Kwantitatieve toepassingen in de bedrijfskunde

Noordhoff Uitgevers

Dr. A. Buijs
Mr. Drs. J.W. Wijbenga
Drs. E.T. Thijssen

Kwantitatieve toepassingen in de bedrijfskunde

Kwantitatieve toepassingen in de bedrijfskunde

Prof. dr. A. Buijs

met medewerking van

Drs. E.T. Thijssen

Mr. drs. J.W. Wijbenga

Vierde druk

Noordhoff Uitgevers Groningen | Houten

Ontwerp omslag: G2K Designers Groningen / Amsterdam
Omslagillustratie: PhotoDisc

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan: Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB Groningen, e-mail: info@noordhoff.nl

Deze uitgave is gedrukt op FSC-papier.

3 / 13

ISBN (ebook) 978 90 01 84765 4
ISBN 978 90 01 11001 7
NUR 123

Woord vooraf

Er zijn heel wat studierichtingen met cursusmodules over onderwerpen uit de wereld van bedrijfskunde en management. Daarbij staat dikwijls het maken van keuzes en het onderbouwen van beslissingen centraal. Dit boek biedt een aantal methoden en instrumenten die kunnen dienen als nuttig gereedschap voor een beslisser bij het afwegen van verschillende keuzes. Daarbij ligt het accent op een aantal methodes uit de wiskunde en statistiek.

Bij sommige universitaire en HBO-studies wordt tegenwoordig nogal kritisch aangekeken tegen cursussen met een duidelijk kwantitatieve inslag. Waarom zou je nog aan onderwijs in kwantitatieve methoden doen? 'We hebben toch computers? Je kunt toch in noodgevallen een specialist raadplegen?', zijn veelgehoorde opmerkingen. Soms lijken dergelijke vragen te maskeren dat men onderwerpen met een kwantitatieve inslag maar lastig vindt. Studenten vinden zulke onderwerpen maar moeilijk is dan de bijgedachte.
Toch denk ik dat het goed en ook belangrijk is om studenten wel degelijk te laten kennismaken met de toepassingen van wiskunde en statistiek zoals die in dit boek worden gepresenteerd. In de eerste plaats maakt de student kennis met een aantal aanpakken om problemen te kunnen oplossen. Ook is het bestuderen van kwantitatieve onderwerpen sowieso een training in systematisch en logisch denken.

Ten opzichte van de derde druk zijn enkele zaken gewijzigd.
- De uitvoerige behandeling van de Simplexmethode bij lineair programmeren is volledig verwijderd. Het vervelende en tijdrovende rekenwerk van deze methode wordt tegenwoordig met software gedaan. Wij hebben gekozen voor de *solver* van Excel. Hierdoor is het accent verlegd van de rekenkundige facetten naar de interpretatie van de gevonden resultaten.
- Nieuw in dit boek is hoofdstuk 11 over inkoopmanagement. Dit is met name opgenomen omdat hiermee kan worden geïllustreerd dat in een deelgebied van management allerlei kwantitatieve methodes worden benut als hulpmiddel om problemen op te lossen.
- Ieder hoofdstuk begint met een openingscasus. De bedoeling van deze cases is dat de student gevoel krijgt voor de problematiek die in dat specifieke hoofdstuk aan de orde komt. In het opgavenboek zal doorgaans het geformuleerde probleem van de openingscasus nader worden uitgewerkt in enkele opgaven die daarop voortbouwen.
- In sommige hoofdstukken zijn delen aangepast, zoals het formuleren van modellen in hoofdstuk 1 en de toevoeging van AHP in hoofdstuk 2.

Verder is de opzet van het opgavenboek en de uitwerkingen gewijzigd. In deze nieuwe editie zijn de opgaven en uitwerkingen in één boek samengebracht. Ook zijn per hoofdstuk enkele nieuwe opgaven toegevoegd.

Aan deze editie is meegewerkt door mr.drs. Jan W. Wijbenga, die met name het onderwerp lineair programmeren heeft bewerkt. Hij heeft ook veel werk verzet bij het maken van de uitwerkingen voor het opgavenboek.
Hoofdstuk 11 over inkoopmanagement is geschreven door drs. Elliott T. Thijssen van *Significant Consultancy*. Beide heren wil ik danken voor hun inzet.

Ik hoop dat de vernieuwingen in deze editie door gebruikers van dit boek als verbeteringen zullen worden ervaren. Opmerkingen van allerlei aard zijn uiteraard van harte welkom.

Bilthoven 2005, Arie Buijs

Inhoud

Modellen, kansen en toepassingen

1

OPENINGSCASUS

Met Trendtrips naar Lesbos

Touroperator Trendtrips is ieder jaar actief met het ontwikkelen van reizen. Na het bedenken van zo'n reis moet er van alles gebeuren. Bernard en Romana zijn beiden inkoopmanager bij Trendtrips en buigen zich over een nieuwe reis naar Lesbos. Hiervoor moeten zij ruim van tevoren hotelkamers en vliegtuigstoelen inkopen. Gelet op de prijzen bij de concurrenten, denken ze een week Lesbos te kunnen aanbieden voor 599 euro. Van hotel Alma in het dorpje Petra is inmiddels een offerte binnengekomen voor de levering van vijftig hotelkamers. Voor het hoogseizoen offreert dit hotel een inkoopprijs 190 euro per kamer per week.
En luchtvaartmaatschappij Air Mokum biedt de retourtickets aan voor 220 euro.
'Wat denk je van die offertes Bernard?' vraagt Romana. 'Tja, de winstgevendheid hangt wel af van het werkelijke aantal boekingen en dat weten we pas volgend jaar', antwoordt Bernard.
Een eerste berekening wordt gemaakt alsof wekelijks alle vijftig reizen worden geboekt. Dat levert dan opbrengst minus kosten, dus per boeking
599 – (190 + 220) = 189 euro. Dat maal 50 maakt 9.450 per week. Heel mooi dus.
'Ja, maar stel dat we minder dan 50 boekingen krijgen, dan moeten we de laatste reizen via de last-minute aaanbiedingen verkopen. Dat kan voor 249 euro, dus dan hebben we een verlies van 161 euro per reis.' 'Maar Bernard, mijn gevoel zegt dat dit juist een heel erg gewilde bestemming wordt. Als er meer dan vijftig klanten zijn, kunnen we kamers en tickets bijboeken. Wel tegen een hoger tarief, namelijk 230 + 280 = 510 euro. Dus ook dan verdienen we nog 89 euro.

Laten we eens wat berekeningen doen op basis van een inkoop van vijftig kamers en tickets. We kunnen eens kijken wat de opbrengst is bij dertig boekingen of vijftig of zeventig.'
Deze berekening maakt veel discussie los. Stel dat je niet vijftig maar veertig reizen inkoopt, of juist zestig, wat is het gevolg dan voor de weekopbrengst indien later blijkt dat veertig, vijftig of zestig reizen worden verkocht?
Het schijnt dat een kamer niet 190, maar 180 euro kost bij een inkoop van minimaal zestig kamers per week. Moeten we daar dan gebruik van maken en het risico aanvaarden dat er te weinig boekingen komen?
En als we nou eens een kansverdeling voor de vraag naar reizen zouden kunnen opstellen, zouden we dan met kansrekening onze inkoopbeslissing kunnen nemen?
Kortom, welke beslissing moeten Bernard en Romana hier nemen om voor Trendtrips een zo hoog mogelijke opbrengst te realiseren en welke factoren moet hierbij worden betrokken?

Bij het nemen van beslissingen spelen vaak heel wat factoren een rol, waardoor het niet eenvoudig is een afweging te maken om te komen tot de beste keuze. Beslissingen nemen is misschien wel de belangrijkste taak die een leidinggevende functionaris heeft, of het nou bij de overheid, semi-overheid of bij het bedrijfsleven is.
Diverse methoden uit de wiskunde en de statistiek kunnen van nut zijn bij het vinden van verstandige oplossingen voor problemen. Veel methoden kunnen worden geschaard onder het vakgebied *Operations Research*, maar we bespreken ook enkele onderwerpen die daarbuiten liggen, zoals financiële toepassingen, inkoopmanagement en kwaliteitszorg.
In dit hoofdstuk maken we in paragraaf 1.2 allereerst enkele algemene opmerkingen over methoden van probleemoplossing en het opstellen van een model.
In paragraaf 1.3 volgt een overzicht van een aantal onderwerpen uit de kansrekening. Met enkele voorbeelden illustreren we hoe men de kansrekening in de praktijk kan toepassen. In de paragraaf 1.4 ten slotte, volgt een interessant toepassingsgebied van kansrekening. Dit betreft het thema bedrijfszekerheid, dat gaat over vraagstukken met betrekking tot de levensduur van machines en onderdelen daarvan.

1.1 Operations Research

Over het tijdstip waarop het vakgebied Operations Research ontstond, zijn de meningen verdeeld. Sommigen zien het toepassen van allerlei metingen in industriële productieprocessen als een eerste belangrijke mijlpaal. We denken dan aan het werk van Frederic Taylor aan het begin van de 20ste eeuw. Hoofddoel hierbij was het opvoeren van de doelmatigheid van productieprocessen. Andere auteurs zien met name veel ontwikkeling in de periode rond de Tweede Wereldoorlog toen teams van wetenschappers werkten aan grote militaire projecten. Planningsprocedures zoals PERT kwamen in die tijd tot ontwikkeling. Nog weer anderen beschouwen het ontstaan van de techniek van lineaire programmering als de serieuze start van het vakgebied.

Duidelijk is dat in de loop van de jaren een accentverschuiving is opgetreden voor wat betreft de focus op de toepassingsgebieden van Operations Research. We onderscheiden vier fasen:

Eerste fase van Operations Research
In de vroegste fase stond het bevorderen van *efficiency* centraal. Bijvoorbeeld bij productiemethoden met gebruik van een lopende band was meting van bewerkingstijden en de onderlinge afstemming daarvan heel belangrijk. Later werden allerlei optimaliseringsmodellen belangrijk, vooral bij complexe processen die zonder een modelmatige aanpak moeilijk zijn te doorgronden. Het gaat dan bijvoorbeeld om planningsvraagstukken bij ingewikkelde projecten, waarbij met name het accent ligt op netwerkplanning. Ook problemen waarbij een 'beste' oplossing moet worden gevonden terwijl men rekening moet houden met allerlei (lineaire) randvoorwaarden, vallen in die categorie.

Tweede fase van Operations Research
Als tweede fase kan men de periode vanaf de jaren zestig en zeventig van de 20^{e} eeuw beschouwen, met haar sterke aandacht op van *kwaliteit* van producten en diensten. Deze ontwikkeling werd mede bevorderd door de sterke opkomst van de Japanse industrie in die tijd. Men constateerde dat het er niet alleen om gaat dat de *hoeveelheid* output van een proces zo groot mogelijk is, maar vooral dat de juiste kwaliteit wordt geproduceerd. Aan afgekeurde eindproducten hebben we niets. Bovendien is het belangrijk dat het productieproces zelf zodanig is georganiseerd dat het maken van fouten wordt voorkomen of dat fouten reeds in een vroeg stadium van het productieproces worden ontdekt.

Derde fase van Operations Research
Als derde fase kan men de ontwikkeling naar *schaalverkleining* zien. Terwijl de eerste toepassingen van Operations Research juist komen uit de wereld van de massaproductie, komt het accent nu te liggen op klantgerichtheid, produceren wat de klant eist, het reduceren van voorraden en just-in-time levering van goederen. Problemen worden niet meer geïsoleerd bekeken, maar meer als onderdeel van een groter geheel. *Supply Chain Management* is een bekende term die bij deze benadering past.

Vierde fase van Operations Research
Bij de laatste fase ten slotte, zijn we in het begin van de 21ste eeuw beland, in een tijd waarin wordt verwacht dat de grote ondernemingen minder belangrijk worden. Het accent komt op *innovatie*, creativiteit, klein ondernemerschap en self-employment te liggen. Wat de gevolgen van deze ontwikkelingen zijn voor de modelmatige studie van productieprocessen, moet nog worden afgewacht.

Innovatie

1.2 Problemen oplossen

In organisaties – of deze nu op het maken van winst gericht zijn of niet – stuit men soms op complexe problemen die om een oplossing vragen. Dit kunnen zowel problemen zijn van operationele aard, die vooral met de dagelijkse gang van zaken te maken hebben, als problemen met een strategisch karakter. In het laatste geval betreft het meer de langetermijnbeslissingen. Typische voorbeelden van beslissingsproblemen zijn:

- Hoe moet het voorraadbeleid van een autodealer worden geformuleerd?
- Hoeveel kassa's moeten worden geopend bij een grote manifestatie teneinde de wachtrijen aan de kassa binnen redelijke proporties te houden?
- Welke keuze is het verstandigst als een bedrijf overweegt een kleine, middelgrote of zeer grote fabriek te bouwen voor een nieuw product waarvoor de toekomstige afzet nog onzeker is?

In al deze gevallen wil je aan de hand van de karakteristieken van het probleem uiteindelijk tot een rationele beslissing komen, die veelal een keuze is uit een aantal alternatieve oplossingen. Hierbij ligt in dit boek de focus op het benutten van de hulpmiddelen die ons vanuit de wiskunde en de statistiek worden aangereikt. In de volgende paragrafen gaan we hier nader op in.

1.2.1 Probleemoplossing in hoofdlijnen

Bij het toepassen van een oplossingsmethode op een bedrijfskundig probleem, kunnen we een aantal stappen onderscheiden.

1 **Allereerst moet het probleem worden geformuleerd. Wat willen we eigenlijk bereiken? Misschien zijn we op zoek naar een deugdelijk productieplan of willen we weten hoeveel voorraad we moeten aanhouden van een product. Belangrijk is dat wordt vastgesteld welke factoren een rol spelen bij de beheersing van het probleem. Dat zullen soms factoren van buitenaf zijn (die in het algemeen niet beïnvloedbaar zijn), zoals de vraag naar een product. Meestal spelen ook factoren binnen een organisatie een belangrijke rol, bijvoorbeeld hoeveel personeel er moet worden ingezet voor een bepaalde taak. Interne factoren zijn vaak wél beïnvloedbaar.**
2 **Heel belangrijk is dat een doel wordt geformuleerd waarmee we kunnen afmeten of een voorgestelde oplossing deugdelijk is. Zo zal men bij veel vraagstukken als doel tegenkomen dat de winst maximaal moet zijn of dat de kosten juist zo laag mogelijk moeten zijn. Ook denkbaar zijn doelen waarbij een bepaald serviceniveau wordt geëist. Denk bijvoorbeeld aan een servicenummer waarvoor men eist dat bellers geen wachttijd mogen hebben langer dan vijf minuten. Meestal werken we met één duidelijke doelstelling, maar er zijn ook technieken waarbij meer dan één doel tegelijkertijd wordt nagestreefd.**
3 **Een volgende stap is het formuleren van de samenhang tussen alle factoren die het probleem kenmerken. Er kan dan een model worden opgesteld, waarbij we al in gedachten houden dat een model doorgaans een vereenvoudigde weergave van de werkelijkheid is. Pas vanaf dat moment komt de echte probleemanalyse aan de orde.**
4 **Op basis van de geformuleerde doelstellingen kunnen we oplossingen zoeken voor het probleem, waarbij met name de beïnvloedbare variabelen een rol spelen. In deze fase moeten keuzes worden aangegeven en moet bekeken worden welke keuzes leiden tot de beste oplossing van het probleem, gegeven de gekozen doelstelling.**
5 **Is er eenmaal een (optimale) oplossing gevonden, dan moeten we een evaluatie maken van de gevolgde procedure. Dit kan met zich mee-**

brengen dat we het geformuleerde probleem en de daarmee samenhangende modelkeuze kritisch moeten bekijken. Soms leidt dit tot een aangepaste formulering van het probleem, waarna de gehele procedure van probleemoplossing opnieuw moet worden gevolgd.

6 De laatste fase is de implementatie van de gevonden oplossing. De werkelijke voordelen van een gevonden oplossing worden pas zichtbaar als men alles in de praktijk brengt. Soms blijkt dat bij de praktische toepassing van een aanpak die op papier superieur lijkt, nog allerlei factoren naar voren komen waarmee aanvankelijk geen rekening is gehouden. Aanpassing van het model is dan wellicht noodzakelijk.

De hier geschetste aanpak heeft een zeer globaal karakter. Wij richten ons vooral op een aantal karakteristieke problemen waarbij toepassing van kwantitatieve methoden centraal staat bij het vinden van oplossingen.

1.2.2 Toepassing van kwantitatieve methoden

Het gebruik van kwantitatieve methoden bij het analyseren van bedrijfskundige problemen is een ontwikkeling die in de twintigste eeuw op gang is gekomen. Ongeveer vanaf de Tweede Wereldoorlog spreekt men van een vakgebied dat wel wordt aangeduid als *Operations Research* (Amerikaans) of *Operational Research* (Engels). Ook wordt wel de naam *Management Science* gebruikt. Van dit vakgebied is het nogal moeilijk om precies aan te geven wat er wel en wat er niet onder valt. Het gaat er in ieder geval om dat de wiskundige en statistische methoden die van het etiket Operations Research voorzien zijn, gekenmerkt worden door de toepassingsmogelijkheden daarvan in bedrijven en organisaties. Hoe dan ook, er is een aantal methoden (zoals lineair programmeren, aanpak van wachttijdproblemen) dat traditioneel tot de Operations Research wordt gerekend. Dergelijke onderwerpen vormen de hoofdmoot van dit boek.
Daarnaast komen enkele kwantitatieve toepassingen aan de orde die niet tot de Operations Research gerekend worden, zoals kwaliteitszorg (hoofdstuk 8), inkoopmanagement (hoofdstuk 11) en financiële toepassingen (hoofdstuk 12).

De verschillende methoden die wij in de volgende hoofdstukken presenteren hebben het karakter van gereedschap, van een hulpmiddel waarvan men gebruik kan maken om gecompliceerde beslissingsproblemen doorzichtig te maken. Het zou beslist onjuist zijn om te denken dat het gebruik van kwantitatieve oplossingsmethoden de hoogst bereikbare wijsheid is. De échte problemen liggen veelal op een ander vlak. Zo zal men bij het in modelvorm brengen van een bepaald probleem heel wat moeilijkheden kunnen tegenkomen.
En bij het gebruik van kansen moet de onderzoeker soms een (subjectieve) invulling geven. Als dergelijke zaken vastliggen, dan levert de rekentechniek vaak wel een optimale oplossing. De vraag is echter of deze gevonden oplossing ook in de praktijk de juiste is. Dat hangt af van verschillende factoren, zoals: is het model een correcte weergave van de werkelijkheid? En zijn de kansen wel correct gespecificeerd?

Het is in de praktijk vaak nodig om niet één optimale oplossing te berekenen, er zal met het model 'gespeeld' moeten worden. Door ietwat verschillende specificaties voor het model te proberen, kan men onderzoeken hoe gevoelig de optimale oplossing is voor deze veranderingen.

1.2.3 **Soorten modellen**

Voordat we een bedrijfskundig probleem kunnen oplossen, moet er een formulering worden gemaakt van het probleem en van de relevante factoren die ermee verband houden.
Door middel van de constructie van een *model* willen we de werkelijkheid zo goed mogelijk beschrijven. Een model kan doorgaans niet alle karakteristieken van de werkelijkheid weergeven. Meestal beperkt men zich tot de karakteristieken die relevant zijn voor het probleem dat men bestudeert.
We onderscheiden twee hoofdsoorten modellen:
a fysieke modellen;
b abstracte modellen.

Ad a Fysieke modellen
Met fysieke modellen wordt een weergave van de werkelijkheid gemaakt waarbij de uiterlijke kenmerken van model en werkelijkheid overeenstemmen. Voorbeelden hiervan zijn maquettes, modelvliegtuigen, enzovoort. Vrijwel altijd zijn fysieke modellen te beschouwen als een schaalverandering van het object dat bestudeerd wordt.

Ad b Abstracte modellen
Bij abstracte modellen gaat het er niet om dat de werkelijkheid op een natuurgetrouwe wijze wordt afgebeeld, maar is het meer de bedoeling dat men de relevante eigenschappen zodanig beschrijft dat een buitenstaander een goede indruk kan krijgen van de samenhang tussen de verschillende onderdelen die van belang zijn.
We kennen hierbij de *verbale* modellen, waarbij door het geven van een omschrijving duidelijk wordt gemaakt hoe de werkelijkheid in elkaar zit, maar meestal gebeurt het door middel van *wiskundige* modellen.
Bij wiskundige modellen worden door middel van wiskundige uitdrukkingen de relaties tussen de verschillende karakteristieken van de werkelijkheid tot uitdrukking gebracht. Meestal kan men met een wiskundig model op een overzichtelijke wijze rekenen, waardoor het onder andere mogelijk wordt om verschillende ingrepen in de werkelijkheid te analyseren.

Deterministische modellen

Stochastische modellen

De wiskundige modellen onderscheiden we in een aantal typen. Er kan een tweedeling worden gemaakt in *deterministische* modellen en *stochastische* modellen. Bij stochastische modellen is er een aantal variabelen in het spel met een stochastisch of kanskarakter. Bij een deterministisch model daarentegen zijn er uitsluitend deterministische variabelen. Het hier aangebrachte onderscheid is van belang voor de manier waarop berekeningen met het model worden gemaakt.
Een voorbeeld van werken met een deterministisch model is de techniek van het lineair programmeren (zie de hoofdstukken 3 en 4). Bij voorraadproblemen kennen we zowel deterministische als stochastische modellen (zie hoofdstuk 6).

Heel belangrijk is het dat bij een gegeven probleem een wiskundige formulering wordt opgesteld waardoor het probleem overzichtelijker wordt en met beschikbare oplostechnieken kan worden aangepakt. Laten we eens een eenvoudig voorbeeld nemen.

■ Voorbeeld 1.1 Een kraam op het feestterrein

Ruud de Boer is een ondernemende student die bij een groot eendaags evenement op het feestterrein een klein kraampje wil huren voor de verkoop van ijs. De huur voor deze verkoopplek bedraagt 1500 euro voor die dag. Vooralsnog overweegt Ruud één soort ijs te verkopen. Dat zijn Liknums, met een inkoopprijs van €0,70. Ruud overweegt een verkoopprijs van €1,95 per Liknum. Hoeveel bedraagt de winst of het verlies van Ruud? Laten we eenvoudig beginnen. We geven het aantal Liknums dat Ruud inkoopt aan met X. Voor ons is X op dit moment een onbekend getal. We noemen dat een variabele. Stel dat Ruud alle ingekochte Liknums verkoopt. We zien dat de opbrengst voor Ruud per ijsje $1{,}95 - 0{,}70 =$ €1,25 bedraagt. Bij X Liknums is zijn winst dus $X \times (1{,}95 - 0{,}70) =$ €1,25 X euro. Daarvan moet hij nog de huur van de verkoopplek betalen. Hiermee rekening houdend bedraagt zijn winst W dus:

$$W = X \times (1{,}95 - 0{,}70) - 1500$$

Als we X weten, kunnen we de winst W berekenen. Dus W is ook een variabele. De waarde van W hangt af van de keuze van X. In dit geval is de relatie tussen W en X een lineaire functie.
We kunnen ons bijvoorbeeld afvragen hoeveel Liknums Ruud moet verkopen om precies quitte te draaien. Dan moet W de waarde 0 krijgen. We lossen dan de volgende vergelijking op:

$$0 = X \times (1{,}95 - 0{,}70) - 1500$$

Dat levert ons $X = 1\,200$. Deze waarde noemt men het *break-even point*. Dat is dus de waarde van X waarbij precies de kosten van de huur van de kraam worden terugverdiend. Zou X hier groter dan 1200 worden gekozen, dan maakt Ruud winst, bij een lagere waarde draait hij verlies.

Het probleem is tot nu toe erg eenvoudig en ook weinig realistisch. We kunnen allerlei uitbreidingen proberen.
Stel dat de opslagruimte van ijsjes beperkt is tot 3 000 stuks en dat de leverancier geen orders aanneemt voor minder dan 600 ijsjes. Dit kunnen we weergeven door te formuleren:
$X \leq 3\,000$ en $X \geq 600$ oftewel $600 \leq X \leq 3\,000$

Nadere onderhandelingen leveren dat Ruud meer producten mag verkopen, namelijk ook frisdrank en bier. Blikjes bier hebben een inkoopprijs van €0,80 en worden verkocht voor €2,00. Blikjes frisdrank kosten €0,60 inkoop en Ruud verkoopt deze voor €1,50. Het aantal blikjes bier noemen we Y en het aantal blikjes frisdrank geven we aan met de variabele Z.
Bier en frisdrank moeten worden bewaard in een grote koelkast. Hierin passen 2 000 blikjes bier of 2 500 blikjes fris, of een passende combinatie daarvan.

We kunnen die informatie om te beginnen verwerken in de winstfunctie:

$$W = X \times (1{,}95 - 0{,}70) + Y \times (2{,}00 - 0{,}80) + Z \times (1{,}50 - 0{,}60) - 1500$$

Ingewikkelder is het om de beperking van de opbergcapaciteit van de koelkast als vergelijking te formuleren. Stel dat we de inhoud van de koelkast (geheel willekeurig) 1 000 noemen.
Als we alleen bier zouden inkopen dan geldt:

$0{,}50\ Y \leq 1\,000$ (Ga maar na, bij Y is 2 000 is de koelkast vol.)

Als we alleen fris inkopen geldt

$0{,}40\ Z \leq 1\,000$

Bier en fris moeten in dezelfde koelkast, dus de realiteit is dat we deze twee voorwaarden moeten samennemen tot

$0{,}50Y + 0{,}40\ Z \leq 1\,000$

Hieruit blijkt dat allerlei combinaties van Y en Z mogelijk zijn. Uiteraard moeten Y en Z wél minstens nul zijn.

Aangezien Ruud er naar streeft een zo hoog mogelijke winst te halen, kunnen we het probleem tot dusver weergeven door het volgende model:

Maximaliseer:

$$W = X \times (1{,}95 - 0{,}70) + Y \times (2{,}00 - 0{,}80) + Z \times (1{,}50 - 0{,}60) - 1500$$

Onder de voorwaarden

$$\begin{aligned} X &\leq 3\,000 \\ X &\geq 600 \\ 0{,}50Y + 0{,}40\ Z &\leq 1\,000 \\ Y &\geq 0 \\ Z &\geq 0 \end{aligned}$$

Deze schrijfwijze komt overeen met een lineair programmeringsprobleem (zie hiervoor hoofdstuk 3 en 4).
Veel uitbreidingen zijn nu mogelijk. Wat is er te zeggen van de *vraag* naar ijs, bier en fris? Is deze onbeperkt, kortom kan Ruud alles verkopen wat hij ingekocht heeft? Ligt de vraag vast? Is de vraag naar die producten wellicht een kansvariabele die afhangt van het aantal bezoekers en het weer? We stellen in dit boek diverse varianten aan de orde.

1.2.4 Enkele voorbeelden van oplosmethoden

Het formuleren van een model is uiteraard niet voldoende om een besliskundig probleem op te lossen. Met het model is er wel een beeld totstandgekomen van de karakteristieken van de werkelijkheid. Onderdeel van een geformuleerd beslissingsprobleem dient altijd een doelstellingsfunctie te zijn. Pas als de doelstellingen duidelijk worden aangegeven, kan een aanvang worden gemaakt met het zoeken naar een zo gunstig mogelijke oplossing. Afhankelijk van het soort model waarmee gewerkt wordt, zijn er verschillende oplosmethoden te onderscheiden.

Analytische methode

Als eerste soort noemen we de analytische methode. Bij toepassing van de analytische methode worden de wiskundige vergelijkingen uit het model bewerkt, zodat er een oplossing kan worden berekend.

■ Voorbeeld 1.2

Bij de bestudering van voorraadproblemen wordt uitgezocht hoe groot een periodieke voorraadaanvulling moet zijn bij een onderneming. Hierbij spelen twee soorten kosten een rol. Enerzijds zijn er de bewaarkosten (of opslagkosten) van de producten die in voorraad worden gehouden. Als we per jaar heel veel kleine voorraadaanvullingen doen, dan houden we het voorraadniveau in het magazijn laag, dus dat lijkt gunstig. Anderzijds hebben de dan een heleboel keer per jaar het gedoe van een bestelling plaatsen en afgeleverd krijgen. Dus die kosten nemen juist toe. Bij het eenvoudigste voorraadprobleem[1] wordt gezocht naar de optimale bestelgrootte Q, die ertoe leidt dat het geheel aan voorraadkosten (bijvoorbeeld op jaarbasis berekend) zo laag mogelijk wordt. Analyse van het deterministische voorraadmodel leidt tot de vergelijking voor de voorraadkosten K:

$$K = V \times P + \frac{V}{Q} \times B + \frac{Q}{2} \times P \times R$$

De kunst is nu om voor de variabele Q hier een dusdanige waarde te kiezen dat de waarde van K zo laag mogelijk wordt.

Een tweede oplossingsmethode is de iteratieve methode. De iteratieve methode is in feite een speciaal soort wiskundige methode waarbij een oplossing niet door in één stap wordt gevonden, maar waarbij een aantal oplossingen van het probleem elkaar opvolgt tot de optimale oplossing bereikt is. Een typisch voorbeeld hiervan treffen we aan bij de ingewikkelder lineaire programmeringsproblemen.
Eerst worden achtereenvolgens een aantal niet-optimale tussenoplossingen berekend. Het zoeken naar een nog betere mogelijkheid stopt indien een oplossing is bereikt waarvan men kan aantonen dat er geen verbetering meer mogelijk is.
Als een derde soort oplossingsmethode van besliskundige problemen noemen we de *simulatiemethode*.
Bij het gebruik van de simulatiemethode vindt er een nabootsing van de werkelijkheid plaats. Dat kan vaak bij modellen waarbij variabelen spelen die een kanskarakter hebben. Door middel van bepaalde lotingsmethoden worden er dan waarden toegekend aan de betrokken variabelen. Een voorbeeld van de simulatiemethode vinden we bij ingewikkelde wachttijdproblemen. Hierbij kan soms een groot aantal kansvariabelen meespelen, waardoor het bijna niet mogelijk is om via berekeningen een oplossing te vinden. Simulatie van het probleem kan dan uitkomst bieden. Vaak wordt er bij het toepassen van deze methode de gebruikgemaakt van een computer.

1.3 Enkele begrippen uit de kansrekening

Kansrekening wordt op ruime schaal toegepast binnen de besliskunde. Om die reden bespreken we hier enkele begrippen uit dit vakgebied beknopt.
Allereerst geven we aandacht aan het kansbegrip en aan enkele kansregels. Vervolgens bespreken we kansvariabelen in het algemeen en als speciale gevallen de normale verdeling, de binomiale verdeling en de Poisson-verdeling.

1 Zie voor de betekenis van de symbolen hoofdstuk 6.

1.3.1 Kansdefinities en kansregels

Wat is eigenlijk een kans? In de kansrekening wordt gesproken van *gebeurtenissen* die door verzamelingen worden weergegeven. Met $P(A)$ wordt aangegeven de kans op het optreden van de gebeurtenis A. De gebeurtenis A kunnen we weergeven door een verzameling A, die te beschouwen is als een deelverzameling van S, de verzameling van alle mogelijke uitkomsten. Er zijn verschillende manieren waarop het begrip 'kans' kan worden geïntroduceerd. We noemen hier:

1 *de kansdefinitie van Laplace*
2 *de experimentele kansdefinitie*
3 *de axiomatische kansdefinitie*
4 *de subjectieve kansdefinitie.*

Laplace

Ad 1 De kansdefinitie van Laplace
Bij de kansdefinitie van Laplace is het uitgangspunt dat alle mogelijke uitkomsten gelijke kans hebben. De kans $P(A)$ op gebeurtenis A wordt dan gegeven als:

$$P(A) = \frac{\text{aantal gunstige gevallen}}{\text{totaal aantal gevallen}}$$

Experimentele kansdefinitie

Ad 2 De experimentele kansdefinitie
Bij deze benaderingswijze wordt de kans op een gebeurtenis bepaald door een reeks experimenten te doen waarbij telkens wordt vastgesteld of de gebeurtenis al dan niet is opgetreden. De kans $P(A)$ op de gebeurtenis A wordt dan gegeven door $P(A) = \lim\limits_{n \to \infty} \dfrac{n_A}{n}$ waarbij n_A het aantal gunstige gevallen is en n het aantal pogingen.

Axiomatische kansdefinitie

Ad 3 De axiomatische kansdefinitie
Bij toepassing van de axiomatische definitie wordt het begrip 'kans' gedefinieerd als een functie $P(\ )$ die aan gebeurtenissen een getal toevoegt. Dit getal noemen we de kans op de betreffende gebeurtenis. De functie $P(\ )$ moet voldoen aan bepaalde eisen: de kansaxioma's.

Subjectieve kansdefinitie

Ad 4 De subjectieve kansdefinitie
Bij toepassing van de subjectieve definitie begeven we ons enigszins op glad ijs. Kansen worden in dat geval niet door middel van een experiment of iets dergelijks bepaald, maar op subjectieve gronden toegekend aan allerlei gebeurtenissen. Vaak moeten we deze definitie gebruiken als er uitspraken moeten worden gedaan over eenmalige gebeurtenissen die in de toekomst liggen.

■ Voorbeeld 1.3
Een investeerder die overweegt een fabriek in Brazilië te bouwen, moet bij het nemen van een beslissing hierover overwegen welke financiële resultaten er van zo'n investering te verwachten zijn. Omdat er veel onzekere factoren in het spel zijn, lijkt het verstandig om niet één vaste getalwaarde op te geven als verwachte uitkomst, maar een reeks mogelijke uitkomsten waaraan kansen verbonden zijn (zie hiervoor hoofdstuk 2: Beslissingstheorie). Dit is een voorbeeld van het werken met subjectieve kansen.

■ **Voorbeeld 1.4**
In een bijkantoor van een bank overweegt men een extra loket aan te brengen om ongewenst lange wachttijden te bestrijden. Om deze beslissing te onderbouwen, wordt gedurende een bepaalde periode onderzocht hoe het aankomstpatroon van de klanten is en hoe het patroon van de bedientijd van de klanten is. In een dergelijke situatie worden er kansverdelingen van de aankomsttijden en bedieningstijden bepaald op basis van waargenomen uitkomsten. Dit is een voorbeeld van het experimenteel bepalen van kansen.

Kansregels

Om te kunnen rekenen met allerlei kansen gebruiken we een aantal kansregels. We noemen:

1 Voor de verzameling S van *alle* uitkomsten geldt: $P(S) = 1$. Met andere woorden, de kans om bij een loting een uitkomst te verkrijgen die behoort tot de verzameling S bedraagt 1, want S is alles.
2 Voor de lege verzameling geldt: $P(\varnothing) = 0$.

Complementregel

3 Verder kennen we de *complementregel*: $P(\overline{A}) = 1 - P(A)$. Hierbij is $\overline{A}$ de verzameling niet-A, dat is de verzameling elementen van S die niet tot de verzameling A behoren.

Algemene optelregel

4 De *algemene optelregel* voor kansen luidt:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Hierbij staat $A \cup B$ voor de vereniging van de verzamelingen A en B en de schrijfwijze $A \cap B$ geeft de doorsnede van de verzamelingen A en B aan.

Speciale optelregel

5 De *speciale* optelregel geldt voor de samenvoeging van twee verzamelingen A en B indien deze een lege doorsnede hebben. Er geldt dan:

$$P(A \cup B) = P(A) + P(B)$$

Voorwaardelijke kans

6 De voorwaardelijke kans $P(A|B)$ is gedefinieerd als:

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

Hierbij symboliseert $P(A|B)$ de kans op de gebeurtenis A onder de voorwaarde van het optreden van de gebeurtenis B.

7 Twee gebeurtenissen A en B worden *onderling onafhankelijk* genoemd indien $P(A|B) = P(A)$. Dat wil zeggen, het al dan niet optreden van de gebeurtenis B is niet van invloed op de kans op gebeurtenis A. Met behulp van de definitie van voorwaardelijke kansen is te zien dat in dit geval geldt:

$$P(A \cap B) = P(A) \times P(B)$$

Dit wordt wel de *speciale productregel* genoemd.

■ **Voorbeeld 1.5**
Een importeur van een automerk handelt in drie modellen, namelijk de Astra, de Beta en de Celia. Van de geïmporteerde auto's bestaat 50% uit Astra's, 20% uit Beta's en 30% uit Celia's.

a Er komt een (willekeurige) bestelling binnen. Hoe groot is de kans dat dit voor een Astra is?
De gegevens maken duidelijk dat $P(\text{Astra}) = 0{,}50$.

b Hoe groot is de kans dat dit een Astra of een Celia is?
Omdat de doorsnede van de verzameling Astra's en Celia's leeg is, geldt de speciale optelregel: $P(\text{Astra} \cup \text{Celia}) = 0{,}50 + 0{,}30 = 0{,}80$.

c Van alle Astra's is 20% zwart, van de Beta's is dat 30% en van de Celia's is 10% zwart. Hoe groot is de kans dat een willekeurige auto zwart is?
We maken nu gebruik van de definitie van voorwaardelijke kansen:

$$P(\text{zwart}) = P(\text{zwart} \cap \text{Astra}) + P(\text{zwart} \cap \text{Beta}) + P(\text{zwart} \cap \text{Celia}) =$$
$$P(\text{zwart}|\text{Astra}) \times P(\text{Astra}) + P(\text{zwart}|\text{Beta}) \times P(\text{Beta}) + P(\text{zwart}|\text{Celia}) \times$$
$$P(\text{Celia}) = 0{,}20 \times 0{,}50 + 0{,}30 \times 0{,}20 + 0{,}10 \times 0{,}30 = 0{,}19$$

Dus 19% van alle auto's is zwart.

Regel van Bayes

Regel van Bayes
In nauw verband met de definitie van voorwaardelijke kansen staat de regel van Bayes. Voor een tweetal gebeurtenissen *A* en *B* definiëren we:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ en } P(B|A) = \frac{P(B \cap A)}{P(A)}$$

We kunnen deze formules ook schrijven als:

$$P(A \cap B) = P(A|B)P(B) \text{ en}$$
$$P(B \cap A) = P(B|A)P(A)$$

Aangezien $A \cap B$ en $B \cap A$ twee schrijfwijzen voor dezelfde doorsnede zijn, zijn de rechterkanten van deze twee uitdrukkingen gelijk:

$$P(B|A)P(A) = P(A|B)P(B), \text{ dus}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \text{ (regel van Bayes)}$$

Deze uitdrukking wordt wel de 'stelling van Bayes' genoemd. Indien de kansen $P(A)$, $P(B)$ en $P(A|B)$ gegeven zijn, dan kan $P(B|A)$ worden berekend.

■ Voorbeeld 1.6
Met de gegevens van voorbeeld 1.5 kunnen we met de regel van Bayes uitrekenen hoe groot de kans is dat een willekeurige auto een Celia is, gegeven de informatie dat deze zwart is. Dus:

$$P(\text{Celia}|\text{zwart}) = \frac{P(\text{Celia} \cap \text{zwart})}{P(\text{zwart})} = \frac{P(\text{zwart}|\text{Celia}) \times P(\text{Celia})}{P(\text{zwart})} =$$
$$\frac{0{,}10 \times 0{,}30}{0{,}19} = 0{,}158$$

We zien hier dat er aanvankelijk een kans van 0,30 was dat een willekeurige auto een Celia is. De informatie dat vaststaat dat de auto zwart is, verandert echter iets aan de kans. Omdat relatief weinig Celia's zwart zijn, is de kans op een Celia gedaald van 0,30 naar 0,16.

De regel van Bayes kan in de praktijk worden toegepast bij het verwerken van onvolledige informatie die beschikbaar is gekomen met betrekking tot het optreden van een bepaalde gebeurtenis.

A priori kans

Men spreekt in dit verband wel van de *a priori kans* $P(B)$ die na het optreden van de gebeurtenis A vervangen wordt door de voorwaardelijke kans $P(B|A)$: 'de kans op B onder de voorwaarde dat gebeurtenis A is opgetreden'.

■ **Voorbeeld 1.7**

In een fabriek staat een vulmachine opgesteld waarmee verpakkingen van een product van inhoud worden voorzien. Een controleur gaat regelmatig na of de afgeleverde verpakkingen aan de gewichtseisen voldoen. Als dit niet het geval is, dan moet de vulmachine worden stilgezet en het vulgewicht opnieuw worden ingesteld. Op grond van langdurige ervaring weet de controleur dat er een kans van 0,98 is dat de machine goed staat afgesteld, terwijl er een kans van 0,02 is dat de machine ontregeld is.
Er geldt dus voor de afstelling van de machine:
$P(goed) = 0{,}98$
$P(fout) = 0{,}02$

Dit zijn de zogenoemde a priori kansen.
Bij de controle van het vulgewicht van een verpakking worden er bepaalde keuringsgrenzen gehanteerd. Indien het vulgewicht binnen de keuringsgrenzen valt, gaat men ervan uit dat de machine goed is ingesteld. Is er een gewicht buiten de keuringsgrenzen waargenomen, dan wordt de vulmachine gestopt en opnieuw ingesteld.
Stel dat we (op basis van ervaring) beschikken over de volgende informatie:
De kans dat een vulgewicht buiten de keuringsgrenzen valt als de machine correct is ingesteld, bedraagt 0,04. De kans op een vulgewicht buiten de keuringsgrenzen bij een ontregelde vulmachine blijkt 0,90 te zijn. We kunnen dit als volgt met voorwaardelijke kansen weergeven:
$\mathrm{P(afkeuren|goed)} = 0{,}04$
$\mathrm{P(afkeuren|fout)} = 0{,}90$

Als de controleur vaststelt dat het gewicht van de door hem willekeurig gekozen verpakking buiten de keuringsgrenzen valt, hoe groot is dan de kans dat de machine ontregeld is?
Gevraagd is hier dus: $P(fout|\mathrm{afkeuren})$
Met behulp van de regel van Bayes berekenen we de gevraagde kans. Allereerst berekenen we voor een willekeurig exemplaar de totale kans op het aantreffen van een vulgewicht buiten de keuringsgrenzen.
Hiervoor geldt:

$$P(\mathrm{afkeuren}) = P(\mathrm{afkeuren|goed}) \times P(\mathrm{goed}) + P(\mathrm{afkeuren|fout}) \times P(\mathrm{fout})$$
$$= 0{,}04 \times 0{,}98 + 0{,}90 \times 0{,}02 = 0{,}0572$$

Voor de gevraagde kans vinden we dan:

$$P(\mathrm{fout|afkeuren}) = \frac{P(\mathrm{afkeuren|fout})P(\mathrm{fout})}{P(\mathrm{afkeuren})} = \frac{0{,}90 \times 0{,}02}{0{,}0572} = 0{,}315$$

Opmerkingen

- Op basis van de berekeningen is komen vast te staan dat de a priori kans op fout van 0,02 in het licht van de beschikbaar gekomen informatie leidt tot de veel grotere kans op fout van 0,315.
- Uit deze berekening blijkt dan tevens dat de kans dat de machine goed is ingesteld nog steeds vrij groot is, namelijk 0,685 (= 1 − 0,315). Als de vulmachine inderdaad wordt stopgezet op basis van één gecontroleerde verpakking waarvan de inhoud is gecontroleerd, dan gebeurt dit stopzetten dus vaak ten onrechte. Het is daarom het overwegen waard om na het aantreffen van een gewicht buiten de keuringsgrenzen nog enkele extra verpakkingen te keuren. Op basis van de hierbij aangetroffen resultaten kunnen we door toepassing van de regel van Bayes vervolgens een soortgelijke berekening maken van P(goed) en P(fout).

1.3.2 Kansvariabelen

In nauwe relatie tot de kansrekening staat het begrip *kansvariabele* of stochastische variabele. Bij een kansvariabele gaat het om een grootheid waarvan de uitkomsten (reële) getallen zijn en waarbij de kansrekening een rol speelt bij het vaststellen hoe vaak een bepaalde uitkomst kan optreden. We maken een onderscheid in twee categorieën, namelijk *discrete* kansvariabelen, die we met het symbool $\underline{k}$ aangeven, en *continue* kansvariabelen, waarvoor we het symbool $\underline{x}$ gebruiken.
Bij discrete kansvariabelen is de verzameling getallen die als uitkomst kan optreden in beginsel beperkt tot een eindig aantal.[2] Voorbeelden van een discrete kansvariabele zijn: het aantal klanten dat in een winkel aanwezig is, het aantal geslaagden bij een examen waaraan vijftig personen meedoen, gokspelletjes waarbij een aantal verschillende uitbetalingen mogelijk is, enzovoort.

Discrete kansvariabelen

In beginsel wordt een discrete kansvariabele volledig beschreven door een lijstje te maken van alle denkbare uitkomsten met vermelding van de daarbij behorende kansen.

■ Voorbeeld 1.8

Bij een gokspelletje zijn de mogelijke uitbetalingen €0, €5, €10 en €20. De kansen op het optreden daarvan zijn respectievelijk 0,50, 0,30, 0,15 en 0,05. De kansen worden als volgt genoteerd: $P(\underline{k} = 0) = 0{,}50$, $P(\underline{k} = 5) = 0{,}30$, $P(\underline{k} = 10) = 0{,}15$ en $P(\underline{k} = 20) = 0{,}05$.
We hebben hier te maken met de kansvariabele $\underline{k}$ die het aantal uit te betalen euro's aangeeft. Merk op dat $\underline{k}$ de *naam* van de variabele aangeeft. Daarentegen geeft k zonder streepje één van de mogelijke *uitkomsten* aan van de variabele $\underline{k}$.

Continue kansvariabelen

Naast discrete kansvariabelen kennen we continue kansvariabelen. Bij continue variabelen zijn er in principe oneindig veel uitkomsten die als uitkomst van een kansexperiment kunnen optreden. Typische voorbeelden hiervan zijn het gewicht van een voorwerp, de lengte van een persoon en de tijdsduur die er verstrijkt tussen twee opeenvolgende binnenkomsten van een klant in een winkel.

2 Er zijn voorbeelden van discrete variabelen te geven waarbij het aantal mogelijke uitkomsten aftelbaar oneindig is.

Dergelijke variabelen kunnen in beginsel oneindig veel verschillende waarden laten zien. In de praktijk is er echter de beperking van de meetapparatuur. Iedere meetmethode, hoe nauwkeurig ook, leidt tot het gebruik van afgeronde waarden bij het waarnemen van een continue kansvariabele.
Een continue kansvariabele $\underline{x}$ wordt beschreven door de kansdichtheid $f(x)$. Dit is een functie die aan bepaalde eisen voldoet waardoor kansen als een oppervlakte onder de grafiek kunnen worden berekend (meestal door middel van integreren). Kenmerkend voor iedere kansdichtheid is dat de totale oppervlakte onder de grafiek de waarde 1 moet hebben (de som van de kansen moet 1 zijn).

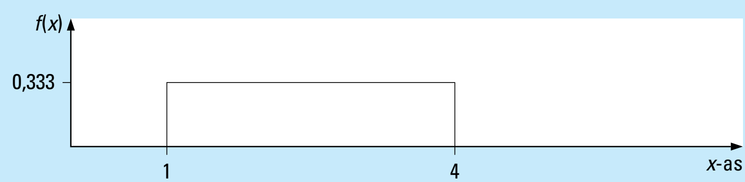
Een typisch voorbeeld van een continue kansvariabele is de uniforme of rechthoekige verdeling. Bij deze verdeling kunnen uitkomsten worden waargenomen van een interval (a, b) op de x-as. De kansdichtheid $f(x)$ heeft overal dezelfde waarde op dit interval en daarbuiten is de waarde nul. De grafiek is dus overal even hoog op het interval (a,b).

■ **Voorbeeld 1.9**
Bij een benzinestation kan de tijd die een auto nodig heeft om te tanken worden beschreven door een rechthoekige verdeling met een benedengrens 1 minuut en bovengrens 4 minuten. Daarmee is het interval (a, b) gelijk aan (1,4). Omdat de breedte van dit interval gelijk is aan 3 heeft de kansdichtheid de waarde 1/3. Dit levert:

$$f(x) = \begin{cases} 0 & \text{voor } x < 1 \text{ of } x > 4 \\ 1/3 & \text{voor } 1 \leq x \leq 4 \end{cases}$$

als rechthoekige verdeling op het interval (1,4).

Figuur 1.1 **De rechthoekige verdeling**



De kans dat een willekeurige klant de benodigde tijdsduur tussen de 2 en 3 minuten en 30 seconden bezig is, kunnen we berekenen als de oppervlakte onder de grafiek tussen 2 en 3,5 is. Formeel kan dit door de integraal te berekenen:

$$P(2 \leq x \leq 3{,}5) = \int_2^{3,5} f(x)dx = \int_2^{3,5} \frac{1}{3}dx = \frac{1}{3}x\Big|_2^{3,5} = 0{,}5$$

Een nuttige grootheid bij het werken met kansvariabelen is de verdelingsfunctie of cumulatieve kans functie $F(\ )$. In het discrete geval is deze gedefinieerd als:

$$F(k) = P(\underline{k} \leq k) = \sum_{m \leq k} P(\underline{k} = m)$$

En in het continue geval geldt een soortgelijke definitie:

$$F(x) = P(\underline{x} \leq x) = \int_{-\infty}^{x} f(x)\mathrm{d}x$$

De verdelingsfunctie geeft de kans aan dat de variabele een waarde aanneemt beneden of gelijk aan een bepaalde grenswaarde.

Verwachtingswaarde en variantie

Voor kansvariabelen is het van belang dat men een indruk kan krijgen van de gemiddelde waarde van de mogelijke uitkomsten. We noemen dat de *verwachtingswaarde* $E(\underline{k})$.

Verwachtingswaarde

Voor een discrete kansvariabele $\underline{k}$ hanteren we de volgende definitie:

$$E(\underline{k}) = \sum k\, P(\underline{k} = k)$$

Met deze formule wordt er een gewogen gemiddelde berekend van alle uitkomsten k. De bijbehorende kansen doen dienst als wegingscoëfficiënten.

■ Voorbeeld 1.10

Op grond van ervaring is komen vast te staan dat het aantal machinestoringen per dag in een fabriek kan worden weergegeven door de kansvariabele $\underline{k}$, waarvoor geldt $P(\underline{k} = 0) = 0{,}2$; $P(\underline{k} = 1) = 0{,}4$; $P(\underline{k} = 2) = 0{,}3$ en $P(\underline{k} = 3) = 0{,}1$.
De verwachtingswaarde van het aantal storingen bedraagt dus:
$E(\underline{k}) = 0 \times 0{,}2 + 1 \times 0{,}4 + 2 \times 0{,}3 + 3 \times 0{,}1 = 1{,}3$ storingen per dag

Een ietwat andere toepassing van het begrip verwachtingswaarde vinden we in het volgende voorbeeld.

■ Voorbeeld 1.11

Een marktkoopman van pluimvee weet dat de vraag naar kalkoenen op de wekelijkse marktdag kan worden beschouwd als een kansvariabele $\underline{k}$ die waarden aanneemt tussen 21 en 30, waarbij alle uitkomsten 21, 22, … 30 dezelfde kans hebben om voor te komen.[3]
Omdat de som van de kansen gelijk aan 1 moet zijn, volgt hieruit dat al die kansen gelijk aan 0,10 zijn. De verwachtingswaarde is eenvoudig te berekenen met:

$$E(\underline{k}) = \sum k\, P(\underline{k} = k) = 25{,}50$$

Stel nu dat de koopman op een marktdag begint met een voorraad $b = 24$ kalkoenen. Hoe groot is de kans dat de (onbekende) vraag hoger is dan de beschikbare voorraad b? Dit is eenvoudig te beantwoorden. We moeten dan berekenen $P(\underline{k} > 24) = P(\underline{k} \geq 25) = P(\underline{k} = 25) + P(\underline{k} = 26) + \ldots + P(\underline{k} = 30)$.
Dus dat is hier gelijk aan $6 \times 0{,}10 = 0{,}60$.
Per verkochte kalkoen verdient de marktkoopman €7. Hoe groot is de verwachte winst die hij misloopt omdat hij (afhankelijk van de werkelijk optredende vraag) soms niet volledig aan de vraag kan voldoen?

3 We zouden dit de discrete versie van de rechthoekige verdeling kunnen noemen (vergelijk met voorbeeld 1.9).

De misgelopen winst is nul indien de vraag 24 of lager is, want dan kan aan alle klanten geleverd worden. Daarboven is de gemiste winst $7 \times (k - 24)$. Per uitkomst k moeten we dit vermenigvuldigen met de kans dat dit voorkomt. We berekenen dus de som:
$7 \times (25 - 24) \times 0{,}10 + 7 \times (26 - 24) \times 0{,}10 + \ldots + 7 \times (30 - 24) \times 0{,}10 =$
$0{,}70 + 1{,}40 + 2{,}10 + 2{,}80 + 3{,}50 + 4{,}20 =$ €14,70

Verliesfunctie

We kunnen dit een verliesfunctie noemen. Deze geeft voor een gekozen waarde van b de verwachtingswaarde aan van het verlies dat ontstaat als de gevraagde hoeveelheid hoger uitkomt dan de beschikbare capaciteit of voorraad. Ook hadden we eerst de verwachtingswaarde van het aantal niet te leveren kalkoenen kunnen berekenen (dat is 2,10 stuks – ga dit na). Vervolgens dient dit verwachte aantal vermenigvuldigd te worden met 7, het misgelopen bedrag aan winst per kalkoen.

Variantie

De mate van spreiding die de afzonderlijke uitkomsten van de variabele $\underline{k}$ vertonen ten opzichte van de verwachtingswaarde kan worden uitgedrukt door middel van de variantie. De definitie hiervoor luidt:

$$\mathbf{Var}(\underline{k}) = \sum (k - E(\underline{k}))^2 P(\underline{k} = k)$$

Met deze formule wordt een gewogen gemiddelde berekend van de kwadratische afwijkingen ten opzichte van $E(\underline{k})$. De kansen $P(\underline{k} = k)$ doen dienst als wegingscoëfficiënten.

■ Voorbeeld 1.12

Voor het aantal machinestoringen per dag $\underline{k}$ uit voorbeeld 1.10 berekenen we de variantie.
$Var(\underline{k}) = \Sigma (k - E(\underline{k}))^2\, P(\underline{k} = k)$ levert na invulling van $E(\underline{k}) = 1{,}3$ en de kansen $P(\underline{k} = k)$ het volgende:

$$Var(k) = (0 - 1{,}3)^2 \times 0{,}2 + (1 - 1{,}3)^2 \times 0{,}4 + (2 - 1{,}3)^2 \times 0{,}3 + (3 - 1{,}3)^2 \times 0{,}1 = 0{,}81$$

In samenhang met het begrip variantie kennen we als maatstaf voor spreiding de standaarddeviatie. Deze wordt meestal aangegeven met het symbool σ en is gedefinieerd als de wortel uit de variantie. Dus $\sigma(k) = \sqrt{Var(\underline{k})}$.

Voor de gegevens uit voorbeeld 1.12 vinden we dus:

$$\sigma(\underline{k}) = \sqrt{0{,}81} = 0{,}9$$

Een hoge waarde van σ betekent dat de uitkomsten veel spreiding ten opzichte van elkaar hebben. Bij een lage σ liggen de uitkomsten juist dicht bij elkaar.

Ook voor continue kansvariabelen zijn er definities te geven van verwachtingswaarde en variantie. Uiteraard zijn deze begrippen nu niet door middel van sommaties te definiëren maar moet er gebruikgemaakt worden van integralen. We definiëren voor de continue kansvariabele $\underline{x}$, die gegeven is door de kansdichtheid $f(x)$, de verwachtingswaarde $E(\underline{x})$ als volgt:

$$E(\underline{x}) = \int x f(x)\, dx$$

Als f(x) bekend is, kan men hiermee het gemiddelde van de uitkomsten van $\underline{x}$ berekenen.

■ Voorbeeld 1.13

Aan een machine moet regelmatig onderhoud worden verricht. De tijdsduur van een willekeurige onderhoudsbeurt kunnen we beschouwen als een kansvariabele $\underline{x}$, waarvoor als kansdichtheid (die op basis van langdurige ervaring is vastgesteld) geldt:

$f(x) = 0$ voor $x < 0$ of $x > 2$
$f(x) = 1 - \frac{1}{2}x$ voor $0 \leq x \leq 2$

De tijdsduur $\underline{x}$ wordt weergegeven in uren.
Hoe groot is de verwachtingswaarde van de tijdsduur van een onderhoudsbeurt? We passen de definitie van verwachtingswaarde voor continue kansvariabelen toe. Dit leidt voor de hier toepasselijke kansdichtheid tot:

$$\begin{aligned} E(\underline{x}) &= \int_0^2 x\,(1 - \tfrac{1}{2}x)\mathrm{d}x \\ &= (\tfrac{1}{2}x^2 - \tfrac{1}{6}x^3)\Big|_0^2 = \tfrac{2}{3} \text{ uur} \end{aligned}$$

De verwachtingswaarde van de tijdsduur luidt dus 40 minuten.

Voor de variantie geldt bij continue kansvariabelen de volgende definitie:

$$Var(\underline{x}) = \int (x - E(\underline{x}))^2 f(x)\mathrm{d}x \quad \text{of} \quad Var(\underline{x}) = E(\underline{x}^2) - E(\underline{x})^2$$

In sommige gevallen is deze laatste vorm handiger toepasbaar.

■ Voorbeeld 1.14

Voor de kansvariabele uit het vorige voorbeeld berekenen we de variantie:

$$Var(\underline{x}) = \int_0^2 (x - \tfrac{2}{3})^2 f(x)\mathrm{d}x = \int_0^2 (x - \tfrac{2}{3})^2(1 - \tfrac{1}{2}x)\mathrm{d}x = \tfrac{2}{9}$$

Voor de praktijk van groot belang zijn de opteleigenschappen van verwachtingswaarde en variantie.
Indien er gegeven zijn twee kansvariabelen $\underline{x}$ en $\underline{y}$ met verwachtingswaarden $E(\underline{x})$ en $E(\underline{y})$ en varianties $Var(\underline{x})$ en $Var(\underline{y})$, dan geldt voor de somvariabelen $\underline{z} = \underline{x} + \underline{y}$ dat
$E(\underline{z}) = E(\underline{x}) + E(\underline{y})$ en
$Var(\underline{z}) = Var(\underline{x}) + Var(\underline{y})$.
De variabele $\underline{z}$ is hierbij de som van een trekking van de variabele $\underline{x}$ en van een trekking van de variabele $\underline{y}$. De opteleigenschap voor *varianties* geldt alleen als we de trekkingen kunnen beschouwen als onderling onafhankelijke kansvariabelen. Een veel voorkomende toepassing hiervan ontmoeten we als er een aantal malen een waarneming gedaan wordt van *dezelfde* kansvariabele. Voor de som van de verkregen uitkomsten gelden de bovengenoemde optelregels.

■ **Voorbeeld 1.15**
Bij een handel in veevoeder is het aantal kilogrammen dat in een week van een product wordt afgezet te beschouwen als een kansvariabele $\underline{x}$ met verwachtingswaarde 2000 kg/week en standaarddeviatie 400 kg. Voor een periode van vier weken moet er voldoende voorraad worden ingekocht. Wat kunnen we zeggen over de nog onbekende afzet in die vier weken? We symboliseren de afzet in een periode van vier weken met de kansvariabele $\underline{x}_{\text{totaal}}$. Deze kansvariabele is gedefinieerd als

$\underline{x}_{\text{totaal}} = \underline{x}_1 + \underline{x}_2 + \underline{x}_3 + \underline{x}_4$ waarbij $\underline{x}_i$ de afzet in week *i* aangeeft.

Voor de verwachtingswaarde van $\underline{x}_{\text{totaal}}$ geldt:

$$\begin{aligned} E(\underline{x}_{\text{totaal}}) &= E(\underline{x}_1) + E(\underline{x}_2) + E(\underline{x}_3) + E(\underline{x}_4) \\ &= 2\,000 + 2\,000 + 2\,000 + 2\,000 = 8\,000 \text{ kg} \end{aligned}$$

Voor de variantie van $\underline{x}_{\text{totaal}}$ geldt:

$$\begin{aligned} Var(\underline{x}_{\text{totaal}}) &= Var(\underline{x}_1) + Var(\underline{x}_2) + Var(\underline{x}_3) + Var(\underline{x}_4) \\ &= 400^2 + 400^2 + 400^2 + 400^2 = 4 \times 400^2 \end{aligned}$$

Dus $\sigma(\underline{x}_{\text{totaal}}) = 800$ kg.

De optelregel voor varianties geldt als we mogen aannemen dat we de omzetten $\underline{x}_i$ in de verschillende weken als onderling onafhankelijke kansvariabelen mogen aanmerken. (Merk op dat we de varianties bij elkaar tellen en niet de standaarddeviaties!)

1.3.3 Enkele specifieke kansverdelingen

In kort bestek bespreken we hier enkele verdelingen van kansvariabelen die voor de praktijk van belang kunnen zijn. Achtereenvolgens komen de normale verdeling, de binomiale verdeling en de Poissonverdeling aan de orde.

Normale verdeling
Gegeven is een (continue) kansvariabele $\underline{x}$ met als dichtheidsfunctie:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma 2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

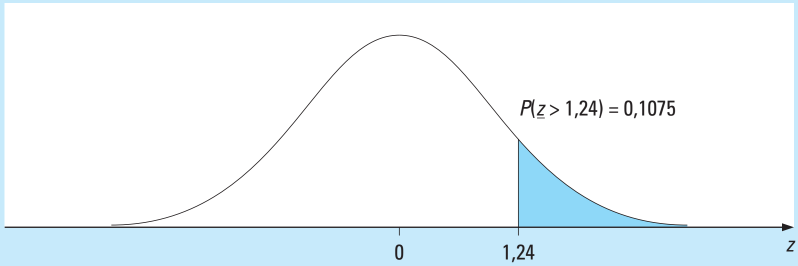
Normale verdeling

Deze kansvariabele hoort bij de een *normale* verdeling. In de formule komen twee parameters voor, namelijk μ en σ. Van μ kan worden aangetoond dat dit de verwachtingswaarde van de variabele $\underline{x}$ aangeeft, terwijl σ de standaarddeviatie van $\underline{x}$ is.

Standaardnormale verdeling

Bij continue kansvariabelen moeten kansen als oppervlakten worden berekend. Dat zijn dus integralen. Omdat de berekening daarvan in dit geval geen eenvoudige zaak is, wordt er bij het bepalen van de kansen gewerkt met een tabel: de tabel van de *standaardnormale* verdeling. Deze verdeling verkrijgt men door voor de parameter μ de waarde 0 te kiezen en voor de parameter σ de waarde 1.

Figuur 1.2 **De standaardnormale verdeling**



In de tabel A van de standaardnormale verdeling (zie pagina 340) kunnen kansen worden afgelezen. Deze tabel is zodanig geconstrueerd dat daarin zogenaamde rechteroverschrijdingskansen (staartoppervlakten) kunnen worden afgelezen. Voor de overschrijdingskans van $z = 1{,}24$ lezen we in de tabel 0,1075 af. Dus $P(z \geq 1{,}24) = 0{,}1075$.
De tabel van de standaardnormale verdeling kan ook voor willekeurige normale verdelingen worden toegepast. Een kansvariabele $\underline{x}$ met een normale verdeling met verwachtingswaarde μ en standaarddeviatie σ moet getransformeerd worden tot een variabele z, waarvoor geldt $\mu = 0$ en $\sigma = 1$.
In het volgende voorbeeld geven we aan hoe de tabel in een dergelijk geval kan worden gebruikt.

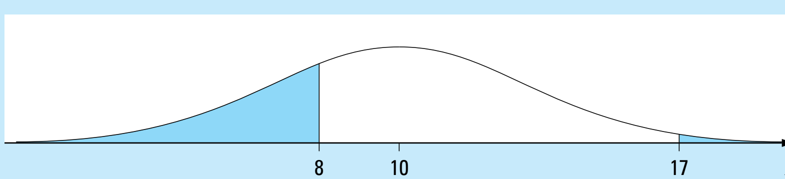
Voorbeeld 1.16

De kansvariabele $\underline{x}$ is normaal verdeeld met $\mu = 10$ en $\sigma = 4$. We bepalen $P(\underline{x} > 17)$ en $P(\underline{x} < 8)$. Om voor een willekeurige normale verdeling kansen te kunnen bepalen, moet eerst de transformatie naar de standaardnormale verdeling worden toegepast.

We definiëren: $\underline{z} = \dfrac{\underline{x} - \mu}{\sigma}$.

Dit levert hier: $\underline{z} = \dfrac{\underline{x} - 10}{\sigma}$.

Figuur 1.3 **Een willekeurige normale verdeling**



Ter bepaling van $P(\underline{x} > 17)$ berekenen we $z = \dfrac{17 - 10}{4} = 1{,}75$.

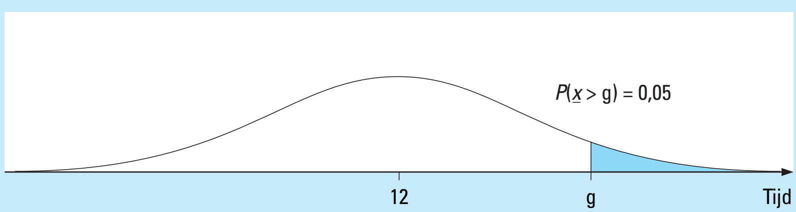
De overschrijdingskans van 1,75 blijkt volgens de tabel van de standaardnormale verdeling 0,0401 te zijn.
Er geldt dus $P(\underline{x} > 17) = 0{,}0401$.

Voor $P(\underline{x} < 8)$ gaan we op dezelfde wijze te werk. We bepalen:
$z = \dfrac{8 - 10}{4} = -0{,}5$. De tabel van de normale verdeling bevat alleen maar positieve grenswaarden. Om de vraag $P(\underline{z} < -0{,}5)$ op te lossen, maken we gebruik van de symmetrische gedaante van de normale verdeling. Er geldt:
$P(\underline{z} < -0{,}5) = P(\underline{z} > +0{,}5)$.
In de tabel vinden we bij $+0{,}5$ de waarde 0,3085.
Dus $P(\underline{x} < 8) = 0{,}3085$.

Voorbeeld 1.17
Het aantal boekingen naar een vakantiebestemming op het eiland Lesbos kan worden weergegeven door de kansvariabele $\underline{x}$ die een normale verdeling heeft met $\mu = 52$ en $\sigma = 8$. Hoe groot is de kans dat er minder dan veertig boekingen zijn? We berekenen de z-waarde. Deze is $z = -1{,}50$ (ga dit na). Hierbij hoort een kans van 0,0668. Dus er is 6,68% kans dat er minder dan veertig reizigers zijn. Eigenlijk zouden we hier nog rekening moeten houden met het feit dat het aantal boekingen geen echte continue variabele is, want het is altijd een geheel getal. Daarom zouden we de uitspraak 'minder dan 40' in de continue wereld moeten weergeven met $P(\underline{x} < 39{,}50)$. De z-waarde wordt dan $12{,}5/8 = -1{,}56$. Dus de kans wordt dan 0,0594.

Voorbeeld 1.18
De tijdsduur $\underline{x}$ die een tandarts nodig heeft voor de behandeling van een patiënt mogen we beschouwen als een kansvariabele met een normale verdeling waarvoor geldt $\mu = 12$ minuten en $\sigma = 4$ minuten.
Hoeveel tijd moet de tandarts reserveren om met kans 0,95 een patiënt binnen de gereserveerde tijdsduur te kunnen behandelen? We zoeken nu een grenswaarde g, die zodanig gekozen is dat er een rechteroverschrijdingskans van 0,05 is.

Figuur 1.4 **Het voorbeeld van de behandeltijd**



De waarde g is te berekenen door te werken vanuit de tabel van de standaardnormale verdeling.
Bij 0,05 lezen we af: $z = 1{,}64$ of $1{,}65$. Dus we kiezen voor $z = 1{,}645$.

Deze waarde van z wordt berekend als $z = \dfrac{g - 12}{4} = 1{,}645$.

Omdat z bekend is, vinden we $g = 18{,}58$.
De betrokken tandarts moet dus ruim $18\frac{1}{2}$ minuut reserveren.

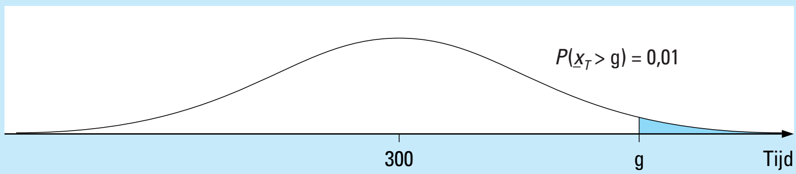
Hoeveel tijd moet de betrokken tandarts reserveren om 25 patiënten achtereenvolgens te behandelen indien geëist wordt dat er een kans 0,99 is dat de patiënten binnen de gekozen tijdsduur afbehandeld zijn?

Het gaat nu om de totaaltijd $\underline{x}_T$ die de behandeling van 25 patiënten vergt. Hiervoor geldt dat dit een kansvariabele met een normale verdeling is, waarvoor geldt:

$E(\underline{x}_T) = 25 \times 12 = 300$ minuten en
$\sigma_{\underline{x}T} = \sqrt{25} \times 4 = 20$ minuten.

De te reserveren tijd g is vervolgens te berekenen door gebruik te maken van de tabel van de standaardnormale verdeling.

Figuur 1.5 **Terugzoeken met de tabel**



Bij een overschrijdingskans 0,01 lezen we af $z = 2{,}33$.

Dus: $z = \dfrac{g - 300}{20} = 2{,}33$

We vinden dan $g = 346{,}6$ minuten.
Binnen deze tijdsduur zijn de 25 patiënten met kans 0,99 behandeld. (Opmerking. Bij deze oplossing gaan we ervan uit dat de betrokken tandarts onmiddellijk aan de volgende patiënt kan beginnen als hij klaar is met een behandeling.)

Centrale limietstelling

Centrale limietstelling

Een krachtig instrument bij het bestuderen van een serie kansvariabelen is de centrale limietstelling. Met behulp hiervan kunnen er uitspraken gedaan worden over het gemiddelde of de som van een reeks uitkomsten van willekeurige kansvariabelen. De centrale limietstelling zegt dat dit gemiddelde of deze som beschouwd mag worden als een kansvariabele met een normale verdeling, mits het aantal termen waaruit dit gemiddelde of deze som is opgebouwd niet al te klein is. Het bijzondere van deze stelling is dat de oorspronkelijke variabelen een willekeurige verdeling mogen hebben.

■ Voorbeeld 1.19

Voor de veevoederhandel uit voorbeeld 1.15 bekijken we de afzet in een periode van 25 weken. Als we de omzetten $\underline{x}_i$ in de 25 weken mogen beschouwen als onderling onafhankelijke kansvariabelen, dan geldt voor de variantie van de variabele $\underline{x}_{totaal}$:

$$\begin{aligned} Var(\underline{x}_{totaal}) &= Var(\underline{x}_1) + Var(\underline{x}_2) + \ldots + Var(\underline{x}_{25}) \\ &= 25 \times (400)^2 \end{aligned}$$

dus $\sigma_{\underline{x}totaal} = 5 \times 400 = 2\,000$

Bovendien geldt:

$$E(\underline{x}_{totaal}) = E(\underline{x}_1) + E(\underline{x}_1) + \ldots + (\underline{x}_1)$$
$$= 25 \times 2\,000 = 50\,000$$

Omdat het aantal termen hier vrij groot is, mogen we de variabele $\underline{x}_{totaal}$ beschouwen als een normaal verdeelde kansvariabele met $\mu = 50\,000$ kg en $\sigma = 2\,000$ kg.

Normale verliesfunctie

Normale verliesfunctie

Een bijzondere variant op de normale verdeling vinden we bij de normale verliesfunctie. We bespreken deze verdeling aan de hand van een voorbeeld dat op voorbeeld 1.11 lijkt.

■ **Voorbeeld 1.20**

In een cafetaria bereidt men iedere dag een aantal kilo's ijs met de bedoeling dat de beschikbare hoeveelheid zo goed mogelijk overeenstemt met de gevraagde hoeveelheid op de bewuste dag. Men weet dat de vraag beschreven kan worden door een kansvariabele $\underline{x}$ die een normale verdeling volgt met $\mu = 20$ kilo en $\sigma = 4$ kilo.
Stel nu dat op zekere dag een hoeveelheid van $b = 22$ kilo wordt bereid. Hoe groot is de kans dat op die dag niet aan alle klanten kan worden geleverd? Met andere woorden, hoe groot is de kans dat op die dag meer dan 22 kilo ijs wordt gevraagd?
Het antwoord is eenvoudig. Met de normale verdeling kan men vaststellen dat bij de grens $g = 22$ een z-waarde hoort van:

$$z = \frac{22 - 20}{4} = 0{,}50$$

De tabel van de normale verdeling levert dan als overschrijdingskans:
$P(\underline{x} > 22) = P(\underline{z} > 0{,}50) = 0{,}3085$

De baas van de cafetaria vraagt zich af hoeveel winst hij daarmee misloopt. Per kilo ijs verdient hij namelijk €8. Het werkelijke bedrag hangt af van het verschil tussen de voor levering beschikbare 22 kilo ijs en de feitelijke gevraagde hoeveelheid x (mits deze groter is dan 22 kilo). Dus soms (als de vraag lager is dan 22 kilo) is de gemiste opbrengst nul, soms is deze positief en in dat geval afhankelijk van de vraag x (die hoger moet zijn dan 22).

Wat is de verwachte of gemiddelde hoeveelheid ijs die niet geleverd kan worden? We definiëren de variabele $\underline{y} = \underline{x} - 22$. Als $\underline{x}$ een discrete verdeling zou hebben, dan konden we in beginsel alle mogelijke uitkomsten van $\underline{x}$ die groter zijn dan 22 opnoemen, daarbij de kans bepalen en vervolgens iedere x vermenigvuldigen met de kans $P(\underline{x} = x)$. Hier is echter sprake van een continue verdeling, dus dat gaat niet. Zo'n verwachtingswaarde wordt dan een integraal:

$$E(\underline{y}) = \int_{22} (x - 22) f(x) \mathrm{d}x$$

Omdat we willen kunnen werken met normale verdelingen die gekenmerkt worden door allerlei waarden van μ en σ is er een gestandaardiseerde functie ontwikkeld om deze integraal op te kunnen lossen. We doen dit met de zogenaamde Normale Verliesfunctie (of Normal Loss Function). We duiden die aan

met NL(z). Voor een normale verdeling met μ een 0 en $\sigma = 1$ geeft deze aan hoe groot de waarde is van de verliesfunctie. In tabel D (pagina 343) hebben we hiervoor enkele waarden weergegeven.
Voor een toepassing zoals hier met $\mu = 20$, $\sigma = 4$ en grenswaarde $g = 22$ vinden we $z = 0{,}50$. Bij deze z levert de tabel NL(0,50) = 0,1978.
Op basis van $\sigma = 4$ levert dit: $E(\underline{y}) = 4 \times 0{,}1978 = 0{,}7912$ kilo ijs. Aangezien de winst €8 per kilo bedraagt, is de gemiste winst $8 \times 0{,}7912 =$ €6,3296.
Voor allerlei voorgenomen productieniveaus b zou men op deze manier de verwachtingswaarde van de gevraagde, maar niet beschikbare hoeveelheid ijs kunnen berekenen. Uiteraard wordt deze verwachtingswaarde lager naarmate b hoger is.

Moet men bij de cafetaria dan voortaan nog meer ijs bereiden? Dat is moeilijk te zeggen. Aangezien de gemiddelde vraag per dag gelijk is aan 20 kilo, is er bij een productie van 22 kilo reeds sprake van een vrij behoorlijk gemiddeld overschot (zie hierover verder hoofdstuk 6).

Binomiale verdeling

Binomiale verdeling

Een verdeling die in de praktijk nogal eens voorkomt, is de binomiale verdeling. Hierbij is sprake van een discrete kansvariabele $\underline{k}$ die het aantal successen aangeeft dat optreedt in een serie lotingen van omvang n.
Bij de binomiale verdeling gaat het erom dat er per loting een vaste kans π is op het waarnemen van een succes en een kans $1 - \pi$ op het waarnemen van een mislukking. Per trekking zijn de uitkomsten ‘succes’ en ‘mislukking’ de enige twee mogelijke resultaten. Een voorbeeld van een binomiale verdeling vinden we bij het 30 maal gooien met een eerlijke dobbelsteen waarbij we vaststellen hoeveel zessen er gegooid worden. Het aantal zessen dat verschijnt, kan dan worden weergegeven door de binomiaal verdeelde kansvariabele $\underline{k}$ waarvoor geldt $n = 30$, succeskans $\pi = \frac{1}{6}$ en kans op mislukking $1 - \pi = \frac{5}{6}$.

Voor de kans op k successen geldt de volgende formule:[4]

$$P(\underline{k} = k) = \binom{n}{k}\pi^k(1-\pi)^{n-k} \text{ voor } (k = 0, \ldots, n)$$

Hiervoor is (uiteraard) aan te tonen dat $\sum_{k=0}^{n} P(\underline{k} = k) = 1$.

Voor de verwachtingswaarde $E(\underline{k})$ van het aantal successen geldt:

$$E(\underline{k}) = n\pi$$

Voor de variantie $Var(\underline{k})$ vinden we:

$$Var(\underline{k}) = n\pi(1-\pi)$$

Voor kleine waarden van n kunnen de kansen rechtstreeks worden berekend met de formule van de binomiale verdeling.
Voor grotere waarden van n ($n \geq 20$) worden de kansen bij een binomiale verdeling in het algemeen berekend door toepassing van een be-

4 De uitdrukking $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ geeft het aantal mogelijke volgorden aan.

Normale benadering

naderingsmethode. Indien hierbij $n\,\pi \geq 5$ en $n(1 - \pi) \geq 5$, dan wordt de normale benadering + gebruikt. Indien er geldt: $n\,\pi \leq 5$ of $n(1 - \pi) \leq 5$, dan wordt er gebruikgemaakt van de Poisson-benadering.
Bij toepassing van de normale benadering wordt er gekozen voor de normale verdeling met $\mu = n\,\pi$ en $\sigma = \sqrt{n\pi(1 - \pi)}$ als de verdeling die het beste aansluit bij de binomiale verdeling met succeskans π en aantal trekkingen n. Het handige van de normale benadering is dat de kansen (als oppervlakten) kunnen worden afgelezen in de tabel van de standaardnormale verdeling. Hierbij moet dan nog wel rekening worden gehouden met de continuïteitscorrectie. Deze correctie moet worden aangebracht, omdat we de kansen van een discrete verdeling (de binomiale in dit geval) berekenen door toepassing van een continue verdeling.

■ Voorbeeld 1.21

Een afnemer van een grote partij goederen die bestempeld zijn als 'tweede soort' wil door het controleren van een steekproef van een aantal exemplaren uit die partij een indruk krijgen van het percentage defecten in de gehele partij. De fabrikant die de partij goederen heeft aangeboden, stelt dat het percentage defecte exemplaren ongeveer 20% bedraagt. Als de afnemer een steekproef van tien stuks neemt, is het waargenomen aantal defecten in deze steekproef te beschrijven door een binomiale kansverdeling.
Als we als uitgangspunt nemen dat de fabrikant gelijk heeft, dan mogen we als 'succeskans' (dat is hier de kans op het aantreffen van een defect exemplaar) de waarde $\pi = 0{,}2$ kiezen.

We vinden dan (met $n = 10$, $\pi = 0{,}2$ en $(1 - \pi) = 0{,}8$):

$P(\underline{k} = 0) = \binom{10}{0}0{,}2^0 0{,}8^{10} = 0{,}1074$

$P(\underline{k} = 1) = \binom{10}{1}0{,}2^1 0{,}8^9 = 0{,}2684$

$P(\underline{k} = 2) = \binom{10}{2}0{,}2^2 0{,}8^8 = 0{,}3020$

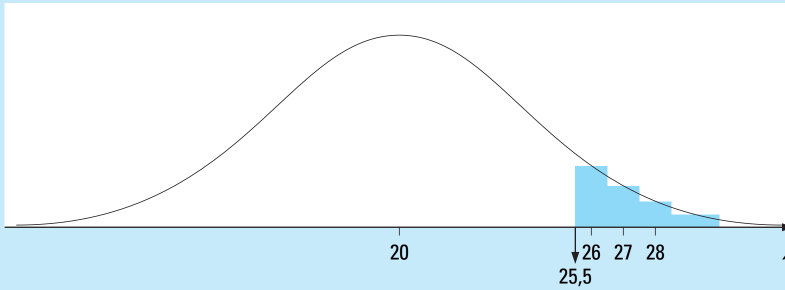
Op deze manier zijn de kansen door toepassing van de formule uit te rekenen. Bij een steekproef van 100 stuks is het aantal defecten uiteraard eveneens te beschrijven door een binomiale kansverdeling. De kansen worden dan echter berekend door toepassing van de normale benadering. We willen berekenen hoe groot de kans is dat er in de steekproef meer dan 25 defecte exemplaren worden aangetroffen.
Het aantal defecte exemplaren in een steekproef van 100 stuks wordt in beginsel beschreven door de kansvariabele $\underline{k}$, die binomiaal verdeeld is met $n = 100$, $\pi = 0{,}2$ en dus $(1 - \pi) = 0{,}8$.
Voor de verwachtingswaarde en de variantie van deze variabele vinden we dus:

$$E(\underline{k}) = n\pi = 100 \times 0{,}2 = 20$$
$$Var(\underline{k}) = n\pi(1 - \pi) = 100 \times 0{,}2 \times 0{,}8 = 16$$

Gevraagd is $P(\underline{k} > 25)$.
Deze kans berekenen we door gebruik te maken van de normale verdeling met $\mu = 20$ en $\sigma = \sqrt{16} = 4$.

We zoeken $P(\underline{k} > 25) = P(\underline{k} \geq 26)$.
Bij toepassing van de normale benadering worden kansen van de discrete binomiale verdeling in zekere zin beschouwd als kolommen. De verzameling kolommen vanaf $k = 26$ beslaat op de x-as het gebied vanaf $x = 25{,}5$, van

Figuur 1.6 **De normale benadering van de binomiale verdeling**



daar dat we moeten berekenen P($x > 25{,}5$). We vinden hiervoor (met behulp van de tabel van de standaardnormale verdeling):

$$z = \frac{25{,}5 - 20}{4} = 1{,}38 \text{ (afgerond)}$$

Aflezen bij punt 1,38 levert P($\underline{z} > 1{,}38$) = 0,0838.
Dus P($\underline{k} > 25$) = 0,0838.

Poisson-verdeling

Poisson-verdeling

Een discrete kansvariabele $\underline{k}$ volgt een Poisson-verdeling als de kans op k successen gedefinieerd is als:

$$P(\underline{k} = k) = \frac{\mu^k e^{-\mu}}{k!} \text{ voor } k = 0, 1, 2, \ldots$$

Om te komen tot een Poisson-verdeling kan men denken aan een binomiale verdeling, waarbij n (het aantal pogingen) zeer groot wordt en waarbij π (de succeskans per keer) naar nul nadert.
Indien tegelijkertijd $n\pi$ constant gehouden wordt, dan geldt dat de limietverdeling die aldus ontstaat een Poisson-verdeling is. De Poisson-verdeling komt in de praktijk nogal eens voor. Als voorbeelden kunnen we denken aan het aantal verkeersongelukken per uur in een bepaalde plaats, het aantal tikken op een geigerteller bij het meten van radioactiviteit, het aantal weeffouten in een m^2 stof, enzovoort. Om de kansen uit te rekenen bij een Poisson-verdeling kunnen we gebruikmaken van de formule (als de verwachtingswaarde μ gegeven is, kunnen we de kansen berekenen), van de Poisson-tabel (zie pagina 341) of kunnen we de normale benadering toepassen (voor $\mu > 8$). Een kenmerkende eigenschap van de Poisson-verdeling is dat de verwachtingswaarde en de variantie aan elkaar gelijk zijn, $E(\underline{k}) = Var(\underline{k}) = \mu$.

■ **Voorbeeld 1.22**

Een firma in bouwmaterialen verhuurt afvalcontainers aan particulieren. Deze containers worden in beginsel telkens voor één dag verhuurd. De vraag per dag mag worden beschouwd als een kansvariabele met een Poisson-verdeling waarvoor geldt $\mu = 2$.
Het bedrijf heeft drie containers beschikbaar voor de verhuur. Op basis van deze gegevens is het aantal verhuurdagen per jaar te berekenen. We bekijken eerst de situatie per dag. Uit de tabel van de Poisson-verdeling blijkt dat:

$P(\underline{k} = 0 | \mu = 2) = 0{,}1353$
$P(\underline{k} = 1 | \mu = 2) = 0{,}2707$
$P(\underline{k} = 2 | \mu = 2) = 0{,}2707$ en
$P(\underline{k} = 3 | \mu = 2) = 0{,}3233$.

Per dag vinden we aldus voor de verwachtingswaarde van het aantal verhuurde containers:

$$\begin{aligned} E(\underline{k}) &= 0 \times P(\underline{k} = 0) + 1 \times P(\underline{k} = 1) + 2 \times P(\underline{k} = 2) + 3 \times P(\underline{k} = 3) \\ &= 0 \times 0{,}1353 + 1 \times 0{,}2707 + 2 \times 0{,}2707 + 3 \times 0{,}3233 \\ &= 1{,}782 \text{ per dag} \end{aligned}$$

Merk op dat bij een vraag naar meer dan drie containers per dag er slechts drie stuks verhuurd kunnen worden. Per jaar zijn er 250 dagen waarop verhuur plaatsvindt. Het verwachte aantal verhuurdagen per jaar is daarom:

$$E(\underline{k}_{\text{totaal}}) = 250 \times 1{,}782 = 445{,}5$$

Immers, de verwachtingswaarde van de som is gelijk aan de som van de verwachtingswaarden.
Voor een dag verhuur rekent het bedrijf €100. We gaan de jaarwinst berekenen op basis van de geformuleerde verwachtingswaarde.
Hiervoor is verder gegeven dat er aan afschrijving en rente per container jaarlijks €6.000 moet worden opgebracht. De variabele kosten zijn €20 per dag van verhuur. Per dag van verhuur levert een verhuurde container een bijdrage van €100 – €20 = €80. Het verwachte aantal verhuurdagen bedraagt 445,5. Per jaar levert dit een winstbijdrage W van:

$$\begin{aligned} W &= 445{,}5 \times 80 - 3 \times 6\,000 \\ &= 17.640 \end{aligned}$$

Het bedrijf overweegt een vierde container aan te schaffen. Op basis van de kansverdeling van de vraag maken we een analyse of een dergelijke uitbreiding zinvol is. Voor de vraag per dag vinden we:

$P(\underline{k} = 0 | \mu = 2) = 0{,}1353$
$P(\underline{k} = 1 | \mu = 2) = 0{,}2707$
$P(\underline{k} = 2 | \mu = 2) = 0{,}2707$
$P(\underline{k} = 3 | \mu = 2) = 0{,}1804$
$P(\underline{k} = 4 | \mu = 2) = 0{,}1429$

De verwachtingswaarde van het aantal verhuurde containers per dag is:

$$E(\underline{k}) = 0 \times 0{,}1353 + 1 \times 0{,}2707 + 2 \times 0{,}2707 + 3 \times 0{,}1804 + 4 \times 0{,}1429 = 1{,}925 \text{ container per dag}$$

Op jaarbasis vinden we dus:

$$E(\underline{k}_{\text{totaal}}) = 250 \times 1{,}925 = 481{,}25$$

Voor de verwachtingswaarde van de jaarwinst vinden we:

$$W = 481{,}25 \times 80 - 4 \times 6\,000 = €14.500$$

Dit bedrag ligt lager dan het berekende bedrag bij drie containers. De extra opbrengst die af en toe behaald wordt met de verhuur van de vierde container weegt dus niet op tegen het extra bedrag aan rente en afschrijving dat jaarlijks moet worden opgebracht.

Poisson-verdeling als benadering van de binomiale verdeling
Bij de bespreking van de binomiale verdeling zagen we dat voor $n > 20$ de kansen bij een binomiale verdeling in het algemeen niet rechtstreeks, maar via een benaderingsmethode moeten worden bepaald. Indien $n\,\pi > 5$ en $n\,(1 - \pi) > 5$ dan kan de normale verdeling worden toegepast. Is daarentegen $n\,\pi < 5$, terwijl $n > 20$ is, dan kan gebruik worden gemaakt van de Poisson-benadering.[5]

■ **Voorbeeld 1.23**
Het aantal ondeugdelijke exemplaren in een grote partij goederen bedraagt 2%. Van 75 willekeurige producten wordt de kwaliteit gecontroleerd. Hoe groot is de kans dat hierbij twee of meer ondeugdelijke exemplaren worden aangetroffen?
In beginsel is het aantal ondeugdelijke exemplaren te beschouwen als een kansvariabele $\underline{k}$ die binomiaal verdeeld is met $n = 75$ en $\pi = 0{,}02$.
Gevraagd is: $P(\underline{k} \geq 2) = 1 - P(\underline{k} = 0) - P(\underline{k} = 1)$.
Om de kansen te bepalen, passen we de Poisson-benadering toe. We berekenen eerst de verwachtingswaarde:

$$E(\underline{k}) = n\pi = 75 \times 0{,}02 = 1{,}5$$

Deze waarde kiezen we als parameter bij de Poisson-verdeling. In de tabel zoeken we dan op:

$P(\underline{k} = 0 \mid \mu = 1{,}5) = 0{,}2231$
$P(\underline{k} = 1 \mid \mu = 1{,}5) = 0{,}3347$
Dus $P(\underline{k} \geq 2 \mid \mu = 1{,}5) = 1 - 0{,}2231 - 0{,}3347 = 0{,}4422$.

1.3.4 Een vervangingsprobleem

Als illustratie van het gebruik van de kansrekening als hulpmiddel bij beslissingsproblemen geven we een voorbeeld van een zogenoemd vervangingsprobleem. Het betreft hier goederen die na verloop van tijd uitvallen en vervangen moeten worden. Het zal blijken dat er verschillende strategieën gevolgd kunnen worden die elk gekenmerkt worden door een verschillend kostenniveau.

■ **Voorbeeld 1.24**
In een grote stad bevinden zich totaal 10 000 straatlantaarns. In iedere straatlantaarn bevindt zich een tl-buis, die het na verloop van tijd begeeft. Als een tl-buis in een straatlantaarn is stukgegaan, dient deze uiteraard na verloop van niet al te lange tijd vervangen te worden door een nieuwe. De gemiddelde levensduur van de lampen volgt uit een berekening die gebaseerd is op de gegevens uit tabel 1.1.

5 Als $n(1 - \pi) < 5$ dan moet de rol van π en $(1 - \pi)$ worden verwisseld. In plaats van successen worden dan pechgevallen geteld.

Tabel 1.1 **Uitgevallen lampen**

Aantal maanden *k*	**Percentage uitvallers na *k* maanden**
1	5%
2	30%
3	70%
4	90%
5	100%

Voor lampen die na één maand zijn uitgevallen, stellen we dat die een gemiddelde levensduur hebben van een halve maand. Analoog is voor lampen die langer dan twee maanden zijn meegegaan maar aan het eind van de derde maand zijn uitgevallen, de gemiddelde levensduur $2\frac{1}{2}$ maand. Aldus is de gemiddelde levensduur te berekenen aan de hand van tabel 1.2.

Tabel 1.2 **Levensduur van lampen**

Maand *k*	**Kans op uitvallen in maand *k***	**Gemiddelde branddduur $\underline{t}$**
1	0,05	0,5
2	0,25	1,5
3	0,40	2,5
4	0,20	3,5
5	0,10	4,5

$$\begin{aligned} E(\underline{t}) &= \sum t P(\underline{t} = t) \\ &= 0{,}5 \times 0{,}05 + 1{,}5 \times 0{,}25 + 2{,}5 \times 0{,}40 + 3{,}5 \times 0{,}20 + 4{,}5 \times 0{,}10 \\ &= 0{,}025 + 0{,}375 + 1{,}00 + 0{,}70 + 0{,}45 = 2{,}55 \text{ maanden} \end{aligned}$$

Bij de vervanging van de lampen kan men verschillende strategieën overwegen. Als eerste mogelijkheid zou men kunnen denken aan het maandelijks vervangen van alleen de defecte lampen. Een bedrijf dat de lampen vervangt, biedt aan om deze werkzaamheden te verrichten voor €15 per lamp. Een tweede strategie zou kunnen zijn een volledige vervanging van alle lampen iedere maand. Het voordeel hiervan is dat de kosten van vervanging per lamp gerekend aanzienlijk lager zijn, omdat het bedrijf dat de vervangingen verricht dan niet hoeft te zoeken naar de defecte lampen. Het bedrijf is bereid de vervanging van alle lampen te doen voor €7,50 per stuk. Voor 10 000 lampen levert dit een rekening op van €75.000 per maand.
Welke keuze dient de gemeente hier te maken? Om een verantwoorde beslissing te kunnen nemen, moet er worden nagegaan hoeveel lampen er gemiddeld per maand vervangen dienen te worden als de methode gekozen wordt waarbij alleen de defecte lampen vervangen worden.
Als we op zekere dag beginnen met 10 000 nieuwe lampen, dan kunnen we maand voor maand nagaan hoeveel lampen (V) er vervangen moeten worden.
Na één maand: aantal uitvallers

$$V_1 = 0{,}05 \times 10\,000 = 500.$$

Na twee maanden zijn er dus 9 500 oude lampen en 500 nieuwe lampen. We vinden: aantal uitvallers

$$V_2 = 0{,}05 \times 500 + 0{,}25 \times 10\,000 = 2525.$$

In de derde maand zijn er drie leeftijdscategorieën, namelijk:
Er zijn dan nog 10 000 – 500 – 2 500 = 7 000 lampen van twee perioden geleden.
Er zijn nog 500 – 25 = 475 lampen van één periode oud en 2 525 nieuwe lampen.
Het aantal uitvallers na drie maanden (V_3) bedraagt:

$$V_3 = 0{,}40 \times 10\,000 + 0{,}25 \times 500 + 0{,}05 \times 2\,525 = \\ = 4\,000 + 125 + 126 = 4\,251.$$

Het aantal te vervangen lampen gaat zich – bij gelijkblijvende kansverdeling van de levensduur – geleidelijk stabiliseren tot een waarde N^*.
Indien er iedere periode een aanvulling (dus een aantal uitvallers) ten bedrage van N^* is, dan kunnen we N^* als volgt berekenen:

$$10\,000\ N^* \text{ (de aanvulling)} + 0{,}95\ N^* \text{ (één periode oud)} + \\ + 0{,}70\ N^* \text{ (twee perioden oud)} + 0{,}30 N^* \text{ (drie perioden oud)} \\ + 0{,}10\ N^* \text{ (vier perioden oud)}$$

Dus 1 000 = 3,05 N^* dus N^* = 3 279.
Op lange termijn moeten er dus iedere maand 3 279 lampen vervangen worden. Aangezien iedere vervanging €15 kost, bedragen de lasten bij toepassing van deze methode 3 279 × 15 = €49.185 per maand. Uit deze berekening blijkt dat het een stuk voordeliger is om alleen de defecte lampen te vervangen, vergeleken met de totale maandelijkse vervanging, die €75.000 bleek te kosten.

Een derde mogelijkheid bij het vervangen van de lampen zou de volgende kunnen zijn:
- Eenmaal per twee maanden worden alle lampen vervangen.
- In de tussenliggende maand worden alleen de uitvallers vervangen.

We berekenen de kosten voor de strategie over een periode van twee maanden. De vervanging van de 10 000 lampen kost €75.000. Na een maand zijn hiervan 500 lampen uitgevallen. Het vervangen hiervan kost €15 × 500 = €7.500. De totale kosten per twee maanden bedragen dus €82.500. Op maandbasis levert dit een kostenpost van €41.250. Deze derde strategie is daarom te verkiezen boven de twee eerder genoemde.
Is hiermee de optimale strategie bepaald? Dat is op voorhand niet te zeggen, omdat we ons hebben beperkt tot de keuze van de drie strategieën. (In het opgavenboek wordt gevraagd nog een vierde strategie te onderzoeken.)

1.4 Bedrijfszekerheid

Bij veel machines en apparaten die wij gebruiken is het van groot belang dat we erop kunnen vertrouwen dat ze op een betrouwbare wijze functioneren. Een defect of weigering kan ertoe leiden dat een productieproces geruime tijd moet worden stilgelegd omdat door het uitvallen van een klein onderdeel een groot systeem onbruikbaar is geworden. Bovendien kunnen er aanzienlijke kosten voor vervanging of reparatie zijn. Verder is het denkbaar dat anderszins grote schade wordt veroorzaakt. Als voorbeelden kan men denken aan het weigeren van beademingsapparatuur in ziekenhuizen, technische problemen met motoren van vliegtuigen of het uitvallen van vitale onderdelen van kerncentrales.

Bedrijfszekerheid
Reliability

Een interessante bedrijfskundige toepassing van kansrekening betreft de bedrijfszekerheid of *reliability* van machines en onderdelen. Hieronder verstaan we de kans dat een product blijft werken gedurende een vastgestelde periode. In sommige gevallen heeft men een indruk kunnen krijgen van de kansen op niet-functioneren op basis van ervaring met bijvoorbeeld soortgelijke systemen.

Als we met $R(t)$ de kans aanduiden dat een product na de gespecificeerde periode t nog functioneert en met $D(t)$ de kans dat het product die periode defect is geraakt, dan geldt de volgende formule:

$$R(t) + D(t) = 1$$

We laten hier enkele toepassingen zien van het berekenen van kansen om de bedrijfszekerheid van een product te bepalen. Bij het beschouwen van dit soort problemen dient men soms een onderscheid te maken tussen:

a de kans dat een onderdeel of systeem uitvalt zodat het systeem gedurende een geheel tijdvak niet (meer) functioneert;
b de kans dat een onderdeel minstens gedurende een gegeven *tijdsduur* wél functioneert.

Met name is daarbij van belang of in een gegeven situatie reparatie mogelijk is (bijvoorbeeld omdat een reserveonderdeel kan worden geïnstalleerd) dan wel dat de weigering een definitief karakter heeft (zoals bij het neerstorten van een vliegtuig, het uitvallen van een pacemaker of het niet meer functioneren van de zonnepanelen van een satelliet).

1.4.1 Bedrijfszekerheid van losse onderdelen

Dikwijls is een product opgebouwd uit diverse onderdelen die allemaal een kans hebben om defect te raken. Als een onderdeel het begeeft, is dan ook het gehele product defect? Dat hangt af van de constructie van het product.

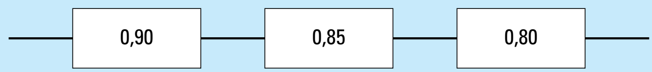
Als de onderdelen in serie geschakeld zijn, dan zal inderdaad het geheel al niet meer functioneren als slechts een onderdeel weigert. Een voorbeeld hiervan is de traditionele kerstverlichting, waarvoor geldt dat alle lampjes uitvallen indien één lampje stuk is (of loszit). Als $R(t)$, kortweg R, de kans is dat een lampje nog werkt na een vastgestelde periode t, dan is R^n de kans dat de gehele keten nog werkt na die periode. Voor een keten van vijftig identieke kerstlampjes betekent dit dat de kans dat deze uitvalt $1 - R^{50}$ bedraagt. Dus bij een kans 0,99 dat een lampje het blijft doen, is er een kans $1 - 0{,}99^{50} = 1 - 0{,}6050 = 0{,}3950$ dat de keten uitvalt.

Een geheel andere opzet krijgen we indien onderdelen parallel geschakeld zijn. In dat geval wordt van een onderdeel dat uitvalt de taak overgenomen door een parallel geschakeld onderdeel. We zouden dit de back-up kunnen noemen.

Serieschakeling
Parallelschakeling

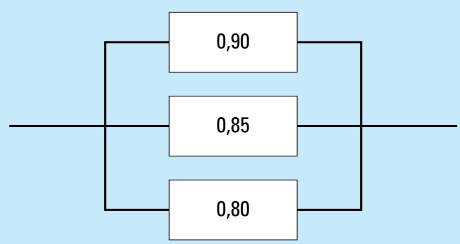
Dergelijke problemen kunnen op een handige manier grafisch worden weergegeven. In figuur 1.7 zien we een eenvoudige serieschakeling, in figuur 1.8 een parallelschakeling en in figuur 1.9 een combinatie van beide.

Figuur 1.7 **Serieschakeling**



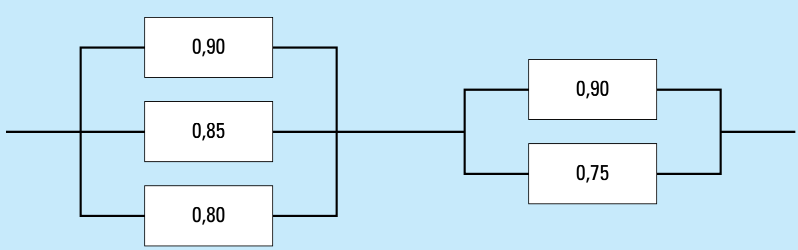
Hier geldt: $R(\text{systeem}) = 0{,}90 \times 0{,}85 \times 0{,}80 = 0{,}612$

Figuur 1.8 **Parallelschakeling**



Dus $R(\text{systeem}) = 0{,}90 + 0{,}10 \times 0{,}85 + 0{,}10 \times 0{,}15 \times 0{,}80 = 0{,}9970$

Figuur 1.9 **Combinatie van serie- en parallelschakeling**



Hier geldt: $R(\text{systeem}) = 0{,}9970 \times (0{,}90 + 0{,}10 \times 0{,}75) = 0{,}9721$

Bij de berekeningen zoals weergegeven in de figuren kunnen we het volgende aantekenen.

- De gebruikte kansregels zijn gebaseerd op onderlinge onafhankelijkheid. Dus als een systeem twee onderdelen X en Y bevat, dan wordt het uitvallen van onderdeel Y niet beïnvloed door het wel of niet uitvallen van onderdeel X.

- In de praktijk worden kansen op wel of niet functioneren van een onderdeel uiteraard vooraf bepaald. Meestal zal men hierbij gebruikmaken van praktijkervaring uit het verleden. Hierbij moeten we erop letten dat de kansen gebaseerd zijn op 'normale' condities. Voor auto-onderdelen zou dit bijvoorbeeld kunnen inhouden dat de levensduurgarantie van onderdelen uitgaat van normaal rijgedrag en normale condities van de wegen.
- Als onderdelen in serie zijn geschakeld, dan geldt de eenvoudige regel: naarmate het aantal onderdelen groter wordt, wordt de bedrijfszekerheid kleiner.

■ Voorbeeld 1.25

Op grond van ervaring is bekend dat de kans 0,05 (5%) is dat van een overheadprojector tijdens een presentatie een lamp uitvalt. Stel nu dat in het apparaat een reservelamp is ingebouwd en dat we door middel van een schakelaar deze reservelamp in werking kunnen zetten zodra de oorspronkelijke lamp uitvalt. Hoe groot is dan de kans dat er toch nog problemen komen met de projector?

Een dergelijke kans is te berekenen indien we weten:

a hoe groot de kans is dat de reservelamp werkt en

b hoe groot de kans is dat de schakelaar deugdelijk functioneert.

Stel dat die kans voor de reservelamp 0,95 bedraagt en voor de schakelaar 0,99, dan kunnen we berekenen:

$$P(\text{projector functioneert}) = 0{,}95 + 0{,}05 \times 0{,}99 \times 0{,}95 = 0{,}997025$$

Dus er is een kans $1 - 0{,}997025 = 0{,}002975$ dat de projector ondanks de aanwezigheid van een reservelamp niet werkt.

In voorbeeld 1.25 hebben we te maken met een vorm van parallelle schakeling (voor een deel in combinatie met een serieschakeling). Dit houdt in dat het niet functioneren van een onderdeel in beginsel wordt opgevangen door een ander onderdeel dat dezelfde functie heeft of dat kan worden ingeschakeld om diezelfde functie te vervullen.

■ Voorbeeld 1.26

Een viermotorig vliegtuig kan nog in de lucht blijven als twee van de vier motoren tijdens de vlucht uitvallen. Stel dat iedere motor een kans heeft van 0,005 om tijdens een vlucht te weigeren. Hoe groot is dan de kans dat het vliegtuig neerstort? We kunnen dit uitrekenen met de binomiale verdeling. Het gaat dus goed als alle vier de motoren het doen ($0{,}995^4$) of als drie van de vier het doen ($4 \times 0{,}995^3 \times 0{,}005$) en ook als twee van de vier het doen ($6 \times 0{,}995^2 \times 0{,}005^2$). Samen levert dit: $0{,}9801495 + 0{,}0197015 + 0{,}0001485 = 0{,}9999995$. De kans op neerstorten is dus $1 - 0{,}9999995 = 0{,}0000005$. Minder dan 1 op de miljoen dus.

Twee motoren bevinden zich aan de linkervleugel en twee aan de rechtervleugel. Indien drie of vier motoren uitvallen stort het vliegtuig neer. Indien echter twee motoren uitvallen aan dezelfde kant van het vliegtuig dan stort het vliegtuig ook neer. Hoe groot is nu de kans dat het vliegtuig neerstort tijdens een vlucht?

Van de zes varianten waarop twee van de vier motoren kunnen uitvallen vervallen er nu twee. Dus de laatste kans in de sommatie wordt nu 0,0000990. Opgeteld wordt de kans dan 0,99995. De kans op neerstorten is nu 0,00005. Dat is toch honderd keer zo groot als de eerder berekende kans.

1.4.2 Kansverdeling van een uitvaltijdstip

De sleutelvariabele bij het bestuderen van bedrijfszekerheid is het tijdstip *t* waarop een onderdeel of een ander systeem uitvalt. Het gaat dan om de kansverdeling van de levensduur. Voor de weergave van de variabele $\underline{t}$ die het tijdstip van uitvallen aangeeft, kan men verschillende kansverdelingen in gedachten nemen. We noemen hier:

1. normale verdeling
2. negatief-exponentiële verdeling
3. gammaverdeling
4. Weibull-verdeling, die regelmatig bij levensduurproblemen wordt gebruikt.

Ad 1 Normale verdeling
Over de *normale verdeling* is eerder in dit hoofdstuk al het nodige gezegd. Men zou zich kunnen indenken dat bij sommige producten, bijvoorbeeld gloeilampen, zo'n verdeling een redelijke beschrijving kan geven voor de levensduur. Ingewikkelder wordt het indien een systeem bestaat uit een groter aantal onderdelen, elk met een normaal verdeelde levensduur. Als het systeem al uitvalt zodra een van die onderdelen niet meer werkt, dan komen we bij de eerder genoemde aanpak voor serieschakelingen. Veel moeilijker wordt het indien een groter aantal onderdelen defect moet zijn alvorens het systeem uitvalt. Met name is het dan niet altijd eenvoudig om de levensduur van het gehele systeem door middel van een kansvariabele weer te geven. We gaan hier niet op in.

Ad 2 Negatief-exponentiële verdeling
De negatief-exponentiële verdeling wordt in de praktijk veel gebruikt. Deze is gerelateerd aan de Poisson-verdeling. Indien we in een vast gekozen tijdsinterval het aantal gebeurtenissen kunnen beschouwen als een Poisson-verdeelde variabele, dan is de duur van de wachttijd tussen twee gebeurtenissen te beschrijven als een continue variabele $\underline{t}$, die een negatief-exponentiële verdeling volgt. Deze verdeling kan bijvoorbeeld van pas komen bij het bestuderen van storingen in het productieproces. De verdeling kan als volgt weergegeven worden:

Negatief-exponentiële verdeling

$$f(t) = \lambda e^{-\lambda t} \qquad \text{voor } t \geq 0 \text{ en}$$
$$f(t) = 0 \qquad \text{voor } t < 0$$

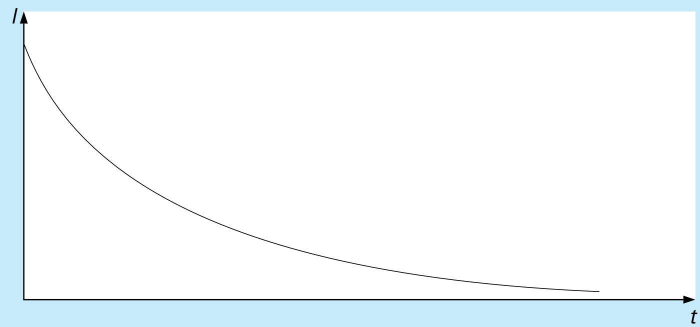
De grootheid λ komt overeen met de parameter μ van de Poisson-verdeling. Voor de verwachtingswaarde van de negatief-exponentiële verdeling geldt de volgende formule:

$$E(\underline{t}) = \frac{1}{\lambda}$$

Deze uitdrukking zegt zoveel als: indien het gemiddelde aantal storingen per uur Poisson-verdeeld is met $\mu = 4$ per uur, dan is na het optreden van een storing de gemiddelde wachttijd $E(\underline{t})$ tot de *volgende* sto-

ring gelijk aan $1/\lambda = 0{,}25$ uur. Ofwel, gemiddeld is er ieder kwartier een storing.
Figuur 1.10 geeft de dichtheidsfunctie van een negatief-exponentiële verdeling aan. Kansen moeten in beginsel worden uitgerekend als integralen (= oppervlakken).

Figuur 1.10 **De negatief-exponentiële verdeling**



■ **Voorbeeld 1.27**

Van een computersysteem is gegeven dat dit af en toe uitvalt wegens een storing. Hierbij is het aantal storingen per jaar (van 2 500 werkuren) Poisson-verdeeld met $\mu = 10$. Uitgedrukt als aantal storingen per uur komt dit overeen met $\mu = 0{,}004$. Dit impliceert dat de lengte van het tijdsinterval (uitgedrukt in uren) tussen twee storingen kan worden beschreven door een negatief-exponentiële verdeling met $\lambda = 0{,}004$. We kunnen nu bijvoorbeeld berekenen dat het systeem langer dan 500 uur achter elkaar blijft werken:

$$P(\underline{t} > 500) = \int_{500}^{\infty} 0{,}004e^{-0{,}004t}dt = 0 - (-e^{-0{,}004 \times 500}) = 0{,}1353$$

Evenzo kunnen we de kans berekenen dat het systeem korter dan 300 uur werkt:

$$P(\underline{t} < 300) = \int_{0}^{300} 0{,}004e^{-0{,}004t}dt = -e^{-0{,}004 \times 300} - (-1) = -0{,}3012 + 1 = 0{,}6988$$

Voor de verwachte tijdsduur tussen twee storingen geldt hier dus
$E(\underline{t}) = 1/\lambda = 250$ uur. En dat resultaat hoort te kloppen met onze intuïtie.

MTBF

In de literatuur over bedrijfszekerheid spreekt men wel over $1/\lambda$ als *MTBF (mean time between failures)*. Daaraan gekoppeld is de grootheid 'beschikbaarheid' *(availability)*. Daarbij moeten we een systeem in gedachten nemen dat op gezette tijden stilvalt en dan gedurende enige tijd gerepareerd moet worden. Als we met MTR (mean time to repair) de verwachte reparatietijd aangeven dan geldt het volgende:

$$\text{beschikbaarheid} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTR}}$$

■ **Voorbeeld 1.28**
Het aantal storingen in een fabriek is Poisson-verdeeld met $\mu = 10$ per jaar (van 2.500 bedrijfsuren). De MTBF is dus 250 uur. Als gegeven is dat de gemiddelde reparatietijd vijf uur is, dan volgt:

$$\text{beschikbaarheid} = \frac{250}{250 + 5} = 0{,}98$$

Bij de planning van grote orders moet met dit gegeven rekening worden gehouden.

Ad 3 Gammaverdeling
De gammaverdeling is verwant aan de negatief-exponentiële verdeling. Terwijl bij de negatief-exponentiële verdeling de wachttijd tot de *eerstvolgende* gebeurtenis wordt bestudeerd, zou men natuurlijk ook geïnteresseerd kunnen zijn in de wachttijd tot de n-de gebeurtenis. Als we voor een gefixeerd tijdsinterval het aantal op te treden gebeurtenissen $\underline{k}$ als een Poisson-verdeelde variabele mogen beschouwen, dan kan men de bedrijfszekerheid van het systeem met behulp van de gammaverdeling bepalen. De variabele $\underline{t}$ geeft het tijdstip aan dat het systeem uitvalt, dat is dus op het moment van het registreren van de gebeurtenis (storing) nummer n:

Gammaverdeling

$$\begin{aligned} P(\underline{t} > t) &= P(\underline{k} < n \text{ in tijdvak van lengte } t) \\ &= \text{som van de kansen op 0, 1, 2 tot } n-1 \text{ gebeurtenissen} \\ &= \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \end{aligned}$$

Uit deze uitdrukking kunnen we rechtstreeks de kansdichtheid en de verdelingsfunctie van de gammaverdeling bepalen (zie Statistiek om mee te werken, hoofdstuk 7).

■ **Voorbeeld 1.29**
Een zaktelefoontje dat regelmatig opgeladen dient te worden valt uit zodra er voor de tiende maal een gesprek mee wordt gevoerd. Het aantal gesprekken is Poisson-verdeeld met $\mu = 2$ per dag. Hoe groot is de kans dat het telefoontje pas na precies zeven dagen uitvalt? Het aantal gesprekken in zeven dagen kan men als een Poisson-verdeling met $\mu = 14$ zien. Voor de bedrijfszekerheid vinden we hier dus: $P(\underline{t} > t) = P(\underline{k} < 10)$ in tijdvak van lengte t.
Hieruit volgt:

$$\sum_{k=0}^{10-1} \frac{(14)^k e^{-14}}{k!} = 0{,}1093$$

In voorbeeld 1.30 zien we een geval dat net iets anders is.

■ **Voorbeeld 1.30**
Een satelliet is voorzien van vijf zonnepanelen. Ieder zonnepaneel heeft een levensduur die kan worden beschreven door de negatief-exponentiële verdeling met $\lambda = 1$. Dus de gemiddelde levensduur van een paneel is $1/\lambda = 1$ jaar. Zodra drie zonnepanelen uitgevallen zijn, houdt de satelliet op te functioneren. Hoe groot is de kans dat de satelliet na twee jaar nog functioneert? We gaan ervan uit dat de levensduren van de panelen onderling onafhankelijk zijn.

Voor een zonnepaneel is de kans om te overleven na twee jaar eenvoudig te berekenen. Dit kan rechtstreeks met de integraal:

$$P(\underline{t} > 2) = \int_2^{\infty} e^{-t} dt = 0 - (-e^{-2}) = 0{,}1353$$

of met de formule:

$$P(\underline{t} > 2) = e^{-\lambda t} = 0{,}1353 \text{ (met } \lambda = 1 \text{ per jaar en } t = 2\text{)}.$$

Dus de kans op uitvallen is 0,8647.

De satelliet functioneert nog na twee jaar indien het totaal aantal uitgevallen zonnepanelen op z'n hoogst twee bedraagt. We berekenen deze kansen met behulp van de binomiale verdeling met 'succeskans' $\pi = 0{,}8647$.
Dus van het aantal uitvallers $\underline{k}$ vinden we:

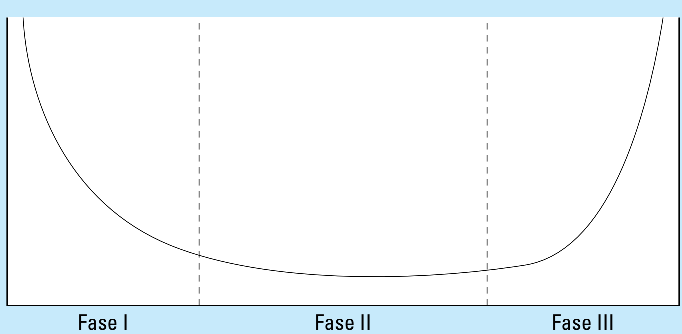
$$P(\underline{k} \leq 2) = + P(\underline{k} = 1) + P(\underline{k} = 2)$$
$$= 0{,}8647^0 \times 0{,}1353^5 + 5 \times 0{,}8647^1 \times 0{,}1353^4 + 10 \times 0{,}8647^2 \times 0{,}1353^3 = 0{,}0199$$

Merk op dat we hier geen gammaverdeling hebben!

Ad 4 Weibull-verdeling
Zijn nu de verdelingen die we hiervoor genoemd hebben goede kandidaten om een beschrijving te geven van de levensduur van een systeem? Ja en nee. In sommige gevallen waarvan hier voorbeelden gegeven zijn, lijkt het heel goed te doen om met bijvoorbeeld de normale verdeling of de negatief-exponentiële verdeling de levensduur aan te geven van een onderdeel of van een systeem. Anderzijds laat de werkelijkheid zich niet zomaar dwingen om zich overeenkomstig een voorgeschreven model te gedragen.
Bij verscheidene systemen lijkt het erop dat de levensduur kan worden weergegeven door een grafiek die enigszins de vorm van een badkuip heeft. Het idee hierbij is dat we het aantal uitvallers weergeven door de zogeheten *failure rate*. Dat is eigenlijk het aantal uitvallers per tijdseenheid, gedeeld door het aantal aanwezige elementen aan het begin van dat tijdvak. Vaak blijkt dat de *failure rate* vrij groot is aan de beginperiode van een systeem, omdat dan exemplaren met een constructiefout uitval-

Failure rate

Figuur 1.11 **Een *failure rate*-functie**



len, daarna volgt een laag deel van de grafiek met een geringe *failure rate* en aan het eind loopt die failure rate sterk op omdat de oude exemplaren versleten raken. Zelfs op mensen lijkt dit denkbeeld van toepassing, aangezien de sterftecijfers onder kleine kinderen en bejaarden het hoogst zijn.

We kunnen dit alles wat formeler benaderen door voor de grootheid $R(t)$ een kansvariabele $\underline{t}$ in te voeren die beschrijft op welk tijdstip t een systeem uitvalt nadat het in de periode tussen $t = 0$ en t heeft gefunctioneerd. Voor $\underline{t}$ geldt een kansdichtheid $f(t)$. Voor de bedrijfszekerheid R(t) kunnen we de volgende formule gebruiken:

$$R(t) = P(\underline{t} > t) = \int_t^\infty f(x)\mathrm{d}x = 1 - F(t)$$

Het is dus eenvoudig vast te stellen dat de afgeleide van R gelijk is aan minus de kansdichtheid. Dus $R'(t) = -f(t)$.

We zouden ons kunnen afvragen in welke mate de kans op uitvallen van het systeem gedurende een tijdsinterval $(t, t + \Delta t)$ samenhang vertoont met de kans $R(t)$, dat is dus de kans dat het systeem tot aan tijdstip t goed heeft gefunctioneerd. Dat is een soort voorwaardelijke kans, waarbij het aantal uitvallers gerelateerd wordt aan het aantal aanwezigen. Dit kan worden weergegeven door de volgende uitdrukking:

$$\frac{F(t + \Delta t) - F(t)}{R(t)}$$

Als we de uitdrukking boven de deelstreep delen door Δt en de limiet naar 0 nemen, dan verschijnt de afgeleide van de verdelingsfunctie $F(t)$, dat is dus de kansdichtheid $f(t)$.

Failure rate-functie

Deze moet nog gedeeld worden door $R(t) = 1 - F(t)$. Het resultaat wordt wel de *failure rate-functie* $Z(t)$ genoemd. Deze grootheid geeft dus iets weer van de kans op het uitvallen van een systeem gegeven de kans dat het systeem tot aan tijdstip t heeft gefunctioneerd:

$$Z(t) = \frac{f(t)}{1 - F(t)}$$

Uit deze uitdrukking blijkt dat indien de kansdichtheid $f(t)$ gegeven is (en dus $F(t)$), daarmee $Z(t)$ te bepalen is. Aangetoond kan worden dat het omgekeerde ook waar is: indien een *failure rate*-functie $Z(t)$ gegeven is, dan ligt daarmee $f(t)$ vast.

Een kansverdeling die soms voor de dichtheid $f(t)$ wordt gebruikt is de zogeheten Weibull-verdeling. Deze wordt weergegeven door de volgende uitdrukking:

$$f(t) = \alpha\beta t^{\beta - 1}e^{-(\alpha t)\beta} \text{ voor } t > 0$$

In deze uitdrukking zien we twee parameters, α en β, die nader ingevuld moeten worden. De parameter β bepaalt vooral de vorm van de verdeling, terwijl α meer de *positie* van de verdeling aangeeft. Voor de waarde $\beta = 1$ ontstaat de negatief-exponentiële dichtheidsfunctie.
Bij $\beta > 1$ kunnen verdelingen ontstaan die in meer of mindere mate overeenkomen met een (ietwat scheve) normale verdeling.

Wiskundeliefhebbers kunnen voor de Weibull-verdeling vaststellen dat een failure rate-functie resulteert die eruitziet als:

$$Z(t) = \alpha\beta t^{(\beta - 1)}$$

Voor $\beta = 1$ leidt dit tot een constante *failure rate*-functie. Dat is dus het geval bij het Poisson-proces, dus bij negatief-exponentieel verdeelde tussentijden. Bij $\beta > 1$ stijgt $Z(t)$ en bij $\beta < 1$ daalt deze. Dat wordt dus niet een vorm zoals de badkuipgrafiek van figuur 1.11.