

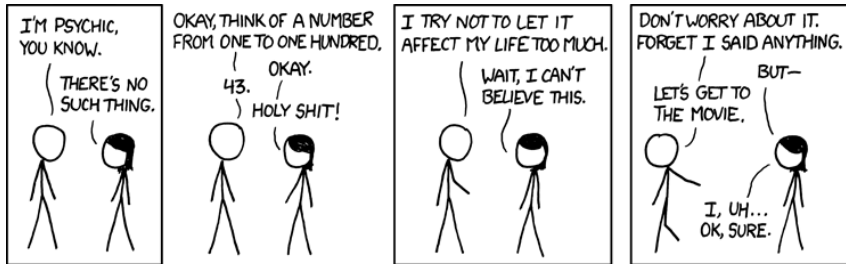
Statistiek voor A.I.

College 3

Dinsdag 18 September 2012

2 Deductieve statistiek

Kansrekening



THIS TRICK MAY ONLY WORK 1% OF THE TIME, BUT WHEN IT DOES, IT'S TOTALLY WORTH IT.

Vragen: voorspellen

Een charlatan zegt te kunnen voorspellen of een ongeboren kind een jongen of een meisje is. Een consult kost €100 en bij een foute voorspelling krijgt de klant het geld terug.

Wat is de kans dat de charlatan aan 10 klanten slechts €100 verdient?

Hoe leert een spamfilter?

Vragen: Simpsons paradox

Bij een zekere universiteit solliciteren gedurende een zekere maand 40 mannen en 40 vrouwen. Het percentage ♀ dat aangenomen wordt is 50 en voor ♂ is dat 60. Men spreekt er schande van.

Vervolgens verdeelt men de sollicitaties naar faculteit.

Bij de Faculteit Bètawetenschappen is het percentage ♀ dat aangenomen wordt 40 en voor ♂ is dat 30.

Bij de Faculteit Geesteswetenschappen is het percentage ♀ dat aangenomen wordt 80 en voor ♂ is dat 70. Men spreekt zijn goedkeuring uit.

Hoe kan dat?

Vragen: verjaardag

Wat is de kans dat twee van jullie op dezelfde dag jarig zijn?

Vragen: 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)?

Vragen: cirkels



Een computer genereert willekeurige punten in de grote cirkel. Wat is de kans dat het punt in de kleine cirkel valt?

In tegenstelling tot de vorige vragen is bij dit experiment het aantal uitkomsten oneindig.

Vragen: natuurlijke getallen

Wat is de kans dat iemand uit een zak gevuld met de natuurlijke getallen een 7 trekt?

Vragen: magneet

Een metalen balletje wordt geworpen op een langwerpige magneet van 1 m. waarvan de magnetische kracht gelijkmatig toeneemt naar de rechter kant.

Wat is de kans dat het balletje op de rechter helft van de magneet terecht komt?

Vragen: lengte

Een lineaal wordt op een willekeurig punt in tweeën gebroken.

Wat is de kans dat een van de delen kleiner is dan $\frac{1}{4}$?

Merk op: de uitkomstenruimte is het interval $(0, l)$, waarbij l de lengte van de lineaal is.

Vragen: telefoontjes

Zij X de tijdsduur tussen twee opeenvolgende telefoontjes in een alarmcentrale.
Wat voor kansverdeling heeft X en kan, bijvoorbeeld, de kans berekend worden dat de tijdsduur tussen twee opeenvolgende telefoontje groter dan 5 minuten is?

Vragen: het drie gevangenen probleem

Er zijn drie gevangenen, A , B en C , die weten dat er twee van hen willekeurig gekozen en vervolgens terechtgesteld zullen worden. De overlevingskans voor ieder van hen is dus $\frac{1}{3}$. A vraagt aan Rita, de cipier, om één gevangene ongelijk A te noemen die terechtgesteld wordt. Rita zegt: B .

Wat is nu de kans dat A overleeft?

Er lijken twee antwoorden mogelijk:

- A krijgt geen nieuwe informatie, hij wist toch al dat B of C terechtgesteld zou worden, dus zijn overlevingskans blijft $\frac{1}{3}$.
- Eerst waren er drie mogelijkheden: A of B of C overleeft. Nu zijn er twee mogelijkheden: A of C overleeft. De kans dat A overleeft is $\frac{1}{2}$.

Wat is de juiste redenering?

Vandaag:

- Kansexperimenten
- Discrete stochasten
- Stochast en populatie
- Axioma's
- Eigenschappen kansen

Kansexperimenten

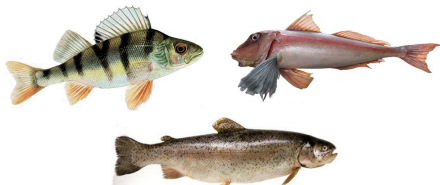
Definitie Een *kansexperiment* heeft een aantal mogelijke uitkomsten.

De *uitkomstenruimte* (*sample space*) is de verzameling S van alle mogelijke uitkomsten. (Het boek gebruikt Ω in plaats van S .)

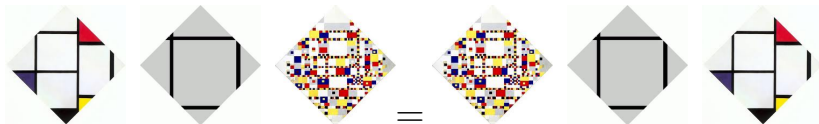
Een *gebeurtenis* (*event*) is een deelverzameling van de uitkomstenruimte.

Voorbeeld

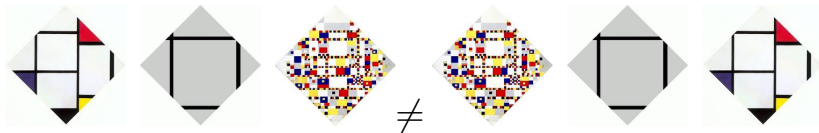
- Twee dobbelstenen gooien en het aantal ogen tellen: $S = \{2, 3, \dots, 12\}$. *Even gooien* is een gebeurtenis.
- Testen op een ziekte: $S = \{\text{ziek}, \text{gezond}\}$.
- De lengte van alle mannen in een orkest meten.
- De mogelijke uitkomsten van de verkiezingen op 12 September 2012. *De PvdA wint* is een gebeurtenis.
- Een vis uit een vijver hengelen: $S =$



Voorbeeld Drie willekeurige schilderijen uit een depot halen:



Drie willekeurige schilderijen aan de muur hangen:



Geordende versus niet geordende uitkomsten:

Voorbeeld

- Twee willekeurige getallen onder de 2 kiezen ongeacht de volgorde:
 $S = \{(0, 0), (0, 1), (1, 1)\}$.
- Een willekeurig getal onder de 2 kiezen, en dan nogmaals:
 $S = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.
- Een commissie van drie parlementsleden kiezen:
 S bestaat uit alle *verzamelingen* van drie parlementsleden.
- Een commissie van drie parlementsleden kiezen, bestaande uit een voorzitter, secretaris en penningmeester:
 S bestaat uit alle *rijtjes* van drie parlementsleden.

Axioma's

Axioma's: waarom

Vraag: Wat is de kans op een gebeurtenis?

Doel: Kansen kunnen berekenen en vergelijken. Inductief redeneren.

Axioma's: de drie axioma's van de kansrekening

Definitie De *kans* op gebeurtenis A wordt aangeduid met $P(A)$.

Definitie Het aantal elementen in A wordt aangeduid met $\#A$.

Axioma's

- Voor elke gebeurtenis A geldt $P(A) \geq 0$.
- Voor de uitkomstenruimte S geldt $P(S) = 1$.
- Als $A_1, A_2, \dots$ onderling disjunct zijn ($A_i \cap A_j = \emptyset$), dan geldt

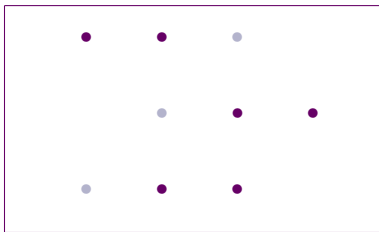
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Stelling Als S een eindige uitkomstenruimte is waarin alle uitkomsten dezelfde kans hebben, dan geldt voor elke gebeurtenis A :

$$P(A) = \frac{\#A}{\#S} = \frac{\text{aantal elementen van } A}{\text{aantal elementen van } S}.$$

Kansen toekennen

Kansen toekennen



$$P(A) = \frac{\#A}{\#S} = \frac{3}{9}.$$

Kansen toekennen

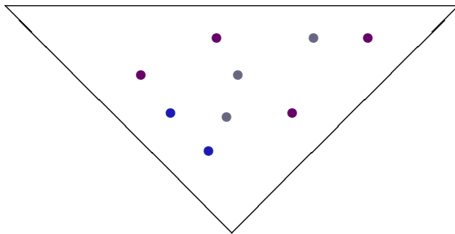
Voorbeeld S bestaat uit zes schoeisels:



Bij het willekeurig pakken van een schoeisel is de kans op een vis $\frac{1}{3}$.
De kans op een slipper is $\frac{2}{3}$.

Voorbeeld De kans op een blauwe of grijze bal bij het trekken van 1 bal uit een vaas met 2 blauwe, 3 grijze en 4 paarse ballen:

$$P(B \text{ of } G) = \frac{\# \text{ blauwe en grijze ballen}}{\# \text{ ballen}} = \frac{5}{9}.$$



Voorbeeld Uit een groep van 3 vrouwen en 2 mannen wordt een commissie van drie mensen (willekeurig) gekozen. De kans op een commissie van 2 vrouwen en 1 man is $\frac{6}{10} = \frac{3}{5}$:

$$S = \{ \{ \text{♀}, \text{♀}, \text{♀} \}, \\ \{ \text{♂}, \text{♀}, \text{♀} \}, \{ \text{♂}, \text{♀}, \text{♀} \}, \{ \text{♂}, \text{♀}, \text{♀} \}, \star \\ \{ \text{♂}, \text{♀}, \text{♀} \}, \{ \text{♂}, \text{♀}, \text{♀} \}, \{ \text{♂}, \text{♀}, \text{♀} \}, \star \\ \{ \text{♂}, \text{♂}, \text{♀} \}, \{ \text{♂}, \text{♂}, \text{♀} \}, \{ \text{♂}, \text{♂}, \text{♀} \} \}.$$

Kansen toekennen: geordend

Voorbeeld Uit een groep van 3 vrouwen en 2 mannen wordt een commissie bestaande uit een voorzitter, secretaris en penningmeester gekozen. De kans op een commissie waarbij de voorzitter en secretaris vrouwen zijn en de penningmeester een man, is $\frac{12}{60} = \frac{1}{5}$:

$$S = \{ \begin{array}{l} \text{♀♀♀, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂,} \\ \text{♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀,} \\ \text{♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀,} \\ \text{♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂,} \star \\ \text{♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀, ♂♀♀,} \\ \text{♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀, ♀♂♀,} \\ \text{♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂, ♀♀♂,} \star \\ \text{♂♂♀, ♂♂♀, ♂♂♀, ♂♂♀, ♂♂♀, ♂♂♀, ♂♂♀, ♂♂♀,} \\ \text{♂♀♂, ♂♀♂, ♂♀♂, ♂♀♂, ♂♀♂, ♂♀♂, ♂♀♂, ♂♀♂,} \\ \text{♀♂♂, ♀♂♂, ♀♂♂, ♀♂♂, ♀♂♂, ♀♂♂, ♀♂♂, ♀♂♂} \end{array} \}.$$

Discrete stochasten

Definitie Gegeven een uitkomstenruimte S is een *stochast* (*stochastische variabele*, *random variable*) een functie die aan elk punt in S een numerieke waarde (een waarde in $\mathbb{R}$) toevoegt. Stochasten worden aangeduid met X, Y, Z . $\text{rng}(X)$ is het beeld van X : $\text{rng}(X) = \{X(s) \mid s \in S\}$.

Als S aftelbaar is dan heet X *discreet* en is de *kansverdeling* (*verdeling*, *distributie*, *distributiefunctie*) van X de functie f (m in boek) gegeven door:

$$f(i) = P(X = i) = P(\{s \in S \mid X(s) = i\}).$$

Er geldt:

- $f : \text{rng}(X) \rightarrow [0, 1]$,
- $\sum_{i \in \text{rng}(X)} f(i) = \sum_{i \in \text{rng}(X)} P(\{s \in S \mid X(s) = i\}) = \sum_{s \in S} P(s) = 1$,
- Voor elke gebeurtenis $A \subseteq \text{rng}(X)$:

$$P(A) = \sum_{i \in A} f(i) = \sum_{\{s \in S \mid X(s) \in A\}} P(s).$$

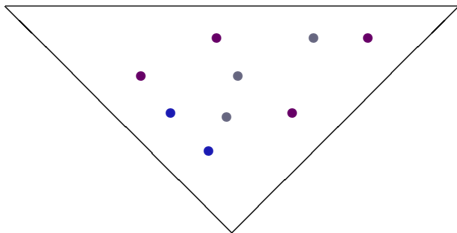
Voorbeeld

- Een dobbelsteen gooien waarbij de stochast X de identiteit is. De mogelijke waarden van X zijn dus 1, 2, 3, 4, 5, 6. De verdeling f van X wordt gegeven door $f(1) = \dots = f(6) = \frac{1}{6}$.

Compacte notatie: " $P(X = 2)$ " i.p.v. " $P(2 \text{ gooien})$ ".

- De kans op een zekere ziekte is 0.6. Testen op de ziekte: de mogelijke waarden van X zijn 0 (positieve uitslag) en 1 (negatieve uitslag). De verdeling f van X wordt gegeven door $f(0) = 0.6$ en $f(1) = 0.4$.

Voorbeeld Een bal trekken uit een vaas met 2 blauwe, 3 grijze en 4 paarse ballen. Bij het trekken van een paarse bal is de winst €1, bij een blauwe €2 en bij een grijze €-3. X is de winst na het trekken van een bal.



De mogelijke waarden van X zijn 1, 2 en -3. De verdeling van X wordt gegeven door:

$$f(1) = P(X = 1) = \frac{4}{9}, f(2) = P(X = 2) = \frac{2}{9}, f(-3) = P(X = -3) = \frac{3}{9}.$$

$$P(X = 1 \text{ of } X = 2) = \frac{6}{9}.$$

Antwoord op een vraag: voorspellen

Voorbeeld Een charlatan zegt te kunnen voorspellen of een ongeboren kind een jongen of een meisje is. Een consult kost €100 en bij een foute voorspelling krijgt de klant het geld terug. Wat is de kans dat de charlatan aan 10 klanten slechts €100 verdient, aangenomen dat bij elke klant de kans dat de charlatan juist raadt $\frac{1}{2}$ is?

De uitkomstenruimte bestaat uit alle rijtjes van tien 0-en en 1-en, waarbij 0 staat voor *foute voorspelling* en 1 voor *goede voorspelling*. De kans op elk rijtje is even groot en van de 2^{10} rijtjes zijn er 10 die precies één 1 bevatten.

De gevraagde kans is dus $\frac{10}{1024} = 0.0098$.

Stochast en populatie

Er bestaat het volgende verband tussen een populatie en een kansexperiment.

Definitie Beschouw een eindige populatie en een numerieke parameter met eindig veel mogelijke waarden: $X_1, \dots, X_n$.

Laat f_i de frequentie (het aantal) van de elementen in de populatie met waarde X_i zijn.

Beschouw het kansexperiment bestaande uit het willekeurig kiezen van een element uit de populatie, waarbij de stochast X de waarde van de parameter voor het gekozen element is (precies: de uitkomstenruimte is $\{X_1, \dots, X_n\}$ en X is daarop de identiteit).

Dan is X een stochast met mogelijke mogelijke waarden $X_1, \dots, X_n$ en verdeling:

$$P(X = X_i) = \frac{f_i}{N}.$$

X is de bij de populatie en de parameter *behorende* stochast.

- Graaf Buffon wierp in de 18e eeuw een munt 4040 maal. Het resultaat: 2048 kop. De relatieve frequentie van kop: $\frac{2048}{4040} = 0.50069$.
- Karl Pearson wierp rond 1900 24.000 maal een munt. Het resultaat: 12,012 kop. De relatieve frequentie van kop: 0.5005.
- Tijdens WO II wierp John Kerrich een munt 10,000, waarvan 5067 kop. De relatieve frequentie van kop: 0.5067.

Eigenschappen van kansen

Definitie $\bar{A}$ is het complement van A ($\tilde{A}$ in het boek).

Stelling $P(\bar{A}) = 1 - P(A)$.

Bewijs Omdat A en $\bar{A}$ disjunct zijn en $S = A \cup \bar{A}$, geldt (vorige stelling)

$$P(S) = P(A) + P(\bar{A}).$$

Omdat $P(S) = 1$ (axioma) volgt hieruit dat

$$P(A) + P(\bar{A}) = 1.$$

Dat geeft

$$P(\bar{A}) = 1 - P(A).$$

Voorbeeld Uit een groep van 15 mogelijke daders wijst een leugenachtige getuige willekeurig twee personen als de gezochte overvallers aan. Wat is de kans dat de gekozen twee niet uit de nummers 2 en 7 bestaat?

S bestaat uit alle verzamelingen van 2 personen uit de groep. Dat zijn $\frac{15 \cdot 14}{2}$ ongeordende paren.

De kans dat $A = \{2, 7\}$ niet het gekozen paar is is
 $P(\bar{A}) = 1 - P(A) = 1 - \frac{2}{15 \cdot 14} = 1 - 0.0095 = 0.9905$.

Stelling $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Bewijs

$A \setminus B$, $B \setminus A$ en $A \cap B$ zijn onderling disjunct.

Omdat $A \cup B = A \setminus B \cup B \setminus A \cup A \cap B$ geldt

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B).$$

Omdat $A = A \setminus B \cup A \cap B$ geldt $P(A) = P(A \setminus B) + P(A \cap B)$. Dus $P(A \setminus B) = P(A) - P(A \cap B)$.

Op dezelfde wijze: $P(B \setminus A) = P(B) - P(A \cap B)$.

Hieruit volgt:

$$\begin{aligned} P(A \cup B) &= \\ P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) &= \\ P(A) + P(B) - P(A \cap B). \end{aligned}$$

Voorbeeld Het percentage mensen met dyslexie is 20%, met dyscalculi 4% en 0.9% heeft dyslexie én dyscalculi.

De kans dat iemand dyslexie of dyscalculi heeft is

$$P(\text{dyslexie}) + P(\text{dyscalculi}) - P(\text{dyslexie en dyscalculi}) = 0.2 + 0.04 - 0.009 = 0.231.$$

Stelling $P(A \cup B \cup C) =$
 $P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$

Stelling

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= P(A_1) + \dots + P(A_n) \\ &\quad - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i < j < h \leq n} P(A_i \cap A_j \cap A_h) \\ &\quad \vdots \\ &\quad \pm P(A_1 \cap \dots \cap A_n). \end{aligned}$$

Eigenschappen van kansen: vereniging

Voorbeeld Van drie ziektes Z_1 , Z_2 , Z_3 is het percentage mensen onderzocht dat aan een van die ziektes of een combinatie daarvan lijdt:

	Z_1	Z_2	Z_3	$Z_1 \& Z_2$	$Z_1 \& Z_3$	$Z_2 \& Z_3$	$Z_1 \& Z_2 \& Z_3$
percentage	1.7	2.1	13.4	1.1	1.4	1.9	0.9

Wat is de kans dat iemand een (of meer) van die ziektes heeft?

$$P(Z_1 \cup Z_2 \cup Z_3) = 0.017 + 0.021 + 0.133 - 0.011 - 0.014 - 0.019 + 0.009 = 0.137.$$

Finis