

STOCHASTIEK MET DE TI-83

Guido Herweyers, KHBO Campus Oostende

1. Inleiding

Stochastiek is een vak dat vaak theoretisch wordt aangebracht met te veel aandacht voor het manuele rekenwerk. Bijgevolg is er te weinig tijd voor het verwerven van inzicht en praktijk in deze, voor de student vaak moeilijke, materie.

De Grafische rekenmachine TI-83 kan hier verandering in brengen:

- boxplots en histogrammen geven een beter globaal beeld van de data dan het klassieke rekenkundig gemiddelde en de standaardafwijking van niet-grafische rekenmachines,
- snelle simulatie van toevalsexperimenten verduidelijkt het kansbegrip en opent de weg naar theoretische verdelingen,
- om na te gaan of de data afkomstig zijn uit een normaal verdeelde populatie kunnen we ze ook uitzetten op normaal waarschijnlijkheidspapier,
- de klassieke discrete en continue verdelingen zijn, zowel grafisch als numeriek, ter beschikking. De statistische tabellen zijn voortaan overbodig zoals de logaritme-tabellen van destijds,
- bij het toetsen van hypothesen verschuift de aandacht van het significantieniveau α naar de meer informatieve overschrijdingskans of “P-waarde”, die bovendien grafisch kan worden voorgesteld als een gearceerde oppervlakte.

In onderstaande uiteenzetting beperk ik mij tot voorbeelden voor het secundair onderwijs, uitgewerkt met de veelzijdigheid van de TI-83.

Het is hier niet de bedoeling een handleiding te geven van de TI-83 op statistisch gebied.

Met de TI-83 hebben we een heus statistisch softwarepakket ter beschikking, waarmee we ook data kunnen downloaden van het Internet via graph-link. De statistische mogelijkheden gaan heel wat verder dan het programma voor het secundair onderwijs, zodat dit toestel ook verder kan gebruikt worden in het hoger onderwijs. Een overzicht van de statistieksyntax op het einde kan hier een idee van geven.

Er bestaat (vaak Engelstalige) goede literatuur en de ondersteuning op het Internet is zeer omvangrijk.

2. Beschrijvende statistiek

Een fabrikant van telecommunicatiemateriaal ontving klachten in verband met een zwak geluid bij telefoonverbindingen op lange afstand. Deze fabrikant had honderden elektronische versterkers geleverd die op regelmatige afstand langsheen de telefoonlijn werden gemonteerd. Als hoofdverdachte werd de versterkingsfactor van de versterkers aangeduid. In de verkoopsvoorwaarden werd een gemiddelde versterkingsfactor van 10 dB (decibel), een toelaatbare minimale versterkingsfactor van 7,75 dB en maximaal 12,25 dB vermeld. We noemen 10 dB de nominale waarde, de specificatiegrenzen zijn 7,75 en 12, 25 dB.

Om de klacht te onderzoeken werden er lukraak 120 versterkers verzameld die tot hetzelfde productielot of populatie behoorden als diegene waarover de klacht werd uitgebracht. Het resultaat van de 120 gemeten versterkingsfactoren vind je in onderstaande tabel

8,1	10,4	8,8	9,7	7,8	9,9	11,7	8,0	9,3	9,0
8,2	8,9	10,1	9,4	9,2	7,9	9,5	10,9	7,8	8,3
9,1	8,4	9,6	11,1	7,9	8,5	8,7	7,8	10,5	8,5
11,5	8,0	7,9	8,3	8,7	10,0	9,4	9,0	9,8	10,7
9,3	9,7	8,7	8,2	8,9	8,6	9,5	9,4	8,8	8,3
8,4	9,1	10,1	7,8	8,1	8,8	8,0	9,2	8,4	7,8
7,9	8,5	9,2	8,7	10,2	7,9	9,8	8,3	9,0	9,6
9,9	10,6	8,6	9,4	8,8	8,2	10,5	9,7	9,1	8,0
8,7	9,8	8,5	8,9	9,1	8,4	8,1	9,5	8,7	9,3
8,1	10,1	9,6	8,3	8,0	9,8	9,0	8,9	8,1	9,7
8,5	8,2	9,0	10,2	9,5	8,3	8,9	9,1	10,3	8,4
8,6	9,2	8,5	9,6	9,0	10,7	8,6	10,0	8,8	8,6

De data worden eerst ingevoerd in de lijst ELEKT, met 1-Var Stats LELEKT verkrijgen we dan de statistische kengetallen van de data:

LELEKT	L1	L2	1
8.1	-----	-----	
10.4			
8.8			
9.7			
7.8			
9.9			
11.7			
ELEKT = {8.1, 10.4, ...			

```

1-Var Stats
X̄=9.0325
ΣX=1083.9
ΣX²=9879.15
Sx=.8639524672
σx=.8603451342
↓n=120

```

```

1-Var Stats
↑n=120
minX=7.8
Q1=8.35
Med=8.9
Q3=9.6
maxX=11.7

```

De minimale waarde is groter dan 7,75 en het maximum is kleiner dan 12,25 zodat al de data binnen de vooropgestelde specificatiegrenzen vallen. Maar het gemiddelde 9,03 ligt duidelijk te laag. De mediaan is kleiner dan het gemiddelde, we verwachten dus een positief scheve verdeling met de langste staart naar rechts. Een boxplot en een histogram verduidelijken de situatie:

```

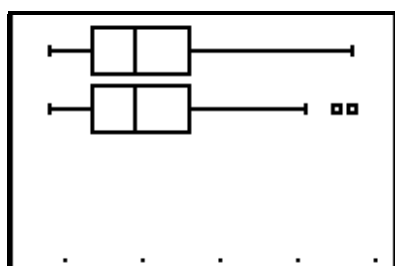
Plot2 Plot3
Off Off
Type: [ ] [ ] [ ]
Xlist: ELEKT
Freq: 1

```

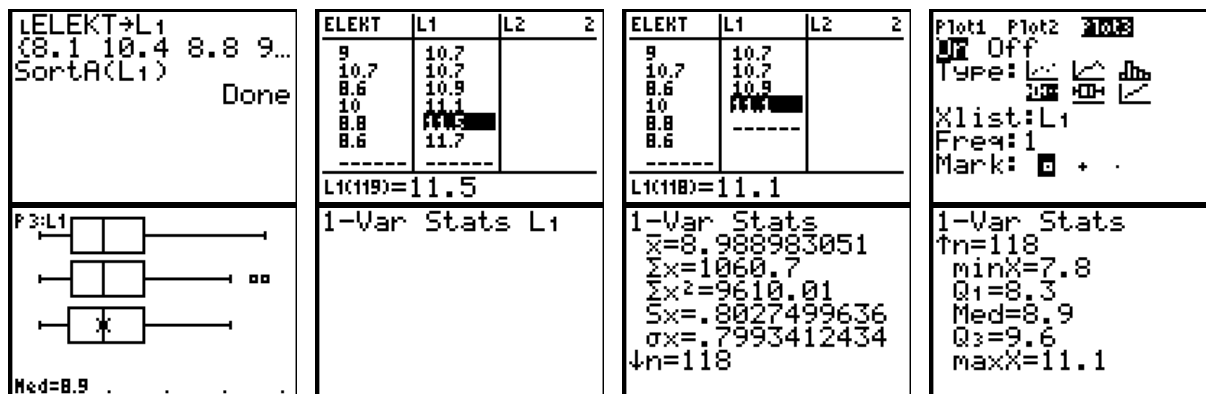
```

Plot1 Plot3
Off Off
Type: [ ] [ ] [ ]
Xlist: ELEKT
Freq: 1
Mark: [ ] + .

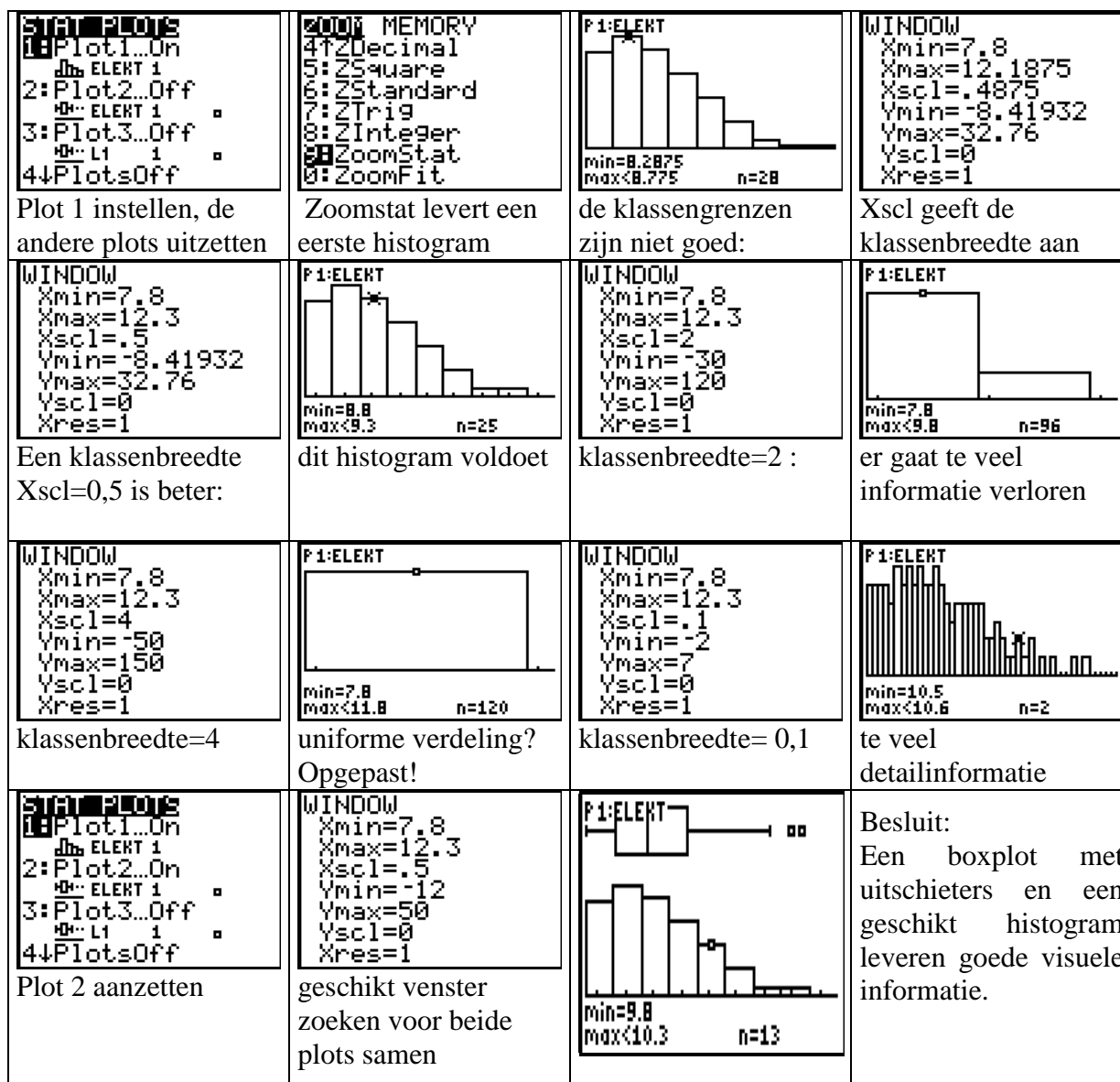
```



De gewone boxplot (Plot1, gebaseerd op de 5 kengetallen minX, Q1, Med, Q3, maxX) en de boxplot met uitschieters (Plot2) werden onder elkaar getekend. Traceren van de tweede plot levert 11,1 als grootste niet-uitschieter, de uitschieters zijn 11,5 en 11,7=maxX. De boxplot met uitschieters, gelegen beneden $Q1-1,5(Q3-Q1)$ of boven $Q3+1,5(Q3-Q1)$, levert een beter beeld van de scheefheid van de verdeling. Om het effect van de uitschieters na te gaan kopiëren we de lijst ELEKT naar de lijst L1, vervolgens rangschikken we L1 van klein naar groot. Na het verwijderen van de uitschieters uit L1 berekenen we hiervan de kengetallen en voegen we de boxplot met uitschieters (Plot3) toe aan bovenstaande boxplots:



We zien dat het minimum, eerste kwartiel, mediaan en tweede kwartiel ongewijzigd blijven bij weglaten van de uitschieters. Het rekenkundig gemiddelde echter daalt van 9,03 naar 8,99 en de steekproefstandaardafwijking Sx daalt van 0,86 naar 0,80 d.i. een daling van 7% ! In de volgende plaatjes maken we een histogram van de lijst ELEKT en bestuderen we de invloed van de klassenbreedte op het histogram:



De klacht was dus terecht: de verdeling is niet symmetrisch, er zijn te veel versterkers met een versterkingsfactor kleiner dan 10 dB. Merk op dat de klassieke steekproefstandaardafwijking S_x hier niet zo betekenisvol is: ze wordt sterk beïnvloed door de twee uitschieters en één getal kan onmogelijk de spreiding links en rechts t.o.v. het gemiddelde vertegenwoordigen bij een asymmetrische verdeling. Het rekenkundig gemiddelde en S_x blijven wel goede kengetallen voor symmetrische verdelingen, zeker bij normale verdelingen waar dan de 68-95-99,7 regel van toepassing is.

3. Simulatie

3.1 Het principe

Simulatie van toevalsexperimenten, ook Monte Carlo methode genoemd, is een krachtige techniek om kansen van gebeurtenissen of kansverdelingen van stochastische veranderlijken benaderend te bepalen. Het principe is eenvoudig: herhaal het toevalsexperiment een groot aantal keren, de relatieve frequentie van het optreden van een gebeurtenis benadert dan de kans op die gebeurtenis. Het toeval wordt gesimuleerd door een generator van toevalsgetallen:

<pre>MATH NUM CPX [233] 1:rand 2:nPr 3:nCr 4:! 5:randInt(6:randNorm(7:randBin(</pre>	<pre>rand .2783088265 .2752142947 .1217901355 .052589806 .7223793157 .0125655621</pre>	<pre>rand(5) (.9854088797 .6...</pre>	<pre>2→A:10→B:A+(B-A) rand 3.584736224 5.186910813 2.936257888 3.923146411 8.824758769</pre>
	de functie rand levert een lukraak getal uit het interval [0,1)	rand(5) levert een lijst van 5 dergelijke toevalsgetallen	A+(B-A).rand wordt een lukraak getal tussen A en B
<pre>randInt(1,6) 6 3 1 randInt(1,6,5) (3 6 4 4) (6 5 5 4 3)</pre>	<pre>randNorm(5,1) 3.819464731 6.153741637 5.125583362 randNorm(5,1,10) (5.083869612 5...</pre>	<pre>randBin(10,.5) 7 6 3 randBin(10,.5,7) (5 3 5 6 6 6 7)</pre>	<pre>randInt(1,6,4)+r randInt(1,6,4) (9 4 8 7) seq(randInt(1,6) +randInt(1,6),X, 1,4) (6 8 7 7)</pre>
een geheel getal tussen 1 en 6 of dergelijke lijst	getallen uit een normale populatie met $\mu=5$ en $\sigma=1$	aantal keer munt bij 10 keer werpen van een muntstuk	werp 2 dobbelstenen en noteer de som van het aantal ogen, doe dit 4 keer

De rekenmachine genereert een reeks toevalsgetallen uitgaande van een bepaalde startwaarde of “zaadje”, dergelijke startwaarde legt de reeks volledig vast. Als startwaarde kies je een natuurlijk getal dat je opslaat in **rand**, dit is erg belangrijk wanneer je een ganse klas laat simuleren; zorg ervoor dat elke student een andere startwaarde neemt indien je nadien de individuele resultaten wenst te groeperen.

3.2 Het verjaardagprobleem

In een klas zitten 15 leerlingen (geen tweelingen). Bepaal de kans dat er minstens 2 van de 15 leerlingen dezelfde verjaardag hebben.

We veronderstellen dat een jaar bestaat uit 365 dagen, die als verjaardagen allemaal even waarschijnlijk zijn.

<pre>randInt(1,365,15) (47 127 159 155...</pre>	<pre>randInt(1,365,15))→L1:SortA(L1):L 1 (1 5 18 44 50 1...</pre>	<p>Je kan de gesorteerde lijst L1 nu manueel overlopen om na te gaan of er twee gelijke getallen in zijn. Via de lijst der voorwaartse differenties van L1 volstaat het echter om nullen op te sporen:</p>																																																																								
<p>simulatie van de verjaardagen van 15 personen</p>	<p>sorteren van 15 verjaardagen in lijst L1, in stijgende volgorde</p>																																																																									
<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> <th>Z</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <td>5</td> <td></td> <td></td> <td></td> </tr> <tr> <td>18</td> <td></td> <td></td> <td></td> </tr> <tr> <td>44</td> <td></td> <td></td> <td></td> </tr> <tr> <td>50</td> <td></td> <td></td> <td></td> </tr> <tr> <td>101</td> <td></td> <td></td> <td></td> </tr> <tr> <td>102</td> <td></td> <td></td> <td></td> </tr> <tr> <td colspan="4">L2 = ΔList(L1)</td> </tr> </tbody> </table>	L1	L2	L3	Z	1				5				18				44				50				101				102				L2 = ΔList(L1)				<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> <th>Z</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>5</td> <td>13</td> <td></td> <td></td> </tr> <tr> <td>18</td> <td>26</td> <td></td> <td></td> </tr> <tr> <td>44</td> <td>6</td> <td></td> <td></td> </tr> <tr> <td>50</td> <td>51</td> <td></td> <td></td> </tr> <tr> <td>101</td> <td>1</td> <td></td> <td></td> </tr> <tr> <td>102</td> <td>5</td> <td></td> <td></td> </tr> <tr> <td colspan="4">L2(1)=4</td> </tr> </tbody> </table>	L1	L2	L3	Z	1	0			5	13			18	26			44	6			50	51			101	1			102	5			L2(1)=4				<pre>L2=0 (0 0 0 0 0 0 0 ... sum(L2=0) 0</pre>
L1	L2	L3	Z																																																																							
1																																																																										
5																																																																										
18																																																																										
44																																																																										
50																																																																										
101																																																																										
102																																																																										
L2 = ΔList(L1)																																																																										
L1	L2	L3	Z																																																																							
1	0																																																																									
5	13																																																																									
18	26																																																																									
44	6																																																																									
50	51																																																																									
101	1																																																																									
102	5																																																																									
L2(1)=4																																																																										
	<p>$L2(i) = L1(i+1) - L1(i)$</p>	<p>geen 2 dezelfde verjaardagen!</p>																																																																								

Het laatste plaatje vraagt om een toelichting: beschouw bijvoorbeeld de uitdrukking $\{1,5,7,12,5\}=5$. Dit is hetzelfde als de lijst van gelijkheden $\{1=5 \ 5=5 \ 7=5 \ 12=5 \ 5=5\}$. De TI-83 werkt dit uit tot de lijst $\{0,1,0,0,1\}$, hierbij staat **0** voor onwaar en **1** voor waar. Nu levert $\text{sum}(\{0,1,0,0,1\})$ als resultaat 2, dit is de som van de elementen uit de lijst $\{0,1,0,0,1\}$, of meteen het aantal elementen uit de oorspronkelijke lijst $\{1,5,7,12,5\}$ die gelijk zijn aan 5.

<pre>randInt(1,365,15))→L1:SortA(L1):s um(ΔList(L1)=0)</pre>	<pre>0 0 0 1 0</pre>	<pre>0 0 0 1 0 1 0</pre>
---	----------------------	--------------------------

In het totaal simuleerden we 18 bijeenkomsten van 15 lukraak gekozen personen, 4 keer waren er minstens 2 personen met dezelfde verjaardag. Als benadering voor de gevraagde kans vinden we dus de relatieve frequentie $4/18=0,22$. Op 2 decimalen nauwkeurig is de exacte kans 0,25 (zie verder). Het is niet moeilijk om een programma te schrijven dat als input het aantal simulaties N vraagt en als output de relatieve frequentie RELFREQ geeft:

<pre>prgmVERJAAR N=10 RELRFREQ= .2 Done</pre>	<pre>prgmVERJAAR N=20 RELRFREQ= .25 Done</pre>	<pre>prgmVERJAAR N=100 RELRFREQ= .32 Done</pre>
---	--	---

Een groter aantal simulaties impliceert niet noodzakelijk een betere kansbenadering

Ziehier het verjaardagprogramma:

<pre>PROGRAM:VERJAAR :Input "N=",N :0→S :For(X,1,N) :randInt(1,365,1 5)→L1 :SortA(L1) :If sum(ΔList(L1</pre>	<pre>PROGRAM:VERJAAR :If sum(ΔList(L1)=0)≠0 :S+1→S :End :Disp "RELFREQ=" ,S/N :</pre>	<p>Oefening: pas het programma aan zodat je na elke simulatie van een bijeenkomst de ogenblikkelijke relatieve frequentie laat zien, dus niet alleen op het einde.</p>
--	--	--

Voor dit probleem kunnen we ook de exacte kans berekenen:

```
1-(365 nPr 15)/3
65^15
.2529013198
```

Hier viel het nog mee om de exacte kans te berekenen. Dit is echter niet altijd zo, zodat simulatie zich dan zeker opdringt. Zie volgend voorbeeld.

3.3 Het probleem van de garderobejuffrouw

6 heren geven hun hoeden af bij de garderobe. De garderobejuffrouw geeft de hoeden later lukraak terug. Hoe groot is de kans dat er minstens één heer zijn eigen hoed krijgt?

Hetzelfde probleem vinden we terug bij volgende opgave:

Een secretaris typt 6 brieven en adresseert 6 briefomslagen. Dan steekt hij de brieven lukraak in de briefomslagen. Hoe groot is de kans dat er minstens één brief in de juiste envelop komt?

Het probleem is nu de opgave te vertalen of te coderen in een wiskundig model dat zich leent tot simulatie.

Nummer de heren (brieven) van 1 tot 6, nummer de hoeden (briefomslagen) ook van 1 tot 6 en zorg ervoor dat heer (brief) met nummer x hoort bij hoed (briefomslag) met nummer x .

Een lukrake keuze van de garderobejuffrouw levert een bijectie van $\{1,2,3,4,5,6\}$ (de heren) naar $\{1,2,3,4,5,6\}$ (de hoeden).

Minstens één heer krijgt zijn eigen hoed wanneer de gekozen permutatie van $\{1,2,3,4,5,6\}$ minstens één “vast punt” heeft, bijvoorbeeld 4 gaat naar 4. Zo een permutatie kunnen we voorstellen door een rijtje van beelden, horende bij argumenten in natuurlijke volgorde.

Zo levert de permutatie $(2,4,3,1,6,5)$ één vast punt op: 1 gaat naar 2, 2 naar 4, 3 naar 3, 4 naar 1, 5 naar 6 en 6 naar 5; het element 3 van de rij $(2,4,3,1,6,5)$ staat op zijn natuurlijke plaats.

Om dit nu te simuleren volstaat het lukraak 6 (verschillende) getallen tussen 0 en 1 te genereren en te onderzoeken of er getallen bij zijn die op hun natuurlijke plaats staan:

Zo levert de rij $(0,87 ; 0,45 ; 0,92 ; 0,68 ; 0,38 ; 0,15)$ of de permutatie $(5, 3, 6, 4, 2, 1)$ één vast punt.

```
rand(6)→L1:L1→L2
:SortA(L1):sum(L
2-L1=0)
```

```
000000000112
```

```
00001110142
```

Als relatieve frequentie vinden we $13/18 = 0,72$
De exacte kans is gelijk aan (moeilijke berekening!):

$$1 - \left(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} \right) = 0,6319$$

De veralgemening voor n heren met n hoeden ligt voor de hand; we krijgen telkens een benadering van $1-1/e = 0,6321$. De gevraagde kans is dus steeds ongeveer $2/3$, onafhankelijk van n (met $n > 2$).

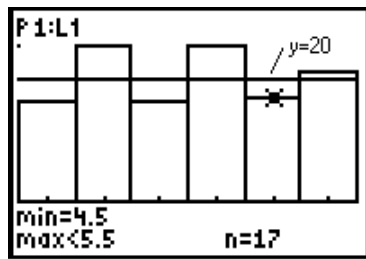
Met bovenstaande simulatie kunnen we echter meer: benader de kans dat er juist 2 heren hun eigen hoed krijgen, enz.

3.4 Kansverdeling van een stochastische veranderlijke

a) Werp lukraak 120 keer met een dobbelsteen. Welke frequentieverdeling (of kansverdeling van de stochast X = aantal ogen bij werpen van een dobbelsteen) verwacht je?

```
randInt(1,6,120)
→L1
{6 3 5 4 4 1 1 ...
```

```
WINDOW
Xmin=.5
Xmax=6.5
Xscl=1
Ymin=-7
Ymax=30
Yscl=5
Xres=1
```



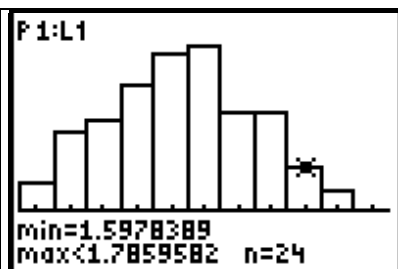
b) Kies lukrake getallen X en Y uit het interval $[0,1]$. Bepaal de kansverdeling van de stochastische veranderlijke $Z=X+Y$. Voor een continue stochastische veranderlijke komt dit neer op het bepalen van de dichtheidsfunctie van Z :

```
rand(500)+rand(500)→L1:L1
{.8234594431 .5...
```

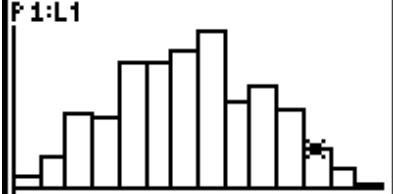
500 keer simuleren

```
STAT PLOTS
1:Plot1...On
  ▾ L1 1
2:Plot2...Off
  * ELERT 1
3:Plot3...Off
  * L1 1
4↓PlotsOff
```

Plot 1 definiëren, andere plots uitzetten



histogram met ZoomStat

<pre>WINDOW Xmin=0 Xmax=2.1 Xscl=.15 Ymin=-27 Ymax=90 Yscl=0 Xres=1</pre>	 <pre>F1:L1 min=1.65 max<1.8 n=18</pre>	<pre>mean(L1) .9941865843 stdDev(L1) .4057412964 √(1/6) .4082482905</pre>
<p>eenvoudiger klassengrenzen en klassenbreedte kiezen</p>	<p>als wiskundig model voldoet een gelijkbenige driehoek</p>	<p>$E(Z) = E(X) + E(Y) = 2 \cdot 0,5 = 1$ $Var(Z) = Var(X) + Var(Y) = 2 \cdot (1/12) = 1/6$</p>

Merk op dat een histogram van de 500 sommen niet alleen een beeld geeft van de frequentieverdeling maar ook van de kansverdeling, via aanpassing van de schaal op de verticale as.

Het gemiddelde van de 500 sommen is een goede schatting voor het gemiddelde van de stochast Z of $E(Z)$, terwijl de steekproefstandaardafwijking een schatting is voor de standaardafwijking van Z of $\sigma_Z = \sqrt{Var(Z)}$.

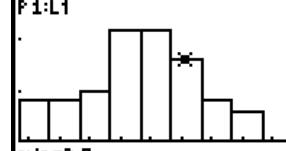
3.5 Verdeling van het steekproefgemiddelde.

Werp 9 dobbelstenen gelijktijdig en bepaal het gemiddeld aantal ogen \bar{X}

Bepaal de verdeling van \bar{X} .

Hiertoe gebruiken we het volgende programma:

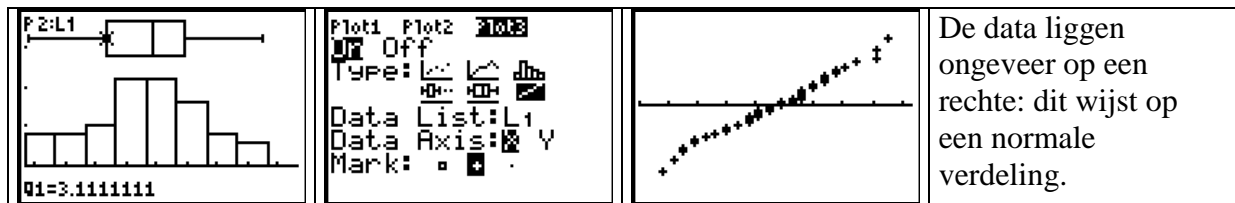
```
PROGRAM: NORMAAL
: Input " GROOTTE STEEKPR=", N
: Disp " AANTAL"
: Input "HERHALINGEN=", K
: ClrList L1
: ClrHome
: For(I,1,K)
: mean(randInt(1,6,N))→L1(I)
: End
: 1-Var Stats L1
: Disp "GEMID=",  $\bar{x}$ 
: Disp "STANDAFW=",  $S_x$ 
```

<pre>PrgmNORMAAL GROOTTE STEEKPR= 9 AANTAL HERHALINGEN=50</pre>	<pre>GEMID= 3.508888889 STANDAFW= .566152895 Done</pre>	<pre>WINDOW Xmin=2 Xmax=5.5 Xscl=.25 Ymin=-7 Ymax=25 Yscl=5 Xres=1</pre>	 <pre>F1:L1 min=3.8 max<4.1 n=50</pre>
---	---	--	---

Het gemiddelde van de 50 gemiddelden is een goede schatting voor $E(\bar{X}) = 3,5$

De steekproefstandaardafwijking van de 50 gemiddelden is 0,566, een schatting voor de theoretische standaardafwijking van \bar{X} . Inderdaad: $Var(\bar{X}) = Var(X)/9 = (35/12)/9 = 0,324$, de vierkantswortel hieruit levert als theoretische standaardafwijking 0,569.

Kunnen we een normale verdeling gebruiken als model voor de 50 gemiddelden?
 Het histogram en de boxplot wijzen al in die richting, uitzetten van de data op normaal waarschijnlijkheidspapier levert een nieuwe controle die ook bruikbaar is voor minder data:



Dit was meteen ook een illustratie van de centrale limietstelling.

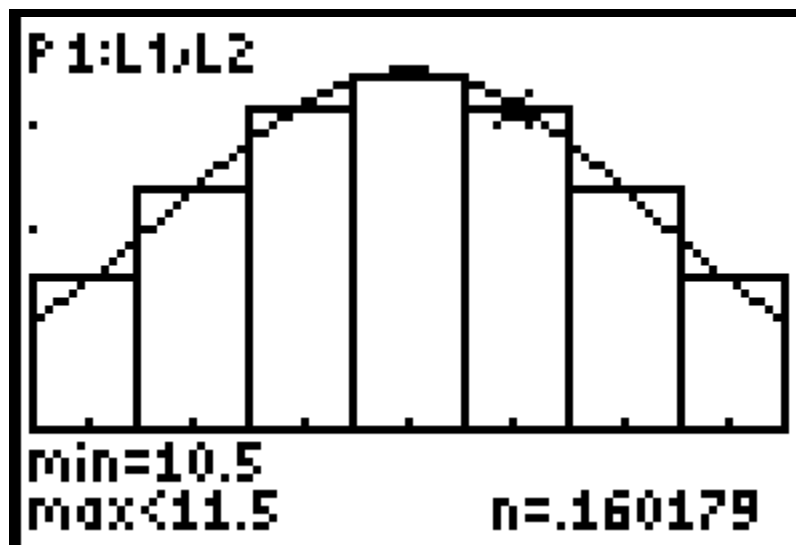
4. De binomiale en de normale verdeling

4.1 De continuïteitscorrectie

Een Laplace-muntstuk (kans op kop = kans op munt) wordt 20 keer geworpen. Bij 9, 10 of 11 keer munt wint men een prijs. Bereken de kans dat men een prijs wint,

- exact met de binomiale verdeling
- benaderend met de normale verdeling

<pre> binompdf(20,0.5, 9)+binompdf(20,0. .5,10)+binompdf(20,0.5,11) .4965553284 </pre>	<pre> binomcdf(20,0.5, 11)-binomcdf(20, 0.5,8) .4965553281 </pre>	<pre> normalcdf(8.5,11 .5,10,√(5)) .497665188 </pre>	<p>Definieer nu L1=seq(X,X,1,20) L2= binompdf(20,0.5,L1)</p>																																
<p>gebruik de binomiale kansfunctie</p>	<p>of de binomiale verdelingsfunctie</p>	<p>volgorde: grenzen,gemid,stdafw</p>																																	
<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> <th>Z</th> </tr> </thead> <tbody> <tr><td>6</td><td>.03696</td><td></td><td></td></tr> <tr><td>7</td><td>.07393</td><td></td><td></td></tr> <tr><td>8</td><td>.12013</td><td></td><td></td></tr> <tr><td>9</td><td>.16018</td><td></td><td></td></tr> <tr><td>10</td><td>.1762</td><td></td><td></td></tr> <tr><td>11</td><td>.16018</td><td></td><td></td></tr> <tr><td>12</td><td>.12013</td><td></td><td></td></tr> </tbody> </table> <p>L2(10) = .160179138...</p>	L1	L2	L3	Z	6	.03696			7	.07393			8	.12013			9	.16018			10	.1762			11	.16018			12	.12013			<pre> Plot1 Plot2 Plot3 Off Off Off Type: L1 L2 L3 Xlist: L1 Freq: L2 </pre>	<pre> Plot1 Plot2 Plot3 \Y1=normalpdf(X, 10,√(5)) \Y2= \Y3= \Y4= \Y5= \Y6= </pre>	<p>WINDOW</p> <p>Xmin=6.5 Xmax=13.5 Xscl=1 Ymin=0 Ymax=.2 Yscl=.05 Xres=1</p>
L1	L2	L3	Z																																
6	.03696																																		
7	.07393																																		
8	.12013																																		
9	.16018																																		
10	.1762																																		
11	.16018																																		
12	.12013																																		

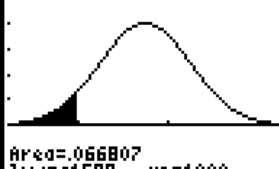


Verklaring van de continuïteitscorrectie

4.2 Automatische vulmachine

Een machine vult pakken met suiker. De massa X van die pakken is normaal verdeeld met gemiddelde $\mu = 1015$ g en standaardafwijking $\sigma = 10$ g.

- Welk percentage van de pakken bevat minder dan 1000 g?
- Stel dat het mogelijk is om de afstelling van het vulapparaat (d.w.z. μ) te veranderen zonder dat σ verandert. Hoe moet μ dan gekozen worden opdat slechts 1% van de pakken een massa zou hebben kleiner dan 1000 g?

<pre>WINDOW Xmin=985 Xmax=1045 Xscl=20 Ymin=-.02 Ymax=.05 Yscl=.01 Xres=1</pre>	<pre>DISTR 0:Z10 1:ShadeNorm(2:Shade_t(3:ShadeX²(4:ShadeF(</pre>	<pre>ShadeNorm(-10^99 ,1000,1015,10)</pre>	
<p>eerst nadenken over het venster</p>	<p>met ENTER wordt de grafiek getekend</p>		<p>kans= 6,68%</p>
<pre>DISTR DRAW 1:normalpdf(2:normalcdf(3:invNorm(4:tpdf(5:tcdf(6:X²pdf(7:X²cdf(</pre>	<pre>normalcdf(-10^99 ,1000,1015,10) .0668072287</pre>		
<p>enkel numeriek</p>	<p>de transformatie naar de standaardnormale verdeling is overbodig</p>		
<pre>DISTR DRAW 1:normalpdf(2:normalcdf(3:invNorm(4:tpdf(5:tcdf(6:X²pdf(7:X²cdf(</pre>	<pre>invNorm(0.01,0,1) -2.326347877 1000-Ans*10 1023.263479</pre>	<pre>normalcdf(-10^99 ,1000,1023.3,10) .009903053</pre>	
<p>nu opgave (b)</p>	<p>$\mu = 1023,3$ g</p>	<p>controle</p>	

Opgave (b) illustreert dat de transformatie naar de standaardnormale verdeling soms noodzakelijk blijft om een probleem op te lossen.

4.3 Een passingsprobleem.

De diameter X van de naald van een diskdrive is normaal verdeeld met gemiddelde 0,525 en standaardafwijking 0,0003. De diameter Y van de kraag er omheen is normaal verdeeld met gemiddelde 0,526 en standaardafwijking 0,0004.

Bereken de kans dat de naald niet in de kraag past.

De essentie van de oplossing is het vinden van de verdeling van de stochast $V = Y - X$.

De gevraagde kans is immers $P(Y \leq X)$ of $P(V \leq 0)$. Ga na dat V normaal verdeeld is met gemiddelde 0,001 en standaardafwijking 0,0005. We vinden $P(V \leq 0) = 2,3\%$

```
normalcdf(-10^99
,0,0.001,0.0005)
.022750062
```

5. Toetsen van hypothesen

5.1 Toetsen met de normale verdeling

De TI-83 genereert toevalsgetallen uit een uniforme verdeling op het interval [0,1) met gemiddelde 0,5 en standaardafwijking $\sqrt{(1/12)} = 0,2887$

Om dit na te gaan genereren we 100 dergelijke getallen en bepalen we hun gemiddelde:

```
rand(100)→L1
(.5393177412 .0...
mean(L1)
.5567973908
```

We krijgen 0,5568 als gemiddelde. Is dit verdacht?

Veronderstel dat de getallen inderdaad afkomstig zijn uit een populatie met $\sigma = 0,2887$.

Als nulhypothese nemen we $\mu=0,5$

De alternatieve hypothese is $\mu \neq 0,5$

Als toetsingsgrootte nemen we het gemiddelde \bar{X} ; onder de nulhypothese is deze normaal verdeeld met gemiddelde 0,5 en standaardafwijking 0,02887. De tweezijdige test levert dan:

```
2*normalcdf(0.55
68.10^99,0.5,0.0
2887)
.0491323378
```

De overschrijdingskans of P-waarde is $P = 4,9\%$, dit is significant op het 5%-niveau zodat je geneigd bent om de nulhypothese te verwerpen. P is echter niet erg klein, daarom nemen we nog eens een steekproef.

```
(.7684378848 .9...
mean(L1)
.5158715641
2*normalcdf(0.51
59.10^99,0.5,0.0
2887)
.5818086205
```

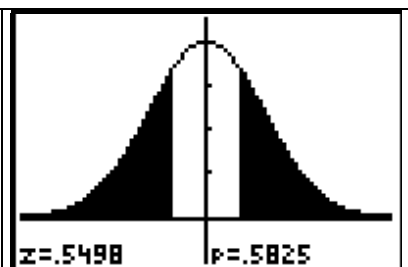
Als overschrijdingskans vinden we 58,2%, dit is helemaal niet uitzonderlijk. We verwerpen de nulhypothese dus niet.

```
EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
```

Hier toetsen we met $Z = (\bar{X} - 0,5)/0,02887$

```
Z-Test
Inpt: Data Stats
μ₀: .5
σ: .2887
List: L1
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw
```

de optie Draw kiest zelf een geschikt venster



$z = (0,515872 - 0,5)/0,02887$ verklaar het kleine verschil in de P-waarde

5.2 Toetsen met de binomiale verdeling

Je geeft een meerkeuzetoets van 10 meerkeuzevragen. Elke vraag bestaat uit 3 alternatieven. De score op elke vraag is 1 (juist) of 0 (verkeerd). Een student haalt 8/10 op die toets en beweert gegokt te hebben bij elke vraag. Geloof je hem?

Nulhypothese: kans op succes $p = 1/3$

Alternatieve hypothese: $p > 1/3$

Toetsingsgrootheid: De behaalde score X

Onder de nulhypothese geldt dat X binomiaal verdeeld is met parameters $n=10$ en $p = 1/3$.

De eenzijdige overschrijdingskans is

$$P(X=8) + P(X=9) + P(X=10) = P(X \leq 10) - P(X \leq 7) = 0,0034$$

Dit is zeer klein en een sterk bewijs tegen de nulhypothese; we besluiten dat de student niet kan gegokt hebben.

<pre>binompdf(10,1/3, 8)+binompdf(10,1 /3,9)+binompdf(1 0,1/3,10) .0034039526</pre>	<pre>sum(seq(binompdf (10,1/3,X),X,8,1 0) .0034039526</pre>	<pre>binomcdf(10,1/3, 10)-binomcdf(10, 1/3,7) .0034039526</pre>
---	---	---

Andere aanpak: simuleer het lukraak invullen van een toets 100 keer en observeer hoe vaak je een 8, 9 of 10 haalt. De relatieve frequentie is een schatting voor de P-waarde.

Kunnen we hier de test ook laten uitvoeren door de TI-83?

<pre>EDIT CALC TESTS 1:Z-Test... 2:T-Test... 3:2-SampZTest... 4:2-SampTTest... 5:1-PropZTest... 6:2-PropZTest... 7:ZInterval...</pre>	<pre>1-PropZTest P0: .3333333333... x:8 n:10 PROP>P0 <P0 >P0 Calculate Draw</pre>	<pre>1-PropZTest PROP>.33333 z=3.130495168 P=8.7262644E-4 P=.8 n=10</pre>
---	---	--

De P-waarde is hier slechts $1/4$ van de correcte P-waarde, waar ligt de oorzaak?

Deze laatste binomiaaltoets werkt als volgt:

- De stochast $X/10$ wordt benaderd door een normale verdeling, *wat hier niet toegestaan is aangezien* $10 \cdot (1/3) < 5$, deze wordt dan herleid tot de standaardnormale verdeling Z die gebruikt wordt als toetsingsgrootheid: $Z = (X/10 - 1/3) / \sqrt{(1/3) \cdot (2/3) / 10}$, dit levert hier $z = (0,8 - 1/3) / \sqrt{(1/45)} = 3,13$
- Er wordt geen continuïteitscorrectie toegepast (wel niet zo erg in dit geval)

6. De TI-83 Statistieksyntax

```
MATH NUM CPX 1234
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(
```

- 1: **rand**[(aantal keer)]
- 5: **randInt**(ondergrens, bovengrens[,aantal keer])
- 6: **randNorm**(μ , σ [, aantal keer])
- 7: **randBin**(n , p [,aantal keer])

```
0:1/2 DRAW
1:normalpdf(
2:normalcdf(
3:invNorm(
4:tpdf(
5:tcdf(
6: $\chi^2$ pdf(
7: $\chi^2$ cdf(
```

```
0:1/2 DRAW
8:Fpdf(
9:Fcdf(
0:binompdf(
A:binomcdf(
B:poissonpdf(
C:poissoncdf(
D:geometpdf(
```

```
0:1/2 DRAW
9:Fcdf(
0:binompdf(
A:binomcdf(
B:poissonpdf(
C:poissoncdf(
D:geometpdf(
E:geometcdf(
```

- 1: **normalpdf**(x, [, μ , σ) (standaard is $\mu=0$ en $\sigma=1$)
 - 2: **normalcdf**(ondergrens, bovengrens[, μ , σ])
 - 3: **invNorm**(oppervlakte [, μ , σ])
 - 4: **tpdf** (x, df)
 - 5: **tcdf**(ondergrens, bovengrens, df)
 - 6: **χ^2 pdf**(x, df)
 - 7: **χ^2 cdf** (ondergrens, bovengrens, df)
 - 8: **Fpdf**(x, df teller, df noemer)
 - 9: **Fcdf**(ondergrens, bovengrens, df teller, df noemer)
 - 0: **binompdf**(n, p [, x]) binompdf(n, p) genereert de kansverdeling in een lijst
 - A: **binomcdf**(n, p [,x])
 - B: **poissonpdf**(λ , x)
 - C: **poissoncdf**(λ , x)
 - D: **geometpdf**(p, x)
 - E: **geometcdf**(p, x)
- pdf: probability density function
c: cumulative
df: degrees of freedom

```
DISTR 0345
1:ShadeNorm(
2:Shade_t(
3:Shade $\chi^2$ (
4:ShadeF(
```

- 1: **ShadeNorm**(ondergrens, bovengrens [, μ , σ])
- 2: **Shade_t**(ondergrens, bovengrens, df)
- 3: **Shade χ^2** (ondergrens, bovengrens, df)
- 4: **ShadeF**(ondergrens, bovengrens, df teller, df noemer)

*Vuistregel syntax: bij continue verdelingen: parameter(s) laatst
bij discrete verdelingen: parameter(s) eerst*

7. Slotbeschouwing

In bovenstaande tekst heb je veel plaatjes gezien, zij brengen de stochastiek tot leven. Door de snelheid van de rekenmachine ontstaat er ruimte voor meer oefeningen, parameteronderzoek en simulaties, die uiteindelijk hopelijk tot meer inzicht zullen leiden. Weg met de statistische tabellen, de transformatie naar de standaardnormale verdeling blijft echter wel belangrijk (standaardscore, betrouwbaarheidsintervallen, toetsen van hypothesen: de Z-test heeft als grafisch voordeel dat het venster vastligt en niet telkens manueel moet bepaald worden ,).

Breng echter variatie in het toetsen van hypothesen door eens een andere toetsingsgrootheid te kiezen dan bij een ingebouwde test, ook dit leidt tot inzicht.

Als je een TI-83 wil gebruiken voor je statistieklessen, dan moet elke student beschikken over dat toestel, dit vermijdt frustraties zowel bij de student als de docent.

Studenten zijn vlot in de omgang met de hedendaagse technologie; maak er gebruik van om de statistiek en de wiskunde in het algemeen een nieuwe elan te geven.

Ook de brug naar de fysica, mechanica, economie, chemie, elektriciteitsleer, ... wordt gelegd; als voorbeeld bij uitstek hebben we nu de mogelijkheid om een goed wiskundig model te zoeken voor de beschrijving van experimentele data (curve fitting).

Dit alles werkt erg motiverend voor student en docent.

De eigenaars van een TI-92 of een TI89 moeten niet treuren met hun minder krachtige statistiek mogelijkheden: statistieksoftware met dezelfde mogelijkheden als de TI-83 kan je downloaden via het Internet: <ftp://ftp.ti.com/pub/graph-ti/calc-apps/92/math/stat/infstat.92g>, voor de handleiding neem je infstat.txt, voor de TI-89 of TI-92 Plus vervang je 92 door 89 in het adres.

8. Literatuur

- [1] G. Barret, *Statistics with the TI-83*, 1997, Meridian Creative Group, USA.
- [2] D. Yates, D. Moore, G. McCabe, *The Practice of Statistics, TI-83 Graphing Calculator Enhanced*, 1999, W.H. Freeman and Company, New York .
- [3] P. Drijvers, M. Doorman, *De TI-83 kennismaken en toepassen*, 1997, Wolters-Noordhoff, Groningen.
- [4] J. Breeman, P. Drijvers, W. Hoekstra, L. van Gastel, P. Schadron, G. Stroomer, *Masterclass Grafische Rekenmachine*, T³-Europe, APS, Utrecht.