STATISTIEK VOOR HET SECUNDAIR ONDERWIJS

Kansrekening voor de derde graad

Werktekst voor de leerling

Prof. dr. Herman Callaert

Hans Bekaert
Cecile Goethals
Lies Provoost
Marc Vancaudenberg

Kansrekening voor de derde graad

Deze tekst sluit aan op de tekst: "Kansrekening voor de tweede graad".

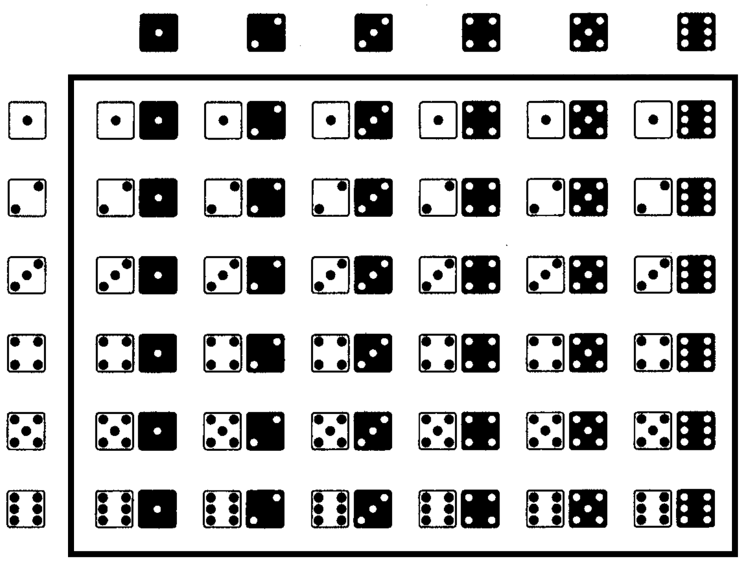
6. Samengestelde experimenten

6.1. De uitkomstenverzameling van samengestelde experimenten

Bij een samengesteld experiment voer je meerdere keren na elkaar een enkelvoudig experiment uit en je noteert telkens het resultaat. Al die resultaten samen, ***in de volgorde waarin ze zijn uitgevoerd***, vormen dan ***één*** van de mogelijke uitkomsten van het samengestelde experiment.
Voorbeelden maken duidelijk waarover het gaat.

6.1.1. Dobbelstenen gooien

Gooi een dobbelsteen en gooi die daarna nog eens. Stel je voor dat je dit echt doet. Eén mogelijke uitkomst van dit experiment is (3,5). Let op: dit is geen verzamelingennotatie {3,5} waarbij je zegt dat er een 3 en een 5 tussen de uitkomsten zat. Neen, (3,5) is de notatie voor een geordend tweetal die zegt: de eerste worp leverde een 3 en de tweede een 5 [en niet omgekeerd]. Je kan voor dit experiment ook twee dobbelstenen gelijktijdig gooien, een witte en een zwarte. Je noteert dan het resultaat telkens in dezelfde volgorde: eerst de witte en dan de zwarte.



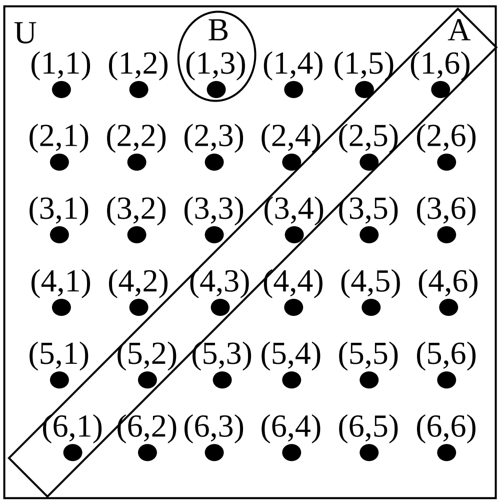
De uitkomstenverzameling bestaat nu uit 36 verschillende uitkomsten waarbij elke uitkomst een geordend tweetal is:

U = {(1,1),(1,2),(1,3), … ,(2,1),(2,2), … ,(6,6)}.

Als A de gebeurtenis is "de som van de ogen is 7" en B de gebeurtenis "eerst een 1 en dan een 3" dan is [bekijk de figuur]:

A = {(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)}

B = {(1,3)}.



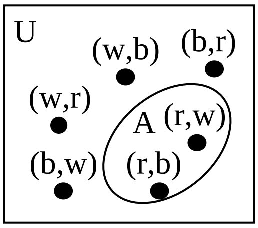
6.1.2. Kaartjes trekken

In een doos liggen 3 gekleurde kaartjes (b = blauw, r = rood, w = wit).
Het experiment is als volgt. Trek een kaartje en noteer de kleur. Leg het niet terug in de doos. Trek daarna opnieuw een kaartje en noteer opnieuw de kleur.

Stel je voor dat je dit echt doet. Eén van de mogelijke uitkomsten is (r,w) wat betekent dat je eerst het rode kaartje getrokken hebt en daarna het witte. De uitkomst (r,r) is niet mogelijk want als je eerst het rode kaartje trekt, dan kan je de tweede keer geen rood hebben want je legt het kaartje niet terug.

De uitkomstenverzameling voor dit experiment is
U = {(b,r),(b,w),(r,b),(r,w),(w,b),(w,r)}.

Als A de gebeurtenis is: “het eerste kaartje is rood”,
dan is A = {(r,b),(r,w)} [bekijk de figuur].



Opdracht 17

Gooi een muntstuk en noteer of het op kruis (k) valt of op munt (m). Gooi het daarna nog eens en noteer terug of je kruis of munt hebt.

- Wat is de uitkomstenverzameling van dit experiment? Gebruik de juiste notatie.

- De gebeurtenis A is “twee keer hetzelfde” en B is “minstens één keer munt”. Schrijf A en B in de juiste notatie. Teken ook een figuur.

6.2. Het kansmodel van samengestelde experimenten

Vooraleer je kan rekenen met kansen moet je eerst het kansmodel (= alle mogelijke uitkomsten met hun kansen) kennen. Dat was zo voor enkelvoudige experimenten en dat is ook zo voor samengestelde experimenten.

Elke uitkomst van een samengesteld experiment bestaat uit een opeenvolging van resultaten van een enkelvoudig experiment ***waarbij de volgorde van belang is***.

Om geen enkele uitkomst te vergeten, kan je gebruik maken van een kansboom waarbij je stap na stap enkelvoudige experimenten uitvoert. Daarbij heb je twee grote systemen:

1. de kans van een uitkomst bij een volgende stap hangt af van wat er bij de vorige stap gebeurde. Een typisch voorbeeld is: “trekken zonder terugleggen” waarbij de doos waaruit je trekt bij elke stap anders is (want wat je eruit haalt, leg je niet terug).

2. de kans van een uitkomst bij een volgende stap hangt niet af van wat er bij de vorige stap gebeurde [onafhankelijkheid]. Een voorbeeld is: “trekken met terugleggen” waar je bij elke stap uit identiek dezelfde doos trekt (want wat je getrokken hebt, leg je er terug in).

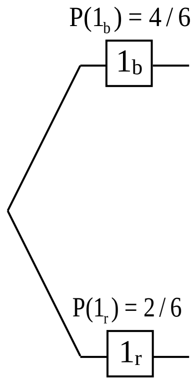
6.2.1. Trekken zonder terugleggen

In een doos zitten zes ballen die identiek zijn behalve de kleur: er zijn 4 blauwe (b) en 2 rode (r).

Jij zal nu twee keer na elkaar een enkelvoudig experiment uitvoeren. Trek lukraak een bal en noteer de kleur. Leg hem niet terug. Trek dan opnieuw lukraak een bal en noteer de kleur.

Aangezien het experiment uit twee stappen bestaat is het goed dat je in de notatie expliciet aangeeft over welke stap het gaat. Bij de eerste trekking een blauwe bal vinden, noteer je als 1_b en de kans die daarbij hoort is $P(1_b) = 4/6$.
Op dezelfde manier zie je dat $P(1_r) = 2/6$.

De eerste stap kan je nu al gebruiken om een kansboom te starten.



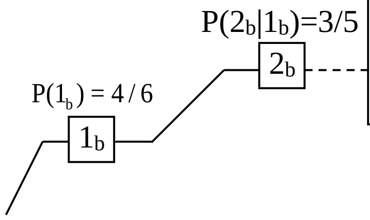
Wat je bij de tweede trekking kan vinden en met welke kans, hangt af van wat je de eerste keer getrokken hebt.

Als je de eerste keer een blauwe bal had, dan heb je nu een doos waarin 3 blauwe en 2 rode ballen zitten. De kans om nu blauw te trekken, is een ***voorwaardelijke*** kans, gegeven dat je de eerste keer blauw had. Bij de notatie van voorwaardelijke kansen gebruik je het verticale streepje | (= gegeven dat). De kans om de tweede keer blauw te trekken (2_b) gegeven dat je de eerste keer blauw had is $P(2_b|1_b) = 3/5$.

Het samengestelde experiment heeft maar twee stappen. De uitkomst die je hierboven gevonden hebt gaat over een doorsnede: de eerste keer blauw en tegelijkertijd de tweede keer blauw. Dat kan je noteren als “1_b en 2_b”.

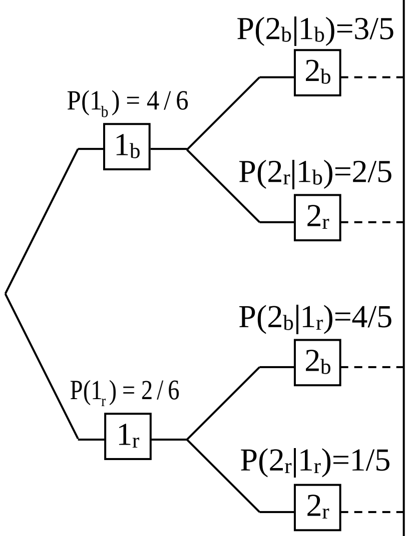
De kans van “1_b en 2_b” bereken je met de productregel: $P(1_b \text{ en } 2_b) = P(1_b) \cdot P(2_b|1_b) = \frac{4}{6} \cdot \frac{3}{5} = \frac{12}{30}$.

Je kan nu al een eerste stukje van de kansboom tekenen, samen met een eerste uitkomst en haar kans. Voor “1_b en 2_b” kan je ook de notatie (b,b) gebruiken want (,) is de notatie voor een ***geordend*** tweetal.



Uitkomst	Kans
“1_b en 2_b” = (b,b)	$P((b,b)) = \frac{4}{6} \cdot \frac{3}{5} = \frac{12}{30}$

De volledige kansboom zie je hieronder. Bekijk hem nauwkeurig en zorg ervoor dat je hem goed begrijpt. Alle kansen zijn uitgedrukt in dertigsten zodat je ze eenvoudig met elkaar kan vergelijken. Zo zie je ook snel dat de som van alle kansen gelijk is aan 1.



Uitkomst	Kans
“1_b en 2_b” = (b,b)	$P((b,b)) = \frac{4}{6} \cdot \frac{3}{5} = \frac{12}{30}$
“1_b en 2_r” = (b,r)	$P((b,r)) = \frac{4}{6} \cdot \frac{2}{5} = \frac{8}{30}$
“1_r en 2_b” = (r,b)	$P((r,b)) = \frac{2}{6} \cdot \frac{4}{5} = \frac{8}{30}$
“1_r en 2_r” = (r,r)	$P((r,r)) = \frac{2}{6} \cdot \frac{1}{5} = \frac{2}{30}$

De kansboom helpt je om het volledige kansmodel van een samengesteld experiment te vinden. Dat is hier:

uitkomst	(b,b)	(b,r)	(r,b)	(r,r)
kans	12/30	8/30	8/30	2/30

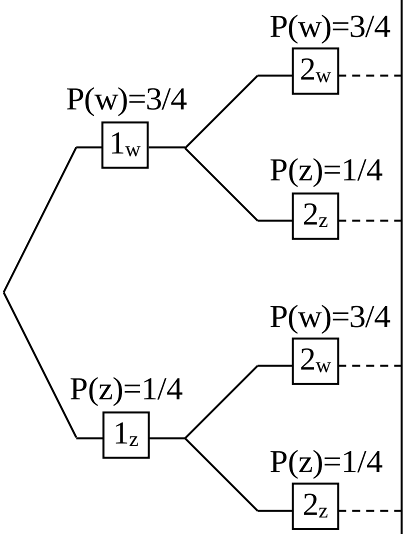
6.2.2. Trekken met terugleggen

In een doos zitten 4 kaartjes die volledig aan elkaar gelijk zijn behalve hun kleur. Er zijn er 3 witte en 1 zwart. Het samengestelde experiment bestaat uit 2 stappen: trek lukraak een eerste kaartje en noteer de kleur. Leg het kaartje terug in de doos. Trek dan opnieuw lukraak een kaartje en noteer de kleur.

Bij de eerste trekking kan je ofwel wit (1_W) hebben ofwel zwart (1_Z). Aangezien je lukraak trekt is $P(1_W) = 3/4$ en $P(1_Z) = 1/4$.

Bij de tweede trekking sta je daar terug met dezelfde doos. Het is alsof je helemaal opnieuw begint. De uitkomst van de tweede trekking hangt niet af van het kaartje dat je de eerste keer getrokken hebt. Als je eerst wit (1_W) had en de tweede keer zwart (2_Z) dan is $P(2_Z|1_W) = P(2_Z)$ [onafhankelijkheid]. In de kansboom kan je alle voorwaardelijke kansen vervangen door gewone kansen. Bovendien trek je telkens uit dezelfde doos zodat de kans op zwart (of op wit) elke keer dezelfde is. Daarom kan je de notatie vereenvoudigen en gewoon P(z) schrijven in plaats van $P(1_Z)$ en $P(2_Z)$. Op dezelfde manier kan je bij elke stap de kans op wit noteren als P(w) want ook die verandert niet.

De volledige kansboom zie je hieronder. Let erop dat je de verkorte kansnotatie in zo'n kansboom alleen maar mag gebruiken als de opeenvolgende stappen onafhankelijk zijn en je bij elke stap dezelfde procedure herhaalt.



Uitkomst	Kans
"1_W en 2_W" = (w,w)	$P((w,w)) = \frac{3}{4}\cdot\frac{3}{4} = \frac{9}{16}$
"1_W en 2_Z" = (w,z)	$P((w,z)) = \frac{3}{4}\cdot\frac{1}{4} = \frac{3}{16}$
"1_Z en 2_W" = (z,w)	$P((z,w)) = \frac{1}{4}\cdot\frac{3}{4} = \frac{3}{16}$
"1_Z en 2_Z" = (z,z)	$P((z,z)) = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}$

Het volledige kansmodel kan je kort samenvatten in een tabel. Bemerk opnieuw dat de som van de kansen gelijk is aan 1.

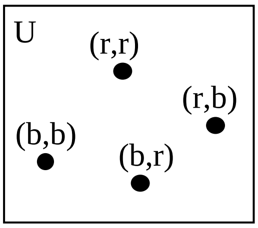
uitkomst	(w,w)	(w,z)	(z,w)	(z,z)
kans	9/16	3/16	3/16	1/16

6.3. Rekenen met kansen bij samengestelde experimenten

Van zodra je het kansmodel kent (alle mogelijke uitkomsten samen met hun kansen), verloopt rekenen met kansen bij samengestelde experimenten hetzelfde als bij enkelvoudige experimenten.

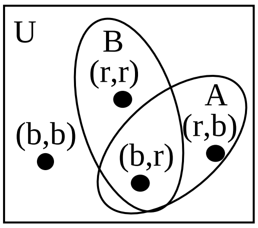
Het volledige kansmodel voor het experiment met de blauwe en rode ballen (twee keer trekken zonder terugleggen) heb je vroeger reeds opgesteld:

uitkomst	(b,b)	(b,r)	(r,b)	(r,r)
kans	12/30	8/30	8/30	2/30



De ***uitkomstenverzameling*** U = {(b,b),(b,r),(r,b),(r,r)} kan je, zoals vroeger, grafisch voorstellen.

Ook ***gebeurtenissen*** kan je verduidelijken op een figuur.



Als A de gebeurtenis is: “de ballen hebben een verschillende kleur”, dan is A = {(b,r),(r,b)}.

Als B de gebeurtenis is: “de tweede bal is rood”, dan is B = {(b,r),(r,r)}.

Kans van een gebeurtenis

Zoals vroeger is de kans van de gebeurtenis A gelijk aan de som van de kansen van alle uitkomsten die in A zitten. In A zitten (b,r) en (r,b) en daarom is P(A) = 8/30 + 8/30 = 16/30.
Op eenzelfde manier zie je dat B opgebouwd is uit (b,r) en (r,r) zodat P(B) = 8/30 + 2/30 = 10/30.

Complement

De kans dat de ballen dezelfde kleur hebben kan je rechtstreeks berekenen, maar je kan ook de rekenregel voor het complement gebruiken:
$P(A^c) = 1–P(A) = 1–16/30 = 14/30$.

Vereniging

De kans van “A of B” is de kans op ofwel een verschillende kleur, ofwel de tweede rood, ofwel zowel een verschillende kleur als de tweede rood. Hieraan voldoen de uitkomsten [bekijk de figuur] (r,b), (r,r) en (b,r) zodat P(A of B) = 8/30 + 2/30 + 8/30 = 18/30.

Doorsnede

De kans van “A en B” is de kans dat de ballen een verschillende kleur hebben en tegelijkertijd moet de tweede rood zijn. Hieraan voldoet alleen [bekijk de figuur] (b,r) zodat P(A en B) = 8/30.

Voorwaardelijke kans

De kans dat de tweede bal rood is, gegeven dat de ballen een verschillende kleur hebben, is een voorwaardelijke kans waarvoor je de rekenregel gebruikt: $P(B\,|\,A)=\dfrac{P(A\text{ en }B)}{P(A)}=\dfrac{8/30}{16/30}=\dfrac{8}{16}=\dfrac{1}{2}$.

Onafhankelijkheid

Bemerk dat P(A en B) = 8/30 terwijl $P(A)\cdot P(B)=(16/30)\cdot(10/30)=8/45$.
Uit $P(A\text{ en }B)\neq P(A)\cdot P(B)$ volgt dat A (verschillende kleur) en B (tweede bal is rood) geen onafhankelijke gebeurtenissen zijn. Dit volgt uit de criteria voor onafhankelijkheid.

Je kan ook een ander criterium gebruiken en opmerken dat P(B|A) = 1/2 terwijl P(B) = 1/3 wat opnieuw aantoont dat A en B niet onafhankelijk zijn.

Opdracht 18

Deze opdracht gaat over het samengestelde experiment waar je met terugleggen twee keer na elkaar een kaartje trekt uit een doos met 3 witte (w) kaartjes en 1 zwart (z).

1. Schrijf in de juiste notatie:
   - de uitkomstenverzameling U
   - de gebeurtenis A: “beide kaartjes hebben dezelfde kleur”
   - de gebeurtenis B: “het eerste kaartje is wit”

   en stel U, A en B ook voor op een figuur.

2. Geef het volledige kansmodel voor dit experiment. Gebruik daarvoor de tabel die vroeger werd opgesteld.

3. Bereken P(A) en P(B) en zeg hoe je dat doet.

4. Zeg in woorden wat “A of B” betekent en bereken P(A of B) [bekijk de figuur].

5. Zeg in woorden wat “A en B” betekent en bereken P(A en B) [bekijk de figuur].

6. Zeg in woorden wat P(B|A) betekent. Bereken ook deze kans en zeg welke kansregel je hiervoor gebruikt.

7. Controleer of de gebeurtenissen A en B onafhankelijk zijn van elkaar. Gebruik hiervoor twee verschillende criteria.

7. De veralgemeende productregel

7.1. De Lotto

Wat is de kans dat de 6 nummers die jij op je formulier hebt aangekruist, getrokken worden?

Je kan stap na stap beschrijven wat er moet gebeuren.

Uit de 45 balletjes wordt lukraak een eerste balletje getrokken. Als er op dat balletje één van de zes nummers staat die jij hebt aangekruist, is dat al een goed begin. Succes hebben bij deze eerste stap noteer je als 1_s. De kans op succes is $P(1_s) = 6/45$ want 6 van de 45 nummers zijn goed.

Als het eerste balletje goed was, dan wil je nu ook dat het tweede goed is. Bij de Lotto trekt men zonder terugleggen en dus zijn er in totaal nu nog 44 balletjes. Daartussen zitten er 5 met een nummer dat jij hebt aangekruist en dat nog niet getrokken is. Je moet nu succes hebben bij het tweede balletje, gegeven dat je succes had bij het eerste. De kans dat zoiets gebeurt, is dus een voorwaardelijke kans en $P(2_s|1_s) = 5/44$ omdat er tussen de resterende 44 balletjes 5 goede zitten.

Bij de derde trekking zijn er nog 43 balletjes en daarvan zijn er nog 4 goede. Je moet nu succes hebben bij de derde trekking, gegeven dat je de eerste keer succes had en de tweede keer succes. De voorwaardelijke kans is $P(3_s|1_s \text{ en } 2_s) = 4/43$.

Je kan nu op dezelfde manier verder redeneren. Je vindt dan:

$P(4_s|1_s \text{ en } 2_s \text{ en } 3_s) = 3/42$
$P(5_s|1_s \text{ en } 2_s \text{ en } 3_s \text{ en } 4_s) = 2/41$
$P(6_s|1_s \text{ en } 2_s \text{ en } 3_s \text{ en } 4_s \text{ en } 5_s) = 1/40$.

Alle zes getrokken nummers zijn goed als het eerste goed is ***en*** het tweede goed ***en*** … ***en*** het zesde goed. Het gaat hier over zes gebeurtenissen die tegelijkertijd moeten waar zijn.

$P(1_s \text{ en } 2_s \text{ en } 3_s \text{ en } 4_s \text{ en } 5_s \text{ en } 6_s)$

$$= P(1_s) \cdot P(2_s|1_s) \cdot P(3_s|1_s en 2_s) \cdot P(4_s|1_s en 2_s en 3_s) \cdot P(5_s|1_s en 2_s en 3_s en 4_s) \cdot P(6_s|1_s en 2_s en 3_s en 4_s en 5_s)$$

$$= \frac{6}{45} \cdot \frac{5}{44} \cdot \frac{4}{43} \cdot \frac{3}{42} \cdot \frac{2}{41} \cdot \frac{1}{40} = \frac{1}{8\,145\,060} \cong 0.000\,000\,1$$

Deze kans is ongeveer even groot als de kans dat je binnen de volgende 12 maanden doodgebliksemd wordt. Blijkbaar is met de Lotto spelen plezanter.

Als systeem zie je dat de regel voor de kans van "en, en, en …" gebeurtenissen niet moeilijk is. Je moet gewoon volgens een vast patroon werken. Je start met de kans van de eerste gebeurtenis. Die vermenigvuldig je met de voorwaardelijke kans van de tweede gebeurtenis gegeven wat ervoor staat (de eerste). Dat ken je al uit de productregel voor de doorsnede van twee gebeurtenissen want daar heb je geleerd dat $P(A \text{ en } B) = P(A) \cdot P(B|A)$.

Dan ga je verder en je vermenigvuldigt met de voorwaardelijke kans van de derde gebeurtenis gegeven wat ervoor staat (de eerste en de tweede).
Dat wordt: P(A en B en C) = $P(A) \cdot P(B|A) \cdot P(C|A \text{ en } B)$. Enzovoort…

Als je te maken hebt met gebeurtenissen die allemaal onafhankelijk van elkaar zijn, dan worden alle voorwaardelijke kansen terug gewone kansen. Je krijgt dan dat de kans van een doorsnede gelijk is aan het product van de kansen.

Productregel voor de doorsnede van n gebeurtenissen $A_1, A_2, \ldots A_n$

- **Algemene productregel bij n gebeurtenissen**

$P(A_1 \text{ en } A_2 \text{ en } A_3 \text{ en } \ldots \text{ en } A_n)$

$$= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \text{ en } A_2) \cdot \ldots \cdot P(A_n|A_1 \text{ en } A_2 \text{ en } A_3 \text{ en } \ldots \text{en } A_{n-1})$$

- **Productregel voor n onafhankelijke gebeurtenissen**

$P(A_1 \text{ en } A_2 \text{ en } A_3 \text{ en } \ldots \text{ en } A_n)$

$$= P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n)$$

7.2. Het verjaardagenprobleem

Selecteer op de speelplaats een groepje van 60 leerlingen die daar toevallig rondlopen en waarvan je niet weet op welke dag zij verjaren.

Zeg dan: “In een jaar zijn er 365 dagen en jullie zijn hier maar met een groepje van 60. En toch durf ik wedden dat er hier minstens twee leerlingen op dezelfde dag verjaren”.

Voor hoeveel durf je dan wedden? Of beter: hoe groot is de kans dat je die weddenschap wint?

Opdracht 19

Deze opdracht gaat over het verjaardagenprobleem waarbij je op zoek gaat naar de kans dat er bij een lukraak gekozen groepje leerlingen minstens twee op dezelfde dag verjaren.

Je beweert dat er minstens 2 leerlingen op dezelfde dag verjaren. Dat kunnen er dus 2 of 3 of 4 of 5 of … zijn. Om al die verschillende mogelijkheden niet te moeten uittellen kan je beter overstappen op het complement. Zoek eerst de kans dat iedereen een verschillende verjaardag heeft. Het complement betekent dan dat er minstens twee op dezelfde dag verjaren.

Vooraleer met 60 leerlingen te werken begin je eens met een groepje van 4. Zet die 4 leerlingen op een rij en eis nu dat elke leerling een verjaardag heeft die verschilt van alle leerlingen die vóór hem staan. Je zou met een aangepaste notatie kunnen werken zoals: 3_v betekent dat de derde leerling in de rij een verjaardag heeft die verschilt van die van de eerste en de tweede.

1. De gebeurtenis “2_v en 3_v en 4_v” betekent dat er tegelijkertijd aan drie gebeurtenissen voldaan is. Zeg in woorden wat die gebeurtenissen betekenen:
   - 2_v betekent dat …
   - 3_v betekent dat …
   - 4_v betekent dat …

2. Zeg in woorden:
   - voor de verjaardagen van die 4 leerlingen betekent “2_v en 3_v en 4_v” dat …
   - P(2_v en 3_v en 4_v) is de kans dat …
   - 1– P(2_v en 3_v en 4_v) is de kans dat …

3. Welke regel gebruik je om P(2_v en 3_v en 4_v) te berekenen? Schrijf die regel op en gebruik hem om P(2_v en 3_v en 4_v) te vinden. Toon duidelijk alle tussenstappen en motiveer je berekeningen.

4. Hoe groot is de kans dat er bij 4 lukraak gekozen leerlingen minstens 2 op dezelfde dag verjaren?

5. Gebruik de voorgaande manier van werken om snel te berekenen hoe groot de kans is dat er bij 6 lukraak gekozen leerlingen minstens 2 op dezelfde dag verjaren.

Wat de kansen zijn om samenvallende verjaardagen te vinden in lukraak gekozen groepjes die groter en groter worden zie je in de onderstaande tabel.

Groep	Kans	Groep	Kans	Groep	Kans	Groep	Kans	Groep	Kans	Groep	Kans
1	0	11	0.14	21	0.44	31	0.73	41	0.90	51	0.974
2	0.003	12	0.17	22	0.48	32	0.75	42	0.91	52	0.978
3	0.008	13	0.19	23	0.51	33	0.77	43	0.92	53	0.981
4	0.016	14	0.22	24	0.54	34	0.80	44	0.93	54	0.984
5	0.027	15	0.25	25	0.57	35	0.81	45	0.94	55	0.986
6	0.040	16	0.28	26	0.60	36	0.83	46	0.95	56	0.988
7	0.056	17	0.32	27	0.63	37	0.85	47	0.95	57	0.990
8	0.074	18	0.35	28	0.65	38	0.86	48	0.96	58	0.992
9	0.095	19	0.38	29	0.68	39	0.88	49	0.97	59	0.993
10	0.117	20	0.41	30	0.71	40	0.89	50	0.97	60	0.994

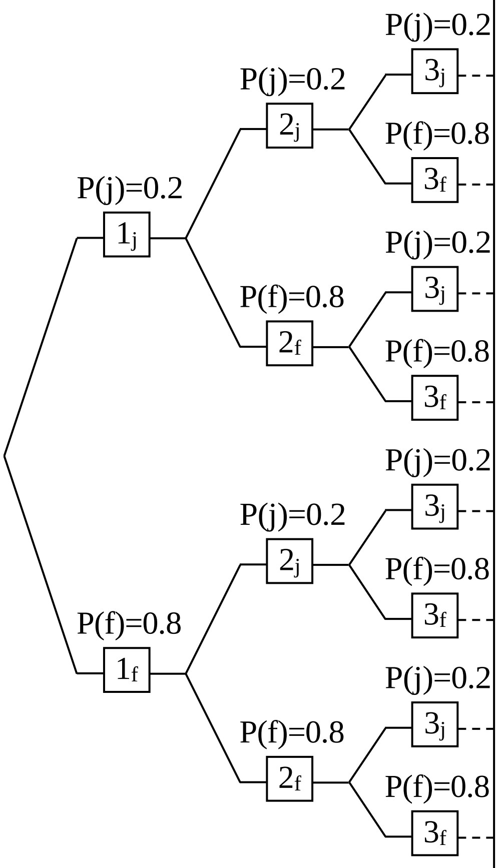
Neem een willekeurig groepje van 12 personen. De kans dat er minstens twee personen op dezelfde dag verjaren is 0.17, wat niet moeilijker is dan een zes gooien met een eerlijke dobbelsteen. Bij 23 personen heb je al kans 0.51. Dat is ongeveer zoals kruis gooien met een muntstuk. En bij een groep van 60 personen heb je 99.4 % kans om je weddenschap te winnen. Spelen dus!

7.3. Raden op het examen

De veralgemeende productregel gebruik je ook wanneer je met een kansboom het kansmodel opstelt van een samengesteld experiment waarbij je meer dan 2 stappen zet.

> Een onderdeel van een examen bestaat uit 3 meerkeuzevragen met elk 5 mogelijke antwoorden waarvan er maar 1 juist is.
> Een leerling heeft helemaal niet gestudeerd en besluit om bij elke vraag lukraak te raden. Het antwoord op een volgende vraag hangt dan ook niet af van het antwoord op een vorige vraag. Elke vraag is als een nieuw begin waarbij lukraak geraden wordt.
>
> Wat is de kans dat de leerling minstens 2 vragen van de 3 juist heeft?

Voor dit vraagstuk zie je dat de leerling 3 keer na elkaar iets doet, namelijk een antwoord geven op een meerkeuzevraag. Elke keer is dat antwoord ofwel juist ofwel fout. Bij elke nieuwe vraag is het antwoord onafhankelijk van het vorige want er wordt telkens opnieuw geraden. Dit betekent dat de opeenvolgende stappen in dit samengesteld experiment onafhankelijk zijn. Zo is bijvoorbeeld de kans om bij de tweede vraag fout te antwoorden, gegeven dat het antwoord op de eerste vraag fout was [notatie $P(2_f|1_f)$], gelijk aan de kans om op de tweede vraag fout te antwoorden [notatie $P(2_f)$]. Bovendien is bij elke vraag de kans op een fout antwoord dezelfde namelijk 4/5. Je kan dus in de kansboom de kans op een fout antwoord overal noteren als P(f) en de kans op een juist antwoord als P(j).



Uitkomst	Kans
"1_j en 2_j en 3_j" = (j,j,j)	P((j,j,j)) = (0.2)·(0.2)·(0.2) = 0.008
"1_j en 2_j en 3_f" = (j,j,f)	P((j,j,f)) = (0.2)·(0.2)·(0.8) = 0.032
"1_j en 2_f en 3_j" = (j,f,j)	P((j,f,j)) = (0.2)·(0.8)·(0.2) = 0.032
"1_j en 2_f en 3_f" = (j,f,f)	P((j,f,f)) = (0.2)·(0.8)·(0.8) = 0.128
"1_f en 2_j en 3_j" = (f,j,j)	P((f,j,j)) = (0.8)·(0.2)·(0.2) = 0.032
"1_f en 2_j en 3_f" = (f,j,f)	P((f,j,f)) = (0.8)·(0.2)·(0.8) = 0.128
"1_f en 2_f en 3_j" = (f,f,j)	P((f,f,j)) = (0.8)·(0.8)·(0.2) = 0.128
"1_f en 2_f en 3_f" = (f,f,f)	P((f,f,f)) = (0.8)·(0.8)·(0.8) = 0.512

De kansboom voor dit experiment ziet eruit zoals bij "trekken met terugleggen".

Bemerk dat je hier de veralgemeende productregel voor ***onafhankelijke*** gebeurtenissen hebt gebruikt bij het berekenen van de kansen: P(A en B en C) = P(A)·P(B)·P(C).

De kansboom helpt je om geen enkele uitkomst te vergeten en om hun kansen te bepalen. Zo vind je het volledige kansmodel:

uitkomst	(j,j,j)	(j,j,f)	(j,f,j)	(j,f,f)	(f,j,j)	(f,j,f)	(f,f,j)	(f,f,f)
kans	0.008	0.032	0.032	0.128	0.032	0.128	0.128	0.512

Zoals altijd is ook hier de som van alle kansen gelijk aan 1.

Opdracht 20

Deze opdracht gaat over "raden op het examen".

1. Schrijf in de juiste notatie:
   - de uitkomstenverzameling U
   - de gebeurtenis A: "minstens 2 vragen juist"
   - de gebeurtenis B: "het antwoord op de eerste vraag is juist"

   en stel U, A en B ook voor op een figuur.

2. Een leerling die zomaar raadt op dat examen heeft weinig kans om van de drie vragen er ten minste twee juist te hebben. Hoe groot is die kans (geef ook aan hoe je die kans vindt)?

3. Wat is de kans dat het antwoord op de eerste vraag juist is (motiveer)?

4. Wat is de kans dat het antwoord op de eerste vraag juist is en dat er minstens twee vragen juist beantwoord zijn (motiveer)?

5. Wat is de kans dat er minstens twee vragen juist beantwoord zijn, gegeven dat het antwoord op de eerste vraag juist is (motiveer)?

6. Wat is de kans dat het antwoord op de eerste vraag juist is gegeven dat er minstens twee vragen juist beantwoord zijn (motiveer)?

Opdracht 21

In een doos zitten 4 kaartjes die volledig aan elkaar gelijk zijn behalve hun kleur. Er zijn er 3 witte en 1 zwart. Het samengestelde experiment bestaat uit 3 stappen: trek lukraak een eerste kaartje en noteer de kleur. Leg het kaartje terug in de doos. Trek opnieuw lukraak een kaartje, noteer de kleur en leg het terug. Trek ten slotte een derde kaartje en noteer de kleur.

1. Gebruik een kansboom om het volledige kansmodel op te stellen voor dit samengestelde experiment. Je mag daarbij de kansboom die vroeger werd opgesteld voor 2 trekkingen uit die doos als start gebruiken en hem verder aanvullen. Vat daarna het kansmodel samen in een tabel.

2. Schrijf in de juiste notatie:
   - de uitkomstenverzameling U
   - de gebeurtenis A: "het derde kaartje is wit"
   - de gebeurtenis B: "de eerste twee kaartjes zijn zwart"

   en stel U, A en B ook voor op een figuur.

3. Bereken P(A) en P(B) en zeg hoe je dat doet.

4. Zeg in woorden wat "A of B" betekent en bereken P(A of B).

5. Zeg in woorden wat "A en B" betekent en bereken P(A en B).

6. Gebruik de bovenstaande berekeningen om de somregel te illustreren voor "A of B".

7. Zeg in woorden wat P(B|A) betekent. Bereken ook deze kans en zeg welke kansregel je hiervoor gebruikt.

8. Controleer of de gebeurtenissen A en B onafhankelijk zijn van elkaar. Gebruik hiervoor twee verschillende criteria.

8. De wet van de totale kans

8.1. Roken en geslacht

Een bevolking bestaat uit 40 % mannen (M) en 60 % vrouwen (V). Bij de mannen zijn er 35 % rokers (R) en bij de vrouwen zijn er 30 % rokers. Men trekt lukraak een persoon uit deze bevolking. Wat is de kans dat deze persoon een roker is?

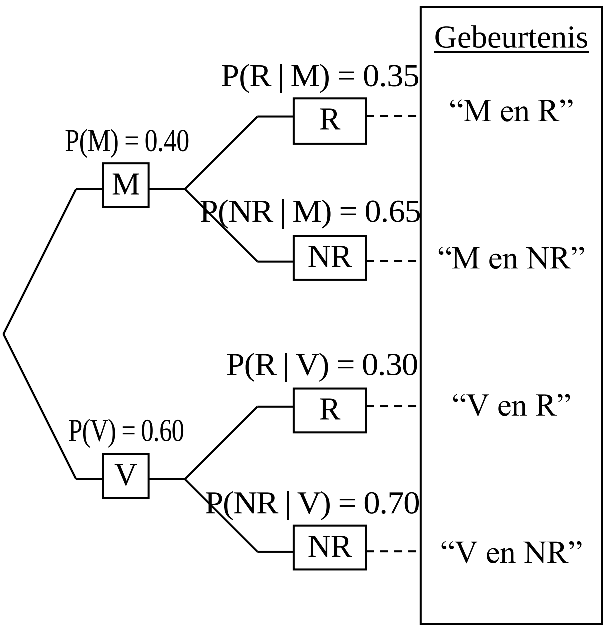
Dit vraagstuk heeft een bepaalde structuur die je nodig hebt om de wet van de totale kans te kunnen toepassen:

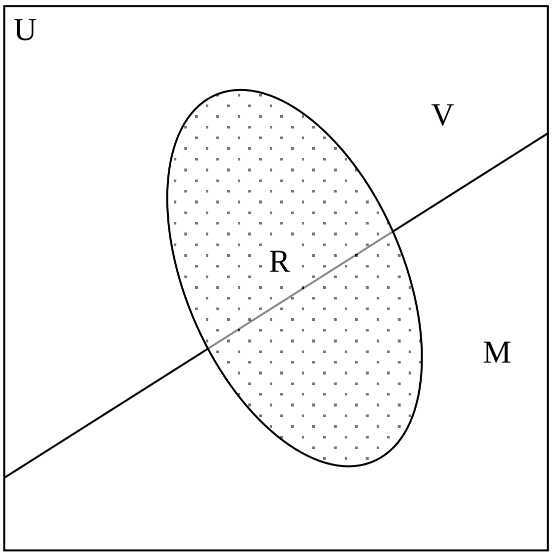
1. De uitkomstenverzameling is verdeeld in een aantal deelverzamelingen (gebeurtenissen) die niet overlappen (disjunct zijn) en waarvan de kans gekend is. [Hier is de bevolking opgedeeld in mannen en vrouwen met P(M) = 0.40 en P(V) = 0.60].

2. Van een bepaalde eigenschap ken je de voorwaardelijke kansen, gegeven de gebeurtenissen uit punt 1. [Hier is de eigenschap “roker zijn” met P(R|M) = 0.35 en P(R|V) = 0.30].

3. De vraag gaat over de (gewone, niet voorwaardelijke) kans van die eigenschap [de kans op een roker].

Voor dit vraagstuk heb je vroeger een kansboom opgesteld voor gebeurtenissen die doorsneden zijn: “M en R”, “M en NR”, “V en R”, “V en NR”.





Nu gaat de vraag over rokers. Als je alle mannelijke rokers samen neemt met alle vrouwelijke rokers dan heb je alle rokers: R = “M en R” ***of*** “V en R” want je hebt de vereniging nodig [bekijk de figuur].

Bemerk dat M en V disjunct zijn. “M en R” is een stukje van M terwijl “V en R” een stukje van V is. Dat betekent dat ook “M en R” en “V en R” disjunct zijn. Je kan dan de somregel voor de vereniging van disjuncte gebeurtenissen gebruiken: P(R) = P(“M en R” ***of*** “V en R”) = P(M en R) + P(V en R).

De productregel zegt dat P(M en R) = $P(R|M)\cdot P(M)$ en P(V en R) = $P(R|V)\cdot P(V)$ zodat de kans op een roker gelijk is aan: $P(R) = P(R|M)\cdot P(M) + P(R|V)\cdot P(V)$.

Voor de ***totale kans*** van R loop je dus over alle disjuncte deelverzamelingen waarin je de uitkomstenverzameling U hebt opgedeeld. Daarbij maak je ***het totaal*** (= de som) van:
(de voorwaardelijke kans gegeven die deelverzameling) maal (de kans van die deelverzameling).

Wet van de totale kans

$$P(R) = P(R|M)\cdot P(M) + P(R|V)\cdot P(V)$$

De kans op een roker is gelijk aan:
(de kans op een roker gegeven dat het een man is) maal (de kans op een man)
+ (de kans op een roker gegeven dat het een vrouw is) maal (de kans op een vrouw)

Als je uit de bevolking van dit vraagstuk lukraak een persoon trekt dan is de kans dat het een roker is gelijk aan $P(R) = P(R|M)\cdot P(M) + P(R|V)\cdot P(V) = (0.35)\cdot(0.40) + (0.30)\cdot(0.60) = 0.32 = 32\ \%$.

8.2. Huwelijksgeluk

Aan een grote groep gehuwde mensen vroeg men hoe gelukkig zij waren binnen hun huwelijk.

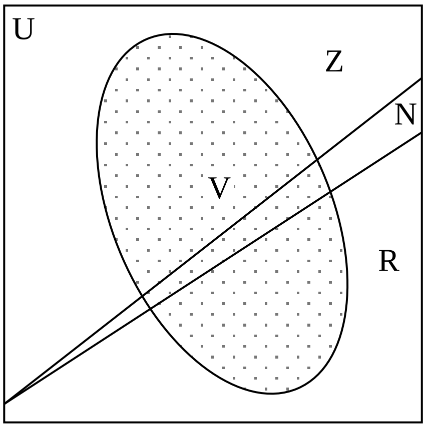
De antwoorden varieerden van *zeer* gelukkig (Z = 60 %) over *redelijk* gelukkig (R = 35 %) tot *niet* gelukkig (N = 5 %).

Bij de groep die zeer gelukkig was, kwamen de meeste antwoorden van mannen (58 %), bij de groep die redelijk gelukkig was, waren de vrouwen in de meerderheid (56 %) en bij de groep die niet gelukkig was, waren er evenveel mannen als vrouwen.

Als je uit al deze mensen lukraak een persoon kiest, wat is dan de kans dat het een vrouw is?

Begin met het vraagstuk te ontleden: wat is er allemaal gegeven en wat is er gevraagd? Op die manier ontdek je welke kansregel van toepassing is.

De uitkomstenverzameling kan op basis van het antwoord verdeeld worden in 3 disjuncte deelverzamelingen (gebeurtenissen): Z, R en N [bekijk de figuur]. Van elk van deze gebeurtenissen ken je ook de kans:
P(Z) = 0.60, P(R) = 0.35 en P(N) = 0.05.



Verder heb je een andere eigenschap: het geslacht. Wat de vrouwen betreft ken je de voorwaardelijke kans om een vrouw te ontmoeten binnen de verschillende antwoordcategorieën:
P(V|Z) = 0.42 [want er zijn daar 58 % mannen], P(V|R) = 0.56 en P(V|N) = 0.50.

Je hebt nu alle elementen die nodig zijn om de wet van de totale kans toe te passen:

$$\begin{aligned} P(V) &= P(V|Z)\cdot P(Z) + P(V|R)\cdot P(R) + P(V|N)\cdot P(N) \\ &= (0.42)\cdot(0.60) + (0.56)\cdot(0.35) + (0.50)\cdot(0.05) = 0.473 \end{aligned}$$

Opdracht 22

Het geboortegewicht (in gram) kan je als volgt in drie klassen opdelen:

L = *laag geboortegewicht* $(gebgw < 2500)$

N = *normaal geboortegewicht* $(2500 \leq gebgw < 4500)$

H = *hoog geboortegewicht* $(gebgw \geq 4500)$

Van de kinderen die in Vlaanderen geboren worden hebben 7 % een laag, 92 % een normaal en 1 % een hoog geboortegewicht. Bij de baby's met een laag geboortegewicht zijn er 46 % jongens, bij hen met een normaal geboortegewicht zijn er dat 51 %, en bij hen met een hoog geboortegewicht zijn het er 71 %.

- Gebruik gepaste symbolen om alle kansen die in de opgave gegeven zijn in de juiste notatie voor te stellen.

- Als je lukraak uit de databank een baby kiest, wat is dan de kans dat het een jongen is? Welke kansregel gebruik je hier? Motiveer waarom je deze kansregel mag gebruiken.

9. De regel van Bayes

9.1. Roken en geslacht

> Een bevolking bestaat uit 40 % mannen (M) en 60 % vrouwen (V).
> Bij de mannen zijn er 35 % rokers (R) en bij de vrouwen zijn er 30 % rokers.
>
> Men trekt lukraak een persoon uit deze bevolking en zegt dat het een roker is. Wat is de kans dat die persoon een man is?

Begin terug met het vraagstuk te ontleden.

De uitkomstenverzameling kan op basis van het geslacht verdeeld worden in 2 disjuncte deelverzamelingen (gebeurtenissen): M en V. Van elk van deze gebeurtenissen ken je ook de kans: P(M) = 0.40 en P(V) = 0.60.

Van de andere eigenschap (het rookgedrag) ken je de voorwaardelijke kansen per geslachtsgroep: P(R|M) = 0.35 en P(R|V) = 0.30.

Bij de gegevens heb je voorwaardelijke kansen *van het rookgedrag gegeven het geslacht*. De vraag gaat over een voorwaardelijke kans *van het geslacht gegeven het rookgedrag*. De regel van Bayes helpt je om "omgekeerde" voorwaardelijke kansen te berekenen.

De regel voor een voorwaardelijke kans zegt dat $P(M \mid R) = \dfrac{P(M \text{ en } R)}{P(R)}$. Verder zegt de productregel dat P(M en R) = P(M) $\cdot$ P(R|M). Samen kom je tot: $P(M \mid R) = \dfrac{P(R \mid M) \cdot P(M)}{P(R)}$.

P(M) en P(R|M) is gegeven. P(R) is niet gegeven maar het vraagstuk bevat genoeg informatie om P(R) te berekenen met de wet van de totale kans: P(R) = P(R|M) $\cdot$ P(M) + P(R|V) $\cdot$ P(V).

Regel van Bayes

$$P(M \mid R) = \frac{P(R \mid M) \cdot P(M)}{P(R \mid M) \cdot P(M) + P(R \mid V) \cdot P(V)}$$

Voor dit vraagstuk is

$$P(M \mid R) = \frac{P(R \mid M) \cdot P(M)}{P(R \mid M) \cdot P(M) + P(R \mid V) \cdot P(V)} = \frac{(0.35) \cdot (0.40)}{(0.35) \cdot (0.40) + (0.30) \cdot (0.60)} = 0.4375 \cong 0.44$$

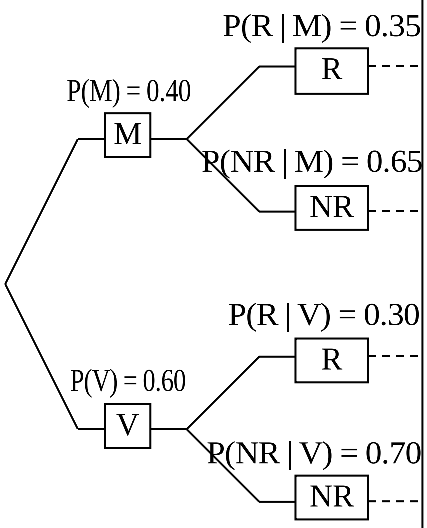
9.2. Kansbomen en kruistabellen

"Omgekeerde" voorwaardelijke kansen kan je rechtstreeks met de regel van Bayes berekenen. Je kan ze ook vinden met kansbomen en kruistabellen. Dat werkt als volgt:

- je gebruikt een kansboom om vanuit de gegeven (voorwaardelijke) kansen te komen tot kansen van doorsneden, zoals $P(M)\cdot P(R|M) = P(M \text{ en } R)$
- de kansen van doorsneden schrijf je in een kruistabel waar je zowel op rijen als op kolommen kan conditioneren. Met de goede keuze kom je zo aan de gevraagde voorwaardelijke kans.

Als voorbeeld werk je met de gegevens van bovenstaand vraagstuk. Alles wat hieronder staat heb je al ontmoet in de tekst voor de tweede graad.

De kansboom met bijhorende gebeurtenissen en kansen ziet er als volgt uit:



Gebeurtenis	Kans
"M en R"	$P(M \text{ en } R) = (0.40)\cdot(0.35) = 0.14$
"M en NR"	$P(M \text{ en } NR) = (0.40)\cdot(0.65) = 0.26$
"V en R"	$P(V \text{ en } R) = (0.60)\cdot(0.30) = 0.18$
"V en NR"	$P(V \text{ en } NR) = (0.60)\cdot(0.70) = 0.42$

Nu kan je een kruistabel opstellen en daarbij ook de randtotalen berekenen. Je hebt hierbij alleen maar de resultaten uit de kansboom nodig.

	R	NR	totaal
M	0.14	0.26	0.40
V	0.18	0.42	0.60
totaal	0.32	0.68	1

Tenslotte kan je voorwaardelijk werken per kolom, bijvoorbeeld gegeven dat je met rokers te maken hebt. Je moet dan de cellen van die kolom delen door het kolomtotaal.

	gegeven R
M	0.14/0.32 = 0.44
V	0.18/0.32 = 0.56
totaal	0.32/0.32 = 1

Er is 44 % kans dat het een man is als er gegeven is dat die persoon rookt: P(M|R) = 0.44.

9.3. Diagnostische testen

Als je je laat testen op AIDS en de test geeft een negatief resultaat, dan moet je gelukkig zijn. Inderdaad, in de medische wereld zegt men dat het resultaat negatief is wanneer de test geen sporen van de ziekte heeft kunnen ontdekken. Als de arts zegt dat het resultaat positief is, dan moet je je zorgen maken.

Diagnostische testen zijn niet perfect. Soms geven zij een foutief resultaat. Als een test een positief resultaat geeft (T+) terwijl je de ziekte niet hebt (Z–), dan spreekt men van een ***vals positief***. Als een test niets gevonden heeft (T–) terwijl je echt ziek bent (Z+), dan heb je met een ***vals negatief*** te maken.

De ***sensitiviteit*** van een test is ***de kans*** dat de test een positief resultaat geeft als je de ziekte ook echt hebt: P(T+|Z+). Dat betekent dat die test gevoelig is (sensitief is) voor de ziekte en positief reageert als je de ziekte hebt.

De ***specificiteit*** van een test is ***de kans*** dat de test een negatief resultaat geeft als je de ziekte niet hebt: P(T–|Z–). Dat betekent dat de test specifiek afgestemd is op die bepaalde ziekte (bijvoorbeeld AIDS). Als je dus geen AIDS hebt maar wel griep, dan mag de test niet positief uitslaan (je zou nogal schrikken).

De ***prevalentie*** van een ziekte is ***de kans*** dat een lukraak getrokken persoon de ziekte heeft. De prevalentie is een maat voor de zeldzaamheid (of voor het veelvuldig voorkomen) van een ziekte in de bevolking.

In België is 1 op 1000 mensen seropositief.

Een test om te ontdekken of iemand seropositief is (en AIDS zal krijgen) heeft een sensitiviteit van 99 % en een specificiteit van 98 %.

Jij wordt lukraak gekozen om de test te ondergaan en je kan je ogen niet geloven: de test is positief! Wat is de kans dat je seropositief bent?

De volgende grootheden zijn gegeven:

- prevalentie: P(Z+) = 0.001 zodat ook (complementregel) P(Z–) = 0.999
- sensitiviteit: P(T+|Z+) = 0.99 zodat ook (complementregel) P(T–|Z+) = 0.01
- specificiteit: P(T–|Z–) = 0.98 zodat ook (complementregel) P(T+|Z–) = 0.02.

De vraag gaat over P(Z+|T+) wat een voorwaardelijke kans van de ziektetoestand is gegeven het testresultaat. De gegevens gaan over voorwaardelijke kansen van het testresultaat gegeven de ziektetoestand. Je bent hier duidelijk in een context om de regel van Bayes toe te passen.

P(Z+|T+) heet ***de voorspellende waarde van een positieve test***.

Volgens de regel van Bayes geldt:

$$\begin{aligned}P(Z+\,|\,T+) &= \frac{P(T+\,|\,Z+)\cdot P(Z+)}{P(T+\,|\,Z+)\cdot P(Z+)+P(T+\,|\,Z-)\cdot P(Z-)}\\ &= \frac{(0.99)\cdot(0.001)}{(0.99)\cdot(0.001)+(0.02)\cdot(0.999)} = 0.05\end{aligned}$$

Dit resultaat is onverwacht. Je werkt met een bijna perfecte test (met een sensitiviteit van 99 % en een specificiteit van 98 %). Toch heb je maar 5 % kans dat je de ziekte echt hebt als die test positief reageert. Hoe komt dat?

Als je nauwkeurig naar de formule kijkt dan zie je daarin 2 dingen:

- grootheden die eigenschappen van de test karakteriseren (in farmaceutische firma’s werkt men hard om zeer goede testen te ontwikkelen)
- grootheden die te maken hebben met de frequentie van de ziekte in de bevolking (en daarover heeft men bij de ontwikkeling van diagnostische testen geen controle).

Moraal van het verhaal: grote programma’s opzetten waar mensen lukraak gescreend worden voor een zeldzame ziekte is weinig zinvol.

Als je met een even goede test een ziekte screent waarbij 1 op 5 mensen de ziekte heeft, dan heeft een lukraak gekozen persoon 93 % kans dat hij die ziekte heeft als de test positief reageert want:

$$\begin{aligned}P(Z+\,|\,T+) &= \frac{P(T+\,|\,Z+)\cdot P(Z+)}{P(T+\,|\,Z+)\cdot P(Z+)+P(T+\,|\,Z-)\cdot P(Z-)}\\ &= \frac{(0.99)\cdot(0.20)}{(0.99)\cdot(0.20)+(0.02)\cdot(0.80)} = 0.93\end{aligned}$$

Opdracht 23

P(Z–|T–) heet ***de voorspellende waarde van een negatieve test****. Het is de kans dat je de ziekte niet hebt als de test niets ontdekt en dus een negatief resultaat geeft.*

- Zoek de voorspellende waarde van een negatieve test als de test een sensitiviteit van 98 % heeft en een specificiteit van 96 %. Onderstel dat de prevalentie 3 % is.

9.4. Drugscontrole op school

Vind je het goed dat, zoals het in sommige landen gebeurt, de politie naar school mag komen om lukraak leerlingen te testen of ze niet onder invloed van drugs zijn?

Als zo'n test positief reageert, ben je dan onder invloed? Dat hangt af van twee dingen:

- hoe goed is de test (sensitiviteit en specificiteit)?
- hoeveel leerlingen zitten onder invloed van drugs op school (prevalentie)?

Onderstel dat men een zeer goede test gebruikt met:

- 99.9 % sensitiviteit zodat de kans dat de test positief reageert als je echt onder invloed van drugs bent gelijk is aan P(T+|D+) = 0.999.
- 99.9 % specificiteit zodat de kans dat de test negatief reageert als je niet onder invloed van drugs bent gelijk is aan P(T–|D–) = 0.999.

Om de rol van de prevalentie te illustreren kan je hier eens met aantallen werken (je kan daarbij aan relatieve frequentie denken om tot kansuitspraken te komen).

Onderstel dat 1 op 1000 leerlingen onder invloed van drugs op school zit. Dan is P(D+) = 0.001 (prevalentie).

Denk nu aan een grote groep van 1 000 000 leerlingen. Als er maar 1 op 1000 onder invloed is, dan zijn er in die populatie 1000 leerlingen onder invloed en 999 000 niet.

Wat doet de test bij die 1000 leerlingen onder invloed (D+)? De sensitiviteit zegt dat de test bij 999 van die leerlingen positief (T+) reageert en bij 1 leerling negatief (T–).

Wat doet die test bij de 999 000 leerlingen die niet onder invloed zijn (D–)? Uit de specificiteit volgt dat de test bij $(0.999)\cdot(999\ 000)$ = 998 001 leerlingen negatief reageert en bij $(0.001)\cdot(999\ 000)$ = 999 leerlingen positief.

Je kan dit allemaal samenvatten in een kruistabel:

	T+	T–	totaal
D+	999	1	1 000
D–	999	998 001	999 000
totaal	1 998	998 002	1 000 000

Nu kan je voorwaardelijk werken per kolom, gegeven dat de test positief reageert. Je moet dan de cellen van die kolom delen door het kolomtotaal.

	gegeven T+
D+	999 / 1 998 = 0.50
D–	999 / 1 998 = 0.50
totaal	1 998 / 1 998 = 1

De kans dat een lukraak gekozen leerling echt onder invloed van drugs is als de (zeer nauwkeurige) test bij die leerling positief reageert, is 50 %. Je kan dus evengoed een muntstuk opgooien.

Opdracht 24

Werk met dezelfde gegevens zoals hierboven maar onderstel dat 1 op 10 leerlingen onder invloed van drugs op school zit.

- Als men nu lukraak een leerling kiest en de test reageert positief, wat is dan de kans dat die leerling echt onder invloed is?