Landmålingens fejlteori
Transformation af stokastiske variable
Lektion 4

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

Institut for Matematiske Fag
Aalborg Universitet

1. maj 2009

Transformation af stokastisk variabel

Lad X være en stokastisk variabel, og g en **transformation** af X, hvor $g : \mathbb{R} \mapsto \mathbb{R}$.
Da er $Y = g(X)$ en ny stokastisk variabel.

Sætning 1

Hvis $Y = g(X)$, hvor X er en stokastisk variabel med tæthed-/sandsynlighedsfunktion f, så gælder

$$\begin{aligned} \mathbb{E}(Y) &= \int_{-\infty}^{\infty} g(x)f(x)dx, \quad \text{hvis } X \text{ kontinuert.} \\ \mathbb{E}(Y) &= \sum_{x \in V} g(x)f(x)dx, \quad \text{hvis } X \text{ diskret.} \end{aligned}$$

Eksempel - variansen

Hvis vi definerer g således: $g(X) = [X - \mathbb{E}(X)]^2 = (X - \mu)^2$ kan vi se at $g(X)$ er den transformation af X vi kalder variansen:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, \quad \text{hvis } X \text{ kontinuert.}$$

$$\mathbb{E}[g(X)] = \sum_{x \in V} g(x)f(x) = \sum_{x \in V} (x-\mu)^2 f(x), \quad \text{hvis } X \text{ diskret.}$$

Dvs.

$$\mathbb{V}(X) = \mathbb{E}((X-\mu)^2).$$

Varians et udtryk for den forventede variation omkring middelværdien, når vi observerer X.

Lineær transformation af stokastisk variabel

Sætning 2

Hvis X er en stokastisk variabel, og $a, b \in \mathbb{R}$, så gælder

$$\begin{aligned} \mathbb{E}(aX+b) &= a\mathbb{E}(X)+b, \\ \mathbb{V}(aX+b) &= a^2\mathbb{V}(X). \end{aligned}$$

Bevis 2

Vises for kontinuert (analog for diskret). Brug Sætning 1.

Middelværdi:

$$\begin{aligned} \mathbb{E}(aX+b) &= \int_{-\infty}^{\infty}(ax+b)f(x)dx \\ &= a\int_{-\infty}^{\infty}xf(x)+b\int_{-\infty}^{\infty}f(x)dx \\ &= a\mathbb{E}(X)+b. \end{aligned}$$

Lineær transformation af stokastisk variabel

Sætning 2

Hvis X er en stokastisk variabel, og $a, b \in \mathbb{R}$, så gælder

$$\begin{aligned} \mathbb{E}(aX+b) &= a\mathbb{E}(X)+b, \\ \mathbb{V}(aX+b) &= a^2\mathbb{V}(X). \end{aligned}$$

Bevis 2

Vises for kontinuert (analog for diskret). Brug Sætning 1.

Varians:

$$\begin{aligned} \mathbb{V}(aX+b) &= \mathbb{E}([aX+b-\mathbb{E}(aX+b)]^2) \\ &= \mathbb{E}([aX+b-a\mathbb{E}(X)-b]^2) \\ &= \mathbb{E}([a(X-\mathbb{E}(X))]^2) \\ &= \mathbb{E}(a^2[X-\mathbb{E}(X)]^2) \\ &= a^2\mathbb{E}([X-\mathbb{E}(X)]^2) \\ &= a^2\mathbb{V}(X). \end{aligned}$$

Eksempel - standardisering

Fra tidligere har vi at $X \sim \mathcal{N}(\mu, \sigma^2)$ kan standardiseres:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Da μ og σ blot er konstanter som a og b i Sætning 2 kan vi kontrollere middelværdi og varians af Z:

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{X - \mu}{\sigma}\right) = \frac{\mathbb{E}(X) - \mu}{\sigma} = 0$$

$$\mathbb{V}(Z) = \mathbb{V}\left(\frac{X - \mu}{\sigma}\right) = \frac{\mathbb{V}(X)}{\sigma^2} = 1.$$

Sum af stokastiske variable

Sætning 4 **middelværdi af sum**

Hvis $X_1, X_2, \ldots, X_n$ er stokastiske variable med endelig middelværdier $\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n)$, så gælder

$$\mathbb{E}\big(X_1 + X_2 + \cdots + X_n\big) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$$

eller

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i).$$

Bemærk: Sætning 4 er sand både når $X_1, X_2, \ldots, X_n$ er afhængige og uafhængige stokastiske variable (se definition på næste slide).

Uafhængige stokastiske variable

Definition

To stokastiske variable X_1 og X_2 er **uafhængige**, hvis

$$P(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}) = P(\{X_1 \in B_1\}) \cdot P(\{X_2 \in B_2\})$$

hvor $B_1 \subseteq \mathbb{R}$ og $B_2 \subseteq \mathbb{R}$.

Uafhængige stokastiske variable

Definition

To stokastiske variable X_1 og X_2 er **uafhængige**, hvis

$$P(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}) = P(\{X_1 \in B_1\}) \cdot P(\{X_2 \in B_2\})$$

hvor $B_1 \subseteq \mathbb{R}$ og $B_2 \subseteq \mathbb{R}$.

Generelt:

n stokastiske variable $X_1, X_2, \ldots, X_n$ er **uafhængige**, hvis

$$P(\{X_1 \in B_1\} \cap \ldots \cap \{X_n \in B_n\}) = P(\{X_1 \in B_1\}) \cdot \cdots \cdot P(\{X_n \in B_n\})$$

hvor $B_1, \ldots, B_n \subseteq \mathbb{R}$.

Dvs. definitionen svarer til den tidligere definition for uafhængige hændelser i udfaldsrummet.

Produkt af uafhængige stokastiske variable

Sætning 5 **middelværdi af produkt af uafhængige variable**
Hvis $X_1, X_2, \ldots, X_n$ er **uafhængige** stokastiske variable med endelig middelværdier $\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n)$, så gælder

$$\mathbb{E}\big(X_1 \cdot X_2 \cdot \cdots \cdot X_n\big) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \cdots \cdot \mathbb{E}(X_n)$$

eller

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

Sum af uafhængige stokastiske variable

Sætning 6 **varians af sum af uafhængige variable**
Hvis $X_1, X_2, \ldots, X_n$ er **uafhængige** stokastiske variable med endelig varians $\mathbb{V}(X_1), \mathbb{V}(X_2), \ldots, \mathbb{V}(X_n)$, så gælder

$$\mathbb{V}\big(X_1 + X_2 + \cdots + X_n\big) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + \cdots + \mathbb{V}(X_n)$$

eller

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Kovarians og korrelation

Definition 2

Lad X og Y være to stokastiske variable med middelværdi μ_X og μ_Y og varians σ_X^2 og σ_Y^2. **Kovariansen** af X og Y er

$$\mathrm{Cov}(X, Y) = \mathbb{E}([X - \mu_X][Y - \mu_Y]) = \mathbb{E}(XY) - \mu_X \mu_Y.$$

Bemærk

$$\mathrm{Cov}(X, X) = \mathbb{E}([X - \mu_X][X - \mu_X]) = \mathbb{E}([X - \mu_X]^2) = \mathbb{V}(X).$$

Korrelationskoefficienten for X og Y defineres som

$$\rho = \rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- også kaldet korrelationen mellem X og Y.

Sætninger

Sætning 7

Lad ρ være korrelationskoefficient for to stokastiske variable X og Y. Da gælder

$$-1 \leq \rho \leq 1.$$

Hvis X og Y er **uafhængige**, så er $\rho = 0$ og $\mathrm{Cov}(X, Y) = 0$.

Sætninger

Sætning

Hvis X og Y er to stokastiske variable, så gælder

$$\mathbb{V}(X \pm Y) = \mathbb{V}(X) + \mathbb{V}(Y) \pm 2\text{Cov}(X, Y).$$

Sum af uafhængige normalfordelte

Sætning 8 **sum af uafhængige normalfordelte**
Hvis X_1 og X_2 er **uafhængige** stokastiske variable, hvor

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \;\text{ og }\; X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2),$$

så er

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Sum af uafhængige normalfordelte

Sætning 8 **sum af uafhængige normalfordelte**

Hvis X_1 og X_2 er **uafhængige** stokastiske variable, hvor

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad \text{og} \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2),$$

så er

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Generelt:

Hvis $X_1, X_2, \ldots X_n$ er **uafhængige** stokastiske variable, hvor

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \text{for} \quad i = 1, 2, \ldots, n,$$

så er

$$X_1 + X_2 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

eller

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Sum af uafhængige normalfordelte

Special tilfælde af Sætning 8

Hvis $X_1, X_2, \ldots X_n$ er **uafhængige identiske** fordelte stokastiske variable (*i engelsk litteratur "iid": Independent Identically Distributed*), hvor

$$X_i \sim \mathcal{N}(\mu, \sigma^2) \quad \text{for} \quad i = 1, 2, \ldots, n,$$

så er

$$X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2).$$

Sum af uafhængige normalfordelte

Vi har altså $\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$ for uafhængige X'er. Dvs

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\mu \quad \text{og} \quad \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2.$$

Derfor har gennemsnittet af X'erne: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ følgende middelværdi og varians:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mu = \mu$$

$$\mathbb{V}(\bar{X}) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Altså:

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \qquad \text{eller} \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Den centrale grænseværdisætning

Den centrale grænseværdisætning

Hvis $X_1, X_2, \ldots X_n$ er **uafhængige identiske** fordelte (iid) stokastiske variable med ens middelværdi μ og ens varians σ^2, så gælder der for store n

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\rightarrow}{\sim} \mathcal{N}(0, 1).$$

Den centrale grænseværdisætning

Den centrale grænseværdisætning

Hvis $X_1, X_2, \ldots X_n$ er **uafhængige identiske** fordelte (iid) stokastiske variable med ens middelværdi μ og ens varians σ^2, så gælder der for store n

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \stackrel{\rightarrow}{\sim} \mathcal{N}(0, 1).$$

Dvs ligegyldigt hvilken fordeling X'erne har, vil deres gennemsnit $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ nærme sig en normalfordeling når n bliver stor:

$$\bar{X} \stackrel{\rightarrow}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Vilkårlig transformation af X

Vilkårlig transformation:
Lad X være en stokastisk variabel med

$$\mathbb{E}(X) = \mu \quad \text{og} \quad \mathbb{V}(X) = \sigma^2.$$

$Y = g(X)$: vilkårlig **differentiabel transformation** af X.

Middelværdien af $Y = g(X)$: (ofte vanskelig at beregne)

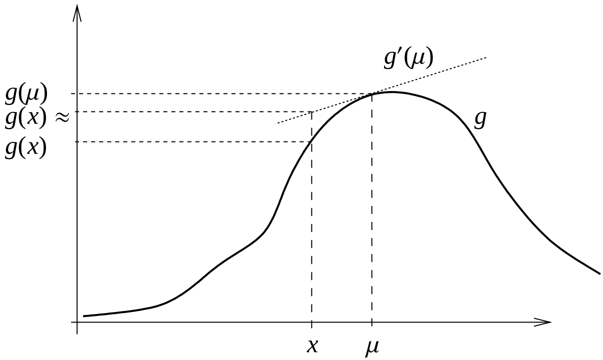
$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Linearisering - én variabel

Lineær approximation af g omkring μ:

$$\begin{aligned} Y = g(X) &\approx g(\mu) + g'(\mu)(X-\mu) \\ &= g'(\mu)X - g'(\mu)\mu + g(\mu) \\ &= aX + b, \end{aligned}$$

hvor $a = g'(\mu)$ og $b = -g'(\mu)\mu + g(\mu)$.



Linearisering - én variabel

Dvs. $Y = aX + b$ hvor $a = g'(\mu)$ og $b = -g'(\mu)\mu + g(\mu)$.

Hermed

$$\begin{aligned} \mathbb{E}(Y) &\approx a\mathbb{E}(X) + b \\ &= g'(\mu)\mu - g'(\mu)\mu + g(\mu) \\ &= g(\mu) \\ \mathbb{V}(Y) &\approx a^2\mathbb{V}(X) \\ &= g'(\mu)^2\sigma^2, \end{aligned}$$

hvor approximationen er god, hvis σ er lille.

Linearisering - generelt

Lad $X_1, X_2, \ldots, X_n$ være **uafhængige** stokastiske variable med

$$\mathbb{E}(X_1) = \mu_1,\ \mathbb{E}(X_2) = \mu_2,\ \ldots\ ,\ \mathbb{E}(X_n) = \mu_n,$$

$$\mathbb{V}(X_1) = \sigma_1^2,\ \mathbb{V}(X_2) = \sigma_2^2, \ldots\ ,\ \mathbb{V}(X_n) = \sigma_n^2.$$

$Y = g(X_1, X_2, \ldots, X_n)$: vilkårlig **differentiabel transformation** af $X_1, X_2, \ldots, X_n$.

Linearisering - generelt

Lineær approximation af g omkring $\mu_1, \mu_2, \ldots, \mu_n$:

$$\begin{aligned} Y &= g(X_1, X_2, \ldots, X_n) \\ &\approx g(\mu_1, \mu_2, \ldots, \mu_n) + \\ &\quad \frac{\partial g}{\partial X_1}(X_1 - \mu_1) + \frac{\partial g}{\partial X_2}(X_2 - \mu_2) + \cdots + \frac{\partial g}{\partial X_n}(X_n - \mu_n) \\ &= \frac{\partial g}{\partial X_1} X_1 + \frac{\partial g}{\partial X_2} X_2 + \cdots + \frac{\partial g}{\partial X_n} X_n \\ &\quad - \frac{\partial g}{\partial X_1}\mu_1 - \frac{\partial g}{\partial X_2}\mu_2 - \cdots - \frac{\partial g}{\partial X_n}\mu_n + g(\mu_1, \mu_2, \ldots, \mu_n), \end{aligned}$$

hvor de partielle afledede beregnes i $\mu_1, \mu_2, \ldots, \mu_n$.

Fejlforplantningsloven

Hermed **fejlforplantningsloven**

$$\begin{aligned}
\mathbb{E}(Y) &\approx g(\mu_1, \mu_2, \ldots, \mu_n) \\
\mathbb{V}(Y) &\approx \left(\frac{\partial g}{\partial X_1}\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial X_2}\right)^2 \sigma_2^2 + \cdots + \left(\frac{\partial g}{\partial X_n}\right)^2 \sigma_n^2,
\end{aligned}$$

hvor approximationen er god, hvis $\sigma_1, \sigma_2, \ldots, \sigma_n$ er små.