Statistisch Priesterschap

Peter Grünwald

Reünie 12-12 2006 n.a.v. 60-jarig bestaan

David van Dantzig (1900-1959)

- Een van de oprichters van het CWI
- “Godfather van de statistiek in Nederland”

Statistisch Priesterschap

- In 1957 publiceerde Van Dantzig het artikel

  Statistical Priesthood *Statistica Neerlandica 11*,1-16.

- Dit is een bespreking van het boek *The Foundations of Statistics* van L. Savage.
- Savage spreekt van “kansen” in situaties waarin dat volgens Van Dantzig niet kan…

Kansloze Situaties

- In de geest van Van Dantzig, laat ik hier zien dat de uitspraak “deze gebeurtenis heeft X procent kans” soms betekenisloos is

1. 3-Gevangenen Probleem (1959)
   – Laat zien dat eenduidige “kans” soms niet bestaat

Kansloze Situaties

- In de geest van Van Dantzig, laat ik hier zien dat de uitspraak “deze gebeurtenis heeft X procent kans” soms betekenisloos is

1. 3-Gevangenen Probleem (1959)
   – Laat zien dat eenduidige “kans” soms niet bestaat
2. 3-Deuren Probleem (1970)
   – Laat zien dat onze intuitie hierover vaak verkeerd is

Kansloze Situaties

- In de geest van Van Dantzig, laat ik hier zien dat de uitspraak “deze gebeurtenis heeft X procent kans” soms betekenisloos is

1. 3-Gevangenen Probleem (1959)
   – Laat zien dat eenduidige “kans” soms niet bestaat
2. 3-Deuren Probleem (1970)
   – Laat zien dat onze intuitie hierover vaak verkeerd is
3. 1-Gevangene Probleem (2004-nu)
   – Laat zien wat de desastreuze gevolgen hiervan zijn !

Voorbeeld No. 0: Dobbelsteen

1. Ik gooi een eerlijke dobbelsteen; ik zie de uitkomst (een getal tussen 1 en 6) maar u niet
2. Ik vertel u ofwel “de uitkomst is even” ofwel “de uitkomst is oneven”
3. Stel dat ik u vertel “de uitkomst is even”. Wat is volgens u dan de kans dat er “4” is gegooid?
   - U zegt: er zijn nog drie mogelijkheden over. Die hebben allemaal gelijke kans, dus: **de kans op “4” is nu 1/3**

Conditionele Kansen

- Eerst was de kans 1/6
- U past deze kans aan omdat u nieuwe informatie heeft; dit heet conditioneren
- De kans is nu 1/3 geworden
- We zeggen: “de conditionele kans op “X=4”, gegeven dat “X is even”, is 1/3”

Conditionele Kansen

- Eerst was de kans 1/6
- U past deze kans aan omdat u nieuwe informatie heeft; dit heet conditioneren
- De kans is nu 1/3 geworden
- We zeggen: "de conditionele kans op "X=4", gegeven dat "X is even", is 1/3"

$$P(X=4 \mid X \in \{2,4,6\}) = \frac{P(X=4)}{P(X \in \{2,4,6\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

Het Drie Gevangenen Probleem

Gardner 1959

- Er zijn drie gevangenen, **A**, **B** en **C**
- Twee van hen worden willekeurig uitgekozen en zullen worden terechtgesteld. De gevangenen weten dit.
- **A** wordt met kans 2/3 geëxecuteerd, dus hij overleeft met kans 1/3.

Het Drie Gevangenen Probleem

- Rita, de cipier, komt langs. A vraagt haar of zij misschien kan zeggen of B of C wordt terechtgesteld
- De cipier zegt: **B**
- Het lijkt alsof de cipier A geen nieuwe informatie over zijn eigen overlevingskans geeft
  - A wist toch al dat B of C zou worden terechtgesteld

Het Drie Gevangenen Probleem

- Rita, de cipier, komt langs. A vraagt haar of zij misschien kan zeggen of B of C wordt terechtgesteld
- De cipier zegt: **B**
- Het lijkt alsof de cipier A geen nieuwe informatie over zijn eigen overlevingskans geeft
  - A wist toch al dat B of C zou worden terechtgesteld
- Dus volgens deze redenering blijft de kans dat A overleeft 1/3.

Het Drie Gevangenen Probleem

- Maar: met conditioneren wordt de kans 1/2 !
- Er waren eerst drie mogelijkheden met gelijke kans:
  - A overleeft, B overleeft, C overleeft.
- Nadat de cipier zegt "B wordt terechtgesteld", zijn er nog twee mogelijkheden over:
  - A overleeft, C overleeft.
- Dus de kans dat A overleeft is 1/2 !

Het Drie Gevangenen Probleem

- Maar: met conditioneren wordt de kans 1/2 !
- Er waren eerst drie mogelijkheden met gelijke kans:
  - A overleeft, B overleeft, C overleeft.
- Nadat de cipier zegt "B wordt terechtgesteld", zijn er nog twee mogelijkheden over:
  - A overleeft, C overleeft.
- Dus de kans dat A overleeft is 1/2 !

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

a priori kans

$$P(A \mid A \text{ of } C) = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

conditionele kans

Wat gaat er hier mis?

- Als we conditioneren, gaat de kans dat A overleeft flink omhoog, wat de cipier ook antwoordt!

Wat gaat er hier mis?

- Als we conditioneren, gaat de kans dat A overleeft flink omhoog, wat de cipier ook antwoordt!
  - Dat is vast niet goed!

Wat gaat er hier mis?

- Als we conditioneren, gaat de kans dat A overleeft flink omhoog, wat de cipier ook antwoordt!
  - Dat is vast niet goed!
- Een andere redenering zegt: de kans dat A overleeft blijft gelijk, wat de cipier ook antwoordt!

Wat gaat er hier mis?

- Als we conditioneren, gaat de kans dat A overleeft flink omhoog, wat de cipier ook antwoordt!
  - Dat is vast niet goed!
- Een andere redenering zegt: de kans dat A overleeft blijft gelijk, wat de cipier ook antwoordt!
  - Bij nader inzien is dat ook niet helemaal goed…

Wat gaat er hier mis?

- Als we conditioneren, gaat de kans dat A overleeft flink omhoog, wat de cipier ook antwoordt!
  - Dat is vast niet goed!
- Een andere redenering zegt: de kans dat A overleeft blijft gelijk, wat de cipier ook antwoordt!
  - Bij nader inzien is dat ook niet helemaal goed…
- Het juiste antwoord is:
  je kunt niet meer zeggen wat de kans is
  - Tenzij je extra aannames doet over de psyche van de cipier

Terug naar de Dobbelstenen

- Ik gooi met dobbelsteen die u niet ziet
  - Ik vertel u ofwel "de uitkomst lag tussen 1 en 3"
    ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 1 en 3" zegt u: kans op '4' is 0.
  - Als ik zeg "tussen 4 en 6" zegt u: kans op '4' is 1/3.

Terug naar de Dobbelstenen

- Ik gooi met dobbelsteen die u niet ziet
  - Ik vertel u ofwel "de uitkomst lag tussen 1 en 3"
    ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 1 en 3" zegt u: kans op '4' is 0.
  - Als ik zeg "tussen 4 en 6" zegt u: kans op '4' is 1/3.
- Stel dat we dit spel 6000 keer herhalen. Dan:
  1. Zal ik ongeveer 3000 keer zeggen "tussen 4 en 6"
  2. In ongeveer 1000 van die 3000 gevallen is de uitkomst '4'.

Terug naar de Dobbelstenen

- Ik gooi met dobbelsteen die u niet ziet
  - Ik vertel u ofwel "de uitkomst lag tussen 1 en 3"
    ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 1 en 3" zegt u: kans op '4' is 0.
  - Als ik zeg "tussen 4 en 6" zegt u: kans op '4' is 1/3.
- Stel dat we dit spel 6000 keer herhalen. Dan:
  1. Zal ik ongeveer 3000 keer zeggen "tussen 4 en 6"
  2. In ongeveer 1000 van die 3000 gevallen is de uitkomst '4'.
- De uitkomst is '4' **in ongeveer 1 op de 3 van de gevallen waarin u *zegt* "de kans op '4' is 1/3"**
  - Uw kansuitspraak is **geijkt**

Variatie

- Ik vertel u ofwel "de uitkomst lag tussen 1 en **4**" ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 4 en 6" zegt u: kans op "4" is 1/3.
- Stel dat we dit spel vaak herhalen.
  - Steeds als de uitkomst 4 is, heb ik een **keuze** wat ik ga vertellen

.

Variatie

- Ik vertel u ofwel "de uitkomst lag tussen 1 en 4" ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 4 en 6" zegt u: kans op "4" is 1/3.
- Stel dat we dit spel vaak herhalen.
  - Steeds als de uitkomst 4 is, heb ik een **keuze** wat ik ga vertellen
  - Ik kan het bijvoorbeeld zo doen: als de uitkomst 4 is zeg ik altijd "tussen 1 en 4" en **nooit** "tussen 4 en 6".
  - Uw uitspraak "kans op 4 is 1/3" is dan dus *niet* geijkt

Variatie

- Ik vertel u ofwel "de uitkomst lag tussen 1 en 4" ofwel "de uitkomst lag tussen 4 en 6"
  - Als ik zeg "tussen 4 en 6" zegt u: kans op "4" is 1/3.
- Stel dat we dit spel vaak herhalen.
  - Steeds als de uitkomst 4 is, heb ik een **keuze** wat ik ga vertellen
  - Ik kan het bijvoorbeeld zo doen: als de uitkomst 4 is zeg ik altijd "tussen 1 en 4" en **nooit** "tussen 4 en 6".
  - Uw uitspraak "kans op 4 is 1/3" is dan dus *niet* geijkt
- **Zolang u niet weet wat voor strategie ik volg, is het onmogelijk voor u een "correcte" kans te bepalen!**

Keuze en Overlap

- Ook 3-gevangenen probleem wordt veroorzaakt doordat cipier (soms) een keuze heeft in wat zij gaat vertellen
  - Als A overleeft (B en C geëxecuteerd), kan de cipier kiezen of zij B of C zegt
- De echte kansen zijn niet te bepalen als je niet weet wat voor strategie de cipier dan volgt

Keuze en Overlap

- Ook 3-gevangenen probleem wordt veroorzaakt doordat cipier (soms) een keuze heeft in wat zij gaat vertellen
  - Als A overleeft (B en C geëxecuteerd), kan de cipier kiezen of zij B of C zegt
- De echte kansen zijn niet te bepalen als je niet weet wat voor strategie de cipier dan volgt
  - Als de cipier een muntje gooit om te kiezen of zij B of C zegt, dan blijft A's overlevingskans 1/3
  - Als zij altijd B zegt als dat kan, en ze zegt B, wordt het 1/2

Keuze en Overlap

- Als er overlap is in de mogelijke antwoorden
  - dan werkt conditioneren niet
  - andere manieren om de kansen aan te passen (bijvoorbeeld negeren van de informatie) werken ook niet
  - dus kansen zijn niet meer goed te bepalen

3-Deuren (Quizmaster) Probleem

Monty Hall 1970

Google: Monty Hall 248000 hits;
David van Dantzig 610 hits

3-Deuren Probleem

- In de studio zijn drie deuren. Achter één deur staat een auto, achter beide andere deuren een geit. U gaat voor een van de deuren staan. Monty Hall opent een van de twee andere deuren, en laat zien dat er een geit achter zit. U mag nu nog wisselen naar de de deur die nog dicht is. Is dit verstandig?

3-Deuren Probleem

- Veranderen van deur is zeer verstandig.
- Vrijwel iedereen denkt echter in eerste instantie dat het niets uitmaakt
  - voor beide dichte deuren geldt immers dat de kans dat de prijs erachter zit, gelijk is aan ½ ?

3-Deuren Probleem

- Veranderen van deur is zeer verstandig.
- Vrijwel iedereen denkt echter in eerste instantie dat het niets uitmaakt
  - voor beide dichte deuren geldt immers dat de kans dat de prijs erachter zit, gelijk is aan ½ ?
- Wiskundig gezien is het 3-Gevangenen Probleem equivalent met het 3-Deuren Probleem. Toch:
  - In het 3-gevangenen probleem heeft vrijwel iedereen de juiste intuitie
  - In het 3-deuren probleem heeft vrijwel iedereen de verkeerde intuitie

Het 1-gevangene probleem

Den Haag 2004

- In 2004 werd verpleegkundige Lucia de B. in hoger beroep veroordeeld tot levenslang voor 7 moorden en 3 pogingen tot moord
- Zij zit al 5 jaar in de gevangenis, maar heeft nooit schuld bekend
- Ton Derksen, auteur van het boek “Lucia de B., Reconstructie van een gerechtelijke dwaling” heeft de zaak aangekaart bij de Commissie Posthumus-II (Evaluatie Afgesloten Strafzaken)

Het 1-gevangene probleem

- In deze zaak heeft statistiek een grote rol gespeeld
- Er waren veel meer “incidenten” tijdens Lucia’s diensten, dan tijdens diensten zonder Lucia
- Volgens de statisticus van het hof is de kans dat iemand bij toeval een dergelijk incidenten-patroon zou meemaken, kleiner dan 1 op 342.000.000

Het 1-gevangene probleem

Preciezer:

- De statisticus deed een “nulhypothese toets” met significantieniveau 1/10000
- Dwz: als hij iets ziet met een “*overschrijdingskans”* < 1/10000, dan zal hij zeggen “dat is geen toeval”
  - Hij vindt vervolgens een kans die nog veel kleiner is dan de gekozen drempel, dus hij claimt: “geen toeval”

“De” kans bestaat niet

- De statisticus belooft in feite:
  als we zijn methode herhaaldelijk toepassen, dan zeggen we maximaal 1 op de 10000 keer
  “dat kan geen toeval zijn” terwijl het wèl toeval is.
- Maar wat betekent “herhaaldelijk toepassen” hier?

“De” kans bestaat niet

- De statisticus belooft in feite:
  als we zijn methode herhaaldelijk toepassen, dan zeggen we maximaal 1 op de 10000 keer
  “dat kan geen toeval zijn” terwijl het wèl toeval is.
- Maar wat betekent “herhaaldelijk toepassen” hier?
- Doen we de uitspraak ‘wel/geen toeval’
  - Telkens als een verpleegkundige in het Juliana Kinderziekenhuis zoiets meemaakt als Lucia?
  - Of telkens als een verpleegkundige ergens in Nederland/op de wereld zoiets meemaakt?
  - Of telkens als er een **rechtzaak** is waarin het OM een nulhypothese toets gebruikt?

“De” kans bestaat niet

- Zonder een precieze context aan te geven, is de uitspraak “het kan geen toeval zijn want de kans is 1 op 342 miljoen” simpelweg betekenisloos
- Rechter was hierop geattendeerd maar begreep dit niet!

“De” kans bestaat niet

- In de drie jaar dat Lucia op de afdeling werkte, waren er 6 sterfgevallen. In de drie jaar ervoor, waren er 7.
  - Er is dus inderdaad iets mis met de statistische analyse
- Statistici beginnen zich te roeren!
  - R. Gill (winnaar Van Dantzig-prijs, oud MC/CWI’er) stuurde samen met mij een brief naar Commissie Posthumus (zie interview Vrij Nederland 6-12-06)
  - W. van Zwet (winnaar Van Dantzig-prijs, oud MC/CWI’er) geeft lezingen
  - (Statisticus van het hof is ook oud MC/CWI’er)

Dank voor Uw Aandacht!

- Zie ook “De verantwoordelijkheden van de statisticus”, D. van Dantzig, *Statistica 7*, 1954.

Keuze en Overlap

- Conditionele kansen hebben alleen betekenis als er geen overlap is in de informatie die verteld zou kunnen worden
  - de verzameling van gebeurtenissen waarop je kunt conditioneren moet **partitie** van uitkomstenruimte zijn
  - equivalent: je moet conditioneren op een gebeurtenis die je kunt schrijven als $Y = \cdot$ voor een fixed random variable Y
  - In maattheoretische kansrekening zijn conditionele kansen alleen *gedefinieerd* in het geval waarin ze betekenis hebben, maar niet velen weten dit