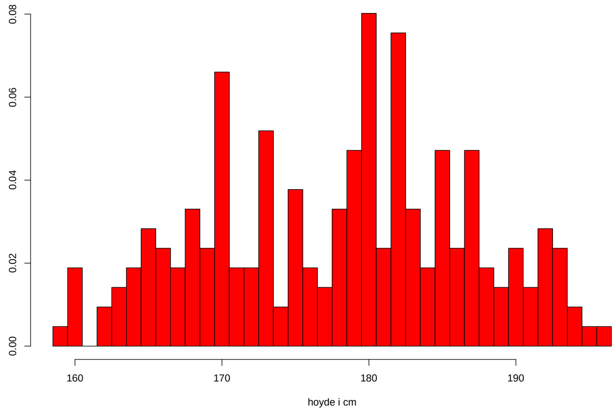
Kapittel 3: Stokastiske variable og sannsynlighetsfordelinger

TMA4245 Statistikk (B, K1, I)

3.1, 3.2, 3.3 foreleses torsdag 15.januar



Eksempel: kast med to terninger

- I et eksperiment kaster vi to terninger og registerer antall øyne på hver terning.
- Utfallsrom S={(1,1),(1,2),(1,3),...,(1,6), (2,1),...,(2,6),...,(6,6)}
- Hva hvis vi nå kun er interessert i summen av øynene på de to terningene?

		Første terning					
		1	2	3	4	5	6
	1	1,1	1,2	1,3	1,4	1,5	1,6
	2	2,1	2,2	2,3	2,4	2,5	2,6
Andre	3	3,1	3,2	3,3	3,4	3,5	3,6
terning	4	4,1	4,2	4,3	4,4	4,5	4,6
	5	5,1	5,2	5,3	5,4	5,5	5,6
	6	6,1	6,2	6,3	6,4	6,5	6,6

Sum to terninger forts.

- Med utfallet (1,1) assosierer vi 2
- Med utfallene (1,2) og (2,1) assosierer vi 3
- Med utfallene (1,3),(2,2),(3,1) assosierer vi 4
- . . .
- Med utfallet (6,6) assosierer vi 12.

		Første terning					
		1	2	3	4	5	6
	1	1,1	1,2	1,3	1,4	1,5	1,6
	2	2,1	2,2	2,3	2,4	2,5	2,6
Andre	3	3,1	3,2	3,3	3,4	3,5	3,6
terning	4	4,1	4,2	4,3	4,4	4,5	4,6
	5	5,1	5,2	5,3	5,4	5,5	5,6
	6	6,1	6,2	6,3	6,4	6,5	6,6

Sum to terninger forts.

- Situasjonen karakteriseres ved
  - vi er ikke interessert i selve enkeltutfallet
  - vi er kun interessert i tallverdien som assosieres til utfallet.
- Vi lar X være summen av to terninger, dette er et eksempel på en stokastisk variabel.

		Første terning					
		1	2	3	4	5	6
	1	1,1	1,2	1,3	1,4	1,5	1,6
	2	2,1	2,2	2,3	2,4	2,5	2,6
Andre	3	3,1	3,2	3,3	3,4	3,5	3,6
terning	4	4,1	4,2	4,3	4,4	4,5	4,6
	5	5,1	5,2	5,3	5,4	5,5	5,6
	6	6,1	6,2	6,3	6,4	6,5	6,6

3.1 Stokastisk variabel

DEF 3.1 Stokastisk variabel: En stokastisk variabel (SV) er en funksjon som assosierer et reelt tall med hvert element i utfallsrommet.

Engelsk: random variable.

Diskret stokastisk variabel

DEF 3.2 Diskret utfallrom: Hvis utfallsrommet inneholder et endelig antall mulige utfall eller en uendelig sekvens med så mange elementer som det er hele tall, så kalles et et diskret utfallsrom. En stokastisk variabel som har et diskret utfallsrom kalles en diskret stokastisk variabel.

- Sum av to terninger.
- .
- .
- .

Kontinuerlig stokastisk variabel

DEF 3.3 Kontinuerlig utfallrom: Hvis utfallsrommet inneholder et uendelig antall mulige utfall (f.eks. reelle tall) så kalles det et kontinuerlig utfallsrom. En stokastisk variabel som har et kontinuerlig utfallsrom kalles en kontinuerlig stokastisk variabel.

- Lengde av diskoskast.
- Svovelinnholdet på et sted i en sølvåre.
- Levetiden til en lyspære.
- .
- .

To terninger, igjen ...

- Vi ønsker å si noe om sannsynligheten til X="summen av to terninger", men vi har til nå kun definert sannsynlighet for hendelser.

$$
\begin{aligned}
A_2 &= \{e \in S | X(e) = 2\} \\
A_3 &= \{e \in S | X(e) = 3\} \\
A_4 &= \{e \in S | X(e) = 4\} \\
A_5 &= \{e \in S | X(e) = 5\} \\
\cdots & \quad \cdots \\
A_{11} &= \{e \in S | X(e) = 11\} \\
A_{12} &= \{e \in S | X(e) = 12\}
\end{aligned}
$$

To terninger forts.

- $P(X = 2) = P(A_2) =$
- $P(X = 3) = P(A_3) =$
- $P(X = 4) = P(A_4) =$
- $\ldots$
- $P(X = 11) = P(A_{11}) =$
- $P(X = 12) = P(A_{12}) =$

		Første terning					
		1	2	3	4	5	6
	1	1,1	1,2	1,3	1,4	1,5	1,6
	2	2,1	2,2	2,3	2,4	2,5	2,6
Andre	3	3,1	3,2	3,3	3,4	3,5	3,6
terning	4	4,1	4,2	4,3	4,4	4,5	4,6
	5	5,1	5,2	5,3	5,4	5,5	5,6
	6	6,1	6,2	6,3	6,4	6,5	6,6

To terninger forts.

- Presentere i tabell:

x	2	3	4	5	6	7
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$

x	8	9	10	11	12
$P(X=x)$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3.2 Diskrete sannsynligetsfordelinger

DEF 3.4: Paret $(x, f(x))$ er sannsynlighetsfordelingen til den diskrete stokastiske variabelen X, dersom for alle mulige utfall x:

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x)$ (punktsannsynlighet)

Notasjon

- Stor bokstav X angir den stokastiske variabelen.
- Liten bokstav x angir numerisk verdi den stokastiske variabelen kan anta.
- Hver verdi av den stokastiske variabelen representerer en (eller flere) hendelse/r.

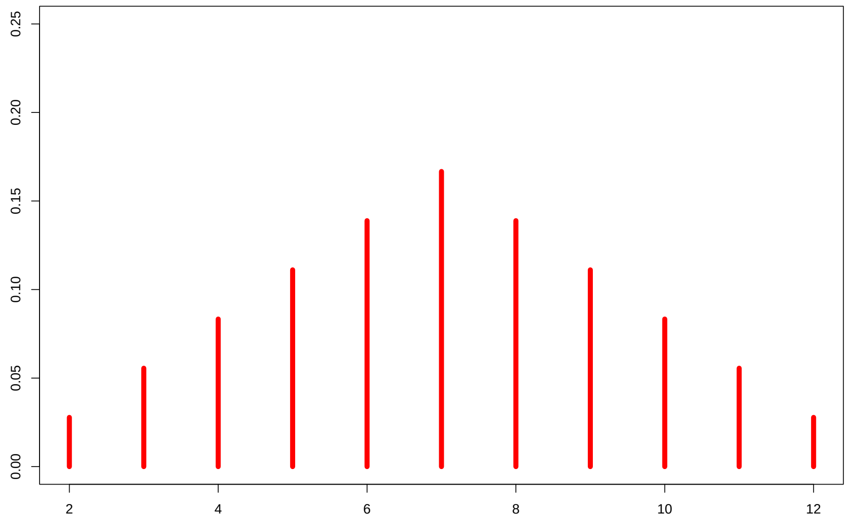
Sum to terninger og DEF 3.4

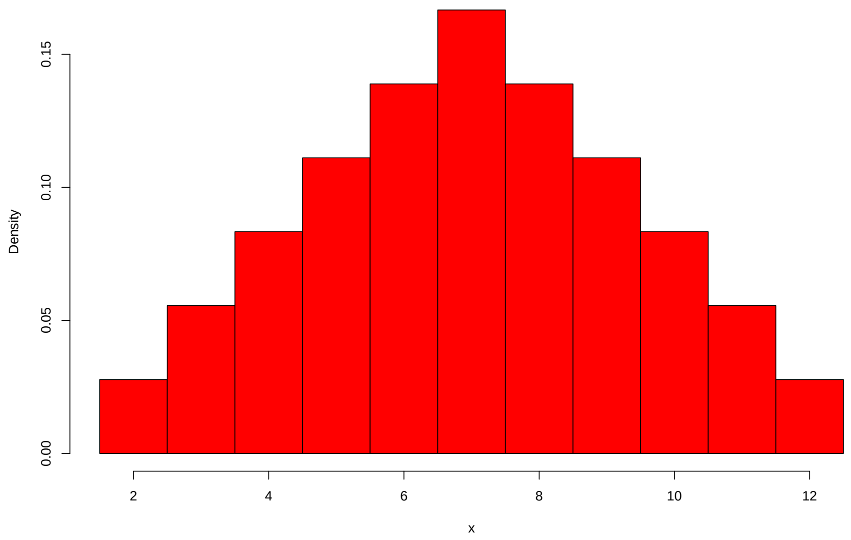
- $f(x)$
- $\sum_x f(x)$
- $P(X = x) = f(x)$ (punktsannsynlighet)

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Visualisering

- Stolpediagram og sannsynlighetshistogram





Kumulativ fordeling, diskret

DEF 3.5: Den kumulative fordelingen $F(x)$ til en diskret stokastisk variabel X med sannsynlighetsfordeling $f(x)$ er

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

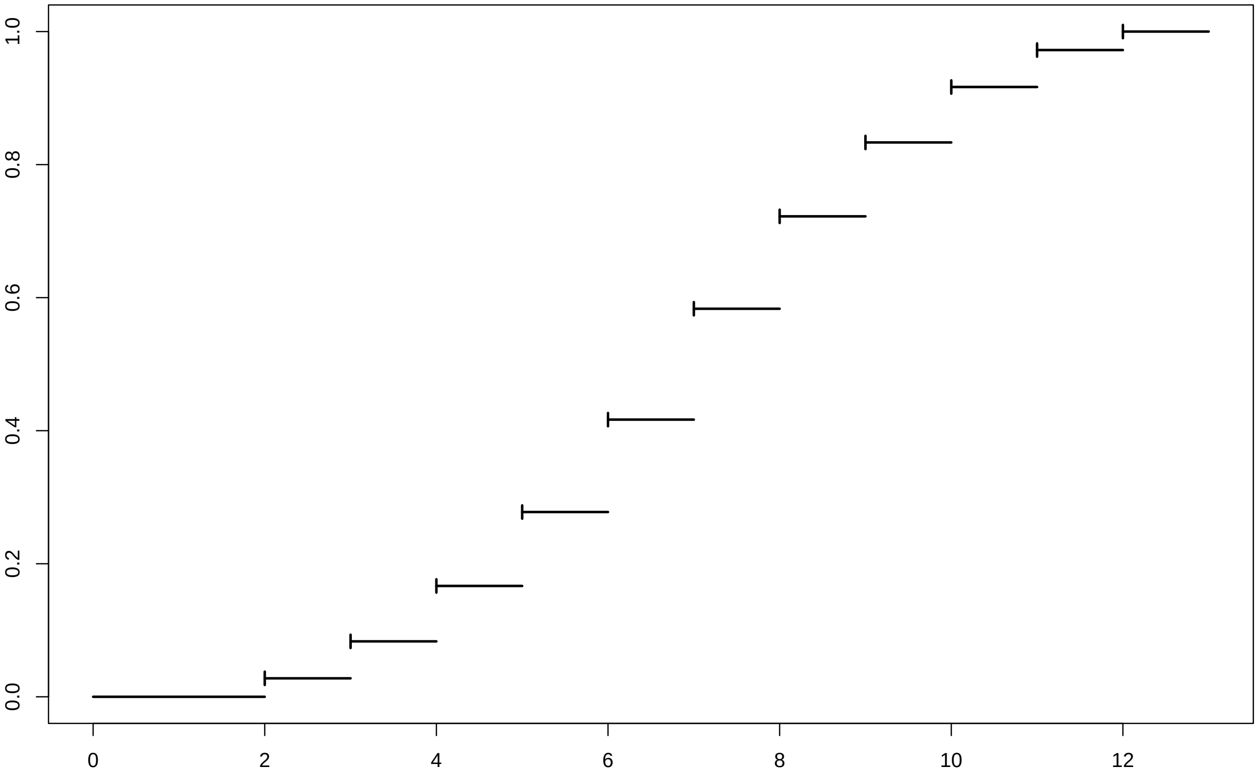
for $-\infty < x < \infty$.

- Definert for alle verdier x, ikke bare de mulige verdiene til den diskrete stokastiske variabelen.

Kumulativ fordeling: sum to terninger

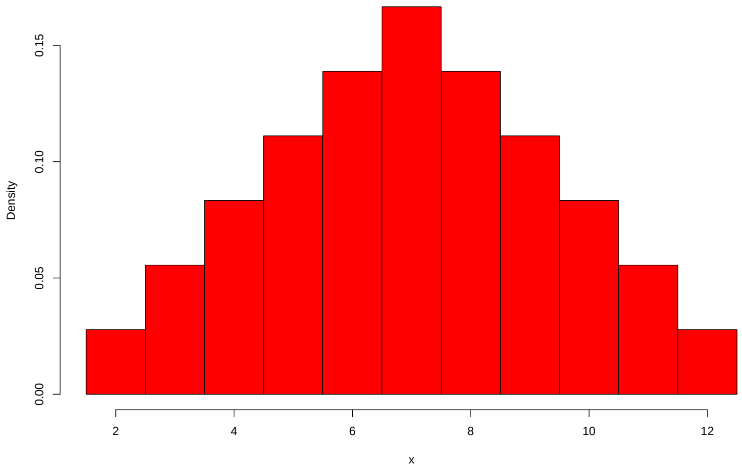
$$
\begin{array}{ll}
x \le 1 & F(X \le 1) = 0 \\
x \le 2 & F(X \le 2) = f(2) \\
x \le 3 & F(X \le 3) = f(2) + f(3) \\
x \le 4 & F(X \le 4) = f(2) + f(3) + f(4) \\
\dots & \dots \\
x \le 10 & F(X \le 10) = f(2) + f(3) + \cdots + f(10) \\
x \le 11 & F(X \le 11) = f(2) + f(3) + \cdots + f(11) \\
x \le 12 & F(X \le 12) = 1 \\
x \le 13 & F(X \le 13) = 1
\end{array}
$$

Kumulativ fordeling: sum to terninger

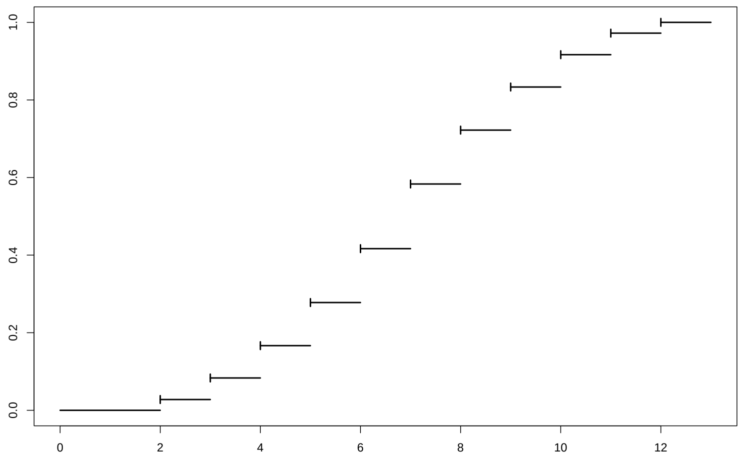


$$P(a < X \leq b)$$

Punktsannsynlighet:

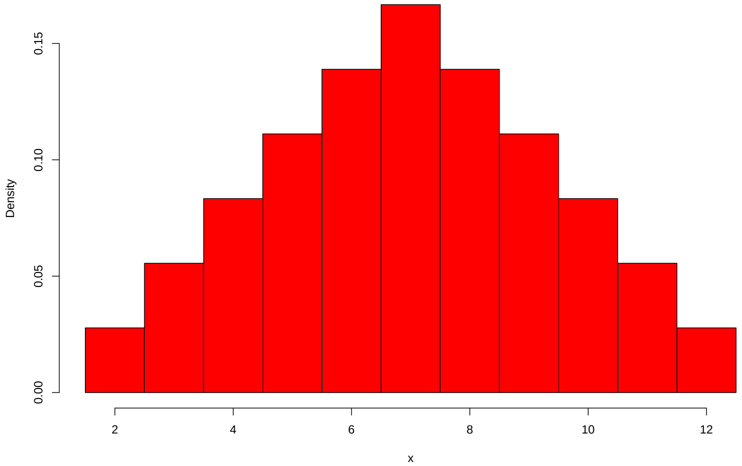


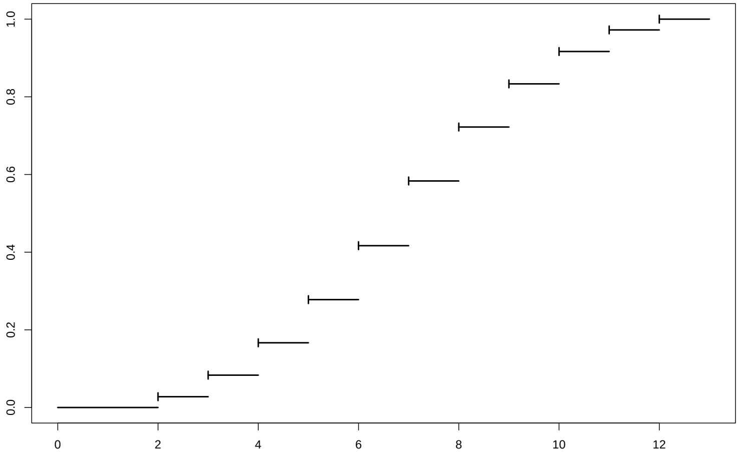
Kumulativ fordeling:



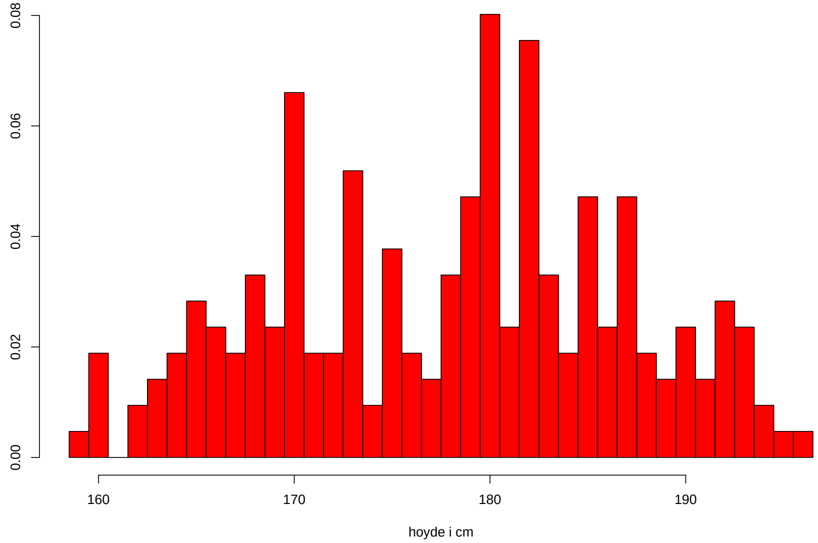
$$P(a < X \leq b) = F(b) - F(a)$$

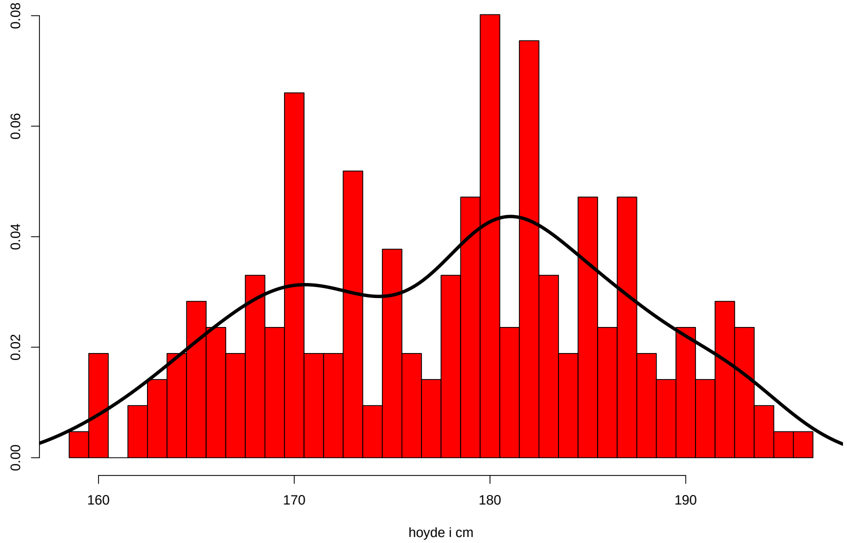
$$\begin{aligned} P(a < X \leq b) &= 1 - P[(X \leq a) \cup (X > b)] \\ &= 1 - [P(X \leq a) + P(X > b)] \\ &= [1 - P(X > b)] - P(X \leq a) \\ &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$





Høyden til TMA4245 studenter

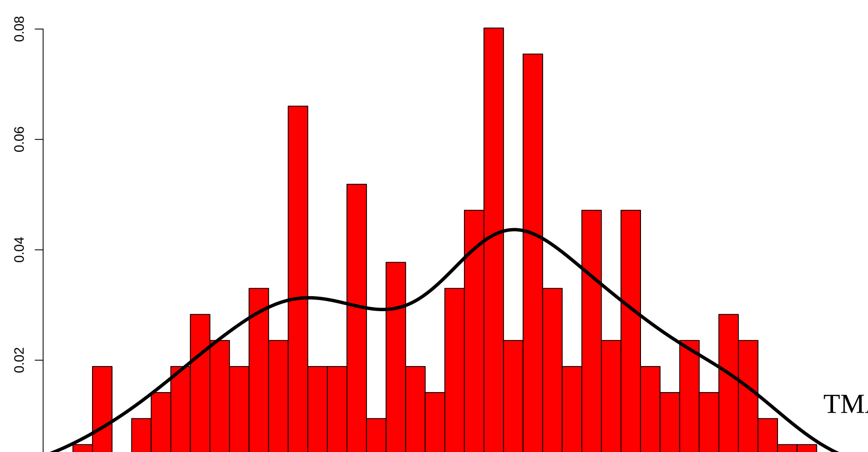




3.3 Kontinuerlige sannsynlighetsfordelinger

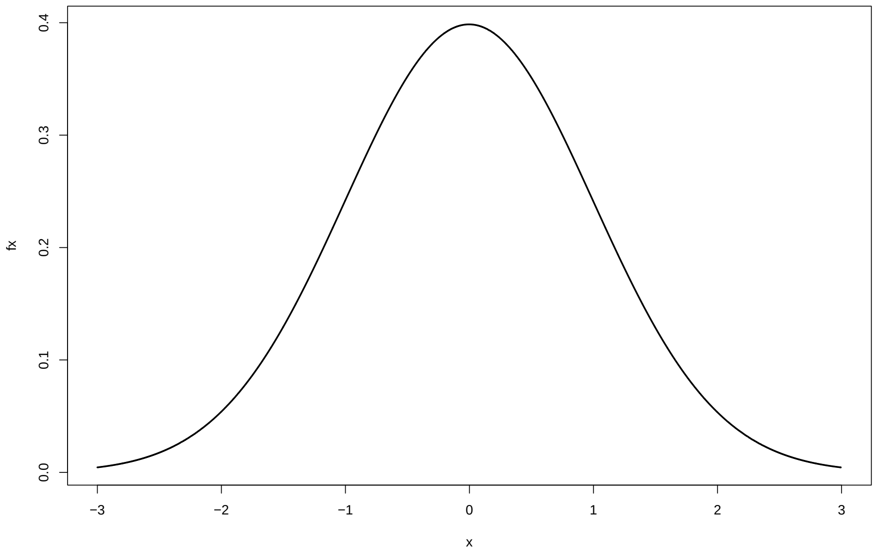
DEF 3.6: Funksjonen $f(x)$ er sannsynlighetstettheten til den kontinuerlige stokastiske variabelen X, dersom

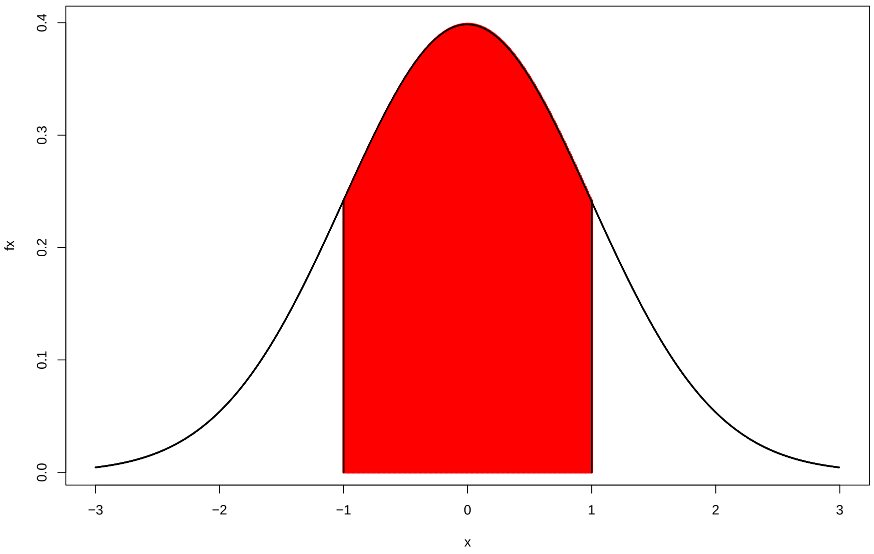
1. $f(x) \geq 0$ for alle $x \in R$ (reelle tall)
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$.



Tolkning av sannsynlighetstetthet

1. $f(x) \geq 0$ for alle $x \in R$ (reelle tall)
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$.



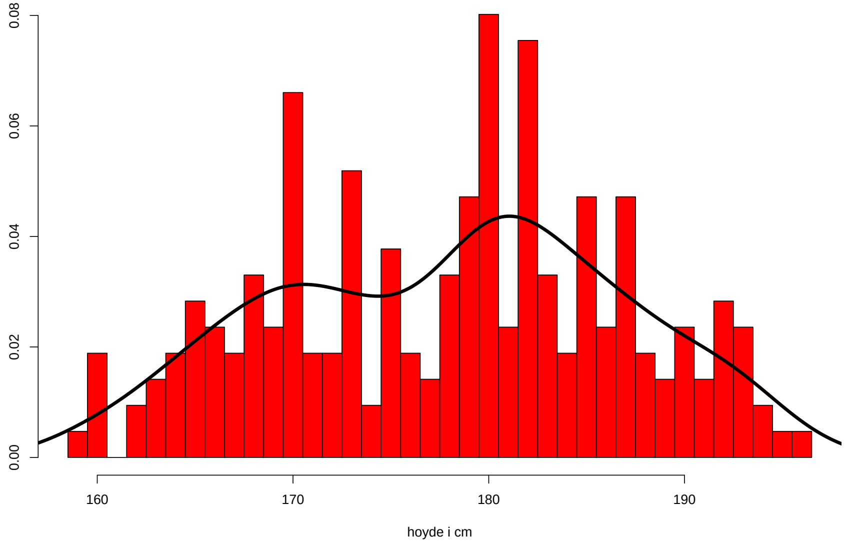


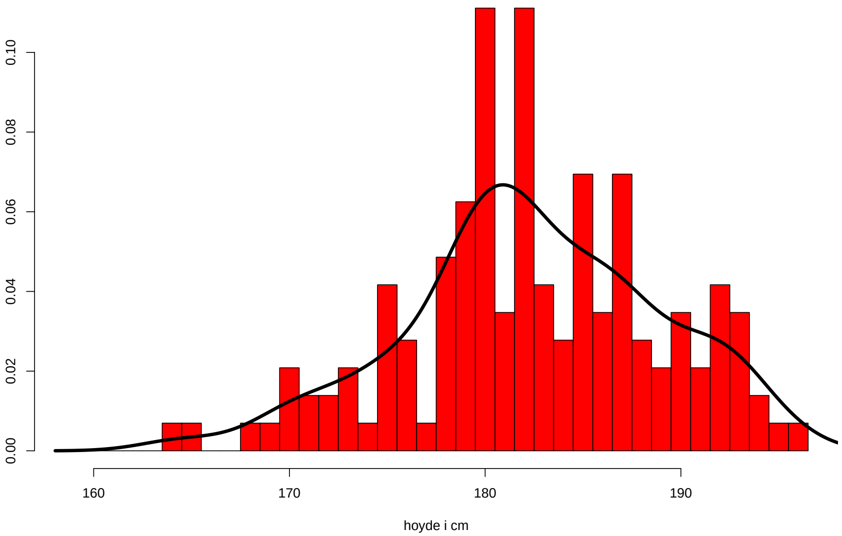
Tetthet og punktsannsynlighet

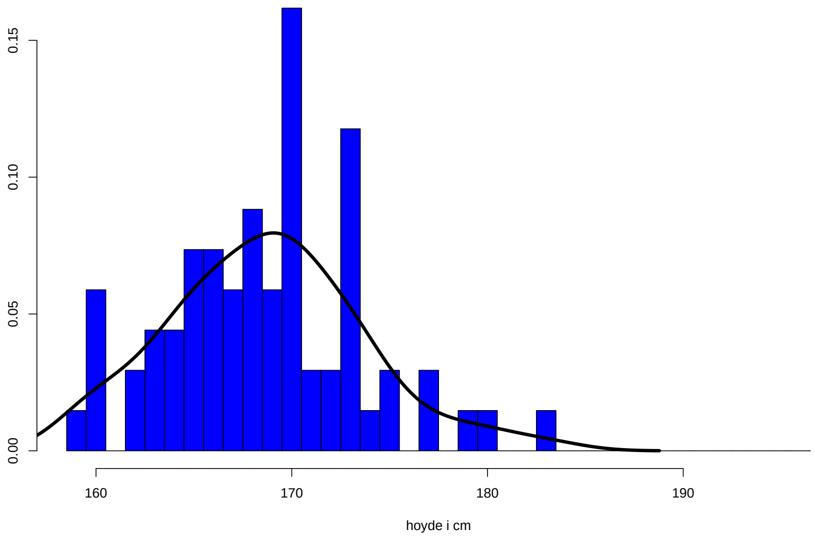
- Ser på et lite område $[x, x+\varepsilon]$
  - $P(x < X \leq x+\varepsilon) \approx f(x) \cdot \varepsilon$
  - Hva hvis $\varepsilon \to 0$?

- For en kontinuerlig stokastisk variabel har vi at $P(X = x) = 0$.
  - Sannsynligheten for et spesielt punkt er 0, og vi kan ikke bruke punktsannsynligheter til å beskrive kontinuerlige stokastiske variable.
  - Dermed $P(a < X \leq b) = P[(a < X < b) \cup (X = b)]$
    $= P(a < X < b) + P(X = b) = P(a < X < b)$

Noen tettheter







Kumulativ fordeling, kontinuerlig

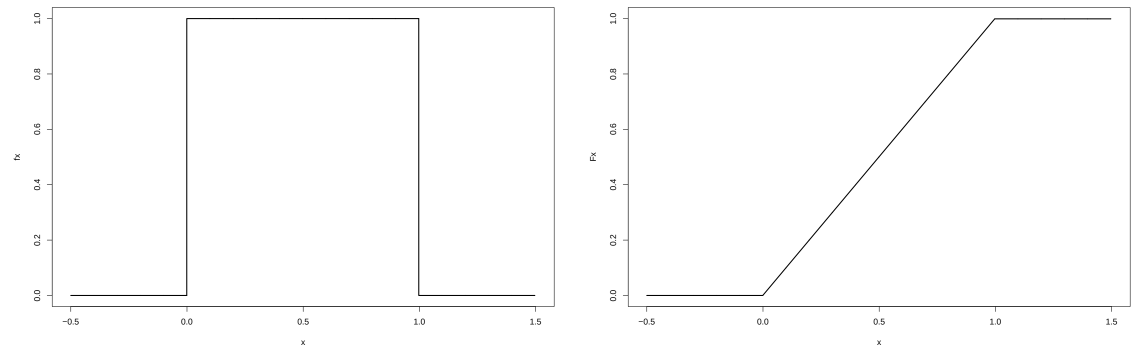
DEF 3.7: Den kumulative fordelingen $F(x)$ til en kontinuerlig stokastisk variabel X med sannsynlighetstetthet $f(x)$ er

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

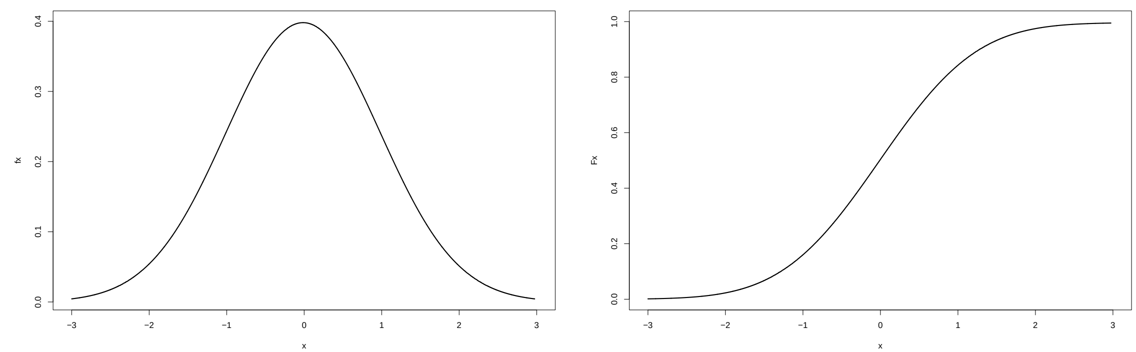
for $-\infty < x < \infty$.

Fordelinger (kapittel 6)

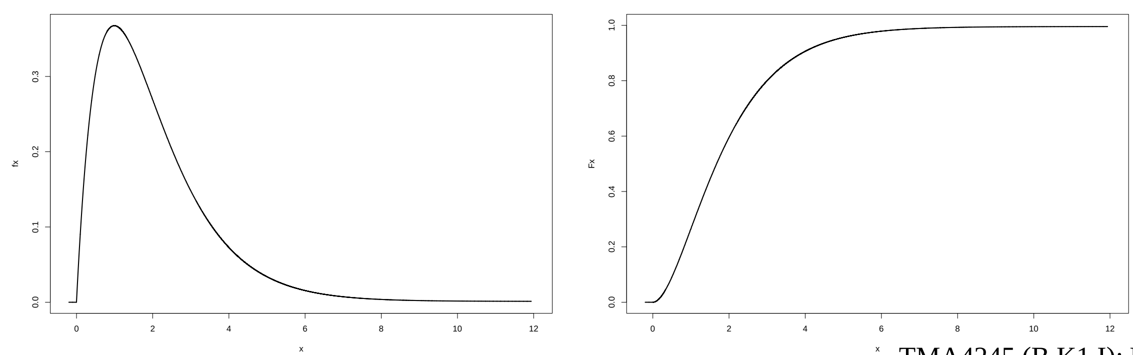
Uniform



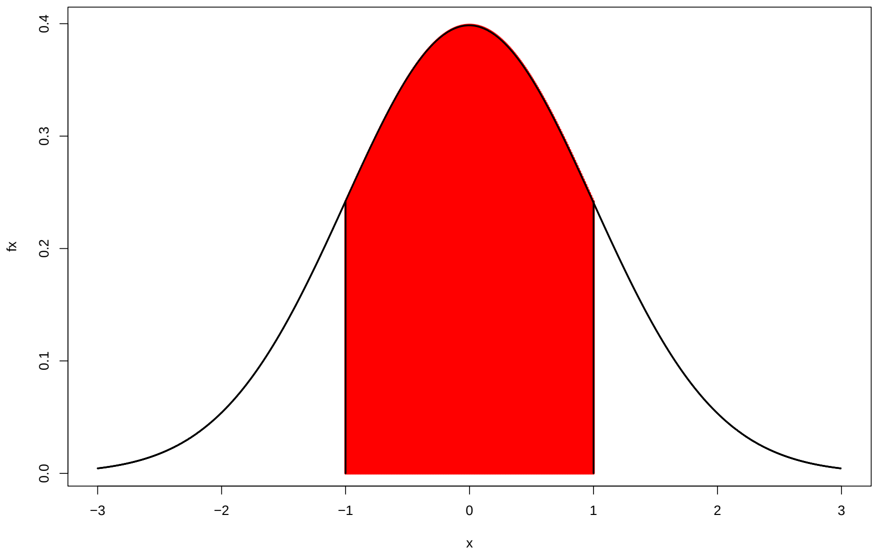
Normal

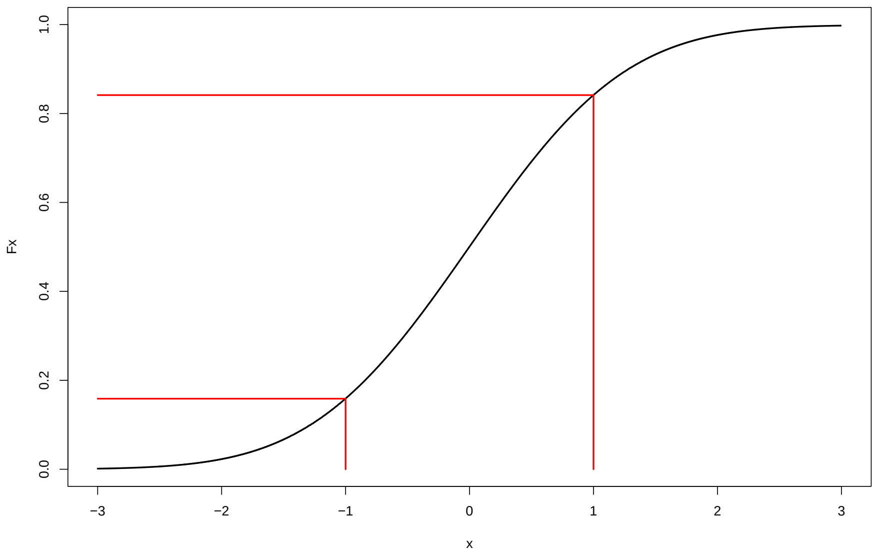


Gamma



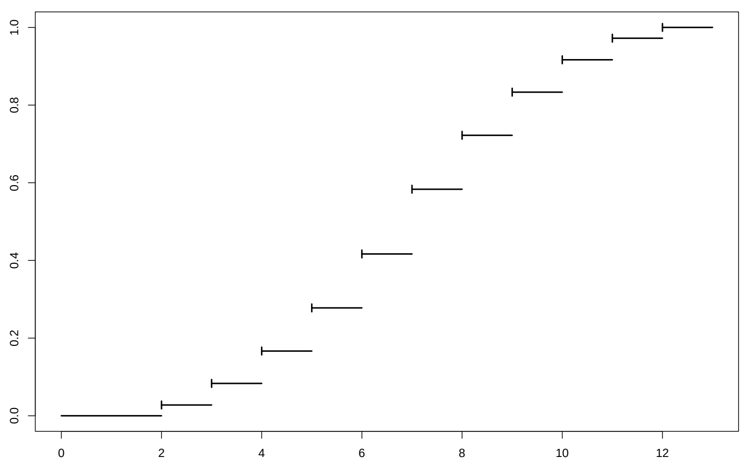
$$P(a < X \leq b)$$

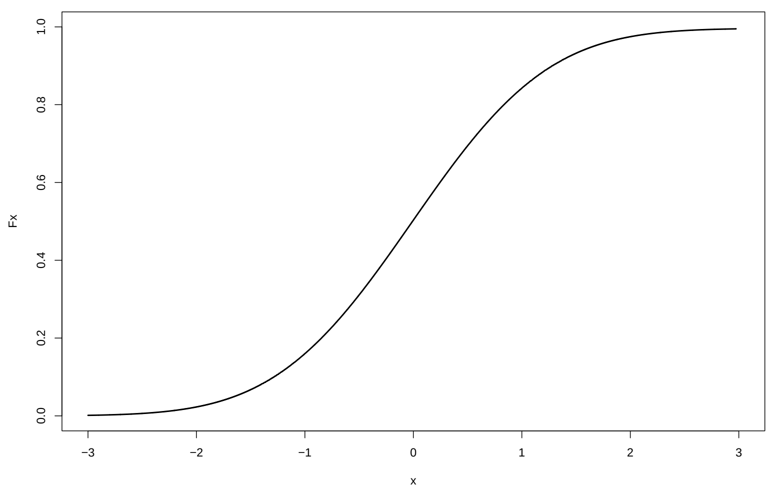




Egenskaper til kumulativ fordeling

- Gjelder for både diskrete og kontinuerlige fordelinger!
- F(x) er voksende
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $P(a < X \leq b) = F(b) - F(a)$





Kontinuerlige sammenhenger

$$\int_{-\infty}^{x} f(t)dt$$

$f(x)$ $F(x)$

$$\frac{dF(x)}{dx}$$

Oppsummering

Diskret stokastisk variabel	**Kontinuerlig stokastisk variabel**
X Mulige verdier x: Endelig eller tellbart mange	X Mulige verdier x: Intervall eller hele R
Eksempel: $\{0, 1, \ldots, n\}$	Eksempel: $[0, 1]$ eller $[0, \infty)$
Sannsynlighetsfordeling: $f(x) = P(X = x)$ for alle mulige x	Sannsynlighetsfordeling: $f(x)$ definert for alle reelle x ved $P(a < X < b) = \int_a^b f(x)dx$
Kumulativ fordeling: $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ definert for alle reelle x	Kumulativ fordeling: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$ definert for alle reelle x
Hvis mulige verdier er heltall: $f(x) = F(x) - F(x-1)$	Hvis f er kontinuerlig i x: $f(x) = F'(x)$

Togforsinkelsen

- Deler av oppgave 1, eksamen desember 2003.