Statistiek deel 1 week 3

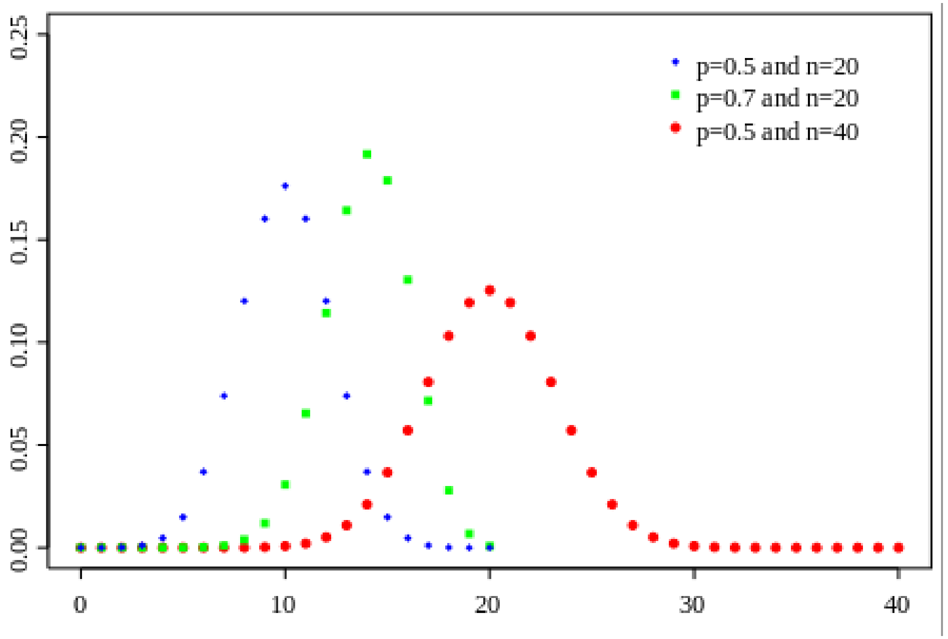
- Herhaling Binomiale verdeling
- Geometrische verdeling
- Hypergeometrische verdeling
- Continue kansverdelingen
- Uniforme verdeling
- Operaties met kansvariabelen

Binomiale verdeling H6.1 + 6.2

De **binomiale verdeling** beschrijft een kansexperiment waarbij je net als bij de Bernoulliverdeling twee mogelijke uitkomsten hebt, (0 en 1) maar dit vervolgens n keer uitvoert, telkens met dezelfde slaagkans p. De kansvariabele $\underline{k}$ is het aantal keren dat je succes hebt ($\underline{k} = 1$ met kans p). Het wordt beschreven door een kansvariabele $\underline{k}$ waarvoor geldt:

$$f(k) = P(\underline{k} = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{voor} \quad k = 0, 1, 2, \dots, n$$

$p \in [0,1]$ is een willekeurig, vast getal (Bv. $p = \frac{1}{2}$ voor een eerlijke munt).



Omdat de trekkingen onafhankelijk zijn is verwachtingswaarde de som van de verwachtingswaarden van n keer het Bernoulli-experiment (zie rekenregels):

$$\mu = E(\underline{k}) = np$$

Hetzelfde geldt voor de variantie:

$$\sigma^2 = E\left((\underline{k} - \mu)^2\right) = np(1-p),$$

$$\sigma = \sqrt{np(1-p)}.$$

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

Er zijn twee parameters: n en p

Notatie: $\underline{k} \sim bin(n; p)$

Geometrische verdeling. 6^{+}.2

Vraag: Hoe vaak moet je gemiddeld met een dobbelsteen gooien om een 6 te gooien?

Antwoord: 6, want $6 \times \frac{1}{6} = 1$

Klopt dit?

Wat is de kans op gooien van de eerste 6?

Kans op 1e 6 bij de eerste worp: $\frac{1}{6}$

Kans op 1e 6 bij de tweede worp: $\frac{5}{6} \cdot \frac{1}{6}$

Kans op 1e 6 bij de derde worp: $\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$

……

Kans op 1e 6 bij de k-de worp: $\left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$

Kansexperiment: Gooi een dobbelsteen totdat de eerste 6 verschijnt

Kansvariabele: $\underline{k}$ is het aantal keer gooien waarop de eerste 6 verschijnt.

Kansverdeling: $f(k) = P(\underline{k} = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$, voor alle $k \geq 0$

Geometrische verdeling

Beschrijft een kansexperiment met twee mogelijke uitkomsten, (0 en 1) (een Bernoulli experiment) dat telkens wordt uitvoerd, steeds met dezelfde slaagkans p.
De kansvariabele $\underline{k}$ is het aantal keren dat je het experiment moet doen om succes te hebben.
(De kans op k successen bij n keer uitvoeren wordt gegeven door $\underline{k} \sim \mathrm{b}in(n; k)$)

Notatie: $\underline{k} \sim \mathrm{geom}(p)$, er is één parameter: p. De kansfunctie is

$$f(k) = P(\underline{k} = k) = (1-p)^{k-1} p \text{ voor } \; k = 1, 2, 3, \ldots$$

$p \in [0,1]$ is een willekeurig, vast getal. De verwachtingswaarde is

$$E(\underline{k}) = \sum_{\text{alle } k} k f(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

Geometrische verdeling. 6⁺.2 (Afleiding niet voor tentamen)

$$f(k) = P(\underline{k} = k) = (1-p)^{k-1}p \text{ voor } \; k = 1, 2, 3, \ldots$$

De som van alle kansen is 1:

$$\sum_{k=1}^{\infty} (1-p)^{k-1}\, p = 1$$

dus $\sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{1}{p}$

Differentieer naar p: $-\sum_{k=1}^{\infty}(k-1)(1-p)^{k-2} = -\frac{1}{p^2}$

dus ($k = 1$ geeft 0): $\sum_{k=2}^{\infty}(k-1)(1-p)^{k-2} = \frac{1}{p^2}$

dus (vervang k door $k+1$): $\sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p}$

dus $\boldsymbol{E(\underline{k}) = \frac{1}{p}}$

Op soortgelijke manier: $\boldsymbol{\mathrm{Var}(\underline{k}) = \frac{1-p}{p^2}}$

dus $\boldsymbol{\sigma(\underline{k}) = \frac{\sqrt{1-p}}{p}}$

Uit te rekenen op de GR: Geometpdf/cdf(p , k)

Hypergeometrische verdeling. 3⁺.2, 6⁺.1

Een vaas bevat a rode en b witte knikkers. Uit de vaas worden n knikkers getrokken (zonder teruglegging). $\underline{k}$ is het aantal rode knikkers. Notatie: $\underline{k} \sim \mathrm{hyp}(a; b; n)$. De kansfunctie is

$$f(k) = P(\underline{k} = k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}$$

Er geldt:

$$E(\underline{k}) = np \quad \text{met} \quad p = \frac{a}{a+b}$$

$$\mathrm{Var}(\underline{k}) = np(1-p)\frac{N-n}{N-1} \quad \text{met} \quad N = a+b \quad \text{(totaal aantal knikkers)}$$

$$\sigma(\underline{k}) = \sqrt{\mathrm{Var}(\underline{k})}$$

Als n veel kleiner is dan N (vuistregel: $n < 0{,}1N$), dan is $\frac{N-n}{N-1} \approx 1$ en is er nauwelijks verschil tussen

trekken zonder terugleggen $\qquad \underline{k} \sim \mathrm{hyp}(a; b; n)$

en trekken met terugleggen $\qquad \underline{k} \sim \mathrm{bin}(n; p = a/(a+b))$

Hypergeometrische verdeling. 3⁺.2, 6⁺.1

Een vaas bevat a rode en b witte knikkers. Uit de vaas worden n knikkers getrokken (zonder teruglegging). $\underline{k}$ is het aantal rode knikkers. Notatie: $\underline{k} \sim hyp(a; b; n)$ De kansfunctie is

$$f(k) = P(\underline{k} = k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}$$

Voorbeeld: Uit een groep met 7 adelborsten en 11 cadetten wordt willekeurig een commissie twee personen samengesteld. Wat is de kans dat de commissie uit een adelborst en een cadet bestaat?

$a = 7, b = 11, n = 2$, $\underline{k}$ is het aantal adelborsten in de commissie, $k = 1$

$$P(\underline{k} = 1) = \frac{\binom{7}{1}\binom{11}{2-1}}{\binom{18}{2}} = 0{,}5032$$

Continue kansvariabelen

Bij een **continue kansvariabele** is de uitkomst van het kansexperiment een reëel getal (kommagetal), bijvoorbeeld: het aantal liter brandstof dat op een dag door een eenheid wordt verbruikt, de maximale temperatuur op een dag in een koelcel, de beladingsgraad van een truck.

Bij continue variabelen wordt alleen gewerkt met kansen dat de continue stochast in een bepaald interval ligt:

$$P(a \leq \underline{x} \leq b)$$

Het maakt daarbij niet uit of je ≤ of < gebruikt, de kansen blijven gelijk, want de kans dat precies één specifieke waarde wordt aangenomen is altijd 0: $P(\underline{x} = a) = 0$

Er wordt altijd gewerkt met een cumulatieve kansverdeling (**cdf, nooit pdf**), de **verdelingsfunctie**:

$$F(x) = P(\underline{x} \leq x)$$

Hierdoor is

$$P(a \leq \underline{x} \leq b) = F(b) - F(a)$$

F is een functie die niet daalt, van waarde 0 (in $-\infty$, of de kleinste waarde) tot 1 (in ∞, of de grootste waarde).

Continue kansvariabelen

De afgeleide van de verdelingsfunctie

$$F(x) = P(\underline{x} \leq x)$$

heet ook wel de **kansdichtheidsfunctie**

$$f(x) = F'(x)$$

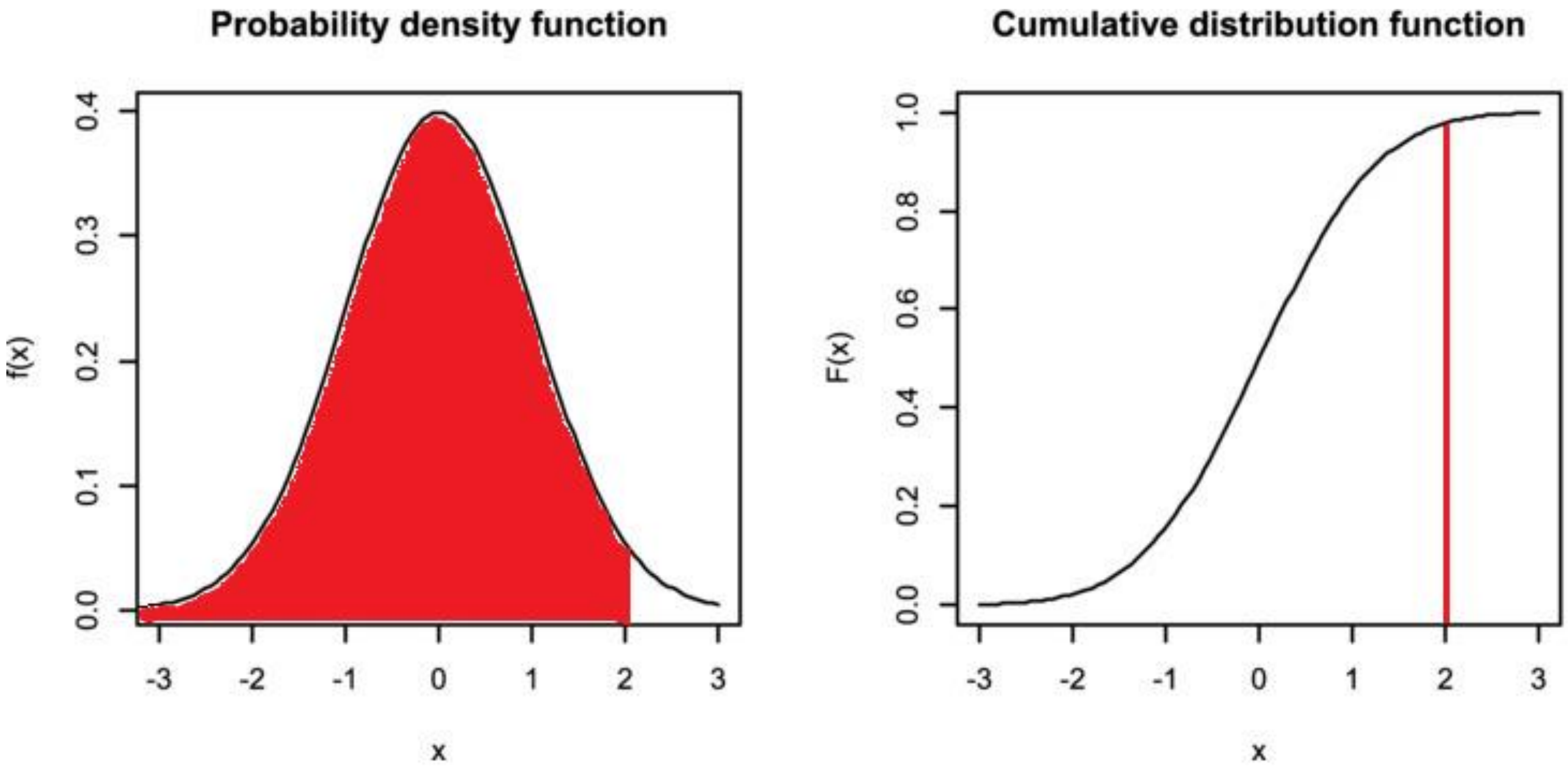
f is een functie die overal ≥ 0 is.
Je kunt hem opvatten als de continue vervanger van de kansfunctie bij discrete variabelen.

Belangrijke eigenschap: Kansen zijn integralen van de kansdichtheidsfunctie:

$$P(a \leq \underline{x} \leq b) = \int_a^b f(x)dx$$

De oppervlakte onder de grafiek van f tussen $x = a$ en $x = b$ is dus de kans dat de waarde van de kansvariabele ligt tussen a en b.

Continue kansvariabelen



$$F(2) = P(\underline{x} \leq 2) = \int_{-\infty}^{2} f(x)dx$$

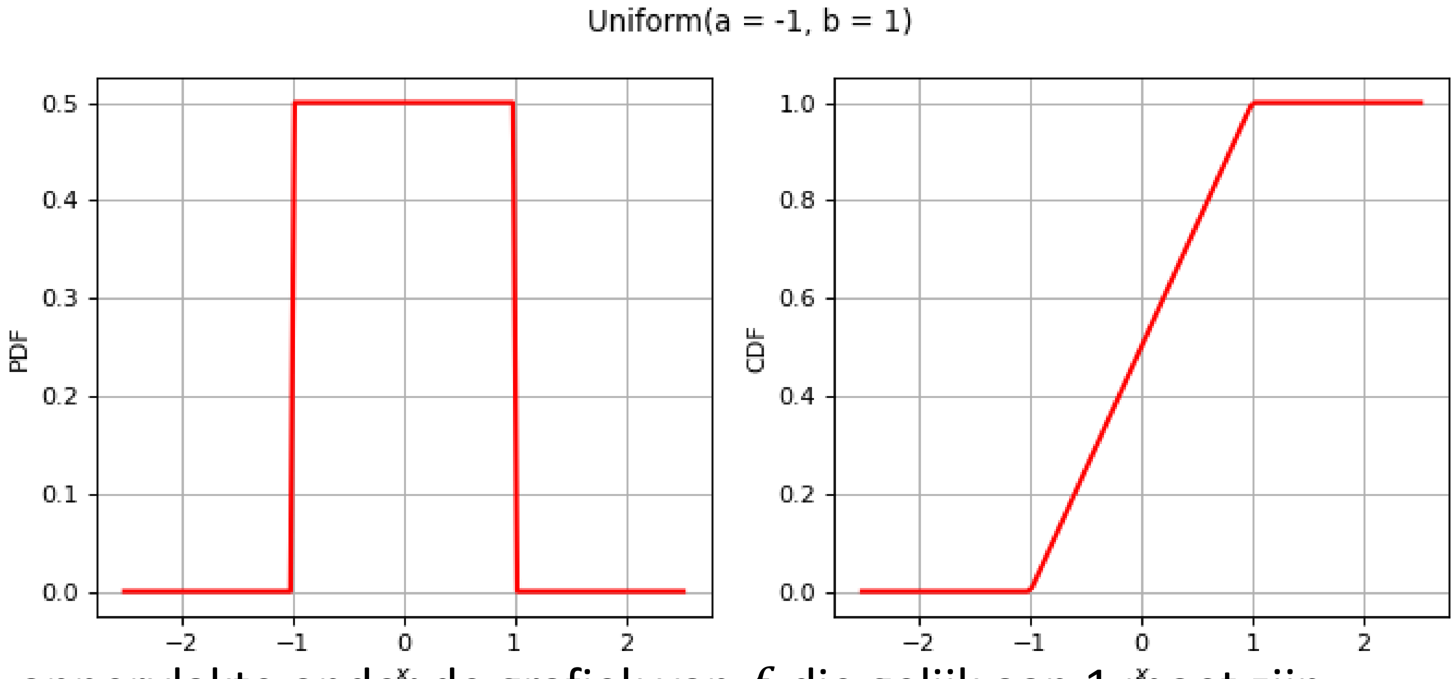
Continue kansvariabelen: μ en σ

De definities van verwachtingswaarde, variantie en standaarddeviatie gaan ook analoog aan die bij discrete kansvariabelen maar dan met integratie i.p.v. sommatie:

Discreet	Continu
$\mu = E(\underline{k}) = \sum_{k=1}^{n} k f(k)$	$\mu = E(\underline{x}) = \int_{-\infty}^{\infty} x f(x) dx$
$\mathrm{Var}(\underline{k}) = E\left((\underline{k} - \mu)^2\right) =$ $= E(\underline{k}^2) - E(\underline{k})^2 = \sum_{k=1}^{n} k^2 f(k) - \mu^2$	$\mathrm{Var}(\underline{x}) = E\left((\underline{x} - \mu)^2\right) =$ $= E(\underline{x}^2) - E(\underline{x})^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$
$\sigma = \sqrt{\mathrm{Var}(\underline{k})}$	$\sigma = \sqrt{\mathrm{Var}(\underline{x})}$

Uniforme verdeling H3⁺.2, 6⁺.1

De eenvoudigste continue verdeling is de **uniforme verdeling**. Er zijn twee parameters, a en b, met $a \leq b$, en de kans dat $\underline{x}$ ergens tussen a en b ligt is overal hetzelfde (=uniform). Dat betekent dat de kansdichtheid tussen a en b constant is, en 0 daarbuiten:



$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elders} \end{cases}$$

$$\mathrm{F}(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$

Notatie: $x \sim \text{uniform}(a; b)$

De hoogte van de functie $\left(\frac{1}{b-a}\right)$ volgt uit de oppervlakte onder de grafiek van f die gelijk aan 1 moet zijn.

Als $c \leq d$ in het interval $[a, b]$ liggen, dan is

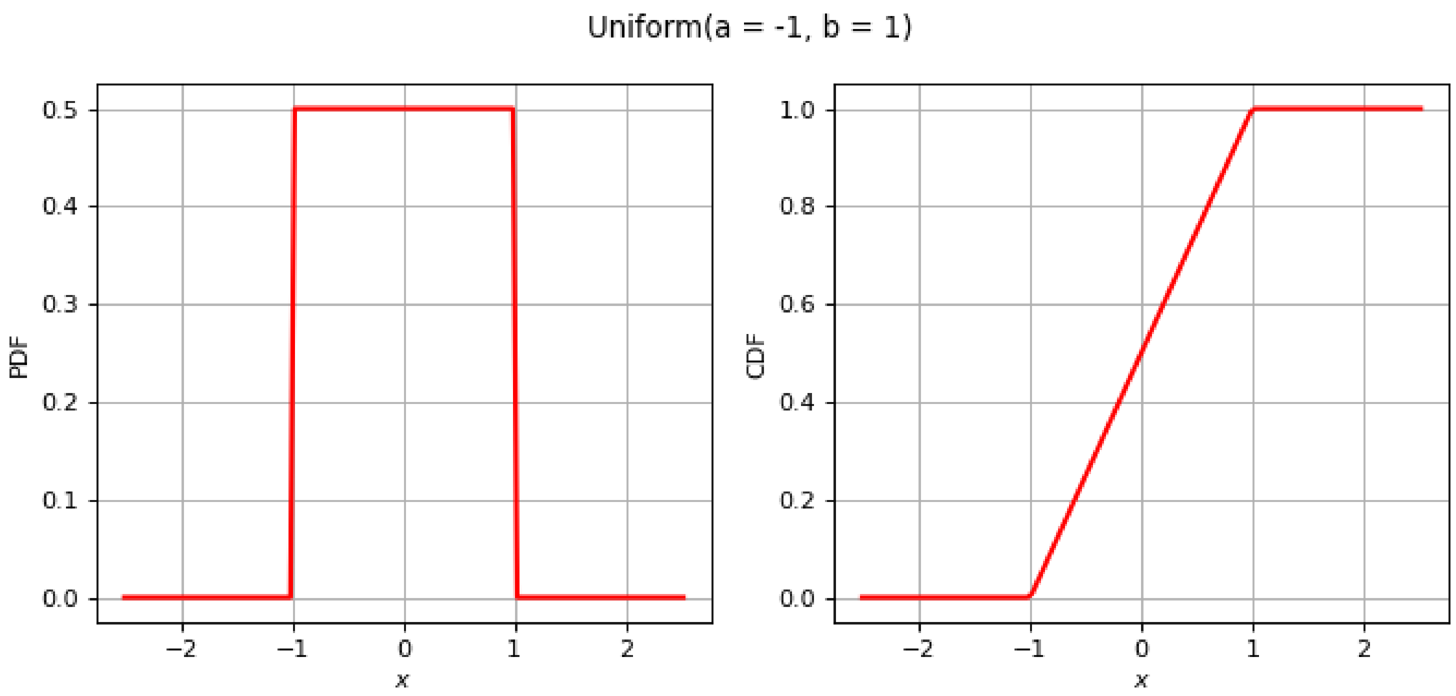
$$P(c \leq \underline{x} \leq d) = F(d) - F(c) = \frac{d-c}{b-a},$$

Dit is de lengte van $[c, d]$ gedeeld door de lengte van $[a, b]$, dat is ook logisch.

Uniforme verdeling H3⁺.2, 6⁺.1

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{elders} \end{cases}$$

$$\mathrm{F}(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$



De verwachtingswaarde is

$$E(\underline{x}) = \frac{a+b}{2}$$

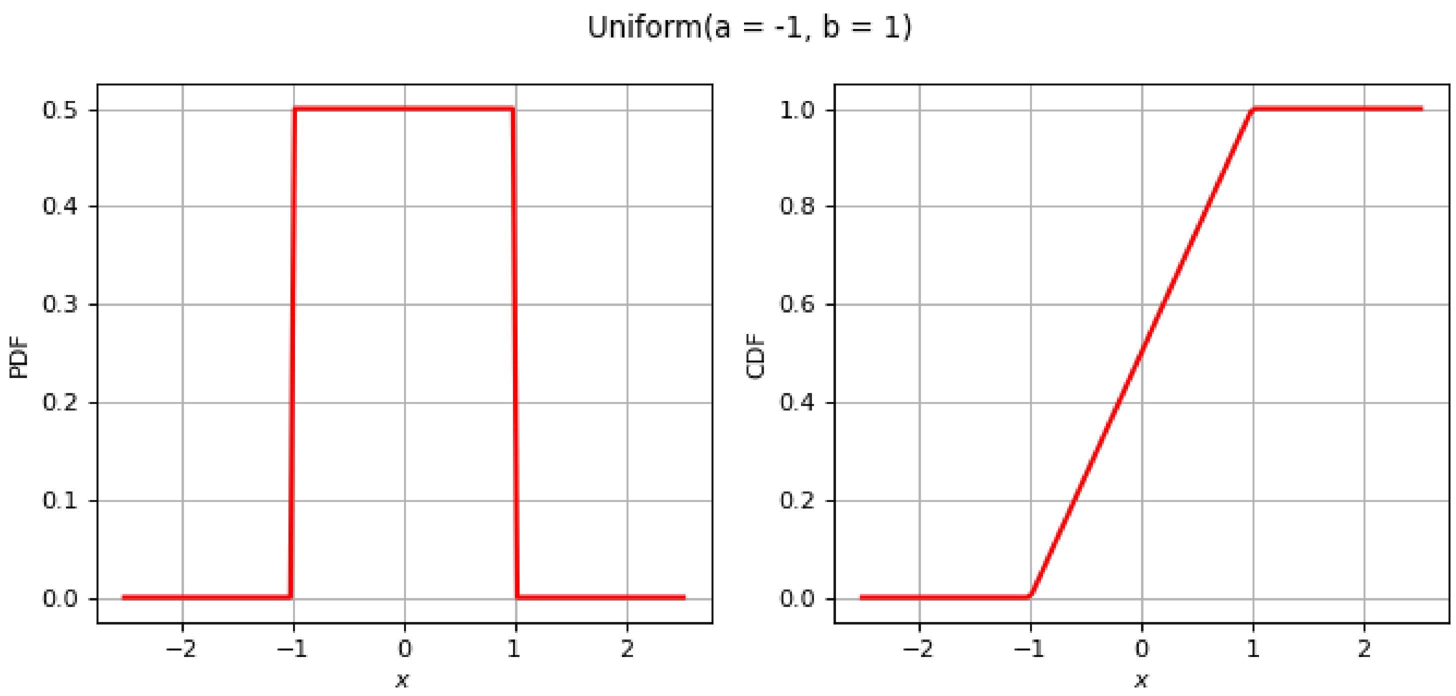
$$\mathrm{Var}(\underline{x}) = \frac{(b-a)^2}{12}$$

$$\sigma(\underline{x}) = \frac{|b-a|}{\sqrt{12}}$$

Uniforme verdeling H3⁺.2, 6⁺.1

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elders} \end{cases}$$

$$\mathrm{F}(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$



Berekening:

$$E(\underline{x}) = \int_{-\infty}^{\infty} x f(x) dx = \int_{a}^{b} \frac{x}{b-a} dx = \left[\frac{1}{2}x^2\right]_{x=a}^{x=b} = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Operaties met continue kansvariabelen H4.4, 4.5

Operatie	$\underline{x}$	$\mu(\underline{x})$	$\mathrm{Var}(\underline{x})$	$\sigma(\underline{x})$
Vermenigvuldigen met vast getal α	$\alpha\underline{x}$	$\alpha\mu(\underline{x})$	$\alpha^2\mathrm{Var}(\underline{x})$	$\lvert\alpha\rvert\sigma(\underline{x})$
Vast getal β erbij tellen	$\underline{x}+\beta$	$\mu(\underline{x})+\beta$	$\mathrm{Var}(\underline{x})$	$\sigma(\underline{x})$
Twee onafhankelijke stochasten optellen	$\underline{x}+\underline{y}$ als $\underline{x}, \underline{y}$ onafhankelijk	$\mu(\underline{x})+\mu(\underline{y})$	$\mathrm{Var}(\underline{x})+\mathrm{Var}(\underline{y})$	$\sqrt{\sigma(\underline{x})^2+\sigma(\underline{y})^2}$
Twee onafhankelijke stochasten aftrekken	$\underline{x}-\underline{y}$ als $\underline{x}, \underline{y}$ onafhankelijk	$\mu(\underline{x})-\mu(\underline{y})$	$\mathrm{Var}(\underline{x})+\mathrm{Var}(\underline{y})$	$\sqrt{\sigma(\underline{x})^2+\sigma(\underline{y})^2}$
Som van resultaten van herhaald experiment (zelfde stochast)	$\underline{x}_1+\underline{x}_2+\cdots+\underline{x}_n$	$n\cdot\mu(\underline{x})$	$n\cdot\mathrm{Var}(\underline{x})$	$\sqrt{n}\sigma(\underline{x})$
Gemiddelde van resultaten van herhaald experiment	$\frac{\underline{x}_1+\underline{x}_2+\cdots+\underline{x}_n}{n}$	$\mu(\underline{x})$	$\frac{1}{n}\mathrm{Var}(\underline{x})$	$\frac{1}{\sqrt{n}}\sigma(\underline{x})$

Stochasten $\underline{x}$ en $\underline{y}$ zijn onafhankelijk als voor de kansdichtheidsfuncties geldt: $f(x,y)=g(x)\cdot h(y)$

Let op: $\underline{x}_1+\underline{x}_2+\cdots+\underline{x}_n$ is niet hetzelfde als $n\underline{x}$, want een stochast $\underline{x}$ kan bij elke aanroep een andere waarde opleveren.