Exercise 8.6.1. *For three variables $X = (X_1, X_2, X_3)$ with density $f(x_1, x_2, x_3)$. Show that $1 \perp\!\!\!\perp 2|3$ and $1 \perp\!\!\!\perp 3$ implies that $1 \perp\!\!\!\perp 2$. In the special case when $X \sim N_3(0, \Sigma)$ is multivariate Gaussian this can be shown by expressing $1 \perp\!\!\!\perp 2|3$ and $1 \perp\!\!\!\perp 3$ in terms of the restrictions on the covariance matrix (what are those?) and concluding $\Sigma_{12} = 0$.*

$\underline{1 \perp\!\!\!\perp 2 \mid 3}$, $X_1 \perp\!\!\!\perp X_2 \mid X_3$ (1)

$$f(x_1, x_2 \mid x_3) = f(x_1 \mid x_3)\, f(x_2 \mid x_3)$$

$X_1 \perp\!\!\!\perp X_3$: $f(x_1 \mid x_3) = f(x_1)$ (2) $\Sigma_{13} = 0$

$$f(x_1, x_2) = \int f(x_1, x_2, x_3)\, dx_3$$

$$= \int f(x_1, x_2 \mid x_3)\, f(x_3)\, dx_3$$

(1) $$= \int \underbrace{f(x_1 \mid x_3)}\, f(x_2 \mid x_3)\, f(x_3)\, dx_3$$

(2) $$= \int \underline{f(x_1)}\, \underbrace{f(x_2 \mid x_3)\, f(x_3)}\, dx_3$$

$$= f(x_1)\int f(x_2, x_3)\,dx_3$$

$$= f(x_1)\cdot f(x_2).$$

$X \sim N(0, \Sigma)$

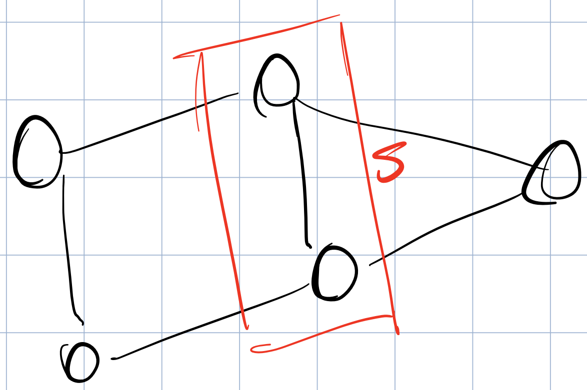
$$\Sigma_{12} - \underline{\Sigma_{13}}\,\Sigma_{33}^{-1}\,\Sigma_{32} = 0 \quad (2) \qquad (X_1 \perp\!\!\!\perp X_2 \mid X_3)$$

$$\Sigma_{12} - 0\,\cancel{\Sigma_{33}^{-1}\,\Sigma_{32}}^{\,0} = 0 \qquad (1)$$

$$\longrightarrow \Sigma_{12} = 0 \xRightarrow{\text{f. Gaussian}} \underline{X_1 \perp\!\!\!\perp X_2}.$$

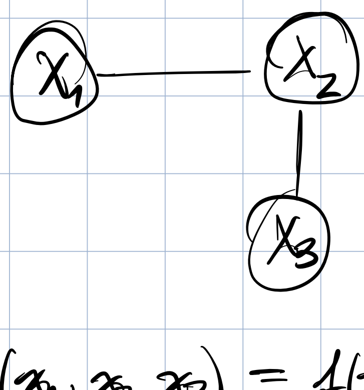
Exercise 8.6.2. *Consider all binary distributions that factorize $f(x_1, x_2, x_3) = \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)$ for $(x_1, x_2, x_3) \in \{0,1\}^3$. Prove the Hammerley-Clifford theorem in this very special case.*

Theorem: (H-C) $f > 0$, (F) $\Longleftrightarrow$ (G) $\Longleftrightarrow$ (P)



(G)

$X_i \perp\!\!\!\perp X_j \mid X_{\{-i,-j\}}$ (P)



(G) $\Longleftrightarrow$ (P)

$$\frac{f(x_1, x_2, x_3)}{f(x_2)} = f(x_1, x_3 \mid x_2)\cancel{f(x_2)} = \frac{\phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)}{f(x_2)}$$

$$f(x_2) = \int \underbrace{\phi_{12}(x_1, x_2)}\ \underbrace{\phi_{23}(x_2, x_3)}\, dx_1\, dx_3$$

$$= \underbrace{\left(\int \phi_{12}(x_1, x_2)\, dx_1 \right)}_{h(x_2)} \underbrace{\left(\int \phi_{23}(x_2, x_3)\, dx_3 \right)}_{g(x_2)}$$

$$f(x_1, x_3 \mid x_2) = \frac{\phi_{12}(x_1, x_2)}{h(x_2)} \cdot \frac{\phi_{23}(x_2, x_3)}{g(x_2)}$$

$$\therefore \ (F) \Longrightarrow (P).$$

Exercise 8.6.3. *Show that the set $\mathbb{S}^m_+$ is convex. Also, show that intersection of two convex sets must be also convex.*

$A, B \in \mathbb{S}^m_+, \quad C = \lambda A + (1-\lambda)B, \quad \lambda \in [0,1]$

C is symmetric: $C^T = (\lambda A + (1-\lambda)B)^T$

$$= \lambda A^T + (1-\lambda)B^T$$

$$= \lambda A + (1-\lambda)B = C \in \mathbb{S}^m_+$$

C is pd.: $x^T C x = x^T(\lambda A + (1-\lambda)B)x$

$$= \lambda \underset{\substack{\vee \\ 0}}{x^T A x} + (1-\lambda)\underset{\substack{\vee \\ 0}}{x^T B x} \in \mathbb{S}_+ . \square$$

U, V convex, $W = U \cap V$, $x, y \in W$, $\lambda \in [0,1]$

$z = \lambda x + (1-\lambda)y$: if $\begin{cases} x, y \in U \Longrightarrow z \in W \subseteq U \\ x, y \in V \Longrightarrow z \in W \subseteq V \end{cases}$

$\lambda x + (1-\lambda)y \in U$, because $x, y \in U$

$\in V$, because $x, y \in V$

$\Longrightarrow \lambda x + (1-\lambda)y \in W.$ $\square$

Exercise 8.6.5. *Suppose we compare a graph G with G_0 obtained from G by removing a single edge. What difference in likelihood is needed so that the AIC/BIC criteria prefer the bigger model?*

$$G_0 = G \setminus \{x_i, x_j\}$$

$$AIC(G) = -2\log L + 2k$$

$$AIC(G_0) = -2\log \hat{L} + 2(k-1)$$

$$AIC(G) - AIC(G_0) < 0$$

$$-2\log\frac{L}{\hat{L}} + 2 < 0$$

$$\underline{\log}\frac{L}{\hat{L}} > 1,$$

$$BIC(G) = -2\log L + \log(n)k$$

$$BIC(G) - BIC(G_0) = -2\log\frac{L}{\hat{L}} + \log(n) < 0$$

$$\log\frac{L}{\hat{L}} > \log\left(\frac{1}{\sqrt{n}}\right) \implies \frac{L}{\hat{L}} > \frac{1}{\sqrt{n}}.$$