

Statistiek voor A.I.

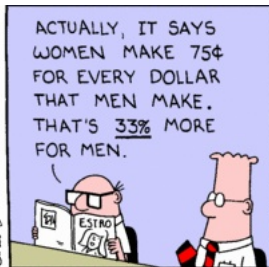
College 5

Dinsdag 25 September 2012

2 Deductieve statistiek

Kansrekening

Percentages



Vragen: blikkie



Kinderen worden slanker als ze anderhalf jaar lang limonade met zoetstof drinken in plaats van met suiker. Dat blijkt uit onderzoek van de VU bij ruim 600 Nederlandse kinderen van 5 tot 11 jaar.

Het onderzoek verscheen op 21 September 2012 in The New England Journal of Medicine. Daar verscheen tegelijkertijd een Amerikaanse studie met een vergelijkbaar resultaat.

Samen vormen de studies een krachtig bewijs dat het vermijden van suikerhoudende dranken helpt tegen overgewicht.

Vragen: zwendel

Een beurshandelaar ontvangt zeven weken lang elke Maandag een brief die voorspelt of de koers van de euro die week zal stijgen of dalen, en elke week blijkt die voorspelling correct.

Op de achtste Maandag wordt hem meegedeeld dat hij een voorspelling van de koers van de euro voor de komende week kan kopen. Moet hij hierop ingaan?

Vragen: missie

Uit een leger van 200 soldaten worden er willekeurig 3 gekozen die op een geheime missie zullen gaan. Wat is de kans dat de jongste soldaat in het leger deel van die missie zal uitmaken?

De missie heeft 1 leider. Wat is de kans dat de jongste soldaat in het leger de leider van die missie wordt?

Vandaag:

- Continue stochasten
- Cumulatieve distributiefuncties
- Permutaties en combinaties.

Continue stochasten

Antwoord op een vraag: lengte

Voorbeeld Een lineaal wordt op een willekeurig punt in tweeën gebroken. Wat is de kans dat tenminste één van de delen kleiner is dan een kwart van de lengte?

Zij X de continue stochast die de lengte van het linker deel is. X heeft een uniforme verdeling, dat wil zeggen dat de kansdichtheid constant is. De uitkomstenruimte is het interval $(0, l)$, waarbij l de lengte van de lineaal is.

De kansdichtheid is

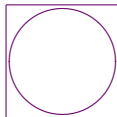
$$f(x) : (0, l) \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = \frac{1}{l}.$$

De gevraagde kans is

$$P(X < \frac{l}{4} \text{ of } X > l - \frac{l}{4}) = \int_0^{\frac{l}{4}} \frac{1}{l} dx + \int_{l-\frac{l}{4}}^l \frac{1}{l} dx = 2 \cdot \int_0^{\frac{l}{4}} \frac{1}{l} dx = 2 \cdot \frac{l}{4l} = \frac{1}{2}.$$

Met behulp van continue stochasten kunnen grootheden geschat worden.

Voorbeeld Een computer produceert willekeurige getallen in $[0,1]$. Een paar (x, y) van twee dergelijke getallen kan opgevat worden als een punt in een rechthoek met zijdes van lengte 1:



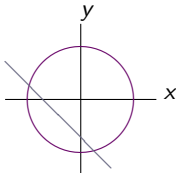
De straal van de cirkel is $\frac{1}{2}$. Dus de kans dat een punt in de cirkel valt is

$$\frac{\text{oppervlakte cirkel}}{\text{oppervlakte rechthoek}} = \frac{\pi}{4}.$$

Als de computer veel willekeurige paren (x, y) genereert, dan is het percentage dat in de cirkel valt een benadering van $\frac{\pi}{4}$. Zo kan via de kansrekening π benaderd worden.

Het begrip “willekeurig” hangt soms af van de manier waarop we de uitkomstenruimte representeren.

Voorbeeld Een cirkel in het zand, waaroverheen een tak willekeurig valt. De tak is zo lang dat die altijd de cirkel raakt of op twee punten snijdt. Wat is de kans dat de lengte l van de koorde (dat deel van de tak dat binnen de cirkel valt) groter dan $\sqrt{3}$ is?



Het antwoord hangt af van de manier waarop de koorde gegeven wordt! De volgende twee representaties geven een verschillend antwoord:

- De coördinaten van het middelpunt van de koorde. Dit geeft $P(X > \sqrt{3}) = \frac{1}{4}$.
- De poolcoördinaten van het middelpunt van de koorde. Dit geeft $P(X > \sqrt{3}) = \frac{1}{2}$.

Cumulatieve distributiefuncties

Definitie De *cumulatieve distributiefunctie* (*verdelingsfunctie*) van een stochast X is de functie $F : \mathbb{R} \rightarrow [0, 1]$ zodat

$$F(r) = P(X \leq r).$$

Indien X discreet is met waarden $X_1, X_2, \dots$, dan geldt

$$F(r) = \sum_{\{i | X_i \leq r\}} P(X = X_i).$$

Indien X continu is met kansdichtheid f , dan geldt

$$F(r) = \int_{-\infty}^r f(x) dx \quad \text{en} \quad F'(x) = f(x) \quad (f \text{ is de afgeleide van } F).$$

Merk op: $f(r)$ is geen kans, maar $F(r)$ is dat wel.

F is een niet-dalende functie: als $r \leq s$, dan $F(r) \leq F(s)$.

Voorbeeld Zij X de continue stochast bij het willekeurig kiezen van een punt in $[0,1]$, dus met kansdichtheid:

$$f(x) : [0, 1] \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = 1.$$

De cumulatieve distributiefunctie van X is

$$F(r) = P(X \leq r) = \int_0^r 1 dx = \int_0^r 1 dx = r.$$

Stelling De cumulatieve distributiefunctie van een uniform verdeelde stochast is linear.

Voorbeeld Een beurshandelaar twijfelt of hij een zeker product zal kopen. Als de kans groter dan 0.9 is dat de waarde van het product morgen niet meer dan €2 is, dan wil hij het product niet kopen.

Zij X de stochast behorende bij de waarde van het product morgen. Op grond van een statistisch economisch model is bekend dat X kansdichtheid f heeft:

$$f : [0, \infty) \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = e^{-x}$$

Dat geeft

$$F(2) = P(X \leq 2) = \int_0^2 f(x) = -e^{-x} \Big|_0^2 \approx -0.13 + 1 = 0.87.$$

Dus de handelaar koopt het product.

Soms is het makkelijker direct met de cumulatieve distributiefunctie te werken dan met de kansdichtheid.

Voorbeeld Er wordt een willekeurig getal X in $[0,1]$ gekozen. A betaalt aan B €1 als $X^2 \leq 0.6$, en B aan A €3 als $X^2 \geq 0.4$. Is de verwachte winst van beide gelijk?

De cumulatieve distributiefunctie van $Y = X^2$ wordt gegeven door:

$$F(r) = P(Y \leq r) = P(X^2 \leq r) = P(X \leq \sqrt{r}) = \sqrt{r}.$$

Daaruit volgt:

$$F(x) = \begin{cases} 0 & \text{als } x \leq 0 \\ \sqrt{x} & \text{als } 0 \leq x \leq 1 \\ 1 & \text{als } x > 1. \end{cases}$$

De kans dat A moet betalen is $P(Y \leq 0.6) = F(0.6) = \sqrt{0.6} = 0.77$. De verwachte winst van A is $3 \cdot 0.23 - 0.77 = -0.08$. Geen eerlijk spel dus.

Soms is het makkelijker de kansdichtheid m.b.v. de cumulatieve distributiefunctie te berekenen dan andersom.

Voorbeeld Er wordt een willekeurig getal X in $[0,1]$ gekozen. A betaalt aan B €1 als $X^2 \leq 0.6$, en B aan A €3 als $X^2 \geq 0.4$.

De cumulatieve distributiefunctie van $Y = X^2$ wordt gegeven door:

$$F(x) = \begin{cases} 0 & \text{als } x \leq 0 \\ \sqrt{x} & \text{als } 0 \leq x \leq 1 \\ 1 & \text{als } x > 1. \end{cases}$$

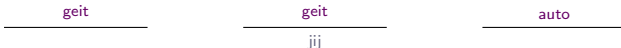
Uit $F'(x) = f(x)$ volgt dat de kansdichtheid van Y de volgende is:

$$f(x) : [0, 1] \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = \frac{1}{2\sqrt{x}}.$$

Permutaties en combinaties

Antwoord op een vraag: Monty Hall

Achter één van drie gesloten deuren staat een auto, achter de andere twee een geit. Jij gaat voor een deur staan. Bijv:



De quizmaster Monty Hall opent een van de twee deuren waar jij niet voor staat en waarachter een geit staat:



Jij mag blijven staan of voor de andere gesloten deur gaan staan. Vervolgens win je dat wat achter jouw deur staat. Is het beter altijd van deur te veranderen (indien je geen geit wilt)?

Het is beter om van deur te veranderen. Bij de *verander-strategie* is de kans dat je een auto wint groter dan bij de *verander-niet-strategie*. Om dit in te zien helpt het om een boomdiagram te tekenen, zie Voorbeeld 4.6 op bladzijde 136 in het boek.

Op hoeveel mogelijke manieren kan een gebeurtenis plaatsvinden?

ofwel

Uit hoeveel elementen bestaat de (eindige) uitkomstenruimte?

Definitie Gegeven een natuurlijk getal $n > 0$, dan is n -*faculteit* het getal

$$n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1,$$

dat wordt genoteerd als $n!$.

Voor $n = 0$ geldt $n! = 0! = 1$.

Voorbeeld

- Het aantal 5-letterige woorden die alleen letters in $\{a, b, c, d, e, f, g\}$ bevatten is $7^5 = 16807$.
- Het aantal 5-letterige woorden die alleen letters in $\{a, b, c, d, e, f, g\}$ bevatten en waarin geen letter meerdere keren voorkomt is $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = \frac{7!}{2!} = \frac{7!}{(7-5)!} = 2520$.
- Het aantal manieren waarop we 5 letters uit $\{a, b, c, d, e, f, g\}$ kunnen kiezen is $\frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{5!} = \frac{7!}{(7-5)!5!} = 21$.

Voorbeeld Beschouw een organisatie die uit 20 mensen bestaat.

- Het aantal manieren om alle leden een verschillend nummer ≤ 20 toe te kennen is $20! > 10^{18}$.
- Het aantal manieren om een commissie bestaande uit een voorzitter, secretaris en penningmeester te kiezen is
$$20 \cdot 19 \cdot 18 = \frac{20!}{17!} = \frac{20!}{(20-3)!} = 6840.$$
- Het aantal manieren om een commissie van 3 mensen te kiezen is
$$\frac{20 \cdot 19 \cdot 18}{3!} = \frac{20!}{17!3!} = \frac{20!}{(20-3)!3!} = 1140.$$

Voorbeeld Beschouw 7 mensen.

- Geordend: Het aantal manieren waarop die een rij kunnen vormen is $7! = 5040$.
- Geordend: Het aantal manieren waarop een rij bestaande uit van 4 van de 7 mensen gevormd kan worden is $7 \cdot 6 \cdot 5 \cdot 4 = \frac{7!}{3!} = \frac{7!}{(7-4)!} = 840$.
- Ongeordend: Het aantal manieren om een verzameling van 4 mensen uit de 7 te kiezen is $\frac{7 \cdot 6 \cdot 5 \cdot 4}{4!} = \frac{7!}{3!4!} = \frac{7!}{(7-4)!4!} = 35$.

Stelling Het aantal *permutaties* van n elementen: Het aantal manieren om n elementen te ordenen is n faculteit:

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 = n!$$

Stelling Het aantal manieren om uit n elementen rijtjes (geordend) van lengte k te vormen is

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+2) \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Stelling Het aantal *combinaties* van k elementen uit n element: Het aantal manieren om verzamelingen (ongeordend) van k elementen uit n elementen te kiezen is n boven k :

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+2) \cdot (n-k+1)}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}.$$

De getallen $\binom{n}{k}$ heten *binomiaalcoëfficiënten* (*binomial coefficients*).

Definitie Voor elke $n \in \mathbb{N}$ geldt $\binom{n}{0} = \binom{n}{n} = 1$.

Voorbeeld Binomiaalcoëfficiënten:

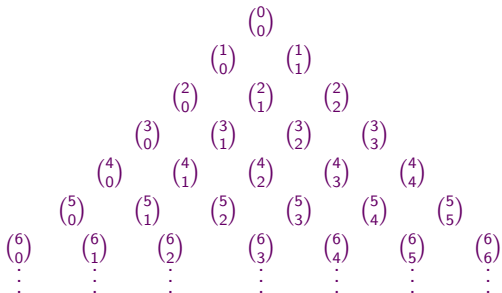
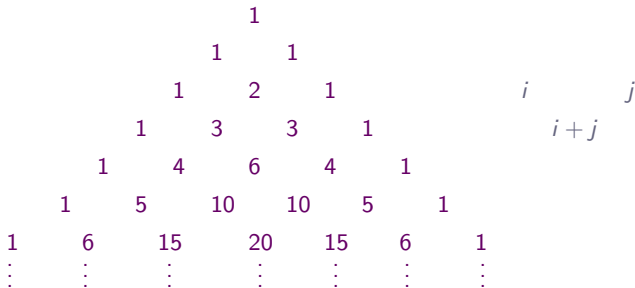
k	0	1	2	3	4	5	6
n							
5	1	5	10	10	5	1	
10	1	10	45	120	210	252	210
20	1	20	190	1140	4845	15504	38760
100	1	100	4950	$1.6 \cdot 10^5$	$3.9 \cdot 10$	$7.5 \cdot 10^7$	$1.1 \cdot 10^9$

Stelling (Stirlings formule)

Voor grote n geldt dat $n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$. Ofwel:

$$\lim_{n \rightarrow \infty} \frac{\left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}}{n!} = 1.$$

Permutaties en combinaties: Pascals driehoek en binomiaalcoëfficiënten



Voorbeeld Beschouw een vereniging met 15 leden. In de volgende vragen is het zo dat niemand lid is van meer dan één commissie.

- Het aantal manieren om 1 commissie van 3 en 1 commissie van 5 leden te kiezen is $\binom{15}{3} \binom{12}{5} = \binom{15}{5} \binom{10}{3}$.
- Het aantal manieren om 2 commissies van 3 leden te kiezen is $\frac{\binom{15}{3} \binom{12}{3}}{2} = \frac{\binom{15}{6} \binom{6}{3}}{2}$.
- Het aantal manieren om 2 commissies te kiezen, beide bestaande uit een voorzitter, secretaris en penningmeester, is $\frac{\frac{15!}{(15-3)!} \frac{12!}{(12-3)!}}{2}$.
- Het aantal manieren om een feestcommissie en een liefdadigheidscommissie, beide van 3 leden, te kiezen is $\binom{15}{3} \binom{12}{3} = \binom{15}{6} \binom{6}{3}$.
- Het aantal manieren om een feestcommissie en een liefdadigheidscommissie te kiezen, beide bestaande uit een voorzitter, secretaris en penningmeester, is $\frac{15!}{(15-6)!}$.
- Het aantal manieren om een feestcommissie en een liefdadigheidscommissie te kiezen, waarbij de feestcommissie bestaat uit een voorzitter, secretaris en penningmeester, en de liefdadighediscommissie uit 3 leden, is $\frac{15!}{(15-3)!} \binom{12}{3} = \binom{15}{3} \frac{12!}{(12-3)!}$.

Antwoord op een vraag: missie

Uit een leger van 200 soldaten worden er willekeurig 3 gekozen die op een geheime missie zullen gaan. Wat is de kans dat de jongste soldaat in het leger deel van die missie zal uitmaken?

$$\frac{\binom{199}{2}}{\binom{200}{3}}$$

De missie heeft 1 leider. Wat is de kans dat de jongste soldaat in het leger de leider van die missie wordt?

$$\frac{\binom{199}{2}}{200 \binom{199}{2}}$$

Voorbeeld Stel dat de loten in een loterij bestaan uit 7 cijfers in $\{1, \dots, 9\}$.

- De kans op een specifiek lot is $\frac{1}{9^7} \approx 10^{-7}$.
- De kans op een lot dat precies de cijfers 2, ..., 5 bevat is $\frac{4^7}{9^7} \approx 0.003$.
- De kans op een lot bestaande uit verschillende cijfers is $\frac{9!}{2!9^7} = \frac{9!}{(9-7)!9^7} \approx 0.038$.
- De kans op een lot waarin 1 cijfer precies 2 maal voorkomt en alle andere cijfers verschillend zijn is

$$\frac{9 \cdot \binom{7}{2} \cdot \frac{8!}{3!}}{9^7} \approx 0.266.$$

- De kans op een lot waarin 1 cijfer precies 3 maal voorkomt en alle andere cijfers verschillend zijn is

$$\frac{9 \cdot \binom{7}{3} \cdot \frac{8!}{4!}}{9^7} \approx 0.066.$$

Voorbeeld Stel dat de loten in een loterij bestaan uit 7 cijfers in $\{1, \dots, 9\}$ en €25 kosten. De loterij bevat 1 prijs van €10.000.025, die elke maand wordt uitgereikt. Mevrouw A. koopt 40 jaar lang elke maand 1 lot.

- De kans op de prijs is $\frac{1}{9^7} = 0.00000021$.
- De kans dat A. nooit wint is: $0.99999979^{40 \cdot 12} \approx 0.99$.
- De kans dat A. precies 1 maal wint is:
 $40 \cdot 12 \cdot 0.99999979^{40 \cdot 12 - 1} \cdot 0.00000021 \approx 0.00009$.
- Haar verwachte winst is:
 $\text{€}12 \cdot 40 \cdot (0.00000021 \cdot 10.000.000 - 0.99999979 \cdot 25) \approx \text{€}-10.992$.

Antwoord op een vraag: zwendel

Een beurshandelaar ontvangt zeven weken lang elke Maandag een brief die voorspelt of de koers van de euro die week zal stijgen of dalen, en elke week blijkt die voorspelling correct. Op de achtste Maandag wordt hem meegedeeld dat hij een voorspelling van de koers van de euro voor de komende week kan kopen. Moet hij hierop ingaan?

Wanneer de voorspelling het resultaat is van de worp van een munt, is de kans op zeven correcte voorspellingen $\frac{1}{2}^7 = 0.008$, zeer klein.

Echter, stel dat het bedrijf in de eerste week aan $2^7 = 128$ mensen een brief stuurt, waarbij in de helft een stijging en in de andere helft een daling wordt voorspeld. De tweede Maandag stuurt het bedrijf aan de 64 mensen voor wie de eerste voorspelling correct was, weer een voorspelling, op dezelfde wijze als op de eerste Maandag, enz. Aan het einde van zeven weken is er één persoon voor wie alle zeven voorspellingen correct zijn.

Niet doen dus.

Zie op You Tube: Derren Brown - the System.

Finis