

Statistiek voor A.I.

College 6

Donderdag 27 September

2 Deductieve statistiek

Kansrekening

Vraag: Afghanistan

In het leger wordt uit een groep van 6 vrouwelijke en 14 mannelijke soldaten een eenheid van 7 soldaten samengesteld die naar Afghanistan zal worden uitgezonden.

Behalve voor de staf en de leden is de samenstelling van de eenheid geheim.

Er lekt uit dat de eenheid tenminste 6 mannen bevat. Wat is dan de kans dat het tevens een vrouw bevat?

Vraag: James Bond

Twee geheimagenten worden door de tegenpartij tot het volgende gedwongen. Voor hen staat een tafel met 4 martini's en 12 likeuren.

Agent 006 moet 2 likeuren kiezen en 1 daarvan opdrinken. Vervolgens moet agent 007 óf de andere door 006 gekozen likeur óf 1 martini opdrinken.

Er wordt hen gezegd dat 4 van de likeuren vergiftigd zijn en 1 van de martini's.

Agent 006 kiest een likeur, drinkt, en valt dood neer. Wat moet agent 007 doen: een likeur kiezen of een martini?

Vandaag :

- Permutaties en combinaties
- Belangrijke verdelingen
- Voorwaardelijke kansen

Permutaties en combinaties

Antwoord op een vraag: verjaardag

Voorbeeld Wat is de kans dat in een groep van 22 mensen tenminste twee mensen op dezelfde dag jarig zijn?

Het aantal verschillende mogelijke verdelingen van de verjaardagen is $365^{22} \approx 10^{56}$.

Het aantal mogelijke verdelingen waarbij iedereen op een verschillende dag jarig is, is $\frac{365!}{(365-22)!} = \frac{365!}{343!}$.

De gevraagde kans is $1 - \frac{365!}{(365-22)!365^{22}} \approx 0.52$.

Voorbeeld Bij poker krijgt een speler 5 kaarten uit 52. Is de kans op *full house* (3 kaarten met dezelfde waarde en 2 andere met dezelfde waarde) groter dan op *four of a kind* (4 kaarten met dezelfde waarde)?

Het aantal mogelijke delingen is $\binom{52}{5}$.

$$P(\text{four-of-a-kind}) = \frac{13 \cdot \binom{4}{4} \binom{48}{1}}{\binom{52}{5}} = \frac{13 \cdot 48}{\binom{52}{5}} = 0.00024.$$

$$P(\text{full house}) = \frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}}{\binom{52}{5}} = 0.0014.$$

Verwachtingswaarde en variantie

Definitie Voor een discrete stochast X met eindig veel waarden $X_1, \dots, X_k$ is de *verwachtingswaarde* van X :

$$E(X) = \sum_{i=1}^k P(X = X_i) X_i.$$

Stelling Bij het beschouwen van een eindige populatie en een parameter met eindig veel waarden $X_1, \dots, X_k$, is de verwachtingswaarde van de bijbehorende stochast X gelijk aan het gemiddelde μ van de parameter in de populatie: $E(X) = \mu$.

Bewijs Zij N het aantal elementen in de populatie en f_i de frequentie (het aantal) van de elementen in de populatie met waarde X_i .

Dan geldt dat

$$\mu = \frac{\sum_{i=1}^k f_i X_i}{N} = \sum_{i=1}^k \frac{f_i X_i}{N} = \sum_{i=1}^k P(X = X_i) X_i = E(X).$$

Verwachtingswaarde en variantie

Definitie Voor een discrete stochast X met eindige veel waarden $X_1, \dots, X_k$ is de *variantie* van X :

$$\text{Var}(X) = E((X - E(X))^2) = \sum_{i=1}^k P(X = X_i)(X_i - E(X))^2.$$

Stelling Bij het beschouwen van een eindige populatie en een parameter met eindig veel waarden $X_1, \dots, X_k$, is de variantie van de bijbehorende stochast X gelijk aan de variantie σ^2 van de parameter in de populatie: $\text{Var}(X) = \sigma^2$.

Bewijs Zij N het aantal elementen in de populatie en f_i de frequentie (het aantal) van de elementen in de populatie met waarde i . Omdat de gehele populatie beschouwd wordt, en niet een steekproef, geldt:

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

Uit de vorige stelling volgt dat voor het gemiddelde μ van de populatie geldt dat $\mu = E(X)$. Daaruit volgt dat

$$\sigma^2 = \frac{\sum_{i=1}^k f_i (X_i - \mu)^2}{N} = \sum_{i=1}^k \frac{f_i (X_i - E(X))^2}{N} = \sum_{i=1}^k P(X = X_i) (X_i - E(X))^2 = \text{Var}(X)$$

Belangrijke verdelingen

Belangrijke verdelingen: uniform

Definitie Een discrete stochast heeft een *uniforme verdeling* als alle uitkomsten gelijke kans hebben, d.w.z. als de kansverdeling constant is:

$$\forall x \in S \forall y \in S : f(x) = f(y).$$

Definitie Een continue stochast heeft een *uniforme verdeling* als de kansdichtheid een constante functie is:

$$\forall x \in S \forall y \in S : f(x) = f(y).$$

Voorbeeld

- De stochast waarvan de waarden de uitkomsten bij het werpen van een dobbelsteen zijn, heeft een uniforme verdeling: $f(i) = P(X = i) = \frac{1}{6}$.
- De stochast die de som van de getallen is bij het werpen van twee dobbelstenen heeft geen uniforme verdeling, omdat bijv.

$$P(X = 2) = \frac{1}{36} \quad P(X = 4) = \frac{3}{36} = \frac{1}{12}.$$

Voorbeeld De volgende continue stochasten zijn uniform verdeeld:

- De continue stochast bij het willekeurig kiezen van een punt in $[0,1]$ heeft kansdichtheid:

$$f(x) : [0, 1] \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = 1.$$

- De continue stochast bij het willekeurig kiezen van een punt in een reëel interval $[r, s]$ heeft kansdichtheid:

$$f(x) : [r, s] \rightarrow \mathbb{R}_{\geq 0} \quad f(x) = \frac{1}{s - r}.$$

- De continue stochast bij het willekeurig kiezen van een punt in een rechthoek $[r, s] \times [u, v]$ heeft kansdichtheid:

$$f(x, y) : [r, s] \times [u, v] \rightarrow \mathbb{R}_{\geq 0} \quad f(x, y) = \frac{1}{(s - r) \cdot (v - u)}.$$

Belangrijke verdelingen: uniform

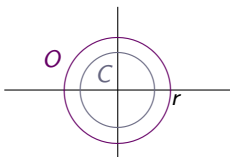
Als een experiment bestaat uit het willekeurig kiezen van een punt in een oppervlak (lijnstuk), dan is de bijbehorende stochast uniform verdeeld.

Omdat moet gelden $\int_S f(x, y) dx dy = 1$ ($\int_S f(x) dx = 1$) en f constant is, geldt

$$f(x, y) = \frac{1}{\text{grootte oppervlak}} \quad (f(x) = \frac{1}{\text{lengte lijnstuk}}).$$

Antwoord op een vraag: cirkels

Voorbeeld Zij X de continue stochast bij het willekeurig kiezen van een punt in een cirkel O met straal r :



Dan is X uniform verdeeld, met de volgende kansdichtheid:

$$f(x, y) : \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq r^2\} \rightarrow \mathbb{R}_{\geq 0} \quad f(x, y) = \frac{1}{\pi r^2}.$$

Voor elke cirkel C met straal $s \leq r$ geldt dat

$$P(X \in C) = \frac{\text{oppervlak } C}{\text{oppervlak } O} = \frac{\pi s^2}{\pi r^2} = \frac{s^2}{r^2}.$$

Belangrijke verdelingen: uniform

Voorbeeld Een stochast die als waardes de tijdsduur (in dagen) tussen twee opeenvolgende emissies van een isotoop heeft:

$$f(t) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \quad f(t) = 7e^{-7t},$$

is niet uniform verdeeld, omdat $\frac{1}{5}e^{-\frac{x}{5}}$ geen constante functie is.

De kans dat de tijdsduur tussen twee opeenvolgende emissies tussen de 2 en 3 dagen ligt is

$$P(2 \leq X \leq 3) = \int_2^3 7e^{-7t} dt = -\frac{1}{e^{-7t}} \Big|_2^3 = \frac{1}{e^{14}} - \frac{1}{e^{21}} \approx 9 \cdot 10^{-7}.$$

Definitie Een *Bernoulli experiment* is een experiment met twee uitkomsten, die doorgaans met *succes* en *falen*, of 1 en 0, worden aangeduid.

De kans op succes wordt vaak aangeduid met p , en de kans op falen met q . Dus $q = (1 - p)$.

Een *Bernoulli verdeling* is de verdeling van een stochast behorende bij een Bernoulli experiment.

Voorbeeld Dit zijn voorbeelden van Bernoulli experimenten:

- Het (1 maal) werpen van een munt: $p = q = 0.5$.
- Het willekeurig beantwoorden van een multiple choice vraag die vijf mogelijke antwoorden heeft waarvan er één correct is: $p = 0.2$. Dus $q = 0.8$.
- Het bepalen of een individu wel of niet met een virus is besmet.

Definitie Een *Bernoulli proces* of een *binomiaal experiment* is een experiment dat bestaat uit het een aantal maal herhalen van een Bernoulli experiment, waarbij de stochast het aantal malen succes is.

Als het Bernoulli experiment n maal wordt herhaald dan kan de bijbehorende stochast dus de waarden $0, 1, \dots, n$ aannemen, die staan voor respectievelijk 0 maal succes, 1 maal succes, $\dots$, n maal succes.

De verdeling van de stochast heet een *binomiale verdeling*. Als de stochast X heet, dan staat $P(X = k)$ voor de kans op k successen.

Stelling Wanneer X een binomiale verdeling heeft, waarbij het onderliggende Bernoulli experiment n maal wordt herhaald, dan geldt dat

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (\text{de kans op } k \text{ maal succes}).$$

Belangrijke verdelingen: binomiaal

Voorbeeld X is het aantal malen K bij het 4 maal werpen van een zuivere munt. De stochast X heeft dus een binomiale verdeling, waarbij $n = 4$, $p = 0.5$.

$$P(X = 0) = \binom{4}{0}(0.5)^4 = 0.0625 \quad P(X = 1) = \binom{4}{1}(0.5)^4 = 0.25$$

$$P(X = 2) = \binom{4}{2}(0.5)^4 = 0.375 \quad P(X = 3) = \binom{4}{3}(0.5)^4 = 0.25$$

$$P(X = 4) = \binom{4}{4}(0.5)^4 = 0.0625.$$

De kans om 4 maal K te gooien is $P(X = 4) = (0.5)^4 = 0.0625$.

De kans om 4 maal M te gooien is $P(X = 0) = (0.5)^4 = 0.0625$.

De kans om precies 2 maal K te gooien is $P(X = 2) = \binom{4}{2}(0.5)^4$.

De kans om ten hoogste 2 maal K te gooien is

$$\sum_{i=0}^2 P(X = i) = \sum_{i=0}^2 \binom{4}{i} \cdot (0.5)^4 = 0.0625 + 0.25 + 0.375 = 0.6875.$$

Voorbeeld X is het aantal malen succes bij het willekeurig beantwoorden van drie multiple choice vragen die alledrie vijf mogelijke antwoorden hebben, waarvan één correct. De verdeling van de stochast X is binomiaal, waarbij $p = 0.2$, $q = 0.8$ en $n = 3$:

$$P(X = 0) = \binom{3}{0}(0.2)^0(0.8)^3 = 0.512 \quad P(X = 1) = \binom{3}{1}(0.2)(0.8)^2 = 0.384$$

$$P(X = 2) = \binom{3}{2}(0.2)^2(0.8) = 0.096 \quad P(X = 3) = \binom{3}{3}(0.2)^3(0.8)^0 = 0.008.$$

Belangrijke verdelingen: binomiaal

Voorbeeld Een bevolkingsonderzoek onder 2000 mensen waarbij bepaald wordt of iemand wel of niet op de PvdA heeft gestemd. X is het aantal mensen in de steekproef dat op de PvdA stemt. X heeft een binomiale verdeling, waarbij $p = 0.26$ en $q = 0.74$ en $n = 2000$.

$$P(X = 2000) = \binom{2000}{2000} (0.26)^{2000} \approx 0$$

$$P(X = 1999) = \binom{2000}{1999} (0.26)^{1999} (0.74) \approx 0$$

$$P(X = 1998) = \binom{2000}{1998} (0.26)^{1998} (0.74)^2 \approx 0$$

$$\vdots$$

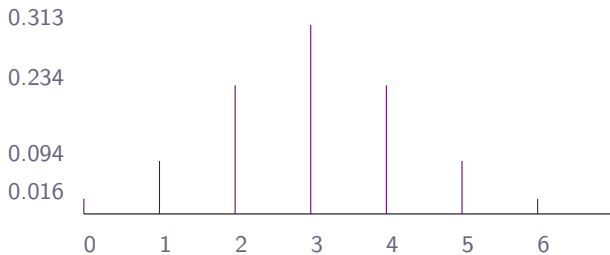
$$P(X = 520) = \binom{2000}{520} (0.26)^{520} (0.74)^{1480}$$

$$\vdots$$

$$P(X = 5) = \binom{2000}{5} (0.26)^5 (0.74)^{1995} \approx 10^{-250}$$

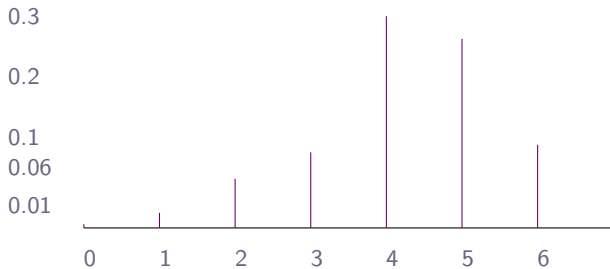
$$P(X = 0) = \binom{2000}{0} (0.74)^{2000} \approx 10^{-262}.$$

Belangrijke verdelingen: binomiaal



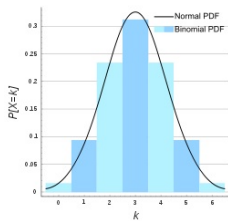
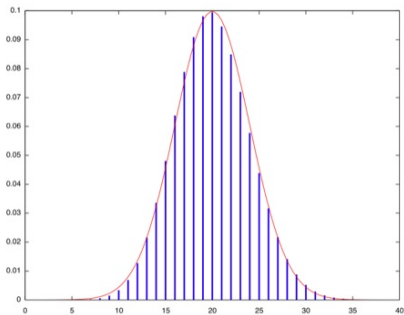
De binomiale verdeling voor $n = 6$ en $p = 0.5$

Belangrijke verdelingen: binomiaal

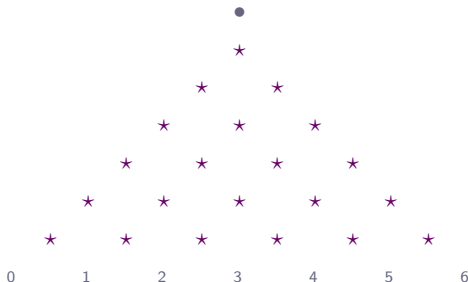


De binomiale verdeling voor $n = 6$ en $p = 0.7$

Belangrijke verdelingen: binomiaal



Belangrijke verdelingen: het bord van Galton en de binomiale verdeling

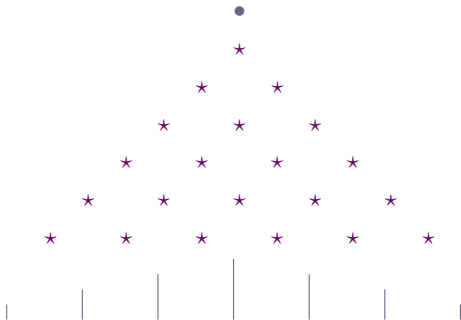


Het grijze balletje valt bij elke spijker met gelijke kans naar links of naar rechts.

De kans dat de bal in vakje k valt is $\binom{6}{k} \left(\frac{1}{2}\right)^6$.

De kans dat bij een Galton bord van diepte n het balletje in vakje k valt is $\binom{n}{k} \left(\frac{1}{2}\right)^n$.

Belangrijke verdelingen: het bord van Galton en de binomiale verdeling

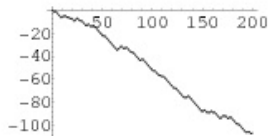
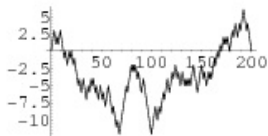
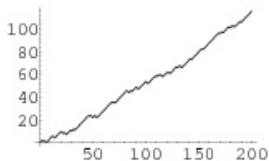


De stapels ballen die zich beneden vormen als er herhaaldelijk een balletje boven het bord wordt losgelaten vormen de binomiale verdeling $\binom{n}{k} \left(\frac{1}{2}\right)^n$.

Belangrijke verdelingen: 1-dimensionale random walks en de binomiale verdeling

Voorbeeld Langs een lijn wordt telkens willekeurig een stap naar links of naar rechts genomen. Het aantal stappen X naar rechts heeft een binomiale verdeling: de kans dat er k stappen naar rechts worden genomen is $P(X = k) = \binom{n}{k}(0.5)^n$.

Als Y het aantal stappen naar links is, dan is de afgelegde afstand t.o.v. het beginpunt $X - Y$. Dus een positieve afstand is naar rechts gericht en een negatieve afstand naar links. Hier zijn de grafieken voor $X - Y$ bij drie experimenten waarbij $n = 200$ en respectievelijk $p = 0.9$, $p = 0.5$ en $p = 0.1$:

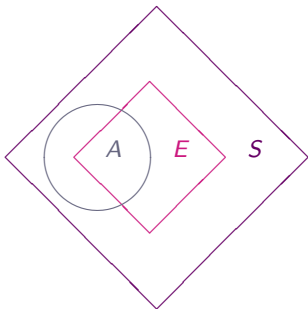


Voorwaardelijke kansen

Voorwaardelijke kansen

Vraag: Wat is de kans op gebeurtenis A gegeven dat gebeurtenis E (*evidence*) heeft plaatsgevonden?

Intuïtie: De nieuwe uitkomstenruimte is E en de nieuwe kans is $\frac{\#(A \cap E)}{\#E}$.



Voorwaardelijke kansen

Definitie Gegeven een discrete stochast X en gebeurtenissen A en E , is de *voorwaardelijke (conditional) kans op A gegeven E* :

$$P(A|E) \equiv_{\text{def}} \frac{P(A \cap E)}{P(E)}.$$

Definitie Gegeven een continue stochast X met kansdichtheid f , is de *voorwaardelijke kansdichtheid gegeven E* :

$$f(x|E) = \begin{cases} \frac{f(x)}{P(E)} & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

$P(A|E)$ wordt dan zo berekend:

$$P(A|E) = \int_A f(x|E)dx = \int_{A \cap E} \frac{f(x)}{P(E)}dx = \frac{P(A \cap E)}{P(E)}.$$

Stelling In beide gevallen geldt:

$$P(A \cap E) = P(A|E)P(E) = P(E|A)P(A).$$

Voorwaardelijke kansen

Voorbeeld Een munt wordt drie maal geworpen. Wat is de kans op 2 maal K gegeven dat er de eerste maal K geworpen is. X is het aantal malen K , Y is de waarde van de eerste worp.

kkk kkm	mmm mmk
kmk kmm	mkm mkk

$$P(X = 2 | Y = K) = \frac{P(X = 2 \cap Y = K)}{P(Y = K)} = \frac{\binom{2}{1}(0.5)^3}{0.5} = 0.5.$$

Vergelijk: de kans op 2 maal K is $P(X = 2) = \binom{3}{1}(0.5)^3 = 0.25$.

Voorbeeld De burens hebben 80 cd's die willekeurig worden afgespeeld, 20 van Beethoven, 12 met strijkkwartetten, waarvan 2 met strijkkwartetten van Beethoven.

Ik hoor muziek bij de burens, een strijkkwartet, noem deze gebeurtenis S . Wat is de kans dat het een strijkkwartet van Beethoven (B) is?

$$P(B|S) = \frac{P(B \cap S)}{P(S)} = \frac{\frac{2}{80}}{\frac{12}{80}} = \frac{1}{6}.$$

Vergelijk: $P(B) = \frac{20}{80} = 0.25$.

Antwoord op een vraag: Afghanistan

In het leger wordt uit een groep van 6 vrouwelijke en 14 mannelijke soldaten een eenheid van 7 soldaten samengesteld die naar Afghanistan zal worden uitgezonden. Behalve voor de staf en de leden is de samenstelling van de eenheid geheim.

Er lekt echter uit dat de eenheid tenminste 6 mannen bevat. Wat is de kans dat het tevens een vrouw bevat? (Y is het aantal mannen in de eenheid en X het aantal vrouwen.)

$$P(X = 1 | Y \geq 6) = \frac{P(X = 1 \cap Y \geq 6)}{P(Y \geq 6)} = \frac{\frac{\binom{6}{1}\binom{14}{6}}{\binom{20}{7}}}{\frac{\binom{14}{6}\binom{6}{1} + \binom{14}{7}}{\binom{20}{7}}} = \frac{\binom{6}{1}\binom{14}{6}}{\binom{14}{6}\binom{6}{1} + \binom{14}{7}} = \frac{18018}{21450} = 0.84.$$

Vergelijk: de kans dat de eenheid precies 1 vrouw bevat is

$$P(X = 1) = \frac{\binom{6}{1}\binom{14}{6}}{\binom{20}{7}} = \frac{18018}{77520} = 0.23.$$

Voorbeeld Twee geheimagenten worden door de tegenpartij tot het volgende gedwongen. Voor hen staat een tafel met 4 martini's en 12 likeuren. Agent 006 moet 2 likeuren kiezen en 1 daarvan opdrinken. Vervolgens moet agent 007 óf de andere door 006 gekozen likeur óf 1 martini opdrinken. Er wordt hen verteld dat 4 likeuren vergiftigd zijn en 1 martini.

Agent 006 kiest 2 likeuren, drinkt 1 daarvan, en valt dood neer. Wat moet agent 007 doen: de andere likeur óf een martini kiezen?

Stel dat $l_1, \dots, l_{12}$ de likeuren zijn en dat $L = \{l_1, l_2, l_3, l_4\}$ de verzameling van vergiftigde likeuren is. Y is de likeur die 006 kiest, en X die die voor 007 overblijft.

De kans dat X vergiftigd is gegeven dat Y dat is, is:

$$P(X \in L | Y \in L) = \frac{P(X \in L \cap Y \in L)}{P(Y \in L)} = \frac{\frac{4 \times 3}{12 \times 11}}{\frac{4 \times 3 + 4 \times 8}{12 \times 11}} = \frac{4 \times 3}{4 \times 3 + 4 \times 8} = 0.27.$$

Als 007 de martini kiest is de kans dat die vergiftigd is 0.25, kleiner dan de kans dat de likeur vergiftigd is.

Daarom drinkt James Bond altijd martini's.

Finis