Data-Science 1

kansverdelingen

Karel de Grote Hogeschool

13/03/2024

Inhoud

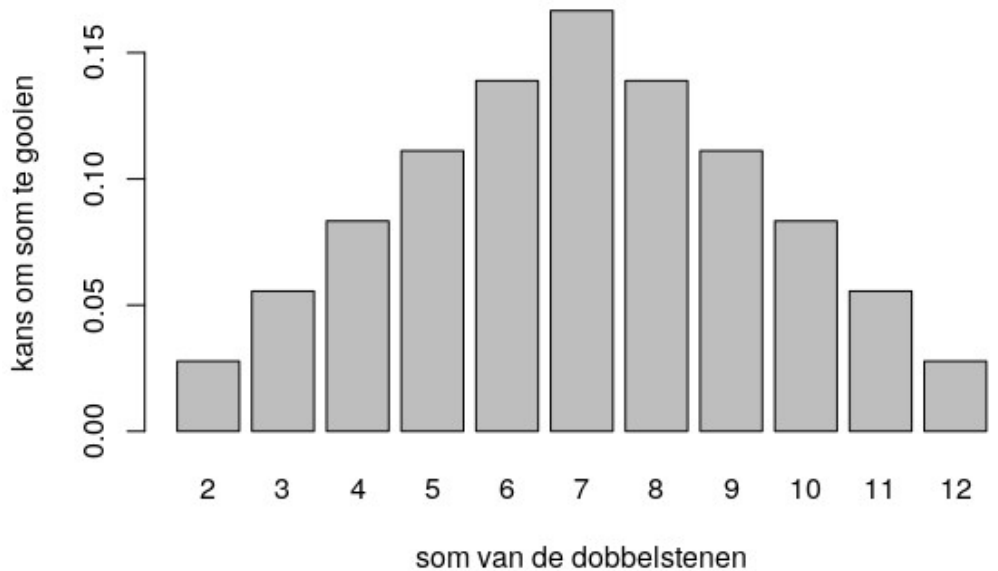
- discrete kansverdelingen
- gemiddelde en standaardafwijking
- continue kansverdelingen
  - normaalverdeling
  - χ^2 verdeling
  - F verdeling

Kansverdelingen

Voorbeeld

- gooi met 2 dobbelstenen
- wat is de kans dat ik 2, 3, 4, ... gooi?

Voorbeeld



waarde	kans
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

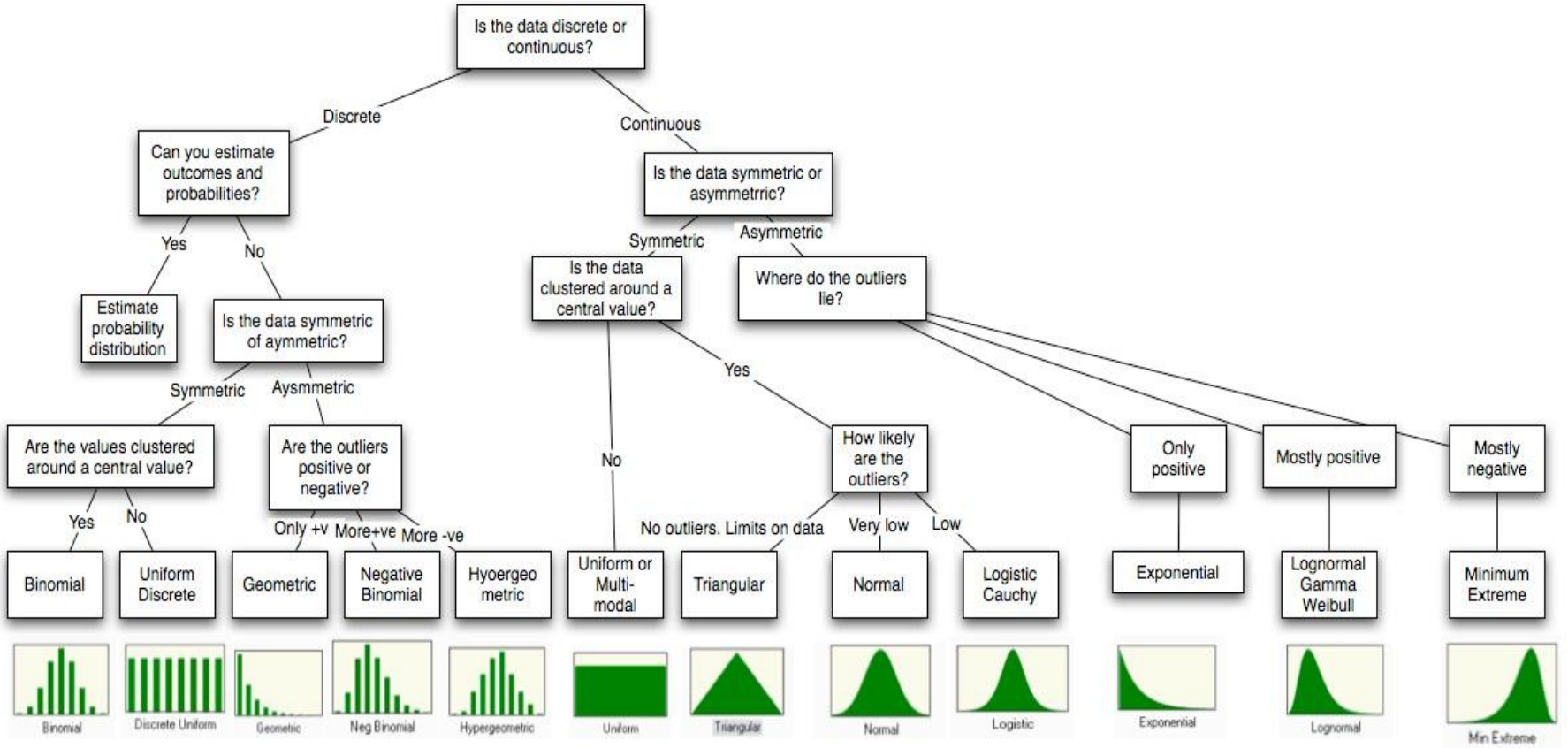
Kansverdeling

- let op: deze kansen werden niet bepaald door effectief te rollen met dobbelstenen
- de kansen geven weer wat we zouden verwachten als we met dobbelstenen zouden rollen

Kansverdeling

- is steeds een theoretisch model
  - geeft voor iedere mogelijke output, de kans dat dit voorkomt
  - voor te stellen in een tabel en/of barplot
- is dus steeds afhankelijk van de situatie
- geeft relatieve frequenties van een (theoretische) oneindige steekproef
- er zijn 2 soorten: discrete en continue kansverdelingen

Kansverdeling



Gemiddelde en standaardafwijking

2

Herhaling

- wat is een gemiddelde?
- wat is een standaardafwijking?
- hoe bereken ik deze?
- wat als ik relatieve frequenties als input heb?

waarde	absolute frequentie	relatieve frequentie
5	15	0,3
8	25	0,5
9	10	0,2
totaal	50	1,0

De verwachte waarden

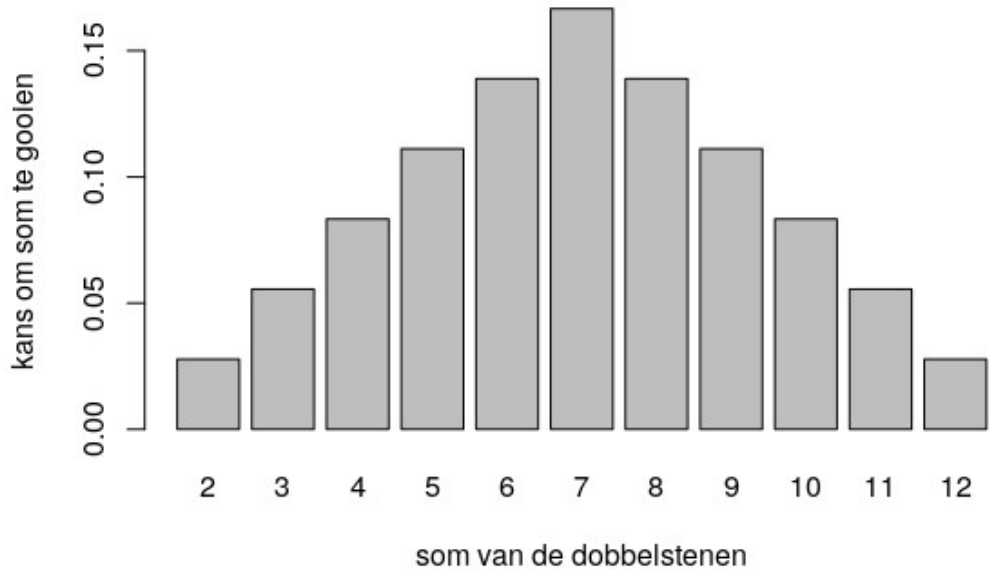
vervang rel. frequenties door kansen

$$\bar{x} = \sum_{i=1}^{n} x_i \cdot f_i \quad \Rightarrow \quad \mu = \sum_{i=1}^{n} x_i \cdot P(x_i)$$

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i} \quad \Rightarrow \quad \sigma = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(x_i)}$$

Voorbeeld

$\mu = 7$
$\sigma = 5{,}8333$



waarde	kans
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Normaalverdeling

4

Context

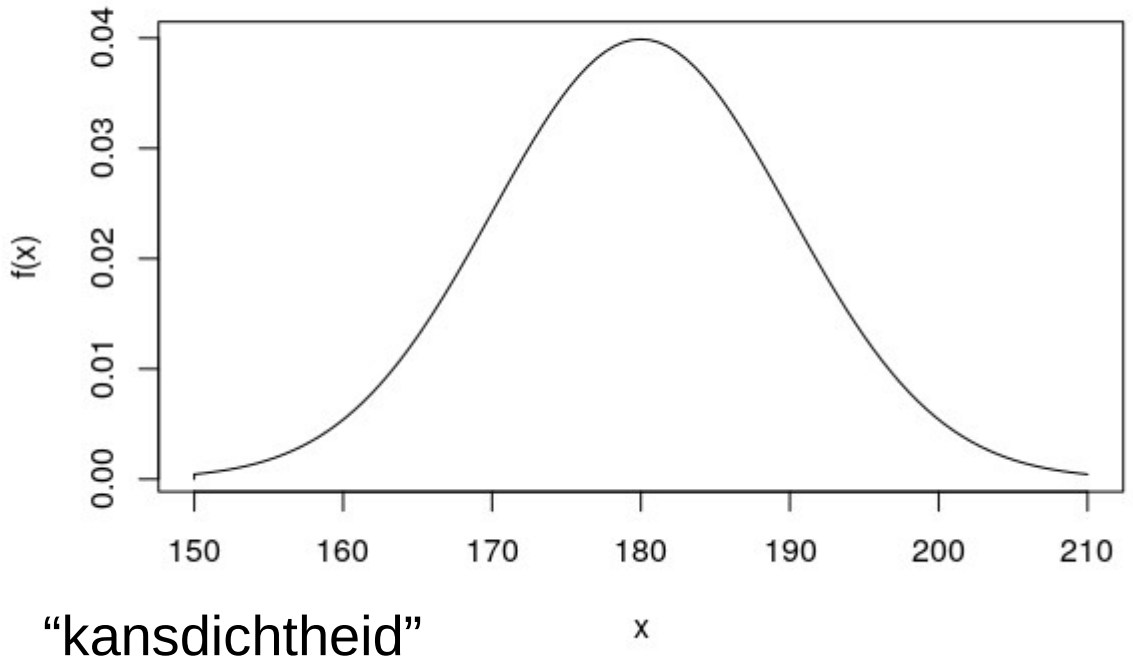
- je beschouwt een continue variabele
- je kent de verwachte waarde en de verwachte standaardafwijking
- de waarden zijn symmetrisch verdeeld rond de verwachte waarde
- je vraagt je af wat de kans is om een waarde te vinden tussen 2 grenzen

Voorbeeld

- lengte studenten 1e jaar
- we verwachten gemiddelde 180cm en standaardafwijking 10cm
- wat is de kans dat iemand exact 182,456532cm lang is?
- wat is dan de kans dat iemand tussen 175 en 180cm groot is?

De normaalverdeling

- geeft <u>niet</u> de kans om waarde x te vinden...

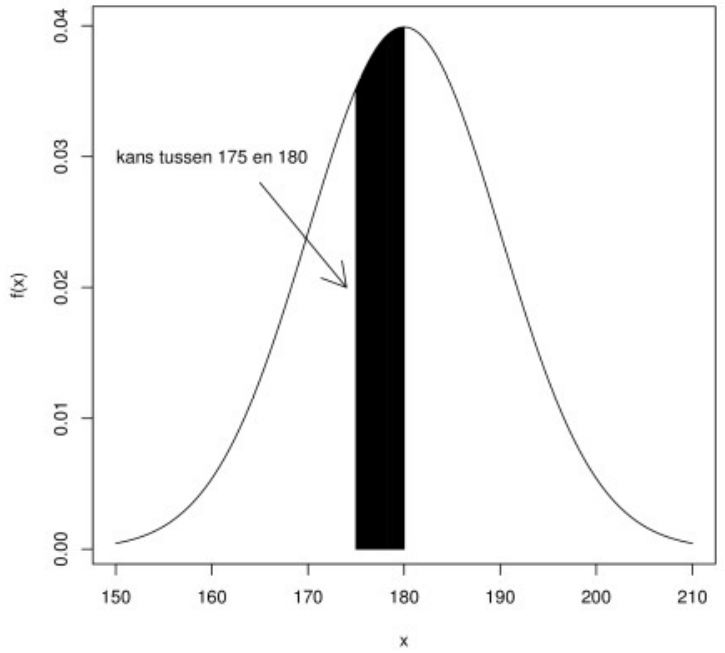


$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{\frac{-(x-\mu)^2}{2 \cdot \sigma^2}}$$

- “kansdichtheid”

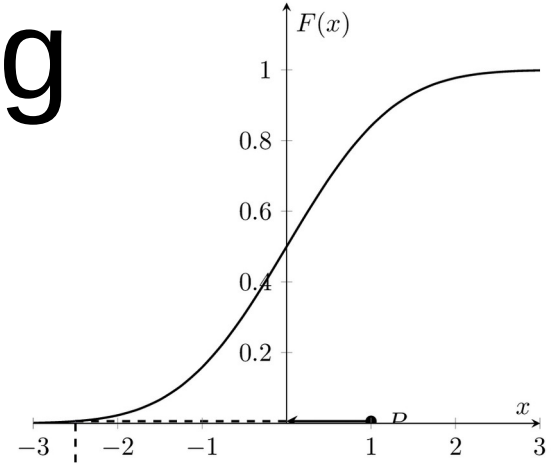
De normaalverdeling

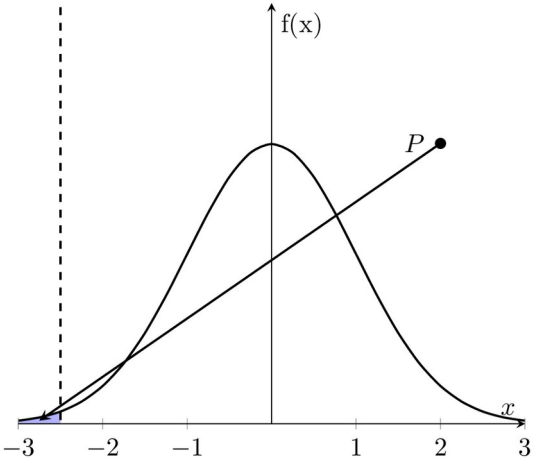
- kans om een waarde tussen twee grenzen te vinden =



De cumulatieve verdeling

- we gebruiken de cumulatieve verdeling om de oppervlakte te berekenen
- cumulatieve verdeling
  = oppervlakte van -oneindig tot x
- er is geen formule voor de cumulatieve verdeling...
- we gebruiken “geogebra”
  - wat is de kans om lengte tussen 175 en 180 te vinden?
  - wat is de kans om een lengte kleiner dan 160 te vinden?
  - wat is de kans om een lengte groter dan 190 te vinden?

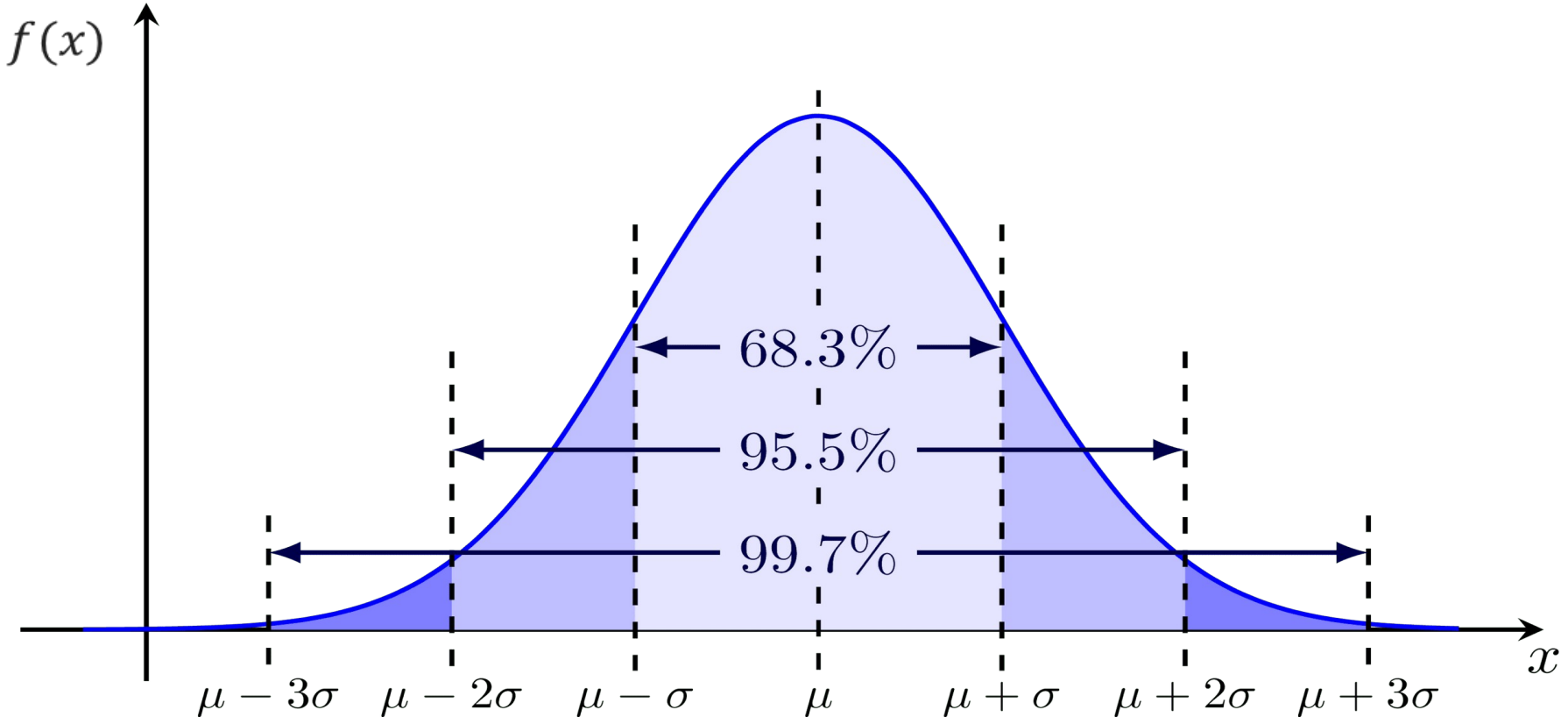




Opmerking

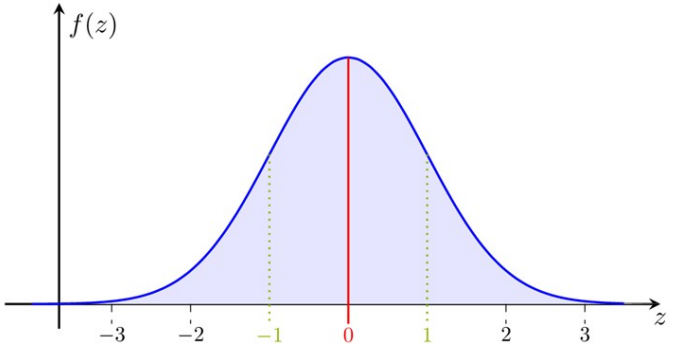
- normaalverdeling is niet helemaal juist
  - lengtes kunnen niet negatief worden...
  - wat is de kans dat je een negatieve lengte vindt volgens de verdeling?
  - kunnen we dit verwaarlozen?

Eigenschappen



De standaardnormaalverdeling

- stel dat gegevens normaal verdeeld zijn
- zet deze om naar Z-scores
- de standaardnormaalverdeling geeft de verdeling van deze Z-scores
  - gemiddelde = 0
  - standaardafwijking = 1
- voorbeeld: meet lengtes van alle studenten en zet ze om naar Z-scores
  - 95,5% kans dat Z-score tussen -2 en +2 zal liggen
  - iemand met Z-score +3 is dus uitzonderlijk lang



Oefeningen

Oefeningen

- zie Canvas