Landmålingens fejlteori
Diskrete stokastiske variable
Lektion 2

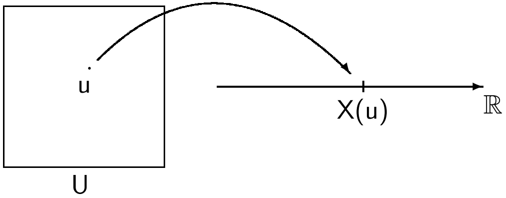
Torben Tvedebrink - tvede@math.aau.dk
http://www.math.aau.dk/∼tvede/teaching/L4

Institut for Matematiske Fag
Aalborg Universitet

28. april 2009

Definition

En reel funktion defineret på et udfaldsrum (med sandsynlighedsfordeling) kaldes en **stokastisk variabel**.



Eksempler

Forsøg	Stokastisk variabel	Type
kast med terning	# øjne	diskret
kast med to terninger	$\sum$ øjne	diskret
familie i Danmark	# børn	diskret
1. gangsfødende	alder	diskret
fødsel	fødselsvægt	kontinuert
hustande i Danmark	indkomst	kontinuert
mænd i Danmark	højde	kontinuert

Diskret stokastisk variabel: antal værdier endelig eller tællelig.

Kontinuert stokastisk variabel: antager værdier i en delmængde af reelle tal.

Definitioner

Definition 1

Lad U være et udfaldsrum med sandsynlighedsfordeling P. Lad X være en funktion givet ved $X : U \mapsto V$, hvor V er endelig eller tællelig delmængde af $\mathbb{R}$. Funktionen X kaldes en **diskret stokastisk variabel**.

Definition

En funktion $f : V \mapsto \mathbb{R}$ givet ved

$$f(x) = P(X = x), \ x \in V,$$

kaldes en **sandsynlighedsfunktion** for X.

Egenskaber for sandsynlighedsfunktionen $f(x)$

Der gælder

$$f(x) \geq 0, x \in V,$$

da $f(x)$ er en sandsynlighed, og

$$\sum_{x \in V} f(x) = 1,$$

da $\sum_{x \in V} f(x) = \sum_{x \in V} P(X = x) = P(X \in V) = 1.$

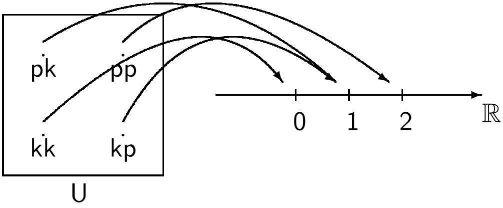
Eksempel

Kast 2 gange med symmetrisk mønt.
Udfaldsrum U:

(plat, krone)	(plat, plat)
(krone, krone)	(krone, plat)

Lad den **diskret stokastisk variabel** X være

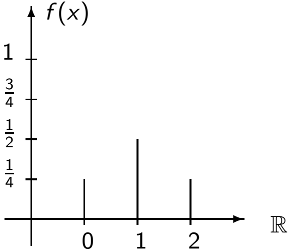
X : antal plat i to kast.
Dvs. $V = \{0, 1, 2\}$



Eksempel - fortsat

Sandsynlighedsfunktion for X:

$$f(x) = P(X = x) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{4} & x = 0, 2 \end{cases}$$



Bemærk $f(x) \geq 0$ for $x = 0, 1, 2$ og $f(0) + f(1) + f(2) = 1$.

Middelværdi og varians

Definition 2

Middelværdien af en diskret stokastisk variabel X er

$$\mu = \mathbb{E}(X) = \sum_{x \in V} x f(x) < \infty,$$

- også kaldet **forventede værdi**.

Middelværdi og varians

Definition 2

Middelværdien af en diskret stokastisk variabel X er

$$\mu = \mathbb{E}(X) = \sum_{x \in V} x f(x) < \infty,$$

- også kaldet **forventede værdi**.

Definition 3

Variansen af en diskret stokastisk variabel X er

$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}((X-\mu)^2) = \sum_{x \in V} (x-\mu)^2 f(x) < \infty,$$

hvor σ kaldes **spredning** eller **standard afvigelse** - mål for forventet variabilitet i X.

Middelværdi og varians

Bemærk at enheden for middelværdi er den samme som for X selv:

$$\mu[\text{enhed}] = \sum_{x \in V} x[\text{enhed}]\ f(x)$$

Hvor imod variansen ikke har samme enhed:

$$\sigma^2[\text{enhed}]^2 = \sum_{x \in V} (x[\text{enhed}] - \mu[\text{enhed}])^2\ f(x)$$

Derfor er spredningen (eller standard afvigelsen) ofte nemmere at forholde sig til idet enheden er denne samme som for X:

$$\sigma[\text{enhed}] = \sqrt{\sigma^2[\text{enhed}]^2}$$

Fx. ved vinkel målinger vil enheden på variansen være gon^2 mens spredningen vil have gon som enhed.

Eksempel - fortsat

Antal plat ved 2 mønstkast.

Middelværdien af X:

$$\mu = \mathbb{E}(X) = \sum_{x \in V} x f(x) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

Variansen af X:

$$\sigma^2 = \mathbb{V}(X) = \sum_{x \in V} (x - \mu)^2 f(x) = (0-1)^2 \frac{1}{4} + (1-1)^2 \cdot \frac{1}{2} + (2-1)^2 \cdot \frac{1}{4} = \frac{1}{2}.$$

Variansformel*

Vi har $\mathbb{E}(X) = \mu$ og $\mathbb{V}(X) = \sigma^2$. Da gælder

$$
\begin{aligned}
\sigma^2 = \mathbb{V}(X) &= \sum_{x \in V} (x-\mu)^2 f(x) \\
&= \sum_{x \in V} (x^2 f(x) + \mu^2 f(x) - 2x\mu f(x)) \\
&= \sum_{x \in V} x^2 f(x) + \mu^2 \sum_{x \in V} f(x) - 2\mu \sum_{x \in V} x f(x) \\
&= \sum_{x \in V} x^2 f(x) + \mu^2 - 2\mu^2 \\
&= \sum_{x \in V} x^2 f(x) - \mu^2 \\
&= \mathbb{E}(X^2) - \mu^2
\end{aligned}
$$

Eksempel - fortsat

Variansen bestemt ved brug af variansformlen

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x \in V} x^2 f(x) = 0^2 \frac{1}{4} + 1^2 \frac{1}{2} + 2^2 \frac{1}{4} = \frac{3}{2} \\
\mu^2 &= 1^2 = 1 \\
\sigma^2 &= \mathbb{E}(X^2) - \mu^2 = \frac{1}{2}.
\end{aligned}
$$

Senere i kurset skal vi se at med 95% sikkerhed fåes en observation i intervallet $\mu \pm 2\sigma$, dvs. σ er et udtryk for forventet variation i ens data.

Binomialfordeling

Forsøg med **to mulige udfald**: succes eller fiasko, hvor

$$P(\text{"succes"}) = p, \text{ hvor } 0 \leq p \leq 1.$$

Dvs.

$$P(\text{"fiasko"}) = 1 - p.$$

Gentag forsøget n gange, og lad

$$X : \text{antal sucesser.}$$

Sandsynlighed for x successer

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x \in \{1, 2, \ldots, n\}.$$

Som tidligere bestemmes $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

Egenskaber

Definition 4

En stokastisk variabel X er **binomialfordelt** med antalsparameter $n \in \mathbb{N}$ og sandsynlighedsparameter p, hvor $0 \leq p \leq 1$, hvis sandsynlighedsfunktionen for X er

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x \in \{0, 1, 2, \ldots, n\}.$$

Notation: $X \sim b(n, p)$.

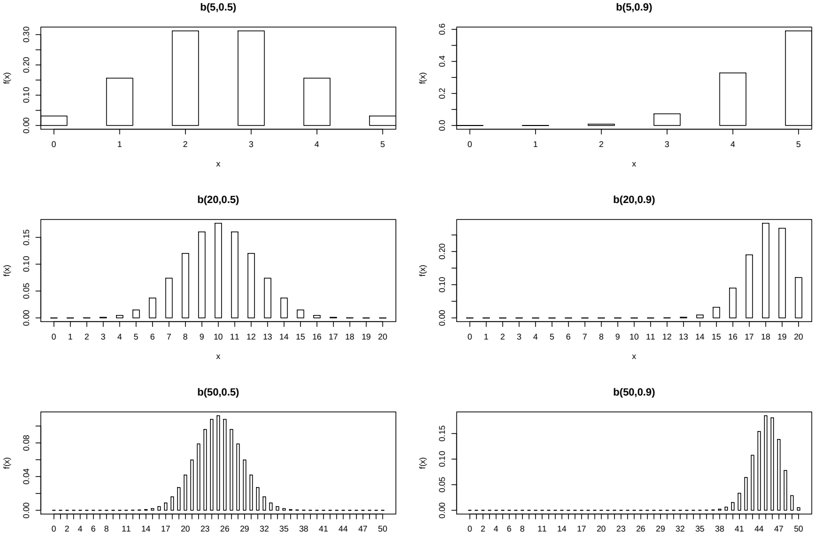
Sætning 1

Hvis $X \sim b(n, p)$, så er **middelværdi** og **varians** af X hhv.

$$\begin{aligned} \mathbb{E}(X) &= np, \\ \mathbb{V}(X) &= np(1-p). \end{aligned}$$

Sandsynlighedsfunktioner

Forskellige binomialfordelinger



Binomialfordeling

Eksempel: Kvalitetskontrol.
BILKA har mulighed for at afvise et parti batterier, hvis de ikke opfylder BILKA's accept-politik:

- Udtag en stikprøve på 20 batterier.
- Hvis et eller flere batterier ikke virker, kasseres hele partiet.

Antag BILKA modtager et parti, hvor 10% ikke virker.

Binomialfordeling

Eksempel: Kvalitetskontrol.
Lad

$$X \text{ : antal defekte batterier i stikprøven.}$$

Dvs.

$$X \sim b(20, 0.1).$$

Hvad er sandsynligheden for hele partiet kasseres?

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{20}{0} 0.1^0 0.9^{20} = 1 - 0.9^{20} = 0.8784.$$

Binomialfordeling

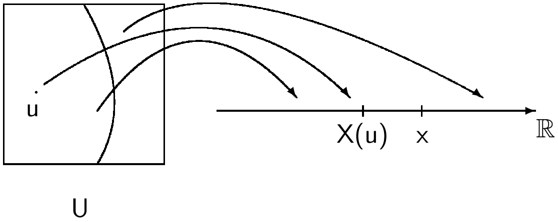
Eksempel: Kvalitetskontrol.
Hvad er sandsynligheden for højst 2 defekte batterier?

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= \binom{20}{0} 0.1^0 0.9^{20} + \binom{20}{1} 0.1^1 0.9^{19} + \binom{20}{2} 0.1^2 0.9^{18} \\ &= 0.1216 + 0.2702 + 0.28518 \\ &= 0.6769. \end{aligned}$$

Fordelingsfunktion

Lad X være en stokastisk variabel (diskret eller kontinuert) givet ved

$$X : U \mapsto \mathbb{R}.$$



Lad x være et reelt tal, og betragt hændelsen

$$\{u \in U | X(u) \leq x\}.$$

Hvad er sandsynligheden for denne hændelse?

Definition 1

Fordelingsfunktion for en stokastisk variabel X er

$$F(x) = P(\{u \in U | X(u) \leq x\}) = P(X \leq x).$$

Bemærk $F : \mathbb{R} \mapsto [0, 1]$.

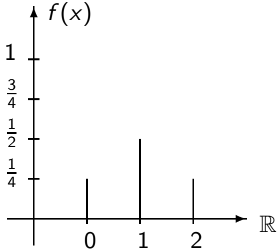
Fordelingen af en stokastisk variabel defineres ofte udfra fordelingsfunktionen.

Eksempel - fortsat

Kast 2 gange med symmetrisk mønt. X er antal plat i 2 kast.

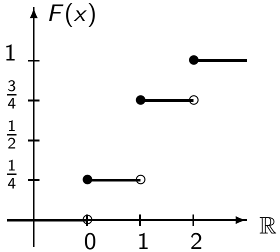
Sandsynlighedsfunktion for X:

$$f(x) = P(X = x) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{4} & x = 0, 2 \end{cases}$$



Fordelingsfunktion for X:

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$



Sætning 1

Lad F være fordelingsfunktion for en stokastisk variabel X. Da gælder

a. $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$,
b. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$,
c. F er ikke-aftagende,
d. $\lim_{x \to \infty} F(x) = 1$,
e. $\lim_{x \to -\infty} F(x) = 0$.

Der gælder

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - P(X < x).$$

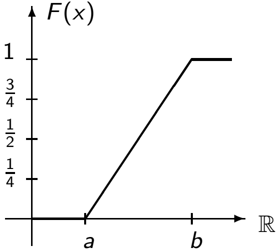
Hvis $P(X = x) = 0$, så springer F ikke i x, ellers angiver $P(X = x)$ størrelsen af **springet**.

Eksempel

Ligefordelingen

En stokastisk variabel X er **ligefordelt** på intervallet $]a;b[$ (eller $[a;b]$), hvis **fordelingsfunktionen** er givet som

$$F(x) = P(X \le x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$



Fx. Hvis X ligefordelt på $]1,3[$, så er

$$\begin{aligned} P(X \le 1) &= 0 \\ P(X \le 2) &= \frac{2-1}{3-1} = 0.5 \\ P(X \le 3.5) &= 1 \\ P(X > 2.5) &= 1 - P(X \le 2.5) = 1 - \frac{2.5-1}{3-1} = 0.25. \end{aligned}$$

Definition (p-fraktil)

Lad F være fordelingsfunktion for en stokastisk variabel X.

For $p \in]0;1[$ defineres **p-fraktilen**, x_p, som det (de) tal, der opfylder

$$F(x_p) = P(X \leq x_p) = p.$$

Dvs.

$$x_p = F^{-1}(p).$$

medianen, m defineres som 50%-fraktilen, dvs. $m = x_{1/2} = F^{-1}\left(1/2\right)$.