

Stokastiske variabler, punktsannsynlighet og sannsynlighetstetthet

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

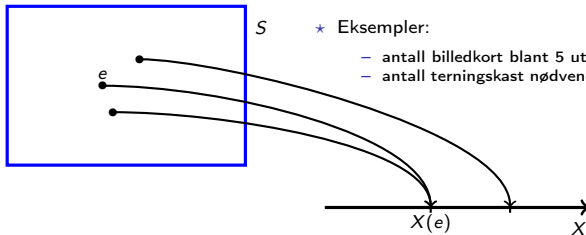
Stokastiske variabler

Definisjon (Stokastiske variabler)

En stokastisk variabel (SV) er en reell funksjon definert på et utfallsrom.

★ La X være en stokastisk variabel definert på et utfallsrom S

- $X(e) \in \mathbb{R}, e \in S$
- $X : S \rightarrow \mathbb{R}$



★ Eksempler:

- antall billedkort blant 5 utdelte kort
- antall terningskast nødvendig for å få 2 seksere

★ Notasjon: Bruker store bokstaver sent i alfabetet til å betegne SV, typisk X, Y, Z, U, V , eventuelt $X_1, X_2, \dots$

★ Man skiller mellom tre situasjoner:

- $X(e)$ kan ta endelig mange verdier, f.eks. 0, 1, 2, 3, 4, 5
- $X(e)$ kan ta tellbart uendelig mange verdier, f.eks. 1, 2, $\dots$
- $X(e)$ kan ta ikke-tellbart uendelig mange verdier, f.eks. $[-2, 3]$

★ De to første situasjonene gir en diskret SV, og i den siste en kontinuerlig SV

For diskret SV: Punktsannsynlighet

- ★ Husk: Vi har bare definert sannsynligheter for hendelser
- ★ Ønsker nå å angi sannsynligheter relatert til SV
 - hva skal vi mene med $P(X = 3)$?

- ★ For et tall x la

$$A_x = \{e \in S | X(e) = x\}$$

- ★ Vi lar da

$$P(X = x) = P(A_x) = P(\{e \in S | X(e) = x\})$$

- ★ Tilsvarende definerer vi

$$P(X \leq 3) = P(\{e \in S | X(e) \leq 3\})$$

$$P(a < X \leq b) = P(\{e \in S | a < X(e) \leq b\})$$

Definisjon (Punktsannsynlighet)

En funksjon $f(x)$ kalles punktsannsynlighet for en SV X dersom, for alle x ,

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

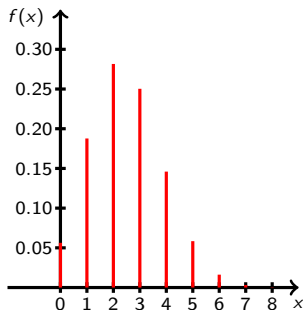
- ★ Merk at vi vil ha
 - $f(x) \geq 0$ for alle x
 - $\sum_x f(x) = 1$

Stolpediagram og sannsynlighetshistogram

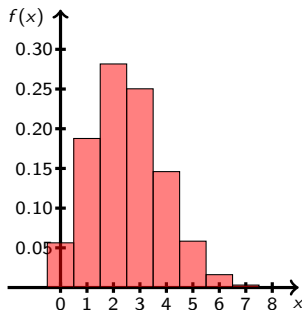
- ★ Eksempel: La X være en SV med punktsannsynlighet

$$f(x) = \binom{10}{x} \cdot 0.25^x \cdot 0.75^{10-x}, x = 0, 1, \dots, 10$$

Stolpediagram



Sannsynlighetshistogram



- ★ Merk:

$$\begin{aligned} P(X \leq 3) &= P(\{e \in S | X(e) \leq 3\}) = P(A_0 \cup A_1 \cup A_2 \cup A_3) \\ &= P(A_0) + P(A_1) + P(A_2) + P(A_3) = f(0) + f(1) + f(2) + f(3) \\ P(2 \leq X \leq 3) &= f(2) + f(3) \end{aligned}$$

For kontinuertlig SV: Sannsynlighetstetthet

- ★ Merk: punktsannsynligheter kan ikke brukes til å angi sannsynligheter for kontinuertlige SV
- ★ Eksempel:
 - Anta at vi slipper ei (infinitesimal) lita kule ned på intervallet $[0, 1]$ på en slik måte at alle posisjoner på intervallet er «like sannsynlige».
 - La X være posisjonen til kula.
 - Hva er da $P(X = 0.58)$?
 - Hvis vi tenker uniform sannsynlighetsmodell får vi



$$P(X = 0.58) = \frac{g}{m} = \frac{1}{\infty} = 0$$

Definisjon (Sannsynlighetstetthet)

Sannsynlighetstettheten $f(x)$ for en kontinuertlig SV X er gitt ved

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

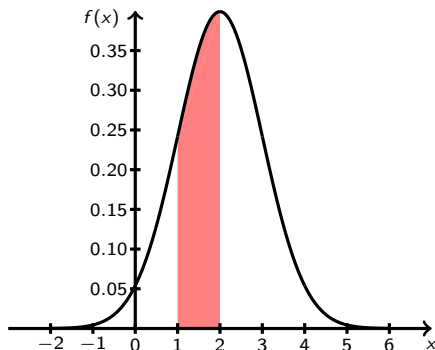
for alle $a < b$.

- ★ Merk at vi vil ha
 - $f(x) \geq 0$ for alle $x \in \mathbb{R}$
 - $\int_{-\infty}^{\infty} f(x) dx = 1$

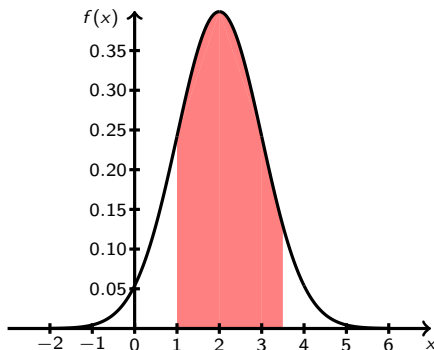
Sannsynlighetstetthet

- ★ Eksempel: La X være en SV med sannsynlighetstetthet

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}, -\infty < x < \infty$$



$$P(1 < X \leq 2) = \int_1^2 f(x) dx$$



$$P(1 < X \leq 3.5) = \int_1^{3.5} f(x) dx$$

- ★ Merk: For kontinuerlige SV får vi

$$P(X \leq x) = P(X < x) + P(X = x) = P(X < x)$$

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

Oppsummering

- ★ Har definert begrepene

- stokastisk variabel
- diskret og kontinuert stokastisk variabel
- punktsannsynlighet, $f(x)$
- sannsynlighetstetthet, $f(x)$

- ★ Har diskutert

- hvilke egenskaper har $f(x)$
- hvordan finne sannsynligheter fra $f(x)$

- ★ Merk:

- samme notasjon for punktsannsynlighet og sannsynlighetstetthet, $f(x)$
- fellesbetegnelse for punktsannsynlighet og sannsynlighetstetthet er *(sannsynlighets)fordeling*
- bruker noen ganger $f_X(x)$ for fordelingen for X
- hvis X har fordelingen $f(x)$ skriver vi

$$X \sim f(x)$$