Landmålingens fejlteori
Repetition - Fordeling af slutfejl
Lektion 8

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

Institut for Matematiske Fag
Aalborg Universitet

15. maj 2008

Fordeling af slutfejl

Betragt en situation hvor vi ved at summen af vores observationer skal være lig en kendt værdi x_0. Dette kunne eksempelvis være:

- $x_0 = 200$ gon, hvor vores målinger er vinkler i en trekant.
- $x_0 = 0$ mm, hvor vi nivellerer i et *lukket* net, dvs. slut- og startpunkt er det samme.
- $x_0 = H_2 - H_1$, hvor målinger er for at bestemme kote til et punkt (eksempel fra sidste gang).
- ...

I dag viser vi generelt hvorledes slutfejlen r fordeles.

Model

Vi har $X_1, \dots, X_n$ uafhængige stokastiske variable.

Der gælder at $\mathbb{E}(X_i) = \mu_i$ for alle $i = 1, \dots, n$.

Målingerne er ikke nødvendigvis af samme nøjagtighed, dvs $\mathbb{V}(X_i) = \sigma_i^2$.
Vi bruger derfor vægtrelationen så vægtene p_i opfylder

$$\sigma_0^2 = p_1\sigma_1^2 = p_2\sigma_2^2 = \cdots = p_n\sigma_n^2.$$

Det betyder således at

$$\sigma_i^2 = \frac{\sigma_0^2}{p_i} \quad \text{for alle } i = 1, \dots, n.$$

Fra vores viden om forsøget ved vi at middelværdierne skal summere sammen til x_0,

$$\mu_1 + \mu_2 + \cdots + \mu_n = x_0.$$

Fx. skal vinklerne i en trekant $\alpha + \beta + \gamma = 200$ gon.

Slutfejlen r

Vi observerer $x_1, \dots, x_n$ ved opmåling (fx. vinkler eller højde forskelle). Der skal ligeledes gælde at disse værdier summere sammen til x_0. Pga målefejl er dette *ikke muligt* - derfor indføres slutfejlen r:

$$x_1 + x_2 + \cdots + x_n + r = x_0.$$

Fx. skal vores målinger af vinklerne i en trekant opfylde de tilsammen giver 200 gon. Hvis vi har observeret

x_1	70 gon
x_2	90 gon
x_3	43 gon
Sum	203 gon

Kan vi bestemme r til at være

$$r = x_0 - (x_1 + x_2 + x_3) = 200 - 203 = -3 \text{ gon}.$$

Estimat af μ_i

Til at estimere μ_i kan vi bruge den direkte observation x_i samt den indirekte observation $x_0 - (x_1 + \cdots x_{i-1} + x_{i+1} + \cdots x_n) = x_i + r$.

Vi kan bestemme variansen af de to bidrag til vores estimat:

$$\mathbb{V}(X_i) = \sigma_i^2$$
$$\mathbb{V}(x_0 - (X_1 + \cdots + X_{i-1} + X_{i+1} + \cdots + X_n)) = \sigma_+^2 - \sigma_i^2,$$

hvor $\sigma_+^2 = \sigma_1^2 + \cdots + \sigma_n^2$.

Vægte

Vægtrelationen skal være opfyldt:

$$p_1\sigma_i^2 = p_2(\sigma_+^2 - \sigma_i^2)$$

Vi vælger

$$p_1 = \frac{\sigma_+^2}{\sigma_i^2} \quad \text{og} \quad p_2 = \frac{\sigma_+^2}{\sigma_+^2 - \sigma_i^2}$$

Det betyder at $p_+ = p_1 + p_2$ er givet ved

$$p_+ = \frac{\sigma_+^2}{\sigma_i^2} + \frac{\sigma_+^2}{\sigma_+^2 - \sigma_i^2} = \frac{\sigma_+^2(\sigma_+^2 - \sigma_i^2)}{\sigma_i^2(\sigma_+^2 - \sigma_i^2)} + \frac{\sigma_+^2\sigma_i^2)}{\sigma_i^2(\sigma_+^2 - \sigma_i^2)} = \frac{(\sigma_+^2)^2}{\sigma_i^2(\sigma_+^2 - \sigma_i^2)}.$$

Estimat af μ_i - fortsat

Til at estimere μ_i anvendes det vægtede gennemsnit $\bar{x}_i^*$. På forrige slide bestemte vi vægtene p_1 og p_2 til observationerne x_i og $x_i + r$,

$$\begin{aligned}\bar{x}_i^* &= \frac{p_1}{p_+}x_i + \frac{p_2}{p_+}(x_i + r)\\ &= \frac{p_1 + p_2}{p_+}x_i + \frac{p_2}{p_+}r\\ &= x_i + \frac{p_2}{p_+}r.\end{aligned}$$

Vi kigger nu nærmere på $\frac{p_2}{p_+}$,

$$\frac{p_2}{p_+} = \frac{\frac{\sigma_+^2}{\sigma_+^2 - \sigma_i^2}}{\frac{(\sigma_+^2)^2}{\sigma_i^2(\sigma_+^2 - \sigma_i^2)}} = \frac{\sigma_+^2}{\sigma_+^2 - \sigma_i^2}\frac{\sigma_i^2(\sigma_+^2 - \sigma_i^2)}{(\sigma_+^2)^2} = \frac{\sigma_i^2}{\sigma_+^2}.$$

Dvs. slutfejlen fordeles efter andelen af den totale varians.

Estimat af μ_i - take home message

Altså - til at estimere μ_i anvendes det vægtede gennemsnit $\bar{x}_i^*$ som er givet ved

$$\bar{x}_i^* = x_i + \frac{\sigma_i^2}{\sigma_+^2} r.$$

Fordi $\sigma_i^2 = \frac{\sigma_0^2}{p_i}$ (fra vægtrelationen) kan udtrykket simplificeres yderligere,

$$\bar{x}_i^* = x_i + \frac{\frac{\sigma_0^2}{p_i}}{\frac{\sigma_0^2}{p_1} + \cdots + \frac{\sigma_0^2}{p_n}} r = x_i + \frac{\frac{1}{p_i}}{\frac{1}{p_1} + \cdots + \frac{1}{p_n}} r.$$

Fordeling af slutfejl i nivellement

Ved nivellering har vi tidligere vist at variansen på en højdemåling over en længde ℓ er $\mathbb{V}(H) = \sigma_k^2 \ell$, hvor σ_k er kilometerspredningen.

Derfor er vægtene

$$p_i = \frac{1}{\ell_i} \quad \Leftrightarrow \quad \frac{1}{p_i} = \ell_i.$$

Dvs vores estimat for μ_i (højdeforskel i):

$$\bar{x}_i^* = x_i + \frac{\frac{1}{p_i}}{\frac{1}{p_1} + \cdots + \frac{1}{p_n}} r = x_i + \frac{\ell_i}{\ell_1 + \cdots + \ell_n} r.$$

Fra eksemplet med koten (lektion 7) havde vi $\mu_1 = \mu_2 = \mu$. Estimatet for μ er således

$$\bar{x}_1^* = x_1 + \frac{\ell_1}{\ell_1 + \ell_2} r.$$

Fordelingen af slutfejl i en trekant

På samme måde kan vi for en trekant let bestemme hvorledes slutfejlen skal fordeles. Antag vi måler vinklerne α, β og γ med samme nøjagtighed σ_v^2, men med et forskelligt antal satser.

Observation	Satser	Varians på middelsats	Vægt
x_α	k_α	$\frac{\sigma_v^2}{k_\alpha}$	k_α
x_β	k_β	$\frac{\sigma_v^2}{k_\beta}$	k_β
x_γ	k_γ	$\frac{\sigma_v^2}{k_\gamma}$	k_γ

Estimatet af fx. β er derfor givet ved det vægtede gennemsnit $\bar{x}_\beta^*$,

$$\bar{x}_\beta^* = x_\beta + \frac{\frac{1}{k_\beta}}{\frac{1}{k_\alpha} + \frac{1}{k_\beta} + \frac{1}{k_\gamma}} r.$$

Hvis antallet af satser er ens for alle vinkler får vi $\bar{x}_\beta^* = x_\beta + \frac{1}{3} r$.

Dobbeltmålinger

Udgangspunktet er $2n$ målinger hvor målingerne to og to måler samme størrelse, altså n forskellige størrelser i alt.
Eks.: siderne i et polygon hvor hver sidelængde måles to gange.

$$\begin{array}{ccccccc} X_{11} & X_{12} & X_{21} & X_{22} & \ldots & X_{n1} & X_{n2} \\ \multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} & \ldots & \multicolumn{2}{c}{\downarrow} \\ x_{11} & x_{12} & x_{21} & x_{22} & \ldots & x_{n1} & x_{n2} \end{array}$$

De n forskellige størrelser har sande værdier $\mu_1, \ldots, \mu_n$, og vi antager således

$$\mathbb{E}(X_{11}) = \mathbb{E}(X_{12}) = \mu_1 \quad \ldots \quad \mathbb{E}(X_{n1}) = \mathbb{E}(X_{n2}) = \mu_n.$$

Ydermere antages målingerne uafhængige af hinanden og foretaget med samme målekvalitet, dvs. ens varians σ^2.

Estimat af σ^2

Idet X_{i1} og X_{i2} måler samme størrelse μ siger deres differens $Y_i = X_{i1} - X_{i2}$ noget om målekvaliteten.

Da målekvaliteten antages at være ens for alle $2n$ målinger, kan alle n par anvendes til at estimere σ^2.

Vi ved

$$\mathbb{E}(Y_i) = \mathbb{E}(X_{i1}) - \mathbb{E}(X_{i2}) = \mu - \mu = 0$$
$$\mathbb{V}(Y_i) = \mathbb{V}(X_{i1}) + \mathbb{V}(X_{i2}) = \sigma^2 + \sigma^2 = 2\sigma^2.$$

Dvs. vi kender den *sande* middelværdi af Y_i, $\mathbb{E}(Y_i) = \mu_{Y_i} = 0$. Derfor er

$$\hat{s}^2 = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \mu_{Y_i})^2 = \frac{1}{2n}\sum_{i=1}^{n} y_i^2 = \frac{1}{2n}\sum_{i=1}^{n}(x_{i1} - x_{i2})^2$$

et estimat af σ^2, som er variansen på den enkelte måling.

Approximativt konfidensinterval for σ^2

Idet vi kan estimere σ^2 vha af $\hat{s}^2$ kan vi ligeledes konstruere et konfidensinterval for σ^2.

Vi har ikke gennemgået teorien for et eksakt konfidensinterval, men nedenstående er et approximativt 95%-konfidensinterval for σ,

$$\left[\hat{s} - 1.96\frac{\hat{s}}{\sqrt{2n}} \;;\; \hat{s} + 1.96\frac{\hat{s}}{\sqrt{2n}}\right].$$