Waarskynlikheid 2 – Diskrete Modelle

Pieta van Deventer

In ’n vorige bydrae het ons die heel basiese beginsels van waarskynlikhede asook die teltegnieke, nl. permutasies en kombinasie wat nuttig te pas kan kom in die berekening van waarskynlikhede bespreek.[1] Daar is ook verwys na die volgende probleem: Ek het elektroniese seine wat uit 4 onafhanklike pulse elk bestaan. Indien met enige sein, 0, 1, 2, 3 of 4 van die pulse foutief kan deurgaan en die waarskynlikheid is 0.05 dat ’n puls in ’n sein foutief kan wees, wat is die waarskynlikheid dat die sein in sy totaliteit foutief of korrek sal deurgaan? Elke puls, nl. pulse 1, 2, 3 en 4 het ’n spesifieke funksie en is noodsaaklik om korrek deur te gaan vir die sein om in sy totaliteit foutloos te wees. Indien enige van die 4 pulse foutief is, is die sein foutief. ’n Verdere vraag sou kon wees: Wat is die waarskynlikheid dat 98 van die volgende 100 seine korrek sal versend? Onderskei baie duidelik op die verskil tussen ’n sein en die pulse waaruit die sein bestaan.

In die volgende tabel word die eerste drie kolomme se ontleding in die res van die kolomme uiteengesit. Beskou O as die simbool vir ’n puls wat nie foutief is nie, d.w.s Okay en dus reg versend is. ’n Foutiewe puls word deur $\overline{\mathrm{O}}$ aangedui. Gebruik X as die veranderlike vir die aantal foutiewe pulse in ’n enkele sein wat versend word. Dus kan X enige een van die waardes $0, 1, 2, 3$ of 4 wees. Beskou die ry met uitkoms $(\mathrm{O}, \mathrm{O}, \mathrm{O}, \mathrm{O})$ wat beteken al vier pulse was reg en geen foutiewe pulse is gestuur nie, d.w.s. $x = 0$. Al vier onafhanklike pulse is dus korrek met waarskynlikheid dat $P(X = 0)$. Dus is $P(\mathrm{O} \cap \mathrm{O} \cap \mathrm{O} \cap \mathrm{O}) = P(\mathrm{O}) \cdot P(\mathrm{O}) \cdot P(\mathrm{O}) \cdot P(\mathrm{O}) = 0.95 \times 0.95 \times 0.95 \times 0.95$ in die tabel soos aangedui. Ons beskou die pulse se uitkomstes as onafhanklik van mekaar, sodat die waarskynlikheid van die vier pulse se uitkomstes gesamentlik as ’n direkte produk van die afsonderlike waarskynlikhede uitgedruk kan word. Dus word die waarskynlikheid dat $x = 0$ aangedui in kolom een. In kolom 4 word aangedui dat daar 1 uitkoms/sein was waar $x = 0$. In die laaste 5 kolomme word daar dus telling gehou van die aantal kere wat $x = 0$ of 1 of 2 of 3 of 4 is. Die waarde van X dui op die aantal suksesse. Suksesse synde hier foutiewe pulse. Dit is duidelik dat ’n “mooi” uitkoms in statistiek nie noodwendig op ’n sukses dui nie. ’n Sukses word behaal wanneer enige gebeurtenis wat as ’n sukses gedefinieer is plaasvind.

Beskou die ry met uitkoms $(\mathrm{O}, \mathrm{O}, \mathrm{O}, \overline{\mathrm{O}})$ wat ons geriefshalwe as (0,0,0,1) uitdruk. Die ander uitkomstes word op ooreenkomstige wyse in die tabel aangedui. Hier is daar een foutiewe puls met drie korrekte pulse met waarskynlikheid soos aangedui. Dus word in kolom 5 aangedui dat daar 1 uitkoms was waar $x = 1$. So kan ’n mens van een potensiële seinuitkomste tot die volgende potensiële seinuitkomste beweeg tot by die laaste een waar al vier pulse foutief is, d.w.s. waar $x = 4$. Let op dat die “volgorde” waarin die pulse in ’n sein foutief is, geen verskil aan die betrokke waarskynlikheid maak nie. As voorbeeld:

[1] https://www.litnet.co.za/wp-content/uploads/2020/06/2020-06-27-Teltegnieke-en-waarskynlikhede-1.pdf

$P(X=1)=$ die waarskynlikheid dat óf die eerste óf die tweede óf die derde óf die vierde puls foutief is

$$
\begin{aligned}
&= P(\bar{O},O,O,O)+P(O,\bar{O},O,O)+P(O,O,\bar{O},O)+P(O,O,O,\bar{O})\\
&= 0.05\cdot 0.95\cdot 0.95\cdot 0.95+0.95\cdot 0.05\cdot 0.95\cdot 0.95\\
&\quad +0.95\cdot 0.95\cdot 0.05\cdot 0.95+0.95\cdot 0.95\cdot 0.95\cdot 0.05\\
&= 4\times 0.05\times 0.95^3\\
&= \binom{4}{1}\times 0.05\times 0.95^3\\
&= 0.171475
\end{aligned}
$$

Bereken die totale van kolomme 4-8. Dit gee vir ons dat X op een manier 'n 0 kan wees, op vier maniere 'n 1, op ses maniere 'n 2, op vier maniere 'n 3 en op een manier 'n 4 kan wees.

1	2	3		4	5	6	7	8
$P(X=x)$	Resultaat van 4 pulse	x		Tellinghouerkolomme vir aantal foutiewe pulse				
				0	1	2	3	4
$0.95^4\times0.05^0$	0 0 0 0	0		1				
$0.95^3\times0.05^1$	0 0 0 1	1			1			
$0.95^3\times0.05^1$	0 0 1 0	1			1			
$0.95^2\times0.05^2$	0 0 1 1	2				1		
$0.95^3\times0.05^1$	0 1 0 0	1			1			
$0.95^2\times0.05^2$	0 1 0 1	2				1		
$0.95^2\times0.05^2$	0 1 1 0	2				1		
0.95×0.05^2	0 1 1 1	3					1	
$0.95^3\times0.05^1$	1 0 0 0	1			1			
$0.95^2\times0.05^2$	1 0 0 1	2				1		
$0.95^2\times0.05^2$	1 0 1 0	2				1		
$0.95^1\times0.05^3$	1 0 1 1	3					1	
$0.95^2\times0.05^2$	1 1 0 0	2				1		
$0.95^1\times0.05^3$	1 1 0 1	3					1	
$0.95^1\times0.05^3$	1 1 1 0	3					1	
$0.95^0\times0.05^4$	1 1 1 1	4						1
				1	4	6	4	1
				$=\binom{4}{0}$	$=\binom{4}{1}$	$=\binom{4}{2}$	$=\binom{4}{3}$	$=\binom{4}{4}$

Die heel laaste ry wys nou hoe ons die syfers 1, 4, 6, 4, 1 uit eerste beginsels, d.w.s met behulp van ons kennis van kombinasies, kan aflei. Dit beantwoord die vraag: Op hoeveel maniere kan daar x items uit 4 foutief wees, ongeag die volgorde vir $x=0,1,2,3$ of 4? Dis dan die kombinasies soos in die laaste ry uiteengesit.

Hierdie tabel kan nou soos volg opgesom word:

$$
\begin{aligned}
P(X=x)=f(x)&=\binom{4}{x}0.05^{x}0.95^{4-x} \quad && x=0,1,2,3,4 \\
&=0 && \text{andersins}
\end{aligned}
$$

Hierdie uitdrukking word gelees as: Die waarskynlikheid dat die eienskap X die waarde x sal aanneem waar x slegs die waardes $0,\ldots,4$ kan aanneem. Weereens as voorbeeld:

$$
\begin{aligned}
P(X=3)=f(3)&=\binom{4}{3}0.05^{3}0.95^{4-3} \\
&=0.000\ 475
\end{aligned}
$$

Dus is die waarskynlikheid dat 3 pulse in ’n sein foutief kan wees gelyk aan 0.000 475.

Indien geen pulse defektief is nie gaan die sein korrek deur en die waarskynlikheid daarvoor is dus

$$
\begin{aligned}
P(X=0)=f(0)&=\binom{4}{0}0.05^{0}0.95^{4-0} \\
&=0.814\ 506\ 25
\end{aligned}
$$

Onthou hierde waarde. Dit kom nou-nou effe afgerond weer voor.

As verdere voorbeeld

$$
\begin{aligned}
P(X\le 2)&=f(0)+f(1)+f(2) \\
&=\binom{4}{0}0.05^{0}0.95^{4-0}+\binom{4}{1}0.05^{1}0.95^{4-1}+\binom{4}{2}0.05^{2}0.95^{4-2} \\
&=0.814\ 506\ 25+0.171\ 475+0.013\ 537\ 5 \\
&=\ 0.999\ 518\ 8
\end{aligned}
$$

Die waarskynlikheid dat ’n enkele sein defektief is, is die waarskynlikheid 1 of meer pulse foutief is, of anders gestel dat

$$
\begin{aligned}
P(X\ge 1)&=f(1)+f(2)+f(3)+f(4) \\
&=1-f(0)=1-P(X=0) \\
&=1-0.814\ 506\ 25 \\
&=0.185\ 493\ 75
\end{aligned}
$$

In die algemeen is

$$
\begin{aligned}
P(X=x)=f(x)&=\binom{n}{x}p^{x}(1-p)^{n-x} \quad && x=0,1,2,3,\ldots,n \\
&=0 && \text{andersins}
\end{aligned}
$$

So ’n funksie staan bekend as ’n waarskynlikheidsmassaverdelingfunksie of kortweg waarskynlikheidsmassafunksie. Soms word dit ook ’n verdelingsfunksie genoem, maar dis nie korrek nie. (Die term "verdelingsfunksie" self verwys na die kumulatiewe waarskynlikheid,

nl.

$$P(X \le x) = F(x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y} \quad \text{vir x} = 0,1,2,\ldots,n$$
$$= 0 \quad \text{andersins}$$

Verwys $P(X \le 2)$ hierbo. Let op die bokas gebruikskonvensie van F vir die verdelingsfunksie.)

Hierdie diskrete (diskreet omdat X slegs diskrete waardes kan aannneem) waarskynlik-heidsmodel staan bekend as die binomiaalwaarskynlikheidsmodel.

Beskou die algemene uitdrukking hierbo.

- Hierdie funksie staan bekend as die binomiaalwaarskynlikheidsmassaverdelingsfunksie of kortweg die binomiaalverdeling. Let op dat die wettige waardes van X baie duidelik gespesifiseer móét word en deel van die funksie is. Sonder hierdie spesifikasie is die waarskynlikheidsmassafunksie van geen waarde nie.
- n is die vaste aantal herhalings van 'n eksperiment.
- p is die vaste waarskynlikheid van sukses van herhaling tot herhaling.
- n en p staan as die parameters van die binomiaalverdeling bekend. As die parameters van enige waarskynlikheidsverdeling bekend is, kan enige ter sake waarskynlikheid bereken word.

Kom ons neem die probleem verder, as nog 'n voorbeeld. Wat is die waarskynlikheid dat uit 100 seine ten minste 98 seine foutloos is, of anders gestel, nie meer as 2 foutief is nie? In hierdie geval is $n = 100$ en $p \doteq 0.8145$, en X dui op die aantal foutlose seine uit 100 seine gestuur en $p = 0.8145$ is die waarskynlikheid dat 'n enkele sein korrek versend word. Nou is

$$P(X = x) = f(x) = \binom{100}{x} 0.8145^x (1-0.8145)^{100-x} \quad x = 0,1,2,3,\ldots,100$$
$$= 0 \quad \text{andersins}$$

Dit volg dus dat

$$\begin{aligned} P(X \ge 98) &= P(X = 98) + P(X = 99) + P(X = 100) \\ &= \sum_{x=98}^{100} \binom{100}{x} 0.8145^x (1-0.8145)^{100-x} \quad \text{of} \quad 1\text{-}P(X \le 97) = F(97) \\ &= 3.15224E\text{-}07 + 2.79616E\text{-}08 + 1.22775E\text{-}09 \\ &= 3.44414E\text{-}07 = 0.000\ 000\ 314 \end{aligned}$$

Dis duidelik dat die 0.05 foutwaarskynklikheid van 'n enkele puls hopeloos te groot is. Die sisteem sal verbeter moet word.

Hiermee het ons ’n waarskynlikheidsmodel ontwikkel. Wat meer is, indien u die vorige artikel gelees het, is u nou al met ’n verskeidenheid waarskynlikheidsmodelle bekend, sonder dat u dit eens agtergekom het en weet u reeds meer as menige ander mense oor hierdie tipe modelle.

- Die diskrete uniforme waarskynlikheidsmodel waar alle waarskynlikhede dieselfde (uniform) is:

$$P(X=x)=f(x)=\frac{1}{n} \quad x=1,2,3,\ldots,n$$
$$=0 \quad \text{vir enige ander waarde van } x$$

Stel die enkele parameter $n=6$ en u het die uitkomstes van ’n onsydige (eerlike) dobbelsteen gemodeleer.

- Die Bernoulli-waarskynlikheidsverdeling:

$$P(X=x)=f(x)=p^{1-x}(1-p)^{x} \quad x=0 \text{ of } 1, \text{ ook geskryf as } x=0,1$$
$$=0 \quad \text{vir enige ander waarde van } x$$

Stel die parameters van die binomiaalverdeling hierbo as $n=1$ en $p=p$. Dit beskryf die waarskynlikheidverdeling van ’n enkele eksperiment met waarskynlikheid p vir ’n suksesvolle uikoms. As $x=0$ volg dat $P(X=0)=p$ en op dieselfde wyse, as $x=1$ volg dat $P(X=1)=1-p$. Die binomiaalverdeling verskaf die waarskynlikhede vir die som van die aantal suksesvolle uitkomstes van n Bernoulli-eksperimente.

- Die hipergeometriese waarskynlikheidsverdeling

$$P(X=x)=f(x)=\begin{cases}\dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & x=0,1,\ldots,\min(K;n)\\ 0 & \text{andersins}\end{cases}$$

Hierdie diskrete waarskynlikheidverdeling is die lottoprobleem van die vorige artikel.

Hier is daar drie parameters, nl. N wat die totale aantal beskikbare items om uit te kies aandui. K dui op die aantal suksesvolle items beskikbaar in die populasie van N items en n is die aantal items wat gekies word. Soos gewoonlik dui X op die aantal suksesse waarop gehoop word. Die bogrens vir X hierbo is tegnies nie 100% korrek nie, maar werk baie lekker in die praktyk en is baie makliker om te interpreteer as die teoretiese bogrens wat gewoonlik daarmee geassosieer word.

- Die Poisson-waarskynlikheidsverdeling

  Beskou nou die binomiaalverdeling waar n baie groot is en p baie klein is. Dink aan die onsydige dobbelsteen. Die verwagte aantal sesse in 600 gooie van die dobbelsteen is …? Ja, u is reg. Dis ongeveer 100 sesse oftewel 100 suksesse. U het dit verkry deur die berekening van $100\times\frac{1}{6}=n\times p$ wat as die verwagte waarde van die binomiaalverdeling met $n=600$ en $p=\frac{1}{6}$ bekend staan. Stel nou $\lambda=np,$ of dan $p=\frac{\lambda}{n}$ in die binomiaalverdeling se funksie. Dan is dit nie moeilik nie om te bewys dat

$$\begin{aligned} g(x) &= \lim_{x\to\infty} P(X=x) = \lim_{x\to\infty} f(x) \\ &= \lim_{n\to\infty}\binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}. \\ \therefore g(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x=0,1,2,3,\ldots \\ &= 0 \qquad \text{andersins} \end{aligned}$$

Hierdie waarskynlikheidsmassaverdelingsfunksie staan bekend as die Poisson-verdeling in kort. Die een voordeel van die Poisson-verdeling is dat ’n mens net nodig het om te weet wat die **verwagte** of **gemiddelde** koers van voorkoms van die gebeurtenis ter sprake in ’n gespesifiseerde kort “tydperk” is, wat dan my waarde van λ is. Tydperk moet hier in ’n breë sin verstaan word. Verwys die voorbeeld wat volg. Dit kan dan ook onder die gegewe voorwaardes as ’n benadering vir die binomiaalverdeling gebruik word met $\lambda=np.$ Om voorsiening te maak vir verskillende “tyds”-eenhede word die Poisson-verdeling meer algemeen as volg gedefinieer – en dis hoe dit eintlik oorspronklik in eie reg ontstaan het – nie as ’n benadering vir die binomiaalverdeling nie:

$$\begin{aligned} g(x) &= \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x=0,1,2,3,\ldots \\ &= 0 \qquad \text{andersins} \end{aligned}$$

Die gemiddelde aantal ernstige slaggate per 5 km op ’n sekere pad tussen A en B word beraam op 3. Wat is die waarskynlikheid dat daar oor ’n ewekansige gekose stuk van die pad oor ’n afstand van 7 kilometer meer as 6 slaggate sal wees?

Let nou op dat $\lambda=3$ en die eenhede van t is in 5 km per eenhede gegee. Nou volg dat $\lambda t=3\cdot\frac{7}{5}=4.2$ slaggate gemiddeld per 7 kilometer verwag word.

$$\begin{aligned} P(X>6 \text{ oor } 7\text{ km}) &= 1-P(X\le 6 \text{ oor } 7\text{ km}) \\ &= 1-\left(P(X=0)+P(X=1)+P(X=1)+\ldots+P(X=6)\right) \\ &= 1-\sum_{x=0}^{6}\frac{(4.5)^x e^{-4.5}}{x!} \\ &= 0.0111+0.0500+0.1125+0.1687+0.1898+0.1708+0.1281 \\ &= 0.8314 \end{aligned}$$

Die antwoord kan natuurlik maklik m.b.v. Excel of menige ander sagteware verkry word.

Daar is nog ’n menigte ander diskrete waarskynlikheidsverdelings waaronder die geometriese/meetkundige asook die negatief binomiaal waarskynlikheidsverdeling ens. wat in die praktyk gebruik word. Benewens die diskrete waarskynlikheidsverdelings is daar dan ook die kontinue waarskynlikheidsdigtheidsfunksies waarmee hierdie kort reeks oor waarskynlikhede by ’n volgende aanbieding dan voorlopig afgesluit sal word.

Die verskil tussen diskrete en kontinue verdelings is tweërlei, nl. diskrete veranderlikes vs kontinue veranderlikes, maar net so belangrik is die feit dat die diskrete verdelings baie spesifieke scenarios aanspreek en dus beperkend is teenoor die kontinue verdelings wat oor ’n baie wyer spektrum toepassings het. Hierbenewens kan gewoonlik onder redelike matige voorwaardes die meeste diskrete verdelings deur kontinue verdelings benader word. ’n Interessante verdere eienskap van die twee tipes is dat daar dikwels ooreenkomstige verdelings is, nl. diskreet versus kontinu. Dit is wat statistiek so onontbeerlik in vandag se navorsing maak. Dan het ons nog nie oor regressie en vele ander onderwerpe gepraat nie. Bitter min navorsing kan vandag sonder ’n goeie skeut statistiek gedoen word.