Korrelasjonsmodeller

Arne Bang Huseby

Standard normalfordelte variable

Gitt en vektor, $\boldsymbol{X}$, av n stokastiske variable $X_1, \ldots, X_n$, der vi i første omgang antar at:

$$X_i \sim N(0,1),\ i = 1, \ldots, n. \tag{1}$$

Vi skal nå se på ulike korrelasjonsmodeller for $X_1, \ldots, X_n$, og hvordan disse kan representeres effektivt i forbindelse med simuleringer.

Vi ser først på det generelle tilfellet der C betegner korrelasjonsmatrisen for $\boldsymbol{X}$ (som i dette tilfelle også er det samme som kovariansmatrisen siden alle varianser er lik 1). Anta så at C kan skrives på formen:

$$C = MM^T, \tag{2}$$

der $M = M^{n \times m}$ er en passende matrise. I så fall kan vektoren $\boldsymbol{X}$ representeres som:

$$\boldsymbol{X} = M \cdot \boldsymbol{U} \tag{3}$$

der $\boldsymbol{U} = (U_1, \ldots, U_m)$ er en vektor av m uavhengige standard normalfordelte stokastiske variable.

Dette følger fordi lineærtransformasjoner av multinormalfordelte variable er multinormalfordelte og fordi:

$$
\begin{aligned}
\text{Cov}(M\boldsymbol{U}) &= E[(M\boldsymbol{U})(M\boldsymbol{U})^T] \qquad (4)\\
&= ME[\boldsymbol{U}\boldsymbol{U}^T]M^T\\
&= MIM^T\\
&= MM^T = C,
\end{aligned}
$$

der vi har benyttet at kovariansmatrisen til en vektor $\boldsymbol{Y}$ der $E\boldsymbol{Y} = \boldsymbol{0}$ er lik $E[\boldsymbol{Y} \cdot \boldsymbol{Y}^T]$.

Av spesiell interesse er den såkalte *Cholesky-dekomposisjonen*. Siden en korrelasjonsmatrise alltid er positivt definit (eller i det minste ikke-negativt definit), finnes det en nedre triangulær matrise $L = L^{n \times n}$ slik at:

$$C = L \cdot L^T \tag{5}$$

Vektoren $\boldsymbol{X}$ kan dermed representeres som $\boldsymbol{X} = L \cdot \boldsymbol{U}$, der $\boldsymbol{U}$ nå er en vektor av n uavhengige standard normalfordelte variable.

Bemerk at det alltid finnes en og bare en nedre triangulær $n \times n$-matrise L som tilfredsstiller betingelsen (5).

For å vise entydighet, setter vi opp ligningene som følger av (5). La $C = \{c_{i,j}\}_{i,j=1}^{n}$. Elementene i matrisen $L = \{\ell_{i,j}\}_{i,j=1}^{n}$ må da tilfredsstille følgende ligninger:

$$\ell_{i,i} = [c_{i,i} - \sum_{k=1}^{i-1} \ell_{i,k}^2]^{1/2}, \tag{6}$$

$$\ell_{j,i} = \frac{1}{\ell_{i,i}}[c_{i,j} - \sum_{k=1}^{i-1} \ell_{i,k}\ell_{j,k}], \tag{7}$$

$$j = i+1, i+2, \ldots, n,$$

der $i = 1, 2, \ldots, n$.

Dersom ligningene (6) og (7) løses fortløpende for $i = 1, 2, \ldots, n$, vil høyresidene av ligningene kun inneholde ledd som er bestemt tidligere i prosessen. Såsant vi har at:

$$c_{i,i} \geq \sum_{k=1}^{i-1} \ell_{i,k}^2,\ i = 1, 2, \ldots, n, \tag{8}$$

så vil ligningssettet dermed ha en entydig løsning.

Betingelsen (8) vil være oppfylt dersom matrisen C er ikke-negativt definit.

Det er en nær sammenheng mellom Cholesky-dekomposisjon og det å finne betingede fordelinger i multinormalfordelingen. For å se dette starter vi med å innføre vi følgende subvektorer:

$$\boldsymbol{X}_k = (X_1, \dots, X_k),\ k = 1, \dots, n.$$

$$\boldsymbol{U}_k = (U_1, \dots, U_k),\ k = 1, \dots, n.$$

Vi lar videre C_k betegne korrelasjonsmatrisen til subvektoren $\boldsymbol{X}_k,\ k = 1, \dots, n$.

Vi blokkdeler så korrelasjonsmatrisen for $\boldsymbol{X}_{k+1}$ som følger:

$$C_{k+1} = \left[\begin{array}{c|c} C_k & C_{k,k+1} \\ \hline C_{k+1,k} & 1 \end{array}\right], \qquad (9)$$

der:

$$\begin{aligned} C_{k+1,k} &= C^T_{k,k+1} \qquad (10) \\ &= (\mathrm{Cov}(X_{k+1}, X_1), \ldots, \mathrm{Cov}(X_{k+1}, X_k)). \end{aligned}$$

Av elementære egenskaper ved multinormal-fordelingen følger det nå at:

$$X_{k+1}|\boldsymbol{X}_k = \boldsymbol{x}_k \sim N(a_{k+1,k}\boldsymbol{x}_k, b_{k+1,k}), \quad (11)$$

der:

$$a_{k+1,k} = C_{k+1,k}C_k^{-1}, \quad (12)$$

og

$$b_{k+1,k} = (1 - C_{k+1,k}C_k^{-1}C_{k,k+1})^{1/2}. \quad (13)$$

Bemerk at uttrykket for $b_{k+1,k}$ alltid er veldefinert fordi $1 - C_{k+1,k} C_k^{-1} C_{k,k+1} \geq 0$. Dette følger fordi:

$$
\begin{aligned}
0 \leq \mathrm{Var}(X_{k+1} - C_{k+1,k} C_k^{-1} \boldsymbol{X}_k) \\
&= 1 + C_{k+1,k} C_k^{-1} C_k C_k^{-1} C_{k,k+1} \\
&\quad -2\,\mathrm{Cov}(X_{k+1}, C_{k+1,k} C_k^{-1} \boldsymbol{X}_k) \\
&= 1 + C_{k+1,k} C_k^{-1} C_{k,k+1} \\
&\quad -2 C_{k+1,k} C_k^{-1} C_{k,k+1} \\
&= 1 - C_{k+1,k} C_k^{-1} C_{k,k+1}
\end{aligned}
$$

Basert på disse relasjonene kan vi nå konstruere en modell for $\boldsymbol{X}$ som følger:

$$
\begin{aligned}
X_1 &= U_1 \qquad (14)\\
X_2 &= a_{2,1}\boldsymbol{X}_1 + b_{2,1}U_2\\
&\cdots\\
X_n &= a_{n,n-1}\boldsymbol{X}_{n-1} + b_{n,n-1}U_n
\end{aligned}
$$

Bemerk at X_k kun avhenger av $U_1, \ldots, U_k$ for $k = 1, \ldots, n$. Dette betyr at det finnes en triangulær matrise, L, slik at $\boldsymbol{X} = L\boldsymbol{U}$. Ved entydigheten av Cholesky-dekomposisjonen, må denne dermed være lik L.

Vi kan også bruke (14) til å finne L rekursivt. La derfor L_k betegne Cholesky-dekomposisjonen av C_k. Dvs. vi har:

$$\boldsymbol{X}_k = L_k \boldsymbol{U}_k,\ k = 1, \ldots, n. \tag{15}$$

Siden $X_1 = U_1$, er selvsagt $L_1 = 1$. Anta så at vi har funnet L_k og skal finne L_{k+1}. De k første radene av L_{k+1} finner vi da umiddelbart ved å ta radene i L_k og føye til en null ytterst til høyre i hver rad. Den siste raden finnes til slutt av (14) ved å sette inn $\boldsymbol{X}_k = L_k \boldsymbol{U}_k$.

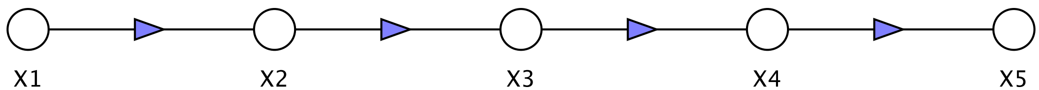
Skrevet ut ved hjelp av blokkdelte matriser får vi dermed følgende rekursjonsformel:

$$L_{k+1} = \left[\begin{array}{c|c} L_k & 0 \\ \hline a_{k+1,k} L_k & b_{k+1,k} \end{array}\right], \qquad (16)$$

for $k = 2, \ldots, n$.

Markov-kjeder

En vanlig brukt avhengighetsstruktur er Markov-kjeder, dvs. sekvenser av stokastiske variable der enhver variabel i sekvensen avhenger av de foregående kun via den forrige variablen.



Influensdiagram for en Markov-kjede

Vi skal nå konstruere en enkel additiv Markov-kjede. La $U_1, \ldots, U_n$ være n uavhengige standard normalfordelte stokastiske variable, og definer:

$$
\begin{aligned}
X_1 &= U_1 \\
X_2 &= aX_1 + bU_2 \\
\cdots \\
X_n &= aX_n + bU_n
\end{aligned}
\tag{17}
$$

der konstantene a og b velges slik at kjeden får den ønskede korrelasjonsstruktur.

Variansene til variablene i Markov-kjeden blir nå gitt ved:

$$\mathrm{Var}(X_1) = \mathrm{Var}(U_1) \tag{18}$$
$$\mathrm{Var}(X_k) = a^2 \mathrm{Var}(X_{k-1}) + b^2 \mathrm{Var}(U_k),$$
$$k = 2, \ldots, n.$$

Vi krever så at alle variablene i kjeden skal ha varians lik 1. Dette leder til følgende betingelse på a og b:

$$a^2 + b^2 = 1. \tag{19}$$

Kovariansen mellom to påhverandre følgende ledd i kjeden blir gitt ved:

$$\begin{aligned}\mathrm{Cov}(X_k, X_{k-1}) &= \mathrm{Cov}(aX_{k-1} + bU_k, X_{k-1}) \\ &= a\,\mathrm{Cov}(X_{k-1}, X_{k-1}) \\ &= a\,\mathrm{Var}(X_{k-1}) \\ &= a. \qquad (20)\end{aligned}$$

for $k = 2, 3, \ldots, n$.

Vi krever videre at:

$$\text{Cov}(X_k, X_{k-1}) = \rho,\ k = 2, 3, \ldots, n. \qquad (21)$$

Av dette, samt av (19), følger det dermed at:

$$a = \rho. \qquad (22)$$

$$b = (1 - \rho^2)^{1/2} \qquad (23)$$

Bemerk spesielt at både a og b er veldefinerte for alle mulige korrelasjonskoeffisienter, dvs. for alle ρ som er slik at:

$$-1 \leq \rho \leq 1. \qquad (24)$$

Vi ser nå litt nærmere på korrelasjonsstrukturen i Markov-kjeden. Dersom $j < k$, har vi at:

$$\begin{aligned} \mathrm{Cov}(X_k, X_j) &= \mathrm{Cov}(aX_{k-1} + bU_k, X_j) \\ &= a\,\mathrm{Cov}(X_{k-1}, X_j). \end{aligned} \tag{25}$$

Ved å gjenta denne operasjonen tilstrekkelig mange ganger, og til sist kombinere dette med (20), ser vi at vi får følgende sammenheng for alle $j < k$:

$$\mathrm{Cov}(X_k, X_j) = a^{k-j} \tag{26}$$

Innsatt $a = \rho$ blir korrelasjonsmatrisen til $\boldsymbol{X}$ dermed gitt ved:

$$C = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \quad (27)$$

Denne strukturen er den første av flere eksempler der korrelasjonen mellom to variable er avhengig av “avstanden” mellom variablene. I dette tilfellet er avstandsbegrepet knyttet til tidsavstand i kjeden.

Mer spesifikt kan vi innføre følgende avstandsmål mellom variablene:

$$d(X_i, X_j) = |i - j|,\ i, j = 1, \ldots, n. \tag{28}$$

Kovariansen mellom to variable kan da uttrykkes på formen:

$$\mathrm{Cov}(X_i, X_j) = \rho^{d(X_i, X_j)},\ i, j = 1, \ldots, n. \tag{29}$$

Siden alle avstander i dette tilfellet er heltall, så er (29) definert både for positive og negative verdier av ρ. Imidlertid gir nok avstandsfortolkningen mest mening når $0 < \rho < 1$. I dette tilfellet avtar korrelasjonen monotont med økende avstand.

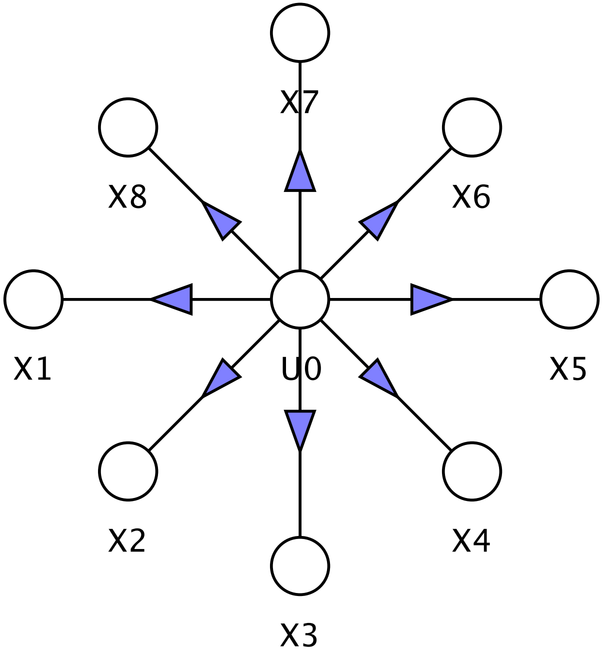
Matrisen C er opplagt positivt definit siden vi faktisk har vist at den er en virkelig korrelasjonsmatrise. Ved å ekspandere (17), ser vi at relasjonen mellom $\boldsymbol{X}$ og $\boldsymbol{U}$ i dette tilfellet blir på form:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a & b & 0 & \cdots & 0 \\ a^2 & ab & b & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a^{n-1} & a^{n-2}b & a^{n-3}b & \cdots & b \end{bmatrix} \boldsymbol{U}, \quad (30)$$

der den triangulære matrisen i dette uttrykket svarer til Cholesky-dekomposisjonen av C.

Stjerne-modeller

Den neste modellen vi skal se på, vil vi kalle en stjernemodell. En slik modell egner seg godt for å beskrive en situasjon der alle variablene avhenger av en felles bakenforliggende variabel, si U_0.



Influensdiagram for en stjerne-modell

La denne gang $\boldsymbol{U} = (U_0, U_1, \ldots, U_n)$ være en vektor av $n+1$ uavhengige standard normalfordelte stokastiske variable, og definer:

$$\begin{aligned} X_1 &= a_1 U_0 + b_1 U_1 \\ X_2 &= a_2 U_0 + b_2 U_2 \\ &\ldots \\ X_n &= a_n U_0 + b_n U_n \end{aligned} \tag{31}$$

der konstantene $a_i, b_i, i = 1, \ldots, n$ velges slik at modellen får den ønskede korrelasjonsstruktur.

Variansen til X_i'ene blir dermed:

$$\begin{aligned} \mathrm{Var}(X_i) &= a_i^2\,\mathrm{Var}(U_0) + b_i^2\,\mathrm{Var}(U_i), \quad (32) \\ &= a_i^2 + b_i^2 \\ i &= i, \ldots, n. \end{aligned}$$

Som sist krever vi at alle X_i'ene skal ha varians lik 1. Dvs. vi må ha at:

$$a_i^2 + b_i^2 = 1,\ i = i, \ldots, n. \qquad (33)$$

Kovariansen mellom to variable X_i og X_j blir:

$$\begin{aligned}
&\text{Cov}(X_i, X_j) \\
&\qquad = \text{Cov}(a_i U_0 + b_i U_i, a_j U_0 + b_j U_j) \\
&\qquad = a_i a_j \,\text{Cov}(U_0, U_0) \\
&\qquad = a_i a_j \,\text{Var}(U_0) \\
&\qquad = a_i a_j \qquad (34)
\end{aligned}$$

for alle $i \neq j$.

Vi krever nå at kovariansene skal kunne skrives på følgende form:

$$\mathrm{Cov}(X_i, X_j) = \rho_i \rho_j,\ i \neq j. \tag{35}$$

Av dette, samt av (33), følger det dermed at:

$$a_i = \rho_i,\ i = 1, \ldots, n, \tag{36}$$

$$b_i = (1 - \rho_i^2)^{1/2},\ i = 1, \ldots, n. \tag{37}$$

Også denne gang er a_i'ene og b_i'ene veldefinerte for alle mulige korrelasjonskoeffisienter, dvs. for alle ρ_i som er slik at:

$$-1 \leq \rho_i \leq 1. \tag{38}$$

Totalt får vi at korrelasjonsmatrisen til $\boldsymbol{X}$ er gitt ved:

$$C = \begin{bmatrix} 1 & \rho_1\rho_2 & \rho_1\rho_3 & \cdots & \rho_1\rho_n \\ \rho_2\rho_1 & 1 & \rho_2\rho_3 & \cdots & \rho_2\rho_n \\ \rho_3\rho_1 & \rho_1\rho_3 & 1 & \cdots & \rho_3\rho_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_n\rho_1 & \rho_n\rho_2 & \rho_n\rho_3 & \cdots & 1 \end{bmatrix} \tag{39}$$

I likhet med Markov-modellen kan også denne strukturen gis en tolkning der korrelasjonen mellom to variable er avhengig av “avstanden” mellom variablene.

For å se hvordan dette kan gjøres, antar vi at $0 < \rho_i \leq 1$, $i = 1, \ldots, n$. La så ρ være et passende tall slik at $0 < \rho < 1$, og innfør så:

$$\alpha_i = \frac{\ln(\rho_i)}{\ln(\rho)},\ i = 1, \ldots, n. \tag{40}$$

Vi har med andre ord at:

$$\rho_i = \rho^{\alpha_i},\ i = 1, \ldots, n. \tag{41}$$

Størrelsen α_i kan nå fortolkes som “avstanden” fra X_i til sentrum av stjernen. Avstanden mellom to variable blir summen av deres respektive avstander inn til sentrum. Dvs.:

$$d(X_i, X_j) = \begin{cases} \alpha_i + \alpha_j & i \neq j \\ 0 & i = j \end{cases} \tag{42}$$

Med dette avstandsmålet får vi igjen at:

$$\mathrm{Cov}(X_i, X_j) = \rho^{d(X_i, X_j)}, \; i, j = 1, \ldots, n. \tag{43}$$

En spesielt enkel situasjon oppstår når $\alpha_i = 1/2$, for $i = 1, \ldots, n$. I dette tilfellet blir $d(X_i, X_j) = 1$ for alle $i \neq j$. Korrelasjons-matrisen til $\boldsymbol{X}$ blir dermed:

$$C = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix} \tag{44}$$

Bemerk at vi i denne situasjonen har at $a_i = \rho^{1/2}$ og at $b_i = (1 - \rho)^{1/2}$, $i = 1, \ldots, n$.

Matrisen C er igjen opplagt positivt definit siden vi faktisk har vist at den er en virkelig korrelasjonsmatrise. Av (31), ser vi at relasjonen mellom $\boldsymbol{X}$ og $\boldsymbol{U}$ i dette tilfellet blir på form:

$$\boldsymbol{X} = \begin{bmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ a_2 & 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_n & 0 & 0 & \cdots & b_n \end{bmatrix} \boldsymbol{U}. \tag{45}$$

Siden dimensjonen til $\boldsymbol{U}$ (dvs. $n+1$) i dette tilfellet er større enn dimensjonen til $\boldsymbol{X}$ (dvs. n), blir matrisen over ikke triangulær.

Dette kan vi imidlertid få til ved å føye til en ekstra variabel X_0 i $\boldsymbol{X}$. Vi kan f.eks. la denne svare til sentrum av stjernen ved simpelthen å la $X_0 = U_0$. Med en slik utvidelse av $\boldsymbol{X}$, får vi at:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_1 & b_1 & 0 & \cdots & 0 \\ a_2 & 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_n & 0 & 0 & \cdots & b_n \end{bmatrix} \boldsymbol{U}. \tag{46}$$

Denne matrisen er triangulær, og svarer til Cholesky-dekomposisjonen av korrelasjonsmatrisen for den utvidede versjonen av $\boldsymbol{X}$.

Korrelasjonsmatrisen for den utvidede versjonen av $\boldsymbol{X}$ blir:

$$C = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_n \\ \rho_1 & 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_n \\ \rho_2 & \rho_2\rho_1 & 1 & \cdots & \rho_2\rho_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_n & \rho_n\rho_1 & \rho_n\rho_2 & \cdots & 1 \end{bmatrix} \qquad (47)$$

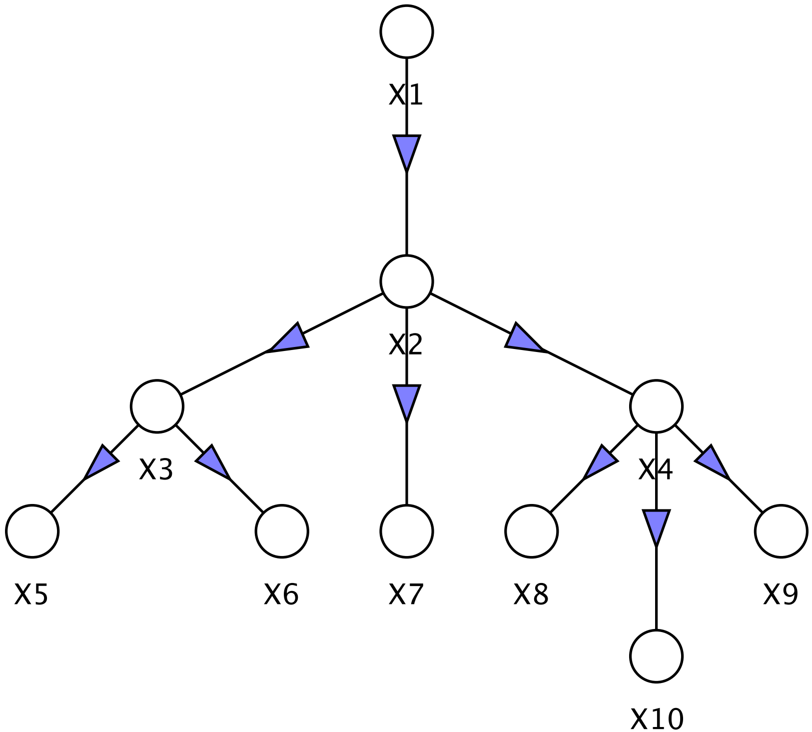
Bemerk at avstandstolkningen videreføres i den utvidede modellen ved at $d(X_0, X_i) = \alpha_i$, $i = 1, \ldots, n$. Dermed holder det også at $\text{Cov}(X_i, X_j) = \rho^{d(X_i, X_j)}$, $i, j = 0, 1, \ldots, n$.

Tre-modeller

Den neste modellen generaliserer både Markov-modellen og stjerne-modellen. Denne modellen kalles en tre-modell, fordi avhengighetsstrukturen grafisk kan illustreres som et (rettet) tre. En tre-modell har følgende viktige egenskaper:

- Modellen inneholder en "rot"-variabel som ikke har noen forgjengere

- Enhver annen variabel har nøyaktig én direkte forgjenger



Influensdiagram for en tre-modell

Vi starter med å definere begrepet *forgjenger* i en tre-modell. En variabel X_i sies å være en forgjenger til en annen variabel X_j hvis tre-modellen inneholder en *rettet sti* fra X_i til X_j. Spesielt sies X_i å være en *direkte forgjenger* til X_j dersom det eksisterer en rettet sti av lengde 1 fra X_i til X_j.

Indeksmengden av forgjengere til variablen X_i betegnes med $P(i)$, mens indeksmmengden av direkte forgjengere betegnes med $p(i)$. En variabel X_i sies å være en *rot* dersom $P(i) = p(i) = \emptyset$.

I en tre-modell skal det altså eksistere nøyaktig en variabel X_i som er slik at $P(i) = p(i) = \emptyset$. Videre skal $|p(j)|$ være lik 1 for alle de øvrige variablene. Dersom $p(j) = \{i\}$, vil vi for enkelhets skyld droppe mengdesymbolene, og i stedet skrive $p(j) = i$.

Vi antar at vi har indeksert variablene i tre-modellen, $\boldsymbol{X} = (X_1, \ldots, X_n)$, på en slik måte at modellens eneste rotvariabel er X_1, og slik at dersom $p(i) = \{j\}$, så er $j < i$. Av dette følger det også at mengden $P(i)$ kun består av indekser som er lavere enn i.

Som tidligere lar vi $\boldsymbol{U} = (U_1, \ldots, U_n)$ være en vektor av n uavhengige standard normalfordelte stokastiske variable, og definer:

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= a_2 X_{p(2)} + b_2 U_2 \\ &\cdots \\ X_n &= a_n X_{p(n)} + b_n U_n \end{aligned} \tag{48}$$

der konstantene $a_i, b_i, i = 2, \ldots, n$ velges slik at modellen får den ønskede korrelasjonsstruktur.

Variansen til X_i'ene blir dermed:

$$\text{Var}(X_i) = a_i^2 \text{Var}(X_{p(i)}) + b_i^2 \text{Var}(U_i), \quad (49)$$

$$i = i, \ldots, n.$$

Som sist krever vi at alle X_i'ene skal ha varians lik 1. Dette betyr at:

$$a_i^2 + b_i^2 = 1, \; i = i, \ldots, n. \qquad (50)$$

Kovariansen mellom X_i og $X_{p(i)}$ blir:

$$
\begin{aligned}
&\mathrm{Cov}(X_i, X_{p(i)}) \\
&\qquad = \mathrm{Cov}(a_i X_{p(i)} + b_i U_i, X_{p(i)}) \\
&\qquad = a_i \,\mathrm{Cov}(X_{p(i)}, X_{p(i)}) \\
&\qquad = a_i \,\mathrm{Var}(X_{p(i)}) \\
&\qquad = a_i \qquad (51)
\end{aligned}
$$

for alle $i > 1$.

Vi krever nå at kovariansen mellom en variabel og dens direkte forgjenger skal være:

$$\text{Cov}(X_i, X_{p(i)}) = \rho_i,\ i = 1, \ldots, n. \qquad (52)$$

Av dette, samt av (50), følger det dermed at:

$$a_i = \rho_i,\ i = 1, \ldots, n, \qquad (53)$$

$$b_i = (1 - \rho_i^2)^{1/2},\ i = 1, \ldots, n. \qquad (54)$$

Vi bemerker at også denne gang er a_i'ene og b_i'ene veldefinerte for alle mulige korrelasjonskoeffisienter.

Vi skal nå se på hvordan vi i tre-modellen kan beregne Cov(X_j, X_i) for et vilkårlig variabel-par. For å få til dette, ser vi på to tilfeller:

- Tilfelle 1: $j \in P(i)$

- Tilfelle 2: $j \notin P(i)$

Tilfelle 1. $j \in P(i)$. La indeksmengden til den rettede stien fra X_j til X_i være $\{i_0, \ldots, i_s\}$, der $i_0 = j$ og $i_s = i$. Vi har da ved (48) at:

$$
\begin{aligned}
\mathrm{Cov}(X_j, X_i) &= \mathrm{Cov}(X_j, a_i X_{p(i)} + b_i U_i) \\
&= \mathrm{Cov}(X_j, a_i X_{i_{s-1}} + b_i U_i) \\
&= a_i \,\mathrm{Cov}(X_j, X_{i_{s-1}}) \qquad (55)
\end{aligned}
$$

Gjentagelse av dette argumentet samt antagelsen om at alle X_i'ene har varians lik 1, gir til slutt at:

$$\mathrm{Cov}(X_j, X_i) = a_{i_1} a_{i_2} \cdots a_{i_s}. \qquad (56)$$

Tilfelle 2. $j \notin P(i)$. Vi velger i dette tilfellet k som den høyeste index i $P(j) \cap P(i)$. Pr. definisjon vil det da finnes en rettet sti fra X_k til X_j og en rettet sti fra X_k til X_i. Vi antar at indeksmengdene for disse to stiene er henholdsvis $\{j_0, \ldots, j_s\}$ og $\{i_0, \ldots, i_t\}$, der $j_0 = i_0 = k$, og der $j_s = j$ og $i_t = i$. Ved et tilsvarende argument som over kan det da vises at:

$$\text{Cov}(X_j, X_i) = a_{j_1} a_{j_2} \cdots a_{j_s} a_{i_1} a_{i_2} \cdots a_{i_t}. \quad (57)$$

På samme måte som tidligere kan også denne modellen gis en avstandstolkning. For å se dette, antar vi igjen at $0 < \rho_i \leq 1$, $i = 1, \ldots, n$, lar så ρ være et passende tall slik at $0 < \rho < 1$, og innfører så:

$$\alpha_i = \frac{\ln(\rho_i)}{\ln(\rho)},\ i = 1, \ldots, n. \tag{58}$$

eller ekvivalent:

$$\rho_i = \rho^{\alpha_i},\ i = 1, \ldots, n. \tag{59}$$

Størrelsen α_i fortolkes som avstanden fra X_i til $X_{p(i)}$. Avstanden mellom to variable blir summen av deres respektive avstander inn den av variablenes felles forgjenger som har høyest index.

Med dette avstandsmålet får vi igjen at:

$$\mathrm{Cov}(X_i, X_j) = \rho^{d(X_i,X_j)},\ i,j = 1,\ldots,n. \tag{60}$$