# Kap 3: Stokastiske variable og sannsynsfordelingar

- Stokastisk variabel: **Forrige tirsdag**
- Diskret sannsynsfordeling: **Forrige tirsdag**
- Kontinuerleg sannsynsfordeling:**I går**
- Kummulativ sannsynsfordeling:**I går**
- Diskret simultanfordeling **I går**
- Kontinuerleg simultanfordeling **I dag**
- Marginal sannsynsfordeling **I går og i dag**
- Betinga sannsynsfordeling **I går og i dag**
- Statistisk uavhengig **I dag**

## Stokastisk variabel

Ein *stokastisk variabel er ein funksjon som assosierer eit reelt tal med kvart element i utfallsrommet.* $X : S \rightarrow \Re$

**Notasjon:**

- $X$, $Y$: Store bokstavar på stokastiske variable
- $x$, $y$: Små bokstavar på utfall (datapunkt)

## Diskret stokastisk variabel:

Dersom den stokastiske variabelen gjev telbart antall utfall.

## Kontinuerleg stokastisk variabel

Dersom den kontinuerlege variabelen har utfall på kontinuerleg skala.

# Sannsynsfordeling

## Diskret

Paret $(x, f(x))$ blir kalla sannsynsfordelinga til den diskret stok. var. $X$ dersom

- $0 \leq f(x)$
- $\sum_{\forall x} f(x) = 1$ (summen over alle mogelege $x$)
- $f(x) = P(X = x)$

## Kontinuerleg

Funksjonen $f(x)$ definert for alle reelle tal $x \in \Re$ blir kalla sannsynsfordelinga til den kontinuerlege stok. var. $X$ dersom

- $0 \leq f(x)$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$

# Kummulativ sannsynsfordeling

## Definisjon

Den *kummulative fordelinlgsfunksjonen* $F(x)$ *for ein stok. var.* $X$ *er:*

$$F(x) = P(X \leq x)$$

- *Diskret:* $F(x) = \sum_{t \leq x} f(t)$
- *Kontinuerleg* $F(x) = \int_{-\infty}^{x} f(t)dt$

## Simultan sannsynsfordeling diskret stok. var. (Def.3.8)

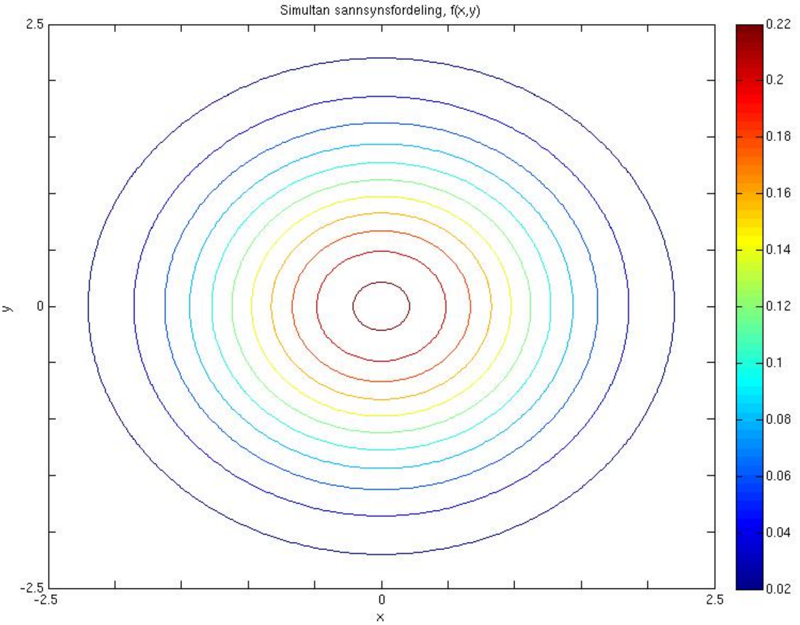
Funksjonen $f(x, y)$ er ei simultan sannsynsfordeling for dei diskret stokastiske variablane $X$ og $Y$ dersom

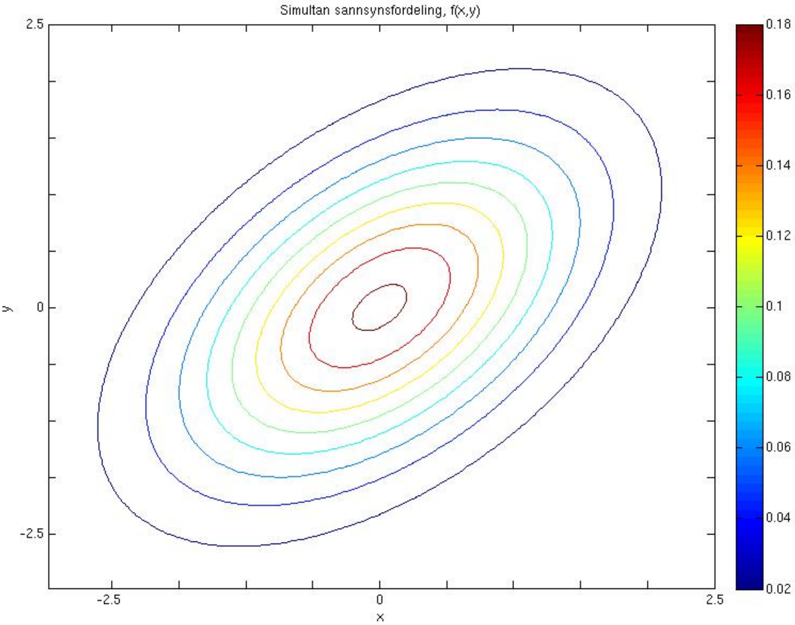
1. $f(x, y) \geq 0$ for alle $(x, y)$
2. $\sum_{\forall x} \sum_{\forall y} f(x, y) = 1$
3. $P(X = x, Y = y) = f(x, y)$

## Simultan sannsynsfordeling kontinuerleg stok. var. (Def.3.9)

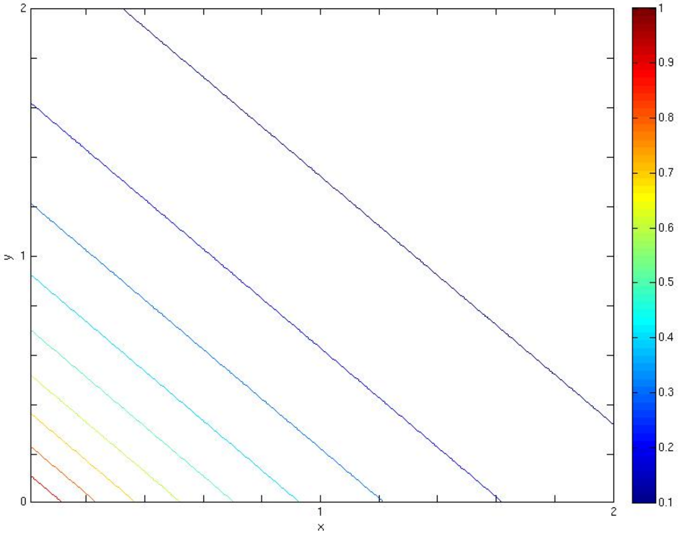
Funksjonen $f(x, y)$ er ei simultan sannsynsfordeling for dei kontinuerlege stokastiske variablane $X$ og $Y$ dersom

1. $f(x, y) \geq 0$ for alle $(x, y)$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) = 1$
3. $P((X, Y) \in A) = \int \int_A f(x, y) dxdy$


Simultan sannsynsfordeling, f(x,y)
2.5
0
-2.5
-2.5
0
2.5
x
y
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02


Simultan sannsynsfordeling, f(x,y)
2.5
0
-2.5
y
-2.5
0
2.5
x
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02

# Simultanfordeling ventilar eksempel



# Marginalfordeling

## Definisjon

Dersom $f(x, y)$ er simltanfordelinga til $(X, Y)$ er *marginalfordelingane til* $X$ *og* $Y$ *hhv:*

### for diskrete stok.var:

- $g(x) = \sum_{\forall y} f(x, y)$, *og*
- $h(y) = \sum_{\forall x} f(x, y)$.

### for kontinuerlege stok.var:

- $g(x) = \int_{-\infty}^{\infty} f(x, y)dy$, *og*
- $h(y) = \int_{-\infty}^{\infty} f(x, y)dx$.

# Betinga fordeling

## Definisjon 3.11

La $X$ og $Y$ vere to stokastiske variable (kont. eller diskrete) med simultan sannsynsfordelin $f(x, y)$ og marginal sannsynsfordeling hhv $g(x)$ og $h(y)$.
Den *betinga fordelinga til* $X$ *gjeve at* $Y = y$ *er*

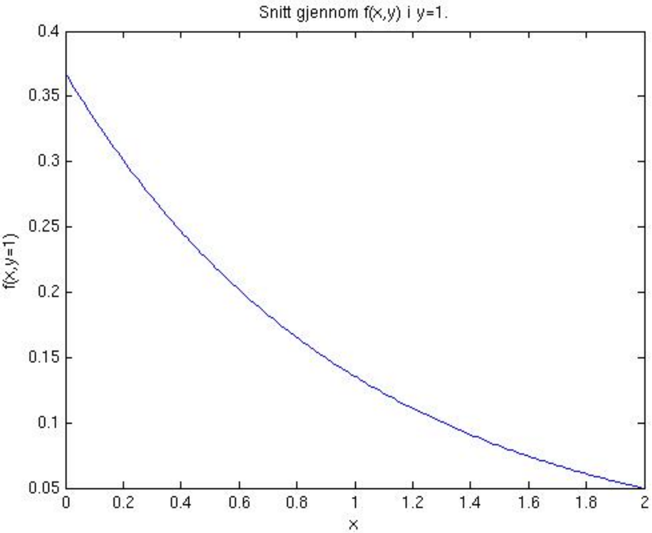
$$f(x|y) = f(x, y)/h(y)$$

*(*$h(y) > 0$*)*
*Tilsvarande er den* betinga fordelinga til $Y$ gjeve at $X = x$

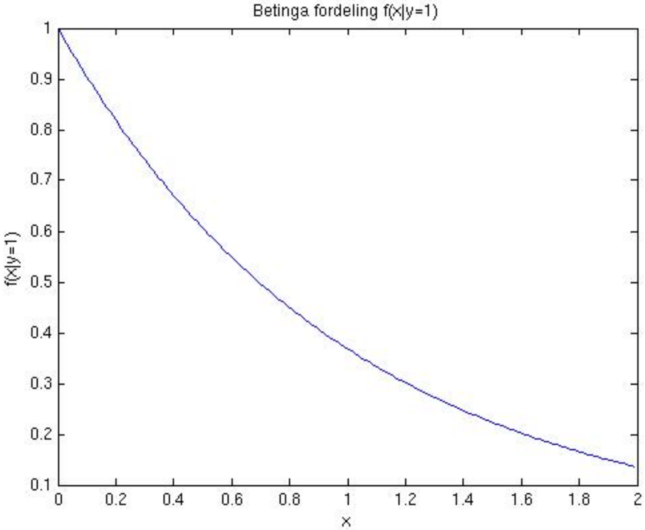
$$f(y|x) = f(x, y)/g(x)$$

$(g(x) > 0)$

# Snitt gjennom simultanfordeling i $Y = 1$



# To ventilar, betinga fordeling



# Statistisk uavhengig

**Definisjon 3.12**

La $X$ og $Y$ vere to stokastiske variable med simultan sannsynsfordelin $f(x, y)$ og marginal sannsynsfordeling hhv $g(x)$ og $h(y)$.
$X$ og $Y$ er *statistisk uavhengige dersom, og berre dersom*

$$f(x, y) = g(x) \cdot h(y)$$

# Generelt statistisk uavhengighet

**Definisjon 3.13**

$X_1, X_2, \ldots, X_n$ er *statistisk uavhengige dersom*

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n)$$

# Dagens tekst

Kap 3: Stokastiske variable og sannsynsfordelingar

- Stokastisk variabel: $X : S \to \Re$
- Diskret sannsynsfordeling: $f(x)$ slik at $P(X = x) = f(x)$
- Kontinuerleg sannsynsfordeling:
  $f(x)$ slik at $P(a < X < b) = \int_a^b f(x)$
- Kummulativ sannsynsfordeling: $F(x)$ slik at $P(X \leq b) = F(b)$
- Diskret simultanfordeling $P(X = x \cap Y = y) = f(x, y)$
- Kontinuerleg simultanfordeling
  $P((X, Y) \in A) = \int\int_A f(x, y)dxdy$
- Marginal sannsynsfordeling: For ein stok.var. aleine
- Betinga sannsynsfordeling $f(x|y) = f(x, y)/h(y)$
- Statistisk uavhengig dersom $f(x_1, x_2) = f_1(x_1)f_2(x_2)$

### Viktige omgrep

- Forventningsverdi
- Varians / standard avvik
- Kovarians / korrelasjon

### Viktige rekneregler

Forventning til lineærkombinasjon:

Varians til lineærkombinasjon:

## Definisjon 4.1

La $X$ vere ein stok.var. med sannsynsfordeling $f(x)$. Forventningsverdien til $X$ er då

$$\mu = E(X) = \sum_{\forall x} xf(x)$$

dersom $X$ er diskret, og

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

dersom $X$ er kontinuerleg.

# Forventningsverdi av funksjon av stokastisk variabel

**Teorem 4.1**

For ein funksjon g(X);

$$\mu_{g(X)} = E(g(X)) = \sum_x g(x)f(x)$$

for diskret $X$, og

$$\mu_{g(X)} = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

for kontinuerleg $X$.