Föreläsning 03

Statistisk inlärning och dataanalys (Kungliga Tekniska Högskolan)

Föreläsning 3

3.1 Betingat oberoende

I föregående föreläsning introducerade vi fleradimensionellafördelningar för stokastiska vektorer $\boldsymbol{X} = (X_1, \ldots, X_m)$ där $X_1, \ldots, X_m$ inte nödvändigtvis är oberoende. Istället kan de vara beroende på komplexa sätt (exempelvis, glödlamporna i Exempel 2.9). En av de främsta anledningarna för att använda multivariatfördelningar är för att modellera sådana beroende. I statistiken föredrar vi ofta att identifiera när det definitivt inte finns något beroende. Därför bildar vi vanligtvis modeller baserade på *de observerbara oberoenderelationarna* inom systemet; dvs. oberoenderelationarna som vi kan anta med hjälp av apriori kunskap om variablerna eller som vi lära oss från data (med hjälp av tekniker vi ska se senare i kursen). Dessa oberoenderelationer kan ha väldigt mycket komplexare former än bara $X_i \perp\!\!\!\perp X_j$. Till exempel kan man betrakta oberoende som inte existerar i den simultan fördelningen men bara existerar i en betingad fördelning.

I det följande betecknar vi en mängd $\{1, \ldots, m\}$ för ett positivt heltal m med $[m]$. Om vi har en fördelning $f_{\boldsymbol{X}}(x_1, \ldots, x_m)$ och $A, B \subseteq [m]$ skriver vi $f_{\boldsymbol{X}_{A\cup B}}(\boldsymbol{x}_A, \boldsymbol{x}_B)$ för marginalfördelningen av $\boldsymbol{X}_{A\cup B}$. Exempelvis, om $m = 8$ och $A = \{3, 5\}$ och $B = \{4, 6, 7\}$ är

$$\begin{aligned} f_{\boldsymbol{X}_{A\cup B}}(\boldsymbol{x}_A, \boldsymbol{x}_B) &= f_{\boldsymbol{X}_A, \boldsymbol{X}_B}(\boldsymbol{x}_A, \boldsymbol{x}_B), \\ &= f_{X_3, X_5, X_4, X_6, X_7}(x_3, x_5, x_4, x_6, x_7), \\ &= f_{X_3, X_4, X_5, X_6, X_7}(x_3, x_4, x_5, x_6, x_7) \end{aligned}$$

Det vill säga att man kan permutera variabler i funktionen som man vill och det är fortfarande samma funktion.

Låt $A, B, C \subset [m]$ vara disjunkta delmängder av $[m]$ där $A, B \neq \emptyset$. Vi säger $\boldsymbol{X}_A$ *är oberoende av* $\boldsymbol{X}_B$ *given* $\boldsymbol{X}_C$ (kortfattat, A *är oberoende av* B *givet* C) om

$$f_{\boldsymbol{X}_{A\cup B}|\boldsymbol{X}_C}(\boldsymbol{x}_A, \boldsymbol{x}_B|\boldsymbol{x}_C) = f_{\boldsymbol{X}_A|\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_C) f_{\boldsymbol{X}_B|\boldsymbol{X}_C}(\boldsymbol{x}_B|\boldsymbol{x}_C)$$

för alla $\boldsymbol{x}_A, \boldsymbol{x}_B$ och alla $\boldsymbol{x}_C$ där $f_{\boldsymbol{X}_C}(\boldsymbol{x}_C) > 0$. I sådant fall skriver vi $X_A \perp\!\!\!\perp X_B|X_C$ (eller $A \perp\!\!\!\perp B|C$). Om A är inte oberoende av B given C säger vi att A *är beroende av* B *given* C och skriver $A \not\perp\!\!\!\perp B|C$.

Exempel 3.1. Låt $(X_1, X_2, X_3) \sim \mathrm{N}(0, \Sigma)$ där

$$\Sigma = \begin{bmatrix} -1 & 16 & -8 \\ 16 & -4 & 2 \\ -8 & 2 & 62 \end{bmatrix}.$$

Det följer från Sats 2.4 att $X_1 \not\perp\!\!\!\perp X_3$ eftersom

$$\mathrm{Cov}[X_1, X_3] = \Sigma_{1,3} = -8 \neq 0.$$

Med hjälp av Sats 2.3 har vi dock att $\boldsymbol{X}_{\{1,3\}}|X_2 = x_2 \sim \mathrm{N}(\mu', \Sigma')$ där

$$\mu' = 0 + \Sigma_{\{1,3\},2}\Sigma_{2,2}^{-1}[x_2] = \begin{bmatrix} 16 \\ 2 \end{bmatrix} \left[-\tfrac{1}{4}\right] \left[x_2\right] = \begin{bmatrix} -4x_2 \\ -\frac{x_2}{2} \end{bmatrix},$$

och

$$\begin{aligned}\Sigma' &= \Sigma_{\{1,3\},\{1,3\}} - \Sigma_{\{1,3\},2}\Sigma_{2,2}^{-1}\Sigma_{2,\{1,3\}},\\ &= \begin{bmatrix}-1 & -8\\ -8 & 62\end{bmatrix} - \begin{bmatrix}16\\ 2\end{bmatrix}\begin{bmatrix}-\frac{1}{4}\end{bmatrix}\begin{bmatrix}16 & 2\end{bmatrix},\\ &= \begin{bmatrix}63 & 0\\ 0 & 63\end{bmatrix}.\end{aligned}$$

Det följer att

$$\begin{aligned}f_{\boldsymbol{X}_{\{1,3\}}|X_2}(x_1,x_3|x_2) &= \frac{1}{(2\pi)(63)}\exp\left(-\frac{1}{2}\left(\begin{bmatrix}-4x_2 & -\frac{x_2}{2}\end{bmatrix} - \begin{bmatrix}x_1 & x_3\end{bmatrix}\right)\begin{bmatrix}\frac{1}{63} & 0\\ 0 & \frac{1}{63}\end{bmatrix}\left(\begin{bmatrix}-4x_2\\ -\frac{x_2}{2}\end{bmatrix} - \begin{bmatrix}x_1\\ x_3\end{bmatrix}\right)\right),\\ &= \frac{1}{(2\pi)(63)}\exp\left(-\frac{1}{2(63)}\left((4x_2+x_1)^2 + ((1/2)x_2+x_3)^2\right)\right),\\ &= \frac{1}{\sqrt{2\pi(63)}}\exp\left(-\frac{1}{2}(4x_2+x_1)^2\right)\frac{1}{\sqrt{2\pi(63)}}\exp\left(-\frac{1}{2}((1/2)x_2+x_1)^2\right),\\ &= f_{X_1|X_2}(x_1|x_2)f_{X_3|X_2}(x_3|x_2),\end{aligned}$$

eftersom

$$\begin{aligned}&X_1|X_2 = x_2 \sim \mathrm{N}(0+\Sigma_{1,2}\Sigma_{2,2}^{-1}x_2, \Sigma_{1,1}-\Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}) = \mathrm{N}(-4x_2, 63), \text{ och}\\ &X_3|X_2 = x_2 \sim \mathrm{N}(-\frac{x_2}{2}, 63).\end{aligned}$$

Eftersom detta gäller för godtyckliga värden x_2 får vi att $X_1 \perp\!\!\!\perp X_3|X_2$.

Exempel 3.1 visar att betingat oberoende är verkligen oberoende som händer in betingade fördelningar. Detta innebär att vi kan generalisera Sats 2.4 för oberoende i multivariatnormalfördelningar till en karaktärisering av betingat oberoende i sådana fördelningar; nämligen, om $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ gäller $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{X}_C$ om och endast om $\mathrm{Cov}[\boldsymbol{X}_A, \boldsymbol{X}_B|\boldsymbol{X}_C] = \boldsymbol{0}$. För att kontrollera $\mathrm{Cov}[\boldsymbol{X}_A, \boldsymbol{X}_B|\boldsymbol{X}_C] = \boldsymbol{0}$ skulle vi naturligtvis beräkna först covariansmatrisen av $\boldsymbol{X}_{A\cup B}|\boldsymbol{X}_C = \boldsymbol{x}_C$:

$$\Sigma_{A\cup B} - \Sigma_{A\cup B,C}\Sigma_{C,C}^{-1}\Sigma_{C,A\cup B}.$$

Enligt den nästa satsen kan vi bara beräkna rangen av en viss submatris av Σ istället.

Sats 3.1. *Låt $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ och $A, B, C \subseteq [m]$ disjunkta. Matrisen $\Sigma_{A\cup C, B\cup C}$ har rang $|C|$ om och endast om $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{X}_C$.*

Bevis. Med hjälp av Sats 2.3 har vi att den betingade fördelningen för $\boldsymbol{X}_{A\cup B}$ givet $\boldsymbol{X}_C = \boldsymbol{x}_C$ är

$$\mathrm{N}\left(\boldsymbol{\mu}_{A\cup B} + \Sigma_{A\cup B,C}\Sigma_{C,C}^{-1}(\boldsymbol{x}_C - \boldsymbol{\mu}_C), \Sigma_{A\cup B,A\cup B} - \Sigma_{A\cup B,C}\Sigma_{C,C}^{-1}\Sigma_{C,A\cup B}\right).$$

Enligt definitionen av betingad oberoende skulle vi vilja se att $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B$ i denna fördelning för alla $\boldsymbol{X}_C = \boldsymbol{X}_C$. Vi ska se i Sats 3.2 att $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B$ om och endast om $f_{\boldsymbol{X}_A|\boldsymbol{X}_B}(\boldsymbol{x}_A|\boldsymbol{x}_B) = f_{\boldsymbol{X}_A}(\boldsymbol{x}_A)$. Vi har dock att denna ekvivalens av fördelningar stämmer om och endast om dessa två normalfördelningar har den samma covariansmatrisen; det vill säga,

$$\Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A} = \Sigma_{A,A}.$$

Eftersom $\Sigma_{B,B}^{-1}$ är positivt definit är det full rang. Så denna ekvivalens av matriser stämmer om och endast om $\Sigma_{A,B} = 0$. (Vi har just nu bevisade formellt att $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{X}_C$ om och endast om $\mathrm{Cov}[\boldsymbol{X}_A, \boldsymbol{X}_B|\boldsymbol{X}_C] = \boldsymbol{0}$ som vi notiserade i den ovanstående diskussionen.)

Det följer att $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{X}_C$ om och endast om

$$\boldsymbol{0} = \left(\Sigma_{A\cup B,A\cup B} - \Sigma_{A\cup B,C}\Sigma_{C,C}^{-1}\Sigma_{C,A\cup B}\right)_{A,B} = \Sigma_{A,B} - \Sigma_{A,C}\Sigma_{C,C}^{-1}\Sigma_{C,B}.$$

Formeln till höger är Schur komplementet $\Sigma_{A\cup C,B\cup C}/\Sigma_{C,C}$ av matrisen

$$\Sigma_{A\cup C,B\cup C} = \begin{bmatrix} \Sigma_{A,B} & \Sigma_{A,C} \\ \Sigma_{C,B} & \Sigma_{C,C} \end{bmatrix}.$$

Har använder vi Schurkomplementet $M/D = A - BD^{-1}C$ för matrisen

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

istället för Schurkomplementet $M/A = D - CA^{-1}B$ som används i föreläsning 2. För Schurkomplementet M/D har vi den *Guttman rang additivitet formeln* som säger att

$$\operatorname{rang}(M) = \operatorname{rang}(D) + \operatorname{rang}(A - BD^{-1}C).$$

I vårt fall blir formeln

$$\operatorname{rang}(\Sigma_{A\cup C,B\cup C}) = \operatorname{rang}(\Sigma_{C,C}) + \operatorname{rang}(\Sigma_{A,B} - \Sigma_{A,C}\Sigma_{C,C}^{-1}\Sigma_{C,B}).$$

Eftersom $\Sigma_{C,C}$ är positivt definit har $\Sigma_{C,C}$ rang $|C|$. Därför följer det att Schurkomplementet $\Sigma_{A\cup B,A\cup B}/\Sigma_{C,C} = \mathbf{0}$ om och endast om $\operatorname{rang}(\Sigma_{A\cup C,B\cup C}) = |C|$. □

Exempel 3.2. Vi återgår till vårt glödlampsexempel, där $\boldsymbol{X} = (X_1, X_2, X_3)$ har fördelningen

$$f_{(X_1,X_2,X_3)}(x_1,x_2,x_3) = \frac{1}{4}e^{-(x_2+x_3)/2} \qquad \text{för } 0 < x_1 < x_2 < +\infty \text{ och } 0 < x_1 < x_3 < +\infty,$$

såg vi i Exempel 2.9 att

$$f_{(X_2,X_3)|X_1}(x_2,x_3 \mid x_1) = \frac{1}{4}e^{-(x_2+x_3-2x_1)/2}$$

och

$$f_{X_2|X_1}(x_2 \mid x_1) = \frac{1}{2}e^{-(x_2-x_1)/2}.$$

På samma sätt har vi att

$$f_{X_3|X_1}(x_3 \mid x_1) = \frac{1}{2}e^{-(x_3-x_1)/2}.$$

Det följer att

$$f_{\boldsymbol{X}_{\{2,3\}}|X_1}(x_2,x_3|x_1) = f_{X_2|X_1}(x_2|x_1)f_{X_3|X_1}(x_3|x_1);$$

det vill säga att $X_2 \perp\!\!\!\perp X_3|X_1$. Det är också en bra övning att kontrollera att $X_2 \not\perp\!\!\!\perp X_3$.

Som vi såg i beviset för Sats 3.1 kan det vara hjälpsamt att ha olika sätt att verifiera betingat oberoende:

Sats 3.2. *Låt $\boldsymbol{X} = (X_1,\dots,X_m)$ har täthetsfunktion (eller sannolikhetsfunktion) $f_{\boldsymbol{X}}(x_1,\dots,x_m) > 0$ för alla simultana utfall $(x_1,\dots,x_m)$. Låt $A,B,C \subseteq [m]$ vara disjunkta. Följande är likvärdiga:*

(a) $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B|\boldsymbol{X}_C$,

(b) $f_{\boldsymbol{X}_A|\boldsymbol{X}_{B\cup C}}(\boldsymbol{x}_A|\boldsymbol{x}_B,\boldsymbol{x}_C) = f_{\boldsymbol{X}_A|\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_C)$ *för alla* $\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C$, *och*

(c) $f_{\boldsymbol{X}_A|\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_B,\boldsymbol{x}_C) = f_{\boldsymbol{X}_A|\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}'_B,\boldsymbol{x}_C)$ *för alla* $\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C$ *och alla* $\boldsymbol{x}'_B \neq \boldsymbol{x}_B$.

Bevis. Vi ska bevisa att (a) och (b) är ekvivalent i övning 3, uppgift 8(b). Så visar vi bara att (b) och (c) är ekvivalena. Anta att (b) stämmer. Då har vi att

$$f_{\boldsymbol{X}_A|\boldsymbol{X}_{B\cup C}}(\boldsymbol{x}_A|\boldsymbol{x}_B,\boldsymbol{x}_C) = f_{\boldsymbol{X}_A|\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_C)$$

för alla $\boldsymbol{x}_A$ och $\boldsymbol{x}_C$ och en given $\boldsymbol{x}_B$. På samma sätt har vi också

$$f_{\boldsymbol{X}_A|\boldsymbol{X}_{B\cup C}}(\boldsymbol{x}_A|\boldsymbol{x}'_B,\boldsymbol{x}_C) = f_{\boldsymbol{X}_A|\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_C)$$

för alla $\boldsymbol{x}_A$ och $\boldsymbol{x}_C$ och en given $\boldsymbol{x}'_B \neq \boldsymbol{x}_B$. Det följer att (c) stämmer.

Nu anta att (c) stämmer. Per definition av betingade fördelningar har vi

$$\frac{f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)}{f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C)} = \frac{f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}'_B,\boldsymbol{x}_C)}{f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}'_B,\boldsymbol{x}_C)},$$

eller

$$f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}'_B,\boldsymbol{x}_C) = f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}'_B,\boldsymbol{x}_C)f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C).$$

Om vi integrerar med avseende på variablen $\boldsymbol{x}'_B$ får vi

$$\begin{aligned}
\int f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}'_B,\boldsymbol{x}_C)d\boldsymbol{x}'_B &= \int f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}'_B,\boldsymbol{x}_C)f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C)d\boldsymbol{x}'_B,\\
f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)\left(\int f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}'_B,\boldsymbol{x}_C)d\boldsymbol{x}'_B\right) &= \left(\int f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}'_B,\boldsymbol{x}_C)d\boldsymbol{x}'_B\right) f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C),\\
f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)f_{\boldsymbol{X}_C}(\boldsymbol{x}_C) &= f_{\boldsymbol{X}_A,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_C)f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C),\\
\frac{f_{\boldsymbol{X}_A,\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_B,\boldsymbol{x}_C)}{f_{\boldsymbol{X}_B,\boldsymbol{X}_C}(\boldsymbol{x}_B,\boldsymbol{x}_C)} &= \frac{f_{\boldsymbol{X}_A,\boldsymbol{X}_C}(\boldsymbol{x}_A,\boldsymbol{x}_C)}{f_{\boldsymbol{X}_C}(\boldsymbol{x}_C)},\\
f_{\boldsymbol{X}_A|\boldsymbol{X}_{B\cup C}}(\boldsymbol{x}_A|\boldsymbol{x}_B,\boldsymbol{x}_C) &= f_{\boldsymbol{X}_A|\boldsymbol{X}_C}(\boldsymbol{x}_A|\boldsymbol{x}_C).
\end{aligned}$$

□

Villkor (2) i Sats 3.2 passar bra ihop med kedjereglen från den föregående föreläsningen. I sin mer allmänna form blir kedjereglen följande:

Sats 3.3 (Kedjereglen)**.** *Låt $\boldsymbol{X} = (X_1,\ldots,X_m)$ ha täthetsfunktion (eller sannolikhetsfunktion) $f_{\boldsymbol{X}}(x_1,\ldots,x_m)$. Då har vi att*

$$f_{\boldsymbol{X}}(x_1,\ldots,x_m) = \prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{[i-1]}}(x_i|\boldsymbol{x}_{[i-1]}).$$

Bevis. Per definition av en betingad fördelning har vi att

$$\prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{[i-1]}}(x_i|\boldsymbol{x}_{[i-1]}) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}_{[m-1]}}(\boldsymbol{x}_{[m-1]})}\cdot\frac{f_{\boldsymbol{X}_{[m-1]}}(\boldsymbol{x}_{[m-1]})}{f_{\boldsymbol{X}_{[m-2]}}(\boldsymbol{x}_{[m-2]})}\cdots\frac{f_{\boldsymbol{X}_{[2]}}(\boldsymbol{x}_{[2]})}{f_{X_1}(x_1)}f_{X_1}(x_1) = f_{\boldsymbol{X}}(\boldsymbol{x}).$$

□

Exempel 3.3. För glödlamporna i Exempel 2.9 har vi

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_2,X_3,X_1}(x_2,x_3,x_1) = f_{X_2|\boldsymbol{X}_{\{1,3\}}}(x_2|x_1,x_3)f_{X_3|X_1}(x_3|x_1)f_{X_1}(x_1).$$

Vi såg i Exempel 3.2 att $X_3 \perp\!\!\!\perp X_2|X_1$. Med hjälp av Sats 3.2 (2) kan vi förenkla faktoriseringen av fördelningen som

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_2|X_1}(x_2|x_1)f_{X_3|X_1}(x_3|x_1)f_{X_1}(x_1).$$

3.2 Hierarkiska modeller

Hittills har vi bara betraktat modeller $(X_1,\ldots,X_m)$ med parameterar $\theta_1,\ldots,\theta_d$ som är fixa. Det kan också vara användbart att betrakta modeller för vilka parameterarna kan variera enligt någon fördelning. För sådana modeller kan kedjereglen hjälpa oss att beräkna interessanta värden – exempelvis, väntevärdet.

Exempel 3.4. Anta att en insekt lägger ett stort antal ägg och varje ägg överlever med sannolikheten p. (Vi antar att äggens överlevnad är oberoende.) Vi skulle vilja beräkna väntevärdet av antalet ägg som överlever. Detta väntevärde beror naturligtvis på antal ägg produceras.

Låt N vara antalet ägg som produceras och X antalet ägg som överlever. Vi kan modellera N med en Poissonfördelning med parameter λ; dvs. $N \sim \mathrm{Po}(\lambda)$. Den betingade fördelningen $X|N = n$ modelleras som $\mathrm{Bin}(n,p)$. Så den *hierarkiska modellen* är

$$\begin{aligned}
X|N = n &\sim \mathrm{Bin}(n,p),\\
N &\sim \mathrm{Po}(\lambda).
\end{aligned}$$

Notera att $X|N=n$ har sannolikhetsfunktion $f_{X|N}(x|n)$.

För att beräkna väntevärdet $\mathrm{E}[X]$ kan vi börja med den simultana fördelningen $f_{X,N}(x,n)$ och hitta marginalfördelningen $f_X(x)$. Kedjereglen säger att

$$f_{X,N}(x,n) = f_{X|N}(x|n)f_N(n) = \binom{n}{x}p^x(1-p)^{n-x}\frac{\lambda^n e^{-\lambda}}{n!}.$$

Därför kan vi beräkna

$$\begin{aligned}
f_X(x) &= \sum_{n=0}^{\infty}\binom{n}{x}p^x(1-p)^{n-x}\frac{\lambda^n e^{-\lambda}}{n!}, \\
&= \sum_{n=x}^{\infty}\binom{n}{x}p^x(1-p)^{n-x}\frac{\lambda^n e^{-\lambda}}{n!}, && (\tbinom{n}{x}=0 \text{ om } n<x) \\
&= \frac{\lambda^x}{\lambda^x}\sum_{n=x}^{\infty}\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\frac{\lambda^n e^{-\lambda}}{n!}, \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!}\sum_{n=x}^{\infty}\frac{(1-p)^{n-x}\lambda^{n-x}}{(n-x)!}, \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!}\sum_{k=0}^{\infty}\frac{(\lambda(1-p))^{n-x}}{k!}, && (\text{låt } k=n-x) \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!}e^{(1-p)\lambda}, \\
&= \frac{(\lambda p)^x e^{-\lambda p}}{x!}.
\end{aligned}$$

Därför är $X \sim \mathrm{Po}(\lambda p)$ och vi har att $\mathrm{E}[X] = \lambda p$.

Låt X och Y vara stokastiska variabler. Den betingade fördelningen $f_{X|Y}(x|y)$ har väntevärdet

$$\mathrm{E}[X|Y=y] = \int x f_{X|Y}(x|y)dx.$$

Eftersom vi kan betrakta $\mathrm{E}[X|Y=y]$ som en funktion av y, dvs. $g(y) = \mathrm{E}[X|Y=y]$, kan vi definiera den stokastiska variablen $\mathrm{E}[X|Y] = g(Y)$. I Exempel 3.4 ville vi att beräkna $\mathrm{E}[X]$. Följande låter oss att göra det utan att beräkna marginalfördelningen.

Sats 3.4 (Lagen om Total Förväntan)**.** *If X och Y är stokastiska variabler har vi att*

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]].$$

Bevis. Om $f_{(X,Y)}(x,y)$ är den simultana fördelningen av X och Y har vi att

$$\begin{aligned}
\mathrm{E}[X] &= \int\int x f_{(X,Y)}(x,y)dxdy, \\
&= \int\int x f_{X|Y}(x|y)f_Y(y)dxdy, \\
&= \int\left(\int x f_{X|Y}(x|y)dx\right)f_Y(y)dy, \\
&= \int \mathrm{E}[X|Y=y]f_Y(y)dy, \\
&= \mathrm{E}[g(Y)],
\end{aligned}$$

där $g(y) = \mathrm{E}[X|Y=y]$. □

Exempel 3.5. För den hierarkiska modellen i Exempel 3.4 har vi

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|N]], \\
&= \mathrm{E}[Np], && (X|N=n \sim \mathrm{Bin}(n,p)) \\
&= p\,\mathrm{E}[N], \\
&= \lambda p. && (N \sim \mathrm{Po}(\lambda))
\end{aligned}$$

Vi kan bestämma formler för varianser på samma sätt. Variansen av $X|Y = y$ är

$$\mathrm{Var}[X|Y = y] = \mathrm{E}[X^2|Y = y] - \mathrm{E}[X|Y = y]^2.$$

På samma sätt som $\mathrm{E}[X|Y]$ kan vi definiera funktionen $g(y) = \mathrm{Var}[X|Y = y]$. Då har vi den stokastiska variablen $\mathrm{Var}[X|Y] = g(Y)$.

Sats 3.5 (Lagen om Total Varians)**.** *Låt X och Y vara stokastiska variabler. Då har vi att*

$$\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X|Y]] + \mathrm{Var}[\mathrm{E}[X|Y]].$$

Bevis. Per definitionen av $\mathrm{Var}[X]$ har vi

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[(X - \mathrm{E}[X|Y] + \mathrm{E}[X|Y] - \mathrm{E}[X])^2],$$

där vi subtraherade och adderade $\mathrm{E}[X|Y]$. Om vi expanderar kvadraten får vi

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X|Y])^2] + \mathrm{E}[(\mathrm{E}[X|Y] - \mathrm{E}[X])^2] + 2\,\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])].$$

Vi ska visa att den sista termen är lika med 0. Med hjälp av Sats 3.4 har vi att

$$\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])] = \mathrm{E}[\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])|Y]].$$

I den betingade fördelningen $X|Y = y$ är X stokastisk. Det följer att $\mathrm{E}[X|Y]$ och $\mathrm{E}[X]$ är konstanter i

$$\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])|Y],$$

och $\mathrm{E}[X|Y] - \mathrm{E}[X]$ också är en konstant. Därför har vi

$$\begin{aligned}
\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])|Y] &= (\mathrm{E}[X|Y] - \mathrm{E}[X])\,\mathrm{E}[(X - \mathrm{E}[X|Y])|Y], \\
&= (\mathrm{E}[X|Y] - \mathrm{E}[X])(\mathrm{E}[X|Y] - \mathrm{E}[X|Y]), \\
&= (\mathrm{E}[X|Y] - \mathrm{E}[X])(0), \\
&= 0.
\end{aligned}$$

Så följer det att

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[(X - \mathrm{E}[X|Y])^2] + \mathrm{E}[(\mathrm{E}[X|Y] - \mathrm{E}[X])^2] + 2\,\mathrm{E}[(X - \mathrm{E}[X|Y])(\mathrm{E}[X|Y] - \mathrm{E}[X])], \\
&= \mathrm{E}[(X - \mathrm{E}[X|Y])^2] + \mathrm{E}[(\mathrm{E}[X|Y] - \mathrm{E}[X])^2].
\end{aligned}$$

På samma sätt (med hjälp av Sats 3.4) ser vi också att

$$\begin{aligned}
\mathrm{E}[(X - \mathrm{E}[X|Y])^2] &= \mathrm{E}[\mathrm{E}[(X - \mathrm{E}[X|Y])^2|Y]], \\
&= \mathrm{E}[\mathrm{Var}[X|Y]],
\end{aligned}$$

och

$$\mathrm{E}[(\mathrm{E}[X|Y] - \mathrm{E}[X])^2] = \mathrm{Var}[\mathrm{E}[X|Y]].$$

□

Exempel 3.6. För den hierarkiska modellen

$$\begin{aligned}
X|N = n &\sim \mathrm{Bin}(n, p), \\
N &\sim \mathrm{Po}(\lambda).
\end{aligned}$$

från Exempel 3.4 ger Satsen 3.5

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[\mathrm{Var}[X|N]] + \mathrm{Var}[\mathrm{E}[X|N]], \\
&= \mathrm{E}[Np(1-p)] + \mathrm{Var}[Np], \\
&= p(1-p)\,\mathrm{E}[N] + p^2\,\mathrm{Var}[N], \\
&= \lambda p(1-p) + \lambda p^2, \\
&= \lambda p.
\end{aligned}$$

Resultatet stämmer eftersom vi har redan sett att $X \sim \mathrm{Po}(\lambda p)$.

I generellt sett kan vi göra detsamma. Anta att X har fördelningen som beror på parametrarna $\theta_1, \ldots, \theta_d$ men $\theta_1, \ldots, \theta_d$ vilken i sin tur varierar enligt fördelningar $f_{\Theta_i}(\theta_i)$. Eftersom parameterarna är nu stokastiska behöver vi vara försiktiga med deras beroenderelationer. Exempelvis, om $\Theta_1, \ldots, \Theta_d$ är oberoende får vi den resulterande *hierarkiska modellen*

$$\begin{aligned}
&X|\Theta_1 = \theta_1, \ldots, \Theta_d = \theta_d \sim f_{X|\Theta_1=\theta_1,\ldots,\Theta_d=\theta_d}(x|\theta_1, \ldots, \theta_d),\\
&\Theta_1 \sim f_{\Theta_1}(\theta_1),\\
&\vdots\\
&\Theta_d \sim f_{\Theta_d}(\theta_d).
\end{aligned}$$

Marginalfördelningen $f_X(x)$ kan beräknas som

$$\begin{aligned}
f_X(x) &= \int_\Omega f_{(X,\Theta_1,\ldots,\Theta_d)}(x, \theta_1, \ldots, \theta_d) d\theta_1 \cdots d\theta_d,\\
&= \int_\Omega f_{X|\Theta_1,\ldots,\Theta_d}(x|\theta_1, \ldots, \theta_d) \prod_{i=1}^{d} f_{\Theta_i|\boldsymbol{\theta}_{[i-1]}}(\theta_i|\theta_1, \ldots, \theta_{i-1}) d\theta_1 \cdots d\theta_d,\\
&= \int_\Omega f_{X|\Theta_1,\ldots,\Theta_d}(x|\theta_1, \ldots, \theta_d) f_{\Theta_1}(\theta_1) \cdots f_{\Theta_d}(\theta_d) d\theta_1 \cdots d\theta_d. \qquad \text{(kedjereglen)}
\end{aligned}$$

Om $\Theta_1 \not\perp\!\!\!\perp \Theta_2$, exempelvis, har vi istället

$$f_X(x) = \int_\Omega f_{X|\Theta_1,\ldots,\Theta_d}(x|\theta_1, \ldots, \theta_d) f_{\Theta_1}(\theta_1) f_{\Theta_2|\Theta_1}(\theta_2|\theta_1) f_{\Theta_3}(\theta_3) \cdots f_{\Theta_d}(\theta_d) d\theta_1 \cdots d\theta_d.$$

3.3 Grafiska modeller

Låt oss nu tänka baklänges, om vi har en fördelning $(X_1, \ldots, X_m)$ som vi vet inte så mycket om kan vi använda betingat oberoende (tillsammans med kedjereglen) för att konstruera en hierarki där några variabler fyller rollen som parametrarna till andra variabler i systemet.

Anta att $(X_1, \ldots, X_m)$ har täthetsfunktion (eller sannolikhetsfunktion) $f_{\boldsymbol{X}}(x_1, \ldots, x_m)$. Det följer från kedjereglen att

$$f_{\boldsymbol{X}}(x_1, \ldots, x_m) = \prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{[i-1]}}(x_i|x_1, \ldots, x_{i-1}).$$

Anta också att det finns en delmängd $\mathrm{pa}(i) \subseteq [i-1]$ för alla $i = 1, \ldots, m$ så att

$$X_i \perp\!\!\!\perp \boldsymbol{X}_{[i-1]\setminus \mathrm{pa}(i)} | \boldsymbol{X}_{\mathrm{pa}(i)}.$$

Enligt Sats 3.2 har vi då att

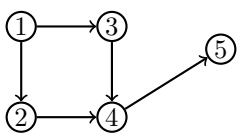
$$f_{\boldsymbol{X}}(x_1, \ldots, x_m) = \prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{\mathrm{pa}(i)}}(x_i|\boldsymbol{x}_{\mathrm{pa}(i)}).$$

Vi kan representera hierarkin med en graf. En *graf* är ett ordnat par $\mathcal{G} = (V, E)$ där V är en mängd som kallas *noderna* av grafen $\mathcal{G}$ och E är en delmängd av $V \times V = \{(i, j) : i, j \in V\}$ som kallas *kanterna* av $\mathcal{G}$. Man brukar rita graferna eftersom det ger oss en bra bild att arbeta med. Om $(i, j) \in E$ ritar vi $i \to j$.

Exempel 3.7. Betrakta grafen $\mathcal{G} = ([5], E)$ där

$$E = \{(1, 2), (1, 3), (2, 4), (3, 4), (4, 5)\}.$$

Då ritar vi $\mathcal{G}$ som

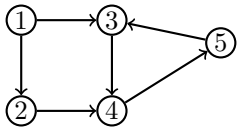


Om E innehåller både (i,j) och (j,i) ritar vi

och vi säger i och j förenas av en *oriktad kant* $i-j$. Om en graf $\mathcal{G}=(V,E)$ har inga oriktade kanter (dvs. om $(i,j)\in E$ då $(j,i)\notin E$) kallar vi $\mathcal{G}$ för en *riktad graf*. Till exempel, grafen i Exempel 3.7 är en riktad graf.

Vi ska använda riktade grafer för att representera hierarkier som håller för $(X_1,\dots,X_m)$ men för att göra det behöver vi ett språk för att beskriva hur man tar sig igenom en graf. En *stig* i en graf $\mathcal{G}=(V,E)$ är en ändlig sekvens av nodar $s=(v_1,\dots,v_t)$ där $v_1,\dots,v_t$ är olika och antingen $(v_i,v_{i+1})\in E$ eller $(v_{i+1},v_i)\in E$ för alla $i=1,\dots,t-1$. Stigen är *riktad* (från v_1 till v_t) om $(v_{i+1},v_i)\notin E$ för alla $i=1,\dots,t-1$. En *cykel* är en stig i en graf förutom vi låter $v_1=v_t$. En *riktad cykel* är en riktad stig förutom vi låter $v_1=v_t$.

Exempel 3.8. Betrakta grafen $\mathcal{G}=([5],E)$ som ritas



Stigar i $\mathcal{G}$ inkluderar $s_1=(1,2,4,5)$, $s_2=(3,4,2,1)$ och $s_3=(1,2,4,5,3)$. Stigarna s_1 och s_3 är riktade men inte s_2. Cyklarna i $\mathcal{G}$ inkluderar $c_1=(1,2,4,3,1)$, $c_2=(1,2,4,5,3,1)$ och $c_3=(5,3,4,5)$. Cykeln c_3 är riktad men inte c_1 eller c_2.

En riktad graf utan riktade cykler kallas för en *acyklisk riktad graf*. Till exempel, grafen i Exempel 3.7 är en acyklisk riktad graf men grafen i Exempel 3.8 är inte en acyklisk riktad graf. För att återvända till modellen $(X_1,\dots,X_m)$ vars täthetsfunktion (eller sannolikhetsfunktion) faktoriseras som

$$f_{\boldsymbol{X}}(x_1,\dots,x_m)=\prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{\mathrm{pa}(i)}}(x_i|\boldsymbol{x}_{\mathrm{pa}(i)})$$

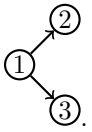
kan vi representera den underliggande hierarkin med en acyklisk riktad graf $\mathcal{G}=([m],E)$ där

$$E=\{(k,i):k\in\mathrm{pa}(i)\}.$$

Exempel 3.9. Vi såg i Exempel 3.3 att den simultana fördelningen av (X_1,X_2,X_3) för våra glödlampor faktoriseras som

$$f_{(X_1,X_2,X_3)}(x_1,x_2,x_3)=f_{X_2|X_1}(x_2|x_1)f_{X_3|X_1}(x_3|x_1)f_{X_1}(x_1).$$

Så vi kan representera fördelningen med grafen



Å andra sidan, givet en acyklisk riktad graf $\mathcal{G}=([m],E)$ låt

$$\mathrm{pa}_{\mathcal{G}}(i)=\{k\in[m]:(k,i)\in E\}.$$

Mängden $\mathrm{pa}_{\mathcal{G}}(i)$ kallas *föräldrar* av i i $\mathcal{G}$. Vi säger att en fördelning $\boldsymbol{X}=(X_1,\dots,X_m)$ är *Markovsk* till en acyklisk riktad graf $\mathcal{G}=([m],E)$ om

$$f_{\boldsymbol{X}}(x_1,\dots,x_m)=\prod_{i=1}^{m} f_{X_i|\boldsymbol{X}_{\mathrm{pa}_{\mathcal{G}}(i)}}(x_i|\boldsymbol{x}_{\mathrm{pa}_{\mathcal{G}}(i)}).$$

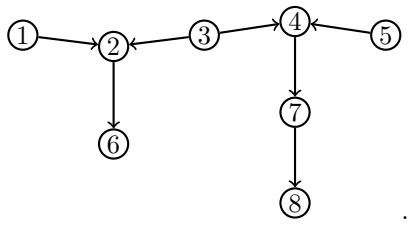
Om $\boldsymbol{X}$ är Markov till $\mathcal{G}$ kan $\mathcal{G}$ fångar olika betingat oberoende i fördelningen. Låt $A,B,C\subseteq[m]$ vara disjunkta delmängder av $[m]$ med $A,B\neq\emptyset$. Vi säger att A och B är *d-sammanhängande* givet C om det finns ett stig $(v_1,\dots,v_t)$ i $\mathcal{G}$ så att

1. $v_1\in A$ och $v_t\in B$,
2. om $v_{i-1}\to v_i\to v_{i+1}$, $v_{i-1}\leftarrow v_i\to v_{i+1}$, eller $v_{i-1}\leftarrow v_i\leftarrow v_{i+1}$ är kanter i $\mathcal{G}$ då $v_i\notin C$, och

3. om $v_{i-1} \to v_i \leftarrow v_{i+1}$ är kanter i $\mathcal{G}$ då $v_i \in C$ eller det finns ett riktad stig $(w_1, \ldots, w_\ell)$ från w_1 till w_ℓ i $\mathcal{G}$ där $w_1 = v_i$ och $w_\ell \in C$.

Om A och B är inte d-sammanhängande given C i $\mathcal{G}$ säger vi att A och B är *d-separerade* given C i $\mathcal{G}$. Om A och B är d-separerade given C i $\mathcal{G}$ skriver vi $A \perp_{\mathcal{G}} B|C$. Om A och B är d-sammanhängande given C i $\mathcal{G}$ skriver vi $A \not\perp_{\mathcal{G}} B|C$.

Exempel 3.10. Betrakta den acykliska riktade grafen $\mathcal{G} = ([8], E)$



Då har vi, exempelvis, att

$$1 \perp_{\mathcal{G}} 5|\emptyset, \qquad \{1,2,6\} \perp_{\mathcal{G}} \{5,7\}|\{3,4,8\}, \qquad \{2,3\} \perp_{\mathcal{G}} 5|\{1,6\},$$

och

$$1 \not\perp_{\mathcal{G}} 5|\{6,8\}, \qquad 1 \not\perp_{\mathcal{G}} 5|\{2,7\}, \qquad \{2,6\} \not\perp_{\mathcal{G}} \{5,7\}|\{1,8\}.$$

Enligt nästa sats kan d-separationer berätta för oss många olika och användbara betingat oberoende som håller i modellen när modellen är Markovsk till en acyklisk riktad graf.

Sats 3.6. *Om $\boldsymbol{X} = (X_1, \ldots, X_m)$ är Markovsk till en acyklisk riktad graf $\mathcal{G} = ([m], E)$ då $X_A \perp\!\!\!\perp X_B|X_C$ när A och B är d-separerad given C i $\mathcal{G}$.*

Beviset för denna sats är ganska komplicerat och mer passande för en kurs på masternivå. Vi kan dock använda satsen redan nu. Till exempel om vi har en fördelning $\boldsymbol{X} = (X_1, \ldots, X_8)$ som är Markovsk till grafen $\mathcal{G}$ i Exempel 3.10 vet vi att

- $X_1 \perp\!\!\!\perp X_5$,
- $\boldsymbol{X}_{\{1,2,6\}} \perp\!\!\!\perp \boldsymbol{X}_{\{5,7\}}|X_3$, och
- $\boldsymbol{X}_{\{2,3\}} \perp\!\!\!\perp X_5|\boldsymbol{X}_{\{1,6\}}$.

En bra representation av $\boldsymbol{X} = (X_1, \ldots, X_m)$ med en acyklisk riktad graf kan därmed avslöja komplexa betingat oberoende struktur i modellen. Det är naturligt att fråga, "Vilken graf är den rätta grafen för att representera min modell?" Generellt sett kan $(X_1, \ldots, X_m)$ vara Markovsk till många olika acykliska riktade grafer. Till exempel, om vi bestämma att $(X_1, \ldots, X_m)$ är Markovsk till en viss graf $\mathcal{G}$ kan det redan finns andra grafer som $(X_1, \ldots, X_m)$ är Markovsk till också. Vi säger att två acykliska riktade grafer $\mathcal{G} = ([m], E)$ och $\mathcal{H} = ([m], E')$ är *Markovekvivalena* om de har samma d-separationer; dvs.

$$A \perp_{\mathcal{G}} B|C \qquad \text{om och endast om} \qquad A \perp_{\mathcal{H}} B|C$$

för alla $A, B, C \subseteq [m]$. Det kan vara svårt att bestämma alla d-separationer i två grafer $\mathcal{G}$ och $\mathcal{H}$ och verifiera att de är lika. Lyckligtvis finns det ett enklare sätt att kolla om två grafer är Markovekvivalent. För att beskriva denna metod behöver vi två definitioner:

Skelettet av en acyklisk riktad graf $\mathcal{G} = ([m], E)$ är den underliggande oriktade grafen; dvs. den oriktade grafen $G = ([m], E')$ där

$$E' = E \cup \{(i,j) : (j,i) \in E\}.$$

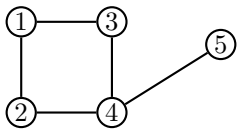
Exempel 3.11. Skelettet av grafen $\mathcal{G} = ([5], E)$ med

$$E = \{(1,2), (1,3), (2,4), (3,4), (4,5)\}$$

från Exempel 3.7 är den oriktade grafen $G = ([5], E')$ där

$$E = \{(1,2), (2,1), (1,3), (3,1), (2,4), (4,2), (3,4), (4,3), (4,5), (5,4)\}.$$

Vi ritar G som



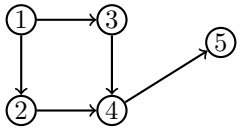
En *v-struktur* i en graf $\mathcal{G} = ([m], E)$ är ett stig (i, j, k) i $\mathcal{G}$ där $(i, j), (k, j) \in E$ och $(i, k), (k, i) \notin E$.

Exempel 3.12. Grafen $\mathcal{G} = ([5], E)$ i Exempel 3.7 innehåller bara en v-struktur: $(2, 4, 3)$. Vi noterar att $(2, 4, 3)$ skulle inte vara en v-struktur om $(2, 3)$ eller $(3, 2)$ var också en kant i $\mathcal{G}$.

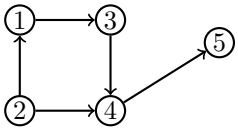
Det visar sig att skelettet och v-stukturer räcker för att bestämma Markovekvivalens.

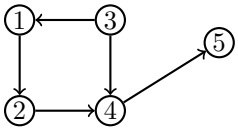
Sats 3.7. *Två acykliska riktade grafer är Markovekvivalena om och endast om de har samma skelettet och de har samma v-strukturer.*

Exempel 3.13. Grafen $\mathcal{G} = ([5], E)$ i Exempel 3.7:



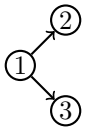
är Markovekvivalent till två andra acykliska riktad grafer:



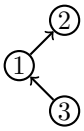


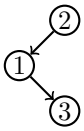
Markovekvivalensklassen av en acyklisk riktad graf $\mathcal{G}$ är mängden av alla grafer som är Markovekvivalent till $\mathcal{G}$. Exempelvis har grafen $\mathcal{G}$ i Exempel 3.13 en Markovekvivalensklass med tre element.

Exempel 3.14. Vi såg i Exempel 3.9 att den simultana fördelningen (X_1, X_2, X_3) av våra tre glödlampor är Markovsk till grafen



Det följer från Sats 3.7 att modellen kan också representeras med två andra hierarkier:





eftersom alla tre har det samma skelettet och ingen v-strukturer. Den första grafen är intuitivt det bästa val för att representera systemet eftersom vi vet att det finns en *kausal effekt* av det första glödlampan på de andra två; nämligen, vi slår på glödlampor två och tre bara efter glödlampa ett har brunnit ut. Men bara baserad på fördelningen kan vi inte bestämma oss att de andra grafer inte är den riktiga *kausalmodellen* (därinne; vi tolkar kanten $i \to j$ att menar X_i har en kausal effekt på X_j). Det vill säga att om vi har bara ett stickprov $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ från fördelingen av glödlamporna $\boldsymbol{X} = (X_1, X_2, X_3)$ och vi vet inte kausalrelationarna mellan glödlamporna kan vi bara använda stickprovet för att lära oss $X_2 \perp\!\!\!\perp X_3 | X_1$. Baserad på bara denna observation kan vi representera modellen med tre olika grafer. Vi kan inte dock bestämma vilken graf är den rätta kausalmodellen för systemet. (Är det glödlampa 1 som har en kausal effekt på 2 och 3 eller har 3 en kausal effekt på 2 och 2 en kausal effekt på 3?) Att välja den rätta grafen för att representera det underliggande kausala systemet är ett av huvudfokusen av underområdet av artificiell intelligens som kallas *kausal inferens*.