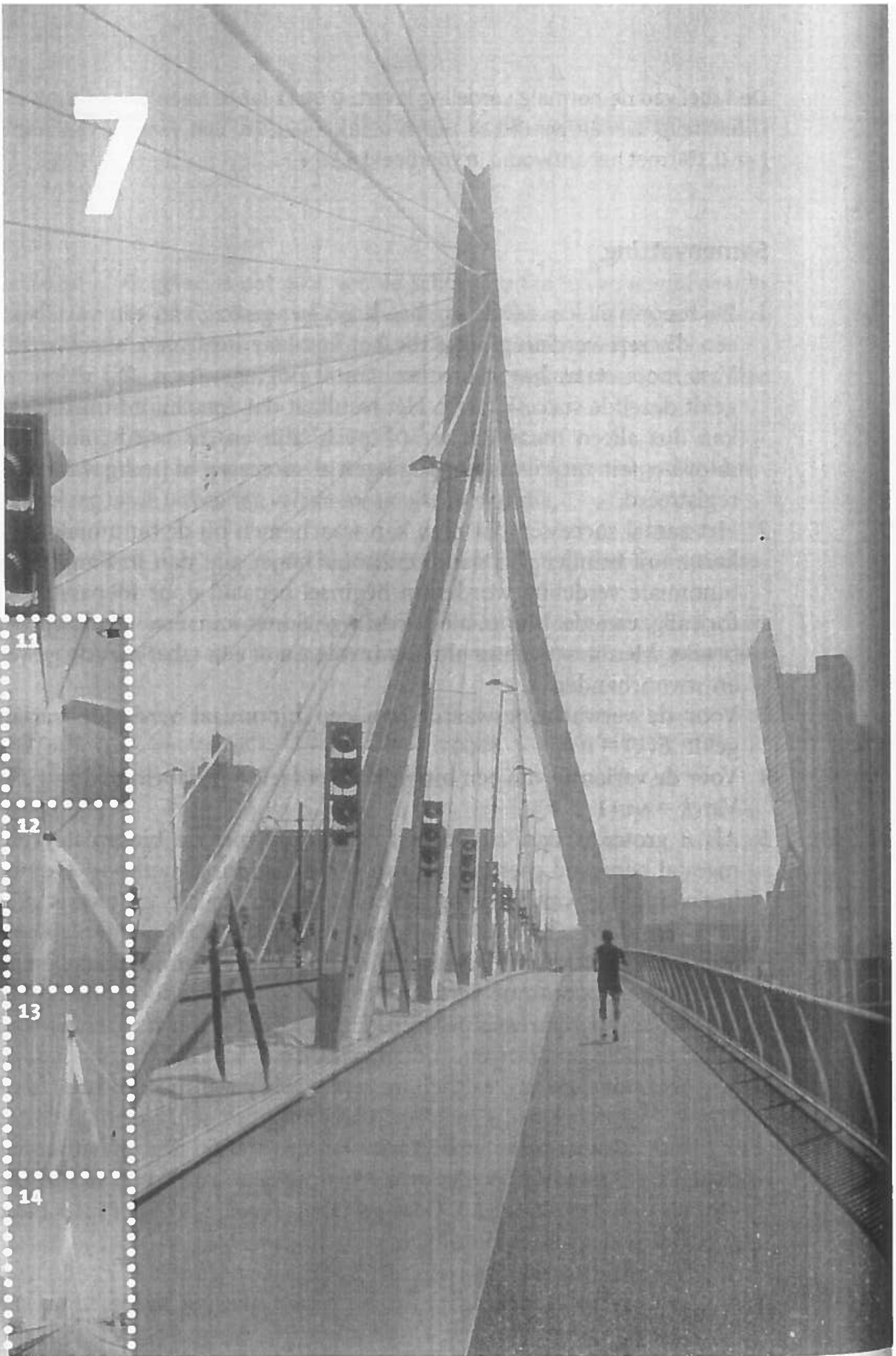
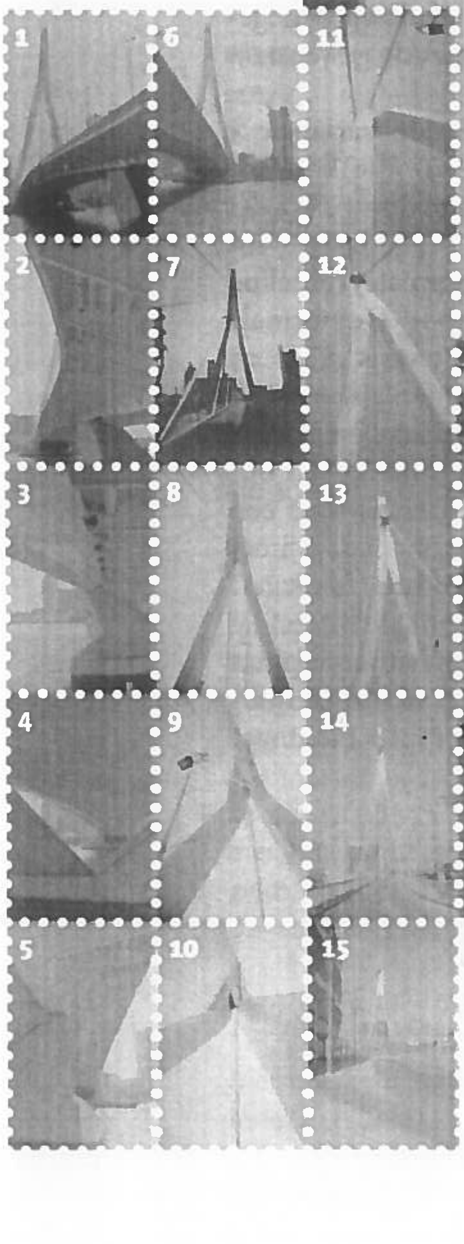
Statistiek om mee te werken

Prof. dr. A. Buijs

Zevende druk

Stenfert Kroese


7


1
6
11
2
7
12
3
8
13
4
9
14
5
10
15

De poissonverdeling

Beschikbare bedden

In een groot regioziekenhuis zijn diverse voorzieningen beschikbaar voor het geval er tijdens een weekend plotseling patiënten moeten worden opgenomen. Zo zijn er op de intensive-care-afdeling standaard drie bedden beschikbaar voor slachtoffers van verkeersongelukken of straatgeweld. Ook dient men soms mensen met acute ziekteverschijnselen op deze afdeling te plaatsen.
Het afdelingshoofd van de intensive-care-afdeling maakt zich de laatste tijd nogal zorgen over de beschikbare capaciteit van deze drie bedden, want recentelijk is het regelmatig voorgekomen dat deze bedden reeds bezet waren wanneer weer een nieuw slachtoffer werd binnengebracht. Om deze reden deed het hoofd een verzoek om voortaan in de weekenden een of twee extra bedden beschikbaar te stellen op de intensive care.
De leiding van het ziekenhuis reageerde vrij negatief op dit verzoek. Men stelde: 'Uit de financiële administratie van het afgelopen jaar is gebleken dat gemiddeld slechts anderhalf bed bezet werd in de weekenden, dus die opmerking over capaciteitsgebrek lijkt ons erg vergezocht ...'

Het afdelinghoofd stelde deze problematiek aan de orde op de wekelijkse afdelingsvergadering. De collega's waren het met hem eens: de laatste maanden hadden ze diverse malen een patiënt per ambulance naar een andere stad moeten doorsturen wegens plaatsgebrek. Eén keer was zo'n patiënt zelfs onderweg overleden... Maar hoe viel deze informatie te rijmen met een bezettingsgraad van slechts 50%? Want dan zou de capaciteit toch ruim voldoende moeten zijn. 'Misschien heeft het iets te maken met de onzekere aantallen,' zei nog iemand, 'maar ja, dat maakt het juist onmogelijk om een verstandige calculatie te maken van de behoefte aan bedden'. Een student die ter afronding van haar studie stage liep bij deze afdeling stond op en zei: 'Heeft hier dan niemand statistiek gehad bij z'n opleiding? Ik denk namelijk dat we met enkele berekeningen met de poissonverdeling deze situatie heel goed kunnen analyseren.' ◀

OPENINGSCASUS

In dit hoofdstuk maken we kennis met een speciale verdeling, de *poissonverdeling*, die kan worden gebruikt voor allerlei praktische toepassingen. Het is een discrete kansverdeling die – evenals de binomiale verdeling – het aantal 'successen' telt. Soms wordt de poissonverdeling de verdeling voor *zeldzame gebeurtenissen* genoemd. Daarmee bedoelen we gebeurtenissen die ieder voor zich een zeer kleine kans hebben om voor te komen, maar door het grote aantal deelnemers worden toch voortdurend realisaties (of successen) waargenomen. Als voorbeeld kunnen we denken aan vliegtuigcrashes. Per vlucht is de kans vrijwel nul op een ongeluk, maar doordat er jaarlijks miljoenen vluchten worden uitgevoerd wordt men toch af en toe geconfronteerd met zo'n ongeluk.

In dit hoofdstuk bespreken we in paragraaf 7.1 eerst de hoofdlijnen van de poissonverdeling en vervolgens wordt in paragraaf 7.2 en 7.3 de relatie gelegd met de normale verdeling en de binomiale verdeling. Verder wordt in paragraaf 7.4 het gebruik van Excel toegelicht. In paragraaf 7.5 volgt een overzicht over het gebruik van de juiste kansverdeling voor een gegeven situatie.

Tot slot wordt in paragraaf 7⁺ aandacht gegeven aan de negatief-exponentiële verdeling. Dat is een continue verdeling die sterk verbonden is met de poissonverdeling. Deze wordt veel gebruikt bij de studie van wachttijdproblemen.

7.1 Poissonverdeling: enkele basisbegrippen

We schetsen in deze paragraaf allereerst in grote lijnen hoe een variabele met een poissonverdeling kan worden omschreven. Vervolgens geven we aan hoe met een formule of een tabel kansen kunnen worden bepaald.

7.1.1 Omschrijving van de poissonverdeling

In dit hoofdstuk houden we ons bezig met de poissonverdeling. Deze kansverdeling is vernoemd naar de Fransman Simeon Poisson, die in 1837 voor het eerst over deze verdeling publiceerde. Evenals de binomiale verdeling heeft de poissonverdeling betrekking op het tellen van successen. Typische voorbeelden van variabelen met een poissonverdeling zijn:

- het aantal klanten dat per uur bij een winkel binnenkomt;
- het aantal binnenkomende telefoongesprekken per minuut;
- het aantal verkeersongelukken per dag in een bepaalde stad;
- het aantal tikken op een geigerteller bij het meten van de radioactiviteit van een preparaat;
- het aantal weeffouten per strekkende meter textiel;
- het aantal bacteriën in een liter water.

Aantal successen of registraties

Al deze voorbeelden hebben gemeen dat er een *aantal* gebeurtenissen, successen of registraties wordt geteld. Bij een binomiale verdeling is de steekproefomvang n gegeven, en is de succeskans π bekend (en bij iedere poging hetzelfde). Bij de poissonverdeling is er in beginsel geen steekproefomvang.

Toch bekijkt men soms de poissonverdeling door de bril van de binomiale verdeling, door te denken aan een zeer grote n en een zeer kleine π. Bij het tellen van verkeersongelukken zou men het beeld kunnen vormen van een zeer groot aantal verkeersdeelnemers die per persoon een zeer kleine kans hebben op een verkeersongeluk op een gegeven dag. Tezamen zorgen al die deelnemers dagelijks echter toch voor een bepaald aantal ongelukken.
Bij een radioactief preparaat kan men denken aan praktisch oneindig veel deeltjes materie. Elk deeltje heeft een uiterst kleine kans op het uitzenden van straling gedurende een bepaalde meetperiode. Tezamen zorgen die deeltjes er echter voor dat een aantal registraties wordt gedaan op een geigerteller.
Vaak speelt de *tijd* een rol bij de poissonverdeling. Als we verkeersongelukken tellen, dan hangt het aantal geregistreerde ongelukken uiteraard af van de duur van de waarnemingsperiode. Als we bijvoorbeeld het aantal ongelukken per week tellen, dan zal dit wel ongeveer zeven maal het aantal verwachte ongelukken per dag zijn. Voor het gemak spreken we vaak over de tijd als parameter bij de poissonverdeling. Andere voorbeelden zijn echter denkbaar. Zo hangt het aantal bacteriën in een monster water uiteraard af van de hoeveelheid water en hangt het aantal te tellen weeffouten in een rol stof af van het aantal onderzochte meters.

7.1.2 De poissonformule

Omdat we ons bezighouden met het tellen van een aantal successen (registraties) gedurende een periode van waarneming, is het direct in te zien dat de kansvariabele $\underline{k}$ die het aantal registraties weergeeft alleen niet-negatieve waarden kan aannemen.
Gegeven is een tijdsinterval van lengte t en een parameter λ (die het gemiddelde of verwachte aantal successen per tijdseenheid aangeeft). Voor de kans op een aantal 'successen' k geldt dan de volgende formule:

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

In plaats van λt schrijft men wel μ. De formule wordt dan:

$$P(\underline{k}=k)=\frac{\mu^k}{k!}\,e^{-\mu} \quad \text{voor} \quad k=0, 1, 2, \ldots \qquad [7.1]$$

Als bijvoorbeeld voor een bankkantoor gegeven is dat het aantal klanten volgens een possionverdeling binnenkomt en dat er gemiddeld $\lambda=3$ klanten per kwartier binnenkomen, dan is het aantal klanten per *uur* (= vier kwartier) volgens de poissonverdeling verdeeld met $\mu=\lambda t=3\times 4=12$. We schrijven dit als $\underline{k}\sim$ Poisson ($\mu=12$). Door de keuze van μ (of λ), die de *poissonparameter* wordt genoemd, is de kansvariabele $\underline{k}$ volledig gekarakteriseerd.

Poissonparameter

Het aantal successen moet uiteraard minstens 0 zijn. De poissonformule is geldig voor elke gehele waarde van k (met $k\geqq 0$). Hiermee zien we een belangrijk verschilpunt met de binomiale verdeling. Bij de binomiale verdeling kan het waargenomen aantal successen nooit groter zijn dan n (het aantal pogingen). Bij de poissonverdeling kennen we geen maximum aantal pogingen. In principe kunnen we voor elke (mogelijk zeer grote) waarde van k de kans op optreden daarvan uitrekenen. Door invullen in de formule zal ech-

μ is het gemiddelde

ter blijken dat bij voldoende grote k de kans verwaarloosbaar klein wordt. Als bijvoorbeeld het gemiddelde aantal verkeersongelukken in een bepaalde stad 5 bedraagt (dat wil zeggen $\mu = 5$ in termen van de poissonverdeling), dan kan het best eens voorkomen dat op zekere dag tien of twaalf of zelfs vijftien ongelukken gebeuren.
Het aantal van 40 ongelukken op zekere dag lijkt echter uitgesloten. Dat wordt ook bevestigd door de formule van de poissonverdeling indien we $k = 40$ invullen.

▸ *Voorbeeld 7.1* Bij een bankkantoor is vastgesteld dat het aantal klanten dat per kwartier binnenkomt, beschreven kan worden door een poissonverdeling met $\mu = 3$. Hoe groot is dan de kans dat in een willekeurige periode van een kwartier precies twee klanten binnenkomen?
We bestuderen dus de kansvariabele $\underline{k}$, die het aantal klanten per kwartier aangeeft. Toepassing van de formule levert:

$$P(\underline{k} = 2) = \frac{3^2}{2!} e^{-3} = \frac{9}{2} \cdot 0{,}0498 = 0{,}2240$$ ◂

Uiteraard is ook bij de poissonverdeling de som van de kansen gelijk aan 1. Een bewijs hiervan is opgenomen in appendix A.
Voor kleine waarde van μ is de poissonverdeling een typisch voorbeeld van een scheve verdeling.

De formule van de poissonverdeling laat ons zien dat er slechts één parameter een rol speelt bij het berekenen van de kansen op de verschillende uitkomsten van $\underline{k}$. Dat is de parameter μ ($= \lambda t$), waarvan men kan aantonen dat deze juist gelijk is aan de *verwachtingswaarde* en de *variantie* van $\underline{k}$. Het gelijk zijn van verwachtingswaarde en variantie is één van de meest opvallende eigenschappen van de poissonverdeling:

$$E(\underline{k}) = \mu \text{ en} \tag{7.2}$$

$$\mathrm{Var}(\underline{k}) = \mu \tag{7.3}$$

Hieruit volgt onmiddellijk dat voor de standaarddeviatie geldt $\sigma = \sqrt{\mu}$.
Deze eigenschappen kunnen worden afgeleid door de definities van verwachtingswaarde en variantie toe te passen (zie appendix A voor de afleiding van deze resultaten).
Door middel van één parameter (μ) is een variabele met poissonverdeling daarom volledig gekarakteriseerd. Het berekenen van de kansen kan bij een poissonverdeling op verschillende manieren worden aangepakt, namelijk:
- rechtstreeks met de formule (zoals in voorbeeld 7.1);
- met een poissontabel;
- met behulp van een benadering met de normale verdeling.

Hierna werken we dat nader uit.

7.1.3 De poissontabellen

Van de poissonverdeling zijn twee soorten tabellen beschikbaar, waarin bij gegeven μ de kansen af te lezen zijn. Allereerst is een 'gewone' poissontabel opgenomen. In deze tabel zijn de kansen op precies k successen, dus $P(\underline{k} = k)$ bij een gegeven μ af te lezen. Direct verwant met deze tabel is de tabel met de cumulatieve kansen van de poissonverdeling. In deze tabel kan men de kans aflezen op k of minder successen, dus $P(\underline{k} \leqq k)$ bij gegeven μ. De tabellen zijn weergegeven als tabel C4 en C5 in de appendix.

▸ *Voorbeeld 7.2* Het aantal alarmmeldingen $\underline{k}$ dat per uur binnenkomt bij een centrale meldkamer blijkt een poissonverdeling te volgen met een gemiddelde van $\mu = 1{,}5$ per uur.
Dus $\underline{k} \sim \text{Poisson}\,(\mu = 1{,}5)$.
De kans dat in een willekeurig uur geen enkele melding binnenkomt, wordt weergegeven door $P(\underline{k} = 0)$. Deze volgt dan uit:

$$P(\underline{k} = 0 \mid \mu = 1{,}5) = \frac{(1{,}5)^0}{0!} e^{-1{,}5} = e^{-1{,}5} = 0{,}2231$$

De kans op bijvoorbeeld precies drie alarmmeldingen in een uur bedraagt:

$$P(\underline{k} = 3 \mid \mu = 1{,}5) = \frac{(1{,}5)^3}{3!} e^{-1{,}5} = 0{,}1255$$

We kunnen deze kansen ook aflezen in de tabel. De gewone poissontabel (tabel C4) levert bij $\mu = 1{,}5$ de gevraagde kansen: $P(\underline{k} = 0) = 0{,}2231$ en $P(\underline{k} = 3) = 0{,}1255$.

Ook de cumulatieve poissontabel had kunnen worden gebruikt. We vinden dan
$P(\underline{k} \leqq 0) = 0{,}2231$ en
$P(\underline{k} = 3) = P(\underline{k} \leqq 3) - P(\underline{k} \leqq 2) = 0{,}9344 - 0{,}8088 = 0{,}1256$.
Merk op dat dit laatste antwoord een klein verschil laat zien met het rechtstreeks berekende antwoord. Dat ligt aan afrondingen bij de tabelwaarden. ◂

▸ *Voorbeeld 7.3* Bij het bankkantoor uit voorbeeld 7.1 komen per kwartier gemiddeld $\mu = 3$ klanten binnen. Hoe groot is de kans dat in een half uur precies vier klanten binnenkomen?

We definiëren $\underline{k}$ als het aantal klanten in een half uur.
Vanuit het gezichtspunt $\mu = \lambda t$ zou men kunnen zeggen dat λ het gemiddeld aantal klanten per kwartier is en t het aantal kwartieren. Dus per half uur geldt $\mu = 3 \times 2 = 6$.

In de tabel is te zien bij $\mu = 6$ dat $P(\underline{k} = 4 \mid \mu = 6) = 0{,}1339$.

Merk op dat deze kans *niet* moet worden berekend door de kans op $k = 2$ klanten in het eerste kwartier en dezelfde kans in het tweede kwartier samen te nemen. Dat is ook logisch, want vier klanten in een half uur kunnen op allerlei manieren tot stand komen, zoals nul in het eerste kwartier en vier in het tweede kwartier, of een in het eerste kwartier en drie in het tweede kwartier, maar ook (2, 2), (3, 1) en (4, 0). Een aardige oefening in dit verband is om de kansen op de hier genoemde vijf gevallen afzonderlijk te gaan bepalen. Dit kan door in de tabel in de kolom bij $\mu = 3$ te kijken en dan door middel van vermenigvuldiging van kansen de kansen op de paren te berekenen. Optelling van deze vijf kansen moet dan weer 0,1339 opleveren.

Hoe groot is de kans dat er meer dan drie klanten in een kwartier binnenkomen? Dat is dus:

$$P(\underline{k} > 3) = P(\underline{k} \geqq 4) = \sum_{k=4}^{\infty} P(\underline{k} = k)$$

Omdat de poissonverdeling in principe oneindig veel termen bevat, is het handig de vraag om te keren.

$$\begin{aligned} P(\underline{k} \geqq 4) &= 1 - P(\underline{k} \leqq 3) \\ &= 1 - P(\underline{k} = 0) - P(\underline{k} = 1) - P(\underline{k} = 2) - P(\underline{k} = 3) \\ &= 1 - 0{,}0498 - 0{,}1494 - 0{,}2240 - 0{,}2240 \\ &= 0{,}3528 \end{aligned}$$

Deze kansen zijn af te lezen in de tabel van de poissonverdeling bij $\mu = 3$. We hadden $P(\underline{k} \leqq 3)$ ook rechtstreeks kunnen aflezen in de tabel van de cumulatieve kansen van de poissonverdeling bij $\mu = 3$, want dan zien we $P(\underline{k} \leqq 3 \mid \mu = 3) = 0{,}6472$. De gevraagde kans is dus dan $1 - 0{,}6472 = 0{,}3528$. ◂

In voorbeeld 7.3 hebben we impliciet gebruikgemaakt van een belangrijke eigenschap van variabelen met een poissonverdeling, namelijk:

Eigenschap

De som van variabelen met een poissonverdeling levert opnieuw een variabele met een poissonverdeling.

De μ van de somverdeling is de optelling van de μ's van de oorspronkelijke variabelen. (Voor het bewijs van deze eigenschap verwijzen we naar appendix A.) Dus als $\underline{k}$ en $\underline{m}$ onderling onafhankelijke kansvariabelen met een poissonverdeling zijn met als parameters $\mu_{\underline{k}}$ en $\mu_{\underline{m}}$, dan heeft de kansvariabele $\underline{k} + \underline{m}$ een poissonverdeling met verwachtingswaarde $\mu_{\underline{k}} + \mu_{\underline{m}}$. Deze eigenschap kan uiteraard worden uitgebreid voor optelling van een grotere serie onderling onafhankelijke variabelen met een poissonverdeling.

7.2 Benadering met behulp van de normale verdeling

De verwachtingswaarde en variantie zijn van groot belang bij het berekenen van kansen bij een poissonverdeling met grote μ. Indien μ groot is, bepalen we de kansen namelijk niet meer met behulp van de formule of met behulp van een poissontabel. Voor $\mu \geqq 10$ kunnen we door toepassing van de normale benadering de kansen berekenen op een manier die volkomen overeenkomt met de normale benadering van de binomiale verdeling. We behandelen deze methode in een voorbeeld.

▸ *Voorbeeld 7.4* Bij een bankkantoor kan het aantal binnenkomende klanten per tijdseenheid beschreven worden door een poissonverdeling. Per kwartier geldt dat gemiddeld $\mu = 3$ klanten binnenkomen. Per dag is de bank tien uur geopend. Hoe groot is de kans dat op een willekeurige dag meer dan 140 klanten komen?
Een periode van tien uur is gelijk aan 40 kwartieren. Per kwartier komt een aantal klanten binnen volgens een poissonverdeling met een gemiddelde van $\lambda = 3$. Dus het aantal klanten per dag is een variabele $\underline{k}$ met een poissonverdeling met verwachtingswaarde $3 \times 40 = 120$. Bij een dergelijke grote waarde van μ (namelijk $\mu \geqq 10$) passen we de normale benadering toe bij de bepaling van kansen.

Voor de variabele $\underline{k}$ met een poissonverdeling geldt hier:

$E(\underline{k}) = 120$ en $Var(\underline{k}) = 120$

We benaderen de variabele $\underline{k}$ door een variabele $\underline{x}$ die normaal verdeeld is met $\mu = 120$ en $\sigma^2 = 120$. We bepalen $P(\underline{k} > 140)$ door een geschikte grens te kiezen bij de normaal verdeelde variabele $\underline{x}$. Hierbij moeten we rekening houden met een *continuïteitscorrectie*. Te bepalen is derhalve $P(\underline{x} > 140{,}5)$.

Continuïteitscorrectie

Voor de verdeling van $\underline{x}$ geldt:

$\underline{x} \sim N(\mu = 120, \sigma = \sqrt{120})$, dus $\sigma = 10{,}95$

We vinden dan $z = \dfrac{140{,}5 - 120}{10{,}95} = 1{,}87$.

Toepassing van de tabel van de normale verdeling levert dan $P(\underline{x} > 140{,}5) = 0{,}0307$. In termen van de discrete poissonverdeling betekent dit voor de kans op meer dan 140 klanten per dag: $P(\underline{k} > 140) = 0{,}0307$.
Merk op bij dit voorbeeld dat hierbij de veronderstelling is gemaakt dat voor alle kwartieren van die dag een poissonverdeling met dezelfde μ geldt. In de praktijk zul je op een dag vermoedelijk periodes aantreffen die steevast wat drukker of wat stiller zijn. ◂

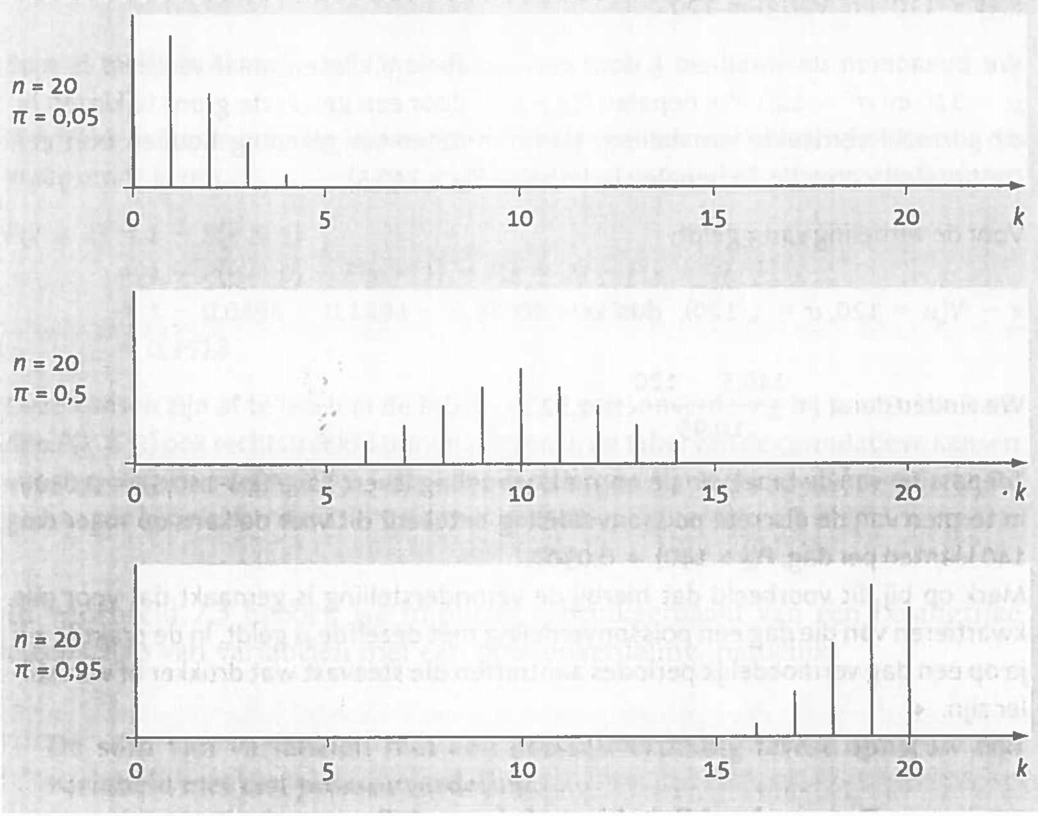
7.3 Toepassing bij de binomiale verdeling

De poissonverdeling kan in een bepaald speciaal geval handig worden gebruikt om kansen te berekenen voor een variabele die eigenlijk binomiaal verdeeld is. Het gaat dan om gevallen waarbij het aantal pogingen n altijd zeer groot is en de succeskans ofwel π ofwel zeer klein is (bijna nul) ofwel zeer groot is (bijna 1).

7.3.1 Een binomiale verdeling met grote n en zeer kleine π

In de inleiding hebben we bij de poissonverdeling een beeld opgeroepen van een binomiale verdeling met zeer grote n en zeer kleine π. Dit beeld gaan we gebruiken om hiervoor de kansen te kunnen berekenen, waarbij we gebruikmaken van een benaderingsmethode met de poissonverdeling.
We hebben reeds in hoofdstuk 6 kennisgemaakt met de binomiale verdeling. Een motief voor het kiezen van een benadering voor de binomiale verdeling was gelegen in het feit dat bij grote n de berekeningen met de binomiaalformule uitermate onaantrekkelijk worden. Het was daarom plezierig de normale verdeling te kunnen gebruiken.
Dit is echter alleen toegestaan indien aan bepaalde voorwaarden is voldaan, namelijk indien het aantal pogingen $n \geqq 20$ en tevens moet nog gelden $n\pi \geqq 5$ en $n(1 - \pi) \geqq 5$. Dus indien $n\pi < 5$ of indien $n(1 - \pi) < 5$, mag de normale benadering niet worden toegepast. Dit wordt eigenlijk gemakkelijk zichtbaar als je voor zo'n situatie een grafiek van de binomiale verdeling maakt, want dan blijkt dat de binomiale verdeling een nogal scheve vorm heeft, zodat op voorhand al te zien is dat daarbij geen mooie symmetrische normale verdeling past. Dit is geïllustreerd in afbeelding 7.1.

Afbeelding 7.1 Grafieken van een binomiale verdeling



In afbeelding 7.1 is de binomiale verdeling weergegeven voor $n = 20$ en respectievelijk $\pi = 0{,}05$, $\pi = 0{,}50$ en $\pi = 0{,}95$. Duidelijk is te zien dat de grafieken voor $\pi = 0{,}05$ en $\pi = 0{,}95$ zeer scheef zijn. Voor $n = 20$ en $\pi = 0{,}05$ blijkt dat nagenoeg aan één van de fundamentele eigenschappen van de poissonverdeling voldaan is. De variantie en de verwachtingswaarde zijn namelijk bijna gelijk aan elkaar. Immers, voor $n = 20$ en $\pi = 0{,}05$ geldt bij de binomiale verdeling:

$$E(\underline{k}) = n\pi = 20 \times 0{,}05 = 1$$
$$\text{Var}(\underline{k}) = n\pi(1 - \pi) = 20 \times 0{,}05 \times 0{,}95 = 0{,}95$$

Maar eigenlijk komt dit effect nog scherper tot uiting als we met een veel hogere n en een veel kleinere π werken. Bijvoorbeeld bij $n = 1\,000$ en $\pi = 0{,}001$ zou gelden:

$$E(\underline{k}) = n\pi = 1\,000 \times 0{,}001 = 1$$
$$\text{Var}(\underline{k}) = n\pi(1 - \pi) = 1\,000 \times 0{,}001 \times 0{,}999 = 0{,}999$$

Hier verschillen dus de verwachting en de variantie bijna helemaal niets.

Poisson-benadering

Om de kansen bij dergelijke binomiale verdelingen te bepalen, is het soms handig om gebruik te maken van de *poissonbenadering*. Door de poissonverdeling met $\mu = 1$ te kiezen (dat wil zeggen met dezelfde verwachtingswaarde als de hierboven besproken binomiale verdeling) kunnen we alle gewenste kansen aflezen in de tabel van de poissonverdelingen. Zouden we bijvoorbeeld de kans op nul successen, $P(\underline{k} = 0)$, willen weten bij de binomiale verdeling met $n = 20$ en $\pi = 0{,}05$, dan lezen we bij $\mu = 1$ in de poissontabel af dat $P(\underline{k} = 0 \mid \mu = 1) = 0{,}3679$.

7.3.2 Rolverwisseling

Ook bij een binomiale verdeling met een zeer *hoge* succeskans (π bijna 1) kunnen we de poissonbenadering toepassen. We moeten dan de rol van π en $(1 - \pi)$ verwisselen. In plaats van 'successen' gaan we dan 'missers' tellen. Als de succeskans bijvoorbeeld 0,95 bedraagt, dan geldt voor de kans op een misser $\pi = 0{,}05$. Bij $\mu = n\pi$ kunnen de kansen op een bepaald aantal missers dan weer worden afgelezen in de poissontabel.

π bijna 1, dus $1 - \pi$ bijna 0

Als we $P(\underline{k} = 18)$ moeten bepalen bij een binomiale verdeling met $n = 20$ en $\pi = 0{,}95$, dan moeten we bedenken dat de gebeurtenis '18 successen' overeenkomt met de gebeurtenis '2 missers'. We moeten dus $P(\underline{k} = 2)$ bepalen bij een binomiale verdeling met $n = 20$ en $\pi = 0{,}05$. We kiezen dan als benadering de poissonverdeling met $\mu = 1$ $(= n\pi = 20 \times 0{,}05)$. Bij $\mu = 1$ levert de tabel dan $P(\underline{k} = 2 \mid \mu = 1) = 0{,}1839$.

Het is interessant deze laatste uitkomst te vergelijken met de 'echte' binomiale kans. Bij $n = 20$ en $\pi = 0{,}05$ vinden we in de binomiale tabel $P(\underline{k} = 2) = P(\underline{k} \leqq 2) - P(\underline{k} \leqq 1) = 0{,}9245 - 0{,}7358 = 0{,}1887$. Hier is toch nog sprake van een duidelijk verschil. Men kan echter nagaan dat de benadering beter wordt naarmate n groter is. En juist bij hoge n (met zeer kleine π) is deze benadering zinvol, zoals blijkt uit het volgende voorbeeld waarbij berekeningen met de binomiale verdeling in feite onmogelijk zouden zijn.

▸ *Voorbeeld 7.5* Van een medische test is bekend dat deze bij 0,2% van alle proefpersonen tot een allergische reactie leidt. Bij een opleidingsinstituut wordt bij 1 500 studenten deze medische test afgenomen. Hoe groot is de kans dat meer dan zes studenten een allergische reactie vertonen?
Het aantal probleemgevallen $\underline{k}$ is binomiaal verdeeld met $n = 1\,500$ en $\pi = 0{,}002$. Dus $\underline{k}$ – Bin($n = 1\,500$, $\pi = 0{,}002$). Omdat $n \geqq 20$ mogen we de binomiale verdeling gaan benaderen door een andere verdeling.
Omdat $n\pi = 3$ (dus kleiner dan 5) moeten we hiervoor de poissonverdeling kiezen. Welke poissonverdeling? Uiteraard de poissonverdeling met dezelfde verwachtingswaarde. We kiezen dus als benadering de variabele $\underline{k}$ die een poissonverdeling heeft met $\mu = 3$. Gevraagd is $P(\underline{k} > 6)$ oftewel $P(\underline{k} \geqq 7)$. We keren deze uitspraak om:

$P(\underline{k} \geqq 7) = 1 - P(\underline{k} \leqq 6)$

We lezen bij $\mu = 3$ af: $P(\underline{k} \leqq 6 \mid \mu = 3) = 0{,}9665$

Dus $P(\underline{k} \geqq 7) = 1 - 0{,}9665 = 0{,}0335$.

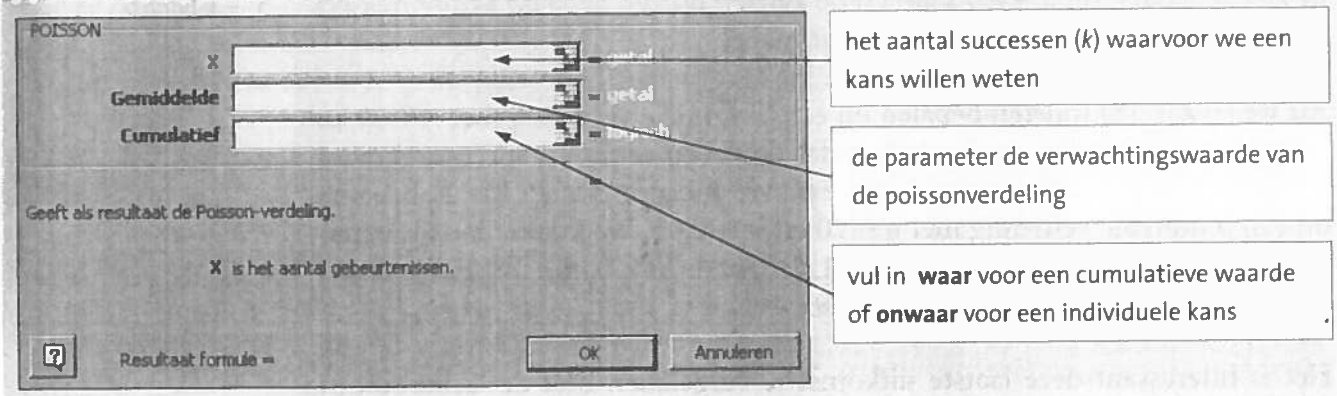
Merk op dat in dit geval de variantie van de hier toepasselijke binomiale verdeling gelijk is aan $\text{Var}(\underline{k}) = n\pi(1 - \pi) = 1\,500 \times 0{,}002 \times 0{,}998 = 2{,}994$. Dat is dus nagenoeg gelijk aan 3, waaruit blijkt dat deze binomiale verdeling vrijwel voldoet aan de poissoneigenschap $E(\underline{k}) = \text{Var}(\underline{k}) = \mu$. ◂

7.4 Werken met Excel

Ook voor het berekenen van kansen voor een poissonverdeling biedt het programma Excel mogelijkheden. Een voordeel ten opzichte van het gebruik van tabellen is, dat we in Excel bij *iedere* waarde van μ de bijbehorende kans kunnen bepalen.

In de rubriek ‘Statistisch’ kiest men de functie ‘POISSON’. We zien dan een dialoogbox waarin we drie gegevens moeten invullen, namelijk X, μ en een aanduiding waarmee we aangeven of we een ‘losse’ kans willen weten of een cumulatieve kans.

Afbeelding 7.2 Dialoogvenster POISSON



Dus als we *waar* invullen krijgen we de kans op een aantal successen kleiner of gelijk aan k, en bij *onwaar* krijgen we de kans op precies k successen.

▸ *Voorbeeld 7.6* Op de intensive-care-afdeling die is besproken in de openingscasus wordt gesuggereerd dat het aantal patiënten dat in een weekend binnenkomt, een kansvariabele is met een poissonverdeling. Men besloot daarom de gegevens van 50 voorgaande weekenden te verzamelen. Hierbij bleek dat per weekend gemiddeld 1,65 kandidaten werden aangemeld om te worden opgenomen. Met Excel kunnen we de diverse kansen berekenen.

Patiënten moeten worden doorgestuurd als er meer dan drie gevallen in een weekend optreden. Als kansen voor het aantal bezette bedden $\underline{k}$ vinden we met Excel de waarden in afbeelding 7.3.

Afbeelding 7.3 Poissonkansen

	A	B	C
1		k	P
2		0	19,20%
3		1	31,69%
4		2	26,14%
5		3	14,38%
6	meer dan	3	8,59%

Merk op dat $P(\underline{k} > 3)$ wordt bepaald via $1 - P(\underline{k} \leq 3)$, in Excel is dat ‘=1-POISSON(3;1,65;waar).’
De gemiddelde bezettingsgraad van de bedden is dus direct te bepalen:

$$0 \times 0{,}1920 + 1 \times 0{,}3169 + 2 \times 0{,}2614 + 3 \times (0{,}1438 + 0{,}0859) = 1{,}5287$$

Opmerking:
Hoe is het nu mogelijk dat voor de verzamelde gegevens bleek dat er gemiddeld 1,65 patiënten per weekend binnenkwamen, terwijl de financiële administratie spreekt van 1,50 per weekend? En de berekening van de gemiddelde bezettingsgraad die we hiervoor hebben weergegeven, levert 1,5287 personen.
Het antwoord is dat af en toe (noodgedwongen) een aantal patiënten niet kon worden opgenomen omdat de drie bedden al vol waren. Dus het gemiddelde aantal daadwerkelijk *opgenomen* patiënten is lager dan het aantal *aangeboden* patiënten. ◂

7.5 Wanneer welke kansverdeling?

In de hoofdstukken 4 tot en met 7 hebben we kennisgemaakt met een aantal verschillende kansverdelingen. Soms kunnen er misverstanden ontstaan over de correcte toepassing van de verschillende verdelingen, omdat in een aantal gevallen de ene kansverdeling moet worden gebruikt als benadering van een andere verdeling. We geven hieronder enkele overwegingen om deze problemen aan te kunnen pakken.

7.5.1 Wat voor variabele hebben we eigenlijk?

Bij het bepalen van de juiste berekeningsmethode onderscheiden we twee stappen, weergegeven door de vragen:
a Wat is de theoretisch correcte omschrijving van de variabele? Met andere woorden: Met wat voor variabele hebben we te maken?
b Gegeven het soort variabele, met welke rekenkundige methodes moeten kansen worden berekend?

Wanneer we in het kader van een onderzoek een variabele bestuderen, is een eerste vraag of een waarneming een afzonderlijk getal x_i oplevert, of dat de variabele een totaaltelling geeft van het aantal gebeurtenissen of successen. Indien dat laatste het geval is, dan hebben we vaak te maken met de binomiale verdeling (als de steekproefomvang n bekend is en de succeskans π bij iedere trekking hetzelfde is) of met de poissonverdeling. Deze laatste is soms herkenbaar als je het beeld in gedachten kunt nemen van een extreem grote n gecombineerd met een uitzonderlijk lage succeskans π. Andere kansverdelingen zijn echter denkbaar, bijvoorbeeld als het aantal elementen van de populatie (N) eindig is. Als dan *tevens* lotingen *zonder* teruglegging plaatsvinden, moet de hypergeometrische verdeling worden gebruikt.

Als elke waarneming een reëel getal x_i oplevert, dan kan men vaak gebruikmaken van de normale kansverdeling. Twee redenen hiervoor kunnen zijn: in de eerste plaats is het mogelijk dat het op de een of andere manier bekend is dat de verdeling van de variabele *zelf* een normale verdeling volgt. In de tweede plaats is het mogelijk dat een verdeling eigenlijk niet precies die mooie symmetrische gedaante van de normale verdeling heeft, maar dat er uitsluitend berekeningen met betrekking tot een steekproefgemiddelde moeten worden verricht. In dat geval komt de centrale limietstelling te hulp.

7.5.2 Hoe worden de kansen berekend?

Nadat is vastgesteld welk type variabele we bestuderen, komt de tweede vraag aan de orde, namelijk: 'Hoe rekenen we de kansen uit?'

De variabele volgt een normale verdeling

Als reeds gegeven (of aannemelijk) is dat een variabele een normale verdeling volgt, dan moet de tabel van de standaardnormale verdeling worden toegepast om allerlei kansen te berekenen. Als uitspraken moeten worden gedaan over het steekproefgemiddelde $\underline{\bar{x}}$ van n trekkingen, dan volgt dit gemiddelde eveneens een normale verdeling met standaarddeviatie:

$$\frac{\sigma}{\sqrt{n}}$$

Eindige-populatie-correctie

De enige complicatie die we bij steekproefgemiddelden kunnen tegenkomen, is de *eindige-populatie-correctie* (*epc*). Als de steekproef zonder teruglegging geschiedt én als bovendien de steekproefomvang n meer is dan 10 % van de populatieomvang N, dan moet de epc worden toegepast wanneer wordt vastgesteld wat de standaarddeviatie is van het steekproefgemiddelde.

De variabele volgt een willekeurige continue verdeling

Om voor een variabele $\underline{x}$ met een willekeurige continue kansverdeling kansen te kunnen berekenen, moeten we in principe op de hoogte zijn van de formulering van de kansdichtheid. De methoden uit hoofdstuk 4 kunnen dan worden toegepast. Twee speciale gevallen komen in dit boek aan de orde, namelijk de rechthoekige verdeling (paragraaf 4^+) en de negatief-exponentiële verdeling (paragraaf 7^+).
We stuiten op een belangrijk punt als we het steekproefgemiddelde van een serie van n trekkingen bestuderen. Zoals aangegeven is in hoofdstuk 5, bij de bespreking van de centrale limietstelling, mag voor een steekproefgemiddelde heel vaak worden verondersteld dat dit een kansvariabele is met een normale verdeling.
Dus: zelfs als we weten dat een variabele $\underline{x}$ zelf niet normaal verdeeld is, kan voor het berekenen van kansen voor het steekproefgemiddelde $\underline{\bar{x}}$ van een serie van n trekkingen van deze variabele toch de normale verdeling worden gebruikt.

De variabele volgt een binomiale verdeling

Een binomiale verdeling is vastgelegd door de parameters n en π. Voor het berekenen van de kansen maken we allereerst het onderscheid n klein ($n \leqq 20$) en n groot ($n > 20$).
Als n klein is, moeten de kansen van een binomiaal verdeelde variabele $\underline{k}$ in principe berekend worden door toepassing van de kansformule. In sommige gevallen (als π gunstig uitvalt) kan direct een tabel geraadpleegd worden. Als n groot is, worden de kansen meestal bepaald met behulp van een benaderingsmethode. Indien geldt $n\pi \geqq 5$ en $n(1-\pi) \geqq 5$, dan gebruiken we de normale verdeling als hulpmiddel. Geldt echter $n\pi < 5$, dan moet de poissonverdeling worden gebruikt met $\mu = n\pi$. Een hiermee verwante aanpak geldt als $n(1-\pi) < 5$. In dat geval moeten we echter eerst een rol-

verwisseling toepassen, waarbij we 'pechgevallen' in plaats van 'successen' tellen. In dat geval gebruiken we de poissonverdeling met $\mu = n(1 - \pi)$.

De variabele volgt een poissonverdeling
Bij een poissonverdeling zijn er in principe drie manieren om de berekeningen uit te voeren, namelijk:
- de kansformule;
- de kanstabel;
- de normale benadering.

De keuze van de methode van aanpak verloopt analoog aan de gang van zaken bij de binomiale verdeling. Dus bij een kleine waarde van μ dient men de kansformule of de tabel te gebruiken, terwijl vanaf $\mu = 10$ meestal de benadering met de normale verdeling zal worden gebruikt.

De variabele volgt een willekeurige discrete verdeling
In dit geval dient er in beginsel te worden gewerkt met de eigen kansfunctie van de variabele. Wel kan men bij het werken met de som (of het gemiddelde) van een aantal trekkingen uit zo'n verdeling weer te maken krijgen met de normale verdeling als een verdedigbare benaderingsmethode. Als voorbeeld noemen we hier de som van het aantal ogen na 100 worpen met een eerlijke dobbelsteen.

Tabel 7.1 Welke methode gebruiken we bij het berekenen van de kansen?

De variabele heeft een:		**Kansen berekenen met:**
normale verdeling		tabel van de normale verdeling met z-waarde
continue verdeling		bij steekproefgemiddelden de normale verdeling met $\frac{\sigma}{\sqrt{n}}$
binomiale verdeling		
(i)	$n < 20$	binomiale formule of de tabel
(ii)	$n \geqq 20$ $n\pi \geqq 5$ en $n(1 - \pi) \geqq 5$	normale benadering met $\mu = n\pi$ en $\sigma = \sqrt{n\pi(1 - \pi)}$
(iii)	$n \geqq 20$ $n\pi < 5$	poissonbenadering met $\mu = n\pi$
(iv)	$n \geqq 20$ $n(1 - \pi) < 5$	poissonbenadering na rolverwisseling van succes en pech
poissonverdeling		
	$\mu < 10$	tabel of formule
	$\mu \geqq 10$	normale benadering met μ en $\sigma = \sqrt{\mu}$

7^+ Twee verwante verdelingen

In deze paragraaf bespreken we een tweetal kansverdelingen die een nauw verband hebben met de poissonverdeling, namelijk de negatief-exponentiële verdeling en de gammaverdeling.

$7^+.1$ De negatief-exponentiële verdeling

Bij de poissonverdeling tellen we het aantal successen of registraties gedurende een bepaald (tijds)interval. Bij het binnenkomen van klanten in een winkel zou men ook een verwant probleem kunnen bestuderen, namelijk: hoeveel tijd verstrijkt er tussen twee binnenkomsten van klanten?

Om deze vraag op te lossen, denken we ons in, dat we een stopwatch indrukken bij binnenkomst van de laatste klant. We meten de tijd die voorbijgaat tot de volgende klant binnenkomt. De poissonverdeling die het aantal binnenkomende klanten per gefixeerde tijdsperiode beschrijft, is gekenmerkt door een parameter λ (verwachtingswaarde per minuut). Als t het aantal minuten is (de waarnemingsperiode), dan wordt het aantal binnenkomende klanten beschreven door de poissonverdeling met parameter λt.

Relatie met poisson

Veronderstel dat de laatste klant zojuist is binnengekomen. Hoe groot is dan de kans dat we minstens t minuten moeten wachten op de volgende klant? Als we minstens t minuten moeten wachten, dan houdt dit in dat er gedurende t minuten niemand binnenkomt. Er worden dus $k = 0$ klanten geteld binnen dat tijdvak. Bij een vast gekozen tijdsduur t kunnen we de kans hierop uitrekenen met de poissonverdeling. Dat is:

$$P(\underline{k} = 0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

Noem $\underline{t}$ de wachttijd op de eerstvolgende klant. Er geldt dan:

$$P(\underline{t} > t) = P(\underline{k} = 0 \text{ in } t \text{ minuten}) = e^{-\lambda t}$$

Als er in t minuten niemand binnenkomt, dan moeten we *dus* langer dan t minuten wachten. Er geldt dus $P(\underline{t} > t) = e^{-\lambda t}$. Derhalve geldt voor de kans dat we hoogstens t minuten moeten wachten:

$$P(\underline{t} \leqq t) = 1 - e^{-\lambda t}$$

Omdat hier een kans met een $\leqq$-teken staat, herkennen we de schrijfwijze van de verdelingsfunctie. Hiermee kunnen we de kansdichtheid van $\underline{t}$ bepalen. De verdelingsfunctie $F(t)$ is gedefinieerd als:

$$F(t) = P(\underline{t} \leqq t) = 1 - e^{-\lambda t}$$

Kansdichtheid

De kansdichtheid is de afgeleide van de verdelingsfunctie.

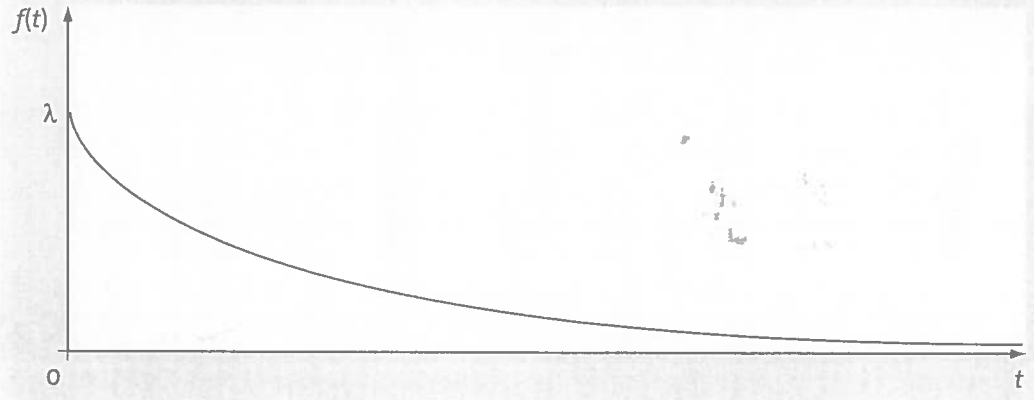
Dus $f(t) = F'(t) = \lambda e^{-\lambda t}$ voor $t \geqq 0$ en $\lambda > 0$.

Men kan eenvoudig nagaan dat de integraal van $f(t)$ de waarde 1 heeft.

$$\int_0^\infty f(t)\mathrm{d}t = \int_0^\infty \lambda e^{-\lambda t}\mathrm{d}t = -e^{-\lambda t}\Big|_0^\infty = 0 - (-1) = 1$$

Voor de verwachtingswaarde van $\underline{t}$ geldt:

Afbeelding 7.4 Negatief-exponentiële verdeling



$$E(\underline{t}) = \int_0^\infty t f(t)\mathrm{d}t = \int_0^\infty t\lambda e^{-\lambda t}\mathrm{d}t = \frac{1}{\lambda}$$

Dit resultaat sluit aan bij onze intuïtie. Als er op een centrale gemiddeld $\lambda = 10$ telefoontjes per uur binnenkomen, dan is de gemiddelde wachttijd tot het eerstvolgende telefoontje $E(\underline{t}) = \frac{1}{10}$ uur $= 6$ minuten.

Voor de variantie van de wachttijd $\underline{t}$ geldt: $\mathrm{Var}(\underline{t}) = 1/\lambda^2$.

Daarom vinden we voor de standaarddeviatie $\sigma_{\underline{t}} = 1/\lambda$ (zie appendix A).
Bij de negatief-exponentiële verdeling blijkt dus dat verwachtingswaarde en *standaarddeviatie* aan elkaar gelijk zijn.

Samengevat:
Voor een variabele $\underline{t}$ met een negatief-exponentiële verdeling geldt:

$$f(t) = \lambda e^{-\lambda t} \quad \text{voor } t \geqq 0 \text{ en } \lambda > 0$$
$$\quad = 0 \quad \text{voor } t < 0$$

$$E(\underline{t}) = \frac{1}{\lambda} \quad \text{en} \quad \mathrm{Var}(\underline{t}) = \frac{1}{\lambda^2}, \quad \text{dus} \quad \sigma_{\underline{t}} = \frac{1}{\lambda}$$

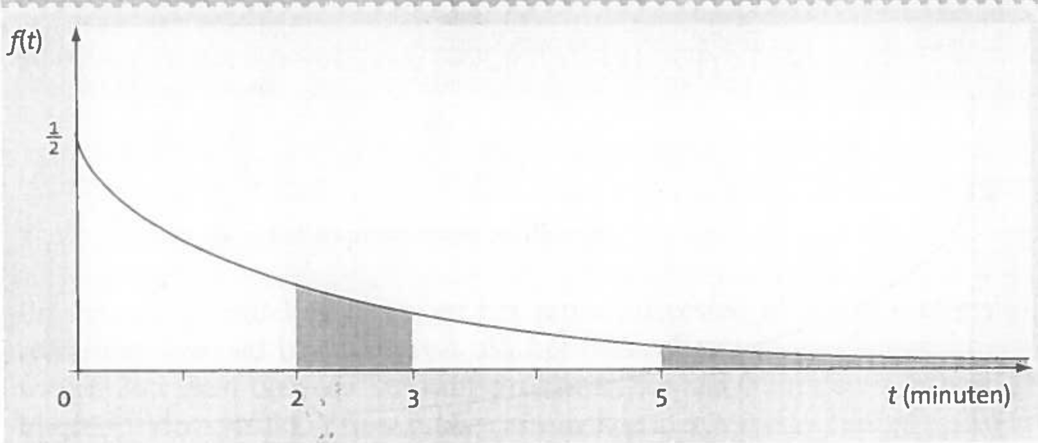
Voorbeeld 7.7 De klanten van een winkel komen volgens een poissonverdeling de winkel binnen met een gemiddelde van 30 per uur. Hoe groot is de kans dat tussen twee binnenkomsten meer dan vijf minuten verstrijken, zie afbeelding 7.5?
Gegeven is $\lambda = 30$ per uur. Omgerekend naar minuten geldt dus $\lambda = \frac{1}{2}$ per minuut.
De wachttijd is dus negatief-exponentieel verdeeld met als kansdichtheid:

$f(t) = \frac{1}{2}e^{-\frac{1}{2}t}$ voor $t \geqq 0$

Zoals dat gebruikelijk is bij continue kansvariabelen, moeten kansen door middel van integratie worden bepaald.

$$P(\underline{t} > 5) = \int_5^\infty \tfrac{1}{2} e^{-\frac{1}{2}t}\,dt = -e^{-\frac{1}{2}t}\Big|_5^\infty = 0 - (-e^{-2\frac{1}{2}}) = 0{,}0821$$

Afbeelding 7.5 Wachttijd tot de volgende klant



Hoe groot is de kans dat we langer dan twee minuten en korter dan drie minuten moeten wachten op de binnenkomst van de volgende klant?
Voor $P(2 < \underline{t} < 3)$ vinden we:

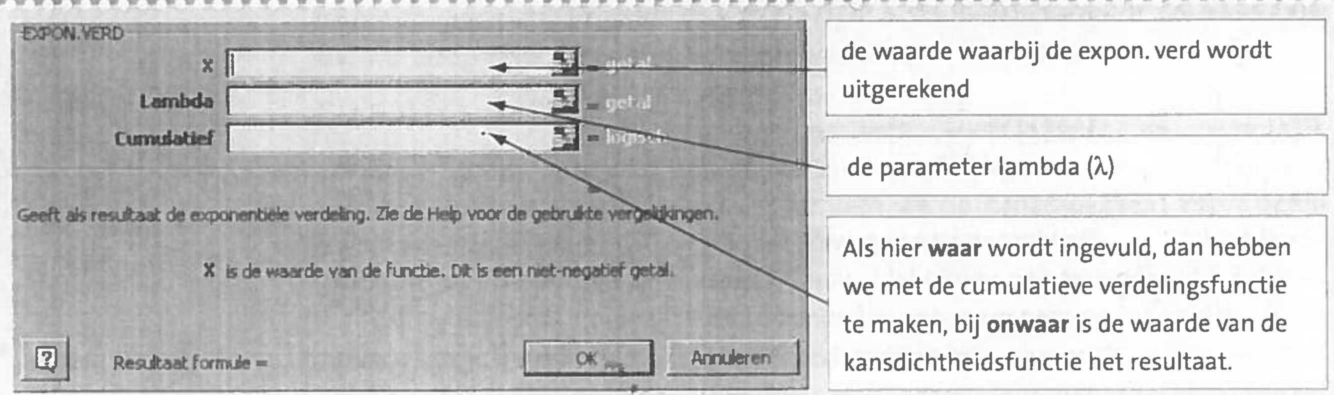
$$\int_2^3 \tfrac{1}{2} e^{-\frac{1}{2}t}\, dt = -e^{-\frac{1}{2}t}\Big|_2^3 = -e^{-\frac{3}{2}} - (-e^{-1}) = -0{,}2231 + 0{,}3679 = 0{,}1448$$

Toepassing van de poissonverdeling en de negatief-exponentiële verdeling komen we regelmatig tegen bij het bestuderen van de theorie over wachttijdproblemen. Hierbij speelt het patroon van aankomst en bediening van klanten een belangrijke rol. Met name voor de aankomsttijden wordt vaak verondersteld dat de tijd die verstrijkt tussen twee opeenvolgende klanten beschreven kan worden door de negatief-exponentiële verdeling. ◂

Gebruik Excel

De negatief-exponentiële verdeling is als 'EXPON.VERD' in Excel te vinden. Ook nu moeten drie parameters worden opgegeven, zie afbeelding 7.6.

Afbeelding 7.6 Dialoogvenster EXPON.VERD



We bekijken de gegevens van voorbeeld 7.7.
$P(\underline{t} > 5) = 1 - P(\underline{t} \leq 5)$ dus in Excel: '=1 − EXPON.VERD(5;0,5;Waar)' en we vinden: 0,0821.
Met dezelfde aanpak zijn ook voor $t = 1, 2, 3$ en 4 de kansen bepaald, zie afbeelding 7.7.

Afbeelding 7.7 Enkele kansen met Excel

	A	B	C
1		P(t<x)	P(t>x)
2	1	0,3935	0,6065
3	2	0,6321	0,3679
4	3	0,7769	0,2231
5	4	0,8647	0,1353
6	5	0,9179	0,0821

=EXPON.VERD(1; 0,5; waar)

Gebruik van een tabel
Voor het berekenen van kansen met een negatief-exponentiële verdeling kan desgewenst ook met een tabel worden gewerkt. Dat gaat als volgt.
De tabel levert bij een gegeven grenswaarde T rechteroverschrijdingskansen voor een negatief-exponentiële verdeling met parameter λ. Bereken nu eerst T maal λ. In de tabel staat naast de waarde van $\lambda \times T$ de kans vermeld dat de kansvariabele $\underline{t}$ een waarde aanneemt die groter is dan T.
Ter controle: in voorbeeld 7.7 zochten we de kans $P(\underline{t} > 5)$ voor de negatief-exponentiële verdeling met $\lambda = 0{,}5$. Dus $\lambda \times T = 0{,}5 \times 5 = 2{,}5$. Dit levert de kans 0,0821 die in voorbeeld 7.7 reeds door middel van het berekenen van een integraal was gevonden.

Een 'geheugenloos' proces
Van de negatief-exponentiële verdeling zegt men wel dat deze de kansen voor een geheugenloos proces weergeeft. Op elk tijdstip t zijn de kansen voor een daaropvolgend tijdvak onafhankelijk van eerder wel of niet geregistreerde gebeurtenissen.

Dit betekent bijvoorbeeld bij een wachttijdprobleem dat de kans dat binnen een minuut de eerste klant binnenkomt niet afhangt van het feit dat reeds geruime tijd op een klant gewacht is. Dit blijkt uit het volgende:
We definiëren T_1 en T_2 met $T_2 > T_1$ en bekijken $P(\underline{t} > T_2)$. Als het eerstvolgende 'succes' na T_2 moet vallen, dan kunnen we schrijven:

$$P(\underline{t} > T_2) = P(\underline{t} > T_2 \mid \underline{t} > T_1)\, P(\underline{t} > T_1)$$

Dus $P(\underline{t} > T_2 \mid \underline{t} > T_1) = \dfrac{P(\underline{t} > T_2)}{P(\underline{t} > T_1)} = e^{-\lambda(T_2 - T_1)}$.

De kans dat er niets gebeurt tussen T_1 en T_2 hangt dus uitsluitend af van het tijds*verschil* $T_2 - T_1$ en niet van de reeds gewachte tijd T_1.

Tabel 7.2 Rechteroverschrijdingskansen van de negatief-exponentiële verdeling

λ × T	Kans	λ × T	Kans	λ × T	Kans
0,10	0,9048	2,10	0,1255	4,10	0,0166
0,20	0,8187	2,20	0,1108	4,20	0,0150
0,30	0,7408	2,30	0,1003	4,30	0,0136
0,40	0,6703	2,40	0,0907	4,40	0,0123
0,50	0,6065	2,50	0,0821	4,50	0,0111
0,60	0,5488	2,60	0,0743	4,60	0,0101
0,70	0,4966	2,70	0,0672	4,70	0,0091
0,80	0,4493	2,80	0,0608	4,80	0,0082
0,90	0,4066	2,90	0,0550	4,90	0,0074
1,00	0,3679	3,00	0,0498	5,00	0,0067
1,10	0,3329	3,10	0,0450	5,20	0,0055
1,20	0,3012	3,20	0,0408	5,40	0,0045
1,30	0,2725	3,30	0,0369	5,60	0,0037
1,40	0,2466	3,40	0,0334	5,80	0,0030
1,50	0,2231	3,50	0,0302	6,00	0,0025
1,60	0,2019	3,60	0,0273	6,20	0,0020
1,70	0,1827	3,70	0,0247	6,40	0,0017
1,80	0,1653	3,80	0,0224	6,60	0,0014
1,90	0,1496	3,90	0,0202	6,80	0,0011
2,00	0,1353	4,00	0,0183	7,00	0,0009

7^+.2 De gammaverdeling

Bij de opzet van de negatief-exponentiële verdeling was het uitgangspunt dat de wachttijd $\underline{t}$ werd bestudeerd tot het eerstvolgende succes. Als het aantal successen $\underline{k}$ voor een gefixeerd tijdsinterval t een poissonverdeling volgt met parameter λt, dan is de kansdichtheid van $\underline{t}$ af te leiden.
Een wat algemener geval ontmoeten we als we de vraag stellen: hoe is de kansverdeling van de wachttijd t tot het n-de succes?
Deze kansvariabele volgt een zogenaamde *gamma-verdeling*. Voor de afleiding van de gammaverdeling volgen we een redenering die overeenkomt met de afleiding van de negatief-exponentiële verdeling. Noem $\underline{t}$ de wachttijd tot het n-de succes. Voor een willekeurig gekozen tijdsduur t geldt dat:

$$P(\underline{t} > t) = P(\text{hoogstens } n-1 \text{ successen in het interval } [0, t])$$

$$= \sum_{k=0}^{n-1} (\lambda t)^k \frac{e^{-\lambda t}}{k!}$$

Dus $F(t) = P(\underline{t} \leqq t) = 1 - \sum_{k=0}^{n-1} (\lambda t)^k \frac{e^{-\lambda t}}{k!}$

Om de kansdichtheid te bepalen, nemen we de afgeleide naar t. Dit levert na het differentiëren van de sommatie:

$$f(t) = \lambda e^{-\lambda t} - \sum_{k=1}^{n-1} \left\{ \lambda(\lambda t)^{k-1} \frac{e^{-\lambda t}}{(k-1)!} - \lambda \frac{(\lambda t)^k}{k!} \right\}$$

$$= \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}$$

Eigenlijk kunnen we de gamma-verdeling beschouwen als de optelling van n onderling onafhankelijke trekkingen uit de negatief-exponentiële verdeling. Daarom geldt voor de gamma-verdeelde variabele $\underline{t}$:

$$E(\underline{t}) = \frac{n}{\lambda} \text{ en } \sigma_{\underline{t}} = \frac{1}{\lambda} \sqrt{n}$$

De gamma-verdeling wordt onder andere gebruikt bij sommige voorraadproblemen. We gaan hier niet verder op in.

Samenvatting

1 Een kansvariabele $\underline{k}$ die een poissonverdeling volgt, heeft het karakter van successen of registraties tellen. De laagste waarde die $\underline{k}$ kan aannemen is nul, maar er is geen bovengrens zoals bij de binomiale verdeling.
2 De kansen voor een variabele met een poissonverdeling worden berekend met de poissonformule. Deze formule kent slechts één parameter, namelijk μ. Deze μ is zowel gelijk aan de verwachtingswaarde van de verdeling als aan de variantie.
3 Kansen voor de poissonverdeling worden berekend met:
   a de poissonformule;
   b de tabel met de afzonderlijke poissonkansen;
   c de tabel met de cumulatieve poissonkansen;
   d de benaderingsmethode met de normale verdeling indien μ groter dan 10 is. Deze normale verdeling heeft verwachtingswaarde μ en standaarddeviatie $\sigma = \sqrt{\mu}$. Bij toepassing van de normale verdeling moet weer rekening worden gehouden met de continuïteitscorrectie.
4 De poissonverdeling kan worden toegepast als een benadering van de binomiale verdeling indien n zeer groot is en π, de succeskans, zeer klein. Kansen voor dergelijke binomiale verdelingen kunnen (als benadering) worden berekend met de poissontabel.
5 De poissonverdeling geeft veelal het aantal successen weer dat wordt waargenomen in een tijdvak van een bepaalde lengte: de waarnemingsperiode. Stel λ is de verwachtingswaarde van de verdeling per eenheid van tijd. Het aantal successen in een periode van t tijdseenheden volgt dan een poissonverdeling met verwachtingswaarde $\mu = \lambda \times t$.