

Appendix

A.1 Erlang verdeling verdeling met parameters n en λ

Voor o.o. discrete s.v.-en X en Y geldt

$$\begin{aligned} P(X + Y = z) &= P(X = x_1 \text{ en } Y = z - x_1) + P(X = x_2 \text{ en } Y = z - x_2) + \dots \\ &= P(X = x_1)P(Y = z - x_1) + P(X = x_2)P(Y = z - x_2) + \dots \\ &= \sum_{i=1}^{\infty} P(X = x_i)P(Y = z - x_i). \end{aligned}$$

Zo geldt voor o.o. continue s.v.-en X en Y dat de kansdichtheid van $X + Y$ gegeven wordt door

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx.$$

In het volgende voorbeeld wordt de kansdichtheid bepaald van de som van twee o.o. s.v.-en die ieder exponentieel verdeeld zijn met parameter 1.

Voorbeeld A.1.1 De o.o. s.v.-en X en Y hebben als kansdichtheid

$$f_X(x) = \begin{cases} e^{-x} & \text{als } x \geq 0 \\ 0 & \text{als } x < 0 \end{cases}$$

en

$$f_Y(y) = \begin{cases} e^{-y} & \text{als } y \geq 0 \\ 0 & \text{als } y < 0. \end{cases}$$

De kansdichtheid van $X + Y$ in het punt $z > 0$ is gelijk aan

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_0^z e^{-x}e^{-(z-x)}dx \\ &= \int_0^z e^{-z}dx = ze^{-z}. \end{aligned}$$

□

APP.2

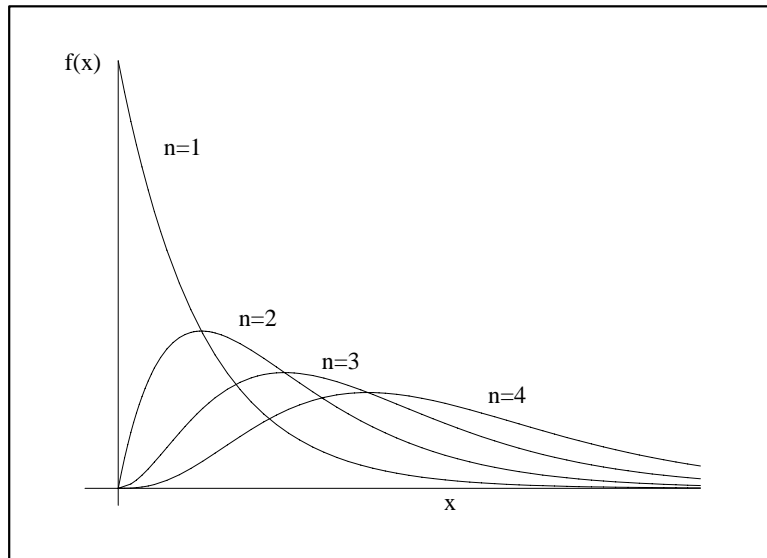
Evenzo kunnen we de kansdichtheid van $X + Y$ berekenen als X en Y o.o. s.v.-en zijn, ieder exponentieel verdeeld met parameter λ . We krijgen als kansdichtheid van $Z = X + Y$ (voor $z > 0$)

$$(A.1.1) \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx \\ = \lambda^2 \int_0^z e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z}.$$

Voorbeeld A.1.2 Een lokettist van een postkantoor moet twee wachtende personen bedienen. We kunnen de totale bedieningsduur modelleren met de som $X + Y$ van twee o.o. exponentiële verdeling, ieder met parameter λ . De kansdichtheid van de totale bedieningsduur wordt dus gegeven door (A.1.1). \square

De kansverdeling in (A.1.1) is een speciaal geval (met $n = 2$) van de Erlang verdeling. We zeggen dat een s.v. X een Erlang verdeling heeft met parameters n en λ als

$$(A.1.2) \quad f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \text{ voor } x \geq 0 \text{ (en } f_X(x) = 0 \text{ voor } x < 0).$$



De Erlang verdeling ontstaat als de verdeling van de som $X_1 + \dots + X_n$ als X_1, \dots, X_n o.o. exponentieel verdeeld zijn, ieder met de parameter λ . Met behulp van deze interpretatie kunnen we gemakkelijk de verwachting en variantie van de Erlang verdeling berekenen. Laat X_1, X_2, \dots, X_n o.o. s.v.-en zijn, ieder met een exponentiële verdeling met parameter λ . Dan is

$$E(X_1 + X_2 + \dots + X_n) = nE(X_1) = \frac{n}{\lambda}$$

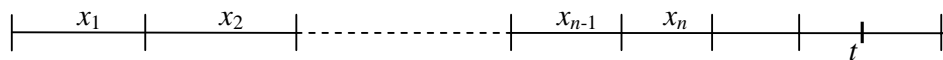
en

$$\text{var}(X_1 + X_2 + \dots + X_n) = n \text{var}(X_1) = \frac{n}{\lambda^2}.$$

Dus als X een Erlang verdeling heeft met parameters n en λ dan geldt

$$E(X) = \frac{n}{\lambda} \text{ en } \text{var}(X) = \frac{n}{\lambda^2}.$$

Bekijken we nu eens een rij klanten X_1, X_2, \dots die bediend moeten worden (o.o. s.v.-en ieder met een exponentiële verdeling met parameter λ). Stel dat we de kans willen uitrekenen dat de totale bedieningsduur van de klanten X_1, X_2, \dots, X_n ten hoogste gelijk is aan t , dan is dat hetzelfde als dat in het tijdsvak $[0, t]$ tenminste n klanten worden bediend. Het aantal klanten dat in het tijdsvak $[0, t]$ bediend wordt is juist Poisson-verdeeld met parameter $\mu = \lambda t$. (We bewijzen dit niet.)



We kunnen daarom $P(X_1 + X_2 + \dots + X_n \leq t)$ op twee manieren uitrekenen, via de Erlang verdeling

$$P(X_1 + X_2 + \dots + X_n \leq t) = \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx$$

of via de Poisson-verdeling

$$\begin{aligned} P(X_1 + X_2 + \dots + X - n \leq t) &= P(Y \geq n) = 1 - P(Y \leq n-1) \\ &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \end{aligned}$$

waarbij Y een s.v. is met een Poisson-verdeling met parameter $\mu = \lambda t$.

Voorbeeld A.1.3 Er zijn nog drie wachtenden voor u. Veronderstel dat de bedieningstijden (in minuten) o.o. s.v.-en zijn ieder met een exponentiële verdeling met parameter $\lambda = \frac{1}{2}$. (Dat correspondeert met een verwachte bedieningstijd van 2 minuten.) Veronderstel verder dat de bediening van de eerste van deze drie juist ingaat. Bereken de kans dat je 10 minuten of langer moet wachten.

We moeten dan de kans $P(X_1 + X_2 + X_3 \geq 10)$ berekenen, waarbij X_1, X_2, X_3 o.o. s.v.-en zijn, ieder met een exponentiële verdeling met parameter $\lambda = \frac{1}{2}$. Met behulp van de Erlang verdeling krijgen we door herhaalde partiële integratie

$$\begin{aligned} P(X_1 + X_2 + X_3 \geq 10) &= \int_{10}^{\infty} \frac{\left(\frac{1}{2}\right)^3 x^2 e^{-x/2}}{2!} dx \\ &= -\frac{1}{8} x^2 e^{-x/2} \Big|_{10}^{\infty} + \frac{1}{4} \int_{10}^{\infty} x e^{-x/2} dx \end{aligned}$$

APP.4

$$\begin{aligned}
 &= -\frac{1}{8}x^2e^{-x/2}\Big|_{10}^{\infty} - \frac{1}{2}xe^{-x/2}\Big|_{10}^{\infty} + \frac{1}{2}\int_{10}^{\infty}e^{-x/2}dx \\
 &= \frac{1}{8}10^2e^{-10/2} + \frac{1}{2}10e^{-10/2} + e^{-10/2} = \frac{37}{2}e^{-5} = 0.125.
 \end{aligned}$$

Gebruikmakend van de Poisson-verdeling gaat de berekening als volgt. De gebeurtenis $\{X_1 + X_2 + X_3 \geq 10\}$ is hetzelfde als "het aantal klanten bediend in $[0, 10)$ is minder dan 3" en dus ten hoogste 2. Het aantal klanten bediend in $[0, 10)$ is Poisson-verdeeld met parameter $\mu = \frac{1}{2} \times 10 = 5$, want $\lambda = \frac{1}{2}$ en $t = 10$. Laat Y een s.v. zijn met een Poisson-verdeling met parameter $\mu = 5$. Dan geldt dus

$$P(X_1 + X_2 + X_3 \geq 10) = P(Y \leq 2) = 0.125,$$

waarbij het laatste resultaat volgt uit de tabel van de Poisson-verdeling. Merk overigens op dat de drie termen $\frac{1}{8}10^2e^{-10/2}$, $\frac{1}{2}10e^{-10/2}$, $e^{-10/2}$ juist gelijk zijn aan $P(Y = 2)$, $P(Y = 1)$, $P(Y = 0)$. \square

Een dergelijke reeks van bedieningen, waarbij de bedieningsduren exponentieel verdeeld zijn, noemen we wel een **Poisson-proces**. Ook voor aankomsten van klanten, van elektronische boodschappen etc. wordt het Poisson-proces gehanteerd. De periodes tussen aankomsten zijn o.o. en exponentieel verdeelde s.v.-en. We spreken van een Poisson-proces met parameter λ als λ de parameter is van de exponentieel verdeelde s.v.-en. De totale aankomsttijd van n klanten heeft een Erlang verdeling met parameters n en λ , terwijl het aantal aankomsten in een tijdsvak $[0, t]$ een Poisson-verdeling heeft met parameter $\mu = \lambda t$.

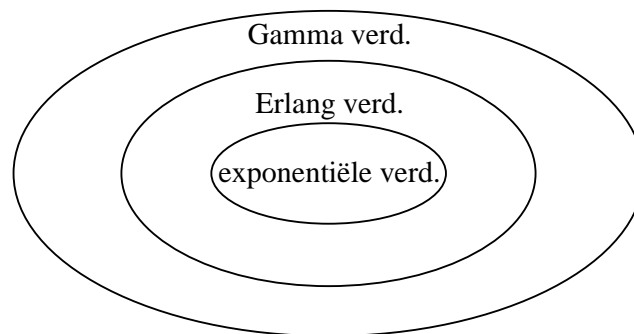
A.2 Gamma verdeling met parameters p en λ

De Gamma verdeling is een directe generalisatie van de Erlang verdeling door de beperking tot gehele waarden weg te laten. De kansdichtheid wordt gegeven door

$$(A.2.1) \quad f(x; p, \lambda) = \frac{\lambda^p x^{p-1} e^{-\lambda x}}{\int_0^{\infty} t^{p-1} e^{-t} dt} \text{ voor } x \geq 0 \text{ (en } f(x; p, \lambda) = 0 \text{ voor } x < 0).$$

De integraal $\int_0^{\infty} t^{p-1} e^{-t} dt$ wordt afgekort tot $\Gamma(p)$, de zogenaamde Gamma functie, waaraan de verdeling zijn naam ontleent.

Als p geheel is, zeg $p = n$, dan kan met partiële integratie berekend worden dat $\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt = (n-1)!$ en geeft (A.2.1) dus juist de kansdichtheid van een Erlang verdeling met parameters n en λ , zie (A.1.2). De parameters p en λ mogen alle positieve waarden aannemen. De parameter p kan dus ook bijvoorbeeld $\frac{5}{3}$ zijn. We hebben dus de volgende situatie:



Een ander speciale subklasse van de Gamma verdelingen is die met $\lambda = \frac{1}{2}$ en $p = \frac{n}{2}$, waarbij n een geheel getal is. Deze verdelingen spelen een belangrijke rol in de statistiek en worden de *chi-kwadraat verdelingen* genoemd.

De klasse van Gamma verdelingen is een rijke klasse van verdelingen, waarmee we veel verschijnselen kunnen modelleren. Dat biedt mogelijkheden als bijv. een exponentiële verdeling niet in voldoende mate de gegevens beschrijft. De ruimere klasse van Gamma verdelingen kan dan uitkomst bieden. Nadeel kan zijn dat de modelbeschrijving daardoor ingewikkelder wordt. Verder ontstaat wanneer de parameters onbekend zijn en geschat moeten worden een grotere schattingsfout. Zoals altijd moet een evenwicht gevonden worden tussen een zo nauwkeurig mogelijke beschrijving en de eenvoud van het model.

Voorbeeld A.2.1 Veronderstel dat de hoeveelheid regenval in mm gedurende de maand januari verdeeld is volgens een Gamma verdeling met $\lambda = 0.11$ en $p = 6.7$. Gevraagd is de kans dat de hoeveelheid regenval meer is dan 90 mm. Laat X de s.v. zijn die de hoeveelheid regenval aangeeft. Dan wordt gevraagd $P(X > 90)$. Deze kans is dan gelijk aan $\int_{90}^{\infty} f(x, 0.11, 6.7) dx$. Deze integraal kan bijvoorbeeld met behulp van Maple berekend worden. Het antwoord is 0.11. \square

Zonder bewijs vermelden we: als X een Gamma verdeling heeft met parameters p en λ , dan geldt $EX = \frac{p}{\lambda}$ en $\text{var}(X) = \frac{p}{\lambda^2}$. Als X en Y o.o. s.v.-en zijn, X met een Gamma verdeling met parameters p en λ en Y met een Gamma verdeling met parameters q en λ , dan is $X + Y$ Gamma verdeeld met parameters $p + q$ en λ . Ook dit bewijzen we niet.

A.3 Weibull verdeling met parameters c en λ

De kansdichtheid van de Weibull verdeling wordt gegeven door

$$(A.3.1) \quad f(x; c, \lambda) = c\lambda^c x^{c-1} e^{-(\lambda x)^c} \text{ voor } x \geq 0 \text{ (en } f(x; c, \lambda) = 0 \text{ voor } x < 0).$$

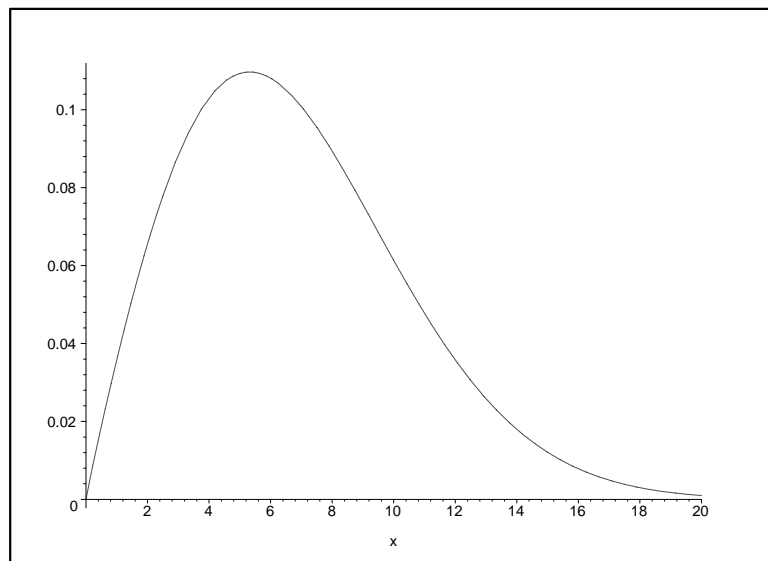
Nemen we $c = 1$ dan krijgen we de exponentiële verdeling. De klasse van Weibull verdelingen omvat dus de klasse van exponentiële verdelingen. Deze klasse biedt,

APP.6

net als de Gamma verdelingen, de mogelijkheid een betere modelbeschrijving te krijgen als de exponentiële verdeling niet toereikend is.

De Weibull verdeling wordt o.a vaak gebruikt om de levensduur van apparaten te modelleren. Een geheel ander voorbeeld waar de Weibull verdeling gebruikt wordt is de windsnelheid, die van belang is bij de opbrengst van windmolens.

Voorbeeld A.3.1 Veronderstel dat de windsnelheid een Weibull verdeling heeft met $c = 1.95$ en $\lambda = 0.13$. De kansdichtheid ziet er zo uit.



De top ligt bij 5.32, maar de verdeling is niet symmetrisch rond deze top. Dat sluit mooi aan bij de werkelijkheid: een windsnelheid van $5.32 + 6 = 11.32$ is heel wel mogelijk, maar een windsnelheid van $5.32 - 6$ uiteraard niet. Met deze kansverdeling kunnen we de kans berekenen dat de windsnelheid groter is dan 5.32. Deze kans is $\int_{5.32}^{\infty} 1.95 \cdot 0.13^{1.95} x^{0.95} e^{-(0.13x)^{1.95}} dx$. Numerieke berekening met Maple geeft als antwoord 0.61. Zoals het plaatje al laat zien, ligt rechts van de top meer dan de helft van de kansmassa. \square

Zonder bewijs vermelden we: als X een Weibull verdeling heeft met parameters

c en λ , dan geldt $EX = \frac{\Gamma\left(1 + \frac{1}{c}\right)}{\lambda}$ en $\text{var}(X) = \frac{1}{\lambda^2} \left\{ \Gamma\left(1 + \frac{2}{c}\right) - \left[\Gamma\left(1 + \frac{1}{c}\right) \right]^2 \right\}$. Hierin is $\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt$, de Gamma functie.

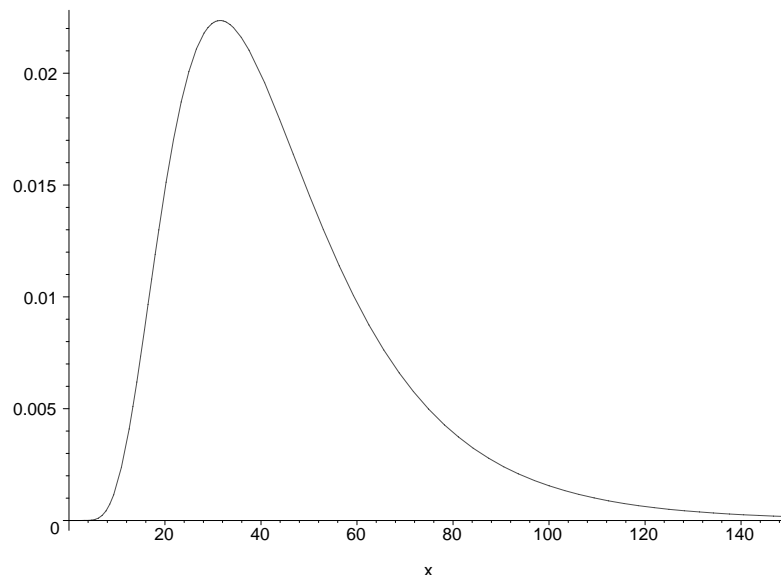
A.4 Lognormale verdeling met parameters μ en σ^2

Een s.v. X heeft een lognormale verdeling als $\ln(X)$ normaal verdeeld is (X neemt alleen positieve waarden aan; \ln staat voor de natuurlijke logaritme). De

lognormale verdeling is een scheve verdeling op $(0, \infty)$ die bijvoorbeeld bij inkomensverdelingen nogal eens gebruikt wordt. De parameters μ en σ^2 zijn *niet* de verwachting en variantie van X zelf, maar de verwachting en variantie van $\ln(X)$, die dus normaal verdeeld is. Wat de verwachting en variantie van X zelf betreft geldt $EX = e^{\mu + \frac{1}{2}\sigma^2}$ en $\text{var}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$.

In veel situaties waarin we de lognormale verdeling als model hanteren gaan we over op de nieuwe s.v. $Y = \ln(X)$, die dan dus $N(\mu, \sigma^2)$ verdeeld is. Vervolgens passen we hierop de theorie van de normale verdeling toe.

Voorbeeld A.4.1 Veronderstel dat in een bepaalde populatie het inkomen (in duizenden euro's) gemodelleerd kan worden met een lognormale verdeling met parameters $\mu = 3.7$ en $\sigma^2 = 0.25$. Deze kansdichtheid ziet er zo uit.



De verwachting is gelijk aan $e^{3.7+0.25/2}$. Merk op dat de top van de verdeling ligt bij 31.5.

Gevraagd is het percentage inkomens in deze populatie boven € 60 000. Dit wordt gegeven door $P(X > 60)$ met X een lognormale verdeling met parameters $\mu = 3.7$ en $\sigma^2 = 0.25$. Definieer $Y = \ln(X)$, dan is $Y \sim N(3.7, 0.25)$ en de gevraagde kans is $P(Y > \ln(60))$. Standardiseren geeft $P(X > 60) = 1 - \Phi\left(\frac{\ln(60)-3.7}{0.5}\right) = 1 - \Phi(0.789) = 0.215$. In deze populatie heeft 21.5% een inkomen boven € 60 000. \square