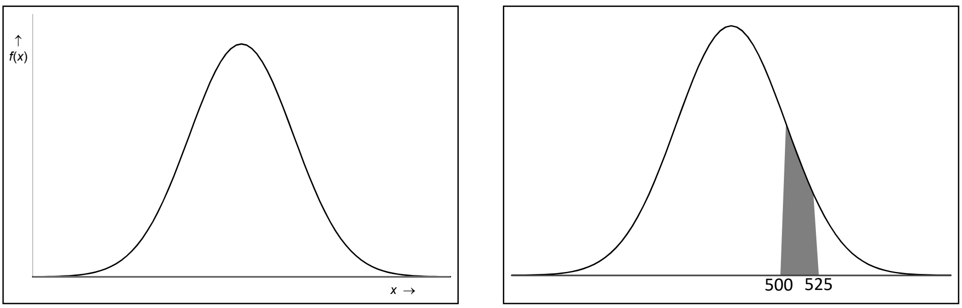
Bijlage 2: De Normale verdeling

De Normale verdeling is een *continue* kansverdeling, in tegenstelling tot de Binomiale verdeling die *discreet* is. Bij een continue kansvariabele is het aantal mogelijke uitkomsten zo groot, dat de kans op elke afzonderlijke uitkomst miniem is. De kans dat bijvoorbeeld een tornado in de Verenigde Staten $500.000.000 aan schade aanricht is niet te onderscheiden van de kans op een schade van $500.000.001 of $499.999.999. Omdat er zoveel uitkomsten mogelijk zijn wordt bij continue kansverdelingen gewerkt met *kansdichtheden* of *dichtheidsfuncties*, waarbij kansen verkregen kunnen worden door middel van oppervlakten.

Op die manier kan de kans dat de schade X van de tornado tussen \$500 en \$525 miljoen ligt, dus P[500 miljoen ≤ X ≤ 525 miljoen], berekend worden door de oppervlakte onder de dichtheidsfunctie te berekenen tussen x = 500 miljoen en x= 525 miljoen.

Laten we bijvoorbeeld aannemen dat de tornadoschade *Normaal* verdeeld is. De aanduiding 'Normaal' betekent niet normaal in de gewone betekenis dat een tornado een normale zaak is, maar verwijst in de statistiek naar een verdeling waarvoor geldt dat kleine afwijkingen ten opzichte van het gemiddelde een grote kans hebben om voor te komen maar dat grotere afwijkingen een steeds kleinere kans hebben. De dichtheids-functie van een Normale verdeling is klokvormig, zoals hieronder.

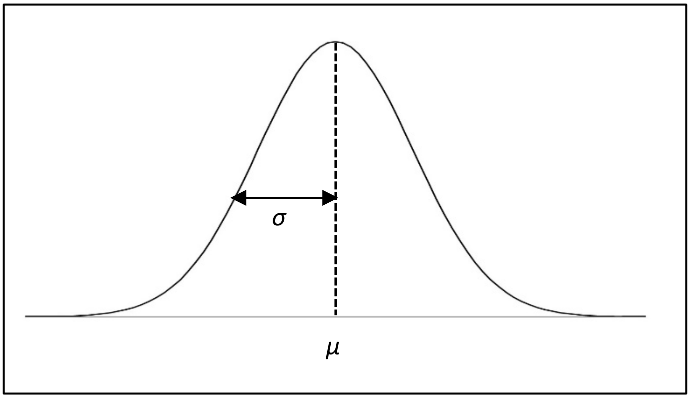


Figuur B2.1 De kansdichtheid bij de Normale verdeling.

De kans dat de tornadoschade tussen \$500 en \$525 miljoen ligt is dan gelijk aan het lichtgrijze oppervlak in rechter grafiek. De oppervlakte onder deze kromme is te berekenen via integreren van de dichtheidsfunctie van de normale verdeling:

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

waarbij μ de verwachtingswaarde is en σ de standaardafwijking, een maat voor de spreiding van de uitkomsten, zie onderstaande Figuur B2.2. Uit de grafiek van een normale verdeling kun je een schatting van het gemiddelde μ en de standaardafwijking σ aflezen.



Figuur B2.2 De Normale verdeling met parameters μ en σ.

Omdat de normale verdeling zo veel voorkomt kunnen de kansen ook via de grafische rekenmachine direct berekend worden, als je de grenzen van het gebied (in het voorbeeld 500 en 525) opgeeft en aangeeft wat het gemiddelde en de standaardafwijking zijn (bijvoorbeeld 400 en 100).

Voor de normale verdeling bestaat een aantal vuistregels:

68% van de waarnemingen liggen tussen $\mu - \sigma$ en $\mu + \sigma$
95% van de waarnemingen liggen tussen $\mu - 2\sigma$ en $\mu + 2\sigma$

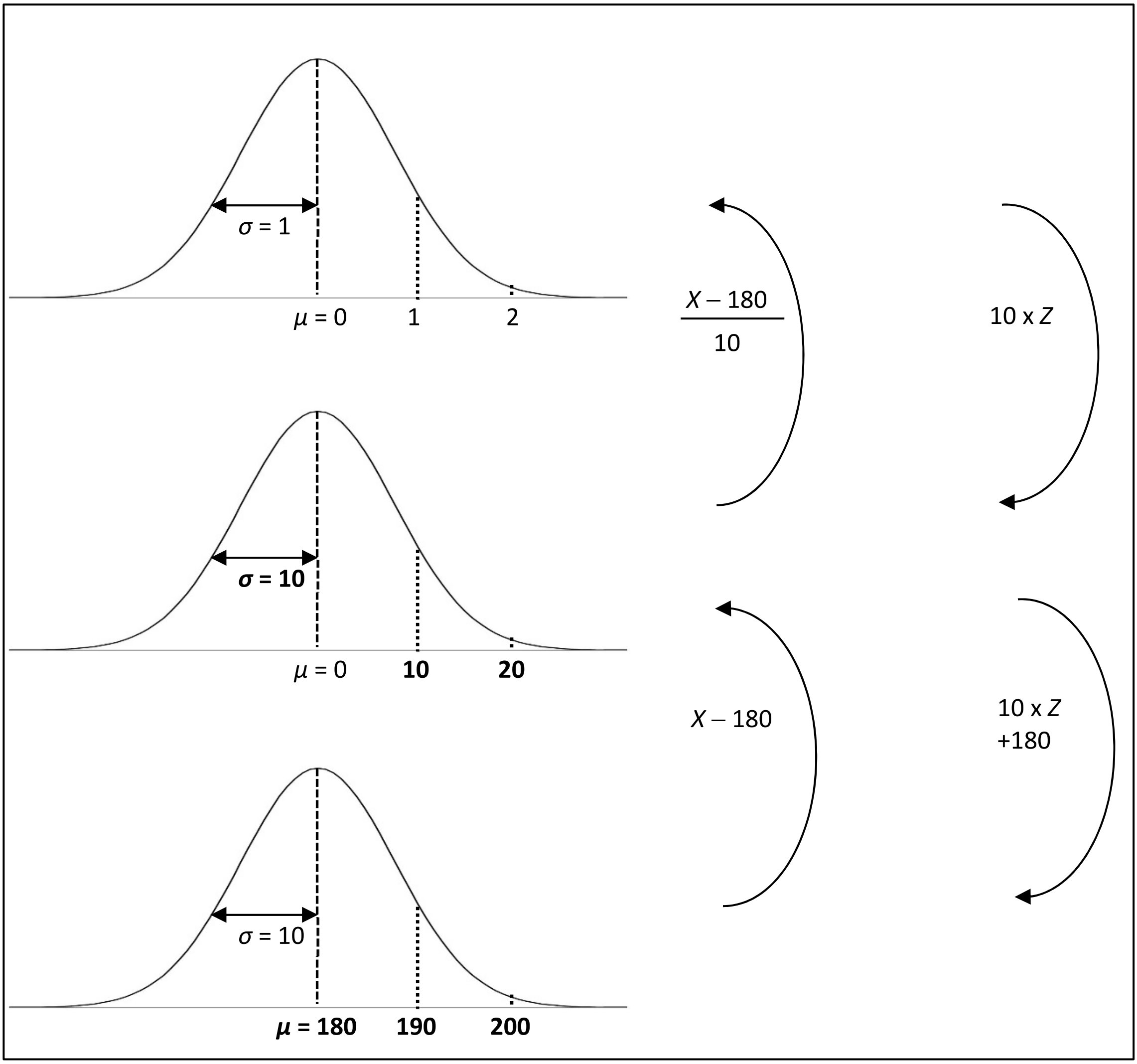
Bij de benadering van een binomiale verdeling door een normale verdeling gebruiken we de volgende formules om het gemiddelde μ en de standaardafwijking σ te berekenen:

$\mu = n \cdot p$ en $\sigma = \sqrt{n \cdot p \cdot (1-p)}$

De standaardnormale verdeling

Een Normale verdeling wordt dus bepaald door twee parameters, de verwachtingswaarde μ en de standaardafwijking σ en als deze bekend zijn dan kun je eenvoudig allerlei kansen berekenen. Bijvoorbeeld: als X Normaal verdeeld is met $\mu = 180$ en $\sigma = 10$, dan kun je met je grafische rekenmachine of Excel de kansen $P(X \geq 200)$ en $P(190 \leq X \leq 200)$ uitrekenen.

Soms is het makkelijker, zoals in de analyse van Hoofdstuk 7, om te werken met een standaardnormale verdeling; dit is een Normale verdeling met $\mu = 0$ en $\sigma = 1$. De reden hiervoor is dat voor een standaardnormaal verdeelde kansvariabele Z een vaste tabel met kansen gebruikt kan worden (in Hoofdstuk 7 is dit de kans $P[Z \leq 2{,}576]=0{,}995$). Met herschalingen van Z kan vervolgens de gevraagde kans worden berekend, zoals we hieronder illustreren voor $P[190 \leq X \leq 200]$.



Figuur B2.3 Het gebruik van schaling om naar de standaardnormale verdeling te komen.

Uit dit diagram blijkt dat P[190 ≤ X ≤ 200] gelijk is aan P[1 ≤ Z ≤ 2], reken dit maar na met je grafische rekenmachine of Excel.

Ook voor andere waarden van μ en σ geldt het schalingsprincipe van Figuur B2.3 zodat $Z = (X – \mu)/\sigma$ altijd standaardnormaal is verdeeld. Andersom, met een standaardnormaal verdeelde kansvariabele Z leidt de bewerking $X = \mu + \sigma Z$ tot een Normaal verdeelde X met verwachting μ en standaardafwijking σ.