

Landmålingens fejlteori

Repetition - Transformation af stokastiske variable

Lektion 5

Torben Tvedebrink - tvede@math.aau.dk
<http://www.math.aau.dk/~tvede/teaching/L4>

Institut for Matematiske Fag
Aalborg Universitet

5. 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

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

$$\mathbb{E}(Y) = \sum_{x \in V} g(x)f(x), \quad \text{hvis } X \text{ diskret.}$$

Eksempel - variansen

Fx. er variansen defineret som

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \text{ hvis } X \text{ kontinuert.}$$

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

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

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

Sætning 4 Middelværdi af sum

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

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

eller

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

Uafhængige stokastiske variable

Definition

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

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

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

Bemærk, $X_1, X_2, \dots, X_n$ i Sætning 4 behøver ikke nødvendigvis at være uafhængige.

Sum af uafhængige stokastiske variable

Sætning 6 varians af sum af uafhængige

Hvis $X_1, X_2, \dots, X_n$ er **uafhængige** stokastiske variable med endelig varians $\mathbb{V}(X_1), \mathbb{V}(X_2), \dots, \mathbb{V}(X_n)$, så gælder

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

eller

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

Tidligere så vi

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

uanset om variablene er uafhængige eller ej.

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

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

Bemærk

$$\text{Cov}(X, X) = \mathbb{V}(X).$$

Korrelationskoefficienten for X og Y defineres som

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y}$$

- også kaldet korrelationen mellem X og Y . Der gælder

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

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

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

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

så er

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

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

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

så er

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

Dvs.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

eller

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

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 $b = -g'(\mu)\mu + g(\mu)$. 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 $X_1, X_2, \dots, X_n$ være **ufhængige** stokastiske variable med

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

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

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

Linearisering - generelt

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

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

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

Fejlförplantningsloven

Hermed **fejlförplantningsloven**

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

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

Eksempel - Længdemål med stålbånd

