Data Science 2
Kansen

Karel de Grote
Hogeschool

Quote van de week

"If there is a 50-50 chance that something can go wrong, then 9 times out of 10 it will."

Paul Harvey (1918-2009)

Agenda

1. Wat is een kans?
2. Verzamelingenleer en kansen
3. Kruistabellen en kansen
4. Rekenen met kansen
   - Algemeen
   - De somregel
   - De productregel
   - Afhankelijke gebeurtenissen
5. Kansen en IT
6. Kans in de media

Cursus/ Boek Tekst op Canvas: *"Kansen.pdf"*

canvas Oefeningen

1

Wat is een kans?

Wat is een kans?

Een experiment dat verschillende uitkomsten produceert ondanks dezelfde beginsituatie

- bv: gooi een dobbelsteen, bepaal de spin van een elektron, test hoe lang de voeding van een server werkt, ...
- Toch zijn er regelmatigheden in de resultaten als je veel metingen doet:
  - bepaalde waarden komen meer voor dan andere
  - alle waarden komen ongeveer evenveel voor
  - het gemiddelde van de waarden situeert zich rond een bepaalde waarde
  - …

Wat is een kans?

- Verschillende interpretaties van het begrip ‘kans’ mogelijk:
  - *resultaat uit afgelijnd experiment*

    trek een balletje uit een zak met 20 rode en 30 blauwe balletjes. Wat is de kans dat het een rood balletje is?
  - *veralgemening van experiment naar populatie*

    van de 20 willekeurig geteste iPads waren er 2 stuk. Hoeveel teruggebrachte iPads mogen we in de winkel verwachten?
  - *kans van een individuele meting*

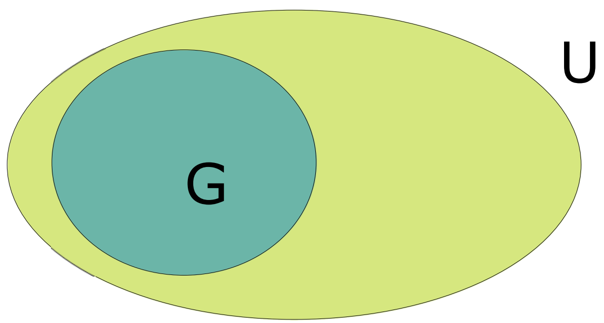
    een patiënt wil een operatie laten ondergaan. Wat is de kans dat deze operatie zal slagen?

Verzamelingenleer en kansen

2

Kansen met verzamelingenleer

- Laplace
  - **verzameling** met mogelijke uitkomsten: U
  - elke uitkomst is **even waarschijnlijk** (uniforme verdeling)
  - **gebeurtenis** = verzameling van gewenste uitkomsten: G
  - kans dat gebeurtenis optreedt **P(G)** = #G / #U

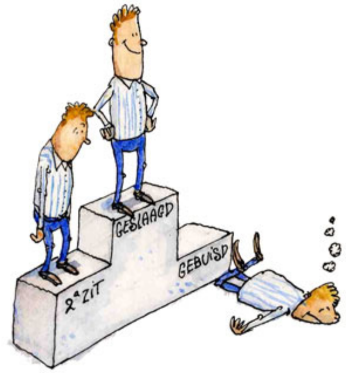


Kansen met verzamelingenleer

- Voorbeeld 1

Gebeurtenis		#G	P(G)
2	(1,1)	1	0,0278
3	(1,2); (2,1)	2	0,0556
4	(1,3); (2,2); (3,1);	3	0,0833
5	(1,4); (2,3), (3,2); (4,1)	4	0,1111
6	(1,5); (2,4); (3,3); (4,2); (5,1)	5	0,1389
7	(1,6); (2,5); (3,4); (4,3); (5,2); (6,1)	6	0,1667
8	(2,6); (3,5); (4,4); (5,3); (6,2)	5	0,1389
9	(3,6); (4,5); (5,4); (6,3)	4	0,1111
10	(4,6); (5,5); (6,4)	3	0,0833
11	(5,6); (6,5)	2	0,0556
12	(6,6)	1	0,0278
	#U (= TOTAAL)	36	1

Kansen met verzamelingenleer



- Voorbeeld 2:

  U = { studenten INF1 v.h. jaar 2013-2014 },
  #U = 210

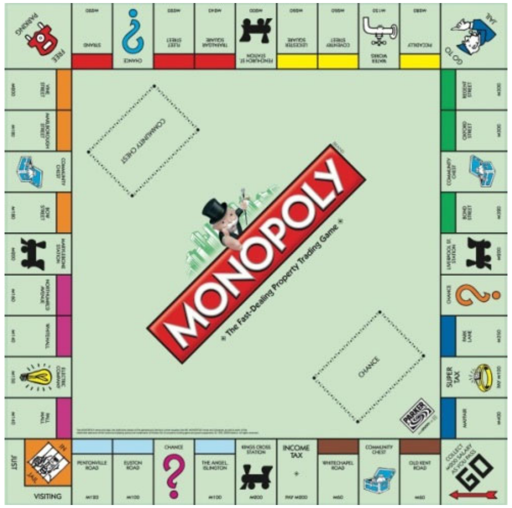
  G = { geslaagde studenten v.h. jaar 2013-2014 },
  #G = 70

  – Wat is de kans dat een willekeurig gekozen student uit U geslaagd is?
  – Wat is de kans dat een willekeurig gekozen student dit jaar geslaagd is?
  – Wat is de kans dat jij geslaagd bent?

Kansen met verzamelingenleer

- Andere voorbeelden Laplace:

- Maar…



Heeft elk veld evenveel kans om ‘bezocht’ te worden gedurende het spel?

3

Kruistabellen en kansen

Kruistabellen en kansen

- Je kan kansen soms ook gemakkelijk aflezen uit kruistabellen:

	Wit merk (White label)	Geen wit merk (Private label)	
Slechte koeling	1498	1513	3011
Goede koeling	504	6485	6989
	2002	7998	10000

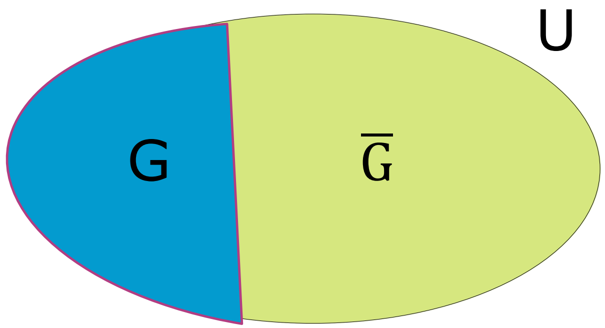
- Wat is de kans dat een pc van een wit merk een slechte koeling heeft?
- U = { pc's van wit merk }, G = { slechte koeling }
- kans is dus: 1498/2002 = 0,75

4

Rekenen met kansen

- Algemeen
- De somregel
- De productregel
- Afhankelijke gebeurtenissen

Rekenen met kansen



- Kans van **tegengestelde gebeurtenis**:

$$P(\overline{G}) = 1 - P(G)$$

is soms eenvoudiger te bepalen/berekenen

Opmerkingen:

- de ***tegengestelde gebeurtenis*** wordt ook wel de ***complementaire gebeurtenis*** genoemd
- G en $\overline{G}$ zijn **uitsluitende gebeurtenissen**

Rekenen met kansen

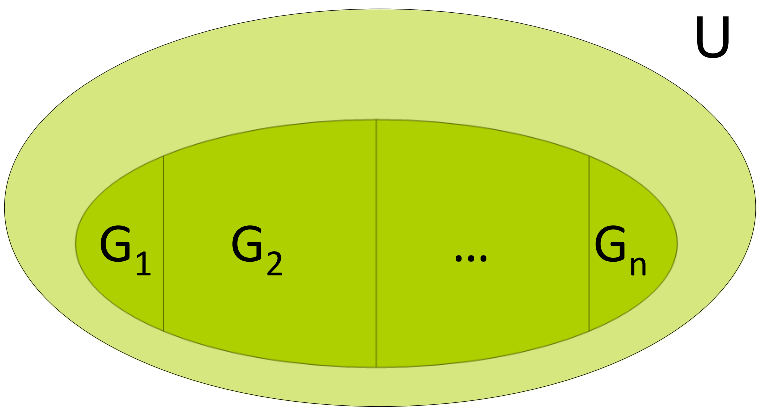
- Voorbeeld: kans dat je met 2 dobbelstenen minstens 4 gooit:
  - U = {...}
  - P(G) = kans dat je 4, 5,..,12 gooit
  - P(G) = 1 - P($\overline{G}$)
  - P($\overline{G}$) = kans dat je 2 of 3 gooit
  - $\overline{G}$ = {...}
  - P($\overline{G}$) is dus ...
  - dus: 1-P($\overline{G}$) = ...

Gebeurtenis		#G
2	(1,1)	1
3	(1,2); (2,1)	2
4	(1,3); (2,2); (3,1);	3
5	(1,4); (2,3), (3,2); (4,1)	4
6	(1,5); (2,4); (3,3); (4,2); (5,1)	5
7	(1,6); (2,5); (3,4); (4,3); (5,2); (6,1)	6
8	(2,6); (3,5); (4,4); (5,3); (6,2)	5
9	(3,6); (4,5); (5,4); (6,3)	4
10	(4,6); (5,5); (6,4)	3
11	(5,6); (6,5)	2
12	(6,6)	1
	#U (= TOTAAL)	36

De Somregel

- Gebeurtenissen kunnen bestaan uit **deelgebeurtenissen**
- Stel dat de deelgebeurtenissen G_i (i=1,..n) niet overlappen, het zijn **uitsluitende deelgebeurtenissen** (ze kunnen niet samen voorkomen of m.a.w. $G_i \cap G_j = \varnothing \ \forall\ i=1..n,\ j=1..n$ waarbij $i \neq j$)
- We vragen ons af wat de kans is dat 1 van deze gebeurtenissen optreedt:

$$P(G) = P(G_1 \ \underline{\text{OF}}\ G_2 \ \underline{\text{OF}}\ \ldots \ \underline{\text{OF}}\ G_n)$$
$$= P(G_1 \cup G_2 \cup \ldots \cup G_n)$$
$$= P(G_1) + P(G_2) + \ldots + P(G_n)$$



De Somregel

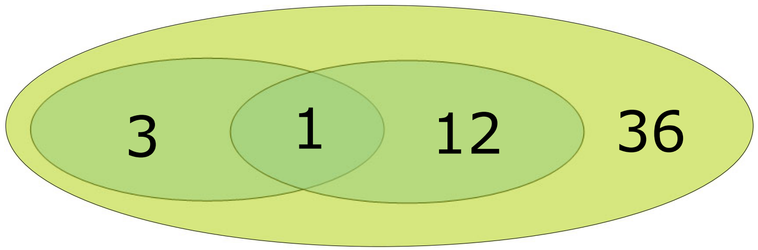
- Voorbeeld: boek kaarten, kies een kaart.
  Wat is de kans dat de kaart een aas of een 4 is?

  - kans dat kaart een aas is = 4/52
  - kans dat kaart een 4 is = 4/52
  - kans dat kaart aas OF 4 is = 4/52 + 4/52 = 8/52
    = 0,154
    = 15,4%

Uitsluitende deelgebeurtenissen!

De Somregel

- Stel dat de deelgebeurtenissen overlappen (niet uitsluitend zijn)
  Stel: enkel 2 deelgebeurtenissen
  - $P(G) = P(G_1 \cup G_2) = P(G_1) + P(G_2) - P(G_1 \cap G_2)$



- Voorbeeld: boek kaarten
  - kies 1 kaart
  - wat is de kans dat deze een aas is of een harten kaart?

$G = G_{aas} \cup G_{harten}$ maar doorsnede is niet leeg ()

$\#G_{aas} = 4, \#G_{harten} = 13, \#(G_{aas}\ \ en\ \ G_{harten}) = 1$

  - kans van G is nu (4+13-1) / 52 = 16/52 = 30,8%

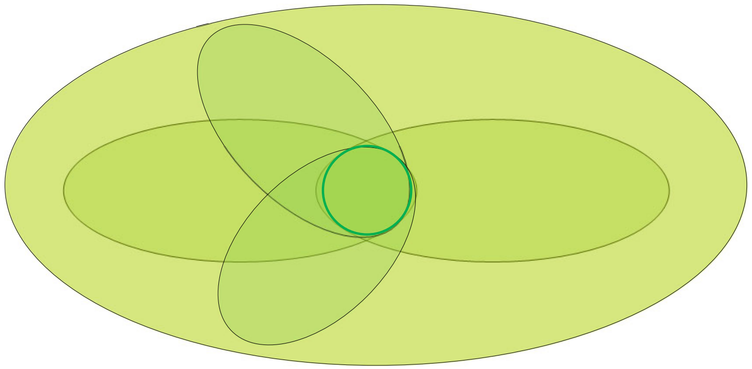
De Productregel

Een gebeurtenis bestaat weer uit deelgebeurtenissen

Wat is de kans dat alle deelgebeurtenissen <u>tegelijk</u> optreden?

- Als de deelgebeurtenissen "**onafhankelijk**" zijn, dan geldt:

$$
\begin{aligned}
P(G) &= P(G_1 \ \underline{\text{EN}}\ G_2 \ \underline{\text{EN}}\ \ldots \ \underline{\text{EN}}\ G_n) \\
&= P(G_1 \cap G_2 \cap \ldots \cap G_n) \\
&= P(G_1) \times P(G_2) \times \ldots \times P(G_n)
\end{aligned}
$$



De Productregel

- Voorbeeld: counter-strike
  - 3 spelers
    - speler1 schiet 1 keer op 5 raak
    - speler2 schiet 1 keer op 4 raak
    - speler3 schiet 1 keer op 3 raak
  - 1 terrorist probeert door tunnel te geraken
  - spelers kunnen elk 1 keer schieten

Wat is de kans dat de terrorist levend door de tunnel raakt?

De Productregel

Wat is de kans dat de terrorist levend door de tunnel raakt?

Dit is de kans dat speler1 mist (G_1) EN speler2 mist (G_2) EN speler3 mist (G_3)

– $P(G) = P(G_1 \text{ EN } G_2 \text{ EN } G_3)$
$= P(G_1 \cap G_2 \cap G_3)$
$= P(G_1) \times P(G_2) \times P(G_3)$

$= 4/5 \times 3/4 \times 2/3 = 24/60$
$= 0{,}4 = 40\%$

⇨ Terrorist heeft 40% kans dat hij door de tunnel raakt

Afhankelijke gebeurtenissen

- De gebeurtenissen kunnen ook **afhankelijk** zijn van elkaar.
  - In dit geval is $P(G_1 \cap G_2 \cap \ldots \cap G_n)$
    $\neq P(G_1) \times P(G_2) \times \ldots \times P(G_n)$
  - Voorbeeld: groep studenten
    - Kans dat iemand een meisje is, is 0,48
    - Kans dat iemand een bril draagt is 0,2
    - Wat is de kans dat een willekeurige persoon een meisje met een bril is?

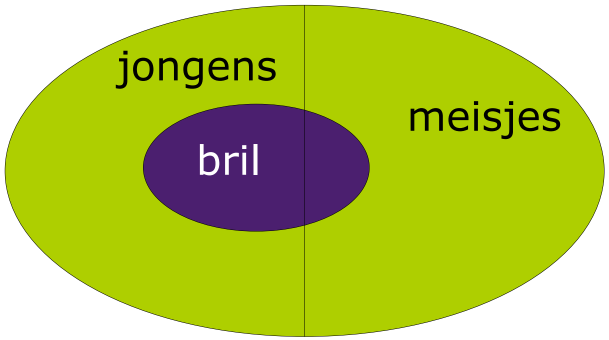
Probleem: misschien zijn alle brildragers jongens.

in dit geval is $P(G) = 0$ terwijl $P(G_{meisje}) \times P(G_{bril}) = 0{,}096$

er is dus een afhankelijkheid tussen jongens en brildragers

Afhankelijke gebeurtenissen

- Om de kans te berekenen dat een student een brildragend meisje is, gebruiken we volgende formule:
    - $P(G_{meisje} \cap G_{bril}) = P(G_{bril} \mid G_{meisje}) \times P(G_{meisje})$
  - hierbij is $P(G_{bril} \mid G_{meisje})$ de kans dat de persoon een bril draagt, ***gegeven dat*** het een meisje is (= een ***voorwaardelijke kans***).
- Als er dus geen meisjes zijn met bril, dan is dit 0
- Stel dat 10% van de meisjes een bril draagt, dan is $P(G_{meisje} \cap G_{bril})$ = 0,1 x 0,48 = 0,048

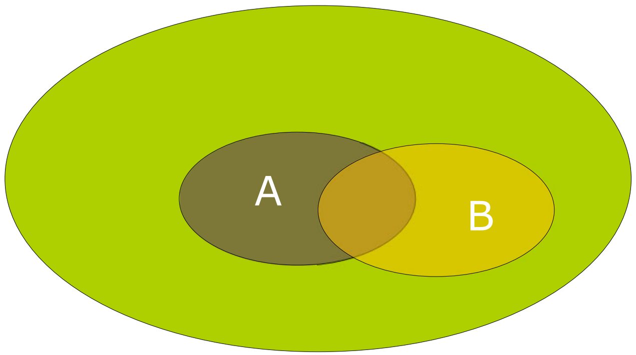


Afhankelijke gebeurtenissen

- Je kan de vorige formule $P(A \cap B) = P(\mathrm{A} \mid B) . P(B)$

  ook omdraaien om P(A|B) te definiëren: $P(\mathrm{A} \mid B) = \dfrac{P(A \cap B)}{P(B)}$

- Opmerking : $P(\mathrm{A} \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B \cap A)}{P(B)} = \dfrac{P(\mathrm{B} \mid A) . P(A)}{P(B)}$



Afhankelijkheid

- Wanneer zijn 2 gebeurtenissen **onafhankelijk**?

  Antwoord: wanneer $P(A \mid B) = P(A)$
  Dus:

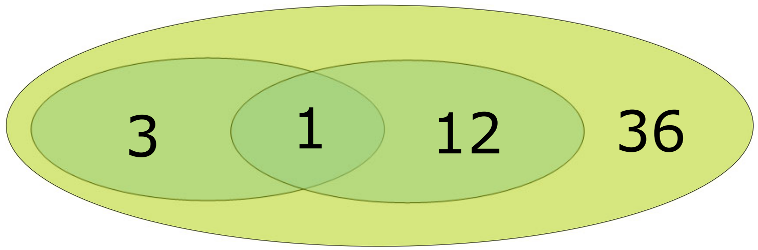
  $$P(A) \cdot P(B) = P(A \mid B) \cdot P(B) = P(A \ en \ B) = P(A \cap B)$$

- Wat kan je besluiten mbt afhankelijk/onafhankelijk wanneer $P(A \cap B) = 0$ wetende dat $P(A) \neq 0$ en $P(B) \neq 0$?

  $P(A \cap B) = 0 \Rightarrow P(A \mid B) \text{ x } P(B) = 0$ maar $P(B) \neq 0 \Rightarrow P(A \mid B) = 0$

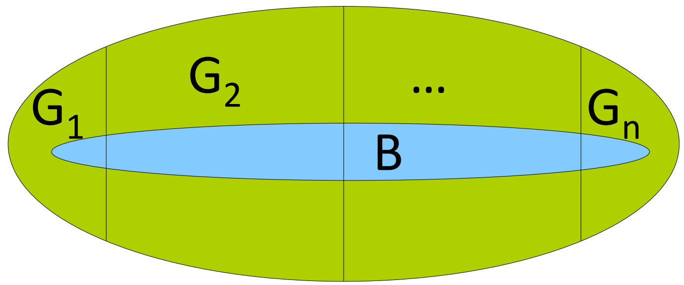
Afhankelijkheid

- Herneem het voorbeeld van het willekeurig nemen van een kaart uit een boek kaarten
  - Zijn de gebeurtenissen ‘de kaart is een aas’ en ‘de kaart is een harten’ onafhankelijk?
  - P(een aas). P(een harten) ? P(een aas en een harten)
  - P(een aas) = 4 / 52
  - P(een harten) = 13 / 52
  - P(een aas en een harten) = 1 / 52
  - P(een aas). P(een harten) = 4 / 52 . 13 / 52 = 1 /52
  - Besluit ?



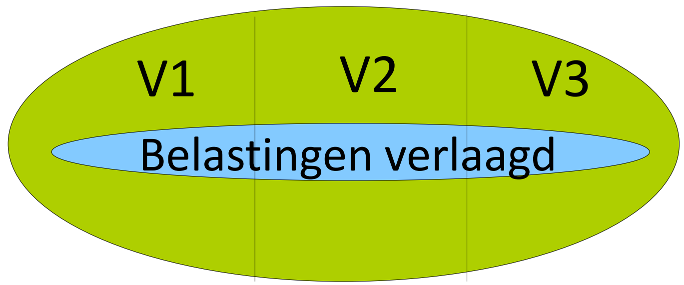
Wet van de totale kans

- Stel dat G_1, G_2, ..., G_n elkaar niet overlappen dan kan je volgende formule opstellen:



$$P(B) = \sum_{i=1}^{n} P\,(B \text{ en } G_i). = \sum_{i=1}^{n} P\,(B \mid G_i).\ P(G_i)$$

- Voorbeeld:
  - 3 politici hebben resp. 30%, 20% en 50% kans dat ze verkozen worden (V1, V2, V3) –als minister van financiën-
  - de kans dat de politici de belastingen verlagen is resp. 50%, 40% en 30%
  - wat is de kans dat de belastingen verlaagd worden?



$$P(B_{verlaagd}) \quad = P(B_{verlaagd} \mid V_1).P(V_1) + P(B_{verlaagd} \mid V_2).P(V_2) + P(B_{verlaagd} \mid V_3).P(V_3)$$

$$= 0{,}5\ .\ 0{,}3 + 0{,}4\ .\ 0{,}2 + 0{,}3\ .\ 0{,}5 = 0{,}38$$

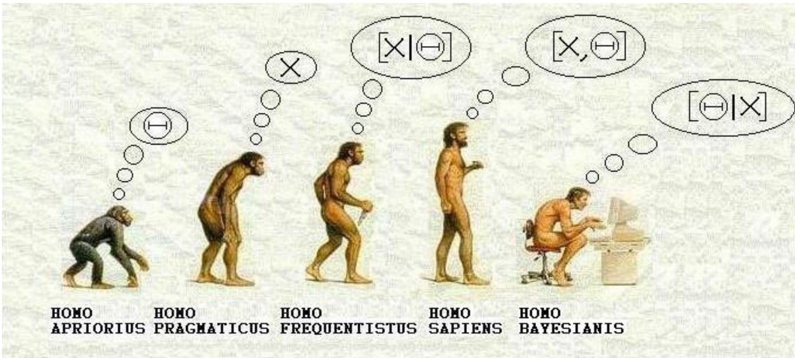
Wet van Bayes

- Stel dat $\{G_1, G_2, ..., G_n\}$ een partitie is van G
  dan zegt de wet van Bayes:

$$P(G_k \mid B) = \frac{P(B \mid G_k).P(G_k)}{P(B)} \quad want\text{: } P(G_k \mid B).P(B) = P(B \cap G_k) = P(B \mid G_k).P(G_k)$$

- Stel nu dat na de verkiezingen de belastingen verlaagd worden. Hoeveel kans is er dan dat dat komt <u>doordat</u> politicus 3 verkozen werd?

$$P\left(V_3 \mid B_{verlaagd}\right) = \frac{P\left(B_{verlaagd} \mid V_3\right).P(V_3)}{P\left(B_{verlaagd}\right)}$$



Wet van Bayes - oefening

Bob pendelt elke dag naar zijn werk

- Kans om te laat te komen:
  - trein (10%), bus (20%), auto (40%)
- Op zekere dag komt Bob te laat. Wat is de kans dat hij met de auto kwam?

  Stel: alle voertuigen met evenveel kans gebruikt

$$P(auto \mid telaat) = \frac{P(telaat \mid auto).P(auto)}{P(telaat)}$$

$$\begin{aligned} P(telaat) \quad &= P(telaat \mid auto).P(auto) && = 0{,}4\ .\ 1/3 \\ &\quad +P(telaat \mid bus).P(bus) && \quad + 0{,}2\ .\ 1/3 \\ &\quad +P(telaat \mid trein).P(trein) && \quad + 0{,}1\ .\ 1/3 \end{aligned}$$

Wet van Bayes - oefening

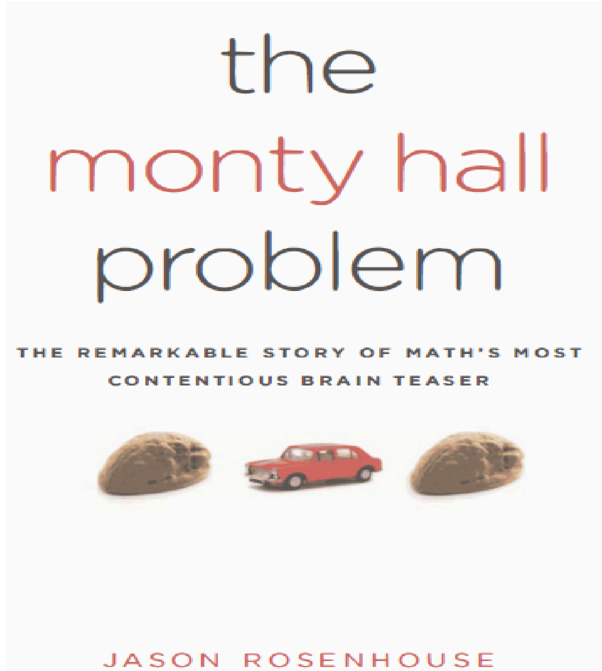
- Stel: kans op gekozen voertuig:
  - auto (10%), trein (80%), bus (10%)

$$P(auto \mid telaat) = \frac{P(telaat \mid auto).\,P(auto)}{P(telaat)}$$

$$\begin{aligned} P(telaat) \quad &= P(telaat \mid auto).\,P(auto) &&= 0{,}4\ .\ 0{,}1 \\ &\quad +P(telaat \mid bus).\,P(bus) &&\quad +0{,}2\ .\ 0{,}1 \\ &\quad +P(telaat \mid trein).\,P(trein) &&\quad +0{,}1\ .\ 0{,}8 \end{aligned}$$

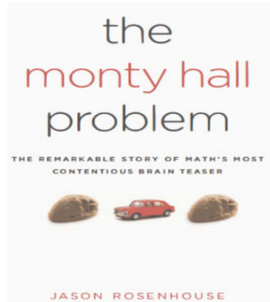
Voorbeeld - Monty Hall problem

Tijdens een spelprogramma, mag de winnende kandidaat kiezen uit 3 afgeschermde ruimtes A, B en C. In één van deze drie ruimtes staat een auto, de twee andere ruimtes zijn leeg. De kandidaat kiest één van de 3 ruimtes waarop de spelleider één van de andere twee ruimtes opent en toont dat deze leeg is. De kandidaat kan nu bij zijn oorspronkelijke keuze blijven of mag nog kiezen voor de andere nog afgeschermde ruimte. Wat zou de kandidaat het best doen om zo veel mogelijk kans te maken om de auto te winnen?



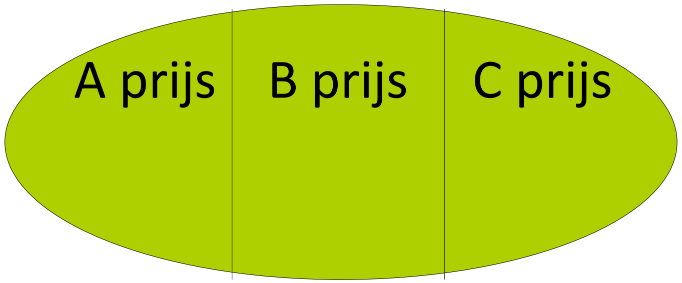
Stel de kandidaat kiest voor ruimte A en de spelleider opent de lege ruimte B. (redenering is analoog wanneer de kandidaat een andere ruimte kiest en de spelleider een andere lege ruimte toont)

Voorbeeld - Monty Hall problem



Beschouw de volgende gebeurtenissen:

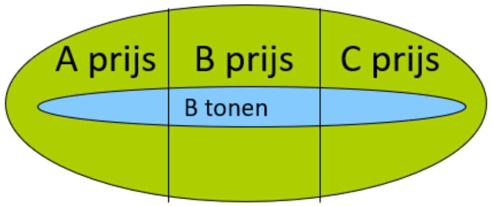
- A prijs (ruimte A bevat de prijs) ⇨ P(A prijs) = 1/3
- B prijs (ruimte B bevat de prijs) ⇨ P(B prijs) = 1/3
- C prijs (ruimte C bevat de prijs) ⇨ P(C prijs) = 1/3

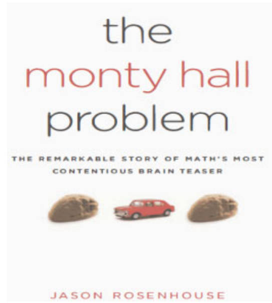


De kandidaat kiest voor ruimte A

Voorbeeld - Monty Hall problem

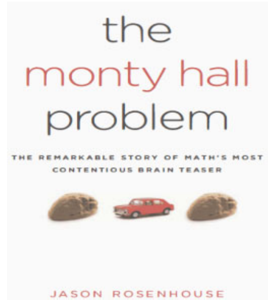
Beschouw dan de gebeurtenis *B tonen*
(de spelleider toont de lege ruimte B):





- De volgende voorwaardelijke kansen kunnen bepaald worden:
  - P(B tonen | A prijs) = 0,5
  - P(B tonen | B prijs) = 0
  - P(B tonen | C prijs) = 1
- P(B tonen) = P(B tonen | A prijs). P(A prijs)
  + P(B tonen | B prijs). P(B prijs)
  + P(B tonen | C prijs). P(C prijs)
  = 0,5 . 1/3 + 0 . 1/3 + 1 . 1/3
  = 0.5

Voorbeeld - Monty Hall problem



- Wet van Bayes toepassen:

$$\begin{aligned} P(\text{A prijs} \mid \text{B tonen}) &= P(\text{B tonen} \mid \text{A prijs}) \,.\, P(\text{A prijs}) \,/\, P(\text{B tonen}) \\ &= (0{,}5 \,.\, 1/3\,) \,/\, 0{,}5 \\ &= 1/3 \end{aligned}$$

$$\begin{aligned} P(\text{C prijs} \mid \text{B tonen}) &= P(\text{B tonen} \mid \text{C prijs}) \,.\, P(\text{C prijs}) \,/\, P(\text{B tonen}) \\ &= (1 \,.\, 1/3\,) \,/\, 0{,}5 \\ &= 2/3 \end{aligned}$$

⇨ De kandidaat heeft de meeste kans om de auto te winnen door zijn keuze te veranderen en te kiezen voor ruimte C

Kansen en IT

- Spam filter

5

Spam filter

Spam is de merknaam van een bepaald soort ingeblikt vlees (op de markt gebracht door Hormel sinds 1937). Stond gedurende WO II als vast onderdeel op het menu van de Amerikaanse soldaat.
Monthy Python gebruikte spam in een sketch -1970- (zie https://youtu.be/anwy2MPT5RE) om het toen actuele verbod op sluikreclame op televisie aan de kaak te stellen. Werd daardoor het symbool van ongewenste reclame en later van ongewenste e-mails . De 1ste spam mail werd in 1978 via ARPAnet verzonden (zie http://www.themarysue.com/first-spam-email/)

Spam filter wil weten wat de kans is dat een email spam (= "slechte kwaliteit")
of ham (= "goede kwaliteit") is

- als P_{spam}/P_{ham} groter is dan bepaalde grens ==> spam
- spam??? --> filmpje (youtube – bayesian filter)
- Toepassing op de wet van Bayes !

Spam filter

- Gegeven: email met woorden w_1, w_2, ..., w_n
- Gevraagd: kans dat dit spam/ham is?
  - $P(SPAM \mid w_1$ en w_2 en ... en $w_n)$
  - $P(HAM \mid w_1$ en w_2 en ... en $w_n)$
- we weten

$$P\left(SPAM \mid w_1, w_2, \ldots, w_n\right) = \frac{P\left(w_1, w_2, \ldots, w_n \mid SPAM\right).P(SPAM)}{P\left(w_1, w_2, \ldots, w_n\right)}$$

$$P\left(HAM \mid w_1, w_2, \ldots, w_n\right) = \frac{P\left(w_1, w_2, \ldots, w_n \mid HAM\right).P(HAM)}{P\left(w_1, w_2, \ldots, w_n\right)}$$

Spam filter

- om te weten of een mail spam is, berekenen we dus:

$$\frac{P(SPAM \mid w_1, w_2, \ldots, w_n)}{P(HAM \mid w_1, w_2, \ldots, w_n)} = \frac{P(w_1, w_2, \ldots, w_n \mid SPAM).P(SPAM)}{P(w_1, w_2, \ldots, w_n \mid HAM).P(HAM)}$$

- we "leren"

$$P(w_1, w_2, \ldots, w_n \mid SPAM) \approx \prod_{i=1}^{n} P\,(w_i \mid SPAM)$$

$$P(w_1, w_2, \ldots, w_n \mid HAM)$$

$$P(SPAM)$$

$$P(HAM) = 1 - P(SPAM)$$

Ter info: In september 2018 was 53,5% van de totale e-mail trafiek spam, in maart 2021 45,1% (https://www.statista.com/statistics/420391/spam-email-traffic-share/)

Kans in de media

6

Kans in de media –
Machine Learning

Kunnen we een computer ooit gezond verstand bijbrengen?

Tom Sercu past bij IBM 'deep learning' toe op spraakherkenning. © Mandy Demuth

Artificiële intelligentie dringt steeds dieper door in ons leven. Maar we weten nog steeds niet wat 'gezond verstand' eigenlijk is, of hoe we dat aan een computersysteem moeten doorgeven, zegt de jonge onderzoeker Tom Sercu van IBM. DOMINIQUE DECKMYN

Het zijn opwindende tijden, voor wie werkt aan artificiële intelligentie. Sinds drie à vier jaar volgt de ene doorbraak op de andere: auto's besturen zichzelf, smartphones begrijpen gesproken instructies steeds beter en het AlphaGo-programma versloeg de wereldkampioen Go.

'Je kunt nu tegen je telefoon spreken en die telefoon zal je 95% van de tijd juist begrijpen als je een zoekopdracht geeft', zegt Tom Sercu. 'Die technologie is pas de laatste drie jaar echt doeltreffend geworden. Hetzelfde geldt voor automatische vertalingen: iets als Google Translate is pas sinds drie à vier jaar echt goed beginnen te werken. En de doorbraak komt van *deep learning*. Dat is wat er onder de motorkap zit.'

In 2013 studeerde Tom Sercu (26) nog in Gent, maar hij kwam terecht aan New York University in het lab van Yann LeCun. Die is inmiddels de AI-baas van Facebook, terwijl Sercu zelf nu bij IBM werkt, in het legendarische T.J. Watson Research Center, waar de transistor ontwikkeld is. Hij past er *deep learning* toe op spraakherkenning, een van de gebieden waar momenteel enorme vorderingen worden gemaakt.

'Bij *deep learning* wordt een "input" gemapt op een "output". Bijvoorbeeld in de spraakherkenning: de input is een audiosignaal, de output is een uitgeschreven tekst. Bij beeldherkenning is de input een afbeelding, de output is een "label" van wat er op de afbeelding te zien is, bijvoorbeeld "auto". Er zijn ontelbare posities en variaties in afbeeldingen die wij allemaal als een auto zouden herkennen. Maar voor een computer is een afbeelding gewoon een lijst met cijfers. Om van die lijst over te gaan naar een begrip van wat er op de afbeelding te zien is, laten we het systeem leren uit enorm veel voorbeelden.'

En dat werkt steeds beter, dankzij snellere computerchips en de beschikbaarheid van grote hoeveelheden gegevens. Maar de technologie heeft ook zijn beperkingen. 'Een neuraal netwerk dat je "getraind" hebt, dat werkt in de situatie waarvoor het is getraind', zegt Sercu. 'Buiten die setting gaat het falen. Wat een AI-systeem niet kan, is op een algemene manier omgaan met de wereld.'

En wij, mensen, kunnen dat wel?

'Ons hele leven lang experimenteren wij, interageren wij met de fysieke wereld. Wij slaan daarbij enorm veel feiten op in ons geheugen. En op elk moment kunnen we die informatie bovenhalen en hergebruiken. Ik heb hier een pen in mijn hand en een glas. Als ik ze allebei op de grond laat vallen, dan weet ik wat er gebeurt: de pen zal onbeschadigd zijn, het glas zal in stukken barsten, het water zal op de grond liggen. Dat is een intuïtie over de fysieke wereld, waar ik op elk moment over kan beschikken. En we hebben ook intuïtie over hoe een persoon zich zal gedragen in een specifieke sociale situatie. Dergelijke redeneringen zijn heel simpel voor ons. We hebben nog geen idee hoe we dat in een computer gaan inbrengen.'

Wat ontbreekt er dan nog?

'Een AI-systeem mist een geheugen en het vermogen om te redeneren over de wereld. Die twee samen, geheugen en redeneren, komen in de buurt van wat we *common sense* noemen – gezond verstand.

Voor wanneer is dan een computer die op een menselijke manier intelligent is?

'Dat is onmogelijk te zeggen. Vlak na de oorlog, toen men de eerste computers bouwde met vacuümbuizen, dachten onderzoekers al dat artificiële intelligentie net om de hoek lag. Zelfs de beste onderzoekers zeggen dat het waanzin is om verder dan vijf jaar vooruit te kijken. We weten niet wat de obstakels zijn die we gaan tegenkomen. In ieder geval is het onwaarschijnlijk dat we die twee problemen, geheugen en redeneren, in de komende vijf jaar oplossen. En het zou goed kunnen dat er dan nog altijd een essentieel element ontbreekt. Binnen de *deep learning*-gemeenschap wordt er de jongste twee jaar nagedacht over hoe we een neuraal netwerk kunnen uitrusten met een geheugen. Maar dat zit nog in een erg vroeg stadium. En redeneren is nog iets verder weg, denk ik.'

Futuroloog en Google-topman Raymond Kurzweil voorspelt nochtans dat in 2045 computers slimmer zullen zijn dan mensen.

'Goh, als je met onderzoekers spreekt die echt aan AI werken, zul je weinig mensen vinden die het met hem eens zijn. Een van de veronderstellingen van Kurzweil is dat de kracht van computers exponentieel blijft groeien. Maar we zitten op dit moment al dicht bij de fysieke limieten van hoe snel we een processor kunnen laten draaien. En zelfs als die exponentiële groei er wel was, zit je nog met die onopgeloste problemen waarover ik het had.'

Veel mensen maken zich zorgen dat hun job binnenkort door een robot of computer wordt gedaan. Realistisch?

'De automatisering is in opmars sinds de industriële revolutie: machines nemen jobs over van mensen. De informatietechnologie heeft al veel jobs in de administratie vergemakkelijkt. Ik geloof eigenlijk niet dat AI op dat vlak een echt breekpunt is. Ik denk dat er wel een kans is dat de trend de komende vijf jaar versnelt, maar het is in wezen hetzelfde fenomeen.'

Mijn collega's en ik wachten inmiddels al sinds de jaren 90 op de computer die onze interviews automatisch uittikt. Komt dat er nog van?

'Het Nederlands is misschien een te kleine taal. Je zult je interviews in het Engels moeten afnemen. Maar in het Engels is het voor binnenkort.'

Zullen we dan in de toekomst ook Engels moeten leren om met onze computer te praten?

'Nee, dat is een tijdelijke situatie, omdat er niet genoeg data beschikbaar zijn voor het Nederlands. Misschien zou het goed zijn als bijvoorbeeld de Vlaamse Gemeenschap wat geld investeert in een grote verzameling data in het Nederlands, dan kunnen wij onderzoekers dat gebruiken om betere spraakherkenningssystemen te maken.'

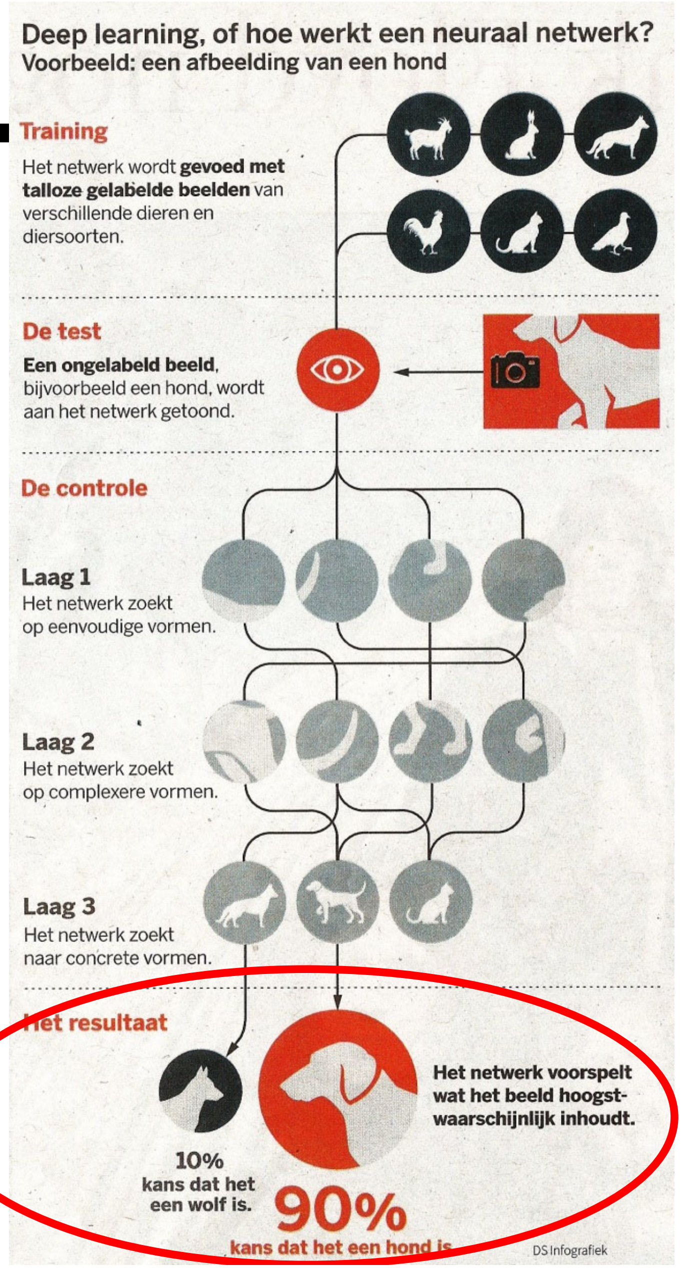
Google, Apple, Amazon en anderen werken aan digitale assistenten die ons zullen helpen bij ons werk en andere taken. Gelooft u daarin?

'Daar zit zeker toekomst in, maar het zal geleidelijk komen. Als je nu experimenteert met digitale assistenten als Siri, Alexa, en Facebook M, dan zie je dat ze erg beperkt zijn. Als je buiten hun script gaat, kunnen ze je niet helpen. Dat zal verbeteren. Vergelijk het met hoe zoeken op Google de afgelopen tien jaar is verbeterd. Tien jaar geleden moest je een zoekopdracht intikken als een reeks steekwoorden. Als je vandaag in Google je vraag formuleert in een zin met spelfouten, krijg je toch nog het juiste resultaat. Zo zal ook AI verbeteren en in ons leven doordringen, geleidelijk meer en meer.'

'Wat een AI-systeem niet kan, is op een algemene manier omgaan met de wereld'

'We zitten op dit moment al dicht bij de fysieke limieten van hoe snel we een processor kunnen laten draaien'

De Standaard 19-20 november 2016



Kans in de media

CIJFER van de dag

1 / 10[22]

Wetenschappers van de University of Rochester zijn vrij zeker dat er buitenaards leven is. Ze hebben berekend dat de kans dat we alleen zijn in dit universum één op tien miljard biljoen is.

Metro 5 maart 2016

Wees kritisch: het gaat om een schatting (waarschijnlijk geheel of gedeeltelijk) gebaseerd op de ***vergelijking van Drake***. Denk o.a. aan de *Fermiparadox* (de grote statistische waarschijnlijkheid van het bestaan van intelligent buitenaards leven staat in schril contrast met een gebrek aan bewijs daarvoor)

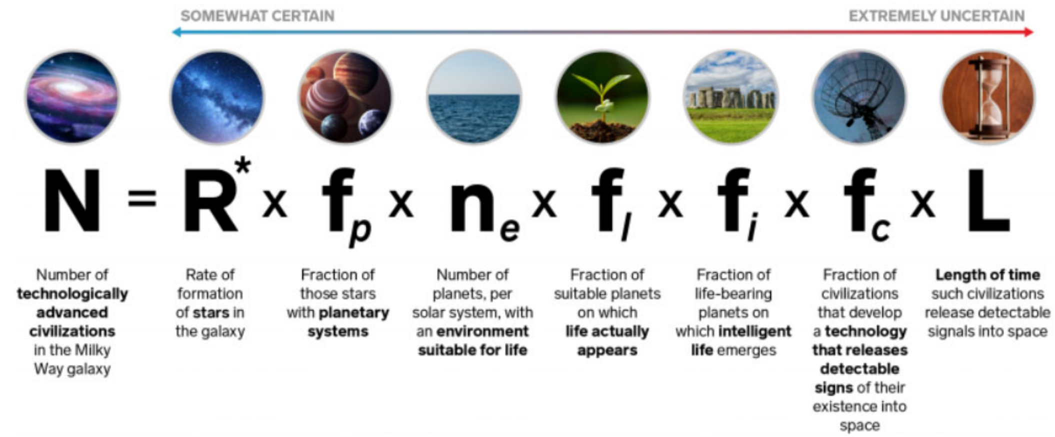
Scientists use this formula, called the Drake equation, to explore the chance we'll make contact with intelligent aliens

Dave Mosher, Jenny Cheng

03 jul 2018

https://www.businessinsider.nl/drake-equation-formula-alien-life-calculation-2018-7?international=true&r=US

- **Nearly half of Americans believe aliens have visited Earth, according to a poll.**
- **The Drake equation explores the chances that detectable alien civilizations exist using seven variables.**
- **While some predictions that use the equation are optimistic, a comprehensive new study suggests a strong likelihood that we're alone in the Milky Way galaxy.**
- **There's also a roughly 38% chance that humans are completely alone in the visible universe.**



De Morgen – 3 januari 2017

Kans in de media

Wees kritisch: Het artikel vermeld geen kansen, wel percentages waarmee de kansen werden verhoogd.

Wetenschappelijk onderzoek heeft inmiddels uitgewezen dat de molecule valproaat het risico op ontwikkelingsstoornissen met 40 procent doet toenemen. De kans op geboorteafwijkingen neemt toe met 11 procent.

De grootteorde van deze kansen wordt verzwegen, wat echter wel crusiaal is:

0,000001 ⇨ 0,0000014

OF

0,1 ⇨ 0,14

OF

?

‘Jarenlang hebben de dokters de risico’s geminimaliseerd’

Artsen verzekerden Nathalie Raemdonck (47) dat het geen kwaad kon om een epilepsiemiddel te slikken tijdens haar zwangerschap. Pas vele jaren later ontdekte ze dat haar zonen hierdoor autisme hebben gekregen. Samen met drie andere gezinnen stapt ze nu naar de rechter.

SARA VANDEKERCKHOVE

40%

De molecule valproaat verhoogt het risico op ontwikkelingsstoornissen met 40 procent.

11%

De kans op geboorteafwijkingen neemt toe met 11 procent.

Nathalie Raemdonck met haar zoon Robin, die aan autisme lijdt.

De Block: ‘Moeten nieuwe gevallen voorkomen’

‘Zwangere vrouwen moeten deze medicijnen mijden’

MAGGIE DE BLOCK, MINISTER VAN VOLKSGEZONDHEID (OPEN VLD)

‘Wie is hier verantwoordelijk voor? Het is nu aan Justitie om dat uit te maken’

NATHALIE RAEMDONCK

Slachtofferfonds?

Facebook / Metro – 16 januari 2017

Kans in de media

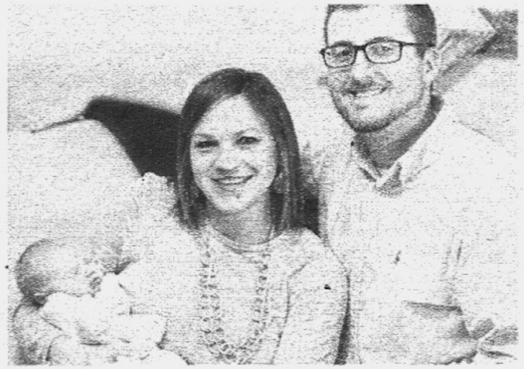
Wat is de kans dat 3 personen op dezelfde dag jarig zijn?

Zijn dit onafhankelijke gebeurtenissen?

In het artikel staat als kans 1 op 133.000

Kom je tot hetzelfde resultaat?

Verklaar een eventueel verschil.



Facebook / Hillary Cadenhead Gardner

Het hele gezin is op dezelfde dag jarig

Groot feest ten huize Gardner uit Mississippi. Het Amerikaanse koppel Luke en Hillary zijn sinds 18 december de kersverse ouders van hun zoon Cade. Opvallend: die dag vieren ook moeder en vader hun verjaardag. Volgens experts gaat het om een uitzonderlijk geval. De kans dat zowel de ouders als het kind op dezelfde dag jarig zijn, is 1 op 133.000. Het stel is in ieder geval in de wolken. Aan CBS News laten ze al grappend weten dat ze gaan proberen om in de toekomst opnieuw op 18 december een kind op de wereld te zetten.

16 april 2018

Kans in de media

Hoe groot is de kans dat deze jobs in België zullen verdwijnen?

De studiedienst van ING België onderzocht volgens de Oxford-methode de vervangbaarheid van liefst 426 jobs in België. Elk beroep in ons land werd beoordeeld op basis van drie factoren die een hinderpaal kunnen vormen voor automatisering: creativiteit, sociale intelligentie en niet-routineuze taken.

Verkoopspecialisten informatie- en communicatietechnologie (ICT)	18,0%
Systeemanalisten	1,1%
Softwareontwerpers	8,6%
Web- en multimediaontwerpers	21,0%
Applicatieprogrammeurs	48,0%
Software- en applicatieontwikkelaars en –analisten, niet elders geclassificeerd	17,7%
Ontwerpers en beheerders van databanken	3,0%
Systeembeheerders	3,0%
Netwerkspecialisten	17,7%
Databank- en netwerkspecialisten, niet elders geclassificeerd	17,7%

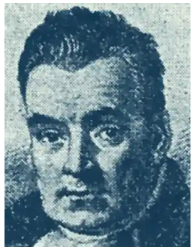
https://www.standaard.be/automatisering-jobs
https://www.standaard.be/extra/pdf/automatiseringskansen.pdf
https://www.ing.be/Assets/Documents/INGFocusArbeidsmarktNL.pdf

Probability in the media – Bayes' Theorem

The obscure maths theorem that governs the reliability of Covid testing

Tom Chivers

Sun 18 Apr 2021 07.00 BST



…Bayes's theorem is written, in mathematical notation, as P(A|B) = (P(B|A)P(A))/P(B). It looks complicated.
But you don't need to worry about what all those symbols mean: it's fairly easy to understand when you think of an example.
Imagine you undergo a test for a rare disease. The test is amazingly accurate: if you have the disease, it will correctly say so 99% of the time; if you don't have the disease, it will correctly say so 99% of the time. But the disease in question is *very* rare; just one person in every 10,000 has it. This is known as your "prior probability": the background rate in the population.
So now imagine you test 1 million people. There are 100 people who have the disease: your test correctly identifies 99 of them. And there are 999,900 people who don't: your test correctly identifies 989,901 of them.
But *that* means that your test, despite giving the right answer in 99% of cases, has told 9,999 people that they have the disease, when in fact they don't. So if you get a positive result, in this case, your chance of *actually having the disease* is 99 in 10,098, or just under 1%. If you took this test entirely at face value, then you'd be scaring a lot of people, and sending them for intrusive, potentially dangerous medical procedures, on the back of a misdiagnosis.
Without knowing the prior probability, you don't know how likely it is that a result is false or true. If the disease was not so rare – if, say, 1% of people had it – your results would be totally different. Then you'd have 9,900 false positives, but also 9,990 true positives. So if you had a positive result, it would be more than 50% likely to be true.
This is not a hypothetical problem. One review of the literature found that 60% of women who have annual mammograms for 10 years have at least one false positive; another study found that 70% of prostate cancer screening positives were false. An antenatal screening procedure for foetal chromosomal disorders which claimed "detection rates of up to 99% and false positive rates as low as 0.1%" would have actually returned false positives between 45% and 94% of the time, because the diseases are so rare, according to one paper….

https://www.theguardian.com/world/2021/apr/18/obscure-maths-bayes-theorem-reliability-covid-lateral-flow-tests-probability

Vragenlijst

7

Vragenlijst

- Download het bestand *vragenlijst 21-22.xlsx* van Canvas
- Exporteer het excel-bestand als een csv bestand
- Plaats *vragenlijst 21-22.csv* in je Python workspace
- Lees de data in en plaats het in het dataframe ***studenq***

```
>>> import pandas as pd
>>> studenq = pd.read_csv('vragenlijst 21-22.csv', delimiter=';',
decimal='.')
```

Vragenlijst

Veronderstel dat we met de vragenlijst ***ALLE*** studenten van INF1 hebben ondervraagd.

1.a Wat is de kans dat een willekeurig gekozen student 2 uur of minder uren wiskunde had in het laatste jaar secondair onderwijs?

1.b Wat is de kans dat een willekeurig gekozen student 3 uur of meer uren wiskunde had?

Vragenlijst

Veronderstel dat we met de vragenlijst ***ALLE*** studenten van INF1 hebben ondervraagd.

2. Wat is de kans dat een willekeurig gekozen student als schrijfhand rechts heeft of meer dan 3 stukken fruit eet?

Vragenlijst

Veronderstel dat we met de vragenlijst ***ALLE*** studenten van INF1 hebben ondervraagd.

3. Wat is de kans dat een willekeurig gekozen rechtshandige student minder dan 2 stukken fruit eet?

Vragenlijst

Veronderstel dat we met de vragenlijst ***ALLE*** studenten van INF1 hebben ondervraagd.

4. Zijn de gebeurtenissen '*geloven in 1 of meer complottheorieën*' en '*rechter schrijfhand*' onafhankelijk?

Oefeningen

canvas

Oefening

Karel de Grote
Hogeschool