Hoofdstuk 6

Kansverdelingen

6.1 Discrete stochasten

6.1.1 De Bernoulli verdeling

Een Bernoulli experiment is een experiment met slechts twee mogelijke uitkomsten, die we "succes" (S) en "mislukking" (M) noemen.

Stellen we $P(S) = p$ en $P(M) = q$, dan geldt steeds dat $p + q = 1$.

Indien we succes coderen met 1 en mislukking met 0, geeft dit een stochast X met als kansverdeling :

x_i	0	1
$P(X = x_i)$	q	p

We zeggen dat X Bernoulli verdeeld is met parameter p.

We noteren : $X \sim \text{Be}(p)$. Toon aan dat $E(X) = p$ en $Var(X) = pq$

6.1.2 De binomiale verdeling

In paragraaf (3.2) en (4.2) bestudeerden we de stochast X = "aantal keer kop bij twee keer werpen van een muntstuk". Het experiment bestaat uit twee onafhankelijke herhalingen van het Bernoulli experiment "werp een muntstuk", waarbij X het totaal aantal successen telt (met succes = kop). De stochast X noemen we binomiaal verdeeld.

Beschouw n *onafhankelijke* herhalingen van een Bernoulli experiment, d.w.z. dat de kans op succes p telkens ongewijzigd blijft. Stel X het totaal aantal successen. We zeggen dat X binomiaal verdeeld is met parameters n en p.

We noteren : $X \sim B(n, p)$.

Om de kansverdeling van X te berekenen, merken we eerst op dat X de waarden $0, 1, 2, \ldots, n$ kan aannemen. De gebeurtenis $X = x$ is de verzameling van alle rijen bestaande uit x keer S en $(n-x)$ keer M. Elk van deze rijen heeft kans $p^x \cdot q^{n-x}$ omwille van de onafhankelijkheid en de productregel van kansen.

Er zijn $\binom{n}{x}$ dergelijke rijen. Voor $x \in \{0,1,2,...,n\}$ geldt :

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} .$$

Dit levert inderdaad een kansverdeling want $P(X = x) \geq 0$ en omwille van het binomium van Newton geldt : $\sum_{x=0}^{n} \binom{n}{x} \cdot p^x \cdot q^{n-x} = (p+q)^n = 1$.

Voorbeeld : Werp 20 keer een dobbelsteen. De kans dat hierbij

- juist 4 keer een zes geworpen wordt, is : $\binom{20}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{16} = 0.202$,
- juist 8 keer "zes of drie" : $\binom{20}{8}\left(\frac{1}{3}\right)^8\left(\frac{2}{3}\right)^{12} = 0.148$
- en hoogstens 4 keer een zes : $P(X \leq 4) = \sum_{x=0}^{4} \binom{20}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{20-x} = 0.769$

Je kan deze kansen met de **TI-83** berekenen via :

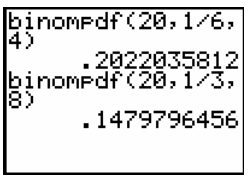
2nd[DISTR] Ø:binompdf(en **2nd[DISTR] A:binomcdf(**

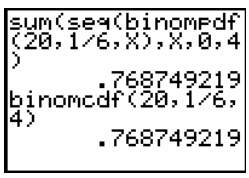
Voor $X \sim B(n,p)$ is :

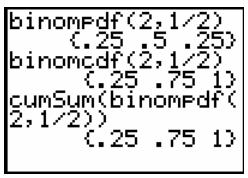
binompdf(n,p,x) = $P(X = x)$; de kansfunctie,

binomcdf(n,p,x) = $P(X \leq x) = \sum_{k=0}^{x} P(X = k)$; de cumulatieve kansfunctie.

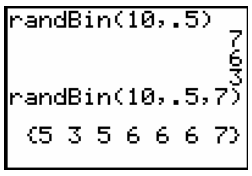
binompdf(n,p) genereert de kansverdeling van X in een lijst (d.i. de lijst van getallen $P(X = x)$ met $x = 0,1,2,\ldots,n$) en **binomcdf(n,p)** levert de cumulatieve kansverdeling op van X in een lijst (d.i. de lijst van getallen $P(X \leq x)$ met $x = 0,1,2,\ldots,n$).

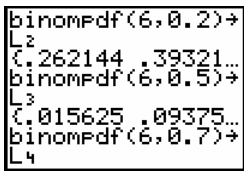


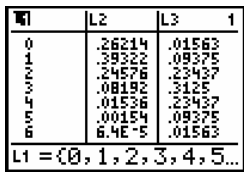




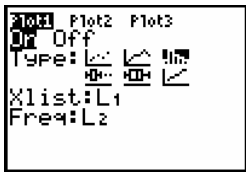
Een lukrake waarde van de binomiaal verdeelde stochast X verkrijg je met **randBin(n,p)** en een lijst van *k* zo'n waarden met **randBin(n,p,k)**. Zo simuleert **randBin(10,.5)** het aantal keer kop bij tien keer werpen van een muntstuk.

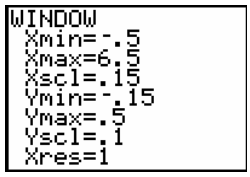


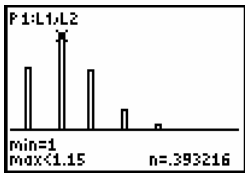


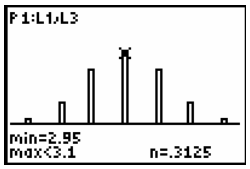


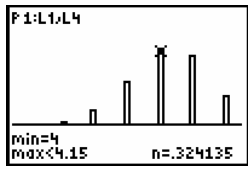
In wat volgt werden de binomiale kansverdelingen gegenereerd voor $n = 6$ en $p = 0.2, 0.5, 0.7$. Op de volgende wijze kun je ze grafisch voorstellen :

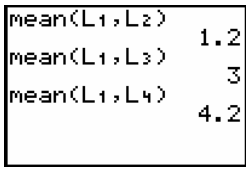












We verkrijgen een symmetrische kansverdeling als $p = 1/2$. Men kan aantonen dat $P(X = x)$ maximaal is voor $x = \text{int}[(n+1)p]$; d.i. het grootste geheel getal kleiner of gelijk aan $(n+1)\,p$.

Kun je aan de hand van het laatste plaatje een formule voor $E(X)$ voorspellen ?

We berekenen $E(X)$ en $Var(X)$ voor $X \sim B(n, p)$.

Er geldt dat $X = X_1 + X_2 + ... + X_n$ waarbij $X_i = 1$ of 0 naargelang het i-de Bernoulli experiment al dan niet succes oplevert met $X_i \sim \text{Be}(p)$ voor elke i.

Bijgevolg is : $E(X) = E(X_1) + E(X_2) + ... + E(X_n) = p + p + ... + p = np$.

Omwille van de onafhankelijkheid van $X_1, X_2, \ldots X_n$ geldt ook :

$$Var(X) = Var(X_1) + Var(X_2) + ... + Var(X_n) = pq + pq + ... + pq = npq$$

6.2 Continue stochasten

6.2.1 De uniforme verdeling

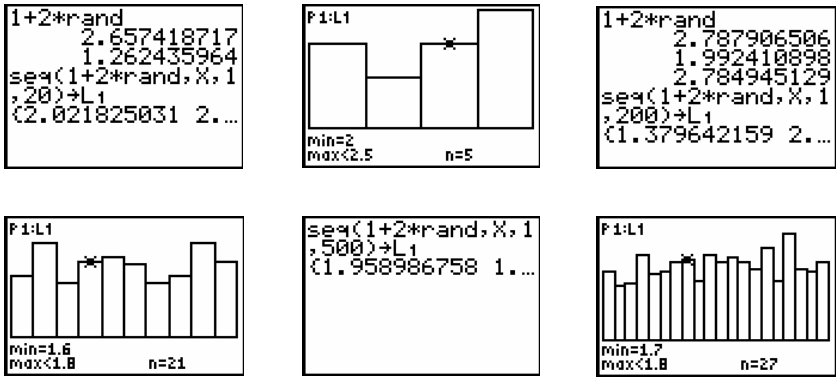
Kies lukraak een reëel getal X uit het interval $[a,b]$. Aangezien X eender welke waarde kan aannemen in een interval noemen we X een *continue* stochast.

Het heeft geen zin te vragen naar $P(X = x)$. Deze kans is steeds nul voor een continue stochast. De kansverdeling van X wordt niet meer vertolkt door een staafdiagram zoals bij een discrete stochast.

We zijn geïnteresseerd in kansen zoals $P(c \le X \le d)$.

Als voorbeeld genereren we met de **TI-83** lukraak eerst 20 getallen, vervolgens 100 getallen en tenslotte 500 getallen uit het interval $[1,3]$ (zie ook paragraaf 3.2).

We tekenen daarbij telkens een histogram van de (relatieve) frequentieverdeling.



Naarmate je meer en meer getallen genereert, kun je de klassenbreedte steeds kleiner maken. Je verwacht op de lange duur ook voor elke klasse dezelfde klassenfrequentie. Bij het laatste histogram kunnen we $P(1.7 \le X < 1.8)$ schatten met de relatieve frequentie 27 / 500 = 0.054.

Dit is de *oppervlakte* van het rechthoekje boven het interval [1.7,1.8] indien we verticaal de *relatieve frequentiedichtheid* uitzetten, d.i. de relatieve frequentie gedeeld door de klassenbreedte.

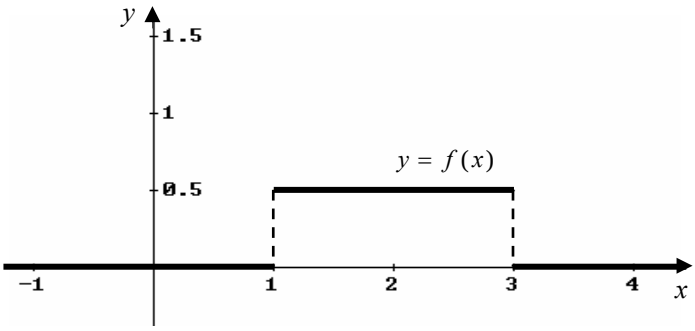
Voor het interval [1.7,1.8] is de relatieve frequentiedichtheid 0.054 / 0.1 = 0.54.

Een schatting van $P(1.6 \le X < 1.8)$ is $\frac{26+27}{500} = 0.106$ of de totale oppervlakte boven het interval $[1.6, 1.8]$.

Genereer in gedachten 1000, 2000, 4000, 8000, … getallen en halveer hierbij telkens de klassenbreedte. Je verwacht dat de getande histogramvorm meer en meer overgaat naar een horizontale lijn.

De kans op *elk* interval wordt de oppervlakte boven dit interval.

Op de lange duur verkrijgen we zo de kansverdeling van de continue stochast X :



We zeggen dat X *uniform verdeeld* is op het interval [1,3]. De kansverdeling wordt beschreven door de *(kans)dichtheidsfunctie* $f(x)$ van X .

Zo is $P(1.5 \le X \le 2) = 0.25$ de *oppervlakte* boven het interval $[1.5, 2]$. Dit is in overeenstemming met onze intuïtie : $P(1.5 \le X \le 2) = \dfrac{\text{lengte interval [1.5,2]}}{\text{lengte interval [1,3]}}$.

6.2.2 Verwachtingswaarde

Algemeen geldt voor een continue stochast X met dichtheidsfunctie $f(x)$:

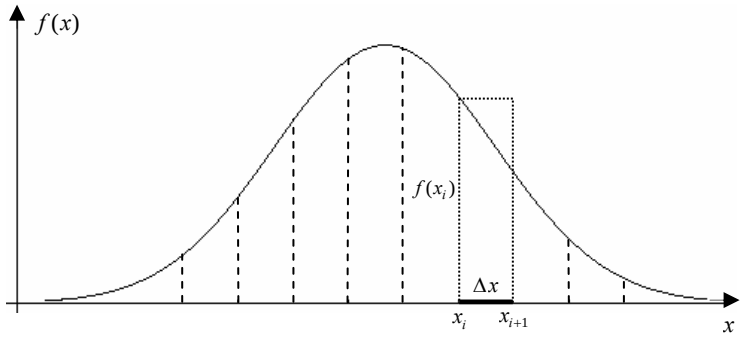
$$\begin{cases} f(x) \ge 0 \text{ voor elke } x \\ \int_{-\infty}^{+\infty} f(x)dx = 1 \end{cases}$$

Merk op dat $P(X = a) = 0$ en $P(a \le X \le b) = \int_a^b f(x)dx$. Bijgevolg geldt :

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b).$$

Hoe berekenen we de verwachtingswaarde van een continue stochast X met gegeven dichtheidsfunctie $f(x)$?

Verdeel de x-as in equidistante punten $\ldots, x_i, x_{i+1}, \ldots$. Beschouw de populatie met populatiestochast X. De kans $P(x_i \le X \le x_{i+1})$ is de relatieve frequentie van de populatiedata tussen x_i en x_{i+1}. Deze kans is ook de oppervlakte boven het interval $[x_i, x_{i+1}]$ onder de dichtheidsfunctie $f(x)$. Die oppervlakte kunnen we benaderen door $f(x_i)\Delta x$.



We vervangen voor elke i alle data tussen x_i en x_{i+1} door x_i. Zo verkrijgen we een discrete dataset van getallen $\ldots, x_i, x_{i+1}, \ldots$ met als relatieve frequenties $\ldots, f(x_i)\Delta x, \ldots$.

Het gemiddelde van deze data is : $\sum_i x_i \cdot f(x_i) \cdot \Delta x$.

Naarmate we Δx laten verkleinen wordt onze benadering steeds beter. In de limiet voor $\Delta x \to 0$ krijgen we het exacte populatiegemiddelde.

Vandaar de algemene definitie van de verwachtingswaarde of het gemiddelde van een continue stochast X met dichtheidsfunctie $f(x)$:

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)\,dx = \mu .$$

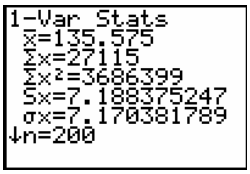
Algemener geldt : $E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x)\,dx$. Hieruit volgt :

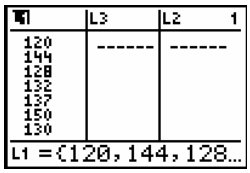
$$Var(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\,dx = \sigma^2 .$$

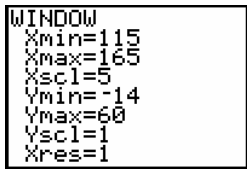
Reken na dat $E(X) = \frac{a+b}{2}$ en $Var(X) = \frac{(b-a)^2}{12}$ voor een uniform verdeelde stochast X op het interval $[a,b]$. Hierbij is $\mu = E(X)$ te voorspellen als "evenwichtspunt" van de massa die "continu" verdeeld is op de x-as volgens de dichtheidsfunctie $f(x)$.

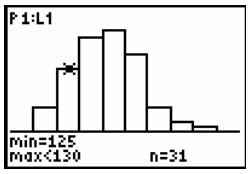
6.2.3 De normale verdeling

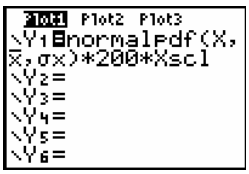
We beschouwen opnieuw als populatie de lengtes van 200 kinderen van 10 jaar (zie paragraaf 5.8).

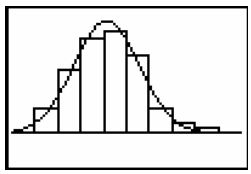












Het histogram van de data neemt ongeveer een “klokvorm” aan.

Als *wiskundig model* voor de relatieve frequentieverdeling van deze populatie gebruiken we een *normale verdeling*, waarvoor we als gemiddelde en standaardafwijking respec-tievelijk het gemiddelde en de populatiestandaar-dafwijking van onze 200 data nemen.

De totale oppervlakte onder de normale dichtheidsfunctie is 1, vandaar dat we de factor 200`Xscl` moeten toevoegen om een grafiek te krijgen die de histogram van de frequenties benadert (zie plaatjes hierboven).

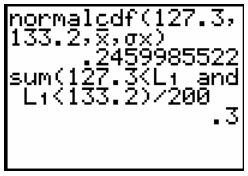
Het werken met dit model biedt als voordeel dat de klokcurve vast is, wat niet het geval is voor een histogram bij wijziging van de klassenbreedte (experimenteer hiermee).

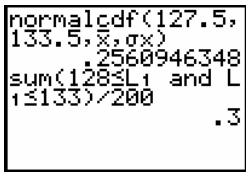
Via dit model kunnen we dan ook kansen berekenen of proporties voorspellen van populatiedata gelegen in een willekeurig interval.

Stel X de lengte van een lukraak gekozen kind, wat is dan de kans volgens ons model dat die lengte gelegen is tussen 127.3 en 133.2 cm ?

Of , anders geformuleerd, welke proportie of fractie van de kinderen hebben een lengte tussen deze grenzen ?

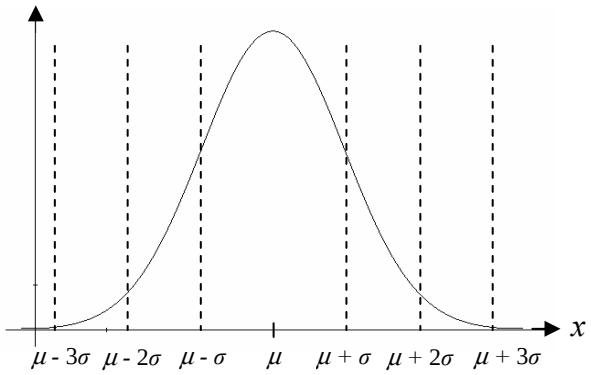
Een berekening via ons model geeft $P(127.3 < X < 133.2) = 0.246$. De exacte waarde is 0.3. Als we echter bedenken dat onze 200 data discreet zijn, krijgen we een betere continue benadering met $P(127.5 < X < 133.5) = 0.256$.





De normale verdeling is de belangrijkste continue verdeling. Een stochast X is normaal verdeeld met parameters μ en σ (met μ reëel en σ >0) indien de dichtheidsfunctie van X gegeven wordt door :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{voor } x \in \mathbb{R}$$

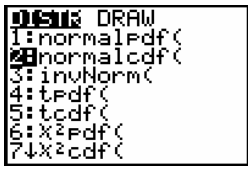


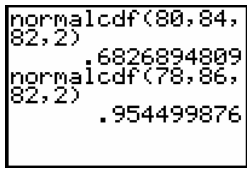
We noteren $X \sim N(\mu, \sigma)$. Er geldt :

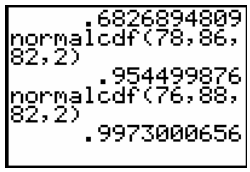
- $E(X) = \mu$ en $Var(X) = \sigma^2$
- $f(x)$ is symmetrisch rond $x = \mu$
- $\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = 0$
- $f(x)$ bereikt een maximum voor $x = \mu$
- De grafiek van $f(x)$ heeft buigpunten voor $x = \mu \pm \sigma$
- <u>De 68-95-99.7 regel</u> : de oppervlakte onder de kromme begrepen tussen

  $\mu \pm \sigma$ is 0,68269 $\approx$ 68%
  $\mu \pm 2\sigma$ is 0,95450 $\approx$ 95%
  $\mu \pm 3\sigma$ is 0,99730 $\approx$ 99,7%

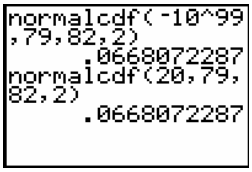
Stel dat de massa X van een groep studenten normaal verdeeld is met gemiddelde $\mu = 82$ kg en standaardafwijking $\sigma = 2$ kg. We controleren de "68-95-99,7 regel" als volgt :

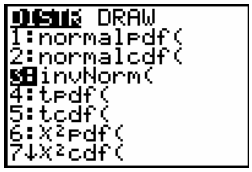


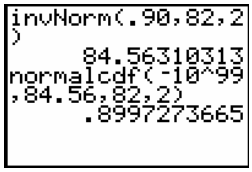




We bepalen achtereenvolgens het percentage van de studenten met een massa kleiner dan 78 kg en de massa waaronder 90% van de massa's van de studenten gelegen zijn. Deze massa noemen we het 90ste percentiel van de verdeling.







De volgende transformatie is belangrijk in de statistiek :

Voor een stochast X met gemiddelde $E(X) = \mu$ en variantie $Var(X) = \sigma^2$ geldt dat de stochast $Z = \dfrac{X - \mu}{\sigma}$ als gemiddelde $E(Z) = 0$ en als variantie $Var(Z) = 1$ heeft.

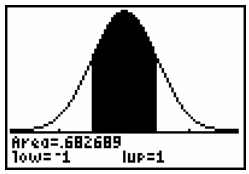
Men kan bovendien aantonen dat Z normaal verdeeld is als X normaal verdeeld is.

M.a.w. als $X \sim N(\mu;\sigma)$, is $Z = \dfrac{X - \mu}{\sigma} \sim N(0;1)$.

Men zegt dat Z *standaard normaal verdeeld* is.

De dichtheidsfunctie van Z is : $f(z) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$ voor $z \in \mathbb{R}$.

Dat de oppervlakte onder de kromme begrepen tussen $\mu \pm \sigma$ gelijk is aan 68 % kunnen we grafisch voorstellen met :



2nd[DISTR]<DRAW> 1:ShadeNorm(-1,1).

Voor het bovenstaand voorbeeld waarbij $X \sim N(82;2)$ de massa van de studenten voorstelt, geldt : $\frac{X-82}{2} \sim N(0;1)$.

Na een lukrake keuze van een student uit onze populatie krijgen we een concrete massa x met bijbehorende $z = \frac{x-\mu}{\sigma}$. Het getal z noemt men de z-score of standaardscore.

Dit getal vertelt ons hoeveel standaardafwijkingen σ de grootheid x verwijderd is van het gemiddelde μ (zie de onderstaande tabel) :

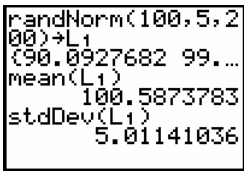
x	$\mu-3\sigma$	$\mu-2\sigma$	$\mu-\sigma$	μ	$\mu+\sigma$	$\mu+2\sigma$	$\mu+3\sigma$
z	-3	-2	-1	0	1	2	3

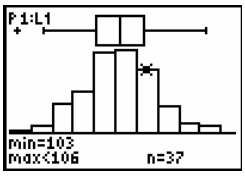
Je kunt de z-score gebruiken om data te vergelijken die afkomstig zijn van twee verschillende normale verdelingen.

6.2.4 Onderzoek of data afkomstig zijn van een normale verdeling

a) Je beschikt over voldoende data

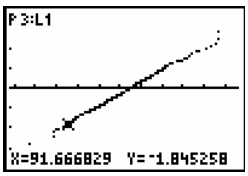
Als illustratie genereren we 200 getallen uit een normaal verdeelde populatie met gemiddelde 100 en standaardafwijking 5 :





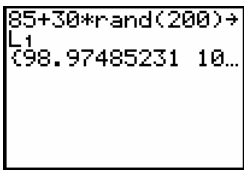
Een histogram wijst al op een normale verdeling en de box-plot met uitschieters toont een symmetrische verdeling (met één uitschieter). De interkwartielafstand is ongeveer 1/3 van de lengte van de box-plot. Het steekproefgemiddelde, 100.59, en de steekproefstandaardafwijking, 5.01, van onze data zijn schattingen voor het populatiegemiddelde 100 en de populatiestandaardafwijking 5.

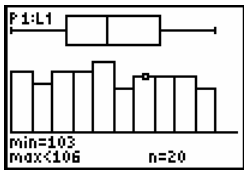
A.h.v. een *waarschijnlijkheidsgrafiek* kunnen we controleren of de data afkomstig zijn van een normale verdeling :

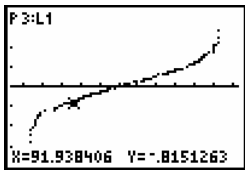


De (getransformeerde) data liggen nagenoeg op een rechte. Dit wijst op een normale verdeling. Merk op dat er op de uiteinden minder punten zijn dan rond het centrum.

Ter vergelijking genereren we 200 getallen uit een uniforme verdeling op het interval [85, 115].

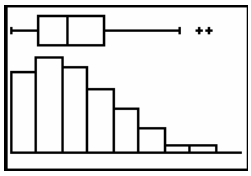
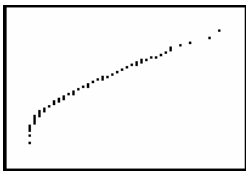






Het histogram wijst op een uniforme verdeling en de box-plot op een symmetrische verdeling waarbij de interkwartielafstand ongeveer de helft van de lengte van de box-plot is. De normale waarschijnlijkheidsgrafiek wijst wel op een symmetrische verdeling maar wijkt duidelijk af van een rechte op de uiteinden. Merk op dat de punten gelijkmatig verdeeld zijn over de kromme.

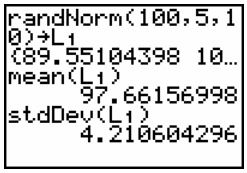
Voor een scheve verdeling gebruiken we de lijst **ELEKT** van 120 versterkingsfactoren van versterkers (zie 4.7) :

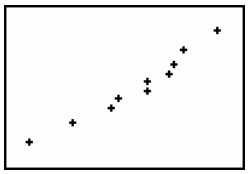



De normale waarschijnlijkheidsgrafiek is niet symmetrisch, wijkt af van een rechte en de punten liggen minder dicht aan de rechterzijde; dit wijst op een scheve verdeling naar rechts.

b) Je beschikt over weinig data

Als illustratie genereren we slechts 10 getallen uit een normaal verdeelde populatie met gemiddelde 100 en standaardafwijking 5.

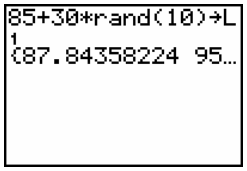


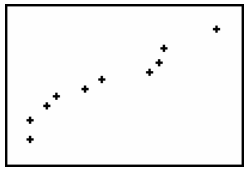


Het steekproefgemiddelde, 97.66, en de steekproefstandaardafwijking, 4.21, van onze data zijn (minder goede) schattingen voor het populatiegemiddelde 100 en de populatiestandaardafwijking 5.

Een histogram of een box-plot zijn hier niet zo zinvol. De normale waarschijnlijkheidsgrafiek toont dat de punten niet te veel van een rechte afwijken zodat we kunnen aanvaarden dat de data afkomstig zijn van een normale populatie.

Ter vergelijking genereren we 10 getallen uit een uniforme verdeling op het interval [85, 115].

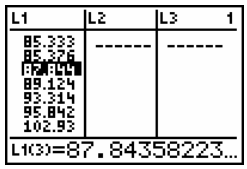


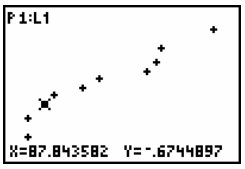


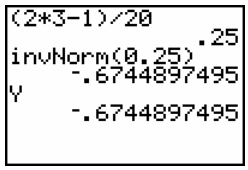
De punten van de normale waarschijnlijkheidsgrafiek vertonen eerder geen lineair verband.
We bekijken even hoe een normale waarschijnlijkheidsgrafiek tot stand komt. We nemen hiervoor de tien data van het laatste voorbeeld.

- Rangschik de data van klein naar groot met **SortA(L1)**. We noteren de gerangschikte lijst met $x_1, x_2, \ldots, x_n$.
- Bereken, bij benadering, voor elke x_i de relatieve cumulatieve frequentie c_i (= de proportie van de data kleiner dan of gelijk aan x_i) met de formule $c_i = \dfrac{2i-1}{2n}$.
- Zoek de corresponderende waarde z_i uit de standaard normale verdeling met dezelfde relatieve cumulatieve frequentie als x_i m.b.v. $z_i =$ **invNorm(** c_i **)**.
- Als de data x_i afkomstig zijn uit een normale verdeling met gemiddelde μ en standaardafwijking σ, geldt voor de corresponderende z_i dat $z_i = \dfrac{x_i - \mu}{\sigma}$. Dit is een lineair verband !

We illustreren dit voor het derde punt ($i = 3$ en $n = 10$) in onze normale waarschijnlijkheidsgrafiek waarbij we de gegeven data x_i horizontaal uitzetten en de corresponderende z_i verticaal.







Opgepast ! Hoe meer data, hoe betrouwbaarder de visuele controle op normale verdeling met een waarschijnlijkheidsgrafiek is.

Met tien data kan je voorspelling eerder foutief zijn dan met 200 data (test dit uit d.m.v. enkele simulaties).

Dit is steeds zo in de statistiek. Met meer steekproefdata doe je betrouwbaarder uitspraken over de populatie. Maar die luxe heb je niet vaak.

6.2.5 De centrale limietstelling

Wat kunnen we zeggen over de verdeling van een som van stochasten ?

Een lineaire combinatie van *onafhankelijke, normaal verdeelde* stochasten is steeds normaal verdeeld.

Als $X_i \sim N(\mu_i, \sigma_i)$, is $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ ook normaal verdeeld.

M.b.v. de eigenschappen van de operatoren E en Var vinden we :

$$E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$$

$$Var(Y) = a_1^2 \cdot Var(X_1) + a_2^2 \cdot Var(X_2) + \ldots + a_n^2 \cdot Var(X_n)$$

Als de stochasten niet normaal verdeeld zijn, helpt de *centrale limietstelling* ons verder.

Als $X_1, X_2, \ldots, X_n$ *onafhankelijke* stochasten zijn met *dezelfde* verdeling (continu of discreet, maar met eindige variantie), zal $S_n = X_1 + X_2 + \ldots + X_n$ *bij benadering* normaal verdeeld zijn voor *n voldoende groot*.

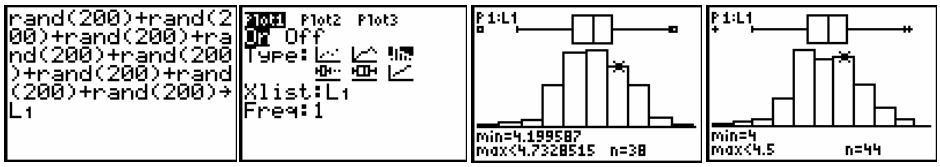
Deze normale verdeling heeft een gemiddelde gelijk aan $E(S_n)$ en een variantie gelijk aan $Var(S_n)$.

Bijgevolg zal ook het steekproefgemiddelde $\overline{X} = \frac{S_n}{n}$ van een steekproef uit een populatie met een *willekeurige* verdeling, nagenoeg normaal verdeeld zijn voor *n* voldoende groot.

Stel μ het gemiddelde en σ de standaardafwijking van de populatie. Voor *n* voldoende groot geldt : $\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$. Dit is een belangrijk resultaat.

We illustreren de centrale limietstelling met twee voorbeelden :

a) Stel $X_1, X_2, \ldots, X_8$ onafhankelijke en uniform verdeelde stochasten op het interval [0,1]. Onderzoek de verdeling van $Y = X_1 + X_2 + \ldots + X_8$.

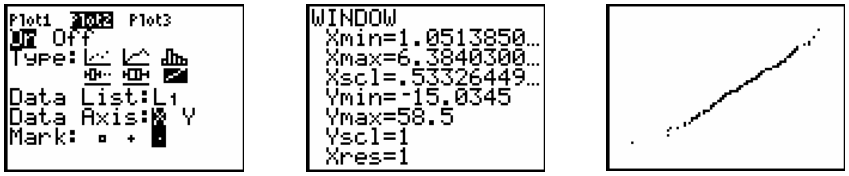


In **L1** bevinden zich 200 lukraak gegenereerde waarden van de stochast *Y*.

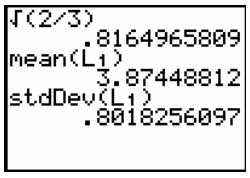
Een histogram van de data in **L1** geeft reeds een idee van de verdeling van *Y* (observeer de invloed van de klassenbreedte). De normale verdeling is klaarblijkelijk een goed model.

Ook de box-plot met uitschieters wijst op een redelijk symmetrische verdeling.

We onderzoeken met een normale waarschijnlijkheidsgrafiek of de data afkomstig zijn uit een normale verdeling.



De getransformeerde data liggen ongeveer op een rechte; een normaal model is aannemelijk. Uitschieters ontdek je hier als punten op de uiteinden die wat afgelegen zijn van de andere punten.



Het gemiddelde, 3.87, en de steekproef-standaardafwijking, 0.80, van de data in **L1** zijn schattingen voor respectievelijk $E(Y) = 4$ en

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{Var(X_1) + Var(X_2) + \ldots + Var(X_8)}$$
$$= \sqrt{8 \cdot \frac{1}{12}} = 0.816$$

b) De normale verdeling als benadering van de binomiale verdeling.

Voor een binomiaal verdeelde stochast $X \sim B(n, p)$ geldt dat $X = X_1 + X_2 + \cdots + X_n$ waarbij $X_i = 1$ of 0 naargelang het i-de Bernoulli experiment al dan niet succes oplevert ($X_i \sim$ Be(p) voor elke i).

We mogen omwille van de centrale limietstelling verwachten dat X bij benadering normaal verdeeld zal zijn voor n voldoende groot.

Een eerlijk muntstuk (kans op kop = kans op munt) wordt 20 keer geworpen.

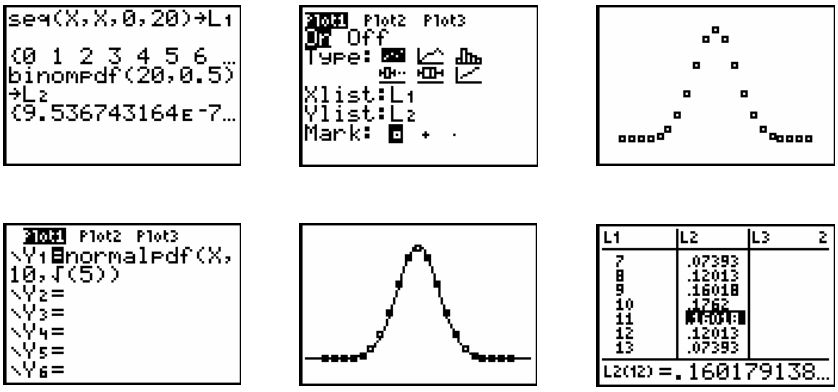
Stel X het aantal keer munt.

X is binomiaal verdeeld met parameters $n = 20$ en kans op succes $p = 0.5$.

Voor het gemiddelde en de variantie vinden we :

$\mu = np = 10$ en $\sigma^2 = npq = 5$.

We vergelijken de binomiale kansfunctie met de benaderende normale dichtheidsfunctie (met hetzelfde gemiddelde en dezelfde variantie).

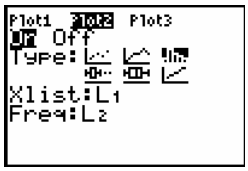


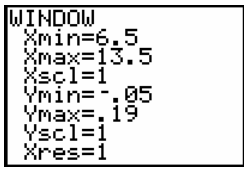
We zien dat de discrete punten van de kansfunctie van de binomiale verdeling goed aansluiten bij de continue dichtheidsfunctie van de normale verdeling.

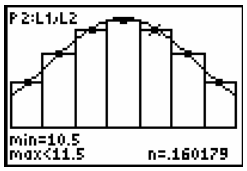
Merk op dat $P(X = 11) = 0.16$ voor $X \sim B(20, 0.5)$.

Deze kans kunnen we ook voorstellen als oppervlakte van de rechthoek getekend op de basis 10.5–11.5 en met hoogte 0.16.

Op deze wijze kunnen we de kansverdeling van de binomiale verdeling voorstellen door een *kanshistogram* met totale oppervlakte 1.







Stel dat we een prijs winnen bij 9, 10 of 11 keer munt.

De kans op een prijs is exact :

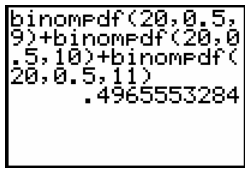
(i) $P(9 \le X \le 11) = P(X = 9) + P(X = 10) + P(X = 11)$
(met de kansfunctie **binompdf)**

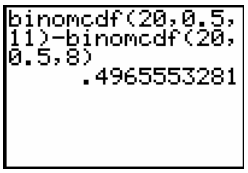
(ii) $P(9 \le X \le 11) = P(X \le 11) - P(X \le 8)$
(met de cumulatieve kansfunctie **binomcdf**)

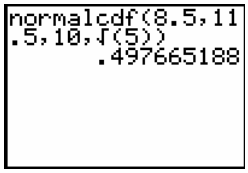
Om deze kans te benaderen met behulp van de normaal verdeelde stochast $Y \sim N\left(10, \sqrt{5}\right)$ gebruiken we de *continuïteitscorrectie* :

$$P(9 \le X \le 11) \approx P(8.5 \le Y \le 11.5) .$$

Dit ligt voor de hand als we kijken naar het kanshistogram waarbij we de totale oppervlakte van 3 rechthoeken moeten benaderen door een oppervlakte onder de dichtheidsfunctie van *Y*.







Als vuistregel kunnen we stellen dat $B(n, p)$ goed kan benaderd worden door $N\left(np, \sqrt{npq}\right)$ van zodra $np > 5$ en $nq > 5$ (met $q = 1 - p$).

De **TI-83** kan echter exact rekenen met de binomiale verdeling voor $n \le 999$. Een benadering door een normale verdeling is pas echt nodig voor $n \ge 1000$.

6.3 Opdrachten

1. Iemand wedt dat er bij 12 worpen van een muntstuk precies 6 keer kop verschijnt. Bereken de winstkans.

2. Bereken de kans om bij 5 worpen van een dobbelsteen minstens 2 keer een zes te werpen.

3. Een fabriek maakt schroeven die met kans 0.001 defect zijn. Bereken de kans dat er bij een levering van 500 schroeven minstens 2 defect zijn.

4. Hoe groot is de kans dat in 6000 worpen met een dobbelsteen minstens 1050 keer een zes geworpen wordt ?

5. Mark speelt tafeltennis tegen Luc. Mark is de betere speler en wint een spel tegen Luc met kans 0.6. Er wordt afgesproken om een tornooi in te richten van $n = 1,3,5,7,\ldots,2k+1$ spelen. Overwinnaar is diegene die meer dan de helft van het aantal spelen *n* wint. Hoe groot mag *n* zijn opdat Luc zou winnen met een kans groter dan 0.25 ?

6. Onder al de personen die een plaats reserveren bij een vliegtuigmaatschappij komen er 4 % niet opdagen. De vliegtuigmaatschappij weet dat en verkoopt 75 kaarten voor 73 plaatsen. Hoe groot is de kans dat al de passagiers een plaats hebben ?

7. Stel dat $X \sim B(25, 0.3)$.

   a) Bereken exact $P(7 \le X \le 12)$.
   b) Bereken $P(7 \le X \le 12)$ met een normale benadering en continuïteits-correctie.
   c) Bereken $P(7 \le X \le 12)$ met een normale benadering zonder continuïteitscorrectie en vergelijk met (b).

8. In een klas met 45 studenten zijn de examenresultaten (max. score 70) van wiskunde (bij benadering) normaal verdeeld met gemiddelde 47 en standaardafwijking 3. De examenresultaten elektronica (max. score 50) zijn normaal verdeeld met gemiddelde 27 en standaardafwijking 2. Een student haalt 52 op 70 voor wiskunde en 31 op 50 voor elektronica. Voor welk vak presteerde hij het best ?

9. De levensduur van een bepaald type gloeilampen is, gemeten in uren, verdeeld volgens $N(1000,120)$. Bereken de kans dat een lukraak gekozen lamp minder dan 800 uur brandt.

10. Tijdens een zomervakantie organiseerde men het spel “kapers op de kust”. Door een aantal dagelijkse opdrachten komt er op het einde van de week een overwinnaar uit de bus die mag deelnemen aan een spel om één van de drie tentoongestelde wagens te winnen (waaronder een cabriolet). De kandidaat moet eerst één sleutel kiezen uit 6 sleutels, waaronder de 3 sleutels van de te winnen wagens en 3 andere sleutels.

    Vervolgens mag de kandidaat proberen met zijn sleutel één van de drie wagens te starten. Als dit niet lukt, mag hij nog proberen één andere wagen te starten met die sleutel. De kandidaat wint de wagen die start.

    a) Bereken de kans om een wagen te winnen.
    b) Na 8 weken en 8 keer spelen werden er 6 wagens gewonnen. Bereken de kans om 6 keer een wagen te winnen op 8 keer spelen. Simuleer ook 8 keer spelen.

11. Voor een bepaalde groep mensen zijn de intelligentiequotiënten verdeeld volgens $N(115,13)$. Bereken het percentage van de groep met een I.Q. tussen 130 en 140.

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

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

13. Stel dat X en Y onafhankelijke stochasten zijn met $X \sim N(30,4)$ en $Y \sim N(14,3)$. Bereken $P(W > 50)$ als $W = X + Y$.

14. De massa (in gram) van pakjes koffie is verdeeld volgens $N(102,3)$. We doen 36 pakjes in een doos. Bepaal de verdeling van de massa in de doos.

15. De stochasten *U, V, W* zijn onafhankelijk en alle drie verdeeld volgens $N(\mu,\sigma)$ .Geef de verdeling van de volgende stochasten :

    a) $U+V+W$ b) $2U$ c) $U-V$
    d) $2U+V$ e) $U+2$ f) $V-\mu$
    g) $\dfrac{W-\mu}{\sigma}$ h) $\dfrac{W}{2}$ i) $2U-V-W$

16. De snelheden van de wagens die op een bepaalde plaats van de autosnelweg passeren, zijn normaal verdeeld. Observaties tonen dat 95% van de wagens daar trager rijden dan 120 km/h en 10% trager dan 85 km/h.

    a) Vind de gemiddelde snelheid van de wagens.
    b) Zoek de proportie van de wagens die sneller rijden dan 100 km/h.

17. Veronderstel dat de massa van de jongens verdeeld is volgens $N(58,6.4)$ en die van de meisjes volgens $N(52,5.9)$. Bereken de kans dat een lukraak gekozen meisje minder weegt dan een lukraak gekozen jongen.

18. Als X en Y onafhankelijke stochasten zijn met $X \sim N(2,5)$ en $Y \sim N(3,4)$, bereken dan :

    a) $P(X+Y<7)$ b) $P(2X-3Y>-10)$
    c) $P(X>Y)$ d) $P\left(\left|X-Y\right|<5\right)$

19. De massa (in kg) van een groep zware mannen is verdeeld volgens $N(100,10)$.

    a) Hoe groot is de kans dat de massa X van één lukraak gekozen man meer dan 4 afwijkt van 100 ?
    b) Hoe groot is de kans dat het steekproefgemiddelde $\overline{X}$ van de massa's van 25 lukraak gekozen mannen (met "terugleggen") meer dan 4 afwijkt van 100 ?

20. Kies lukraak 12 getallen uit het interval [0,1]. Bereken de kans dat de som van deze 12 getallen meer dan 8 oplevert.