

MULTI-LAYER HIERARCHICAL BAYESIAN PROBABILISTIC INTERACTION
MODELING WITH INFORMATIVE PRIOR PROBABILITY FOR GENE
EXPRESSION

BY

XIAOHUAN XUE

A Thesis Submitted to the Graduate Faculty of
WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF ARTS

Mathematics and Statistics

May 2019

Winston-Salem, North Carolina

Approved By:

James Norris, Ph.D., Advisor

Edward Allen, Ph.D., Chair

David John, Ph.D.

Dedication

To Dr. Han Du.
She is the fire burning in my heart.
When I was in the wilderness,
insignificant, inadequate and invisible,
She found me.

Acknowledgments

I would like to express my deep gratitude to Professor James Norris for being my thesis advisor and his patient guidance and precious advice. I have learned a lot in doing research and scientific writing from him. His commitment to research, kindness to students, and contribution to community, make him a role model to me.

I would like to thank Dr. David John and Dr. Edward Allen for being on my thesis committee. I am also grateful for Dr. Gloria Muday and the Center of Molecular Signaling at Wake Forest University for the research assistantship and freeing me from the financial concerns, which has been important to me during my graduate study.

Assistance provided by the faculty and staffs at Wake Forest University was greatly appreciated. Particularly I would like to thank Dr. Jennifer Erway, Dr. Miaohua Jiang and Professor Jule Connolly. I would like to express my great appreciation to Dr. Leigh Watson at the Department of Psychology for her kindness and insights.

My friends, (Vikki) Kehan Sui, Yutong Yang, and Mengxuan Li made my world a better place by being in it. My years at Wake are special and precious because of you. My conversations with Mengxuan are filled with entertaining thoughts and laughter. Mengxuan, you are invincible. I also would like to thank Xiawei Dong, Nan Ji and Yue Zhao for being by my side even when we are on different pieces of the planet.

I am grateful to my former advisor, Professor Qun Lin at the Chinese Academy of Sciences, who envisioned the bigger picture of math to me, heals the wounds in my heart, trusts me and loves me unconditionally.

Table of Contents

Dedication	ii
Acknowledgments	iii
Abstract	v
Chapter 1 Introduction	1
Chapter 2 Hierarchical Modeling of Gene Expressions for Multiple Laboratories and Replicates	5
2.1 Introduction	5
2.2 Likelihood for children with at least one parent for the DAG	6
2.2.1 Likelihood for multiple independent labs	12
2.3 Likelihood for multiple hierarchical labs as well as reps within labs ..	14
2.3.1 Likelihood for more general cases	23
Chapter 3 Prior Probability of Directed Acyclic Graph	31
3.1 Adjusted Equal Likelihood Approximation of Prior Probability of DAG	32
3.2 Typical Non-Edge Approximation of Prior Probability and Its Sensitivity	34
3.2.1 Approximation of Odds Ratio and Log Odds Ratio	35
3.2.2 Odds Ratio and Log Odds Ratio for Typical Non-Edge Approx-	
imation	36
3.3 Improved Typical Non-Edge Approximation of Prior Probability and	
Its Sensitivity	39
3.4 Example of A DAG with Four Nodes	44
3.5 Some Remarks of the Three Approximation Methods	48
Chapter 4 Conclusions	50
Bibliography	53
Curriculum Vitae	56

Abstract

Rigorous and expansive biological experiments of genes involve not only multiple replicates of sparse time data developed within a given laboratory but also, potentially, replicates generated by multiple laboratories. The posterior probability of a directed acyclic graph (DAG) of our models of gene associations given the hierarchical time-course data is proportional to the product of the prior probability of the DAG and the likelihood of the data given the DAG. From such data modeling, protein or gene interaction posterior probabilities are computed based on hierarchical structures. A result based on multiple replicates in a single laboratory is developed first, then extended it to replicates from multiple laboratories. Rather than assuming equal priors for DAGs, three methods to estimate the prior probabilities of DAGs are presented. Their sensitivity based on different assumptions and additional information are discussed here. At the same time, the odds ratio of two estimates under the same setting are calculated. The odds ratio often has a much more concise form and is easier to apply in practical computations than the priors themselves.

Chapter 1: Introduction

Modeling gene interactions based on time-course gene expression data has been a challenging and active research focus [1, 2, 3]. There are many approaches to the construction of gene interactions, and this paper outlines rigorous hierarchical Bayesian posterior probability cotemporal modeling [4, 5, 6, 7]. Our model building subunit is a Directed Acyclic Graph (DAG) [1, 8], which is a finite directed graph with no directed cycles, and where the nodes represent genes and the directed edges represent the interactions of the genes [1]. From a Bayesian's perspective [9, 10], the posterior probability of a DAG given the hierarchical time-course data, is proportional to the product of the prior probability of the DAG and the likelihood of the data given the DAG [9, 11], which are also the two main foci of this paper.

Time-course data sets of gene interaction are often collected in different laboratories by different researchers. There are many ways to combine the data from replicates and laboratories; often, however, the random differences that occur between data replications have a hierarchical structure, i.e., the parameters come from a higher level distribution. In practical modeling, the variation within the individual replicates, the variation from replicate to replicate within the individual lab and the variation from lab to lab combine to make a hierarchical structure [12, 13]. This multi-level data is most appropriate to give an unified and more comprehensive statistical approach [6, 7]. For the case where all the replicates are conducted under similar conditions in only one laboratory, and where the slope parameters of each of the replicates are assumed to come from a common normal distribution with slope mean of zero then a closed form solution has been given [6, 7] for the likelihood of a DAG utilizing Gaussian-based regression hierarchical likelihood and noninformative

uniform priors.

In Chapter 2, we first give a closed form solution for the likelihood of a DAG with a non-zero slope mean. We then show how to combine the data collected from different laboratories, with each laboratory generating several replicates. The slope parameters of each replicate in a laboratory are assumed to come from a common Normal distribution with a mean parameter related to the laboratory; also, the slope parameters related to laboratories come from a new common Normal distribution, i.e. adding an additional layer to the hierarchical structure of [6]. The closed form solution of the likelihood given a DAG is obtained by integrating over the within rep parameters, the rep to rep within lab parameters, and the lab to lab parameters in the Bayesian hierarchical structure. This yields a likelihood that is just conditional on the DAG.

Directed Acyclic Graphs (DAGs) are widely used in Bayesian networks and Bayesian statistical inferences [15, 16] since the definition of DAG, a finite directed graph with no directed cycles, is perfectly compatible with calculating probability, or likelihood from a Bayesian perspective. The acyclic structure of a DAG allows the factorization of joint likelihood of all children in a DAG into the product of each child's likelihood in the DAG. In our case of gene expression data [1, 16], a DAG represents the probabilistic interactions between genes. The genes are represented by nodes and their interactions are represented by directed edges. If all prior probabilities are equal, the likelihood of a given DAG is proportional to its relative posterior probability. Then, Bayesian averaging is used [19] with many intelligent searching algorithms such as Metropolis-Hastings algorithm [6, 7] and genetic algorithms [18]. To estimate the probability of a certain connection between two nodes, rigorous and fine-tuned genetic algorithms are used to search the DAG space to find DAGs with high relative likelihood. Then, Bayesian modeling averaging precisely combines all these results

to obtain posterior probabilities of edges. If all DAGs' priors are equal, the relative posterior probability of a specific interaction between two nodes is the ratio of the sum of likelihoods of all visited DAGs with the specific interaction and the sum of likelihoods of all visited DAGs. Unless all DAG priors are equal, since the posterior probability of a DAG given time-course data is proportional to the product of the prior of the DAG and the likelihood of the data. The estimation of the prior probability is indispensable in getting the posterior. Some results, [5, 6, 7, 18], have utilized noninformative uniform priors of DAGs; Chapter 3 gives several methods to estimate the nonuniform prior probabilities of DAGs. Chapter 3 first gives three methods to approximate and estimate the prior probabilities of DAGs and how the prior probabilities could change based on different assumptions and additional information. For the three approximation methods given in Chapter 3, the adjusted equal prior approximation methods requires more information about prior probabilities of specific pairs from specialists while the probabilities of the rest of the pairs are assumed to have a specific equal prior. Thus this version incorporates the empirical information of specialists to improve the prior probability while still maintaining relatively low computational complexity. Chapter 3 also shows that the typical non-edge approximation method improves the approximation of the prior probability by bringing in a typical probability for a non-edges to make the approximation more flexible and precise; this might require more computations but can give more accuracy. The improved typical non-edge approximation requires more information to adjust the probability of each pair without specific prior information and thus it has more flexibility and accuracy but it increases the computation inevitably. At the same time, Chapter 3 also shows that the odds ratio of two prior estimates under the same setting often has a much more concise form and is easier in practical computations than are the priors themselves, since a lot of common terms of the priors can be canceled in the

ratio. An example is given to illustrate the three approximation methods of the prior probability of a DAG.

Multiplying the relative likelihood we obtained in Chapter 2 and a relative prior probability developed in Chapter 3, yields the relative posterior probability which is then utilized along with Bayesian model averaging to obtain the estimated posterior probabilities of gene associations.

Chapter 2: Hierarchical Modeling of Gene Expressions for Multiple Laboratories and Replicates

2.1 Introduction

The fundamental structure used for our modeling is the directed acyclic graph (DAG). $P(DAG|Q)$ is the posterior probability of a DAG given the entire multi-level time-course data Q . By Bayes Theorem [9, 10, 11], for a given DAG we have

$$P(DAG|Q) = \frac{f(Q|DAG)P(DAG)}{f(Q)} \propto f(Q|DAG)P(DAG) \quad (2.1)$$

where $f(Q|DAG)$ is the likelihood of a DAG.

For a pair of nodes A and B in a DAG then if there is a directed edge from A to B , then A is the parent of B . The non-cyclic structure of a DAG allows the factorization of the joint likelihood of all children in a DAG into the product of each child's likelihood. For a specific ordering of the children given a DAG, if child k have child 1, \dots , child $k - 1$ as possible parents, we letting y_i represent the data of child i for all replicates, by Bayes Theorem,

$$\begin{aligned} f(Q|DAG) &= f(y_1, y_2, \dots, y_k|DAG) \\ &= f(y_1)f(y_2|y_1)f(y_3|y_1, y_2) \cdots f(y_k|y_1, y_2, \dots, y_{k-1}) \end{aligned} \quad (2.2)$$

If a child j is not a parent of child i , then child j does not influence the likelihood of child i . We then have

$$f(Q|DAG) = f(y_1|\text{parents of child 1}) \cdots f(y_k|\text{parents of child k}) \quad (2.3)$$

2.2 Likelihood for children with at least one parent for the DAG

Consider first a single laboratory. With the assumption of hierarchical structure of parameters of the replicates in a single laboratory, we now calculate the likelihood for children with at least one parent. In our calculation of likelihood, let $(y_{ic})_{t \times 1}$ be the data vector of any given child c from the i^{th} replicate and let t be the number of time points. Let $(x_{ic})_{t \times k_c}$ be the $t \times k_c$ matrix consisting of the data of the parents of child c , where k_c is the the number of parents of child c . Set the $t \times k_c$ matrix \bar{x}_c to be the average over all replicates of c 's parents' time-course data. Let $(\beta_{ic})_{k_c \times 1}$ be replicate i 's slope vector relating c 's relationship to its parents. Let σ_c^2 and β_c be the common within replicate error variance and the mean child slope vector of child c . For given β_{ic} and σ_c^2 , each y_{ic} is independent with the distribution

$$y_{ic} | \beta_{ic}, \sigma_c^2 \sim N_t(x_{ic}\beta_{ic}, \sigma_c^2 I) \quad (2.4)$$

Given σ_c^2 , each β_{ic} is independent with distribution

$$\beta_{ic} | \sigma_c^2 \sim N_{k_c}(\beta_c, g\sigma_c^2 B_c^{-1}) \quad (2.5)$$

where $B_c = \bar{x}_c^T \bar{x}_c$. Note that $B_c^T = B_c$. It is worth noting that the case that $\beta_c = 0$ was calculated in [6].

If $\tau_c = \frac{1}{\sigma_c^2}$, and we assume τ_c follows a gamma distribution,

$$\tau_c \sim \text{Gamma}\left(\frac{v_0}{2}, \frac{v_0 \sigma_0^2}{2}\right). \quad (2.6)$$

Then, by definition, the distribution of σ_c^2 is the inverse gamma distribution [10]

$$\sigma_c^2 \sim \text{Inverse-Gamma}\left(\frac{v_0}{2}, \frac{v_0 \sigma_0^2}{2}\right). \quad (2.7)$$

To find $f(Q|DAG)$, we first calculate $f(y_{ic}|\sigma_c^2)$ by

$$f(y_{ic}|\sigma_c^2) = \int_{\beta_{ic}} f(y_{ic}, \beta_{ic}|\sigma_c^2)d\beta_{ic} = \int_{\beta_{ic}} f(y_{ic}|\beta_{ic}, \sigma_c^2)f(\beta_{ic}|\sigma_c^2)d\beta_{ic} \quad (2.8)$$

Then we can compute $f(Q|DAG)$ by

$$f(y_c|DAG) = \int_{\sigma_c^2} f(y_{1c}|\sigma_c^2) \cdots f(y_{rc}|\sigma_c^2)f(\sigma_c^2)d\sigma_c^2 \quad (2.9)$$

$$= \int_{\tau_c} f(y_{1c}|\tau_c) \cdots f(y_{rc}|\tau_c)f(\tau_c)d\tau_c \quad (2.10)$$

The reason that we can calculate the likelihood by those two steps, which is the spirit of the calculations in this section, is

$$\begin{aligned} f(y_c|DAG) &= f(y_{1c}, \cdots, y_{rc}|DAG) \\ &= \int_{\sigma_c^2} \int_{\beta_{1c}} \cdots \int_{\beta_{rc}} f(y_{1c}, \cdots, y_{rc}, \beta_{1c}, \cdots, \beta_{rc}, \sigma_c^2)d\beta_{1c} \cdots d\beta_{rc}d\sigma_c^2 \\ &= \int_{\sigma_c^2} \int_{\beta_{1c}} \cdots \int_{\beta_{rc}} f(y_{1c}, \cdots, y_{rc}|\beta_{1c}, \cdots, \beta_{rc}, \sigma_c^2)f(\beta_{1c}, \cdots, \beta_{rc}|\sigma_c^2) \\ &\quad f(\sigma_c^2)d\beta_{1c} \cdots d\beta_{rc}d\sigma_c^2 \\ &= \int_{\sigma_c^2} \int_{\beta_{1c}} \cdots \int_{\beta_{rc}} \left[\prod_{i=1}^r f(y_{ic}|\beta_{ic}, \sigma_c^2)f(\beta_{ic}|\sigma_c^2) \right] f(\sigma_c^2)d\beta_{1c} \cdots d\beta_{rc}d\sigma_c^2 \\ &= \int_{\sigma_c^2} \left[\int_{\beta_{1c}} f(y_{1c}|\beta_{1c}, \sigma_c^2)f(\beta_{1c}|\sigma_c^2)d\beta_{rc} \cdots \int_{\beta_{rc}} f(y_{rc}|\beta_{rc}, \sigma_c^2)f(\beta_{rc}|\sigma_c^2)d\beta_{rc} \right] \\ &\quad f(\sigma_c^2)d\sigma_c^2 \\ &= \int_{\sigma_c^2} f(y_{1c}|\sigma_c^2) \cdots f(y_{rc}|\sigma_c^2)f(\sigma_c^2)d\sigma_c^2 \end{aligned} \quad (2.11)$$

In the calculation, because given β_{ic} and σ_c^2 , each y_{ic} is independent and given σ_c^2 ,

each β_{ic} is independent,

$$f(y_{1c}, \dots, y_{rc} | \beta_{1c}, \dots, \beta_{rc}, \sigma_c^2) f(\beta_{1c}, \dots, \beta_{rc} | \sigma_c^2) = \prod_{i=1}^r f(y_{ic} | \beta_{ic}, \sigma_c^2) f(\beta_{ic} | \sigma_c^2) \quad (2.12)$$

By the distributions we know in (2.4) and (2.5),

$$\begin{aligned} & f(y_{ic} | \beta_{ic}, \sigma_c^2) f(\beta_{ic} | \sigma_c^2) \\ &= |2\pi\sigma_c^2 I|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{ic} - x_{ic}\beta_{ic})^T (\sigma_c^2 I)^{-1} (y_{ic} - x_{ic}\beta_{ic})\right\} \cdot |2\pi g\sigma_c^2 B_c^{-1}|^{-\frac{1}{2}} \\ & \quad \exp\left\{-\frac{1}{2}(\beta_{ic} - \beta_c)^T \frac{B_c}{g\sigma_c^2} (\beta_{ic} - \beta_c)\right\} \quad (2.13) \\ &= |2\pi\sigma_c^2 I|^{-\frac{1}{2}} |2\pi g\sigma_c^2 B_c^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_c^2} [(y_{ic} - x_{ic}\beta_{ic})^T (y_{ic} - x_{ic}\beta_{ic}) + \right. \\ & \quad \left. (\beta_{ic} - \beta_c)^T \frac{B_c}{g} (\beta_{ic} - \beta_c)]\right\} \end{aligned}$$

Denote

$$D_{ic} = x_{ic}^T x_{ic} + \frac{B_c}{g} = x_{ic}^T x_{ic} + \frac{1}{g} \bar{x}_c^T \bar{x}_c \quad (2.14)$$

and we expand and collect the terms with β_{ic} to construct a probability density function of a Multivariate Normal distribution to integrate out β_{ic} . The bracket in the exponential part of (2.13) can be written as

$$\begin{aligned}
& (y_{ic} - x_{ic}\beta_{ic})^T(y_{ic} - x_{ic}\beta_{ic}) + (\beta_{ic} - \beta_c)^T \frac{B_c}{g} (\beta_{ic} - \beta_c) \\
&= y_{ic}^T y_{ic} - 2y_{ic}^T x_{ic} \beta_{ic} + \beta_{ic}^T x_{ic}^T x_{ic} \beta_{ic} + \beta_{ic}^T \frac{B_c}{g} \beta_{ic} - 2\beta_c^T \frac{B_c}{g} \beta_{ic} + \beta_c^T \frac{B_c}{g} \beta_c \\
&= \beta_{ic}^T (x_{ic}^T x_{ic} + \frac{B_c}{g}) \beta_{ic} - 2(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) \beta_{ic} + y_{ic}^T y_{ic} + \beta_c^T \frac{B_c}{g} \beta_c \\
&= \beta_{ic}^T D_{ic} \beta_{ic} - 2(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1} D_{ic} \beta_{ic} + y_{ic}^T y_{ic} + \beta_c^T \frac{B_c}{g} \beta_c \\
&= [\beta_{ic} - ((y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1})^T]^T D_{ic} [\beta_{ic} - ((y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1})^T] \\
&\quad + y_{ic}^T y_{ic} + \beta_c^T \frac{B_c}{g} \beta_c - (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1} (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g})^T
\end{aligned} \tag{2.15}$$

The probability density function of a Multivariate Normal distribution [10] with mean vector μ and variance-covariance matrix Σ is

$$|2\pi\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\} \tag{2.16}$$

Let $\mu_{ic} = ((y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1})^T$, then the probability density function of a Multivariate Normal distribution is given by

$$\int_{\beta_{ic}} \frac{1}{|2\pi\sigma_c^2 D_{ic}^{-1}|^{\frac{1}{2}}} \exp\{-\frac{1}{2\sigma_c^2} [\beta_{ic} - \mu_{ic}]^T D_{ic} [\beta_{ic} - \mu_{ic}]\} d\beta_{ic} = 1 \tag{2.17}$$

For a single replicate i , by (2.13) and (2.17), we have

$$\begin{aligned}
f(y_{ic}|\sigma_c^2) &= \int_{\beta_{ic}} f(y_{ic}, \beta_{ic}|\sigma_c^2) d\beta_{ic} \\
&= \int_{\beta_{ic}} f(y_{ic}|\beta_{ic}, \sigma_c^2) f(\beta_{ic}|\sigma_c^2) d\beta_{ic} \\
&= |2\pi\sigma_c^2 I|^{-\frac{1}{2}} |gB_c^{-1}|^{-\frac{1}{2}} |D_{ic}^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_c^2} \left[y_{ic}^T y_{ic} + \beta_c^T \frac{B_c}{g} \beta_c \right. \right. \\
&\quad \left. \left. - (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1} (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g})^T \right] \right\}
\end{aligned} \tag{2.18}$$

Applying (2.18) to all of the r replicates gives

$$\begin{aligned}
&f(y_{1c}|\sigma_c^2) \cdots f(y_{rc}|\sigma_c^2) \\
&= |2\pi\sigma_c^2 I|^{-\frac{r}{2}} |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_c^2} \left[\sum_{i=1}^r y_{ic}^T y_{ic} + r\beta_c^T \frac{B_c}{g} \beta_c - \right. \right. \\
&\quad \left. \left. \sum_{i=1}^r (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g}) D_{ic}^{-1} (y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g})^T \right] \right\}
\end{aligned} \tag{2.19}$$

Since $f(y_c|DAG) = \int_{\tau_c} f(y_{1c}|\tau_c) \cdots f(y_{rc}|\tau_c) f(\tau_c) d\tau_c$ where $\tau_c = \frac{1}{\sigma_c^2} \sim \Gamma(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2})$, the next step is integration over τ_c by constructing a probability density function of a gamma distribution $\Gamma(\alpha, \beta)$ for τ_c

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \tau_c^{\alpha-1} e^{-\beta\tau_c} \tag{2.20}$$

$$\begin{aligned}
& f(y_c|DAG) \\
&= \int_{\tau_c} |2\pi\sigma_c^2 I|^{-\frac{r}{2}} |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{\tau_c}{2} \left[\sum_{i=1}^r y_{ic}^T y_{ic} + r\beta_c^T \frac{B_c}{g} \beta_c - \right. \right. \\
&\quad \left. \left. \sum_{i=1}^r \left(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g} \right) D_{ic}^{-1} \left(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g} \right)^T \right] \right\} \frac{\left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}}}{\Gamma\left(\frac{v_0}{2}\right)} \tau_c^{\frac{v_0}{2}-1} \exp\left\{-\frac{v_0\sigma_0^2}{2} \tau_c\right\} d\tau_c \\
&= \int_{\tau_c} \frac{\lambda^{\frac{v_0+rt}{2}}}{\Gamma\left(\frac{v_0+rt}{2}\right)} \tau_c^{\frac{v_0+rt}{2}-1} \exp\{-\tau_c \lambda\} d\tau_c \frac{\Gamma\left(\frac{v_0+rt}{2}\right)}{\Gamma\left(\frac{v_0}{2}\right)} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr}{2}} \lambda^{-\frac{v_0+rt}{2}} \\
&\quad |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}} \\
&= \frac{\Gamma\left(\frac{v_0+rt}{2}\right)}{\Gamma\left(\frac{v_0}{2}\right)} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr}{2}} \lambda^{-\frac{v_0+rt}{2}} |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}}
\end{aligned} \tag{2.21}$$

where

$$\lambda = \frac{1}{2} \left[v_0\sigma_0^2 + \sum_{i=1}^r y_{ic}^T y_{ic} + r\beta_c^T \frac{B_c}{g} \beta_c - \sum_{i=1}^r \left(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g} \right) D_{ic}^{-1} \left(y_{ic}^T x_{ic} + \beta_c^T \frac{B_c}{g} \right)^T \right] \tag{2.22}$$

When we set $v_0 = 1, \sigma_0^2 = 1, g = 1$,

$$\lambda = \frac{1}{2} \left[1 + \sum_{i=1}^r y_{ic}^T y_{ic} + r\beta_c^T B_c \beta_c - \sum_{i=1}^r \left(y_{ic}^T x_{ic} + \beta_c^T B_c \right) D_{ic}^{-1} \left(y_{ic}^T x_{ic} + \beta_c^T B_c \right)^T \right] \tag{2.23}$$

and

$$f(y_c|DAG) = \frac{\Gamma\left(\frac{v_0+rt}{2}\right)}{\Gamma\left(\frac{v_0}{2}\right)} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr}{2}} \lambda^{-\frac{v_0+rt}{2}} |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}} \tag{2.24}$$

Now we consider the special case when children have no parents, [6] gives the contribution of a parentless child's for r replicates as

$$(2\pi)^{-\frac{rt}{2}} e^{-\frac{r(t-1)}{2}} \tag{2.25}$$

Then for a DAG with $m+n$ children with m of those children, y_1, y_2, \dots, y_m , having at least one parent and n children, $y_{m+1}, y_{m+2}, \dots, y_{m+n}$, having no parents, the relative likelihood given the DAG is

$$f(Q|DAG) \tag{2.26}$$

$$= f(y_1, y_2, \dots, y_m, y_{m+1}, y_{m+2}, \dots, y_{m+n}|DAG) \tag{2.27}$$

$$= f(y_1|DAG) \cdots f(y_m|DAG) f(y_{m+1}|DAG) \cdots f(y_{m+n}|DAG) \tag{2.28}$$

$$= \prod_{c=1}^m \left[\frac{\Gamma(\frac{v_0+rt}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr}{2}} \lambda^{-\frac{v_0+rt}{2}} |gB_c^{-1}|^{-\frac{r}{2}} \prod_{i=1}^r |D_{ic}^{-1}|^{-\frac{1}{2}} \right] \cdot ((2\pi)^{-\frac{rt}{2}} e^{-\frac{r(t-1)}{2}})^n \tag{2.29}$$

2.2.1 Likelihood for multiple independent labs

For multiple independent labs, in each lab j , we can compute $f(y_{jc}|DAG)$, for children with at least one parent in the laboratory, by our calculation of several replicates in one lab

$$f(y_{jc}|DAG) = \frac{\Gamma(\frac{v_0+rt}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr_j}{2}} \lambda_j^{-\frac{v_0+r_jt}{2}} |gB_{jc}^{-1}|^{-\frac{r_j}{2}} \prod_{i=1}^{r_j} |D_{jic}^{-1}|^{-\frac{1}{2}} \tag{2.30}$$

$$\lambda_j = \frac{1}{2} \left[v_0\sigma_0^2 + \sum_{i=1}^{r_j} y_{jic}^T y_{jic} + r\beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} - \sum_{i=1}^{r_j} \left(y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \right) D_{jic}^{-1} \left(y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \right)^T \right] \tag{2.31}$$

where $B_{jc} = \bar{x}_{jc}^T \bar{x}_{jc}$, $D_{jic} = x_{jic}^T x_{jic} + \frac{1}{g} \bar{x}_{jc}^T \bar{x}_{jc}$.

Assuming that all those labs are independent, we have

$$\begin{aligned}
& f(y_c|DAG) \\
&= f(y_{1c}, y_{1c}, \dots, y_{Jc}|DAG) \\
&= f(y_{1c}|DAG)f(y_{2c}|DAG) \cdots f(y_{Jc}|DAG) \tag{2.32} \\
&= \prod_{j=1}^J \left[\frac{\Gamma(\frac{v_0+r_j t}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr_j}{2}} \lambda_j^{-\frac{v_0+r_j t}{2}} |gB_{jc}^{-1}|^{-\frac{r_j}{2}} \prod_{i=1}^{r_j} |D_{jic}^{-1}|^{-\frac{1}{2}} \right]
\end{aligned}$$

For the case when children have no parents, $f(Q|DAG)$ can be factored into the product of the likelihoods of the individual children and the contribution of a parentless child's for each of the r_j replicates is

$$(2\pi)^{-\frac{r_j t}{2}} e^{-\frac{r_j(t-1)}{2}} \tag{2.33}$$

When there are J laboratories with replicates r_1, \dots, r_J respectively, we have

$$\prod_{j=1}^J (2\pi)^{-\frac{r_j t}{2}} e^{-\frac{r_j(t-1)}{2}} = (2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j} \tag{2.34}$$

Then for a DAG with $m+n$ children in which m of those children, y_1, y_2, \dots, y_m , have at least one parent and n children, $y_{m+1}, y_{m+2}, \dots, y_{m+n}$, have no parents, the relative likelihood given the DAG is

$$f(Q|DAG) \tag{2.35}$$

$$= f(y_1, y_2, \dots, y_m, y_{m+1}, y_{m+2}, \dots, y_{m+n}|DAG) \tag{2.36}$$

$$= f(y_1|DAG) \cdots f(y_m|DAG) f(y_{m+1}|DAG) \cdots f(y_{m+n}|DAG) \tag{2.37}$$

$$= \prod_{c=1}^m \left[\prod_{j=1}^J \left[\frac{\Gamma(\frac{v_0+r_j t}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{tr_j}{2}} \lambda_j^{-\frac{v_0+r_j t}{2}} |gB_{jc}^{-1}|^{-\frac{r_j}{2}} \prod_{i=1}^{r_j} |D_{jic}^{-1}|^{-\frac{1}{2}} \right] \right] \tag{2.38}$$

$$\cdot ((2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j})^n \tag{2.39}$$

From the calculation, we can see that to get the likelihood for multiple independent labs, we have to assume that the labs are independent which is valid if the laboratories are very different with little commonality. Often, though, there is commonality between laboratories in which case it is usually advisable to extend hierarchy by having the β'_{ic} s coming from a common distribution.

2.3 Likelihood for multiple hierarchical labs as well as reps within labs

We have calculated the likelihood for children with at least one parent with the assumption of hierarchical structure of parameters of the replicates in a single laboratory. Now we consider the case of hierarchical structure of parameters of the replicates in multiple laboratories and of parameters of laboratories. In our calculation of the likelihood, let $(y_{jic})_{t_j \times 1}$ be the data vector of any given child c from the i^{th} replicate of the j^{th} laboratory, and t_j is the number of time points in laboratory j . Let $(x_{jic})_{t_j \times k_c}$ be the $t_j \times k_c$ matrix consisting of the data of the parents of child c , where k_c is the the number of parents of child c . Let the $t_j \times k_c$ matrix \bar{x}_{jc} be the average over all replicates of c 's parents' time-course data for laboratory j . In a laboratory j , let $(\beta_{jic})_{k_c \times 1}$ be replicate i 's slope vector relating c 's relationship to its parents for replicate i within laboratory j . Furthermore for laboratory j , let σ_{jc}^2 and β_{jc} be the within replicate error variance and mean slope vector for parentd of child c . Given β_{jic}, β_{jc} and σ_{jc}^2 , each y_{jic} is independent with distribution

$$y_{jic} | \beta_{jic}, \sigma_{jc}^2, \beta_{jc} \sim N_t(x_{jic}\beta_{jic}, \sigma_{jc}^2 I) \quad (2.40)$$

Given σ_{jc}^2 and β_{jc} , each β_{jic} is independent with distribution

$$\beta_{jic} | \sigma_{jc}^2, \beta_{jc} \sim N_{k_c}(\beta_{jc}, g\sigma_{jc}^2 B_{jc}^{-1}) \quad (2.41)$$

where $B_{jc} = \bar{x}_{jc}^T \bar{x}_{jc}$. Note that $B_{jc}^T = B_{jc}$.

When only given σ_{jc}^2 , each β_{jc} is independent with distribution

$$\beta_{jc}|\sigma_{jc}^2 \sim N_{k_c}(\beta_c, g\sigma_{jc}^2 B_c^{-1}) \quad (2.42)$$

where $B_c = \bar{x}_c^T \bar{x}_c$. By calculating B_c^T , we also have $B_c^T = B_c$. But for simplicity, we first calculate the special case when $\beta_c = 0$, i.e., the distribution of β_{jc} given σ_{jc}^2 in this section is

$$\beta_{jc}|\sigma_{jc}^2 \sim N_{k_c}(0, g\sigma_{jc}^2 B_c^{-1}); \quad (2.43)$$

We will then show how to use a similar procedure to compute the result for general β 's.

If we denote $\tau_{jc} = \frac{1}{\sigma_{jc}^2}$, and we assume τ_{jc} follows a gamma distribution

$$\tau_{jc} \sim \text{Gamma}\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right) \quad (2.44)$$

By the definition of inverse gamma function [10], we have the prior distribution of σ_{jc}^2

$$\sigma_{jc}^2 \sim \text{Inverse-Gamma}\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right) \quad (2.45)$$

Since we have

$$f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) = \int_{\beta_{jic}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic})f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc})d\beta_{jic} \quad (2.46)$$

to integrate over β_{jic} , we need to find $f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic})f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc})$. By (2.40) and (2.41), we have

$$\begin{aligned}
& f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic})f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc}) \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_{jic} - x_{jic}\beta_{jic})^T(\sigma_{jc}^2 I)^{-1}(y_{jic} - x_{jic}\beta_{jic})\right\} \cdot |2\pi g\sigma_{jc}^2 B_{jc}^{-1}|^{-\frac{1}{2}} \\
& \quad \exp\left\{-\frac{1}{2}(\beta_{jic} - \beta_{jc})^T \frac{B_{jc}}{g\sigma_{jc}^2}(\beta_{jic} - \beta_{jc})\right\} \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |2\pi g\sigma_{jc}^2 B_{jc}^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2}[(y_{jic} - x_{jic}\beta_{jic})^T(y_{jic} - x_{jic}\beta_{jic}) + \right. \\
& \quad \left. (\beta_{jic} - \beta_{jc})^T \frac{B_{jc}}{g}(\beta_{jic} - \beta_{jc})]\right\}
\end{aligned} \tag{2.47}$$

furthermore, by expanding the terms and collecting the terms with β_{jic} , the bracket in the exponential part of (2.47) can be written as

$$\begin{aligned}
& (y_{jic} - x_{jic}\beta_{jic})^T(y_{jic} - x_{jic}\beta_{jic}) + (\beta_{jic} - \beta_{jc})^T \frac{B_{jc}}{g}(\beta_{jic} - \beta_{jc}) \\
&= y_{jic}^T y_{jic} - 2y_{jic}^T x_{jic}\beta_{jic} + \beta_{jic}^T x_{jic}^T x_{jic}\beta_{jic} + \beta_{jic}^T \frac{B_{jc}}{g}\beta_{jic} - 2\beta_{jc}^T \frac{B_{jc}}{g}\beta_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}\beta_{jc} \\
&= \beta_{jic}^T (x_{jic}^T x_{jic} + \frac{B_{jc}}{g})\beta_{jic} - 2(y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})\beta_{jic} + y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}\beta_{jc} \\
&= \beta_{jic}^T D_{jic}\beta_{jic} - 2(y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})D_{jic}^{-1}D_{jic}\beta_{jic} + y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}\beta_{jc} \\
&= [\beta_{jic} - ((y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})D_{jic}^{-1})^T]^T D_{jic} [\beta_{jic} - ((y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})D_{jic}^{-1})^T] \\
& \quad + y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}\beta_{jc} - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})D_{jic}^{-1}(y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T
\end{aligned} \tag{2.48}$$

where

$$D_{jic} = x_{jic}^T x_{jic} + \frac{B_{jc}}{g} = x_{jic}^T x_{jic} + \frac{1}{g} \bar{x}_{jc}^T \bar{x}_{jc} \tag{2.49}$$

By denoting $\mu_{jic} = ((y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1})^T$, we have the probability density function of a Multivariate Normal distribution for parameter β_{jic}

$$\int_{\beta_{jic}} \frac{1}{|2\pi\sigma_{jc}^2 D_{jic}^{-1}|^{\frac{1}{2}}} \exp\{-\frac{1}{2\sigma_{jc}^2} [\beta_{jic} - \mu_{jic}]^T D_{jic} [\beta_{jic} - \mu_{jic}]\} d\beta_{jic} = 1 \quad (2.50)$$

Then we can get $f(y_{jic}|\sigma_{jc}^2, \beta_{jc})$ by (2.47), (2.48) and (2.50),

$$\begin{aligned} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) &= \int_{\beta_{jic}} f(y_{jic}, \beta_{jic}|\sigma_{jc}^2, \beta_{jc}) d\beta_{jic} \\ &= \int_{\beta_{jic}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic}) f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc}) d\beta_{jic} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \\ &\quad - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T]\} \end{aligned} \quad (2.51)$$

Our next step would be get $f(y_{jic}|\sigma_{jc}^2)$ by integrating out β_{jc}

$$\begin{aligned} f(y_{jic}|\sigma_{jc}^2) &= \int_{\beta_{jc}} f(y_{jic}, \beta_{jc}|\sigma_{jc}^2) d\beta_{jc} \\ &= \int_{\beta_{jc}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) f(\beta_{jc}|\sigma_{jc}^2) d\beta_{jc} \end{aligned} \quad (2.52)$$

Since $\beta_{jc}|\sigma_{jc}^2 \sim N_{k_c}(\beta, g\sigma_{jc}^2 B_c^{-1})$, we have

$$\begin{aligned}
& f(y_{jic}|\sigma_{jc}^2, \beta_{jc})f(\beta_{jc}|\sigma_{jc}^2) \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \right. \\
&\quad \left. - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T]\right\} \cdot |2\pi g\sigma_{jc}^2 B_c^{-1}|^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2} \beta_{jc}^T \frac{B_c}{g\sigma_{jc}^2} \beta_{jc}\right\} \tag{2.53} \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |2\pi g\sigma_{jc}^2 B_c^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \right. \\
&\quad \left. - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T + \beta_{jc}^T \frac{B_c}{g} \beta_{jc}]\right\}
\end{aligned}$$

Furthermore, by expanding and collecting the terms containing β_{jc} , the bracket in the exponential part of equation (2.53) can be written as

$$\begin{aligned}
& y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T \\
& + \beta_{jc}^T \frac{B_c}{g} \beta_{jc} \\
& = y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - 2 y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} \\
& - \beta_{jc}^T \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} + \beta_{jc}^T \frac{B_c}{g} \beta_{jc} \\
& = \beta_{jc}^T \left(\frac{B_{jc}}{g} \right) \beta_{jc} - \beta_{jc}^T \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} + \beta_{jc}^T \frac{B_c}{g} \beta_{jc} - 2 y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} \\
& + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \\
& = \beta_{jc}^T \left(\frac{B_{jc}}{g} + \frac{B_c}{g} - \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \right) \beta_{jc} - 2 (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) \beta_{jc} \\
& + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \\
& = \left[\beta_{jc} - \left((y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} \right)^T \right]^T H_{jic} \left[\beta_{jc} - \left((y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} \right)^T \right] \\
& + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g})^T
\end{aligned} \tag{2.54}$$

where

$$H_{jic} = \frac{B_{jc}}{g} + \frac{B_c}{g} - \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \tag{2.55}$$

Note that the probability density function of a Multivariate Normal distribution [10] with mean vector μ and variance-covariance matrix Σ is

$$|2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right\} \tag{2.56}$$

Let $\mu_{jc} = \left((y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} \right)^T$. We have the probability density function of a

Multivariate Normal distribution

$$\int_{\beta_{jc}} \frac{1}{|2\pi\sigma_{jc}^2 H_{jic}^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [\beta_{jc} - \mu_{jc}]^T H_{jic} [\beta_{jc} - \mu_{jc}]\right\} d\beta_{jc} = 1 \quad (2.57)$$

and thus

$$\begin{aligned} & f(y_{jic}|\sigma_{jc}^2) \\ &= \int_{\beta_{jc}} f(y_{jic}, \beta_{jc}|\sigma_{jc}^2) d\beta_{jc} \\ &= \int_{\beta_{jc}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) f(\beta_{jc}|\sigma_{jc}^2) d\beta_{jc} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |gB_c^{-1}|^{-\frac{1}{2}} |H_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} \right. \\ &\quad \left. - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g})^T]\right\} \end{aligned} \quad (2.58)$$

Applying the result to all the replicates in one laboratory gives

$$\begin{aligned} & \prod_{i=1}^I f(y_{jic}|\sigma_{jc}^2) \\ &= \prod_{i=1}^I |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |gB_c^{-1}|^{-\frac{1}{2}} |H_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} \right. \\ &\quad \left. - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g})^T]\right\} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{I}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \exp\left\{-\frac{1}{2\sigma_{jc}^2} \left[\sum_{i=1}^I y_{jic}^T y_{jic} \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g})^T\right]\right\} \end{aligned} \quad (2.59)$$

Since

$$f(y_{jc}|DAG) = \int_{\tau_{jc}} f(y_{j1c}|\tau_{jc}) \cdots f(y_{jIc}|\tau_{jc}) f(\tau_{jc}) d\tau_{jc} \quad (2.60)$$

where $\tau_{jc} = \frac{1}{\sigma_c^2} \sim \Gamma(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2})$, the next step is integration over τ_{jc} by constructing a probability density function of a gamma distribution $\Gamma(\alpha, \beta)$ for τ_{jc}

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \tau_{jc}^{\alpha-1} e^{-\beta\tau_{jc}} \quad (2.61)$$

By (2.59) and (2.60),

$$\begin{aligned} & f(y_{jc}|DAG) \\ &= \int_{\tau_{jc}} f(y_{j1c}|\tau_{jc}) \cdots f(y_{jIc}|\tau_{jc}) f(\tau_{jc}) d\tau_{jc} \\ &= \int_{\tau_{jc}} |2\pi\sigma_{jc}^2 I|^{-\frac{I}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \exp\left\{-\frac{\tau_{jc}}{2} \left[\sum_{i=1}^I y_{jic}^T y_{jic}\right.\right. \\ &\quad \left.- \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g})^T\right. \\ &\quad \left.- \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic}\right\} \cdot \frac{(\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}}}{\Gamma(\frac{v_0}{2})} \tau_{jc}^{\frac{v_0}{2}-1} \exp\left\{-\frac{v_0\sigma_0^2}{2} \tau_{jc}\right\} d\tau_{jc} \\ &= \int_{\tau_{jc}} (2\pi)^{-\frac{It}{2}} \tau_{jc}^{\frac{It}{2}} \exp\left\{-\frac{\tau_{jc}}{2} [v_0\sigma_0^2 + \sum_{i=1}^I y_{jic}^T y_{jic} - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic}\right. \\ &\quad \left.- \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g})^T]\right\} \tau_{jc}^{\frac{v_0}{2}-1} d\tau_{jc} \\ &\quad \frac{(\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}}}{\Gamma(\frac{v_0}{2})} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \\ &= \int_{\tau_{jc}} \frac{\lambda_j^{\frac{v_0+It}{2}}}{\Gamma(\frac{v_0+It}{2})} \tau_{jc}^{\frac{v_0+It}{2}-1} \exp\{-\tau_{jc}\lambda_j\} d\tau_{jc} \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} \\ &\quad |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \quad (2.62) \\ &= \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0\sigma_0^2}{2}\right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \end{aligned}$$

where

$$\begin{aligned} \lambda_j = & v_0 \sigma_0^2 + \sum_{i=1}^I y_{jic}^T y_{jic} - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \\ & - \sum_{i=1}^I \left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g} \right) H_{jic}^{-1} \left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jic}}{g} \right)^T \end{aligned} \quad (2.63)$$

Then we have $f(y_{jc}|DAG)$ for a single lab j

$$\begin{aligned} & f(y_{jc}|DAG) \\ &= \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2} \right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \end{aligned} \quad (2.64)$$

and thus

$$\begin{aligned} & f(y_c|DAG) \\ &= f(y_{1c}|DAG) f(y_{2c}|DAG) \cdots f(y_{Jc}|DAG) \\ &= \prod_{j=1}^J \left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2} \right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \right) \\ &= \left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2} \right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} |gB_c^{-1}|^{-\frac{I}{2}} \right)^J \left(\prod_{j=1}^J \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \right) \prod_{j=1}^J \prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \end{aligned} \quad (2.65)$$

For the case when children have no parents, $f(Q|DAG)$ can be factored into the product of the likelihoods of the individual children and the contribution of a parentless child's for r_i replicates is

$$(2\pi)^{-\frac{r_i t}{2}} e^{-\frac{r_i(t-1)}{2}} \quad (2.66)$$

When there are J labs with reps r_1, \dots, r_J respectively, we have

$$\prod_{j=1}^J (2\pi)^{-\frac{r_j t}{2}} e^{-\frac{r_j(t-1)}{2}} = (2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j} \quad (2.67)$$

Then for a DAG with $m + n$ children and with m of those children, y_1, y_2, \dots, y_m , having at least one parent and n children, $y_{m+1}, y_{m+2}, \dots, y_{m+n}$, having no parents, the relative likelihood given the DAG is

$$\begin{aligned}
& f(Q|DAG) \\
&= f(y_1, y_2, \dots, y_m, y_{m+1}, y_{m+2}, \dots, y_{m+n}|DAG) \\
&= f(y_1|DAG) \cdots f(y_m|DAG) f(y_{m+1}|DAG) \cdots f(y_{m+n}|DAG) \\
&= \prod_{c=1}^m \left[\left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \right) \left(\frac{v_0\sigma_0^2}{2} \right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} |gB_c^{-1}|^{-\frac{I}{2}} \right]^J \left(\prod_{j=1}^J \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \right) \prod_{j=1}^J \prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \\
& \cdot ((2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j})^n
\end{aligned} \tag{2.68}$$

2.3.1 Likelihood for more general cases

In the previous calculations, we assumed each β_{jc} given σ_{jc}^2 is independent in a distribution

$$\beta_{jc} | \sigma_{jc}^2 \sim N_{k_c}(0, g\sigma_{jc}^2 B_c^{-1}) \tag{2.69}$$

where $B_c = \bar{x}_c^T \bar{x}_c$. We now calculate the likelihood for a general β while keeping the distributions of other parameters the same:

$$\beta_{jc} | \sigma_{jc}^2 \sim N_{k_c}(\beta_c, g\sigma_{jc}^2 B_c^{-1}) \tag{2.70}$$

$$y_{jic} | \beta_{jic}, \sigma_{jc}^2, \beta_j \sim N_t(x_{jic} \beta_{jic}, \sigma_{jc}^2 I) \tag{2.71}$$

$$\beta_{jic} | \beta_{jc}, \sigma_{jc}^2 \sim N_{k_c}(\beta_{jc}, g\sigma_{jc}^2 B_{jc}^{-1}) \tag{2.72}$$

If we denote $\tau_{jc} = \frac{1}{\sigma_{jc}^2}$, and we assume τ_{jc} follows a gamma distribution

$$\tau_{jc} \sim \text{Gamma}\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right) \tag{2.73}$$

then we know the distribution of σ_{jc}^2 by the definition of inverse gamma distribution,

$$\sigma_{jc}^2 \sim \text{Inverse-Gamma}\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right) \quad (2.74)$$

Then we can compute $f(y_{jic}|\sigma_{jc}^2, \beta_{jc})$ by integrating over β_{jic} , since the distributions of $f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic})$ and $f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc})$ are given in (2.71) and (2.72), thus the distribution of $f(y_{jic}|\sigma_{jc}^2, \beta_{jc})$ can be computed by the same procedure as (2.47), (2.28), (2.50), and (2.51),

$$\begin{aligned} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) &= \int_{\beta_{jic}} f(y_{jic}, \beta_{jic}|\sigma_{jc}^2, \beta_{jc}) d\beta_{jic} \\ &= \int_{\beta_{jic}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}, \beta_{jic}) f(\beta_{jic}|\sigma_{jc}^2, \beta_{jc}) d\beta_{jic} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \right. \\ &\quad \left. - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T]\right\} \end{aligned} \quad (2.75)$$

Our next step would be getting $f(y_{jic}|\sigma_{jc}^2)$ by integrating over β_{jc} as (2.76), but now as $\beta_{jc}|\beta_c, \sigma_{jc}^2 \sim N_{k_c}(\beta_c, g\sigma_{jc}^2 B_c^{-1})$, we assume $\beta_c \neq 0$

$$\begin{aligned} f(y_{jic}|\sigma_{jc}^2) &= \int_{\beta_{jc}} f(y_{jic}, \beta_{jc}|\sigma_{jc}^2) d\beta_{jc} \\ &= \int_{\beta_{jc}} f(y_{jic}|\sigma_{jc}^2, \beta_{jc}) f(\beta_{jc}|\beta_c, \sigma_{jc}^2) d\beta_{jc} \end{aligned} \quad (2.76)$$

Since $\beta_{jc}|\sigma_{jc}^2 \sim N_{k_c}(\beta_c, g\sigma_{jc}^2 B_c^{-1})$, we have

$$\begin{aligned}
& f(y_{jic}|\sigma_{jc}^2, \beta_{jc})f(\beta_{jc}|\beta_c, \sigma_{jc}^2) \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \right. \\
&\quad \left. - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T]\right\} \cdot |2\pi g\sigma_{jc}^2 B_c^{-1}|^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2} (\beta_{jc} - \beta_c)^T \frac{B_c}{g\sigma_{jc}^2} (\beta_{jc} - \beta_c)\right\} \tag{2.77} \\
&= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |2\pi g\sigma_{jc}^2 B_c^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} \right. \\
&\quad \left. - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T + (\beta_{jc} - \beta_c)^T \frac{B_c}{g} (\beta_{jc} - \beta_c)]\right\}
\end{aligned}$$

By again denoting

$$H_{jic} = \frac{B_{jc}}{g} + \frac{B_c}{g} - \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \tag{2.78}$$

we expand and collect the terms containing β_{jc} to construct a probability density function of a Multivariate Normal distribution, the bracket in the exponential part of (2.77) can be written as

$$\begin{aligned}
& y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} - (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g}) D_{jic}^{-1} (y_{jic}^T x_{jic} + \beta_{jc}^T \frac{B_{jc}}{g})^T \\
& + (\beta_{jc} - \beta_c)^T \frac{B_c}{g} (\beta_{jc} - \beta_c) \\
& = y_{jic}^T y_{jic} + \beta_{jc}^T \frac{B_{jc}}{g} \beta_{jc} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - 2 y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} \\
& - \beta_{jc}^T \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} + \beta_{jc}^T \frac{B_c}{g} \beta_{jc} - 2 \beta_c^T \frac{B_c}{g} \beta_{jc} + \beta_c^T \frac{B_c}{g} \beta_c \\
& = \beta_{jc}^T \left(\frac{B_{jc}}{g} \right) \beta_{jc} - \beta_{jc}^T \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} + \beta_{jc}^T \frac{B_c}{g} \beta_{jc} - 2 y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} \beta_{jc} \\
& - 2 \beta_c^T \frac{B_c}{g} \beta_{jc} + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} + \beta_c^T \frac{B_c}{g} \beta_c \\
& = \beta_{jc}^T \left(\frac{B_{jc}}{g} + \frac{B_c}{g} - \frac{B_{jc}}{g} D_{jic}^{-1} \frac{B_{jc}}{g} \right) \beta_{jc} - 2 \left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right) \beta_{jc} \\
& + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} + \beta_c^T \frac{B_c}{g} \beta_c \\
& = \left[\beta_{jc} - \left(\left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right) H_{jic}^{-1} \right)^T \right]^T H_{jic} \left[\beta_{jc} - \left(\left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right) H_{jic}^{-1} \right)^T \right] \\
& + y_{jic}^T y_{jic} - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} + \beta_c^T \frac{B_c}{g} \beta_c \\
& - \left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right) H_{jic}^{-1} \left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right)^T
\end{aligned} \tag{2.79}$$

The probability density function of a Multivariate Normal distribution [10] with mean vector μ and variance-covariance matrix Σ is

$$|2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right\} \tag{2.80}$$

By denoting $\mu_{jic} = \left(\left(y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g} \right) H_{jic}^{-1} \right)^T$, we have the probability density

function of a Multivariate Normal distribution

$$\int_{\beta_{jc}} \frac{1}{|2\pi\sigma_{jc}^2 H_{jic}^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [\beta_{jc} - \mu_{jic}]^T H_{jic} [\beta_{jc} - \mu_{jic}]\right\} d\beta_{jc} = 1 \quad (2.81)$$

$$\begin{aligned} & f(y_{jic} | \sigma_{jc}^2) \\ &= \int_{\beta_{jc}} f(y_{jic}, \beta_{jc} | \sigma_{jc}^2) d\beta_{jc} \\ &= \int_{\beta_{jc}} f(y_{jic} | \sigma_{jc}^2, \beta_{jc}) f(\beta_{jc} | \sigma_{jc}^2) d\beta_{jc} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |gB_c^{-1}|^{-\frac{1}{2}} |H_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_c^T \frac{B_c}{g} \beta_c \right. \\ &\quad \left. - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g})^T]\right\} \\ & \quad (2.82) \end{aligned}$$

Applying (2.82) to all the replicates in one laboratory gives

$$\begin{aligned} & \prod_{i=1}^I f(y_{jic} | \sigma_{jc}^2) \\ &= \prod_{i=1}^I |2\pi\sigma_{jc}^2 I|^{-\frac{1}{2}} |gB_{jc}^{-1}|^{-\frac{1}{2}} |D_{jic}^{-1}|^{\frac{1}{2}} |gB_c^{-1}|^{-\frac{1}{2}} |H_{jic}^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{jc}^2} [y_{jic}^T y_{jic} + \beta_c^T \frac{B_c}{g} \beta_c \right. \\ &\quad \left. - y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} - (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g})^T]\right\} \\ &= |2\pi\sigma_{jc}^2 I|^{-\frac{I}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}\right) \exp\left\{-\frac{1}{2\sigma_{jc}^2} \left[\sum_{i=1}^I y_{jic}^T y_{jic} \right. \right. \\ &\quad \left. \left. + I\beta_c^T \frac{B_c}{g} \beta_c - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g})^T\right]\right\} \\ & \quad (2.83) \end{aligned}$$

Since

$$f(y_{jc}|DAG) = \int_{\tau_{jc}} f(y_{j1c}|\tau_{jc}) \cdots f(y_{jIc}|\tau_{jc}) f(\tau_{jc}) d\tau_{jc} \quad (2.84)$$

where $\tau_{jc} = \frac{1}{\sigma_{jc}^2} \sim \Gamma(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2})$, the next step is integration over τ_{jc} by constructing a probability density function of a gamma distribution $\Gamma(\alpha, \beta)$ for τ_{jc}

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \tau_{jc}^{\alpha-1} e^{-\beta\tau_{jc}} \quad (2.85)$$

By (2.73), (2.83) and (2.84)

$$\begin{aligned} & f(y_{jc}|DAG) \\ &= \int_{\tau_{jc}} f(y_{j1c}|\tau_{jc}) \cdots f(y_{jIc}|\tau_{jc}) f(\tau_{jc}) d\tau_{jc} \\ &= \int_{\tau_{jc}} |2\pi\sigma_{jc}^2 I|^{-\frac{I}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \exp\left\{-\frac{\tau_{jc}}{2} \left[\sum_{i=1}^I y_{jic}^T y_{jic} \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g})^T + I \beta_c^T \frac{B_c}{g} \beta_c \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \right] \right\} \cdot \frac{(\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}}}{\Gamma(\frac{v_0}{2})} \tau_{jc}^{\frac{v_0}{2}-1} \exp\left\{-\frac{v_0\sigma_0^2}{2} \tau_{jc}\right\} d\tau_{jc} \\ &= \int_{\tau_{jc}} (2\pi)^{-\frac{I}{2}} \tau_{jc}^{\frac{I}{2}} \exp\left\{-\frac{\tau_{jc}}{2} \left[v_0\sigma_0^2 + \sum_{i=1}^I y_{jic}^T y_{jic} - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta_c^T \frac{B_c}{g})^T + I \beta_c^T \frac{B_c}{g} \beta_c \right] \right\} \\ &\quad \tau_{jc}^{\frac{v_0}{2}-1} d\tau_{jc} \cdot \left[\frac{(\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}}}{\Gamma(\frac{v_0}{2})} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \int_{\tau_{jc}} \frac{\lambda_j^{\frac{v_0+It}{2}}}{\Gamma(\frac{v_0+It}{2})} \tau_{jc}^{\frac{v_0+It}{2}-1} \exp\{-\tau_{jc}\lambda_j\} d\tau_{jc} \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} (\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} \\
&\quad |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \\
&= \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} (\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_j^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right)
\end{aligned} \tag{2.86}$$

where

$$\begin{aligned}
\lambda_{jc} &= v_0\sigma_0^2 + \sum_{i=1}^I y_{jic}^T y_{jic} + I\beta^T \frac{B_c}{g} \beta - \sum_{i=1}^I y_{jic}^T x_{jic} D_{jic}^{-1} x_{jic}^T y_{jic} \\
&\quad - \sum_{i=1}^I (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta^T \frac{B_c}{g}) H_{jic}^{-1} (y_{jic}^T x_{jic} D_{jic}^{-1} \frac{B_{jc}}{g} + \beta^T \frac{B_c}{g})^T
\end{aligned} \tag{2.87}$$

Then we have $f(y_{jc}|DAG)$ for a single lab j

$$\begin{aligned}
&f(y_{jc}|DAG) \\
&= \frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} (\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_{jc}^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right)
\end{aligned} \tag{2.88}$$

Then for all J laboratories, we have

$$\begin{aligned}
&f(y_c|DAG) \\
&= f(y_{1c}|DAG) f(y_{2c}|DAG) \cdots f(y_{Jc}|DAG) \\
&= \prod_{j=1}^J \left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} (\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} \lambda_{jc}^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |gB_c^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \left(\prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right) \right) \\
&= \left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} (\frac{v_0\sigma_0^2}{2})^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} |gB_c^{-1}|^{-\frac{I}{2}} \right)^J \left(\prod_{j=1}^J \lambda_{jc}^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \right) \prod_{j=1}^J \prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}}
\end{aligned} \tag{2.89}$$

For the case when children have no parents, $f(Q|DAG)$ can be factored into the product of the likelihoods of the individual children and the contribution of a

parentless child's for r_i replicates is

$$(2\pi)^{-\frac{r_i t}{2}} e^{-\frac{r_i(t-1)}{2}} \quad (2.90)$$

When there are J labs with reps r_1, \dots, r_J respectively, we have

$$\prod_{j=1}^J (2\pi)^{-\frac{r_j t}{2}} e^{-\frac{r_j(t-1)}{2}} = (2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j} \quad (2.91)$$

Then for a DAG with $m+n$ children with m of those children, y_1, y_2, \dots, y_m , having at least one parent and n children, $y_{m+1}, y_{m+2}, \dots, y_{m+n}$, having no parents, the relative likelihood given the DAG is

$$\begin{aligned} & f(Q|DAG) \\ &= f(y_1, y_2, \dots, y_m, y_{m+1}, y_{m+2}, \dots, y_{m+n}|DAG) \\ &= f(y_1|DAG) \cdots f(y_m|DAG) f(y_{m+1}|DAG) \cdots f(y_{m+n}|DAG) \\ &= \prod_{c=1}^m \left[\left(\frac{\Gamma(\frac{v_0+It}{2})}{\Gamma(\frac{v_0}{2})} \left(\frac{v_0 \sigma_0^2}{2} \right)^{\frac{v_0}{2}} (2\pi)^{-\frac{It}{2}} |gB_c^{-1}|^{-\frac{I}{2}} \right)^J \left(\prod_{j=1}^J \lambda_{jc}^{-\frac{v_0+It}{2}} |gB_{jc}^{-1}|^{-\frac{I}{2}} |H_{jic}^{-1}|^{\frac{I}{2}} \right) \prod_{j=1}^J \prod_{i=1}^I |D_{jic}^{-1}|^{\frac{1}{2}} \right] \\ & \cdot \left((2\pi)^{-\frac{t}{2} \sum_{j=1}^J r_j} e^{-\frac{(t-1)}{2} \sum_{j=1}^J r_j} \right)^n \end{aligned} \quad (2.92)$$

Chapter 3: Prior Probability of Directed Acyclic Graph

Recall that we have discussed in Chapter 2, in the hierarchical modeling of protein interaction or gene expression, that the model is represented by a Directed Acyclic Graph (DAG). A DAG is a graph consisting of nodes and directed edges without cycles, where the nodes may represent genes and a directed edge between two nodes would represent an association between that pair of proteins or genes. Let $P(DAG|Q)$ denote the posterior probability of a DAG given the full multi-level time-course data Q . By Bayes Theorem, the posterior probability of a given DAG is

$$P(DAG|Q) = \frac{f(Q|DAG)P(DAG)}{f(Q)} \propto f(Q|DAG)P(DAG) \quad (3.1)$$

To compute the relative posterior probability of a DAG, we need to know the likelihood of a DAG which is denoted as $f(Q|DAG)$, and, if a uniform prior is not assumed, the prior probability of a DAG which is denoted as $P(DAG)$. The computation of $f(Q|DAG)$ has been discussed in the previous chapter and here we focus on the prior probability $P(DAG)$.

For the prior probability of a DAG, the requirement of acyclicity greatly impedes the formation of useful informative prior probabilities [8, 10]. To simplify the situation, we reduce a DAG in the models to directed graph to give approximations to the prior probability of a given Directed Acyclic Graph, which means we only consider Directed but not necessarily acyclic graphs in our discussion in the approximations to prior probabilities for DAGs.

For any given pair of nodes A and B, three possible states must be either: an edge from A to B ($A \rightarrow B$), an edge from B to A ($A \leftarrow B$), or no edge between them ($A \cdots B$). In this paper, three approximation methods are given to model the prior

probability of a DAG based on one fundamental and obvious assumption

$$P(A \longrightarrow B) + P(A \longleftarrow B) + P(A \cdots B) = 1 \quad (3.2)$$

3.1 Adjusted Equal Likelihood Approximation of Prior Probability of DAG

In this section we first examine the assumption that the prior probability of all those three edge/non-edge states are equally likely, in which case

$$P(A \longrightarrow B) = P(A \longleftarrow B) = P(A \cdots B) = \frac{1}{3} \quad (3.3)$$

Applying this equally likely assumption to all pairs of nodes in the graph and with the simplifying assumption of independence between edges, for every DAG we have the following approximated prior probability of a given graph with $l = \binom{N}{2}$ pairs where N is the number of nodes in a graph,

$$P_0(DAG) = \frac{1}{3} \cdots \frac{1}{3} = \left(\frac{1}{3}\right)^l \quad (3.4)$$

It is worth noting that even for a moderate number, N , of genes, the number of pairs $\binom{N}{2}$ can still be large. This is the setting where all DAGs have the same prior probability, i.e. equally likely prior DAGs resulting in

$$P(DAG|Q) \propto f(Q|DAG)P(DAG) \propto f(Q|DAG) \quad (3.5)$$

for a given DAG.

If instead, suppose a DAG has exactly k pairs of genes whose prior probability for the three cases, \longrightarrow , \longleftarrow , and \cdots , are not $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$, then for those k pairs of

genes. Further let p_1, p_2, \dots, p_k denote their known prior probabilities for those k pairs states in the DAG, we have a better approximations of the prior probability of the graph that differs from $(\frac{1}{3})^l$. Specifically, it is

$$P_1(DAG) = p_1 p_2 \cdots p_k \cdot \frac{1}{3} \cdots \frac{1}{3} = p_1 p_2 \cdots p_k \cdot \left(\frac{1}{3}\right)^{l-k} \quad (3.6)$$

The ratio, $R(DAG) = \frac{P_1(DAG)}{P_0(DAG)}$, of these two different approximated prior probabilities of the DAG, which is all that is needed for prior's input into relative posterior probability according to $P(DAG|Q) \propto f(Q|DAG)P(DAG)$ gives insights. We can see that the ratio is only related to k and the specified prior probabilities p_1, p_2, \dots, p_k of those pairs

$$R(DAG) = \frac{P_1(DAG)}{P_0(DAG)} = \frac{p_1 p_2 \cdots p_k \cdot \left(\frac{1}{3}\right)^{l-k}}{\left(\frac{1}{3}\right)^n} = \frac{p_1 p_2 \cdots p_k}{\left(\frac{1}{3}\right)^k} \quad (3.7)$$

Thus we have the adjusted prior probability of the DAG related to the equally likely prior probability model is

$$P_1(DAG) = \frac{p_1 p_2 \cdots p_k}{\left(\frac{1}{3}\right)^k} P_0(DAG) \quad (3.8)$$

The ratio of different approximated prior probabilities is related to the edges with updated information, all the edges still with a $\frac{1}{3}$ probability will not matter in terms of ratio. Since $P(DAG|Q) \propto f(Q|DAG)P(DAG)$, we can see for posterior probability, only the relative values (e.g. ratios) of the priors $P(DAG)$ and the likelihoods $f(Q|DAG)$ of the DAGs matter as far as the posterior probability of respective DAGs is concerned. Thus the equally likely prior probability $P_0(DAG)$ can be left out for determination of the relative posterior probability of one DAG versus another.

3.2 Typical Non-Edge Approximation of Prior Probability and Its Sensitivity

In the previous section, we assumed that for most of the gene pairs, the three potential cases of an edge or a non-edge were equally likely with the probability $\frac{1}{3}$ when approximating the prior probability. However, in practical cases, a typical non-edge state is often more common than a typical directed state, and thus we might improve our approximation by assuming that the probability of a non-edge for a given typical pair, without further information, is a^* , a^* might be obtained by specialists or empirical information. In this section, we also assume that the probabilities of the two directed states are equally likely. In other words, for a typical node pair A,B, the probabilities of three states are

$$P(A \cdots B) = a^*, P(A \longrightarrow B) = P(A \longleftarrow B) = \frac{1}{2}(1 - a^*) \quad (3.9)$$

It is worth noticing that the equally-likely approximation of prior probability in the previous section is a special case in our new setting when $a^* = \frac{1}{3}$ since the assumption was the three potential connections of an edge or a non-edge were equally likely with the probability $\frac{1}{3}$. In this section with the typical edge approximation, we have more flexibility in estimating the prior probability since the possible typical probability a^* can vary from 0 to 1.

By simplifying the assumption that edges are independent, we may get the prior probability of the graph by multiplying together the prior probabilities of each edge respectively. Suppose that in a DAG with l pairs, there are k of the l pairs with known prior probabilities for their observed edge/non-edge state, labeled as: p_1, p_2, \cdots, p_k . There are $m + n$ of the l pairs which information is unknown and n of which are non-edges and m directed edges, and $m + n + k = l$. Then we have the approximated

prior probability of that DAG is

$$P(DAG) = p_1 p_2 \cdots p_k \cdot (a^*)^n \cdot \left(\frac{1 - a^*}{2} \right)^m \quad (3.10)$$

Since the approximated prior of a DAG is the product of the probabilities of each edge, which is a number between 0 and 1, and when it comes to DAGs with many nodes, $P(DAG)$ often is small and close to zero, $1 - P(DAG)$ would be close to 1. Thus the odds $\frac{P(DAG)}{1 - P(DAG)}$ can be approximated by $P(DAG)$. Sometimes we might have more than one model to estimate the prior probability of a given DAG, the odds ratio or log odds ratio can be applied to measure the sensitivity of two models.

3.2.1 Approximation of Odds Ratio and Log Odds Ratio

For a given DAG and two estimates, $P_0(DAG)$ and $P_1(DAG)$, of its prior probability by two different modelings, their odds are $\frac{P_0(DAG)}{1 - P_0(DAG)}$ and $\frac{P_1(DAG)}{1 - P_1(DAG)}$ respectively, since when $P(DAG)$ is close to zero

$$\frac{P(DAG)}{1 - P(DAG)} \approx P(DAG) \quad (3.11)$$

The odds ratio can be approximated by the ratio of prior probability

$$\frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} \approx \frac{P_1(DAG)}{P_0(DAG)} \quad (3.12)$$

Since the log function is continuous in its domain and $\log(1 - P(DAG))$ is approximately zero, the log odds ratio can be approximated by the log ratio of prior probabilities

$$\log \frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} \approx \log \frac{P_1(DAG)}{P_0(DAG)} \quad (3.13)$$

Since $\log(1 + x) \sim x$ when $x \rightarrow 0$, if $\frac{P_1(DAG)}{P_0(DAG)} \rightarrow 1$, which might not be always achieved practically, then $\frac{P_1(DAG)}{P_0(DAG)} - 1 \rightarrow 0$. We can furthermore approximate $\log \frac{P_1(DAG)}{P_0(DAG)}$ by $\frac{P_1(DAG)}{P_0(DAG)} - 1$, and we can have a theoretical equation

$$\log \frac{P_1(DAG)}{P_0(DAG)} = \log \left(\left(\frac{P_1(DAG)}{P_0(DAG)} - 1 \right) + 1 \right) \sim \frac{P_1(DAG)}{P_0(DAG)} - 1 \quad (3.14)$$

3.2.2 Odds Ratio and Log Odds Ratio for Typical Non-Edge Approximation

By equations (3.11), (3.12), (3.13), and (3.14), we can see the ratio of prior probabilities is crucial in approximating both odds ratios and log odds ratios. Our next step would be finding the ratio of prior probabilities under the typical edge approximation modeling.

When coming to the beginning of our typical edge approximation, for a given DAG and a typical non-edge without further information, we might have two estimates, a^* and b^* , of its probability, the probabilities of the two directed case of that edge are still equally likely, i.e. for any given edge between node A and node B, the probabilities of three potential cases respectively are

$$P_0(A \cdots B) = a^*, P_0(A \longrightarrow B) = P_0(A \longleftarrow B) = \frac{1}{2}(1 - a^*) \quad (3.15)$$

$$P_1(A \cdots B) = b^*, P_1(A \longrightarrow B) = P_1(A \longleftarrow B) = \frac{1}{2}(1 - b^*) \quad (3.16)$$

We still assume that the DAG has $m + n + k$ pairs, where k edges have known prior probabilities p_1, p_2, \dots, p_k , and $m + n$ edges in which information is unknown where n are non-edges and m are directed edges. Then we have the approximated prior probability of that DAG under different typical edge probability estimations are

$$P_0(DAG) = p_1 p_2 \cdots p_k \cdot (a^*)^n \cdot \left(\frac{1-a^*}{2}\right)^m \quad (3.17)$$

$$P_1(DAG) = p_1 p_2 \cdots p_k \cdot (b^*)^n \cdot \left(\frac{1-b^*}{2}\right)^m \quad (3.18)$$

Then we have the ratio of the prior probabilities of a DAG with different probability in a typical pair

$$\frac{P_1(DAG)}{P_0(DAG)} = \frac{p_1 p_2 \cdots p_k \cdot (b^*)^n \cdot \left(\frac{1-b^*}{2}\right)^m}{p_1 p_2 \cdots p_k \cdot (a^*)^n \cdot \left(\frac{1-a^*}{2}\right)^m} = \left(\frac{b^*}{a^*}\right)^n \cdot \left(\frac{1-b^*}{1-a^*}\right)^m \quad (3.19)$$

which implies the ratio $\frac{P_1(DAG)}{P_0(DAG)}$ is irrelevant to the pairs with specified prior probabilities. Given our discussion in odds ratio and log odds ratio, with our assumptions under the typical edge approximations, we have more specific and concise expressions of the odds ratios

$$\begin{aligned} \frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \frac{P_1(DAG)}{P_0(DAG)} \\ &= \left(\frac{b^*}{a^*}\right)^n \cdot \left(\frac{1-b^*}{1-a^*}\right)^m \end{aligned} \quad (3.20)$$

The log odds ration can be simplified as

$$\begin{aligned} \log \frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \log \frac{P_1(DAG)}{P_0(DAG)} \\ &= \log \left[\left(\frac{b^*}{a^*}\right)^n \cdot \left(\frac{1-b^*}{1-a^*}\right)^m \right] \\ &= n \log \left(\frac{b^*}{a^*}\right) + m \log \left(\frac{1-b^*}{1-a^*}\right) \\ &= n [\log b^* - \log a^*] + m [\log(1-b^*) - \log(1-a^*)] \end{aligned} \quad (3.21)$$

From (3.19), (3.20), and (3.21), we can see the second term $m \log \left(\frac{1-b^*}{1-a^*} \right) = m [\log(1-b^*) - \log(1-a^*)]$ will be close to 0 in the settings for relatively small a^* , b^* and m , which is the number of directed edges in the DAG, is not large. As for the first term, $n \log \left(\frac{b^*}{a^*} \right) = n [\log b^* - \log a^*]$ will normally be considerably larger, than the second term when a^* , b^* , and m are small. Thus, we have a more concise approximation of the log odds ratio in this setting,

$$\begin{aligned}
\log \frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \log \frac{P_1(DAG)}{P_0(DAG)} \\
&= \log \left[\left(\frac{b^*}{a^*} \right)^n \cdot \left(\frac{1-b^*}{1-a^*} \right)^m \right] \\
&= n \log \left(\frac{b^*}{a^*} \right) + m \log \left(\frac{1-b^*}{1-a^*} \right) \\
&\approx n \log \left(\frac{b^*}{a^*} \right) \\
&= n [\log b^* - \log a^*]
\end{aligned} \tag{3.22}$$

By the same reasoning, we can also have a more concise approximation of the odds ratio when a^* , b^* , and m are small

$$\begin{aligned}
\frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \frac{P_1(DAG)}{P_0(DAG)} \\
&= \left(\frac{b^*}{a^*} \right)^n \cdot \left(\frac{1-b^*}{1-a^*} \right)^m \\
&\approx \left(\frac{b^*}{a^*} \right)^n
\end{aligned} \tag{3.23}$$

Thus in this setting we have a concise summary of the sensitivity of the prior estimates if the P(typical non-edge) had been b^* rather than a^* .

3.3 Improved Typical Non-Edge Approximation of Prior Probability and Its Sensitivity

In the typical edge approximation, we assumed that the probabilities of a non-edge for any given pairs without further information are the same value a^* and the probabilities of the two directed cases of that edge are equally likely. We might improve the assumption by assuming the probability of a non-edge for a pair $\{A, B\}$ without further information is different by a multiple c_α , where the multiple c_α is subject to the specific nodes $\{A, B\}$, rather than the same value for all pairs of nodes, i.e.

$$P(A \cdots B) = c_\alpha a^*, P(A \longrightarrow B) = P(A \longleftarrow B) = \frac{1}{2}(1 - c_\alpha a^*) \quad (3.24)$$

Since the typical edge approximation of prior probability in the previous section is a special case in our new setting when all the multiples are one, i.e. $c_i = 1$ for all i pair of nodes. In this section with the improved typical edge approximation, we have more flexibility in estimating the prior probability when the possible probability $c_i a^*$ between the i th pair can be adjusted by a multiple c_i with respect to the specific pairs of nodes.

For a DAG with $m+n+k$ pairs, in which are k edges have known prior probabilities p_1, p_2, \dots, p_k , and there are $m+n$ edges in which information is unknown with n non-edges and m directed edges, the prior probability for a non-edge pair is different by a multiple c_i , then the prior probability of the DAG is

$$P_0(DAG) = p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1 - c_{n+1} a^*}{2} \right) \cdots \left(\frac{1 - c_{n+m} a^*}{2} \right) \quad (3.25)$$

With our assumptions in this section, we have more flexibility in modeling the prior probability between a pair of nodes. However, it is still different from specifying a prior probability, $p_1, p_2, p_3, \dots, p_{\binom{N}{2}}$, to each pair of nodes in the DAG, since we are

assuming the cases of directed edges between a pair of nodes are equally likely, i.e. $P(A \longrightarrow B) = P(A \longleftarrow B)$.

For a given DAG and a typical non-edge without further information, we might have two estimations of its probability, $c_\alpha a^*$ and $c'_\alpha b^*$, the probabilities of the two directed cases of that edge are still equally likely, i.e. for any given edge between node A and node B, the probabilities of three potential cases respectively are

$$P_0(A \cdots B) = c_\alpha a^*, P_0(A \longrightarrow B) = P_0(A \longleftarrow B) = \frac{1}{2}(1 - c_\alpha a^*) \quad (3.26)$$

$$P_1(A \cdots B) = c'_\alpha b^*, P_1(A \longrightarrow B) = P_1(A \longleftarrow B) = \frac{1}{2}(1 - c'_\alpha b^*) \quad (3.27)$$

For the same DAG has $m+n+k$ pairs, where k edges have known prior probabilities p_1, p_2, \dots, p_k , and $m+n$ edges in which information is unknown where n are non-edges and m are directed edges, we have the approximated prior probability of that DAG under different typical non-edge probability estimations are

$$P_0(DAG) = p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1 - c_{n+1} a^*}{2} \right) \cdots \left(\frac{1 - c_{n+m} a^*}{2} \right) \quad (3.28)$$

$$P_1(DAG) = p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot \left(\frac{1 - c'_{n+1} b^*}{2} \right) \cdots \left(\frac{1 - c'_{n+m} b^*}{2} \right) \quad (3.29)$$

Then we have the ratio of prior probabilities of the given DAG under improved typical non-edge approximation is

$$\begin{aligned} \frac{P_1(DAG)}{P_0(DAG)} &= \frac{p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot \left(\frac{1 - c'_{n+1} b^*}{2} \right) \cdots \left(\frac{1 - c'_{n+m} b^*}{2} \right)}{p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1 - c_{n+1} a^*}{2} \right) \cdots \left(\frac{1 - c_{n+m} a^*}{2} \right)} \\ &= \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*} \right)^n \cdot \left(\frac{1 - c'_{n+1} b^*}{1 - c_{n+1} a^*} \right) \cdots \left(\frac{1 - c'_{n+m} b^*}{1 - c_{n+m} a^*} \right) \end{aligned} \quad (3.30)$$

Equation (3.30) shows that the accuracy of c_k and c'_k is important in the estimation of prior probability and odds ratios.

We can see $\left(\frac{1-c'_{n+1}b^*}{1-c_{n+1}a^*}\right) \cdots \left(\frac{1-c'_{n+m}b^*}{1-c_{n+m}a^*}\right)$ will probably be close to 1 in most settings for relatively small $c_{n+j}a^*$ and $c'_{n+j}b^*$, $j = 1, 2, \dots, m$, and the number m of directed edges in the DAG is not large. As for the term $\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*}\right)^n$ will normally be considerably larger than the second term when $c_i a^*$, $c'_i b^*$, $i = 1, 2, \dots, n$, and m are small. Thus we have a more concise approximation of the log odds ratio

$$\begin{aligned}
\frac{P_1(DAG)}{P_0(DAG)} &= \frac{p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot \left(\frac{1-c'_{n+1}b^*}{2}\right) \cdots \left(\frac{1-c'_{n+m}b^*}{2}\right)}{p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1-c_{n+1}a^*}{2}\right) \cdots \left(\frac{1-c_{n+m}a^*}{2}\right)} \\
&= \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*}\right)^n \cdot \left(\frac{1-c'_{n+1}b^*}{1-c_{n+1}a^*}\right) \cdots \left(\frac{1-c'_{n+m}b^*}{1-c_{n+m}a^*}\right) \quad (3.31) \\
&\approx \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*}\right)^n
\end{aligned}$$

Given the ratio of prior probabilities and our discussions in the approximations of odds ratio and log odds ratio, we have more specific and concise expressions of those ratios

$$\begin{aligned}
\frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \frac{P_1(DAG)}{P_0(DAG)} \\
&= \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*}\right)^n \cdot \left(\frac{1-c'_{n+1}b^*}{1-c_{n+1}a^*}\right) \cdots \left(\frac{1-c'_{n+m}b^*}{1-c_{n+m}a^*}\right) \quad (3.32) \\
&\approx \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*}\right)^n
\end{aligned}$$

$$\begin{aligned}
\log \frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \log \frac{P_1(DAG)}{P_0(DAG)} \\
&= \log \left[\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*} \right)^n \cdot \left(\frac{1 - c'_{n+1} b^*}{1 - c_{n+1} a^*} \right) \cdots \left(\frac{1 - c'_{n+m} b^*}{1 - c_{n+m} a^*} \right) \right] \\
&\approx \log \left[\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \left(\frac{b^*}{a^*} \right)^n \right] \\
&= \log \left(\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \right) + n \log \left(\frac{b^*}{a^*} \right)
\end{aligned} \tag{3.33}$$

From the concise approximations of the odds ratio and log odds ratio above, we can see that mainly the probabilities of the non-edges of different estimations matter.

Furthermore, we might also write the prior probability of a directed edge as a multiple of a typical edge, then every possible connection without specified prior probability is measured in terms of a typical probability a^*

$$P(A \cdots B) = c_\alpha a^*, P(A \longrightarrow B) = P(A \longleftarrow B) = \frac{1}{2}(1 - c_\alpha a^*) = d_\alpha a^* \tag{3.34}$$

For the DAG with $m + n + k$ pairs, the prior probability for any pair is different by a multiple,

$$\begin{aligned}
P_0(DAG) &= p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1 - c_{n+1} a^*}{2} \right) \cdots \left(\frac{1 - c_{n+m} a^*}{2} \right) \\
&= p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot (d_1 a^*) \cdots (d_m a^*)
\end{aligned} \tag{3.35}$$

For a given DAG and a typical non-edge without further information, we might have two estimations of its typical probability, a^* and b^* , the probabilities of the two directed cases of that edge are still equally likely and expressed in terms of multiples of typical probabilities, i.e. for any given edge between node A and node B, the probabilities of three potential cases respectively are

$$P_0(A \cdots B) = c_\alpha a^*, P_0(A \longrightarrow B) = P_0(A \longleftarrow B) = \frac{1}{2}(1 - c_\alpha a^*) = d_\alpha a^* \tag{3.36}$$

$$P_1(A \cdots B) = c'_\alpha b^*, P_1(A \longrightarrow B) = P_1(A \longleftarrow B) = \frac{1}{2}(1 - c'_\alpha b^*) = d'_\alpha b^* \quad (3.37)$$

For the same DAG with $m + n + k$ pairs, we have that the approximated prior probability of that DAG under different typical edge probability estimations are

$$\begin{aligned} P_0(DAG) &= p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot \left(\frac{1 - c_{n+1} a^*}{2} \right) \cdots \left(\frac{1 - c_{n+m} a^*}{2} \right) \\ &= p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot (d_1 a^*) \cdots (d_m a^*) \end{aligned} \quad (3.38)$$

$$\begin{aligned} P_1(DAG) &= p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot \left(\frac{1 - c'_{n+1} b^*}{2} \right) \cdots \left(\frac{1 - c'_{n+m} b^*}{2} \right) \\ &= p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot (d'_1 b^*) \cdots (d'_m b^*) \end{aligned} \quad (3.39)$$

Then the ratio of prior probabilities of the given DAG is

$$\begin{aligned} \frac{P_1(DAG)}{P_0(DAG)} &= \frac{p_1 p_2 \cdots p_k \cdot (c'_1 b^*) \cdots (c'_n b^*) \cdot (d'_1 b^*) \cdots (d'_m b^*)}{p_1 p_2 \cdots p_k \cdot (c_1 a^*) \cdots (c_n a^*) \cdot (d_1 a^*) \cdots (d_m a^*)} \\ &= \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \cdot \frac{d'_1 d'_2 \cdots d'_m}{d_1 d_2 \cdots d_m} \cdot \left(\frac{b^*}{a^*} \right)^{n+m} \end{aligned} \quad (3.40)$$

Given the ratio of prior probabilities and our discussions in the approximations of odds ratio and log odds ratio, we have more specific and concise expressions of those ratios if all the c_k, d_k, c'_k, d'_k terms are accurate.

$$\begin{aligned} \frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} &\approx \frac{P_1(DAG)}{P_0(DAG)} \\ &= \frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \cdot \frac{d'_1 d'_2 \cdots d'_m}{d_1 d_2 \cdots d_m} \cdot \left(\frac{b^*}{a^*} \right)^{n+m} \end{aligned} \quad (3.41)$$

$$\begin{aligned}
\log \frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} &\approx \log \frac{P_1(DAG)}{P_0(DAG)} \\
&= \log \left[\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \cdot \frac{d'_1 d'_2 \cdots d'_m}{d_1 d_2 \cdots d_m} \cdot \left(\frac{b^*}{a^*} \right)^{n+m} \right] \\
&= \log \left(\frac{c'_1 c'_2 \cdots c'_n}{c_1 c_2 \cdots c_n} \right) + \log \left(\frac{d'_1 d'_2 \cdots d'_m}{d_1 d_2 \cdots d_m} \right) + (m+n) \log \left(\frac{b^*}{a^*} \right)
\end{aligned} \tag{3.42}$$

3.4 Example of A DAG with Four Nodes

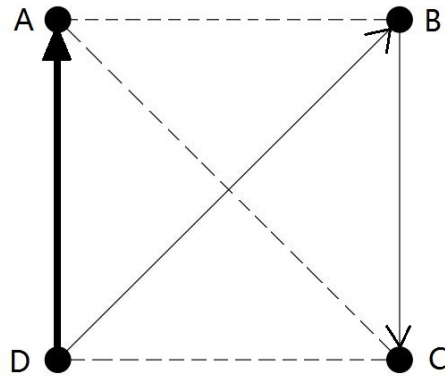


Figure 3.1: A DAG with Four Nodes

To illustrate different methods in approximating the prior probability of a DAG, we come up with an example of a DAG with four nodes. Here we have $N = 4$ nodes and $l = \binom{4}{2} = 6$ pairs. Then for the model with assumptions of equally likely connections to all pairs of nodes, the prior probability of the given DAG is

$$P_0(DAG) = \left(\frac{1}{3} \right)^6 \tag{3.43}$$

If we can have the probability that there is a directed edge from D to A is high

from the knowledge of biology, say the probability is 0.9, and for the rest of the edges the probabilities are equally likely for the three cases \longrightarrow , \longleftarrow , and \cdots . That is to say, $k = 1$, then $P(D \longrightarrow A) = p_1 = 0.9$ in our previous notation. The prior probability of the given DAG under the adjusted equal likelihood model is

$$P_1(DAG) = p_1 \cdot \left(\frac{1}{3}\right)^5 = 0.9 \cdot \left(\frac{1}{3}\right)^5 \quad (3.44)$$

The ratio of the above and the equally likely approximations of the prior probability of the DAG is

$$R(DAG) = \frac{P_1(DAG)}{P_0(DAG)} = \frac{p_1 \cdot \left(\frac{1}{3}\right)^{6-1}}{\left(\frac{1}{3}\right)^6} = \frac{p_1}{\left(\frac{1}{3}\right)^1} = 0.9 \times \frac{1}{3} = 0.3 \quad (3.45)$$

Thus we have the adjusted indifferent prior probability of the DAG related to the equally likely prior probability model is

$$P_1(DAG) = R(DAG) \times P_0(DAG) = 0.3P_0(DAG) \quad (3.46)$$

For the DAG with four nodes and $l = \binom{4}{2} = 6$ pairs, there is $k = 1$ pair with known prior probability $P(D \longrightarrow A) = p_1 = 0.9$, 5 pairs for which specific prior information is unknown and 3 of which are non-edges and 2 directed edges. The probability of a non-edge for a given pair without further information is $a^* = 0.1$, then the probability of a particular directed edge for a given pair without further information is $\frac{1}{2}(1 - a^*) = 0.45$.

Then we have the approximated prior probability of that DAG by typical edge approximation is

$$P_0(DAG) = p_1 \cdot (a^*)^3 \cdot \left(\frac{1 - a^*}{2}\right)^2 = 0.9 \times 0.1^3 \times 0.45^2 = 1.8225 \times 10^{-4} \quad (3.47)$$

If the prior probability of a non-edge for a given pair without further information is $b^* = 0.2$, then the prior probability of a directed edge for a given pair without further information is $\frac{1}{2}(1 - b^*) = 0.4$. Then we have another approximated prior probability of the DAG

$$P_1(DAG) = p_1 \cdot (b^*)^3 \cdot \left(\frac{1 - b^*}{2}\right)^2 = 0.9 \times 0.2^3 \times 0.4^2 = 1.152 \times 10^{-3} \quad (3.48)$$

Then we have the following

$$\frac{P_1(DAG)}{P_0(DAG)} = 6.32098765432 \quad (3.49)$$

$$\frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} = \frac{0.00115332863}{0.00018228322} = 6.32712454052 \quad (3.50)$$

$$\left(\frac{b^*}{a^*}\right)^n = \left(\frac{0.2}{0.1}\right)^3 = 8 \quad (3.51)$$

We can see

$$\frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} \approx \frac{P_1(DAG)}{P_0(DAG)} \quad (3.52)$$

If a^* and b^* are close and very small we can also have

$$\frac{\frac{P_1(DAG)}{1 - P_1(DAG)}}{\frac{P_0(DAG)}{1 - P_0(DAG)}} \approx \frac{P_1(DAG)}{P_0(DAG)} \approx \left(\frac{b^*}{a^*}\right)^n \quad (3.53)$$

Next we examine the improved typical edge approximation of the DAG. $P(D \rightarrow A) = p_1 = 0.9$, the probability of a typical non-edge for a given pair without further

information is $a^* = 0.1$, then we need to specify the probability of each edge with a multiple,

$$P_0(A \cdots B) = a^*, P_0(A \cdots C) = 1.1a^*, P_0(C \cdots D) = 1.2a^* \quad (3.54)$$

$$P_0(B \longrightarrow C) = \frac{1}{2}(1 - 0.9a^*), P_0(D \longrightarrow B) = \frac{1}{2}(1 - 1.3a^*) \quad (3.55)$$

Then we have that the approximated prior probability of that DAG by improved typical edge approximation is the product of these because of the above pairwise relationships.

$$\begin{aligned} P_0(DAG) &= P_0(D \longrightarrow A)P_0(A \cdots B)P_0(A \cdots C)P_0(C \cdots D)P_0(B \longrightarrow C)P_0(D \longrightarrow B) \\ &= 2.351349 \times 10^{-4} \end{aligned} \quad (3.56)$$

If the probability of a typical non-edge for a given pair without further information is $b^* = 0.12$, then we need to specify the probability of each edge with a respective multiple,

$$P_1(A \cdots B) = b^*, P_1(A \cdots C) = 1.1b^*, P_1(C \cdots D) = 1.12b^* \quad (3.57)$$

$$P_1(B \longrightarrow C) = \frac{1}{2}(1 - 0.8b^*), P_1(D \longrightarrow B) = \frac{1}{2}(1 - 1.4b^*) \quad (3.58)$$

Then we have the approximated prior probability of that DAG by improved typical edge approximation is

$$\begin{aligned} P_1(DAG) &= P_1(D \longrightarrow A)P_1(A \cdots B)P_1(A \cdots C)P_1(C \cdots D)P_1(B \longrightarrow C)P_1(D \longrightarrow B) \\ &= 3.8600412 \times 10^{-4} \end{aligned} \quad (3.59)$$

Then we have the following

$$\frac{P_1(DAG)}{P_0(DAG)} = 1.64162835887 \quad (3.60)$$

$$\frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} = \frac{0.00038615317}{0.0002351902} = 1.64187610708 \quad (3.61)$$

$$\frac{c'_1 c'_2 \cdots c'_n \left(\frac{b^*}{a^*}\right)^n}{c_1 c_2 \cdots c_n} = \frac{1 \times 1.1 \times 1.12}{1 \times 1.1 \times 1.2} \left(\frac{0.12}{0.1}\right)^3 = 1.6128 \quad (3.62)$$

We can see if a^* and b^* are close to zero, the following approximation is true and it is a better approximation when n , the number of non-edges, is relatively large.

$$\frac{\frac{P_1(DAG)}{1-P_1(DAG)}}{\frac{P_0(DAG)}{1-P_0(DAG)}} \approx \frac{P_1(DAG)}{P_0(DAG)} \approx \frac{c'_1 c'_2 \cdots c'_n \left(\frac{b^*}{a^*}\right)^n}{c_1 c_2 \cdots c_n} \quad (3.63)$$

3.5 Some Remarks of the Three Approximation Methods

For the three approximation methods given in this chapter, the adjusted equal probability approximation methods requires more information about prior probabilities of specific pairs from specialists while the probabilities of the rest of the pairs remain equally likely, it incorporates the empirical information of specialists to improve the prior probability while still keeps a relatively low computation complexity. The typical non-edge approximation method improves the approximation of the prior probability by bringing in a typical probability of non-edges to make the approximation more flexible and precise, this might requires more computations but gives more accuracy if original estimates are correct. For the improved typical edge approximation, it requires more information to adjust the probability of each pair without specific prior information and thus it has more flexibility and accuracy but it increases the computation inevitably. We can see at a lot of times, the typical edge approximations is the middle ground between accuracy and computation complexity but all three

approximation methods can be used according to specific situations with different requirements and purposes.

Chapter 4: Conclusions

To obtain the relative posterior probability of a DAG, we need to know the likelihood of a DAG which is denoted as $f(Q|DAG)$, and if a uniform prior is not assumed, an estimate of the prior probability of the DAG, which is denoted as $P(DAG)$. By Bayes Theorem, for a given DAG we have

$$P(DAG|Q) = \frac{f(Q|DAG)P(DAG)}{f(Q)} \propto_{DAG} f(Q|DAG)P(DAG) \quad (4.1)$$

The computation of $f(Q|DAG)$ has been discussed in Chapter 2 and we focused on the prior probability $P(DAG)$ in Chapter 3.

In Chapter 2, the relative likelihood of a DAG is computed for multiple replicates within a single lab and then separately for multiple replicates with each of multiple laboratories assuming extended hierarchical structures. A result based on multiple replicates in a single laboratory is given first, then the result for multiple laboratories and multiple replicates within laboratories is given by adding an additional layer to the hierarchy structure. In the latter there is also a more general setting given for multiple laboratories for possible future work in adding another level beyond laboratories. The main ideas used in chapter 2 involved expressing $f(Q|DAG)$ in terms of slope and variance parameters with given distributions and integrating the product of density functions to eliminate the parameters one by one. Specifically, much of the efforts in integrating over a parameter are used in collecting all the terms containing the parameter and then constructing a probability density function. By allowing multiple labs, more information can be integrated to obtain better relative likelihoods with DAGs.

Chapter 3 first gives three methods to approximate and estimate the prior prob-

abilities of DAGs and their sensitivity based on different assumptions and additional information. For the three approximation methods given in Chapter 3, the adjusted equal likelihood approximation methods require more information about prior probabilities of specific pairs from specialists while the probabilities of the rest of the pairs remain equally likely. It incorporates the empirical information of specialists to improve the prior probability while still keeping a relatively low computation complexity. The typical non-edge approximation method improves the approximation of the prior probability by bringing in a typical probability of non-edges to make the approximation more flexible and precise. This requires more computations than equally likely prior estimation but gives more accuracy if the knowledge is correct. For the improved typical non-edge approximation, it requires more information to adjust the probability of each pair without specific prior information and thus it has more flexibility and perhaps accuracy but it increases the computation. At the same time, Chapter 3 also shows that the odds ratio of two estimating terms under the same setting often has a much more concise form and is easier in practical computations than the prior probabilities themselves. Specifically, many of the common terms of the priors can be canceled in the ratio.

Unfortunately even with a relatively small number of genes, for example 12, it would be prohibitive to get the likelihood associated with every possible DAG [8, 15]. Thus in practical applications, we have to use a fine-tuned genetic searching algorithm over the DAG space to search for DAGs with high relative posterior probability. In previous work, we have assumed equal priors for DAGs and only a single lab and have obtained estimated posterior probabilities of edges through the search and use of Bayesian model averaging. This new research in Chapter 3 on non-equal priors for DAGs, allows us to better approximate the relative posterior of a DAG by taking the product of the prior of the DAG and the likelihood (density function of the data) for

the DAG. This research's first part is an extension from a single laboratory to multiple laboratories. This allows us to get an improved relative posterior probability of a DAG by allowing the relative likelihood to be both over replicates within a laboratory and over laboratories. The product of the new likelihood and the new prior for a DAG is potentially a much better estimator of the relative posterior probability of the DAG than was our previous estimator.

Bibliography

- [1] H. Bolouri, Computational Modeling of Gene Regulatory Networks A Primer, Imperial College Press, 2008.
- [2] E.E. Allen, J.S. Fetrow, L.W. Daniel, S.J. Thomas, D.J. John, "Algebraic Dependency Models of Protein Signal Transduction Networks from Time-Series Data", J. Theoretical Biology, vol. 238, no. 2, pp. 317-330, Jan. 2006.
- [3] R. Laubenbacher, B. Stigler, "A Computational Algebra Approach to the Reverse Engineering of Gene Regulatory Networks", J. Theoretical Biology, vol. 229, no. 4, pp. 523-537, Aug. 2004.
- [4] 5. N. Friedman, M. Linial, I. Nachman, D. Peer, "Using Bayesian Networks to Analyze Expression Data", J. Computational Biology, vol. 7, no. 3, pp. 601-620, 2000.
- [5] D.J. John, J.S. Fetrow and J.L. Norris Continuous cotemporal probabilistic modeling of systems biology networks from sparse data. IEEE/ACM Transactions on computational biology and bioinformatics journal. (2011) 8 (2) 1208-1222.
- [6] K.L. Patton, Bayesian Interaction and Associated Networks from Multiple Replicates of Sparse Time-Course Data. Master's Thesis, Dept. of Math., Wake Forest University, May 2012.
- [7] K. L. Patton, D. J. John, J. L. Norris, D. R. Lewis and G. K. Muday, "Hierarchical Probabilistic Interaction Modeling for Multiple Gene Expression Replicates," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 11, no. 2, pp. 336-346, March-April 2014.

- [8] Gross, Jonathan L.; Yellen, Jay; Zhang, Ping (2013), Handbook of Graph Theory (2nd ed.), CRC Press.
- [9] Peter D. Hoff, A First Course in Bayesian Statistical Methods, Springer-Verlag New York 2009.
- [10] Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). Bayesian data analysis. Chapman and Hall/CRC.
- [11] Morris H. DeGroot and Mark J. Schervish. Probability and Statistics: Fourth Edition. Pearson.
- [12] J.S. Clark, M. Wolosin, M. Dietze, I. Ibanez, S. LaDeau, M. Welsh, B. Kloppe, "Tree Growth Inference and Prediction from Diameter Censuses and Ring Widths", Ecological Applications, vol. 17, no. 7, pp. 1942-1953, 2007.
- [13] A. Gelman, J. Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge Univ. Press, 2007.
- [14] K.L. Patton, D.J. John and J.L. Norris Bayesian probabilistic network modeling from multiple independent replicates. BMC Bioinformatics (2012) 13 (Supplement 9) 1-13.
- [15] Pearl, Judea (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.
- [16] Shmulevich, Ilya; Dougherty, Edward R. (2010), Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks, Society for Industrial and Applied Mathematics.
- [17] Fragoso, Tiago and Bertoli, Wesley and Louzada, Francisco. (2015). Bayesian

Model Averaging: A Systematic Review and Conceptual Classification. *International Statistical Review*. 86. 10.1111/insr.12243.

- [18] B.A. LaPointe, D.J. John, J.L. Norris, E.E. Allen, A.F. Harkey, J.K. Muhlemann and G. Muday, "A BCHC genetic algorithm model of cotemporal hierarchical *Arabidopsis thaliana* gene interactions," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 2701-2708.
- [19] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, "Bayesian Model Averaging: A Tutorial (with Comments by M. Clyde David Draper and E.I. George and a Rejoinder by the Authors)", *Statistical Science*, vol. 14, no. 4, pp. 382-417, 1999.
- [20] Y. Wang, T. Joshi, X.-S. Zhang, D. Xu, L. Chen, "Inferring Gene Regulatory Networks From Multiple Microarray Datasets", *Bioinformatics*, vol. 22, no. 19, pp. 2413-2420, 2006.

Curriculum Vitae :

XIAOHUAN XUE

Email: xuex17@wfu.edu

Department of Mathematics and Statistics
Wake Forest University, Winston-Salem, NC 27109

EDUCATION

Wake Forest University M.A. in Mathematics (Thesis Track) Advisor: Dr. James Norris	<i>Sep. 2017 - May 2019</i>
Chinese Academy of Sciences Academy of Mathematics and Systems Science Master of Science Advisor: Qun Lin (Academician of CAS, SIAM Fellow, TWAS Fellow)	<i>Sep. 2014 - Jul. 2017</i>
Taiyuan Normal University Bachelor of Science	<i>Sep. 2009 - Jul. 2013</i>

EXPERIENCE

Tutor at Math Center <i>Wake Forest University</i>	<i>Sep. 2017 - May 2019</i>
Research Assistant <i>Math Department and Center for Molecular Signaling</i> <i>Wake Forest University</i>	<i>Jan. 2018 - May 2019</i>
Teaching Assistant <i>Wake Forest University</i> Elementary Probability and Statistics, Instructor: Dr. Nicole Dalzell Calculus I, Instructor: Dr. Frank Moore	<i>June 2018 - Aug. 2018</i>
Research Assistant <i>The National Laboratory of Scientific and Engineering Computing</i> <i>Academy of Math and Systems Science, Chinese Academy of Sciences</i>	<i>Jul. 2013 - Aug. 2017</i>
Coauthor Calculus Fast Lane (The Third Edition), <i>Science Press, Beijing, 2013.</i>	<i>2013</i>
Invited Speaker <i>International Workshop on Application and Computation of Mathematics</i> Academy of Math and Systems Science, Chinese Academy of Sciences, Beijing	<i>July 2014</i>

SELECTED AWARDS

Graduate Academic Scholarship , Chinese Academy of Sciences	<i>2014, 2015, 2016</i>
First Prize , Chinese Mathematics Contest for College Students (CMC)	<i>2011, 2012</i>
First Prize , National English Contest for College Students (NECCS)	<i>2011, 2012</i>