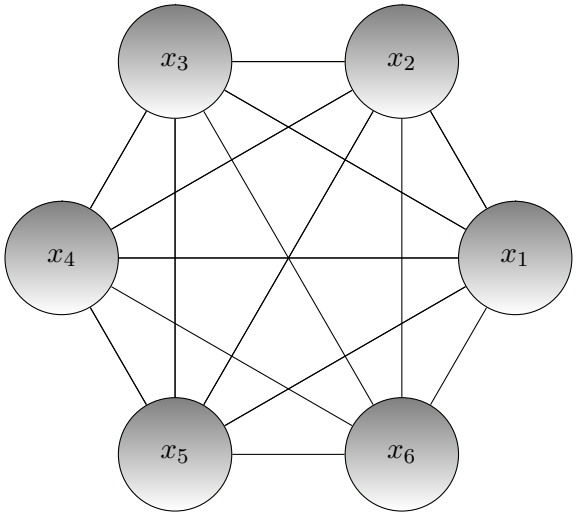
October 29, 2018

1 Gene network



- Each node corresponds to a gene
- Each edge corresponds to a pair of genes which are 'co-expressed'

'co-expressed' meaning that their expression levels are highly correlated. The weights represent the strength of the connection between two genes.

True correlation between gene m,n ρ_{mn} and correlation by observed data is $r_{mn} = corr(x_m, x_n)$ by Fisher's z-transformation $w_{mn} = arctanh(r_{mn})$ aproximately normal distribution .

$$w_{mn} \sim N(arctanh(\rho_{mn}), \frac{1}{N-3}) \tag{1}$$

Let E =$\{e_{mn}\}$ be a set of true edges in the network.

The number of Gene G is larger and that most paris ar not co-expressed, and the number of possible edge $K = G(G-1)/2$ is very larger than the number of true edges in the network $|E|$

So in this gene network model we have some assumption

- This graphical model is undirected model because each node means correlation
- The network can be cycle??
- The network is sparse ($|E| << K$)

- The weight w_{mn} follows L_2N model

L_2N model is a three component mixture model. First the 'null' component follow normal distribution with mean 0, which means majority of correlation between genes are 0. And the 'not null' components which mean strong positive/negative correlation follow log-normal distribution each.

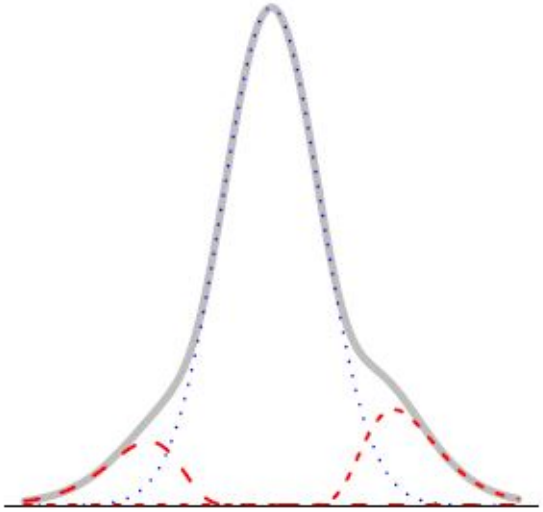


Figure 1: L_2N model

$$w_{mn}|e_{mn} \notin E \sim N(0, \sigma^2) \tag{2}$$

$$w_{mn}|[w_{mn} > 0, e_{mn} \in E] \sim LogNormal(\theta_1, \kappa_1^2) \tag{3}$$

$$-w_{mn}|[w_{mn} < 0, e_{mn} \in E] \sim LogNormal(\theta_2, \kappa_2^2) \tag{4}$$

$\sigma^2 = 1/(N-3) + \sigma_0^2$,where $1/(N-3)$ from Fisher's z-transformation and σ_0^2 from random effect model. Let C_0, C_1, C_2 means the distribution in the mixture model C_0 the null component, C_1, C_2 the 2 not null components.

$$Pr(e_{mn} \in C_j|w_{mn}) = \frac{P_j f_j(w_{mn})}{P_0 f_0(w_{mn}) + P_1 f_1(w_{mn}) P_2 f_2(w_{mn})} \tag{5}$$

$\mathbf{b}_{mn}$ =$(b_{mn0}, b_{mn1}, b_{mn2})$ means the indicator vector $b_{mnj} = 0$ when the $Pr(e_{mn} \in C_j|w_{mn})$ has the highest probablity and 0 for the other two. Make adjacency $G \times G$ matrix $\mathbf{A}$, where $\mathbf{A}_{mn} = 1 - b_{mn0}$,

2 Bayesian Dirichlet

Bayesian network learning($P(B|D)$) can be decomposed to structure learning($P(\mathcal{G}|D)$) and parameter learning($P(\theta|\mathcal{G}, D)$)

$$P(B|D) = P(\mathcal{G}, \theta|D) = P(\mathcal{G}|D) * P(\theta|\mathcal{G}, D) \tag{6}$$

$$P(\mathcal{G}|D) \propto P(D|\mathcal{G})P(\mathcal{G}) \tag{7}$$

under assumption

- Multinomial sample
- Dirichlet
- Parameter independence
  1. $P(\theta|\mathcal{G}) = \prod_{i=1}^{n} P(\theta_i, \mathcal{G})$ - global parameter independence
  2. $P(\theta_i|\mathcal{G}) = \prod_{j=1}^{q_i} P(\theta_{ij}, \mathcal{G})$ for all $i = 1, .., n$ - local parameter independence
- Parameter modularity

  Given two directed acyclic graphs, $\mathcal{G}$ and$\mathcal{G}'$, such that $P(\mathcal{G}) > 0$ and $P(\mathcal{G}') > 0$, if X_i has the same parents in $\mathcal{G}$ and $\mathcal{G}'$, then $P(\theta_{ij}|\mathcal{G}) = P(\theta_{ij}|\mathcal{G}')$

if samples are iid we can evaluate $P(D|\mathcal{G})$ by taking the product $P(x_i|D, \mathcal{G})$

$$P(x_l|D, \mathcal{G}) = \prod_{i}^{n}\prod_{j}^{q_i}\prod_{k}^{r_i} \frac{\alpha_{ijk:l} + n_{ijk:l}}{\sum_k(\alpha_{ijk:l} + n_{ijk:l})} \tag{8}$$

$$P(D|\mathcal{G}) = \prod_{l}^{N}\prod_{i}^{n}\prod_{j}^{q_i}\prod_{k}^{r_i} \frac{\alpha_{ijk:l} + n_{ijk:l}}{\sum_k(\alpha_{ijk:l} + n_{ijk:l})} \tag{9}$$

$$= \prod_{i}^{n}\prod_{j}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k}^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \tag{10}$$

$$P(\mathcal{G})P(D|\mathcal{G}) = P(\mathcal{G}) \prod_{i}^{n}\prod_{j}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k}^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \tag{11}$$

We called Bayesian Dirichlet scoring fuction

- If we set $\alpha_{ijk} = 1$, then it is called the K2 metric
- We can retrict hyperparameters to ensure likelihood equivalence

  $\alpha_{ijk} = \alpha P(X_i = k, Pa(X_i) = j|\mathcal{G})$ this is BDe

- Typically, uniformative hyperparamethers are used
  $\alpha_{ijk} = \frac{\alpha}{r_i q_i}$ this is BDeu

3 Hill Climbing

A bayesian network, DAG $\mathcal{G} = (V, E)$, $V = (X_1, ..., X_n)$ represents node, E represents conditional independence(Structure)

Under the observed dataset D we can get the structure by

$$\mathcal{G} = \arg\max_{\mathcal{G} \in \mathcal{G}^n} f(\mathcal{G} : D) \tag{12}$$

$f(\mathcal{G} : D)$ is a scoring metric(BIC or e Bayesian Dirichlet equivalent...) if scoring metric f is decomposable then

$$f(\mathcal{G} : D) = \sum_{i=1}^{n} f_D(X_i, Pa\mathcal{G}(X_i)) \tag{13}$$

$Pa\mathcal{G}(X_i)$ is the parents of X_i under the structure $\mathcal{G}$

- Addition of $X_j \to X_i : f_D(X_i, Pa\mathcal{G}(X_i) \cup X_j) - f_D(X_i, Pa\mathcal{G}(X_i))$
- Deletion of $X_j \to X_i : f_D(X_i, Pa\mathcal{G}(X_i) - X_j) - f_D(X_i, Pa\mathcal{G}(X_i))$
- Reversal of $X_j \to X_i : [f_D(X_i, Pa\mathcal{G}(X_i) - X_j) - f_D(X_i, Pa\mathcal{G}(X_i))] + [f_D(X_j, Pa\mathcal{G}(X_j) \cup X_i) - f_D(X_j, Pa\mathcal{G}(X_j))]$

First, calculate the score of the input structure. Second add the arc which does not violate the acycle and get a arc which has highest score. and do it for the deletion and reversal. Third change the structure which has the highest score if it is larger than original score. Repeat this process we can get the structure which has the local maximum score.

When we want to get the structure by HC algorithm we input the Data D and initial structure $\mathcal{G}_0$ and metric f, but the estimated structure is local maximum.