FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

STATISTIEK I(A+B)

Prof. dr. David Vyncke

Bachelor Economische Wetenschappen
Bachelor Toegepaste Economische Wetenschappen
Bachelor Toegepaste Economische Wetenschappen:
Handelsingenieur
Academiejaar 2019–2020

UNIVERSITEIT
GENT

Voorwoord

De diverse economische problemen waarmee beleidsmakers in de publieke en private sector geconfronteerd worden, bevatten vaak een belangrijke post onzekerheid. Economische grootheden zijn immers niet altijd makkelijk meetbaar of (gedeeltelijk) onvoorspelbaar. Anderzijds moeten beslissingen of aanbevelingen in de huidige algemeen economische of bedrijfseconomische context steeds vaker gebaseerd zijn op feiten, het zogeheten Evidence Based Decision Making. Maar in hoeverre kunnen vaststellingen in een steekproef met een beperkte hoeveelheid gegevens veralgemeend worden naar feiten voor de hele populatie, een productgroep, een bepaald type consumenten?

Gegeven de onzekerheid die met dergelijke beslissingen gepaard gaat, is een grondige kennis van de statistiek onontbeerlijk voor een (toegepast) economist of handelsingenieur. Statistiek is bovendien een taal die in alle disciplines gebruikt wordt om wetenschappelijke resultaten te communiceren. Deze cursustekst heeft de ambitie om studenten in de academische opleidingen (Toegepaste) Economische Wetenschappen en Handelsingenieur vertrouwd te maken met de elementaire statistische begrippen en de statistische methoden die gebruikt worden ter ondersteuning van beleidsbeslissingen, met een kritische kijk op de mogelijkheden en de beperkingen van de verschillende methoden. Dit moet de studenten in staat stellen op een correcte manier relevante informatie en data te verzamelen, deze met de gepaste technieken kritisch te analyseren en samen te vatten, indien nodig de juiste toets te selecteren en uit te voeren en een statistisch gefundeerde en genuanceerde conclusie te formuleren.

Het vak Statistiek I is opgesplitst in twee delen. In een eerste deel, Statistiek I(A), komen de beschrijvende statistiek en kansrekening aan bod. Daarnaast wordt ook aandacht besteed aan eigenschappen van kansvariabelen en worden verscheidene bekende verdelingen bestudeerd. Het tweede deel, Statistiek I(B), behandelt voornamelijk de verklarende statistiek.

Inhoudsopgave

Hoofdstuk 7

Continue verdelingen

In dit hoofdstuk bestuderen we enkele continue kansdichtheden meer in detail.

7.1 Uniforme verdeling

De eenvoudigste continue verdeling is de **uniforme verdeling**. De kansvariabele X is uniform verdeeld over het interval $[a, b]$ als de kansdichtheid constant is over $[a, b]$, nl.

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elders} \end{cases} \tag{7.1}$$

Voor de cumulatieve verdelingsfunctie van $X \sim$Unif(a, b) vinden we

$$F_X(x) = \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a}, \quad x \in [a, b] \tag{7.2}$$

en dus

$$F(a) = \frac{a-a}{b-a} = 0, \quad F(b) = \frac{b-a}{b-a} = 1$$

De verdeling met $a = 0$ en $b = 1$ noemen we de **standaard uniforme verdeling**. Figuur 7.1 toont de kansdichtheid en verdelingsfunctie van deze verdeling.

Voor de verwachtingswaarde en de variantie vinden we

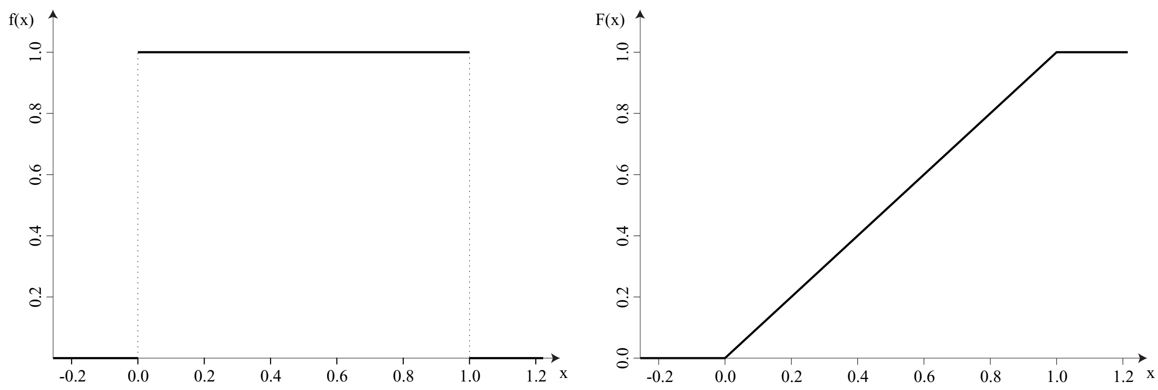
$$\mathrm{E}(X) = \frac{a+b}{2} \tag{7.3}$$

en

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12} \tag{7.4}$$

Dit volgt uit

$$\mathrm{E}(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}$$



Figuur 7.1: Kansdichtheid en verdelingsfunctie van de standaard uniforme verdeling

en

$$\begin{aligned}
\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 &= \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{1}{b-a}\frac{b^3-a^3}{3} - \frac{(a+b)^2}{4} = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} \\
&= \frac{4a^2+4ab+4b^2}{12} - \frac{3a^2+6ab+3b^2}{12} = \frac{a^2-2ab+b^2}{12} \\
&= \frac{(a-b)^2}{12} = \frac{(b-a)^2}{12}
\end{aligned}$$

Voorbeeld 7.1 (Vroegmarkt)
Als de prijs X van een bepaald product op de vroegmarkt uniform verdeeld is tussen 3 en 7 euro, wat is dan de kans dat je minstens 5 euro betaalt? Uit

$$\mathrm{P}(X \geq 5) = 1 - \mathrm{P}(X < 5) = 1 - \mathrm{P}(X \leq 5) = 1 - \frac{5-3}{7-3} = 1 - \frac{1}{2}$$

volgt dat deze kans precies 50% bedraagt.

7.2 Exponentiële verdeling

Een andere vaak voorkomende verdeling is de exponentiële verdeling. De kansvariabele X heeft een **exponentiële verdeling** met parameter $\lambda > 0$ als

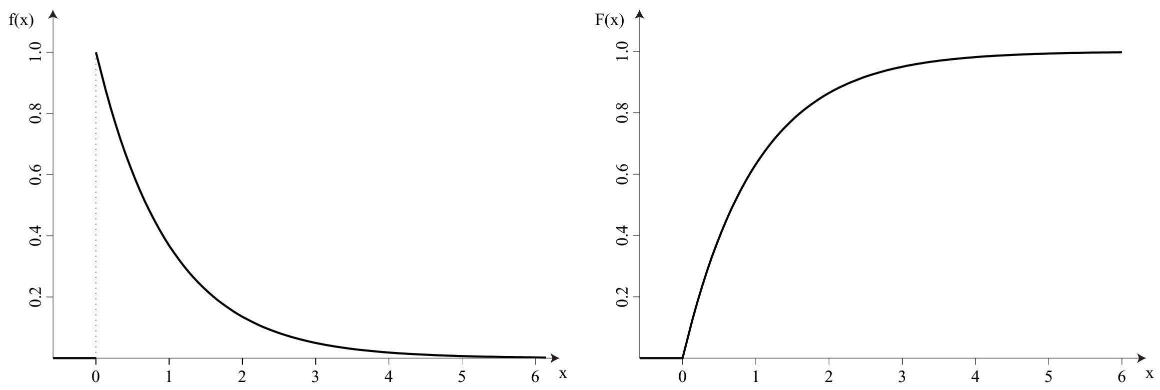
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{elders} \end{cases} \tag{7.5}$$

Voor de cumulatieve verdelingsfunctie van $X \sim \mathrm{Exp}(\lambda)$ vinden we

$$F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = \int_{-\lambda x}^0 e^z dz = 1 - e^{-\lambda x}, \quad x \geq 0 \tag{7.6}$$

en $F_X(x) = 0$ voor $x < 0$. Daaruit volgt dat

$$F(0) = 1 - 1 = 0, \quad \lim_{x \to +\infty} F(x) = 1 - 0 = 1$$



Figuur 7.2: Kansdichtheid en verdelingsfunctie van een exponentieel verdeelde variabele

Figuur 7.2 geeft de kansdichtheid en verdelingsfunctie weer van een exponentieel verdeelde kansvariabele met $\lambda = 1$.

Als X exponentieel verdeeld is met parameter λ, dan geldt

$$\mathrm{E}(X) = \frac{1}{\lambda} \tag{7.7}$$

en

$$\mathrm{Var}(X) = \frac{1}{\lambda^2} \tag{7.8}$$

De berekening van deze formules is een eenvoudige toepassing van de integraalrekening en wordt dan ook als een oefening gelaten. Inverteren van vergelijking (7.6) geeft ons de kwantielen

$$Q_X(p) = -\frac{1}{\lambda}\ln(1-p) \tag{7.9}$$

Voorbeeld 7.2 (Wachtrij)
In een postkantoor komt gemiddeld om de 6 minuten een klant over de vloer. Als we aannemen dat de tijd X tussen twee aankomsten exponentieel verdeeld is met parameter λ en er is net een klant binnengekomen, hoe groot is dan de kans dat de volgende klant over minder dan 3 minuten binnenkomt? We vinden

$$\lambda = \frac{1}{\mathrm{E}(X)} = \frac{1}{6}$$

en dus

$$\mathrm{P}(X \leq 3) = F(3) = 1 - e^{-\frac{1}{6}3} \approx 0.393$$

De tijd waarbinnen de volgende klant met 95% kans zal binnenkomen, bedraagt dan weer $Q_X(0.95) = -6\ln(1-0.95) \approx 18$ minuten.

Eigenschap 7.3 (Verband met Poisson verdeling)
Beschouwen we een kansvariabele $X \sim \text{Poisson}(\lambda)$. De intensiteit λ stelt daarbij het gemiddeld aantal gebeurtenissen per tijdseenheid voor. De tijd T tussen twee gebeurtenissen is dan exponentieel verdeeld met dezelfde parameter λ. We kunnen dit inzien als volgt: de kans dat de tijd T tussen twee gebeurtenissen groter is dan t is dezelfde als de kans dat in

een interval van lengte t zich geen gebeurtenissen voordoen. Aangezien het gemiddeld aantal gebeurtenissen per tijdseenheid gelijk is aan λ, zullen er zich gemiddeld λt gebeurtenissen voordoen in een interval van lengte t. Het aantal gebeurtenissen X_t in een interval van lengte t is bijgevolg Poisson(λt) verdeeld. Er geldt dus

$$F_T(t) = 1 - \mathrm{P}(T > t) = 1 - \mathrm{P}(X_t = 0) = 1 - \frac{e^{-\lambda t}(\lambda t)^0}{0!} = 1 - e^{-\lambda t}$$

waarin we onmiddellijk de exponentiële verdelingsfunctie herkennen.

Voorbeeld 7.4 (Wachtrij)
In een postkantoor komen gemiddeld 10 klanten per uur over de vloer. Als we aannemen dat het aantal klanten per uur Poisson verdeeld is en er is net een klant binnengekomen, hoe groot is dan de kans dat de volgende klant over minder dan 3 minuten (0.05 uur) binnenkomt? We vinden

$$\mathrm{P}(T \leq 0.05) = 1 - e^{-10\cdot 0.05} \approx 0.393,$$

precies zoals in voorbeeld 7.2.

Een andere belangrijke eigenschap van de exponentiële verdeling is de geheugenloosheid.

Eigenschap 7.5 (Geheugenloosheid)
Indien X exponentieel verdeeld is, dan geldt voor $t_1 > 0$ en $t_2 > 0$ dat

$$\mathrm{P}(X > t_1 + t_2 | X > t_1) = \mathrm{P}(X > t_2) \tag{7.10}$$

Dit volgt eenvoudig uit

$$\begin{aligned}
\mathrm{P}(X > t_1 + t_2 | X > t_1) &= \frac{\mathrm{P}(X > t_1 + t_2, X > t_1)}{\mathrm{P}(X > t_1)} \\
&= \frac{\mathrm{P}(X > t_1 + t_2)}{\mathrm{P}(X > t_1)} \\
&= \frac{e^{-\lambda(t_1+t_2)}}{e^{-\lambda t_1}} \\
&= e^{-\lambda t_2} \\
&= \mathrm{P}(X > t_2)
\end{aligned}$$

De exponentiële verdeling heeft dus net als de geometrische verdeling geen geheugen. De voorwaardelijke kans om nog eens een tijd t_2 te wachten als men al een tijd t_1 heeft staan wachten, is gelijk aan de onvoorwaardelijke kans om een tijd t_2 te wachten. De exponentiële verdeling is dus als het ware vergeten dat men al een tijd t_1 stond te wachten. De exponentiële verdeling is de enige continue verdeling die deze eigenschap heeft.

7.3 Normale verdeling

7.3.1 Definitie en kenmerken

De belangrijkste verdeling in de (verklarende) statistiek is ongetwijfeld de normale verdeling. Deze verdeling werd door Adolphe Quetelet in de sociale statistiek geïntroduceerd. De kansvariabele X heeft een **normale verdeling** met parameters $\mu \in \mathbb{R}$ en $\sigma > 0$ als

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R} \tag{7.11}$$

Wanneer we deze kansdichtheid uitzetten in een grafiek krijgen we een zogenaamde Gauss-curve of klokcurve. Figuur 7.3 toont deze klokcurve voor verschillende waarden van μ bij gelijke σ (links) en voor verschillende waarden van σ bij gelijke μ (rechts). In de eerste grafiek zien we dat een wijziging van μ een verschuiving van de curve veroorzaakt. Vergroten van σ zorgt er voor dat de curve lager en breder wordt. De parameter μ is dus een locatieparameter, terwijl σ een maat voor de spreiding is. Dit wordt bevestigd door de uitdrukkingen voor verwachtingswaarde en variantie:

$$\mathrm{E}(X) = \mu \tag{7.12}$$

en

$$\mathrm{Var}(X) = \sigma^2 \tag{7.13}$$

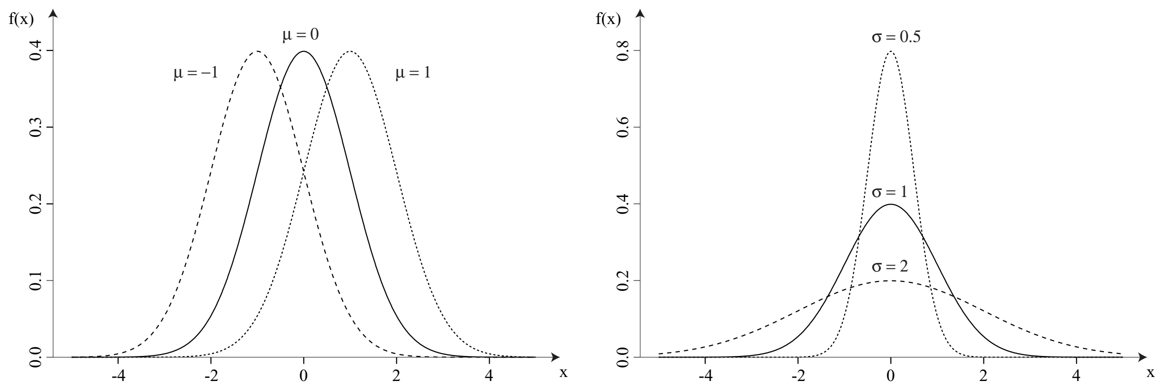
Uitdrukking (7.12) volgt meteen uit de vaststelling dat de normale kansdichtheid $f_X(x)$ symmetrisch is rond μ, m.a.w. $f_X(\mu + x) = f_X(\mu - x)$. Men kan immers eenvoudig aantonen dat een kansvariabele waarvan de kansdichtheid een symmetrie-as $x = a$ heeft, verwachtingswaarde a heeft (eigenschap 4.22). Merk op dat ook de modus en de mediaan van de normale verdeling gelijk zijn aan μ.

Omwille van het verband tussen de parameters enerzijds en de verwachtingswaarde en variantie anderzijds, wordt de normale verdeling genoteerd als

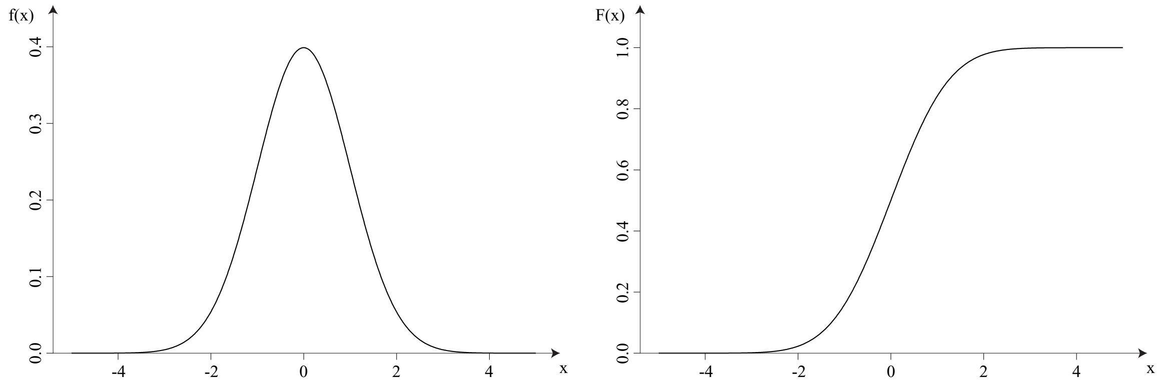
$$X \sim N(\mu, \sigma^2)$$

7.3.2 Standaard normale verdeling

De normale verdeling met parameters $\mu = 0$ en $\sigma = 1$ noemen we de **standaard normale verdeling**. Een kansvariabele met een standaard normale verdeling wordt doorgaans aangeduid met Z en de bijhorende kansdichtheid met $\phi(z)$. Uit (7.11) volgt dat deze kansdichtheid



Figuur 7.3: Normale kansdichtheid voor verschillende waarden van μ en σ



Figuur 7.4: Kansdichtheid en verdelingsfunctie van $Z \sim N(0,1)$

gelijk is

$$\phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \tag{7.14}$$

De verdelingsfunctie $F_Z(z)$ noteren we met $\Phi(z)$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \tag{7.15}$$

en de kwantielfunctie $Q_Z(p)$ met $\Phi^{-1}(p)$ of kortweg z_p. Voor de integraal in (7.15) is geen exacte oplossing gekend. De verdelingsfunctie $\Phi(z)$ kan bijgevolg niet exact berekend worden. Met behulp van numerieke methoden kunnen wel heel nauwkeurige benaderingen berekend worden. Op het rekentoestel kan dat met het commando `normalcdf(onder,boven)` waarbij we `onder` gelijkstellen aan `-1E99` en `boven` aan z. De bijhorende kansdichtheid berekenen we met het commando `normalpdf(z)`.

Uit de symmetrie van de kansdichtheid volgt nog een eenvoudig verband tussen de verdelingsfunctie voor negatieve en positieve waarden.

Eigenschap 7.6 (Verdelingsfunctie voor negatieve waarden)
Voor elke $z \in \mathbb{R}$ geldt

$$\Phi(-z) = 1 - \Phi(z) \tag{7.16}$$

Immers, aangezien de standaard normale kansdichtheid symmetrisch is rond $x = 0$, geldt

$$\Phi(-z) = \mathrm{P}(Z \leq -z) = \int_{-\infty}^{-z} \phi(x)dx = \int_{z}^{+\infty} \phi(x)dx = \mathrm{P}(Z \geq z)$$

Anderzijds is

$$\mathrm{P}(Z \geq z) = 1 - \mathrm{P}(Z < z) = 1 - \mathrm{P}(Z \leq z) = 1 - \Phi(z)$$

waaruit (7.16) onmiddellijk volgt.

Met deze eigenschap en eigenschap 4.7 vinden we nog

$$\mathrm{P}(-z \leq Z \leq z) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 \tag{7.17}$$

voor $Z \sim N(0,1)$ en elke $z \geq 0$. Met $z = 0$ wordt dit

$$0 = \mathrm{P}(0 \leq Z \leq 0) = 2\Phi(0) - 1$$

en dus

$$\Phi(0) = 0.5 \tag{7.18}$$

Voorbeeld 7.7 (Standaard normale kansen)
Voor $Z \sim N(0,1)$ geldt

$$\begin{aligned}\mathrm{P}(Z \leq 1.96) &= \Phi(1.96) = \texttt{normalcdf(-1E99,1.96)} = 0.9750\\ \mathrm{P}(Z \leq -1.96) &= \Phi(-1.96) = 1 - \Phi(1.96) = 0.0250\end{aligned}$$

Met het commando `invNorm(p)` berekenen we de kwantielen $\Phi^{-1}(p)$ van de standaard normale verdeling.

Voorbeeld 7.8 (Standaard normale kwantielen)
Uit voorbeeld 7.7 weten we dat $\mathrm{P}(Z \leq 1.96) = 0.9750$. Het 0.975-kwantiel zou dus 1.96 moeten zijn. Dit wordt bevestigd met het rekentoestel:

$$z_{0.975} = \Phi^{-1}(0.975) = \texttt{invNorm(0.975)} = 1.95996$$

De kwantielen voor $p < 0.5$ kunnen we ook berekenen uit de kwantielen voor $p > 0.5$.

Eigenschap 7.9 (Kwantielfunctie voor $p < 0.5$)
Voor elke $p \in]0,1[$ geldt

$$\Phi^{-1}(1-p) = -\Phi^{-1}(p) \tag{7.19}$$

Immers, uit vergelijking (7.16) weten we dat

$$1 - \Phi(z) = \Phi(-z)$$

Stellen we hierin $\Phi(z) = p$, en dus $z = \Phi^{-1}(p)$ aangezien Φ continu is, dan geldt

$$1 - p = \Phi(-\Phi^{-1}(p))$$

Van beide leden de kwantielfunctie Φ^{-1} berekenen, geeft dan

$$\Phi^{-1}(1-p) = \Phi^{-1}(\Phi(-\Phi^{-1}(p))) = -\Phi^{-1}(p)$$

aangezien $\Phi^{-1}(\Phi(x)) = x$.

Voorbeeld 7.10 (Standaard normale kwantielen)
Uit voorbeeld 7.8 weten we dat het 0.975-kwantiel van $Z \sim N(0,1)$ gelijk is aan 1.96 (afgerond). Het 0.025-kwantiel is dan

$$z_{0.025} = \Phi^{-1}(0.025) = \Phi^{-1}(1-0.975) = -\Phi^{-1}(0.975) = -z_{0.975} = -1.96$$

7.3.3 Willekeurige normale verdeling

Om de verdelingsfunctie van een willekeurige normale verdeelde kansvariabele te bepalen, kunnen we een eenvoudig verband tussen deze verdelingsfunctie en de standaard normale verdelingsfunctie gebruiken.

Eigenschap 7.11 (Lineaire transformatie)
De lineaire transformatie $Y = aX + b$ van een normaal verdeelde kansvariabele $X \sim N(\mu, \sigma^2)$ is eveneens normaal verdeeld, nl.

$$Y \sim N(a\mu + b, a^2\sigma^2) \tag{7.20}$$

Uit eigenschap 4.16 weten we immers dat

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Invullen van de normale kansdichtheid geeft dan

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{1}{2}\left(\frac{y-a\mu-b}{a\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma'} e^{-\frac{1}{2}\left(\frac{y-\mu'}{\sigma'}\right)^2} \end{aligned}$$

waarin we de normale kansdichtheid met parameters $\mu' = a\mu + b$ en $\sigma' = |a|\sigma$ herkennen.

Eigenschap 7.12 (Omzetten naar standaard normale verdeling)
Voor een willekeurige normaal verdeelde kansvariabele $X \sim N(\mu, \sigma^2)$ is de *gestandaardiseerde* kansvariabele

$$Z = \frac{X-\mu}{\sigma} \tag{7.21}$$

standaard normaal verdeeld. Dit volgt meteen uit eigenschap 7.11 met $a = \frac{1}{\sigma}$ en $b = \frac{-\mu}{\sigma}$.

Eigenschap 7.12 laat toe om de verdelingsfunctie van een willekeurige normaal verdeelde kansvariabele $X \sim N(\mu, \sigma^2)$ te berekenen in functie van Φ:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R} \tag{7.22}$$

De kwantielen van X volgen uit

$$Q_X(p) = \mu + \sigma\Phi^{-1}(p), \quad 0 < p < 1 \tag{7.23}$$

Er is dus een lineair verband tussen de kwantielen van een normaal verdeelde kansvariabele en deze van een standaard normaal verdeelde kansvariabele. De cumulatieve verdelingsfunctie en de kwantielfunctie van een normaal verdeelde kansvariabele kan ook rechtstreeks met het rekentoestel berekend worden d.m.v. de commando's `normalcdf(-1E99,x,`μ`,`σ`)` en `invNorm(p,`μ`,`σ`)`. Merk op dat de laatste parameter niet de variantie is maar de standaardafwijking.

Voorbeeld 7.13 (Kansen en kwantielen van een willekeurige normale verdeling)
Het intelligentiequotiënt (IQ) heeft in de meeste intelligentietesten een normale verdeling met parameters $\mu = 100$ en $\sigma = 15$. Hoe groot is de kans dat iemand minstens 130 scoort op een dergelijke test? Noteren we de score met $X \sim N(100, 15^2)$, dan

$$P(X \geq 130) = 1 - P(X \leq 130) = 1 - P\left(\frac{X-100}{15} \leq \frac{130-100}{15}\right) = 1 - P(Z \leq 2)$$

met $Z \sim N(0,1)$. Daaruit volgt

$$P(X \geq 130) = 1 - 0.9772 = 1 - \texttt{normalcdf(-1E99,2)} = 0.0228$$

wat overeenstemt met

$$1 - \texttt{normalcdf(-1E99,130,100,15)} = \texttt{normalcdf(130,1E99,100,15)} = 0.0228$$

De kans dat iemand een score haalt tussen 70 en 85 is

$$\begin{aligned} P(70 \leq X \leq 85) &= P\left(\frac{70-100}{15} \leq \frac{X-100}{15} \leq \frac{85-100}{15}\right) \\ &= P(-2 \leq Z \leq -1) = \Phi(-1) - \Phi(-2) = 0.1359 \end{aligned}$$

wat precies evenveel is als de kans op een score tussen 115 en 130 [ga zelf na].

Welke score moet men minstens halen om bij de top 1% te horen? Noteren we deze score met s, dan is $P(X \leq s) = 0.99$ en dus $s = Q_X(0.99)$. In termen van Z wordt dit

$$P\left(\frac{X-100}{15} \leq \frac{s-100}{15}\right) = P\left(Z \leq \frac{s-100}{15}\right) = 0.99$$

en dus

$$\Phi^{-1}(0.99) = \frac{s-100}{15}$$

Bijgevolg is

$$Q_X(0.99) = s = 100 + 15\Phi^{-1}(0.99) \approx 135$$

zoals ook meteen volgt uit vergelijking (7.23).

De ongelijkheid van Chebyshev (4.32) geeft ons een (ruime) afschatting voor een willekeurige kansvariabele X, louter op basis van de verwachtingswaarde en de variantie. Wanneer we nu de bijkomende veronderstelling maken dat X normaal verdeeld is, dan geldt

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= P(-1 \leq Z \leq 1) = 68.3\% \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= P(-2 \leq Z \leq 2) = 95.4\% \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= P(-3 \leq Z \leq 3) = 99.7\% \end{aligned}$$

De kans dat een normaal verdeelde variabele een waarde aanneemt die verder dan 3 standaardafwijkingen van het gemiddelde verwijderd is, is dus slechts 0.3%.

7.3.4 Centrale limietstelling

De **centrale limietstelling** (CLS) is één van de basisresultaten uit de statistiek. Eenvoudig gezegd stelt de CLS dat de som van n onafhankelijke kansvariabelen X_i voor $n \to +\infty$ normaal verdeeld is, ongeacht de verdelingen van de kansvariabelen X_i.

Stelling 7.14 (Centrale limietstelling)
Beschouw n onafhankelijke kansvariabelen $X_1, X_2, \ldots, X_n$ met een willekeurige kansverdeling of -dichtheid en met $\mathrm{E}(X_i) = \mu_i$ en $\mathrm{Var}(X_i) = \sigma_i^2$. Onder bepaalde technische voorwaarden geldt voor $Y_n = X_1 + X_2 + \cdots + X_n$ dat

$$\lim_{n\to+\infty} \mathrm{P}\left(\frac{Y_n - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq x\right) = \Phi(x) \tag{7.24}$$

Voor voldoende grote waarden van n is Y_n dus bij benadering normaal verdeeld met verwachtingswaarde $\mathrm{E}(Y_n) = \sum_{i=1}^n \mu_i$ en variantie $\mathrm{Var}(Y_n) = \sum_{i=1}^n \sigma_i^2$. Indien de kansvariabelen $X_1, X_2, \ldots, X_n$ alle dezelfde verdeling hebben met $\mathrm{E}(X_i) = \mu$ en $\mathrm{Var}(X_i) = \sigma^2$, dan geldt

$$\lim_{n\to+\infty} \mathrm{P}\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \tag{7.25}$$

Opmerkelijk aan de centrale limietstelling is dat, op de technische voorwaarden na, geen enkele voorwaarde opgelegd wordt aan de verdeling van de kansvariabelen X_i. In de praktijk zullen we uiteraard nooit oneindig veel variabelen optellen, maar zoals in de stelling aangegeven, is de benadering geldig voor voldoende grote waarden van n. Wat de betekenis is van “voldoende groot” hangt wel af van de verdeling van de kansvariabelen X_i. Indien de kansvariabelen alle dezelfde verdeling hebben, dan kunnen we de volgende vuistregels hanteren:

- de kansdichtheid wijkt niet te sterk af van een normale kansdichtheid: $n \geq 5$
- de kansverdeling of -dichtheid vertoont geen al te grote pieken: $n \geq 12$
- voor andere continue kansvariabelen die in de praktijk voorkomen: $n \geq 30$

In hoofdstuk 6 zagen we dat de binomiale verdeling goed benaderd kon worden door de Poisson verdeling voor n groot en p klein. Volgens de centrale limietstelling kunnen we de binomiale verdeling echter ook benaderen door een normale verdeling.

Eigenschap 7.15 (Normale benadering van de binomiale verdeling)
De kansvariabele $X \sim B(n,p)$ is voor n voldoende groot benaderend normaal verdeeld,

$$B(n,p) \approx N(np, np(1-p)) \tag{7.26}$$

We kunnen dat inzien als volgt: stel dat $X_1, X_2, \ldots, X_n$ onderling onafhankelijk en Bernoulli verdeeld zijn met parameter p. De som van deze variabelen is dan per definitie binomiaal verdeeld met parameters n en p. Volgens de centrale limietstelling is de som echter ook bij benadering normaal verdeeld met verwachtingswaarde $n\mathrm{E}(X_i) = np$ en variantie $n\mathrm{Var}(X_i) = np(1-p)$.

Als vuistregel voor de minimale waarde voor n stelt men meestal dat zowel np als $n(1-p)$ minstens 5 moeten zijn. Voor de Poisson verdeling kunnen we een gelijkaardige benadering formuleren [ga zelf na]. De benadering van een discrete kansvariabele X door een continue kansvariabele Y stelt ons wel voor een probleem. Hoe benaderen we bijvoorbeeld $\mathrm{P}(X=c)$? Voor een continue kansvariabele is deze kans immers altijd 0. Analoog hebben we $\mathrm{P}(X \leq c) \neq \mathrm{P}(X < c)$ terwijl $\mathrm{P}(Y \leq c) = \mathrm{P}(Y < c)$. We voeren daarom eerst een zogenaamde **continuïteitscorrectie** uit:

$$\mathrm{P}(a \leq X \leq b) \approx \mathrm{P}\left(a - \frac{1}{2} \leq Y \leq b + \frac{1}{2}\right) \tag{7.27}$$

De kans $\mathrm{P}(X=c)$ wordt dan benaderd door

$$\mathrm{P}(X=c) = \mathrm{P}(c \leq X \leq c) \approx \mathrm{P}\left(c - \frac{1}{2} \leq Y \leq c + \frac{1}{2}\right)$$

Voorbeeld 7.16 (Continuïteitscorrectie)
Een examen statistiek bestaat uit 20 meerkeuzevragen met telkens 4 mogelijke antwoorden. Indien elke vraag slechts 1 juist antwoord heeft en de student kiest lukraak een antwoord, hoe groot is dan de kans dat de student hoogstens 5 vragen juist beantwoordt? Noteren we het aantal juist beantwoorde vragen met X, dan is X binomiaal verdeeld met parameters $n = 20$ en $p = \frac{1}{4}$, $X \sim B(20, 0.25)$. Aangezien $np = 5 \geq 5$ en $n(1-p) = 15 \geq 5$ kunnen we deze verdeling benaderen met een normale verdeling met parameters $\mu = np = 5$ en $\sigma^2 = np(1-p) = 3.75$, $Y \sim N(5, 3.75)$. Daaruit volgt

$$\mathrm{P}(X \leq 5) = 0.6172 \quad \text{en} \quad \mathrm{P}(Y \leq 5) = 0.5$$

Passen we de continuïteitscorrectie toe, dan vinden we

$$\mathrm{P}(X \leq 5) \approx \mathrm{P}\left(Y \leq 5 + \frac{1}{2}\right) = 0.6019$$

De kans dat de student precies 5 vragen juist beantwoordt, kunnen we benaderen door

$$\mathrm{P}(X = 5) \approx \mathrm{P}\left(5 - \frac{1}{2} \leq Y \leq 5 + \frac{1}{2}\right) = 0.2037$$

wat dicht bij de exacte kans $\mathrm{P}(X = 5) = 0.2023$ ligt.

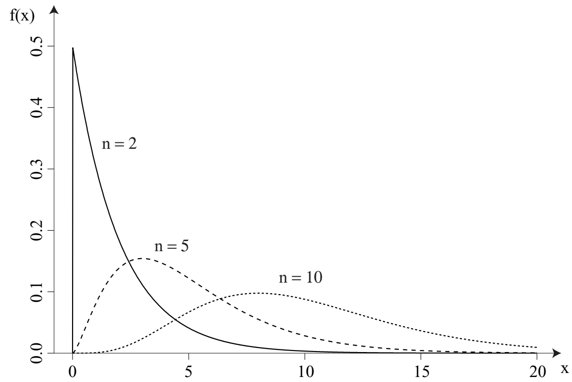
7.4 χ^2-verdeling

In de verklarende statistiek speelt ook de **chi-kwadraat** (χ^2) verdeling een belangrijke rol. Een kansvariabele X heeft een χ^2-verdeling met parameter n, genoteerd $X \sim \chi^2_n$, als

$$f_X(x) = \frac{2^{-n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2}, \quad x > 0$$

waarbij de gammafunctie Γ gedefinieerd is als

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx, \quad t > 0$$



Figuur 7.5: Kansdichtheid van de χ^2_n verdeling voor verschillende n

De parameter n noemt men het aantal **vrijheidsgraden** van de verdeling. Figuur 7.5 toont de kansdichtheid voor verschillende waarden van n. De χ^2-verdeling is dus een rechtsscheve verdeling.

Eigenschap 7.17 (Karakterisatie van de χ^2 verdeling)
Een kansvariabele is χ^2_n-verdeeld als en slechts dan als de variabele kan geschreven worden als de som van de kwadraten van n onafhankelijke standaard normaal verdeelde variabelen $Z_1, Z_2, \ldots, Z_n$:

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2_n \tag{7.28}$$

met $Z_i \sim N(0,1)$ onafhankelijk.

Voor de verwachtingswaarde en variantie van $X \sim \chi^2_n$ geldt

$$\mathrm{E}(X) = n \quad \text{en} \quad \mathrm{Var}(X) = 2n \tag{7.29}$$

Zoals bij de normale verdeling zijn numerieke methoden nodig om de verdelingsfunctie $F(x)$ te berekenen. Op het rekentoestel kan dat met het commando `χ²cdf(onder,boven,n)` waarbij we `onder` gelijkstellen aan 0 en `boven` aan x. De bijhorende kansdichtheid $f(x)$ berekenen we met het commando `χ²pdf(x,n)`.

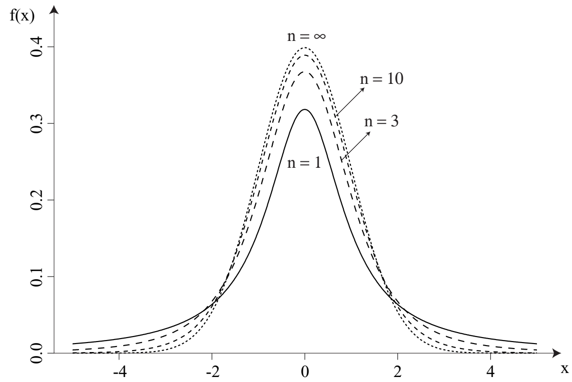
De kwantielfunctie vinden we door de verdelingsfunctie numeriek te inverteren. Het programma `INVCHI2` berekent het p-kwantiel $\chi^2_{n,p}$:

```
Input "df=",N
Input "P(X≤x)=",P
solve(χ²cdf(0,X,N)-P,X,1,0,1E99)
```

7.5 t-verdeling

Een andere belangrijke verdeling uit de verklarende statistiek is de **t-verdeling**. Een kansvariabele X heeft een t-verdeling met parameter n, genoteerd $X \sim t_n$, als

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}$$



Figuur 7.6: Kansdichtheid van de t_n-verdeling voor verschillende n

De parameter n noemt men ook hier het aantal vrijheidsgraden. Figuur 7.6 toont de kansdichtheid voor verschillende waarden van n. Als het aantal vrijheidsgraden oneindig groot wordt, vinden we de standaard normale dichtheid terug. De t-verdeling is dus een symmetrische verdeling met zwaardere staarten dan de normale verdeling.

Eigenschap 7.18 (Karakterisatie van de t-verdeling)
Een kansvariabele is t-verdeeld als en slechts dan als de variabele kan geschreven worden als het quotiënt van een standaard normaal verdeelde kansvariabele X en de wortel uit een onafhankelijke χ^2_n-verdeelde kansvariabele Y gedeeld door n:

$$\frac{X}{\sqrt{Y/n}} \sim t_n \tag{7.30}$$

met $X \sim N(0,1)$ en $Y \sim \chi^2_n$ onafhankelijk.

Voor de verwachtingswaarde en variantie van $X \sim t_n$ geldt

$$\mathrm{E}(X) = 0 \tag{7.31}$$

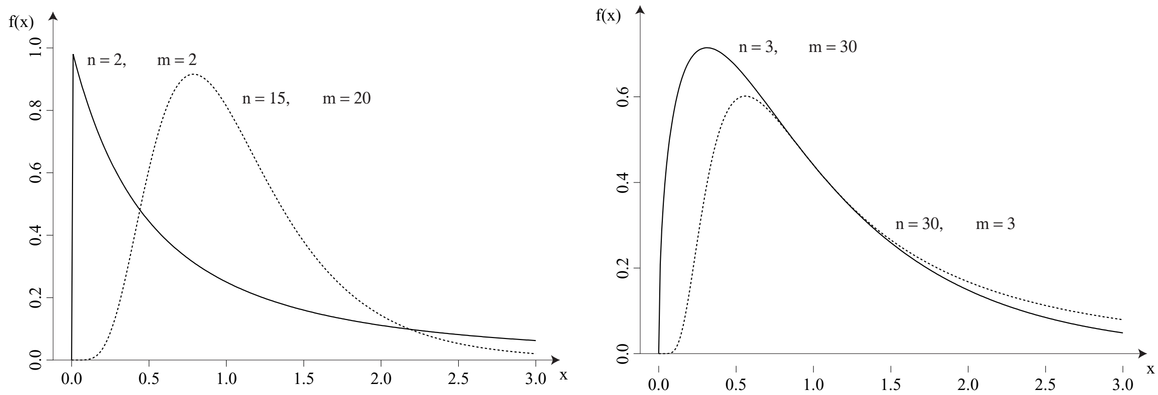
en

$$\mathrm{Var}(X) = \frac{n}{n-2}, \quad n > 2 \tag{7.32}$$

De variantie is dus steeds groter dan bij de standaard normale verdeling.

Ook hier zijn weer numerieke methoden nodig om de verdelingsfunctie $F(x)$ te berekenen. Op het rekentoestel kan dat met het commando `tcdf(onder,boven,n)` waarbij we `onder` gelijkstellen aan `-1E99` en `boven` aan x. De bijhorende kansdichtheid berekenen we met het commando `tpdf(x,n)` en het p-kwantiel $t_{n,p}$ met `invT(p,n)`.

Voorbeeld 7.19 (Kwantielen van de t-verdeling)
Uit voorbeeld 7.8 weten we dat het 0.975-kwantiel van de standaard normale verdeling gelijk is aan 1.96. Aangezien de t-verdeling zwaardere staarten heeft, zal het 0.975-kwantiel groter zijn dan 1.96. Voor $n = 10$ bijvoorbeeld vinden we $t_{10,0.975} =$ `invT(0.975,10)` $= 2.23$.



Figuur 7.7: Kansdichtheid van de $F_{n,m}$-verdeling voor verschillende n en m

7.6 F-verdeling

De **F-verdeling** ten slotte is een verdeling met 2 parameters. Een kansvariabele X heeft een F-verdeling met vrijheidsgraden n en m, genoteerd $X \sim F_{n,m}$, als

$$f_X(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} n^{n/2} m^{m/2} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, \quad x > 0$$

Figuur 7.7 toont de kansdichtheid voor verschillende waarden van n en m. De F-verdeling is dus een rechtsscheve verdeling.

Eigenschap 7.20 (Karakterisatie van de F-verdeling)
Een kansvariabele is F-verdeeld als en slechts dan als de variabele kan geschreven worden als het quotiënt van twee onafhankelijke χ^2-verdeelde kansvariabelen X en Y, elk gedeeld door het aantal vrijheidsgraden:

$$\frac{X/n}{Y/m} \sim F_{n,m} \tag{7.33}$$

met $X \sim \chi^2_n$ en $Y \sim \chi^2_m$ onafhankelijk.

Voor de verwachtingswaarde en variantie van $X \sim F_{n,m}$ geldt

$$\mathrm{E}(X) = \frac{m}{m-2}, \quad m > 2 \qquad \text{en} \qquad \mathrm{Var}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, \quad m > 4 \tag{7.34}$$

De verdelingsfunctie kunnen we berekenen met het commando `Fcdf(onder,boven,n,m)` waarbij we `onder` gelijkstellen aan 0 en `boven` aan x. Het commando `Fpdf(x,n,m)` geeft de bijhorende kansdichtheid. Voor het p-kwantiel $F_{n,m,p}$ gebruiken we met het programma `INVF`:

```
Input "df1=",N
Input "df2=",M
Input "P(X≤x)=",P
solve(Fcdf(0,X,N,M)-P,X,1,0,1E99)
```

7.7 Bivariate normale verdeling

Om dit hoofdstuk af te sluiten, beschouwen we nog een multivariate uitbreiding van de normale verdeling, meer bepaald de bivariate normale verdeling. Een kansvariabele (X,Y) heeft een bivariate normale verdeling als, voor $(x,y) \in \mathbb{R}^2$,

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \tag{7.35}$$

met μ_X en μ_Y reële getallen, $\sigma_X > 0$, $\sigma_Y > 0$ en $-1 \le \rho \le 1$. Figuur 7.8 toont de kansdichtheid voor verschillende waarden van ρ.

De marginale kansdichtheden van X en Y zijn respectievelijk

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}$$

en

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}$$

De kansvariabelen X en Y zijn dus allebei normaal verdeeld

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2) \tag{7.36}$$

Verder vinden we

$$\text{Cov}(X,Y) = \rho\sigma_X\sigma_Y \tag{7.37}$$

De bivariate verdeling wordt dan ook genoteerd als

$$(X,Y) \sim BVN(\vec{\mu}, \Sigma) \tag{7.38}$$

met

$$\vec{\mu} = (\mu_X, \mu_Y), \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \tag{7.39}$$

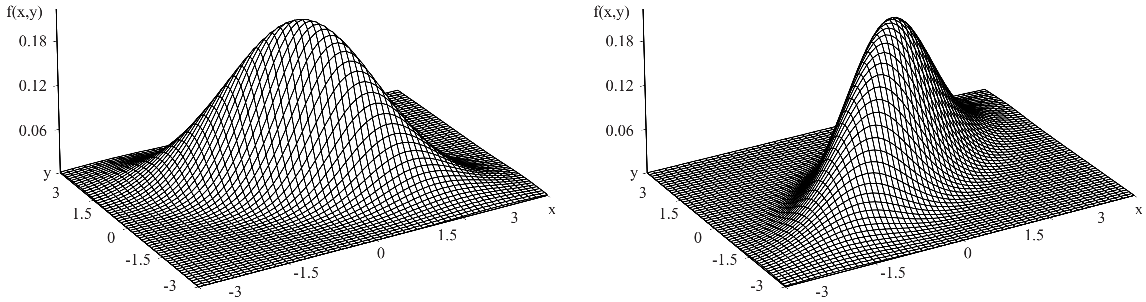
De parameter ρ speelt een belangrijke rol bij de afhankelijkheid tussen X en Y. We vinden immers

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} = \frac{\rho\sigma_X\sigma_Y}{\sigma_X\sigma_Y} = \rho \tag{7.40}$$

De parameter ρ van de bivariate normale verdeling is dus precies gelijk aan de correlatie $\rho(X,Y)$. Uit eigenschap 5.20 weten we dat de correlatie tussen onafhankelijke variabelen gelijk is aan 0. Voor bivariate normale variabelen geldt ook het omgekeerde.

Eigenschap 7.21 (Bivariate normale verdeling en onafhankelijkheid)
Het koppel (X,Y) heeft een bivariate normale verdeling met $\rho = 0$ als en slechts als X en Y onafhankelijk en normaal verdeeld zijn. Immers, met $\rho = 0$ geldt

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\} \\ &= f_X(x)f_Y(y) \end{aligned}$$



Figuur 7.8: Bivariate normale kansdichtheid met $\rho = -0.7$ (links) en $\rho = 0.7$ (rechts)

voor alle $(x, y) \in \mathbb{R}^2$.

Bij eigenschap 5.20 merkten we op dat ongecorreleerde kansvariabelen niet noodzakelijk onafhankelijk zijn. Bij de bivariate normale verdeling geldt dit dus wel. Omgekeerd is de voorwaarde van onafhankelijkheid in eigenschap 7.21 ook van belang. Indien de kansvariabelen X en Y niet onafhankelijk zijn, zijn ze samen niet noodzakelijk bivariaat normaal verdeeld.

Een andere interessante eigenschap van de bivariate normale verdeling, en bij uitbreiding van de multivariate normale verdeling, is de volgende.

Eigenschap 7.22 (Lineaire combinatie van bivariate normale kansvariabelen)
Indien (X, Y) een bivariate normale verdeling $BVN(\overrightarrow{\mu}, \Sigma)$ heeft, dan is elke lineaire combinatie $aX + bY$ eveneens normaal verdeeld. De parameters volgen uit formules (5.16) en (5.31),

$$\mathrm{E}(aX + bY) = a\mu_X + b\mu_Y \quad \text{en} \quad \mathrm{Var}(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$$

zodat $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$.

Voorbeeld 7.23 (Value-at-Risk)
Voor het bepalen van de efficiënte portefeuilles in voorbeeld 5.24 gebruikten we enkel de verwachtingswaarden en (co)varianties. We legden m.a.w. geen specifieke verdeling op aan de rendementen. Willen we echter de kans berekenen dat het rendement van de portefeuille negatief kan zijn en dat de belegger dus verlies maakt, dan moeten we wel een verdeling opleggen aan de variabelen. Een klassieke veronderstelling die daarbij gemaakt wordt, is dat de rendementen multivariaat normaal verdeeld zijn.

Stel dat de belegger wil investeren in aandelen uit de technologiesector (R_2) en in vastgoed (R_4). We nemen daarbij aan dat de rendementen (R_2, R_4) bivariaat normaal verdeeld zijn met

$$\overrightarrow{\mu} = (0.15, 0.11), \quad \Sigma = \begin{pmatrix} 0.26^2 & 0.05 \cdot 0.26 \cdot 0.125 \\ 0.05 \cdot 0.26 \cdot 0.125 & 0.125^2 \end{pmatrix}$$

Indien de belegger 80% van de middelen investeert in aandelen en 20% in vastgoed, hoe groot is dan de kans dat hij verlies zal lijden? Het rendement van de portefeuille $R_p = 0.8R_2 + 0.2R_4$ is wegens eigenschap 7.22 normaal verdeeld met

$$\mathrm{E}(R_p) = 0.8 \cdot 0.15 + 0.2 \cdot 0.11 = 0.142$$

en

$$\mathrm{Var}(R_p) = 0.8^2 0.26^2 + 0.2^2 0.125^2 + 2 \cdot 0.8 \cdot 0.2 \cdot 0.05 \cdot 0.26 \cdot 0.125 = 0.0444$$

Voor de kans $\mathrm{P}(R_p \leq 0)$ vinden we dan

$$\mathrm{P}(R_p \leq 0) = \mathrm{P}\left(\frac{R_p - 0.142}{\sqrt{0.0444}} \leq \frac{0 - 0.142}{\sqrt{0.0444}}\right) = \Phi\left(\frac{-0.142}{\sqrt{0.0444}}\right) \approx 0.25$$

De kans op een negatief rendement is dus niet gering.

Tot slot beschouwen we nog een speciaal geval van eigenschap 7.22.

Eigenschap 7.24 (Som van onafhankelijke normaal verdeelde kansvariabelen)
De som Y_n van onafhankelijke en normaal verdeelde kansvariabelen $X_1, X_2, \ldots, X_n$ is opnieuw normaal verdeeld, m.a.w. voor

$$X_i \sim N(\mu_i, \sigma_i^2) \quad \text{o.o.}$$

geldt dat

$$Y_n = X_1 + \ldots + X_n \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

De limietverdeling van de centrale limietstelling is in dit geval dus exact voor alle n.

Oefeningen hoofdstuk 7

Bekende continue verdelingen

Oefening 7.1
Een geneesmiddel heeft een effect op een normale mens waarvan de duur exponentieel verdeeld mag worden verondersteld. De gemiddelde werking van het geneesmiddel is 30 uur. Wat is de kans dat bij een persoon het geneesmiddel nog steeds werkt na 32 uur?

Oefening 7.2
Zij T de exponentieel verdeelde kansvariabele die de tijd voorstelt (uitgedrukt in minuten) die verstrijkt vanaf openingstijd, totdat de eerste klant een winkel binnentreedt. De verwachtingswaarde van T is $\frac{5}{2}$. Bepaal

(a) de cumulatieve verdelingsfunctie;

(b) $\mathrm{P}(\mathrm{T} \leq 3)$, $\mathrm{P}(\mathrm{T} > 4)$, $\mathrm{P}(3 < \mathrm{T} \leq 4)$;

(c) de variantie van T.

Oefening 7.3
De continue kansvariabele X is uniform verdeeld in het interval $[-\sqrt{3}, +\sqrt{3}]$.

(a) Bereken de kansen dat $|X| \geq \frac{3}{2}$ en $|X| \geq 2$.

(b) Bepaal de verwachtingswaarde, alsook de standaardafwijking van X.

Oefening 7.4
Het aantal uur per dag dat een student op Facebook zit, is $N(2.5, 4)$ verdeeld. Wat is de kans dat een student minstens 3 uur per dag op Facebook zit?

Oefening 7.5
De duur van een telefoongesprek volgt een exponentiële verdeling met een verwachtingswaarde van 5 minuten. Wat is de kans dat:

(a) een gesprek minstens 5 minuten zal duren;

(b) een gesprek tussen de 5 en 6 minuten zal duren?

Oefening 7.6
De tijdsduur nodig voor het uitvoeren van een bepaalde proef is uniform verdeeld in het interval [30',40']. Als 5 laboranten deze proef uitvoeren, bepaal dan de kans dat minstens 2 van hen minder dan 33' nodig hebben.

Oefening 7.7
Zij X een exponentieel verdeelde variabele met verwachtingswaarde 2. Bepaal:

(a) $P(X > 3)$ en $P(2 < X < 4)$;

(b) a zodat $P(X > a) = 0.6$.

Oefening 7.8
Een eilandengroep in de Stille Oceaan wordt regelmatig geteisterd door overvloedige overstromingen. Stel dat de tijd die verloopt tussen 2 overstromingen exponentieel verdeeld is met een verwachtingswaarde van 2 jaar.

(a) Bereken de kans dat er minstens 5 jaar is tussen 2 opeenvolgende overstromingen.

(b) Geef de kans dat het gebied hoogstens 5 keer gedurende de volgende 20 jaar onder water zal staan.

Oefening 7.9
De levensduur T van een systeem is exponentieel verdeeld met verwachtingswaarde 1000u.

(a) Bepaal de kwantielen $Q_T(0.25)$, $Q_T(0.50)$ en $Q_T(0.75)$.

(b) Bereken $\dfrac{Q_T(0.75) - Q_T(0.25)}{\sigma_T}$.

(c) Wat is de kans dat het systeem langer dan 60 dagen in werking blijft?

Oefening 7.10
Een grote transportfirma wil een verzekeringscontract afsluiten met een nieuwe verzekeringsmaatschappij. Deze laatste neemt aan dat de kans dat een vrachtwagen van de firma langer dan 12 jaar bruikbaar is gelijk is aan 5% (deze variabele is exponentieel verdeeld). Bereken de kans dat een vrachtwagen een levensduur heeft tussen 5 en 8 jaar.

Oefening 7.11
De lengte van personen is normaal verdeeld met verwachtingswaarde $\mu = 167$ cm en standaardafwijking $\sigma = 3$ cm. Wat is het percentage van de bevolking met lengte

(a) groter dan 167 cm;

(b) groter dan 170 cm;

(c) begrepen tussen 161 en 173 cm?

Oefening 7.12
Een machine vult dozen die verondersteld worden 1 pond suiker te bevatten. We veronderstellen dat het gewicht van de gevulde dozen normaal verdeeld is met verwachtingswaarde 16.5 ons en standaardafwijking 0.5 ons (1 pond = 16 ons). Hoeveel procent van de dozen zal minder wegen dan 1 pond? Hoeveel procent van de gevulde dozen zal een gewicht bevatten dat begrepen is tussen 16.3 ± 0.2 ons?

Oefening 7.13
Men heeft 4 onafhankelijke variabelen X, Y, Z en U die elk standaard normaal verdeeld zijn. Bepaal de kans dat

(a) $X^2 + Y^2 < 7.378$

(b) $\dfrac{X^2 + Y^2}{Z^2 + U^2} < 19$

Oefening 7.14
Een koffieautomaat bevat bekertjes die elk maximaal 100 ml kunnen bevatten. De hoeveelheid koffie die de automaat per keer levert is normaal verdeeld met parameters 90 ml (=μ) en 4.3 ml (=σ). Bereken de kans dat het bekertje overloopt.

Oefening 7.15
De kansvariabele X heeft een normale verdeling met verwachtingswaarde 24 en standaardafwijking 6. Bereken a als gegeven is:

(a) $\mathrm{P}(X > a) = 0.05$;

(b) $\mathrm{P}(X < a) = 0.025$

(c) $\mathrm{P}(30 < X < a) = 0.1$;

(d) $\mathrm{P}(24 - a < X < a + 24) = 0.95$

(e) $\mathrm{P}(Y \geq -36) = a^2$ waarbij $Y = -2X - 5$.

Oefening 7.16
Een fabrikant van digitale hoogtemeters garandeert zijn klanten gratis vervanging in geval van defect binnen de eerste twee jaar en vervanging tegen halve prijs bij defect binnen 5 jaar (en meer dan 2 jaar). De levensduur van een digitale hoogtemeter heeft een exponentiële verdeling met verwachtingswaarde 10 jaar. Bereken de kans op een defect

(a) binnen de eerste twee jaar;

(b) binnen 5 jaar (maar meer dan 2 jaar).

Als de verkoopprijs 200 euro bedraagt waarvan 125 euro produktiekosten zijn,

(c) bereken dan de verwachte winst op 100 verkochte hoogtemeters.

Oefening 7.17
Beschouw de onderling onafhankelijke veranderlijken X_1, X_2, X_3 en Y waarbij elke X_i een $N(0,4)$-verdeling en Y een χ_9^2-verdeling heeft. Bepaal a zodat

(a) $\mathrm{P}(X_1 < \sqrt{Y}) = a$

(b) $\mathrm{P}(8X_1^2 \geq X_2^2 + X_3^2) = a$

(c) $\mathrm{P}(X_1 < a\sqrt{Y}) = 97.5\%$

(d) $\mathrm{P}(aX_1^2 \geq X_2^2 + X_3^2) = 5\%$

Oefening 7.18
Wie een IQ heeft dat behoort tot de hoogste 2% van de populatie kan lid worden van de internationale Mensa organisatie. Wat is de laagste IQ waarmee men nog toegelaten wordt als we weten dat het gemiddelde 100 en de standaardafwijking 15 is en het IQ normaal verdeeld is over de populatie?

Oefening 7.19
De variabele X is χ^2-verdeeld met 24 vrijheidsgraden.

(a) Bereken $\mathrm{P}(X = 13.85)$.

(b) Als $Y \sim \chi_2^2$ onafhankelijk is van X, wat is dan de kans dat $X > -Y + 38.885$?

Oefening 7.20
Het gewicht van een pilletje is normaal verdeeld met verwachtingswaarde 3g en standaardafwijking 0.5g. Bepaal de kans dat

(a) de inhoud van een doosje met 100 pilletjes minder dan 285g weegt;

(b) er op 10 pilletjes juist 4 zijn waarvan het gewicht begrepen is tussen 2.56g en 3.53g.

Oefening 7.21
De variabele X is normaal verdeeld met parameters 53 (=μ) en 90 (=σ^2). Bereken in de volgende uitdrukkingen de ontbrekende waarden:

(a) $\mathrm{p}_1 = \mathrm{P}(X > 40)$ en $\mathrm{p}_2 = \mathrm{P}(60 < X < 65)$;

(b) $\mathrm{P}(X > a) = 0.95$.

Oefening 7.22
Beschouw de onafhankelijke variabelen $X \sim N(2,9)$ en $Y \sim N(3,16)$. Bepaal a zodat

(a) $\mathrm{P}(X + Y < 8) = a$

(b) $\mathrm{P}(X + Y < a) = 0.1$

(c) $\mathrm{P}(X - Y < 3) = a$

Oefening 7.23
De variabele X is normaal verdeeld met verwachtingswaarde 0 en $\mathrm{P}(X^2 \leq 15) = 0.95$. Wat is de standaardafwijking van X?

Oefening 7.24
De variabelen X_1, X_2, X_3, X_4 en X_5 zijn onderling onafhankelijk en standaardnormaal verdeeld. Bepaal a zodat

(a) $\mathrm{P}(X_1^2 + X_2^2 \leq X_3^2 + X_4^2 + X_5^2) = a$

(b) $\mathrm{P}(X_1^2 + X_2^2 \leq X_3^2 - X_4^2 + X_5^2) = a$

(c) $\mathrm{P}(X_1^2 - 2X_2^2 \leq 2X_3^2 - X_4^2 - X_5^2) = a$

(d) $\mathrm{P}(X_1^2 - aX_2^2 \leq aX_3^2 - X_4^2 - X_5^2) = 0.9$

Oefening 7.25
Stel dat $X \sim N(2,4)$ en $Y \sim N(0,4)$ onafhankelijk zijn. Bepaal

(a) $\mathrm{P}(X^2 - 4X + Y^2 < 0)$

(b) $\mathrm{P}(\sqrt{(X-2)^2} < Y)$

Oefening 7.26
Stel dat $X \sim N(0,2)$. Bepaal $\mathrm{E}(X^4)$.

Oefening 7.27
De levensduur T van een elektrische component is exponentieel verdeeld met gemiddelde waarde 500u. Men beschikt over 20 lampen. Zij hun levensduur t_i voor $i = 1\ldots20$. Bepaal benaderd de kans dat de som van de levensduren van deze 20 lampen groter is dan 1 jaar (waarbij verondersteld wordt dat de lampen dag en nacht branden).

Oefening 7.28
Beschouw 4 onafhankelijke veranderlijken X_1, X_2, X_3, X_4 die standaard normaal verdeeld zijn. Bepaal de kans dat

(a) $X_1 + 2X_2 + 3X_3 + 4X_4 < 3$;

(b) $\mid X_1 + X_2 + X_3 \mid^3 < 27$.

Oefening 7.29
Decimale getallen worden vaak benaderd door het dichtstbijzijnde gehele getal. Er worden 48 reële getallen willekeurig gekozen. De som S van deze 48 getallen wordt benaderd door de som S' van de overeenkomstige gehele getallen. Veronderstel dat de afrondingsfouten onafhankelijke kansvariabelen zijn die uniform verdeeld zijn over het interval $[\frac{-1}{2}, \frac{1}{2}[$. Bepaal benaderd de kans dat de som S' van de gehele getallen niet meer dan 2 eenheden afwijkt van de werkelijke som S.

Oplossingen

[7.1] 0.3442

[7.2] (a)

$$F(t) = \begin{cases} 0 & \text{als } t \leq 0 \\ 1 - \exp(\frac{-2}{5}t) & \text{als } t > 0 \end{cases}$$

(b) resp. 0.6988; 0.2019; 0.09930; (c) 6.25

[7.3] (a) resp. 0.1340; 0; (b) resp. 0; 1

[7.4] 0.4013

[7.5] (a) 0.3679; (b) 0.0667

[7.6] 0.4718

[7.7] (a) resp. 0.2231; 0.2325; (b) 1.0217

[7.8] (a) 0.0821; (b) 0.0671

[7.9] (a) resp. 287.682u; 693.147u; 1386.294u; (b) 1.0986; (c) 0.2369

[7.10] 0.1513

[7.11] (a) 50%; (b) 15.87%; (c) 95.45%

[7.12] resp. 15.87%; 28.81%

[7.13] (a) 0.975; (b) 0.95

[7.14] 0.0100204

[7.15] (a) 33.87; (b) 12.24; (c) 33.395; (d) 11.76; (e) ±0.2798

[7.16] (a) 0.1813; (b) 0.2122; (c) 4703.63 euro

[7.17] (a) 0.9161; (b) 0.6667; (c) 1.5081; (d) 0.1080

[7.18] 130.81

[7.19] (a) 0; (b) 0.05

[7.20] (a) 0.001350; (b) 0.0574

[7.21] (a) resp. 0.9147; 0.1273; (b) 37.3955

[7.22] (a) 0.7257; (b) −1.4078; (c) 0.7881

[7.23] 1.976

[7.24] (a) 0.6464; (b) 0.3536; (c) 0.5443; (d) 13.7427

[7.25] (a) 0.3935; (b) 0.25

[7.26] 12

[7.27] 0.7104

[7.28] (a) 0.7081; (b) 0.9167

[7.29] 0.6827