



**STATISTIEK** VOOR HET SECUNDAIR ONDERWIJS

Populatiemodellen en normaal verdeelde populaties

1. De wereld van de kansmodellen

*Werktekst voor de leerling*

Prof. dr. Herman Callaert

Hans Bekaert  
Cecile Goethals  
Lies Provoost  
Marc Vancaudenberg

# De wereld van de kansmodellen

<b>1. De realiteit en het model</b> .....	<b>1</b>
<b>2. Kansmodellen</b> .....	<b>1</b>
<b>3. Kansmodellen voor experimenten met discrete uitkomsten</b> .....	<b>3</b>
3.1. Een kansmodel voorstellen door een vaas .....	3
3.2. Een kansmodel voorstellen door een staafdiagram .....	5
3.3. Een kansmodel voorstellen door een tabel.....	7
3.4. De kansverdeling .....	10
<b>4. Kansmodellen voor experimenten met continue uitkomsten</b> ....	<b>11</b>
4.1. Een eerste voorbeeld .....	11
4.2. Ongelijkheidstekens bij continue kansmodellen .....	13
4.3. De kansdichtheid.....	13
4.4. Een tweede voorbeeld.....	14
4.4.1. Ik wil de grootste zijn.....	14
4.4.2. Het maximum van 2 lukrake getallen uit $[0 ; 1]$ .....	15
4.5. Kansen berekenen met de GRM .....	16
<b>5. Populatiemodellen</b> .....	<b>18</b>
5.1. Populatiemodel vanuit een wiskundige definitie .....	19
5.2. Populatiemodel vanuit een abstract experiment .....	20
5.3. Populatiemodel vanuit concrete opmetingen .....	21
5.4. Afspraak woordgebruik.....	24
<b>6. De binomiale en de normale</b> .....	<b>25</b>
6.1. Een beroemd discreet populatiemodel: de binomiale .....	25
6.2. Een beroemd continu populatiemodel: de normale.....	26

## 1. De realiteit en het model

Als je een echte dobbelsteen in je handen krijgt, en men vraagt je of dat een “eerlijke” dobbelsteen is, dan kan je dat eigenlijk nooit met 100 % zekerheid weten. Maar je hebt wel een goed idee over wat een eerlijke dobbelsteen zou moeten zijn. En om dit uit te leggen stap je over op een ideaal model. Je zegt dan dat het een dobbelsteen moet zijn die met kans  $1/6$  op één valt, met kans  $1/6$  op twee, enz. Je weet tegelijkertijd dat er geen enkele echte dobbelsteen bestaat (of die nu gemaakt is uit hout, plastic, metaal, ...) die de perfectie van een wiskundig model kan halen.

Zelfs al is de werkelijkheid niet te vangen in ideale modellen, toch heb je dikwijls zo'n modellen nodig om de wereld beter te begrijpen.

## 2. Kansmodellen

Bij een dobbelsteen kan je naar de resultaten kijken die jij vindt als je vandaag 10 keer gooit. Dat is bijvoorbeeld 4 4 6 1 4 5 3 5 6 3. Je weet dat je bijna zeker iets anders zal vinden als je met diezelfde dobbelsteen morgen terug 10 keer gooit.

Over die dobbelsteen kan je ook iets helemaal anders zeggen, iets dat niet afhangt van vandaag of morgen. Je kan zeggen wat jij “globaal” verwacht te vinden, als je heel veel keren zou gooien. Bij 6 miljoen worpen bijvoorbeeld verwacht je “ongeveer” 1 miljoen keer één, “ongeveer” 1 miljoen keer twee, enz. Deze verwachting geeft een eigenschap van die dobbelsteen weer, en die verandert niet van vandaag op morgen.

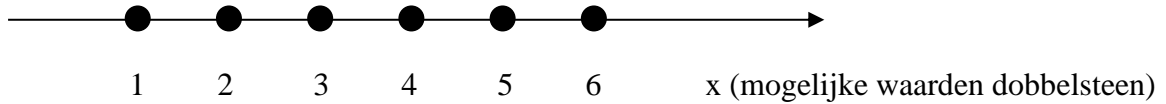
Je kan je verwachting van wat er “in the long run” zal gebeuren formuleren als een “ideaal model”, dat **vooraf** zegt wat er te verwachten is. Je spreekt dan over een eerlijke dobbelsteen, waarbij de getallen 1 2 3 4 5 6 allemaal met dezelfde kans tot jou komen, namelijk met kans  $1/6$ .

Je hebt nu voor het gedrag van die dobbelsteen een “kansmodel” opgesteld.

**Een kansmodel is een geïdealiseerd wiskundig model dat je gebruikt om experimenten te bestuderen waarbij het toeval een rol speelt.**

Experimenten waarbij het toeval een rol speelt kan je in twee soorten klasseren.

Sommige experimenten hebben uitkomsten die “**discreet**” zijn. Bij zo’n experimenten bestaat de verzameling van alle mogelijke uitkomsten uit getallen die je op de getallenas kan plaatsen met “ruimte” ertussen. Een dobbelsteen gooien is een voorbeeld van een experiment met discrete uitkomsten. De mogelijke uitkomsten zijn hierbij  $\{ 1, 2, 3, 4, 5, 6 \}$ .



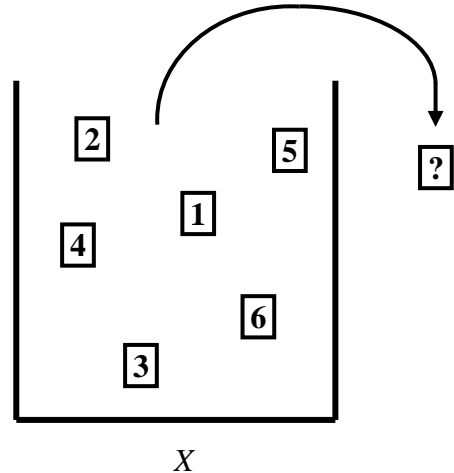
Andere experimenten hebben uitkomsten die als “**continu**” behandeld worden. Op de getallenas wordt de verzameling van alle mogelijke uitkomsten in zo’n geval voorgesteld als een aaneengesloten gebied. Voor de geboortegewichten van Vlaamse kinderen (in kg) kan je bijvoorbeeld het lijnstuk nemen van nul tot zeven. Gewicht, lengte, tijd... zijn allemaal voorbeelden van grootheden die als continu behandeld worden.



### 3. Kansmodellen voor experimenten met discrete uitkomsten

#### 3.1. Een kansmodel voorstellen door een vaas

Het gedrag van een eerlijke dobbelsteen kan je voorstellen met een **vaasmodel**. Dat is een vaas met daarin kaartjes waarop nummers staan. Je onderstelt dat al die kaartjes eenzelfde vorm en gewicht hebben en dat je die perfect door elkaar kan mengen. Probeer je nu voor te stellen wat er zou gebeuren als jij uit die vaas lukraak een kaartje zou trekken.



Figuur 1

Het is niet eenvoudig om goed te begrijpen wat zo'n hoofdletter juist betekent. Met de hoofdletter  $X$  bedoel je de volledige figuur. Op die figuur zie je alle kaartjes die je uit die vaas kan trekken. Er is ook een pijl getekend die aangeeft dat je lukraak uit die vaas moet trekken. Dat betekent dat elk kaartje dezelfde kans heeft om getrokken te worden.

Kan je vooraf zeggen welk getal je zal vinden als je een kaartje uit die vaas zal trekken?

Schrik niet, maar het antwoord is zowel neen als ja, en het is belangrijk dat je het verschil tussen die twee antwoorden goed begrijpt.

Als je denkt aan dat ene kaartje, dat je nu blindelings trekt, dan weet je natuurlijk niet welk getal daar zal op staan. Je bent immers niet helderziende, of toch?

Maar kan je dan niets zeggen?

Toch wel, en je kan zelfs heel veel zeggen. Als je naar het vaasmodel kijkt dan weet je dat er op dat kaartje nooit een 7 zal staan. Je weet wel dat er bijvoorbeeld een 4 kan op staan. Maar je weet nog meer. Je weet zelfs wat de kans is dat er een 4 op zal staan. Die kans is  $1/6$ .

Eigenlijk weet je dus alles, maar je weet het enkel in de voorwaardelijke wijs.

Je moet nu als volgt leren denken: “wat **zou** ik vinden als ik uit deze vaas een kaartje **zou gaan trekken**?”. Het antwoord hierop wordt je gegeven door het model dat hier de naam  $X$  (hoofdletter) heeft gekregen.  $X$  vertelt je: het zal een 1 zijn met kans  $1/6$ , of een 2 met kans  $1/6$ , of een 3, 4, 5 of 6, elk met kans  $1/6$ .

**Het zijn enkel die getallen met die kansen die je bij dit model kan krijgen.** Geen andere getallen en geen andere kansen.

Denk nu eens aan een dobbelsteen die “eerlijk” op elk zijvlak valt met kans  $1/6$ , maar waarbij jij de cijfers (het aantal ogen) op die zijvlakken hebt veranderd. Op 3 zijvlakken staat het cijfer 1, op twee zijvlakken het cijfer 3 en op het overblijvende zijvlak het cijfer 6. Voor het verdere verhaal zullen we aan deze dobbelsteen een kleur geven en hem de rode dobbelsteen noemen.

### Opdracht 1

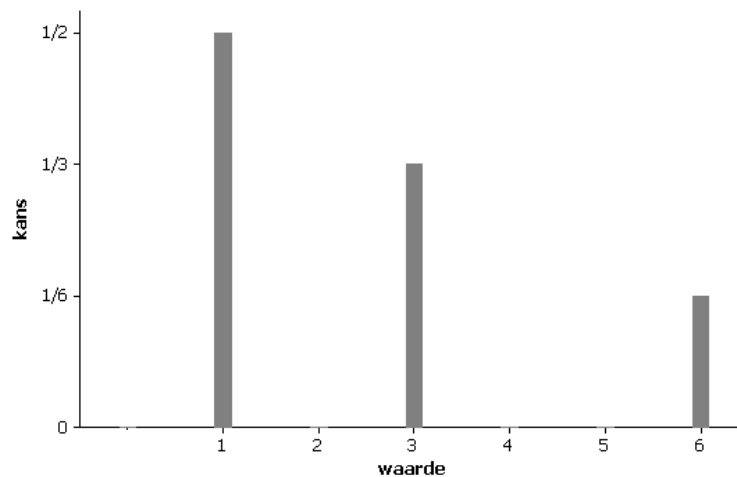
- Zeg in woorden welke kaartjes je in de vaas moet steken om een vaasmodel voor deze rode dobbelsteen te tekenen. Geef dit vaasmodel de naam  $X$ .
- Zeg in woorden wat je met  $X$  juist bedoelt. Doe dat door de volgende zin te vervolledigen. Je mag hierbij kiezen uit: “één”, “drie”, “zes”, “een derde”, “een zesde”, “getal”, “kansmodel”, “experiment”.

$X$  is het ..... dat een ..... beschrijft waarbij je als uitkomst een .....

kan hebben met kans  $1/2$ , een drie met kans ....., en een ..... met kans .....
- Maak nu een vaasmodel voor die rode dobbelsteen  $X$  waarbij er 600 000 kaartjes in de vaas zitten. Je moet die vaas niet tekenen, maar hoe zou jij die vaas vullen?
- Wat is het verschil tussen het vaasmodel dat jij in het begin hebt getekend voor die rode dobbelsteen en het vaasmodel met die 600 000 kaartjes? Zitten er in beide vazen evenveel kaartjes? Als je kijkt naar alle mogelijke verschillende uitkomsten, is dat dan hetzelfde voor beide vazen? En als je lukraak een kaartje trekt, heb je dan bij beide vazen dezelfde kans op eenzelfde uitkomst?

### 3.2. Een kansmodel voorstellen door een staafdiagram

Voor die rode dobbelsteen  $X$  heb je twee vazen op een verschillende manier met kaartjes gevuld, en toch hebben die vazen identieke modeleigenschappen. Je kan zeker nog aan veel andere vazen denken die je op een goede manier met kaartjes kan vullen en die ook allemaal het gedrag van die rode dobbelsteen weerspiegelen. Al die verschillende vazen stellen eenzelfde kansmodel voor en je kan al die vazen samenvatten in één en dezelfde grafiek, namelijk in een staafdiagram voor kansmodellen.



Kansmodel  $X$  voor de rode dobbelsteen

Figuur 2

Bij een staafdiagram voor kansmodellen zet je alle **verschillende** uitkomsten die kunnen voorkomen op de x-as. Boven elke uitkomst teken je een staafje. De lengte van dat staafje is de **kans** van die uitkomst. Omdat de uitkomsten discreet numeriek zijn moet de ijk gerespecteerd worden. Daarom is het handig om tussenliggende waarden, die niet als uitkomst voorkomen, toch op de x-as aan te duiden (zonder staafje erboven).

Let goed op het verschil tussen het vaasmodel en het staafdiagram.

-Bij het *vaasmodel* bepaal je de kansen door extra kaartjes toe te voegen. Als het getal één drie keer meer kans moet hebben dan het getal zes, dan stop je in die vaas drie keer zoveel kaartjes waar een één op staat als waar een zes op staat. Door zo te werken kan je daarna “lukraak” uit die vaas trekken. De juiste getallen komen dan met de juiste kansen tot jou.

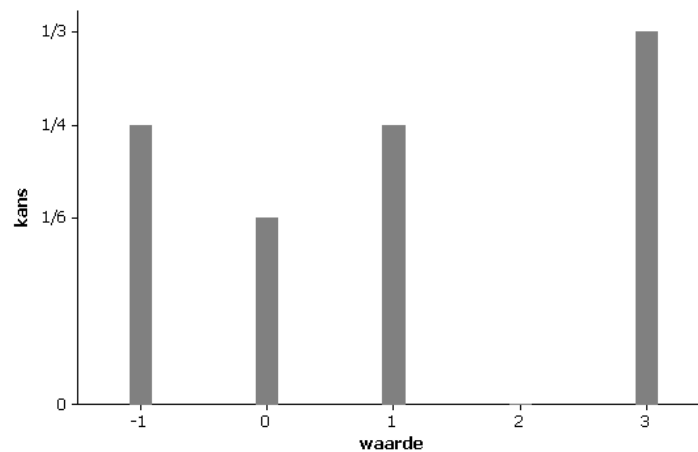
-Het *staafdiagram* toont je enkel de verschillende uitkomsten, en die worden maar één keer (met een staafje erboven) aangeduid. Op de figuur zie je alleen boven de getallen 1, 3 en 6 een staafje. Dat betekent dat alleen die getallen als uitkomst van je experiment kunnen optreden. De uitkomstenverzameling is dus  $\{1, 3, 6\}$ . Maar in deze uitkomstenverzameling mag je helemaal niet lukraak trekken, want anders zou elk van die drie getallen met dezelfde kans (namelijk kans  $\frac{1}{3}$ ) te voorschijn komen. Hoe die getallen dan wel tot jou komen wordt bepaald door de lengte van de staafjes. In dit voorbeeld staat boven het getal één een staafje met lengte  $\frac{1}{2}$  en dat betekent dat je te maken hebt met een experiment waarbij het getal één met kans  $\frac{1}{2}$  tot jou komt. Het getal drie treedt op met kans  $\frac{1}{3}$  en het getal zes komt met kans  $\frac{1}{6}$  voor.

### Opdracht 2

Als je met de rode dobbelsteen 450 keer zou gooien hoeveel keren zou je dan (ongeveer) een drie gezien hebben? Verklaar je redenering.

### Opdracht 3

- Beschrijf in woorden het kansmodel  $X$  dat hieronder met een staafdiagram wordt voorgesteld.



Kansmodel  $X$

Figuur 3

- Beschrijf voor dit staafdiagram een vaasmodel met zo weinig mogelijk kaartjes.



### 3.3. Een kansmodel voorstellen door een tabel

In het Engels heet kans “probability”. Het is de **P** van **Probability** die je gebruikt bij een kansuitspraak. De notatie ziet er uit als  $P(\dots)$ . Wat je tussen die haakjes schrijft moet te maken hebben met iets wat aan het toeval onderhevig is, en wat dus met een bepaalde kans kan gebeuren.

Soms zie je de notatie  $P(5)$  en dan bedoelt men “de kans om bij het gooien van een dobbelsteen een 5 te hebben”. De bedoeling is goed, maar de notatie is fout. Tussen de haakjes staat alleen het getal 5, en dat zegt niets over het kansmodel, dat is doodgewoon een vast getal.

Tussen de haakjes moet je zeggen over welk kansmodel het gaat en van welke uitkomst je de kans wil weten. Als het bijvoorbeeld over een klassieke dobbelsteen gaat dan kan je schrijven:

$P(\text{bij het gooien van een eerlijke dobbelsteen is het resultaat } 5)$ . Je kan die lange zin ook veel korter maken. Stel het kansmodel waarmee je hier werkt (namelijk het model van de eerlijke dobbelsteen) voor door de hoofdletter  $X$ . Schrijf dan gewoon:  $P(X \text{ levert het resultaat } 5)$ .

Of nog korter:  $P(X = 5)$ .

Soms wil je gewoon de kans opschrijven dat de dobbelsteen op één van zijn mogelijke uitkomsten valt, of dat nu een 5, een 6 of wat dan ook is. In dat geval heb je een algemene notatie nodig om “een uitkomst” voor te stellen. Daarvoor gebruik je een kleine letter  $x$  wanneer je het kansmodel voorstelt door een grote letter  $X$ . Je schrijft dan  $P(X = x)$  om aan te duiden dat het kansmodel  $X$  op één van zijn mogelijke waarden valt. Dat is gewoon maar een algemene notatie. Als je echt een kans wil uitrekenen dan moet je zeggen wat  $x$  is, bijvoorbeeld 6. En dan wordt in die situatie  $P(X = x)$  gelijk aan  $P(X = 6)$ . Met een kleine letter  $y$  bedoel je natuurlijk een uitkomst van het kansmodel dat je de naam grote  $Y$  gegeven hebt. Als algemene notatie gebruik je dan  $P(Y = y)$ .

Je kan ook willen weten wat de kans is om hoogstens een 4 te gooien. Zoals verwacht noteer je dat als  $P(X \leq 4)$ . Wanneer  $X$  een eerlijke dobbelsteen is dan is die kans gelijk aan  $4/6$ .

#### Opdracht 4

Lees  $P(X < 3)$  in woorden. Hoe groot is die kans bij een eerlijke dobbelsteen?

Bij het opgooien van een muntstuk kan je zeggen dat munt gelijk is aan nul en dat kruis één is. Je hebt dan een experiment dat je met een eenvoudig kansmodel kan beschrijven. Geef aan dit kansmodel de naam  $X$ . Onderstel dat een muntstuk bij opgooien altijd “eerlijk” is.

**Opdracht 5**

- Stel  $X$  voor met een vaasmodel
  
- Stel  $X$  voor met een staafdiagram
  
- Vul de volgende zinnen aan met: “kansmodel  $X$ ”, “muntstuk”, “kans”, “kruis”, “munt”, “waarde”, “uitkomst”. Bij de/het kies je op de juiste manier tussen “de” en “het”.
  - $P(X = 1)$  is de/het .....dat de/het ..... de/het ..... één aanneemt
  - $P(X = 1)$  is de/het ..... dat de/het ..... op ..... valt
  - $P(X = 1) = 1/2$  : de/het ..... levert een ..... één met .....1/2.

Nu je vertrouwd bent met de notatie kan je een kansmodel eenvoudig in tabelvorm voorstellen. Voor het opgooien van een muntstuk heb je volgende tabel.

het kansmodel $X$	de uitkomsten van het model $X$	$x$	0	1
	de bijhorende kansen voor dit model $X$	$P(X=x)$	$\frac{1}{2}$	$\frac{1}{2}$

**Opdracht 6**

Kijk eens naar het experiment dat beschreven wordt in figuur 3 en vergelijk dat met de onderstaande tabel. Vind je op beide plaatsen dezelfde informatie terug? Denk eraan dat een uitkomst van het model  $X$  genoteerd wordt door een kleine letter  $x$ .

het kansmodel $X$	de uitkomsten van het model $X$	$x$	- 1	0	1	3
	de bijhorende kansen voor dit model $X$	$P(X=x)$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$

- In de tabel staat het cijfer nul. Wat is dat hier? Kan je dat in woorden uitleggen?
- Schrijf een formule waarin de  $4/12$  van de tabel voorkomt. Gebruik de juiste notatie. Schrijf daarna die formule ook in woorden.
- Als je het verschil tussen hoofdletters en kleine letters goed begrijpt, dan kan je in de bovenstaande tabel de extra uitleg weglaten. Als je op die manier het kansmodel  $X$  van de rode dobbelsteen opschrijft, dan krijg je (vul de getallen zelf in):

$x$		6
$P(X=x)$		1/6

Kansmodel  $X$  voor de rode dobbelsteen

### 3.4. De kansverdeling

Een kansmodel ken je volledig als je alle mogelijke uitkomsten samen met hun bijhorende kansen kent.

Elke beschrijving van alle uitkomsten samen met hun kansen noem je een **kansverdeling**. Het woord kansverdeling gebruik je voor kansmodellen met **discrete uitkomsten**. Daar kan je immers alle uitkomsten en hun kansen opsommen. Soms stel je zo'n kansmodel voor door een vaas. Je kan ook een staafdiagram tekenen. En in andere gevallen gebruik je een tabel.

Je zal verderop zien dat je bij continue uitkomsten het woord **kansdichtheid** moet gebruiken om een kansmodel aan te duiden.

#### Opdracht 7

$x$	- 1	0	1	3
$P(X=x)$	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$

- Kijk naar de getallen in bovenstaande tabel die op dezelfde rij staan als waarop ook  $P(X=x)$  staat. Zou daar, bij een ander experiment, ook eens een negatief getal kunnen tussen staan, of bijvoorbeeld het cijfer 3 ? Waarom?
- Tel in de tabel alle kansen samen. Hoeveel vind je? Zou je een ander totaal kunnen vinden bij een andere kansverdeling? Waarom?
- Welke algemene eigenschappen heb je hier ontdekt voor elke kansverdeling?

## 4. Kansmodellen voor experimenten met continue uitkomsten

Voor een experiment met continue uitkomsten kan je geen kansmodel maken met behulp van een vaas, een staafdiagram of een tabel. In al die gevallen moet je immers elke mogelijke uitkomst kunnen opnoemen. Je moet ook kunnen zeggen wat de kans is die bij elk van die uitkomsten hoort.

Als uitkomsten een gebied bestrijken, zoals alle getallen in het interval  $[0; 1]$ , dan kan je die getallen niet meer afzonderlijk opnoemen. Probeer dat maar eens. Neem als eerste getal nul. Wat is nu het volgende getal? Wat is het getal dat na nul komt? Dat gaat dus niet.

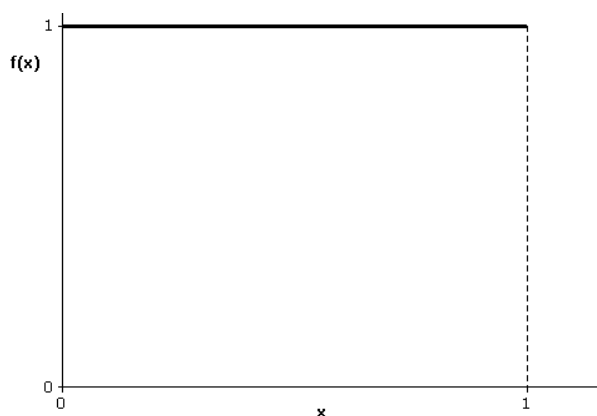
### 4.1. Een eerste voorbeeld

Bij experimenten met continue uitkomsten kan je de afzonderlijke uitkomsten niet opnoemen. Je kan dus ook niet meer werken met de kans om één of ander getal uit te komen. Hoe moet het dan wel?

Werken met een continu kansmodel betekent overstappen van “getallen” op “intervallen”. Voor een experiment waarbij de mogelijke uitkomsten alle getallen zijn in het interval  $[0; 1]$  kan je naar deelintervallen kijken. Je kan terechtkomen in  $[0; 1/4]$  of in  $[3/5; 1]$  of in  $[0.55; 0.65]$  enz.

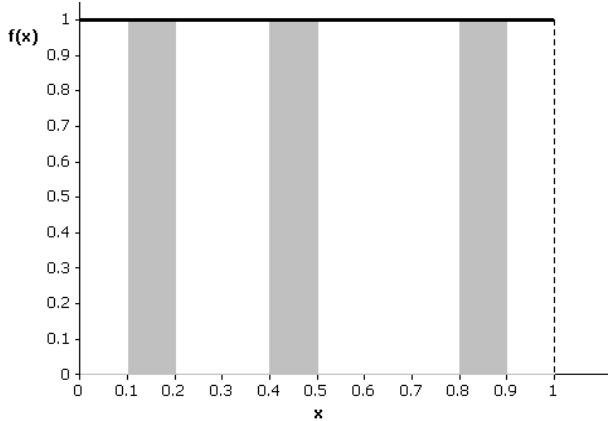
Je moet nu ook een manier vinden om de kans te bepalen dat je in een bepaald interval terechtkomt. Dat doe je met oppervlakten.

Een heel eenvoudige situatie krijg je als je “lukraak trekken” wil nabootsen. In dat geval moet je er voor zorgen dat er geen enkel deelinterval bevoor- of benadeeld wordt. Als intervallen even groot zijn dan moet je er met dezelfde kans in terechtkomen. Boven die intervallen moeten dan oppervlakten staan die even groot zijn. Als een interval drie keer groter is dan een ander, dan moet je daar ook met drie keer meer kans in terechtkomen. Boven dat interval moet dan een oppervlakte staan die drie keer groter is. Aan al deze eisen kan je tegelijkertijd voldoen door boven het interval  $[0; 1]$  de grafiek te tekenen van de constante functie  $f(x) = 1$ . Zo krijg je een kansmodel voor “lukraak trekken uit  $[0; 1]$ ”. Kijk maar.

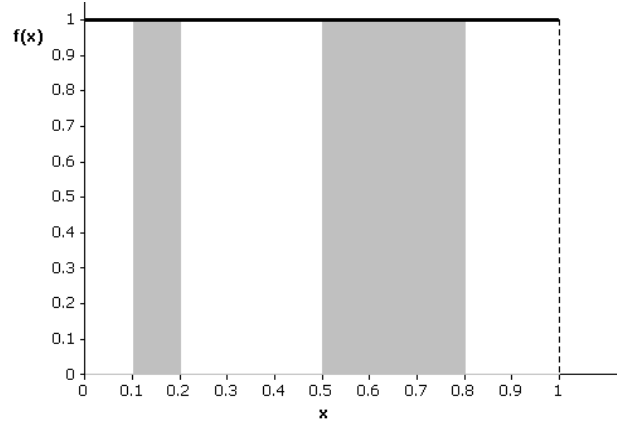


Figuur 4

Wat de maatgetallen betreft (en vanaf nu zullen we alleen met maatgetallen werken) is voor elk deelinterval in  $[0; 1]$  de oppervlakte boven dat interval gelijk aan de lengte van dat deelinterval. De oppervlakte van een rechthoek is basis  $\times$  hoogte, en de hoogte is hier overal gelijk aan 1.



Figuur 5



Figuur 6

De deelintervallen  $[0.1; 0.2]$ ,  $[0.4; 0.5]$  en  $[0.8; 0.9]$  zijn allemaal even groot. De oppervlakte boven die intervallen is gelijk aan 0.1 en dat is ook de kans om in zo'n interval terecht te komen.

Het interval  $[0.5; 0.8]$  is drie keer groter dan het interval  $[0.1; 0.2]$ . De kans om in  $[0.5; 0.8]$  terecht te komen is 0.3 en dat is ook drie keer groter dan 0.1.

Bemerkt dat het kansmodel, dat hier door de functie  $f(x) = 1$  gestuurd wordt, je alles vertelt over hoe je in deelintervallen van  $[0; 1]$  terechtkomt. Voor alle deelintervallen die je maar wil. Zo heb je bijvoorbeeld een kans van 0.245 om in  $[0.646; 0.891]$  terecht te komen want de oppervlakte boven het interval  $[0.646; 0.891]$  is gelijk aan 0.245.

### Opdracht 8

Vervolledig de volgende zin met een keuze uit:

“x-as”, “y-as”, “hoogte”, “oppervlakte”, “functiewaarde”.

De mogelijke waarden die ik met een continu kansmodel kan uitkomen zie ik op de .....

Om te weten wat de kans is om in een bepaald interval terecht te komen kijk ik naar de .....

## 4.2. Ongelijkheidstekens bij continue kansmodellen

Bij het berekenen van de kans om in een interval terecht te komen moet je naar de oppervlakte boven dat interval kijken. Op figuur 6 zie je dat de oppervlakte boven  $[0.5; 0.8]$  gelijk is aan 0.3. Maar dat is natuurlijk ook de oppervlakte boven  $[0.5; 0.8[$  of boven  $]0.5; 0.8]$  of boven  $]0.5; 0.8[$  want de oppervlakte van het lijnstuk dat je weglaat is toch nul.

Dit betekent dat de volgende kansen allemaal aan elkaar gelijk zijn:

$$P(0.5 \leq X \leq 0.8) = 0.3$$

$$P(0.5 \leq X < 0.8) = 0.3$$

$$P(0.5 < X \leq 0.8) = 0.3$$

$$P(0.5 < X < 0.8) = 0.3$$

Let op. De intervallen  $[0.5; 0.8]$ ,  $[0.5; 0.8[$ ,  $]0.5; 0.8]$  en  $]0.5; 0.8[$  zijn wel degelijk 4 verschillende intervallen maar de kans, bij een continu kansmodel  $X$ , om in gelijk welk van die 4 intervallen terecht te komen is telkens dezelfde.

### Opdracht 9

Bij een continu kansmodel is  $P(X \leq x)$  gelijk aan  $P(X < x)$ . Is dat ook waar voor een discreet kansmodel? Staaf je antwoord met een voorbeeld.

## 4.3. De kansdichtheid

Bij het voorgaande experiment, lukraak trekken uit  $[0; 1]$ , zou je voor de volledigheid de functie  $f(x)$  als volgt kunnen definiëren:

$f(x) = 0$  voor alle  $x$ -waarden tussen min oneindig en nul (kleiner dan 0)

$f(x) = 1$  voor alle  $x$ -waarden tussen nul en één

$f(x) = 0$  voor alle  $x$ -waarden tussen één en plus oneindig (groter dan 1)

of in wiskundige notatie: 
$$f(x) = \begin{cases} 0 & \text{voor } -\infty < x < 0 \\ 1 & \text{voor } 0 \leq x \leq 1 \\ 0 & \text{voor } 1 < x < +\infty \end{cases}$$

Op deze manier definieer je  $f(x)$  op de volledige getallen, voor  $x$ -waarden van  $-\infty$  tot  $+\infty$ . Je hebt hierbij  $f(x)$  gelijk aan nul gesteld over al de gebieden waarin je toch niet kan terechtkomen. Dikwijls moet je zo'n uitgebreide definitie niet geven. Je weet voldoende als je  $f(x)$  kent over het gebied waarin je experiment "leeft" en waar je dus echt kan terechtkomen.

Bemerk dat in dit voorbeeld de functie  $f(x)$  nergens negatief is want ze is ofwel nul ofwel één. Bovendien is de totale oppervlakte onder de grafiek gelijk aan één. Dat zie je in figuur 4.

Elke functie  $f(x)$  die nergens negatief is en waarbij de totale oppervlakte onder de grafiek gelijk is aan één stelt een **continu** kansmodel  $X$  voor. Zo'n functie heet een **kansdichtheid** of een **dichtheidsfunctie** of kortweg een **dichtheid**. Als je de dichtheid van  $X$  kent, dan ken je  $X$  volledig. Je weet dan in welk interval de uitkomsten kunnen vallen en je weet ook, voor alle mogelijke deelintervallen, met welke kans je daarin terechtkomt.

Modellen waarbij “alles dezelfde kans” heeft krijgen de naam “**uniform**”.

- Bij discrete modellen spreek je van “**discreet uniform**”. Dat is bijvoorbeeld het geval bij een eerlijke dobbelsteen waar elke mogelijke uitkomst dezelfde kans  $1/6$  heeft.
- Bij continue modellen spreek je over “**continu uniform**”. Daar hebben alle intervallen die even lang zijn dezelfde kans om erin terecht te komen. Zo is bijvoorbeeld het kansmodel voor “lukraak” trekken uit  $[0; 1]$  een *continu uniform kansmodel op  $[0; 1]$* .

#### 4.4. Een tweede voorbeeld

Een ander continu kansmodel moet je nu eens zelf proberen te ontdekken.

##### 4.4.1. Ik wil de grootste zijn

Steven en Pol gaan monopoly spelen. Om te zien wie er mag beginnen kunnen zij met een dobbelsteen gooien. Wie het grootste getal gooit mag beginnen. Als zij hetzelfde getal gooien dan mag de jongste beginnen (dat is Pol). Maar er is een probleem... zij hebben geen dobbelsteen.

Steven weet raad. Hij neemt zijn GRM, drukt **MATH**, loopt naar PRB en drukt 1: rand en **ENTER**. Zo trekt hij lukraak een getal tussen 0 en 1. Hij vindt 0.3631... Dan neemt Pol die GRM en doet juist hetzelfde. Hij vindt 0.8267... Pol wint.

Het getal van Pol is het grootste. Het is het maximum van die 2 gevonden getallen. Jij moet nu, in de volgende opdracht, op speurtocht gaan naar het gedrag (kansmodel) van zo'n maximum.

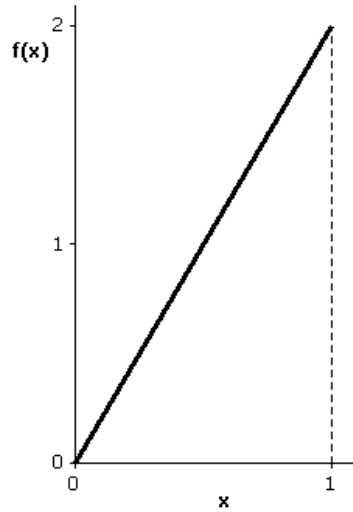
#### Opdracht 10

- Moet je voor “het maximum” hier werken met een discreet of met een continu kansmodel? Waarom?
- Tussen welke waarden zal gelijk welk gevonden maximum hier terechtkomen? Waarom?
- Nu je weet in welk gebied je met dat maximum terechtkomt, moet je op zoek gaan naar het passende kansmodel. Je krijgt hierbij een tip over de vorm van de dichtheidsfunctie. Die vorm is een lijnstuk dat begint in het punt  $(0, 0)$ . Je moet op zoek naar de vergelijking van een rechte. Lukt het je nu om het kansmodel voor dat maximum te vinden? Geef aan dat kansmodel de naam  $X$ . Toon je redenering.



### 4.4.2. Het maximum van 2 lukrake getallen uit [ 0 ; 1 ]

Het kansmodel voor **het maximum** van 2 getallen die, onafhankelijk van elkaar, getrokken zijn uit een continu uniform kansmodel op [ 0 ; 1 ] zie je in de volgende figuur.

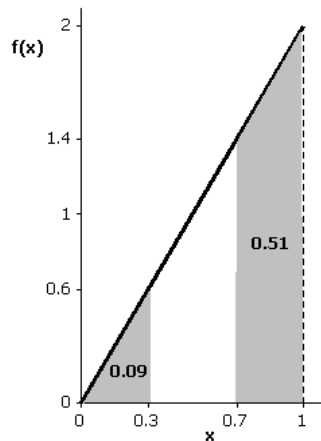


Figuur 7

Boven het interval [ 0 ; 1 ] zie je een grafiek die je herkent als een stuk van de rechte  $y = 2x$  (een rechte door de oorsprong en met richtingscoëfficiënt 2). Let op! Je experiment “leeft” in dit voorbeeld enkel op [ 0 ; 1 ]. Dat betekent dat de dichtheidsfunctie hier enkel op [ 0 ; 1 ] gelijk is aan  $f(x) = 2x$ . Buiten [ 0 ; 1 ] is  $f(x)$  gelijk aan nul. Dus:

$$f(x) = \begin{cases} 0 & \text{voor } -\infty < x < 0 \\ 2x & \text{voor } 0 \leq x \leq 1 \\ 0 & \text{voor } 1 < x < +\infty \end{cases}$$

Als je naar de figuur kijkt dan zie je dat dit nieuwe kansmodel helemaal geen nabootsing is van “lukraak trekken” uit [ 0 ; 1 ]. Deelintervallen die even groot zijn hebben niet meer dezelfde kans om erin terecht te komen. Vergelijk maar eens [ 0 ; 0.3 ] met [ 0.7 ; 1 ], twee deelintervallen die beide lengte 0.3 hebben .



Figuur 8

De kans om in  $[0; 0.3]$  terecht te komen wordt gegeven door de oppervlakte die boven  $[0; 0.3]$  staat. Je ziet daar een driehoek met basis 0.3 en hoogte 0.6 zodat de oppervlakte 0.09 is. Je valt dus met kans 0.09 in het interval  $[0; 0.3]$ .

De oppervlakte boven  $[0.7; 1]$  kan je opsplitsen in een rechthoek waarop een driehoek staat. De oppervlakte van de rechthoek is  $\text{basis} \times \text{hoogte} = 0.3 \times 1.4 = 0.42$ . De oppervlakte van de driehoek is, zoals daarnet, gelijk aan 0.09. Samen heb je dus een kans van 0.51 om in  $[0.7; 1]$  terecht te komen.

Voor dit kansmodel  $X$  geldt dus dat  $P(0.7 \leq X \leq 1) = 0.51$ .

In woorden lees je deze kansuitspraak als:

- de kans dat de resultaten van het kansmodel  $X$  terechtkomen in het interval  $[0.7; 1]$  is 0.51
- met een kans van 51 % levert het model  $X$  uitkomsten die minstens 0.7 en hoogstens 1 zijn.

Het is belangrijk dat je vooraf goed aangeeft met welk model  $X$  je aan het werken bent. Je hebt gezien dat er nogal een verschil is tussen een model  $X$  dat gedefinieerd wordt door figuur 4 en een model  $X$  dat een dichtheid heeft zoals in figuur 7.

### Opdracht 11

Het model  $X$  van figuur 7 heeft als dichtheidsfunctie  $f(x) = 2x$  voor  $0 \leq x \leq 1$ . Voor  $x = 1/4$  is de functiewaarde gelijk aan  $1/2$ . Kan je  $f(1/4) = 1/2$  in woorden zeggen? Kan je hierbij het woord “kans” gebruiken? Hoe zeg je dat dan? Kijk daarna ook eens wat er gebeurt bij  $x = 3/4$ . Welke les trek je hieruit?

## **4.5. Kansen berekenen met de GRM**

Bij elk continu model  $X$  kan je de kans bepalen om terecht te komen in gelijk welk door jou gekozen interval. Je berekent daarvoor de oppervlakte boven dat interval, tot aan de dichtheidsfunctie. Zo'n berekening wordt snel ingewikkeld en dan komt je GRM goed van pas.

Bemerk dat je bij deze berekeningen werkt met de echte dichtheidsfunctie en dat je dus de echte kansen van het kansmodel berekent.

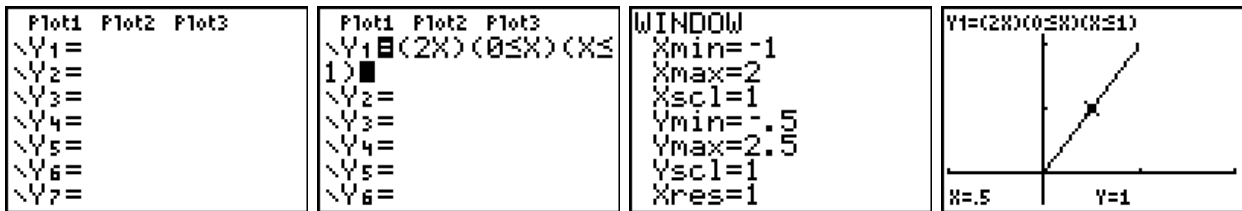
### Opdracht 12

Als voorbeeld controleer je voor figuur 8 of je ook met je GRM aan  $P(0.7 \leq X \leq 1) = 0.51$  komt. Zorg ervoor dat je daarbij “de oppervlakte onder de curve” grafisch voorstelt.

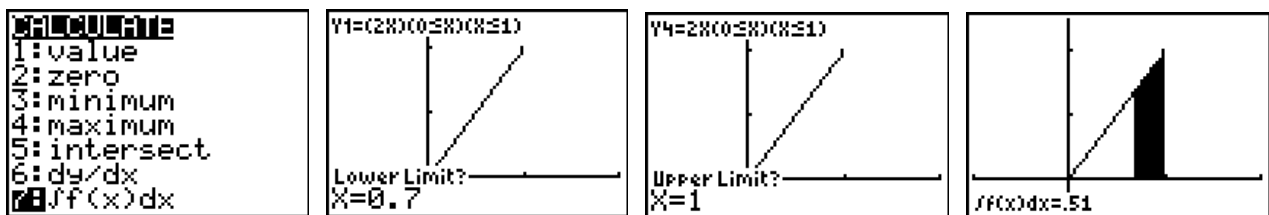
Controleer eerst of alle statistische plots af staan. Druk  $\text{[2nd][STAT PLOT]}$ , tik dan  $\text{[4]}$  en  $\text{[ENTER]}$ .

Druk nu op de toets  $\text{[Y=]}$  en vul bij  $Y_1=$  (of op een andere  $Y=$  plaats die nog vrij is) het functievoorschrift in. Je wil de functie  $f(x) = 2x$  enkel op het interval  $[0; 1]$ . Daarom zeg je wat de grenzen van  $x$  zijn, namelijk  $0 \leq x$  en  $x \leq 1$ . Voor het ongelijkheidsteken druk je  $\text{[2nd][TEST]}$  en dan  $\text{[6]}$ .

Vooraleer de figuur te tekenen kan je best de vensterinstellingen aanpassen. Druk op  $\text{[WINDOW]}$  en vul de waarden in die je hier ziet. Druk daarna op  $\text{[TRACE]}$ . Je GRM tekent nu de dichtheidsfunctie.



Om de oppervlakte boven het interval  $[0.7; 1]$  te berekenen ga je als volgt te werk. Vertrek vanuit het scherm waarop je de grafiek ziet. Druk dan  $\text{[2nd][CALC]}$  en  $\text{[7]}$ . Je berekent nu (met een bepaalde integraal) de oppervlakte onder de curve. Daarbij moet je het beginpunt van het interval aangeven. Je GRM vraagt naar de Lower Limit en dan tik je  $0.7$  en  $\text{[ENTER]}$ . Daarna wordt de Upper Limit (eindpunt) gevraagd en dan tik je  $1$  en  $\text{[ENTER]}$ . Je krijgt nu de gearceerde oppervlakte boven het interval  $[0.7; 1]$  en je kan onderaan aflezen dat deze oppervlakte gelijk is aan  $0.51$ . Dat is inderdaad de kans om in  $[0.7; 1]$  terecht te komen.



**Opdracht 13**

Gebruik je GRM om de kans te berekenen dat hetzelfde kansmodel  $X$  in het interval  $[0; 0.3]$  terechtkomt. Je kent het resultaat al want het staat in figuur 8. Hint: als je eerst het gearceerde deel op je GRM wil wissen, kan dat met  $\text{[2nd][DRAW]}$  1:ClrDraw.

## 5. Populatiemodellen

In een studie over M&M-snoepjes maakt men een onderscheid tussen “alle M&M-snoepjes die door de fabrikant gemaakt worden” en “de snoepjes in jouw zakje M&M’s”. In de statistiek spreek je over een **populatie** als je “de totale verzameling” bedoelt (dus alle M&M-snoepjes). Meestal heb je geen tijd of geld om een volledige populatie te bestuderen en daarom bekijk je enkel een klein deeltje van die populatie. Zo’n deeltje noem je een **steekproef** en het is belangrijk dat die steekproef op een goede manier getrokken wordt. Zo kan je een idee krijgen over hoe de hele populatie eruitziet.

Steekproefuitkomsten zijn aan het toeval onderhevig. Als je vandaag een steekproef trekt uit een populatie en morgen terug een steekproef uit diezelfde populatie dan zal je (bijna zeker) andere uitkomsten hebben.

Een populatie is niet aan het toeval onderhevig. Het is een vaste verzameling die je (meestal) niet volledig kent.

**Vraag:** als een populatie vast is, waarom beschrijft men dan een populatie met een kansmodel?

**Antwoord:** denk aan een *vaste* populatie als aan een *vast* mechanisme dat getallen naar jou stuurt. Welke getallen je kan krijgen en met welke kans hangt af van hoe die *vaste* populatie eruitziet. Een populatiemodel is dus niets anders dan een kansmodel dat zegt welke uitkomsten met welke kans tot jou komen wanneer je uit die populatie trekt.

Populatiemodellen kan je op verschillende manieren ontmoeten.

- Je kan vertrekken vanuit de wiskundige definitie van kansmodellen en dan verdere eigenschappen bestuderen. Je hebt hiervoor geen concrete context nodig.
- Je kan denken aan concrete experimenten waarvan je de eigenschappen in een model vastlegt zonder dat je het experiment echt uitvoert.
- Je kan binnen een concrete studie beschikken over steekproefgegevens samen met bijkomende informatie over de context van die studie. Van hieruit krijg je een idee over hoe de volledige populatie er zou kunnen uitzien. Dit gebruik je als basis om een populatiemodel op te stellen.

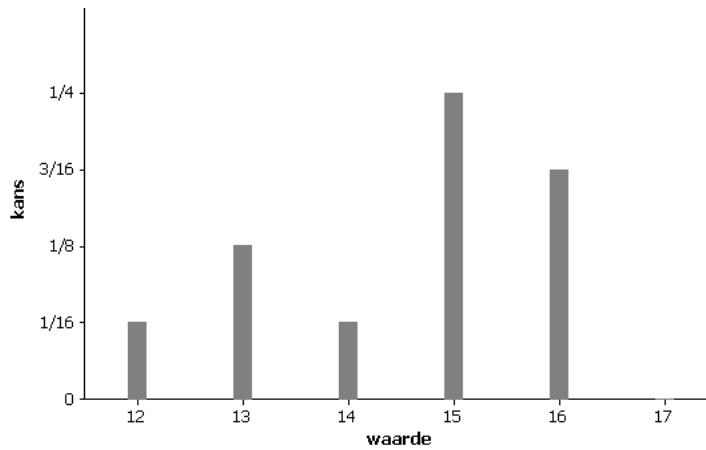
Hieronder zie je enkele voorbeelden.

### 5.1. Populatiemodel vanuit een wiskundige definitie

Voor de volgende opdracht heb je geen verhaal nodig over munten, dobbelstenen of wat dan ook. Vanuit de wiskundige definitie kan je met een kansmodel werken. Als je dit model gebruikt om je populatie te beschrijven dan leg je vast welke uitkomsten mogelijk zijn en met welke kansen.

**Opdracht 14**

- Het volgende staafdiagram stelt een kansmodel  $X$  voor waarbij de uitkomstenverzameling gelijk is aan  $\{12, 13, 14, 15, 16, 17\}$  maar het staafje boven 17 is weggevallen. Hoe hoog moet dat staafje zijn? Waarom?



Kansmodel  $X$

Figuur 9

- Maak een kansverdeling in tabelvorm voor het kansmodel  $X$  dat je uit figuur 9 kan afleiden. Let op de juiste notatie.


- Vul in (en toon je berekening):  $P(X \leq 14) = \dots\dots\dots$

## 5.2. Populatiemodel vanuit een abstract experiment

Hierboven heb je gedacht aan experimenten zoals:

- trek lukraak een kaartje uit een vaas waarin 6 kaartjes zitten met daarop een 1,2,3,4,5,6.
- gooi een eerlijke dobbelsteen.

Je hebt nog nooit met zo’n vaas gewerkt en een dobbelsteen heb je ook niet gegooid. En toch heb je een kansmodel opgesteld. Je hebt gedacht aan alle mogelijke uitkomsten en aan de kansen waarmee die uitkomsten tot jou zouden komen als je zo’n experiment echt zou doen.

Je hebt “het gooien van een eerlijke dobbelsteen” toen voorgesteld met een vaasmodel.

### Opdracht 15

Maak een kansverdeling in tabelvorm voor het kansmodel  $X$  dat het gooien van een eerlijke dobbelsteen beschrijft. Let op de juiste notatie.


Je hebt vroeger ook gedacht aan experimenten zoals:

- trek lukraak een kaartje uit een vaas waarin 600 000 kaartjes zitten: 300 000 kaartjes met het cijfer 1, 200 000 kaartjes met het cijfer 3 en 100 000 kaartjes met het cijfer 6.
- gooi een rode dobbelsteen die “eerlijk” op elk zijvlak valt met kans  $1/6$ , maar waarbij de cijfers (het aantal ogen) op die zijvlakken zijn veranderd. Op 3 zijvlakken staat het cijfer 1, op twee zijvlakken het cijfer 3 en op het overblijvende zijvlak het cijfer 6.

Ook hier heb je niet met een vaas gewerkt en zo’n rode dobbelsteen heb je nog nooit gezien. Maar je was blijkbaar toch in staat om onderstaande kansverdeling op te stellen.

$x$	1	3	6
$P(X=x)$	$3/6$	$2/6$	$1/6$

“Het gooien van de rode dobbelsteen” of “het lukraak trekken van kaartjes uit die vaas” kan je beschouwen als een onderliggende populatie die getallen naar jou stuurt, exact zoals aangegeven in de kansverdeling.

Als je uit de populatie van de rode dobbelsteen zou gaan trekken, dan verwacht je een 1 met kans  $3/6$ , een 2 met kans  $2/6$  en een 6 met kans  $1/6$ .

### 5.3. Populatiemodel vanuit concrete opmetingen

In echte statistische studies krijg je het populatiemodel niet zomaar gegeven als een wiskundig model of op basis van theoretische overwegingen over hoe een populatie zich zou gedragen.

Om een ongekende populatie te beschrijven maak je gebruik van informatie die je haalt uit (grote) steekproeven. Maar alleen een hoop getallen is niet genoeg. Je moet ook de context kennen: wat is de bedoeling van de studie, wat is er juist opgemeten, wat weet je uit vroegere studies, ... ?

Al die informatie gebruik je dan om een kansmodel op te stellen met de bedoeling:

- NIET om je gevonden steekproef te benaderen
- WEL om te dienen als een model voor de ongekende populatie.

#### Opdracht 16

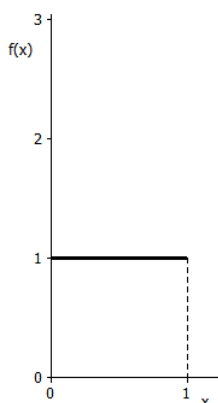
In deze opdracht moet je vanuit een steekproef een zinvol voorstel formuleren om de onderliggende populatie met een kansmodel te beschrijven.

Dit is de eerste keer dat je zoiets doet. Daarom is het goed dat je eerst een uitgewerkt voorbeeld doorloopt. Doe dat actief, met een eigen steekproef die je zelf trekt.

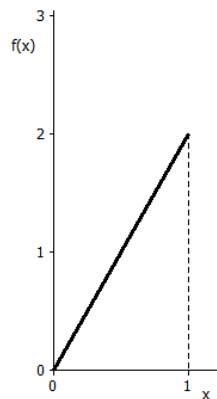
Zorg dat de programma's `FREQCONT`, `HISDICH` en `ZOEKPOP` in je GRM staan en dat er voldoende werkgeheugen (RAM) vrij is.

#### Gegeven.

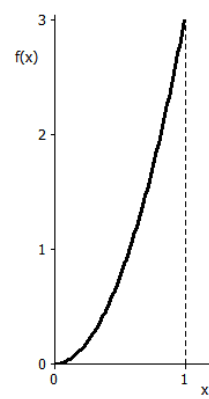
Uit voorafgaand onderzoek samen met de context van de studie weet je al dat het gaat over **continue** gegevens uit een onderliggende populatie die kan beschreven worden met één van de volgende 3 dichtheidsfuncties:



*Populatie A*  
 $f(x) = 1$   
 $0 \leq x \leq 1$



*Populatie B*  
 $f(x) = 2x$   
 $0 \leq x \leq 1$



*Populatie C*  
 $f(x) = 3x^2$   
 $0 \leq x \leq 1$

Gevraagd.

Gebruik de gegeven informatie samen met een steekproef uit de ongekende populatie om te bepalen met welk model je die populatie zinvol kan beschrijven.

Oplossing.

Het programma ZOEKPOP trekt een lukrake steekproef uit de te ZOEKEN POPulatie.

Druk **[PRGM]**, loop naar ZOEKPOP en druk 2 keer **[ENTER]**. Vul in dat je een steekproef van grootte  $n = 500$  wil trekken en druk **[ENTER]**. De steekproefresultaten komen terecht in lijst [L1].

Er is gegeven dat het gaat om continue gegevens. Je weet dat je die grafisch kan voorstellen met een histogram. Daarvoor heb je een frequentietabel met klassenindeling nodig. Om die op te stellen moet je het totale gebied kennen waarin je getallen zijn terecht gekomen. Je kan dan dat gebied opdelen in een aantal deelintervallen en zo de klassengrenzen bepalen.

Om het maximum en minimum van de getrokken getallen te vinden druk je **[STAT]**, je loopt naar CALC en drukt dan 1:1-Var Stats. Bij List: vul je [L1] in, FreqList: laat je blanco en dan loop je naar Calculate en druk je **[ENTER]**.

```

1-Var Stats
List:L1
FreqList:
Calculate

```

In dit voorbeeld is  $\min X = 0.12$  en  $\max X = 0.99$ . In jouw steekproef zal dat iets anders zijn.

De gevonden getallen liggen tussen 0 en 1 en dit interval ga je nu opdelen, bijvoorbeeld in 10 deelintervallen (klassen). De klassengrenzen, namelijk 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 kan je met de hand intikken in lijst [L2] maar je kan ook als volgt tewerk gaan:

```

1-Var Stats
n=500
minX=.12172362
Q1=.6352507407
Med=.793189988
Q3=.907971851
maxX=.99949785

```

druk **[2nd][LIST]**, loop naar OPS en druk 5:seq(. Vul dan het venster in zoals aangegeven, loop naar Paste, druk **[ENTER]** en dan **[STO▶]** en **[2nd][L2]** en **[ENTER]**.

```

seq
Expr:X
Variable:X
start:0
end:1
step:0.1
Paste

```

Nu de klassen vastgelegd zijn moet je zoeken hoeveel getallen in die klassen terechtkomen. Dat doet het programma FREQCONT dat een FREquentietabel met klassenindeling opstelt voor CONTinue gegevens. Druk **[PRGM]**, loop naar FREQCONT en druk 2 keer **[ENTER]**. Je hebt nu een frequentietabel met klassengrenzen in [L2] en bijhorende frequenties in [L3]. Dit is de informatie die het programma HISDICH nodig heeft om een HISTogram op de DICHtheidsschaal te tekenen. Dat betekent een histogram waarbij de totale oppervlakte gelijk is aan 1, juist zoals bij een dichtheidsfunctie (waar de totale oppervlakte onder de curve ook gelijk is aan 1).

Teken nu zo'n histogram. Druk **[PRGM]**, loop naar HISDICH en druk 2 keer **[ENTER]**. Als je het keuzeschermb ziet, kies je 1 (voor histogram) en druk je **[ENTER]**.

```

1 = Histogram
2 = Veelhoek
3 = Beide
4 = Histo+Norm.
5 = Stop
nummer+ENTER
?

```

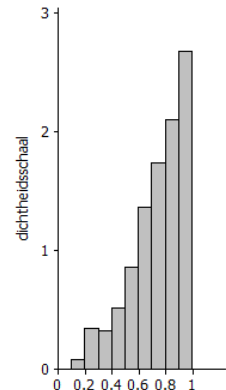
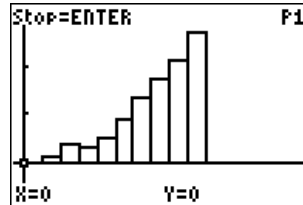


Kijk nu goed naar het histogram en let op

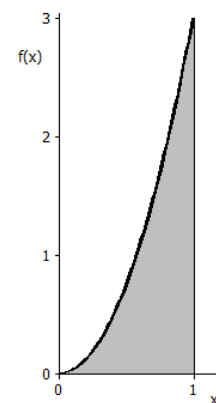
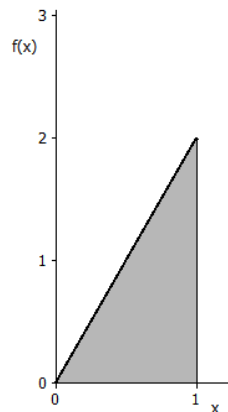
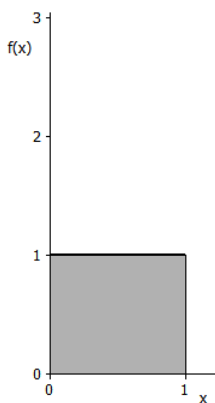
- de ligging op de x-as (hier ligt inderdaad alles tussen 0 en 1)
- de globale vorm (als een “volle” figuur).

**Nota**

In het venster van je GRM is de fysische lengte van 1 eenheid op de x-as meer dan 3 keer de fysische lengte van 1 eenheid op de y-as. Als dit voor verwarring zorgt, dan kan je het histogram overtekenen op dezelfde schaal als die waarop de dichtheidsfuncties zijn getekend. Je moet daar niets voor berekenen want de klassengrenzen staan nog altijd in [L2] en de bijhorende hoogtes vind je in [L6]. Beide figuren zie je hiernaast voor de resultaten van dit voorbeeld.



Je moet nu de globale vorm van je gevonden histogram vergelijken met de globale vorm van de opgegeven dichtheidsfuncties, die er als “volle” figuur uitzien zoals hieronder:



Ondanks de schommelingen in het histogram (eigen aan een toevallige steekproef – kijk maar naar de histogrammen van verschillende medeleerlingen) sluit de globale vorm van het histogram het beste aan bij de derde figuur (populatie C).

Druk **[ENTER]** om de figuur te verlaten en tik daarna 4 en **[ENTER]** om te stoppen.

**Besluit.**

Op basis van de extra informatie die je vooraf hebt gekregen, samen met de resultaten van je steekproef, besluit je om met  $f(x) = 3x^2$  voor  $0 \leq x \leq 1$  te werken als model voor de populatie in dit onderzoek.

**Opdracht 17**

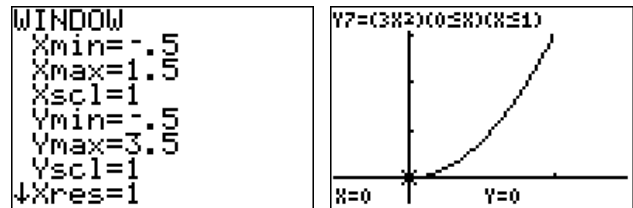
Een continue populatie  $X$  heeft als dichtheidsfunctie  $f(x) = 3x^2$  voor  $0 \leq x \leq 1$ . Wat is de kans om in het interval  $[0 ; 0.45]$  terecht te komen als je uit deze populatie trekt?

Bereken die kans met je GRM en zorg ervoor dat je daarbij “de oppervlakte onder de curve” grafisch voorstelt. Ga tewerk zoals in opdracht 12.

Controleer eerst of alle statistische plots af staan. Druk  $\boxed{2nd}$ [STAT PLOT], tik dan  $\boxed{4}$  en  $\boxed{ENTER}$ .

Druk nu op de toets  $\boxed{Y=}$  en vul bijvoorbeeld bij  $Y_2=$  het functievoorschrift in. Let erop dat je ook de grenzen van  $x$  aangeeft.

Vooraleer de figuur te tekenen kan je best de vensterinstellingen aanpassen. Druk op  $\boxed{WINDOW}$  en vul de waarden in die je hier ziet. Druk daarna op  $\boxed{TRACE}$ .



Werk verder zoals in opdracht 12. Bekijk de gearceerde oppervlakte onder de curve en vul de volgende kansuitspraak aan (rond af op 2 cijfers na de komma):

$$P(0 \leq X \leq 0.45) = \dots\dots\dots$$

**5.4. Afspraak woordgebruik**

Bij continue kansmodellen heb je oneindig veel mogelijke uitkomsten. Als je dan een vraag stelt over 9 % van die uitkomsten dan gaat het natuurlijk niet over een “aantal” van die uitkomsten want 9 % van oneindig veel blijft oneindig veel.

Wat je bedoelt is een interval waarin je met een bepaalde kans terechtkomt. Waar liggen bijvoorbeeld de 9 % kleinste uitkomsten bij het kansmodel  $X$  dat  $f(x) = 3x^2$ ,  $0 \leq x \leq 1$  als dichtheidsfunctie heeft? Om op deze vraag te antwoorden zoek je een interval waar de kleinere waarden terechtkomen. Je hebt zopas gevonden dat  $P(0 \leq X \leq 0.45) = 0.09$ . De 9 % kleinste uitkomsten liggen hier dus in  $[0 ; 0.45]$ .

## 6. De binomiale en de normale

### 6.1. Een beroemd discreet populatiemodel: de binomiale

Een volledige studie van het binomiale kansmodel is voor later. Hier bekijk je een voorbeeld. Wat je geleerd hebt over discrete kansmodellen kan je al gebruiken.

Denk aan een vaas met 3 kaartjes waarop een 1 staat en 2 kaartjes met een 0. Jij trekt lukraak een kaartje uit die vaas. Als je een 1 trekt, noem je dat “succes”, een 0 is “mislukking”. Je legt dat kaartje terug in de vaas en dan trek je opnieuw lukraak. Dat herhaal je 4 keer. Dan komt de vraag: hoeveel successen heb je gezien bij die 4 herhalingen?

De voorwaarden voor een binomiale kansverdeling zijn als volgt:

- je legt vooraf vast hoeveel keren (genoteerd als  $n$ ) je een experiment zal herhalen met maar 2 mogelijke uitkomsten: succes en mislukking
- de kans op succes is bij elke herhaling dezelfde (in dit voorbeeld is de succeskans  $3/5 = 0.6$ )
- het resultaat bij een volgende trekking hangt niet af van de vorige (onafhankelijk)
- de vraag is: “hoeveel successen had je bij  $n$  dergelijke herhalingen”?

Bij een binomiaal kansmodel is  $X =$  aantal successen bij  $n$  herhalingen.

Wat is de kansverdeling van  $X$  (= aantal successen) bij het trekken van 4 kaartjes uit die vaas? Daarvoor moet je alle mogelijke uitkomsten samen met hun kansen kennen.

De uitkomsten zijn “het aantal successen”. Als je brute pech hebt, dan heb je bij die 4 trekkingen geen enkele keer een 1 en is het totaal aantal successen gelijk aan 0. Supergelukkig ben je als je telkens een 1 trekt en 4 successen hebt. En alles daartussen kan natuurlijk ook.

De binomiale kansverdeling  $X$  ziet er in dit voorbeeld uit als:

$x$	0	1	2	3	4
$P(X=x)$					

waarbij je de kansen nog moet invullen. Doe dit vanuit volgende opdracht.

#### Opdracht 18

```

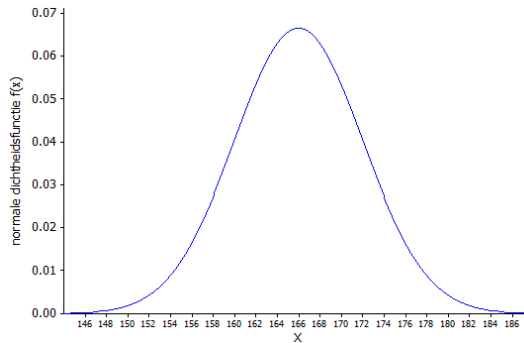
binompdf
trials:4
p:0.6
x Value:L1
Paste
    
```

- Zet de waarden { 0 1 2 3 4 } in [L1]. Druk [2nd][DISTR], loop naar A:binompdf( en druk [ENTER]. Vul in zoals aangegeven, loop naar Paste, druk [ENTER] en dan [STO] en [2nd][L2] en [ENTER]. De binomiale kansen staan nu in [L2]. Vul ze in.
- Wat is de kans om hier minstens 3 successen te hebben? Schrijf je antwoord als een kansuitspraak in formulevorm met de juiste notatie.

## 6.2. Een beroemd continu populatiemodel: de normale

Een volledige studie van het normale kansmodel komt aan bod in een volgende tekst. Hier bekijk je een voorbeeld. Wat je geleerd hebt over continue kansmodellen kan je hier al gebruiken.

In Vlaanderen zijn 17-jarige meisjes gemiddeld 166 cm groot. Lengte is een continue veranderlijke en als je een kansmodel wil opstellen voor de totale populatie (= de lengte van al die meisjes) dan moet je werken met een dichtheidsfunctie. Die ziet er zo uit:



De klokvormige curve is een normale dichtheidsfunctie. Je ziet dat de functie nergens negatief wordt en het is ook zo dat de totale oppervlakte onder de curve gelijk is aan 1. Dat betekent dat de getekende functie  $f(x)$  geschikt is om een continu kansmodel  $X$  te beschrijven. In dit voorbeeld is  $X$  het populatiemodel van de lengte van 17-jarige meisjes in Vlaanderen.

### Opdracht 19

Een continu kansmodel is volledig bepaald door zijn dichtheidsfunctie. Als je  $f(x)$  kent dan weet je voor elk deelinterval wat de kans is om daarin terecht te komen. Dat weet je van vroeger.

Deze eigenschap kan je nu toepassen op dat normale kansmodel  $X$ . Als je lukraak een meisje van 17 ontmoet, wat is dan de kans dat zij niet kleiner zal zijn dan 160 cm maar ook niet groter dan 172 cm? Of wat is voor dit kansmodel  $P(160 \leq X \leq 172)$  ?

Pas eerst de vensterinstellingen aan (druk **WINDOW**) zoals aangegeven. Druk dan **2nd**[DISTR] en loop met het pijltje **▸** naar DRAW. Druk nu **1:ShadeNorm(** , vul in zoals aangegeven, loop naar Draw en druk **ENTER**). Bekijk de figuur. Wat betekent de gearceerde oppervlakte? Vul nu de volgende kansuitspraak in en zeg in woorden wat die formule betekent.

```

WINDOW
Xmin=140
Xmax=190
Xscl=1
Ymin=-.02
Ymax=.08
Yscl=1
↓Xres=1

ShadeNorm
lower:160
upper:172
μ:166
σ:6
Draw
    
```

$P(160 \leq X \leq 172) = \dots\dots\dots$

In woorden:  $\dots\dots\dots$