Simultanfordeling, betinget fordeling og uavhengige stokastiske variabler

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland
Institutt for matematiske fag
Norges teknisk-naturvitenskapelige universitet

Simultanfordeling

Definisjon (Simultanfordeling)

Hvis vi har to diskrete SV X og Y er deres simultanfordeling (simultan punktsannsynlighet) $f(x,y)$ gitt ved

$$f(x,y) = P(X = x, Y = y).$$

Hvis vi har to kontinuerlige SV X og Y er deres simultanfordeling (simultan sannsynlighetstetthet) $f(x,y)$ gitt ved

$$P((X,Y) \in A) = \iint_A f(x,y)dx\,dy,$$

der $A \subseteq \mathbb{R}^2$.

- ⋆ Merk:
  - noen ganger skriver vi $f_{XY}(x,y)$ i stedet for $f(x,y)$ for å understreke at det er simultanfordelingen for de SV X og Y
  - når X og Y er diskrete SV må $f(x,y)$ oppfylle

$$f(x,y) \geq 0$$
$$\sum_x \sum_y f(x,y) = 1$$

  - når X og Y er kontinuerlige SV må $f(x,y)$ oppfylle

$$f(x,y) \geq 0$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\,dy = 1$$

Marginalfordeling

- La $f_{XY}(x, y)$ være simultanfordelingen til de to stokastiske variablene X og Y
  - hva er da (marginal)fordelingen til X, $f_X(x)$?
- Husk regneregelen: Hvis $A_1, A_2, \ldots, A_n$ er parvis disjunkte hendelser har vi at

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

- Hvis X og Y er diskrete SV gir denne at

$$f_X(x) = P(X = x) = P\left(\bigcup_{y}(X = x, Y = y)\right) = \sum_{y} P(X = x, Y = y) = \sum_{y} f_{XY}(x, y)$$

$$f_Y(y) = P(Y = y) = P\left(\bigcup_{x}(X = x, Y = y)\right) = \sum_{x} P(X = x, Y = y) = \sum_{x} f_{XY}(x, y)$$

- Tilsvarende har vi for kontinuerlige SV, men må da integrere i stedet for å summere

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)dx$$

Betinget fordeling

Betinget sannsynlighet: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
Multiplikasjonssetningen: $P(A \cap B) = P(A|B)P(B)$
Bayes regel: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Definisjon (Betinget fordeling)

La X og Y være SV med simultanfordeling $f_{XY}(x,y)$. Den betingede fordelingen for Y gitt $X = x$ er da

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \textit{hvis} \quad f_X(x) > 0,$$

og den betingede fordelingen for X gitt $Y = y$ er

$$f_{X|Y}(y|x) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \textit{hvis} \quad f_Y(y) > 0.$$

- ⋆ Multiplikasjonssetningen for fordelinger

$$f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$$
$$f_{XY}(x,y) = f_{X|Y}(x|y)f_Y(y)$$

- ⋆ Bayes regel for fordelinger

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

- ⋆ Setningen om total sannsynlighet

$$f_X(x) = \sum_y f_{X|Y}(x|y)f_Y(y) \quad \text{(diskret SV)}$$
$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y) \quad \text{(kontinuerlig SV)}$$

Uavhengige stokastiske variabler

Hendelsene A og B er uavhengige hvis
$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

Definisjon (Uavhengige stokastiske variabler)

To stokastiske variabler X og Y er uavhengige hvis og bare hvis

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \ \text{ for alle } \ (x, y).$$

- Hvis X og Y er uavhengige og $f_X(x) > 0$ får vi også

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y)$$

- Tilsvarende hvis X og Y er uavhengige og $f_Y(y) > 0$ får vi

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x)$$

- Hvis vi har n uavhengige stokastiske variabler $X_1, X_2, \ldots, X_n$ har vi

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Oppsummering

- Vi har definert begrepene
  - simultanfordeling, $f_{XY}(x, y)$
  - betinget fordeling, $f_{X|Y}(x|y)$
  - uavhengige stokastiske variabler
- Har diskutert
  - hvordan finne marginalfordeling $f_X(x)$ fra en simultanfordeling $f_{XY}(x, y)$
  - multiplikasjonssetning for fordelinger
  - Bayes regel for fordelinger
  - setningen om total sannsynlighet for fordelinger