Landmålingens fejlteori
Lektion 5
Fejlforplantning

Rasmus Waagepetersen - rw@math.aau.dk

Institut for Matematiske Fag
Aalborg Universitet

Repetition: Varians af linear kombination

Antag $X_1, X_2, \ldots, X_n$ er *uafhængige* stokastiske variable, og Y er en linearkombination af $X_1, X_2, \ldots, X_n$:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

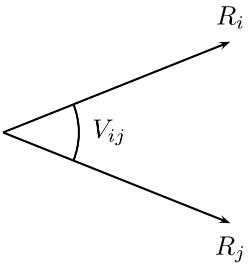
Da er variansen for Y givet ved

$$\mathbb{V}\mathrm{ar}(Y) = a_1^2 \mathbb{V}\mathrm{ar}(X_1) + a_2^2 \mathbb{V}\mathrm{ar}(X_2) + \cdots + a_n^2 \mathbb{V}\mathrm{ar}(X_n).$$

Specielt:

$$\mathbb{V}\mathrm{ar}(aX) = a^2 \mathbb{V}\mathrm{ar} X$$

Eksempel: Vinkelberegning

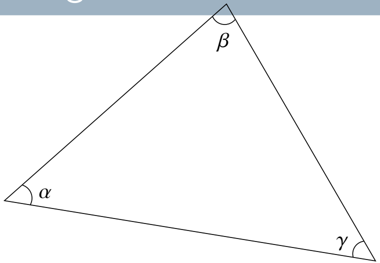


Vinkler bestemmes som differensen mellem to retningsbestemmelser. Fx. er $V_{ij} = R_j - R_i$ hvor både R_j og R_i er uafhængige stokastiske variable.

Vi antager retningerne er målt med samme nøjagtighed, dvs $\mathbb{V}\mathrm{ar}(R_j) = \mathbb{V}\mathrm{ar}(R_i) = \sigma_R^2$. Variansen på V_{ij} er givet ved

$$\sigma_V^2 = \mathbb{V}\mathrm{ar}(V_{ij}) = \mathbb{V}\mathrm{ar}(R_j - R_i) = \mathbb{V}\mathrm{ar}(R_j) + (-1)^2\mathbb{V}\mathrm{ar}(R_i) = 2\sigma_R^2.$$

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

$$\mathbb{V}\mathrm{ar}(X_\alpha) = 0.3 \quad \mathbb{V}\mathrm{ar}(X_\beta) = 0.1 \quad \mathbb{V}\mathrm{ar}(X_\gamma) = 0.3$$

Antag vi har målt $x_\alpha = 61$ $x_\beta = 72$ $x_\gamma = 65$. Da er slutfejlen

$$r = 200 - 61 - 72 - 65 = 2$$

Estimation af α:

Sæt $Y_\alpha = 200 - X_\beta - X_\gamma$. Da har Y middelværdi α og varians $0.1 + 0.3 = 0.4$.

Vægtet gennemsnit (vægte 4 henholdsvis 3 for X_α og Y_α):

$$\begin{aligned}\bar{x}^* &= \frac{4}{3+4}x_\alpha + \frac{3}{3+4}y_\alpha \\ &= \frac{4}{3+4}x_\alpha + \frac{3}{3+4}(r + x_\alpha) \\ &= x_\alpha + \frac{3}{3+4}r = 61.8571\end{aligned}$$

NB: her fordeles slutfejlen ikke ligeligt !

Varians af vægtet gennemsnit (idet vinkelmålingerne antages uafhængige):

$$\mathbb{V}\mathrm{ar}\bar{X}^* = \left(\frac{4}{3+4}\right)^2 \mathbb{V}\mathrm{ar}X_\alpha + \left(\frac{3}{3+4}\right)^2 \mathbb{V}\mathrm{ar}Y_\alpha = 0.1714$$

Alternativ beregning

$$\sigma_0^2 = 4 * 0.3 = 3 * 0.4 = 1.2$$

$$\mathbb{V}\mathrm{ar}\bar{X}^* = \frac{\sigma_0^2}{p_1 + p_2} = \frac{1.2}{7} = 0.1714$$

Fejlforplantning

Landmåling involverer ofte bestemmelse af størrelser som ikke kan måles direkte, men kan beregnes ud fra andre målinger:

- Vinkler - vha differenser af retningsmålinger.
- Arealer - vha vinkler og længder.
- Længder - vha trigonometriske relationer.
- ...

I resten af kurset gennemgår vi hvorledes fejlene på de målbare størrelser *forplanter* sig til fejlen på den interessante ikke-målbare størrelse.

Eksempelvis kan arealet, T, af en trekant bestemmes ved

$$T = \frac{1}{2} ab \sin C,$$

hvor længdemålingerne a og b samt vinklen C måles med usikkerhed.

Mere teknisk: Vi vil finde tilnærmede udtryk for (teoretiske) middelværdi og varians for de ikke-målbare størrelser på baggrund af middelværdi og varians for de målbare størrelser.

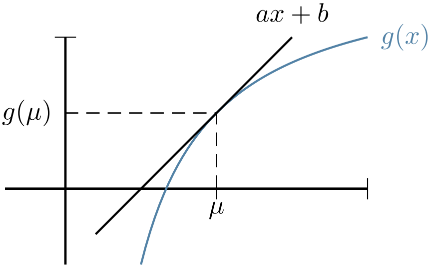
Repetition: Linearisering

Vi har tidligere set, hvordan vi finder tilnærmede udtryk for middelværdi og varians for Y, når Y *ikke* er en lineær funktion af én stokastisk 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: Flere variable

- Antag $X_1, X_2, \ldots, X_n$ er n uafhængige stokastiske variable.
- Middelværdi og varians for X_i er hhv. μ_i og σ_i^2.
- Antag Y er en funktion af $X_1, X_2, \ldots, X_n$:

$$Y = g(X_1, X_2, \ldots, X_n),$$

  hvor g er differentiabel.
- Vi ønsker at finde (tilnærmede) udtryk for middelværdi og varians for Y.
- Hvis g *ikke* er lineær i $X_1, \ldots, X_n$ kan vi ikke anvende de sædvanlige udtryk.

Linearisering: Flere variable

Løsningen er at linearisere g omkring punktet $(\mu_1, \mu_2, \ldots, \mu_n)$:

$$Y \approx g(\mu_1, \mu_2, \ldots, \mu_n) + \frac{\partial g}{\partial X_1}(X_1 - \mu_1) + \cdots + \frac{\partial g}{\partial X_n}(X_n - \mu_n),$$

hvor vi anvender notationen

$$\frac{\partial g}{\partial X_1} \equiv \frac{\partial g(x_1, x_2, \ldots, x_n)}{\partial x_i}\Big|_{x_1=\mu_1, x_2=\mu_2, \ldots, x_n=\mu_n},$$

dvs. $\frac{\partial g}{\partial X_1}$ betegner den i'te partielle afledede evalueret i punktet $(\mu_1, \mu_2, \ldots, \mu_n)$.

I praksis kender vi typisk ikke $\mu_1, \ldots, \mu_n$. I stedet anvender vi estimater, fx. gennemsnit af en eller flere målinger af μ_i.

Eksempel

For $n = 2$ svarer den lineære approksimation til at tilnærme en funktion med et tangent-plan.

$$g(x,y) = \frac{3}{4}y^2 + \frac{1}{24}y^3 - \frac{1}{32}y^4 - x^2$$

Tilnærm med tangent-plan i punktet (2,4):

$$g(2,4) = 8/3 \quad \frac{\partial g}{\partial x}(2,4) = -4 \quad \frac{\partial g}{\partial y}(2,4) = 0$$

$$g(x,y) \approx 8/3 - 4(x-2)$$

Fejlforplantningsloven

Vi kan approximere variansen på Y

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(Y) &\approx \mathbb{V}\mathrm{ar}\left(g(\mu_1,\mu_2,\dots,\mu_n) + \frac{\partial g}{\partial X_1}(X_1-\mu_1) + \cdots + \frac{\partial g}{\partial X_n}(X_n-\mu_n)\right) \\
&= \mathbb{V}\mathrm{ar}\Bigg(\underbrace{g(\mu_1,\mu_2,\dots,\mu_n) - \frac{\partial g}{\partial X_1}\mu_1 - \cdots - \frac{\partial g}{\partial X_n}\mu_n}_{\text{konstant}} + \\
&\qquad\qquad \frac{\partial g}{\partial X_1}X_1 + \cdots + \frac{\partial g}{\partial X_n}X_n\Bigg) \\
&= \mathbb{V}\mathrm{ar}\left(\frac{\partial g}{\partial X_1}X_1 + \cdots + \frac{\partial g}{\partial X_n}X_n\right) \\
&= \left(\frac{\partial g}{\partial X_1}\right)^2 \mathbb{V}\mathrm{ar}(X_1) + \cdots + \left(\frac{\partial g}{\partial X_n}\right)^2 \mathbb{V}\mathrm{ar}(X_n)
\end{aligned}
$$

Den simple fejlforplantningslov

Antag $X_1, X_2, \ldots, X_n$ er n *uafhængige* stokastiske variable, hvor $\mathbb{V}\mathrm{ar}(X_1) = \sigma_1^2, \ldots, \mathbb{V}\mathrm{ar}(X_n) = \sigma_n^2$. Lad $Y = g(X_1, X_2, \ldots, X_n)$, hvor g er en differentiabel funktion. Et tilnærmet udtryk for variansen for $\mathbb{V}\mathrm{ar}(Y) = \sigma_Y^2$ er

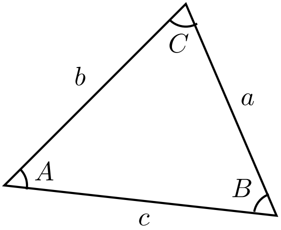
$$\sigma_Y^2 \approx \left(\frac{\partial g}{\partial X_1}\right)^2 \sigma_1^2 + \cdots + \left(\frac{\partial g}{\partial X_n}\right)^2 \sigma_n^2$$

Bemærk: For en lineær transformation $Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$ gælder $\frac{\partial g}{\partial X_1} = a_1, \ldots, \frac{\partial g}{\partial X_n} = a_n$. I dette tilfælder giver ovenstående resultatet det samme som den sædvanlige variansregel for en linearkombination af stokastiske variabel.

Bemærk: X_i's bidrag til den samlede varians (for Y) afhænger af

1. Variansen σ_i^2 for X_i.
2. Hvor sensitiv Y er overfor ændringer i X_i - målt ved $|\frac{\partial g}{\partial X_i}|$

Fejlforplantning ved arealbestemmelse



Arealet T kan bestemmes på flere måder:

- $T = \frac{1}{2}ab \sin C \quad (1)$
- $T = \frac{1}{2}ac \sin B \quad (2)$
- $T = \frac{1}{2}bc \sin A \quad (3)$

Fejlforplantning anvendt på (1)

Vi analyserer arealudtrykket (1): $T = \frac{1}{2} ab \sin C$.

Jf. fejlforplantningsloven gælder der,

$$\sigma_T^2 \approx \left(\frac{\partial T}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial T}{\partial b}\right)^2 \sigma_b^2 + \left(\frac{\partial T}{\partial C}\right)^2 \sigma_C^2$$

De partielt afledte er:

$$\frac{\partial T}{\partial a} = \frac{1}{2} b \sin C = \frac{\frac{1}{2} ab \sin C}{a} = \frac{T}{a}$$
$$\frac{\partial T}{\partial b} = \frac{1}{2} a \sin C = \frac{\frac{1}{2} ab \sin C}{b} = \frac{T}{b}$$
$$\frac{\partial T}{\partial C} = \frac{1}{2} ab \cos C \frac{1}{\omega} = \frac{1}{2} ab \sin C \frac{\cos C}{\sin C} \frac{1}{\omega} = T \frac{\cos C}{\sin C} \frac{1}{\omega} = \frac{T}{\tan C} \frac{1}{\omega}, \ [\omega = \frac{200}{\pi}]$$

Sidste omskrivning gælder idet

$$\tan x = \frac{\sin x}{\cos x} \quad \Leftrightarrow \quad \frac{1}{\tan x} = \frac{\cos x}{\sin x}.$$

Estimat af T

I et eksempel fra noterne er følgende oplysninger givet:

$$a = 115.53\,\text{m}, \quad b = 152.17\text{m} \qquad \sigma_a = \sigma_b = 1\text{cm}.$$
$$C = 93.273\,\text{gon} \qquad \sigma_C = 0.002\text{gon}.$$

Dvs. vi kan regne estimatet for T som

$$T = \frac{1}{2}ab\sin C = \frac{1}{2}115.53\text{m} \times 152.17\text{m} \times \sin 93.273 = 8741.072\text{m}^2$$

Variansen σ_T^2

Variansen σ_T^2 på estimatet T er fra forrige slide påvirket af σ_a, σ_b og σ_C på følgende måde,

$$\begin{aligned}\sigma_T^2 &\approx \left(\frac{T}{a}\right)^2 \sigma_a^2 + \left(\frac{T}{b}\right)^2 \sigma_b^2 + \left(\frac{T}{\tan C}\frac{1}{\omega}\right)^2 \sigma_C^2 \\ &= \left(\frac{8741\text{m}^2}{115.53\text{m}}\right)^2 (0.01\text{m})^2 + \left(\frac{8741\text{m}^2}{152.17\text{m}}\right)^2 (0.01\text{m})^2 + \left(\frac{8741\text{m}^2}{\tan 93.273}\frac{1}{\omega}\right)^2 (0.002\text{gon})^2 \\ &= 0.57\text{m}^4 + 0.33\text{m}^4 + 0.0008\text{m}^4 \\ &= 0.9033\text{m}^4.\end{aligned}$$

Standard-afvigelsen er:

$$\sigma_T = \sqrt{\sigma_T^2} = 0.9504\text{m}^2.$$

Eksempel - Extended edition

Antag nu at målene i trekanten er målt således:

	a	b	c	A	B	C
1	115.5603	152.1643	181.1362	43.7505	62.9538	93.2993
2	115.5397	152.1410	181.1246	43.7485	62.9498	93.3008
3	115.5527	152.1700	181.1312		62.9514	93.2730
4	115.5350		181.1141		62.9511	
5	115.5341		181.1138			
6	115.5431					
7	115.5519					
8	115.5300					
$\bar{x}$	115.5434	152.1584	181.1240	43.7495	62.9515	93.2910
$\mathbb{V}\text{ar}(\bar{X})$	$\frac{\sigma_a^2}{8}$	$\frac{\sigma_b^2}{3}$	$\frac{\sigma_c^2}{5}$	$\frac{\sigma_A^2}{2}$	$\frac{\sigma_B^2}{4}$	$\frac{\sigma_C^2}{3}$

Estimater af arealet T

Fra tidligere kan arealet T beregnes på mindst tre måder (1)-(3). Hvis vi anvender gennemsnitsmålingerne fra forrige slide har vi:

$$(1)\ T_1 = \frac{1}{2}ab\sin C = 115.5434 \times 152.1584 \times \sin 93.2910 = 8741.681\text{m}^2$$

$$(2)\ T_2 = \frac{1}{2}ac\sin B = 115.5434 \times 181.1240 \times \sin 62.9515 = 8741.376\text{m}^2$$

$$(3)\ T_3 = \frac{1}{2}bc\sin A = 152.1584 \times 181.1240 \times \sin 43.7495 = 8741.710\text{m}^2$$

Vægte til estimat af T vha $\bar{x}^*$

Tidligere så vi hvordan σ_T^2 blev bestemt for (1) med data fra noternes eksempel. Nedenfor bestemmes $\sigma_{T_1}^2$, $\sigma_{T_2}^2$ og $\sigma_{T_3}^2$:

$$\sigma_{T_1}^2 \approx \left(\frac{8741}{115.54}\right)^2 \frac{0.01^2}{8} + \left(\frac{8741}{152.16}\right)^2 \frac{0.01^2}{3} + \left(\frac{8741}{\tan 93.29}\frac{1}{\omega}\right)^2 \frac{0.002^2}{3}$$
$$= 0.1818502\text{m}^4.$$

$$\sigma_{T_2}^2 \approx \left(\frac{8741}{115.54}\right)^2 \frac{0.01^2}{8} + \left(\frac{8741}{181.12}\right)^2 \frac{0.01^2}{5} + \left(\frac{8741}{\tan 62.95}\frac{1}{\omega}\right)^2 \frac{0.002^2}{4}$$
$$= 0.1262968\text{m}^4.$$

$$\sigma_{T_3}^2 \approx \left(\frac{8741}{152.16}\right)^2 \frac{0.01^2}{3} + \left(\frac{8741}{181.12}\right)^2 \frac{0.01^2}{5} + \left(\frac{8741}{\tan 43.75}\frac{1}{\omega}\right)^2 \frac{0.002^2}{2}$$
$$= 0.2126244\text{m}^4.$$

Vægte til estimat af T vha $\bar{x}^*$

Vægtene i det vægtede gennemsnit $\bar{x}^*$ skal opfylde vægtrelationen,

$$p_1\sigma_{T_1}^2 = p_2\sigma_{T_2}^2 = p_3\sigma_{T_3}^2.$$

Fx. kan vi vælge $p_1 = 1$, hvilket medfører at

$$p_2 = \frac{\sigma_{T_1}^2}{\sigma_{T_2}^2} = \frac{0.1819}{0.1263} = 1.4399 \quad \text{og} \quad p_3 = \frac{\sigma_{T_1}^2}{\sigma_{T_3}^2} = \frac{0.1819}{0.2126} = 0.8553$$

Således er $p_+ = 1 + 1.4399 + 0.8553 = 3.2952$ og estimatet af T,

$$\begin{aligned} \bar{x}^* &= \frac{p_1}{p_+}T_1 + \frac{p_2}{p_+}T_2 + \frac{p_3}{p_+}T_3 \\ &= \frac{1}{3.2952}8741.681 + \frac{1.4399}{3.2952}8741.376 + \frac{0.8553}{3.2952}8741.710 \\ &= 8741.55\text{m}^2 \end{aligned}$$

Konfidensinterval for T

For at udregne konfidensinterval for T har vi brug for at kende variansen af $\bar{X}^*$.

Det er nemt hvis T_1, T_2 og T_3 kan betragtes som realisationer af uafhængige stokastiske variable - men dette er ikke tilfældet ! (hvorfor ?)

Vender tilbage til udregning af variansen i lektion 6.

Varians af afhængige variable

Betragt X_1 og X_2 som ikke antages at være uafhængige. Dette betyder at $\mathbb{V}\mathrm{ar}(X_1 + X_2)$ bliver mere kompliceret.

Lad $\mathbb{E}(X_1) = \mu_1$ og $\mathbb{E}(X_2) = \mu_2$:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(X_1 + X_2) &= \mathbb{E}\left[\left[(X_1 + X_2) - (\mu_1 + \mu_2)\right]^2\right] \\
&= \mathbb{E}\left[\left[(X_1 - \mu_1) + (X_2 - \mu_2)\right]^2\right] \\
&= \mathbb{E}\left[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)\right] \\
&= \mathbb{E}[(X_1 - \mu_1)^2] + \mathbb{E}[(X_2 - \mu_2)^2] + 2\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= \mathbb{V}\mathrm{ar}(X_1) + \mathbb{V}\mathrm{ar}(X_2) + 2\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]
\end{aligned}
$$

Middelværdien

$$\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]$$

kaldes *kovariansen* mellem X_1 og X_2.

Hvis X_1 og X_2 er uafhængige gælder

$$\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \mathbb{E}[X_1 - \mu_1]\mathbb{E}[X_2 - \mu_2] = 0$$

hvorved $\mathbb{V}\mathrm{ar}(X_1 + X_2) = \mathbb{V}\mathrm{ar}(X_1) + \mathbb{V}\mathrm{ar}(X_2)$

Kovarians

Definition: Kovarians

Antag X_1 og X_2 er to stokastiske variable med middelværdier hhv. μ_1 og μ_2. Kovariansen mellem X_1 og X_2 er da defineret som

$$\mathbb{C}\mathrm{ov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)].$$

Notation: Kovariansen mellem to stokastiske variable X_1 og X_2 betegnes ofte σ_{12}.

Egenskab: Kovarians og uafhængighed

Hvis X_1 og X_2 er uafhængige, så er $\mathbb{C}\mathrm{ov}(X_1, X_2) = 0$.

Fortolkning af kovarians

Kovarians > 0: $(X_1 - \mu_1)$ og $(X_2 - \mu_2)$ har overvejende samme fortegn.

Kovarians < 0: $(X_1 - \mu_1)$ og $(X_2 - \mu_2)$ har overvejende modsat fortegn.

Dvs. kovarians er et mål for om X_1 og X_2 er positivt eller negativt associerede.

Kovarians: Regneregler

Kovariansen mellem a_1X_1 og a_2X_2 er

$$\mathbb{C}\text{ov}(a_1X_1, a_2X_2) = a_1a_2\mathbb{C}\text{ov}(X_1, X_2).$$

Det medfører at variansen for $a_1X_1 + a_2X_2$ er

$$\begin{aligned}\mathbb{V}\text{ar}(a_1X_1 + a_2X_2) &= \mathbb{V}\text{ar}(a_1X_1) + \mathbb{V}\text{ar}(a_2X_2) + 2\mathbb{C}\text{ov}(a_1X_1, a_2X_2)\\ &= a_1^2\mathbb{V}\text{ar}(X_1) + a_2^2\mathbb{V}\text{ar}(X_2) + 2a_1a_2\mathbb{C}\text{ov}(X_1, X_2).\end{aligned}$$

Specialtilfælde:

$$\mathbb{V}\text{ar}(aX) = \mathbb{C}\text{ov}(aX, aX) = a \cdot a \cdot \mathbb{C}\text{ov}(X, X) = a^2\mathbb{V}\text{ar}X$$

Kovarians for summer:

$$\mathbb{C}\text{ov}(X, Y + Z) = \mathbb{C}\text{ov}(X, Y) + \mathbb{C}\text{ov}(X, Z)$$

Antag $\mathbb{C}\mathrm{ov}(X_1, X_2) = 2$. Så er kovarians mellem $3X_1$ og $4X_2$ 12 gange så stor.

Kovarians afhænger af skalering og er derfor ikke et velegnet mål for graden af afhængighed mellem to stokastiske variable - vi kan få vilkårlig stor kovarians ved blot at gange de stokastiske variable med store konstanter.

Kovarians afhænger også af enheder. Kan virke underligt at måle 'afhængighed' i f.eks. gon^2.

Korrelation

Definition: Korrelation

Antag X_1 og X_2 er to stokastiske variable med varianser σ_1^2 and σ_2^2. Korrelation $\mathbb{C}\mathrm{orr}(X_1, X_2)$ mellem de stokastiske variable X_1 og X_2 er defineret som

$$\mathbb{C}\mathrm{orr}(X_1, X_2) = \frac{\mathbb{C}\mathrm{ov}(X_1, X_2)}{\sqrt{\mathbb{V}\mathrm{ar}[X_1]\mathbb{V}\mathrm{ar}[X_2]}}.$$

Ofte skrives korrelation som $\rho = \mathbb{C}\mathrm{orr}(X_1, X_2)$.

Med notationen $\mathbb{V}\mathrm{ar}[X_i] = \sigma_i^2$ og $\mathbb{C}\mathrm{ov}(X_1, X_2) = \sigma_{12}$ er korrelationen mellem X_1 og X_2

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}.$$

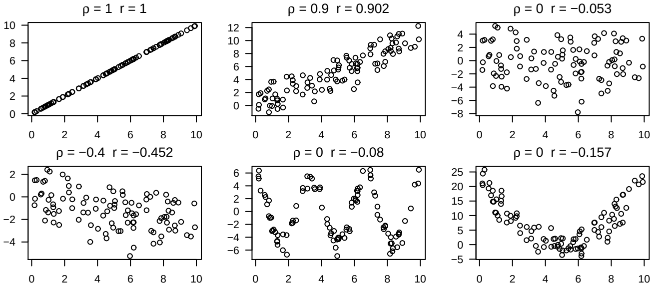
Hvis vi kender korrelation og varianserne har vi kovariansen som:

$$\sigma_{12} = \sigma_1 \sigma_2 \rho.$$

Korrelation: Egenskaber og Eksempler

- korrelation måler kovarians i forhold til varianser
- korrelation uafhængig af enheder
- $-1 \leq \rho \leq 1$
- Mål for graden af *lineær* sammenhæng.
- $\rho = -1$ og $\rho = 1 \Rightarrow$ perfekt lineær sammenhæng.
- Uafhængighed $\Rightarrow \rho = 0$.

Eksempler: ρ er korrelationen i populationen og r er den estimerede korrelation for de viste stikprøver.



NB: de sidste (lidt patologiske) eksempler viser, at $\mathbb{C}\mathrm{ov}(X_1, X_2) = 0$ eller $\mathbb{C}\mathrm{orr}(X_1, X_2) = 0$ ikke medfører uafhængighed.

Eksempel: $\mathbb{V}\mathrm{ar} X_1 = 2$, $\mathbb{V}\mathrm{ar} X_2 = 3$ $\rho = 0.5$

$$\mathbb{V}\mathrm{ar}(X_1 + X_2) = 2 + 3 + 2\sqrt{2}\sqrt{3} \cdot 0.5 = 7.45$$

Dvs. positiv korrelation *øger* varians af sum.

Med $\rho = -0.5$ fås

$$\mathbb{V}\mathrm{ar}(X_1 + X_2) = 2 + 3 - 2\sqrt{2}\sqrt{3} \cdot 0.5 = 2.55$$

Dvs. negativ korrelation *mindsker* varians af sum.

Fejlforplantning: Afhængige målinger

Tidligere har vi set $Y = g(X_1, X_2)$ hvor g er en transformation af X_1 og X_2.

En lineær approximation af Y omkring punktet (μ_1, μ_2) er givet ved:

$$Y = g(X_1, X_2) \approx g(\mu_1, \mu_2) + \frac{\partial g}{\partial X_1}(X_1 - \mu_1) + \frac{\partial g}{\partial X_2}(X_2 - \mu_2)$$

Hvis X_1 og X_2 er *uafhængige* har vi set, at

$$\mathbb{V}\mathrm{ar}(Y) \approx \left(\frac{\partial g}{\partial X_1}\right)^2 \mathbb{V}\mathrm{ar}(X_1) + \left(\frac{\partial g}{\partial X_2}\right)^2 \mathbb{V}\mathrm{ar}(X_2)$$

Hvordan ser det ud, hvis X_1 og X_2 er *afhængige*, dvs. når $\mathbb{C}\mathrm{ov}(X_1, X_2) \neq 0$?

Fejlforplantning - fortsat

Hvis $\mathbb{C}\mathrm{ov}(X_1, X_2) = \sigma_{12} \neq 0$ bliver variansen af Y:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(Y) &\approx \mathbb{V}\mathrm{ar}\left(g(\mu_1, \mu_2) + \frac{\partial g}{\partial X_1}(X_1 - \mu_1) + \frac{\partial g}{\partial X_2}(X_2 - \mu_2)\right) \\
&= \mathbb{V}\mathrm{ar}\left(\underbrace{g(\mu_1, \mu_2) - \frac{\partial g}{\partial X_1}\mu_1 - \frac{\partial g}{\partial X_2}\mu_2}_{\text{konstant}} + \frac{\partial g}{\partial X_1}X_1 + \frac{\partial g}{\partial X_2}X_2\right) \\
&= \mathbb{V}\mathrm{ar}\left(\frac{\partial g}{\partial X_1}X_1 + \frac{\partial g}{\partial X_2}X_2\right) \\
&= \left(\frac{\partial g}{\partial X_1}\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial X_2}\right)^2 \sigma_2^2 + 2\frac{\partial g}{\partial X_1}\frac{\partial g}{\partial X_2}\mathbb{C}\mathrm{ov}(X_1, X_2) \\
&= \left(\frac{\partial g}{\partial X_1}\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial X_2}\right)^2 \sigma_2^2 + 2\frac{\partial g}{\partial X_1}\frac{\partial g}{\partial X_2}\sigma_{12}
\end{aligned}$$

Variansen af afhængige variable

Varians for linearkombination af afhængige SV

Antag $X_1, X_2, \ldots, X_n$ er stokastiske variable, med varianser $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. Antag desuden, at kovariansen mellem X_i og X_j er $\mathbb{C}\mathrm{ov}(X_i, X_j) = \sigma_{ij}$.
Variansen for linearkombinationen er

$$\begin{aligned}\mathbb{V}\mathrm{ar}(a_0 + a_1 X_1 + \cdots + a_n X_n) &= a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 + \sum_{i\neq j} a_i a_j \sigma_{ij} \\ &= \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i<j} a_i a_j \sigma_{ij}\end{aligned}$$

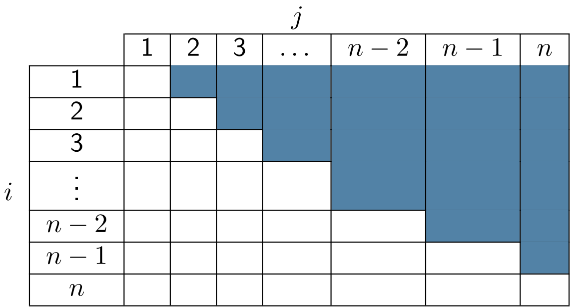
Bemærk, at summen $\sum_{i<j} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}$ betyder at vi lader i gennemløbe alle index $1, \ldots, n$ og j skal hver gang være strengt større end i.

$\sum_{i<j}$

Variansen er altså

$$\mathbb{V}\mathrm{ar}(a_0 + a_1 X_1 + \cdots + a_n X_n) = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 + 2\sum_{i<j} a_i a_j \sigma_{ij}.$$

Summen $\sum_{i<j}$ svarer til de *mørke* celler nedenfor. **Bemærk**: $\sigma_{ij} = \sigma_{ji}$.



Vilkårlig transformation

Er transformationen af $Y = g(X_1, \dots, X_n)$ en vilkårlig differentiabel funktion laver vi som tidligere en lineær approximation af g omkring middelværdierne af X_i'erne $(\mu_1, \dots, \mu_n)$,

$$g(X_1, \dots, X_n) \approx g(\mu_1, \dots, \mu_n) + \frac{\partial g}{\partial X_1}(X_1 - \mu_1) + \dots + \frac{\partial g}{\partial X_n}(X_n - \mu_n)$$

Dette svarer til at

$$g(X_1, \dots, X_n) \approx a_0 + a_1 X_1 + \dots + a_n X_n,$$

hvor

$$a_0 = g(\mu_1, \dots, \mu_n) - \frac{\partial g}{\partial X_1}\mu_1 - \dots - \frac{\partial g}{\partial X_n}\mu_n$$

$$a_i = \frac{\partial g}{\partial X_i} \quad i = 1, \dots, n$$

Vi har altså

$$g(X_1, \dots, X_n) \approx a_0 + a_1 X_1 + \cdots + a_n X_n,$$

hvor

$$\begin{aligned} a_0 &= g(\mu_1, \dots, \mu_n) - \frac{\partial g}{\partial X_1}\mu_1 - \dots - \frac{\partial g}{\partial X_n}\mu_n \\ a_i &= \frac{\partial g}{\partial X_i} \quad i = 1, \dots, n \end{aligned}$$

Dvs. et tilnærmet udtryk for variansen, σ_Y^2, af Y er

$$\begin{aligned} \sigma_Y^2 &\approx \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j \sigma_{ij} \\ &= \left(\frac{\partial g}{\partial X_1}\right)^2 \sigma_1^2 + \cdots + \left(\frac{\partial g}{\partial X_n}\right)^2 \sigma_n^2 + 2 \sum_{i<j} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} \sigma_{ij} \end{aligned}$$

Eksempel: X_a og X_b betegner målinger af sidelængderne a og b (kateterne) i en retvinklet trekant. Et estimat for hypotenusen c er da

$$\sqrt{x_a^2 + x_b^2}$$

Antag X_a og X_b har samme spredning 0.05 og korrelation $\rho = 0.2$.

Med $c = \sqrt{a^2 + b^2}$ fås

$$\frac{\partial c}{\partial a} = \frac{a}{\sqrt{a^2 + b^2}} \quad \frac{\partial c}{\partial b} = \frac{b}{\sqrt{a^2 + b^2}}$$

Dermed fås en tilnærmet spredning på estimatet af c som

$$\left(\frac{x_a}{\sqrt{x_a^2 + x_b^2}}\right)^2 0.05^2 + \left(\frac{x_b}{\sqrt{x_a^2 + x_b^2}}\right)^2 0.05^2 + 2 \cdot 0.05^2 \cdot 0.2 \frac{x_a}{\sqrt{x_a^2 + x_b^2}} \frac{x_b}{\sqrt{x_a^2 + x_b^2}}$$

Med $x_a = 3.1$ og $x_b = 4.5$ er estimatet for c 5.46 og variansen for estimatet af c bliver 0.00296718.

Hvis $\rho = 0$ bliver variansen kun 0.0025.