepalen de kans dat alle r mensen verschillende verjaardagen hebben. Dan zijn we voor de gunstige uitkomsten in geval II, want een verjaardag van één persoon mag niet meer het verjaardag van een andere persoon zijn. Er zijn dus $\binom{365}{r}r!$ gunstige uitkomsten (d.w.z. alle verjaardagen zijn verschillend). Bij elkaar genomen is de kans dat twee mensen op dezelfde dag jarig zijn dus

$$p = 1 - \frac{\binom{365}{r}r!}{365^r}.$$

Hier zijn een paar waarden van p voor verschillende grootten r van de groep:

$$\begin{array}{ll} r = 2 \Rightarrow p = 0.003, & r = 5 \Rightarrow p = 0.027, \\ r = 10 \Rightarrow p = 0.117, & r = 15 \Rightarrow p = 0.253, \\ r = 20 \Rightarrow p = 0.411, & r = 23 \Rightarrow p = 0.507, \\ r = 25 \Rightarrow p = 0.569, & r = 30 \Rightarrow p = 0.706, \\ r = 50 \Rightarrow p = 0.970, & r = 70 \Rightarrow p = 0.999. \end{array}$$

In Figuur III.2 zie je de functie, die de kans op twee mensen met dezelfde verjaardag afhankelijk van de grootte r van de groep aangeeft. Omdat veel mensen het verrassend vinden dat de kans al voor $r = 23$ groter dan 0.5 is, noemt men dit ook het *verjaardagsparadox*. Er laat zich aantonen dat in het algemeen voor $r \approx \sqrt{n}$ geldt dat r grepen uit n objecten met kans $\frac{1}{2}$ twee dezelfde resultaten opleveren.



Figuur III.2: Kans op dezelfde verjaardag bij r mensen

BELANGRIJKE BEGRIPPEN IN DEZE LES

- relatieve frequentie
- permutaties van n elementen
- geordende en ongeordende grepen
- grepen met en zonder terugleggen
- binomiaalcoëfficiënt
- verjaardagsparadox

OPGAVEN

55. We dobbelen met twee dobbelstenen. Bepaal de kansen voor de volgende uitkomsten:

- (i) De som van de twee getallen is 5.
- (ii) Beide dobbelstenen tonen een oneven getal.
- (iii) De eerste dobbelsteen toont een kleiner getal dan de tweede.
- (iv) De som van de twee getallen is even.
- (v) De som van de twee getallen is minstens 4.
- (vi) De som van de twee getallen is of even of minstens 4 (of allebei).

De absolute waarde van het verschil van de twee getallen ligt tussen 0 en 5. Geef de kansverdeling $P(k)$ aan dat bij een worp met twee dobbelstenen de absolute waarde van het verschil precies k is.

56. Bij het *Poker* spel krijg je 5 kaarten uit een kaartspel met 52 kaarten. Verschillende combinaties van kaarten hebben een bijzondere waarde:

- (i) tweeling: twee kaarten van dezelfde soort (bijvoorbeeld twee boeren),
- (ii) dubbele tweeling: twee verschillende tweelingen (bijvoorbeeld twee vrouwen en twee azen),
- (iii) drieling: drie kaarten van dezelfde soort,
- (iv) vierling: vier kaarten van dezelfde soort,
- (v) full house: een tweeling en een drieling,
- (vi) straight: vijf kaarten in de goede volgorde (bijvoorbeeld 9, 10, boer, vrouw, heer),
- (vii) straight flush: een straight van dezelfde kleur.

Bepaal voor elke van deze combinaties de kans en breng de combinaties hierdoor in een volgorde van opstijgende waarde.

57. Bij het skaat spel krijg je 10 kaarten uit een kaartspel met 32 kaarten. Om een solo te spelen wil je van één kleur veel kaarten hebben en een andere kleur helemaal niet. We vereenvoudigen het probleem door de boeren als gewone kaarten te tellen. Hoe groot is de kans dat je een blad krijgt waarbij je van elke kleur minstens 2 kaarten hebt (zo iets is meestal een slecht blad)?

Als je een solo speelt, krijg je er nog 2 kaarten bij en leg je aansluitend weer 2 van je 12 kaarten weg. Hoe groot is de kans, dat je nu alleen nog met 2 kleuren zit (wat meestal een goed blad is)?

58. Een groep van 18 personen verdeeld zich in een restaurant over drie tafels van 4, 6 en 8 plaatsen. Hoeveel verschillende arrangements zijn er, als de plaatsing aan een tafel geen rol speelt?
59. Je spreekt met een vriend af om op de volgende dag in de rij te staan om kaarten voor Bruce Springsteen (of AC/DC of David Helfgott) te kopen. Op een gegeven moment staan jullie allebei in de rij, maar hebben elkaar niet gezien. Hoe groot is de kans, dat in een rij van n mensen precies r mensen tussen jullie staan? Hoe groot is de kans dat jullie elkaar kunnen zien als er 1000 mensen in de rij staan en je aanneemt dat je je vriend onder de 100 mensen naast je kunt herkennen?

Les 12 Kansverdelingen

We hebben in het begin gesteld dat we de kans voor een zekere gunstige uitkomst berekenen als het aantal gunstige uitkomsten gedeelt door het totaal aantal mogelijke uitkomsten. Maar vaak is het handig, dat we verschillende uitkomsten samenvatten en dit als een nieuwe soort uitkomst bekijken. Bijvoorbeeld kunnen we bij het werpen van twee dobbelstenen de som van de twee geworpen getallen als uitkomst nemen. Als we met $P(s)$ de kans op de som s noteren, zien we (door de mogelijke gevallen na te gaan) makkelijk in, dat $P(2) = 1/36, P(3) = 2/36, P(4) = 3/36, P(5) = 4/36, P(6) = 5/36, P(7) = 6/36, P(8) = 5/36, P(9) = 4/36, P(10) = 3/36, P(11) = 2/36, P(12) = 1/36$. Hieruit laat zich bijvoorbeeld snel aflezen, dat de kans op het dobbelen van een som die een priemgetal is, gelijk is aan $(1 + 2 + 4 + 6 + 2)/36 = 5/12$.

Om ook voor dit soort algemenere situaties over kansen te kunnen praten, hebben we een algemener begrip dan de relatieve frequenties nodig, namelijk het begrip van een *kansverdeling*, waarvan de relatieve frequenties een belangrijk speciaal geval zijn.

Het algemeen principe van een kansverdeling is nog altijd redelijk voor de hand liggend, we eisen alleen maar eigenschappen die heel natuurlijk zijn:

Zij Ω de verzameling van mogelijke uitkomsten. We willen nu graag aan elke deelverzameling $A \subseteq \Omega$ een kans $P(A)$ toewijzen. Hiervoor hebben we een functie

$$P : \mathcal{P}(\Omega) := \{A \subseteq \Omega\} \rightarrow \mathbb{R}$$

nodig, die op de *machtsverzameling* van Ω , d.w.z. de verzameling van alle deelverzamelingen van Ω , gedefinieerd is. We noemen zo'n functie $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ een *kansverdeling* als P aan de volgende eisen voldoet:

- (i) $P(A) \geq 0$ voor alle $A \subseteq \Omega$,
- (ii) $P(\Omega) = 1$,
- (iii) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.

De eerste eigenschap zegt alleen maar, dat kansen niet negatief mogen zijn, en de tweede eigenschap beweert, dat alle mogelijke uitkomsten inderdaad in Ω liggen. De derde eigenschap is een soort van additiviteit, die zegt dat we de kansen voor uitkomsten die niet overlappen (dus niets met elkaar te maken hebben) gewoon mogen optellen. We hadden in principe ook nog kunnen eisen, dat $P(A) \leq 1$ is voor alle $A \subseteq \Omega$, maar dit kunnen we inderdaad uit (i)-(iii) afleiden en willen graag zo zuinig als mogelijk met onze eisen zijn.

12.1 Discrete kansverdelingen

We hebben tot nu toe alleen maar naar voorbeelden gekeken, waarbij de verzameling Ω van mogelijke uitkomsten eindig is. In deze situatie spreken we van *discrete* kansverdelingen, in tegenstelling tot *continue* kansverdelingen die we in de volgende paragraaf gaan behandelen.

Een belangrijk voorbeeld van een discrete kansverdeling hebben we al gezien, namelijk de *gelijkverdeling* die vaak ook *Laplace-verdeling* heet: Elke mogelijke uitkomst $w \in \Omega$ moet dezelfde kans hebben (vandaar de naam), dan is $P(w) = \frac{1}{|\Omega|}$ voor elke $w \in \Omega$. Hieruit volgt met eigenschap (iii) dat $P(A) = \frac{|A|}{|\Omega|}$ en dit is precies de relatieve frequentie.

We gaan nu een aantal voorbeelden bekijken waarin we het tellen van uitkomsten toepassen en daarbij verschillende belangrijke discrete kansverdelingen tegen komen.

Voorbeeld 1: Bij de lotto 6 uit 49 worden uit een vaas met 49 ballen 6 ballen getrokken en vervolgens in opstijgende volgorde gebracht. Omdat de volgorde hier geen rol speelt en zonder terugleggen getrokken wordt, zijn we in het geval *IV* (volgens de lijst uit de vorige les). Het aantal mogelijke uitkomsten is dus $\binom{49}{6}$. We willen nu de kans bepalen dat we bij ons 6 kruisjes k goede getallen hebben waarbij $0 \leq k \leq 6$. De k goede getallen kunnen we op $\binom{6}{k}$ manieren uit de 6 juiste getallen kiezen. Maar ook voor de verkeerd aangekruisde getallen moeten we nog iets zeggen, want we willen *precies* k goede getallen hebben, dus mogen we niet per ongeluk nog een verder goed getal krijgen. We moeten dus onze $6 - k$ resterende getallen uit de $49 - 6 = 43$ verkeerde getallen kiezen en hiervoor zijn er $\binom{43}{6-k}$ mogelijkheden. Het aantal manieren hoe we precies k goede getallen kunnen kiezen is dus $\binom{6}{k} \cdot \binom{43}{6-k}$ en de kans op k goede getallen is dus

$$\frac{\binom{6}{k} \cdot \binom{43}{6-k}}{\binom{49}{6}}.$$

De waarden voor deze kansen zijn:

$k = 0$:	43.6%	(1 in 2.3)
$k = 1$:	41.3%	(1 in 2.4)
$k = 2$:	13.2%	(1 in 7.6)
$k = 3$:	1.8%	(1 in 57)
$k = 4$:	0.1%	(1 in 1032)
$k = 5$:	0.002%	(1 in 54201)
$k = 6$:	0.000007%	(1 in 13983816)

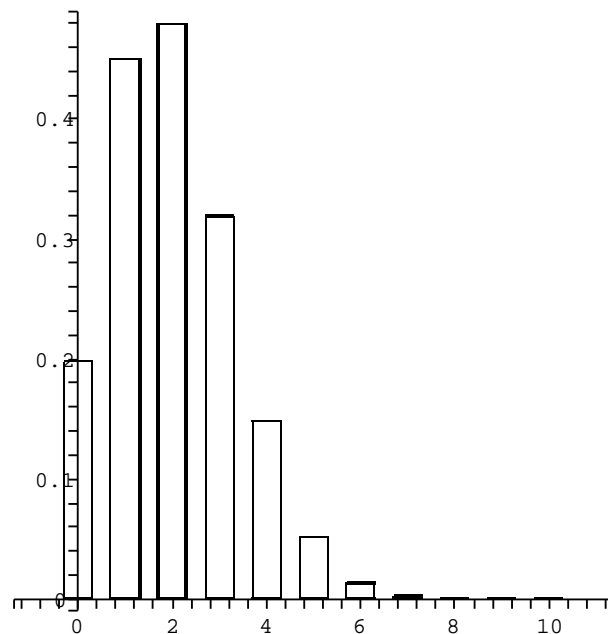
Voorbeeld 2: Bij een kwaliteitstoets kiezen we uit een levering van n stukken een steekproef van m stukken die we testen en niet terugleggen. Dit is bijvoorbeeld het geval als de test het object beschadigt, zo als bij het testen van lucifers. We nemen aan dat de levering s slechte stukken bevat en willen de kans berekenen, dat we in onze steekproef k slechte stukken vinden. Omdat we alleen maar in het aantal slechte stukken geïnteresseerd zijn, maar niet of de eerste of laatste slecht zijn, zijn we weer in het geval *IV*. We kunnen de kans nu net als in het voorbeeld van de lotto berekenen: Er zijn $\binom{s}{k}$ mogelijkheden om k slechte uit de s slechte stukken te vissen, dan zijn er $\binom{n-s}{m-k}$ mogelijkheden om nog $m - k$ goede stukken te kiezen en het totale aantal van mogelijke grepen

is $\binom{n}{m}$. De kans, om k slechte te vinden is dus

$$h(n, m, s; k) := \frac{\binom{s}{k} \cdot \binom{n-s}{m-k}}{\binom{n}{m}}.$$

Omdat dit zo'n belangrijk geval is, heeft deze kansverdeling een eigen naam, ze heet de *hypergeometrische verdeling*.

Ook de kansverdeling die we in Voorbeeld 1 hebben bekeken, is een hypergeometrische kansverdeling, namelijk $h(49, 6, 6; k)$. Figuur III.3 laat een histogram voor de hypergeometrische verdeling $h(1000, 100, 20; k)$ zien: Bij een levering van 1000 stukken, waarvan 2% slecht is maken we een steekproef van 100 stuk en kijken, met welke kans we k slechte stukken vinden. Zo als men dat misschien zou verwachten, is de kans bij $k = 2$ maximaal.



Figuur III.3: Hypergeometrische verdeling $h(1000, 100, 20; k)$

De praktijk van een kwaliteitstoets ziet er natuurlijk eigenlijk iets anders eruit: We weten niet hoeveel slechte stukken er in de levering zitten, maar de leverancier beweert dat het er minder dan s_0 zijn. Wij kennen wel de waarden n , m en k en schatten nu de waarde \hat{s} van s zo dat $h(n, m, \hat{s}; k)$ maximaal wordt. Als onze schatting \hat{s} groter dan s_0 is, zullen we de levering waarschijnlijk niet accepteren.

Een andere toepassing van dit soort schatting vinden we in de ecologie. Als we het aantal vissen in een vijver willen bepalen, kunnen we een

aantal s van vissen markeren en op de volgende dag het aantal k van gemarkeerde vissen in een greep van m vissen bepalen. We schatten dan het aantal \hat{n} van vissen in de vijver zo dat $h(\hat{n}, m, s; k)$ maximaal wordt.

Stel we markeren 1000 vissen en vangen op de volgende dag ook 1000 vissen, waaronder we 100 gemarkeerde vissen vinden. We weten nu dat er minstens nog 900 gemarkeerde vissen in de vijver zitten, dus is $n \geq 1900$. Maar $h(1900, 1000, 1000; 100) \approx 5 \cdot 10^{-430}$, dus deze kans is heel erg klein. Evenzo is de kans op een miljoen vissen heel klein, namelijk $h(10^6, 1000, 1000; 100) \approx 2 \cdot 10^{-163}$. We vinden de maximale waarde voor $\hat{n} = 10000$ en nemen daarom aan dat er ongeveer 10000 vissen in de vijver zijn. Zo'n soort schatting noemen we een *maximum likelihood* schatting, omdat we de parameter n zo kiezen dat de kans $h(n, m, s; k)$ maximaal wordt.

Voorbeeld 3: Als we een kwaliteitstoets uitvoeren waarbij de stukken niet beschadigt worden en we misschien ook iets heel kostbaars testen (bijvoorbeeld het gewicht van een staaf goud) zullen we getoetste stukken waarschijnlijk weer terugleggen. Dan zijn we niet meer in het geval *IV* maar moeten de kans op een andere manier bepalen. We letten nu wel op de volgorde en zijn dus in het geval *I*. Er zijn s^k manieren om k slechte uit de s slechte stukken te kiezen en er zijn $(n-s)^{m-k}$ manieren om $m-k$ goede uit de $n-s$ goede stukken te kiezen. Maar omdat de goede niet van de slechte stukken gescheiden zijn moeten we ook nog tellen hoe we de k slechte stukken op de m grepen kunnen verdelen. Hiervoor zijn er $\binom{m}{k}$ mogelijkheden. Als we de relatieve frequentie van slechte stukken $p := \frac{s}{n}$ noemen vinden we dus voor de kans om k slechte stukken te kiezen:

$$\begin{aligned} b(n, m, s; k) &:= \frac{\binom{m}{k} s^k (n-s)^{m-k}}{n^m} = \binom{m}{k} \frac{s^k}{n^k} \frac{(n-s)^{m-k}}{n^{m-k}} \\ &= \binom{m}{k} p^k (1-p)^{m-k} =: b(m, p; k). \end{aligned}$$

Ook deze kansverdeling is heel fundamenteel en heet de *binomiale verdeling*. We kunnen de binomiale verdeling ook iets anders interpreteren: Stel bij een experiment hebben we iedere keer een kans van p op succes. Dan geeft $b(m, p; k)$ de kans aan dat we bij m pogingen k successen boeken. Merk op dat er bij deze interpretatie geen verzameling van n stukken meer is, waaruit we een steekproef nemen, maar dat we van begin af met de kans p op een succes werken.

Intuïtief zullen we zeggen, dat het voor het geval dat n veel groter is dan m bijna geen verschil maakt of we met of zonder terugleggen trekken, want de kans dat we een element twee keer pakken is heel klein. Er laat zich inderdaad zuiver aantonen, dat voor $n \gg m$ de hypergeometrische verdeling meer en meer op de binomiale verdeling lijkt en in de limiet geldt

$$\lim_{n \rightarrow \infty} h(n, m, np; k) = b(m, p; k).$$

Deze samenhang tussen hypergeometrische en binomiale verdeling wordt meestal de *binomiale benadering* van de hypergeometrische verdeling genoemd. Merk

op dat de binomiale verdeling (behalve van de grootte m van de greep) alleen maar van één parameter afhangt, namelijk het relatieve aantal $p = \frac{s}{n}$ van slechte stukken, terwijl de hypergeometrische verdeling van het totaal aantal n van stukken en het aantal s van slechte stukken afhangt. Dit maakt het natuurlijk veel handiger om met de binomiale verdeling te werken, vooral als je bedenkt dat deze functies vaak in de vorm van tabellen aangegeven worden.

Er laat zich geen algemene regel aangeven, wanneer de binomiale benadering goed genoeg is. Soms leest men iets van $n > 2000$ en $\frac{m}{n} < 0.1$, maar in sommige gevallen heeft de benadering dan al een behoorlijke afwijking. Voor $n = 2000$, $m = 100$, $s = 20$ en $k = 2$ hebben we bijvoorbeeld $h(2000, 100, 20; 2) = 18.95\%$ en de binomiale benadering geeft in dit geval $b(100, \frac{20}{2000}; 2) = 18.49\%$ wat al een tamelijke afwijking is. Als we aan de andere kant naar de kans op 2 goede getallen in de lotto kijken, hebben we $h(49, 6, 6; 2) = 13.24\%$. De binomiale benadering hiervan is $b(6, \frac{6}{49}; 2) = 13.34\%$ en dit is een redelijke benadering terwijl we hier niet aan het criterium voldoen.

De Poisson-verdeling

Vaak willen we bij experimenten de kans weten, dat er bij m pogingen k keer een bepaalde uitkomst plaats vindt. We hebben gezien dat we dit met de binomiale verdeling kunnen beschrijven: Als de kans voor een gunstige uitkomst p is, dan is $b(m, p; k) := \binom{m}{k} p^k (1 - p)^{m-k}$ de kans op k gunstige uitkomsten bij m pogingen.

Voor heel zeldzame gebeurtenissen zullen we verwachten dat er veel pogingen nodig zijn tot dat een gunstige uitkomst optreedt en als de kans p maar nog half zo groot is, zullen we verwachten twee keer zo vaak te moeten proberen. Om voor gebeurtenissen waar p tegen 0 loopt nog een gunstige uitkomst te kunnen verwachten, moeten we dus m zo laten groeien dat $m \cdot p = \lambda$ constant blijft. De constante λ geeft aan hoeveel gunstige uitkomsten we bij m pogingen eigenlijk verwachten.

De vraag is nu wat er met de binomiale verdeling $b(m, p; k)$ gebeurt als we de limiet $p \rightarrow 0$, $m \rightarrow \infty$ bekijken met $p \cdot m = \lambda$. We hebben

$$\begin{aligned} \binom{m}{k} p^k (1 - p)^{m-k} &= \frac{m!}{k!(m-k)!} \frac{\lambda^k}{m^k} \left(1 - \frac{\lambda}{m}\right)^{m-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{m}\right)^m \left(\frac{m}{m} \cdot \frac{m-1}{m} \cdot \dots \cdot \frac{m-k+1}{m}\right) \left(1 - \frac{\lambda}{m}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

omdat $\left(1 - \frac{\lambda}{m}\right)^m \rightarrow e^{-\lambda}$ voor $m \rightarrow \infty$ en $\frac{m-k+1}{m} \rightarrow 1$ en $\left(1 - \frac{\lambda}{m}\right)^{-k} \rightarrow 1$ voor $m \rightarrow \infty$.

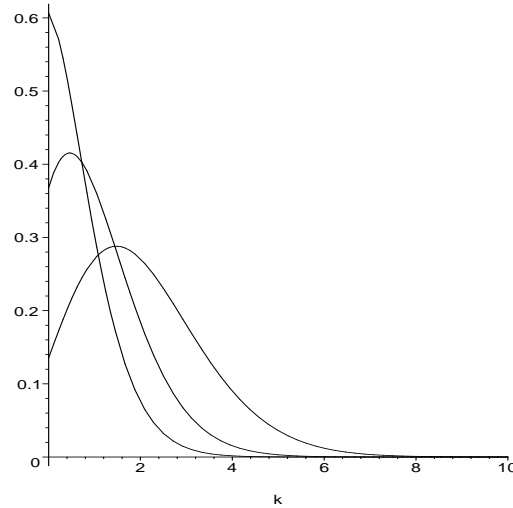
Voor zeldzame gebeurtenissen gaat de binomiale verdeling dus in de limiet tegen de *Poisson-verdeling*

$$P(k) = p_{o\lambda}(k) := \frac{\lambda^k}{k!} e^{-\lambda}.$$

Merk op dat bij de binomiale verdeling het aantal gunstige uitkomsten natuurlijk door het aantal pogingen begrensd is. In de Poisson-verdeling is de enige parameter het aantal verwachte successen λ en we kunnen dus met deze verdeling de kans voor elk aantal gunstige uitkomsten berekenen.

Hoe goed de Poisson-verdeling de binomiale verdeling benadert hangt natuurlijk van de parameters af. Als een vuistregel geldt, dat men de Poisson-benadering mag gebruiken als $p < 0.1$ en $\lambda \leq 5$ of $\lambda \leq 10$, maar hierbij speelt natuurlijk ook weer de benodigde nauwkeurigheid een rol.

De afhankelijkheid van de Poisson-verdeling van de parameter λ kunnen we in Figuur III.4 zien, waar de Poisson-verdelingen voor de parameters $\lambda = 0.5, 1, 2$ als continue functies van k getekend zijn. De kansen worden alleen maar op de punten $k \in \mathbb{N}$ afgelezen.



Figuur III.4: Poisson-verdelingen voor parameters $\lambda = 0.5, 1$, en 2

Omdat $\lim_{k \rightarrow 0} \frac{\lambda^k}{k!} = 1$ is, heeft de Poisson-verdeling in 0 de waarde $e^{-\lambda}$ en we zien dat voor kleinere waarden van λ de grafiek bij een hogere waarde voor $k = 0$ begint maar dan sneller naar 0 toe gaat. Dit klopt ook met onze intuïtie, want als de kans voor een zeldzaam gebeurtenis minder groot is, verwachten we met een hogere waarschijnlijkheid dat het helemaal niet gebeurt. In het plaatje hoort dus de grafiek die bij $e^{-0.5} \approx 0.61$ begint bij de parameter $\lambda = 0.5$, de grafiek die bij $e^{-1} \approx 0.37$ begint hoort bij de parameter $\lambda = 1$, en de grafiek die bij $e^{-2} \approx 0.14$ begint hoort bij de parameter $\lambda = 2$.

Het maximum van de continue Poisson-verdeling laat zich alleen maar door een ingewikkelde functie (de Ψ -functie) beschrijven, voor $\lambda = 1$ ligt het ongeveer bij 0.46 en voor $\lambda = 2$ bij 1.48 . Voor kleine waarden van λ is de grafiek van de Poisson-verdeling dalend, een maximum bestaat alleen maar voor waarden $\lambda \gtrsim 0.562$.

De maximale waarde van de Poisson-verdeling voor $k \in \mathbb{N}$ laat zich wel berekenen. We hebben $\frac{po_{\lambda}(k+1)}{po_{\lambda}(k)} = \frac{\lambda^{k+1}}{(k+1)!} \cdot \frac{k!}{\lambda^k} = \frac{\lambda}{k+1}$. Dit toont aan dat de waarden van po_{λ} voor $k \leq \lambda$ groeien en dan weer dalen. De maximale waarde is bereikt voor het grootste gehele getal $\leq \lambda$. Als λ zelf een geheel getal is, zijn

de waarden voor $k = \lambda - 1$ en $k = \lambda$ hetzelfde.

De Poisson-verdeling is altijd van belang als het erom gaat zeldzame gebeurtenissen te beschrijven. Voorbeelden hiervoor zijn:

- Gevallen met een heel hoge schade voor verzekeringsmaatschappijen.
- Het uitzenden van α -deeltjes door een radioactief preparaat.
- Het aantal drukfouten op een bladzijde.

We kijken naar een voorbeeld: We dobbelen met vier dobbelstenen, dan is de kans om vier 6en te hebben gelijk aan $\frac{1}{6^4}$. Als we nu 1000 keer dobbelen is de parameter $\lambda = m \cdot p = \frac{1000}{1296} \approx 0.77$. De kans om bij de 1000 werpen geen enkele keer vier zessen te hebben is dus $e^{-\lambda} \approx 0.46$, de kans dat het een keer gebeurd is $\lambda e^{-\lambda} \approx 0.36$, de kans op twee keer zo'n werp is $\frac{\lambda^2}{2} e^{-\lambda} \approx 0.14$. De kans op drie of meer keer vier zessen is ongeveer 4.3%.

Merk op dat we altijd het aantal m van grepen kennen en de parameter λ kunnen uitrekenen als we de kans p van gunstige uitkomsten kennen. Vaak komen we in de praktijk het omgedraaide probleem tegen: We kennen het aantal k van gunstige uitkomsten bij een aantal m van pogingen. Hieruit willen we nu de kans p op een gunstige uitkomst schatten. Hiervoor kiezen we de parameter λ zo dat de bijhorende Poisson-verdeling een maximale waarde in k heeft. Dit is weer een *maximum likelihood* schatting.

12.2 Continue kansverdelingen

We hebben tot nu toe alleen maar naar eindige uitkomstenruimten Ω gekeken, d.w.z. naar uitkomstenruimten met $|\Omega| = n < \infty$. Met analoge technieken laten zich ook kansverdelingen op oneindige maar aftelbare ruimten Ω definiëren, d.w.z. op ruimten Ω die in bijectie zijn met de natuurlijke getallen \mathbb{N} . Zo'n bijectie geeft gewoon nummers aan de elementen en we krijgen $\Omega = \{\omega_1, \omega_2, \dots\} = \{\omega_i \mid i \in \mathbb{N}\}$. Door ω_i door het gewone getal i te vervangen kunnen we elke aftelbare ruimte Ω tot de natuurlijke getallen \mathbb{N} terugbrengen en we hoeven dus bij aftelbaar oneindige uitkomstenruimten alleen maar aan de natuurlijke getallen te denken.

De normering $P(\Omega) = 1$ van de kansverdeling komt in dit geval neer op een uitspraak over een oneindige reeks, namelijk $\sum_{i=0}^{\infty} P(i) = 1$. Ook kansverdelingen voor aftelbare uitkomstenruimten noemen we nog *discrete kansverdelingen* omdat we de punten van de natuurlijke getallen als gescheiden punten op de reële lijn beschouwen.

Vaak hebben experimenten echter helemaal geen discrete uitkomsten. Als we bijvoorbeeld naar de wachttijd kijken die we als klant in een rij doorbrengen voordat we geholpen worden, kan de uitkomst een willekeurige tijd t zijn (met misschien een zekere bovengrens). Net zo kunnen we bij een test van het invloed van doping-middelen op de prestatie van kogelstoters willekeurige waarden tussen $10m$ en $25m$ verwachten. In dit voorbeeld leert onze ervaring al een

mogelijke oplossing, hoe we naar discrete uitkomsten terug komen. De prestaties worden namelijk alleen maar tot op centimeters nauwkeurig aangegeven en we vatten dus alle waarden in een zeker interval tot een enkele uitkomst samen.

Maar we kunnen ook kansverdelingen met continue uitkomsten beschrijven. Om het idee hiervan nader toe te lichten, bekijken we als voorbeeld een bekend kansspel, het *Rad van avontuur*. Zo'n rad is in een aantal (even grote) segmenten ingedeeld en op sommige van de segmenten maak je een winst als het rad op dit segment stopt. Als we er n segmenten hebben noemen we deze $1, \dots, n$ en voor elke k met $1 \leq k \leq n$ is de kans dat het rad in het k -de segment stopt gelijk aan $\frac{1}{n}$ (we gaan van een eerlijk rad uit). Maar we kunnen de uitslag dat het rad in het k -de segment stopt ook anders beschrijven, namelijk met behulp van de hoek φ waarop het rad stopt. We hebben namelijk de uitkomst k als voor de hoek φ geldt dat $(k-1)\frac{2\pi}{n} \leq \varphi \leq k\frac{2\pi}{n}$.

Als we nu naar de kans kijken dat het rad van avontuur tussen de hoeken φ_1 en φ_2 stopt dan is deze kans $\frac{\varphi_2 - \varphi_1}{2\pi}$ omdat dit het aandeel van de rand is die tussen de hoeken ligt.

We gaan nu van het rad van avontuur naar het dartspeel over. Ook hier is de kans om een pijltje tussen de hoeken φ_1 en φ_2 te plaatsen gelijk aan $\frac{\varphi_2 - \varphi_1}{2\pi}$, maar dit geldt nu alleen maar omdat de dart schijf een cirkel is. Als we een schijf hebben die niet rond is maar waarvan de straal afhangt van de hoek, dan geven we de straal met een functie $r(\varphi)$ aan. Een segment met een hoek van $\Delta\varphi$ van een cirkel met straal r heeft een oppervlakte van $\frac{\Delta\varphi}{2\pi}\pi r^2 = \frac{1}{2}r^2\Delta\varphi$, dus kunnen we de totale oppervlakte van de schijf door de integraal

$$O = \int_0^{2\pi} \frac{1}{2} r(\varphi) d\varphi$$

berekenen en krijgen de oppervlakte van het segment tussen φ_1 en φ_2 als

$$S = \frac{1}{2} \int_{\varphi_1}^{\varphi_2} r(\varphi)^2 d\varphi.$$

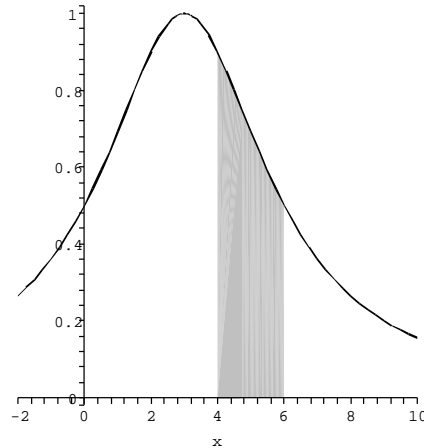
De kans dat een dart-pijltje (bij een toevallige verdeling over de schijf) in het segment tussen φ_1 en φ_2 terecht komen is het aandeel van het segment aan de totale oppervlakte, dus de integraal

$$P(\varphi_1, \varphi_2) = \frac{S}{O} = \frac{1}{2O} \int_{\varphi_1}^{\varphi_2} r(\varphi)^2 d\varphi.$$

Na deze voorbeelden kunnen we het algemene idee voor continue kansverdelingen makkelijk inzien: We beschrijven de kans dat de uitkomst x van een experiment in het interval $[a, b]$ valt als oppervlakte onder de grafiek van een geschikte functie $f(x)$ op het interval $[a, b]$ zo als in Figuur III.5 te zien.

De oppervlakte onder de grafiek is natuurlijk de *integraal* en we krijgen dan voor de kans $P(a \leq x \leq b)$ dat x in het interval $[a, b]$ ligt:

$$P(a \leq x \leq b) = \int_a^b f(t) dt.$$



Figuur III.5: Kans op een uitkomst in een interval als oppervlakte onder de grafiek van een functie.

Als de kans groot is, moet de gemiddelde waarde van $f(x)$ op het interval dus ook groot zijn, als de kans klein is, heeft ook de functie $f(x)$ kleine waarden. Om op deze manier echt een kansverdeling te krijgen, moet de functie $f(x)$ aan de volgende eisen voldoen:

- (i) $f(x) \geq 0$ voor alle $x \in \mathbb{R}$,
- (ii) $\int_{-\infty}^{\infty} f(x) dx = 1$.

De eerste eis zorgt ervoor dat we steeds niet-negatieve kansen krijgen en de tweede eis zegt dat de totale oppervlakte onder de grafiek 1 is en geeft dus de normering van de kansverdeling weer. We noemen een functie $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ die aan deze eisen voldoet een *dichtheidsfunctie*.

Merk op dat we met de definitie van de kans als oppervlakte op een interval automatisch aan de eis voldoen dat $P(A \cup B) = P(A) + P(B)$ als $A \cap B = \emptyset$ (eis (iii) uit de oorspronkelijke definitie van een kansverdeling) want voor niet overlappende deelintervallen $[a, b]$ en $[c, d]$ worden de oppervlakten gewoon bij elkaar opgeteld.

De reden voor de naam *dichtheidsfunctie* ligt in het feit, dat we de kans op een waarde in een interval van breedte Δx voor kleine intervallen kunnen benaderen door $\Delta x \cdot f(x)$. Als we Δx als een *eenheidsinterval* zien, is $f(x)$ de dichtheid van de kansmassa rond x , net zo als we de dichtheid van een stof zien als de massa van een eenheidsvolume van de stof.

In nauw verband met de dichtheidsfunctie $f(x)$ staat de *verdelingsfunctie* $F(a)$, die voor elke waarde van a de kans $P(x \leq a)$ dat de uitkomst hoogstens a is aangeeft. Omdat dit betekent dat $-\infty < x \leq a$, krijgen we deze kans als

oppervlakte onder de grafiek van $f(x)$ tussen $-\infty$ en a , dus (weer als integraal geschreven) als

$$F(a) := \int_{-\infty}^a f(x) dx.$$

De verdelingsfunctie heeft de eigenschappen:

- (i) $\lim_{a \rightarrow -\infty} F(a) = 0$, $\lim_{a \rightarrow \infty} F(a) = 1$.
- (ii) $F(a)$ is stijgend, dus $a_2 \geq a_1 \Rightarrow F(a_2) \geq F(a_1)$.
- (iii) $P(a \leq x \leq b) = F(b) - F(a)$.
- (iv) $F'(a) = f(a)$, dus de afgeleide van $F(a)$ geeft de dichtheidsfunctie.

We gaan nu een aantal belangrijke voorbeelden van continue kansverdelingen bekijken

De uniforme verdeling

Dit is het continue analogo van de discrete gelijkverdeling en wordt soms ook de *homogene verdeling* of *rechthoekverdeling* genoemd. Op een bepaald interval $[a, b]$ (of een vereniging van intervallen) heeft elke punt dezelfde kans en buiten het interval is de kans 0. De normering $\int_{-\infty}^{\infty} f(x) dx = 1$ geeft dan de waarde voor $f(x)$ op het interval $[a, b]$. De dichtheidsfunctie $f(x)$ en verdelingsfunctie $F(x)$ van de uniforme verdeling zijn

$$f(x) = \begin{cases} 0 & \text{als } x < a \\ \frac{1}{b-a} & \text{als } a \leq x \leq b \\ 0 & \text{als } x > b \end{cases} \quad \text{en} \quad F(x) = \begin{cases} 0 & \text{als } x < a \\ \frac{x-a}{b-a} & \text{als } a \leq x \leq b \\ 1 & \text{als } x > b \end{cases}$$

De normale verdeling

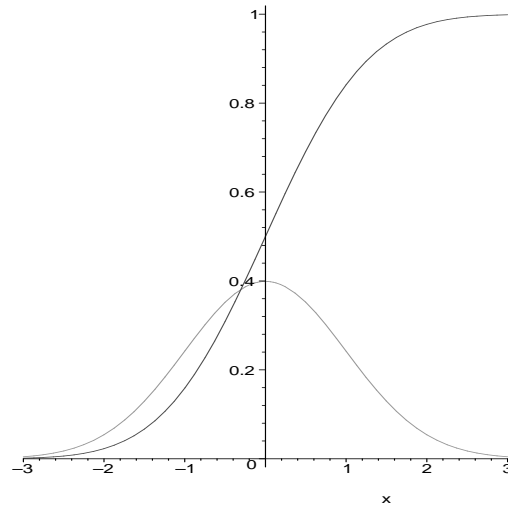
De belangrijkste continue verdeling is de normale verdeling of *Gauss verdeling* die bijvoorbeeld centraal staat in de statistiek maar ook in de theorie van Fourier transformaties. De dichtheidsfunctie heeft de vorm van een klok en is gegeven door

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Omdat de functie $f(x)$ geen primitieve heeft, die we als gesloten functie kunnen schrijven, laat zich in dit geval de verdelingsfunctie alleen maar door de integraal $F(a) = \int_{-\infty}^a f(x) dx$ beschrijven.

De normale verdeling met $\mu = 0$ en $\sigma = 1$ noemen we *standaard-normale verdeling*.

De redenen voor de centrale stelling van de normale verdeling zijn veelvoudig. Een van de redenen wordt geformuleerd als de *Centrale limietstelling* die ruwweg zegt dat de combinatie van een aantal (onafhankelijke) toevallige gebeurtenissen bij benadering tot een normale verdeling leidt. We zullen dit later nader toelichten. Een verdere reden is, dat voor zekere (voldoende grote)



Figuur III.6: Dichtheidsfunctie en verdelingsfunctie voor de standaard-normale verdeling

waarden van de parameters ook sommige discrete kansverdelingen goed door de normale verdeling benaderd worden.

Bijvoorbeeld wordt de binomiale verdeling $b(m, p; k)$ door de normale verdeling met $\mu = mp$ en $\sigma^2 = mp(1 - p)$ benaderd. Deze benadering wordt meestal als redelijk beschouwd als $mp \geq 5$ en $m(1 - p) \geq 5$ geldt.

Voor de Poisson-verdeling geldt iets soortgelijks (omdat deze al een benadering voor de binomiale verdeling is). De Poisson-verdeling met parameter λ wordt benaderd door de normale verdeling met parameters $\mu = \lambda$ en $\sigma^2 = \lambda$. Hierbij wordt vaak de vuistregel $\lambda \geq 5$ voor de toepasbaarheid van de benadering gehandteerd (merk op dat we bij de Poisson-verdeling al veronderstellen dat p klein is, dus is $mp = \lambda \geq 5$ tegenover $m(1 - p) \geq 5$ de sterkere eis).

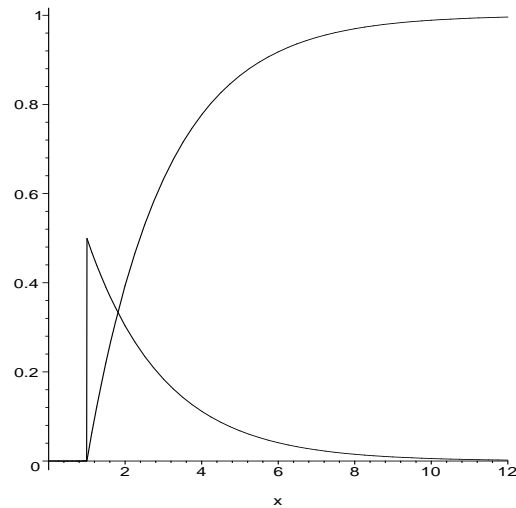
De exponentiële verdeling

Bij het bepalen van de levensduur van dingen als radioactieve preparaten of borden in de kast gaan we ervan uit dat het aantal verdwijnende objecten evenredig is met het aantal objecten die er nog zijn. Dit soort processen voldoet aan een differentiaalvergelijking $f'(x) = \lambda f(x)$ die de oplossing $e^{-\lambda x}$ heeft. De dichtheidsfunctie en verdelingsfunctie die de levensduur van dit soort objecten beschrijft, zijn:

$$f(x) = \begin{cases} 0 & \text{als } x < c \\ \lambda e^{-\lambda(x-c)} & \text{als } x \geq c \end{cases} \quad \text{en} \quad F(x) = \begin{cases} 0 & \text{als } x < c \\ 1 - e^{-\lambda(x-c)} & \text{als } x \geq c \end{cases}$$

Merk op dat de constante factor λ bij de exponentiële functie weer door de normering bepaald is, want

$$\int_c^\infty e^{-\lambda(x-c)} dx = \frac{-1}{\lambda} e^{-\lambda(x-c)} \Big|_c^\infty = \frac{1}{\lambda}.$$



Figuur III.7: Dichtheidsfunctie en verdelingsfunctie voor de exponentiële verdeling met $\lambda = 0.5$ en $c = 1$

BELANGRIJKE BEGRIPPEN IN DEZE LES

- kansverdeling
- gelijkverdeling (Laplace-verdeling)
- hypergeometrische verdeling
- binomiale verdeling
- Poisson-verdeling
- continue kansverdelingen
- dichtheidsfunctie, verdelingsfunctie
- normale verdeling, exponentiële verdeling

OPGAVEN

60. Bij een hockeytoernooi zijn er 18 teams aangemeld. In de eerste ronde worden de teams in twee groepen van 9 teams geloot. Onder de deelnemers zijn 5 teams uit de hoogste klasse. Hoe groot is de kans dat deze 5 teams in dezelfde groep terecht komen? Hoe groot is de kans dat er in een groep 2 en in de andere 3 teams uit de hoogste klasse terecht komen.
61. In een kast liggen n paren schoenen (dus $2n$ schoenen) willekeurig door elkaar. Je grijpt blindelings $k \leq n$ schoenen. Hoe groot is de kans dat je er minstens één passend paar uit vist? Hoe groot is de kans dat je precies één paar uit vist?
62. Een test bestaat uit 10 ja-nee vragen. Iemand die van toeten nog blazen weet, besluit de vragen op goed geluk te beantwoorden (dit betekent dat hij voor elke vraag een kans van $\frac{1}{2}$ op een goed antwoord heeft). Met 6 goede antwoorden ben je in de test geslaagd. Wat is de kans voor onze kandidaat om de test te halen?

63. Volgens een statistiek vinden in Nederland per jaar 3 op de 100.000 mensen een portemonnee met meer dan 1000 €. Wat is de kans dat in en stad als Nijmegen (met 150.000 inwoners) dit geluk (a) 3, (b) 5, (c) 10, (d) hooguit 2 mensen overkomt.
64. De kans dat een eerstejaars student in een bepaald vak afstudeert is 40%. Wat zijn de kansen dat uit een groep van 5 eerstejaars:
- (i) niemand afstudeert,
 - (ii) precies 1 afstudeert,
 - (iii) minstens 3 afstuderen?

Les 13 Verwachtingswaarde en spreiding

13.1 Stochasten

In een paar voorbeelden hebben we al gezien dat we bij een experiment vaak niet zo zeer in een enkele uitkomst geïnteresseerd zijn, maar bijvoorbeeld wel in het aantal uitkomsten van een zeker soort. Zo willen we bij een steekproef weten, hoeveel stukken defect zijn, maar niet of nu het eerste of laatste stuk defect is.

Vaak zijn de uitkomsten waarin we geïnteresseerd zijn veel eenvoudiger dan de uitkomstenruimte zelf, bijvoorbeeld kijken we naar het aantal k van defecte stukken in plaats van alle combinaties van m testresultaten, waarvan k negatief zijn. We kunnen dus zeggen, dat we verschillende uitkomsten die een zekere eigenschap gemeenschappelijk hebben in een cluster samenvatten, Zo'n eigenschap laat zich door een functie

$$X : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto X(\omega)$$

beschrijven, die aan elk element ω van de uitkomstenruimte een waarde $X(\omega)$ toekent. Zo'n functie X noemen we een *random variable* (in het Engels), een *stochastische variabele*, een *kansvariabele* of kort een *stochast*.

In het voorbeeld van de kwaliteitsproef is de stochast dus de functie die aan een rij van testresultaten het aantal negatieve (of positieve) resultaten toekent.

Een ander voorbeeld is het dobbelen met twee dobbelstenen: Als we alleen maar in de som van de geworpen getallen geïnteresseerd zijn, nemen we als stochast de functie $X(\omega_1, \omega_2) := \omega_1 + \omega_2$.

Het belangrijke aan de stochasten is, dat we makkelijk een kansverdeling hiervoor kunnen definiëren: De kans $P(X = x)$ dat de stochast de waarde x aanneemt definiëren we door

$$P(X = x) := \sum_{X(\omega)=x} P(\omega)$$

dus we tellen gewoon de kansen voor alle elementen van Ω op, waar de stochast de waarde x oplevert.

Onbewust hebben we al eerder stochasten op deze manier gebruikt, bijvoorbeeld voor het uitrekenen van de kans dat we met twee dobbelstenen een som van 5 werpen.

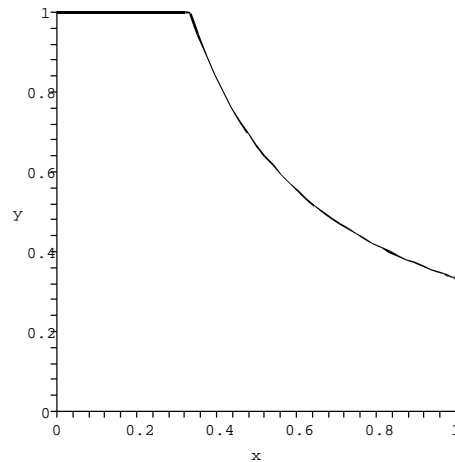
Voor continue kansverdelingen gaat de som over de uitkomsten met $X(\omega) = x$ over in een integraal. Omdat de kans op een enkele uitkomst steeds 0 is, wordt hier de kans bepaald, dat de stochast X een waarde onder een gegeven grens aanneemt. Voor een continue kansverdeling met dichtheidsfunctie $f(x)$ krijgen we:

$$P(X \leq x) = \int_{X(t) \leq x} f(t) dt.$$

Meestal zijn continue stochasten door hun eigen dichtheidsfunctie aangegeven, er geldt dan

$$P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Voorbeeld: Stel we hebben een randomgenerator die toevalsgetallen tussen 0 en 1 volgens de uniforme verdeling voortbrengt. We vragen ons af, wat de kans is dat het product van twee opeenvolgende van die toevalsgetallen kleiner is dan een grens $0 \leq a \leq 1$. De stochast die bij dit probleem hoort is $X(x, y) := x \cdot y$ en omdat we het met de uniforme verdeling te maken hebben, moeten we alleen maar de oppervlakte van het gebied $G = \{(x, y) \in \mathbb{R}^2 \mid x \cdot y \leq a\}$ bepalen. Als $x \leq a$ kan y elke waarde tussen 0 en 1 hebben, maar voor $x \geq a$ hebben we $y \leq \frac{a}{x}$ nodig. De volgende schets laat dit (voor $a = \frac{1}{3}$) zien:



Met behulp van een eenvoudige integratie kunnen we de kansverdeling van deze stochast ook expliciet bepalen, er geldt:

$$P(X \leq a) = \int_0^a dx + \int_a^1 \frac{a}{x} dx = a + a(\log(1) - \log(a)) = a(1 - \log(a)).$$

Voor $a = 0.5$ is deze kans bijvoorbeeld $P(X \leq 0.5) \approx 0.85$ en pas voor $a < 0.187$ is $P(X \leq a) < 0.5$.

13.2 Verwachtingswaarde

Als we in het casino roulette gaan spelen, zijn we er niet in geïnteresseerd of we in het eerste of laatste spel winnen of verliezen en ook niet hoe vaak we winnen of verliezen. Eigenlijk willen we alleen maar weten of we kunnen verwachten dat we aan het eind van de dag met een winst naar huis komen. Als we N keer spelen en bij elke keer 10€ op rood zetten, dan is bij elk spel de kans dat we 10€ winnen gelijk aan $\frac{18}{37}$, want er zijn 18 rode en 18 zwarte getallen en de groene 0. De kans dat we de 10€ verliezen is dus $\frac{19}{37}$. Als we heel vaak spelen, kunnen we verwachten dat we $\frac{18 \cdot N}{37}$ keer winnen en $\frac{19 \cdot N}{37}$ keer verliezen. Dit betekent dat we een verlies van $N \cdot \frac{1}{37} \cdot 10€$ kunnen verwachten.

Uit het perspectief van het casino is dit natuurlijk heel wenselijk. Omdat alle winsten alleen maar op de getallen 1 t/m 36 zijn gebaseerd (dus als je op de 3 getallen 4, 5, 6 zet maak je een winst van 12 keer je inzet) heeft de groene 0 het effect dat het casino gemiddeld een zevenendertigste van alle inzetten wint.

In het voorbeeld van het roulette spel hebben we een stochast gebruikt die het bedrag van de winst of verlies aangeeft. Waar we in geïnteresseerd zijn is de gemiddelde winst die we per spel zullen maken. Dit is het gemiddelde van de mogelijke waarden van de stochast, waarbij elke waarde met zijn kans gewogen wordt. Wat we zo krijgen is de winst die we per spel gemiddeld verwachten, en daarom noemen we dit ook de *verwachtingswaarde*.

Algemeen definiëren we voor een stochast X de verwachtingswaarde $E(X)$ (de E staat voor het Engelse *expectation*) door

$$E(X) := \sum_{x \in X} x \cdot P(X = x) = \sum_{x \in X} x \cdot \left(\sum_{X(\omega)=x} P(\omega) \right) = \sum_{\omega \in \Omega} X(\omega)P(\omega).$$

Voor een stochast X met continue kansverdeling is de verwachtingswaarde met behulp van zijn dichtheidsfunctie $f(x)$ analoog gedefinieerd door de integraal

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

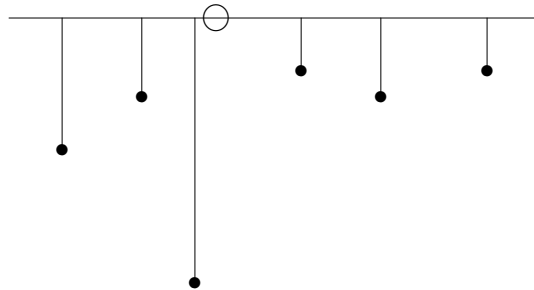
Merk op dat we van een continu verdeelde stochast door samenvatten van de waarden in een deelinterval naar een discreet verdeelde stochast kunnen komen:

Er geldt $P(X \in [x, x + \delta]) = \int_x^{x+\delta} f(t) dt$ en voor kleine δ kunnen we aannemen dat $f(t)$ op het interval $[x, x + \delta]$ bijna constant is, dit geeft

$$P(X \in [x, x + \delta]) \approx \delta \cdot f(x).$$

Als we nu de reële lijn in stukjes $[i \cdot \delta, (i + 1) \cdot \delta]$ van lengte δ onderverdelen en de uitkomsten $x \in [i \cdot \delta, (i + 1) \cdot \delta]$ tot de uitkomst $x = i \cdot \delta$ samenvatten, hebben we alleen maar nog de discrete verzameling $\{i \cdot \delta \mid i \in \mathbb{Z}\}$ van uitkomsten. Voor deze *gediscretiseerde* stochast is de verwachtingswaarde $\sum_{i \in \mathbb{Z}, x=i \cdot \delta} x \cdot P(X \in [x, x + \delta]) \approx \sum_{i \in \mathbb{Z}, x=i \cdot \delta} x \cdot \delta \cdot f(x)$ en dit is juist de discrete benadering van de integraal $\int_{-\infty}^{\infty} x \cdot f(x) dx = E(X)$.

We kunnen de verwachtingswaarde aanschouwelijk zien als het evenwichtspunt van een balk (oneindig lang, zonder gewicht), waar we in het punt x een gewicht van massa $P(x)$ aan hangen. Het evenwichtspunt is dan juist het punt $E(X)$. In het volgende plaatje zijn de gewichten gerepresenteerd door de lengten van de verticale ribben.



Een aantal belangrijke elementaire eigenschappen van de verwachtingswaarde kunnen we meteen uit de definitie aflezen. Als X en Y stochasten zijn, dan geldt:

- (i) $E(X + Y) = E(X) + E(Y)$, dus de som van de verwachtingswaarden van twee stochasten is de verwachtingswaarde van de som van de stochasten.
- (ii) $E(\alpha X) = \alpha E(X)$.
- (iii) $X(\omega) \geq Y(\omega)$ voor alle $\omega \in \Omega \Rightarrow E(X) \geq E(Y)$.

Als we in (i) voor Y de constante stochast $Y(\omega) = c$ nemen, volgt hieruit dat een verschuiving van de stochast om c ook de verwachtingswaarde om c verschuift (omdat de constante stochast verwachtingswaarde c heeft). We kunnen dus een stochast door aftrekken van zijn verwachtingswaarde altijd zo verschuiven dat hij verwachtingswaarde 0 heeft:

$$X_0 := X - E(X) \Rightarrow E(X_0) = E(X - E(X)) = E(X) - E(X) = 0.$$

We gaan nu de verwachtingswaarden van de belangrijkste kansverdelingen berekenen.

Binomiale verdeling

We hebben $P(X = k) = b(m, p; k) = \binom{m}{k} p^k (1 - p)^{m-k}$, dus:

$$\begin{aligned} E(X) &= \sum_{k=0}^m k \binom{m}{k} p^k (1 - p)^{m-k} = \sum_{k=0}^m k \frac{m!}{k!(m-k)!} p^k (1 - p)^{m-k} \\ &= m \cdot p \cdot \sum_{k=1}^m \frac{(m-1)!}{(k-1)!(m-k)!} p^{k-1} (1 - p)^{m-k} \\ &= m \cdot p \cdot \sum_{k=0}^{m-1} \binom{m-1}{k} p^k (1 - p)^{m-1-k} \\ &= m \cdot p \cdot \sum_{k=0}^{m-1} b(m-1, p; k) = m \cdot p. \end{aligned}$$

In de laatste stap hebben we hierbij gebruik van het feit gemaakt, dat de som over de kansen $b(m-1, p; k)$ voor alle waarden van k de totale kans 1 oplevert. De verwachtingswaarde van de binomiale verdeling is dus $m \cdot p$ en dit is precies het verwachte aantal van gunstige uitkomsten als we m pogingen doen bij een kans van p voor een gunstige uitkomst.

Hypergeometrische verdeling

We hebben $P(X = k) = h(n, m, s; k) = \frac{\binom{s}{k} \binom{n-s}{m-k}}{\binom{n}{m}}$, en er geldt: $k \cdot \binom{s}{k} = k \cdot \frac{s!}{k!(s-k)!} = s \cdot \frac{(s-1)!}{(k-1)!(s-k)!} = s \cdot \binom{s-1}{k-1}$ en $\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n}{m} \cdot \frac{(n-1)!}{(m-1)!(n-m)!} =$

$\frac{n}{m} \cdot \binom{n-1}{m-1}$. Hieruit volgt:

$$\begin{aligned} E(X) &= \sum_{k=0}^m k \frac{\binom{s}{k} \cdot \binom{n-s}{m-k}}{\binom{n}{m}} = \sum_{k=1}^m \frac{s \binom{s-1}{k-1} \cdot \binom{n-s}{m-k}}{\frac{n}{m} \binom{n-1}{m-1}} = m \frac{s}{n} \sum_{k=1}^m \frac{\binom{s-1}{k-1} \cdot \binom{n-s}{m-k}}{\binom{n-1}{m-1}} \\ &= m \frac{s}{n} \sum_{k=0}^{m-1} \frac{\binom{s-1}{k} \cdot \binom{n-s}{m-1-k}}{\binom{n-1}{m-1}} = m \frac{s}{n} \sum_{k=0}^{m-1} h(n-1, m-1, s; k) = m \frac{s}{n}. \end{aligned}$$

In de stap naar de laatste regel hebben hierbij k door $k+1$ verplaatst, de som die voor k van 1 tot m loopt, loopt voor $k+1$ van 0 tot $m-1$. In de laatste stap loopt de som over de kansen $h(n-1, m-1, s; k)$ voor alle waarden van k , dus is deze som gelijk aan 1. Het resultaat hadden we ook intuïtief kunnen afleiden, want de kans om bij een greep een van de s slechte stukken uit de totale n stukken te pakken is $\frac{s}{n}$, en als we m keer grijpen zouden we gemiddeld $m \frac{s}{n}$ slechte stukken verwachten.

Poisson-verdeling

We hebben $P(X = k) = p_{o\lambda}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ en maken gebruik van de gelijkheid $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$:

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

Ook hier vinden we het verwachte resultaat, omdat de Poisson-verdeling de limiet van de binomiale verdeling is als $p \rightarrow 0$ gaat en $m \cdot p = \lambda$ constant is.

Uniforme verdeling

We hebben $P(X = x) = \frac{1}{b-a}$ als $a \leq x \leq b$ en 0 anders, dus

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

De verwachtingswaarde is dus het middelpunt van het interval waarop de dichtheidsfunctie niet 0 is.

Exponentiële verdeling

We nemen aan dat we de dichtheidsfunctie zo hebben verschoven dat de beginwaarde $c = 0$ is. Dan is $f(x) = \lambda e^{-\lambda x}$ als $x \geq 0$ en $f(x) = 0$ anders. Dit geeft

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = -x \lambda e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}$$

(merk op dat we hierbij gebruiken dat $\lim_{x \rightarrow \infty} x e^{-x} = 0$ is). Ook hier is het resultaat voor de verwachtingswaarde plausibel, want als λ groter wordt, gaat de functie $f(x)$ sneller naar nul en moeten we dus een kleinere verwachtingswaarde krijgen.

Normale verdeling

In dit geval kunnen we de verwachtingswaarde zonder enig rekenwerk bepalen.

Als we de stochast X met dichtheidsfunctie $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ om μ verschuiven, krijgen we de nieuwe stochast $X_0 := X - \mu$ met dichtheidsfunctie $f_0(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$ die symmetrisch ten opzichte van de y -as is. Hieruit volgt dat $x \cdot f_0(x)$ een oneven functie is, dus is $E(X_0) = \int_{-\infty}^{\infty} x \cdot f_0(x) dx = 0$. Maar dit geeft

$$E(X) = E(X_0 + \mu) = E(X_0) + \mu = \mu$$

dus is de parameter μ in de dichtheidsfunctie van de normale verdeling juist de verwachtingswaarde.

13.3 Spreiding

Als we de verwachtingswaarde van een stochast kennen, weten we wat we op lange termijn gemiddeld kunnen verwachten. Maar vaak willen we toch iets meer weten, bijvoorbeeld hoe ver de daadwerkelijke uitkomsten van de verwachtingswaarde verwijderd zijn. Als we namelijk een stochast X zo verschuiven dat de verwachtingswaarde 0 is, dan heeft ook de stochast αX verwachtingswaarde 0, maar voor $\alpha > 1$ zijn de enkele uitkomsten verder van de verwachtingswaarde verwijderd.

In het model van de balk met gewichten kunnen we het verschil tussen de stochasten X en αX duidelijk zien. Als de gewichten dicht bij het evenwichtspunt zijn, kunnen we de balk makkelijk om dit punt draaien. Als we nu bijvoorbeeld naar de stochast $10 \cdot X$ kijken, worden de afstanden van het evenwichtspunt met 10 vermenigvuldigd. Nu hebben we meer kracht nodig om de balk te draaien. Dit ligt eraan dat het traagheidsmoment van de balk groter geworden is, dit is namelijk gegeven als de som over $m \cdot r^2$ waarbij m de massa in een punt is die afstand r van het draaipunt heeft. Als we het traagheidsmoment naar de stochast vertalen wordt dit

$$Var(X) := \sum_{x \in X} (x - E(X))^2 \cdot P(X = x) = E((X - E(X))^2)$$

en dit noemen we de *variantie* of *spreiding* van X . De variantie is dus de verwachtingswaarde van de kwadratische afstand van de stochast van zijn verwachtingswaarde en is dus een maat ervoor hoe dicht de waarden van een stochast bij de verwachtingswaarde liggen.

Vaak wordt in plaats van de variantie de wortel uit de variantie als maat voor de afwijkingen gebruikt, omdat deze lineair met de stochast verandert (d.w.z. als we X met een factor α vermenigvuldigen, wordt ook de wortel uit de variantie met α vermenigvuldigt). We definiëren dus

$$\sigma_X := \sqrt{Var(X)}$$

en noemen dit de *standaardafwijking* van X .

Voorbeeld: Bij het werpen van een dobbelsteen is de verwachtingswaarde $E(X) = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{7}{2}$. De variantie is dan $Var(X) = \sum_{k=1}^6 (k - \frac{7}{2})^2 \cdot \frac{1}{6} = \frac{35}{12}$ en de standaardafwijking $\sigma_X = \sqrt{\frac{35}{12}} \approx 1.7$.

We hebben boven opgemerkt dat de variantie van een stochast aangeeft hoe sterk de uitkomsten van de verwachtingswaarde afwijken. Deze samenhang tussen verwachtingswaarde en spreiding kunnen we heel expliciet aangeven, namelijk in de *Ongelijkheid van Chebyshev*. Hierbij maken we een afschatting voor de kans dat een uitkomst een grotere afstand dan $a > 0$ van de verwachtingswaarde $E(X)$ heeft.

Volgens de definitie berekenen we de variantie door $Var(X) = \sum_{x \in X} (x - E(X))^2 \cdot P(X = x)$. Als we de som beperken tot de waarden van x met $|x - E(X)| \geq a$, krijgen we

$$Var(X) \geq \sum_{|x-E(X)| \geq a} (x - E(X))^2 \cdot P(X = x) \geq \sum_{|x-E(X)| \geq a} a^2 \cdot P(X = x)$$

en dit is juist $a^2 \cdot P(|X - E(X)| \geq a)$. We hebben dus bewezen:

Ongelijkheid van Chebyshev: Voor een stochast X met verwachtingswaarde $E(X)$ en variantie $Var(X)$ geldt voor elke $a > 0$ de ongelijkheid

$$P(|X - E(X)| \geq a) \leq \frac{1}{a^2} Var(X).$$

Als voorbeeld kunnen we met de ongelijkheid van Chebyshev eens afschatten, wat de kans op het dobbelen van een zes is. We hebben boven gezien dat de verwachtingswaarde bij het dobbelen $\frac{7}{2}$ en de variantie $\frac{35}{12}$ is. De afstand tussen een 6 en de verwachtingswaarde $\frac{7}{2}$ is $\frac{5}{2}$ en volgens de ongelijkheid van Chebyshev geldt $P(|X - E(X)| \geq \frac{5}{2}) \leq \frac{4}{25} \cdot \frac{35}{12} = \frac{7}{15} \approx 0.467$. Omdat deze kans ook het dobbelen van een 1 insluit, moeten we nog door twee delen en schatten de kans op een 6 dus met 23.3% (naar boven) af. Natuurlijk weten we dat de kans in feite $\frac{1}{6} = 16.7\%$ is en dit laat zien dat de afschatting niet eens zo slecht is.

In de statistiek wordt vaak als vuistregel de zo genoemde 2σ -regel gebruikt: Voor een stochast X met standaardafwijking σ_X liggen meestal 95% van de gebeurtenissen in het interval $(E(X) - 2\sigma_X, E(X) + 2\sigma_X)$. De ongelijkheid van Chebyshev geeft aan dat dit interval minstens 75% van de gebeurtenissen bevat, maar voor de meeste kansverdelingen (in het bijzonder voor de normale verdeling) geldt de sterkere uitspraak van de 2σ -regel.

Naast de ongelijkheid van Chebyshev kunnen we een aantal verdere belangrijke eigenschappen voor de variantie van een stochast X meteen uit de definities afleiden:

- (i) $Var(X) = 0$ dan en slechts dan als $X = c$ constant is.
- (ii) $Var(\alpha X) = \alpha^2 Var(X)$ en $\sigma_{\alpha X} = \alpha \cdot \sigma_X$.

(iii) $Var(X+c) = Var(X)$, dus zo als we dit zouden verwachten is de variantie onafhankelijk van een verschuiving van de stochast.

(iv) $Var(X) = E(X^2) - E(X)^2$, want:

$$\begin{aligned} Var(X) &= \sum_{x \in X} (x - E(X))^2 \cdot P(X = x) \\ &= \left(\sum_{x \in X} x^2 \cdot P(X = x) \right) - 2E(X) \left(\sum_{x \in X} x \cdot P(X = x) \right) + E(X)^2 \\ &= E(X^2) - 2E(X) \cdot E(X) + E(X)^2 = E(X^2) - E(X)^2. \end{aligned}$$

Dit is in veel gevallen een handige formule om de variantie van een stochast uit te rekenen.

Vaak is het nuttig een stochast zo te normeren dat hij verwachtingswaarde 0 en variantie 1 heeft. Dit kunnen we met behulp van (ii) en (iii) makkelijk bereiken, want voor $X_0 := \frac{X - E(X)}{\sigma_X}$ geldt $E(X_0) = \frac{1}{\sigma_X}(E(X) - E(X)) = 0$ en $Var(X_0) = Var\left(\frac{X}{\sigma_X}\right) = \frac{1}{\sigma_X^2} Var(X) = 1$.

We gaan nu ook de varianties van de meest belangrijke kansverdelingen berekenen.

Binomiale verdeling

Dit pakken we met de formule $Var(X) = E(X^2) - E(X)^2$ aan:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^m k^2 \binom{m}{k} p^k (1-p)^{m-k} \\ &= m \cdot p \cdot \sum_{k=1}^m k \frac{(m-1)!}{(k-1)!(m-k)!} p^{k-1} (1-p)^{m-k} \\ &= m \cdot p \cdot \sum_{k=0}^{m-1} (k+1) \binom{m-1}{k} p^k (1-p)^{m-1-k}. \end{aligned}$$

De som $\sum_{k=0}^{m-1} (k+1) \binom{m-1}{k} p^k (1-p)^{m-1-k}$ is de verwachtingswaarde van de verschoven stochast $X+1$ voor de parameter $m-1$, dus is de waarde hiervan $(m-1)p+1$. We hebben dus $E(X^2) = mp((m-1)p+1) = mp(mp+(1-p))$ en dus

$$Var(X) = E(X^2) - E(X)^2 = mp(mp+(1-p)) - (mp)^2 = mp(1-p).$$

Hypergeometrische verdeling

Dit is een beetje omslachtig om uit te werken, dus geven we (voor de volledigheid) alleen maar het resultaat aan. Voor een stochast X met $P(X = k) = h(n, m, s; k)$ geldt

$$Var(X) = m \frac{s}{n} \left(1 - \frac{s}{n}\right) \frac{n-m}{n-1}.$$

Als n veel groter is dan m geldt $\frac{n-m}{n-1} \approx 1$ en met $p = \frac{s}{n}$ wordt de variantie van de hypergeometrische verdeling dan benadert door de variantie van de binomiale verdeling.

Poisson-verdeling

We gebruiken weer de formule $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda$. Er geldt:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \sum_{k=1}^{\infty} ((k-1) + 1) \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \left(\sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} \right) + \left(\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \right) \\ &= \lambda^2 e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) + \lambda e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = \lambda^2 + \lambda. \end{aligned}$$

We hebben dus

$$Var(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Dit hadden we ook uit de variantie voor de binomiale verdeling kunnen gokken, want de Poisson-verdeling is de limiet voor $p \rightarrow 0$ met $mp = \lambda$ en bij deze limiet gaat $mp(1-p)$ naar $mp = \lambda$.

Uniforme verdeling

Er geldt

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3(b-a)} (b^3 - a^3) = \frac{1}{3} (a^2 + ab + b^2)$$

dus hebben we

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{3} (a^2 + ab + b^2) - \frac{1}{4} (a^2 + 2ab + b^2) = \frac{1}{12} (a-b)^2.$$

Exponentiële verdeling

Er geldt

$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = -x^2 \lambda e^{-\lambda x} \Big|_0^{\infty} + 2 \int_0^{\infty} x e^{-\lambda x} dx = 2 \int_0^{\infty} x e^{-\lambda x} dx$$

en dit is gelijk aan $\frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}$. We hebben dus

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Normale verdeling

Dit is iets ingewikkelder te berekenen maar de parameters in de normale verdeling zijn zo gekozen dat σ^2 de variantie aangeeft en dus σ de standaardafwijking.

13.4 Covariantie en correlatie

Het is iets moeilijker om iets over de variantie van de som van twee stochasten te zeggen dan dit bij de verwachtingswaarde het geval was. We hebben

$$\begin{aligned} \text{Var}(X + Y) &= E((X + Y)^2) - (E(X + Y))^2 \\ &= E(X^2 + 2X \cdot Y + Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(X \cdot Y) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2E(X \cdot Y) - 2E(X)E(Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2(E(X \cdot Y) - E(X) \cdot E(Y)). \end{aligned}$$

We noemen $E(X \cdot Y) - E(X) \cdot E(Y)$ de *covariantie* van X en Y en noteren dit met $\text{Cov}(X, Y)$. Er geldt dus

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

en dit betekent dat de covariantie aangeeft hoe sterk de variantie van de som van twee stochasten afwijkt van de som van de varianties.

De covariantie laat zich ook beschrijven als de verwachtingswaarde van het product van $(X - E(X))$ en $(Y - E(Y))$, want:

$$\begin{aligned} E((X - E(X)) \cdot (Y - E(Y))) &= E(X \cdot Y - E(X)Y - E(Y)X + E(X)E(Y)) \\ &= E(X \cdot Y) - E(E(X)Y) - E(E(Y)X) + E(E(X)E(Y)) \\ &= E(X \cdot Y) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(X \cdot Y) - E(X)E(Y) = \text{Cov}(X, Y), \end{aligned}$$

dus hebben we

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))).$$

We zullen in de volgende les uitgebreid bediscussiëren wat het betekent dat twee stochasten *onafhankelijk* zijn, maar intuïtief zou men al zeggen, dat de uitkomst van de ene stochast de uitkomst van de andere niet mag beïnvloeden. We zullen twee stochasten X en Y onafhankelijk noemen, als de kans $P(X = x, Y = y)$ op de gecombineerde uitkomst $X = x$ en $Y = y$ gelijk is aan het product $P(X = x) \cdot P(Y = y)$ van de kansen op de aparte uitkomsten en als dit voor alle paren (x, y) geldt.

Stel nu dat X en Y onafhankelijke stochasten zijn, dan geldt:

$$\begin{aligned} E(X \cdot Y) &= \sum_{(x,y) \in X \times Y} x \cdot y \cdot P(X = x, Y = y) \\ &= \sum_{(x,y) \in X \times Y} x \cdot y \cdot P(X = x) \cdot P(Y = y) \\ &= \left(\sum_{x \in X} x \cdot P(X = x) \right) \left(\sum_{y \in Y} y \cdot P(Y = y) \right) = E(X) \cdot E(Y). \end{aligned}$$

We hebben dus gezien:

Voor onafhankelijke stochasten X en Y geldt $E(X \cdot Y) = E(X) \cdot E(Y)$, dus $Cov(X, Y) = 0$ en dus

$$Var(X + Y) = Var(X) + Var(Y) \text{ als } X \text{ en } Y \text{ onafhankelijk zijn.}$$

Waarschuwing: De omkering van deze stelling geldt niet. Twee stochasten kunnen covariantie 0 hebben zonder onafhankelijk te zijn.

We hebben gezien dat de covariantie $Cov(X, Y)$ in zekere zin en maat voor de afhankelijkheid van X en Y is. Er laat zich aantonen dat $|Cov(X, Y)| \leq \sigma_X \sigma_Y$ is, dus de covariantie van twee stochasten is begrensd door het product van de standaardafwijkingen van de stochasten. Met behulp van de standaardafwijkingen kunnen we dus de covariantie op waarden tussen -1 en 1 normeren. We noemen

$$\rho_{X,Y} := \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

de *correlatiecoëfficiënt* van X en Y . De waarde van de correlatiecoëfficiënt ligt tussen -1 en 1 de waarde $\rho_{X,Y} = -1$ treedt alleen maar op voor $Y = -\alpha X + \beta$ met $\alpha > 0$, de waarde $\rho_{X,Y} = 1$ alleen maar voor $Y = \alpha X + \beta$ met $\alpha > 0$. Precies gezegd geeft de correlatiecoëfficiënt dus aan, in hoeverre de stochasten X en Y *lineair* van elkaar afhangen, d.w.z. hoe goed zich Y door $\alpha X + \beta$ laat benaderen. Voor $\rho_{X,Y} > 0$ spreekt men van *positieve afhankelijkheid* voor $\rho_{X,Y} < 0$ van *negatieve afhankelijkheid*.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- stochasten
- verwachtingswaarde
- variantie, standaardafwijking
- covariantie, correlatiecoëfficiënt

OPGAVEN

65. Bij een spel met een dobbelsteen win je $n\text{€}$ als je n dobbelt en n even is en je verliest $n\text{€}$ als n oneven is. Wat is de verwachtingswaarde van je winst/verlies.
66. Bij het skaat spel krijg je 10 kaarten uit een kaartspel met 32 kaarten (8 soorten, 4 kleuren). Wat is de verwachtingswaarde voor het aantal boeren dat je krijgt?
67. In een loterij heb je 70% nieten en 30% winnende lotjes. Iemand beslist zo lang lotjes te kopen tot dat hij een winnende lot krijgt, maar hooguit vijf keer. Wat kan hij voor een uitgave verwachten, als een lot 2€ kost?
68. Je koopt een nieuwe speelautomaat voor je kroeg. In de automaat draaien twee onafhankelijke wielen die in tien even grote segmenten zijn opgedeeld en volgens een gelijkverdeling in een van de segmenten stoppen. De segmenten hebben de nummers 1 t/m 10. Een speler heeft alleen maar de volgende winstmogelijkheden (bij alle andere uitkomsten verliest hij zijn inzet):

- Als beide wielen 10 tonen wint hij 5€.
- Als beide wielen hetzelfde getal maar niet 10 tonen wint hij 2€.
- Als precies een van de wielen 10 toont wint hij 1€.

Je wilt natuurlijk winst met je automaat maken. Wat is de minimale inzet die je per spel moet vragen om een winst te kunnen verwachten?

69. Twee tennissters A en B spelen vaker tegen elkaar en gemiddeld wint A 60% van de sets. De speelsters ontmoeten elkaar op een toernooi in een best-of-five match (dus wie het eerst drie sets wint heeft gewonnen).
- Wat zijn de kansen dat A in 3, 4 of 5 sets wint? Hoe zit het met B ? Wat is de kans dat B überhaupt wint?
 - Bereken de verwachtingswaarde voor het aantal sets die het match duurt.
 - Bereken apart de verwachtingswaarden voor het aantal sets in het geval dat A wint en dat B wint.
 - Bereken de spreiding en de standaardafwijking voor het aantal sets die het match duurt: onafhankelijk van wie er wint, als A wint en als B wint.

Les 14 Voorwaardelijke kansen en de regel van Bayes

Sommige vragen uit de kanstheorie hebben een antwoord dat niet met de intuïtie van iedereen klopt. Een voorbeeld hiervoor is het *Monty-Hall probleem* ook bekend als *Geitenprobleem*:

Bij een TV-show valt er voor de kandidaat een auto te winnen. Het enige wat de kandidaat moet doen is uit drie deuren de goede deur te kiezen waar de auto achter staat. Achter de andere twee deuren zijn er geiten. Nadat de kandidaat een deur heeft gekozen, wordt deze niet meteen geopend, maar de showmaster (die weet waar de auto staat) opent een van de niet gekozen deuren en een geit blaast tegen het publiek (en de kandidaat). De vraag is nu: Is het voor de kandidaat verstandig is om bij zijn keuze te blijven, of is het gunstiger om te wisselen of maakt het niets uit.

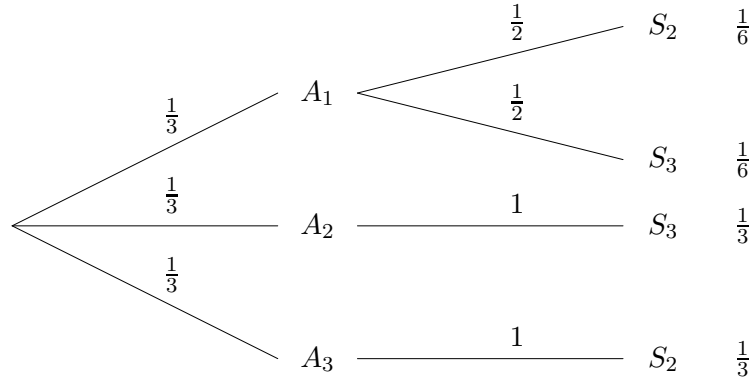
Intuïtief zullen veel mensen denken, dat na het openen van een van de deuren met een geit daarachter de kans 50 : 50 is, dat de auto achter de door de kandidaat gekozen deur staat. Dus zou het niets uitmaken of de kandidaat wisselt of niet. In de VS heeft een journaliste, Marilyn vos Savant, de oplossing voor dit probleem in haar column in de tijdschrift *Parade* gepubliceerd. Deze vrouw heeft een van de hoogste IQ's ter wereld en haar antwoord was dat de kans op de auto groeit als de kandidaat wisselt. Haar column resulteerde in een lawine van boosaardige en verontwaardigde brieven, waaronder veel van wiskundigen, die het antwoord van vos Savant bespottelijk maakten. Als reactie op dit gebeuren werd in Duitsland door de journalist Gero von Randow in de weekkrant *Die Zeit* een artikel gepubliceerd, waarin hij het geitenprobleem en een oplossing met dezelfde conclusie als die van vos Savant voorstelde. Ook hier was de reactie opmerkelijk: Over weken kwamen er brieven binnen, waarin professoren, gepromoveerde en dergelijk 'geleerden' uitlegden waarom de oplossing van vos Savant en von Randow onzin is. Ook hier waren er behoorlijk veel wiskundigen bij.

Hoe zit het nu met de oplossing van het geitenprobleem? De reden waarom veel mensen voor de 50 : 50 oplossing kiezen is dat ze ervan uit gaan, dat de situatie na het openen van een van de deuren door de showmaster onafhankelijk is van wat er eerder is gebeurd. Dit is echter niet het geval! Als de kandidaat een deur met een geit daarachter heeft gekozen, heeft de showmaster geen keuze welke deur hij gaat openen, terwijl hij in het geval dat de kandidaat de deur met de auto heeft gekozen twee mogelijkheden heeft.

We kunnen dit als volgt analyseren: Stel de kandidaat heeft deur 1 gekozen. De auto kan nu achter deur 1, 2 of 3 staan, deze gevallen noemen we A_1 , A_2 en A_3 en we gaan ervan uit dat elk van deze gevallen een kans van $\frac{1}{3}$ heeft. In het geval A_1 kan de showmaster deur 2 of deur 3 openen. Deze gevallen noemen we S_2 en S_3 en omdat er geen verschil tussen de deuren (en de geiten) is, kunnen we aannemen dat S_2 en S_3 dezelfde kans $\frac{1}{2}$ hebben. De kans dat de auto achter deur 1 staat en de showmaster deur 2 opent is dus $\frac{1}{6}$, hetzelfde

geldt voor het openen van deur 3. Maar in het geval A_2 heeft de showmaster geen keuze, hij moet deur 3 openen, dus is de kans voor dit geval $\frac{1}{3}$. Evenzo moet de showmaster in het geval A_3 deur 2 openen, dus is ook hier de kans $\frac{1}{3}$.

Deze situatie kunnen we door het volgende boomdiagram beschrijven:



In het geval dat de showmaster deur 2 heeft geopend is de kans dus twee keer zo groot dat de auto achter deur 3 staat dan dat hij achter deur 1 staat. Hetzelfde geldt voor het geval dat de showmaster deur 3 heeft geopend. In elk geval is het dus verstandig dat de kandidaat van keuze verandert, want hierdoor wordt zijn kans op de auto twee keer zo groot.

We zullen later nog eens op het geitenprobleem terug komen en het antwoord uit de regel van Bayes afleiden. Maar eerst gaan we algemeen naar het probleem kijken dat de kans voor een uitkomst kan veranderen als aanvullende informatie over gerelateerde gebeurtenissen bekend wordt.

14.1 Voorwaardelijke kansen

Het idee dat de kans voor een uitkomst kan veranderen als we aanvullende informatie hebben, is zo natuurlijk dat we er meestal niet over nadenken. Bijvoorbeeld kan de kans op vorst op 30 april over de afgelopen 150 jaar eenvoudig afgelezen worden uit de tabellen van de weerkundige dienst. Als er bijvoorbeeld 10 keer in de afgelopen 150 jaren vorst op 30 april was, kunnen we aannemen dat de kans op vorst op 30 april 2005 ongeveer 6.67% is. Als aanvullende informatie kunnen we gebruiken dat er ook 10 keer vorst op 29 april is geweest en dat er in 5 jaren vorst op 29 en 30 april gevallen is. Zo ver maakt dit nog geen verschil voor de kans op vorst op 30 april 2005. Maar als er inderdaad vorst op 29 april 2005 valt, kunnen we zeggen dat de kans op vorst op 30 april 2005 opeens 50% is, want in 5 van de 10 jaren met vorst op 29 april was er ook vorst op 30 april.

De kans dat er vorst op 30 april valt, gegeven het feit dat er vorst op 29 april is, noemen we een *voorwaardelijke kans*.

Abstract gaan we dit zo beschrijven: Stel we willen de kans van $A \subseteq \Omega$ bepalen onder de voorwaarde dat $B \subseteq \Omega$ plaats vindt. Deze kans definiëren we als de kans dat A en B gebeuren, gegeven het feit dat B gebeurt. Als de kansen door relatieve frequenties gegeven zijn, dus $P(A) = \frac{|A|}{|\Omega|}$, hebben

we $\frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{P(A \cap B)}{P(B)}$ en het laatste nemen we als definitie voor de voorwaardelijke kans:

Voor een kansverdeling P of Ω en $B \subseteq \Omega$ met $P(B) > 0$ noemen we

$$P(A | B) := \frac{P(A, B)}{P(B)} := \frac{P(A \cap B)}{P(B)}$$

de voorwaardelijke kans op A gegeven B .

Notatie: De kans voor het gemeenschappelijke optreden van de gebeurtenissen A en B wordt meestal met $P(A, B)$ in plaats van $P(A \cap B)$ genoteerd.

Om te rechtvaardigen, dat we $P(A | B)$ een *kans* noemen, moeten we even aantonen dat $P(\cdot | B)$ voor $P(B) > 0$ een kansverdeling is, waarbij we natuurlijk gebruiken dat $P(\cdot)$ al een kansverdeling is. (Voor $P(B) = 0$ is het onzin een kans onder de voorwaarde B te bekijken, want B gebeurt nooit.)

(i) $P(A | B) = \frac{P(A \cap B)}{P(B)} \geq 0.$

(ii) $P(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$

(iii) Voor $A_1, A_2 \subseteq \Omega$ met $A_1 \cap A_2 = \emptyset$ geldt $(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B)$. Verder is $(A_1 \cap B) \cap (A_2 \cap B) = \emptyset$ omdat $A_1 \cap B$ een deelverzameling van A_1 en $A_2 \cap B$ een deelverzameling van A_2 is. Daarom geldt:

$$P(A_1 \cup A_2 | B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} = \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} = P(A_1 | B) + P(A_2 | B).$$

Voorbeeld: Hier is een typisch voorbeeld van een vraag die met voorwaardelijke kansen te maken heeft:

Aan 1000 werknemers wordt gevraagd of ze een hoog of een laag salaris hebben. Van de werknemers geven 210 vrouwen aan een hoog salaris te hebben en 360 geven aan een laag salaris te hebben. Van de mannen blijken 210 een hoog en 220 een laag salaris te hebben. Deze gegevens vinden we in het volgende schema terug:

	hoog salaris	laag salaris	som
vrouw	0.21	0.36	0.57
man	0.21	0.22	0.43
totaal	0.42	0.58	1.00

De vraag is nu of vrouwen en mannen dezelfde kans op een hoog salaris hebben. De kans voor een vrouw om een hoog salaris te hebben is de voorwaardelijke kans $P(\text{hoog} | \text{vrouw}) = \frac{P(\text{hoog en vrouw})}{P(\text{vrouw})} = \frac{0.21}{0.57} \approx 0.37$. Voor mannen is de kans $P(\text{hoog} | \text{man}) = \frac{P(\text{hoog en man})}{P(\text{man})} = \frac{0.21}{0.43} \approx 0.49$ dus hebben mannen in dit voorbeeld een behoorlijk grotere kans op een hoog salaris dan vrouwen.

We kunnen voorwaardelijke kansen niet alleen maar voor twee deelverzamelingen van Ω maar ook algemeen voor n deelverzamelingen definiëren. Het

idee hierbij is hetzelfde, we kijken naar de kans van het gemeenschappelijke optreden van de voorwaarden met een gebeurtenis, gedeeld door de kans voor de voorwaarden en krijgen dus:

$$P(A_n | A_1 \cap \dots \cap A_{n-1}) = P(A_n | A_1, \dots, A_{n-1}) = \frac{P(A_1, \dots, A_n)}{P(A_1, \dots, A_{n-1})}.$$

We hebben dus bijvoorbeeld $P(A_3 | A_1, A_2) = \frac{P(A_1, A_2, A_3)}{P(A_1, A_2)}$.

Omgekeerd kunnen we de kans voor het gemeenschappelijke optreden van gebeurtenissen (iteratief) door voorwaardelijke kansen uitdrukken en krijgen zo de zogeheten *kettingregel*:

$$P(A_1, A_2) = P(A_2 | A_1) \cdot P(A_1),$$

$$P(A_1, A_2, A_3) = P(A_3 | A_1, A_2) \cdot P(A_1, A_2) = P(A_3 | A_1, A_2) \cdot P(A_2 | A_1) \cdot P(A_1),$$

en in het algemeen

$$P(A_1, \dots, A_n) = P(A_n | A_1, \dots, A_{n-1}) \cdot P(A_{n-1} | A_1, \dots, A_{n-2}) \cdot \dots \cdot P(A_2 | A_1) \cdot P(A_1).$$

14.2 Regel van Bayes

Omdat de doorsnede $A \cap B$ symmetrisch in A en B is, vinden we uit de definitie voor de voorwaardelijke kans dat

$$P(A | B) \cdot P(B) = P(A \cap B) = P(B | A) \cdot P(A)$$

en dit geeft de eenvoudigste vorm van de *regel van Bayes*, namelijk

$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A)}.$$

De nut van deze regel ligt in het omdraaien van de rollen van voorwaarde en uitkomst. Denk hierbij bijvoorbeeld aan een test op een ziekte. Als de uitslag van de test gegeven is, zijn we geïnteresseerd in de kans dat we de ziekte hebben of niet. Maar bekend is alleen maar de nauwkeurigheid van de test die zegt met welke kans de test bij een gezonde mens het verkeerde resultaat geeft en andersom.

De regel van Bayes wordt vaak op een iets slimmere manier toegepast. Hiervoor wordt de deelverzameling $B \subseteq \Omega$ in verschillende gevallen onderverdeeld die elkaar uitsluiten, dus we schrijven $B = \cup_{i=1}^n B_i$ met $B_i \cap B_j = \emptyset$ als $i \neq j$. Een belangrijk speciaal geval hiervoor is $B = B_1 \cup B_2$ met $B_2 = B \setminus B_1 = B_1^c$. We noemen B_2 het *complement* van B_1 in B .

Er geldt:

$$P(A \cap B) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A | B_i) \cdot P(B_i)$$

en dus

$$P(A | B) = \frac{1}{P(B)} \sum_{i=1}^n P(A | B_i) \cdot P(B_i).$$

In het bijzonder kunnen we in het geval $A \subseteq B$ de *totale kans* $P(A)$ berekenen als

$$P(A) = P(A \cap B) = P(A | B) \cdot P(B) = \sum_{i=1}^n P(A | B_i) \cdot P(B_i)$$

en het belangrijkste geval hiervoor is $B = \Omega$, d.w.z. we delen alle mogelijke uitkomsten in een aantal klassen van uitkomsten op. In het geval van de opsplitsing van Ω in een deelverzameling $B_1 \subseteq \Omega$ en zijn complement $B_2 = \Omega \setminus B_1 = B_1^c$ geeft dit

$$P(A) = P(A | B_1) \cdot P(B_1) + P(A | B_1^c) \cdot P(B_1^c).$$

We kunnen nu de algemene versie van de regel van Bayes formuleren, deze luidt als volgt:

Regel van Bayes: Zij $B \subseteq \Omega$ met $B = \cup_{i=1}^n B_i$ en $B_i \cap B_j = \emptyset$ als $i \neq j$. Verder zij $A \subseteq B$. Dan geldt

$$P(B_j | A) = \frac{P(A | B_j) \cdot P(B_j)}{P(A)} = \frac{P(A | B_j) \cdot P(B_j)}{\sum_{i=1}^n P(A | B_i) \cdot P(B_i)}$$

Om de abstracte concepten duidelijk te maken, passen we de regel van Bayes op een aantal voorbeelden toe.

Voorbeeld 1: De uitkomst van een HIV-test noemen we A als de test positief was en A^c als de test negatief was. Het geïnfecteerd zijn noemen we I en het niet geïnfecteerd zijn I^c . Over de kwaliteit van de test is bekend, dat hij voor geïnfecteerden in 99% van de gevallen een positief resultaat oplevert en voor niet geïnfecteerden in 99.9% van de gevallen een negatief resultaat. We hebben dus $P(A | I) = 0.99$, $P(A^c | I) = 0.01$ en $P(A^c | I^c) = 0.999$, $P(A | I^c) = 0.001$. Verder nemen we aan dat 1 uit 10000 mensen HIV-geïnfecteerd is, dus $P(I) = 0.0001$ en $P(I^c) = 0.9999$. De vraag is nu, hoe groot bij een positieve HIV-test de kans is, inderdaad geïnfecteerd te zijn, dus hoe groot de voorwaardelijke kans $P(I | A)$ is. Met de regel van Bayes hebben we

$$\begin{aligned} P(I | A) &= \frac{P(A | I) \cdot P(I)}{P(A)} = \frac{P(A | I) \cdot P(I)}{P(A | I) \cdot P(I) + P(A | I^c) \cdot P(I^c)} \\ &= \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.001 \cdot 0.9999} \approx 9.0\%. \end{aligned}$$

Deze verrassend lage kans is opmerkelijk maar toch goed te begrijpen. Als we 10000 mensen testen, dan is er gemiddeld 1 HIV-geïnfecteerde mens bij en die krijgt waarschijnlijk ook een positieve test-uitslag. Maar bij de 9999 niet-geïnfecteerden zal de test in 0.1% van de gevallen een (verkeerd) positief resultaat opleveren, dus komen er nog 10 positieve resultaten bij. Als we dus naar de 11 positieve resultaten kijken, is dit alleen maar in één geval veroorzaakt door een geïnfecteerde, maar in 10 gevallen door een test-fout.

Merk op dat er in dit soort vragen vaak verkeerd geargumenteed wordt. Dit vind je zelfs in wetenschappelijke publicaties, bijvoorbeeld in de medicijn of in de rechtsgeleerdheid terug. Denk hier bijvoorbeeld aan een misdadiger

waarbij de schuld door een DNA-analyse wordt bewezen. Het probleem is, dat zelfs bij een test met een hoge nauwkeurigheid het aantal verkeerde uitslagen vaak hoger is dan het aantal van de gezochte zeldzame uitkomsten.

Voorbeeld 2: Een student moet bij een tentamen een multiple-choice vraag met n mogelijkheden oplossen. Als hij voorbereid is, zal zijn antwoord juist zijn, als niet zal hij willekeurig een antwoord gokken en dus een kans van $\frac{1}{n}$ op een juiste antwoord hebben. De kans dat de student voorbereid is, zij p . Voor de docent is het nu interessant om de kans te bepalen, dat de student inderdaad voorbereid was, als hij een juiste antwoord heeft gegeven. Als we een juiste antwoord met J en een voorbereide student met V betekenen hebben we dus:

$$\begin{aligned} P(V | J) &= \frac{P(J | V) \cdot P(V)}{P(J | V) \cdot P(V) + P(J | V^c) \cdot P(V^c)} \\ &= \frac{1 \cdot p}{1 \cdot p + \frac{1}{n}(1 - p)} = \frac{np}{np + (1 - p)}. \end{aligned}$$

Het is duidelijk dat dit voor grote waarden van n dicht bij 1 ligt, want dan is $(1 - p)$ tegen np te verwaarlozen. Maar voor $n = 4$ en $p = 0.5$ hebben we bijvoorbeeld $P(V | J) = \frac{4}{5} = 80\%$ en voor $n = 4$ en $p = 0.2$ geldt al $P(V | J) = \frac{1}{2} = 50\%$. Als de docent dus weet dat gewoon maar een vijfde van de studenten voorbereid is, weet hij ook dat de helft van de goede antwoorden goede gokken zijn.

Voorbeeld 3: In de automatische spraakherkenning gaat het erom, gegeven een akoestisch signaal X het woord w te vinden dat hier het beste bij past, d.w.z. waarvoor de voorwaardelijke kans $P(w | X)$ maximaal is. Hiervoor gebruiken we ook de regel van Bayes en schrijven

$$P(w | X) = \frac{P(X | w) \cdot P(w)}{P(X)}.$$

Omdat we alleen maar aan het woord met de hoogste kans geïnteresseerd zijn, kunnen we de noemer gewoon vergeten, omdat die voor elk woord hetzelfde is. In de teller geeft $P(X | w)$ de kans, dat een zeker woord w tot het signaal X lijdt. Deze kans wordt tijdens het *training* van een systeem bepaald, waarbij een aantal mensen het woord spreekt en uit de zo verkregen signalen een kansverdeling geschat wordt. De kans $P(w)$ is de totale kans dat een woord gesproken wordt. Dit noemen we de a-priori kans voor het woord, en deze kansen worden als relatieve frequenties op heel grote tekst-corpora (bijvoorbeeld 10 jaar NRC Handelsblad) bepaald.

Hetzelfde principe geldt trouwens voor de meeste soorten van patroonherkenning (beeld-herkenning, handschrift-herkenning).

Voorbeeld 4: We komen nog eens terug op het Monty-Hall probleem. Stel de kandidaat heeft deur 1 gekozen, dan nemen we aan dat de showmaster deur 2 heeft geopend (S_2), het geval S_3 geeft een analoog resultaat. We zijn nu geïnteresseerd in de kansen $P(A_1 | S_2)$ en $P(A_3 | S_2)$, dus de voorwaardelijke kansen dat de auto achter deur 1 of deur 3 staat, gegeven het feit dat de

showmaster deur 2 heeft geopend. Er geldt

$$\begin{aligned} P(A_1 | S_2) &= \frac{P(S_2 | A_1) \cdot P(A_1)}{P(S_2 | A_1) \cdot P(A_1) + P(S_2 | A_2) \cdot P(A_2) + P(S_2 | A_3) \cdot P(A_3)} \\ &= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 + 1 \cdot \frac{1}{3}} = \frac{1}{3}. \end{aligned}$$

Evenzo berekenen we de kans $P(A_3 | S_2)$ als

$$\begin{aligned} P(A_3 | S_2) &= \frac{P(S_2 | A_3) \cdot P(A_3)}{P(S_2 | A_1) \cdot P(A_1) + P(S_2 | A_2) \cdot P(A_2) + P(S_2 | A_3) \cdot P(A_3)} \\ &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 + 1 \cdot \frac{1}{3}} = \frac{2}{3}. \end{aligned}$$

We zien dus weer dat het voor de kandidaat verstandig is om naar deur 3 te wisselen, omdat de kans dat de auto daar achter zit twee keer zo groot is.

14.3 Onafhankelijkheid

Nu dat we goed naar voorwaardelijke kansen hebben gekeken kunnen we ook zeggen wat het betekent dat twee uitkomsten onafhankelijk zijn. Intuïtief zullen we zeggen, dat twee uitkomsten A en B onafhankelijk zijn, als de kans voor A niet ervan afhangt of B optreedt of niet. Met de voorwaardelijke kans kunnen we dit zo formuleren:

*Twee uitkomsten A en B heten onafhankelijk als $P(A) = P(A | B)$.
Equivalent hiermee is dat $P(A \cap B) = P(A) \cdot P(B)$.*

De equivalentie van de twee formuleringen volgt uit de definitie van de voorwaardelijke kans, want $P(A \cap B) = P(A | B) \cdot P(B)$ geeft $P(A) = P(A | B) \Leftrightarrow P(A \cap B) = P(A | B) \cdot P(B) = P(A) \cdot P(B)$. Omdat ook $P(A \cap B) = P(B | A) \cdot P(A)$ geldt, volgt hieruit ook dat $P(A) = P(A | B) \Leftrightarrow P(B) = P(B | A)$, dus het maakt niets uit welke voorwaardelijke kans we bekijken.

Een eenvoudig voorbeeld zijn de soorten en kleuren in een kaartspel. De kans om uit een kaartspel met 52 kaarten een aas te trekken is $\frac{1}{13}$, de kans om een kaart van kleur klaver te trekken is $\frac{1}{4}$. De doorsnede van de uitkomsten *aas* en *klaver* is alleen maar de kaart *klaver aas* en de kans om deze kaart te trekken is $\frac{1}{52} = \frac{1}{13} \cdot \frac{1}{4}$. Omdat we ook elke andere soort of kleur hadden kunnen kiezen, toont dit aan, dat de soorten en de kleuren onafhankelijk zijn.

In een ander voorbeeld kijken we naar een familie met twee kinderen. We vragen ons af of de uitkomsten

A : er is een meisje en een jongen B : er is hoogstens een meisje

onafhankelijk zijn. Als we m voor een meisje en j voor een jongen schrijven, zijn de mogelijkheden voor de twee kinderen (m, m) , (m, j) , (j, m) en (j, j) . We

zien makkelijk dat $P(A) = \frac{1}{2}$ en $P(B) = \frac{3}{4}$, maar $P(A \cap B) = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$. Dus zijn de uitkomsten A en B niet onafhankelijk.

Als we de familie nu van twee naar drie kinderen uitbreiden maar dezelfde uitkomsten bekijken, is de situatie veranderd. De mogelijkheden voor de drie kinderen zijn nu (m, m, m) , (m, j, m) , (j, m, m) , (j, j, m) , (m, m, j) , (m, j, j) , (j, m, j) en (j, j, j) . In dit geval is $P(A) = \frac{3}{4}$, $P(B) = \frac{1}{2}$ en $P(A \cap B) = \frac{3}{8} = P(A) \cdot P(B)$, dus zijn de uitkomsten nu inderdaad onafhankelijk.

Aan de hand van dit voorbeeld zien we, dat soms uitkomsten kanstheoretisch onafhankelijk zijn, die we in het echte leven niet onafhankelijk zouden noemen.

De onafhankelijkheid van uitkomsten A en B heeft ook nuttige consequenties voor de complementen A^c en B^c . Er geldt namelijk dat met (A, B) ook de paren (A, B^c) , (A^c, B) en (A^c, B^c) onafhankelijk zijn. Dit kunnen we met behulp van een paar eenvoudige manipulaties van de betrokken verzamelingen uit $P(A \cap B) = P(A) \cdot P(B)$ concluderen:

$$P(A \cap B^c) = P(A \cup B) - P(B) = P(A) + P(B) - P(A \cap B) - P(B) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A)(1 - P(B)) = P(A) \cdot P(B^c).$$

Dit werkt evenzo voor $P(A^c \cap B)$.

$$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A) \cdot P(B) = (1 - P(A))(1 - P(B)) = P(A^c) \cdot P(B^c).$$

We kunnen het begrip van onafhankelijkheid ook naar stochasten uitbreiden: Voor twee stochasten X, Y zij $A_x := \{\omega \in \Omega \mid X(\omega) = x\}$ en $B_y := \{\omega \in \Omega \mid Y(\omega) = y\}$. We noemen de uitkomsten A_x en B_y onafhankelijk als $P(A_x \cap B_y) = P(A_x) \cdot P(B_y)$. Maar in de taal van stochasten heet dit dat

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

en we noemen twee stochasten X, Y onafhankelijk als dit voor alle paren (x, y) geldt.

Tot nu toe hebben we het alleen maar over de onafhankelijkheid van *twee* uitkomsten gehad. Als we meerdere uitkomsten bekijken, zijn er verschillende mogelijkheden om hun onafhankelijkheid te definiëren:

- (1) We noemen de n uitkomsten A_1, \dots, A_n *paarsgewijs onafhankelijk* als $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ voor alle $i \neq j$.
- (2) We noemen n uitkomsten A_1, \dots, A_n *onafhankelijk* als $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$ voor elke deelverzameling $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$.

Als we de begrippen op deze manier definiëren is het duidelijk dat onafhankelijke uitkomsten ook paarsgewijs onafhankelijk zijn. Het omgekeerde geldt niet, wat aan het volgende tegenvoorbeeld duidelijk wordt:

We dobbelen met twee dobbelstenen en bekijken de kansen van de volgende uitkomsten:

A_1 : de eerste dobbelsteen toont een oneven getal,

A_2 : de tweede dobbelsteen toont een oneven getal,

A_3 : de som van de getallen is even.

We hebben $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$ en $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$, dus zijn de uitkomsten paarsgewijs onafhankelijk. Maar $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2)$ omdat de som van twee oneven getallen even is, dus is $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}$ en dus zijn de drie uitkomsten niet onafhankelijk.

We zouden bij de definitie van onafhankelijkheid voor meerdere uitkomsten ook kunnen hopen dat het voldoende is om $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n)$ te eisen, maar het volgende tegenvoorbeeld laat zien dat hieruit niet eens volgt dat de A_i paarsgewijs onafhankelijk zijn: We werpen een munt drie keer en kijken naar de volgende uitkomsten:

A_1 : de eerste worp toont kop,

A_2 : er valt vaker kop dan munt,

A_3 : de laatste twee worpen leveren hetzelfde resultaat.

Door naar de mogelijke uitkomsten te kijken zien we dat $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$ en dat $P(A_1 \cap A_2 \cap A_3) = \frac{1}{8}$. Aan de andere kant hebben we $P(A_1 \cap A_2) = \frac{3}{8}$, dus zijn A_1 en A_2 niet (paarsgewijs) onafhankelijk. De andere paren zijn wel onafhankelijk, want $P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$.

14.4 Bernoulli-model

Een belangrijke toepassing van de onafhankelijkheid van uitkomsten is de herhaalde uitvoering van een experiment. We nemen aan dat we in de uitkomstenruimte Ω een deelverzameling $A \subseteq \Omega$ van gunstige uitkomsten hebben. Bij de eenmalige uitvoering van het experiment is de kans op een gunstige uitkomst gegeven door $p = \frac{|A|}{|\Omega|}$. De kans voor een ongunstige uitkomst is dan $1 - p$. Als we het experiment twee keer uitvoeren is de kans dat we twee gunstige uitkomsten hebben de kans van de doorsnede van een gunstige uitkomst bij de eerste keer en een gunstige uitkomst bij de tweede keer. Omdat we ervan uitgaan dat het eerste en het tweede experiment onafhankelijk zijn, kunnen we de kans voor de doorsnede als product van de enkele kansen berekenen, dus als $p \cdot p = p^2$.

Merk op dat de eis dat herhalingen van een experiment onafhankelijk zijn een voorwaarde voor de opzet van het experiment is. Als je bijvoorbeeld de kans wilt bepalen waarmee een vaccinatie tot de uitbraak van een ziekte leidt mag je bij het herhalen van het experiment geen mensen nemen die al bij de vorige keer gevaccineerd zijn, omdat deze een hoger aantal antilichamen hebben en dus een kleinere kans lopen dat de ziekte uitbreekt.

Als we ervan uitgaan dat het herhalen van een experiment onafhankelijke uitkomsten heeft, dan is de kans dat we bij m herhalingen k keer een gunstige uitkomst hebben gegeven door de binomiale verdeling:

$$b(m, p; k) = \binom{m}{k} p^k (1 - p)^{m-k}.$$

De kans dat de eerste k uitkomsten gunstig zijn is namelijk p^k en de kans dat de laatste $m - k$ uitkomsten ongunstig zijn is $(1 - p)^{m-k}$. Nu kunnen we de gunstige uitkomsten nog op $\binom{m}{k}$ manieren over de m experimenten verdelen.

De beschrijving van uitkomsten door onafhankelijke herhaling van een experiment noemen we het *Bernoulli-model*.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- voorwaardelijke kans
- regel van Bayes
- onafhankelijkheid, paarsgewijs onafhankelijk
- Bernoulli-model

OPGAVEN

70. Een socioloog wil de kans bepalen dat mensen een keer een winkeldiefstal hebben gepleegd. Omdat mensen op een rechtstreekse vraag waarschijnlijk niet eerlijk zouden antwoorden heeft hij de volgende opzet verzonnen: Elke persoon krijgt 10 kaarten waarvan op 4 de vraag staat:
Heb je ooit een winkeldiefstal gepleegd?
 en op de andere 6 de vraag
Heb je nog nooit een winkeldiefstal gepleegd?
 De mensen worden nu gevraagd om toevallig één van de tien kaarten te trekken, het (waarheidsgetrouwe) antwoord op een briefje te schrijven en alleen maar dit briefje aan de onderzoeker te geven. Zo hoeft niemand om zijn anonimiteit te vrezen.
 Bij 1000 testpersonen krijgt de onderzoeker 516 keer het antwoord *ja* en 484 keer het antwoord *nee*. Hoe kan hij nu de gezochte kans berekenen en wat is deze kans?
71. Er wordt met twee dobbelstenen gedobbeld. Gegeven de informatie dat de twee dobbelstenen verschillende getallen tonen (bijvoorbeeld in een spel waar je bij gelijke getallen nog een keer dobbelt), wat is de kans dat de som oneven is?
72. In een zak zitten drie munten, waarvan twee eerlijk zijn maar de derde heeft twee kop-zijden. Er wordt blindelings een munt getrokken, vervolgens wordt deze munt twee keer geworpen, waarbij twee keer kop valt. Bepaal de kans, dat de getrokken munt een eerlijke munt is.
 Hoe zit het met het geval dat in de zaak een miljoen in plaats van drie munten zitten, waarvan weer één oneerlijk is. Nu werp je twintig keer in plaats van twee keer en krijgt twintig keer het resultaat kop. Hoe groot is nu de kans dat de getrokken munt een eerlijke munt is.
73. In sommige studies is er na het eerste semester een advies aan de studenten die weliswaar niet bindend is. Neem aan dat in een (zware) studie gemiddeld 40% van de studenten vroegtijdig afhaken. Het blijkt dat van de afhakende studenten 90% een negatief studieadvies kregen, terwijl slechts 1% van de studenten die afstuderen een negatief advies hadden. Wat is de kans dat een student met negatief studieadvies wel in dit vak zou afstuderen?

74. Bij een rechtbank zal een leugendetector geraadpleegd worden. Het is bekend dat voor een schuldige verdachte de detector in 90% van de gevallen het juiste resultaat (schuldige) geeft en voor een onschuldige verdachte in 99% van de gevallen het resultaat onschuldig. Uit de statistieken van de belastingdienst is bekend dat 5% van de burgers in hun belastingaangifte ernstig bedriegen. Bij een verdachte geeft de leugendetector aan dat de man/vrouw schuldig is. Wat is de kans, dat de verdachte toch onschuldig is?

Les 15 Schatten en simuleren

15.1 Maximum likelihood schatting

Tot nu toe hebben we meestal naar voorbeelden gekeken waar we van een kansverdeling zijn uitgegaan en dan voorspellingen hebben gemaakt. In de praktijk komen we echter vaak een iets andere situatie tegen. We weten dat er iets volgens een zekere kansverdeling zal gebeuren, maar deze hangt van een parameter af die we niet kennen. Bijvoorbeeld kunnen we aannemen dat de kans p waarmee een machine defecte stukken produceert constant is, maar dat we de waarde van p niet kennen. Als we nu in een steekproef defecte stukken tellen, kunnen we het aantal defecte stukken door de binomiale (of hypergeometrische) verdeling beschrijven. Wat we nu nodig hebben is een *schatting* voor de kans p , gegeven de aantallen van defecte stukken in een paar steekproeven. Neem aan dat we altijd een steekproef van m stukken nemen, dan vinden we in de verschillende steekproeven k_1, k_2, \dots, k_n defecte stukken. We kunnen nu op verschillende manieren een waarde voor p schatten, bijvoorbeeld:

- simplistisch: We schatten $p = \frac{k_1}{m}$, dus we nemen aan dat de eerste steekproef typisch was en negeren de anderen (dit kunnen we natuurlijk ook met k_3 of k_n in plaats van k_1 doen).
- optimistisch: We schatten $p = \frac{k_{min}}{m}$, waarbij k_{min} de minimale waarde van de k_i is.
- pessimistisch: We schatten $p = \frac{k_{max}}{m}$, waarbij k_{max} de maximale waarde van de k_i is.
- pragmatisch: We schatten $p = \frac{1}{n} \cdot \frac{\sum_{i=1}^n k_i}{m}$, dus we nemen het gemiddelde van de relatieve frequenties in de enkele steekproeven.

Een algemene methode om parameters van kansverdelingen te schatten is gebaseerd op het volgende argument: Voor elke keuze van een parameter (of meerdere parameters) heb je een kansverdeling, die aan een waargenomen resultaat een zekere kans geeft. In het voorbeeld is dit $P_p(X = k) = b(m, p; k) = \binom{m}{k} p^k (1-p)^{m-k}$. Bij onafhankelijke herhaling kunnen we de kans voor een rij observaties als product van de kansen voor de aparte observaties berekenen, in het voorbeeld hebben we dus

$$P_p(k_1, \dots, k_n) = \prod_{i=1}^n \binom{m}{k_i} p^{k_i} (1-p)^{m-k_i}.$$

De kans voor de observaties is nu een functie van de parameter p en we noemen deze functie de *aannemelijkheidsfunctie* of *likelihood* functie. We maken nu een schatting voor p door te zeggen, dat we p zo kiezen dat de aannemelijkheidsfunctie aan maximale waarde heeft, dus dat de kans voor onze observatie maximaal wordt. Deze methode van schatting noemt men de *meest aannemelijke* of *maximum likelihood* schatting van de parameter.

Om een maximum likelihood schatting uit te werken, moeten we in principe de functie $P_p(k_1, \dots, k_n)$ naar p afleiden en de nulpunten van de afgeleide bepalen. Omdat de kans een product van de enkele kansen is, zal het afleiden een hele hoop termen opleveren, want we moeten altijd de productregel toepassen. Hier is het volgende trucje vaak erg handig: In plaats van het maximum van $P_p(k_1, \dots, k_n)$ te berekenen, bepalen we het maximum van $\log(P_p(k_1, \dots, k_n))$. Dit zit namelijk op dezelfde plek, omdat de logaritme een monotone functie is. De (negatieve) logaritme van de kans noemt men soms ook de *score* van een uitkomst.

We gaan nu de maximum likelihood schatting voor een aantal kansverdelingen uitwerken:

Binomiale verdeling: In n steekproeven van grootte m_1, \dots, m_n vinden we k_1, \dots, k_n gunstige uitkomsten. We hebben

$$P_p(k_1, \dots, k_n) = \prod_{i=1}^n \binom{m_i}{k_i} p^{k_i} (1-p)^{m_i-k_i}.$$

We definiëren nu

$$\begin{aligned} L(p) &= \log(P_p(k_1, \dots, k_n)) = \sum_{i=1}^n \left(\log\left(\binom{m_i}{k_i}\right) + \log(p^{k_i}) + \log((1-p)^{m_i-k_i}) \right) \\ &= \sum_{i=1}^n \log\left(\binom{m_i}{k_i}\right) + \sum_{i=1}^n k_i \log(p) + \sum_{i=1}^n (m_i - k_i) \log(1-p). \end{aligned}$$

De afgeleide (met betrekking tot p) hiervan is

$$L'(p) = \frac{1}{p} \left(\sum_{i=1}^n k_i \right) - \frac{1}{1-p} \left(\sum_{i=1}^n (m_i - k_i) \right).$$

We hebben

$$\begin{aligned} L'(p) = 0 &\Leftrightarrow (1-p) \left(\sum_{i=1}^n k_i \right) = p \left(\sum_{i=1}^n (m_i - k_i) \right) \Leftrightarrow \sum_{i=1}^n k_i = p \left(\sum_{i=1}^n m_i \right) \\ &\Leftrightarrow p = \frac{\sum_{i=1}^n k_i}{\sum_{i=1}^n m_i}. \end{aligned}$$

Dit betekent dat we de parameter als de relatieve frequentie van gunstige uitkomsten in alle steekproeven bij elkaar kiezen. Dit komt op de pragmatische keuze neer, maar we hebben nu een betere onderbouwing voor onze keuze. Het is namelijk de parameter die de observaties het beste verklaart.

Poisson-verdeling: Een zeldzaam gebeurtenis zien we k_1, \dots, k_n keer gebeuren. We hebben

$$P_\lambda(k_1, \dots, k_n) = \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}.$$

We definiëren nu

$$\begin{aligned} L(\lambda) &= \log(P_\lambda(k_1, \dots, k_n)) = \sum_{i=1}^n (\log(\lambda^{k_i}) - \log(k_i!) - \lambda) \\ &= \sum_{i=1}^n k_i \log(\lambda) - \sum_{i=1}^n \log(k_i!) - n\lambda. \end{aligned}$$

De afgeleide hiervan is

$$L'(\lambda) = \frac{1}{\lambda} \left(\sum_{i=1}^n k_i \right) - n$$

en we hebben

$$L'(\lambda) = 0 \Leftrightarrow \lambda = \frac{1}{n} \left(\sum_{i=1}^n k_i \right).$$

De schatting voor de verwachtingswaarde λ van de Poisson-verdeling is dus het rekenkundig gemiddelde van de aantallen geobserveerde zeldzame gebeurtenissen. Ook dit klopt met onze intuïtie, dat we na een aantal pogingen aannemen, dat we vervolgens ook weer gebeurtenissen met ongeveer hetzelfde gemiddelde zullen krijgen.

Normale verdeling: De normale verdeling wordt in de opgaven behandeld.

Exponentiële verdeling: Voor een gebeurtenis dat volgens een exponentiële verdeling optreedt maken we de observaties x_1, \dots, x_n . Er geldt

$$f_\lambda(x_1, \dots, x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(\sum_{i=1}^n x_i)}.$$

Merk op dat we het hier met een dichtheidsfunctie voor de kansverdeling te maken hebben. Maar we kunnen aannemen, dat we steeds een klein interval rond een geobserveerde waarde bekijken, dan is de kans voor een observatie in het interval $[x, x + \delta]$ gegeven door $P_\lambda(X \in [x, x + \delta]) = \lambda e^{-\lambda x} \delta$. Maar δ heeft als constante factor geen invloed op het maximum van de functie, dus kunnen we meteen naar de dichtheidsfunctie kijken.

We definiëren nu

$$L(\lambda) = \log(f_\lambda(x_1, \dots, x_n)) = \log(\lambda^n) - \lambda \left(\sum_{i=1}^n x_i \right) = n \log(\lambda) - \lambda \left(\sum_{i=1}^n x_i \right).$$

De afgeleide hiervan is

$$L'(\lambda) = \frac{n}{\lambda} - \left(\sum_{i=1}^n x_i \right)$$

en we hebben

$$L'(\lambda) = 0 \Leftrightarrow \frac{1}{\lambda} = \frac{1}{n} \left(\sum_{i=1}^n x_i \right).$$

De schatting voor de verwachtingswaarde $\frac{1}{\lambda}$ van de exponentiële verdeling is dus weer het rekenkundig gemiddelde van de observaties.

Hypergeometrische verdeling: In de tweede les hebben we al het voorbeeld bekeken dat we het aantal vissen in een vijver willen bepalen. Het idee hiervoor is, dat we s vissen markeren en dan kijken hoeveel gemarkeerde vissen we in een (latere) steekproef van m vissen vinden. De kans dat we er k gemarkeerde vissen in vinden is gegeven door de hypergeometrische verdeling

$$h(n, m, s; k) = \frac{\binom{s}{k} \binom{n-s}{m-k}}{\binom{n}{m}}$$

waarbij n het onbekende aantal vissen in de vijver is. In dit voorbeeld gaan we niet de logaritme gebruiken, maar bepalen we het maximum van $h(n, m, s; k)$ als een functie van n op een andere manier. We kijken naar de quotiënt

$$q(n) := \frac{h(n, m, s; k)}{h(n-1, m, s; k)}.$$

Als $q(n) \geq 1$ is, stijgt $h(n, m, s; k)$ van $n-1$ naar n , als $q(n) \leq 1$ is, daalt $h(n, m, s; k)$ van $n-1$ naar n . We hebben

$$\begin{aligned} q(n) &= \frac{\binom{s}{k} \binom{n-s}{m-k}}{\binom{n}{m}} \cdot \frac{\binom{n-1}{m}}{\binom{s}{k} \binom{n-1-s}{m-k}} = \frac{\frac{(n-s)!}{(m-k)!(n-s-m+k)!} \cdot \frac{(n-1)!}{m!(n-1-m)!}}{\frac{n!}{m!(n-m)!} \cdot \frac{(n-1-s)!}{(m-k)!(n-1-s-m+k)!}} \\ &= \frac{(n-s)!(n-1)!(n-m)!(n-1-s-m+k)!}{(n-s-m+k)!(n-1-m)!n!(n-1-s)!} = \frac{(n-s)(n-m)}{(n-s-m+k)n} \\ &= \frac{n^2 - sn - nm + sm}{n^2 - sn - mn + kn}. \end{aligned}$$

We zien dus dat $q(n) \geq 1$ als $sm \geq kn$ en $q(n) \leq 1$ als $sm \leq kn$. Het maximum wordt dus bereikt voor $n = \frac{sm}{k}$, d.w.z. voor $\frac{k}{m} = \frac{s}{n}$. Dit betekent dat we de grootte van de populatie zo schatten dat het relatieve aantal gemarkeerde vissen in onze vangst hetzelfde is als het relatieve aantal in de hele vijver.

In de voorbeelden die we hier hebben behandeld, kunnen we de maximum likelihood schatting expliciet uitrekenen en krijgen meestal een resultaat dat we ook intuïtief hadden verwacht. Voor ingewikkeldere kansverdelingen (bijvoorbeeld met veel parameters) is het vaak niet mogelijk de nulpunten van de partiële afgeleiden expliciet te bepalen. Hier worden dan iteratieve benaderingsmethoden toegepast, bijvoorbeeld het *EM-algoritme* (van **E**xpectation **M**aximization).

Er zijn ook andere schatters dan de maximum likelihood schatter, bijvoorbeeld de *momentenschatters*. Het k -de moment van een stochast X is de verwachtingswaarde $E(X^k)$ van de k -de macht van de stochast. Bij een momentenschatter wordt geprobeerd de parameters van een kansverdeling zo te bepalen dat de momenten van de kansverdeling gelijk zijn aan de momenten die in een steekproef waargenomen zijn.

We zullen ons hier niet verder in verdiepen omdat het probleem van het schatten van parameters van een kansverdeling meer in de statistiek thuis hoort.

15.2 Simulatie

Soms heb je bij experimenten na een aantal observaties een idee erover wat er gebeurt en maak je er een model om de resultaten te beschrijven. De kwaliteit van een model ligt in het vermogen om toekomstige resultaten te kunnen voorspellen en dit is ook de manier hoe een model getoetst wordt. Vaak zijn experimenten zo ingewikkeld of kostbaar dat je bij een aanpassing van het beschrijvende model niet meteen weer veel experimenten kunt of wilt doen. Dan is het handig om het nieuwe model met een simulatie te testen, waarbij je zekere parameters volgens een kansverdeling kiest.

Een andere motivatie voor het simuleren van kansverdeling is dat sommige effecten pas naar heel veel herhalingen van een experiment naar voren komen. Voor een computer is het veel makkelijker om iets 10000 keer te herhalen dan dit in de realiteit te doen, bijvoorbeeld een munt 10000 keer te werpen.

We gaan daarom in deze paragraaf bekijken hoe we voor een aantal kansverdelingen een stochast met gegeven verdelingsfunctie kunnen simuleren.

Randomgenerator

Het startpunt voor alle soorten simulaties is een *toevalsgenerator* of *randomgenerator*. Dit is een procedure (een soort orakel) die een rij getallen u_1, u_2, \dots tussen 0 en 1 produceert, waarbij het i -de getal u_i de uitkomst van een stochast U_i is. De rij stochasten U_1, U_2, \dots moet (bij een ideale randomgenerator) aan de volgende eisen voldoen:

- (1) De kansverdeling op de i -de plek in de rij is de uniforme verdeling op het interval $[0, 1)$, d.w.z. er geldt $P(U_i \leq x) = x$ voor elke i .
- (2) De stochasten U_1, U_2, \dots zijn onafhankelijk, d.w.z. voor elke keuze van indices $i_1 < i_2 < \dots < i_k$ hebben we $P(U_{i_1} \leq x_1, U_{i_2} \leq x_2, \dots, U_{i_k} \leq x_k) = P(U_{i_1} \leq x_1) \cdot P(U_{i_2} \leq x_2) \cdot \dots \cdot P(U_{i_k} \leq x_k) = x_1 \cdot x_2 \cdot \dots \cdot x_k$.

Als de rij U_1, U_2, \dots van stochasten aan deze eisen voldoet, noemt men de getallen u_i *toevalsgetallen*. Helaas kan een praktische implementatie van een randomgenerator nooit perfect aan deze eisen voldoen, men spreekt daarom strikt genomen beter ervan dat een randomgenerator *pseudo-toevalsgetallen* en geen 'echte' toevalsgetallen produceert.

Een veel gebruikte type van randomgeneratoren zijn de *lineaire congruentie modellen*: Kies een getal $m \in \mathbb{N}$, constanten $a, c \in \mathbb{Z}$ en een *zaad* (Engels: seed) k_0 . Vervolgens bereken je iteratief

$$k_{n+1} := (ak_n + c) \bmod m$$

waarbij $x \bmod m$ de rest bij het delen van x door m is. De waarden van de getallen k_n liggen tussen 0 en $m - 1$, hieruit krijgt men toevalsgetallen u_n in het interval $[0, 1)$ door $u_n := \frac{k_n}{m}$ te definiëren.

Omdat k_n alleen maar de waarden $0, 1, \dots, m - 1$ kan hebben, is deze randomgenerator altijd periodiek met een periode van lengte hoogstens m . Maar behalve voor speciale (slechte) waarden van m, a, c en k_0 wordt deze lengte van de periode ook bereikt en levert deze methode een redelijk goede randomgenerator.

Vaak wordt voor m een macht van 2 zoals 2^{32} gekozen, omdat dit op een computer met 32-bit of 64-bit getallen de *modulo* operatie heel efficiënt maakt. In dit geval laat zich aantonen, dat een lineair congruentie model met $a \equiv 1 \pmod{4}$ en een oneven c altijd een periode van maximale lengte m oplevert.

Voordat een randomgenerator voor simulaties wordt gebruikt, is het verstandig om te toetsen of de pseudo-toevalsgetallen die hij oplevert inderdaad redelijk goed gelijkverdeeld en onafhankelijk zijn. Hiervoor zijn er een aantal toetsen, die op methoden uit de statistiek gebaseerd zijn.

Om een eerste indruk te krijgen, kan men de punten (u_{2i-1}, u_{2i}) in het 2-dimensionale vlak plotten en kijken of dit er redelijk toevallig uitziet. Als er hier al een soort structuur of patroon opvalt, is er zeker iets mis met de randomgenerator.

Bij een iets systematischere toets deelt men het interval $[0, 1)$ in d (even grote) deelintervallen, telt hoe veel van de getallen u_1, u_2, \dots, u_n in elk van die deelintervallen ligt en toetst deze verdeling met een χ^2 -toets tegen de gelijkverdeling.

De χ^2 -toets is een standaardtoets uit de statistiek die toetst of de gevonden verdeling te veel of te weinig van de gelijkverdeling afwijkt. Een te grote afwijking geeft evidentie tegen de hypothese dat alle U_i uniform verdeeld zijn, en zeer kleine afwijking laat aan de onafhankelijkheid twijfelen, want het is ook heel onwaarschijnlijk, dat in elk deelinterval precies d/n getallen terecht komen.

In plaats van de gelijkverdeling van de enkele toevalsgetallen kan men met de χ^2 -toets ook de gelijkverdeling van paren of in het algemeen van k -dimensionale vectoren (u_i, \dots, u_{i+k-1}) , $(u_{i+k}, \dots, u_{i+2k-1})$ van de toevalsgetallen toetsen. De k -dimensionale vectoren moeten gelijkverdeeld in de k -dimensionale kubus $[0, 1)^k$ zijn.

De onafhankelijkheid van de stochasten U_i kan ook op diverse manieren getoetst worden. Bijvoorbeeld wordt in de *gap test* een deelinterval $[a, b]$ van $[0, 1)$ gekozen en vervolgens gekeken, hoe lang de stukken van de rij (u_i) zijn die niet in $[a, b]$ liggen. Als we $p := |b - a|$ definiëren, dan is de kans op een stuk van lengte k tussen twee getallen die wel in $[a, b]$ liggen, gelijk aan $p(1 - p)^k$ (dit noemt men een geometrische verdeling met parameter p). De gevonden verdeling van lengtes van stukken kan men nu ook weer tegen de verwachte geometrische verdeling toetsen (bijvoorbeeld met een χ^2 -toets).

We gaan er vanaf nu van uit dat we een (wel getoetste) randomgenerator U_1, U_2, \dots ter beschikking hebben, die elke keer dat we hem gebruiken een toevalsgetal $u_i \in [0, 1)$ oplevert zo dat deze getallen gelijk verdeeld en onafhankelijk zijn.

Er zijn een aantal algemene principes, hoe we een gewenste kansverdeling met behulp van zo'n randomgenerator kunnen simuleren. De meest belangrijke zijn de methode van de *inverse verdelingsfunctie* en de *wegwerp* (rejection)

methode. In principe zijn deze methoden op discrete en continue kansverdelingen toepasbaar, omdat we ook voor discrete kansverdelingen vaak eenvoudig een verdelingsfunctie $F(x)$ en dichtheidsfunctie $f(x)$ kunnen aangeven. Maar voor zekere discrete kansverdelingen zullen we later nog andere (meer directe) methoden aangeven.

Simulatie met behulp van de inverse verdelingsfunctie

Voor een algemene (continue) kansverdeling met dichtheidsfunctie $f(x)$ en verdelingsfunctie $F(x) = \int_{-\infty}^x f(t) dt$ passen we de inverse F^{-1} van de verdelingsfunctie op de uniforme verdeling toe:

Zij U een stochast met uniforme verdeling op $[0, 1]$, dan definiëren we een nieuwe stochast X door

$$X := F^{-1}(U), \text{ dus } U = F(X).$$

Voor de kans $P(X \leq a)$ geldt nu

$$P(X \leq a) = P(F(X) \leq F(a)) = P(U \leq F(a)) = F(a)$$

omdat U uniform verdeeld is. De stochast X heeft dus juist de gewenste verdelingsfunctie $F(x)$.

Voorbeeld 1: We willen een algemene rechthoekverdeling op het interval $[a, b]$ simuleren. De verdelingsfunctie voor deze verdeling is $F(x) = \frac{1}{b-a}(x-a)$ en uit $y = \frac{1}{b-a}(x-a) \Leftrightarrow (b-a)y = (x-a) \Leftrightarrow x = a + (b-a)y$ volgt $F^{-1}(y) = a + (b-a)y$.

We krijgen dus een randomgenerator V_1, V_2, \dots met uniform verdeelde waarden in het interval $[a, b]$ door

$$V_i := a + (b-a)U_i.$$

Dit hadden we natuurlijk ook zonder de inverse van de verdelingsfunctie kunnen bedenken, want door het interval $[0, 1]$ met $b-a$ te vermenigvuldigen, rekken we het op $[0, b-a]$ en door a op te tellen verschuiven we het naar $[a, b]$.

Voorbeeld 2: Ook voor de exponentiële verdeling krijgen op deze manier een simulatie. Na een mogelijke verschuiving op de x -as heeft de exponentiële verdeling de dichtheidsfunctie $f(x) = \lambda e^{-\lambda x}$ en de verdelingsfunctie $F(x) = 1 - e^{-\lambda x}$. Omdat $y = 1 - e^{-\lambda x} \Leftrightarrow -\lambda x = \log(1-y) \Leftrightarrow x = -\frac{1}{\lambda} \log(1-y)$, hebben we $F^{-1}(y) = -\frac{1}{\lambda} \log(1-y)$. Voor een uniform verdeelde stochast U is dus de stochast

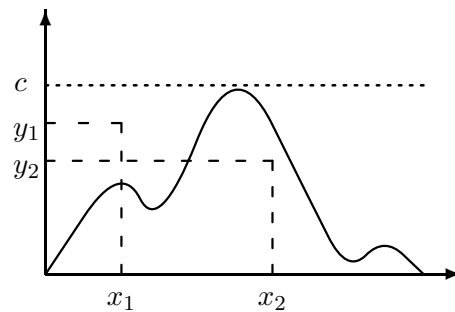
$$X := -\frac{1}{\lambda} \log(1-U) \quad \text{of} \quad X := -\frac{1}{\lambda} \log(U)$$

exponentieel verdeeld met parameter λ , want met U is ook $1-U$ gelijkverdeeld op $[0, 1]$.

Simulatie met behulp van de wegwerp methode

Soms is de inverse F^{-1} van de verdelingsfunctie $F(x)$ van een kansverdeling niet makkelijk te bepalen of zelfs onmogelijk expliciet op te schrijven. Het meest prominente voorbeeld hiervoor is de normale verdeling.

Maar we kunnen een kansverdeling met dichtheidsfunctie $f(x)$ op een eindig interval $[a, b]$ als volgt simuleren: Stel de dichtheidsfunctie is op het interval $[a, b]$ door een waarde c begrensd, d.w.z. $f(x) \leq c$ voor alle $x \in [a, b]$. Dan produceren we twee rijen toevalsgetallen, namelijk een rij (x_i) volgens een gelijkverdeling op $[a, b]$ en een rij (y_i) volgens een gelijkverdeling op $[0, c]$. We accepteren nu alleen maar de x_i voor de indices i waarvoor geldt dat $y_i \leq f(x_i)$ en werpen de andere x_i weg. In het plaatje hieronder zou men bijvoorbeeld x_1 verwerpen maar x_2 accepteren.



Het is niet moeilijk om in te zien dat de geaccepteerde toevalsgetallen x_i de dichtheidsfunctie $f(x)$ hebben, want een waarde $x_i = x$ wordt juist met kans $\frac{f(x)}{c}$ geaccepteerd. Om deze methode zo efficiënt mogelijk te maken, wordt de waarde c zo klein mogelijk gekozen, want de kans op het verwerpen van een getal x_i is juist de oppervlakte tussen de grafiek van $f(x)$ en de lijn $y = c$.

15.3 Simulatie van speciale verdelingen

Voor een aantal belangrijke kansverdelingen geven we nu aan hoe we met behulp van een randomgenerator U_1, U_2, \dots die toevalsgetallen op het interval $[0, 1)$ produceert, een stochast X met deze kansverdeling kunnen simuleren.

Discrete gelijkverdeling

Voor een eindige uitkomstenruimte Ω met $|\Omega| = n$ kunnen we aannemen dat $\Omega = \{0, \dots, n-1\}$. We krijgen een gelijkverdeling op Ω door

$$X := \lfloor n \cdot U_i \rfloor,$$

waarbij we met $\lfloor x \rfloor$ het grootste gehele getal $\leq x$ aangeven.

Bernoulli experiment en binomiale verdeling

We kunnen algemeen de uitkomst van een Bernoulli experiment met kans p simuleren door $X := \lfloor p + U_i \rfloor$, want $p + U_i$ is een gelijkverdeling op het verschoven interval $[p, 1 + p]$ en we hebben een waarde ≥ 1 met kans p .

Voor de binomiale verdeling $b(m, p; k)$ herhalen we m keer een Bernoulli experiment met kans p en krijgen zo als simulatie voor de binomiale verdeling:

$$X := \sum_{i=1}^m \lfloor p + U_i \rfloor.$$

Hypergeometrische verdeling

Om de hypergeometrische verdeling met parameters n , m en s te simuleren volgen we in principe de procedure van een echte proef. We noemen s_i het aantal slechte stukken die voor de i -de greep nog in de verzameling zitten en $p_i = \frac{s_i}{n}$ de kans dat we in de i -de greep een slecht stuk kiezen. Onze stochast X is het aantal slechte stukken die we grijpen. We beginnen dus met $X := 0$, $s_1 := s$ en $p_1 := \frac{s_1}{n} = \frac{s}{n}$ en voeren de volgende procedure voor $i = 1, 2, \dots, m$ uit:

Laat $A_i := \lfloor p_i + U_i \rfloor$ dan geeft $A_i = 1$ aan dat een slecht stuk werd getrokken, en $A_i = 0$ dat geen slecht stuk werd getrokken. We zetten nu $X := X + A_i$, $s_{i+1} := s_i - A_i$ en $p_{i+1} := \frac{s_{i+1}}{n}$.

Poisson-verdeling

Als m groot is kunnen we met behulp van de simulatie van de binomiale verdeling ook de Poisson-verdeling met parameter $\lambda = m \cdot p$ simuleren. Maar hiervoor maken we beter gebruik van het idee van een *Poisson-proces* die gewoon de tijdstippen van gebeurtenissen beschrijft die volgens een Poisson-verdeling optreden. Het cruciale punt is dat de tussentijden tussen twee gebeurtenissen van een Poisson-proces exponentieel verdeeld zijn en de parameter van deze exponentiele verdeling noemen we de *intensiteit* van het Poisson-proces. In het bijzonder heeft voor een Poisson-proces met intensiteit 1 het aantal waarnemingen in het tijdsinterval $[0, \lambda]$ een Poisson-verdeling met parameter λ .

Om een Poisson-verdeling met parameter λ te simuleren, moeten we dus het tijdsinterval $[0, \lambda]$ met exponentieel verdeelde tussentijden overdekken en tellen hoeveel tussentijden er nodig zijn. Hiervoor nemen we onafhankelijke stochasten Y_1, Y_2, \dots die exponentieel verdeeld zijn met parameter 1. Als we nu een stochast X definiëren door de eigenschap

$$\sum_{i=1}^X Y_i \leq \lambda < \sum_{i=1}^{X+1} Y_i$$

dan heeft X een Poisson-verdeling met parameter λ . Maar de Y_i kunnen we zo als boven gezien met behulp van een randomgenerator U_i simuleren door $Y_i := -\log(U_i)$ (de parameter van de exponentiële verdeling is 1), dus is $-\sum_{i=1}^X \log(U_i) \leq \lambda < -\sum_{i=1}^{X+1} \log(U_i)$ en dit is equivalent met

$$\prod_{i=1}^X U_i \geq e^{-\lambda} > \prod_{i=1}^{X+1} U_i.$$

We vermenigvuldigen dus uniform verdeelde toevalsgetallen tussen 0 en 1 tot dat het product kleiner is dan $e^{-\lambda}$, als dit na $X + 1$ getallen gebeurd is X een stochast met Poisson-verdeling met parameter λ .

Normale verdeling

Voor de normale verdeling bestaat er behalve de wegwerpmethode nog een andere mogelijkheid om tot een efficiënte simulatie te komen. Deze methode berust op een van de meest belangrijke stellingen in de kanstheorie:

Centrale limietstelling: Als X_1, X_2, \dots onafhankelijke stochasten zijn met verwachtingswaarde $E(X_i)$ en variantie $Var(X_i)$, dan is de limiet

$$X := \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n (X_i - E(X_i))}{\sqrt{\sum_{i=1}^n Var(X_i)}}$$

onder zwakke verdere voorwaarden aan de X_i een stochast met standaard-normale verdeling.

In het bijzonder wordt aan de voorwaarden voldaan als alle stochasten X_i dezelfde variantie σ^2 hebben, in dit geval convergeert $\frac{1}{\sqrt{n} \cdot \sigma} (\sum_{i=1}^n X_i - E(X_i))$ tegen de standaard-normale verdeling.

Voor de door de randomgenerator (U_i) gesimuleerde uniforme verdeling op $[0, 1)$ hebben we $E(X) = \frac{1}{2}$ en $Var(X) = \int_0^1 (x - \frac{1}{2})^2 dx = \frac{1}{12}$. Als we nu n waarden van de rij (U_i) optellen, hebben we $S_n = \sum_{i=1}^n U_i$ en er geldt $E(S_n) = \frac{n}{2}$ en $Var(S_n) = \frac{n}{12}$. Als benadering van een normale verdeling met verwachtingswaarde $\mu = 0$ en variantie $\sigma^2 = \frac{n}{12}$ krijgen we dus

$$X := \left(\sum_{i=1}^n U_i \right) - \frac{n}{2}.$$

en voor $n = 12$ geeft dit een uitstekende benadering van de standaard-normale verdeling die voor de meeste toepassingen voldoet.

Voorbeeld: Simulatie van het Monty-Hall probleem

We kijken tot slot naar een simulatie van het Monty-Hall probleem, om mensen die de theoretische argumenten niet accepteren door een experiment te kunnen overtuigen. De simulatie volgt de stappen in de show:

- (1) Kies een deur A waar de auto achter staat: $A := \lfloor 3 \cdot U_i \rfloor$ (we noemen de deuren 0, 1 en 2).
- (2) De kandidaat kiest een deur K : $K := \lfloor 3 \cdot U_i \rfloor$.
- (3) De moderator opent een deur M . Hier zijn twee gevallen mogelijk:
 - (i) $A = K$: in dit geval heeft de moderator de keuze tussen $A + 1$ en $A + 2$ (als we nummers van de deuren modulo 3 nemen) we nemen dus $M := A + \lfloor 2 \cdot U_i \rfloor + 1 \pmod 3$.

(ii) $A \neq K$: in dit geval heeft de moderator geen keuze, hij moet de deur M openen met $M \neq A$ en $M \neq K$.

(4) Hier zijn er twee versies:

(A) De kandidaat blijft bij zijn keuze, dus $K' = K$.

(B) De kandidaat wisselt van keuze, dus K' zo dat $K' \neq K$ en $K' \neq M$.

(5) Als $K' = A$ krijgt de kandidaat de auto, anders alleen maar de geit.

Dit kunnen we voor de versies A en B in stap (4) op een computer heel makkelijk 10000 keer doorspelen. Na drie herhalingen voor beide versies krijgen we bijvoorbeeld 3319, 3400 en 3327 successen voor versie A en 6583, 6655 en 6675 successen voor versie B .

Het blijkt dus ook uit het experiment dat het verstandig voor de kandidaat is om van keuze te wisselen.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- maximum likelihood schatting
- simulatie
- randomgenerator, toevalsgetallen
- methode van de inverse verdelingsfunctie
- wegwerp methode
- centrale limietstelling

OPGAVEN

75. Voor een gebeurtenis dat volgens een normale verdeling met dichtheidsfunctie

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

optreedt, zijn de observaties x_1, \dots, x_n gemaakt.

- (i) Bepaal de maximum-likelihood schatting voor de verwachtingswaarde μ als de variantie σ^2 bekend is.
- (ii) Bepaal de maximum-likelihood schatting voor de variantie σ^2 als de verwachtingswaarde μ bekend is.

(Opmerking: Als de verwachtingswaarde μ en de variantie σ^2 onbekend zijn, zijn de waarden uit (i) en (ii) de nulpunten van de partiële afgeleiden van de likelihood-functie en geven dus noodzakelijke voorwaarden voor een maximum van de likelihood-functie. Er laat zich aantonen dat men zo inderdaad een maximum vindt, dus laten zich μ en σ simultaan schatten.)

76. Laten U_1 en U_2 twee uniform verdeelde stochasten op $[0, 1)$ zijn. Laat zien dat $\sqrt{U_1}$ en $\max(U_1, U_2)$ dezelfde verdeling hebben. (Dit geeft een zuinige manier om het maximum van twee uniforme kansverdeling te simuleren.)

77. Een symmetrische *driehoeksverdeling* op het interval $[-1, 1]$ heeft de dichtheidsfunctie $f(x) = 1 - |x| = \begin{cases} 1 + x & \text{als } x < 0 \\ 1 - x & \text{als } x \geq 0 \end{cases}$.

Laten U en V twee stochasten zijn die uniform verdeeld zijn op het interval $[0, 1)$.

- (i) Laat zien dat de stochast $X_1 := U + V - 1$ de boven aangegeven driehoeksverdeling als dichtheidsfunctie heeft.
- (ii) Ga na dat de stochast $X_2 := U - V$ dezelfde kansverdeling als X_1 heeft. (We hebben dus twee manieren om de driehoeksverdeling met behulp van een randomgenerator te simuleren.)
- (iii) Laat zien dat X_1 en X_2 covariantie 0 hebben, d.w.z. dat $E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$ is.
- (iv) Toon aan dat X_1 en X_2 *niet* onafhankelijk zijn.

78. Bedenk en beschrijf een efficiënte simulatie voor het trekken van de lottogetallen.

Opgaven voor Kansrekening

Opgave 1.

Je hebt 4 verschillende wiskunde boeken, 6 psychologie boeken en 2 letterkundige boeken. Hoeveel manieren zijn er om deze twaalf boeken op een boord te plaatsen als:

- (i) je een genie bent en geen orde nodig hebt,
- (ii) je tenminste de wiskunde boeken naast elkaar plaatst,
- (iii) de boeken van elk vakgebied naast elkaar moeten staan?

Opgave 2.

Hoeveel verschillende getallen van 4 cijfers kan je uit de zestien hexadecimale 'cijfers'

$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F\}$ maken?

- (i) Hoeveel van deze getallen zijn 'echte' 4-cijfer getallen, dus hebben de eerste cijfer $\neq 0$?
- (ii) Hoeveel van de getallen uit (i) hebben vier verschillende cijfers?
- (iii) Hoeveel van de getallen uit (ii) eindigen op het cijfer 0?
- (iv) Hoeveel van de getallen uit (ii) hebben opstijgende cijfers?

Opgave 3.

Een Nederlands kentekenplaatje bestaat uit twee groepen van twee letters en een groep van twee cijfers. De groep van cijfers mag voor, tussen of achter de groepen met letters staan. Verder worden bij de letters geen klinkers gebruikt. Bepaal het aantal mogelijke nummerborden.

Opgave 4.

Uit een werkgroep van 8 mannen en 6 vrouwen moet een commissie van 3 mannen en 4 vrouwen gekozen worden. Hoeveel verschillende mogelijkheden bestaan er voor de commissie?

Opgave 5.

Een zekere faculteit heeft 6 hoogleraren, 8 UHD's, 4 UD's en 5 AIO's. In de feestcommissie van de faculteit zitten er 2 hoogleraren, 4 UHD's, 3 UD's en 3 AIO's. Hoeveel mogelijkheden zijn er voor de commissie? Hoe veranderd het aantal als een van de hoogleraren een begenadigde zanger en een van de UHD's (oorspronkelijk) een bierbrouwer is en deze twee per se in de commissie moeten zitten?

Opgave 6.

In een vaas zitten 8 rode, 3 witte en 9 blauwe knikkers. Je trekt drie keer een knikker zonder terugleggen. Bepaal de volgende kansen:

- (i) alle drie getrokken knikkers zijn rood,
- (ii) alle drie getrokken knikkers zijn wit,
- (iii) twee van de getrokken knikkers zijn rood, de derde is blauw,
- (iv) minstens een van de getrokken knikkers is wit,
- (v) bij de getrokken knikkers is een van elke kleur,
- (vi) de knikkers worden in de volgorde rood, wit, blauw getrokken.

Opgave 7.

Bij de schakers A en B wint A een spel gemiddeld met kans $\frac{1}{2}$ en B wint met kans $\frac{1}{3}$, de andere partijtjes eindigen in gelijkspel. Vanavond spelen A en B een match van drie spelen. Wat is de kans dat (a) A alle drie spelen wint, (b) twee spelen in een gelijkspel eindigen, (c) A en B afwisselend winnen, (d) B minstens een spel wint?

Opgave 8.

Een oneerlijke dobbelsteen is zo gemaakt dat 3 drie keer zo vaak valt als 4 en 2 twee keer zo vaak als 5. Verder vallen 1, 2, 3 en 6 even vaak.

- (i) Geef een kansverdeling voor het werpen van deze dobbelsteen aan.
- (ii) Bepaal de kans dat bij twee keer werpen van deze dobbelsteen de som minstens 11 is.

Opgave 9.

In Nijmegen zijn er 800 families met vijf kinderen. Hoeveel families met (a) 3 meisjes, (b) 5 meisjes, (c) 2 of 3 jongens verwacht je? (Je kunt ervan uit gaan dat er even veel jongens als meisjes geboren worden.)

Opgave 10.

Een rad van avontuur heeft vier sectoren waarin het rad met dezelfde kans tot stilstand komt. Het rad wordt gedraaid tot dat het in sector I stopt, maar hooguit 10 keer. Bepaal de kansen voor de volgende gebeurtenissen:

A_i : Het rad stopt bij de i -de draaiing in sector I.

B : Het rad stopt helemaal niet in sector I.

C : Het aantal draaiingen is even.

Opgave 11.

In een vaas zitten 7 witte en 1 rode knikkers. Je trekt herhaald een knikker, bekijkt de kleur en legt hem vervolgens terug. Bepaal de kans dat je bij 8 pogingen precies 3 keer de rode knikker pakt. Gebruik hiervoor (a) de binomiale verdeling, (b) de benadering door de Poisson-verdeling.

Hoe zit het met de resultaten als 15 witte en 1 rode knikker hebt en 16 pogingen doet? En hoe zit het bij 79 witte en 1 rode knikker en 80 pogingen?

Opgave 12.

In vaas I zitten 3 rode en 5 witte knikkers, in vaas II zijn er 4 rode en 2 witte. Er wordt een knikker willekeurig uit vaas I gegrepen en in vaas II gelegd. Vervolgens wordt er een knikker uit vaas II getrokken. Wat is de kans dat deze knikker wit is?

Opgave 13.

In een vaas zitten 3 rode en 2 blauwe knikkers, in een tweede vaas zitten 2 rode en 8 blauwe knikkers. Er wordt eerst een munt geworpen om te bepalen uit welke vaas een knikker getrokken wordt: als kop valt uit de eerste en als munt valt uit de tweede.

- (i) Bepaal de kans dat een rode knikker getrokken wordt.
- (ii) Stel dat je niet hebt gezien of kop of munt gevallen is, maar wel dat een rode knikker getrokken wordt. Wat is de kans dat kop is gevallen, dus dat de knikker uit de eerste vaas is getrokken?

Opgave 14.

Een huis is voorzien met een alarminstallatie. Als er een inbraak is, zal er met 96% kans een alarm komen, maar ook als er geen inbraak is, is er (door aardbevingen of andere storingen) met een kans van 0.1% een alarm. In de woonwijk van het huis is de kans op een inbraak 0.3%. Vannacht is er een alarm. Hoe groot is de kans dat er daadwerkelijk een inbraak plaats vindt?

Opgave 15.

Er wordt twee keer met een eerlijke dobbelsteen gedobbed. De uitkomsten A , B en C zijn:

A : Er wordt twee keer hetzelfde getal gedobbed.

B : Het eerste getal is 1 of 6.

C : Het tweede getal is even.

Zijn A, B, C onafhankelijk? Zijn de uitkomsten paarsgewijs onafhankelijk?

Opgave 16.

De goedkope random-trein vertrekt op een willekeurig tijdstip tussen 10.00 en 10.30 uur. Je beslist zelf ook op een willekeurig tijdstip in dit half uur op het station op te dagen en hooguit 5 minuten te wachten. Als de trein in dit interval niet komt, pak je een taxi om nog op tijd naar het college te komen. Wat is de kans dat je met de trein zult rijden?

Opgave 17.

Een stok wordt willekeurig in twee stukken gehakt. Wat is de kans dat het ene stuk minstens twee keer zo lang is als het andere?

Opgave 18.

Op een stok van $1m$ lengte kies je willekeurig twee punten en hakt vervolgens de stok op deze twee punten door. Wat is de kans dat je uit de drie zo verkregen stukken een driehoek kunt vormen? (Je kunt een driehoek vormen dan en slechts dan als de som van twee stukken altijd groter is dan het derde stuk.)

De net verkregen kans vind je te klein, daarom kies je voor een andere aanpak. Eerst hak je de stok op een willekeurig punt in twee delen en vervolgens pak je het grotere stuk en hak dit nog eens op een willekeurig gekozen punt door. Wat is nu de kans dat je uit de drie stukken een driehoek kunt vormen?

Opgave 19.

De kans dat een student bij het grote lustrumfeest een bier krijgt is 99.2% (soms is het bier op, soms denkt de baas dat de student geen 16 jaar oud is). Een slimme verzekeringsmaatschappij biedt eenmalig een verzekeringspolis, waar je voor een premie van 10 € tegen bierarmoede verzekerd bent. In het geval dat je inderdaad geen bier op het feest krijgt betaalt de verzekering 1000 €. Wat is de verwachte winst van de verzekeringsmaatschappij bij elk afgesloten polis?

Opgave 20.

Er wordt met twee (eerlijke) dobbelstenen gedobbeld. De stochast X beschrijft het maximale getal in een worp. Bereken $P(X = k)$ voor $k = 1, \dots, 6$ en de verwachtingswaarde $E(X)$.

Bekijk hetzelfde probleem voor drie dobbelstenen.

Opgave 21.

Bij een bloedtest van 10 personen is bekend dat 2 een zeker virus in hun bloed hebben. Om het aantal tests in te krimpen wordt te volgende methode toegepast: De 10 personen worden willekeurig in twee groepen van 5 personen ingedeeld. Het bloed van de personen in een groep wordt vermengd en getest. Als het virus in het mengsel gevonden wordt, wordt het bloed van elk persoon in de groep apart getest.

Beschrijf een geschikte ruimte Ω met een kansverdeling P , zo dat het aantal van bloedtests een stochast op deze kansruimte is. Bereken de verwachtingswaarde voor het aantal bloedtests.

Opgave 22.

Een winnaar bij een televisiequiz krijgt een van drie prijzen. De prijzen worden één voor één aangeboden en de winnaar mag of een prijs accepteren (dan worden de verdere prijzen niet meer getoond) of weigeren (dan mag hij op deze prijs niet meer terug komen). Stel dat de winnaar een duidelijke rangschikking voor de prijzen heeft. Er zijn nu twee voor de hand liggende strategieën:

- (A) Kies in ieder geval de eerste prijs.
- (B) Als de tweede prijs aantrekkelijker is dan de eerste, kies de tweede, als niet, kies de derde.

Bepaal voor elk van de twee strategieën de kansen om de meest aantrekkelijke en de minst aantrekkelijke prijs te kiezen.

Opgave 23.

Er wordt twee keer met een dobbelsteen gedobbeld. De stochast X_1 beschrijft het aantal ogen in de eerste worp, de stochast X_2 het aantal ogen in de tweede. Verder geeft $U := \min(X_1, X_2)$ het kleinste en $V := \max(X_1, X_2)$ het grootste van de twee aantallen.

- (i) Zijn U en V onafhankelijk?
- (ii) Bepaal de kansverdeling van U .
- (iii) Bepaal de verwachtingswaarden $E(U)$ en $E(U + V)$.
- (iv) Bepaal de voorwaardelijke kans $P(X_1 = 3 \mid U = 3)$.