1) THEORETICAL BACKGROUND

Multivariate Gaussian model (this is a key assumption for the DAGs)

$$\underline{x}_1, \dots, \underline{x}_n \overset{iid}{\sim} N_q(\underline{0}, \Omega^{-1}) \qquad \underline{x}_i \in \mathbb{R}^q$$

We'll do inference on this

Ω s.p.d. = Precision matrix

$\Omega \overset{iid}{\sim} \text{Wishart}_q(a, U)$ where
- $a > q-1$
- U s.p.d.

$$p(\Omega) = c(a, U) \cdot |\Omega|^{\frac{a-q-1}{2}} \cdot \exp\left\{-\frac{1}{2} \operatorname{tr}(U\Omega)\right\}$$

$$p(\underline{x}_1, \dots, \underline{x}_n \mid \Omega) = \dots = (2\pi)^{-\frac{nq}{2}} \cdot |\Omega|^{\frac{n}{2}} \cdot \exp\left\{-\frac{1}{2} \operatorname{tr}(S\Omega)\right\}$$

where $S = \sum_{i=1}^{n} \underline{x}_i \cdot \underline{x}_i^T = X^T X \in \mathbb{R}^{q \times q}$

where X is our data matrix $n \times q$

$$\underbrace{p(\Omega \mid \underline{x}_1, \dots, \underline{x}_n)}_{\text{posterior}} \propto \underbrace{p(\underline{x}_1, \dots, \underline{x}_n \mid \Omega) \cdot p(\Omega)}_{\text{prior}}$$

$$\propto |\Omega|^{\frac{a+n-q-1}{2}} \cdot \exp\left\{-\frac{1}{2} \operatorname{tr}\left[(U+S) \cdot \Omega\right]\right\}$$

$$\Rightarrow p(\Omega \mid \underline{x}_1, \dots, \underline{x}_n) \sim W_q(a+n, U+S)$$

the posterior is just another Wishart with modified parameters:
- $a +$ sample size
- $U + X^T X$

$$c(a, U) = \frac{|U|^{\frac{a}{2}}}{2^{\frac{aq}{2}} \cdot \Gamma_q\left(\frac{a}{2}\right)} \qquad \text{prior normalizing constant}$$

$$c(a+n, U+S) = \frac{|U+S|^{\frac{a+n}{2}}}{2^{\frac{(a+n)q}{2}} \cdot \Gamma_q\left(\frac{a+n}{2}\right)} \qquad \text{posterior normalizing constant}$$

Model selection through marginal likelihoods:

$$p(\underline{x}_1, \dots, \underline{x}_m) = \int p(\underline{x}_1, \dots, \underline{x}_n | \Omega) \cdot p(\Omega)\, d\Omega = \dots =$$

$$= (2\pi)^{-\frac{mq}{2}} \cdot \frac{c(a, U)}{c(a+m, U+S)} = p(X)$$

Marginal distributions:

$A \subseteq \{1, \dots, q\}$ A any subset of index of the var.

I would like to have the marginal likelihood of X_A data matrix with only the variables of columns A

$$p(X_A) ?$$

$$\underline{x}_1^A, \dots, \underline{x}_m^A \sim ?$$

$$\underline{x}_1^A, \dots, \underline{x}_m^A \,|\, ? \sim N_{|A|}(\underline{0}, (\underline{\Omega_{A|\bar{A}}})^{-1}) \quad (\star)$$

Schur complement of Ω:

$$\Omega_{A|\bar{A}} = \Omega_{AA} - \Omega_{A\bar{A}} (\Omega_{\bar{A}\bar{A}})^{-1} \cdot \Omega_{\bar{A}A} = (\Sigma_{AA})^{-1}$$

where: $\bar{A} = \{1, \dots, q\} \setminus A$ (all columns not in A)

Ω_{AA} is block of Ω with rows of A and columns of A

$$\Omega_{A|\bar{A}} \sim W_{|A|}(a - |\bar{A}|, U_{AA}) \quad (\star\star)$$

Note: since we have the prior on Ω, I'm doing all the computation using Ω instead of $\Sigma = \Omega^{-1}$

$(\star) + (\star\star) \Longrightarrow$ we can proceed as before:

$$\Omega_{A|\bar{A}} \mid \underline{x}_1^A, \dots, \underline{x}_m^A \sim W_{|A|}\left(a+m-|\bar{A}|,\ U_{AA}+S_{AA}\right)$$

$$\Longrightarrow p(\underline{x}_1^A, \dots, \underline{x}_m^A) = p(X^A) =$$

$$= \int \underbrace{p(\underline{x}_1^A, \dots, \underline{x}_m^A \mid \Omega_{A|\bar{A}})}_{\text{likelihood } N_{|A|}} \cdot p(\underbrace{\Omega_{A|\bar{A}}}_{\sim W_{|A|}})\, d\Omega_{A|\bar{A}} =$$

$$= \boxed{(2\pi)^{-\frac{m|A|}{2}} \cdot \frac{c(a-|\bar{A}|,\ U_{AA})}{c(a+m-|\bar{A}|,\ U_{AA}+S_{AA})} =: p(X_A)} \quad \Delta$$

where
$$c(a-|\bar{A}|,\ U_{AA}) = \frac{|U_{AA}|^{\frac{a-|A|}{2}}}{2^{\frac{(a-|\bar{A}|)\cdot|A|}{2}} \cdot \Gamma_{|A|}\left(\frac{a-|\bar{A}|}{2}\right)},$$

$$c(a+m-|\bar{A}|,\ U_{AA}+S_{AA}) = \frac{|U_{AA}+S_{AA}|^{\frac{a+m-|A|}{2}}}{2^{\frac{(a+m-|\bar{A}|)\cdot|A|}{2}} \cdot \Gamma_{|A|}\left(\frac{a+m-|\bar{A}|}{2}\right)}$$

We want to implement the following algorithm:

Input:		Output:
$\boxed{a, U, X}$	$\longrightarrow$	$\boxed{p(X_A)}$

2) DAG

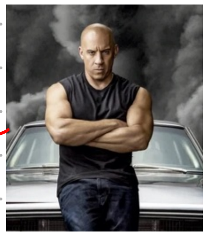
$\mathcal{D} = (V, E)$ $V = \{1, \dots, q\}$ = set of nodes = variables

$E \subseteq V \times V$ = edges

$(u, v) \in E \Longrightarrow \exists\, u \to v$ in $\mathcal{D}$

$pa_{\mathcal{D}}(j) = \{u : (u, j) \in E\}$ = parents of j

$fa_{\mathcal{D}}(j) = j \cup pa_{\mathcal{D}}(j)$ = family of j



Under $\mathcal{D}$ the joint p.d.f. (since we are in Gauss setting) of $X_1, \dots, X_q$ is:

$$p(x_1, \dots, x_q \mid \mathcal{D}) = \prod_{j=1}^{q} p(x_j \mid pa_{\mathcal{D}}(j))$$

x_A is sub-vector of $\underline{x}$ with columns in A.

In a Gauss DAG model:

$$\underline{x}_1, \dots, \underline{x}_n \mid \Omega_{\mathcal{D}} \overset{iid}{\sim} N_q(\underline{0}, \Omega_{\mathcal{D}}^{-1})$$

where $\Omega_{\mathcal{D}}$ is:
- spd (as before)
- Markov wrt $\mathcal{D}$ (satisfies the conditional independence of the DAG)

prior of Ω_D ? I need Wishart + conditional dependencies of D through Markov

Multiple methods to find it:

i) DAG - Wishart prior (Cao et al., Annals of Statistics)

prior of the modified Cholesky decomposition

$$\Omega \longrightarrow (D, L)$$

$\boxed{ii)}$ if the goal is to compute $p(X \mid D)$, i.e. the marginal likelihood of D:

Start from a prior on the parameter of a complete DAG model $\Longrightarrow$ (in the Gauss setting (our)) Ω is Wishart: $\Omega \sim W_q(a, U)$,

then:

$$p(X \mid D) = \prod_{j=1}^{q} \frac{p(X_{fa_D(j)})}{p(X_{pa_D(j)})} \qquad (\star\star\star)$$

$p(X_{fa_D(j)})$ and $p(X_{pa_D(j)})$ are computed as $\triangle$

(Geiger and Heckerman (2002), Annals of Statistics)

(Consonni and La Rocca (2012), Scandinavian Journal of Statistics)

Finding parents and family of j:

(Castelletti and Mascaro)

Search-and-score method:

Starting $\mathcal{D} \longrightarrow$ construct $\mathcal{D}^*$, then accept / reject $\mathcal{D}^*$ based on:

$$p(\text{accept } \mathcal{D}^*) \propto \frac{p(X \mid \mathcal{D}^*)}{p(X \mid \mathcal{D})} \cdot \frac{p(\mathcal{D}^*)}{p(\mathcal{D})} \cdot \frac{q(\mathcal{D} \mid \mathcal{D}^*)}{q(\mathcal{D}^* \mid \mathcal{D})}$$

$$\alpha_{\mathcal{D}^*, \mathcal{D}} = \min \left\{ 1; \frac{p(X \mid \mathcal{D}^*)}{p(X \mid \mathcal{D})} \cdot \frac{p(\mathcal{D}^*)}{p(\mathcal{D})} \cdot \frac{q(\mathcal{D} \mid \mathcal{D}^*)}{q(\mathcal{D}^* \mid \mathcal{D})} \right\}$$

$\mathcal{D}^*$ accepted with prob. $\alpha_{\mathcal{D}^*, \mathcal{D}}$

3) PC-algorithm:

$$X_u \perp\!\!\!\perp X_v \mid X_S \qquad\qquad S \subseteq \{1,\dots,q\} \setminus \{u,v\}$$

Until now we know how to compute only through frequentist method. With Gauss assumpt. $\Rightarrow$ test on partial correlation coefficient.

$X = (X_1, \dots, X_q)$

$\rho_{u,v|S} = \operatorname{corr}(X_u, X_v \mid X_S)$

$A = \{u, v, S\} \xrightarrow[\text{we compute}]{}$ marginal distribution of X_A

$$X_u, X_v \mid X_S \sim N_2(\dots, \dots)$$

we should use these to compute $\rho_{u,v|S}$

But! Testing $\rho_{u,v|S} = 0$ is the same as

testing $\underline{L_{u|v,S}} = 0 \leftarrow$ regression coefficient of u on $\{v, S\}$

Interpretation: $D^{(1)}$:

$$\begin{tikzcd} & S \arrow[dl] \arrow[dr] & \\ v \arrow[rr] & & u \end{tikzcd}$$

$L_{u|v,S}$ measure how much important the $v \to u$ arrow.

$L_{u|v,S} = 0 \Longrightarrow$ no arrow

and it's the same as $p(x_u \mid x_{pa(u)})$

Remember! We are in the skeleton estimation phase, therefore I don't need to know the direction of the arrow, I just want to know if $u \perp\!\!\!\perp v \mid S$ for now.

Let $\mathcal{D}^{(0)}$:

$$\begin{tikzcd} & S \arrow[dl, bend right] \arrow[dr, bend left] & \\ v & & u \end{tikzcd}$$

Compare, in a Bayesian way, $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(0)}$ through Bayes-factor:

$$BF_{01} = \frac{m(X \mid \mathcal{D}^{(0)})}{m(X \mid \mathcal{D}^{(1)})}$$

← marginals that I find with the previous formulas

Based on BF_{01} I decide if I include the $v \to u$ arrow.

$$p(X \mid \mathcal{D}) = \prod_{j=1}^{q} \frac{p(X_{fa_{\mathcal{D}}(j)})}{p(X_{pa_{\mathcal{D}}(j)})} \longrightarrow$$

Notice that since we are in Gauss setting, this should be exactly equal to $p(u \to v)$

$\mathcal{D}^{(1)}$:

$$\begin{tikzcd} & S \arrow[dl, bend right] \arrow[dr, bend left] & \\ v \arrow[rr] & & u \end{tikzcd}$$

$\mathcal{D}^{(0)}$:

$$\begin{tikzcd} & S \arrow[dl, bend right] \arrow[dr, bend left] & \\ v & & u \end{tikzcd}$$

$$\frac{p(X \mid \mathcal{D}^{(0)})}{p(X \mid \mathcal{D}^{(1)})} = \frac{\dfrac{p(X_{fa_0(u)})}{p(X_{pa_0(u)})}}{\dfrac{p(X_{fa_1(u)})}{p(X_{pa_1(u)})}}$$

I need only node u, difference between u v and $u \to v$ will emerge by itself

Next step

i) R function to evaluate $p(X_A)$ (see 1))

ii) DAG $\mathcal{D}$ $\rightsquigarrow$ function to evaluate $p(X \mid \mathcal{D})$ (using $p(X_A)$) (see 2))

iii) Modify the pcalg "skeleton" function to use BF instead of the already implemented pc algorithm function