Structure Learning, Conditional densities and prior distributions

1: Define generative model and priors

Conditional on the parents $pa_G(Y_t^i)$
each index Y_t^i is normally distributed with
mean μ_t^i given by a linear function of the parent variable values

$$Y_t^i|\mathrm{pa}_G(Y_t^i), \tilde{\tau}_i^2 \sim N(\mu_t^i, \tilde{\tau}_i^{-2})$$

$$\mu_t^i = \beta_0^i + \sum_{j=1}^{p_i} \beta^i_{(k_j,\tau_j)} Y^{k_j}_{t-\tau_j}$$

Congugate normal-gamma priors specified for the regression coefficients
$\beta^i_{(k_j,\tau_j)}$ and the conditional precision $\tilde{\tau}_i^2$

$$\tilde{\tau}_i^2 \sim \mathrm{Gamma}(a_\tau, b_\tau)$$

$$\beta_0^i|\tilde{\tau}_i^2 \sim N\left(0, \frac{\nu_i^2}{\tilde{\tau}_i^2}\right)$$

$$\beta^i_{(k_j,\tau_j)}|\tilde{\tau}_i^2, \mathrm{pa}_G(Y_t^i) \sim N\left(0, \frac{\nu_i^2}{\tilde{\tau}_i^2}\right) \quad j = 1, \ldots, p_i$$

2: Learn graph structure and model parameters

Given data $D = y_1, \ldots, y_T$ where y_t denotes the values of the random variables
$Y_t = (Y_t^1, \ldots, Y_t^n)^T$ at time t, learning the structure G and parameters θ, where parameters θ
consist of the coefficients $\beta^i_{(k_j,\tau_j)}$, and the conditional precision $\tilde{\tau}_i^2$

$$P(\theta, G|D) = P(\theta|G, D)P(G|D)$$

3: Score based approach to model selection

Sampling from the full posterior distribution of possible graphs $P(G|D)$ requires
evaluation of the marginal likeihood

$$P(D|G) = \int d\theta P(D|G, \theta)P(\theta|G)$$

4: Apply MCMC

Given a current candidate structure G, the sampling scheme proceeds by proposing a new
structure G' according to a proposal distribution $q_G(G'; G)$.
The proposal G' is accepted with probability

$$\alpha = \min\left\{1, \frac{q_G(G; G')}{q_G(G', G)} \frac{P(D|G')}{P(D|G)} \frac{P(G')}{P(G)}\right\}$$

Set maximum time-lag τ_{max} and impose a maximum size p_{max}
on the allowed parent sets in order to sparsify the networks

$$P(\mathrm{pa}_G(Y_t^i)) = \begin{cases} \left[\sum_{j=0}^{p_{\max}} \binom{n\tau_{\max}}{j}\right]^{-1}, & |\mathrm{pa}_G(Y_t^i)| \le p_{\max}, \\ 0, & \text{otherwise.} \end{cases}$$

We also adopt a uniform proposal density on graphs G'
in the neighbourhood of the current graph G,

$$q_G(G', G) = \begin{cases} \frac{1}{|\mathrm{nhd}(G)|}, & G \in \mathrm{nhd}, \\ 0, & \text{otherwise.} \end{cases}$$

5: Determine summary posterior distribution

From a sample of size S from the posterior distribution $P(G|D)$, distributional estimates for
structural uncertainties can be quantified by taking Δ to be an indicator function for the presence of
a given edge, given the observed data, obtained by averaging over the sample where $G^{(s)}$ is the
s^{th} structure sample.

$$\Pr(\Delta|D) = \sum_{G \in \mathcal{G}} \Pr(\Delta|G, D)P(G|D)$$

$$\approx \frac{1}{S} \sum_{s=1}^{S} \Pr(\Delta|G^{(s)}, D)$$

quantifies the posterior probability $\hat{\pi}$ for the presence of an edge conditional
on the observerational data