Les 6 Poisson processen

Als gebeurtenissen op willekeurige tijdstippen plaats vinden, kunnen we dit opvatten als een soort proces die de gebeurtenissen op een toevallige manier voortbrengt. Voorbeelden van dit soort processen zijn:

- klanten die een winkel binnen lopen,
- inkomende aanvragen bij een telefooncentrale,
- uitvallen van servers van een groot internet-bedrijf,
- emissie van een radioactief preparaat.

Een groot aantal van processen die we als *toevallig* beschouwen, laat zich door een paar heel eenvoudige regels karakteriseren. Als $N(t_1, t_2)$ het aantal gebeurtenissen van een proces in het tijdsinterval $[t_1, t_2]$ aangeeft, zou men bij een proces bijvoorbeeld het volgende kunnen eisen:

(1) De kansverdeling van $N(t, t+h)$ is onafhankelijk van t, d.w.z. de kans voor de gebeurtenissen is invariant onder een verschuiving in de tijd.

(2) Voor $t_1 < t_2 < t_3 < t_4$ zijn $N(t_1, t_2)$ en $N(t_3, t_4)$ onafhankelijk, d.w.z. de gebeurtenissen op niet overlappende intervallen zijn onafhankelijk.

(3) $P(N(t, t+h) = 1) = \lambda h + o(h)$ en $P(N(t, t+h) = 0) = 1 - \lambda h + o(h)$. Hierbij geeft $o(h)$ een term aan die voor $h \to 0$ sneller naar 0 gaat dan h, dus waarvoor geldt dat $\lim_{h\to 0} \frac{o(h)}{h} = 0$. Hieruit volgt dat voor kleine tijdsintervallen de kans op één gebeurtenis in het interval evenredig aan de lengte van het tijdsinterval is, want $\lim_{h\to 0} \frac{\lambda h + o(h)}{h} = \lambda$.

De eisen (1) en (2) betekenen dat de kansverdeling $N(t_1, t_2)$ alleen maar van de lengte $|t_2 - t_1|$ van het tijdsinterval afhangt en onafhankelijk van eerdere gebeurtenissen is. Omdat de kansverdeling $N(t_1, t_2)$ niet van het verleden of de *geschiedenis* van het proces afhangt, noemt men zo'n proces ook *geheugenloos*.

De parameter λ heet de *intensiteit* van het proces. Uit eis (3) kunnen we concluderen dat de kans op twee of meer gebeurtenissen in een tijdsinterval gegeven is door $P(N(t, t+h) \geq 2) = o(h)$ en voor $h \to 0$ gaat deze kans naar 0, want $\lim_{h\to 0} \frac{o(h)}{h} = 0$. Dit betekent in het bijzonder dat er nooit twee gebeurtenissen op hetzelfde tijdstip plaats vinden.

Een proces die aan deze eisen voldoet, noemen we een *Poisson-proces*, de naamgeving zal straks toegelicht worden.

Nemen we als model voor dit soort processen klanten die een winkel binnen lopen, dan kunnen we ons ook voorstellen dat de klanten in een rij staan voor dat ze geholpen worden. Daarom spreekt men hier ook van *wachtrijen*. We kunnen nu verschillende typen van vragen stellen, bijvoorbeeld:

- Wat is de kans dat er binnen 5 minuten twee klanten de winkel binnen lopen?

- Wat is de kans dat er meer dan 5 minuten niemand binnen komt?

De eerste vraag gaat over het aantal mensen die in een rij staan en we zullen zien dat we dit met behulp van de Poisson-verdeling kunnen bepalen. De tweede vraag is over de tussentijden tussen twee gebeurtenissen, en het aardige is dat we uit onze aannamen over onafhankelijkheid kunnen afleiden dat de tussentijden tussen de n-de en $(n+1)$-de gebeurtenis alle onafhankelijk van elkaar zijn en door een *exponentiële verdeling* beschreven worden. In het bijzonder vinden we hier een koppeling tussen de discrete verdeling van de gebeurtenissen en de continue verdeling van de tussentijden.

6.1 Tussentijden bij een Poisson-proces

Om de Poisson-processen te beschrijven kijken we eerst naar de kans dat er tot een tijdstip t helemaal geen gebeurtenis waargenomen wordt. We schrijven $P_0(t) := P(N(0,t) = 0)$ voor deze kans. Omdat de tijdsintervallen $[0,t]$ en $[t, t+h]$ niet overlappen, geldt volgens eis (2) dat

$$P_0(t+h) = P_0(t) \cdot P(N(t, t+h) = 0) = P_0(t)(1 - \lambda h).$$

Hierbij hebben we de $o(h)$ termen weggelaten, omdat we later de limiet $h \to 0$ bekijken, waarbij deze termen sowieso wegvallen. Hieruit volgt

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t)$$

en door hiervan de limiet $h \to 0$ te nemen, krijgen we aan de linkerkant de afgeleide van $P_0(t)$, dus

$$P_0'(t) = -\lambda P_0(t).$$

Zo'n soort vergelijking die een functie en hun afgeleide bevat, heet een *differentiaalvergelijking*. In het algemeen is het enigszins moeilijk om de oplossingen van differentiaalvergelijking te vinden, maar in ons geval is het zelfs mogelijk, alle functies $f(x)$ die aan de differentiaalvergelijking $f'(x) = -\lambda f(x)$ voldoen expliciet aan te geven:

Men gaat (door afleiden) snel na dat $f(x) = e^{-\lambda x}$ inderdaad een oplossing is. Als we nu nog een tweede functie $g(x)$ hebben, waarvoor ook geldt dat $g'(x) = -\lambda g(x)$, dan kunnen we de quotiënt $\frac{g(x)}{f(x)}$ van de twee functies bekijken. Voor de afgeleide van deze quotiënt geldt (met behulp van de quotiëntenregel):

$$\left(\frac{g(x)}{f(x)}\right)' = \frac{g'(x)f(x) - g(x)f'(x)}{g(x)^2} = \frac{-\lambda g(x)f(x) - g(x)(-\lambda f(x))}{g(x)^2} = \frac{0}{g(x)^2} = 0.$$

Maar als een functie afgeleide 0 heeft, is de functie zelf constant, d.w.z. $\frac{g(x)}{f(x)} = c$ en dus $g(x) = c \cdot f(x)$. De oplossingen van de differentiaalvergelijking zijn dus juist de veelvouden van $f(x)$.

We hebben op deze manier gezien dat $P_0(t)$ noodzakelijk van de vorm $P_0(t) = c \cdot e^{-\lambda t}$ is. Als we $t = 0$ invullen, kijken we naar de kans dat er geen gebeurtenis in het interval $[0,0]$ plaats vindt. Maar deze kans is 1 omdat het om een enkele punt gaat, dus is $P_0(0) = 1$ en dus $c = 1$.

De kans dat tot een tijdstip t geen enkel gebeurtenis waargenomen wordt is dus $e^{-\lambda t}$. Als we het tijdstip van de eerste waarneming T noemen, betekent dit dat $P(T > t) = e^{-\lambda t}$ en dus

$$P(T < t) = 1 - e^{-\lambda t}.$$

Dit betekent dat de kansverdeling voor het tijdsinterval tot de eerste waarneming een *exponentiële verdeling met parameter* λ is.

Herinnering: De exponentiële verdeling met parameter λ heeft dichtheidsfunctie $f(x) = \lambda e^{-\lambda x}$ en verdelingsfunctie $F(x) = 1 - e^{-\lambda x}$. Dit betekent dat voor een exponentieel verdeelde stochast X de kans op een uitkomst van hoogstens x gegeven is door $P(X \leq x) = 1 - e^{-\lambda x}$. De verwachtingswaarde van zo'n exponentieel verdeelde stochast X is $E(X) = \frac{1}{\lambda}$.

Onze veronderstelling over de onafhankelijkheid tegenover verschuivingen in de tijd zegt nu, dat we het tijdstip $t = 0$ ook op het tijdstip T van de eerste waarneming kunnen leggen. Dan kunnen we meteen concluderen dat de kansverdeling voor het tijdsinterval tussen de eerste en de tweede waarneming van een gebeurtenis ook een exponentiële verdeling met parameter λ is. Met hetzelfde argument vinden we, dat de kansverdeling voor het tijdsinterval tussen de n-de en de $(n+1)$-de waarneming dezelfde exponentiële verdeling met parameter λ is.

Dat de tussentijden tussen de verschillende gebeurtenissen onafhankelijk van elkaar zijn volgt uit de onafhankelijkheid voor niet overlappende intervallen

Merk op: Omdat de tussentijden bij een Poisson-proces met intensiteit λ exponentieel verdeeld met parameter λ zijn en dus verwachtingswaarde $\frac{1}{\lambda}$ hebben, kunnen we uit de kennis van de gemiddelde tussentijden de intensiteit van het proces bepalen. Als namelijk de tussentijden gemiddeld τ zijn, heeft het Poisson-proces de intensiteit $\frac{1}{\tau}$.

6.2 Aantallen gebeurtenissen bij een Poisson-proces

We gaan nu nog aantonen dat bij een Poisson-proces het aantal gebeurtenissen in een gegeven tijdsinterval door een *Poisson-verdeling* beschreven wordt.

We hebben al gezien dat $P(N(0,t) = 0) = e^{-\lambda t}$. Als afkorting schrijven we nu $P_1(t) := P(N(0,t) = 1)$ voor de kans dat er precies één gebeurtenis tot het tijdstip t plaats vindt. Als er in het tijdsinterval $[0, t+h]$ één waarneming is, is die of in het interval $[0,t]$ of in het interval $[t, t+h]$. Omdat deze twee intervallen niet overlappen, geldt volgens onze onafhankelijkheidseisen voor Poisson-processen dat

$$\begin{aligned} P_1(t+h) &= P_0(t) \cdot P(N(t,t+h) = 1) + P_1(t) \cdot P(N(t,t+h) = 0) \\ &= P_0(t) \cdot \lambda h + P_1(t) \cdot (1 - \lambda h). \end{aligned}$$

Hieruit volgt

$$\frac{P_1(t+h) - P_1(t)}{h} = \lambda P_0(t) - \lambda P_1(t)$$

en door hiervan de limiet $h \to 0$ te nemen en $P_0(t)$ van boven in te vullen, krijgen we de differentiaalvergelijking

$$P_1'(t) = \lambda e^{-\lambda t} - \lambda P_1(t).$$

Ook hier is het niet zo moeilijk om aan te tonen, dat $P_1(t)$ van de vorm $P_1(t) = \lambda t e^{-\lambda t} + c e^{-\lambda t}$ moet zijn, en uit $P_1(0) = 0$ volgt $c = 0$. Er geldt dus

$$P(N(0,t) = 1) = (\lambda t) e^{-\lambda t}.$$

We kunnen nu op een soortgelijke manier doorgaan om aan te tonen dat

$$P(N(0,t) = k) = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t}.$$

Stel we hebben dit voor $k-1$ al gezien, dan schrijven we net als boven $P_k(t) := P(N(0,t) = k)$ voor de kans op precies k gebeurtenissen tot het tijdstip t. Omdat we nooit twee gebeurtenissen in een klein interval h hebben, geldt:

$$\begin{aligned} P_k(t+h) &= P_{k-1}(t) \cdot P(N(t,t+h) = 1) + P_k(t) \cdot P(N(t,t+h) = 0) \\ &= P_{k-1}(t) \cdot \lambda h + P_k(t) \cdot (1 - \lambda h). \end{aligned}$$

Hieruit volgt

$$\frac{P_k(t+h) - P_k(t)}{h} = \lambda P_{k-1}(t) - \lambda P_k(t)$$

en door hiervan de limiet $h \to 0$ te nemen en $P_{k-1}(t)$ van boven in te vullen, krijgen we

$$P_k'(t) = \lambda \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} - \lambda P_k(t).$$

Hieruit volgt $\frac{d}{dt}(e^{\lambda t} P_k(t)) = (P_k'(t) + \lambda P_k(t)) e^{\lambda t} = \lambda \frac{(\lambda t)^{k-1}}{(k-1)!}$ en door integreren krijgen we $e^{\lambda t} P_k(t) = \frac{(\lambda t)^k}{k!} + c$. Er geldt dus $P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} + c e^{-\lambda t}$ en uit $P_k(0) = 0$ volgt weer $c = 0$.

Conclusie: Alles bij elkaar genomen hebben we aangetoond dat voor een Poisson-proces met intensiteit λ het aantal gebeurtenissen in het interval $[0,t]$ een Poisson-verdeling met parameter λt heeft, dus dat

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

Merk op: Een Poisson-verdeling met parameter λt heeft verwachtingswaarde λt. De intensiteit λ van een Poisson-proces is dus het gemiddelde aantal gebeurtenissen in het eenheidstijdsinterval $[0,1]$.

Voorbeeld 1: Klanten komen een winkel binnen volgens een Poisson-proces met intensiteit 3 (per uur). Elke klant blijft twintig minuten in de winkel. We willen de kans berekenen dat twee klanten elkaar ontmoeten. Hiervoor kijken we naar het tijdsinterval tussen twee klanten en we hebben de kans nodig, dat zo'n interval hoogstens 20 minuten is. Maar we weten dat de tussentijden exponentieel met parameter $\lambda = 3$ verdeelt zijn, dus is de kans op een interval $T < \frac{1}{3}uur$ gegeven door $P(T < \frac{1}{3}) = 1 - e^{-3\cdot\frac{1}{3}} = 1 - e^{-1} \approx 0.632$.

Voorbeeld 2: Een interactief systeem kan maximaal 15 transacties per seconde verwerken. In een spitsuur zijn er gemiddeld 10 transacties per seconde die volgens een Poisson-proces binnen komen. Wat is de kans dat het systeem tijdens een spitsuur overbelast raakt?

We hebben een intensiteit van $\lambda = 10$ (transacties per seconde) en willen de kans op meer dan 15 transacties in een tijdsinterval van een seconde bepalen. Dit is het complement van de kans op hoogstens 15 transacties en deze is

$$P(X \leq 15) = (1 + \frac{10}{1} + \frac{10^2}{2!} + \ldots + \frac{10^{15}}{15!})e^{-10} \approx 0.9513.$$

De gezochte kans is dus $1 - P(X \leq 15) \approx 4.87\%$.

We gaan nu eens omgekeerd uit van een proces die niet noodzakelijk een Poisson-proces is, maar waarvan we weten, dat de kansverdeling $P(N(0,t) = k)$ een Poisson-verdeling met parameter λt is. In dit geval is het eenvoudig om aan te tonen dat het tijdsinterval T tot de eerste waarneming exponentieel met parameter λ verdeeld is; in feite hebben we het argument boven al toegepast.

De kans dat T groter dan t is, is namelijk gelijk aan de kans dat in het interval $[0,t]$ geen waarneming ligt, en die is bij een Poisson-verdeling $e^{-\lambda t}$. We hebben dus $P(T < t) = 1 - e^{-\lambda t}$ en de verdeling van het tijdsinterval tot de eerste waarneming van een gebeurtenis is inderdaad een exponentiële verdeling met parameter λ.

We kunnen nog een verder aspect bekijken, dat laat zien dat een Poisson-proces inderdaad toevallige gebeurtenissen beschrijft. Hiervoor nemen we aan dat we weten dat er een gebeurtenis in het interval $[0,t]$ valt. We gaan nu de kansverdeling voor het tijdstip T bepalen, waarop de gebeurtenis plaats vindt. Voor de voorwaardelijke kans $P(T \leq x \mid N(0,t) = 1)$ dat T hoogstens x is, geldt $P(T \leq x \mid N(0,t) = 1) = \dfrac{P(T \leq x, N(x,t) = 0)}{P(N(0,t) = 1)}$ en de teller hiervan kunnen we opsplitsen, namelijk $P(T \leq x, N(x,t) = 0) = P(N(0,x) = 1, N(x,t) = 0) = P(N(0,x) = 1) \cdot P(N(x,t) = 0)$, omdat de twee tijdsintervallen niet overlappen. Hieruit volgt

$$\begin{aligned} P(T \leq x \mid N(0,t) = 1) &= \frac{P(N(0,x) = 1) \cdot P(N(x,t) = 0)}{P(N(0,t) = 1)} \\ &= \frac{\lambda x e^{-\lambda x} \cdot e^{-\lambda(t-x)}}{\lambda t e^{-\lambda t}} = \frac{x}{t}. \end{aligned}$$

en dit betekent dat het tijdstip T uniform op het interval $[0,t]$ verdeeld is. Als we weten dat er een gebeurtenis in het interval $[0,t]$ valt, hebben we dus geen verdere informatie over het tijdstip van de gebeurtenis, elk punt in het interval is even goed.

Samenvattend kunnen we zeggen, dat een Poisson-proces met intensiteit λ gekarakteriseerd is door een van de volgende eigenschappen:

(1) Het aantal gebeurtenissen in een tijdsinterval van lengte t is gegeven door een Poisson-verdeling met parameter λt.

(2) De tussentijden tussen de gebeurtenissen zijn onafhankelijk van elkaar en alle verdeelt volgens een exponentiële verdeling met parameter λ.

Belangrijke begrippen in deze les

- Poisson-proces
- intensiteit
- wachtrijen
- exponentiële verdeling

Opgaven

25. In een zeker gebied treden aardbevingen bij benadering op volgens een Poisson-proces met een gemiddelde van 2 aardbevingen per maand.

    (i) Bereken de kans dat er de komende twee maanden minstens drie aardbevingen optreden.

    (ii) Wat is de kans dat de eerstvolgende aardbeving minstens drie maanden op zich laat wachten?

26. Op een computer systeem komen aanvragen volgens een Poisson-proces binnen, gemiddeld 60 per uur. Bepaal de kansen voor de volgende tussentijden tussen twee op elkaar volgende aanvragen:

    (i) meer dan 4 minuten,

    (ii) minder dan 8 minuten,

    (iii) tussen 2 en 6 minuten.

27. Op een kantoor komen gemiddeld 12 gesprekken per uur binnen. Het aantal gesprekken dat per 10 minuten binnenkomt kan beschouwd worden als een stochast met Poisson-verdeling. Bereken de kans dat er

    (i) meer dan 3,

    (ii) hoogstens 4,

    (iii) meer dan 1 maar hoogstens 4

    klanten geen gehoor krijgen als de telefoniste gedurende 10 minuten afwezig is.

28. Een telefooncentrale kan per minuut maximaal 20 telefoongesprekken aan. Het aantal gesprekken per uur is een Poisson-verdeelde stochast met verwachtingswaarde 600. Bereken de kans dat de telefooncentrale gedurende een bepaalde minuut overbelast zal raken.