Dictaat bij studieonderdeel:

Data-acquisitie & Toegepaste analyse

NS-109B

2015/16

DATA-V

Julius Instituut
Departement Natuur- en Sterrenkunde
Faculteit Bètawetenschappen
Universiteit Utrecht

Eindredactie: Peter van Capel
Aan eerdere versies van dit dictaat zijn grote bijdragen geleverd door Erik Langereis en Wim Westerveld.

Inhoudsopgave

Hoofdstuk 1

Data-acquisitie & Toegepaste analyse

In de natuurkunde kom je voortdurend in aanraking met kwantitatieve (numerieke) gegevens. Zowel het *verwerven* als het *verwerken* van kwantitatieve informatie (respectievelijk het meetproces en de analyse van de verkregen data) is in veel gevallen een complexe aangelegenheid.
Het studie-onderdeel "Data-acquisitie & Toegepaste analyse (DATA)" geeft een solide basis in het verwerven en verwerken van kwantitatieve data, aan de hand van experimenten met een natuurkundige context. Het is samengesteld uit DATA-P (het introductiepracticum natuurkunde), DATA-V (de statistische verwerking van kwantitatieve data) en DATA-M (de beginselen van het softwarepakket *Mathematica*).[1]
In het onderdeel DATA-V komt de inhoud van dit dictaat aan de orde. In DATA-V is in het bijzonder aandacht voor (de gevolgen van) meetonzekerheden en de kritische beoordeling van eindresultaten die door middel van meting of berekening zijn verkregen. De leerdoelen van DATA-V: de student

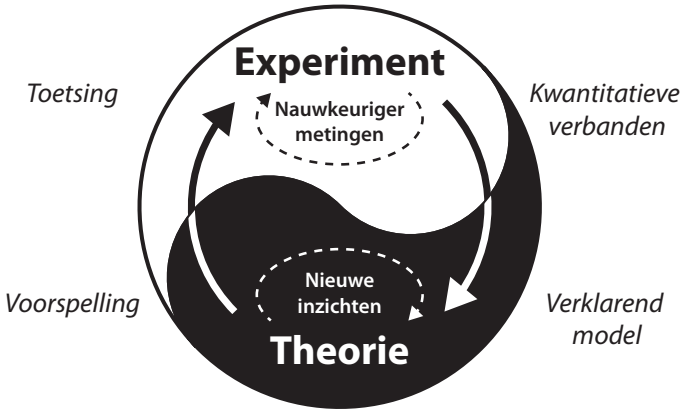
- ziet het belang en de consequenties van het toekennen van onzekerheden aan data.
- kan kwantitatieve gegevens op een statistisch verantwoorde manier verwerken en interpreteren. Belangrijke technieken hierbij zijn elementaire statistiek, propagatie van onzekerheid, het gebruik van kansverdelingsfuncties, het maken van aanpassingen, en het berekenen van overschrijdingskansen.
- begrijpt de beginselen, aannames en beperkingen van de hiervoor genoemde technieken.

Veel van de dataverwerkingstechnieken die in dit dictaat worden behandeld, worden in het *Mathematica*-materiaal geïllustreerd.
Dit hoofdstuk geeft een algemene inleiding over kwantificeren in de empirische wetenschappen en een overzicht van de stof in het vervolg van dit dictaat.

1.1 De empirische wetenschappen

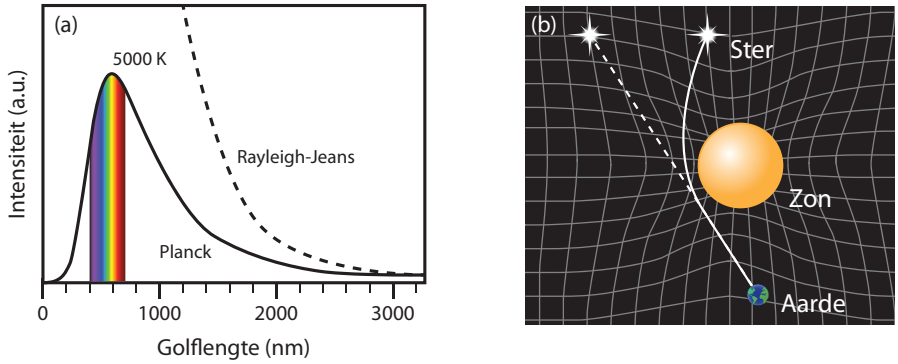
Natuurkunde is een empirische wetenschap (empirie = op basis van waarneming). Uitgangspunt is hierbij dat, hoewel het waarnemen voor iedere waarnemer een unieke en subjectieve gewaarwording is, we ons toch bevinden in een waarnemer-onafhankelijke, objectieve werkelijkheid. Die werkelijkheid ("de natuur") vertoont vaste gedragingen en patronen ("wetmatigheden"). Het doel van de natuurkunde is deze gedragingen te begrijpen.
Het is van belang om je te realiseren dat in dit proces van begrijpen, idee- en theorievorming in voortdurende wisselwerking is met proefondervindelijke waarneming, en dat het één niet zonder het ander kan.

[1] Meer informatie over de leerdoelen van het vak DATA en de rol binnen de leerlijn Experimentele Fysica is te vinden in de Algemene Studenteninstructie bij het Practicum Natuurkunde. Gedetailleerde informatie over inhoud, roostering en beoordeling is te vinden in de Blackboard-omgeving van dit vak.



Figuur 1.1: Schets van het proces van begripsvorming. In beginsel start een wetenschappelijk experiment meestal op een *kwalitatieve* manier: "beschrijf zo nauwkeurig mogelijk wat je hebt waargenomen". Begripsvorming vindt plaats als we de waarnemingen kwantificeren en kunnen verzamelen in *kwantitatieve verbanden* tussen *grootheden*. Wanneer we een logisch principe van werking ontwikkelen dat consistent is met bestaande theoretische modellen kunnen we de gevonden kwantitatieve verbanden verzamelen in een *verklarend model*, essentieel voor verdergaande begripsvorming. Dit model zal *voorspellingen* genereren voor het gedrag van (fysische) systemen. Wanneer vervolgens experimenten worden ontworpen waarin dit gedrag kan worden onderzocht, kan het theoretische model worden *getoetst*. Hiermee is de cirkel rond: een falsificatie van het ontworpen model moet leiden tot nieuwe modellen. Uiteraard is het mogelijk om onafhankelijk van elkaar experimentele resultaten en theoretische modellen intrinsiek te verbeteren (door het gebruik van respectievelijk nauwkeuriger meetmethodes en het gebruik van nieuwe theoretische inzichten of wiskundige technieken), wat eveneens tot nieuwe kennis kan leiden.

Waarnemingen zonder theorie geven hooguit ervaringskennis. Theorie zonder experimentele verificatie geeft aanleiding tot speculatie en leidt niet tot begrip.
In Figuur 1.1 is het proces van begripsvorming geschetst. Dit plaatje impliceert niet dat een natuurkundige slechts óf de theorie óf het experiment beheerst. Het beheersen van een zo groot mogelijk deel van het begripsvormingsproces is juist voordelig in het forceren van wetenschappelijke doorbraken. Het is de combinatie van theorie en experiment in de wetenschappelijke methodiek die sinds de vroege Renaissance tot spectaculaire vooruitgang heeft geleid. De geschiedenis laat zien dat de baanbrekende vooruitgang kan worden geïnitieerd vanuit zowel experiment als theorie.
Enerzijds kan een in eerste instantie onverklaarbaar experimenteel resultaat leiden tot nieuw theoretisch inzicht. De experimenteel waargenomen stralingscurve van de zwarte straler (zie Figuur 1.2 (a)) gaf aan dat bij korte golflengtes steeds minder straling wordt uitgezonden, in complete tegenstelling met de klassieke Rayleigh-Jeans-theorie. Het leidde Max Planck in 1900 tot de aanname dat elektromagnetische straling uit discrete kwanta (fotonen) bestond, wat het begin vormde van de ontwikkeling van de kwantummechanica.
Anderzijds worden spectaculaire nieuwe theorieën (soms jaren) later pas experimenteel bevestigd. Een bekend voorbeeld is de in 1915 door Albert Einstein gepubliceerde algemene relativiteitstheorie, die voorspelt dat licht door de aanwezigheid van een zeer zware massa wordt afgebogen, zogenaamde *gravitational lensing* (zie Figuur 1.2 (b)). De theorie was controversieel, maar dat veranderde snel toen het effect (al) in 1919 experimenteel geverifieerd werd door tijdens een zonsverduistering de afbuiging van sterrenlicht ten gevolge van de zwaartekracht van de zon te meten en te vergelijken met de berekende waarde.
Bij modern onderzoek zoals de zoektocht naar het Higgs-boson is bij uitstek het empirische karakter zichtbaar: de koppeling tussen theoretische modelvorming en de experimentele waarneming. Op theoretische gronden werd voorspeld dat het Higgs-boson een massa zou hebben binnen een bepaald massagebied, waar het uiteindelijk in 2012 "gevonden" is in de metingen met de *Large Hadron Collider* (LHC) van het CERN.
Het ontwikkelen van theorie is het opstellen van een *verklarend model*, dat uitlegt waarom een breed scala aan (samenhangende) experimentele observaties optreedt. Een belangrijk aspect bij theorievorming in de



Figuur 1.2: Twee wetenschappelijke inzichten in de natuurkunde. (a) De experimentele gemeten spectrale emissiecurve van de zwarte straler leidend tot de theorie van de kwantummechanica. (b) De theoretische voorspelling door de algemene relativiteitstheorie over de kromming van ruimtetijd ten gevolge van een zware massa, experimenteel geverifieerd door metingen van de schijnbare positie van sterren (tijdens een zonsverduistering) die duidden op afbuiging van licht door de zwaartekracht van de zon.

empirische wetenschappen is dat de theorie *toetsbaar* is, waarmee bedoeld wordt dat er (in principe) experimenten uitgevoerd moeten kunnen worden waarmee de geldigheid van de theorie aangetoond of weerlegd kan worden.
Omgekeerd zijn experimentele resultaten niet zonder meer te verwerpen op alleen theoretische basis, mits je er zeker van kunt zijn dat deze resultaten correct zijn. Bij twijfel aan experimentele resultaten biedt uitsluitend experimentele verificatie uitkomst. Bij het doen van experimenten is onder andere van belang dat de resultaten *waarnemer-onafhankelijk* en *reproduceerbaar* zijn, wat zoveel wil zeggen dat het niet uitmaakt wie het experiment uitvoert en hoe dat is gedaan. Een cruciaal begrip bij experimenteel werk is *betrouwbaarheid*: hoe nauwkeurig zijn de verkregen resultaten en hoe zeker ben je ervan dat je in het experiment geen fouten hebt gemaakt?
In zowel theorie als experiment is *kwantificeren* essentieel. Dit betekent dat eigenschappen die aan de natuur worden toegekend getalsmatig worden uitgedrukt. Het grote voordeel dat een exacte wetenschap als natuurkunde heeft boven bijvoorbeeld de sociale wetenschappen, is dat de begripsvorming kan worden opgetild van een kwalitatieve beschrijving naar een beschrijving in termen van wiskundige relaties. In dit studieonderdeel zullen we veel nadruk leggen op het proces van kwantificeren van resultaten.

1.2 Onzekerheid

Van meetresultaten kun je nooit zeggen dat ze *exact* zijn. Door allerhande effecten kent *elk* gerapporteerde waarde voor een grootheid een (toevallige) *onzekerheid*, dat wil zeggen een beperkte nauwkeurigheid. Deze onzekerheid hangt samen met een afleesnauwkeurigheid of een eindige meetnauwkeurigheid van het gebruikte meetapparaat, of variaties in de waarde van de grootheid ten gevolge van bijvoorbeeld verstorende omgevingseffecten.
De notatie van een meetresultaat is in zo'n geval $x \pm \sigma_x$, waarbij x de gemeten waarde voorstelt, en σ_x de onzekerheid in de gemeten waarde x. De plus-minus $\pm$ dient om aan te geven dat we verwachten dat de "echte" waarde (die we niet exact kunnen bepalen) een kans kent om in een zekere bandbreedte rondom de gemeten waarde te liggen, waarover later meer.
De onzekerheid wordt in de volksmond ook wel aangeduid met de term meet*fout* (in het Engels wordt deze onzekerheid ook wel aangeduid met *error*). Deze terminologie suggereert onterecht dat de onderzoeker iets verkeerd gedaan heeft. Het correct bepalen of afschatten van de onzekerheid in resultaten is juist een

van de belangrijkste aspecten in de empirische wetenschappen. Het uitsluiten van alternatieve theoretische modellen kan alleen op basis van zorgvuldige evaluatie van kwantitatieve resultaten én hun onzekerheden.

Higgs-boson In de experimentele zoektocht naar het Higgs-boson bij de *Large Hadron Collider* (LHC) van het CERN is consciëntieus rekening gehouden met alle factoren die een bijdrage leveren aan de nauwkeurigheid van de metingen. In het persbericht over het aantonen van het bestaan van het deeltje in het *Compact Muon Solenoid* (CMS) experiment wordt dan ook uitermate zorgvuldig geformuleerd wat de kans zou zijn dat het gemeten signaal een (toevallige) fluctuatie van het achtergrondsignaal zou zijn:

"*CMS observes an excess of events at a mass of approximately 125 GeV with a statistical significance of five standard deviations (5 sigma) above background expectations. The probability of the background alone fluctuating up by this amount or more is about one in three million*". *http://cms.web.cern.ch/node/1381*

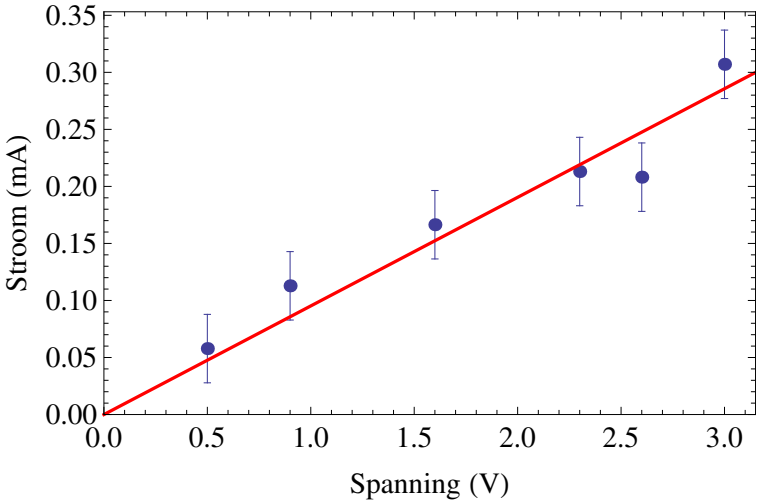
Bij herhaling van de meting en bijvoorbeeld middeling van meerdere meetresultaten kan bij aanwezigheid van toevallige onzekerheden een nauwkeuriger resultaat verkregen worden. Hierbij gaat de statistische wetmatigheid op dat de betrouwbaarheid van een uitspraak of de nauwkeurigheid van een resultaat omhoog gaat naarmate er meer meetresultaten (met elk een toevallige onzekerheid) beschikbaar zijn om die uitspraak of dat resultaat op te baseren. Inzichten vanuit de mathematische statistiek (Hoofdstukken 3 en 4) zijn belangrijk om tot verantwoorde ("beste") resultaten te komen bij een gegeven verzameling meetresultaten.

Gravity Probe B Een recent voorbeeld van extreem nauwkeurige metingen is het *Gravity Probe B* experiment, waarmee na zeer zorgvuldig werk over vele jaren het meesleepeffect (hoe de draaiende aarde lokaal de ruimte-tijd vervormt) voorspeld door de algemene relativiteitstheorie is aangetoond. Er is veel informatie over dit experiment te vinden, zie bijvoorbeeld *http://einstein.stanford.edu/*.

1.3 Dataverwerkingstechnieken

Bij veel indirecte meetmethoden wordt een andere eigenschap gemeten dan die waarover je informatie wilt hebben. Denk aan het eerder genoemde voorbeeld van de constante van Planck. Dat betekent dat er op de één of andere manier een vertaalslag gemaakt moet worden van de meetresultaten (inclusief hun onzekerheden) naar de grootheid waarin je geïnteresseerd bent. Daarnaast moeten de resultaten gecommuniceerd worden, in overzichtelijke en overtuigende vorm. Bij de verwerking en rapportage van resultaten staan je een aantal technieken en hulpmiddelen ter beschikking, die in DATA-V worden geoefend en worden toegepast in DATA-P, latere practicumonderdelen en het afstudeeronderzoek.

Grafische weergave Zeer veel fysische experimenten volgen de volgende methodiek: prepareer een fysisch systeem, varieer één grootheid (de *onafhankelijke* variabele, vaak met verwaarloosbare onzekerheid), en noteer de bijbehorende waarden van een grootheid die hiervan afhangt (de *afhankelijke* variabele). Als voorbeeld geven we metingen van de stroom I door een weerstand R als functie van de spanning V over de weerstand, getoond in Figuur 1.3. Met behulp van een dergelijke grafiek wordt het verband tussen beide grootheden inzichtelijk gemaakt.
Grafieken (en in mindere mate tabellen) zijn essentiële hulpmiddelen in de verwerking en communicatie van kwantitatieve resultaten. Met een doordachte grafische weergave worden de gemeten verbanden vaak in één oogopslag duidelijk. Een efficiënte communicatie stelt wel eisen aan opmaak en inhoud van figuur en tabel. Deze worden in Hoofdstuk 2 besproken.



Figuur 1.3: Gemeten waarden voor de elektrische stroom I uitgezet tegen de ingestelde spanning V. De richtingscoëfficiënt is gelijk aan $1/R$. De lijn geeft een *lineaire aanpassing* aan de data weer. In dit geval is het verschil tussen de gemeten data (V,I) en de aanpassing minimaal voor een waarde $R = 9.8$ kΩ. De nauwkeurigheid waarmee R zo bepaald wordt bedraagt ± 0.6 kΩ.

Aanpassingen Het verband tussen grootheden kan gekwantificeerd worden door een *aanpassing* uit te voeren. Voor de data in Figuur 1.3 is het aangegeven functionele verband lineair, en de richtingscoëfficiënt α gelijk aan $1/R$ (de "wet" van Ohm: $I = V/R$). Doordat de gemeten waarden voor zowel V als I toevallige onzekerheden (kunnen) hebben, is het lineaire verband van de meetresultaten nooit exact (een getrokken rechte lijn zal nooit precies door alle datapunten heen lopen). Wel kan ervoor gezorgd worden dat het "verschil" tussen de gemeten data en een lineaire relatie tussen V en I voor een "beste" waarde van α zo klein mogelijk gemaakt wordt. Het op de juiste, kwantitatieve manier vaststellen van het lineaire verband op basis van de beschikbare data wordt een (lineaire) aanpassing genoemd, en wordt besproken in Hoofdstuk 6. Aan het eind van de cursus komen ook niet-lineaire aanpassingen aan de orde.

Propagatie van onzekerheid Het zal vaak voorkomen dat het meetresultaat voor één grootheid met behulp van een theoretisch verband moet worden omgerekend naar een waarde van een andere grootheid. Als voorbeeld kijken we weer naar de data in Figuur 1.3. Stel nu dat de aanpassing een waarde $\alpha \pm \sigma_\alpha$ oplevert voor de richtingscoëfficiënt. We weten dat $\alpha = 1/R$. Het berekenen van de waarde van R uit de waarde van α is triviaal. Maar hoe je de onzekerheid σ_R uit σ_α berekent is dat niet. Zo zal je eerste (naïeve) gedachte misschien zijn dat $\sigma_\alpha = 1/\sigma_R$. Maar dit kan niet kloppen. Dat zou namelijk betekenen dat de onzekerheid in de weerstandswaarde zeer klein wordt wanneer de richtingscoëfficiënt extreem slecht bepaald is (σ_α zeer groot).
Nog veel complexer wordt het wanneer meerdere resultaten (bijvoorbeeld voor verschillende grootheden), die elk een onzekerheid kennen, gecombineerd moeten worden voor de berekening van een eindantwoord. Hoe de onzekerheden in deze afzonderlijke metingen correct kunnen worden doorgerekend in de onzekerheid van het eindantwoord (zogenaamde *propagatie* van onzekerheid) komt aan de orde in Hoofdstuk 5.

Hoofdstuk 2

Meten: grootheden, eenheden en onzekerheid

2.1 Inleiding

Er is een groot conceptueel verschil tussen "waarnemen" (het observeren, beschrijven en ordenen van fenomenen) en "meten". Wanneer we praten over meten, dan voegen we aan de waarneming toe dat deze *objectief* (waarnemer-onafhankelijk) en *kwantitatief* (uitgedrukt in een numerieke waarde) is. Hiermee wordt het mogelijk resultaten te vergelijken en communiceren met anderen. Door deze extra eisen aan een waarneming komen we meteen bij een belangrijk aspect van experimentele natuurkunde, namelijk *standaardisatie*: eenduidige afspraken over definities en hoeveelheden.
In dit hoofdstuk gaan we in op het stelsel van grootheden en eenheden dat een uiting is van deze standaardisatie. Verder bekijken we het meetproces en hoe dit leidt tot het optreden van toevallige onzekerheden en (mogelijk) systematische effecten. Tot slot staan we uitvoerig stil bij de richtlijnen voor (eind)rapportage van resultaten, zowel in numerieke vorm, grafische weergave als presentatie in tabelvorm.

2.2 Grootheden

Om een natuurkundig verschijnsel te beschrijven en begrijpen introduceren we allereerst een geschikte "taal". De natuurkunde gebruikt als "taal" een eenduidige set fysische *grootheden*. Een *fysische grootheid* van een fysisch systeem of object is een meetbare eigenschap die in de natuurwetenschap aan zo'n systeem of object wordt toegekend.[1]
Om alle fysische grootheden op een zo efficiënt mogelijke manier vast te leggen zijn in het *Système international d'Unités* oftewel *SI-stelsel*, het internationaal afgesproken systeem van grootheden en eenheden, zeven *basisgrootheden* vastgelegd (Tabel 2.1). Op basis van deze basisgrootheden worden *afgeleide grootheden* gedefinieerd als een (functionele) samenstelling van basisgrootheden. Zo is bijvoorbeeld de afgeleide grootheid snelheid het quotiënt van de grootheden lengte en tijd, en de grootheid oppervlakte het kwadraat van de grootheid lengte.

[1] Het is een filosofische vraag of het systeem of object deze toegekende eigenschap ook intrinsiek "bezit", of dat een grootheid slechts een door ons op het systeem geplakt label is om daarmee een deel van het waargenomen gedrag te beschrijven. Deze vraag valt in dezelfde categorie als de vraag of een elektron werkelijk bestaat (omschreven als *realisme*), of slechts een begrip is dat handig is om in een bepaalde situatie waargenomen gedrag te beschrijven of in een theorie te hanteren. Hoewel in de filosofie een belangrijk onderwerp van debat, maakt het voor de dagelijkse toepassing niet zoveel uit welk standpunt je inneemt. In dit vak houden we ons op de vlakte en beschouwen we grootheden en objecten (als elektronen) louter vanuit een experimentele, meettechnische context. Dit weinig controversiële standpunt komt het meest overeen met wat omschreven wordt als *instrumentalisme*.

Tabel 2.1: Lijst van basisgrootheden en -eenheden in het SI-stelsel. Meer informatie over alle grootheden en eenheden binnen het SI-stelsel kan gevonden worden in *si_brochure_8_en.pdf* van het BIPM, ook beschikbaar via Blackboard.

Basisgrootheid	**Symbool**	**Dimensie**	**Basiseenheid**	**Symbool eenheid**
Lengte	$l, x, \cdots$	L	meter	$[l] = \mathrm{m}$
Massa	m	M	kilogram	$[m] = \mathrm{kg}$
Tijd	t	T	seconde	$[t] = \mathrm{s}$
Elektrische stroomsterkte	I, i	I	ampère	$[I] = \mathrm{A}$
Thermodynamische temperatuur	T	Θ	kelvin	$[T] = \mathrm{K}$
Hoeveelheid	n	N	mol	$[n] = \mathrm{mol}$
Lichtsterkte	I_v	J	candela	$[I_v] = \mathrm{cd}$

Grootheden kunnen worden onderscheiden naar *soort*. Twee grootheden zijn van dezelfde soort (of: zijn equivalent) als ze direct met elkaar kunnen worden vergeleken (groter dan, gelijk aan, kleiner dan). Een lengte en een oppervlak zijn dus niet van dezelfde soort; een lengte en een breedte natuurlijk wel (bijvoorbeeld "de lengte van een huis is meestal groter dan de breedte"). De zeven basisgrootheden zijn uiteraard allemaal van verschillende soort.

Een fysische grootheid heeft pas betekenis als deze kwantificeerbaar is, met andere woorden, dat er een meetmethode is te definiëren waarmee haar waarde in principe kan worden bepaald. Het moet zelfs mogelijk zijn de *definitie van de grootheid* in de vorm van een *meetvoorschrift* te geven. Zo'n voorschrift kan betrekking hebben op een zeer gecompliceerde indirecte meting, maar moet in principe tot reproduceerbare resultaten leiden.

Notatie

- Bij opgaaf van kwantitatieve resultaten wordt eerst de grootheid gespecificeerd, vaak met een letter. Voor veel grootheden zijn standaardletters in gebruik, en het helpt om daarbij aan te sluiten. Een fysische grootheid wordt standaard in *schuingedrukte* vorm (*italics*) weergegeven. Als voorbeeld: de versnelling van de zwaartekracht wordt aangegeven als g.

Dimensie Gerelateerd aan de soort van de grootheid is het concept *dimensie*. Alhoewel het begrip dimensie zeker van nut is bij het tot stand komen van een internationaal stelsel van afspraken, is het algemene concept dimensie moeilijk op een zinvolle manier te definiëren. Voordat we verder gaan met het stelsel van eenheden, is het toch belangrijk om kort stil te staan bij dit concept.

Grofweg zou je kunnen stellen dat de dimensie van een grootheid een soort concepteenheid is, waarvoor geen afspraken hoeven te worden gemaakt over hoe deze concepteenheid eigenlijk gestandaardiseerd is (Sectie 2.3). Bijvoorbeeld de grootheid lengte heeft als dimensie L, maar om feitelijk numerieke waarden voor de grootheid lengte te realiseren zullen we vervolgens moeten definiëren welke eenheid we hanteren (de meter, de inch, een lichtjaar, een vadem, etc).

Een voordeel van dimensies is dat je met grootheden kunt werken of rekenen, zonder te hoeven afspreken in welk eenhedenstelsel er eigenlijk gewerkt wordt. Dimensies van grootheden worden gebruikt om onafhankelijk van de gehanteerde eenheden af te leiden welke fysische relatie er tussen grootheden zou kunnen gelden (zoals gesteld in het zogenaamde Buckingham-π-theorema). Zonder sluitende dimensie-analyse kan de gestelde relatie niet juist zijn.

Een probleem met dimensies In het artikel *On the concept of dimension* behandelt W.H. Emerson uitgebreider (de verwarring over) het concept dimensie.[a] Emerson geeft als treffend voorbeeld de grootheid benzinegebruik van een voertuig per gereden afstand, meestal uitgedrukt in liters/kilometer. De dimensie hiervan is $L^3/L = L^2$. Deze grootheid is dan van dezelfde dimensie als de grootheid oppervlakte. De twee grootheden kunnen echter op geen enkele manier zinnig met elkaar vergeleken worden (oftewel, zijn niet van dezelfde *soort*). Bijzonder problematisch is ook het concept van *dimensieloze* grootheden, zoals bijvoorbeeld de hoekmaat (verhouding van twee lengtematen met dimensie L/L = 1).
De conclusie moet zijn dat voor vergelijkbaarheid van grootheden meer vereist is dan het hebben van dezelfde dimensie.

[a] Originele bron: *Metrologia* **42** L21 (2005).

De gehanteerde symbolen voor de dimensie van de basisgrootheden zijn opgenomen in Tabel 2.1. Dimensies van afgeleide grootheden zijn uit te drukken als het product van machten van dimensies van basisgrootheden. Om een voorbeeld te geven: de dimensie van dichtheid (massa per volume-eenheid) is bijvoorbeeld $M\,L^{-3}$.

2.3 Standaarden en eenheden

Dat een fysische grootheid kwantificeerbaar is, leidt er automatisch toe dat er een *waarde* of *hoeveelheid* aan kan worden toegekend. Hiertoe is het noodzakelijk dat er een objectieve *maat* of *standaard* gebruikt wordt. In de "Van Dale" (12^e herziene druk, 1995) is "meten" dan ook gedefinieerd als: "bepalen hoeveel malen een zekere grootheid (de maat) in een voorwerp begrepen is".
Er wordt dus een vaste referentie voor de grootheid, de *standaard*, gekozen. De *getalwaarde* is dan de verhouding tussen de betreffende grootheid en de gekozen standaard. Een meting van een grootheid resulteert aldus in een getal voor de waarde van die grootheid gekoppeld aan de gebruikte standaard. Hiermee wordt de meting communiceerbaar en in principe ook testbaar op reproduceerbaarheid.
Bij een bepaling van de grootheid lengte is de standaard bijvoorbeeld een schuifmaat (zie Figuur 2.1 (a)). De aanduiding op de schuifmaat is meestal in de *eenheid* centimeter of millimeter, die respectievelijk een honderdste of duizendste zijn van de eenheid meter.
Om een goede standaard zoals de schuifmaat te verkrijgen, wordt deze in een ijkprocedure gecalibreerd ten opzichte van een zeer nauwkeurig bekende referentiestandaard. Door heldere internationale afspraken over de definitie van de meter en een deugdelijk beheer van standaarden kunnen we ervan uitgaan dat twee verschillende fysici, die twee verschillende lengtemetingen hebben gedaan, met twee verschillende schuifmaten, toch hun resultaten kwantitatief kunnen vergelijken.
Voor een constructieve vorm van wetenschapsbeoefening moet zoveel mogelijk met algemeen erkende en toegankelijke referentiestandaarden gewerkt worden. De afspraken rondom standaarden in internationaal verband zijn vastgelegd in het eerder genoemde *SI-stelsel*, en worden algemeen in de wetenschapsbeoefening erkend. In Nederland is het beheer van referentiestandaarden in handen van het Nederlands Meetinstituut (NMi, *http://www.nmi.nl/*).
De basiseenheden behorend bij de zeven basisgrootheden zijn weergegeven in Tabel 2.1. Men geeft de eenheid vaak aan met vierkante haken om het symbool voor de grootheid, dus b.v. voor tijd: $[t] = \mathrm{s}$.

2.3.1 Standaardmaten

Een logische eerste stap voor de definitie van zo'n basiseenheid is het vastleggen van fysieke *standaardmaten* voor de genoemde basisgrootheden. Dit is bijvoorbeeld zo voor de kilogram als eenheid van massa. De primaire standaard, dé kilogram, is een platina-iridium blok, beheerd door het *Bureau international des*

poids et mesures (BIPM) in Parijs. Van deze zogenaamde primaire standaard zijn alle nationaal gebruikte standaarden afgeleid. Hoewel fysieke standaarden lang voldeden in praktische zin, zijn met de voortschrijdende eisen aan nauwkeurigheid fundamentele en praktische problemen te constateren. Wanneer bijvoorbeeld de Nederlandse afgeleide standaard moet worden geijkt, moet deze eerst naar Parijs worden gebracht en met de primaire standaard worden vergeleken. Door dit proces kan zowel de primaire als de afgeleide standaard slijten, met alle gevolgen van dien. Daarnaast kun je je bij de mechanische grootheden lengte, tijd en massa nog wel een fysieke maat voorstellen. Maar het is niet zo dat voor alle basiseenheden op eenvoudige wijze een primaire standaard te construeren valt. Hoe zou zo'n maat voor de grootheid elektrische stroomsterkte er bijvoorbeeld uit moeten zien?

2.3.2 Standaardeenheden en meetvoorschriften

Behoudens de kilogram zijn de basiseenheden niet (meer) gekoppeld aan een fysieke maat, maar aan een *meetvoorschrift* voor de *standaardeenheid*. In dit meetproces spelen fundamentele fysische constanten, zoals de constante van Planck h en de elementaire lading e een steeds grotere rol. Deze constanten worden fundamenteel genoemd omdat zij overal en (voor zover wij weten) onveranderlijk hetzelfde zijn. Meestal spreken we in dat geval over de geaccepteerde waarde. Voor een lijst van fundamentele natuurconstanten zie *http://physics.nist.gov/cuu/Constants/*.
Het ideaal hierbij is dan om de basiseenheden te definiëren in termen van alleen fundamentele fysische constanten. Om dit programma te kunnen uitvoeren zijn er twee essentiële elementen vereist, namelijk:

- het ontdekken van geschikte fysische verschijnselen met meetbare grootheden die op een eenvoudige manier van alleen fundamentele natuurconstanten afhangen;
- het ontwikkelen van instrumenten en meettechnieken om deze grootheden met zeer grote nauwkeurigheid vast te leggen.

Een eerste stap is hierbij gezet in 1983 met het numeriek *definiëren* van de lichtsnelheid c. Voor die tijd was de meter (net als de kilogram) een fysieke maat, met identieke problemen. Door het vastleggen van de lichtsnelheid c is, gegeven de definitie van de seconde, direct de meter als eenheid van lengte vastgelegd. De tamelijk ongrijpbare problemen als reproduceerbaarheid en slijtage zijn vervangen door meer praktische problemen als het ontwikkelen van (interferometrische) meetopstellingen en het realiseren van een hoog vacuüm. Door deze stap is de nauwkeurigheid waarmee de meter is vastgelegd na de jaren '80 enkele ordes van grootte verbeterd.
Er zijn grote voordelen aan standaardeenheden boven standaardmaten. Ten eerste zorgt het feit dat iedereen overal deze meting kan uitvoeren ervoor dat primaire standaarden algemeen voorhanden zijn. Een tweede groot praktisch voordeel is dat de weg open ligt om nieuwe standaarden te ontwikkelen die beter recht doen aan de definitie van de standaardeenheid. Bij (toekomstige) nauwkeuriger meetmethoden kan de nauwkeurigheid van de eenheid worden opgevoerd. Zie voor een overzicht van de definities van basiseenheden *http://www1.bipm.org/en/si/*, de site van het BIPM. De keuzen die hierbij zijn gemaakt lijken misschien onnodig complex maar zijn mede een gevolg van de eis van continuïteit met in het verleden gebruikte definities.
Binnen enkele jaren wordt waarschijnlijk besloten ook de elementaire lading e, de Boltzmannconstante k_B, de constante van Avogadro N_A en de constante van Planck h numeriek vast te leggen. Op dat moment kunnen de mol (eenheid van hoeveelheid stof), de ampère (eenheid van elektrische stroomsterkte), de kelvin (eenheid van temperatuur) en de kilogram (eenheid van massa) in termen van meetvoorschriften en waarden van fundamentele natuurconstanten worden vastgelegd. Zie voor meer informatie *http://www1.bipm.org/en/si/*.
In het SI-systeem van eenheden staat vanwege de extreme nauwkeurigheid de definitie van de seconde centraal als eenheid van de grootheid tijd $[t]$. De eenheid seconde is zodanig gedefinieerd dat een pri-

maire standaard kan worden geconstrueerd in de vorm van de cesiumklok. Voor meer informatie zie *http://tycho.usno.navy.mil/cesium.html* of *http://www.npl.co.uk/science-technology/time-frequency/*.

Er wordt nog steeds veel werk verricht om zo goed mogelijke afspraken en primaire standaarden te ontwikkelen. Een voorbeeld is deze nóg nauwkeuriger tijdmeting: *http://www.nist.gov/pml/div688/clock-082213.cfm* en *http://www.sciencemag.org/content/341/6151/1215.full*. De beschreven atoomklok gebaseerd op ytterbium haalt een nauwkeurigheid van 1 op 10^{18} (ter vergelijking: het heelal bestaat ongeveer 4.3×10^{17} s).

Notatie Voor de *typografie* van op te geven resultaten gelden concrete afspraken:

- Als er sprake is van een eenheid moet deze worden vermeld direct na de numerieke uitkomst. Eenheden moeten rechtop (*plain*) worden weergegeven, dus bijvoorbeeld $g = 9.82$ m/s^2. In tegenstelling tot de voorkeursnotatie voor grootheden is het niet toegestaan af te wijken van standaardafkortingen voor eenheden (bedenk dus niet je eigen afkorting zoals "sec" i.p.v. de afgesproken "s" voor "seconde"). Het is toegestaan eenheden voluit weer te geven, maar eenheden ontleend aan namen moeten dan beginnen met een kleine letter. De eenheid van kracht is terecht gekoppeld aan de naam van Newton, en wordt weergegeven als "N" maar voluit als "newton".

- In combinatie met SI-eenheden worden vaak voorvoegsels (*prefixes*) gebruikt om een eenheid wat betreft grootteorde aan te passen bij de gemeten grootheid. Denk hierbij aan nanometer of nm in plaats van 0.000 000 001 m. Ook over het toegelaten gebruik van voorvoegsels zijn afspraken gemaakt. Zie hiervoor *http://physics.nist.gov/cuu/Units/prefixes.html*.

2.3.3 Meten en meetinstrumenten

Bij het vergelijken met een maat als de methode van "kwantificeren van zintuiglijke waarneming" moeten we het gebruik van hulpmiddelen betrekken. Waar de wiskunde het belangrijkste gereedschap is van een theoretisch fysicus, zijn *meetinstrumenten* de gereedschappen van de experimentator.

Meetinstrumenten maken niet alleen een (digitale) aflezing mogelijk van zintuiglijk waarneembare grootheden, maar functioneren ook als "kunstmatige zintuigen" waarmee metingen worden gedaan die buiten het bereik liggen van de menselijke zintuigen. Bij elektromagnetische straling gebruiken we bijvoorbeeld vaak (foto)detectoren, waarmee in principe het gehele elektromagnetische spectrum binnen bereik komt. Bij het gebruik van dergelijke hulpmiddelen functioneren onze zintuigen nog steeds als doorgevers van informatie aan ons zenuwstelsel, maar de kwantificering is dan vaak al instrumenteel gebeurd (denk aan de aflezing van een digitale voltmeter) en soms is informatie al verwerkt met een computerprogramma.

Directe en indirecte metingen We noemen een meting zijnde "vergelijken met een standaard" (een fysieke maat of fysiek object) een *directe* meting. Een lengtebepaling met behulp van een schuifmaat is een typisch voorbeeld van een directe meting. In heel veel gevallen is een meting gecompliceerder.

Een indirecte lengtemeting In de alledaagse praktijk is "meten" kan zelfs een lengtebepaling (waarbij een fysieke maat voorhanden is)al problemen geven. Het probleem is dan het *vergelijken* met deze maat. Hoe meet je bijvoorbeeld de golflengte van zichtbaar licht (in de orde van honderden nanometer), of de afstand van de aarde tot een ster (al gauw 10^{15} kilometer)? Een nauwkeurige schuifmaat is dan wel leuk, maar voor het bepalen van deze afstanden zullen we toch echt anders te werk moeten gaan en op een andere, *indirecte* manier de te meten grootheid moeten vastleggen. Wanneer je de afstand tot een ster wilt bepalen, kun je bijvoorbeeld gebruik maken van het feit dat de aarde een pad om de zon volgt. Hierdoor lijken sterren die niet al te ver van ons af staan ten opzichte van veel verder gelegen achtergrondsterren een cirkelvormig pad te beschrijven. Dit effect wordt wel aangeduid als de *parallax* van deze ster, en wordt gekarakteriseerd door de parallaxhoek (de verplaatsing van de ster aan de hemel uitgedrukt in de hoek). Door de beweging zoals wij die waarnemen te analyseren en te combineren met kennis over de aardbaan kunnen we de parallaxhoek relateren aan de afstand. Je wilt hier een afstand bepalen, maar feitelijk *meet je een hoek*.

Voor het bepalen van de gravitatieconstante van Newton G te bepalen is directe "*maat*" voorhanden waarmee we vergelijken. Om G te bepalen, meten we aan een fysisch proces waarbij deze constante een herkenbare rol heeft. Dit kan bijvoorbeeld de aantrekkingskracht tussen twee massa's zijn, waarbij G aanwezig is in de relatie tussen de massa's m_1 en m_2 en hun onderlinge afstand r. Als we metingen kunnen uitvoeren aan dit proces[2] en waarden voor de daarbij relevante grootheden bepalen, en we kennen ook de *relatie* tussen deze gemeten grootheden, dan kan op een *indirecte* manier de waarde van G worden bepaald.

We spreken in het geval van bovenstaande voorbeelden verband van een *indirecte* meting. Voor indirecte metingen is theorievorming *essentieel*!

Hoe meet men een temperatuur? Je zou kunnen zeggen: "ik gebruik gewoon een thermometer en lees die af". Maar "meten" met een thermometer werkt natuurlijk wel heel anders dan het meten van een lengte met een meetlat. Immers, mijn meetlat is ook echt een maat voor lengte (want mijn meetlat heeft de eigenschap lengte). Bij het meten van een lengte met een meetlat gebruik ik dus gelijkwaardige grootheden die ik onderling vergelijk (hoeveel groter, hoeveel kleiner). Bij het gebruik van een thermometer heeft mijn thermometer ook een temperatuur. Bij het meten van de temperatuur van een object wil ik echter dat mijn thermometer de temperatuur van een object *overneemt* (en die bovendien niet beïnvloedt).
Hierbij is echter geen sprake van een *directe* meting, want met welke maat zou er dan moeten worden vergeleken? Maak ik bijvoorbeeld gebruik van een ouderwetse kwikthermometer dan wordt daar *indirect* gemeten door op een slimme manier gebruik gemaakt van thermische uitzetting: de fysische eigenschap dat materialen kunnen uitzetten/krimpen bij temperatuurverandering. Met de verschillen in uitzettingseigenschappen tussen het glas en het kwik kan daarmee een temperatuurschaal worden vastgelegd. Mijn aflezing op de temperatuurschaal accepteer ik als mijn temperatuurmeting. Dat hier het nodige voorwerk gedaan moet worden, zal geen betoog behoeven. Denk er bijvoorbeeld maar eens over na hoe de afleesschaal op de thermometer moet worden vastgelegd en via welke metingen de uitzettingscoëfficiënten van het glas en het kwik zijn bepaald.

[2] Dit experiment (GRAV) kan gedaan worden in het practicum DATA-P.

2.4 Onzekerheid

De waarde van een fysische grootheid kan, anders dan een mathematische grootheid, niet exact worden vastgesteld[3]. Zowel de *maat*, waaraan de grootheid wordt afgemeten ("vergelijken"), als de *meetprocedure*, leidend tot een waardebepaling, heeft een beperkte nauwkeurigheid.
Met deze uitspraak bedoelen we dat zowel de maat die we gebruiken als de meting niet exact te reproduceren zijn. De experimentator zal er natuurlijk alles aan doen om de grootheid van interesse zo nauwkeurig mogelijk vast te stellen, maar zal ontdekken dat daarbij beperkingen van praktische aard, maar soms ook van fundamentele aard, zijn (zoals bij de kwantummechanische Heisenberg-onzekerheidsrelaties).
De beperkte nauwkeurigheid impliceert dat uit een experiment verkregen informatie moet bestaan uit zowel een numerieke waarde voor de bepaalde grootheid, als een indicatie van de numerieke *onzekerheid* in deze bepaling. De numerieke waarde is hier de "beste schatter", de waarde die naar verwachting het best overeenkomt met de "echte" waarde. De onzekerheid (ook wel *tolerantie* of *marge* genoemd) is een afschatting van de afwijking van de "beste schatter" ten opzichte van de "echte" waarde van de betreffende grootheid.

Een waarde van een grootheid zonder specificatie van onzekerheid is waardeloos!

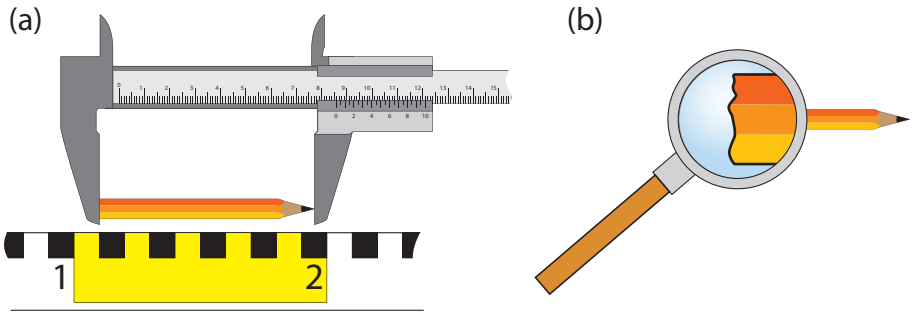
2.4.1 Oorsprong en interpretatie van onzekerheid

We interpreteren de onzekerheid als de kans om de "echte" waarde van een grootheid in een bepaalde bandbreedte rondom de gevonden meetwaarde aan te treffen. Waarom spreken we eigenlijk over een *kans* om de meting aan te treffen? Als iemand een meting uitvoert en daarbij een onzekerheid opgeeft, waarom weet die persoon dan niet 100% zeker dat de meetwaarde in dat bepaalde interval valt?
We beantwoorden deze vraag aan de hand van de bepaling van de lengte van een potlood, geïllustreerd in Figuur 2.1.

Afleesnauwkeurigheid In een experiment wordt de onzekerheid in de meetwaarde in eerste instantie bepaald door de afleesnauwkeurigheid van het meetinstrument. Stel je voor dat je een schoolbordliniaal als meetlat gebruikt, zoals in Figuur 2.1 (a). Je ziet dat de lengte van het potlood "ongeveer" 8.5 cm is. Het *kwalitatieve* begrip "ongeveer" wordt *gekwantificeerd* door middel van de onzekerheid. De opdracht is nu om een *realistische* schatting te maken van de onzekerheid.
Neem een tamelijk gemakzuchtig persoon A, een ervaren onderzoeker B en een persoon C met iets teveel zelfvertrouwen. Persoon A observeert de situatie van een afstand en rapporteert 8.5 ± 0.5 cm. Hoewel het een onderzoeker vrij staat zo'n grote onzekerheid op te geven, kun je je afvragen of dit een *realistische*

[3] Een filosofische vraag is of het een *eigenschap* van een grootheid is om een (exacte) waarde te hebben. Alhoewel dit een interessant discussiepunt is, stellen we nu dat in de empirische wetenschappen altijd getoetst wordt door middel van (het uitvoeren van) metingen. Dat meetproces gaat altijd gepaard met een eindige nauwkeurigheid. Het is voor de experimentator dus niet van essentieel belang of een grootheid nu wel/niet een exacte waarde *heeft* en of deze wel/niet *bekend* kan zijn.

In de kwantummechanica is voorts het meetproces niet onafhankelijk van het bemeten object te zien. Een kwantummechanische grootheid, zoals de spin van een elektron, kan zich namelijk bevinden in een superpositie van twee (of meerdere) meetuitkomsten, zoals "spin up" of "spin down" van het elektron. Wanneer een meting aan de elektronspin gedaan wordt, wordt wel één van deze twee waarden gevonden. De meting "beïnvloedt" hierbij het systeem; één van de mysteries van de kwantummechanica. De kwantummechanische grootheid *heeft* dus geen exacte waarde *totdat* er een meting uitgevoerd wordt. Wel kun je met veel metingen de statistiek achterhalen, dat wil zeggen de kans om één van de twee toestanden experimenteel aan te treffen. Bij het toetsen van de kwantummechanica ligt de nadruk op het experimenteel verifiëren (met een grote, maar eindige zekerheid) van deze statistiek van de elektronspin.



Figuur 2.1: Het verschil tussen (a) de afleesnauwkeurigheid van de meetmethode, zoals bij gebruik van een schoolbordliniaal of schuifmaat, en (b) de onzekerheid in (oftewel onbepaaldheid van) de gemeten grootheid zelf, geïllustreerd voor de lengtebepaling van een potlood.

schatting is. Het is tamelijk duidelijk dat een waarde van 8.1 cm, ruim binnen de onzekerheidsbandbreedte 8.0 - 9.0 cm, onwaarschijnlijk is gegeven deze observatie, omdat je immers "ziet" dat het potlood duidelijk langer is.

De ervaren onderzoeker B maakt gebruik van zogenaamde *interpolatie*. Door de achterkant van het potlood op een maatstreep te leggen, goed naar het uiteinde te kijken en een "denkbeeldige" fijnaanduiding op de liniaal te projecteren, is duidelijk dat 8.5 cm zeer waarschijnlijk is, 8.3 cm niet zo waarschijnlijk en de kans dat het potlood 8.1 cm lang is is verwaarloosbaar klein. Op basis van deze observatie rapporteert persoon B bijvoorbeeld 8.50 ± 0.15 cm.

Persoon C tenslotte gebruikt ook de techniek van interpolatie, maar projecteert een zeer nauwkeurige virtuele schaal op de liniaal. Hij komt tot 8.47 ± 0.01 cm. Hoewel dit antwoord een grote nauwkeurigheid suggereert, kun je er wel wat kanttekeningen bij plaatsen. Is het mogelijk om de achterkant van het potlood tot op 100 micrometer nauwkeurig uit te lijnen met de liniaal? En kan het menselijk oog betrouwbaar een lijn in 100 stukken delen en vervolgens afschatten in welk deel de punt van een potlood zich bevindt?

Het antwoord op de vraag waarom de kans op een "echte" waarde binnen de onzekerheidsbandbreedte geen 100% is, is dus dat in dat geval geen optimale, realistische afschatting van de waarde van de onzekerheid is gemaakt. Je zou het ook anders kunnen formuleren: wanneer de bepaling een aantal keren door verschillende personen wordt herhaald, verwacht je een groot aantal resultaten tussen de 8.35 cm en 8.65 cm (de bandbreedte door persoon B opgegeven), maar gerapporteerde waarden daarbuiten kunnen niet helemaal worden uitgesloten. Wat belangrijk is, is dat iedereen dezelfde interpretatie van het begrip onzekerheid gebruikt. In Hoofdstuk 3 introduceren we een formele definitie van onzekerheid, de *standaarddeviatie*, die op basis van onafhankelijke metingen wordt vastgesteld en zijn oorsprong heeft in de wiskundige statistiek.

Onzekerheid in de grootheid Wanneer de meting herhaald wordt, maar nu met een schuifmaat, kan er veel nauwkeuriger gemeten worden. Een realistische bepaling wordt dan bijvoorbeeld 8.580 ± 0.002 cm. Deze meting laat zien dat de rapportage van persoon C overtuigend fout is, omdat de nauwkeurige meting maar liefst elf onzekerheden (0.11 cm) ver weg ligt (zie ook de volgende sectie over discrepantie en significantie). Ook met de schuifmaat blijft een realistische bepaling van de onzekerheid essentieel.

Maar wat nu als je steeds nauwkeuriger meetapparatuur gaat gebruiken voor de lengtebepaling, bijvoorbeeld een micrometer of microscoop? Bij een voldoende grote meetnauwkeurigheid loop je tegen het feit aan dat de lengte van het potlood gemeten op verschillende plekken niet even groot is; bijvoorbeeld omdat de achterkant van het potlood niet perfect vlak is (zie Figuur 2.1 (b)). Bij elke nieuwe bepaling van de lengte wordt dan een ander resultaat gevonden, afhankelijk van waar of hoe je precies de lengtebepaling uitvoert.

De (statistische) onzekerheid in de gemeten waarde wordt nu niet meer bepaald door de meetnauwkeurigheid, maar komt direct voort uit het feit dat de gemeten grootheid *zelf* met eindige nauwkeurigheid gespe-

cificeerd is. In dit geval rapporteren we de lengte van het potlood als, bijvoorbeeld, 8.5756 ± 0.0003 cm en interpreteren we de onzekerheid als een kans om bij meting de meetwaarde in dat interval aan te treffen. Bij regelmatig herhalen van de meting van de lengte van het potlood zullen er zeker meetwaarden gevonden worden die buiten het afgebakende onzekerheidsinterval vallen.

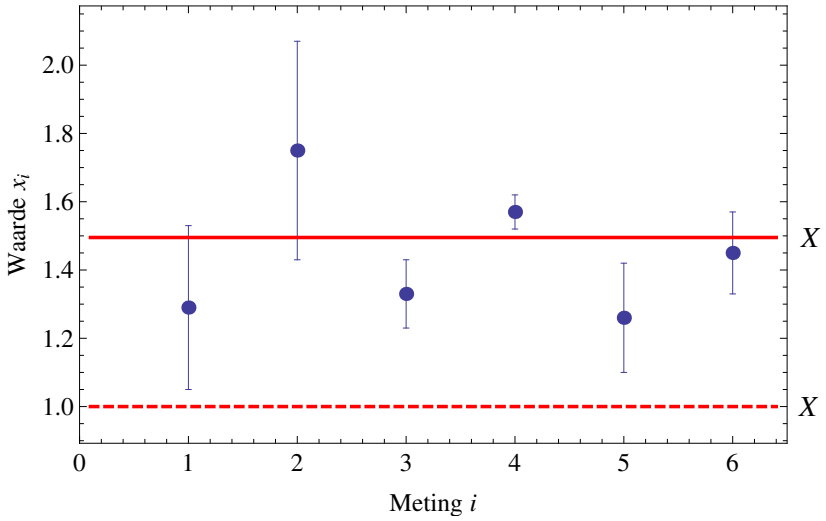
In een meting wil je de onzekerheid in de gemeten grootheid opgeven. Je hebt dan uiteraard meetapparatuur nodig waarvan de afleesnauwkeurigheid niet beperkend is voor het meetresultaat. Met andere woorden, de nauwkeurigheid moet dusdanig goed zijn dat er bij herhaalde meting een verzameling meetwaarden wordt verkregen waaruit de intrinsieke onzekerheid in de grootheid zelf afgeschat kan worden.

Bepaling van de brekingsindex Stel, je wilt de brekingsindex van glas nauwkeurig bepalen door een lichtpuls door het glas te zenden en de reistijd te bepalen. De intrinsieke onzekerheid in deze bepaling wordt veroorzaakt door dichtheidsvariaties in het glas.
Wanneer je nu een stopwatch en meetlint gaat gebruiken om de lichtsnelheid te bepalen, voel je op je klompen aan dat dit gedoemd is te mislukken. Allereerst is de methode volstrekt ongeschikt vanwege de enorm grote waarde van de lichtsnelheid en dus de reactietijd van de experimentator die veel te lang is om de reistijd van het licht te "klokken".
Ten tweede leveren de onzekerheden in de tijdsregistratie en in de lengtebepaling van het blokje een onzekerheid in de brekingsindex op die volledig gedomineerd wordt door de afleesnauwkeurigheid van de meetapparatuur. Het effect van een dichtheidsvariatie op de reistijd van de lichtpuls is vele ordes kleiner dan deze afleesnauwkeurigheid. Je zult uiteindelijk niets over de intrinsieke onzekerheid in de brekingsindex kunnen afleiden. Voor een zinvol experiment moet je nauwkeuriger meetinstrumenten of een andere meettechniek gebruiken.

2.4.2 Toevallige onzekerheid

Wanneer een meting wordt gedaan, worden door de meetmethode en/of -instrumenten, externe verstoringen en soms fundamentele fysische processen toevallige (*stochastische*) effecten in de resultaten geïntroduceerd. Bij herhaling van metingen onder ogenschijnlijk identieke omstandigheden zal blijken dat meetresultaten niet exact met elkaar overeenstemmen - mits we maar voldoende decimalen bekijken. Vanwege het onvoorspelbare element spreken we van *toevallige onzekerheid*, die voor ons van *stochastische* aard is. Figuur 2.2 toont een voorbeeld van dergelijke variatie van meetdata. In figuren wordt een onzekerheid vaak weergegeven met een staafje, dat boven (en onder) het meetpunt uitsteekt en een lengte heeft gelijk aan de numerieke waarde van de onzekerheid.

Geaccepteerde waarde is bekend Stel dat een set metingen $\{x_1, x_2, \cdots, x_n\}$ aan grootheid x wordt vastgelegd (zoals in Figuur 2.2), gebruikmakend van een bepaalde meetmethode. De stochastische effecten zijn er de oorzaak van dat er in deze metingen een afwijking ontstaat ten opzichte van een "echte" waarde X voor grootheid x. De grootte van de afwijking $x_i - X$ voor elke meting i is niet te voorspellen. Er is echter vaak wel iets te zeggen over de *kansverdeling* van x, dat wil zeggen, hoe de gemeten waarden $\{x_1, x_2, \cdots, x_n\}$ gespreid liggen rond de waarde van X. Als we de verdeling die van toepassing is (min of meer) kennen, dan kunnen we daarmee aangeven hoe groot de kans is dat onze meetwaarde x_i in een bepaald interval rond de werkelijke waarde X ligt.



Figuur 2.2: Grafische weergave van een meetserie ter bepaling van grootheid x, waarbij elk punt met bijbehorende *toevallige onzekerheid* is weergegeven. De punten vertonen een stochastische spreiding rond een "echte" waarde $X = 1.5$ van grootheid x. In de situatie dat de "echte" waarde gegeven zou worden door $X' = 1$, duidt de discrepantie tussen stochastische spreiding en de werkelijke waarde op een *systematisch effect* in de metingen.

Meten aan een bekende waarde? Je vraagt je misschien af wat voor nut het heeft om metingen te verrichten aan een grootheid waarvan al een waarde bekend is. Is dat niet een beetje verspilde moeite?
Je kunt metingen aan een bekende waarde gebruiken om te bezien of het experiment en de meetmethode resulteren in de verwachte waarde. De bekende waarde wordt als het ware gebruikt om de meetmethode te *calibreren* en systematische effecten op te sporen (zie de volgende paragraaf). Maar het zou natuurlijk ook zo kunnen zijn dat een nieuwe bepaling met een geavanceerde meetmethode een significant verschillende waarde oplevert van de verwachte waarde, en zo een reden geeft om de geaccepteerde waarde of de theorie waaruit deze is afgeleid aan te passen aan nieuwe inzichten. Overigens zou dit betekenen dat alle voorgangers iets systematisch over het hoofd gezien hebben!

Geaccepteerde waarde is onbekend In de (gebruikelijke) situatie dat er geen waarde X van grootheid x bekend is, wordt de omgekeerde weg bewandeld: een set metingen $\{x_1, x_2, \cdots, x_n\}$ kan gebruikt worden om een *waardeschatting* (voorspelling) te doen voor de "echte" waarde X. De gemiddelde waarde van de metingen is een voorbeeld van een schatting van de "echte" waarde. De (toevallige) variaties in metingen $\{x_1, x_2, \cdots, x_n\}$ hebben tot gevolg dat dit gemiddelde niet exact zal overeenstemmen met de "echte" waarde voor X. Door heel veel metingen te doen zien we echter een verdelingspatroon verschijnen. Naarmate we meer en meer metingen doen kunnen we de "echte" waarde X zo steeds beter schatten.

Het komen tot een beste schatter van een (geaccepteerde) waarde wordt meer formeel behandeld in Hoofdstukken 3 en 4. De centrale boodschap is dat bij toevallige onzekerheden het uitbreiden van de meetserie de statistiek duidelijker in beeld brengt en voor een nauwkeuriger resultaat zorgt.

Wanneer we het in het vervolg over een onzekerheid in een meetresultaat hebben, betreft dit altijd een toevallige onzekerheid (tenzij specifiek anders vermeld).

Materiaaleigenschappen van ultradunne lagen In het fabricageproces van de hedendaagse microprocessor voor computers zijn de afmetingen van de componenten (transistor, condensator, weerstand, etc.) nog maar enkele tientallen nanometer groot. Deze componenten zijn opgebouwd uit verschillende nanometer-dunne lagen van verschillende materialen (zoals geleiders, halfgeleiders en isolatoren), die tezamen de functionaliteit geven aan de component.
Het spreekt voor zich dat er bepaalde specificaties zijn voor de eigenschappen van deze materialen (zoals de waarde van de weerstand, brekingsindex, diëlektrische constante, etc.) om de component (en uiteindelijk microprocessor) te laten functioneren als gewenst. De crux is nu dat de materiaaleigenschappen voor nanometer-dunne lagen helemaal niet bekend zijn. De reden hiervoor is dat ze van heel veel factoren afhangen, zoals bijvoorbeeld intrinsiek van de laagdikte (bij nanometerdikke lagen is immers geen sprake meer van bulkeigenschappen) en extrinsiek van het gebruikte fabricageproces (bijvoorbeeld de hoeveelheid onzuiverheden in de laag).
Uit de karakterisatie van de dunne lagen worden dus de "geaccepteerde waarden" voor de materiaaleigenschappen gevonden.

2.4.3 Systematische effecten

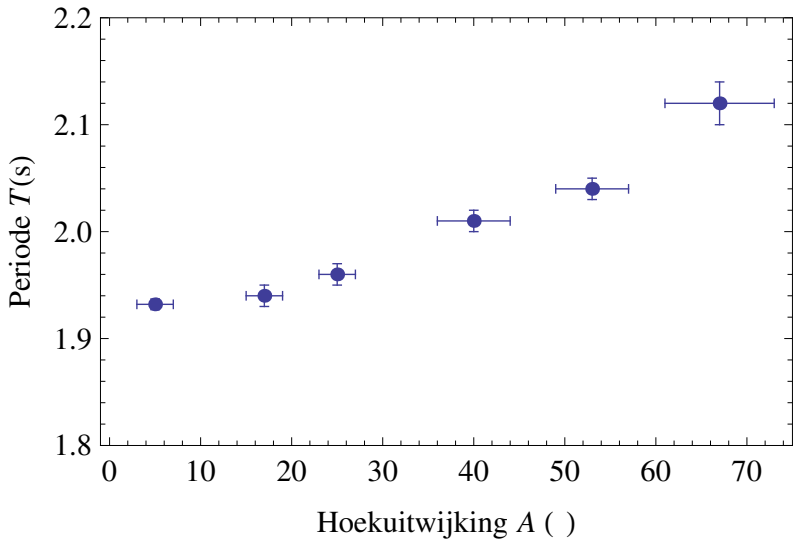
Stel dat een meting aan grootheid x een serie uitkomsten $\{x_1, x_2, \cdots, x_n\}$ oplevert. We veronderstellen dat x een "echte" waarde heeft die we aangeven met X' (analoog aan Figuur 2.2). Een *systematisch* effect zorgt ervoor dat een meetresultaat structureel afwijkt van de "echte" waarde X' van de grootheid. In Figuur 2.2 is een dergelijk systematisch effect weergegeven voor de situatie dat de werkelijke waarde gegeven wordt door $X' = 1$, terwijl de resultaten stochastisch spreiden rondom een waarde $X = 1.5$.

Systematische afwijking Een systematische *afwijking* is een niet-statistische afwijking tussen het meetresultaat en de "echte" waarde. We geven hier twee voorbeelden.

- Een systematische afwijking kan worden geïntroduceerd door een verkeerde ijking van het instrument waarmee gemeten wordt. Wanneer we een schuifmaat gebruiken die zodanig is geijkt dat 1 millimeter aflezing in werkelijkheid correspondeert met 0.99 millimeter, dan zullen we bij elke meting met deze meetlat structureel een 1% te grote lengte rapporteren, zonder daar iets van te merken.

- Een tweede voorbeeld van een systematische afwijking betreft de situatie waarbij in de verwerking van een meetserie een theoretisch model wordt gebruikt dat niet voldoende recht doet aan de realiteit (zie hiervoor ook Sectie 5.2.2). Wanneer de slingertijd T gebruikt wordt om de lokale versnelling van de zwaartekracht g te bepalen, wordt vaak gewerkt met de benadering $T = 2\pi\sqrt{l/g}$ (met l de slingerlengte) voor de periode van de mathematische slinger bij kleine uitslag. Bij een fysische slinger blijkt de periode van de slingering echter wel degelijk afhankelijk te zijn van de grootte van de uitslag van de slinger, zie Figuur 2.3. Op het moment dat de bepaling wordt uitgevoerd met een eindige hoekuitwijking wordt er een systematische *afwijking* (dus van niet-statistische aard) geïntroduceerd in de waarde van T. Bij de meting in Figuur 2.3 is *bij de gegeven onzekerheden* dit systematische effect zichtbaar voor hoeken groter dan $\approx 10^\circ$.

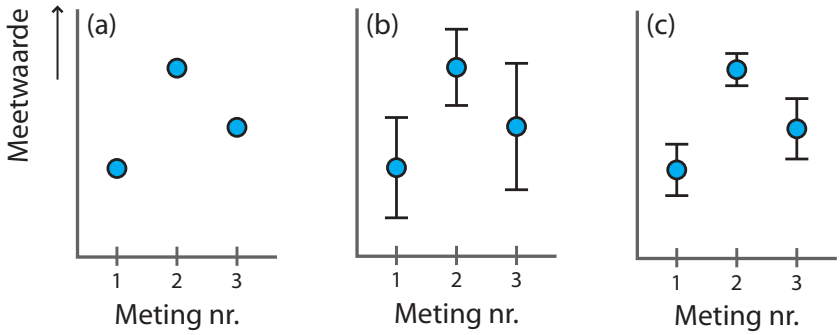
Wanneer systematische afwijkingen opgespoord worden, kan de uitkomst van een meting achteraf gecorrigeerd worden door de grootte van de systematische afwijking te kwantificeren. In het voorbeeld van de schuifmaat kan een correctiefactor voor de metingen bepaald worden na een calibratie van het instrument. Bij de slinger kan een verbeterd (nauwkeuriger) theoretisch model worden gebruikt, waarmee de meetresultaten als functie van de hoekuitwijking gecorrigeerd worden.
Het grote probleem met systematische afwijkingen is dat ze vaak lastig op te sporen zijn. Een diepgaande beheersing van het experiment is vaak noodzakelijk om systematische afwijkingen te elimineren.



Figuur 2.3: Grafische weergave van meetresultaten inclusief onzekerheden voor de periode T (in s) van een slinger als functie van de uitslaghoek A (in graden). Omdat bij grote uitwijkingen een waarde voor de periode wordt gemeten die vele malen de onzekerheid verschilt van de waarde rond 1.93 s bij kleine hoeken, concluderen wij dat T niet onafhankelijk is van A.

Sneller dan het licht? In september 2011 werd de (wetenschappelijke) wereld opgeschrikt door het bericht dat onderzoekers van CERN metingen hadden gedaan waaruit zou blijken dat neutrinos die in de LHC in Genève werden geproduceerd, naar Gran Sasso (Italië) reisden met een snelheid hoger dan de lichtsnelheid. Dit zou grote gevolgen hebben voor de fundamenten van de natuurkunde, bijvoorbeeld voor de geldigheid van de relativiteitstheorie.
De onderzoekers zijn voor de publicatie van hun meetresultaten lang op zoek geweest naar eventuele meettechnische effecten en fouten in dataverwerking die dit zouden kunnen verklaren. In de maanden na publicatie is door een grotere groep onderzoekers verder gezocht. In februari 2012 bleek dat een niet goed aangesloten optische fiber zorgde voor 60 nanoseconde systematische afwijking. Het experiment is inmiddels met verschillende detectoren en in verschillende onderzoekscentra herhaald, en het blijkt dat neutrinos zich keurig aan de "maximumsnelheid" houden. Een mooi voorbeeld van hoe "meten" werkt.
Zie: *http://news.sciencemag.org/scienceinsider/2012/06/once-again-physicists-debunk.html*

Systematische onzekerheid Anders dan bij de systematische *afwijking* speelt bij systematische *onzekerheid* statistiek een rol. Stel dat we een slingerexperiment uitvoeren om de zwaartekrachtsconstante g te bepalen. We hebben hierbij een slingerlengte l bepaald met een relatieve onzekerheid van 1% vanwege een beperkte meetnauwkeurigheid. De gerapporteerde waarde voor l kan dus in de orde van 1% afwijken van de "echte" waarde, maar heeft voor elke slingering wel een constante afwijking ten opzichte van de "echte" waarde. Wanneer een serie metingen aan de slingertijd $\{T_1, T_2, \cdots, T_n\}$ wordt uitgevoerd voor een nauwkeurige bepaling voor g, dan werkt het systematische effect (ten gevolge van de bepaling van l) op identieke wijze door in de berekeningen van elke $g_i = 4\pi^2 l/T_i^2$.
Door het verhogen van het aantal onafhankelijke bepalingen van l met elk een toevallige variatie kan de systematische onzekerheid wél verkleind worden (zie hiervoor Hoofdstuk 3). Stel dat hierdoor de relatieve



Figuur 2.4: De relevantie van de onzekerheid in de datapunten geïllustreerd voor metingen aan dezelfde fysische grootheid, maar gerapporteerd (a) zonder onzekerheid, met relatief (b) grote en (c) kleine onzekerheid (bijvoorbeeld uitgevoerd door verschillende onderzoekers of met verschillende meetinstrumenten). De discussie of het verschil (= de discrepantie) tussen de datapunten *significant* is, wordt in dit geval strikt bepaald door de grootte van de onzekerheid in de datapunten in verhouding tot de spreiding.

onzekerheid in l daalt naar 0.1%, dan daalt dus ook de systematische onzekerheidsbijdrage in g ten gevolge van de onzekerheid in l.

2.4.4 Discrepantie en significantie

Doordat elke meting van een grootheid een eindige nauwkeurigheid heeft, worden in opvolgende metingen nooit identieke meetresultaten verkregen. We definiëren de *discrepantie* als het verschil in twee gemeten waarden. Veel aandacht gaat vervolgens uit naar de vraag of meetresultaten strijdig zijn met elkaar of juist niet. In het eerste geval zeggen we dan dat de resultaten *significant* verschillen[4]. Voor het vaststellen van significante verschillen is niet de absolute discrepantie maatgevend, maar de discrepantie in relatie tot de onzekerheden in de meetresultaten. Figuur 2.4 illustreert dit. In elk van de drie deelfiguren zie je een dataset bestaande uit drie meetwaarden, bijvoorbeeld verkregen door verschillende onderzoekers die elk een eigen meetmethode gehanteerd hebben. De meetwaarden zijn gerapporteerd met (a) geen, (b) relatief kleine en (c) relatief grote onzekerheid (aangegeven met een verticaal balkje).[5] De vraag is nu of we aan de hand van elk van de figuren kunnen stellen dat de metingen *significant* (niet aan toeval toe te schrijven) verschillend zijn of niet? Dit bespreken we kwalitatief aan de hand van de weergaven in Figuur 2.4:

- **Geen onzekerheid gerapporteerd**, zie Figuur 2.4 (a). Alhoewel er verschillende waarden zijn gevonden, kan er eigenlijk geen zinvol antwoord op de vraag geformuleerd worden. Immers om te kunnen stellen of de datapunten *verschillend* zijn hebben we ook informatie nodig over de nauwkeurigheid waarmee de waarden bekend zijn.

- **Relatief grote onzekerheid**, zie Figuur 2.4 (b). Hier is de onzekerheid in de punten gespecificeerd met onzekerheidsbalkjes, die de kans aangeven binnen welke marge de echte waarde zal liggen ten opzichte van de meetwaarde. We kunnen als een strikt provisorisch criterium voor significant verschil aanhouden dat de onzekerheidsbalkjes voor verschillende metingen niet mogen overlappen. Aangezien het verschil tussen meting 2 en 3 door de onzekerheidsbalkjes wordt overbrugd, geldt dat deze

[4] "Significant hetzelfde" is taalkundig niet mogelijk. Aangezien vanwege eindige nauwkeurigheid nooit gesteld kan worden dat twee waarden *hetzelfde zijn*, moet strikt genomen altijd de formulering zijn dat resultaten niet significant verschillen.

[5] In een experiment is de onzekerheid uiteraard niet vrij kiesbaar, maar volgt deze uit de specificaties van de meetapparatuur en de statistische spreiding in de meetwaarde. Voor ons argument zijn ook de gevonden meetwaarden in Fig. 2.4 (a) - (c) gelijk, alhoewel het (zeer) onwaarschijnlijk is om "dezelfde" waarden te vinden bij verschillende meetnauwkeurigheden.

twee metingen niet (significant) verschillend zijn. Meting 1 en 2 zijn dan wel verschillend omdat de onzekerheden niet overlappen.

- **Relatief kleine onzekerheid**, zie Figuur 2.4 (c). Naarmate de opgegeven waarden nauwkeuriger bepaald zijn en de onzekerheden dus kleiner, zijn ook meting 2 en 3 (significant) verschillend. Bij het uitvoeren van nog nauwkeuriger bepalingen (aannemende dat de meetwaarden gelijk blijven bij een nieuwe nauwkeuriger meting) zullen zelfs meting 1 en 3 (significant) gaan verschillen.

In eerste instantie hanteren we een provisorische, strikte eis dat er sprake is van een significant verschil als de onzekerheden niet overlappen. Maar, zoals we eerder hebben gezien, correspondeert het onzekerheidsinterval niet met een exacte bandbreedte. Deze strikte eis is dus eigenlijk niet correct. In hoofdstukken 2 en 4 introduceren dan ook een *statistische aanpak* om een *kwantitatieve uitspraak* te kunnen doen of twee punten al dan niet als verschillend moeten worden beschouwd.

2.4.5 Weergave van onzekerheid

We zullen nu kort ingaan op de weergave van een resultaat met (toevallige) onzekerheid.

Notatie

1. Kwantitatieve resultaten moeten altijd worden gerapporteerd samen met *bijbehorende onzekerheid*. Voor de onzekerheid in de uitkomst van een fysische grootheid x gebruiken we standaard de symbolische notatie δx of σ_x. De complete numerieke informatie wordt dan gegeven als $x \pm \delta x$ of $x \pm \sigma_x$.

2. In de numerieke weergave corresponderen de minst significante cijfers van "beste schatter" en onzekerheid. Voor het eerder gebruikte voorbeeld van de zwaartekrachtsconstante is een correcte weergave: $g = (9.82 \pm 0.03)$ m/s^2. Niet acceptabel zijn bijvoorbeeld $g = (9.82 \pm 0.2739)$ m/s^2 of $g = (9.81846 \pm 0.03)$ m/s^2.

3. Men ziet onzekerheid soms ook wel genoteerd op de volgende manier: $g = 9.818(27)$ m/s^2. De informatie tussen haakjes slaat dan op de onzekerheid in de laatste decimalen van het getal ervóór (dus $g = 9.818 \pm 0.027$ m/s^2). Deze notatie is vooral efficiënt bij zeer precieze gegevens, zoals de lading van een elektron: $q = (-1.602176565 \pm 0.000000035) \times 10^{-19}$ C versus $-1.602176565(35) \times 10^{-19}$ C.

4. In het voorgaande is de onzekerheid gegeven in absolute vorm δx, wat de meest voorkomende vorm van rapportage is. Als alternatief kan ook de *relatieve onzekerheid* worden opgegeven, gedefinieerd als $\delta x/x$. De relatieve onzekerheid wordt opgegeven als $1 \pm \delta x/x$. In het voorbeeld hierboven dus: $g = 9.82(1 \pm 0.003)$ m/s^2.

5. Sterk gelijkend op de relatieve onzekerheid is de *procentuele onzekerheid*, waarbij de onzekerheid wordt genoteerd als *percentage* van de bepaalde waarde. In het voorbeeld hierboven dus: $g = 9.82$ m/s^2 $\pm$ 0.3% of $g = 9.82(1 \pm 0.3\%)$ m/s^2.

Voor nadere informatie over deze afspraken zie *http://physics.nist.gov/Pubs/SP811/contents.html*.

Afrondingsregels Bij verwerking van digitaal gemeten datasets worden in het algemeen resultaten met zeer veel cijfers achter de komma berekend, bijvoorbeeld $g = 9.81846 \pm 0.02739$ m/s^2. Het blijkt echter dat de waarde van de onzekerheid z'elf beperkt nauwkeurig is: de relatieve onzekerheid in de onzekerheid zelf is $\frac{\sigma_{\sigma_x}}{\sigma_x}$ is typisch in de orde van 10% (zie Sectie 3.2 voor een onderbouwing van dit getal).

Tabel 2.2: De 10%-regel en de DATA-vuistregel als afrondingsregels voor de onzekerheid σ in het gebied $[0.095 < \sigma < 0.9499\cdots]$.

Onzekerheid σ	Afronding 10%-regel	Afronding DATA-vuistregel	Onzekerheid σ	Afronding 10%-regel	Afronding DATA-vuistregel
0.095	0.1	0.10	0.255	0.26	0.3
0.10	0.1	0.10	0.26	0.26	0.3
0.11	0.1	0.11	0.28	0.3	0.3
0.12	0.12	0.12	0.30	0.3	0.3
0.13	0.13	0.13	0.32	0.3	0.3
0.14	0.14	0.14	0.34	0.34	0.3
0.15	0.15	0.15	0.36	0.36	0.4
0.16	0.16	0.16	0.38	0.4	0.4
0.17	0.17	0.17	0.40	0.4	0.4
0.18	0.18	0.18	0.44	0.4	0.4
0.19	0.2	0.19	0.48	0.5	0.5
0.20	0.2	0.20	0.52	0.5	0.5
0.21	0.2	0.21	0.56	0.6	0.6
0.22	0.2	0.22	0.60	0.6	0.6
0.23	0.23	0.23	0.70	0.7	0.7
0.24	0.24	0.24	0.80	0.8	0.8
0.25	0.25	0.25	0.90	0.9	0.9
$0.25499\cdots$	0.25	0.25	$0.9499\cdots$	0.9	0.9

Als de onzekerheid relatief maar tot op 10% nauwkeurig bekend is, suggereert het rapporteren van zoveel cijfers achter de komma als in het ruwe resultaat dus een nauwkeurigheid die in werkelijkheid niet bestaat. Er dient dus te worden *afgerond*.
Alleen decimalen die kunnen worden verantwoord, gegeven de onzekerheid, mogen worden opgegeven. Een onzekerheid moet, als we er van uitgaan dat een onzekerheid zelf niet nauwkeuriger bekend is dan 10%, op één tot twee significante cijfers worden afgerond. Het aantal significante cijfers hangt dan af van de afwijking die geïntroduceerd wordt door afronding. We geven hier twee afrondingsregels, die we zullen aanduiden als de *10%-regel* en de *DATA-vuistregel*, en die beiden gebruikt mogen worden.

- **De *10%-regel*** kijkt naar de afrondingsfout die wordt gemaakt bij het afronden van de onzekerheid op één decimaal. Als deze fout groter is dan 10%, dient te worden afgerond op twee significante cijfers. Wanneer bijvoorbeeld een tijdsmeting is gedaan met resultaat $t = 0.96 \pm 0.14$ s, dan verwachten wij dat de onzekerheid 0.14 s niet beter dan 0.014 s (10%) nauwkeurig bekend is. Afronden op één decimaal tot 0.1 s geeft dan een afrondingsfout van 0.04 s in de onzekerheid; veel groter dan 0.014 s. Er wordt dan significante informatie weggegooid. Beter is dus om hier de afronding op twee significante cijfers te houden: de onzekerheid blijft dan 0.14 s en het gerapporteerde resultaat blijft dus $t = 0.96 \pm 0.14$ s.
  Wanneer was gemeten 0.96 ± 0.48 s, dan is de onzekerheid zelf tot op 0.048 s bekend. Afronden op één significant cijfer tot 0.5 s zorgt dan voor een afrondingsfout van 0.02 s, kleiner dan 0.048 s. Er is dan geen (significant) verlies aan informatie. De positie van het laatste significante cijfer van de waarde 0.96 s moet natuurlijk wel sporen met het laatste significante cijfer van de op te geven onzekerheid. Het gerapporteerde resultaat wordt dan $t = 1.0 \pm 0.5$ s.

- **De *DATA-vuistregel*** is afgeleid van de 10%-regel. De 10%-regel leidt soms tot het vreemde resultaat dat een onzekerheid één significant cijfer krijgt (bijvoorbeeld 0.22 wordt 0.2), terwijl een vlakbij liggende waarde op twee wordt afgerond (0.225 wordt 0.23). De DATA-vuistregel legt een grens

bij $0.25499\cdots$: alles eronder tot 0.095 wordt op twee significante cijfers afgerond, alles erboven tot $0.9499\cdots$ op één significant cijfer. Op deze manier wordt de 10%-regel op de meeste intervallen gevolgd.

In Tabel 2.2 is weergegeven hoe de beide afrondingsregels voor de onzekerheid uitpakken in het bereik $[0.095 < \sigma < 0.9499\cdots]$ (een decade). Voor resultaten in andere decaden kunnen de getallen in deze tabel met de vereiste macht van tien worden vermenigvuldigd.

2.5 Grafische weergave

Tabellen en grafieken vormen een belangrijk hulpmiddel in de verwerking van numerieke informatie. Beide hebben tot doel het *effectief* overbrengen van informatie. Afhankelijk van het publiek waarvoor de informatie bedoeld is (jezelf, medestudenten, een breed publiek) en de fase van het onderzoek (meting, verwerking, rapportage) moeten vorm en inhoud van de grafiek worden bepaald. Er zijn echter een aantal algemene eisen te formuleren waaraan altijd moet worden voldaan.

2.5.1 Algemene eisen aan tabellen

Een tabel is een opsomming van meetgegevens. Tabellen zijn vooral nuttig om een helder overzicht te geven van resultaten die niet (goed) in een figuur te vangen zijn, zoals gemeten resultaten voor verschillende grootheden, of wanneer de numerieke waarden van de grootheden expliciet vergeleken moeten worden. Wanneer de resultaten erg gelijkvormig zijn of functionele verbanden vertonen is een figuur meer aansprekend en verhelderend.
In Figuur 2.5 staat een voorbeeld van identieke data in een matig opgemaakte en een overzichtelijke tabel. Het moge duidelijk zijn dat goed nagedacht moet worden over de vormgeving van de tabel om tot een heldere weergave te komen. Een tabel in een tekst voldoet (minimaal) aan de volgende voorwaarden:

- Probeer door een heldere keuze voor tabelinrichting de data op een overzichtelijke wijze weer te geven, zodat de lezer in één oogopslag de strekking van de tabel vat. Plaats in de tabel alleen getallen.
- Geef voldoende informatie over de getallen in de tabel. Dit kan door de kolommen en (mogelijk) de rijen koppen te geven die beschrijven wat er in die kolom of rij te vinden is en in welke eenheid de data gespecificeerd is.
- Geef ook de onzekerheid in de gevonden data en denk hierbij aan het correcte aantal significante cijfers.
- Geef de tabel een bovenschrift (of *caption*) met een korte, maar volledige beschrijving van de inhoud, en eventueel wat toelichting op de relevantie of de herkomst van de gegevens.

2.5.2 Algemene eisen aan grafieken

In Figuur 2.6 hebben we een uitermate slecht en een goed opgemaakt figuur naast elkaar gezet. Je begrijpt nu waarom de opmaak van een figuur zo belangrijk is voor de communicatie van je resultaten. Een korte opsomming van de zaken die een goed opgemaakt figuur kenmerken:

- De asbijschriften (aanduiding van grootheden en eenheden op de assen). Let op de conventies die in Secties 2.2 en 2.3 zijn besproken!
- De ordegroottes zijn zoveel mogelijk verwerkt in de eenheidsaanduiding; dit is beter dan de ordes van grootte in de numerieke waarden weer te geven.

(a)

Tabel 2.3 Materiaaleigenschappen van verschillende materialen.

Fabricageproces		Materiaaleigenschappen	
Materiaal	Dikte	Brekingsindex	Band gap
Al_2O_3	10.34 ± 0.4	1.67 ± 0.03	-
Al_2O_3	109.03 ± 0.15	1.63 ± 0.02	-
HfO_2	11.8 ± 0.5	2.00 ± 0.04	5.77 ± 0.09
HfO_2	60.3 ± 0.3	1.96 ± 0.02	5.72 ± 0.04
Er_2O_3	7.1 ± 0.4	1.78 ± 0.05	4.8 ± 0.3
Ta_3N_5	48.51 ± 0.20	2.68 ± 0.02	2.45 ± 0.02

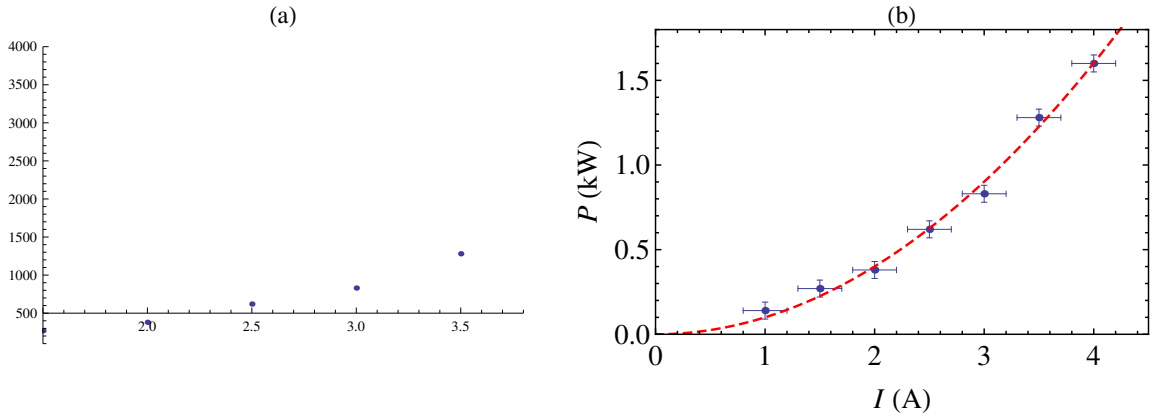
(b)

Tabel 2.3 Optische eigenschappen van Al_2O_3, HfO_2, Er_2O_3 en Ta_3N_5 lagen van verschillende dikten. De brekingsindex bepaald bij een energie van 1.96 eV en de optische band gap zijn verkregen uit de metingen met spectroscopische ellipsometrie.

Fabricageproces		**Materiaaleigenschappen**	
Materiaal	Dikte (nm)	Brekingsindex	Band gap (eV)
Al_2O_3	10.34 ± 0.4	1.67 ± 0.03	- [a]
	109.03 ± 0.15	1.63 ± 0.02	- [a]
HfO_2	11.8 ± 0.5	2.00 ± 0.04	5.77 ± 0.09
	60.3 ± 0.3	1.96 ± 0.02	5.72 ± 0.04
Er_2O_3	7.1 ± 0.4	1.78 ± 0.05	4.8 ± 0.3
Ta_3N_5	48.51 ± 0.20	2.68 ± 0.02	2.45 ± 0.02

[a] buiten het energiebereik van de gebruikte ellipsometer.

Figuur 2.5: Belang van het duidelijk structureren van de informatie in een tabel geïllustreerd voor (a) een matige opmaak en (b) een overzichtelijke opmaak. Het bovenschrift als korte, maar volledige beschrijving van de inhoud van de tabel is een essentieel onderdeel.



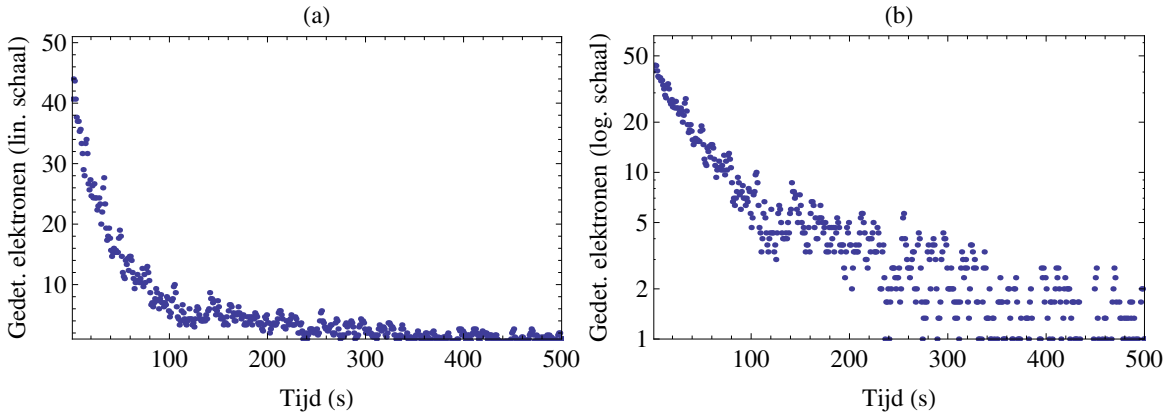
Figuur 2.6: Noodzaak van belang van opmaken van een figuur geïllustreerd voor een (a) uitermate slecht en (b) goed opgemaakt figuur. De aandachtspunten voor het goed opmaken van een figuur staan beschreven in de tekst.

- De asbijschriften en de puntgrootte zijn voldoende groot gekozen. Verder is het belangrijk dat de figuur (als deze wordt opgenomen in een tekst of presentatie) voldoende resolutie heeft.
- Het weergavebereik (*plot range*) is zodanig gekozen dat alle punten goed in beeld komen.
- In de grafiek zijn onzekerheden in de meetwaarden weergegeven. Zo wordt snel duidelijk hoe nauwkeurig de bepaalde resultaten zijn.
- In de grafiek zijn waar mogelijk resultaten van een (theoretisch) model opgenomen als lijn. Op deze manier krijgt de figuur voor de lezer/kijker grote meerwaarde. Er wordt direct duidelijk of het verwachte verband over het meetbereik wordt gereproduceerd en hoe betrouwbaar het behaalde resultaat (op het oog) is.
- In wetenschappelijke literatuur is het gebruikelijk om een rand of *frame* om de figuur te trekken (in *Mathematica* gebeurt dit niet standaard).
- Voeg in een presentatie een duiding van de gegevens in de figuur toe d.m.v. titel van *slide* of onderschrift, en in een rapport of artikel een onderschrift (*caption*). Zo wordt voor de lezer snel het kader van de resultaten geschetst.

In de *Mathematica*-werkboeken wordt uitgebreid besproken hoe de standaardopmaak kan worden aangepast om een figuur te maken dat aan deze voorwaarden voldoet. In het vervolg van deze sectie geven we wat voorbeelden van de effectieve inzet van figuren.

2.5.3 Snelle analyse van meetresultaten

Een grafische weergave van gemeten data in de *meetfase* van een experiment is van groot belang. Dit geldt in het bijzonder als er metingen worden uitgevoerd als functie van een zekere grootheid. In Figuur 2.3 als voorbeeld een meting van de periode T van een slinger als functie van de hoekuitslag A.
Meestal wordt de *onafhankelijke* variabele/grootheid (de grootheid die in het experiment gevarieerd wordt, in Figuur 2.3 de hoekuitslag A) langs de x-as uitgezet en de *afhankelijke* variabele (de grootheid die gemeten wordt, hier de periode T) langs de y-as. Betrouwbaarheidsintervallen oftewel *error bars* kunnen worden toegevoegd voor een nog duidelijker weergave van de meetresultaten. Uitgangspunt is steeds de grafiek zo in te richten dat zo effectief mogelijk informatie wordt overgedragen. Uit deze meting en bij deze onzekerheden kunnen we nu afschatten vanaf welke hoekuitslag er sprake is van een *significant* verschil van de periode gemeten bij “kleine” uitslag.



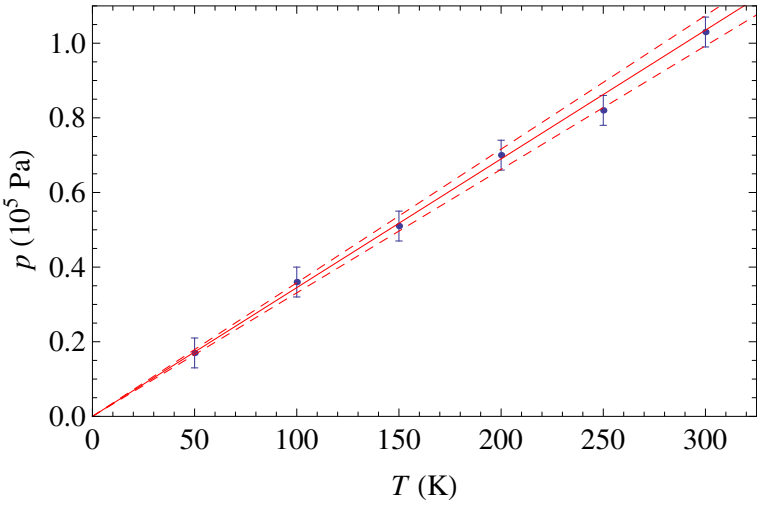
Figuur 2.7: Het belang van een optimale weergave in de verwerking van resultaten bij het vaststellen van kwalitatieve verbanden, geïllustreerd voor het gebruik van (a) lineaire en (b) logaritmische schaalverdeling op de verticale as. In een experiment is het β-verval van geactiveerd zilver weergegeven als functie van de tijd. Uit weergave (a) lijkt het aantal gedetecteerde elektronen exponentieel te vervallen met vervaltijd ≈ 40 s. De logaritmische schaalverdeling in weergave (b) maakt een hoger detailniveau zichtbaar, en laat een tweede vervalsproces zien met karakteristieke vervaltijd ≈ 150 s.

2.5.4 Vaststellen van kwalitatieve verbanden

Bij de *analyse* van een meting heeft grafische weergave vaak een rol bij het vaststellen van (kwalitatieve) verbanden. Dit is in het bijzonder het geval als de grafiek zo kan worden ingericht (eventueel na omrekenen van de gemeten grootheden) dat een *lineair* verband tussen de uitgezette waarden zichtbaar wordt. Het oog onderkent namelijk uitstekend lineaire verbanden. Voor maximaal profijt van de gevoeligheid van het oog worden de schalen zo gekozen dat de in de grafiek geplaatste punten ongeveer op een lijn liggen die een hoek van zo'n 45° met de beide assen maakt. Kleine afwijkingen zijn dan het best te herkennen. Het loont dus de moeite om te bedenken of de meetresultaten in de grafische representatie als een "rechte-lijn-relatie" te brengen zijn.
Als voorbeeld nemen we een experiment waarbij geëmitteerde elektronen (β-verval) afkomstig van geactiveerde zilverkernen worden gedetecteerd met een Geiger-Müllerbuis. Het aantal getelde elektronen als functie van de tijd is weergegeven in Figuur 2.7 (a). Het verval lijkt op het eerste gezicht exponentieel als functie van de tijd. Door nu de intensiteit logaritmisch uit te zetten tegen de tijd als in Figuur 2.7 (b) wordt een rechte-lijn-relatie zichtbaar in de grafiek voor tijden < 100 s. De helling van deze grafiek geeft nauwkeuriger informatie over de betreffende levensduur. Daarnaast is nu te zien dat dit verval feitelijk *dubbel-exponentieel* van aard is (twee vervalsprocessen met verschillende levensduren). De oorzaak ligt in de aanwezigheid van twee zilverisotopen, die vervallen met verschillende vervalsconstanten.

2.5.5 Kwantificeren van verbanden: grafische aanpassing

Wanneer een verband lineair is of (zoals in Figuur 2.7) lineair gemaakt kan worden, is het mogelijk om een "grafische aanpassing" te doen, dat wil zeggen een bepaling van de helling op basis van grafische informatie. Hiermee wordt een geconstateerd kwalitatief verband gekwantificeerd.
In Figuur 2.8 geven we ter illustratie de meting van de gasdruk p van een gas bij constant volume als functie van de absolute temperatuur T. In de analyse gaan wij uit van de ideale gaswet voor constant volume: $p = \alpha T$. Volgens de ideale gaswet geldt dan $\alpha = nk_B$, met n de dichtheid van het gas in deeltjes/m^3 en k_B de zogenaamde constante van Boltzmann ($\simeq 1.38 \times 10^{-23}$ J/K). Een "grafische aanpassing" kan nu als volgt worden uitgevoerd (deze aanpak wordt in de volgende hoofdstukken geformaliseerd):



Figuur 2.8: Gasdruk p tegen (absolute) temperatuur T. De lijnen vormen een "grafische" aanpassing aan de data, uitgaand van het verband $p = \alpha T$. De middelste getrokken lijn is de aanpassing met de "meest waarschijnlijke waarde" $\hat{\alpha}$. De bovenste en onderste gestippelde lijnen zijn grafische schattingen waarmee de onzekerheid in $\hat{\alpha}$ bepaald is. Een extra criterium in dit voorbeeld is dat alle lijnen door de oorsprong moeten gaan.

- Aan de centrale lijn $p = \hat{\alpha}T$ met $\hat{\alpha}$ de beste schatter van α wordt de eis gesteld dat aan beide kanten ongeveer evenveel meetwaarden liggen, waarbij geprobeerd wordt om de totale afstand tot de meetpunten zo klein mogelijk te houden.
- Om de onzekerheid σ_α in de helling te bepalen zoeken we twee lijnen (symmetrisch om de centrale lijn) waarvoor aan de éne kant ongeveer $5\times$ zoveel meetwaarden liggen dan aan de andere zijde. (1σ-criterium waarvoor de tweezijdige overschrijdingskans ongeveer 1/3 is, de "68%-regel", zie Hoofdstuk 4). Hiermee bepalen we de breedte van de verdeling van de punten rondom de gemiddelde lijn.
- Voor het bepalen van de onzekerheid $\sigma_{\hat{\alpha}}$ in de beste schatter $\hat{\alpha}$ kunnen we dit getal nog delen door de wortel van het aantal punten $\sqrt{n}$.

Uiteraard lopen alle lijnen voor dit geval door de oorsprong, maar data mogen voor de grafische aanpassing ook een asafsnede vertonen. De docenten van dit vak hebben de grafische aanpassing handmatig uitgevoerd voor de data in Figuur 2.8. Op basis hiervan komen zij tot $\alpha = (3.45 \pm 0.06) \times 10^2$ J/K m^3. Een analytische uitwerking op basis van de theorie in Hoofdstuk 6 levert op $\alpha = (3.41 \pm 0.04) \times 10^2$ J/K m^3 (met inachtneming van de eisen aan afronding).
Grafisch aanpassen heeft zijn beperkingen. Ten eerste is het goed kwantificeren van onzekerheden vaak moeilijk. Ten tweede zijn niet alle functionele verbanden (eenvoudig) te lineariseren. Ten derde is het moeilijk om in rekening te brengen dat meetpunten ongelijke onzekerheid hebben. Het formuleren van een grondiger aanpak is dus noodzakelijk. Deze aanpak wordt nader behandeld in Hoofdstuk 6.

2.5.6 Rapportage

In de *rapportagefase* ligt de nadruk op het effectief overbrengen van informatie aan de lezer, zodanig dat deze jouw werk begrijpt en op waarde kan schatten. Dit betekent dat bij grafische weergave een *selectie* van meetresultaten wordt gebruikt om de conclusies te onderbouwen. Het kan verleidelijk zijn alles wat je aan waarnemingen hebt te presenteren, maar je kan dan door de overmaat van detail gemakkelijk je doel voorbij schieten. Figuur 2.8 is een goed voorbeeld van een figuur waarin informatie gedoseerd wordt gebracht en waarin overbrengen van het doel van het experiment (het kwantificeren van het verband tussen druk en temperatuur) door de getoonde gegevens effectief wordt ondersteund.

Hoofdstuk 3

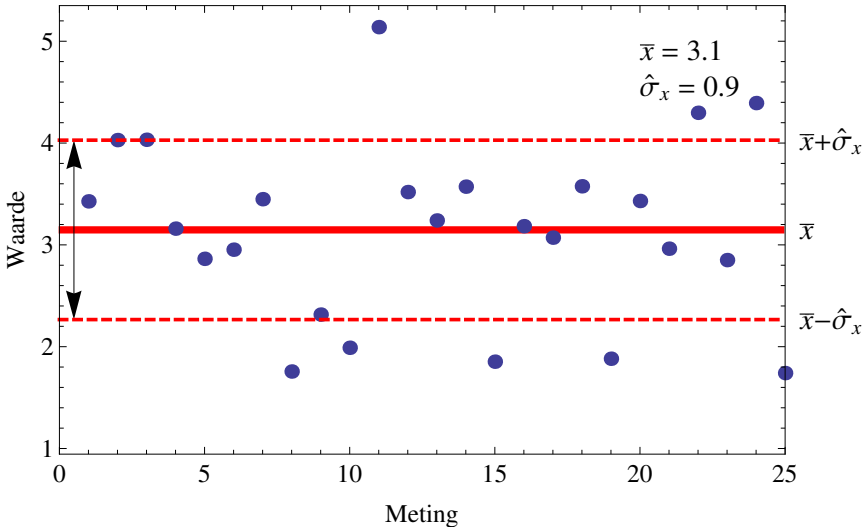
Gemiddelde en spreiding van een gemeten dataset

3.1 Inleiding

In het voorgaande hoofdstuk hebben we gezien hoe enkelvoudige, kwantitatieve meetresultaten moeten worden gerapporteerd in termen van grootheden, eenheden en onzekerheid. In de fysica worden metingen aan dezelfde fysische grootheid onder identieke omstandigheden vaak een aantal malen herhaald.
Bij herhaalde meting zullen (toevallige) variaties in de gemeten waarden optreden. Dit kan zijn vanwege een eindige nauwkeurigheid van het meetinstrument, fluctuaties (ruis) of fundamentele onzekerheden. Een set metingen aan een grootheid zal daarom een spreiding kennen rondom een *"meest waarschijnlijke waarde"* of *"beste schatter"* voor die grootheid. Vaak wordt deze "beste schatter" bepaald door een middeling van de meetresultaten in de set. De spreiding rondom deze "beste schatter" kan als maat dienen voor de nauwkeurigheid van de meting.
Je kunt intuïtief aanvoelen dat naarmate het aantal metingen groter wordt, de nauwkeurigheid waarmee deze "beste schatter" voor de grootheid bekend is, verbetert. We laten hier systematische effecten buiten beschouwing. Immers voor systematische effecten voegen statistische operaties, zoals middeling, niets toe aan de nauwkeurigheid van de bepaling.
In dit hoofdstuk gaan we in op het op een juiste manier bepalen van de meest waarschijnlijke (gemiddelde) waarde oftewel beste schatter voor een gemeten grootheid, uitgaand van een set van meetwaarden. De *standaarddeviatie* wordt bepaald door de spreiding van de waarden rondom de beste schatter. We kijken daarnaast naar de *standaarddeviatie van het gemiddelde*, de onzekerheid in de beste schatter. Hierbij bekijken we uitsluitend *ongewogen* metingen (wanneer onzekerheden in de meetresultaten niet bekend zijn). Wanneer (eventueel variërende) onzekerheden bekend zijn bij de meetwaarden is sprake van *gewogen* middeling. Dit geval komt uitgebreid aan de orde in Hoofdstuk 6.

3.2 Gemiddelde en standaarddeviatie

Aan de fysische grootheid x met "echte" waarde X is een serie van n metingen verricht, met uitkomsten $\{x_1, x_2, \cdots, x_n\}$. Een voorbeeld is gegeven in Figuur 3.1 voor $n = 25$. Doorgaans zal de dataset $\{x_1, x_2, \cdots, x_n\}$ *toevallige spreiding* vertonen (als eerder gezegd laten we systematische effecten buiten beschouwing). We gaan er nu van uit dat deze metingen onderling *statistisch onafhankelijk* zijn, dat wil zeggen dat het resultaat x_i voor meting i $(i = 1, \ldots, n)$ niet afhangt van de overige gemeten x-waarden. Onder die voorwaarde kan een statistisch beste bepaling ("beste schatter") voor X worden verkregen door het nemen



Figuur 3.1: Gemiddelde $\overline{x}$ en standaarddeviatie $\hat{\sigma}_x$ van een set van $n = 25$ metingen aan grootheid x (punten) grafisch weergegeven.

van het *gemiddelde* (*mean*) $\overline{x}$ van deze meetresultaten:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.1}$$

Met de notatie $\overline{x}$ wordt tot uitdrukking gebracht dat er een *middelingsoperatie* is uitgevoerd. Het is niet te verwachten dat het gemiddelde $\overline{x}$ van deze dataset exact zal overeenstemmen met de "echte" waarde X. Wat we wél verwachten is dat in de meeste gevallen het gemiddelde $\overline{x}$ dichter bij X zal liggen dan een willekeurig gekozen meetwaarde x_i uit de set (dat is precies de reden waarom we resultaten middelen!).
De gemeten waarden $\{x_1, x_2, \cdots, x_n\}$ liggen op een bepaalde (toevallige) manier verdeeld rondom het gemiddelde $\overline{x}$. Als maat voor deze spreiding voeren we in het begrip *standaarddeviatie* (*standard deviation*). Een "beste schatting" $\hat{\sigma}_x$ voor de *standaarddeviatie* (SD) van de meetserie $\{x_1, x_2, \cdots, x_n\}$ wordt gegeven door

$$\hat{\sigma}_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2}. \tag{3.2}$$

De notatie $\hat{\sigma}_x$ met het "dakje" $\hat{}$ drukt uit dat de statistisch "beste" schatting voor de standaarddeviatie is verkregen op basis van een dataset $\{x_1, x_2, \cdots, x_n\}$. In definitie Vgl. (3.2) is de som van de *kwadratische* verschillen tussen gemeten punten en het gemiddelde genomen en geen som $\sum(x_i - \overline{x})$ van *lineaire* verschillen, omdat in dit laatste geval negatieve en positieve afwijkingen van het gemiddelde elkaar opheffen.
De *variantie* $\hat{\sigma}_x^2$ van de meetserie $\{x_1, x_2, \cdots, x_n\}$ is het kwadraat van de standaarddeviatie $\hat{\sigma}_x$:

$$\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2. \tag{3.3}$$

Het gemiddelde en de standaarddeviatie zijn statistische gegevens die gebruikt worden om een gemeten verdeling van waarden te karakteriseren.[1] In Figuur 3.1 zijn de waarden voor $\overline{x}$ en $\hat{\sigma}_x$ voor de testverdeling van

[1] De berekening van deze gegevens is wellicht al bekend van statistische functies op eenvoudige zakrekenmachines. Hierbij is $\overline{x}$ zoals hier gedefinieerd gelijk aan de toets $\overline{x}$, en $\hat{\sigma}_x \leftrightarrow \sigma_{[n-1]}$.

25 punten weergegeven. Het is in dit geval duidelijk dat de puntendichtheid in de buurt van het gemiddelde groter is dan ver ervan af, en dat de meeste punten worden gevonden in een band $\pm\hat{\sigma}_x$ rond $\overline{x}$.
De geïntroduceerde concepten *gemiddelde*, *variantie* en *standaarddeviatie* verraden iets over de statistische verdelingsfunctie die schuil gaat achter de meting (Hoofdstuk 4). De (vaak onuitgesproken) aanname hierbij is dat deze verdelingsfunctie gedurende de meting *stabiel* is, met andere woorden: de "echte" waarden van de betrokken grootheden mogen hierbij niet significant verlopen gedurende de meting.

Interpretatie van de standaarddeviatie Wanneer wij een willekeurige meting x_i uit de dataset zouden trekken of (onder dezelfde omstandigheden) nóg een meting x_{n+1} uit zouden voeren, is uit de spreiding die de meetwaarden $\{x_1, x_2, \cdots, x_n\}$ onderling vertonen, af te schatten dat deze meting in een gebied $\pm\hat{\sigma}_x$ rond $\overline{x}$ zal liggen. De standaarddeviatie is dan te interpreteren als een meetnauwkeurigheid, oftewel een maat voor de *onzekerheid* in elke afzonderlijke meting. De set meetgegevens $\{x_1, x_2, \cdots, x_n\}$ verschaft ons dus zowel een verwachte waarde ($\overline{x}$) als een onzekerheid voor een individuele (eventueel toekomstige) meting.
In uitdrukkingen (3.2) en (3.3) staat een voorfactor $\left(\frac{1}{n-1}\right)$ waar je wellicht $\left(\frac{1}{n}\right)$ zou verwachten (zoals voor het gemiddelde, Vgl. (3.1)). In Hoofdstuk 4 geven we een fundamentele onderbouwing van deze voorfactor. We beargumenteren nu voorlopig dat wanneer $n = 1$ (dus een dataset bestaand uit één enkele meetwaarde) de spreiding van meetwaarden onbepaald blijft. Dit wordt tot uitdrukking gebracht in de factor $\left(\frac{1}{n-1}\right)$.
Een andere manier om de voorfactor $\left(\frac{1}{n-1}\right)$ is door te kijken naar Vgl. (3.1) in iets andere vorm: $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$. Hieruit blijkt dan dat het laatste verschil $(x_n - \overline{x})$ bepaald kan worden uit de eerste $n-1$ termen. Met andere woorden, van de n verschillen $(x_i - \overline{x})$ kunnen er slechts $n-1$ waarden "vrij bewegen". Daarom worden de standaarddeviatie (en variantie) gedeeld door $n-1$ in plaats van n. De waarde $n-1$ noemt men het aantal *vrijheidsgraden*, daarover in Hoofdstuk 6 meer.

Onzekerheid in de standaarddeviatie De standaarddeviatie Vgl. (3.2) is uit metingen bepaald en heeft daarom een beperkte nauwkeurigheid. Deze nauwkeurigheid neemt toe naarmate het aantal beschikbare meetresultaten n stijgt. De "standaarddeviatie in de standaarddeviatie" $\hat{\sigma}_{\hat{\sigma}_x}$ wordt dus steeds kleiner voor grotere n. In Hoofdstuk 2 is de aanname gebruikt dat de *relatieve onzekerheid* in de standaarddeviatie $\hat{\sigma}_{\hat{\sigma}_x}/\hat{\sigma}_x$ in de regel niet kleiner is dan 10%, zodat een onzekerheid met hooguit twee significante cijfers mag worden weergegeven. We onderbouwen deze aanname.
In Appendix A wordt bewezen dat de relatieve onzekerheid (voor normaal verdeelde datasets, zie Hoofdstuk 4) bij benadering gegeven wordt door

$$\hat{\sigma}_{\hat{\sigma}_x}/\hat{\sigma}_x \approx \frac{1}{\sqrt{2(n-1)}}. \tag{3.4}$$

Uit Vgl. (3.4) leiden we af dat voor een relatieve nauwkeurigheid van 10% in de onzekerheid al $n \approx 51$ metingen nodig zijn. Een eventuele derde significant cijfer kan pas worden toegevoegd bij een relatieve nauwkeurigheid van 1%, waarvoor $n \approx 5000$ metingen vereist zijn. Dit aantal zal maar in een zeer beperkt aantal fysische meetsituaties gehaald worden.

3.3 De standaarddeviatie van het gemiddelde (SDOM)

De standaarddeviatie is het best te begrijpen als de onzekerheid voor een individuele meting. Anders gezegd, als ik één meting doe aan grootheid x, dan zal deze waarschijnlijk binnen een afstand $\pm\hat{\sigma}_x$ van $\overline{x}$ liggen. Een interessante vraag is nu hoe goed de waarde van $\overline{x}$ de "werkelijke waarde" X van de grootheid x weerspiegelt. We verwachten dat voor grote aantallen metingen $\overline{x}$ minder van X zal afwijken dan een willekeurige meting x_i.

De *standaarddeviatie* $\hat{\sigma}_{\overline{x}}$ *in het gemiddelde* $\overline{x}$ (*standard deviation of the mean* of SDOM) van de meetserie $\{x_1, x_2, \cdots, x_n\}$ wordt gedefinieerd als

$$\hat{\sigma}_{\overline{x}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{3.5}$$

$$\equiv \sqrt{\frac{1}{n}} \hat{\sigma}_x. \tag{3.6}$$

Analoog is de *variantie in het gemiddelde* gelijk aan $\hat{\sigma}^2_{\overline{x}}$, de SDOM in het kwadraat:

$$\hat{\sigma}^2_{\overline{x}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2 . \tag{3.7}$$

De SDOM $\hat{\sigma}_{\overline{x}}$ is de beste schatting voor de (statistische) onzekerheid van $\overline{x}$. Voor de "beste schatter" van de waarde van grootheid x op basis van de meetserie $\{x_1, x_2, \cdots, x_n\}$ rapporteren we nu $\overline{x} \pm \hat{\sigma}_{\overline{x}}$. Merk op dat ook voor deze rapportage met betrekking tot grootheden, eenheden en significantie de regels gelden zoals beschreven in het vorige hoofdstuk.
Wanneer een dataset $\{x_1, x_2, \cdots, x_n\}$ stochastisch ("toevallig") verdeeld is, dan wordt de statistisch bepaalde onzekerheid $\hat{\sigma}_{\overline{x}}$ in het gemiddelde $\overline{x}$ (schatting van de afwijking van de "echte" waarde X) steeds kleiner met toenemend aantal meetresultaten n. Er geldt hierbij ongeveer een omgekeerde evenredigheid met de wortel van het aantal metingen n: $\hat{\sigma}_{\overline{x}} \propto \frac{1}{\sqrt{n}}$. Dit in tegenstelling tot $\hat{\sigma}_x$, die behoudens statistische variaties niet afhankelijk is van de grootte van de dataset.

Mathematische identiteit Uit de definitie Vgl. (3.1) volgt de identiteit

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 &= \frac{1}{n} \left(\sum_{i=1}^{n} x_i^2 - 2\overline{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \overline{x}^2 \right) \qquad (3.8) \\ &= \frac{1}{n} \left(\sum_{i=1}^{n} x_i^2 - 2\overline{x}(n\overline{x}) + n\overline{x}^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\ &\equiv \overline{x^2} - \overline{x}^2. \qquad (3.9) \end{aligned}$$

Hierbij is in Vgl. (3.9) de definitie van $\overline{x^2}$ (het gemiddelde van de gekwadrateerde meetwaarden) ingevoerd. Deze vergelijking laat zien dat $\overline{x^2}$ altijd groter dan of gelijk is aan $\overline{x}^2$. Er geldt dus voor de variantie:

$$\hat{\sigma}^2_{\overline{x}} = \frac{1}{n-1} \left(\overline{x^2} - \overline{x}^2 \right). \tag{3.10}$$

3.4 De weergave van een gemeten frequentieverdeling

Voor het goed omgaan met kwantitatieve meetresultaten is inzicht in de statistiek belangrijk. Tot nu toe hebben we meetseries $\{x_1, x_2, \cdots, x_n\}$ weergegeven met het nummer van de meting op de x-as en de gemeten waarde op de y-as, zie bijvoorbeeld Figuur 3.1. Wanneer de metingen alleen toevallige onzekerheden

Tabel 3.1: In de bovenste tabel is een set van 80 (virtuele) tentamenresultaten weergegeven. Er is beoordeeld op een schaal van 0 tot 10, waarbij ook halftallige resultaten zijn toegestaan. In de onderste tabel staat de bepaalde frequentie van voorkomen van elk cijfer. De statistische eigenschappen van deze dataset: gemiddelde $\overline{x} = 5.90$, standaarddeviatie $\hat{\sigma}_x = 1.4$, standaarddeviatie van het gemiddelde $\hat{\sigma}_{\overline{x}} = 0.16$.

5.5	5	6.5	5.5	6.5	4.5	7	2.5	5	7.5
7	4.5	3.5	7.5	5.5	5.5	5.5	5	6.5	4
6	6.5	6	7.5	3.5	5.5	5	7	6	4.5
4.5	5.5	4	8	5.5	2.5	6.5	6	8.5	7
6	8.5	6	6.5	8	7	8	6.5	3	5
5	5	6.5	4	10	8	3.5	7	5.5	6.5
3.5	6.5	5.5	7	6	6	7	6	6.5	7
8	6	5	5	5.5	7	7	3.5	6	7

Cijfer	**Aantal**	**Cijfer**	**Aantal**	**Cijfer**	**Aantal**
2.5	2	5.0	9	7.5	3
3.0	1	5.5	11	8.0	5
3.5	5	6.0	11	8.5	2
4.0	3	6.5	11	10	1
4.5	4	7.0	12		

bevatten is het nummer van de meting in de statistiek geen relevant gegeven. We kunnen de metingen ook door elkaar gooien; de statistische eigenschappen van de data veranderen hierdoor niet. Liever zouden we laten zien "hoe vaak welke gemeten waarde voorkomt" en "welke typische spreiding de waarden tonen".
In de statistiek wordt de distributie van de dataset $\{x_1, x_2, \cdots, x_n\}$ over de waarden van een grootheid x de *frequentieverdeling* van de dataset genoemd (met frequentie in de betekenis van "mate van voorkomen").
Voor de constructie van een dergelijke verdeling uit een set meetgegevens moeten we in de praktijk onderscheid maken tussen situaties waar de meetuitkomsten x_i in principe discreet zijn (bijvoorbeeld tentamencijfers of telexperimenten bij radioactieve emissie) en situaties waar deze continu zijn (bijvoorbeeld lengtemetingen). Bij een *discrete verdeling* wordt daarbij de *kans* of *waarschijnlijkheid* gegeven om specifieke uitkomsten te vinden. Bij een *continue verdeling* wordt de *kans-* of *waarschijnlijkheidsdichtheid* gegeven, waaruit de waarschijnlijkheid volgt om meetwaarden te vinden in een bepaald *interval* van mogelijke uitkomsten.
We geven voor zowel discreet als continu verdeelde grootheden het recept om de frequentieverdeling te verkrijgen.
Opmerking: veel meetprocedures voor continue grootheden zijn discreet en/of digitaal van aard, en leveren dus eigenlijk discrete uitkomsten op. In veel gevallen zijn de discrete stappen zodanig klein dat ze bij het weergeven van de frequentieverdeling niet relevant zijn. Wanneer de discrete stappen erg klein zijn, is het volgen van het recept voor continu verdeelde uitkomsten toch het meest toepasselijk.

<u>Discreet</u> verdeelde uitkomsten:

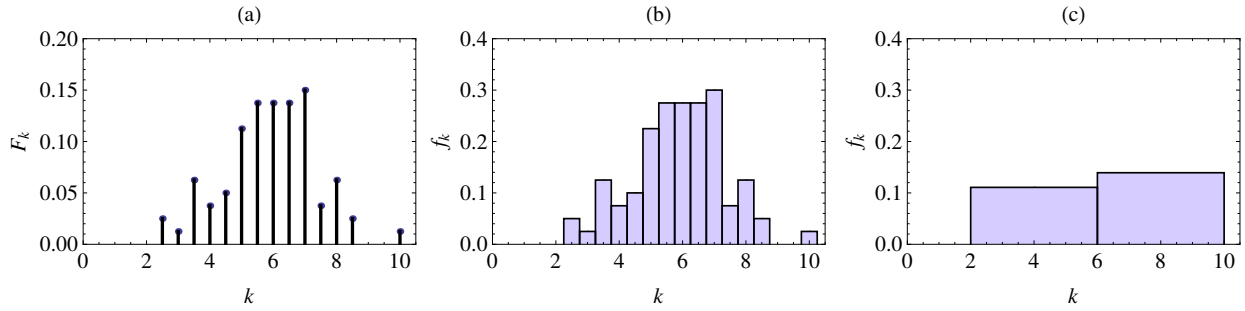
1. bepaal voor elke mogelijke uitslag k hoe vaak deze voorkomt (n_k). De waarden n_k zelf kunnen uiteraard worden uitgezet tegen de betrokken uitkomsten k, maar in veel gevallen wordt eerst n_k genormeerd op het totaal aantal uitslagen $n = \sum n_k$, zodat de *fractie* of *kans* $F_k = n_k/n$ wordt verkregen. De normering met een factor $1/n$ zet absolute aantallen om naar *relatieve frequenties*. De frequentieverdeling is dan *genormaliseerd*.

2. maak een *bar-histogram* van de fracties F_k tegen de mogelijke waarden k, waarbij de fractie of kans F_k wordt gegeven door de "hoogte" van de "*bar*" bij waarde k. Met deze constructie is de som van de *hoogtes* van de *bars* gelijk aan 1.

<u>Continu</u> verdeelde uitkomsten:

1. verdeel het x-bereik in een discreet aantal k_{max}, direct op elkaar aansluitende, intervallen (zogenaamde "*bins*", letterlijk: "bakjes") met een geschikt gekozen intervalgrootte Δx_k. Een vuistregel voor het aantal bins k_{max} is de wortel van het aantal resultaten, $k_{max} = \sqrt{n}$ (uiteraard afgerond op hele aantallen).
2. bepaal n_k, dat is nu het aantal x-waarden in de dataset $\{x_1, x_2, \cdots, x_n\}$ dat ligt in het k^e interval. Vervolgens kan de fractie van het aantal waarden per interval F_k verkregen worden door normering: $F_k = n_k/n$.
3. bepaal vervolgens $f_k = F_k/(\Delta x_k)$, met Δx_k de breedte van de k^e bin. We zeggen dat f_k een *kansdichtheid* is: een *differentiële* grootheid, die alleen betekenis heeft in combinatie met een intervalbreedte.
4. maak een *bin-histogram* van de gevonden kansdichtheid f_k over de bijbehorende intervallen. In dit histogram wordt f_k gegeven door de *hoogte* van de k^e *bin*. Het *oppervlak* beslagen per *bin* is gelijk aan de fractie F_k van het aantal x-waarden dat in het k^e interval ligt. Met deze constructie is het totale *oppervlak* beslagen door het histogram gelijk aan 1.

Een voorbeeld van een dataset is de uitslag van een (nogal slecht gemaakt) tentamen in Tabel 3.1. Hier is ook de frequentie van voorkomen van de (discrete) uitkomsten bepaald. In Figuur 3.2 (a) is het *bar-histogram* van de relatieve frequenties weergegeven. Als ik nu willekeurig een student kies (zonder te weten wat de tentamenuitslag voor deze student is) dan geeft F_k de kans dat deze student de uitslag k heeft behaald. De kans dat een willekeurige student uit de groep van 80 één willekeurige uitslag heeft behaald is uiteraard 1 (elke deelnemer heeft een uitslag). De som van de *hoogten* in de figuur is daarom genormeerd op 1.
Het staat ons in dit geval vrij om net te doen als of deze uitslag continu verdeeld is en de uitslagen te "*binnen*", d.w.z. intervallen (*bins*) te definiëren en te tellen hoeveel uitslagen in de verschillende intervallen liggen. In principe kunnen we de grootte van de intervallen vrij kiezen. Het heeft echter weinig zin om intervallen met een breedte kleiner dan 0.5 te kiezen; dan zijn er intervallen die nooit "gevuld" worden, vanwege het discrete karakter van de uitslagen. Evenzo geven veel te grote intervallen het detailniveau van de verdeling niet goed weer.
We kiezen als voorbeeld intervallen met een breedte van 0.5, rondom de mogelijke uitslagen. Het interval bij bijvoorbeeld uitslag 7.5 loopt dan van 7.25 tot 7.75. Het aantal uitslagen dat we per interval registreren is uiteraard weer gelijk aan de eerder voor het *bar-histogram* geïntroduceerde n_k. Echter bij een *bin-histogram* wordt niet F_k weergegeven maar $f_k = F_k/\Delta x_k$, waarbij Δx_k de breedte van het k^e interval. De "hoogte" van de *bin* is hierbij gelijk aan f_k (zie Figuur 3.2 (b)). Om F_k te vinden voor een waarde in het k^e interval moet dus altijd f_k worden vermenigvuldigd met de breedte van het k^e interval (vergelijk hiervoor de verticale schalen van Figuur 3.2 (a) en (b)). Figuur 3.2 (c) is een voorbeeld van een histogram met een te laag aantal *bins*, waardoor het goed beoordelen van de verdeling feitelijk onmogelijk is. Voor deze tentamenuitslag leidt de vuistregel tot $k_{max} = 9$.
In histogrammen kun je vaak snel zien of de verdeling van de data bijzondere effecten vertoont (de verdeling kan bijvoorbeeld twee pieken hebben). Je kunt dan iets concluderen over relevantie en bruikbaarheid van het gemiddelde, of de typische spreiding van de data. Bij voldoende metingen kan de verdeling gezien worden als een goede voorspelling voor toekomstige resultaten. Voor de tentamenopgave bijvoorbeeld: als een willekeurige student dit tentamen maakt, wat is dan de kans op een 7, of een onvoldoende?
In fysische situaties waarbij zéér grote aantallen metingen worden verzameld, zal de gemeten verdeling steeds meer gaan lijken op een "werkelijke" onderliggende verdeling. In veel voorkomende (fysische) si-



Figuur 3.2: Grafische weergave van verdeling van tentamenuitslagen. (a) Een *bar-histogram* waarin de discrete kansen F_k zijn weergegeven. De som van de *hoogten* is gelijk aan 1. (b) Een *bin-histogram* waarvoor de resultaten eerst zijn ge"*bin*"d (intervalbreedte bedraagt hier 0.5). Weergegeven is f_k, dat is de kans *per eenheid van intervalbreedte*. f_k is een differentiële grootheid. Het totale *oppervlak* beslagen door het *bin-histogram* is gelijk aan 1. (c) Een *bin-histogram* met een intervalbreedte van vier. Door de grote intervalbreedte ten opzichte van de spreiding in resultaten is alle detailinformatie over de verdeling verdwenen.

tuaties kan de verdeling van waarden beschreven worden door een analytische uitdrukking, een *verdelingsfunctie*. Dit onderwerp wordt behandeld in het volgende hoofdstuk.

Hoofdstuk 4

Kansverdelingsfuncties

4.1 Inleiding

Wanneer we een eindige set van n metingen vergaren voor de fysische grootheid x met "echte" waarde X - die in veel gevallen niet bekend is - dan laat de analyse aan het begin van het vorige hoofdstuk zien dat deze metingen informatie bevatten over de meetonzekerheid $\hat{\sigma}_x$ (SD), de beste schatter $\overline{x}$ (het al dan niet gewogen gemiddelde) voor X, en de standaarddeviatie $\hat{\sigma}_{\overline{x}}$ in dit gemiddelde (SDOM). Verder kan op basis van de data een frequentieverdeling worden vastgesteld.
De informatie is echter noodgedwongen beperkt. Naarmate we het aantal metingen n opvoeren (idealiter $n \to \infty$), wordt de informatie over X en de achterliggende statistische verdeling van x steeds nauwkeuriger bekend en kunnen nauwkeuriger uitspraken worden gedaan. In de limiet $n \to \infty$ zal bovendien de gemeten frequentieverdeling gelijk worden aan de "werkelijke" verdeling, die we zullen aanduiden als *limiting distribution* of *(kans)verdelingsfunctie*.
In veel situaties zal ons echter de tijd en/of mogelijkheid ontbreken om zeer uitgebreide meetseries te genereren. In een dergelijke situatie is het een groot voordeel als op basis van fysische of mathematische overwegingen al te zeggen is welke verdelingsfunctie in de meetsituatie van toepassing is. Met die informatie kan een optimale verwerking van de (beperkte) meetserie plaatsvinden die recht doet aan de achterliggende statistiek. In dit hoofdstuk volgt daarom een inleiding tot een aantal standaard verdelingsfuncties.
Met behulp van de verdelingsfunctie, of de verwachtingswaarde X en standaarddeviatie σ_x kunnen *statistische uitspraken* of *voorspellingen* gedaan worden. Hiermee wordt bedoeld dat de kans op waarheid van een gegeven uitspraak of voorspelling kan worden berekend. Het maakt hier niet uit of de verdelingsfunctie is afgeleid op theoretische gronden of afgeleid uit een zeer grote set metingen. In de natuurkunde, maar vooral in medische en sociale wetenschappen, wordt vaak de kans berekend dat een resultaat significant verschilt van een ander (gerelateerd) resultaat. Hiervoor gebruikt men het begrip *overschrijdingskans*. We behandelen een correcte toepassing van dit begrip in de laatste sectie van dit hoofdstuk.[1]

4.2 Verdelingsfuncties

We veronderstellen allereerst dat voor een gegeven meetproces een *statistische verdelingsfunctie* bestaat. De beide grootheden *verwachtingswaarde* (ook wel het *gemiddelde* genoemd) en *standaarddeviatie* (de wortel

[1] Men vindt vaak sensationele mediaberichten die te maken hebben met gezondheid, voeding, politiek, of het verschil tussen mannen en vrouwen, die overduidelijk statistische uitspraken als absolute waarheden bestempelen, en waarbij je je af kunt vragen of er wel sprake is van een volledig aselecte steekproef. Aan de vroegere Engelse premier Disraeli wordt wel de uitspraak "*There are three kinds of lies: lies, damned lies, and statistics*" toegeschreven. Houd daarom de lessen van dit hoofdstuk in gedachte wanneer je dit soort berichten tegenkomt!

van de *variantie*) zijn karakteristieke grootheden van deze verdelingsfunctie. Wanneer de verdelingsfunctie bekend is, kunnen deze karakteristieke eigenschappen direct berekend worden. We geven daarvoor de volgende algemene *identiteiten* in het geval van zowel discrete als continue verdelingen.

Discrete verdelingen Een discrete verdelingsfunctie F_k levert bij elk van de mogelijke uitkomsten k de *kans* Prob(k) op die uitkomst. Een overzicht van de karakteristieke eigenschappen van discrete verdelingen:

$$\text{Kans: Prob}(k) \qquad \equiv F_k \text{ (kans op waarde } k) \tag{4.1}$$

$$\text{Normering: } \sum_{k=1}^{n} F_k \qquad = 1 \tag{4.2}$$

$$\text{Verwachtingswaarde: } K \qquad = \sum_{k=1}^{n} kF_k \tag{4.3}$$

$$\text{Variantie: } \sigma_k^2 \qquad = \sum_{k=1}^{n} (k-K)^2 F_k \tag{4.4}$$

Continue verdelingen Een continue verdelingsfunctie specificeert de *kansdichtheid* $f(x)$ bij een gegeven uitkomst x. Omdat de *kans* op exact die uitkomst x gelijk is aan nul, wordt een kans Prob$(x \leq x' \leq x+\mathrm{d}x)$ vastgelegd op een interval $[x, x+\mathrm{d}x]$. De kans op die uitkomst is dan gelijk aan $f(x)\mathrm{d}x$. De eigenschappen van continue verdelingen samengevat:

$$\text{Kans: Prob}(x \leq x' \leq x+\mathrm{d}x) \qquad \equiv f(x)\mathrm{d}x \text{ (kans op waarde in interval } [x, x+\mathrm{d}x]) \tag{4.5}$$

$$\text{Normering: } \int_{-\infty}^{\infty} f(x)\mathrm{d}x \qquad = 1 \tag{4.6}$$

$$\text{Verwachtingswaarde: } X \qquad = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x \tag{4.7}$$

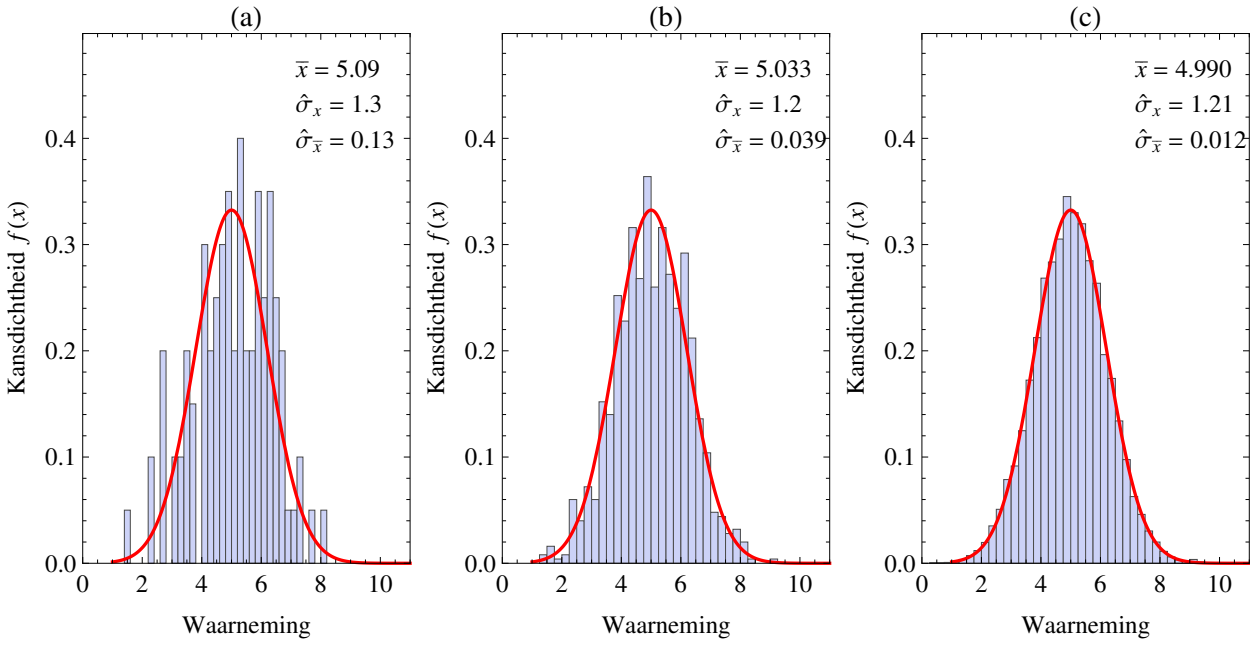
$$\text{Variantie: } \sigma_x^2 \qquad = \int_{-\infty}^{\infty} (x-X)^2 f(x)\mathrm{d}x \tag{4.8}$$

4.3 De relatie tussen gemeten verdelingen en verdelingsfuncties

In het vorige hoofdstuk hebben we aangegeven hoe "beste schattingen" verkregen kunnen worden voor het gemiddelde $\overline{x}$ en de standaarddeviatie $\hat{\sigma}_x$, op basis van een eindige *meetserie* $\{x_1, x_2, \cdots, x_n\}$. In de vorige sectie hebben we de verwachtingswaarde X en de standaarddeviatie σ_x voor een *verdelingsfunctie* wiskundig bepaald. Het onderscheid in de notatie is gemaakt door voor bepalingen op basis van meetseries boven het gemiddelde $\bar{}$ en boven de standaarddeviatie $\hat{}$ te plaatsen. In tegenstelling tot $\overline{x}$ en $\hat{\sigma}_x$, die een onzekerheid kennen, zijn X en σ_x *exact*, gegeven de verdelingsfunctie.

We veronderstellen dat in elke meetsituatie aan de gemeten grootheid een verdelingsfunctie kan worden toegekend. De metingen kunnen we opvatten als een *eindige* set van "trekkingen" uit deze verdelingsfunctie. In deze sectie definiëren we hoe voor een steeds groter wordend aantal metingen de eindige meetserie naar de achterliggende verdelingsfunctie "toegroeit". De verdelingsfunctie kan dan worden opgevat als de *limietverdeling* voor het aantal metingen $n \to \infty$. De uitwerking is uitgevoerd voor continue verdelingen, maar analoge uitdrukkingen en conclusies zijn van toepassing op discrete verdelingen.

In Figuur 4.1 is te zien hoe de frequentieverdeling van een gemeten dataset voor een steeds groter wordend aantal metingen convergeert naar de onderliggende verdelingsfunctie.



Figuur 4.1: Illustratie van de convergentie van een gemeten verdeling naar de limietverdeling (een normale verdeling (Sectie 4.5.1) met $X = 5$ en standaarddeviatie $\sigma_x = 1.2$). De gemeten genormeerde frequentieverdelingen zijn van (a) naar (c) bepaald voor $n = 100$, $n = 1000$ en $n = 10000$, en weergegeven als *bin-histogram*. De doorgetrokken lijn is de eerder genoemde (normale) verdelingsfunctie. De statistische eigenschappen van de dataset zijn ook in de figuren weergegeven. Zo is duidelijk dat voor toenemend aantal n het gemiddelde steeds nauwkeuriger bepaald wordt ($\sigma_{\overline{x}} \Rightarrow 0$), terwijl de standaarddeviatie $\hat{\sigma}_x$ niet verandert. Merk op dat vanwege Vgl. (3.4) de nauwkeurigheid van $\hat{\sigma}_x$ wél toeneemt.

4.3.1 Uitspraken geldig in de statistische limiet

We beschouwen weer een statistisch onafhankelijke dataset $\{x_1, x_2, \cdots, x_n\}$ met (continue) kansverdelingsfunctie $f(x)$. Als "beste" schattingen voor de verwachtingswaarde X en de standaarddeviatie σ_x van deze statistische verdeling hebben we respectievelijk ingevoerd het gemiddelde $\overline{x}$ en de standaarddeviatie $\hat{\sigma}_x$. De standaarddeviatie σ_X (met "beste schatting" $\hat{\sigma}_{\overline{x}}$, de SDOM) van de "echte" waarde X is eveneens een statistische grootheid.
Naarmate het aantal metingen n groter is kunnen we in principe steeds nauwkeuriger uitspraken doen over X, σ_x en σ_X. Om de verdelingsfunctie echt te kunnen achterhalen zou je naar de *statistische limiet* moeten gaan, dat wil zeggen, $n \to \infty$ metingen moeten verzamelen en bekijken hoe de gemeten waarden onderling spreiden. We kunnen wel *statistische uitspraken* (d.w.z. uitspraken die beter gelden naarmate er een groter aantal n van toevallig verdeelde meetresultaten wordt beschouwd) formuleren:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \Rightarrow X \tag{4.9}$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad \Rightarrow \sigma_x, \tag{4.10}$$

$$\hat{\sigma}_{\overline{x}} = \hat{\sigma}_x/\sqrt{n} \qquad \Rightarrow \sigma_X = \sigma_x/\sqrt{n} = 0. \tag{4.11}$$

We gebruiken hier het symbool $\Rightarrow$ dat een "=" wordt in de limiet $n \to \infty$. In de statistische limiet gaat de onzekerheid $\hat{\sigma}_{\overline{x}}$ in $\overline{x}$ naar nul, en wordt $\overline{x}$ gelijk aan X. De standaarddeviatie van de verdeling, die we in een meting identificeren met de onzekerheid voor een enkele meting, gaat naar de constante waarde σ_x.

4.3.2 Bewijs van de voorfactor $\frac{1}{n-1}$ in $\hat{\sigma}_x^2$

In voorgaand hoofdstuk is de variantie $\hat{\sigma}_x^2$ gedefinieerd middels Vgl. (3.3). In deze vergelijking is een voorfactor $\frac{1}{n-1}$ in plaats van $\frac{1}{n}$ opgenomen. Als voorlopige legitimatie gebruikten we toen dat een standaarddeviatie voor één enkel punt ongedefinieerd blijft. We geven nu het bewijs dat het gebruik van $\frac{1}{n-1}$ daadwerkelijk de "beste" schatting van σ_x van de verdelingsfunctie oplevert.
Hiervoor beginnen wij bij de kwadratische term in Vgl. (3.3). Deze werken wij uit tot

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\overline{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 && (4.12)\\
&= \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} x_i x_j\right) && (4.13)\\
&= \left(1 - \frac{1}{n}\right)\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} x_i x_j. && (4.14)
\end{aligned}$$

Voordat we de afleiding afronden, definiëren we analoog aan Vgl. (4.7) in de statistische limiet

$$\overline{x^2} \Rightarrow \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x. \tag{4.15}$$

Uit uitsplitsing van de kwadratische term in Vgl. (4.8) en door gebruik te maken van $\int xXf(x)\mathrm{d}x = X^2$ en

$\int X^2 f(x)\mathrm{d}x = X^2$ volgt nu dat

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - X^2, \text{ oftewel}$$
$$\overline{x^2} \Rightarrow X^2 + \sigma_x^2. \tag{4.16}$$

Terugkomend op Vgl. (4.14) geldt nu in de statistische limiet onder gebruikmaking van Vgl. (4.16):

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\overline{x})^2 &= \left(1-\frac{1}{n}\right)\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}x_i x_j \\ &\Rightarrow \left(1-\frac{1}{n}\right)n\left(X^2+\sigma_x^2\right) - \frac{1}{n}n(n-1)X^2 \\ &= (n-1)\sigma_x^2. \end{aligned} \tag{4.17}$$

Derhalve geldt in de statistische limiet

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \Rightarrow \sigma_x^2. \tag{4.18}$$

Het gebruik van de voorfactor $(n-1)$ in Vgl. (3.3) is dan ook terecht. De bepaling van de variantie $\hat{\sigma}_x^2$ op basis van een *gemeten* dataset $\{x_1, x_2, \cdots, x_n\}$ middels Vgl. (3.3) vormt dus ook de "beste schatter" van de variantie σ_x^2 van de achterliggende verdelingsfunctie.[2]

4.4 Discrete kansverdelingsfuncties

We zullen nu een aantal veel voorkomende verdelingsfuncties onder de loep nemen. We beginnen met twee veel voorkomende discrete verdelingen, de *binomiale verdeling* en de *Poissonverdeling*.

4.4.1 De binomiale verdeling

De binomiale verdeling illustreren we aan de hand van een dobbelexperiment. Bij het werpen met een "zuivere" dobbelsteen verwacht je per worp een kans van $p = \frac{1}{6}$ om een één (*ace*) te vinden. We definiëren dit als een *succes*. De kans q op een andere uitkomst (geen succes) is dan per definitie $q = 1-p = \frac{5}{6}$. We kunnen nu vragen formuleren als: gesteld dat je tien keer werpt ($n_t = 10$), wat is dan de kans op geen enkele *ace* (aantal successen $\nu = 0$)? En de kans op vijf keer een *ace* ($\nu = 5$)? Of de kans op meer dan drie *aces*? Algemener geformuleerd levert dit het volgende op. Aangenomen dat per worp (trekking) mijn kans op succes gegeven wordt door p (kans $q = 1-p$ op geen succes) dan is de kans op ν successen in n_t worpen $(0 \leq \nu \leq n_t)$ gelijk aan

$$B_{n_t,p}(\nu) = \binom{n_t}{\nu} p^{\nu}(1-p)^{n_t-\nu}. \tag{4.19}$$

Vergelijking (4.19) wordt de *binomiale verdelingsfunctie* genoemd. Hierin is $\binom{n_t}{\nu} = \frac{n_t!}{\nu!(n_t-\nu)!}$ een binomiaalcoëfficiënt. Deze coëfficiënt geeft het aantal verschillende manieren waarmee ν successen kunnen worden behaald uit n_t worpen. Als je bijvoorbeeld één enkel succes $(\nu = 1)$ behaalt in n_t worpen, dan kan

[2] De finesse in dit bewijs zit in de overgang (4.12) - (4.13). In de vermenigvuldiging $\sum_i x_i \sum_i x_i$ zijn de gevallen waarbij i in de eerste term gelijk is aan i in de tweede term *niet statistisch onafhankelijk*. Anders gezegd, $\overline{x}$ is niet onafhankelijk van de gemeten data, maar is bepaald uit die data. Daarom is in de statistische limiet $\sum_i x_i \sum_i x_i \not\Rightarrow (nX)(nX) = n^2X^2$, maar $\sum_i x_i \sum_i x_i \Rightarrow n^2X^2 + n\sigma_x^2$!

deze éne keer succes behaald worden bij de *eerste*, óf bij de *tweede*, $\cdots$, óf bij de n_t^{e} worp. Dit geeft in totaal n_t verschillende mogelijkheden, overeenkomend met de binomiaalcoëfficiënt $\binom{n_t}{1} = n_t$.
Middels het *Binomium van Newton* is te controleren dat de totale kans om $\nu = 0$, óf 1, óf 2, $\cdots$, óf n_t successen te behalen gelijk is aan één:

$$\sum_{\nu=0}^{n_t} B_{n_t,p}(\nu) = \sum_{\nu=0}^{n_t} \binom{n_t}{\nu} p^{\nu} q^{n_t-\nu} \tag{4.20}$$

$$= (p+q)^{n_t} \tag{4.21}$$

$$= 1. \tag{4.22}$$

De binomiale verdeling is dus correct *genormeerd* naar Vgl. (4.2). De verwachtingswaarde en de variantie van de binomiale verdeling (respectievelijk (4.3) en (4.4)) worden berekend als:

$$\text{de } \textit{verwachtingswaarde } N_\nu = \sum_{\nu=0}^{n_t} \nu B_{n_t,p}(\nu)$$

$$= n_t p, \tag{4.23}$$

$$\text{de } \textit{variantie } \sigma_\nu^2 = \sum_{\nu=0}^{n_t} (\nu - N_\nu)^2 B_{n_t,p}(\nu)$$

$$= n_t p(1-p). \tag{4.24}$$

Wanneer in een meting n_t trekkingen worden uitgevoerd, kunnen op basis van de uitdrukkingen in Hoofdstuk 3 een gemiddeld aantal successen $\overline{\nu}$, een standaarddeviatie $\hat{\sigma}_\nu$ en een standaarddeviatie van het gemiddelde $\hat{\sigma}_{\overline{\nu}}$ bepaald worden. Naarmate het aantal meetseries n toeneemt, waarbij in elke meetserie n_t trekkingen verricht worden, zal gelden

$$\begin{aligned} &\overline{\nu} \Rightarrow n_t p, \\ &\hat{\sigma}_\nu \Rightarrow \sigma_\nu = \sqrt{n_t p(1-p)}, \\ &\hat{\sigma}_{\overline{\nu}} = \hat{\sigma}_\nu/\sqrt{n} \Rightarrow \sigma_\nu/\sqrt{n} = 0. \end{aligned} \tag{4.25}$$

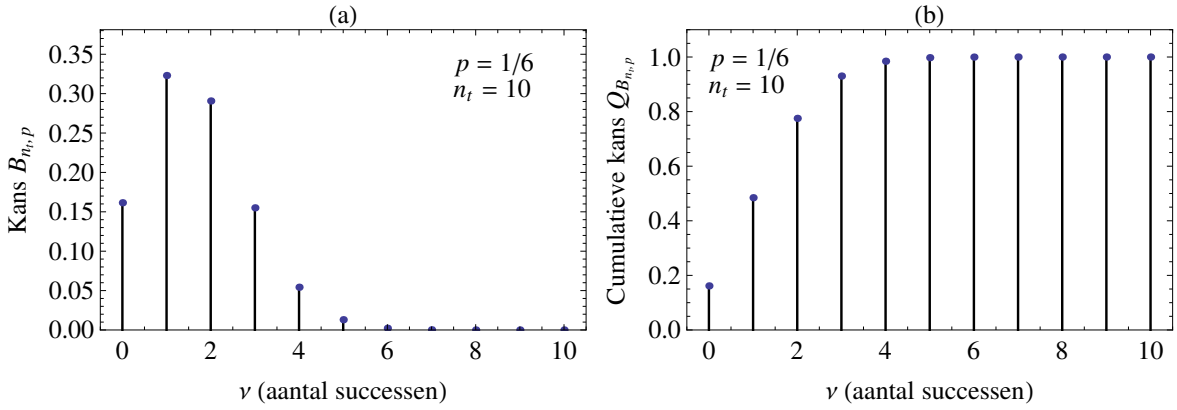
We introduceren hier ook de cumulatieve verdeling $Q_{B_{n_t,p}}(\nu)$ als de som van alle kansen bij waarden $0 \leq \nu' \leq \nu$:

$$Q_{B_{n_t,p}}(\nu) = \sum_{\nu'=0}^{\nu} B_{n_t,p}(\nu'). \tag{4.26}$$

De cumulatieve verdeling is een goed middel om de kans op een aantal successen groter of kleiner dan een zeker aantal te bepalen. Een voorbeeld van discrete resultaten voor een binomiale verdeling (4.19) met $p = 1/6$ en $n_t = 10$ (het eerder aangehaalde voorbeeld van het dobbelexperiment) wordt getoond in Figuur 4.2 (a). Hieruit valt af te lezen dat de kans op nul keer een *ace*, $B_{10,1/6}(\nu = 0)$, gelijk is aan 0.16. De kans op vijf keer een *ace*, $B_{10,1/6}(\nu = 5)$ is gelijk aan (slechts) 0.01.
Voor het bepalen van de kans op meer dan drie keer een *ace* gebruiken we de cumulatieve verdelingsfunctie $Q_{B_{10,1/6}}(\nu)$ die is weergegeven in Figuur 4.2 (b). In deze figuur zien we dat de kans op minder dan of gelijk aan twee successen $Q_{B_{10,1/6}}(\nu = 2)$ gelijk is aan 0.78. De kans op drie of meer keer succes wordt dan gelijk aan $1 - 0.78 = 0.22$.

4.4.2 De Poissonverdeling

Ook de *Poissonverdeling* is discreet van aard. Deze verdelingsfunctie is van groot belang bij *telexperimenten*, zo lang de *gemiddelde telsnelheid* of *rate* een goed gedefinieerde grootheid is gedurende de tijd



Figuur 4.2: (a) Binomiale verdeling $B_{n_t,p}(\nu)$, met $n_t = 10$ en $p = 1/6$. (b) Cumulatieve binomiale verdeling $Q_{B,n_t,p}(\nu)$ voor dezelfde waarden van n_t en p. De cumulatieve verdeling drukt de kans uit op een waarde kleiner dan of gelijk aan ν.

dat geteld wordt. Voorbeelden van metingen waar de Poissonverdelingsfunctie vaak van toepassing is, betreft metingen waarbij deeltjes- of stralingsdetectoren worden gebruikt. Veel van deze detectoren geven per gedetecteerd *event* een (elektronische) puls af. Het aantal in een meetperiode getelde pulsen voldoet dan doorgaans aan de Poissonstatistiek.

Bij gegeven *rate r* verwacht men in een tijdinterval T gemiddeld een aantal $\mu = rT$ (tellingen) te registreren. De Poissonverdelingsfunctie $P_\mu(\nu)$ geeft nu de kans om bij een meting over het tijdinterval T een aantal ν te tellen. Hier is uiteraard ν een geheel getal (≥ 0), maar het gemiddelde μ zal doorgaans geen geheel getal zijn.

De Poissonverdeling kan begrepen worden als een binomiale verdeling met oneindig veel trekkingen, namelijk een trekking in elk tijdsinterval Δt, met de limiet $\Delta t \to 0$. Je voert dan $n_t = T/\Delta t$ trekkingen uit, met in elke trekking is een kans $p = r\Delta t$ op succes. Invullen in de binomiale verdeling Vgl. (4.19) geeft na uitwerken

$$\lim_{\Delta t \to 0} \left[\binom{T/\Delta t}{\nu} (r\Delta t)^\nu (1 - r\Delta t)^{-\nu + T/\Delta t} \right] = \frac{1}{\nu!} \mu^\nu e^{-\mu}. \tag{4.27}$$

Hierbij is gebruik gemaakt van $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$, $\mu = rT$ en $\frac{(T/\Delta t)!}{(T/\Delta t - \nu)!} \sim (T/\Delta t)^\nu$ als $(T/\Delta t) \gg \nu$.
De Poissonverdeling is nu

$$P_\mu(\nu) = \frac{1}{\nu!} \mu^\nu e^{-\mu}. \tag{4.28}$$

De totale kans om een geheel aantal $\nu = (0, 1, 2, \cdots)$ successen te behalen gelijk is aan één (*normering*):

$$\sum_{\nu=0}^{\infty} P_\mu(\nu) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \mu^\nu e^{-\mu} = 1. \tag{4.29}$$

Belangrijke karakteristieken van de Poissonverdeling $P_\mu(\nu)$:

$$\begin{aligned} \text{de } \textit{verwachtingswaarde } N_\nu &= \sum_{\nu=0}^{\infty} \nu P_\mu(\nu) \\ &= \mu, \end{aligned} \tag{4.30}$$

$$\begin{aligned} \text{de } \textit{variantie } \sigma_\nu^2 &= \sum_{\nu=0}^{\infty} (\nu - N_\nu)^2 P_\mu(\nu) \\ &= \mu. \end{aligned} \tag{4.31}$$

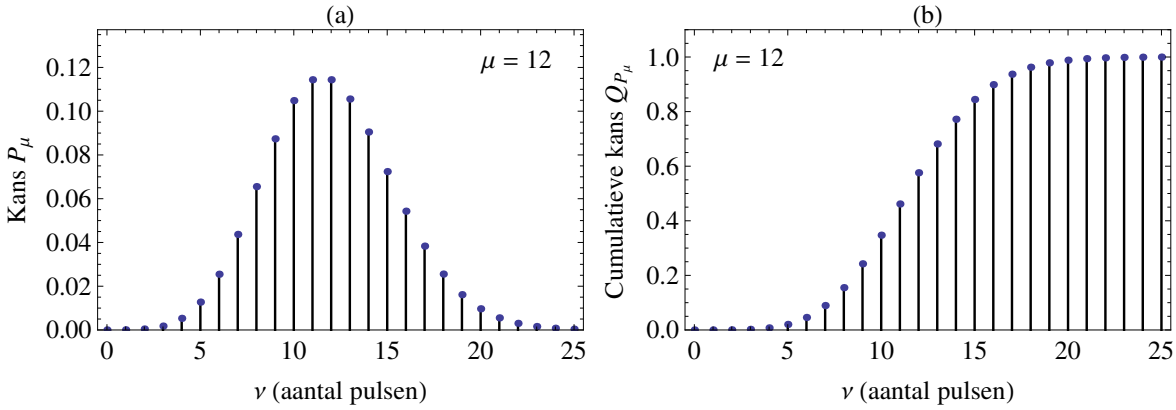
Voor een Poissonverdeling met gemiddelde μ geldt dus dat de standaarddeviatie gelijk is aan $\sqrt{\mu}$. Uit telexperimenten die voldoen aan de voorwaarden voor Poissonstatistiek zijn dus *bij voorbaat* statistische eigenschappen af te leiden. Stel dat in één enkel telexperiment een feitelijk aantal ν wordt geteld. De "beste" schatting van de statistische onzekerheid $\hat{\sigma}_\nu$ in dit getal ν is dan gelijk aan $\sqrt{\nu}$.
Ter vermijding van misverstand: niet het feitelijk getelde aantal ν is hierbij onzeker; we kunnen echt wel goed tellen! Het gaat hierbij om de *beperkte reproduceerbaarheid* van het getelde aantal. Bij herhalingen van het telexperiment onder identieke omstandigheden worden vanwege statistische redenen steeds andere aantallen geteld.
Een ander kenmerk van de Poissonverdeling is voorts dat de som van twee Poissonverdeelde grootheden met respectievelijk gemiddelde μ_1 en μ_2 opnieuw een Poissonverdeling oplevert, nu met gemiddelde $\mu_1 + \mu_2$. Uit deze regel valt af te leiden dat voor twee telexperimenten met getelde aantallen ν_1 en ν_2, de som $\nu_1 + \nu_2$ een "beste" schatting voor de onzekerheid kent van $\sqrt{\nu_1 + \nu_2}$.
Wanneer n identieke telexperimenten worden gedaan, met resultaten $\nu_1, \nu_2, \cdots, \nu_n$, kunnen met behulp van de vergelijkingen in Hoofdstuk 3 gemiddelde $\overline{\nu}$, standaarddeviatie $\hat{\sigma}_\nu$ en standaarddeviatie van het gemiddelde $\hat{\sigma}_{\overline{\nu}}$ worden bepaald. Op basis van Vgl. (4.30) en (4.31) geldt in de statistische limiet $n \to \infty$:

$$\begin{aligned} \overline{\nu} &\Rightarrow \mu \\ \hat{\sigma}_\nu &\Rightarrow \sqrt{\mu} \\ \hat{\sigma}_{\overline{\nu}} &\Rightarrow \sqrt{\mu}/\sqrt{n} = 0. \end{aligned} \tag{4.32}$$

De cumulatieve Poissonverdeling $Q_{P_\mu}(\nu)$ definiëren wij nu analoog aan Vgl. (4.26) voor de binomiale verdeling als de som van alle kansen voor tellingen $0 \leq \nu' \leq \nu$:

$$Q_{P_\mu}(\nu) = \sum_{\nu'=0}^{\nu} P_\mu(\nu'). \tag{4.33}$$

Hier een voorbeeld van een telexperiment. We bekijken het aantal door een ^{90}Sr-bron uitgezonden β-deeltjes met een Geiger-Müller-teller. Over een zeer lange meettijd is de *rate* voor deze detectie bepaald op $r = 24$ pulsen/s. We kunnen nu het verwachte aantal tellingen ν in tijd $T = 0.5$ s berekenen met behulp van de Poissonverdeling. Het gemiddelde aantal tellingen is in dit geval $\mu = rT = 12$ (merk op dat in het algemeen μ geen geheel getal hoeft te zijn). De standaarddeviatie van de verdeling is gelijk aan $\sqrt{12} \sim 3.5$. De verdeling $P_{12}(\nu)$ en de cumulatieve verdeling $Q_{P_{12}}(\nu)$ zijn weergegeven in Figuur 4.3. We kunnen nu bijvoorbeeld de kans berekenen op exact 12 getelde pulsen ($P_{12}(12) = 0.11$) of nul pulsen ($P_{12}(0) = 6 \times 10^{-6}$). Met behulp van de cumulatieve verdeling kun je nu bijvoorbeeld de kans op minder dan 14 pulsen ($Q_{P_{12}}(13) = 0.68$) en de kans op 14 of meer pulsen (Prob$(\nu \geq 14) = 1 - 0.68 = 0.32$) berekenen.
Tot slot: er zijn telsituaties die nadrukkelijk niet aan de voorwaarden voor Poissonstatistiek voldoen. Een voorbeeld is het tellen van het aantal vlakken van een dobbelsteen; hier ontbreekt namelijk het kanselement. Ook zijn er situaties aan te geven waar duidelijk wél sprake is van een kanselement, maar waarbij het de vraag is in hoeverre aan de voorwaarden voor Poissonstatistiek voldaan is. Als voorbeeld: iemand telt elke



Figuur 4.3: (a) Poissonverdeling $P_\mu(\nu)$, met $\mu = 12$. (b) Cumulatieve Poissonverdeling $Q_{P,\mu}(\nu)$ voor dezelfde waarde van μ.

dinsdag het aantal auto's dat tussen 15.00 en 18.00 uur gebruik maakt van de Leuvenlaan op de Uithof. Het is twijfelachtig of hier sprake is van een Poissonverdeling voor het getelde aantal, aangezien er sprake is van beperkte reproduceerbaarheid van het gedrag van de *rate* van telling tot telling (de verschillende tellingen zijn dus hoogstwaarschijnlijk niet op te vatten als "trekkingen" uit identieke Poissonverdelingen).

4.5 Continue verdelingen

We beschouwen nu een (statistische) variabele x die *continu* verdeeld is. Hoewel in de praktijk een veelvoud aan continue verdelingen waargenomen wordt (o.a. exponentieel bij vervalsprocessen, Lorentziaan bij atomaire lijnvormen), concentreren we ons op de *normale verdeling* (ook wel: *Gauss-verdeling*).

4.5.1 De Gauss- of normale verdeling

Door toevallige oorzaken zullen gemeten waarden $\{x_1, x_2, \cdots, x_n\}$ op een continue manier spreiden rond de "echte" waarde X met een standaarddeviatie σ. In veel meetsituaties waarin *random* fluctuaties (ruis) optreden, zullen we te maken hebben met de zogenaamde *normale verdeling* $G_{X,\sigma}(x)$:

$$G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-X)^2}{2\sigma^2}}. \tag{4.34}$$

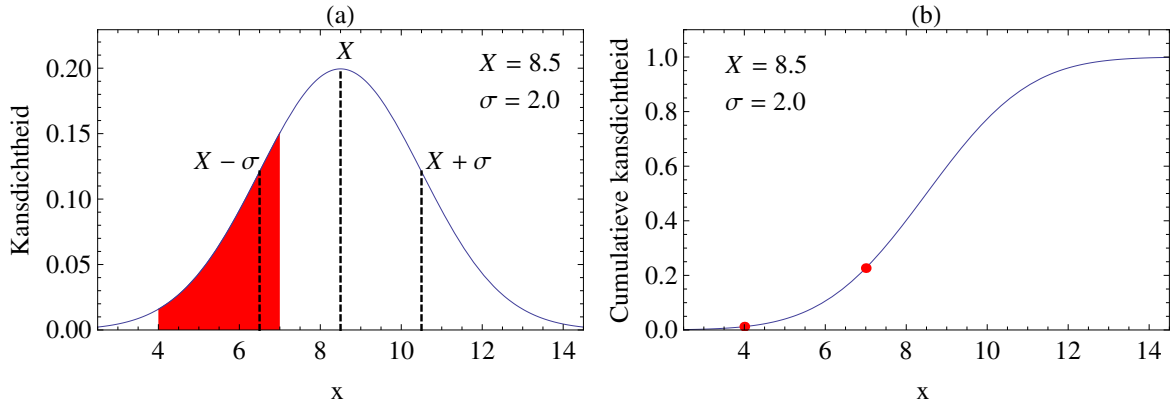
De voorfactor $\frac{1}{\sigma\sqrt{2\pi}}$ is nodig om de goede normalisatie te realiseren (Vgl. (4.6)):

$$\int_{-\infty}^{\infty} G_{X,\sigma}(x)\mathrm{d}x = 1. \tag{4.35}$$

We merken hier nogmaals op dat $G_{X,\sigma}(x)$ een *kansdichtheid* is. De *kans* $\mathrm{Prob}_G[x_1 \leq x \leq x_2]$ op een meetwaarde is gedefinieerd voor een interval $[x_1, x_2]$:

$$\mathrm{Prob}_G[x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} G_{X,\sigma}(x)\mathrm{d}x. \tag{4.36}$$

Wellicht zijn een aantal vuistregels voor kansen bij de normale verdeling al bekend. De kans dat een waarde in het gebied $[X-\sigma, X+\sigma]$ ligt is ongeveer $P_G[X-\sigma, X+\sigma] = 68\%$. De kans dat een waarde binnen



Figuur 4.4: (a) Normale verdeling $Q_{G_{X,\sigma}}(x)$, met $X = 8.5$ en $\sigma = 2$. De maximale hoogte $1/\sigma\sqrt{2\pi}$ wordt bereikt bij waarde X. In de figuur is de kans in de range $4 \leq x \leq 7$ gearceerd aangegeven. De kans om een waarde in dit interval te vinden kan berekend worden met behulp van Vgl. (4.36) en is gelijk aan 0.214. (b) Cumulatieve normale verdeling $Q_{G,X,\sigma}(x)$ voor dezelfde waarden van X en σ. De kans die voor de linker figuur is berekend kan ook met behulp van Vgl. (4.41) worden berekend als $Q_{G_{8.5,2}}(7) - Q_{G_{8.5,2}}(4)$. Beide punten zijn aangegeven als stippen.

$[X-2\sigma, X+2\sigma]$ ligt is ongeveer 95%. De exacte kansen kunnen worden berekend middels numerieke integratie (bijvoorbeeld in *Mathematica*) of via een tabel zoals in Appendix B.
Met behulp van Vgl. (4.7) en (4.8) bepalen wij karakteristieke eigenschappen van de normale verdeling:

$$\begin{aligned} \text{de } \textit{verwachtingswaarde } X_x &= \int_{-\infty}^{\infty} x G_{X,\sigma}(x)\mathrm{d}x \\ &= X, \qquad (4.37) \\ \text{de } \textit{variantie } \sigma_x^2 &= \int_{-\infty}^{\infty} (x - X_x)^2 G_{X,\sigma}(x)\mathrm{d}x \\ &= \sigma^2. \qquad (4.38) \end{aligned}$$

Voor een meetserie $\{x_1, x_2, \cdots, x_n\}$ geldt in de statistische limiet $n \to \infty$:

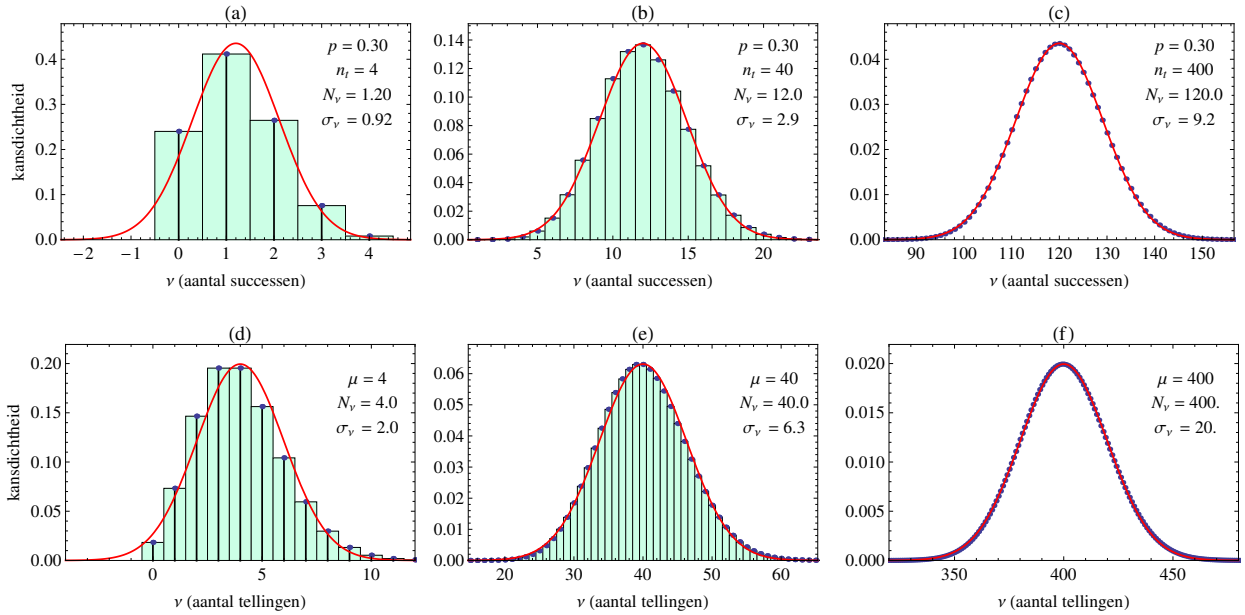
$$\begin{aligned} \overline{x} &\Rightarrow X, \\ \hat{\sigma}_x &\Rightarrow \sigma, \qquad (4.39) \\ \hat{\sigma}_{\overline{x}} &\Rightarrow \sigma/\sqrt{n} = 0. \end{aligned}$$

Als bij de Poissonverdeling kunnen we ook een uitspraak doen over de som van twee Gaussisch verdeelde grootheden. Wanneer bekend is dat de verdelingen van deze grootheden gemiddelden hebben van X_1 respectievelijk X_2, en standaarddeviaties van σ_1 respectievelijk σ_2, is te berekenen dat de *som* van deze grootheden ook Gaussisch verdeeld is:

$$G_{X_1,\sigma_1}(x) + G_{X_2,\sigma_2}(x) = G_{X,\sigma}(x), \qquad (4.40)$$

met gemiddelde $X = X_1 + X_2$ en $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Voor een *verschil* geldt eveneens dat dit Gaussisch verdeeld is, maar dan met $X = X_1 - X_2$ en $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.
De cumulatieve verdelingsfunctie $Q_{G_{X,\sigma}}(x)$ is gedefinieerd als de integraaluitdrukking

$$Q_{G_{X,\sigma}}(x) = \int_{-\infty}^{x} G_{X,\sigma}(x')\mathrm{d}x'. \qquad (4.41)$$

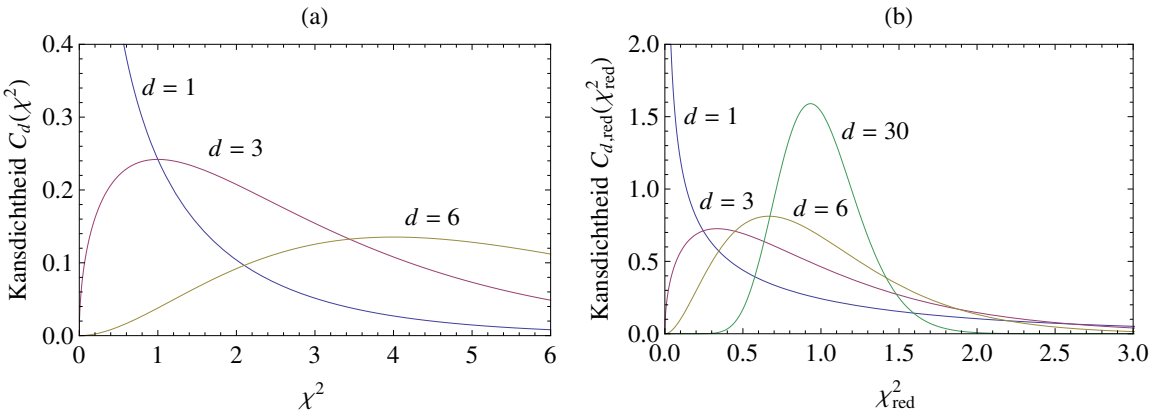


Figuur 4.5: Tendens van binomiaal- en Poissonverdeling naar de normale verdeling voor grote aantallen. (a) - (c): barhistogrammen van binomiaalverdeling voor $p = 0.5$ en respectievelijk $n = 4, 40, 400$, en normale verdeling (lijn) bij gegeven gemiddelde $X = np$ en variantie $\sigma^2 = np(1-p)$. (d) - (f) barhistogrammen van Poissonverdeling voor respectievelijk $\mu = 4, 40, 400$, en normale verdeling (lijn) bij gegeven gemiddelde $X = \mu$ en variantie $\sigma^2 = \mu$. Omdat de discrete verdelingen vergeleken worden met de continue normale verdeling, zijn de discrete verdelingen weergegeven als binhistogram. In (c) en (f) zijn de *bins* weggelaten vanuit het oogpunt van overzichtelijkheid.

Als voorbeeld is in Figuur 4.4 een normale verdeling getoond voor $X = 8.5$ en $\sigma = 2$. Op basis van vergelijking (4.36) bepalen wij de kans op een waarneming in het gebied $4 \leq x \leq 7$ gelijk aan 0.214. De kans op een waarde tussen 4.5 en 12.5 (oftewel $[X - 2\sigma, X + 2\sigma]$) is gelijk aan 0.955, in overeenstemming met eerdergenoemde vuistregel.
De verdelingsfunctie $G_{X,\sigma}(x)$ staat in principe meetwaarden x_i toe in het volledige bereik $(-\infty, \infty)$ - iets dat in een echte meting niet zo handig is vanwege het eindige meetbereik van een meetinstrument. Toch is de normale verdeling in veel meetsituaties bruikbaar. Het meetbereik wordt dan vaak zo gekozen dat X er voldoende centraal in ligt en de waarde σ ten opzichte van het totale meetbereik zeer klein is. Anders gezegd, hoewel er in theorie meetwaarden buiten het toegankelijke meetbereik kunnen vallen, wordt de kans op zo'n waarde in de praktijk verwaarloosbaar gemaakt.

4.5.2 De normale verdeling als limiet van de binomiale en Poissonverdeling

Naarmate het aantal mogelijke uitkomsten ν bij de binomiale- en Poissonverdeling toeneemt (dat wil zeggen, respectievelijk bij een groter aantal trekkingen n_t of een groter aantal tellingen), wordt het discrete karakter van de verdeling minder herkenbaar. Bovendien wordt de kans op nul tellingen verwaarloosbaar klein. Het blijkt dat beide verdelingen voor voldoende grote aantallen tellingen tenderen naar de normale verdeling. Het bewijs hiervoor wordt gegeven door de zogenaamde *centrale limietstelling*, die we hier niet zullen behandelen.
Voor de binomiale verdeling met aantal trekkingen n_t en succeskans p worden gemiddelde en variantie berekend middels Vgl. (4.23) - (4.24). Bij $p = 0.3$ en n_t respectievelijk 4, 40 en 400 resulteert dit in waarden $N_\nu \pm \sigma_\nu$ van 1.20 ± 0.92, 12.0 ± 2.9 en 120.0 ± 9.2. We plotten de binomiale verdelingen voor deze waarden van n_t en p, en de normale verdelingen met bijbehorende waarden van N_ν en σ_ν in Figuur 4.5



Figuur 4.6: (a) Chi-kwadraatverdeling $C(\chi^2,d)$ als functie van χ^2, voor waarden $d = 1$, $d = 3$ en $d = 6$. (b) Gereduceerde chi-kwadraatverdeling $C_{red}(\chi^2_{red},d)$ voor waarden $d = 1$, $d = 3$, $d = 6$ en $d = 30$.

(a) - (c). Het is overduidelijk zichtbaar dat de binomiale verdeling snel nadert naar de normale verdeling.

Eenzelfde analyse kunnen wij doen voor de Poissonverdeling. Met behulp van (4.30) en (4.31) berekenen we bij $\mu = 4, 40, 400$ onzekerheden σ_ν van respectievelijk 2.0, 6.3, en 20. De Poissonverdeling en bijbehorende normale verdelingen zijn weergegeven in Figuur 4.5 (d) - (f). Opnieuw is te zien dat het verschil tussen de Poissonverdeling en de normale verdeling in de limiet verwaarloosbaar wordt.

Het feit dat de discrete binomiale- en Poissonverdeling voor (zeer) grote aantallen mogelijke uitkomsten tenderen naar de normale verdeling zorgt ervoor dat veel experimenten die in fundamenteel opzicht telexperimenten zijn, zoals bijvoorbeeld een aantal elektronen (stroom) of fotonen (lichtintensiteit), bij macroscopische hoeveelheden (bijvoorbeeld $\gg 10^4$) niet alleen als nagenoeg continu mogen worden verondersteld, maar ook als normaal verdeeld. Veel uitwerkingen in de volgende hoofdstukken zijn gebaseerd op de eigenschappen van de normale verdeling, zoals symmetrie en continuïteit.

4.5.3 De chi-kwadraatverdeling

In Hoofdstuk 6 komt de *gereduceerde chi-kwadraat* aan de orde, in het kader van het vergelijken van een dataset met een model. Met het oog hierop introduceren we hier reeds de chi-kwadraatverdeling. Deze verdeling $C_d(\chi^2)$ wordt gegeven door

$$C_d(\chi^2) = \frac{1}{2^{d/2}\Gamma(d/2)} \left(\chi^2\right)^{d/2-1} e^{-\frac{\chi^2}{2}} \quad (\chi^2 \geq 0). \tag{4.42}$$

met d het aantal vrijheidsgraden en $\Gamma(d/2)$ de Gamma-functie. Je hoeft deze verdeling niet in detail te kennen; het gaat in dit college vooral om haar karakteristieke eigenschappen. *Mathematica* kent een commando om deze verdeling aan te roepen.

In Figuur 4.6 (a) is deze verdeling getoond voor een aantal waarden van d. Je ziet dat de verdeling voor zeer lage waarden van d bijzonder asymmetrisch is, wat te verwachten is doordat χ^2 alleen positieve waarden kan hebben. Voor grotere waarden van d wordt de verdeling meer symmetrisch. Je ziet verder dat de piek van de verdelingsfunctie $C_d(\chi^2)$ voor hogere waarden van d de waarde d benadert.

Ook voor deze verdeling zijn karakteristieke eigenschappen te berekenen met behulp van Vgl. (4.7) en (4.8):

$$\begin{aligned} \text{de } \textit{verwachtingswaarde } X_{\chi^2} &= \int_0^\infty \chi^2 C_d(\chi^2) \mathrm{d}\left(\chi^2\right) \\ &= d, \end{aligned} \tag{4.43}$$

$$\begin{aligned} \text{de } \textit{variantie } \sigma^2_{\chi^2} &= \int_0^\infty (\chi^2 - X_{\chi^2})^2 C_d(\chi^2) \mathrm{d}\left(\chi^2\right) \\ &= 2d. \end{aligned} \tag{4.44}$$

Interessant is ook om te kijken naar de gereduceerde-chi-kwadraatverdeling $C_{d,red}(\chi^2_{red})$. In Hoofdstuk 6 definiëren we $\chi^2_{red} = \frac{\chi^2}{d}$. Hiermee is de gereduceerde-chi-kwadraatverdeling via transformatie en renormalisatie af te leiden uit $C_d(\chi^2)$. Zie hiervoor Appendix C. In Figuur 4.6 (b) tonen we $C_{d,red}(\chi^2_{red})$ voor de d-waarden die ook voor (a) zijn gebruikt, maar ook voor een zeer hoge waarde $d = 30$. Deze figuur laat nog duidelijker zien dat bij toenemende d de verdeling van de gereduceerde chi-kwadraat steeds smaller (de standaarddeviatie is $\sqrt{\frac{2}{d}}$) en meer symmetrisch wordt. In de limiet gaat ook deze verdeling naar een normale verdeling! De piek van $C_{d,red}(\chi^2_{red})$ komt steeds dichter bij de verwachtingswaarde 1 te liggen.

4.6 Analyse van strijdigheid: significantieniveau en overschrijdingskans

Wanneer een meting is uitgevoerd en voor een grootheid een beste schatter inclusief onzekerheid is bepaald, moet regelmatig de waarde worden vergeleken met een "verwachte" waarde. Dit kan een waarde zijn uit de literatuur (tabellenboek of artikel), een bepaling uit eenzelfde experiment dat eerder is uitgevoerd, of een bepaling van dezelfde grootheid met een andere meetmethode. Hierbij kan ook de bepaling waarmee vergeleken wordt een aanzienlijke onzekerheid hebben.
Omdat (meet)resultaten een verdeling kennen, is nooit met volledige zekerheid vast te stellen dat het ene resultaat "gelijk is" aan het andere. Wel is de *kans* vast te stellen dat twee resultaten niet in overeenstemming met elkaar zijn. In dit kader wordt dan vaak gezegd dat twee resultaten *significant verschillen*. Een kwantitatieve maat voor de waarschijnlijkheid van een significant verschil is de *overschrijdingskans*. We bespreken een *stappenplan* om te komen tot een overschrijdingskans en een conclusie aan de hand van een concreet voorbeeld voor continu verdeelde grootheden.

Voorbeeld: bepaling van de constante van Planck Met behulp van het foto-elektrisch effect kan een bepaling worden gedaan van de constante van Planck h.[3] Eén student (1) rapporteert op basis van een experiment $\overline{h}_1 \pm \hat{\sigma}_{\overline{h},1} = (6.93 \pm 0.27) \times 10^{-34}$ J s. Een andere student (2) vindt op een andere opstelling de waarde $\overline{h}_2 \pm \hat{\sigma}_{\overline{h},2} = (6.02 \pm 0.18) \times 10^{-34}$ J s. In een tabellenboek wordt vervolgens de geaccepteerde waarde $h_0 = 6.6260693(11) \times 10^{-34}$ J s gevonden.
De vraag is nu of de door studenten 1 en 2 gevonden waarden strijdig zijn met de literatuurwaarde h_0.

4.6.1 Stappenplan voor de kwantitatieve analyse van strijdigheid

Een eerste beeld kan verkregen worden door de waarden te plotten inclusief onzekerheden, zie Figuur 4.7 (a). Op basis hiervan zouden we kunnen zeggen dat het resultaat van student 1 "redelijk" in overeenstemming is met h_0, en het resultaat van student 2 niet. Het is echter noodzakelijk om onze analyse te onderbouwen met *objectieve argumenten*, met andere woorden er is een kwantitatieve maat nodig om strijdigheid te duiden.

[3] Het experiment PLNK kan in het natuurkundepracticum worden gedaan.

Hiertoe onderscheiden we vier stappen; het opstellen van de nulhypothese, het vastleggen van het significantieniveau, het berekenen van de overschrijdingskans, en de verantwoorde conclusie over het wel of niet optreden van strijdigheid.

1. Opstellen nulhypothese De eerste stap is het opstellen van de zogenaamde *nulhypothese*. We volgen onderstaande stappen bij het formuleren van deze hypothese (de uitwerking is voor meting 1 maar geldt uiteraard ook voor meting 2):

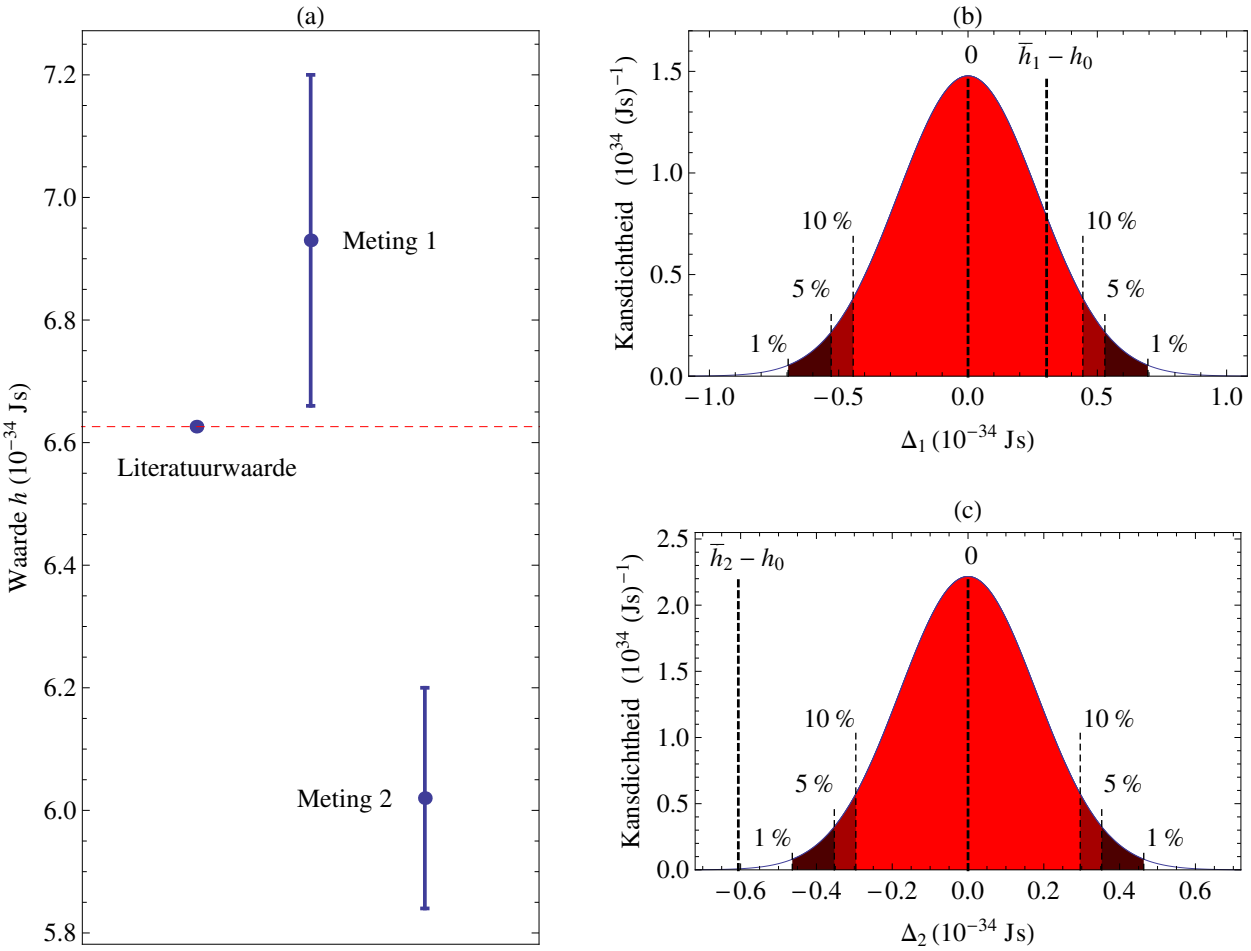
- We veronderstellen dat de referentiewaarde (de geaccepteerde waarde) normaal verdeeld is (Vgl. (4.4))als $G_{h_0,\sigma_{h_0}}(h)$. De referentiewaarde heeft dan een verwachtingswaarde h_0 en een standaarddeviatie σ_{h_0} (die in dit geval zeer klein is).
- We nemen ook aan dat de hypothetische verdelingsfunctie behorend bij meting 1 een normale verdeling $G_{h_1=h_0,\hat{\sigma}_{\overline{h}_1}}(h)$ is. De verdelingsfunctie van de meting heeft een verwachtingswaarde gelijk aan de referentiewaarde h_0 en een standaarddeviatie $\hat{\sigma}_{\overline{h}_1}$ zoals bepaald voor de meting.
- Met behulp van Vgl. (4.36) kunnen we de (hypothetische) verdelingsfunctie van het *verschil* $G_{h_1=h_0,\hat{\sigma}_{\overline{h}_1}}(h) - G_{h_0,\sigma_{h_0}}(h)$ berekenen. Deze verdelingsfunctie is volgens Vgl. (4.40) opnieuw een normale verdeling $G_{\Delta_1,\sigma_{\Delta_1}}(\Delta)$, met verwachtingswaarde $\Delta_1 = h_0 - h_0 = 0$, en standaarddeviatie $\sigma_{\Delta_1} = \sqrt{\sigma_{h_0}^2 + \hat{\sigma}_{\overline{h}_1}^2}$. In dit specifieke geval is de onzekerheid in de literatuurwaarde σ_{h_0} veel kleiner dan de onzekerheid in het meetresultaat $\hat{\sigma}_{\overline{h}_1}$, zodat $\sigma_{\Delta_1} \simeq \hat{\sigma}_{\overline{h}_1}$.

We veronderstellen derhalve dat er géén verschil is tussen de verwachtingswaarde van onze meting en die van de referentiewaarde, vandaar de term nulhypothese. De nulhypothese voor meting 1 (en idem 2) luidt nu als volgt:

- De verdelingsfunctie behorend bij het verschil tussen de verdelingsfunctie behorend bij meting 1 (2) en de verdelingsfunctie behorend bij de referentie is $G_{0,\sigma_{\Delta_1}}(\Delta) \simeq G_{0,\hat{\sigma}_{\overline{h}_1}}(\Delta)$ (en voor meting 2: $G_{0,\sigma_{\Delta_2}}(\Delta) \simeq G_{0,\hat{\sigma}_{\overline{h}_2}}(\Delta)$).

De verschil-verdelingsfuncties $G_{0,\hat{\sigma}_{\overline{h}_1}}(\Delta)$ en $G_{0,\hat{\sigma}_{\overline{h}_2}}(\Delta)$ zijn getoond in respectievelijk Figuur 4.7 (b) en (c). Het daadwerkelijk gevonden verschil tussen de meetwaarden en de verwachtingswaarde van de referentieverdeling, $\overline{h}_1 - h_0$ respectievelijk $\overline{h}_2 - h_0$ (eveneens weergegeven in Figuur 4.7 (b) en (c)), is nu op te vatten als een *trekking* uit deze hypothetische verdelingsfuncties. De *discrepanties* $\overline{h}_1 - h_0$ en $\overline{h}_2 - h_0$ ten opzichte van de verwachtingswaarde nul, zijn onder onze aannames ten gevolge van alléén toevallige variaties.

2. Vastleggen significantieniveau De volgende stap is dat we gegeven de hypothetische verdeling $G_{0,\hat{\sigma}_{\overline{h}_1}}(\Delta)$ een grenswaarde voor de afstand $\overline{h}_1 - h_0$ vastleggen waarboven we het verschil als significant benoemen (redeneringen verlopen uiteraard analoog voor meting 2).
Gegeven de verdeling zal een fractie α' van een grote set metingen binnen een bepaalde bandbreedte rond nul liggen. Voor meting 1 geldt bijvoorbeeld dat $\alpha' \simeq 0.95$ binnen de bandbreedte $[-2\hat{\sigma}_{\overline{h}_1}, +2\hat{\sigma}_{\overline{h}_1}]$. We stellen dat een meetresultaat in overeenstemming met de nulhypothese als onwaarschijnlijk beschouwd kan worden als de meting buiten een bepaalde bandbreedte ligt.
Hiertoe voeren we het *significantieniveau* α in, waarbij geldt $\alpha = 1 - \alpha'$. In Figuur 4.7 hebben we de bandbreedten geplot voor $\alpha = 0.1$, $\alpha = 0.05$ en $\alpha = 0.01$. Afhankelijk van de keuze voor α moeten we dus verschillende conclusies trekken over de betrouwbaarheid van de bepaling. Veel gebruikt is de zogenaamde 2σ-norm, waarvoor $\alpha \approx 0.05$. In het voorbeeld met de constante van Planck leggen we het dubbelzijdig (zie punt 3.) significantieniveau op $\alpha = 0.05$.



Figuur 4.7: (a) Een plot van de literatuurwaarde voor de constante van Planck en de meetwaarden van beide studenten 1 en 2.
(b) Grafische weergave van het resultaat $\overline{h}_1 - h_0$ van student 1 ten opzichte van de hypothetische verdeling van het verschil tussen meting en referentie $G_{0,\hat{\sigma}_{\overline{h}_1}}(\Delta)$. De grenzen van het gebied waarbinnen 90%, 95% en 99% van de resultaten worden verwacht (gearceerde gebieden) zijn met stippellijnen aangegeven. Gegeven de hypothetische verdeling $G_{0,\hat{\sigma}_{\overline{h}_1}}(\Delta)$ ligt het verschil $\overline{h}_1 - h_0$ binnen het gebied waarbinnen 90% van de waarden wordt verwacht. Er is op basis van de gangbare significantieniveaus dus geen reden om de nulhypothese te verwerpen.
(c) Resultaat $\overline{h}_2 - h_0$ van student 2 grafisch weergegeven, analoog aan (b). Dit resultaat ligt ver in de flanken van de hypothetische verdeling van het verschil tussen meting en referentie $G_{0,\hat{\sigma}_{\overline{h}_2}}(\Delta)$. De nulhypothese en het gevonden resultaat zijn derhalve niet in overeenstemming.

In veel medisch onderzoek moet tegenwoordig in het onderzoeksprotocol vóór het uitvoeren van de metingen het gehanteerde significantieniveau al worden vastgelegd. Het vastleggen van het significantieniveau is dan stap 1. in het stappenplan. Dit om te voorkomen dat na afloop van het onderzoek het significantieniveau wordt aangepast om zodoende een significant resultaat te kunnen presenteren. Dit is een vorm van wetenschappelijke fraude.

3. Overschrijdingskans Omdat de verdelingsfunctie bekend is, kunnen we met behulp van Vgl. (4.36) of (4.41) de kans berekenen dat de "getrokken" waarde $\overline{h}_1 - h_0$ op de huidige afstand of zelfs verder van de meest waarschijnlijke waarde nul ligt.[4] Dit noemen we de *overschrijdingskans* P (of *P-value*). Deze overschrijdingskans kan enkelzijdig (P_e) of dubbelzijdig (P_d) gedefinieerd worden.
Voor een algemene Gaussische verdeling $G_{X,\sigma}(x)$ en een gemeten waarde x_m worden enkelzijdige overschrijdingskansen berekend als

$$P_e = \int_{-\infty}^{x_m} G_{X,\sigma}(x)\mathrm{d}x = Q_{G_{X,\sigma}}(x_m), \qquad \text{(linkszijdig, } x_m < X\text{)} \tag{4.45}$$

$$P_e = \int_{x_m}^{\infty} G_{X,\sigma}(x)\mathrm{d}x = 1 - Q_{G_{X,\sigma}}(x_m). \quad \text{(rechtszijdig, } x_m > X\text{)} \tag{4.46}$$

We maken hier onderscheid tussen het geval dat $x_m < X$ (men berekent dan de *linkszijdige* overschrijdingskans) en $x_m > X$ (de *rechtszijdige* overschrijdingskans).
Op statistische gronden kunnen we stellen dat de (absolute) afwijking $|\overline{h}_1 - h_0|$ relevant is en niet of $\overline{h}_1$ nu groter of kleiner is dan h_0, is de dubbelzijdige overschrijdingskans de gangbaar gebruikte maat. Het significantieniveau heeft bij de analyse van strijdigheid dan ook bijna altijd betrekking op de dubbelzijdige overschrijdingskans.
De dubbelzijdige overschrijdingskans kan vanwege de symmetrie van de normale verdeling eenvoudig berekend worden als $P_d = 2P_e$, voor zowel $x_m < X$ als $x_m > X$. De dubbelzijdige kans P_d kan ook direct geformuleerd worden voor een x_m die willekeurig ligt ten opzichte van X als

$$P_d = 1 - \left| \int_{x_m}^{2X - x_m} G_{X,\sigma}(x)\mathrm{d}x \right| \tag{4.47}$$

$$= 1 - \left| Q_{G_{X,\sigma}}(2X - x_m) - Q_{G_{X,\sigma}}(x_m) \right|. \tag{4.48}$$

De bovenstaande integralen zijn eenvoudig te berekenen met behulp van *Mathematica*, zie hiervoor de werkboeken. De overschrijdingskans kan ook worden bepaald met behulp van tabellen die numerieke resultaten voor (een deel van) bovengenoemde integralen bevatten. In Appendix B kun je Tabel B.1 gebruiken om de overschrijdingskansen voor een normale verdeling op te zoeken. Hiervoor wordt eerst de geschaalde discrepantie $t = (x_m - X)/\sigma$ berekend (oftewel, het aantal standaarddeviaties dat waarde x_m van X vandaan ligt). Vervolgens kan met behulp van de tabel bepaald worden met welke (rechts-, links- of dubbelzijdige) overschrijdingskans deze waarde van t correspondeert.
Voor de twee waarden $\overline{h}_1 - h_0$ en $\overline{h}_2 - h_0$ berekenen we $t_1 = 1.15$ respectievelijk $t_2 = -3.33$. De dubbelzijdige overschrijdingskansen worden dan $P_{d,1} = 0.26$ en $P_{d,2} = 0.00076$ (deze laatste waarde staat niet in Tabel B.1 want deze loopt tot $t = 3.09$, je kunt daaruit alleen bepalen dat $P_{d,2} \leq 0.001$).

[4] Je zou misschien denken: bepaal de kans op (trekking van) waarde $\overline{h}_1 - h_0$, en als deze kans kleiner is dan een zeker percentage, concluderen we dat onze meting niet in overeenstemming is met de hypothese. Dit is geen bruikbare methode, omdat de kans op *exact* waarde $\overline{h}_1$ gelijk is aan nul. Voor discrete verdelingen zou dit in principe wel kunnen, maar hier hangt de kans op een meetwaarde af van het aantal mogelijke uitkomsten en/of de breedte van de verdeling (zie bijvoorbeeld Figuur 4.5). Door het bepalen van een verschil gelijk aan of groter dan de waarde $\overline{h}_1 - h_0$ wordt de methode hier ongevoelig voor.

4. Vergelijking en conclusie De laatste stap is dat de gevonden overschrijdingskans wordt vergeleken met het significantieniveau. Als $P > \alpha$ dan wordt het resultaat als niet strijdig met de nulhypothese beschouwd. Als $P < \alpha$, dan ligt de gemeten waarde *buiten* het gebied waarvoor (gegeven de nulhypothese) geldt dat het resultaat als statistisch acceptabel wordt beschouwd. Er is dan een op het gestelde α-niveau een significante afwijking, en de nulhypothese wordt weerlegd. We gaan er derhalve van uit dat er een conflict bestaat tussen beide bepalingen. Dit duidt op systematische (niet-statistische) afwijkingen en geeft mogelijk aanleiding tot een nadere analyse van één en mogelijk beide bepalingen.
Bij de bepaling van van de constante van Planck voor student 1 geldt dat $P_d = 0.26 > \alpha = 0.05$. We concluderen nu dat het verschil $\overline{h}_1 - h_0$ *niet* strijdig is met de (hypothetische) verdeling $G_{0,\hat{\sigma}_{\overline{h}_1}}$ (de nulhypothese). Anders gezegd is de meting $\overline{h}_1 \pm \hat{\sigma}_{\overline{h}_1}$ niet strijdig met de literatuurwaarde h_0.
Bij de meting van student 2 echter is $P_d = 0.00076 < \alpha = 0.05$. Het meetresultaat is zoals dat heet "strijdig op het niveau $\alpha = 0.05$" met de literatuurwaarde, en zelfs strijdig op $\alpha = 0.01$. Het is waarschijnlijk dat de opstelling een systematische afwijking vertoont, of dat de student een fout heeft gemaakt in (de verwerking van) zijn meting. De mogelijke conclusie dat de waarde van de constante van Planck afwijkt van h_0 is hier wat overmoedig, gezien de nauwkeurigheid en de betrouwbaarheid van de literatuurwaarde h_0. Wanneer je zelf serieus onderzoek gaat doen, kun je deze mogelijkheid echter niet zomaar uitsluiten!

Hoofdstuk 5

Propagatie van onzekerheden

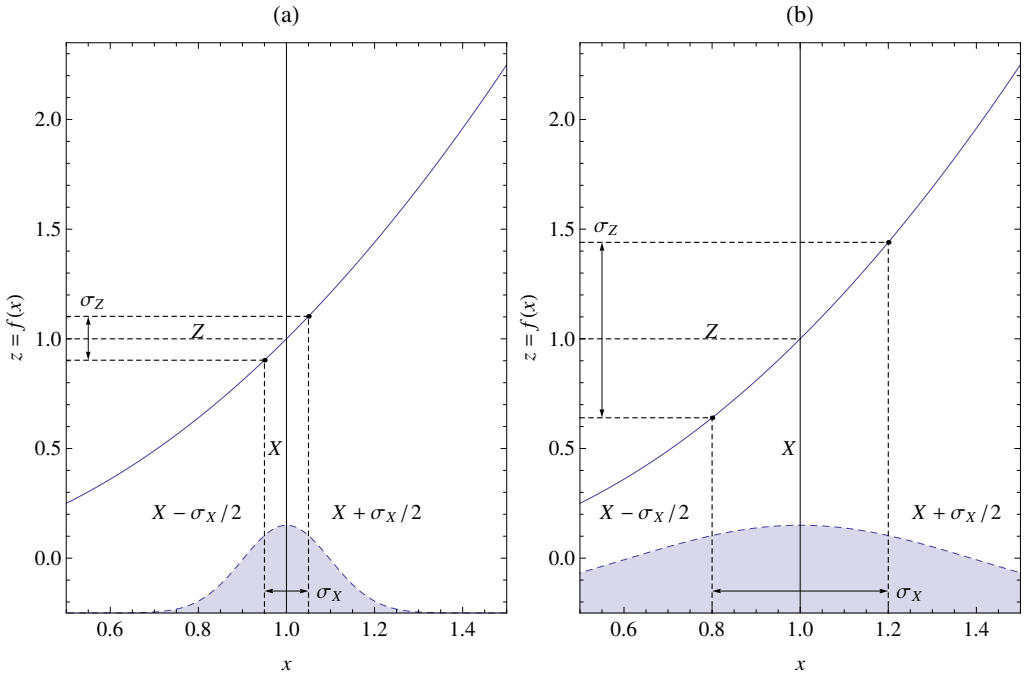
5.1 Inleiding

De "beste schatters" van grootheden In Hoofdstukken 3 en 4 zijn we voor enkele metingen, datasets en verdelingsfuncties diverse "beste schatters" tegengekomen. Het hangt van de situatie af wat als "beste schatter" moet worden beschouwd. Een samenvatting in gebruikte terminologie in eerdere hoofdstukken is gegeven in Tabel 5.1. De reden voor dit overzicht is dat het voor de theorie in dit hoofdstuk niet uitmaakt hoe de beste schatters tot stand zijn gekomen. Het enige dat we aannemen, is dat de beste schatters bekend zijn en dat de scherpst bepaalde onzekerheden gebruikt worden in berekeningen. Vanwege *notationele* redenen duiden we in het vervolg van dit hoofdstuk de "beste schatters" van grootheden x, y, $\cdots$ aan met X, Y, $\cdots$, en de onzekerheid in deze "beste schatters" met σ_X, σ_Y, $\cdots$.

Propagatie van onzekerheid Voor het beantwoorden van een kwantitatieve onderzoeksvraag moeten vaak meetresultaten voor verschillende grootheden $x, y, \cdots$ worden gecombineerd. Uitgangspunt is dat er een *functie* $z = f(x, y, \cdots)$ is en dat grootheden x, y, $\cdots$ bepaald zijn met beste schatters $X \pm \sigma_X$, $Y \pm \sigma_Y$, $\cdots$ in een meetprocedure of anderszins. De vraag die we nu willen beantwoorden is hoe de beste schatter $Z \pm \sigma_Z$ voor grootheid z bepaald wordt. Het ligt voor de hand dat de beste schatter Z voor z bepaald kan worden met behulp van de beste schatters $X, Y, \cdots$ voor $x, y, \cdots$ volgens $Z = f(X, Y, \cdots)$. Maar: hoe bereken je nu de onzekerheid σ_Z in Z?
In dit hoofdstuk leggen we uit hoe deze *propagatie van onzekerheden* in zijn werk gaat. We starten met propagatie van onzekerheid voor functies van een enkele variabele ($z = f(x)$) middels twee op elkaar lijkende methodes, de *variatiemethode* en de *differentieermethode*. Hier wordt aangetoond dat het voor een goede schatting van de onzekerheid σ_Z in Z in de meeste gevallen (gelukkig!) niet nodig is om rekening te houden met de verdelingsfunctie van x. Het is voldoende om de karakteristieke eigenschappen (dat wil zeggen, gemiddelde/verwachtingswaarde en standaarddeviatie) van de verdeling van x te kennen om een goede schatting voor σ_Z te verkrijgen.
Vervolgens kijken we naar het combineren van meetresultaten $z = f(x, y, \cdots)$ voor meerdere grootheden

Tabel 5.1: Beste schatters zoals die gedefinieerd zijn in voorgaande hoofdstukken.

Situatie	Beste schatter	Onzekerheid
Enkele meting	x	$\hat{\sigma}_x$
Meetserie		
Ongewogen middeling	$\overline{x}$	$\hat{\sigma}_{\overline{x}}$
Verdelingsfunctie	X	σ_x



Figuur 5.1: Illustratie van het doorwerken van onzekerheid met de variatiemethode voor één variabele voor de functie $z = f(x) = x^2$. (a) In het punt $X = 1$ met $\sigma_X = 0.1$. (b) In het punt $X = 1$ met $\sigma_X = 0.4$. De variatiemethode is toegepast aan de hand van de symmetrische vorm Vgl. (5.2).

$x, y, \cdots$. Eerst worden voor *statistisch onafhankelijke* grootheden (dat wil zeggen dat behaalde resultaten voor de ene grootheid niet afhangen van de resultaten voor een andere) de variatiemethode en de differentieermethode geformuleerd. Daarna definiëren we een test voor *statistische afhankelijkheid* oftewel *correlatie* van grootheden, en bepalen we hoe de propagatiemethodes aangepast moeten worden voor een correcte bepaling van onzekerheden van afhankelijke grootheden.

Tot slot geven we een algemeen recept voor analyse van meetresultaten. De laatste stap in dit recept is de analyse van de onzekerheid in het eindantwoord voor grootheid z in termen van de *partiële onzekerheden*: de onzekerheidsbijdragen van de gemeten grootheden x, y, $\cdots$. De grootheid met de grootste partiële onzekerheid is de eerste die nauwkeuriger bepaald moet worden om de nauwkeurigheid van de beste schatter Z voor z te verhogen.

5.2 Propagatie van onzekerheid voor functionele verbanden $z = f(x)$

Stel dat een meetserie $\{x_1, \cdots, x_n\}$ is vergaard met gemiddelde X en standaarddeviatie in dit gemiddelde σ_X. We willen nu de "beste schatting" Z met onzekerheid σ_Z bepalen voor de grootheid z, die functioneel afhangt van x als $z = f(x)$. Een illustratie van deze situatie is gegeven in Figuur 5.1 voor $z = f(x) = x^2$. Hier is x bijvoorbeeld de lengte van de zijde van een vierkant, en z het oppervlak van dit vierkant.

5.2.1 De "beste schatter" Z voor grootheid z

Voor het bepalen van de beste schatter Z voor z wordt eenvoudig de functie in het punt X te geëvalueerd:

$$Z = f(X). \tag{5.1}$$

5.2.2 Propagatie van onzekerheid voor $z = f(x)$ middels de variatiemethode

Een goede eerste benadering voor de onzekerheid σ_Z in het berekende resultaat Z wordt verkregen door te kijken hoe een variatie in x over afstand σ_X rond meetresultaat X doorwerkt in z. Dit kan bijvoorbeeld door $f(X+\sigma_X/2)$ en $f(X-\sigma_X/2)$ te berekenen en waaruit als schatting van σ_Z volgt:

$$\sigma_Z \simeq |f(X+\sigma_X/2) - f(X-\sigma_X/2)|\,. \tag{5.2}$$

De absoluutstrepen zorgen ervoor dat σ_Z altijd positief is. Een bepaling van σ_Z via een dergelijke discrete variatie noemen we de *variatiemethode*. In Figuur 5.1 is voor één variabele x het principe weergegeven.

De waarde van σ_Z in de *symmetrische vorm* wordt berekend door evaluatie van de functie $f(x)$ als in Vgl. (5.2) of alternatief

$$\sigma_Z = \frac{1}{2}\,|f(X+\sigma_X) - f(X-\sigma_X)|\,. \tag{5.3}$$

Ook kan gebruik gemaakt worden van de *asymmetrische vormen*

$$\sigma_Z = |f(X+\sigma_X) - f(X)|\,, \tag{5.4}$$
$$\sigma_Z = |f(X) - f(X-\sigma_X)|\,. \tag{5.5}$$

De asymmetrische vormen vereisen één numerieke evaluatie minder ($Z = f(X)$ is namelijk al bekend) en besparen derhalve rekenwerk.

Eén van de aantrekkelijke aspecten van de variatiemethode is het gemak waarmee de evaluaties kunnen worden geprogrammeerd. De variatiemethode kan ook efficiënt worden toegepast wanneer het functionele verband tussen x en z moeilijk of niet analytisch uit te drukken is (bijvoorbeeld in numerieke simulaties). Tot slot kan de variatiemethode een bepaling opleveren die betrouwbaarder is dan de differentieermethode (zie volgende paragraaf) wanneer de functie $f(x)$ meer of minder niet-lineair gedrag vertoont op het domein $[X-\sigma_X, X+\sigma_X]$, zie bijvoorbeeld Figuur 5.3 (d).

Een aspect van de variatiemethode, waaruit blijkt dat een *schatter* van de onzekerheid σ_Z in grootheid z verkregen wordt, is dat afhankelijk van het functionele verband $z = f(x)$ een *asymmetrisch onzekerheidsinterval* bepaald wordt (σ_Z uit Vgl. (5.4) kan immers ongelijk zijn aan σ_Z uit Vgl. (5.5)). Het resultaat van de variatiemethode wordt gebruikelijk wel gerapporteerd in *symmetrische vorm* $Z \pm \sigma_Z$.

Numerieke propagatie van onzekerheid middels variatiemethode We beschouwen een slinger, waarbij een massa m is opgehangen aan een draad met lengte $l = 1.50$ m (zeer nauwkeurig bekend). De slinger krijgt een beginuitwijking $\phi_0 = 35^\circ \pm 5^\circ$ en wordt losgelaten op $t = 0$. De vraag is nu wat de uitwijking $\phi(t)$ is op $t = 12.5$ s. Voor de beantwoording van deze vraag verwaarlozen wij wrijvingseffecten.
De differentiaalvergelijking die de slingerbeweging beschrijft, is

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2} = -\frac{g}{l}\sin\phi, \tag{5.6}$$

met $g = 9.81$ m/s^2 de zwaartekrachtsconstante (bekend verondersteld met verwaarloosbare onzekerheid). In Figuur 2.3 was al te zien dat de slingertijd afhankelijk is van de hoekuitwijking. De reden hiervoor ligt in het feit dat Vgl. (5.6) ter rechterzijde de niet-lineaire sinusterm bevat. Voor kleine hoeken wordt $\sin\phi \simeq \phi$ en wordt de oplossing van deze vergelijking $\phi(t) = \phi_0 \cos 2\pi t/T$, met $T = 2\pi\sqrt{l/g}$. We kunnen echter op basis van Figuur 2.3 vaststellen dat $\phi_0 = 35^\circ$ geen kleine hoek is. Vgl. 5.6 is dan niet *exact* op te lossen, alleen *numeriek*.
We gebruiken *Mathematica* om de slingervergelijking numeriek op te lossen voor lengte $l = 1.50$ m en $\phi_0 = 35^\circ$. De gevonden $\phi(t)$ als functie van t is geplot in Figuur 5.2. Op basis hiervan bepalen we bij $t = 50$ s een waarde $\phi(50) = 21.79^\circ$. Voor het berekenen van de onzekerheid gebruiken we de variatiemethode Vgl. (5.2): we stellen de waarde voor ϕ_0 op respectievelijk $\phi_0 + 1/2\sigma_{\phi_0} = 37.5^\circ$ en $\phi_0 - 1/2\sigma_{\phi_0} = 32.5^\circ$, en lossen het probleem onder deze beginvoorwaarden opnieuw numeriek op.
De resulterende $\phi(t)$ zijn als gestippelde lijnen in Figuur 5.2 getoond. Je ziet dat een verschil in de beginuitwijking leidt tot een andere slingerperiode. De numerieke evaluatie levert voor 37.5° en 32.5° respectievelijk $\phi_+(12.5) = 26.55^\circ$ en $\phi_-(12.5) = 17.41^\circ$ (merk op dat er sprake is van asymmetrie). We berekenen derhalve $\sigma_{\phi(50)} = 9.13^\circ$. Tot slot rapporteren wij (rekening houdend met de significantie) voor de waarde $\phi(12.5) = 22 \pm 9^\circ$. Wanneer de differentieermethode op de lineaire benadering wordt losgelaten, resulteert dit overigens in $\sigma_\phi = 0.51^\circ$.

5.2.3 Propagatie van onzekerheid voor $z = f(x)$ middels de differentieermethode

Bij de variatiemethode wordt gebruik gemaakt van de variatie van f bij een *discrete* variatie van x over afstand $\pm\sigma_X$. We kunnen de resultaten bij $X \pm \sigma_X/2$ ook schrijven als een Taylorbenadering rond steunpunt X:

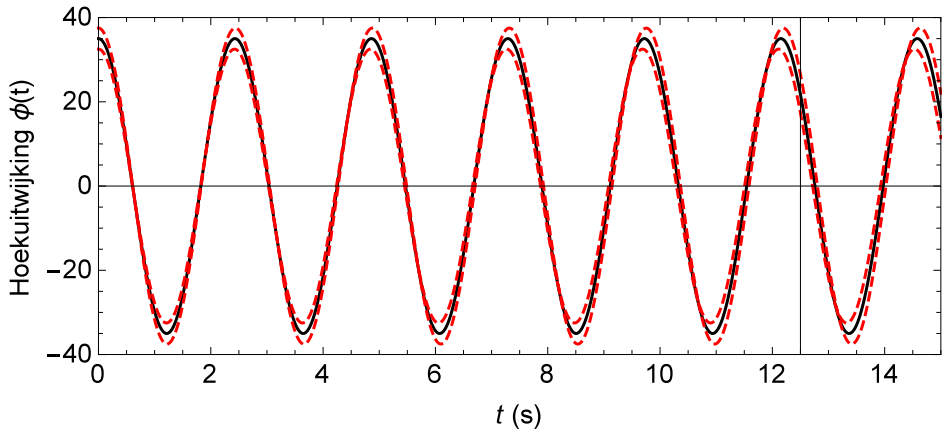
$$f(X + \sigma_X/2) = f(X) + \sigma_X/2 \left(\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right)_{x=X} + O\left((\sigma_X/2)^2 \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2}\right), \tag{5.7}$$

$$f(X - \sigma_X/2) = f(X) - \sigma_X/2 \left(\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right)_{x=X} + O\left((\sigma_X/2)^2 \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2}\right). \tag{5.8}$$

Wanneer nu het gedrag van de functie f op het gebied $[X - \sigma_X/2, X + \sigma_X/2]$ bij benadering *lineair* is, dan zijn de hogere-orde termen in Vgl. (5.7) en (5.8) verwaarloosbaar klein. Substitutie in Vgl. (5.2) levert dan

$$\sigma_Z \simeq \left|\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x=X} \sigma_X. \tag{5.9}$$

De discrete variatie is nu vervangen door een expliciete differentiatie. Bij deze zogenaamde *differentieermethode* wordt aangenomen dat een eerste-orde Taylorbenadering in het punt X het gedrag van $f(x)$ rond het



Figuur 5.2: Variatiemethode in een numeriek probleem. De doorgetrokken lijn is de numerieke oplossing van Vgl. (5.6) voor $\phi_0 = 35^\circ$. De stippellijnen zijn de numerieke oplossingen voor waarden $\phi_0 = 37.5^\circ$ $(= 35^\circ + \frac{1}{2}\sigma_{\phi_0})$ en $\phi_0 = 32.5^\circ$ $(= 35^\circ - \frac{1}{2}\sigma_{\phi_0})$. De onzekerheid in ϕ_0 vertaalt zich in een onzekerheid in hoekuitwijking ϕ op tijdstip $t = 12.5$ s (verticale lijn).

punt X adequaat beschrijft. Dit komt er op neer dat in het punt (X,Z) wordt uitgegaan van de *raaklijn* aan de kromme $z = f(x)$. De afgeleide (helling raaklijn) van de functie in het punt (X,Z) wordt vermenigvuldigd met σ_X om de grootte van de onzekerheid in z te bepalen. De afgeleide van f moet dan wel gedefinieerd zijn in het punt $x = X$! Samenvattend:

$$Z = f(X). \tag{5.10}$$

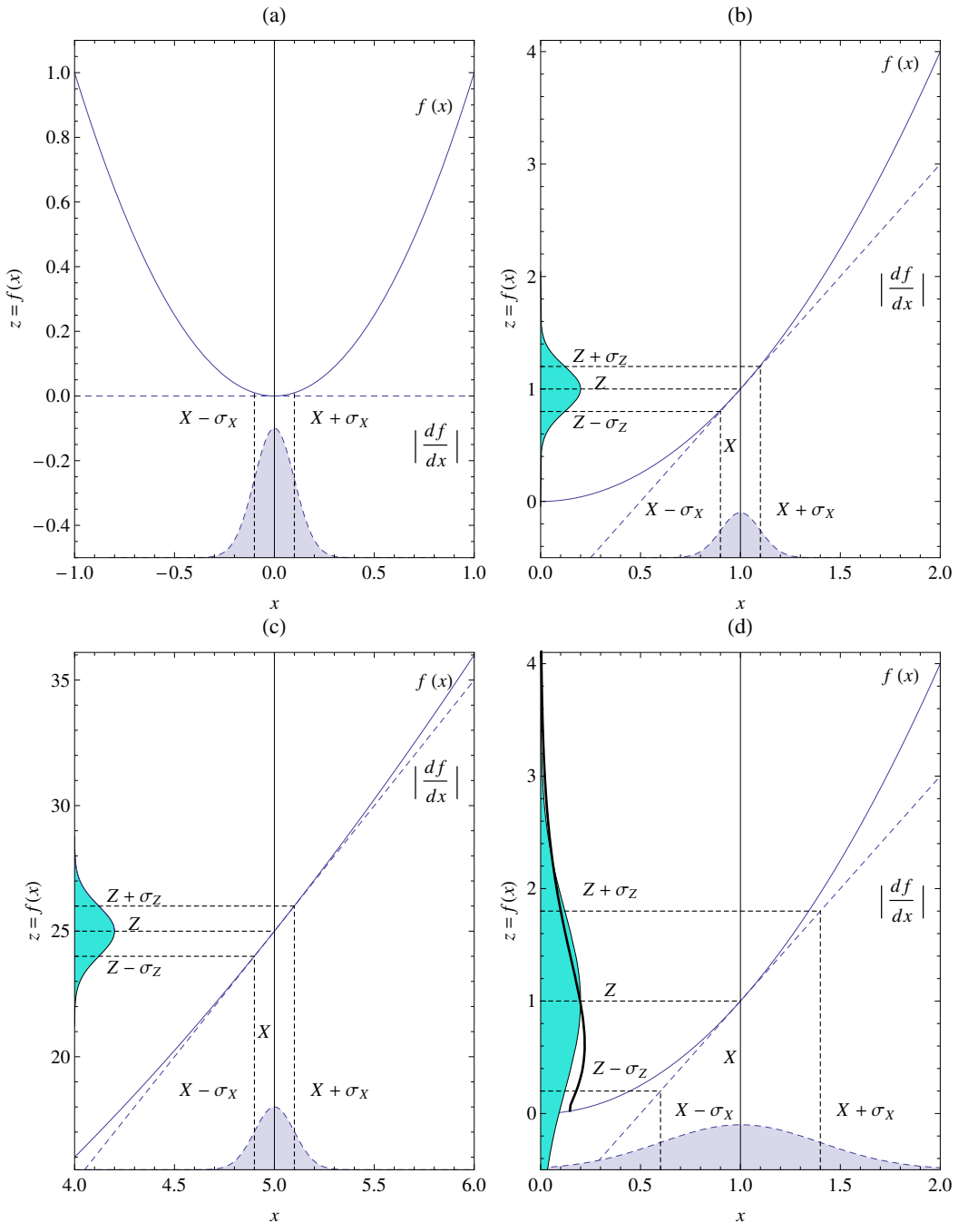
$$\sigma_Z = \left|\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x=X} \sigma_X. \tag{5.11}$$

In Figuur 5.3 (a) - (c) is de differentieermethode geïllustreerd voor $z = x^2$ met waarden $X = 0, 1, 5$ en $\sigma_X = 0.1$. Volgens de differentieermethode Vgl. (5.11) is de onzekerheid σ_Z dan $2X\sigma_X$. Voor de drie waarden van X in (a) - (c) levert dit respectievelijk $\sigma_Z = 0, \sigma_Z = 0.2, \sigma_Z = 1$ op. De waarden voor Z en σ_Z zijn in Figuur 5.3 met stippellijnen weergegeven. Naarmate de waarde van X groter wordt, werkt de onzekerheid in x sterker door in de onzekerheid van z. Voor $X = 0$ vinden we middels de differentieermethode dat een variatie in x niet leidt tot een onzekerheid in z omdat z in dit punt nauwelijks gevoelig is voor x. Overigens laat Figuur 5.3 (a) zien dat een variatie in x wel degelijk tot een (kleine) variatie in z leidt. De benadering door de differentieermethode Vgl. (5.9) verwaarloost deze variatie omdat dit een hogere-orde (niet-lineair) effect is.

We verkrijgen met de differentieermethode dus net als met de variatiemethode een *schatter* van de onzekerheid in grootheid z. Er wordt immers aangenomen dat de functie lineariseerbaar is rond X. Deze aanname is slechts toepasbaar als σ_X voldoende klein is en de functie lokaal gedomineerd wordt door lineair gedrag. Dit aspect komt aan bod in de volgende sectie.

Wanneer de functie f een significante kromming vertoont in het traject $X \pm \sigma_X$ is het niet te verwachten dat de differentieermethode en de variatiemethode *numeriek* identieke resultaten zullen geven. Bij toeval leveren beide methodes voor de voorbeeldfunctie $f(x) = x^2$ numeriek hetzelfde resultaat op, namelijk $\sigma_Z = 0.2$ en $\sigma_Z = 0.8$ voor de situaties in respectievelijk Figuur 5.1 (a) en (b).

De differentieermethode is een zeer efficiënte methode wanneer gebruik kan worden gemaakt van programmeeromgevingen die expliciete differentiatie toelaten, zoals *Mathematica*, en is door de analytische vorm



Figuur 5.3: Illustratie van het doorwerken van onzekerheid met de differentieermethode voor $z = f(x) = x^2$ in de punten: (a) $X = 0$, (b) $X = 1$ en (c) $X = 5$. In al deze gevallen geldt dat $\sigma_X = 0.1$ en de data voor x zijn normaal verdeeld. Voor kleine (relatieve) variaties σ_X is de verdeling van z bij benadering ook normaal verdeeld. (d) $X = 1$ en $\sigma_X = 0.4$. De differentieermethode geeft nu een minder betrouwbare schatting van σ_Z. De dikke zwarte lijn op de z-as is de werkelijke verdelingsfunctie in z op basis van de verdelingsfunctie voor x. Deze is sterk asymmetrisch en inderdaad afgekapt bij $x = 0$.

van de onzekerheid vaak erg inzichtelijk. Voor functies waarbij het differentiëren op problemen stuit moet alsnog worden uitgeweken naar de variatiemethode.

5.2.4 Toepasbaarheid van de propagatiemethodes

In het voorgaande hebben we de standaarddeviatie σ_X gebruikt als een maat voor de breedte van de verdeling van x. Zowel de variatiemethode Vgl. (5.2) (of alternatieven) als de differentieermethode Vgl. (5.11) impliceren dat de *verdeling* van z (bij benadering) gelijkvormig is aan de *verdeling* van x, behoudens een schalingsfactor die gelijk is aan de verhouding tussen σ_X en σ_Z. Voor de differentieermethode is deze schalingsfactor gelijk aan $\left|\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|$. Wanneer de verdeling van z gelijkvormig is aan de verdeling van x is de eerste betrouwbaar te karakteriseren met de bekende Z en σ_Z.
De aanname dat de verdelingsfunctie niet van vorm verandert, is alleen geldig wanneer σ_X zodanig klein is dat de functie $f(x)$ lineair mag worden verondersteld. Zie Figuur 5.3 (a) - (c). Hier hebben we een normaal verdeelde x aangenomen. Figuur 5.3 (d) illustreert wat er gebeurt als niet aan deze voorwaarde voldaan wordt, voor opnieuw $f(x) = x^2$ met $X = 1$ maar nu met $\sigma_X = 0.4$. Net als in Figuur 5.3 (b) blijkt de beste schatter Z voor z gelijk aan 1. De differentieermethode Vgl. (5.11) suggereert een onzekerheid $\sigma_Z = 0.8$. De functie $f(x)$ over het gebied van mogelijke waarden voor x wijkt sterk af van de lineaire benadering, met name voor waarden $x < 1$. De waarde $\sigma_Z = 0.8$ blijkt zelfs met aanzienlijke kans toe te staan dat negatieve waarden voor z worden gevonden, terwijl dit op basis van het werkelijke functionele verband feitelijk onmogelijk is.
De dikkere zwarte lijn die in Figuur 5.3 (d) langs de z-as is getrokken door de verdelingsfunctie (zoals gegeven door de differentieermethode) is de werkelijke verdelingsfunctie voor z. Deze is sterk vervormd ten opzichte van de normale verdeling in x. Met behulp van Vgl. (4.7) en (4.8) voor een arbitraire verdelingsfunctie berekenen we de "echte" waarden $Z_{werk} = 1.14$ en $\sigma_{Z,werk} = 0.82$. De differentieermethode met 1.0 ± 0.8 geeft dus een redelijke benadering van $Z \pm \sigma_Z$ (de variatiemethode geeft identieke resultaten), maar het mag duidelijk zijn dat het gebruik van de karakteristieken X, σ_X in plaats van de volledige verdeling voor x ervoor zorgt dat z onterecht als normaal verdeeld wordt verondersteld.
In veel meetsituaties is de onzekerheid σ_X gelukkig wel zodanig klein dat het gedrag van $f(x)$ in het relevante gebied lineair mag worden verondersteld, en geven de propagatiemethodes een betrouwbare benadering van Z en σ_Z. Houd echter in gedachten dat deze aanname met betrekking tot $f(x)$ vaak stilzwijgend wordt gedaan!

5.2.5 Rekenregels bij eenvoudige bewerkingen met een enkele statistische variabele

Voor eenvoudige functies $f(x)$ levert algebraïsche differentiatie een eenvoudig resultaat en volgen er standaard rekenregels voor propagatie van de onzekerheid σ_X voor een functie $f(x)$. Een aantal van deze rekenregels zijn weergegeven in Tabel 5.2. Bij eenvoudige bewerkingen van de soort $z = f(x)$ ligt het werken met deze rekenregels voor de hand omdat geen expliciete differentiatie van de functie f meer nodig is. Dit bespaart tijd en rekenwerk, en geeft inzicht!

5.3 Propagatie van onzekerheid voor meerdere onafhankelijke grootheden

Naast propagatie van onzekerheden voor één enkele variabele, komt men in de praktijk vaker de situatie tegen waarbij de functie f afhankelijk is van meerdere grootheden $x, y, \cdots$, waarvan de "beste schatters" X, $Y, \cdots$ en onzekerheden $\sigma_X, \sigma_Y, \cdots$ bepaald zijn.
In deze sectie maken we op basis van eenvoudige argumenten eerst simpele (maar niet optimale) afschattingen voor optelling $z = x + y$ en vermenigvuldiging $z = x \times y$. Vervolgens wordt theoretisch uitgewerkt hoe dan wél tot een correcte propagatie van onzekerheid te komen.

Tabel 5.2: Enkele rekenregels voor propagatie van onzekerheid voor één variabele x.

Functie $z = f(x)$	$\left\lvert\frac{df}{dx}\right\rvert$	**σ_Z**	**Consequentie**
$z = f(x) = a + x$	1	σ_X	$\sigma_Z = \sigma_X$
$z = f(x) = ax^n$	$\lvert anx^{n-1}\rvert$	$\lvert anX^{n-1}\rvert\,\sigma_X$	$\lvert\frac{\sigma_Z}{Z}\rvert = n\lvert\frac{\sigma_X}{X}\rvert$
$z = f(x) = a\log(bx)$	$\lvert a\frac{1}{x}\rvert$	$\lvert\frac{a}{X}\rvert\,\sigma_X$	$\sigma_Z = \lvert a\rvert\lvert\frac{\sigma_X}{X}\rvert$
$z = f(x) = ae^{bx}$	$\lvert abe^{bx}\rvert$	$\lvert abe^{bX}\rvert\,\sigma_X$	$\lvert\frac{\sigma_Z}{Z}\rvert = \lvert b\rvert\,\sigma_X$

5.3.1 Provisorische afschatting van propagatie van onzekerheid

Som van twee meetresultaten Stel we hebben voor grootheden x en y twee metingen X en Y gedaan met bijbehorende onzekerheden σ_X respectievelijk σ_Y (bijvoorbeeld x en y zijn lengten). We willen voor een toepassing de resultaten optellen tot $z = x + y$. Wat is nu de onzekerheid σ_Z in Z?
Voor het beantwoorden van deze vraag stellen we eerst een *provisorische regel* op. Voor een optelling van twee onafhankelijke meetresultaten X en Y nemen we aan dat een uitkomst voor x met zekere waarschijnlijkheid ligt in het interval $[X - \sigma_X, X + \sigma_X]$ en een uitkomst voor y in het interval $[Y - \sigma_Y, Y + \sigma_Y]$. De minimale uitkomst van de som bedraagt dan $Z = X + Y - \sigma_X - \sigma_Y$, de maximale $X + Y + \sigma_X + \sigma_Y$ (eigenlijk een soort variatiemethode). Bij optellen formuleren we hiermee een eerste *provisorische regel*:

$$Z = X + Y, \tag{5.12}$$

$$\sigma_Z \approx \sigma_X + \sigma_Y. \tag{5.13}$$

Het zal blijken dat deze σ_Z in het algemeen een *overschatting* is! Dit is niet echt verwonderlijk. In Vgl. (5.13) is als onzekerheid de som van de afzonderlijke onzekerheden σ_X en σ_Y genomen. Maar in datasets zal in ongeveer 50% van de metingen i de waarde x_i te groot uitvallen en y_i te klein, en *vice versa*. Als gevolg hiervan zullen positieve en negatieve uitschieters elkaar (deels) compenseren.

Vermenigvuldiging van twee meetresultaten Ook in geval van een vermenigvuldiging $v = x \times y$ (in bovenstaand voorbeeld corresponderend met het berekenen van een oppervlak) kunnen we een *provisorische regel* formuleren.
De minimale uitkomst voor V uitgaande van meetwaarden $X \pm \sigma_X, Y \pm \sigma_Y$ bedraagt (aannemende dat zowel X als Y positief is) ongeveer $X \times Y - X \times \sigma_Y - Y \times \sigma_X$. De maximale waarde bedraagt ongeveer $X \times Y + X \times \sigma_Y + Y \times \sigma_X$ (we nemen hier aan dat σ_X en σ_Y zo klein zijn dat het product $\sigma_X \times \sigma_Y$ verwaarloosbaar is ten opzichte van termen $X \times Y$, $X \times \sigma_Y$, $Y \times \sigma_X$). Daarmee volgt $\sigma_V = X \times \sigma_Y + Y \times \sigma_X$. Deze uitkomst is makkelijker herkenbaar in haar *relatieve vorm*:

$$V = X \times Y, \tag{5.14}$$

$$\left\lvert\frac{\sigma_V}{V}\right\rvert \approx \left\lvert\frac{\sigma_X}{X}\right\rvert + \left\lvert\frac{\sigma_Y}{Y}\right\rvert. \tag{5.15}$$

We zullen dadelijk zien dat ook deze provisorische regel Vgl. (5.15) een *bovengrens* vastlegt voor de onzekerheid, maar in het algemeen een *overschatting* vormt.

5.3.2 Algemene uitwerking van differentieermethode voor twee (of meer) statistische variabelen

In deze sectie zullen we met behulp van een theoretische uitwerking tot een "betere" schatting voor de onzekerheid dan de voorgaande simpele afschattingen. De uitwerking is voor een functie van twee variabelen $z = f(x,y)$ en twee datasets $x = \{x_1, x_2, \cdots, x_n\}$ en $y = \{y_1, y_2, \cdots, y_n\}$ met standaarddeviaties σ_x en σ_y en standaarddeviaties van het gemiddelde σ_X en σ_Y, maar is eenvoudig te generaliseren naar meer variabelen.

Taylorbenadering rondom het gemiddelde $f(X,Y)$ Wanneer meetwaarden x_i, y_i zijn bepaald, kan voor kleine afwijkingen van $z_i = f(x_i, y_i)$ rondom $f(X,Y)$ analoog aan Vgl. (5.9) een Taylorbenadering uitgevoerd worden in het punt X,Y, maar nu naar twee variabelen:

$$\begin{aligned} z_i &= f(x_i, y_i) \\ &= f(X,Y) + \left(\frac{\partial f}{\partial x}\right)_{x=X,y=Y}(x_i - X) + \left(\frac{\partial f}{\partial y}\right)_{x=X,y=Y}(y_i - Y) \\ &\quad + O((x_i-X)^2, (x_i-X)(y_i-Y), (y_i-Y)^2) \\ &\simeq f(X,Y) + \left(\frac{\partial f}{\partial x}\right)_{X,Y}(x_i - X) + \left(\frac{\partial f}{\partial y}\right)_{X,Y}(y_i - Y). \end{aligned} \tag{5.16}$$

$\left(\frac{\partial f}{\partial x}\right)_{X,Y}$ en $\left(\frac{\partial f}{\partial y}\right)_{X,Y}$ zijn hier de *partiële* afgeleiden van de functie $f(x,y)$ respectievelijk naar x en y geëvalueerd in het punt $(x = X, y = Y)$. In het vervolg geven we de toevoegingen X,Y bij de partiële afgeleiden niet meer weer. Een uitdrukking voor het gemiddelde Z van z vinden we door de uitdrukking voor het gemiddelde Vgl. (3.1) toe te passen op de dataset $\{z_1, \cdots, z_n\}$ en Vgl. (5.16) te substitueren:

$$\begin{aligned} Z &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} f(x_i, y_i) \\ &\simeq \frac{1}{n}\sum_{i=1}^{n}\left(f(X,Y) + \frac{\partial f}{\partial x}(x_i - X) + \frac{\partial f}{\partial y}(y_i - Y)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} f(X,Y) + \frac{\partial f}{\partial x}\frac{1}{n}\sum_{i=1}^{n}(x_i - X) + \frac{\partial f}{\partial y}\frac{1}{n}\sum_{i=1}^{n}(y_i - Y) \qquad (5.17) \\ &= f(X,Y). \qquad (5.18) \end{aligned}$$

Bij de laatste stap is gebruik gemaakt van Vgl. (3.1) voor de gemiddelden X en Y. We vinden dus dat de beste schatter Z zoals verwacht in eerste-orde benadering gelijk is aan $f(X,Y)$.
Op een analoge manier kunnen we nu ook een uitdrukking afleiden voor de variantie σ_z^2 uitgedrukt in σ_x en σ_y door Vgl. (5.16) te substitueren in Vgl. (3.7):

$$\begin{aligned} \sigma_z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(f(x_i,y_i) - Z)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(f(x_i,y_i) - f(X,Y))^2 \\ &\simeq \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x}(x_i - X) + \frac{\partial f}{\partial y}(y_i - Y)\right)^2 \\ &= \frac{1}{n-1}\left(\left(\frac{\partial f}{\partial x}\right)^2\sum_{i=1}^{n}(x_i - X)^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\sum_{i=1}^{n}(x_i - X)(y_i - Y) + \left(\frac{\partial f}{\partial y}\right)^2\sum_{i=1}^{n}(y_i - Y)^2\right) \\ &= \left(\frac{\partial f}{\partial x}\right)^2\sigma_x^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - X)(y_i - Y) + \left(\frac{\partial f}{\partial y}\right)^2\sigma_y^2. \end{aligned} \tag{5.19}$$

In Vgl. (5.19) herkennen we in de eerste en laatste term het gekwadrateerde resultaat van de differentieermethode voor één variabele (zie Vgl. (5.9)). De middelste term is een *correlatieterm*. Deze term heeft (voor grote aantallen n) alleen een waarde ongelijk aan nul wanneer de gemeten waarden x_i en y_i op de één of andere manier gekoppeld (of afhankelijk) zijn. Ga maar na: als x_i en y_i onafhankelijk van elkaar zijn en *random* verdeeld, dan zullen in de statistische limiet de gevallen [x_i ondergemiddeld én y_i bovengemiddeld], en [x_i bovengemiddeld én y_i ondergemiddeld] - resulterend in negatieve termen in de som - evenveel voorkomen als de gevallen [x_i en y_i beide bovengemiddeld], en [x_i en y_i beide ondergemiddeld] - resulterend

in positieve somtermen. In de statistische limiet is de som, en dus de correlatieterm, gelijk aan nul voor onafhankelijke data x_i en y_i.

In het voorgaande hebben we steeds van datasets geëist dat meetpunten *statistisch onafhankelijk* zijn. Hier kunnen we nu het belang van deze eis formuleren. Wanneer de resultaten voor x en y op de een of andere manier gekoppeld of *statistisch afhankelijk* zijn (bijvoorbeeld: als x_i bovengemiddeld, dan y_i met kans beduidend groter dan 50% óók bovengemiddeld), dan is de correlatieterm niet gelijk aan nul en moet voor een berekende onzekerheid in z een correctie worden aangebracht. We werken het onderwerp correlaties verder uit in Sectie 5.4.

Aangezien de variantie voor de beste schatter σ_Z^2 gevonden wordt door een eenvoudige herschaling σ_z^2/n, kunnen we de onzekerheid in de beste schatter van de dataset $\{z_1,\cdots,z_n\}$ uit Vgl. (5.19) berekenen als

$$\begin{aligned}\sigma_Z^2 &= \left(\frac{\partial f}{\partial x}\right)^2\frac{\sigma_x^2}{n} + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i-X)(y_i-Y) + \left(\frac{\partial f}{\partial y}\right)^2\frac{\sigma_y^2}{n}\\ &= \left(\frac{\partial f}{\partial x}\right)^2\sigma_X^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i-X)(y_i-Y) + \left(\frac{\partial f}{\partial y}\right)^2\sigma_Y^2. \qquad (5.20)\end{aligned}$$

Het blijkt dat voor *statistisch onafhankelijke* data het voldoende is om de beste schatters X en Y en de respectievelijke onzekerheden in de beste schatter σ_X, σ_Y te kennen. Voor een functie $z=f(x,y)$ van twee grootheden definiëren wij dan de differentieermethode als

$$\sigma_Z = \sqrt{\left(\frac{\partial f}{\partial x}\sigma_X\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_Y\right)^2} \qquad (5.21)$$

$$\equiv \sqrt{\sigma_{Z,x}^2 + \sigma_{Z,y}^2}. \qquad (5.22)$$

In de laatste regel hebben we de *partiële onzekerheden* in Z ten gevolge van respectievelijk x en y gedefinieerd:

$$\sigma_{Z,x} = \left|\frac{\partial f}{\partial x}\right|\sigma_X \qquad (5.23)$$

$$\sigma_{Z,y} = \left|\frac{\partial f}{\partial y}\right|\sigma_Y. \qquad (5.24)$$

De partiële onzekerheid $\sigma_{Z,x}$ is te begrijpen als de onzekerheidsbijdrage in Z ten gevolge van de grootheid x. Voor de totale onzekerheid worden de partiële onzekerheden ten gevolge van grootheden x en y *kwadratisch opgeteld*, zie Vgl. (5.22).

Voor een functie $f(x_1,\cdots,x_m)$ van een willekeurig aantal m grootheden die onderling statistisch onafhankelijk zijn, generaliseert Vgl. (5.21) tot

$$\sigma_Z = \sqrt{\sum_{i=1}^{m}\left(\frac{\partial f}{\partial x_i}\sigma_{X_i}\right)^2}. \qquad (5.25)$$

In andere woorden: bereken voor elke (statistisch onafhankelijke) variabele x_i afzonderlijk de partiële onzekerheid $\sigma_{Z,x_i} = \left|\frac{\partial f}{\partial x_i}\right|\sigma_{X_i}$ geëvalueerd in het punt $(X_1,X_2,\cdots,X_m)$. De onzekerheid van de beste schatter Z volgt dan uit de kwadratische som Vgl. (5.25) van de partiële onzekerheden.

Tabel 5.3: Enkele rekenregels voor propagatie van onzekerheid voor twee variabelen x, y. Bij vermenigvuldiging(/deling/machtsverheffing) werken we dus met *relatieve* onzekerheden!

Functie $z = f(x,y)$	$\frac{\partial f}{\partial x}$	$\frac{\partial f}{\partial y}$	**Onzekerheid in Z**
$z = f(x,y) = ax + by$	a	b	$\sigma_Z = \sqrt{(a\sigma_X)^2 + (b\sigma_Y)^2}$
$z = f(x) = ax^n y^m$	$anx^{n-1}y^m$	$amx^n y^{m-1}$	$\frac{\sigma_Z}{Z} = \sqrt{(n\frac{\sigma_X}{X})^2 + (m\frac{\sigma_Y}{Y})^2}$

Standaarddeviatie in het gemiddelde In Hoofdstuk 3 is al de uitdrukking voor de standaarddeviatie in het gemiddelde $\hat{\sigma}_{\overline{x}}$ gegeven (Vgl. (3.6)). Deze uitdrukking is met behulp van de differentieermethode voor propagatie van onzekerheid voor meer variabelen af te leiden. Beschouw het gemiddelde $\overline{x}$ als een functie $f = f(x_1, \cdots, x_n)$:

$$f(x_1, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{5.26}$$

We nemen verder aan dat elke waarde x_i bepaald is met eenzelfde waarde voor de standaarddeviatie $\hat{\sigma}_x$, gelijk aan de standaarddeviatie van de dataset. Dan is de partiële onzekerheid σ_{f,x_i} voor f ten gevolge van x_i te berekenen als

$$\sigma_{f,x_i} = \frac{\partial f(x_1, \cdots, x_m)}{\partial x_i}\hat{\sigma}_x \tag{5.27}$$

$$= \frac{1}{n}\hat{\sigma}_x. \tag{5.28}$$

Met behulp van Vgl. (5.25) volgt dan eenvoudig

$$\sigma_{\overline{x}} = \sqrt{\sum_{i=1}^{n}\sigma_{f,x_i}^2} = \sqrt{\sum_{i=1}^{n}\frac{1}{n^2}\hat{\sigma}_x^2} = \frac{1}{n}\hat{\sigma}_x\sqrt{\sum_{i=1}^{n}1}$$

$$= \hat{\sigma}_x/\sqrt{n}. \tag{5.29}$$

5.3.3 Rekenregels bij bewerkingen met twee statistische variabelen

Een aantal eenvoudige rekenregels die kunnen worden afgeleid met behulp van de algemene uitdrukking Vgl. (5.21) staan in Tabel 5.3. Voor complexere functies die bestaan uit meerdere variabelen, is het dikwijls werkzamer om de partiële onzekerheden met *Mathematica* uit te rekenen.

Met deze resultaten in de hand kijken we nog even terug naar de analyse op het niveau van verdelingen in Sectie 5.3. Bij optelling moeten de standaarddeviaties σ_X en σ_Y niet lineair worden opgeteld ($\sigma_Z \neq \sigma_X + \sigma_Y$), maar kwadratisch: $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2} \leq \sigma_X + \sigma_Y$. Evenzo dient bij vermenigvuldiging niet relatief lineair te worden opgeteld ($\sigma_V/V \neq |\sigma_X/X| + |\sigma_Y/Y|$), maar relatief kwadratisch: $\sigma_V/V = \sqrt{(\sigma_X/X)^2 + (\sigma_Y/Y)^2} \leq |\sigma_X/X| + |\sigma_Y/Y|$. In beide gevallen is de onzekerheid in de berekende grootheid dus kleiner dan volgens de *provisorische regels*, en is er sprake van een nauwkeuriger afschatting.

In Sectie 5.4 zullen we zien dat de *provisorische regels* voor propagatie toepasbaar zijn in de situatie dat er sprake is van maximale afhankelijkheid (correlatie) tussen de variabelen.

5.3.4 Meetkundige interpretatie van kwadratische optelling

We kunnen de kwadratische optelling van partiële onzekerheden voor een functie van meerdere onafhankelijke variabelen ook een meetkundige interpretatie geven. Voor een functie van twee variabelen kunnen we een figuur analoog aan Figuur 5.3 tekenen, maar dan uitgebreid naar drie dimensies. Als voorbeeld nemen we het bepalen van de kinetische energie $E(m,v) = \frac{1}{2}mv^2$ voor een auto met massa $m = 1500 \pm 100$ kg en snelheid $v = 30 \pm 8$ m/s. In Figuur 5.4 (a) is deze functie getekend als een vlak in de buurt van $(M,V) = (1500,30)$.
De vector

$$\vec{w} = \begin{pmatrix} m \\ v \\ E(m,v) = \frac{1}{2}mv^2 \end{pmatrix} \tag{5.30}$$

beschrijft alle punten in het vlak $E(m,v) = \frac{1}{2}mv^2$. De partiële afgeleiden

$$\frac{\partial \vec{w}}{\partial m} = \begin{pmatrix} \frac{\partial m}{\partial m} \\ \frac{\partial v}{\partial m} \\ \frac{\partial E}{\partial m} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \frac{1}{2}v^2 \end{pmatrix} \quad \text{en} \tag{5.31}$$

$$\frac{\partial \vec{w}}{\partial v} = \begin{pmatrix} \frac{\partial m}{\partial v} \\ \frac{\partial v}{\partial v} \\ \frac{\partial E}{\partial v} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ mv \end{pmatrix} \tag{5.32}$$

beschrijven de raaklijnen aan de functie in respectievelijk de m- en v-richting. Wanneer we beide raakvectoren schalen met respectievelijk σ_M en σ_V vinden we vectoren waarvan de verandering in de (verticale) E-richting correspondeert met de partiële onzekerheden $\sigma_{E,m} = \frac{\partial E}{\partial m}\sigma_M$ en $\sigma_{E,v} = \frac{\partial E}{\partial v}\sigma_V$. Deze geschaalde raakvectoren zijn weergegeven met de pijlen in Figuur 5.4. Duidelijk is dat $\sigma_{E,v}$ dominant is.
Omdat de variabelen m en v statistisch onafhankelijk zijn, staan deze twee partiële onzekerheden loodrecht op elkaar. De kwadratische som σ_E kan nu worden opgevat als de *norm* (lengte) van de vector $\begin{pmatrix} \sigma_{E,m} \\ \sigma_{E,v} \end{pmatrix}$.

5.3.5 Variatiemethode voor een functie van meerdere onafhankelijke variabelen

Ook de variatiemethode is toe te passen voor meerdere onafhankelijke variabelen. Voor een functie $z = f(x_1,x_2,\cdots,x_m)$ van een willekeurig aantal variabelen die onderling statistisch onafhankelijk zijn wordt de onzekerheid σ_Z in Z gegeven door

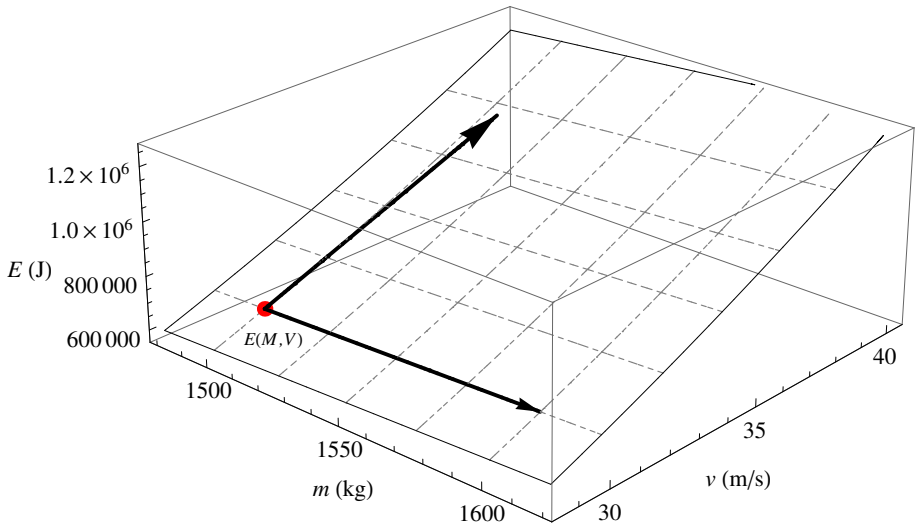
$$\sigma_Z = \sqrt{\sum_{i=1}^{m} \sigma_{Z,x_i}^2}, \quad \text{met} \tag{5.33}$$

$$\sigma_{Z,x_i} = f(X_1,X_2,\cdots,X_i + \frac{1}{2}\sigma_{X_i},\cdots,X_m) - f(X_1,X_2,\cdots,X_i - \frac{1}{2}\sigma_{X_i},\cdots,X_m). \tag{5.34}$$

Hier is σ_{Z,x_i} de partiële onzekerheid in Z ten gevolge van x_i.

5.4 Correlaties

Zoals in Vgl. (5.19) aangegeven, bevat de formule voor de propagatie van onzekerheid voor een functie $f(x,y)$ een term die samenhangt met *statistische afhankelijkheid* of *correlatie* tussen de twee variabelen x en y. In deze sectie zullen we deze correlatie nader beschouwen. Aan statistische afhankelijkheden tussen datasets kunnen verschillende oorzaken/verklaringen ten grondslag liggen. We bespreken daarom eerst enkele voorbeelden van correlatie.



Figuur 5.4: Illustratie van het doorwerken van onzekerheid in de differentieermethode voor een functie $E(m,v)$ van twee variabelen. De functie $E(m,v)$ is als een (gekromd) vlak getekend in de omgeving van gemeten waarden $M = 1500$ kg en $V = 30$ m/s. De partiële onzekerheden $\sigma_{E,m}$ en $\sigma_{E,v}$ zijn gerelateerd aan de raaklijnen aan de functie $E(m,v)$, die na schaling met respectievelijk σ_M en σ_V, zijn weergegeven met pijlen. De verticale component van de vector is gelijk aan de partiële onzekerheidsbijdrage.

Statistische verbanden voor dataparen In sociale en medische wetenschappen worden door middel van vragenlijsten vaak statistische relaties tussen gedragingen gevonden, die evenwel niet altijd eenvoudig zijn te interpreteren. Zo blijkt uit de DATA-V studentenenquête van het jaar 2011-2012 dat studenten die het dictaat duidelijk vinden, het dictaat ook vaker gebruiken (of andersom). Er is zogezegd sprake van een correlatie of statistisch verband tussen de respons op deze twee afzonderlijke vragen.
De sterkte van een dergelijk verband wordt beperkt door het feit dat er een statistische spreiding in de antwoorden zit, er zijn immers ook studenten die als respons geven "dictaat redelijk duidelijk - dictaat zelden gebruikt", "dictaat onduidelijk - dictaat vaak gebruikt", etc. Een valkuil bij dergelijk onderzoek is dat het statistisch verband ten onrechte wordt opgevat als een causaal verband. Het hoeft zeker niet zo te zijn dat studenten het dictaat vaak gebruiken *omdat* ze het duidelijk vinden (alhoewel de docenten dat wel graag willen geloven).

Functioneel geïntroduceerde correlatie In de natuurkunde is een analytische relatie tussen grootheden vaak te identificeren op basis van theoretische overwegingen. Op veel andere gebieden is weliswaar vaak een logische relatie te observeren tussen grootheden, maar is de relatie niet analytisch uit te drukken.
Een voorbeeld is de correlatie tussen de lengte en het gewicht van (volwassen) mensen. Een natuurkundige zal dit verband wellicht direct proberen te duiden door het opstellen van een functioneel verband tussen deze twee grootheden. Afhankelijk van de mate van correlatie en het feit dat er ook statistisch toevallige spreiding in de data zit, is het functionele verband zelf vaak niet eenduidig herkenbaar. Het definiëren van de mate van correlatie is dan een zinvolle (eerste) stap.

Correlatie tussen parameters van een aanpassing In het volgende hoofdstuk zullen we zien dat bij het uitvoeren van een aanpassing aan data, zoals het vinden van de best passende rechte lijn $y = A + Bx$ aan een set meetresultaten (zie Figuur 6.6), de gevonden "beste" waarden van de parameters (A en B) statistisch ge-

zien afhankelijk (kunnen) zijn. Intuïtief is een dergelijke afhankelijkheid goed te begrijpen; wanneer we de waarde van de asafsnijding A iets laten toenemen, dan zal dat gecompenseerd worden door een verandering van de richtingscoëfficiënt B opdat de rechte lijn toch zo goed mogelijk door de punten heen gaat.

De aanwezigheid van correlatie in data is vaak een gegeven bij een bepaalde experimentele procedure en niet noodzakelijk problematisch. Het is echter wel belangrijk om de mate van correlatie te kwantificeren en er vervolgens adequaat mee om te gaan bij propagatie van onzekerheden. Zoals in de vorige sectie formuleren we de propagatie van onzekerheden in aanwezigheid van correlatie in eerste instantie voor twee variabelen. De generalisatie naar meerdere variabelen komt later aan de orde.

5.4.1 Kwantificeren van correlatie voor datasets: covariantie en correlatiecoefficient

We gaan uit van grootheden x en y, waaraan *paarsgewijze* metingen $\{\{x_1,y_1\},\{x_2,y_2\},\cdots,\{x_n,y_n\}\}$ zijn uitgevoerd. We voeren in de *covariantie* tussen de datasets $\{x_1,\cdots,x_n\}$ en $\{y_1,\cdots,y_n\}$ als:

$$\sigma_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - X)(y_i - Y). \tag{5.35}$$

Hierbij is het van belang dat de paarsgewijze samenstelling gehandhaafd blijft! Door bijvoorbeeld de lijst $\{x_1,\cdots,x_n\}$ op grootte te sorteren en $\{y_1,\cdots,y_n\}$ niet, wordt de berekening van de covariantie verstoord. Van de covariantie voor de beste schatter $\sigma_{X,Y} = \sigma_{x,y}/n$ wordt in het algemeen weinig gebruik gemaakt.
De *lineaire correlatiecoëfficiënt* $\rho_{x,y}$ wordt nu gedefinieerd als de verhouding tussen de covariantie en de standaarddeviaties van het gemiddelde voor x en y:

$$\begin{aligned}\rho_{x,y} &= \frac{\sigma_{x,y}}{\sigma_x\sigma_y} \\ &= \frac{\sum_i (x_i - X)(y_i - Y)}{\sqrt{\sum_i (x_i - X)^2 \sum_i (y_i - Y)^2}}.\end{aligned} \tag{5.36}$$

Merk op dat het voor de correlatiecoëfficiënt niet uitmaakt of σ_x, σ_y en $\sigma_{x,y}$ worden gebruikt of de geschaalde versies. Een substitutie in Vgl. (5.20) levert dan de uitdrukking voor propagatie van onzekerheid voor gecorreleerde grootheden:

$$\sigma_Z^2 = \left(\frac{\partial f}{\partial x}\sigma_X\right)^2 + 2\rho_{x,y}\left(\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right)\sigma_X\sigma_Y + \left(\frac{\partial f}{\partial_y}\sigma_Y\right)^2. \tag{5.37}$$

De waarde van $\rho_{x,y}$ is beperkt tot het interval $[-1,1]$. Bij $\rho = 1$ zijn x en y maximaal positief gecorreleerd, en bij $\rho = -1$ maximaal negatief gecorreleerd.
Correlatie tussen twee datasets kan dus onderzocht worden door de covariantie uit te rekenen, en vervolgens de correlatiecoëfficiënt. Wanneer deze laatste significant verschilt van nul, dient in verwerking van de resultaten met correlatie rekening te worden gehouden.
Als voorbeeld bekijken we de functie $f(x,y) = x \times y$ met gemiddelden X en Y en standaarddeviaties σ_X en σ_Y. Wanneer x en y ongecorreleerd zijn, berekenen we als onzekerheid volgens de regels in Sectie 5.3.2

$$\sigma_{f,\rho=0} = \sqrt{(Y\sigma_X)^2 + (X\sigma_Y)^2}. \tag{5.38}$$

In het extreme geval dat y volledig positief gecorreleerd is met x ($\rho_{X,Y} = 1$), berekenen we met Vgl. (5.37):

$$\begin{aligned}\sigma_{f,\rho=+1} &= \sqrt{(Y\sigma_X)^2 + 2XY\sigma_X\sigma_Y + (X\sigma_Y)^2}. \\ &= \sqrt{X^2Y^2\left(\frac{\sigma_X}{X} + \frac{\sigma_Y}{Y}\right)^2} \\ &= |Y\sigma_X + X\sigma_Y|. \end{aligned} \tag{5.39}$$

Dit resultaat is gelijk aan de *provisorische regel* Vgl. (5.15)! In het geval van maximale positieve correlatie is er géén sprake is van uitmiddeling zoals geschetst in Sectie 5.3: als x_i groter dan gemiddeld, dan is zeker y_i óók groter dan gemiddeld. Wanneer X en Y beiden groter zijn dan nul, dan neemt in dit geval de onzekerheid in f dus *toe* voor positieve correlatie.

In het andere extreme geval, $\rho_{X,Y} = -1$, is juist sprake van maximale uitmiddeling: als x_i groter dan gemiddeld, dan is zeker y_i kleiner dan gemiddeld. Voor $X, Y > 0$ leidt negatieve correlatie voor het product dus tot een *afname* in de onzekerheid:[1]

$$\begin{aligned}\sigma_{f,\rho=-1} &= \sqrt{(Y\sigma_X)^2 - 2XY\sigma_X\sigma_Y + (X\sigma_Y)^2}. \\ &= \sqrt{X^2Y^2\left(\frac{\sigma_X}{X} - \frac{\sigma_Y}{Y}\right)^2} \\ &= |Y\sigma_X - X\sigma_Y|. \end{aligned} \tag{5.40}$$

[1] Er geldt niet in het algemeen dat σ_f toeneemt bij positieve correlatie respectievelijk afneemt bij negatieve correlatie. Dit heeft te maken met de tekens van de afgeleiden $\frac{\partial f}{\partial x}$ en $\frac{\partial f}{\partial y}$. Een voorbeeld waarbij σ_f afneemt bij positieve correlatie wordt beschreven in het kader.

Ontstaan van correlatie door functioneel verband Hier is een voorbeeld van correlatie die je creëert (soms zonder het jezelf te realiseren) door statistisch onafhankelijke grootheden in een functionele relatie te combineren.
De polarisatiegraad P van een lichtbundel wordt gedefinieerd als het verschil tussen twee lichtintensiteiten I_h en I_v (gemeten met een polarisator in horizontale en verticale stand) gedeeld door de som van deze beide intensiteiten:

$$P = \frac{I_h - I_v}{I_h + I_v}. \tag{5.41}$$

Na het uitvoeren van meetseries worden de beste schatters voor de lichtintensiteiten voor de polarisatorstanden gegeven door $I_p = 150 \pm 15$ mW en $I_m = 50 \pm 5$ mW. Deze twee grootheden zijn statistisch onafhankelijk. De vraag is nu uiteraard wat de beste schatter van de polarisatiegraad is (natuurlijk met onzekerheid!).
Het lijkt een optie om met de rekenregels in Tabel 5.3 eerst de beste schatters voor de teller $T = |I_h - I_v|$ en noemer $N = I_h + I_v$ uit te rekenen. Dit resulteert in $\hat{T} = 100 \pm 16$ mW en $\hat{N} = 200 \pm 16$ mW. De beste schatter voor de polarisatiegraad $P = T/N$ wordt vervolgens verkregen door de rekenregels voor een quotiënt toe te passen, $\hat{P} = 0.50 \pm 0.09$.
Hoewel deze aanpak logisch lijkt, zit er toch een addertje onder het gras: de teller T en noemer N zijn *gecorreleerd*, omdat ze beiden afhankelijk zijn van grootheden I_p en I_m! Om de onzekerheid σ_P uit de beste schatters $\hat{T}$ en $\hat{N}$ te kunnen berekenen zou eerst de correlatiecoëfficiënt $\rho_{T,N}$ berekend moeten worden. Maar wat doe je als er maar één meting voor I_h en I_v is gedaan, of wanneer de metingen niet paarsgewijs zijn gedaan? Het gebruik van $P = P(T,N)$ voor propagatie is vanwege correlatie tussen T en N dus niet zo'n goed idee.
De transparante en eenvoudige manier om de beste schatter voor P te bepalen is door te stellen dat P berekend wordt uit de *statistisch onafhankelijke grootheden* I_h en I_v. Oftewel $P = P(I_h, I_v)$; dat is dus Vgl. (5.41)! Reken zelf na dat het toepassen van de variatiemethode op Vgl. (5.41) resulteert in een beste schatter $\hat{P} = 0.50 \pm 0.05$. Je ziet dat correlatie in de "foute" methode de onzekerheid in P significant heeft beïnvloed.

Dit voorbeeld illustreert dat het bij propagatie van onzekerheden raadzaam is om waar mogelijk met functionele verbanden te werken die zijn uitgedrukt in statistisch onafhankelijke grootheden. Er hoeft dan geen rekening gehouden te worden met het optreden van correlatie-effecten (bijvoorbeeld door het opdelen van de uitdrukking in afhankelijke deelformules).

5.4.2 Correlatie voor meer dan twee variabelen: de covariantiematrix

Wanneer er sprake is van meer dan twee variabelen, $x_1, \cdots, x_m$, kunnen er tussen alle variabelen correlaties bestaan. Voor elk paar variabelen x_k en x_l kunnen we dan een covariantie $\sigma_{k,l}$ vaststellen. De covarianties kunnen op overzichtelijke manier worden weergegeven in de zogenaamde *covariantiematrix* C:

$$C = \begin{pmatrix} \sigma_{x_1,x_1} = \sigma_{x_1}^2 & \sigma_{x_1,x_2} & \cdots & \sigma_{x_1,x_m} \\ \sigma_{x_2,x_1} & \sigma_{x_2,x_2} = \sigma_{x_2}^2 & \cdots & \sigma_{x_2,x_m} \\ \vdots & \vdots & \ddots & \\ \sigma_{x_m,x_1} & \sigma_{x_m,x_2} & & \sigma_{x_m,x_m} = \sigma_{x_m}^2 \end{pmatrix}. \tag{5.42}$$

Op de hoofddiagonaal vindt men dan de varianties. Een buitendiagonaalelement $\sigma_{k,l}$ is de *covariantie* tussen x_k en x_l en representeert de *statistische correlatie* tussen x_k en x_l. De covariantie $\sigma_{k,l}$ is altijd gelijk aan $\sigma_{l,k}$.

Voor $k \neq l$ kan $\sigma_{k,l}$ negatief zijn, echter gebruikmakend van de relatie $\sigma_{k,l} = \rho_{k,l}\sqrt{\sigma_{k,k}\sigma_{l,l}}$ geldt er altijd dat:

$$|\sigma_{k,l}| \leq \sqrt{\sigma_{k,k}\sigma_{l,l}}. \tag{5.43}$$

Wanneer alle variabelen k en l (min of meer) ongecorreleerd zijn, is de matrix C bij benadering een diagonaalmatrix. De covariantiematrix zal in het volgende hoofdstuk terugkomen, maar dan niet als uitdrukking van correlatie tussen verschillende *datasets*, maar van correlatie tussen *aanpassingsparameters*. De variantie σ_f^2 in de functie $f(x_1, \cdots, x_m)$ ten gevolge van varianties en covarianties in C is bij gebruik van de differentieermethode compact te schrijven als

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_m}\right) C \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_m} \end{pmatrix}. \tag{5.44}$$

5.5 Algemene aanpak bij het verwerken van numerieke informatie

We zijn nu zo ver dat we een redelijk arsenaal van mogelijkheden hebben om kwantitatieve informatie met inbegrip van de onzekerheid te verwerken. Hieronder een algemeen recept voor het berekenen van beste schatters en onzekerheden uit gemeten data.

5.5.1 Stappenplan

Formulering van het model

- Druk de te bepalen grootheid eerst *symbolisch* (in formulevorm) uit in *onafhankelijke* meetresultaten.
- Ga na onder welke voorwaarden deze formulering geldig is. Controleer in hoeverre aan deze voorwaarden voldaan is (in het licht van de aangeleverde meetgegevens).
- Bereken de beste schatter voor de te bepalen grootheid.

Bereken partiële onzekerheden

- Differentieermethode (vooral handig bij *functionele* doorrekening):
  - neem partiële afgeleiden naar de onafhankelijke meetresultaten;
  - vermenigvuldig deze met de bijbehorende onzekerheden;
- Variatiemethode (vooral nuttig bij *numerieke* doorrekening):
  - bereken variatie in uitkomsten bij variatie van beste schatter voor een grootheid over waarde $\pm$ halve standaarddeviatie, voor alle grootheden;

Bij gebruik van *Mathematica*:

- vermijd zoveel mogelijk het expliciet toekennen van numerieke waarden aan variabelen;
- gebruik een *vervangingsoperator* `/.` om numerieke waarden te berekenen;
- hierop aansluitend: vermijd (ook op papier) het gebruik van (afgeronde) numerieke tussenresultaten.

Bereken en rapporteer eindantwoord

- Tel de partiële onzekerheden, afkomstig van de onafhankelijke grootheden, kwadratisch op.
- Betrek, wanneer toch correlaties aanwezig zijn, ook eventuele correlatietermen in de berekening van de onzekerheid in de te bepalen grootheid.
- Rapporteer het eindantwoord volgens de regels gesteld in Hoofdstuk 2, dus inclusief symbool voor grootheid, eenheid en afgerond volgens de eisen voor significantie.

Probeer de opgaven bij dit hoofdstuk zoveel mogelijk uit te werken volgens dit recept.

5.5.2 Ontwerp van een betere meting

Aan het eind van de verwerking is het vaak nuttig om te reflecteren op het behaalde resultaat. In veel gevallen start je met een evaluatie van de beste schatter. Is dit resultaat (gegeven de onzekerheid in de beste schatter) in overeenstemming met de verwachting op basis van de literatuur of het gebruikte model? Een significantietoets (bepalen van overschrijdingskans) is hierbij vaak een nuttig hulpmiddel.
Daarnaast is, met het oog op het verbeteren van de meetmethode of meetprocedure, een terugblik op de behaalde meetresultaten nuttig. Wanneer je de nauwkeurigheid in het eindantwoord wilt verhogen, is de belangrijkste vraag welk van de huidige meetresultaten deze nauwkeurigheid het meest beperkt. Hiervoor moeten de eerder berekende *partiële onzekerheden* worden vergeleken. Vanwege de kwadratische optelling van de partiële onzekerheden is de bijdrage van één grootheid snel dominant. Het meetresultaat dat leidt tot de hoogste partiële onzekerheid moet als eerste verbeterd worden.
Door meerdere metingen te doen kan bijvoorbeeld de beperkende partiële onzekerheid verkleind worden (zie Hoofdstuk 3). Ook is het soms mogelijk de onzekerheid in één enkele meting te verkleinen door betere apparatuur aan te schaffen of de opstelling nog beter te prepareren.

Hoofdstuk 6

Aanpassingen aan data

6.1 Inleiding

6.1.1 Het meten en kwantificeren van functionele afhankelijkheden

In de onderzoekspraktijk worden vaak *functionele afhankelijkheden* gemeten. Eén voorbeeld hiervan hebben we al gegeven in Hoofdstuk 1 (stroom tegen spanning in Figuur 1.3). Een fysisch experiment kent vaak de volgende opzet. Er is een experimenteel toegankelijke grootheid (of variabele) y waaraan gemeten wordt, waarbij de *functionele afhankelijkheid* van deze zogenaamde *afhankelijke* grootheid y van een andere grootheid x getest wordt. We noemen in het vervolg x de *onafhankelijke* grootheid of variabele, omdat deze door de waarnemer in het experiment ingesteld wordt. De hypothese is (op basis van theorie of andersoortige overwegingen) dat y in functioneel verband staat met x middels $y = f(x)$.
De experimentator zal in het experiment de grootheid x instellen op een aantal waarden $x_1, \cdots, x_n$. De waarde y_i van y wordt met standaarddeviatie σ_i bepaald bij elk van de waarden x_i. We nemen in eerste instantie aan dat de waarden x_i met verwaarloosbaar kleine onzekerheid worden vastgelegd.[1] Zo worden combinaties $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ verkregen.
Gedurende het experiment wordt getracht de variaties van alle ándere grootheden die de waarde van y kunnen beïnvloeden te minimaliseren, zodat zij als constant kunnen worden beschouwd. Als de hypothese juist is, zal bij verandering van de waarde van x door de experimentator de waarde van y veranderen zoals vastgelegd door het verband $y = f(x)$. We mogen echter niet vergeten (Hoofdstuk 3) dat de metingen aan y vanwege de meetonzekerheid een statistische spreiding vertonen. Het verband tussen y en x zal dus nooit *exact* worden gereproduceerd.
Het experiment gaat een stap verder wanneer de set metingen $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ gebruikt wordt om een *functioneel verband* tussen de grootheden x en y te *kwantificeren*. Door het verband kwantitatief te maken, wordt het mogelijk om waarden te bepalen van grootheden of constanten $\boldsymbol{p} = (p_1, \cdots, p_r)$ (inclusief onzekerheid) die in het verband tussen x en y voorkomen. We noemen deze grootheden of constanten *parameters*. Het functioneel verband tussen x en y noteren we dan als $y = f_{\boldsymbol{p}}(x)$. De notatie met $\boldsymbol{p}$ als *subscript* aan de functie f drukt uit dat de waarden $\boldsymbol{p}$ gedurende het experiment constant worden verondersteld.
Het doel is nu uit de beschikbare dataset "zo goed mogelijk" de "beste waarden" $\hat{\boldsymbol{p}}$ voor de parameters $\boldsymbol{p}$ te bepalen. Daartoe wordt een *aanpassing* (Engels: *fit*) van de functie $f_{\boldsymbol{p}}(x)$ aan de beschikbare data gemaakt. Aanpassen betekent hier: het zoeken van die numerieke waarden van de modelparameters $\boldsymbol{p}$ waarvoor de functie $f_{\boldsymbol{p}}(x)$ bij waarden x_i van de onafhankelijke grootheid "zo weinig mogelijk" afwijkt van de verkregen meetresultaten y_i.
In dit hoofdstuk behandelen we de theorie achter dit aanpassingsprobleem, en werken we het analytisch uit voor de eenvoudige en vaak voorkomende gevallen $f_{(A)}(x) = A$ (gewogen gemiddelde) en $f_{(A,B)}(x) = A + Bx$

[1] Aan het eind van het hoofdstuk bespreken we kort de consequenties van een niet te verwaarlozen onzekerheid in x_i.

(rechte-lijnaanpassing). De rest van de inleiding op dit hoofdstuk is gewijd aan een bespreking van een concreet geval in kwalitatieve termen.

6.1.2 Voorbeeld: diffractie van licht aan een enkele spleet

Als voorbeeld behandelen we de bepaling van de hoek waaronder interferentieminima achter een smalle spleet worden waargenomen in een optisch diffractie-experiment.[2] De situatie is geschetst in Figuur 6.1 (a). Wanneer licht met een zekere golflengte λ op een smalle spleet met breedte d valt, ontstaat op voldoende afstand achter de spleet een karakteristiek diffractiepatroon. Dit patroon heeft interferentieminima bij specifieke hoeken θ. Men kan laten zien dat op voldoende grote afstand achter de spleet de hoeken θ_m waarbij minima gevonden worden voor waarden m (m heeltallig en $m \neq 0$), gelijk zijn aan:

$$\theta_m = \arcsin\left(\frac{m\lambda}{d}\right). \tag{6.1}$$

In dit voorbeeld geldt dus $y = \theta_m$ en $f = f(m,\lambda,d) = \arcsin\left(\frac{m\lambda}{d}\right)$. Bij het testen van een functioneel verband zullen we ervoor moeten zorgen dat één van de variabelen m,λ,d gecontroleerd wordt gevarieerd, terwijl de andere twee constant worden gehouden.[3]

Het testen van een functioneel verband De keuze voor een afhankelijke grootheid hangt vaak af van de experimentele situatie.

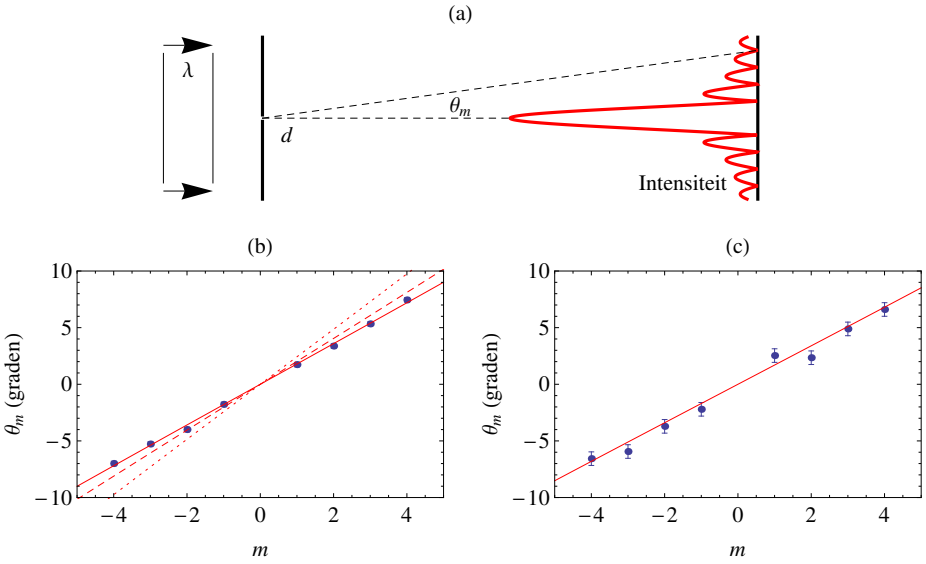
- Stel dat we in ons voorbeeld maar één spleet met breedte d en een lichtbron met maar één golflengte λ tot onze beschikking hebben, dan kunnen we de waarde van θ_m alleen bepalen met m als onafhankelijke variabele: $\theta_m = f_{\lambda,d}(m)$.
- Stel dat wij een spleet tot onze beschikking hadden waarvan de waarde d in te stellen en af te lezen valt, dan hadden wij d als onafhankelijke variabele kunnen nemen en voor een vaste waarde voor m het verband tussen d en θ_m kunnen bepalen: $\theta_m = f_{\lambda,m}(d)$.
- Met een lichtbron met instelbare golflengte hadden we d en m constant kunnen houden en λ kunnen variëren: $\theta_m = f_{m,d}(\lambda)$.

We veronderstellen dat bekend is dat de lichtbron een (diode)laser is met vaste golflengte $\lambda = 635$ nm, met verwaarloosbare onzekerheid. De constante waarde van d is onbekend.
In Figuur 6.1 (b) en (c) zijn voor ons experiment twee (fictieve) experimenteel bepaalde datasets te zien met θ_m als functie van m. Het verschil tussen (b) en (c) is de meetnauwkeurigheid in de waarde van θ_m. In (b) is de meetonzekerheid in elke hoekmeting 0.2°. De *error bars* zijn nauwelijks groter dan het punt. Bij (c) is de andere set metingen te zien, waarbij de meetonzekerheid 0.6° bedraagt. Hoewel duidelijk is dat er een verband is tussen m en θ_m, maakt de grotere meetonzekerheid het moeilijker om het kwantitatieve (werkelijke) verband te herkennen. Dat er een relatie is tussen m en θ_m is echter overduidelijk.

Kwantificeren van het verband We bepalen nu met behulp van de dataset θ_m tegen m in Figuur 6.1 (b) en (c) de waarde van een andere grootheid die voorkomt in het verband Vgl. (6.1). Omdat λ bekend is, kunnen we de spleetbreedte d bepalen. De waarde van d ligt in het algemeen in de orde van tientallen micrometers

[2] Dit experiment is onder de naam FRS1 onderdeel van het natuurkundepracticum.

[3] In veel gevallen is het zo dat de parameters die van belang zijn in het probleem bekend zijn. Het is natuurlijk ook mogelijk dat je wilt testen of een willekeurige variabele überhaupt een rol speelt. Stel dat wij in dit experiment bijvoorbeeld de omgevingstemperatuur T variëren van 20 °C tot 30 °C, dan vinden we behoudens effecten ten gevolge van toevallige onzekerheden geen merkbare verandering in θ_m. De variabele T staat in dit experiment dus niet in een (functioneel) verband met θ_m.



Figuur 6.1: (a) Intensiteitspatroon van licht met golflengte λ achter een enkele spleet met breedte d. De hoeken waarbij het patroon een minimum vertoont duiden we aan met θ_m. (b) Hoek θ_m als functie van de diffractieorde m bij constante waarden $\lambda = 635$ nm. De onzekerheid in alle waarden voor θ_m is 0.2°. Gestippelde, gestreepte en doorgetrokken lijn zijn respectievelijk een slechte, betere en "beste" aanpassing, met waarden $d = 15$ μm, $d = 18$ μm en $d = 20.3$ μm. Uit een aanpassing aan functie Vgl. (6.1), vinden we $d = 20.3 \pm 0.3$ μm met een χ^2_{red} van 1.20. (c) Een nieuwe set metingen θ_m met waarde $\lambda = 635$ nm, als bij (b). De meetwaarden kennen nu echter een onzekerheid van 0.6°. We vinden voor de spleetbreedte $d = 21.4 \pm 1.1$ μm met een χ^2_{red} van 1.20, bij toeval gelijk aan de waarde voor de data in (b).

en is vaak lastig direct te bepalen (b.v. met een microscoop). De hoeken θ_m zijn wél macroscopisch en dus nauwkeurig te meten.

Allereerst is het goed om te laten zien hoe de waarde van θ_m beïnvloed wordt door een keuze voor de waarde van d. In Figuur 6.1 (b) tonen we het verband tussen θ_m en m voor waarden $d = 15$ μm (stippellijn), $d = 18$ μm (gestreepte lijn), en de "beste aanpassing" $d = 20.3 \pm 0.3$ μm (doorgetrokken lijn). De eerste twee zijn overduidelijk geen aanpassingen met optimale waarde voor d; het verschil tussen model en metingen is erg groot. De "beste aanpassing" is gevonden met behulp van de theorie die later in dit hoofdstuk wordt behandeld. Het gebruik van een aanpassing voor deze serie metingen maakt het dus mogelijk de waarde van d vast te stellen met een onzekerheid kleiner dan een micrometer.

Wanneer we d bepalen met behulp van de dataset in Figuur 6.1 (c) vinden we $d = 21.4 \pm 1.1$ μm, een veel onnauwkeuriger resultaat dan voor de dataset in Figuur 6.1 (b). Dit sluit aan bij de intuïtie dat nauwkeuriger metingen moeten leiden tot een nauwkeurigere bepaling van parameters.

Het is goed om op te merken dat er twee manieren zijn om onzekerheden te berekenen: een *interne* die vooral wordt bepaald door de onzekerheden in de meetwaarden, en een *externe* die vooral wordt bepaald door de spreiding van de meetpunten. Hier wordt gerapporteerd met de externe onzekerheid. We duiden deze begrippen in Sectie 6.2.

Kwaliteit van de aanpassing De kwaliteitsmaat die we hanteren voor de overeenkomst tussen model en metingen (consistentie) is de in *gereduceerde chi-kwadraat* χ^2_{red}. Dit getal relateert de spreiding van meetwaarden rondom het model aan de onzekerheid die aan de meetwaarden is toegekend, en zou in de buurt van de statistisch ideale waarde 1 moeten liggen. De waarde voor de data in Figuur 6.1 (b) en (c) blijkt bij toeval ongeveer gelijk te zijn, namelijk $\chi^2_{red} = 1.20$. Met behulp van de methode van de *overschrijdingskans*,

al bekend uit Sectie 4.6 maar nu berekend voor de verdeling van de gereduceerde chi-kwadraat is aan te tonen dat dit voldoende dicht bij 1 is.
Hoewel de onzekerheden in het geval van Figuur 6.1 (c) veel groter zijn, is er nog steeds sprake van consistentie. Je kunt concluderen dat dit komt doordat de puntspreiding rondom het model net als de meetonzekerheid een factor 3 is toegenomen.

6.2 De "beste" aanpassing via de kleinste-kwadratenmethode

In deze sectie formaliseren we de procedure van het doen van een aanpassing zoals geschetst in de vorige sectie. Voorbeelden en illustraties volgen in het *Mathematica*-materiaal en later in dit hoofdstuk.
We beschouwen de dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$. De (onafhankelijke) x_i zijn hierbij instelwaarden met verwaarloosbare onzekerheid en de (afhankelijke) y_i zijn de meetwaarden bij deze instellingen verkregen, bepaald met onzekerheden σ_i. Als in de voorgaande sectie geven we het functionele verband tussen x en y (symbolisch) weer met de functie

$$y = f_{\boldsymbol{p}}(x), \tag{6.2}$$

waarbij de vector $\boldsymbol{p}$ *modelparameters* $(p_1, \cdots, p_r)$ weergeeft. Het doel is waarden $\hat{\boldsymbol{p}} = (\hat{p}_1, \cdots, \hat{p}_r)$ te vinden die het beste recht doen aan de gemeten waarden. Hierbij gaan we uit van de volgende vooronderstellingen:

a. de opgegeven standaarddeviaties σ_i doen recht aan de meetonzekerheden in y_i,

b. het gebruikte model $f_{\boldsymbol{p}}(x)$ is van toepassing (dat wil zeggen, de functie $f_{\boldsymbol{p}}$ beschrijft het verband tussen x en y adequaat),

c. de metingen $y_1, \cdots, y_n$ zijn onderling statistisch onafhankelijk.

6.2.1 Grafische weergave en aanpassing

Een eerste stap in de analyse van een dergelijk probleem is het weergeven van de waarden $\{y_i, \sigma_i\}$ tegen x_i. Een goed opgezette figuur (zie Sectie 2.5) geeft een gevoel voor het functionele verband en kan ook gebruikt worden voor een "grafische aanpassing": een eerste schatting van de (orde van grootte van) "beste" parameterwaarden $\hat{\boldsymbol{p}}$.

6.2.2 De (gewogen) kwadratische norm

Het uitvoeren van een aanpassing kan alleen als er een betrouwbare *norm* of maat is vastgesteld om de afwijkingen tussen dataset en modelwaarden te kwantificeren. In de meeste situaties wordt daarvoor de *kwadratische norm* $S(\boldsymbol{p})$ gebruikt. Deze $S(\boldsymbol{p})$ is *de gewogen som van de gekwadrateerde afwijkingen* d_i *tussen meetwaarden en modelwaarden* als functie van de te bepalen parameters $\boldsymbol{p}$:

$$\begin{aligned} S(\boldsymbol{p}) &= \sum_{i=1}^{n} w_i d_i^2 \\ &= \sum_{i=1}^{n} w_i \left(y_i - f_{\boldsymbol{p}}(x_i)\right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{y_i - f_{\boldsymbol{p}}(x_i)}{\sigma_i}\right)^2. \end{aligned} \tag{6.3}$$

met

$$w_i = \frac{1}{\sigma_i^2} \tag{6.4}$$

het *gewicht* voor elk punt. De bepalingen die een kleine onzekerheid hebben, krijgen een groter gewicht (dus een grotere waarde w_i) dan waarden met een groter onzekerheid. Onder deze norm wordt het oplossen van het aanpassingsprobleem aangeduid als een "kleinste-kwadraten" minimaliseringsprobleem[4]. Een onderbouwing van juist deze norm wordt gegeven in Appendix A. Met deze definitie begrijpen we de kwadratische norm ook als *de som van de kwadratische afstanden van elk punt tot het model uitgedrukt in de onzekerheid van dat punt*.
Wanneer de σ_i onbekend zijn, kiezen we de w_i gelijk aan 1 en spreken we van een *ongewogen* aanpassing. Wanneer standaarddeviaties σ_i bekend zijn bij de metingen y_i is, wordt een *gewogen* aanpassing uitgevoerd.

Grafische rechte-lijnaanpassing Op het Sketchpad Resource Center[a] wordt de kwadratische norm grafisch weergegeven voor een (ongewogen) rechte-lijnaanpassing. Je kunt hier zelf wat experimenteren met de veranderingen in de waarde van deze norm bij verandering van de fitparameters.

[a] http://www.dynamicgeometry.com/JavaSketchpad/Gallery/Other_Explorations_and_Amusements/Least_Squares.html.

6.2.3 Minimalisatie

Wanneer de meetwaarden $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ en de door de waarnemer opgegeven functie f vaststaan, hangt de waarde van de (gewogen) kwadratische som van afwijkingen tussen data en modelfunctie $S(\boldsymbol{p})$ alleen nog af van de waarden van de parameters $\boldsymbol{p}$. We beschouwen de "aanpassing" nu als optimaal, wanneer $S(\boldsymbol{p})$ minimaal is. Een dergelijke bepaling wordt aangeduid als een *kleinste-kwadraten-aanpassing*. Bij dit (diepste) minimum van $S(\hat{\boldsymbol{p}})$ horen vervolgens "beste" uitkomsten $\hat{\boldsymbol{p}} = (\hat{p}_1, \cdots, \hat{p}_r)$ voor de te bepalen parameters.
De geminimaliseerde waarde $S(\hat{\boldsymbol{p}})$ van $S(\boldsymbol{p})$ wordt ook wel aangeduid als de *chi-kwadraat* (χ^2) van de aanpassing:

$$\chi^2 = \sum_{i=1}^{n} w_i \left(y_i - f_{\hat{\boldsymbol{p}}}(x_i)\right)^2. \tag{6.5}$$

De relevante vraag is nu hoe $S(\boldsymbol{p})$ (analytisch of numeriek) te minimaliseren als functie van de parameterwaarden $\boldsymbol{p} = (p_1, \cdots, p_r)$. We bespreken kort twee methoden.

Methode 1: Gedrag van $S(\boldsymbol{p})$ rond het minimum Een eerste oplossingsmethode maakt gebruik van het feit dat $S(\hat{\boldsymbol{p}})$ een *minimum* is van normfunctie $S(\boldsymbol{p})$. In dat geval zijn *alle eerste afgeleiden* van $S(\boldsymbol{p})$ naar de modelparameters $\boldsymbol{p}$ in het punt $\hat{\boldsymbol{p}}$ gelijk aan nul: $\frac{\partial S}{\partial \boldsymbol{p}}|_{\boldsymbol{p}=\hat{\boldsymbol{p}}} = 0$, oftewel

$$\begin{pmatrix} \frac{\partial S}{\partial p_1} \\ \frac{\partial S}{\partial p_2} \\ \vdots \\ \frac{\partial S}{\partial p_r} \end{pmatrix}_{\boldsymbol{p}=\hat{\boldsymbol{p}}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{6.6}$$

[4] Andere normen zijn ongebruikelijk maar evenwel mogelijk, zoals bijvoorbeeld de som van de *absolute* verschillen $|y_i - f_{\boldsymbol{p}}(x_i)|$ tussen y_i en de modelwaarde $f_{\boldsymbol{p}}(x_i)$.

Voor aanpassingen van eenvoudige vorm (zoals rechte-lijnaanpassingen) is het expliciet oplossen van dit stelsel vergelijkingen zonder veel moeite analytisch uit te voeren. In dat geval kunnen formules worden gegeven voor de te bepalen parameters in termen van de beschikbare data. Zie voor voorbeelden de volgende secties.
In meer complexe situaties is het niet aan te bevelen naar analytische oplossingen te zoeken. Het ligt dan meer voor de hand om het minimaliseringsprobleem, deels of geheel, numeriek op te lossen.

Methode 2: Numerieke minimalisatie van $S(\boldsymbol{p})$ Een alternatieve methode is het numeriek zoeken naar een minimum van $S(\boldsymbol{p})$ door *variatie van de waarden van de modelparameters*. Deze methode is meer algemeen toepasbaar.
Er is een verscheidenheid aan numerieke oplossingsmethoden. De meeste maken gebruik van zogenaamde *steepest descent* routines. Er wordt door de gebruiker een set realistische beginwaarden $\boldsymbol{p}_0$ gekozen voor parameters $\boldsymbol{p}$. De $\boldsymbol{p}_0$ worden veelal gekozen op basis van op voorhand beschikbare informatie, bijvoorbeeld via een grafische weergave of gegevens uit de literatuur.
Vervolgens worden (analytisch of numeriek) de afgeleiden $\frac{\partial S}{\partial \boldsymbol{p}}|_{\boldsymbol{p}=\boldsymbol{p}_0}$ bepaald. Op basis van deze informatie kan afgeschat worden in welke richting $\boldsymbol{\Delta p}$ de parameters moeten worden aangepast voor een zo groot mogelijke verlaging van de waarde van $S(\boldsymbol{p})$. Deze procedure wordt vervolgens opnieuw doorlopen met $\boldsymbol{p_1} = \boldsymbol{p_0} + \boldsymbol{\Delta p}$ als nieuwe startwaarden. Dit heet daarom een *iteratieve* aanpak. De procedure stopt wanneer óf de verandering van $S(\boldsymbol{p})$ óf de verandering van $\boldsymbol{p}$ zodanig klein wordt (dat wil zeggen dat het minimum dusdanig dicht benaderd is) dat verder doorzoeken geen significante verandering meer tot gevolg heeft, óf wanneer een door de gebruiker ingesteld maximum aantal iteratiestappen is bereikt.
Bij deze methodes is de keuze van de beginwaarden $\boldsymbol{p}_0$ voor parameters $\boldsymbol{p}$ in het algemeen zeer belangrijk. Het "landschap" van de S-functie kan bij niet-lineaire aanpassingen een veelvoud aan (lokale) minima bevatten, waarvan er in het algemeen echter maar één globaal minimum is (het allerlaagste punt). Bij een keuze die niet dicht genoeg bij dit globale minimum ligt, bestaat de kans dat de numerieke minimalisatie naar een verkeerd minimum itereert, Figuur 6.7.

Lineaire en niet-lineaire aanpassingen De functionele verbanden waarmee aanpassingen worden gedaan vallen uiteen in twee categorieën: *lineaire* en *niet-lineaire* aanpassingen. Bij een *lineaire* aanpassing is de functie $f_{\boldsymbol{p}}(x)$ lineair in elk van de te bepalen parameters $\boldsymbol{p} = (p_1, \cdots, p_r)$. $S(\boldsymbol{p})$ is dan een kwadratische vorm in termen van deze parameters; in het geval van slechts één enkele parameter p een parabool. Een dergelijke kwadratische vorm heeft één enkel minimum en dus een unieke oplossing. Voor lineaire problemen werkt zowel Methode 1 als Methode 2 feilloos. We behandelen deze categorie verder in Secties 6.3 en 6.4.
Het is goed nogmaals te benadrukken dat de functie lineair is in de *parameters*. In het geval van een lineaire aanpassing mag de variabele x dus zonder bezwaar in niet-lineaire vorm aanwezig zijn in het gegeven functionele verband, zoals bijvoorbeeld van de vorm $f_{(p_1,p_2,p_3)}(x) = p_1 + p_2\sqrt{x} + p_3 \log x$.
Bij *niet-lineaire* aanpassingen is de functie niet-lineair in de parameters $\boldsymbol{p} = (p_1, \cdots, p_r)$. Een voorbeeld is $f_{(p_1,p_2)}(x) = p_1 \cos(p_2 x)$. Bij niet-lineaire aanpassingen bestaat de mogelijkheid dat de functie S meerdere minima heeft bij verschillende waarden voor $\boldsymbol{p}$, waarvan er maar één correspondeert met de "beste" aanpassing. De set vergelijkingen die met behulp van Methode 1 wordt opgesteld heeft dan geen unieke oplossing. Het ligt dan voor de hand om numerieke minimalisatiemethoden (Methode 2) te gebruiken, in combinatie met een goede schatting van de beginwaarde $\boldsymbol{p}_0$ van de parameters. We gaan hier wat dieper op in in Sectie 6.5.

6.2.4 De gereduceerde chi-kwadraat

Wanneer r parameters $(p_1, \cdots, p_r)$ worden bepaald uit n meetwaarden $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ én het model recht doet aan de data én de onzekerheden σ_i realistisch de meetonzekerheden weergeven, dan moet

in de statistische limiet ($n \to \infty$) gelden:

$$S(\hat{\boldsymbol{p}}) \Rightarrow n - r, \tag{6.7}$$

waarbij $n-r$ wordt aangeduid als het aantal *vrijheidsgraden* van het aanpassingsprobleem. Net als bij het gewogen gemiddelde kunnen we nu de *gereduceerde chi-kwadraat* χ^2_{red} definiëren als de χ^2 (zie Vgl. (6.5)) geschaald op het aantal vrijheidsgraden:

$$\chi^2_{red} = \frac{\chi^2}{n-r}. \tag{6.8}$$

$$\Rightarrow 1 \tag{6.9}$$

6.2.5 Kwaliteit van de aanpassing

De gereduceerde chi-kwadraat χ^2_{red} is de parameter die gebruikt wordt om te bepalen of er "iets niet klopt" bij de gegeven dataset. We zullen dit vaak aanduiden met het begrip *consistentie*. Wanneer $\chi^2_{red} \simeq 1$, dan kunnen we concluderen dat de statistische spreiding van de meetdata niet in conflict is met de functie f die we gebruiken voor de aanpassing. Nu is een complicerende factor dat een χ^2_{red} die afwijkt van de waarde 1 ook veroorzaakt kan worden door toevallige effecten. Zeker voor kleine aantallen meetwaarden is de spreiding slecht bepaald; één enkele toevallige uitschieter kan de χ^2_{red} dan flink laten afwijken.
Aan de hand van een waarde van χ^2_{red} kun je dus nooit een absolute uitspraak doen over of een dataset consistent is of niet. Het beste dat je kunt doen is een uitspraak doen over de *kans* dat de dataset consistent is. Als deze kans kleiner wordt dan een bepaalde drempelwaarde (met andere woorden, het wordt wel erg onwaarschijnlijk dat de toevallige ligging van de meetwaarden de afwijking van χ^2_{red} van 1 kan verklaren), zullen we concluderen dat er een niet-statistisch effect in de meting aanwezig is. Het is zaak voor de experimentator om dan op zoek te gaan naar deze oorzaak.

Overschrijdingskans van de gereduceerde chi-kwadraat De bovenstaande analyse moet uiteraard worden gekwantificeerd. Dit kan met behulp van de *overschrijdingskans* van de χ^2_{red}. De aanpak lijkt erg op die van de significantietoets die in Sectie 4.6 is behandeld.
Een verschil is dat in dit geval geen gebruik wordt gemaakt van de normale verdeling, maar van de gereduceerde-chikwadraatverdeling Vgl. (C.2). In Sec. 4.5.3 en Appendix C kan meer informatie gevonden worden over deze verdeling. Belangrijk is dat de breedte van de verdeling sterk afhangt van het aantal vrijheidsgraden $d = n - r$.
De nulhypothese is in dit geval dat de verwachtingswaarde voor χ^2_{red} gelijk is aan 1. Als significantieniveau, de (harde) grens of drempelwaarde van de kans waarbij de situatie van waarschijnlijk naar onwaarschijnlijk overgaat, wordt gebruikelijk 5% gehanteerd. Deze waarde geldt voor de *enkelzijdige overschrijdingskans*.
In tegenstelling tot de consistentietoets met de normale verdeling wordt gebruik gemaakt van enkelzijdige (links- of rechtszijdige) overschrijdingskansen omdat ten eerste de gereduceerde-chikwadraatverdeling asymmetrisch is, en ten tweede de situaties $\chi^2_{red} < 1$ en $\chi^2_{red} > 1$ verschillende praktische consequenties hebben.
Wanneer $\chi^2_{red} > 1$ berekenen we de *rechtszijdige* overschrijdingskans dat de gereduceerde chi-kwadraat $\tilde{\chi}^2_{red}$ groter is dan de gevonden waarde χ^2_{red} voor deze dataset (oftewel verder afligt van 1 dan χ^2_{red}):

$$\text{Prob}(\tilde{\chi}^2_{red} \geq \chi^2_{red}) \quad \text{rechtszijdig}. \tag{6.10}$$

Wanneer $\chi^2_{red} < 1$ berekenen we de *linkszijdige* overschrijdingskans dat de gereduceerde chi-kwadraat $\tilde{\chi}^2_{red}$ kleiner is dan de gevonden waarde χ^2_{red} voor deze dataset:

$$\text{Prob}(\tilde{\chi}^2_{red} \leq \chi^2_{red}) \quad \text{linkszijdig} \tag{6.11}$$

$$\equiv 1 - \text{Prob}(\tilde{\chi}^2_{red} \geq \chi^2_{red}). \tag{6.12}$$

De *linkszijdige* kans is dus gelijk aan 1 minus de *rechtszijdige* kans. De overschrijdingskans is numeriek te bepalen met behulp van *Mathematica*, of af te schatten met behulp van de tabel in Appendix C. In Appendix C vind je daarom een tabel met alleen de *rechtszijdige* overschrijdingskansen voor de gereduceerde chi-kwadraat.
Wanneer de overschrijdingskans groot is, dan betekent dat dat de waarde voor χ^2_{red} statistisch gezien erg waarschijnlijk is, en dat er dus geen reden is om de consistentie in de dataset te betwijfelen. Wanneer de kans erg klein is, dan is er vermoedelijk een extra effect opgetreden.
In de tabel in Appendix C zie je dat de waarde van χ^2_{red} waarvoor het significantieniveau wordt bereikt voor toenemend aantal metingen n steeds dichter bij de waarde 1 komt te liggen. Zo ligt het significantieniveau voor $n = 2$ metingen rechtszijdig bij $\chi^2_{red} = 3$, voor $n = 10$ in de buurt van $\chi^2_{red} = 1.8$ en voor $n = 25$ rond $\chi^2_{red} = 1.5$.
Het smaller worden van de verdeling van de gereduceerde chi-kwadraat met toenemende waarde van d impliceert dat voor meer vrijheidsgraden de consistentietoets steeds strenger wordt. De reden is dat de spreiding van de dataset steeds beter wordt vastgelegd. In de limietsituatie dat het aantal metingen naar oneindig gaat, ken je exact de achterliggende statistiek, en moet χ^2_{red} precies op 1 uitkomen.

Inconsistentie Wanneer χ^2_{red} zodanig sterk afwijkt van 1 dat de overschrijdingskans lager uitvalt dan het gehanteerde significantieniveau (5%), dan is aan minstens één van de vooronderstellingen a. - c. (zie inleiding van deze sectie) niet voldaan. Dat wil zeggen:

1. de opgegeven standaarddeviaties σ_i doen *geen* recht aan de meetonzekerheden in y_i. Het kan bijvoorbeeld zijn dat de onzekerheden bepaald zijn op basis van specificaties van een meetinstrument, maar dat in de praktijk de feitelijke nauwkeurigheid wordt beperkt door allerlei andere bronnen van toevallige onzekerheid; de meetonzekerheid is dan *groter* dan verwacht (en $\chi^2_{red} \ll 1$). Het kan ook zijn dat een onzekerheid *kleiner* is dan in eerste instantie geschat; dit komt vooral voor wanneer niet-statistische methoden gebruikt zijn om de onzekerheid af te schatten (dan $\chi^2_{red} \gg 1$).
2. het gebruikte model $f_{\boldsymbol{p}}(x)$ is *niet* van toepassing. Dat wil zeggen, de functie $f_{\boldsymbol{p}}(x)$ beschrijft het verband tussen x en y *niet* adequaat; er moet een andere functie gebruikt worden om de aanpassing te laten slagen. Een andere mogelijkheid is de aanwezigheid van systematische afwijkingen in de meting.
3. de metingen y_i zijn *niet* statistisch onafhankelijk.
4. óf er is iets anders misgegaan in de procedure. Bijvoorbeeld: een punt of zijn onzekerheid is niet correct bepaald (of ingevoerd), de minimalisatieprocedure voor χ^2 is mislukt.

Bij een waarde van χ^2_{red} die significant afwijkt van 1 is het dus zaak alle gemaakte stappen kritisch te bestuderen!

Grafische rechte-lijnaanpassing De PhET-simulatie http://phet.colorado.edu/sims/curve-fitting/curve-fitting_en.html illustreert mooi de relatie tussen meetresultaten, onzekerheden, fitfunctie en χ^2_{red}.

Dataset zonder bekende onzekerheden Wanneer er geen onzekerheden σ_i in y_i worden opgegeven is alleen een *ongewogen* aanpassing (effectief is dan $\sigma_i = 1$ en dus $w_i = 1$) mogelijk. De χ^2_{red} is dan niet direct te interpreteren als een maat voor de kwaliteit van de aanpassing, omdat de waarde $\sigma_i = 1$ in het algemeen niet representatief is voor de meetonzekerheid. Wel geldt nog steeds dat de χ^2 correspondeert met het minimum van de functie $S(\boldsymbol{p})$.

Herschaling en "achteraf" afschatting van onzekerheden Wanneer géén onzekerheid in de meetpunten voorhanden is, kan de gevonden χ^2_{red} gebruikt worden om met terugwerkende kracht (ook wel *a posteriori*) de onzekerheden in de punten y_i af te schatten door *herschaling* van de originele onzekerheden σ_i met een factor $\sqrt{\chi^2_{red}}$. Voor de onzekerheden na herschaling σ_i' geldt dan

$$\sigma_i' = \sqrt{\chi^2_{red}}\sigma_i. \tag{6.13}$$

Wanneer deze herschaling heeft plaatsgevonden, geldt uiteraard dat $\chi^2_{red} = 1$.
Het is wel belangrijk om na herschaling te beschouwen of de nieuwe waarden σ_i' vanuit experimenteel oogpunt aannemelijk zijn. Wanneer de "correcte" meetonzekerheid na herschaling volgens Vgl. (6.13) een orde van grootte afwijkt van de nauwkeurigheid die verwacht werd op basis van instrumentspecificaties of proefmetingen, dan moet eveneens het lijstje punten op p. 84 worden gecontroleerd.

6.2.6 Covarianties en onzekerheden

De covariantiematrix De *onzekerheden* van de bepaalde parameters volgen uit de *covariantiematrix* van de parameters. De *interne* covariantiematrix wordt voor *lineaire* aanpassingen verkregen door de *tweede afgeleiden* van de functie $S(\boldsymbol{p})$ te bepalen, en te evalueren bij de "beste" waarden $\hat{\boldsymbol{p}}$.[5] Er geldt:

$$C_{\boldsymbol{p},int} = 2\left(\frac{\partial^2 S(\boldsymbol{p})}{\partial \boldsymbol{p}^2}\right)^{-1}\Bigg|_{\boldsymbol{p}=\hat{\boldsymbol{p}}} \tag{6.14}$$

$$= 2\begin{pmatrix} \frac{\partial^2 S}{\partial p_1^2} & \frac{\partial^2 S}{\partial p_1 \partial p_2} & \cdots & \frac{\partial^2 S}{\partial p_1 \partial p_r} \\ \frac{\partial^2 S}{\partial p_2 \partial p_1} & \frac{\partial^2 S}{\partial p_2^2} & \cdots & \frac{\partial^2 S}{\partial p_2 \partial p_r} \\ \vdots & \vdots & \ddots & \\ \frac{\partial^2 S}{\partial p_r \partial p_1} & \frac{\partial^2 S}{\partial p_r \partial p_2} & & \frac{\partial^2 S}{\partial p_r^2} \end{pmatrix}^{-1}\Bigg|_{p_1=\hat{p}_1, p_2=\hat{p}_2, \cdots, p_r=\hat{p}_r}. \tag{6.15}$$

Het is snel in te zien dat voor een model lineair in de parameters $\boldsymbol{p}$ de tweede afgeleiden van de kwadratische norm $S(\boldsymbol{p})$ naar $\boldsymbol{p}$ onafhankelijk zijn van $\boldsymbol{p}$. Voor lineaire aanpassingen bevat de (interne) covariantiematrix $C_{\boldsymbol{p},int}$ (Vgl. (6.15)) dan automatisch numerieke waarden. Dan geldt dat

$$C_{\boldsymbol{p},int} = \begin{pmatrix} \sigma_{p_1}^2 & \sigma_{p_1,p_2} & \cdots & \sigma_{p_1,p_r} \\ \sigma_{p_2,p_1} & \sigma_{p_2}^2 & \cdots & \sigma_{p_2,p_r} \\ \vdots & \vdots & \ddots & \\ \sigma_{p_r,p_1} & \sigma_{p_r,p_2} & & \sigma_{p_r}^2 \end{pmatrix}. \tag{6.16}$$

Zoals in Vgl. (5.42) staan op de hoofddiagonaal van deze matrix de *varianties*, en buiten de diagonaal de *covarianties*, die de correlatie tussen verschillende parameters weergeven. De op deze manier bepaalde waarden voor (co)varianties zijn *interne* waarden, dat wil zeggen onafhankelijk van de spreiding van de data rond de modelfunctie. In de uitwerking voor een rechte lijn in Sectie 6.4 maken we de vorm en eigenschappen van de covariantiematrix wat inzichtelijker.

[5] Op Blackboard is extra informatie te vinden over het berekenen van de covariantiematrix voor niet-lineaire aanpassingen. Kort gezegd wordt dan gekeken naar de doorsnede van de S-functie bij de waarde $\chi^2 + 1$.

Naar verwachting is er een statistische correlatie tussen de bepaalde parameters $\hat{\boldsymbol{p}}$, want ze worden immers afgeleid uit dezelfde dataset. Dit betekent dat de buitendiagonaalelementen van de covariantiematrix significant van nul zullen verschillen. De correlatiecoëfficiënt ρ_{p_i,p_j} is hierbij zinvol om de mate van correlatie tussen parameters te kwantificeren:

$$\rho_{p_i,p_j} = \frac{\sigma_{p_i,p_j}}{\sigma_{p_i}\sigma_{p_j}}. \tag{6.17}$$

Wanneer er propagatie van onzekerheid uitgevoerd moet worden met de bepaalde parameters $\hat{\boldsymbol{p}}$ is het meenemen van deze correlatietermen belangrijk.

Interne en externe onzekerheden Wanneer onzekerheden in de meetwaarden zijn gegeven, zijn er twee mogelijkheden om de onzekerheid in de parameters te berekenen: de (*interne maat*) wordt grotendeels bepaald door de absolute onzekerheden in de meetwaarden, en de (*externe maat*) door de (gewogen) spreiding van de meetpunten rond het model.
Zoals gezegd levert de berekening van (co-)varianties middels Vgl. (6.15) *interne* onzekerheden in parameterwaarden $\hat{\boldsymbol{p}}$. De *externe* onzekerheden zijn *geschaalde* versies van de *interne* onzekerheden. Schaling van de interne covariantiematrix vindt plaats met behulp van χ^2_{red}:

$$C_{\boldsymbol{p},ext} = C_{\boldsymbol{p},int} \times \frac{\chi^2}{n-r} \tag{6.18}$$

$$= C_{\boldsymbol{p},int} \times \chi^2_{red}. \tag{6.19}$$

De interne en externe onzekerheid voor parameter p_i zijn gerelateerd via de wortel van χ^2_{red}:

$$\sigma_{\hat{p}_i,ext} = \sqrt{\chi^2_{red}}\sigma_{\hat{p}_i,int}. \tag{6.20}$$

De eis dat $\chi^2_{red} \simeq 1$ in geval van consistentie is dus ook te vertalen in de eis dat de interne en externe covariantiematrices "ongeveer" aan elkaar gelijk zijn. Wanneer $\chi^2_{red} \ll 1$ is $\hat{\sigma}_{int} > \hat{\sigma}_{ext}$ en is er sprake van een *overschatting* van de meetonzekerheden. Voor $\chi^2_{red} \gg 1$ is juist $\hat{\sigma}_{int} > \hat{\sigma}_{ext}$. Er kan dan sprake zijn van een *onderschatting* van de meetonzekerheden, maar er moet ook rekening worden gehouden met systematische effecten.
Het moge duidelijk zijn dat interne onzekerheden van bepaalde parameters alléén betekenis hebben wanneer gewichten $w_i = 1/\sigma_i^2$ voor metingen y_i gedefinieerd zijn. Wanneer een ongewogen aanpassing wordt gedaan (bijvoorbeeld met alle $w_i = 1$) kunnen onzekerheden van deze parameters alléén worden opgegeven als *externe* onzekerheden. De in notebooks aan de orde komende *Mathematica*-aanpassingsroutines geven daarom steeds de *externe* onzekerheden van bepaalde parameters!

Gebruik interne of externe onzekerheid in rapportage Er zijn voor datasets $\{x_1, x_2, \cdots, x_n\}$ met gespecificeerde onzekerheden σ_i dus twéé onzekerheden (intern en extern) te definiëren. Er komen nu twee vragen op:

1. welk van deze onzekerheden is de "juiste" maat voor de onzekerheid in de beste schatters $\hat{\boldsymbol{p}}$, en
2. hoe interpreteren we een verschil tussen de twee onzekerheden?

De eerste vraag kan beantwoord worden door allereerst te kijken naar de betrouwbaarheid van de onzekerheden σ_i in de meetpunten. Zijn deze afkomstig uit een statistische operatie, dan kan een relatieve nauwkeurigheid van 10% (denk aan Vgl. (3.4)) best bereikt worden.[6] Maar soms worden de meetnauwkeurigheden

[6] Er wordt hier verondersteld dat de standaarddeviaties σ_i, van toevallige aard zijn, en dat deze onzekerheden statistisch onafhankelijk zijn.

σ_i ook *a priori* afgeschat op basis van specificaties van meetinstrumenten en moeten ze mogelijk als minder nauwkeurig worden beoordeeld. Wanneer de σ_i onnauwkeurig worden geacht is de *externe* onzekerheid (die ongevoelig is voor de absolute waarde van de σ_i) de beste maat voor de onzekerheid in $\hat{\boldsymbol{p}}$.

Wanneer de waarden van de σ_i wél betrouwbaar worden geacht, moet die nauwkeurigheid afgewogen worden tegen de informatie afkomstig van de spreiding van de meetpunten rond het model. Je mag verwachten dat de spreiding van meetpunten voor het betreffende experiment het best beschrijft hoe nauwkeurig het model gevolgd wordt, en dus hoe nauwkeurig de parameterwaarden kunnen worden vastgelegd. De vraag is alleen hoe goed de spreiding vastgelegd is of kan worden. Wanneer er bijvoorbeeld 100 punten gemeten zijn, kan de spreiding (denk aan Vgl. (3.4)) goed bepaald worden. Wanneer er slechts drie punten gemeten zijn, is de onzekerheid in de spreiding juist erg hoog. In zo'n geval kan ervoor gekozen worden om de interne onzekerheid als maat te hanteren[7].

Kort gezegd zijn externe onzekerheden dus de beste schatting voor de onzekerheden in $\hat{\boldsymbol{p}}$, tenzij er zwaarwegende redenen zijn om voor de interne onzekerheden te kiezen (bijvoorbeeld een zeer klein aantal meetpunten). Het advies is hoe dan ook om zowel de externe als de interne onzekerheid te rapporteren. Vervolgens kan een korte beschouwing worden toegevoegd omtrent de vraag welk van deze twee onzekerheden de betrouwbaarste maat is.

6.3 Het gewogen gemiddelde: een constante functie $y = A$

We beschouwen een dataset $\{y_1, y_2, \cdots, y_n\}$ met onzekerheden $\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$. Aan deze set doen wij de één-parameter-aanpassing $y = A$, een horizontale lijn met een waarde onafhankelijk van x. Deze sectie heeft twee doelen: de introductie van het *gewogen gemiddelde*, en een eenvoudige illustratie van de tamelijk abstracte theorie uit de vorige sectie. Omdat er sprake is van een aanpassing lineair in de parameter A kunnen wij een analytische methode volgen.

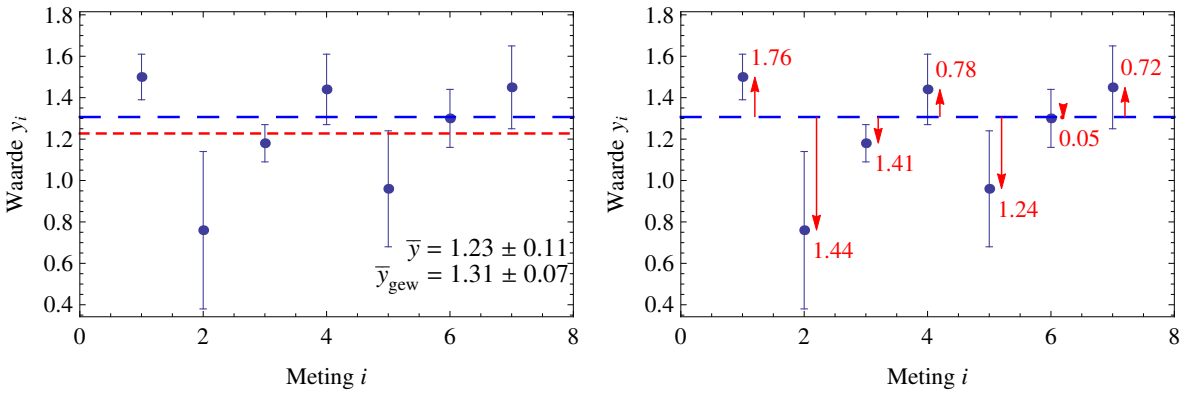
Deze aanpak is een manier om het "gemiddelde" van de dataset te bepalen die beter recht doet aan de gegevens die beter recht doet aan de data dan de standaard middelingsoperatie Vgl. (3.1). Bij de standaardmiddeling heeft σ_i geen rol en tellen de x_i met kleine σ_i (oftewel een hoge betrouwbaarheid) even zwaar mee in het vastleggen van het gemiddelde als de x_i met grote σ_i (lage betrouwbaarheid). Het zal duidelijk zijn dat statistisch gezien dit niet tot een "beste schatter" leidt.

6.3.1 Minimalisatie van de kwadratische norm $S(A)$

We beginnen met het definiëren van de functie $S(A)$ voor de dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$:

$$S(A) \equiv \sum_{i=1}^{n} w_i \left(y_i - A\right)^2 . \tag{6.21}$$

[7] Helaas is het niet zo dat er een algemeen geldend aantal datapunten te definiëren is waarbij we kunnen stellen dat de externe onzekerheid nauwkeuriger bepaald is dan de interne onzekerheid. Afhankelijk van de onzekerheid in de meetnauwkeurigheid en de feitelijk waargenomen spreiding moet dit van geval tot geval (zo mogelijk kwantitatief) bekeken worden.



Figuur 6.2: (a) Illustratie van het meenemen van gewichtsfactoren. De kort-gestreepte lijn is de berekening van het ongewogen gemiddelde $\overline{y}$ met Vgl. (3.1). Wanneer gewogen middeling plaatsvindt, neemt het relatieve belang van metingen 2 en 5 af. Het gewogen gemiddelde $\overline{y}_{gew}$ (de lang-gestreepte lijn) ligt dan ook 0.08 hoger dan $\overline{y}$. Gegeven dat de standaarddeviatie in het gemiddelde 0.11 is, mogen we hier van een aanzienlijke verschuiving spreken. (b) Voor de (gereduceerde) χ^2 is niet de absolute discrepantie tussen meetpunt y_i en gewogen gemiddelde $\overline{y}_{gew}$ relevant, maar de discrepantie in termen van de meetonzekerheid y_i.

De "beste schatting" $\hat{A}$ voor A kan gevonden worden door de eerste orde afgeleide te nemen van S naar A en gelijk te stellen aan nul:

$$\begin{aligned}
\left.\frac{\partial S}{\partial A}\right|_{A=\hat{A}} &= -2\sum_{i=1}^{n} w_i(y_i - \hat{A}) \\
&= 0, \text{ oftewel} \\
\hat{A} &= \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \qquad (6.22) \\
&\equiv \overline{y}_{gew}, \qquad (6.23)
\end{aligned}$$

met w_i als gedefinieerd in Vgl. (6.4). De uitdrukking $\overline{y}_{gew}$ staat bekend als het *gewogen gemiddelde*. Vgl. (6.22) laat zien dat alleen het *relatieve gewicht* van betekenis is: wanneer alle toegekende gewichten w_i met eenzelfde factor worden vermenigvuldigd, blijft $\overline{y}_{gew}$ ongewijzigd. In het bijzondere geval dat alle onzekerheden gelijk zijn ($\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma_y$) reduceert het gewogen gemiddelde tot het ongewogen gemiddelde $\overline{y}$ volgens Vgl. (3.1).

In Figuur 6.2 (a) is geïllustreerd wat de effecten zijn van het meenemen van weging in de berekening van een gemiddelde. Het idee is duidelijk: de punten met een grote onzekerheid liggen naar verwachting verder af van de beste schatter Y voor grootheid y. Als zodanig moeten zij voor de bepaling van die waarde uit een middelingsoperatie minder zwaar tellen.

6.3.2 De interne onzekerheid in een gewogen gemiddelde

Met behulp van de tweede orde afgeleide van S naar A en Vgl. (6.15) kunnen we de *interne* variantie $\sigma^2_{\hat{A},int}$ voor het gewogen gemiddelde $\hat{A} = \overline{y}_{gew}$ vaststellen:

$$\frac{\partial^2 S}{\partial A^2} = 2\sum_{i=1}^{n} w_i; \tag{6.24}$$

$$\begin{aligned}\sigma^2_{A,int} &= 2\left(\frac{\partial^2 S}{\partial A^2}\right)^{-1} \\ &= \frac{1}{\sum_i w_i}; \end{aligned} \tag{6.25}$$

$$\sigma_{A,int} = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}}. \tag{6.26}$$

Deze onzekerheid in $\overline{y}_{gew}$ wordt de *interne* onzekerheid genoemd. De aanduiding *intern* is ontleend aan het feit dat de onzekerheid te bepalen is uit de bekende onzekerheden van punten in de dataset *zonder* de spreiding in de gemeten data y_i te beschouwen. Vgl. (6.26) laat ook zien dat de interne onzekerheid gevoelig is voor de absolute waarden van de meetonzekerheden.
Voor de interne onzekerheid in het gemiddelde $\sigma_{A,int}$ wordt als alle meetresultaten y_i dezelfde onzekerheid σ_y hebben Vgl. (3.6) gevonden:

$$\sigma_{A,int} = \sigma_y/\sqrt{n}. \tag{6.27}$$

6.3.3 De gereduceerde chi-kwadraat

De *chi-kwadraat* of χ^2 is volgens Vgl. (6.28)

$$\chi^2 \equiv \sum_{i=1}^{n}\left(\frac{y_i - \hat{A}}{\sigma_i}\right)^2 \tag{6.28}$$

$$\equiv \sum_{i=1}^{n} w_i \left(y_i - \hat{A}\right)^2. \tag{6.29}$$

De bijdrage van elk meetpunt aan de χ^2 is dus volgens Vgl. (6.28) gelijk aan het kwadraat van de afwijking tussen $\overline{y}_{gew}$ en de meetpunten y_i genormeerd op de onzekerheid σ_i. Zie voor een illustratie Figuur 6.2 (b). Gebruik van de definitie van gewichten Vgl. (6.4) laat zien dat de χ^2 ook gelijk is aan de gewogen som van kwadratische afwijkingen tussen $\overline{y}_{gew}$ en y_i.
Omdat er bij het gewogen gemiddelde sprake is van één aanpassingsparameter is het aantal vrijheidsgraden $n-1$. De *gereduceerde chi-kwadraat* χ^2_{red} is de geschaalde versie van χ^2:

$$\chi^2_{red} \equiv \frac{\chi^2}{n-1} \tag{6.30}$$

$$\equiv \frac{1}{n-1}\sum_{i=1}^{n} w_i \left(y_i - \hat{A}\right)^2. \tag{6.31}$$

6.3.4 De externe onzekerheid in een gewogen gemiddelde

De *externe* onzekerheid wordt berekend als

$$\sigma_{A,ext} \equiv \sqrt{\chi^2_{red}}\sigma_{A,int} \tag{6.32}$$

$$= \sqrt{\frac{1}{n-1}\frac{\sum_{i=1}^{n} w_i \left(y_i - \hat{A}\right)^2}{\sum_{i=1}^{n} w_i}} \tag{6.33}$$

$$\equiv \mathrm{SDOM}_{gew}.$$

De aanduidingen extern geeft aan dat de spreiding "extern" bepaald wordt door in één keer de volledige dataset te bekijken. Aan uitdrukking Vgl. (6.33) is te zien dat alleen de *relatieve gewichten* en de *geconstateerde spreiding* van belang zijn.
In de laatste stap gebruiken we dat de externe onzekerheid geïnterpreteerd kan worden als een *gewogen* versie van de standaarddeviatie van het gemiddelde (SDOM) Vgl. (3.6). Wanneer alle σ_i identiek zijn en gelijk aan σ_y, wordt de "eenvoudige" uitdrukking Vgl. (3.6) voor de ongewogen SDOM $\hat{\sigma}_{\overline{y}}$ teruggevonden. Alhoewel het ongewogen gemiddelde en de SDOM van een dataset $\{y_1, y_2, \cdots, y_n\}$ mathematisch dezelfde resultaten opleveren als het gewogen gemiddelde en de $\mathrm{SDOM}_{\mathrm{gew}}$ van dezelfde dataset met gelijke onzekerheden $\{y_1 \pm \sigma_y, y_2 \pm \sigma_y, \cdots, y_n \pm \sigma_y\}$, is alleen in het laatste geval de interne onzekerheid te bepalen.
Analoog aan Vgl. (3.9) kunnen we afleiden dat met de definitie

$$\overline{y^2}_{gew} = \frac{\sum_{i=1}^{n} \left(w_i y_i^2\right)}{\sum_{i=1}^{n} w_i}, \tag{6.34}$$

geldt dat:

$$\hat{\sigma}^2_{ext} = \frac{1}{n-1}\left(\overline{y^2}_{gew} - \overline{y}^2_{gew}\right). \tag{6.35}$$

6.3.5 Een uitgewerkt voorbeeld

Gesteld, er zijn vijf afzonderlijke bepalingen van een voltage gedaan, met uitkomsten 1.4 ± 0.5 V, 1.2 ± 0.2 V, 1.00 ± 0.25 V, 1.3 ± 0.2 V en 1.0 ± 0.4 V. De opgave is om een beste schatting voor het voltage te vinden, inclusief een bijbehorende onzekerheid. Beschouw hierbij zowel de *interne* als de *externe* onzekerheid. Gegeven is dat de onzekerheden in de gemeten voltages betrouwbaar bepaald zijn, om precies te zijn $\frac{\sigma_{\sigma_i}}{\sigma_i} \leq 1\%$ is. Dit kan bijvoorbeeld wanneer de vijf gevonden waarden voor het voltage elk bepaald zijn uit een uitgebreide meetserie.

Uitwerking Voor een volledige berekening (gewogen gemiddelde, externe en interne onzekerheid, gereduceerde chi-kwadraat) moeten worden geëvalueerd: $\sum_i w_i$, $\sum_i w_i y_i$, $\sum_i w_i y_i^2$ en $\chi^2 = \sum_i w_i (y_i - \overline{y}_{gew})^2$. In tabelvorm worden de resultaten overzichtelijk weergegeven in Tabel 6.1. In Figuur 6.3 zijn de vijf punten met onzekerheid weergegeven, samen met de interne en externe onzekerheden.
De waarde van de gereduceerde chi-kwadraat $\chi^2_{red} \simeq 0.32$ geeft aan dat de meetwaarden zijn gespreid over een kleiner gebied dan de onzekerheden in deze meetwaarden suggereren. Bij zo'n klein aantal punten kan het goed zo zijn dat de externe onzekerheid (als maat voor de spreiding) toevallig zo klein uitvalt.

Tabel 6.1: Rekenschema in tabelvorm voor het bepalen van een gewogen gemiddelde. Merk op dat de laatste kolom pas kan worden gemaakt als het gewogen gemiddelde $\overline{y}_{gew}$ al is berekend. De waarde van χ^2 kan echter al direct berekend worden met behulp van de resultaten $\overline{y^2}_{gew}$ en $\overline{y}_{gew}$ en Vgl. (6.29), zie berekening onder de tabel.

Meting					
y_i	σ_i	$w_i = 1/\sigma_i^2$	$w_i y_i$	$w_i y_i^2$	$w_i \left(y_i - \overline{y}_{gew}\right)^2$
1.4	0.5	4.	5.6	7.84	0.185041
1.2	0.2	25.	30.	36.	0.00568664
1.00	0.25	16.	16.	16.	0.547115
1.3	0.2	25.	32.5	42.25	0.331096
1.0	0.4	6.25	6.25	6.25	0.213717
		$\sum_i w_i = 76.25$	$\sum_i w_i y_i = 90.35$	$\sum_i w_i y_i^2 = 108.34$	$\chi^2 = 1.28266$

$$
\begin{aligned}
\overline{y}_{gew} &= \frac{\sum_i w_i y_i}{\sum_i w_i} &&= 1.18492 \\
\overline{y^2}_{gew} &= \frac{\sum_i w_i y_i^2}{\sum_i w_i} &&= 1.42085 \\
\sigma_{\overline{y}_{gew},int} &= \sqrt{\frac{1}{\sum_i w_i}} &&= 0.11452 \\
\chi^2 &= \left(\overline{y^2}_{gew} - \overline{y}^2_{gew}\right) \sum_i w_i &&= 1.28266 \\
\chi^2_{red} &= \frac{\chi^2}{n-1} &&= 0.32066 \\
\sigma_{\overline{y}_{gew},ext} &= \sqrt{\frac{1}{n-1}\left(\overline{y^2}_{gew} - \overline{y}^2_{gew}\right)} &&\equiv \sigma_{\overline{y}_{gew},int}\sqrt{\chi^2_{red}} \\
& &&= 0.06485
\end{aligned}
$$

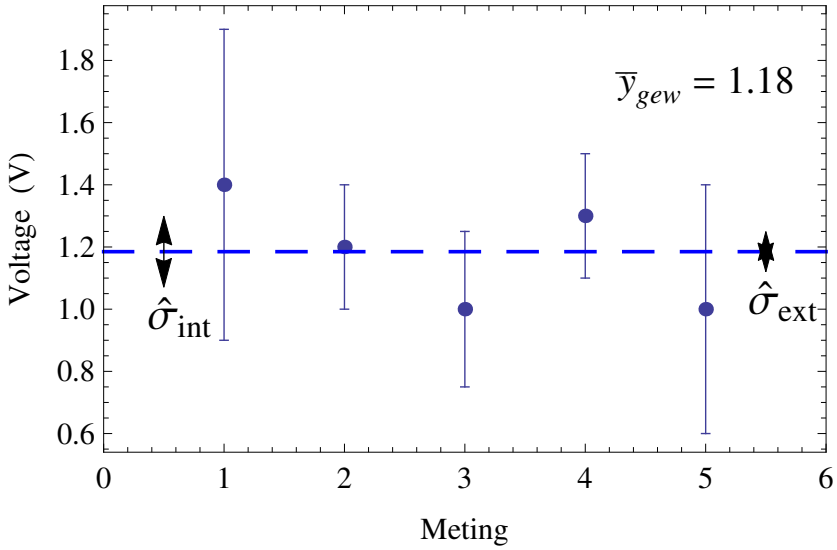
We kwantificeren de waarschijnlijkheid om voor deze dataset de waarde $\chi^2_{red} = 0.32$ te verkrijgen met behulp van de overschrijdingskans voor de gereduceerde chi-kwadraat. Het aantal vrijheidsgraden is in dit geval gelijk aan $d = 4$, immers $d =$ (het aantal datapunten n) $-$ (het aantal parameters) $= 5 - 1$. Verder constateren we dat $\chi^2_{red} = 0.32 < 1$; we dienen dus de *linkszijdige* overschrijdingskans te bepalen.

In de tabel in Appendix C vinden we met behulp van de rij $d = 4$ dat voor $\chi^2_{red} = 0.32$ de (*linkszijdige* overschrijdings)kans ergens tussen $1 - 0.938 = 0.062$ en $1 - 0.809 = 0.191$ ligt. Beide waarden zijn groter dan het significantieniveau $0.05 = 5\%$. De gevonden waarde $\chi^2_{red} = 0.32$ verschilt dus niet significant van de waarde 1, met andere woorden model en data zijn consistent.

We kunnen dus ook niet concluderen dat $\sigma_{\overline{y}_{gew},ext}$ en $\sigma_{\overline{y}_{gew},int}$ strijdig zijn met elkaar. Het lijkt er op dat het verschil tussen interne en externe onzekerheid door statistische effecten wordt verklaard. Gegeven het kleine aantal punten is de spreiding niet scherp bepaald. Wanneer de waarde $\chi^2_{red} = 0.32$ was gevonden voor 20 punten, dan was het resultaat wel erg onwaarschijnlijk geweest (zoek dit zelf na in Appendix C).

De rapportage is nu de laatste stap. We dienen te bepalen of de beste schatter voor het voltage gerapporteerd moet worden met de interne of externe onzekerheid. We merken op dat we hier een keuze hebben omdat beide onzekerheden *niet* significant verschillend zijn, ondanks dat ze wel *wel* numeriek verschillen.

De keuze voor het gebruik van de interne of externe onzekerheid vertaalt nu naar de vraag welke van deze twee onzekerheden *zelf* het nauwkeurigst bepaald is. Voor dit specifieke voorbeeld kun je stellen dat de nauwkeurigheid waarmee de onzekerheid gespecificeerd is ($\frac{\sigma_{\sigma_i}}{\sigma_i} \leq 1\%$) ertoe leidt dat $\sigma_{\overline{y}_{gew},int}$ nauwkeurig bepaald is. Het kleine aantal van vijf datapunten maakt dat de schatter $\sigma_{\overline{y}_{gew},ext}$ voor de (gewogen) spreiding van de data rond het gemiddelde minder goed bepaald is. Vanwege deze redenatie verkiezen we dus om de interne onzekerheid te gebruiken. Uiteindelijk geven wij derhalve als antwoord: $x = 1.18 \pm 0.11$ V.



Figuur 6.3: Vijf voltagemetingen inclusief onzekerheid (punten), en het gewogen gemiddelde $\overline{y}_{gew}$ van deze waarden (stippellijn). Interne en externe onzekerheden in het gewogen gemiddelde zijn met pijlen aangegeven. De kwaliteit van het gewogen gemiddelde van deze dataset wordt uitgedrukt met $\chi^2_{red} = 0.32$.

6.3.6 Nadere analyse van de functie $S(A)$

Voor dit eenvoudige ééndimensionale geval bestuderen we voor nader inzicht de functie S nader. Allereerst herschrijven we $S(A)$ in de vorm van een parabolische functie rond $A = \hat{A}$, met minimum $S(\hat{A})$. Hiervoor definiëren we allereerst $A = \hat{A} + a$. Vervolgens vullen we het rechterlid van deze uitdrukking in in Vgl. (6.21):

$$S(\hat{A}+a) = \sum_i w_i \left(y_i - (\hat{A}+a)\right)^2 \tag{6.36}$$

$$= \sum_i w_i (y_i - \hat{A})^2 + \sum_i w_i a^2 + 2\sum_i w_i y_i a - 2\sum_i w_i \hat{A} a. \tag{6.37}$$

In de eerste somterm herkennen wij $S(\hat{A})$. Verder is met behulp van Vgl. (6.22) voor $\hat{A}$ de derde somterm te herschrijven als $2\hat{A}\sum_i w_i a$. De derde en vierde somterm vallen dan tegen elkaar weg. Tot slot substitueren we $a = (A - \hat{A})$, waarna

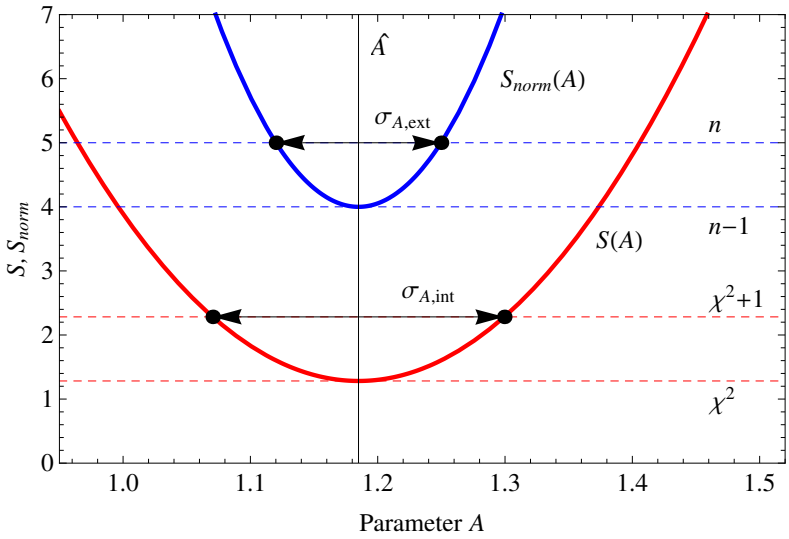
$$S(A) = S(\hat{A}) + \left(A - \hat{A}\right)^2 \sum_i w_i \tag{6.38}$$

$$\equiv \chi^2 + \frac{\left(A - \hat{A}\right)^2}{\sigma^2_{\hat{A},int}}. \tag{6.39}$$

Het blijkt nu dat de *breedte* van de parabolische functie $S(A)$ samenhangt met de interne onzekerheid $\sigma_{A,int}$. Ter illustratie gebruiken we nu de resultaten van het uitgewerkte voorbeeld in deze sectie. Voor deze set data is bepaald $\hat{A} = \overline{x}_{gew} = 1.18$, $\sigma_{A,int} = \hat{\sigma}_{int} = 0.11$, en $\chi^2 = 1.28$. Een plot van de functie $S(A)$ rond het punt $\hat{A}$ is gegeven in Figuur 6.4.
Het blijkt nu na invullen van Vgl. (6.39) dat

$$S(\hat{A} \pm \sigma_{A,int}) = S(\hat{A}) + 1. \tag{6.40}$$

We kunnen het probleem van het bepalen van onzekerheden voor $\hat{A}$ dus herformuleren door te zoeken naar oplossingen van de vergelijking $S(A) = \chi^2 + 1$. De doorsnede van $S(A)$ bij $\chi^2 + 1$ is eveneens aangegeven in Figuur 6.4.



Figuur 6.4: Grafische weergave van de functies $S(A)$ en $S_{norm}(A)$, met daarbij de onzekerheden $\sigma_{A,int}$ en $\sigma_{A,ext}$ in $\hat{A}$. Voor de gebruikte data, zie Figuur 6.3.

Erg inzichtelijk is wat er gebeurt op het moment dat een geschaalde, "externe" versie S_{norm} van de kwadratische norm wordt gedefinieerd door S te normeren op $\chi^2_{red} = \frac{\chi^2}{n-1}$:

$$S_{norm}(A) = \frac{S(A)}{\chi^2_{red}} \tag{6.41}$$

$$= \frac{\chi^2 + \frac{(A-\hat{A})^2}{\sigma_{A,int}}}{\left(\frac{\chi^2}{n-1}\right)} \tag{6.42}$$

$$= (n-1) + \frac{(A-\hat{A})^2}{\sigma^2_{A,ext}}, \tag{6.43}$$

met $\sigma_{A,ext}$ de *externe* onzekerheid in $\hat{A}$. Voor $S_{norm}(A)$ blijkt te gelden dat het minimum gelijk is aan $n-1$, en equivalent aan Vgl. (6.40)

$$S_{norm}(\hat{A} \pm \sigma_{A,ext}) = S_{norm}(\hat{A}) + 1. \tag{6.44}$$

De genormeerde functie $S_{norm}(A)$ en de doorsnijding bij een waarde één boven het minimum $(n-1)$ zijn eveneens weergegeven in Figuur 6.4.

De hierboven geïllustreerde relaties voor deze eenvoudige éénparameter-aanpassing kunnen gegeneraliseerd worden voor meerdere parameters. De relaties Vgl. (6.40) en (6.44) blijken dan eveneens te gelden. Voor twee fitparameters, bijvoorbeeld A en B, beschrijft de relatie $S(A,B) = S(\hat{A},\hat{B}) + 1$ een ellipsvormige doorsnijding van de paraboloïde functie $S(A,B)$. Deze doorsnijding produceert een ellips die *confidence ellipse* wordt genoemd, en die in Sectie 6.6 iets uitgebreider wordt besproken.

6.4 Rechte-lijn aanpassing: $y = A+B\,x$

Als tweede toepassing van de theorie uit Sectie 6.2 werken we het veel voorkomende geval van een lineaire aanpassing aan de functie $y = f_{A,B}(x) = A + Bx$ expliciet uit.[8] De rechte-lijnaanpassing is een bijzonder geval van de bredere set van lineaire aanpassingsproblemen, ook wel aangeduid als *lineaire-kleinste-kwadraten* (LKK) aanpassingen.
Lineaire kleinste-kwadratenaanpassingen hebben twee belangrijke eigenschappen. Ten eerste levert het opstellen van vergelijkingen middels Vgl. (6.6) (Methode 1) een stelsel vergelijkingen dat *lineair* is in de parameters. Dit stelsel vergelijkingen heeft een *unieke* oplossing (behoudens uitzonderingssituaties). Ten tweede is de covariantiematrix $C_{\boldsymbol{p}}$ op te bouwen uit tweede-orde afgeleiden van een kwadratische vorm in de parameters $\boldsymbol{p}$.

6.4.1 Analytische uitwerking van minimalisering voor $y = A + Bx$

Een rechte-lijn relatie tussen x en y van de vorm $y = A + Bx$ is een eenvoudig voorbeeld van een functie lineair in de parameters. De vector van modelparameters is dan gegeven door $\boldsymbol{p} = \{A, B\}$. Om deze modelparameters te bepalen is een passend experiment uitgevoerd met als resultaat de dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$. Omdat we in dit voorbeeld met twee onbekende modelparameters te maken hebben is voor een bepaling zeker vereist dat $n \geqslant 2$. Bij $n = 1$ is het systeem onderbepaald (er zijn oneindig veel lijnen te trekken door één enkel punt). Bij $n = 2$ (met $x_1 \neq x_2$) is het systeem *exact* bepaald: er is maar één oplossing $\hat{A}, \hat{B}$ die precies door beide punten loopt. Voor $n > 2$ hebben we dan echter te maken met een *overgedetermineerd* systeem van relaties (meer vergelijkingen dan onbekenden). Met behulp van de in Sectie 6.2 opgestelde kwadratische norm S is dan met een kleinste-kwadraten aanpassing een beste schatting $\hat{A}, \hat{B}$ voor parameters A en B te verkrijgen. Bovendien kunnen dan ook de onzekerheden van de parameters $\hat{A}, \hat{B}$ worden bepaald.
Het expliciet invullen van de S-functie Vgl. (6.3) met $d_i = y_i - (A + Bx_i)$ leidt nu tot de te minimaliseren functie

$$S(A,B) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left(y_i - (A + Bx_i)\right)^2 \tag{6.45}$$

$$= \sum_{i=1}^{n} w_i \left(y_i - A - Bx_i\right)^2. \tag{6.46}$$

Hierbij zijn gewichten w_i als voorheen vastgelegd als $1/\sigma_i^2$.

Minimalisatie van $S(A,B)$ De minimalisatie van $S(A,B)$ kan analytisch worden uitgevoerd via Vgl. (6.6). Het minimum van $S(A,B)$ wordt gevonden door te eisen dat de eerste orde afgeleiden van $S(A,B)$ naar de parameters A en B gelijk zijn aan nul:

$$\frac{\partial S(A,B)}{\partial A} = 0, \text{ en} \tag{6.47}$$

$$\frac{\partial S(A,B)}{\partial B} = 0. \tag{6.48}$$

[8] De uitwerkingen voor rechte-lijn twee-parameteraanpassingen kunnen worden gegeneraliseerd, als de dataset maar van de vorm $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ is. Een korte uitwerking voor de algemene set functies $y = Af(x) + Bg(x)$ is te vinden in het extra materiaal op Blackboard.

Na expliciete uitvoering van deze differentiatie resteert een stelsel van twee vergelijkingen in A en B in termen van de meetgegevens $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$:[9]

$$A\sum_i w_i + B\sum_i w_i x_i = \sum_i w_i y_i, \tag{6.49}$$

$$A\sum_i w_i x_i + B\sum_i w_i x_i^2 = \sum_i w_i x_i y_i. \tag{6.50}$$

In matrix-vectornotatie wordt dit stelsel vergelijkingen

$$\begin{pmatrix} \sum_i w_i & \sum_i w_i x_i \\ \sum_i w_i x_i & \sum_i w_i x_i^2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum_i w_i y_i \\ \sum_i w_i x_i y_i \end{pmatrix}. \tag{6.51}$$

De oplossing levert "beste" parameterwaarden $\hat{A}$ en $\hat{B}$

$$\hat{A} = \left(\frac{1}{\Delta}\right)\left(\sum_i \left(w_i x_i^2\right)\sum_i \left(w_i y_i\right) - \sum_i \left(w_i x_i\right)\sum_i \left(w_i x_i y_i\right)\right), \tag{6.52}$$

$$\hat{B} = \left(\frac{1}{\Delta}\right)\left(\sum_i \left(w_i\right)\sum_i \left(w_i x_i y_i\right) - \sum_i \left(w_i x_i\right)\sum_i \left(w_i y_i\right)\right), \text{ waarbij} \tag{6.53}$$

$$\Delta = \sum_i \left(w_i\right)\sum_i \left(w_i x_i^2\right) - \left(\sum_i \left(w_i x_i\right)\right)^2. \tag{6.54}$$

De parameter Δ is gelijk aan de determinant van de matrix in Vgl. (6.51). De gevonden oplossing voor $\hat{A}$ en $\hat{B}$ is uniek mits Δ ongelijk is aan nul.

De chi-kwadraat Door invullen van beste schatters $\hat{A}$ en $\hat{B}$ in Vgl. 6.46 vinden we

$$\chi^2 \equiv S\left(\hat{A}, \hat{B}\right) \tag{6.55}$$

$$= \sum_{i=1}^{n} w_i \left(y_i - \hat{A} - \hat{B}x_i\right)^2. \tag{6.56}$$

Deze uitdrukking kan nog iets worden vereenvoudigd tot [10]

$$\chi^2 = \sum_i \left(w_i y_i^2\right) - \hat{A}\sum_i \left(w_i y_i\right) - \hat{B}\sum_i \left(w_i x_i y_i\right). \tag{6.57}$$

[9] Verifieer zelf deze uitdrukkingen door het kwadraat in Vgl. (6.46) uit te schrijven en de uitdrukking partieel te differentiëren naar parameters A en B.

[10] Allereerst werken we de kwadratische term uit tot

$$\chi^2 = \sum_i w_i y_i^2 - \sum_i w_i y_i(\hat{A} + \hat{B}x_i) - \sum_i w_i y_i(\hat{A} + \hat{B}x_i) + \sum_i w_i(\hat{A} + \hat{B}x_i)^2.$$

De derde somterm kan met behulp van Vgl. (6.49) - (6.50) herschreven worden tot

$$\begin{aligned} \sum_i w_i y_i(\hat{A} + \hat{B}x_i) &= \hat{A}\sum_i w_i y_i + \hat{B}\sum_i w_i y_i x_i \\ &= \hat{A}\sum_i w_i(\hat{A} + \hat{B}x_i) + \hat{B}\sum_i w_i(\hat{A} + \hat{B}x_i)x_i \\ &= \sum_i w_i(\hat{A} + \hat{B}x_i)(\hat{A} + \hat{B}x_i). \end{aligned}$$

De derde term blijkt op een minteken na gelijk te zijn aan de vierde somterm; ze vallen derhalve tegen elkaar weg.

Aan de hand van Vgl. (6.9) voor $r = 2$ parameters verwachten wij dat

$$\chi^2 \approx n - 2. \tag{6.58}$$

De gereduceerde χ^2 voor een twee-parameter-aanpassing is

$$\chi^2_{red} = \frac{\chi^2}{n-2}. \tag{6.59}$$

Voor een gewogen aanpassing, waarbij de σ_i in y_i bekend zijn, is nu te evalueren of χ^2_{red} voldoende dicht in de buurt ligt van 1 middels de overschrijdingskans van de gereduceerde chi-kwadraat. Voor een vergelijking met het gekozen significantieniveau kan bepaald worden of het model (rechte lijn) strijdig is met de metingen.
Wanneer géén onzekerheden σ_i in y_i waren opgegeven, kan alleen een ongewogen aanpassing worden uitgevoerd. De waarde van χ^2_{red} kan dan niet direct gebruikt worden voor evaluatie van de kwaliteit van de aanpassing. Wel kan na afloop met behulp van Vgl. (6.13) berekend worden wat de onzekerheid σ_i' in y_i moet zijn om acceptabele overeenstemming te krijgen. Vervolgens rest dan nog de taak om te controleren of deze onzekerheden op experimentele gronden te verdedigen zijn.

6.4.2 De covariantiematrix en de correlatiecoëfficiënt ρ

De interne covariantiematrix Als eerder opgemerkt wordt een rechte-lijnverband nooit exact gereproduceerd in metingen vanwege de statistische spreiding van y_i ten gevolge van de meetonzekerheden σ_i. Derhalve kennen $\hat{A}$ en $\hat{B}$ ook onzekerheden. Hoewel de metingen metingen y_i met onzekerheden σ_i beschouwd kunnen worden als *statistisch onafhankelijk* van elkaar, zijn de bepalingen van A en B niet zonder meer statistisch onafhankelijk van elkaar; ze zijn tenslotte beide ontleend aan dezelfde dataset. We verwachten dan ook een correlatiecoëfficiënt (en dus een covariantie) ongelijk aan nul.
De *interne covariantiematrix* $C_{\boldsymbol{p},int}$ van de parameters $\boldsymbol{p} = (A,B)$ in termen van de *tweede afgeleiden* van de functie $S(A,B)$ is

$$C_{\boldsymbol{p},int} = 2 \begin{pmatrix} \frac{\partial^2 S(A,B)}{\partial A^2} & \frac{\partial^2 S(A,B)}{\partial A \partial B} \\ \frac{\partial^2 S(A,B)}{\partial B \partial A} & \frac{\partial^2 S(A,B)}{\partial B^2} \end{pmatrix}^{-1} \Bigg|_{A=\hat{A},B=\hat{B}} \tag{6.60}$$

$$= \begin{pmatrix} \sigma^2_{A,int} & \sigma_{A,B,int} \\ \sigma_{B,A,int} & \sigma^2_{B,int} \end{pmatrix}. \tag{6.61}$$

Expliciete uitvoer van de dubbele differentiatie en matrixinversie levert

$$\sigma^2_{A,int} = \left(\frac{1}{\Delta}\right) \sum_i \left(w_i x_i^2\right), \tag{6.62}$$

$$\sigma^2_{B,int} = \left(\frac{1}{\Delta}\right) \sum_i w_i, \tag{6.63}$$

$$\sigma_{A,B,int} = \sigma_{B,A,int} = \left(\frac{1}{\Delta}\right) \sum_i w_i x_i. \tag{6.64}$$

We merken hier op dat de onzekerheden *interne* onzekerheden zijn. Dit is in de uitdrukkingen Vgl. (6.62) - (6.63) voor σ_A en σ_B te zien omdat nergens de meetwaarden y_i voorkomen. Met andere woorden, deze schattingen σ_A en σ_B zijn puur gebaseerd op de aan de metingen toegekende onzekerheden σ_i (en *niet* op de geconstateerde spreiding van de metingen y_i t.o.v. de "beste" aanpassing $A + Bx_i$).

De interne onzekerheden $\sigma_{A,int}$ en $\sigma_{B,int}$ hebben alleen betekenis wanneer onzekerheden σ_i worden opgegeven bij y_i. Zonder deze onzekerheden kan alleen een *ongewogen* aanpassing (met $w_i = 1$) worden gemaakt. De onzekerheden in $\hat{A}$ en $\hat{B}$ kunnen dan alleen worden opgegeven als *externe* onzekerheden $\sigma_{A,ext}$ en $\sigma_{B,ext}$ (zie volgende paragraaf), ervan uitgaand dat het model gerechtvaardigd is.

Externe onzekerheden in $\hat{A}$ en $\hat{B}$ We willen uiteraard ook de *feitelijke spreiding* van de meetwaarden rond de gevonden beste rechte-lijn aanpassing in het spel brengen. Met behulp van Vgl. (6.19) formuleren we de *externe* onzekerheden in de bepaalde parameterwaarden $\hat{A}$ en $\hat{B}$:

$$\sigma_{A,ext} = \sqrt{\chi^2_{red}}\sigma_{A,int}, \tag{6.65}$$

$$\sigma_{B,ext} = \sqrt{\chi^2_{red}}\sigma_{B,int}. \tag{6.66}$$

Expliciet invullen van χ^2_{red} en $\sigma_{A,int}$ respectievelijk $\sigma_{B,int}$ laat zien dat $\sigma_{A,ext}$ en $\sigma_{B,ext}$ ongevoelig zijn voor een schaling van de gewichten w_i met een zekere vermenigvuldigingsfactor. Oftewel: indien de onzekerheden σ_i voor alle punten y_i min of meer gelijk zijn, is het *niet* noodzakelijk om gewichten op te geven voor een schatting van de *externe* onzekerheden in de parameters.
De aanwezigheid van de term $n-2$ in de uitdrukking voor χ^2_{red} is wellicht beter te begrijpen in het kader van Vgl. (6.65) - (6.66). Wanneer we slechts twee meetpunten zouden hebben, dan liep een rechte-lijnaanpassing exact door deze punten. In dat geval is de waarde van χ^2 gelijk aan nul, en blijft χ^2_{red} onbepaald. In dat geval blijven ook de externe onzekerheden onbepaald.

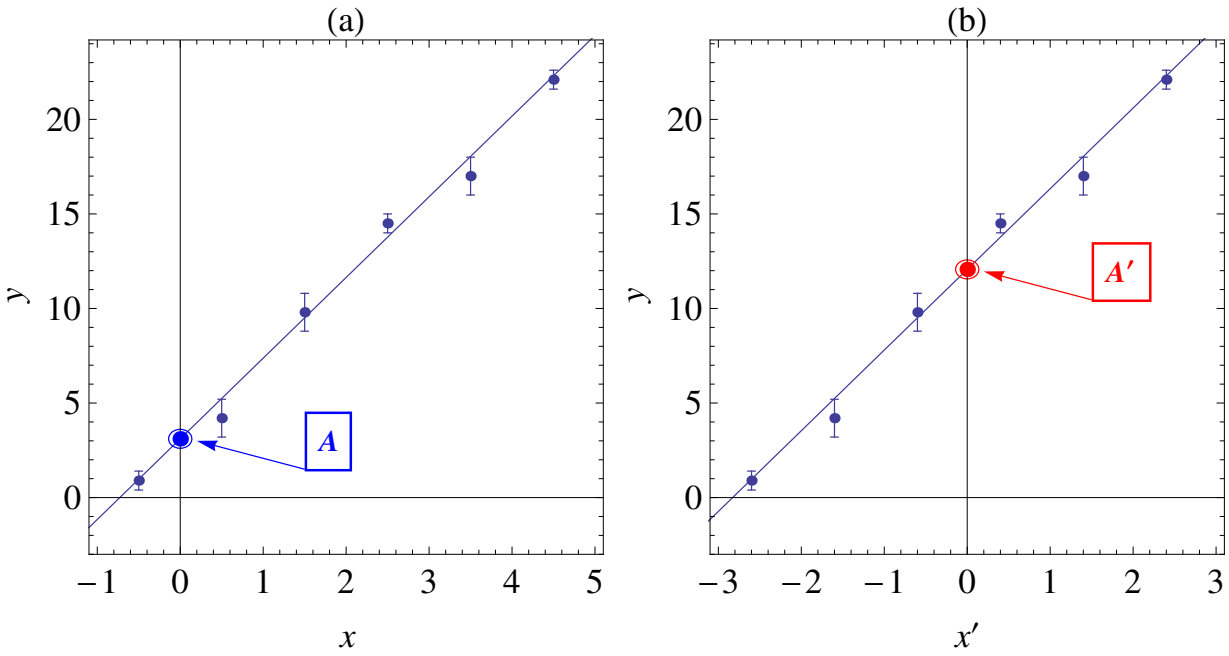
De correlatiecoëfficiënt De *correlatiecoëfficiënt* $\rho_{A,B}$ ($-1 \leqslant \rho_{A,B} \leqslant 1$) van de parameters A, B is gelijk aan

$$\rho_{A,B} \equiv \frac{\sigma_{A,B}}{\sigma_A \sigma_B} \tag{6.67}$$

$$= -\frac{\sum w_i x_i}{\sqrt{\sum w_i \sum w_i x_i^2}}. \tag{6.68}$$

Het maakt hier uiteraard niet uit of interne of externe onzekerheden worden gebruikt. De betekenis van deze correlatie voor een rechte lijn is geïllustreerd in Figuur 6.5 (a). De helling $\hat{B}$ "kantelt" om het zwaartepunt $(\overline{x}_{gew}, \overline{y}_{gew})$ van de metingen, met $\overline{x}_{gew} \equiv \sum w_i x_i / \sum w_i$ en $\overline{y}_{gew} = \frac{\sum w_i y_i}{\sum w_i}$. Een variatie in $\hat{B}$ zal dan doorwerken in de asafsnede $\hat{A}$. Met andere woorden, A en B zijn *statistisch afhankelijk*. In Vgl. (6.68) is dit ook te zien. De sommaties in de noemer zijn per definitie positief. De teller $\sum w_i x_i$, die evenredig is met $\overline{x}_{gew}$, kan positief, negatief of gelijk aan nul uitvallen, en is daarmee een maat voor de sterkte van de correlatie. Met de mate van correlatie tussen parameters moet uiteraard rekening gehouden worden bij propagatie van onzekerheden.
Statistische afhankelijkheid hangt af van de vorm van het probleem, zoals het volgende voorbeeld laat zien. Wanneer de x_i symmetrisch om de oorsprong liggen (in de gewogen zin, dat wil zeggen $\sum w_i x_i = 0$) zijn de beide parameters $\hat{A}, \hat{B}$ vanzelf al statistisch van elkaar *onafhankelijk*. Met andere woorden, we kunnen $\hat{A}, \hat{B}$ onafhankelijk "maken" door alle x-waarden te verminderen met $\overline{x}_{gew}$ en dan een nieuwe aanpassing maken aan metingen $x' = x - \overline{x}_{gew}$ en y:

$$\begin{aligned} y &= A + Bx \\ &= (A + B\overline{x}_{gew}) + Bx' \\ &\equiv A' + B'x'. \end{aligned}$$



Figuur 6.5: (a) Een rechte-lijn aanpassing $y = A + Bx$ resulteert in het algemeen in *statistische correlatie* tussen de bepaalde parameters $\hat{A}$ en $\hat{B}$. Een kleine verandering in helling B correspondeert dan automatisch met een verandering in asafsnede A. (b) Een verschuiving van de x-as naar de x'-as met $x' = x - \overline{x}_{\text{gew}}$, met $\overline{x}_{\text{gew}}$ het gewogen gemiddelde van de x-waarden heft deze correlatie op. In dat geval zijn $\hat{A}'$ en $\hat{B}$ statistisch ontkoppeld.

Tabel 6.2: Rekenschema voor het bepalen van een gewogen lineaire aanpassing van de vorm $y = a + bx$. Afwijkingen $d_i = y_i - (\hat{a} + \hat{b}x_i)$ van de beste rechte lijn in de laatste kolom worden *achteraf* berekend.

Meting			**Aanpassing**						
x_i	y_i	σ_i	$w_i = 1/\sigma_i^2$	$w_i x_i$	$w_i x_i^2$	$w_i x_i y_i$	$w_i y_i$	$w_i y_i^2$	$w_i d_i^2$
0.	0.9	0.5	4.0	0.0	0.0	0.0	3.6	3.24	0.022
1.	4.2	1.0	1.0	1.0	1.0	4.2	4.2	17.64	1.083
2.	9.8	1.0	1.0	2.0	4.0	19.6	9.8	96.04	0.086
3.	14.5	0.5	4.0	12.0	36.0	174.0	58.0	841.0	2.113
4.	17.0	1.0	1.0	4.0	16.0	68.0	17.0	289.0	1.080
5.	22.1	0.5	4.0	20.0	100.0	442.0	88.4	1953.64	0.169
		Totalen	15.0	39.0	157.0	707.8	181.0	3200.56	$\chi^2 = 4.553$

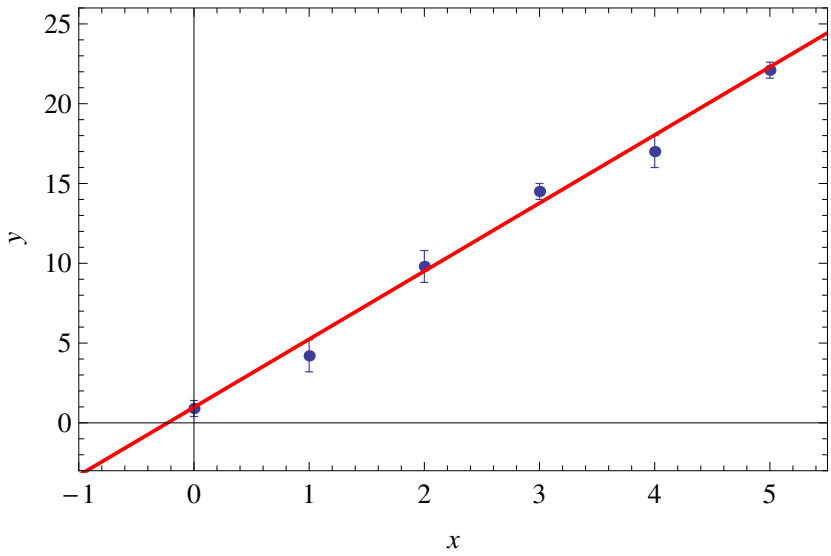
$$
\begin{aligned}
\Delta &= (\textstyle\sum_i w_i)\sum_i \left(w_i x_i^2\right) - \left(\sum_i w_i x_i\right)^2 && = && 834.0 \\
\hat{a} &= \left(\textstyle\sum_i w_i x_i^2 \sum_i w_i y_i - \sum_i w_i x_i \sum_i w_i x_i y_i\right)/\Delta && = && 0.97458 \\
\hat{b} &= \left(\textstyle\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i\right)/\Delta && = && 4.26619 \\
\sigma_{a,int} &= \sqrt{\textstyle\sum_i \left(w_i x_i^2\right)/\Delta} && = && 0.434 \\
\sigma_{b,int} &= \sqrt{\textstyle\sum_i (w_i)/\Delta} && = && 0.134 \\
\rho_{a,b} &= -\frac{\sum_i w_i x_i}{\sqrt{\sum_i w_i \sum_i w_i x_i^2}} && = && -0.804 \\
\chi^2 &= \textstyle\sum_i w_i y_i^2 - \hat{a}\sum_i w_i y_i - \hat{b}\sum_i w_i x_i y_i && \approx && 4.55 \\
\chi^2_{red} &= \chi^2/(n-2) && \approx && 1.07 \\
\sigma_{a,ext} &= \sigma_{a,int}\sqrt{\chi^2_{red}} && = && 0.463 \\
\sigma_{b,ext} &= \sigma_{b,int}\sqrt{\chi^2_{red}} && = && 0.143
\end{aligned}
$$

De herziene situatie is te zien in Figuur 6.5 (b). De helling van deze aanpassing ($\hat{B}'$) is met deze verschuiving uiteraard onveranderd ($\hat{B}' = \hat{B}$). De waarde van $\hat{A}'$ is uiteraard wel anders dan $\hat{A}$, namelijk $A' = A + B\overline{x}_{gew}$ $\left(= \frac{\sum w_i y_i}{\sum w_i} = \overline{y}_{gew}\right)$. Een verandering in helling $\hat{B}'$ vereist nu niet meer dat de asafsnede $\hat{A}'$ verandert. De parameter $\hat{A}'$ is nu *statistisch onafhankelijk* van B'.

6.4.3 Rechte-lijnaanpassing: een uitgewerkt voorbeeld

Voor een rechte-lijnaanpassing van de vorm $A + Bx$ zijn in het voorgaande *formules* verkregen onder de kleinste-kwadraten norm. In eenvoudige situaties (met minder dan 10 datapunten) kan gebruik worden gemaakt van deze formules met *uitwerking in tabelvorm* (handmatig, desnoods op papier). Voor praktische uitwerking van aanpassingen aan grotere datasets kan een *uitwerking in formulevorm* worden geprogrammeerd.
Hier volgt een voorbeelduitwerking in tabelvorm. We voeren een experiment uit waarbij meetresultaten y_i met onzekerheden σ_i worden bepaald als functie van een instelling x_i; resultaten zijn weergegeven in Tabel 6.2. De onzekerheid in de opgegeven instelwaarden x_i is op deze schaal verwaarloosbaar klein. Het is bekend dat er een lineaire relatie $y = a + bx$ geldt voor deze waarnemingen. We gebruiken een gewogen aanpassing om "beste" waarden voor $\hat{a}$ en $\hat{b}$ vast te leggen inclusief de aan deze bepalingen toe te kennen onzekerheid.

Uitwerking Een goede eerste stap is altijd het maken van een plot van de data. We geven de data grafisch weer in Figuur 6.6. Op het eerste gezicht is een lineair verband te verdedigen.

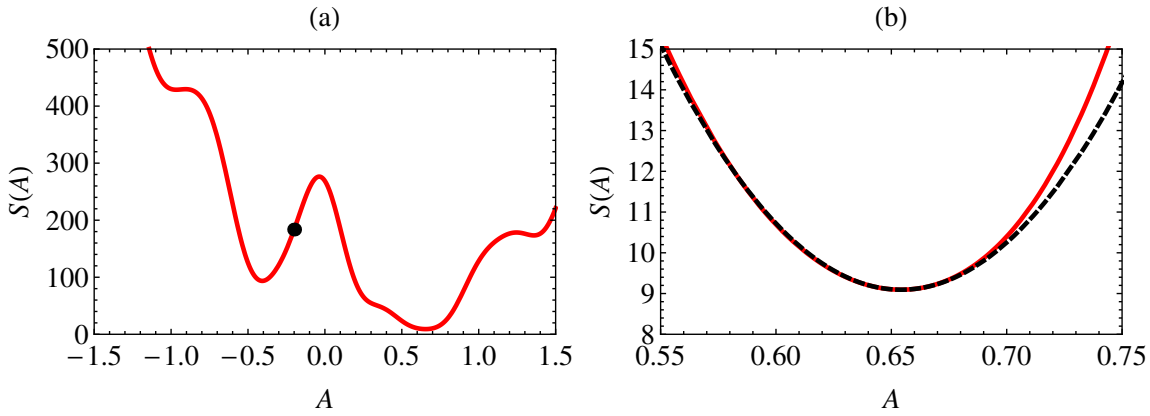


Figuur 6.6: Gemeten data (punten) en beste aanpassing $\hat{a}+\hat{b}x$ zoals bepaald met behulp van Tabel 6.2 (lijn). Bij deze aanpassing vinden we een gereduceerde chi-kwadraat $\chi^2_{red} \simeq 1.07$.

Voor puntenparen $\{\{x_1,y_1,\sigma_1\},\cdots,\{x_n,y_n,\sigma_n\}\}$ moeten voor een dergelijke aanpassing worden geëvalueerd: $\sum_{i=1}^n w_i x_i$, $\sum_{i=1}^n w_i x_i^2$, $\sum_{i=1}^n w_i y_i$, $\sum_{i=1}^n w_i x_i y_i$ en $\sum_{i=1}^n w_i y_i^2$, met $w_i = 1/\sigma_i^2$. Door deze waarden in een tabel uit te zetten kan ook op papier overzichtelijk een aanpassing gedaan worden. Een dergelijk rekenschema is gegeven in Tabel 6.2. Onder de tabel zijn de waarden voor beste schatters, interne en externe onzekerheden weergegeven.
De "beste aanpassing" $y = \hat{a}+\hat{b}x$ is als lijn weergegeven in Figuur 6.6. Deze lijn lijkt goed door de gemeten waarden te lopen. We kunnen "goed" kwantificeren aan de hand van de waarde van $\chi^2_{red} \simeq 1.07$ en het aantal vrijheidsgraden $n-r = 6-2 = 4$. Met behulp van Tabel C.1 in Appendix C schatten we af dat de rechtszijdige overschrijdingskans ligt tussen 0.406 en 0.308, oftewel veel hoger dan het gangbare significantieniveau $P = 0.05$. Er is dus geen reden om aan te nemen dat de beste aanpassing $y = \hat{a}+\hat{b}x$ strijdig is met de meetdata.

6.5 Niet-lineaire aanpassingen

In het oplossen van een aanpassingsprobleem staat het minimaliseren van de functie $S(\boldsymbol{p})$ centraal. In dit hoofdstuk zijn tot dusver alleen eenvoudige aanpassingen (*lineair* in $\boldsymbol{p}$) beschouwd die analytisch oplosbaar bleken via de oplossing van het stelsel Vgl. (6.6) (Methode 1). Hieruit volgden waarden voor de parameters en hun onzekerheden.
Bij het oplossen van een *niet-lineair* aanpassingsprobleem is de situatie veel minder rooskleurig. In dergelijke situaties zijn de te bepalen parameters in *niet-lineaire* vorm aanwezig in het op te lossen stelsel Vgl. (6.6). Meestal bestaan er dan geen analytische oplossingen. In dat geval is het vaak sneller om via Methode 2 (*iteratief*) naar een minimalisering van de functie S toe te werken.
Routines waarmee niet-lineaire aanpassingen kunnen worden uitgevoerd hebben dan ook de mogelijkheid *startwaarden* voor de te bepalen parameters op te geven van waaruit de iteratie wordt gestart. Hierbij is het van groot belang dat deze startwaarden *goed* worden gekozen. Schattingen voor geschikte startwaarden volgen bijvoorbeeld uit: een verwachte waarde op basis van de literatuur, een waarde bepaald in een eerder experiment, of een eerste schatting op basis van een grafische aanpassing. Het kiezen van startwaarden die zo dicht mogelijk bij de "beste waarden" van fitparameters liggen, zorgt ervoor dat de routine sneller een



Figuur 6.7: (a) Voorbeeld S-functie voor een niet-lineaire aanpassing. De functie heeft meerdere minima, waarvan de meeste lokale minima zijn. De keuze van een verkeerde startwaarde (zoals $A = -0.2$, als een punt weergegeven in de figuur) geeft een iteratie naar het verkeerde minimum $A \simeq -0.41$. Voor iteratie naar het globale minimum bij $A \simeq 0.65$ is het vastleggen van de startwaarde tussen ruwweg 0.0 en 1.2 vereist. (b) Nabij het globale minimum is de functie bij benadering parabolisch (stippellijn). Dat betekent dat een onzekerheid kan worden toegekend volgens (bij benadering) dezelfde aanpak als voor lineaire aanpassingen.

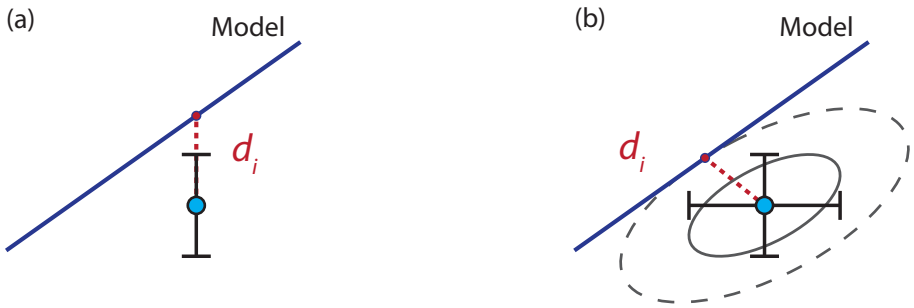
eindresultaat bereikt, en ook dat de kans dat de routine naar een verkeerd minimum (of zelfs helemaal niet) convergeert zo klein mogelijk wordt gehouden. Zie voor een illustratie van dit feit Figuur 6.7 (a).
Voor niet-lineaire aanpassingen is de S-functie niet exact parabolisch. Gelukkig kan in de meeste gevallen het gedrag rond het minimum benaderd worden door een parabool (Figuur 6.7 (b)). Dat betekent dat de concepten voor de berekening van onzekerheden voor lineaire aanpassingen met enige wijzigingen over te zetten zijn naar niet-lineaire aanpassingen.
In het werkcollege wordt uitgebreid aandacht besteed aan het uitvoeren van niet-lineaire aanpassingen met behulp van *Mathematica*.

6.6 Complexere aanpassingen

De vaardigheid tot uitvoeren van lineaire en niet-lineaire aanpassingen is voor een fysicus nuttige bagage om in vaak voorkomende gevallen en experimenten een (fysisch) model aan te passen aan meetgegevens. Het zal je echter niet verbazen dat de behandelde theorie in dit hoofdstuk in veel andere gevallen tekortschiet. Om een idee te geven van situaties waar complexere aanpassingsroutines ingezet moeten worden, bespreken we twee situaties.

6.6.1 Onzekerheden in de instelparameter x

We zijn er tot nog toe van uitgegaan dat de onzekerheid in de onafhankelijke variabele x verwaarloosd kon worden. Echter, ook het vaststellen van de instelparameter x is gekoppeld aan een meetproces (m.a.w. de rol van x en y is in principe gelijkwaardig in de analyse). Dit maakt al duidelijk dat het in de behandeling gemaakte scherpe onderscheid tussen x-waarden en y-waarden in feite niet bestaat: zowel de ene als de andere is een statistische grootheid en daarom behept met onzekerheid.
Wanneer zowel in de x- als in de y-richting onzekerheden moeten worden verdisconteerd laten alle gangbare aanpassingsroutines binnen *Mathematica* het afweten. In het MFpackage is een tweetal routines opgenomen, `LineFit2` en `GenFit2`, waarmee dergelijke aanpassingen kunnen worden uitgevoerd.



Figuur 6.8: Concept van het aanpassen van een model aan data met (a) alleen een onzekerheid σ_{y_i} en (b) met onzekerheden (σ_{x_i},σ_{y_i}) in de gevonden (gecorreleerde) (x_i, y_i)-waarde. Het zoeken naar het best passend model gaat over het minimaliseren van de (totale) afstand d_i voor alle punten (x_i, y_i) samen. Situatie (a) is uitvoerig in dit hoofdstuk besproken. Het minimaliseren bij situatie (b) is aanzienlijk complexer, aangezien de afstand d_i volgt uit de positie waar de ellips raakt aan de lijn en dat dit punt verschuift voor andere parameterwaarden.

Bij aanpassingen waarbij ook een onzekerheid in de onafhankelijke variabele is, komt een nieuw element van complexiteit aan de orde, namelijk *hoe definieer ik de "afstand" tussen een meetpunt en de aan te passen functie als aan het meetpunt in beide richtingen een onzekerheid is toegekend*? De situatie is geschetst in Figuur 6.8. Wanneer er uitsluitend een onzekerheid is in de y-richting, is het voldoende om voor elk punt i de verticale afstand uit te drukken in de onzekerheid σ_{y_i}. In het geval van een onzekerheid $(\sigma_{x_i}, \sigma_{y_i})$ in beide richtingen kun je rondom elk punt een ellips definiëren. Deze zogenaamde *confidence ellipse* voldoet aan de vergelijking

$$\frac{(x-x_i)^2}{\sigma_{x_i}^2} - 2\rho\frac{(x-x_i)(y-y_i)}{\sigma_{x_i}\sigma_{y_i}} + \frac{(y-y_i)^2}{\sigma_{y_i}^2} = r_i^2(1-\rho^2). \tag{6.69}$$

De statistische correlatie tussen x_i en y_i ($\rho \neq 0$) zorgt ervoor dat de hoofdassen van de ellips gedraaid liggen ten opzichte van de x- en y-as. Voor de ellipsvergelijking met $r_i = 1$ hebben de hoofdassen de lengtes $2\sigma_{x_i}$ en $2\sigma_{y_i}$. Net als voor één variabele bakent de ellips bij $r_i = 1$ een gebied af waarbinnen een vaste fractie (die overigens niet gelijk is aan 68%) van de metingen te verwachten is. Meer informatie over het begrip *confidence ellipse* kan gevonden worden in additioneel materiaal op Blackboard, dat buiten de stof van het studieonderdeel DATA-V valt.

De te minimaliseren afstand d_i voor punt i is nu de afstand tussen het punt (x_i, y_i) en een zeker punt (x_0, y_0) waar de modelfunctie raakt aan een ellips voor een zekere waarde r_i, zie Figuur 6.8 (b).

Omdat de positie (x_0, y_0) waar de raaklijn de ellips raakt afhankelijk is van de parameterwaarden, is de S-functie in het algemeen geen kwadratische vorm. Ook voor in beginsel lineaire aanpassingen kun je dus niet de exacte oplossingsmethode van sectie 6.4 gebruiken. Voor een rechte-lijnaanpassing $y = A + Bx$ is de afstand d_i analytisch uit te drukken in termen van de aanpassingsparameters A en B (die ook de richtingscoëfficiënt is van de raaklijn aan de ellips). De S-functie heeft een uniek minimum, en dit is via numerieke oplossing van Vgl. (6.6) te vinden. Voor de rechte-lijnaanpassing is de routine `LineFit2` beschikbaar in het MF-package.

In het algemene geval, waaronder ook niet-lineaire aanpassingen vallen, moeten geavanceerde minimalisatiemethoden worden gebruikt, zoals bijvoorbeeld de routine `GenFit2` in het MF-package. In het helpnotebook bij het MF-package, dat te vinden is op Blackboard, wordt uitgebreider ingegaan op de de aanpak van data met onzekerheid in x- én y-richting.

6.6.2 Combineren van meerdere modellen

Eveneens een complexer aanpassingsprobleem doet zich voor als er sprake is van data betrokken uit verschillende soorten experimenten, maar met gekoppelde modellen. Een simpel voorbeeld is de bepaling van de valversnelling g met twee verschillende methoden:

1. Meting van valtijden voor verschillende hoogten;
2. Meting van maximale hoogten bij verticale lancering met verschillende beginsnelheden.

Wanneer de meetresultaten van deze methoden gecombineerd worden gebruikt, is het mogelijk een "beste" aanpassing te maken door voor beide datasets een S-functie op te stellen en de som $S_1 + S_2$ van deze beide functies te minimaliseren als functie van de modelparameters (waarbij de gezochte valversnelling g een gemeenschappelijke parameter is).

Hoofdstuk 7

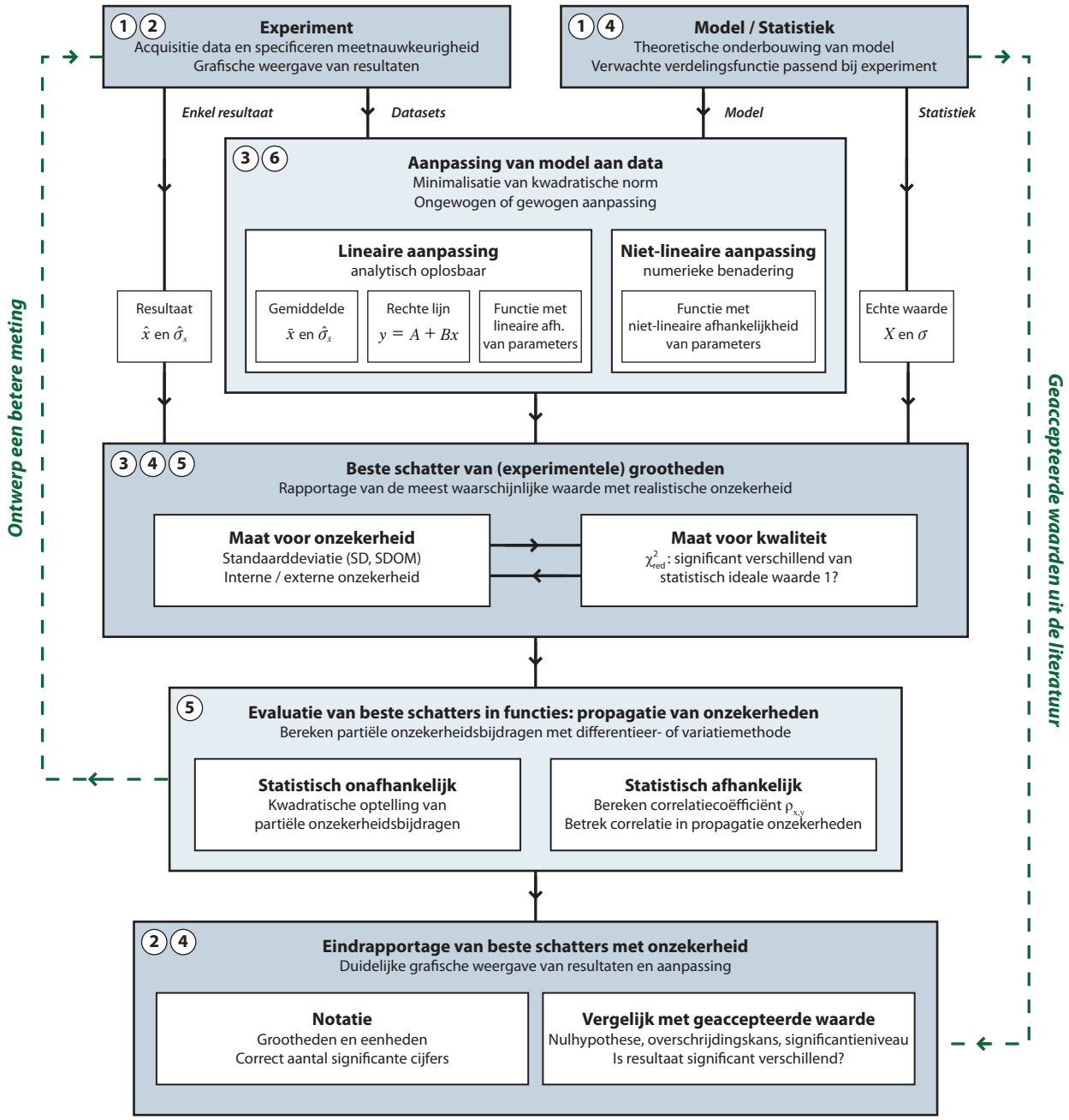
Conclusie

In de empirische wetenschap speelt de wisselwerking tussen experimentele waarneming en theorievorming een essentiële rol. In het studieonderdeel DATA-V is een introductie gegeven in data-analyse: het verantwoord verwerken van kwantitatieve gegevens, het bepalen van en het doorrekenen met onzekerheden, en het op een kritische manier verbinden van conclusies aan meetresultaten. Centraal staat dat conclusies worden onderbouwd met *kwantitatieve* en *objectieve* argumenten.
Een schematisch overzicht van de behandelde stof en een methode van aanpak is weergegeven in Figuur 7.1.
Het schema voor dataverwerking start met de invoer van experimentele waarnemingen en een (vernieuwd?) theoretisch model of verwachte statistiek, waaraan de data getoetst kan worden. Door middel van een aanpassingsroutine worden data en model aan elkaar gekoppeld, waarbij de onzekerheden in de datapunten een cruciale rol spelen in de bepaling van de best passende aanpassing.
Uit de aanpassing volgen beste schatters voor de modelparameters met bijbehorende onzekerheden. Bovendien is er inzicht verkregen in de kwaliteit van de aanpassing door in termen van de gereduceerde chi-kwadraat te kwantificeren hoe de opgegeven onzekerheden zich verhouden tot de waargenomen spreiding rondom het model.
Met behulp van propagatie van onzekerheden kan worden geëvalueerd welke modelparameters de grootste bijdrage leveren aan de uiteindelijke onzekerheid van de beste schatter. Deze informatie kan gebruikt worden om een verbeterd experiment te ontwerpen en daarmee tot een nauwkeuriger bepaling te komen.
De eindrapportage en de discussie of het verkregen resultaat significant verschilt van waarden uit eerdere metingen of modellen, maken het schema compleet. Een belangrijk onderdeel van de eindrapportage is de inzichtelijke (grafische) weergave van de uitkomst van het experiment.
Het moge duidelijk zijn dat je met de opgedane kennis de eerste stappen gezet hebt in de wereld van de data-analyse, waarbij het essentieel is om tot kwantitatieve uitspraken te komen over (de significantie van) een verkregen resultaat.

7.1 Hoe nu verder?

Alhoewel de in dit dictaat besproken methoden breed toepasbaar zijn bij het verwerken van resultaten verkregen met (natuurkundig) onderzoek, zal het je niet verbazen dat er voor complexere experimenten of aanpassingen een breed scala aan statistische technieken en algoritmes beschikbaar is. Voor de echte liefhebber is er op Blackboard extra verdiepingsmateriaal te vinden. We geven hier twee voorbeelden van geavanceerde methoden.

- Bij de dataverwerking zoals in deze cursus gepresenteerd is ervan uitgegaan dat er modelrelaties bekend zijn die kunnen worden aangepast aan meetresultaten (door variatie van vrije modelparameters).



Figuur 7.1: Schematisch overzicht van de inhoudelijke structuur van "Data-acquisitie & Toegepaste analyse: verwerking (DATA-V)". Het onderliggende verband tussen de verschillende kaders definieert een methode van aanpak voor data-analyse. In elk gekleurd kader staat aangegeven in welke hoofdstukken de stof aanbod gekomen is.

Er zijn echter complexe dataverwerkingsvraagstukken waar geen sprake is van modellen die kunnen worden aangepast. Een eenvoudig voorbeeld afkomstig uit de beeldverwerking: beschouw een twee-dimensionaal patroon van punten, met de vraag: is het patroon willekeurig verdeeld?

- Ook als er wel modellen beschikbaar zijn waaraan kan worden aangepast is de in deze cursus gegeven beschrijving erg beperkend. Er is bijvoorbeeld vanuitgegaan dat er bij één enkele instelparameter x_i steeds één enkele meting y_i hoort en dat deze dataparen worden gebruikt in de analyse.

  Maar meer algemeen is het aanpassen veel complexer: hoe als we per instelwaarde x_i meerdere soorten metingen $y_{i,1}, y_{i,2}, \ldots$ uitvoeren die gekoppeld moeten worden met een gemeenschappelijk model? Of andersom: mogelijk zijn er meerdere instelwaarden $x_{i,1}, x_{i,2}, \ldots$ van belang voor de bepaling van een enkele meetwaarde y_i.

In deze situaties is er een veel verdergaande analyse nodig om op verantwoorde wijze "beste waarden" te kunnen extraheren voor relevante parameters (en hun onzekerheden). Standaard applicaties zoals bijvoorbeeld beschikbaar in *Mathematica* zijn hierbij niet toereikend. In de praktijk moet dan vaak voor de specifieke situatie de numerieke data-analyse worden ontwikkeld. Een extreem voorbeeld is de analyse van data verkregen met de Hubble-Space-Telescope (HST). Het ontwikkelen van computationele technieken voor het extraheren van parameters uit dergelijke complexe datasets vereiste een jarenlange inspanning op hoog niveau.

Literatuurlijst

- David C. Baird. *Experimentation: An Introduction to Measurement Theory and Experiment Design*. Oxford University Press, 1998, 212 pages. ISBN: 0133032981.
- Roger Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. John Wiley & Son Ltd. Reprint edition, 1993, 222 pages. ISBN: 0471922951.
- H.J.C. Berendsen. *Goed meten met Fouten*. Bibl. der RUG, 2009, 141 pages. ISBN: 9036707080. Ook beschikbaar als pdf via http://www.hjcb.nl/meten/GoedMeten.pdf.
- Stuart L. Meyer. *Data Analysis for Scientists and Engineers*. Peer Management Consultants Ltd. Reprint edition, 1992, 513 pages. ISBN: 0963502700.
- G.L. Squires. *Practical Physics*. Prentice Hall. 4th edition, 2001, 212 pages. ISBN: 0521779405 (paperback version).
- John R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books. 2nd edition, 1997, 327 pages. ISBN: 093570275X (paperback version).
- Philip R. Bevington and D. Keith Robinson. *Data reduction and error analysis for the physical sciences*. McGraw-Hill. 3rd edition, 2003, 320 pages. ISBN: 0079112439.

Index

Appendix A

Enige bewijzen omtrent gewichten en chi-kwadraat

In Hoofdstukken 3 en 6 zijn enige aannames gemaakt omtrent de definitie van gewichtsfuncties, onzekerheden en de chi-kwadraat als beste maat voor de kwaliteit van een aanpassing. In deze appendix wordt een aantal van deze aannames bewezen. Deze bewijzen zijn grotendeels ontleend aan Bevington, *Data reduction and error analysis for the Physical sciences* (McGraw-Hill, 2003).

A.1 Bepaling van onzekerheid in de standaarddeviatie

In Sectie 3.2, Vgl. (3.4) is gesteld dat de relatieve onzekerheid in de standaarddeviatie

$$\hat{\sigma}_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2} \tag{A.1}$$

gelijk is aan

$$\hat{\sigma}_{\hat{\sigma}_x}/\hat{\sigma}_x = \frac{1}{\sqrt{2(n-1)}} \tag{A.2}$$

We zullen hier uitwerken hoe deze waarde bepaald wordt.

De "foutieve" aanpak De eerste aanpak zou kunnen zijn om, net als we hebben gedaan voor de bepaling van de onzekerheid in het gemiddelde Vgl. (3.6) het probleem te lijf te gaan met propagatie van onzekerheid. Deze aanpak is te vinden bij Vgl. (5.29) op pagina 69. Maar, bij netjes uitwerken zul je het foute resultaat $1/\sqrt{(n-1)}$ vinden, een factor $\sqrt{2}$ ernaast!
De reden hiervoor is dat bij de differentieermethode (zoals in Hoofdstuk 5 vermeld) stilzwijgend aangenomen wordt dat de functiewaarde op het variatiegebied bij benadering *lineair* verandert. Voor Vgl. (5.29), en trouwens ook voor het bewijs in Sectie A.2, gaat de propagatie van onzekerheid goed omdat de functie $\overline{x}_{gew}$ *lineair* is in de variabelen x_i (vanzelfsprekend is de verdelingsfunctie van $\overline{x}_{gew}$ dan gelijkvormig).
Bij Vgl. (A.1) heb je te maken met een sterk *niet-lineaire* functie in x_i; dit lijkt enigszins op de situatie die in Figuur 5.3 (d) wordt geïllustreerd. Je kunt ook berekenen dat de tweede-orde Taylortermen in Vgl. (5.7) - (5.8) niet verwaarloosbaar zijn ten opzichte van de lineaire termen, en daarnaast aanleiding geven tot een asymmetrische verdeling.

De juiste aanpak Voor de juiste aanpak kunnen we dus niet de analyse doen op basis van alleen het gemiddelde en de standaarddeviatie, maar moeten we kijken naar de volledige *verdelingsfunctie*. Het eenvoudigst is om te kijken naar de *variantie*, dat is $\hat{\sigma}_x^2$. Deze variantie (zelf ook een statistische grootheid) blijkt een verdelingsfunctie te kennen gelijk aan de *chi-kwadraatverdeling* (Vgl. (4.42)) $C_{n-1}(\hat{\sigma}_x^2)$. Het bewijs hiervan is wat te lang om in deze appendix op te nemen.

Van de chi-kwadraatverdelingsfunctie $C_{n-1}(\hat{\sigma}_x^2)$ weten we dat deze een verwachtingswaarde $n-1$ heeft, en een variantie $2(n-1)$. De "beste schatter" voor de *relatieve onzekerheid* in een variantie is dan $\sqrt{2/(n-1)}$. Omdat bij machtsverheffen de relatieve onzekerheden met de macht worden vermenigvuldigd (een rekenregel uit Tabel 5.2), wordt de onzekerheid in de standaarddeviatie dan $1/2 \times \sqrt{2/(n-1)} = 1/\sqrt{2(n-1)}$.

A.2 Definitie van gewichten voor het gewogen gemiddelde

We nemen aan dat n metingen zijn gedaan aan grootheid x, met resultaten $\{x_1, \cdots, x_n\}$ en onzekerheden $\{\sigma_1, \cdots, \sigma_n\}$. Hierbij worden de metingen *random* en statistisch onafhankelijk verondersteld.

In Hoofdstuk 6 is het gewogen gemiddelde Vgl. (6.22) ingevoerd:

$$\overline{x}_{gew} = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i. \tag{A.3}$$

Zonder bewijs hebben we aangenomen Vgl. (6.4), oftewel het feit dat het "beste" gewicht w_i dat aan elk meetpunt i moet worden toegekend gelijk is aan $w_i = 1/\sigma_i^2$. Hier volgt een algemeen bewijs voor deze aanname.

Met behulp van de algemene uitdrukking voor propagatie van onzekerheid middels de differentieermethode, Vgl. (5.25), kunnen we een algemene uitdrukking opstellen voor de onzekerheid in $\overline{x}_{gew}$:

$$\begin{aligned}
\sigma_{\overline{x}_{gew}}^2 &\equiv \mathrm{SDOM}_{gew}^2 \\
&= \sum_{i=1}^{n} \left(\frac{\partial \overline{x}_{gew}}{\partial x_i} \sigma_i \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{1}{\sum_{j=1}^{n} w_j} \frac{\partial}{\partial x_i} \sum_{j=1}^{n} w_j x_j \sigma_i \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{\delta_{ij} w_j}{W} \sigma_i \right)^2 \\
&= \frac{\sum_{i=1}^{n} (w_i \sigma_i)^2}{W^2}.
\end{aligned} \tag{A.4}$$

Hierbij is gebruik gemaakt van de zogenaamde *Kronecker-delta* δ_{ij}. Deze heeft waarde 1 wanneer $i = j$ en 0 in alle andere gevallen, en drukt uit dat alleen de term in de som waarbij i gelijk is aan j een waarde ongelijk aan nul heeft. Verder hebben we gesubstitueerd $W = \sum_{j=1}^{n} w_j$.

We wensen nu de gewichten w_i zó vast te stellen dat SDOM_{gew} geïdentificeerd wordt met de kleinst mogelijke onzekerheid in $\overline{x}_{gew}$. Deze eis dat SDOM_{gew} een *minimum* vormt, impliceert dat alle afgeleiden van SDOM_{gew} naar w_j gelijk zijn aan nul, voor alle $j = \{1, \cdots, n\}$. Omdat $\mathrm{SDOM}_{gew} \geq 0$, correspondeert dit met een minimum in SDOM_{gew}^2.

Expliciete differentiatie van Vgl. (A.4) naar w_j levert

$$\begin{aligned}\frac{\partial}{\partial w_j}\mathrm{SDOM}_{gew}^2 &= \frac{\partial}{\partial w_j}\frac{\sum_{i=1}^n (w_i\sigma_i)^2}{W^2} \\ &= \frac{2\delta_{ji}w_i\sigma_i\sigma_i W^2 - 2W\delta_{ji}\sum_{i=1}^n (w_i\sigma_i)^2}{W^4} \\ &\equiv 0 \qquad \text{voor alle } j. \end{aligned} \tag{A.5}$$

De teller van deze uitdrukking vereenvoudigen levert (de noemer is per definitie ongelijk aan nul)

$$w_j\sigma_j^2\left(\sum_{i=1}^n w_i\right) - \sum_{i=1}^n (w_i\sigma_i)^2 = 0 \qquad \text{voor alle } j. \tag{A.6}$$

Dus

$$w_j\sigma_j^2 = \frac{\sum_{i=1}^n (w_i\sigma_i)^2}{\sum_{i=1}^n w_i} \qquad \text{voor alle } j. \tag{A.7}$$

Dit moet wel betekenen dat $w_j\sigma_j^2$ niet van j afhangt. Oftewel $w_j \propto \sigma_j^{-2}$, wat bewezen diende te worden. Hierbij blijft er een vrije (schalings)factor over die geen invloed heeft op SDOM_{gew}.

A.3 Definitie van de interne onzekerheid

In Sectie 6.3 is ingevoerd de *interne* onzekerheid $\hat{\sigma}_{int}$ van het gewogen gemiddelde (zie Vgl. (6.26)). Deze onzekerheid wordt vastgesteld onafhankelijk van de geconstateerde spreiding in meetwaarden $\{x_1, \cdots, x_n\}$. De uitdrukking Vgl. (6.26) kan worden afgeleid door toepassing van de differentieermethode Vgl. (5.25) op de uitdrukking Vgl. (A.3) voor het gewogen gemiddelde, met invullen van de gewichten $w_i = 1/\sigma_i^2$. De expliciete uitwerking voor de gekwadrateerde onzekerheid:

$$\begin{aligned}\sigma_{\overline{x}_{gew}}^2 &= \sum_{i=1}^n \left(\frac{\partial \overline{x}_{gew}}{\partial x_i}\sigma_i\right)^2 \\ &= \sum_{i=1}^n \sigma_i^2 \left(\frac{\partial}{\partial x_i}\sum_{j=1}^n \sigma_j^{-2}x_j \Big/ \sum_{k=1}^n \sigma_k^{-2}\right)^2 \\ &= \sum_{i=1}^n \sigma_i^2 \frac{1}{\left(\sum_{k=1}^n \sigma_k^{-2}\right)^2}\left(\delta_{ij}\sigma_j^{-2}\right)^2 \\ &= \sum_{i=1}^n \sigma_i^{-2} \frac{1}{\left(\sum_{k=1}^n \sigma_k^{-2}\right)^2} \\ &= \frac{1}{\sum_{k=1}^n \sigma_k^{-2}}. \end{aligned} \tag{A.8}$$

De gebruikte *Kronecker-delta* is gedefinieerd in sectie A.2.

A.4 De kwadratische norm en χ^2 als goede maat voor de beste aanpassing

Tot slot nog een korte onderbouwing van het gebruik van de kwadratische norm Vgl. (6.3) bij aanpassingsroutines, en de chi-kwadraat χ^2 bij de beste aanpassing.
We gaan uit van een dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$, dat wil zeggen een serie metingen aan grootheid y met resultaten y_i en onzekerheden σ_i, bij instelwaarden x_i van grootheid x. Tevens veronderstellen we een functie $f_{\boldsymbol{p}}(x) = y(x)$ die het (theoretische) verband tussen grootheden x en y beschrijft, waarbij grootheden of parameterwaarden $\boldsymbol{p}$, die gedurende het experiment constant worden verondersteld, bepaald dienen te worden.
We nemen verder aan dat de metingen y_i aan grootheid y bij instelling x_i van grootheid x normaal verdeeld zijn, met onzekerheid σ_i. In zijn meest algemene vorm is de normale verdeling Vgl. (4.34)

$$G_{X,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-X)^2}{2\sigma^2}\right). \tag{A.9}$$

We veronderstellen nu dat voor elke meting het meetresultaat y_i is getrokken uit een normale verdeling $G_{f_{\boldsymbol{p}}(x_i),\sigma_i}(x)$. Hierbij nemen we derhalve aan dat onder elke meting met instelwaarde x_i een verdeling ligt met gemiddelde $f_{\boldsymbol{p}}(x_i)$ (de voorspelde waarde bij x_i), maar dat door onder andere experimentele effecten de breedte van de verdeling bij elke meting varieert. De kans $\mathrm{d}P_i$ op het vinden van het meetresultaat y_i in een (vaste) bandbreedte $\mathrm{d}y$ gegeven de veronderstelde verdelingsfunctie $G_{f_{\boldsymbol{p}}(x_i),\sigma_i}(x)$ is dan

$$\mathrm{d}P_i = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{-(y_i - f_{\boldsymbol{p}}(x_i))^2}{2\sigma_i^2}\right)\mathrm{d}y. \tag{A.10}$$

De *samengestelde kans* $\mathrm{d}P$ op het vinden van de volledige dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$ gegeven de functie $f_{\boldsymbol{p}}(x)$ (of preciezer, de functiewaarden $f_{\boldsymbol{p}}(x_i)$ bij instelwaarden x_i) is dan gelijk aan de vermenigvuldiging van de individuele kansen $\mathrm{d}P_i$:

$$\begin{aligned}
\mathrm{d}P &= \prod_{i=1}^{n} \mathrm{d}P_i \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{-(y_i - f_{\boldsymbol{p}}(x_i))^2}{2\sigma_i^2}\right)\mathrm{d}y \\
&= \mathrm{d}y^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - f_{\boldsymbol{p}}(x_i))^2}{\sigma_i^2}\right) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i}.
\end{aligned} \tag{A.11}$$

De methode van de *meest waarschijnlijke schatter* (Engels: *Maximum likelihood method*) stelt nu, dat de meest waarschijnlijke waarde voor $f_{\boldsymbol{p}}(x_i)$ die is, waarvoor de kans $\mathrm{d}P$ maximaal wordt. Anders gezegd, als "beste schatter" voor $f_{\boldsymbol{p}}(x_i)$ (of beter gezegd, voor de "beste schatters" voor parameters $\boldsymbol{p}$ in de functie f) veronderstellen we die waarde, waarvoor de hoogste waarschijnlijkheid bestaat op het vinden van de daadwerkelijk gevonden dataset $\{\{x_1, y_1, \sigma_1\}, \cdots, \{x_n, y_n, \sigma_n\}\}$.
Een belangrijke observatie is dat de kans $\mathrm{d}P$ *gemaximaliseerd* wordt, als de term

$$S(\boldsymbol{p}) = \sum_{i=1}^{n} \frac{(y_i - f_{\boldsymbol{p}}(x_i))^2}{\sigma_i^2} \tag{A.12}$$

in de exponentiële functie *geminimaliseerd* wordt. Vgl. (A.12) is de kwadratische norm Vgl. (6.3).
In de afleiding van Vgl. (A.12) is normaal verdeelde data aangenomen. De gebruikte methode laat echter het afleiden van aangepaste uitdrukkingen voor de te minimaliseren norm toe voor een willekeurige maar bekende verdelingsfunctie, zoals de Poissonverdeling.

Appendix B

Tabel met overschrijdingskansen voor de normale verdeling

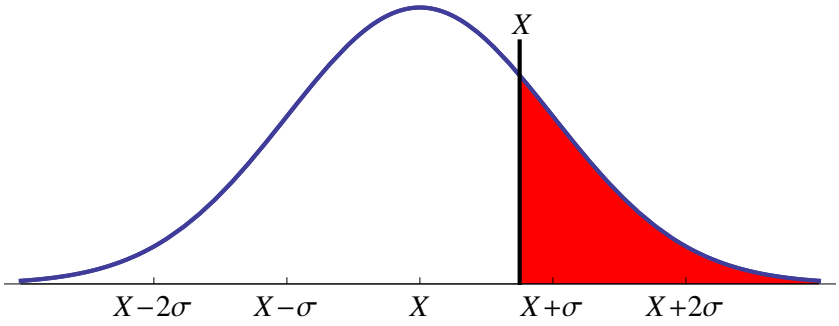
Voor het berekenen van de overschrijdingskansen gaan we uit van een normale verdeling Vgl. (4.34):

$$G_{X,\sigma} = \frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{(x-X)^2}{2\sigma^2}. \tag{B.1}$$

Hierin is X de "echte" waarde en σ de standaarddeviatie. In Tabel B.1 is de *rechtszijdige* overschrijdingskans

$$\int_{X+t\sigma}^{\infty} G_{X,\sigma} dx. \tag{B.2}$$

weergegeven als functie van de parameter $t = (X' - X)/\sigma$. De parameter t is een geschaalde discrepantie, en drukt de gevonden waarde X' uit in het aantal keer de standaarddeviatie dat de waarde afligt van de "echte" waarde X, oftewel $X' = X + t\sigma$. Een illustratie van de berekende integraal (kans) is te zien in Figuur B.1. Uit de overschrijdingskansen in deze tabel zijn de tweezijdige overschrijdingskansen eenvoudig af te leiden.



Figuur B.1: Illustratie van de berekening van de overschrijdingskans voor een gevonden waarde X' ($= X + t\sigma$ met $t = 0.75$) voor een grootheid die normaal verdeeld is rond "echte" waarde X met standaarddeviatie σ. Het oppervlak van het gearceerde gebied komt overeen met een kans van 22.7%, zie Tabel B.1 bij de waarde $t = 0.75$.

Tabel B.1: Tabel met rechtszijdige overschrijdingskansen voor de normale verdeling als functie van de parameter t. Tweezijdige (of eventueel linkszijdige) overschrijdingskansen kunnen eenvoudig uit deze rechtszijdige overschrijdingskansen bepaald worden.
De eerste cel van de rij geeft de eerste twee significante cijfers van t, en de kolom voegt het derde significante cijfer toe. Een voorbeeld: in rij 3, kolom 2 vind je de rechtszijdige overschrijdingskans voor t-waarde $t = 0.2 + 0.01 = 0.21$, wat resulteert in een overschrijdingskans van 41.7%.

t	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.500	0.496	0.492	0.488	0.484	0.480	0.476	0.472	0.468	0.464
0.1	0.460	0.456	0.452	0.448	0.444	0.440	0.436	0.433	0.429	0.425
0.2	0.421	0.417	0.413	0.409	0.405	0.401	0.397	0.394	0.390	0.386
0.3	0.382	0.378	0.374	0.371	0.367	0.363	0.359	0.356	0.352	0.348
0.4	0.345	0.341	0.337	0.334	0.330	0.326	0.323	0.319	0.316	0.312
0.5	0.309	0.305	0.302	0.298	0.295	0.291	0.288	0.284	0.281	0.278
0.6	0.274	0.271	0.268	0.264	0.261	0.258	0.255	0.251	0.248	0.245
0.7	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221	0.218	0.215
0.8	0.212	0.209	0.206	0.203	0.200	0.198	0.195	0.192	0.189	0.187
0.9	0.184	0.181	0.179	0.176	0.174	0.171	0.169	0.166	0.164	0.161
1.0	0.159	0.156	0.154	0.152	0.149	0.147	0.145	0.142	0.140	0.138
1.1	0.136	0.133	0.131	0.129	0.127	0.125	0.123	0.121	0.119	0.117
1.2	0.115	0.113	0.111	0.109	0.107	0.106	0.104	0.102	0.100	0.099
1.3	0.097	0.095	0.093	0.092	0.090	0.089	0.087	0.085	0.084	0.082
1.4	0.081	0.079	0.078	0.076	0.075	0.074	0.072	0.071	0.069	0.068
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
1.6	0.055	0.054	0.053	0.052	0.051	0.049	0.048	0.047	0.046	0.046
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.038	0.037
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.031	0.030	0.029
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
2.0	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018
2.1	0.018	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.014
2.2	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011
2.3	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008
2.4	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006
2.5	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005
2.6	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
2.7	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.8	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001
3.0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Appendix C

Tabel met overschrijdingskansen voor de gereduceerde chi-kwadraat

Zonder bewijs stellen we hier dat de chi-kwadraatverdeling wordt gegeven door (zie ook Sectie 4.5.3 en extra materiaal)

$$C_d(\chi^2) = \frac{1}{2^{d/2}\Gamma(d/2)} \left(\chi^2\right)^{d/2-1} e^{-\frac{\chi^2}{2}} \quad (\chi^2 \geq 0). \tag{C.1}$$

Hierin is d het aantal vrijheidsgraden (aantal datapunten minus aantal aanpassingsparameters) en $\Gamma(d/2)$ de Gamma-functie. De verdeling van de *gereduceerde* chi-kwadraat is hieruit af te leiden middels de transformatie $\chi^2 \rightarrow \chi^2_{red} = \frac{\chi^2}{d}$:

$$C_{d,red}(\chi^2_{red}) = \frac{d}{2^{d/2}\Gamma(d/2)} \left(d\chi^2_{red}\right)^{d/2-1} e^{-d\frac{\chi^2_{red}}{2}} \quad (\chi^2_{red} \geq 0). \tag{C.2}$$

Omdat de (gereduceerde) chi-kwadraat altijd groter is dan nul, hebben de verdelingen alleen waarden ≥ 0 voor $\chi^2 \geq 0$ respectievelijk $(\chi^2_{red} \geq 0)$.
In Tabel C.1 is als functie van het aantal vrijheidsgraden d en de waarde χ^2_{red}) de *rechtszijdige* overschrijdingskans gegeven, dat wil zeggen de kans om $\tilde{\chi}^2_{red}$ aan te treffen groter dan een zekere χ^2_{red} (die bijvoorbeeld is gevonden voor een dataset):

$$\text{Prob}(\tilde{\chi}^2_{red} \geq \chi^2_{red}) = \int_{\chi^2_{red}}^{\infty} C_{d,red}(\chi^2_{red}) \mathrm{d}\left(\chi^2_{red}\right). \tag{C.3}$$

De *linkszijdige* overschrijdingskans kan gevonden worden door middel van $1 - \text{Prob}(\tilde{\chi}^2_{red} \geq \chi^2_{red})$, en is dus gelijk aan de kans om een $\tilde{\chi}^2_{red}$ aan te treffen kleiner dan de referentiewaarde χ^2_{red}.

Tabel C.1: Tabel met overschrijdingskansen voor de gereduceerde chi-kwadraat χ^2_{red} als functie van het aantal vrijheidsgraden $d = n - r$ (aantal datapunten minus aantal aanpassingsparameters).

$d\downarrow \mid \chi^2_{red}\rightarrow$	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**	**2.2**	**2.4**	**2.6**	**2.8**	**3.0**
1	1.000	0.655	0.527	0.439	0.371	0.317	0.273	0.237	0.206	0.180	0.157	0.138	0.121	0.107	0.094	0.083
2	1.000	0.819	0.670	0.549	0.449	0.368	0.301	0.247	0.202	0.165	0.135	0.111	0.091	0.074	0.061	0.050
3	1.000	0.896	0.753	0.615	0.494	0.392	0.308	0.241	0.187	0.145	0.112	0.086	0.066	0.050	0.038	0.029
4	1.000	0.938	0.809	0.663	0.525	0.406	0.308	0.231	0.171	0.126	0.092	0.066	0.048	0.034	0.024	0.017
5	1.000	0.963	0.849	0.700	0.549	0.416	0.306	0.221	0.156	0.109	0.075	0.051	0.035	0.023	0.016	0.010
6	1.000	0.977	0.879	0.731	0.570	0.423	0.303	0.210	0.143	0.095	0.062	0.040	0.025	0.016	0.010	0.006
7	1.000	0.986	0.903	0.756	0.587	0.429	0.299	0.200	0.130	0.082	0.051	0.031	0.019	0.011	0.007	
8	1.000	0.991	0.921	0.779	0.603	0.433	0.294	0.191	0.119	0.072	0.042	0.024	0.014	0.008		
9	1.000	0.994	0.936	0.798	0.616	0.437	0.290	0.182	0.109	0.063	0.035	0.019	0.010	0.005		
10	1.000	0.996	0.947	0.815	0.629	0.440	0.285	0.173	0.100	0.055	0.029	0.015	0.008			
11	1.000	0.998	0.957	0.830	0.640	0.443	0.280	0.165	0.091	0.048	0.024	0.012	0.006			
12	1.000	0.998	0.964	0.844	0.651	0.446	0.276	0.157	0.084	0.042	0.020	0.009				
13	1.000	0.999	0.971	0.856	0.661	0.448	0.271	0.150	0.077	0.037	0.017	0.007				
14	1.000	0.999	0.976	0.867	0.670	0.450	0.267	0.143	0.071	0.033	0.014	0.006				
15	1.000	1.000	0.980	0.878	0.679	0.451	0.263	0.137	0.065	0.029	0.012	0.005				
16	1.000	1.000	0.983	0.887	0.687	0.453	0.258	0.131	0.060	0.025	0.010					
17	1.000	1.000	0.986	0.895	0.695	0.454	0.254	0.125	0.055	0.022	0.008					
18	1.000	1.000	0.988	0.903	0.703	0.456	0.250	0.120	0.051	0.020	0.007					
19	1.000	1.000	0.990	0.910	0.710	0.457	0.246	0.114	0.047	0.017	0.006					
20	1.000	1.000	0.992	0.916	0.717	0.458	0.242	0.109	0.043	0.015	0.005					
21	1.000	1.000	0.993	0.922	0.723	0.459	0.239	0.105	0.040	0.014						
22	1.000	1.000	0.994	0.927	0.729	0.460	0.235	0.100	0.037	0.012						
23	1.000	1.000	0.995	0.932	0.735	0.461	0.231	0.096	0.034	0.011						
24	1.000	1.000	0.996	0.937	0.741	0.462	0.228	0.092	0.032	0.009						
25	1.000	1.000	0.997	0.941	0.747	0.462	0.224	0.088	0.029	0.008						
26	1.000	1.000	0.997	0.945	0.752	0.463	0.221	0.085	0.027	0.007						
27	1.000	1.000	0.998	0.949	0.757	0.464	0.218	0.081	0.025	0.007						
28	1.000	1.000	0.998	0.952	0.762	0.464	0.214	0.078	0.023	0.006						
29	1.000	1.000	0.998	0.956	0.767	0.465	0.211	0.075	0.021	0.005						
30	1.000	1.000	0.999	0.959	0.772	0.466	0.208	0.072	0.020	0.005						