Landmålingens fejlteori
Lektion 4
Vægtet gennemsnit
Fordeling af slutfejl

Rasmus Waagepetersen - rw@math.aau.dk

Institut for Matematiske Fag
Aalborg Universitet

Estimation af varians/spredning

Antag $X_1, \dots, X_n$ stokastiske variable med fælles middelværdi μ og varians σ^2.

Hvis μ er *kendt* estimeres σ^2 ved

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Hvis μ er *ukendt* estimeres σ^2 ved

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

NB: begge estimater er centrale (middelværdirette) men $\hat{S}^2$ har mindst varians (mest præcist).

Dobbeltmålinger: X_{i1} og X_{i2} uafhængige middelværdi μ_i og varians σ^2.

Da har differens $Y_i = X_{i1} - X_{i2}$ kendt middelværdi nul og varians $\sigma_Y^2 = 2\sigma^2$.

σ_Y^2 kan da estimeres ved

$$\hat{S}_Y^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

og σ^2 estimeres ved

$$\hat{S}_Y^2/2$$

Linearisering

Antag X har middelværdi 10 og spredning 0.1.

Lad $Y = \log(X)$ (naturlig logaritme).

Linearisering af $\log(\cdot)$:

$$\log(x) \approx \log(10) + \frac{1}{10}(x - 10) = \frac{1}{10}x + \log(10) - 1$$

Dermed

$$\mathbb{E}Y \approx \log(10) = 2.306 \quad \mathbb{V}\mathrm{ar}Y \approx \left(\frac{1}{10}\right)^2 0.1^2 = 0.0001$$

Fortolkning af afledet

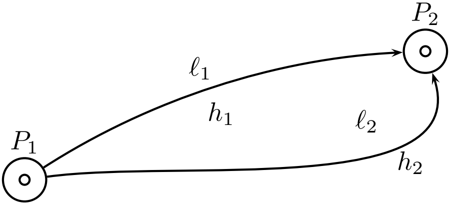
Hvis $Y = h(X)$ måler $h'(\mu)$ hvor sensitiv Y er overfor ændringer i X omkring μ (hvor meget giver en ændring i X anledning til at Y ændres).

Derfor giver det god mening at

$$\mathbb{V}\mathrm{ar} Y \approx (h'(\mu))^2 \mathbb{V}\mathrm{ar} X$$

Vægte — motiverende eksempel

- Højdeforskellen mellem punkterne P_1 og P_2 er opmålt ved et geometrisk nivellement over to forskellige strækninger $\ell_1 > \ell_2$.



- De to estimater h_1 og h_2 er realisationer af stokastiske variable H_1 og H_2 med varianser $\ell_1\sigma_k^2$ og $\ell_2\sigma_k^2$.
- Kan vi kombinere de to estimater h_1 og h_2 for at opnå et bedre estimat under hensyntagen til, at der er større usikkerhed for H_2 end for H_1 ?
- En mulighed: $h = (h_1 + h_2)/2$ - men det kan gøres bedre !

Vægte

Antag $X_1, \dots, X_n$ er uafhængige stokastiske variable med ens middelværdi μ men forskellige varianser $\sigma_1^2, \dots, \sigma_n^2$.

F.eks. målinger med instrumenter af variende kvalitet eller nogle X_i kunne repræsentere gennemsnit af gentagne målinger.

Hvert X_i tildeles en vægt $p_i > 0$ som afspejler målingens kvalitet (stort p_i svarer til god måling).

Vi estimerer da μ ved det vægtede gennemsnit

Definition: Vægtet gennemsnit

For observationer $X_1, \dots, X_n$ med vægte $p_1 > 0, \dots, p_n > 0$ er det **vægtede gennemsnit**

$$\bar{X}^* = \frac{p_1}{\sum_{i=1}^n p_i} X_1 + \cdots + \frac{p_n}{\sum_{i=1}^n p_i} X_n = \frac{1}{\sum_{i=1}^n p_i}(p_1 X_1 + \cdots + p_n X_n).$$

"resulterende" vægte

$$w_i = \frac{p_i}{\sum_{l=1}^{n} p_l}$$

summer sammen til en:

$$\sum_{i=1}^{n} w_i = 1$$

Centralitet af vægtet gennemsnit

Det vægtede gennemsnit

$$\bar{X}^* = \frac{p_1}{\sum_{i=1}^n p_i} X_1 + \cdots + \frac{p_n}{\sum_{i=1}^n p_i} X_n = \frac{1}{\sum_{i=1}^n p_i} (p_1 X_1 + \cdots + p_n X_n)$$

er et centralt estimat af μ, dvs.

$$\mathbb{E}(\bar{X}^*) = \mu$$

for et hvilket som helst valg af vægte $p_1 > 0, \ldots, p_n > 0$.

Optimale vægte

Der er uendelig mange mulige valg af vægte. Hvad er det bedste valg ?

Bedst betyder variansen $\mathbb{V}\mathrm{ar}\bar{X}^*$ mindst mulig !

Sætning: Optimale vægte

Variansen $\mathbb{V}\mathrm{ar}\bar{X}^*$ er minimal hvis **vægtrelationen**

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

er opfyldt.

Vægtrelationen

Antag vægtrelationen er opfyldt:

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

Lad σ_0^2 betegne den fælles værdi af $p_i\sigma_i^2$ (svarer til varians for observation med vægt $p_0 = 1$).

Da ser vi

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}$$

Dvs. vi skal simpelthen vælge vægtene så de er omvendt proportionale med varianserne !

Vægtrelationen: Eksempel

Antag variansen på første måling er dobbelt så stor som på den anden ($\sigma_1^2 = 2\sigma_2^2$).

Et valg af vægte, der opfylder vægtrelationen, er $p_1 = 1$ og $p_2 = 2$ ($\sigma_0^2 = 2\sigma_2$), idet

$$p_1\sigma_1^2 = p_2\sigma_2^2$$
$$\Leftrightarrow 1 \cdot 2\sigma_2^2 = 2\sigma_2^2$$

Der er uendelig mange lige gode valg af vægte!!

Et andet valg er $p_1 = \frac{1}{2}$ og $p_2 = 1$ ($\sigma_0^2 = \sigma_2^2$).

Bemærk: I eksemplet er vægten på den bedste måling dobbelt så stor som vægten på den dårlige måling.

Begge valg af vægte giver samme $\bar{X}^*$:
$p_1/(p_1 + p_2) = 1/(1 + 2) = (1/2)/(1/2 + 1) = 1/3$

Alle vægte der opfylder vægtrelationen giver samme resulterende vægte og dermed samme $\bar{X}^*$.

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}$$

Dermed

$$w_i = \frac{p_i}{\sum_{l=1}^n p_l} = \frac{\frac{\sigma_0^2}{\sigma_i^2}}{\sum_{l=1}^n \frac{\sigma_0^2}{\sigma_l^2}} = \frac{\frac{1}{\sigma_i^2}}{\sum_{l=1}^n \frac{1}{\sigma_l^2}} = \frac{1}{\sum_{l=1}^n \frac{\sigma_i^2}{\sigma_l^2}}$$

Dvs. i sidste ende afhænger $\bar{X}^*$ ikke af σ_0^2.

Med et andet sæt vægte $\tilde{p}_i = \tilde{\sigma}_0^2/\sigma_i^2$ fås nøjagtig samme w_i !

Variansen for det vægtede gennemsnit

Sætning: Variansen af vægtet gennemsnit

Antag vægtrelationen er opfyldt, dvs.

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

Da gælder

$$\mathbb{V}\mathrm{ar}(\bar{X}^*) = \frac{\sigma_0^2}{\sum_{i=1}^n p_i}.$$

Eksempel

X_1, X_2 og X_3 har fælles middelværdi μ og varianser hhv. $3, 4, 1$.

Et optimalt sæt af vægte er $p_1 = 1/3$, $p_2 = 1/4$ og $p_3 = 1$.

Antag $x_1 = 5$, $x_2 = 5.7$ og $x_3 = 4.9$ er observeret. Da er

$$\bar{x}^* = 5.047$$

Endvidere er

$$\mathbb{V}\mathrm{ar}\bar{X}^* = \frac{1}{\frac{1}{3} + \frac{1}{4} + 1} = 0.63$$

Estimat af σ_0^2

Estimat af σ_0^2

Som estimat af σ_0^2 anvendes S_0^2:

$$S_0^2 = \frac{\sum_{i=1}^n p_i(X_i - \bar{X}^*)^2}{n-1}$$
$$= \frac{\sum_{i=1}^n p_i X_i^2 - (\bar{X}^*)^2 \sum_{i=1}^n p_i}{n-1}.$$

Der gælder, at S_0^2 er et centralt estimat for σ_0^2: $\mathbb{E}S_0^2 = \sigma_0^2$.

Estimat af $\mathbb{V}\mathrm{ar}(\bar{X}^*)$

- Vi har, at s_0^2 er et centralt estimat for σ_0^2.
- Variansen for det vægtede gennemsnit er

$$\mathbb{V}\mathrm{ar}(\bar{X}^*) = \frac{\sigma_0^2}{\sum_{i=1}^n p_i}$$

- Et centralt estimat for variansen af det vægtede gennemsnit er derfor

$$\frac{s_0^2}{\sum_{i=1}^n p_i}$$

Eksempel: Geometrisk nivellement

En højdeforskel er nivelleret over n strækninger med *samme* kilometerspredning σ_k. Variansen over en længde ℓ er fra tidligere givet ved $\sigma_k^2 \ell$.

Højdeforskel h	Længde ℓ	Varians på h over ℓ
h_1	ℓ_1	$\sigma_1^2 = \sigma_k^2 \ell_1$
h_2	ℓ_2	$\sigma_2^2 = \sigma_k^2 \ell_2$
$\vdots$	$\vdots$	$\vdots$
h_n	ℓ_n	$\sigma_n^2 = \sigma_k^2 \ell_n$

Forskellige varianser pga. forskellige strækninger ($\ell_i \neq \ell_j$).

Eksempel: Geometrisk nivellement — valg af vægte

Vægtrelationen er opfyldt når

$$\sigma_0^2 = p_1\sigma_1^2 = \cdots = p_n\sigma_n^2$$

Indsættes udtrykket for $\sigma_i^2 = \ell_i\sigma_k^2$ har vi:

$$\sigma_0^2 = p_1\sigma_k^2\ell_1 = \cdots = p_n\sigma_k^2\ell_n$$

Heraf fremgår det at hvis $p_i = \ell_i^{-1}$ er ligheden opfyldt:

$$\begin{aligned}\sigma_0^2 &= \ell_1^{-1}\sigma_k^2\ell_1 = \cdots = \ell_n^{-1}\sigma_k^2\ell_n \\ &= \sigma_k^2 = \cdots = \sigma_k^2.\end{aligned}$$

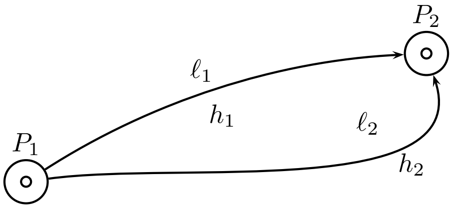
Dvs. vægte kan vælges til $p_i = (\ell_i)^{-1}$ (reciprokke længde).

Bemærk: her svarer σ_0^2 til kilometervariansen σ_k^2.

Hvis σ_k^2 er ukendt kan den estimeres vha. S_0^2 !

Eksempel

- Højdeforskellen mellem punkterne P_1 og P_2 er opmålt ved et geometrisk nivellement over to forskellige strækninger.



Data:

- $h_1 = 347$mm, $\ell_1 = 671$m
- $h_2 = 356$mm, $\ell_2 = 853$m

Estimat af højdeforskel (vægtet gennemsnit):

$$\bar{x}^* = \frac{1/671}{1/671 + 1/853} 347 + \frac{1/853}{1/671 + 1/853} 356 \approx 350,96\text{mm}$$

Antag yderligere at kilometerspredning $\sigma_k = 3\text{mm}/\sqrt{\text{km}}$.

Varians på estimatet:

$$\mathbb{V}\mathrm{ar}\bar{X}^* = \frac{3^2/1000}{1/671 + 1/853} = 3.3801$$

Til sammenligning er varianserne for X_1 og X_2 henholdsvis 6.039 og 7.677.

NB: vi omregner kilometervarians 3^2 mm^2/km til "metervarians" $3^2/1000$ mm^2/m da længderne er angivet i m.

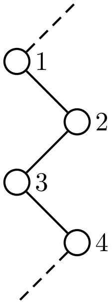
Eksempel - polygonvinkler

Vi måler vinklerne i punkterne 1-4.

Alle vinkler er målt med samme varians σ_v^2.

Lad $\sigma_{m,i}^2$ være variansen af middelsatsen i punkt i.

Punkt	Antal satser	$\sigma_{m,i}^2$
1	2	$\sigma_{m,1}^2 = \frac{\sigma_v^2}{2}$
2	3	$\sigma_{m,2}^2 = \frac{\sigma_v^2}{3}$
3	4	$\sigma_{m,3}^2 = \frac{\sigma_v^2}{4}$
4	6	$\sigma_{m,4}^2 = \frac{\sigma_v^2}{6}$



Eksempel - fortsat

Ifølge vægtrelationen skal $p_i \sigma_{m,i}^2$ være ens for alle i,

$$p_1 \sigma_{m,1}^2 = p_2 \sigma_{m,2}^2 = p_3 \sigma_{m,3}^2 = p_4 \sigma_{m,4}^2.$$

Jf. udtrykkene fra forrige slide

$$p_1 \frac{\sigma_v^2}{2} = p_2 \frac{\sigma_v^2}{3} = p_3 \frac{\sigma_v^2}{4} = p_4 \frac{\sigma_v^2}{6}.$$

Således kan vægtene vælges lig antal satser for hvert punkt:

$$p_1 = 2 \quad p_2 = 3 \quad p_3 = 4 \quad p_4 = 6$$

Med disse vægte gælder $\sigma_0^2 = \sigma_v^2$.

Igen kan σ_v^2 om fornødent estimeres vha. S_0^2.

Fordeling af slutfejl

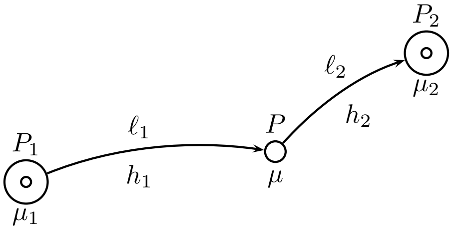
Betragt en situation, hvor vi ved, at summen af vores observationer skal være lig en kendt værdi x_0. Dette kunne eksempelvis være:

- $x_0 = 200$ 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.
- ...

Slutfejlen er afvigelsen mellem x_0 og den faktiske sum af observationer.

Nivellement — bestemmelse af kote

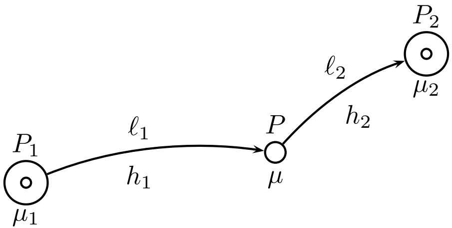
Vi ønsker at beregne koten μ til punktet P med højdeforskelle h_1 til P_1 og h_2 til P_2. Længderne fra P til de to punkter P_1 og P_2 er henholdvis ℓ_1 og ℓ_2.



Punkterne P_1 og P_2 har kendte koter, hhv. μ_1 og μ_2.

To bud på koten i P: $\mu_1 + h_1$ og $\mu_2 - h_2$.

Model



$X_1 = \mu_1 + H_1$ og $X_2 = \mu_2 - H_2$ er uafhængige stokastiske variable med $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mu$ hvor H_1 og H_2 er de stokastiske variable svarende til h_1 og h_2.

Fra før har vi, at $\mathbb{V}\mathrm{ar}(X_1) = \ell_1 \sigma_k^2$ og $\mathbb{V}\mathrm{ar}(X_2) = \ell_2 \sigma_k^2$.

Estimat af μ

Til at estimere μ anvendes det vægtede gennemsnit $\bar{x}^*$ idet varianserne på højdemålingerne ikke nødvendigvis er ens,

$$\bar{x}^* = \frac{p_1}{p_1 + p_2}(\mu_1 + h_1) + \frac{p_2}{p_1 + p_2}(\mu_2 - h_2),$$

hvor vægtene p_1 og p_2 opfylder vægtrelationen $p_1 \ell_1 \sigma_k^2 = p_2 \ell_2 \sigma_k^2$.

Dvs. de reciprokke længder kan anvendes som vægte,

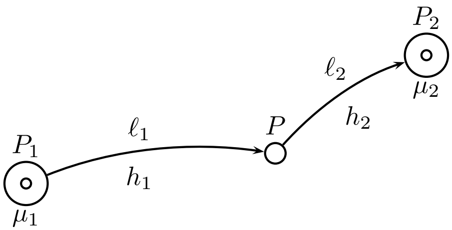
$$p_1 = \ell_1^{-1} \quad \text{og} \quad p_2 = \ell_2^{-1}.$$

Estimatet for μ er altså givet ved:

$$\bar{x}^* = \frac{\ell_1^{-1}}{\ell_1^{-1} + \ell_2^{-1}}(\mu_1 + h_1) + \frac{\ell_2^{-1}}{\ell_1^{-1} + \ell_2^{-1}}(\mu_2 - h_2).$$

Eksempel

- Højdeforskellen mellem punkterne P_1 og P_2 er opmålt ved et geometrisk nivellement over to forskellige strækninger.



Data:

- $\mu_1 = 7431$mm, $h_1 = 247$mm, $\ell_1 = 671$m
- $\mu_2 = 7828$mm, $h_2 = 156$mm, $\ell_2 = 853$m

Estimat af koten i punktet P:

$$\bar{x}^* = \frac{1/671}{1/671 + 1/853}(7431+247) + \frac{1/853}{1/671 + 1/853}(7828-156) \approx 7675,36\text{mm}$$

Estimat af μ - en omskrivning

Vi kan omskrive udtrykket for $\bar{x}^*$ således:

$$\begin{aligned}
\bar{x}^* &= \frac{\ell_1^{-1}}{\ell_1^{-1}+\ell_2^{-1}}(\mu_1+h_1)+\frac{\ell_2^{-1}}{\ell_1^{-1}+\ell_2^{-1}}(\mu_2-h_2) \\
&= \frac{\frac{1}{\ell_1}}{\frac{1}{\ell_1}+\frac{1}{\ell_2}}\frac{\ell_1\ell_2}{\ell_1\ell_2}(\mu_1+h_1)+\frac{\frac{1}{\ell_2}}{\frac{1}{\ell_1}+\frac{1}{\ell_2}}\frac{\ell_1\ell_2}{\ell_1\ell_2}(\mu_2-h_2) \\
&= \frac{\ell_2}{\ell_1+\ell_2}(\mu_1+h_1)+\frac{\ell_1}{\ell_1+\ell_2}(\mu_2-h_2)
\end{aligned}$$

Slutfejlen r

På grund af målefejl er $h_1 + h_2 \neq \mu_2 - \mu_1$. Derfor indføres **slutfejlen** r

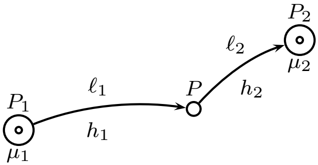
$$r = \mu_2 - \mu_1 - (h_1 + h_2) \quad \Leftrightarrow \quad \mu_2 - h_2 = \mu_1 + h_1 + r.$$

Indsættes dette i udtrykket for $\bar{x}^*$ får vi

$$\begin{aligned}
\bar{x}^* &= \frac{\ell_2}{\ell_2 + \ell_1}(\mu_1 + h_1) + \frac{\ell_1}{\ell_2 + \ell_1}(\mu_2 - h_2) \\
&= \frac{\ell_2}{\ell_2 + \ell_1}(\mu_1 + h_1) + \frac{\ell_1}{\ell_2 + \ell_1}(\mu_1 + h_1 + r) \\
&= (\mu_1 + h_1) + \frac{\ell_1}{\ell_2 + \ell_1} r
\end{aligned}$$

Slutfejlen på estimatet $\bar{x}^*$ af koten μ skal under de anvendte vægte (reciprokke længdemål) fordeles proportionalt med vejlængden.

Eksempel — fortsat



Data:

- $\mu_1 = 7431$mm, $h_1 = 247$mm, $\ell_1 = 671$m
- $\mu_2 = 7828$mm, $h_2 = 156$mm, $\ell_2 = 853$m

Slutfejl:

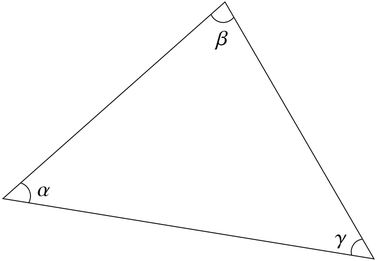
- $\mu_2 - \mu_1 = 7828 - 7431 = 397$, $h_1 + h_2 = 247 + 156 = 403$mm
- Slutfejl $r_n = 397 - 403 = -6$mm

Estimat af koten i punktet P:

$$(\mu_1 + h_1) + \frac{\ell_1}{\ell_2 + \ell_1} r_n = 7431 + 247 + \frac{671}{671 + 853}(-6) = 7431 + 247 - 2,64$$

Højdeforskelen h_1 skal altså nedkorrigeres 2,64mm.

Slutfejl af vinkelmålinger i trekant



Betragt trekanten med sande vinkler α, β og γ. Dvs $\alpha + \beta + \gamma = 200$ gon.
Lad X_α, X_β og X_γ være stokastiske variable (vinkelmålinger) med

$$\mathbb{E}(X_\alpha) = \alpha \quad \mathbb{E}(X_\beta) = \beta \quad \mathbb{E}(X_\gamma) = \gamma$$

og

$$\mathbb{V}\mathrm{ar}(X_\alpha) = \mathbb{V}\mathrm{ar}(X_\beta) = \mathbb{V}\mathrm{ar}(X_\gamma) = \sigma^2$$

Estimat af α

X_α er en direkte observation af α.

$Y_\alpha = 200 - X_\beta - X_\gamma$ er en indirekte observation af α:

$$\mathbb{E}Y_\alpha = 200 - \mathbb{E}X_\beta - \mathbb{E}X_\gamma = \alpha \quad \mathbb{V}\mathrm{ar}Y_\alpha = 2\sigma^2$$

Idet varianserne for de to observationer X_α og Y_α er forskellige anvendes vægtet gennemsnit med $p_1 = 2$ og $p_2 = 1$ (da variansen på Y_α to gange varians på X_α).

Dvs. α estimeres ved

$$\bar{x}^* = \frac{2}{3}x_\alpha + \frac{1}{3}y_\alpha = \frac{2}{3}x_\alpha + \frac{1}{3}(200 - x_\beta - x_\gamma)$$

Slutfejlen r_v på vinkelsummen

Som i foregående eksempel er summen af x_α, x_β og x_γ ikke 200 pga. uundgålige målefejl.

Derfor indføres slutfejl r_v

$$r_v = 200 - (x_\alpha + x_\beta + x_\gamma) \quad \Leftrightarrow \quad 200 - x_\beta - x_\gamma = x_\alpha + r_v,$$

som indsættes i udtrykke for $\bar{x}^*$,

$$\begin{aligned} \bar{x}^* &= \frac{2}{3}x_\alpha + \frac{1}{3}(200 - x_\beta - x_\gamma) \\ &= \frac{2}{3}x_\alpha + \frac{1}{3}(x_\alpha + r_v) \\ &= x_\alpha + \frac{1}{3}r_v. \end{aligned}$$

Dvs. slutfejlen fordeles ligeligt på alle vinkler når vinklerne er målt med samme varians (NB ikke tilfældet i eksamensopgave B).

Eksempel

$$x_\alpha = 91 \quad x_\beta = 28 \quad x_\gamma = 87$$

Slutfejl:

$$r_v = 200 - 91 - 28 - 87 = -6$$

Estimat af α:

$$x_\alpha + \frac{1}{3} r_v = 91 - 2 = 89$$

Antag spredningen på vinkelmålingerne er $\sigma = 0.1$. Da er variansen på estimatet (med $\sigma_0^2 = 2\sigma^2$)

$$\mathbb{V}\mathrm{ar}\bar{X}^* = \frac{2\sigma^2}{3} = \frac{2}{3}\sigma^2$$

og spredningen på estimatet er

$$\sqrt{\frac{2}{3}}\sigma = 0.816\sigma$$