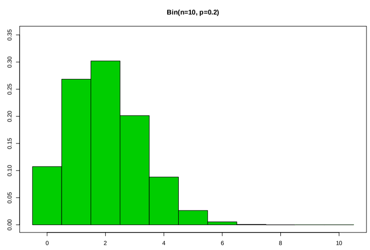
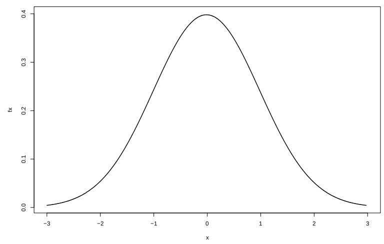
Kapittel 3: Stokastiske variable og sannsynlighetsfordelinger

TMA4240 Statistikk (F2 og E7)

3.1, 3.2, 3.3 foreleses onsdag 25.august, 2004



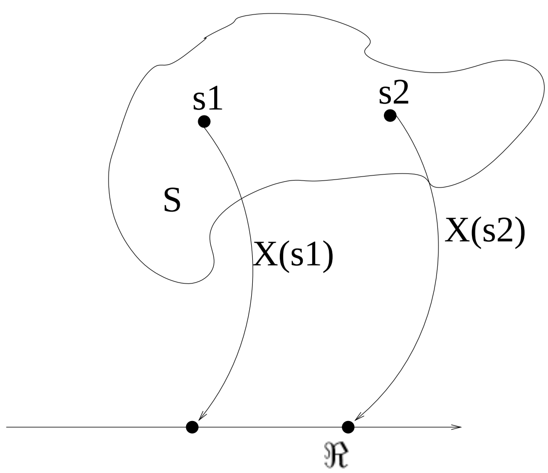


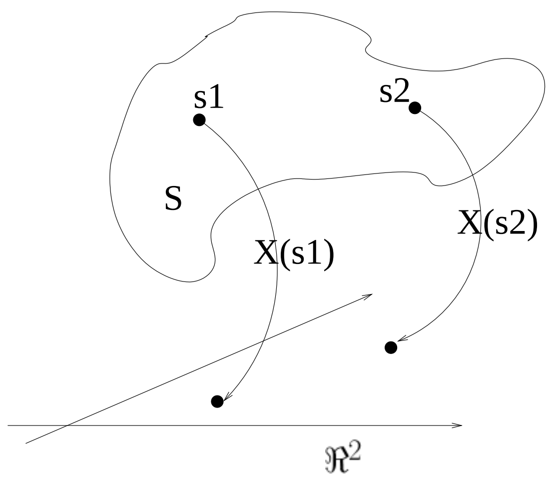
Ole.Petter.Lodoen@math.ntnu.no

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.





3.2 Diskrete sannsynligetsfordelinger

DEF 3.4: $f(x)$ 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)

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 $-\infty < x < \infty$.

Eks. kast med to terninger

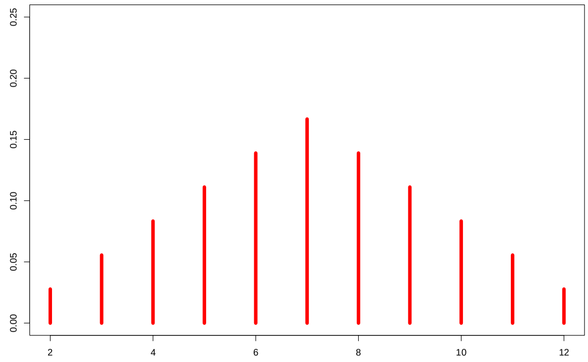
- Situasjonen karakteriseres ved
  - vi er ikke interessert i selve enkeltutfallet
  - vi er kun interessert i tallverdien som assosieres til utfallet.
- Vi lar X være summen av to terninger, dette er et eksempel på en stokastisk variabel.

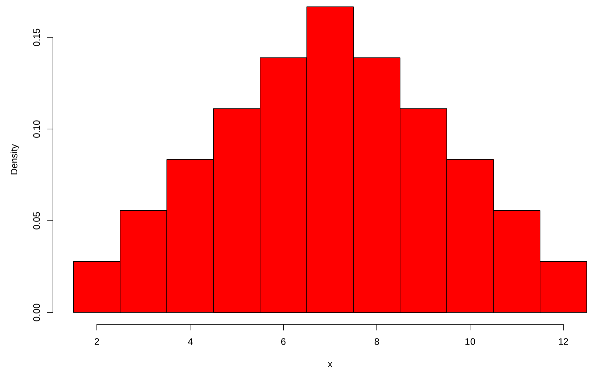
		Første terning					
		1	2	3	4	5	6
	1	1,1	1,2	1,3	1,4	1,5	1,6
	2	2,1	2,2	2,3	2,4	2,5	2,6
Andre	3	3,1	3,2	3,3	3,4	3,5	3,6
terning	4	4,1	4,2	4,3	4,4	4,5	4,6
	5	5,1	5,2	5,3	5,4	5,5	5,6
	6	6,1	6,2	6,3	6,4	6,5	6,6

x	1	2	3	4	5	6	7	8	9	10	11	12	13
$f(x)$	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0

$f(x)$ grafisk

- Stolpediagram og sannsynlighetshistogram

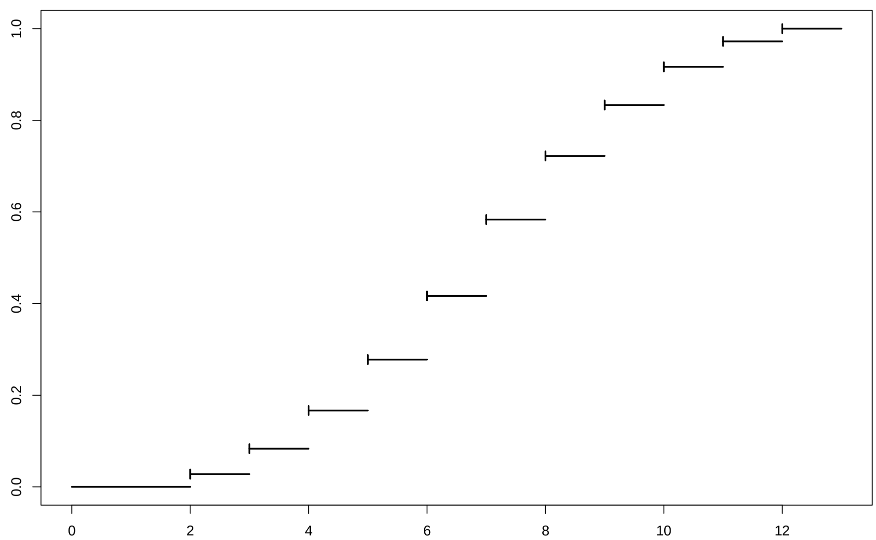




x	1	2	3	4	5	6	7	8	9	10	11	12	13
$f(x)$	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0

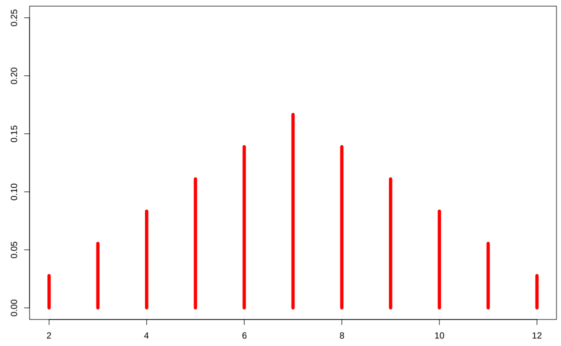
$F(x)$ grafisk

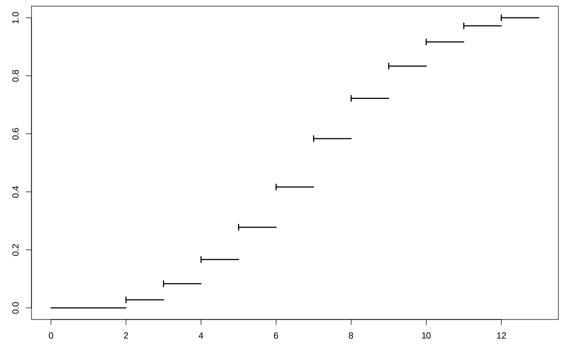
x	1	2	3	4	5	6	7	8	9	10	11	12	13
$f(x)$	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0
$F(x)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1	1



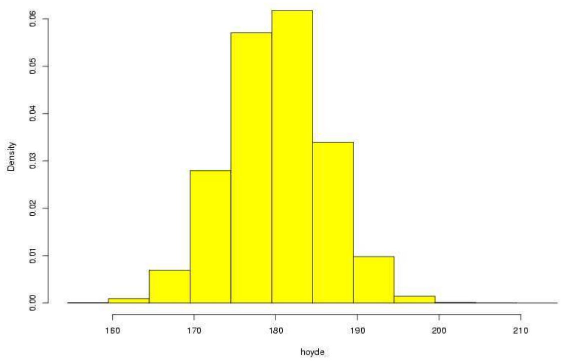
Diskret fordeling

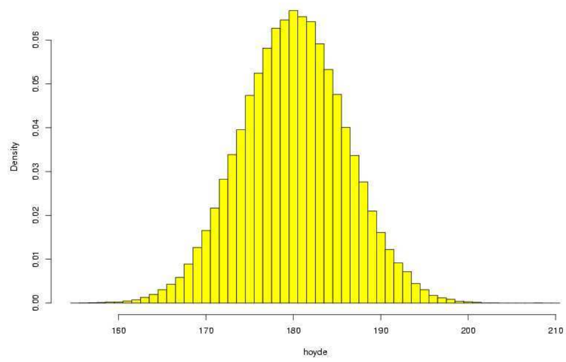
Sannsynlighetsfordeling, $f(x)$, og kumulativ fordeling, $F(x)$.

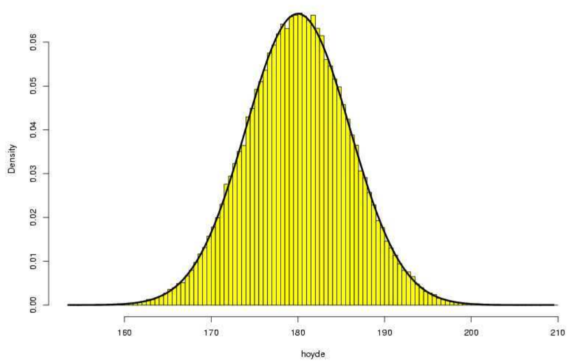




Høyde til mannlige studenter







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. $P(a < X < b) = \int_a^b f(x)dx$.

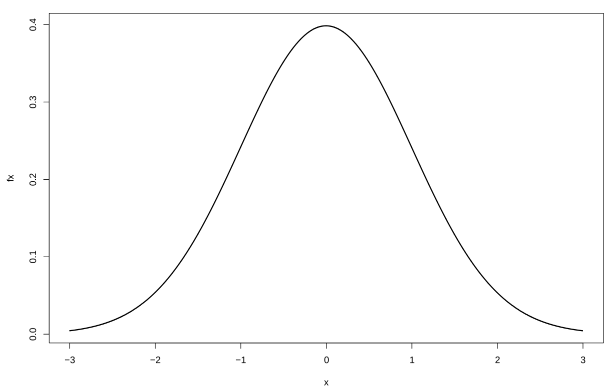
DEF 3.7: Den kumulative fordelingen $F(x)$ til en kontinuerlig stokastisk variabel X med sannsynlighetstetthet $f(x)$ er

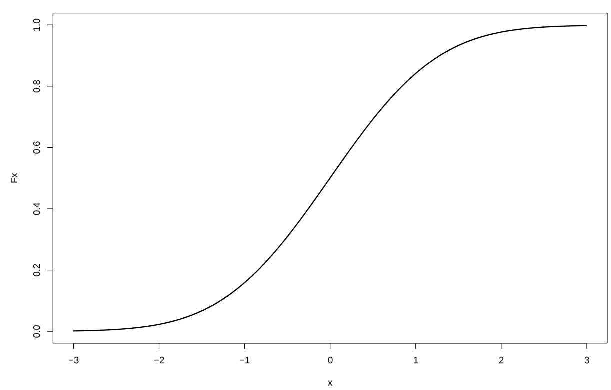
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

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

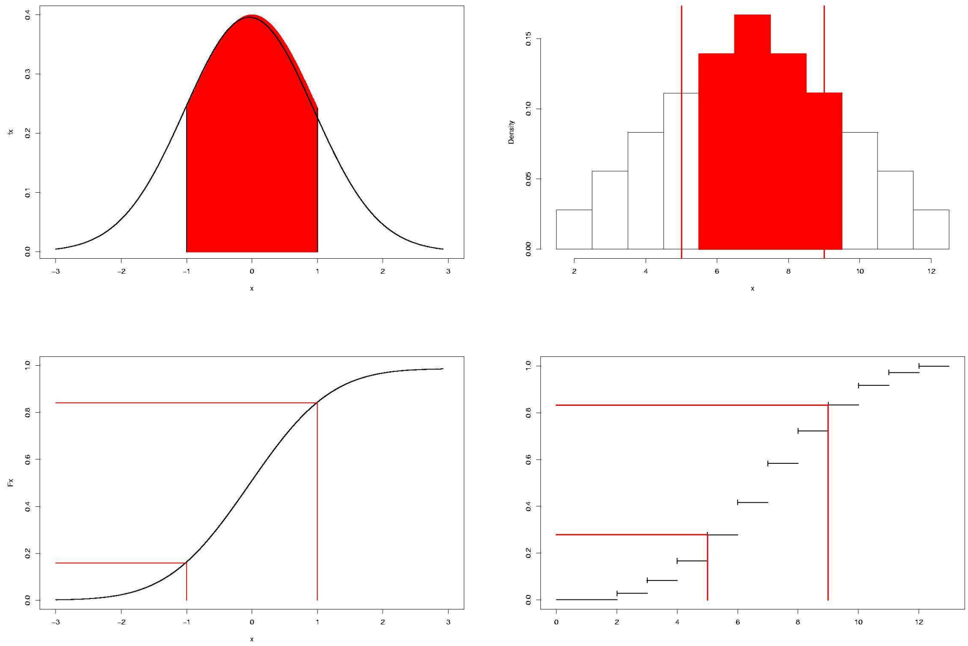
Kontinuerlig fordeling

Sannsynlighetsfordeling, $f(x)$, og kumulativ fordeling, $F(x)$.





$P(a < X \leq b)$



Togforsinkelsen

- Deler av oppgave 1, eksamen desember 2003.
  - I denne oppgaven kan du bruke uten å vise det at

$$\int_0^\infty x^r e^{-ax} dx = \frac{r!}{a^{r+1}} \text{ når } a > 0 \text{ og } r \text{ er et heltall} \geq 0$$

  - Vi betrakter ankomst- og oppholdstider for et bestemt lokaltog på en jernbanestasjon. Toget skal etter rutetabellen ankomme hver hverdag klokka 8:00, men kommer alltid etter dette tidspunktet.
  - La X (minutter) betegne togets forsinkelse på en tilfeldig valgt hverdag. Vi antar at X er en stokastisk variabel med sannsynlighetstetthet

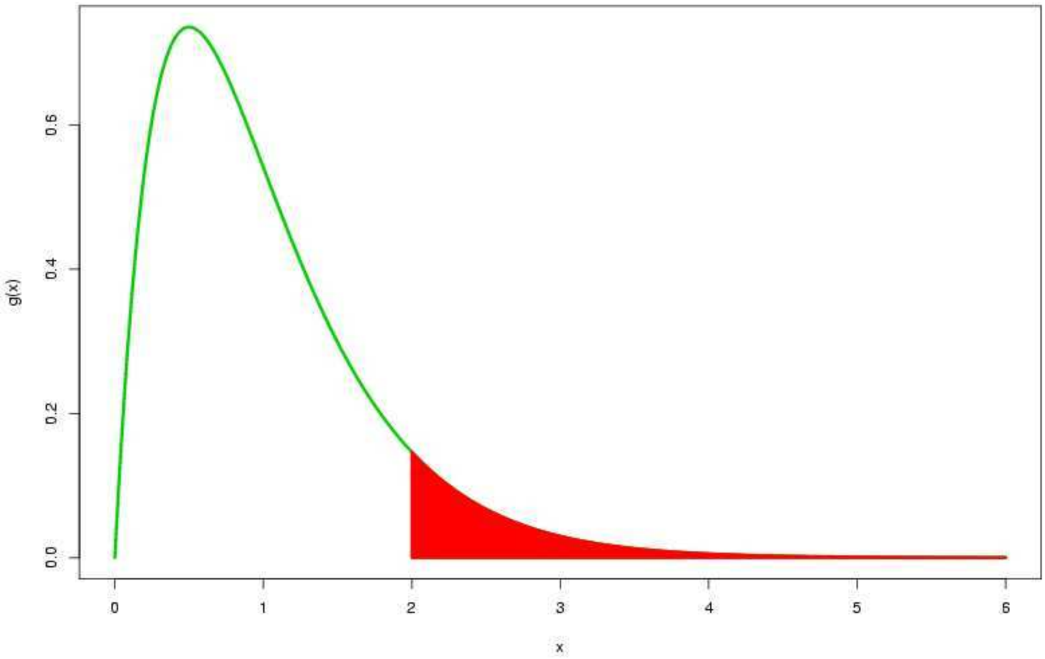
$$g(x) = \begin{cases} kxe^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

der $k > 0$ er en konstant.

i) Vis at $k = 4$.

ii) Vis at sannsynligheten for at toget er mer enn 2 minutter forsinket er tilnærmet lik 0.09.

$g(x)$ og $P(X > 2)$



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
$P(a < X \leq b) = \sum_{x \in (a,b]} f(t)$ $= F(b) - F(a)$	$P(a < X \leq b) = \int_a^b f(x)dx$ $= F(b) - F(a)$
Hvis mulige verdier er heltall: $f(x) = F(x) - F(x-1)$	Hvis f er kontinuerlig i x: $f(x) = F'(x)$