Kansrekenen

Sandra Van Aert

6 oktober 2011

Kansrekening of kanstheorie

- processen of experimenten waarvan de uitkomst onzeker is
- deterministisch proces: zuiver water kookt bij 760 mm luchtdruk en 100 °C
- meeste processen zijn stochastisch of probabilistisch
  - opgooien van dobbelsteen
  - percentage defecte producten op een productielijn gedurende een bepaalde periode
- kansrekening doet uitspraken over de waarschijnlijkheid van bepaalde uitkomsten
- kans = uitdrukking van (on)waarschijnlijkheid

Verschil kansrekenen - statistiek

- kansrekenen bestudeert rechtstreeks populaties en processen
  Voorbeeld : kans berekenen om minstens 20 keer een 6 te gooien wanneer een eerlijke dobbelsteen 100 keer wordt opgegooid
- statistiek bestudeert populaties en processen via steekproefgegevens
  Voorbeeld : eerlijkheid dobbelsteen nagaan op basis van gegevens na een groot aantal keer dobbelsteen op te gooien

Kansexperiment E

- uitkomstenruimte Ω = verzameling van alle mogelijke uitkomsten
- E_1: opgooien muntstuk $\rightarrow \Omega_1 = \{\text{kop}, \text{munt}\}$
- E_2: opgooien dobbelsteen $\rightarrow \Omega_2 = \{1,2,3,4,5,6\}$
- E_3: aantal keer dat dobbelsteen opgegooid moet worden vooraleer een 6 bekomen wordt $\rightarrow \Omega_3 = \{1,2,3,\ldots\}$
- E_4: bedieningstijd van een klant aan bankloket $\rightarrow \Omega_4 = \{t : t > 0\}$
- E_2: opgooien dobbelsteen $\rightarrow \Omega_5 = \{\text{even}, \text{oneven}\}$

Gebeurtenis G

- gebeurtenis G = een verzameling van mogelijke uitkomsten
- E_1: opgooien muntstuk $\rightarrow G_1 = \{\text{kop}\}$
- E_2: opgooien dobbelsteen $\rightarrow G_2 = \{2, 4, 6\}$
- E_3: aantal keer dat dobbelsteen opgegooid moet worden vooraleer een 6 bekomen wordt $\rightarrow G_3 = \{1, 2, 3\}$
- E_4: bedieningstijd van een klant aan bankloket $\rightarrow G_4 = \{t : \text{2 minuten} \leq t \leq \text{5 minuten}\}$

Enkele begrippen

- elementaire gebeurtenis: bevat slechts één uitkomst
- een gebeurtenis G doet zich voor wanneer de uitkomst van het kansexperiment een element is van G
- G_1 en G_2 doen zich samen voor als de uitkomst tot zowel G_1 als G_2 behoort, m.a.w. tot $G_1 \cap G_2$
- mekaar uitsluitende gebeurtenissen kunnen zich niet samen voordoen, d.i. als doorsnede ledig

Enkele begrippen

- gebeurtenis G_1 of G_2 doet zich voor als de uitkomst tot ofwel G_1 ofwel tot G_2 behoort, m.a.w. tot $G_1 \cup G_2$
- het complement G^c van een gebeurtenis G is de verzameling van alle mogelijke uitkomsten die niet in G zitten

Verzamelingenleer

- unie of vereniging $G = G_1 \cup G_2 \cup \cdots \cup G_k = \bigcup_{i=1}^{k} G_i$
  - exhaustief als unie = Ω
  - ene gebeurtenis doet zich voor of de andere
- doorsnede $G = G_1 \cap G_2 \cap \cdots \cap G_k = \bigcap_{i=1}^{k} G_i$
  - disjuncte, mekaar uitsluitende gebeurtenissen als $G_1 \cap G_2 = \emptyset$
  - gebeurtenissen doen zich samen voor: en
- verschil $G = G_1 \setminus G_2$
  - deelverzameling: $G \subseteq G_1$

Verzamelingenleer

- complement $G^c = \Omega \setminus G$
  - $G \cup G^c = \Omega$
  - $G \cap G^c = \emptyset$

- partitie $G_1, G_2, \ldots, G_k \subset \Omega$
  - deelverzamelingen zijn niet leeg: $G_i \neq \emptyset, \quad \forall i$
  - hun unie is Ω: $G_1 \cup G_2 \cup \cdots \cup G_k = \Omega$
  - alle deelverzamelingen zijn disjunct: $G_i \cap G_j = \emptyset, \quad \forall i \neq j$

Partities: voorbeelden

$\Omega = \{1, 2, 3, 4, 5, 6\}$

1. $G_1 = \{1, 3, 5\}$ en $G_2 = \{2, 4, 6\}$
2. $G_1 = \{1\}$, $G_2 = \{3\}$ en $G_3 = \{2, 4, 5, 6\}$
3. $G_1 = \{1\}$, $G_2 = \{2\}, \ldots, G_6 = \{6\}$

Definitie van kans

- kans drukt waarschijnlijkheid of onwaarschijnlijkheid van een gebeurtenis G uit
- functie $P(G)$ die met G een reëel getal associeert
- 3 definities
  - empirische of frequentiedefinitie
  - klassieke definitie van Laplace
  - axiomatische definitie

Empirische of frequentiedefinitie

- herhaal een experiment een groot aantal keer (n)
- noteer de frequentie dat gebeurtenis G zich voordoet: $f_n(G)$

$$P(G) = \lim_{n \to \infty} \frac{f_n(G)}{n}$$

Klassieke definitie van Laplace

- veronderstelt dat de waarschijnlijkheid van alle elementaire gebeurtenissen gekend is
- gemakshalve: alle elementaire gebeurtenissen even waarschijnlijk

$$P(G) = \frac{\text{aantal elementen in } G}{\text{aantal elementen in } \Omega}$$

$$P(\text{even}) = \frac{\text{aantal even uitkomsten}}{\text{aantal elementen in } \Omega} = \frac{3}{6} = \frac{1}{2}$$

Axiomatische definitie

- reële functie $P(\,)$
- axioma 1: $P(G) \geq 0$
- axioma 2: $P(\Omega) = 1$
- axioma 3: indien $G_1, G_2, \ldots$ mekaar uitsluitende gebeurtenissen zijn, dan geldt dat

$$P(G_1 \cup G_2 \cup \ldots) = P(G_1) + P(G_2) + \cdots$$

$\rightarrow$ rekenregels

Rekenregels

- $P(\emptyset) = 0$
- indien $G_1 \subseteq G_2$, dan is $P(G_2 \setminus G_1) = P(G_2) - P(G_1)$ en $P(G_1) \leq P(G_2)$
- $0 \leq P(G) \leq 1$
- $P(G) + P(G^c) = 1$ zodat $P(G^c) = 1 - P(G)$
- $P(G_1 \cup G_2) = P(G_1) + P(G_2) - P(G_1 \cap G_2)$

  (optelregel)
- $$P(G_1 \cup G_2 \cup G_3) = P(G_1) + P(G_2) + P(G_3) - P(G_1 \cap G_2) - P(G_1 \cap G_3) - P(G_2 \cap G_3) + P(G_1 \cap G_2 \cap G_3)$$

  (veralgemeende optelregel)

Illustratie van kansregels

lukraak trekken van een brief verstuurd met De Post

- priorzegel of niet
- aantal dagen onderweg

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

Gebeurtenissen G_1 EN G_2 doen zich voor

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

G_1: levering binnen 2 dagen
G_2: prior

$$P(G_1 \cap G_2) = 0.400 + 0.060 = 0.460$$

Complementregel

G $$P(G) = 0.035 + 0.005 + 0.180 + 0.225 + 0.085 + 0.010 = 0.540$$

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

G^c $$P(G^c) = 1 - P(G) = 1 - 0.540 = 0.460$$

Optelregel: Gebeurtenis G_1 OF G_2 doet zich voor

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

G_1: levering binnen 2 dagen

$$P(G_1) = 0.400 + 0.060 + 0.180 + 0.225 = 0.865$$

G_2: prior

$$P(G_2) = 0.400 + 0.060 + 0.035 + 0.005 = 0.500$$

$$P(G_1 \cup G_2) = 0.865 + 0.500 - 0.460 = 0.905$$

Optelregel: speciaal geval
Elkaar uitsluitende gebeurtenissen

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

G_1: priorzegel en levering na meer dan 2 dagen

$$P(G_1) = 0.035 + 0.005 = 0.040$$

G_2: levering binnen 1 dag

$$P(G_2) = 0.400 + 0.180 = 0.580$$

$$P(G_1 \cup G_2) = 0.040 + 0.580 - 0 = 0.620$$

Voorwaardelijke kans

- wat is de kans dat je met een dobbelsteen een getal gooit dat kleiner is dan of gelijk aan 3, gegeven dat je weet dat het een even getal is?
- **notatie**: $P(G_1 \mid G_2)$ met $G_1 = \{1, 2, 3\}$ en $G_2 = \{2, 4, 6\}$
- **oplossing**:
- hoeveel elementen zijn er kleiner dan of gelijk aan 3 én even? 1
- hoeveel elementen zijn even? 3
- antwoord? $P(G_1 \mid G_2) = \frac{1}{3} = \frac{1/6}{3/6} = \frac{P(G_1 \cap G_2)}{P(G_2)}$

Voorwaardelijke kans

$$\text{kans op } \underbrace{\text{levering} \le 2 \text{ dagen}}_{G_1}\text{, gegeven } \underbrace{\text{prior}}_{G_2}$$

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

$$P(G_1 \cap G_2) = 0.400 + 0.060 = 0.460$$
$$P(G_2) = 0.400 + 0.060 + 0.035 + 0.005 = 0.500$$
$$P(G_1 \mid G_2) = 0.460/0.500 = 0.920$$

Voorwaardelijke kans

- **definitie**: $P(G_1 \mid G_2) = P(G_1 \cap G_2) / P(G_2)$
- productregel:

$$\begin{aligned} P(G_1 \cap G_2) &= P(G_1 \mid G_2) \cdot P(G_2) \\ &= P(G_2 \mid G_1) \cdot P(G_1) \end{aligned}$$

- G_2 bevat negatieve informatie over G_1:
  $P(G_1 \mid G_2) < P(G_1)$
- G_2 bevat positieve informatie over G_1:
  $P(G_1 \mid G_2) > P(G_1)$
- G_2 bevat geen info over G_1 (onafhankelijk):
  $P(G_1 \mid G_2) = P(G_1) \Rightarrow P(G_1 \cap G_2) = P(G_1) \cdot P(G_2)$

Illustratie: positieve informatie

	1	2	3	>3
Prior	0.400	0.060	0.035	0.005
N-Prior	0.180	0.225	0.085	0.010

G_1: levering binnen 2 dagen

$$P(G_1) = 0.400 + 0.060 + 0.180 + 0.225 = 0.865$$

G_2: prior

$$P(G_1 \mid G_2) = 0.460/0.500 = 0.920$$

$0.920 > 0.865$ dus G_2 bevat positieve informatie over G_1

Onafhankelijke gebeurtenissen

- **definitie**: $P(G_1 \mid G_2) = P(G_1)$
- gevolgen:

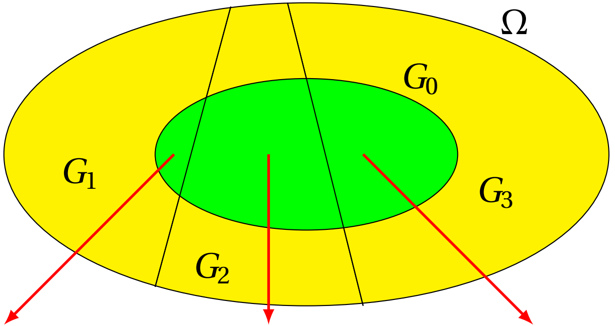
$$\begin{aligned} P(G_1 \cap G_2) &= P(G_1 \mid G_2) \cdot P(G_2) \\ &= P(G_1) \cdot P(G_2) \end{aligned}$$

$$\begin{aligned} P(G_2 \mid G_1) &= P(G_1 \cap G_2) / P(G_1) \\ &= P(G_1) \cdot P(G_2) / P(G_1) \\ &= P(G_2) \end{aligned}$$

Illustratie

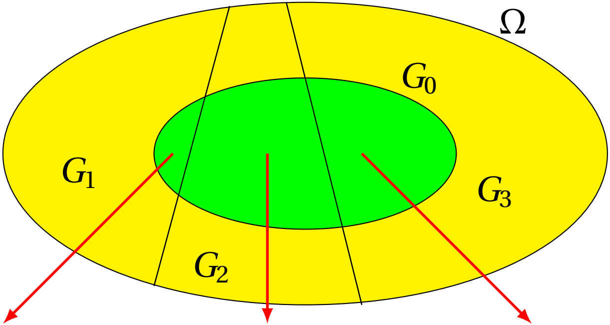
- spel van 52 kaarten
- één kaart wordt lukraak getrokken
- G_1: trekken van een aas
- G_2: trekken van een rode kaart
- $P(G_1) = \dfrac{4}{52} = \dfrac{1}{13}$
- $P(G_1 \mid G_2) = \dfrac{P(G_1 \cap G_2)}{P(G_2)} = \dfrac{2/52}{26/52} = \dfrac{2}{26} = \dfrac{1}{13}$

Stelling van de totale kans



$$(G_1 \cap G_0) \quad \cup \quad (G_2 \cap G_0) \quad \cup \quad (G_3 \cap G_0) \quad = \quad G_0$$

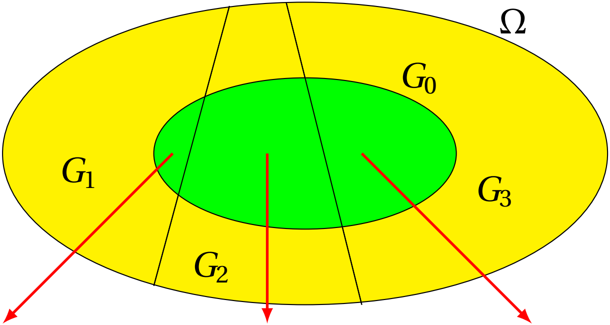
Stelling van de totale kans



$$P\Big((G_1 \cap G_0) \quad \cup \quad (G_2 \cap G_0) \quad \cup \quad (G_3 \cap G_0) \Big) = P(G_0)$$

optelregel voor elkaar uitsluitende
gebeurtenissen toepassen

Stelling van de totale kans



$$P(G_1 \cap G_0) \quad + \quad P(G_2 \cap G_0) \quad + \quad P(G_3 \cap G_0) \;=\; P(G_0)$$

$$\overbrace{P(G_0 \mid G_1) \cdot P(G_1)}\ \overbrace{P(G_0 \mid G_2) \cdot P(G_2)}\ \overbrace{P(G_0 \mid G_3) \cdot P(G_3)}$$

Stelling van de totale kans

$$P(G_0) = \sum_{i=1}^{k} P(G_0 \mid G_i) \cdot P(G_i)$$

Kansregel van Bayes

$$\begin{aligned} P(G_j \mid G_0) &= \frac{P(G_j \cap G_0)}{P(G_0)} && \text{(definitie voorwaardelijke kans)} \\ &= \frac{P(G_0 \mid G_j) \cdot P(G_j)}{P(G_0)} && \text{(productregel)} \\ &= \frac{P(G_0 \mid G_j) \cdot P(G_j)}{\sum_{i=1}^{k} P(G_0 \mid G_i) \cdot P(G_i)} && \text{(stelling van de totale kans)} \end{aligned}$$

Toepassing kansregel van Bayes

- test op HIV-virus
  - sensitiviteit 98%
  - specificiteit 95%

- gebeurtenissen
  - pos. test: positief testresultaat
  - HIV: effectief besmet
  - pos. test^c: negatief testresultaat
  - HIV^c: niet besmet

Vervolg toepassing

- sensitiviteit
  - $P(\text{pos. test} \mid \text{HIV}) = 0.98$
  - $P(\text{pos. test}^{c} \mid \text{HIV}) = 0.02$
- specificiteit
  - $P(\text{pos. test}^{c} \mid \text{HIV}^{c}) = 0.95$
  - $P(\text{pos. test} \mid \text{HIV}^{c}) = 1 - 0.95 = 0.05$
- Wat is de kans dat u besmet bent indien de test voor u positief is?
- $P(\text{HIV} \mid \text{pos. test})$

$$= \frac{P(\text{pos. test}|\text{HIV}) \cdot P(\text{HIV})}{P(\text{pos. test}|\text{HIV}) \cdot P(\text{HIV}) + P(\text{pos. test}|\text{HIV}^{c}) \cdot P(\text{HIV}^{c})}$$

$$= \frac{0.98 \cdot 0.001}{0.98 \cdot 0.001 + 0.05 \cdot 0.999} = 0.0192 \text{ (geen risicogedrag)}$$

Vervolg toepassing

- sensitiviteit
  - $P(\text{pos. test} \mid \text{HIV}) = 0.98$
  - $P(\text{pos. test}^c \mid \text{HIV}) = 0.02$
- specificiteit
  - $P(\text{pos. test}^c \mid \text{HIV}^c) = 0.95$
  - $P(\text{pos. test} \mid \text{HIV}^c) = 1 - 0.95 = 0.05$
- Wat is de kans dat u besmet bent indien de test voor u positief is?
- $P(\text{HIV} \mid \text{pos. test})$

$$= \frac{P(\text{pos. test}|\text{HIV}) \cdot P(\text{HIV})}{P(\text{pos. test}|\text{HIV}) \cdot P(\text{HIV}) + P(\text{pos. test}|\text{HIV}^c) \cdot P(\text{HIV}^c)}$$

$$= \frac{0.98 \cdot 0.10}{0.98 \cdot 0.10 + 0.05 \cdot 0.90} = 0.6853 \text{ (risicogedrag)}$$