

Kap 3: Stokastiske variable og sannsynsfordelingar

- Stokastisk variabel: **onsdag**
- Diskret sannsynsfordeling: **onsdag**
- Kontinuerleg sannsynsfordeling: **onsdag**
- Kummulativ sannsynsfordeling: **onsdag**
- Diskret simultanfordeling **I dag**
- Kontinuerleg simultanfordeling **I dag**
- Marginal sannsynsfordeling **I dag**
- Betinga sannsynsfordeling **I dag**
- Statistisk uavhengig **I dag**

Stokastisk variabel

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

Notasjon:

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

Diskret stokastisk variabel:

Dersom den stokastiske variabelen gjev tellbart antall utfall.

Kontinuerleg stokastisk variabel

Dersom den kontinuerlege variabelen har utfall på kontinuerleg skala.

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 moglege x)
- $f(x) = P(X = x)$

Kontinuerleg

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

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

Definisjon

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

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

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

Kap. 3.4 Simultane sannsynsfordelingar

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

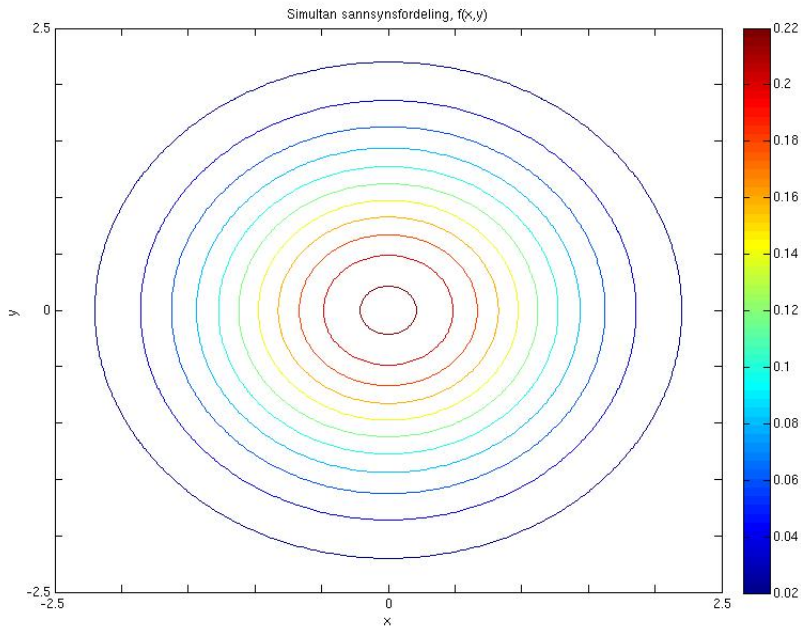
- 1 $f(x, y) \geq 0$ for alle (x, y)
- 2 $\sum_{\forall x} \sum_{\forall y} f(x, y) = 1$
- 3 $P(X = x, Y = y) = f(x, y)$

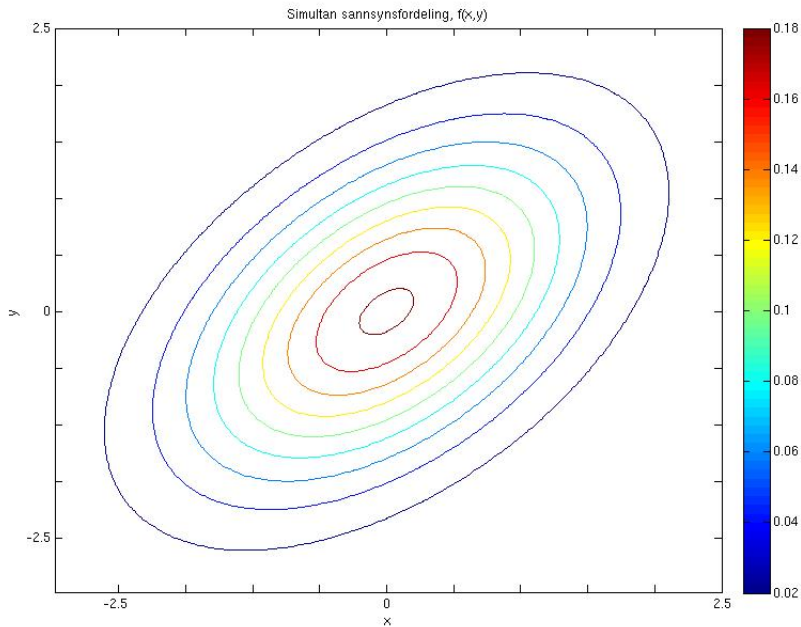
For ein kvar region A i xy -plane: $P(A) = \sum_A f(x, y)$

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

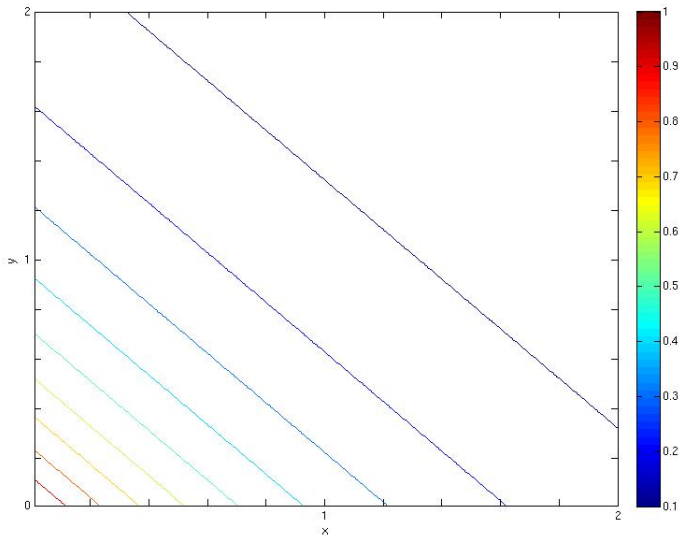
Funksjonen $f(x, y)$ er ei simultan sannsynsfordeling for dei kontinuerlege stokastiske variablane X og Y dersom

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





Simultanfordeling ventilar eksempel



Definisjon 3.10

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

Definisjon 3.11

La X og Y vere to stokastiske variable (kont. eller diskrete) med simultan sannsynsfordeling $f(x, y)$ og marginal sannsynsfordeling hhv $g(x)$ og $h(y)$.

Den *betinga fordelinga til X gjeve at $Y = y$* er

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

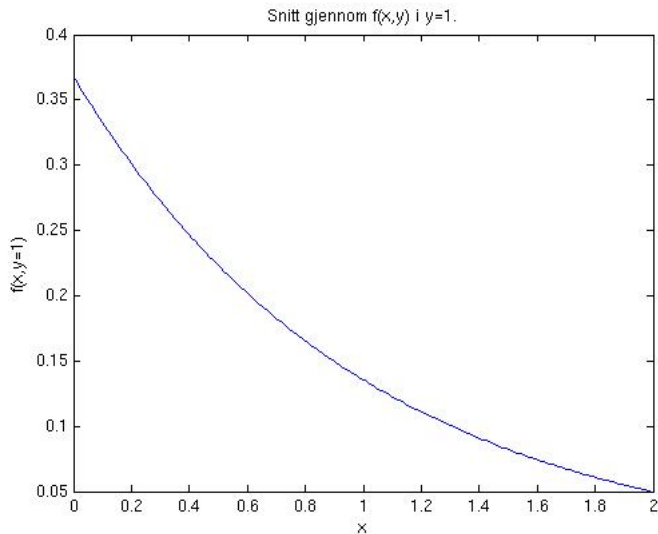
$$(h(y) > 0)$$

Tilsvarende er den betinga fordelinga til Y gjeve at $X = x$

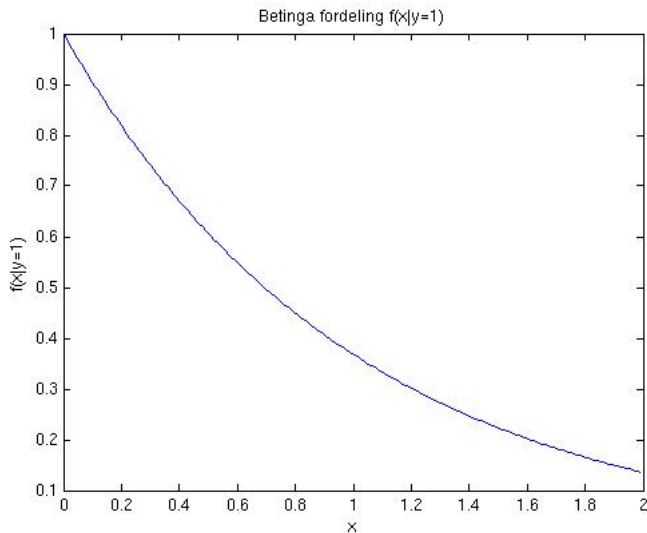
$$f(y|x) = f(x, y)/g(x)$$

$$(g(x) > 0)$$

Snitt gjennom simultanfordeling i $Y = 1$



To ventilar, betinga fordeling



Definisjon 3.12

La X og Y vere to stokastiske variable med simultan sannsynsfordeling $f(x, y)$ og marginal sannsynsfordeling hhv $g(x)$ og $h(y)$.

X og Y er *statistisk uavhengige dersom, og berre dersom*

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

Definisjon 3.13

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

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

Kap 3: Stokastiske variable og sannsynsfordelingar

- Stokastisk variabel: $X : S \rightarrow \mathbb{R}$
- 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
 $P((X, Y) \in A) = \int \int_A f(x, y) dx dy$
- Marginal sannsynsfordeling: For ein stok.var. aleine
- Betinga sannsynsfordeling $f(x|y) = f(x, y)/h(y)$
- Statistisk uavhengig dersom $f(x_1, x_2) = f_1(x_1)f_2(x_2)$