

9 Continue kansverdelingen

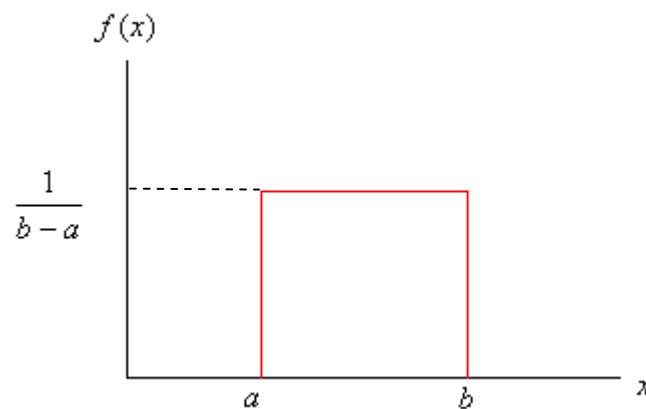
9.1 Uniforme kansverdeling

De eerste continue kansverdeling die hier besproken zal worden is de uniforme kansverdeling. Vergeleken andere kansverdelingen is de praktische toepassing van de uniforme verdeling beperkt.

De uniforme verdeling wordt beschreven door de functie

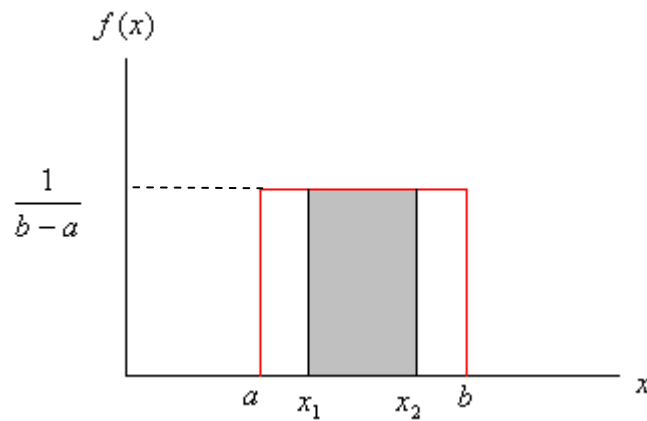
$$f(x) = \frac{1}{b-a} \text{ waar } a \leq x \leq b$$

De uniforme verdeling wordt ook wel rechthoekige verdeling genoemd vanwege de rechthoekige vorm van de kansdichtheidsfunctie $f(x)$ zoals in de onderstaande figuur is te zien.



Figuur 9^a. Uniforme verdeling

Om de kans van elk willekeurig interval te berekenen, kijkt men naar het gebied onder de curve. Wil men bijvoorbeeld de kans berekenen dat X tussen x_1 en x_2 valt, dan kijkt men naar het gebied met als basis $x_2 - x_1$ en als hoogte $1/(b-a)$. De onderstaande figuur geeft het gebied aan dat men wenst te berekenen.



Figuur 9^b. Uniforme verdeling voorbeeld

Kortom,

$$P(x_1 < X < x_2) = \text{basis} \times \text{hoogte} = (x_2 - x_1) \times \frac{1}{b-a}$$

Voorbeeld 9^a

De gemiddelde hoeveelheid bier die er op een drukke avond doorheen gaat in een populaire kroeg is uniform verdeeld met een minimum van 200 liter (= a) en een maximum van 500 liter (= b). De kans dat er op een avond tussen de 250 (= x_1) en 300 (= x_2) liter bier doorheen gaat wordt als volgt berekend:

$$P(250 \leq X \leq 300) = (300 - 250) \times \left(\frac{1}{300} \right) = 0,1667$$

9.2 Normale verdeling

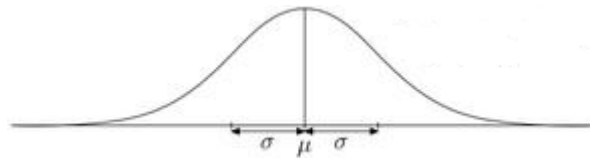
De normale verdeling is de belangrijkste van de kans verdelingen door de cruciale rol die deze vervult in het universum van de statistiek.

De kansdichtheidsfunctie van een normaal verdeelde variabele is

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

Hier is $e = 2,71828$ en $\pi = 3,14159$

De onderstaande figuur geeft een normale verdeling weer. Wat opvalt is dat de curve symmetrisch is door het midden (gemiddelde μ) en de verdeelde variabele ligt tussen $-\infty$ en ∞ .



Figuur 9^c. Symmetrische normale verdeling

Bij de normale verdeling spelen twee parameters een grote rol, het gemiddelde μ en de standaarddeviatie σ . Als μ groter wordt, verschuift de gehele curve naar rechts en als μ kleiner wordt, verschuift de curve naar links.

Bij de standaarddeviatie gebeurt iets anders bij het vergroten en verkleiner. Als σ groter wordt, naar de breedte van de curve toe en als σ kleiner wordt, zal de breedte van de curve afnemen.

9.3 Het berekenen van normale kansverdelingen

Om de kans te berekenen dat een normaal verdeelde variabele in een bepaald interval valt, dient men het gebied in het interval onder de curve te berekenen. Helaas is de formule hiervoor niet zo eenvoudig als voor de uniforme verdeling. Bij deze berekening is een tabel van toepassing waar de normaal verdeelde variabele in gestandaardiseerd is. Standaardiseren gebeurt door middel van door het gemiddelde μ van de normaal verdeelde variabele X af te trekken en deze uitkomst te delen door de standaarddeviatie σ . De noemt men de standaard normaal verdeelde variabele Z . In formulevorm wordt dit

$$Z = \frac{X - \mu}{\sigma}$$

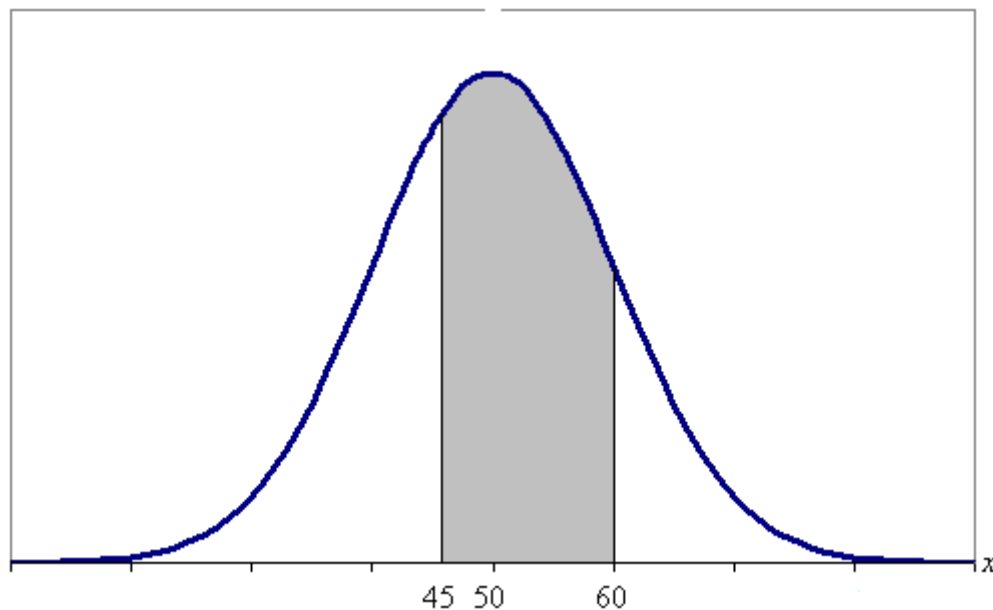
Aan de hand van een voorbeeld wordt hieronder uitgelegd hoe de tabel in Appendix A1 te gebruiken. (dit voorbeeld is erg belangrijk en bevat ook uitleg!)

Voorbeeld 9^b

De tijd die Eva er over doet om in de ochtendspits naar haar werk te komen is normaal verdeeld met een gemiddelde van 50 minuten en een standaarddeviatie van 10 minuten. Men wil weten wat de kans is dat ze er tussen de 45 en 60 minuten over doet. Kortom,

$$P(45 < X < 60)$$

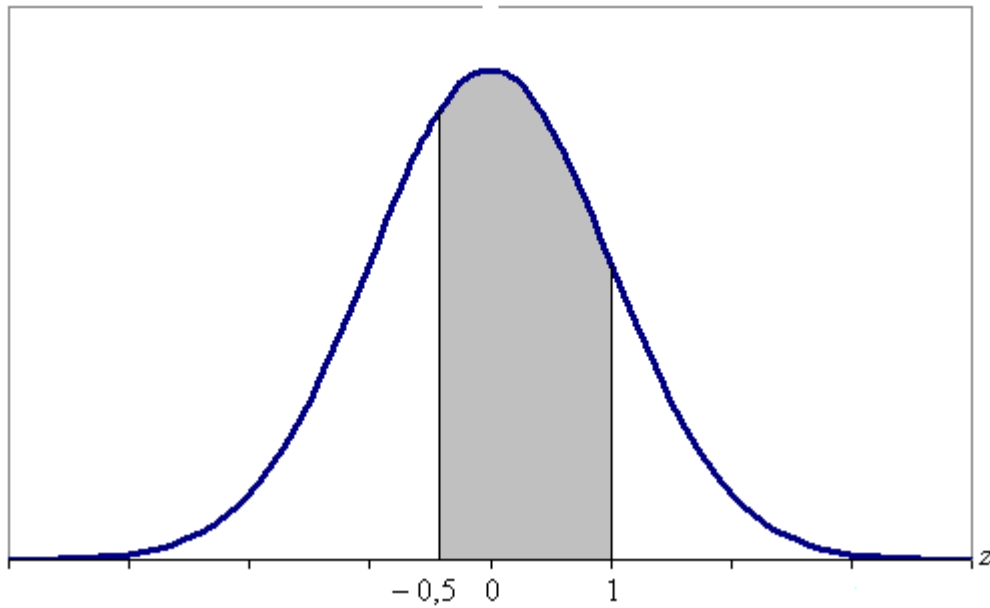
In de onderstaande figuur is het gebied aangegeven dat man wil berekenen.



Eerst zal X gestandaardiseerd worden. Hierbij moet wel rekenen gehouden worden dat als 50 gestandaardiseerd wordt, dat dan hetzelfde gedaan dient te worden met 45 en 60.

$$P(45 < X < 60) = P\left(\frac{45 - 50}{10} < \frac{X - \mu}{\sigma} < \frac{60 - 50}{10}\right) = P(-0,5 < Z < 1)$$

In de onderstaande figuur is te zien wat standaardiseren precies doet. Wat opvalt is dat de variabele X is veranderd in Z , 45 is veranderd in $-0,5$ en 60 is veranderd in 1 . Echter, het gebied van belang is niet veranderd. Met andere woorden, de kans die berekend dient te worden, $P(45 < X < 60)$, is identiek aan $P(-0,5 < Z < 1)$.



De waarden van Z geven de locaties van de waarden van X aan. De waarde $Z = -0,5$ correspondeert met de waarde $X = 45$ en staat synoniem voor een halve standaarddeviatie onder het gemiddelde. De waarde $Z = 1$ correspondeert met de waarde $X = 60$ en staat synoniem voor 1 standaarddeviatie boven het gemiddelde. Het gemiddelde van $Z (= 0)$ correspondeert met het gemiddelde van X . Als het gemiddelde en de standaarddeviatie van een normaal verdeelde variabele bekend zijn, kan de kansbewering van X in een kansbewering van Z veranderd worden. Hiervoor wordt de tabel gebruikt uit Appendix A1.

Deze tabel geeft de kans dat een standaard normaal verdeelde variabele tussen 0 en waarden van z valt. Bijvoorbeeld de kans $P(0 < Z < 2,00)$ is in de tabel te vinden door aan de linkerkant te zoeken naar 2.0 en aan de bovenkant te zoeken naar 0.00. De waarde die hierbij hoort is 0,4772. De kans $P(0 < Z < 2,01)$ is te vinden in dezelfde rij, maar aan de bovenkant bij 0.01. Die kans is dus 0,4778. Doordat de normale curve symmetrisch is en gespiegeld kan worden bij het gemiddelde en het totale gebied onder de curve gelijk is aan 1, kan het volgende geconcludeerd worden:

$$P(Z > 0) = P(Z < 0) = 0,5$$

De grootste waarde in de tabel is 3,09 en $P(0 < Z < 3,09) = 0,4990$. Dit houdt in dat

$$P(Z > 3,09) = P(Z > 0) - P(0 < Z < 3,09) = 0,5 - 0,4990 = 0,01$$

Terug bij voorbeeld 8^a bestaat de kans die men zoekt dus eigenlijk uit de som van 2 kansen:

$$P(-0,5 < Z < 1) = P(-0,5 < Z < 0) + P(0 < Z < 1)$$

De kans aan de rechterkant van de som is makkelijk in de tabel te vinden: $P(0 < Z < 1,00) = 0,3413$. De kans aan de linkerkant $P(-0,5 < Z < 0)$ is gelijk aan $P(0 < Z < 0,5)$ en de corresponderende kans uit de tabel is 0,1915. Dus

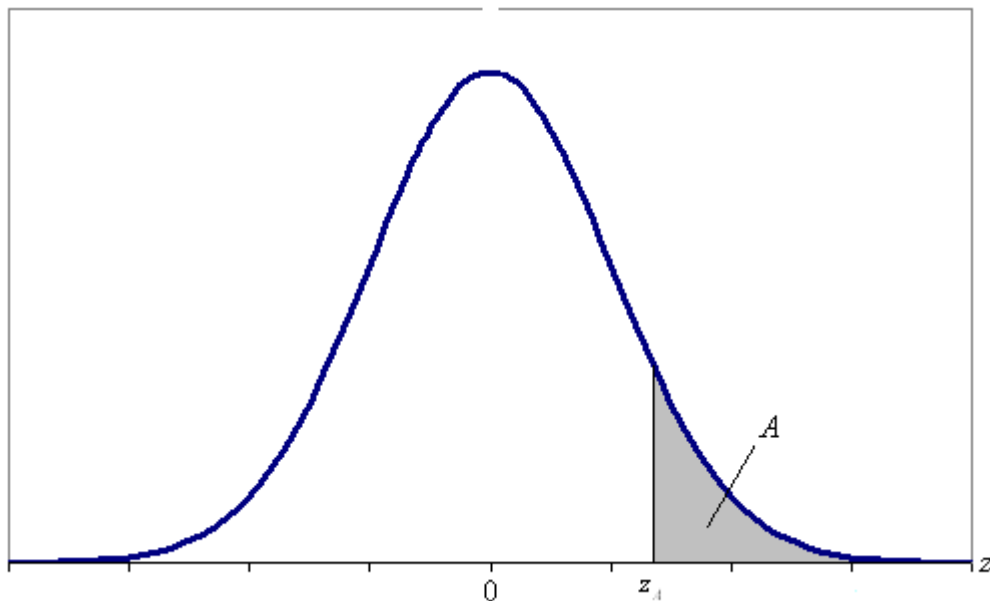
$$P(-0,5 < Z < 1) = 0,1915 + 0,3413 = 0,5328$$

9.4 Waarden van z vinden

In enkele gevallen is het nodig om de waarde van z vast te stellen gegeven een bepaalde kans. Hier wordt de notatie z_A gebruikt om de waarde van z aan te geven op zo'n wijze dat het gebied aan de rechterkant hiervan onder de standaard normale curve A is. Oftewel, z_A is een waarde van een standaard normaal verdeelde variabele op zo'n wijze dat

$$P(Z > z_A) = A$$

De onderstaande figuur geeft dit aan.



Figuur 9^d. $P(Z > z_A) = A$

Om z_A voor elke willekeurige waarde van A vinden dient de tabel van de normale kansverdeling omgekeerd gebruikt te worden. Om dit te kunnen doen moet men een kans specificeren en de corresponderende z -waarde vaststellen. Bijvoorbeeld de waarde $z_{0,025}$. Door het format van de tabel (standaard normaal) moet men eerst het gebied tussen 0 en $z_{0,025}$ vaststellen. Dit is gelijk aan $0,5 - 0,025 = 0,4750$. (aangegeven in 4 decimalen, daar de waarden in de tabel tevens in 4 decimalen weergegeven zijn). Ze z -waarde uit de tabel corresponderend met een kans van 0,4750 is 1,96. Met andere woorden, $z_{0,025} = 1,96$ en dat houdt in dat $P(Z > 1,96) = 0,025$.

9.5 Exponentiële verdeling

Een andere belangrijke continue verdeling is de exponentiële verdeling. Een willekeurige variabele X is exponentieel verdeeld als de kansdichtheidsfunctie wordt gegeven door

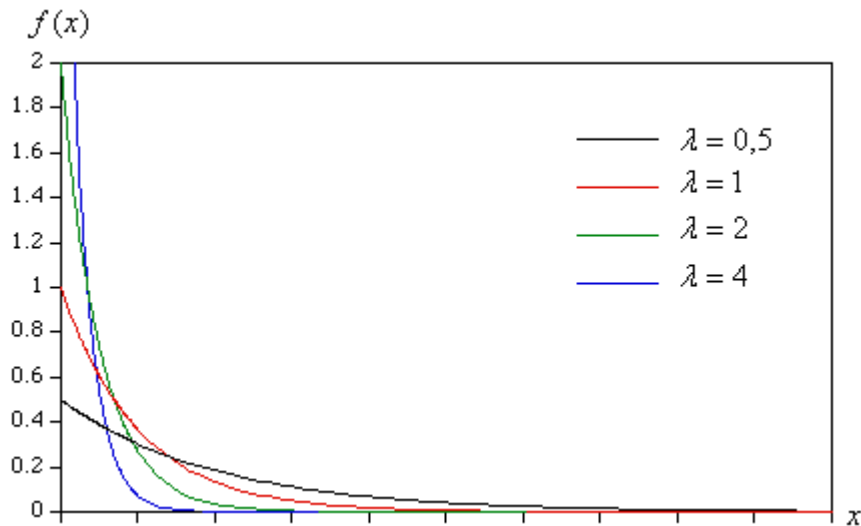
$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

met $e = 2,71828\dots$ en λ is de parameter van de verdeling.

Statistici hebben aangetoond dat het gemiddelde en de standaarddeviatie van een exponentiële willekeurige variabele gelijk zijn aan elkaar:

$$\mu = \sigma = \frac{1}{\lambda}$$

De normale verdeling, die besproken is in de paragrafen hierboven, is een twee-parameter verdeling. Zodra de twee parameters μ en σ bekend zijn, kan de verdeling volledig uitgewerkt worden. De exponentiële verdeling is echter een verdeling waar maar een parameter een rol speelt. Zodra de parameter λ bekend is, kan de verdeling uitgewerkt worden. De onderstaande figuur geeft 4 exponentiële verdelingen weer die corresponderen met 4 verschillende waarden van de parameter λ . Wat hierbij opvalt is dat voor elke willekeurige kansdichtheidsfunctie $f(x)$ geldt $f(0) = \lambda$ en $f(x)$ nadert 0 als x oneindigheid nadert.



Figuur 9^e. Exponentiële verdelingen (bron: content.answers.com)

De kansdichtheidsfunctie van een exponentiële verdeling is makkelijker om mee te werken dan die van de normale verdeling en daardoor kan men formules ontwikkelen om de kans van elke willekeurig bereik van waarden te berekenen. Men kan daardoor de volgende uitspraken doen over kansverdelingen:

Als X een exponentieel verdeelde variabele is, dan geldt

$$P(X > x) = e^{-\lambda x}$$

$$P(X < x) = 1 - e^{-\lambda x}$$

$$P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1) = e^{-\lambda x_1} - e^{-\lambda x_2}$$

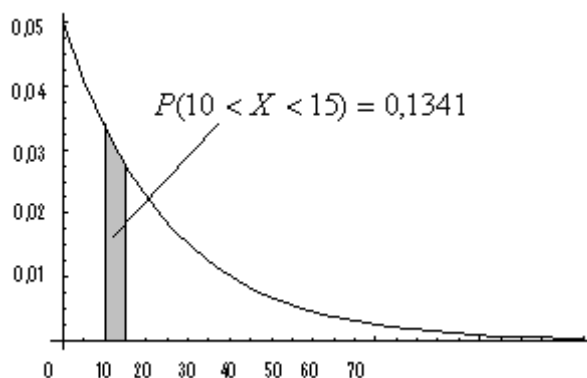
Voorbeeld 9^c

De standby tijd van een mobiele telefoon (gemeten in uren) is exponentieel verdeeld met $\lambda = 0,05$.

Als men het gemiddelde en de standaarddeviatie wil berekenen gaat dat als volgt:

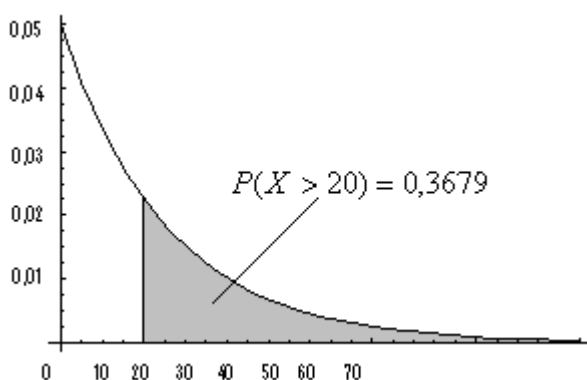
$$\mu = \sigma = \frac{1}{\lambda} = \frac{1}{0,05} = 20 \text{ uur}$$

Men wil achter de kans komen dat een telefoon tussen de 10 en 15 uur kan zonder opgeladen te hoeven worden. In de onderstaande figuur staat aangegeven welk gebied men wil berekenen.



$$\begin{aligned} P(10 \leq X \leq 15) &= e^{-0,05(10)} - e^{-0,05(15)} \\ &= e^{-0,5} - e^{-0,75} \\ &= 0,6065 - 0,4724 \\ &= 0,1341 \end{aligned}$$

Als men wil berekenen wat de kans is dat een mobiele telefoon meer dan 20 uur meegaat zonder op laden gaat dat als volgt. In de onderstaande figuur staat aangegeven welk gebied men wil berekenen.



$$P(X > 20) = e^{-0,05(20)} = e^{-1} = 0,3679$$