Vrije Universiteit Amsterdam

Universiteit van Amsterdam

MSc Bioinformatics and Systems Biology

Literature Review

Comparison of Frequentist and Bayesian Methods for Inference of Gene Regulatory Networks

Olivier MF Martin

supervised by

Huub Hoefsloot
Johan Westerhuis

July 21, 2016

Contents

1 Introduction 3

2 Mathematical background 4

2.1 Mathematical graphs and biological networks 4

2.2 Difference between Bayesian and frequentist approaches . . . 4

2.2.1 Frequentist approaches 5

2.2.2 Bayesian approaches 6

3 Gene regulatory network inference 8

3.1 Definition 8

3.2 Motivation 8

3.3 Difficulties 9

3.4 Methods 10

3.4.1 Association measures 10

3.4.2 Graphical models 10

3.4.3 Regression 12

4 Comparison of frequentist and Bayesian methods 12

4.1 Prediction accuracy 12

4.2 Inference 13

4.3 Integration of prior knowledge 15

4.4 Robustness to noise, small sample size and missing data . . . 16

4.5 Computational time 16

5 Conclusion 17

Acronyms

BMA Bayesian Model Averaging. 7, 14

BN Bayesian Network. 4, 10–13, 16

ChIP Chromatin Immunoprecipitation. 8, 9, 15

DREAM Dialogue on Reverse-Engineering Assessment and Methods. 12, 13, 17

GGM Graphical Gaussian Model. 4, 10, 11, 13

GLASSO Graphical LASSO. 11

GRN Gene Regulatory Network. 3, 4, 8, 9, 15–17

LASSO Least Absolute Shrinkage and Selection Operator. 5, 12, 15, 16

MAP Maximum *a Posteriori*. 6

MCMC Markov Chain Monte Carlo. 6, 11, 13, 14, 16

ML Maximum Likelihood. 5, 6, 10

RNAseq RNA Sequencing. 8, 9

1 Introduction

High-throughput technologies such as microarrays and massive parallel sequencing have resulted in an exponential growth of the amount of data available in biology. In order to make sense of these data, one must find comprehensible and analyzable representations. One representation that has shown to be successful in the field of systems biology is that of a network (Emmert-Streib et al., 2014). Consequently, numerous statistical and computational techniques have been developed in order to infer the structure of biological networks from observational data (Penfold and Wild, 2011).

Network inference methods can be distinguished, on the basis of their statistical approach, into frequentist and Bayesian. These two statistical frameworks have been the object of major debate inside the statistical community during the past century. Statisticians have now recognized that there is an interplay between them and that methods from both toolboxes should be used (Bayarri and Berger, 2004). Nonetheless, frequentist statistics remain the most widely used approach in bioinformatics and systems biology. It is only recently that the potential of Bayesian methods for biological data analysis was recognized. Indeed, biological data is complex, non-linear and noisy, and often only partially observed. Frequentist approaches tend to struggle with data with such characteristics. For these reasons, Bayesian approaches are becoming more and more used (Wilkinson, 2007).

The goal of this literature review is to compare frequentist and Bayesian methods for network inference in terms of predictive accuracy, inference, prior knowledge integration, robustness to noise and small sample size, and computational time. We expect that different methods will not have the same performance for different biological networks. Accordingly, our analysis will mostly limit itself to Gene Regulatory Networks (GRNs) that are by far the most studied in the scientific literature. Nonetheless, we will exceptionally discuss some particularly interesting results concerning different biological networks.

The remainder of the review is organized as follows. We will start by explaining the mathematical background necessary. This includes a mathematical definition a graph and its relationship to biological networks, but also, the main differences between the frequentist and Bayesian approaches. Subsequently, we will describe the network inference problem and the computational methods used to solve it. Finally, we will compare the advantages and disadvantages of frequentist and Bayesian methods.

2 Mathematical background

2.1 Mathematical graphs and biological networks

A graph, or network, is an ordered pair of sets $G = (V, E)$ composed of a set of vertices $V = \{v_1, \ldots, v_p\}$ and a set of edges E connecting vertices. In a directed graph, such as Boolean networks and Bayesian Networks (BNs), edges have a direction and are represented as ordered pairs: $E = \{e_1 = (v_i, v_j), \ldots, e_m = (v_k, v_l)\}$. Likewise, in an undirected graph, such as relevance networks and graphical Gaussical networks, edges have no direction and are represented as unordered pairs: $E = \{e_1 = \{v_i, v_j\}, \ldots, e_m = \{v_k, v_l\}\}$. A graph can also be represented as an adjacency matrix $A(p \times p)$ where $a_{ij} = 1$ if there exists an edge between vertices v_i and v_j $\{e_i, e_j\} \in E$ and $a_{ij} = 0$ if that edge does not exist $\{e_i, e_j\} \notin E$ (Dehmer et al., 2011).

In biological networks, a vertex v_i represents a biomolecule such as a gene in GRNs, a protein in protein-protein interaction networks or a metabolite in metabolic networks. In some cases, a vertex can also represent a functional module of biomolecules (De Smet and Marchal, 2010). Whereas the meaning of a vertex is straightforward, the meaning of an edge is more subtle, and depends on the biological network being studied and the mathematical model being used (Stolovitzky et al., 2007). For example, in a GRN, an undirected edge $\{v_i, v_j\}$ will represent a statistical association between genes v_i and v_j, whereas, a directed edge (v_i, v_j) can have a biological meaning by representing a causal relationship such as gene v_i activates gene v_j. On the other hand, undirected edges in protein-protein interaction networks or metabolic network can also have biological meaning by representing physical interactions between molecules.

The granularity of a GRN depends on its mathematical representation. As an illustration, the vertices in a Boolean network can only represent active or inactive genes. On the other side, vertices in a Graphical Gaussian Models (GGMs) allow genes to take continuous expression levels. Both these models are static, however, dynamic models, such as dynamic BNs, can also include temporal information (Kim et al., 2014).

2.2 Difference between Bayesian and frequentist approaches

The goal of GRN inference is to determine the adjacency matrix $A(p \times p)$ of a GRN using the matrix of expression levels of p genes found in n different samples $X(n \times p)$. One possible manner to classify GRN methods is by the statistical framework: frequentist or Bayesian. In this section, we will describe the differences between these two approaches. Please note that this section is not meant to be a description of a general framework for solving the network inference problem, but rather an introduction to the differences between frequentist and Bayesian approaches. Because of this, we will describe the differences using an illustrative example; we will suppose we are inferring a biological network using linear models as in Bonneau et al. (2006) and Hill et al. (2012b). These models will be further described in section 3.4.3. Figure 1 summarizes the presented workflow.

For any v_i of the p vertices in v, we can assume that the expression levels X^i of gene v_i can be expressed as a linear function of all other genes in the network $X^{\setminus i} = X \setminus X^i$ with regression coefficient vector $\beta^i = (\beta^i_1, \ldots, \beta^i_{p-1})^T$ and random centred Gaussian noise: ϵ: $X^i = X^{\setminus i}\beta^i + \epsilon$ (Margolin and Califano, 2007). If $\beta^i_j \neq 0$, this can be interpreted as the presence of an edge e_{ij} between genes v_i and v_j. In terms of the adjacency matrix A, $a_{ij} = 1$ if $\beta^i_j \neq 0$ else 0. This strategy is then repeated for all genes, thus p times. We will now describe the general frequentist and the Bayesian approach to solving this problem.

2.2.1 Frequentist approaches

Frequentist approaches are often based on Maximum Likelihood (ML) point estimates of a model's parameter given the observed data. The likelihood function $L(\beta^i|X)$ describes the agreement between the model's predictions $X^{\setminus i}\beta^i$ and the observed data X^i. ML attempts to identify the set of parameters $\hat{\beta}^i$ that maximizes the likelihood function in order to obtain the best fitting model: $\hat{\beta}^i = \arg\max_b L(b|X)$ (Raue et al., 2013). This corresponds to solving the ordinary least squares equation: $\hat{\beta}^i = \arg\min_b \left\|X^i - X^{\setminus i}b\right\|^2$. One should note here that there is only one model that is yielded by solving the optimization problem.

Given that biological networks are sparse, it makes sense to solve the linear regression problem using a penalized regression method such as the Least Absolute Shrinkage and Selection Operator (LASSO) in order to shrink a reasonable amount of regression coefficient β^i to zero. This corresponds to adding a penalty parameter to the ordinary least square equation is less than a certain value (Wu et al., 2009). This problem becomes where λ determines the amount of regularization and thus the sparsity of the network and $g(X^{\setminus i})$ is a penalty term such as L1 norm of $\left\|\beta^i\right\|^1$ for LASSO. The penalty term can also be adapted to include prior information about the network as in Studham et al. (2014).

$$\hat{\beta}^i = \arg\min_b \left\| X^i - X^{\setminus i} b \right\|^2 + \lambda g(X^{\setminus i})$$

The significance of the presence of an edge is evaluated by p-values in the frequentist framework. P-values estimate the probability of obtaining a result β^i_j at least as extreme given the null hypothesis H_0, the case where no edges are present: $H_0 = \beta^i_j = 0 \ \forall j$ For a vertex v_i, an edge e_{ij} and corresponding estimated regression coefficient $\hat{\beta}^i_j$, its p-value is defined as $p(\hat{\beta}^i_j \neq 0 | \beta^i_j = 0 \ \forall j)$. P-values can be computed using theoretical distribution of regression coefficient or using permutation tests. Most often, one we define a threshold α such as 0.05 and consider as significant any edge with a p-value inferior to α.

$$a_{ij} = \begin{cases} 1 & \text{if } p(\hat{\beta}^i_j \neq 0 | \beta^i_j = 0 \ \forall j) < \alpha \\ 0 & \text{otherwise} \end{cases}$$

2.2.2 Bayesian approaches

Bayesian approaches are based on the Bayesian theorem. This theorem states that the probability of the model's parameters posterior to observing data $p(\beta^i|X)$ is proportional to the likelihood $p(X|\beta^i)$ times and the probability prior to observing data $p(\beta^i)$. The prior probability $p(\beta^i)$ allows inclusion of knowledge about the biological network such as scarcity or previous evidence for edges (Mukherjee and Speed, 2008) (Hill et al., 2012b) (Santra, 2014). The normalizing coefficient is equal to the probability of observing the data $p(X)$.

$$p(\beta^i|X) = \frac{p(X|\beta^i)p(\beta^i)}{p(X)}$$

The goal of Bayesian approaches is to estimate the posterior probability $p(\beta^i|X)$ by making use of the previous equation. However, practical problems arise because the normalizing factor $p(X)$ is not known and because the likelihood $p(X|\beta^i)$ is usually not analytical making the problem analytically intractable. This can be solved by using sampling methods such as Markov Chain Monte Carlo (MCMC) (Wilkinson, 2007).

Once the posterior distribution is known, it is possible to obtain point estimates for the parameters. Quite similarly to ML methods, it is possible to obtain the Maximum *a Posteriori* (MAP) estimate of the parameters so that these correspond to the maximum posterior probability: $\hat{\beta}^i = \arg\max_b p(b|X)$. Nonetheless, as we will see in section 3.3, network inference problems are ill-defined and the posterior probability may be diffuse with many high-scoring models (Friedman et al., 2000) Hill et al. (2012b). One manner to solve this is to make use of Bayesian Model Averaging (BMA) which averages over the entire model space and allows to estimate the posterior probability of the inclusion of an edge.

$$p(e_{ij} = 1|X) = \sum_{\hat{\beta}^i_j | \hat{\beta}^i_j \neq 0} p(\hat{\beta}^i_j|X)$$

Similarly to the frequentist approach, it is possible to construct a solution network by defining a threshold t so that only edges with high posterior probability are retained.

$$a_{ij} = \begin{cases} 1 & \text{if } p(e_{ij} = 1|X) > t \\ 0 & \text{otherwise} \end{cases}$$

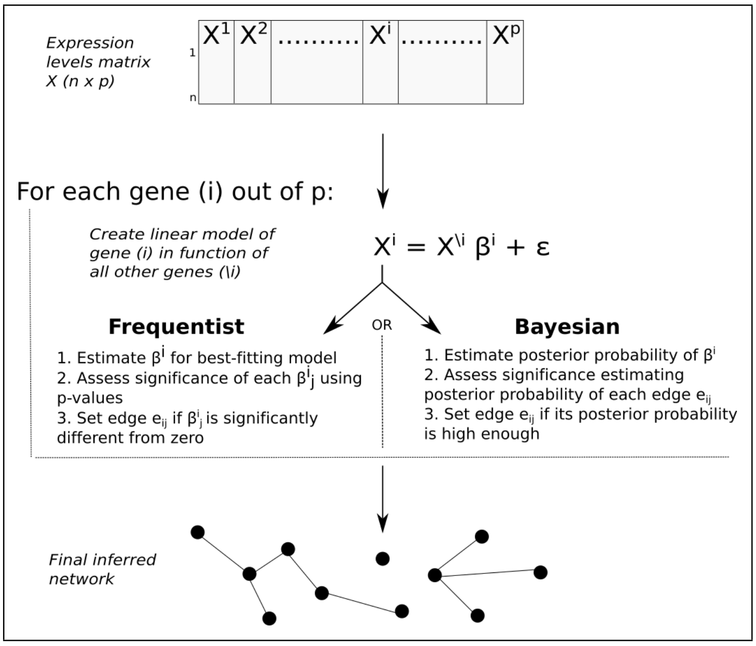
Expression levels matrix X (n x p)
1
n
X¹ X² .... Xⁱ .... Xᵖ
For each gene (i) out of p:
Create linear model of gene (i) in function of all other genes (\i)
Xⁱ = X\ⁱ βⁱ + ε
Frequentist
OR
Bayesian
1. Estimate βⁱ for best-fitting model
2. Assess significance of each βⁱⱼ using p-values
3. Set edge eᵢⱼ if βⁱⱼ is significantly different from zero
1. Estimate posterior probability of βⁱ
2. Assess significance estimating posterior probability of each edge eᵢⱼ
3. Set edge eᵢⱼ if its posterior probability is high enough
Final inferred network

Figure 1: Difference between frequentist and Bayesian approaches in network inference illustrated using linear models. The expression level of genes is measured and arranged as a matrix. For each gene, a linear model is constructed. This model can be fitted using a frequentist or a Bayesian approach. Results can then be summarized as a solution network.

3 Gene regulatory network inference

3.1 Definition

The goal of GRN inference is to determine the structure of a GRN using gene expression data obtained from high-throuput experiments such as microarrays or RNA Sequencing (RNAseq) through statistical and computational methods (Emmert-Streib et al., 2014). In order words, these algorithms try to estimate the adjacency matrix $A(p \times p)$ of a GRN using the matrix of expression levels of p genes found in n different samples $X(n \times p)$. These data can be acquired at steady-state to infer static networks or as a time-series to infer dynamic networks (Margolin and Califano, 2007). Although certain algorithms try to predict characteristics of edges, for example rate parameters or dissociation constants, this will not be considered in this review.

Expression levels need not be the only input. Information concerning the network perturbations can also be included. These can be acquired through gene knockouts, biochemical inhibitions and environmental change experiments (Stolovitzky et al., 2007). Supplementary data acquired from experiments Chromatin Immunoprecipitation (ChIP) or prior biological information such as Gene Ontology terms or protein sequence can also be incorporated. (Santra, 2014) (Studham et al., 2014) (Praveen and Fröhlich, 2013).

3.2 Motivation

Much of the work done in molecular biology in the 20^{th} century was concerned with the thorough enumeration and characterization of biomolecules. Notwithstanding, it is now clear that biological systems are not simply a collection of independent and isolated biomocules. Instead, these work together in intricate biological networks. Understanding these networks becomes thus essential in order to better grasp biological nature (Bruggeman and Westerhoff, 2007). Network inference is a step in this direction since its goal is the elucidation of their structure.

Interactions between biomolecules are traditionally inferred from *in vitro* equilibrium binding experiments. Nonetheless, these experiments are expensive, low-throughput and do not necessarily account for the conditions found *in vivo* (Dehmer et al., 2011). Network inference tries to circumvent this shortcoming by inferring networks using computational methods and data obtained from high-throughput experiments directly acquired from living systems. Given the ever-lowering price of high-throughput technologies such as RNAseq (Check Hayden, 2014) and the advantages of automated and computational network inference algorithms, research concerning network inference is very attractive.

It is important to note that the inferred network is not the final result. Indeed, they are used to solve different biological or medical questions. For example, one can compare physiological with diseased networks using different graph-theoretic tools, for example, to gain a better understanding of cancer (Emmert-Streib et al., 2014). GRNs can also be used to guide drug discovery (Madhamshettiwar et al., 2012) (Lecca and Priami, 2013). Finally, the inferred network can also be used to simulations in order to gain novel insight about the studied GRN (Kim et al., 2014). For more information concerning the applications of GRN, Emmert-Streib et al. (2014) and Barabási et al. (2011) provide a good introduction.

3.3 Difficulties

Network inference is an example of an inverse problem: given a limited amount of data, determine the data-generating model. Inverse problems are frequently ill posed problems, and network inference is no exception to this rule (Villaverde and Banga, 2014). Indeed, the data used for network-inference is high dimensional, in other words, it is composed of many features and a small sample size ($p \gg n$). Given this limited amount of data, it is not possible to identify a unique solution network. Instead, a set of solution networks will explain the data equally well. In order for this problem to become trackable, this requires reducing the search space by making assumptions about the biological network structure (*e.g.* sparseness and scale-freeness), using dimension reduction methods (*e.g.* variable subset selection, regularization or clustering), incorporating prior biological knowledge (*e.g.* ChIP data) or using supervised methods (De Smet and Marchal, 2010) (Marbach et al., 2010) (Linde et al., 2015).

Biological data such as those obtained from microarrays are noisy and not all factors are observed (Wilkinson, 2007). This makes it critical to assess the confidence we have in our results in the form of probabilities. This is handled as p-values in frequentist framework and posterior probabilities in the Bayesian framework. This also makes it desirable to be able to handle missing data in more elegant methods than simple deletion. Moreover, data are heterogeneous (*e.g.* concentrations, sequences, annotations, networks, ontologies) (Santra, 2014). It is, however, not always clear how all this information should be integrated in the learning algorithm.

Biological networks exhibit complex non-linear relationships (Wilkinson, 2007). A realistic model would thus have to be non-linear. However, adequate non-linear modelling is conceptually harder and requires more data than linear models (Linde et al., 2015). Moreover, it has been shown that perturbations around steady-state are linear justifying the linearity assumption (Margolin and Califano, 2007). For these reasons, linear models are often favoured (Werhli et al., 2006). Nonetheless, some recent non-linear frequentist and Bayesian methods were shown to outperform linear methods (Penfold et al., 2012) (Huynh-Thu et al., 2010).

3.4 Methods

3.4.1 Association measures

The first methods described for biological network inference were based on statistical association (Butte and Kohane, 2000). These algorithms access the connectivity of two vertices by measuring their pairwise statistical association using measures such as the linear Pearson's correlation coefficient and non-linear information-theoretic measures such as mutual information (Werhli et al., 2006) (Penfold and Wild, 2011). The resulting network is sometimes referred to a relevance network in the literature (Butte and Kohane, 2000). These methods correspond to a purely frequentist approach. Indeed, statistical association is estimated using pointwise estimators such as ML estimators. Furthermore, edge uncertainty is assessed using p-values. Because association does not imply a direct interaction, but can also indicate an indirect interaction or regulation by a common gene (Schäfer and Strimmer, 2005). Consequently, methods were developed to distinguish direct from indirect associations or confounding. These include using conditional association measures and the data inequality principle (Maetschke et al., 2014). Moreover, statistical association between two variables is symmetric and does not allow to determine the directionality of the interaction. To solve this problem, timed-lagged variants of association measures can be used (Zoppoli et al., 2010).

3.4.2 Graphical models

Graphical models refer to any model that makes uses of a graph to represent the interdependencies of variables. Here we will introduce two of the most widely used graphical models in network inference: BNs and GGMs. Both of these models can be inferred and analyzed using frequentist, Bayesian, or both approaches.

BNs are a succinct representation of a joint probability distribution in the form of a directed acyclic graph. Although they have a purely probabilistic definition, directed edges can represent causal relationships between vertices (Friedman et al., 2000) (Judea Pearl and Russell, 2003). Learning the structure of a BN from data requires the definition of a scoring function (*e.g.* posterior probability of the graph, mutual information) and a search algorithm meant to optimize the score (*e.g.* hill climbing, MCMC) (Friedman et al., 1997). BNs can be static or dynamic. Static BNs have the disadvantage of not being able to model feedback loops or self-regulation. Given the importance of such network motifs in certain biological networks, they cannot be expected to correctly model biological reality. This constraint is however relaxed for dynamic BNs (Readhead and Dudley, 2013). One advantage of BNs is that they are able to easily incorporate missing values in the learning process (Werhli et al., 2006). However, they most often require discretization of continuous values, such as gene expression levels, which results in a loss of information (Dehmer et al., 2011). Finally, it is important to note that BNs do not necessarily make explicit use of Bayesian statistics. For example, the structure of the network can be learned from a frequentist perspective by maximizing the mutual information and the not posterior probability of the graph (Wilkinson, 2007).

GGMs represent a joint probability distribution as an undirected graph. Gene expression levels are assumed to follow a multivariate Gaussian distribution with mean μ and covariance Σ. Although, the number of parameters is much smaller than in BNs, GGMs were shown to perform equally well for observational data (Werhli et al., 2006). Furthermore, they can be cyclic and can thus model feedback loops. Naively, the graph's structure can be learned by estimating the correlation matrix P and setting an edge between vertices v_i and v_j if $\rho_{ij} \neq 0$. However, as seen in 3.4.1, association does not imply a direct interaction. Because of this, structure is learned by estimating the partial correlation matrix given all other genes Π. This corresponds to the inverse of correlation matrix $\Pi = P^{-1}$ for multivariate Gaussian probability distributions. However, the correlation matrix Π is only invertible if there are more samples than features ($n > p$) (Schäfer and Strimmer, 2005). Obtaining reliable small-sample point estimates requires using regularization techniques such such as the Graphical LASSO (GLASSO) as in Friedman et al. (2008) or Li and Jackson (2015), or using the Moore–Penrose pseudoinverse and a bagging procedure as in Schäfer and Strimmer (2005). Both of these techniques allow incorporating prior knowledge through a frequentist or a Bayesian framework.

3.4.3 Regression

Regression methods transform the inference problem into a multiple regression problem for each gene in the network. These rely on the assumption that the expression level dynamics of one gene $\dot{X}^i$ is a function f_i of the expression levels of all other genes in the network $X = (X_1, \dots, X_p)$ (Margolin and Califano, 2007): $\dot{X}^i = f_i(X)$.

Assuming steady state, $\forall i \, \dot{X}^i = 0$, the former equation states that the expression level of each gene is a function of the expression level of all other genes in the network $X^i = f_i(X \setminus X^i) = f_i(X^{\setminus i})$ (Margolin and Califano, 2007). Most often, f_i will be a linear additive model with coefficients β^i with random centred Gaussian noise ϵ: $X^i = X^{\setminus i}\beta^i + \epsilon$. This corresponds to the equation we have studied in 2.2. It would most commonly be solved using ordinary least-squares. However, this assumes, among other things, that there are more samples than genes. This condition will rarely be met in network inference where $p \gg n$. Furthermore, given that biological networks are sparse, a solution to this problem should set a fair amount of coefficients β^i to zero. These concerns can be addressed by using variable subset selection (Hill et al., 2012b) or the LASSO (Wu et al., 2009) (Santra, 2014). Both of these methods have frequentist and Bayesian variants.

Although we have only discussed linear models, non-linear (Penfold et al., 2012) and tree-based regression models (Huynh-Thu et al., 2010) have also been developed and have shown some promising results. For example, the GENIE3 method described in Huynh-Thu et al. (2010) had the best performance in the fourth Dialogue on Reverse-Engineering Assessment and Methods (DREAM) competition (Greenfield et al., 2010).

4 Comparison of frequentist and Bayesian methods

4.1 Prediction accuracy

The DREAM provides a platform for the unbiased assessment of network-inference methods (Stolovitzky et al., 2007). Examples of DREAM competitions are Marbach et al. (2010) and Marbach et al. (2012). Briefly, in these studies, they provide the results of a benchmark on genome-wide networks of biological networks such as *E. coli*, and *in silico* networks. The methods they typically investigate are regression, correlation, information-theoretic and BNs.

The DREAM competitions showed that no individual method outperforms all others. Instead, the authors conclude that ensemble methods, that is aggregating predictions across many models, outperform individual methods. Moreover, in Marbach et al. (2010), the authors reported that there was no correlation between the method class used and the prediction accuracy obtained. In this context, it makes little sense to compare individual methods in terms of accuracy. Nonetheless, and somewhat contradictory, in Marbach et al. (2012), BNs, had below-average performance. They argued that this is because that the algorithms used to optimize the posterior distribution of the graph were too costly for the size of the networks. In any case, these limited results make it difficult make general conclusions concerning Bayesian methods in terms of prediction. It would have been interesting to see results of a simpler method making use of the Bayesian framework such as GGMs. Indeed, their performance is in many cases comparable to that of BNs (Werhli et al., 2006).

It seems important to point out the it is hard to actually assess the predictive power of Bayesian methods in the context of DREAM benchmarks. Indeed, no biological context or information is provided in these assessments. This impedes incorporating prior knowledge which can be argued as one of the main advantages of Bayesian methods. Nonetheless, there is clear evidence from other studies that prior information leads to better predictive power. This is true whether the approach is frequentist (Studham et al., 2014) or Bayesian (Santra, 2014) (Praveen and Fröhlich, 2013) (Mukherjee and Speed, 2008) (Young et al., 2014). To summarize, it does not seem possible to advocate for frequentist or Bayesian methods simply in terms of prediction accuracy. Hence, the choice of the statistical framework should depend on other factors.

4.2 Inference

Frequentist methods assume that a model's parameters are fixed. Simply put, frequentists assume that there is one unique value that is correct for a parameter. Bayesian methods, on the contrary, model their uncertainty concerning parameter values using probability distributions. In other words, certain parameters are more likely than others, but ultimately there is no unique parameter value that is correct. For network inference, the main parameter of interest is the adjacency matrix. Accordingly, a frequentist would say that there is only one true adjacency matrix. Consequently, s/he would try and estimate that parameter. On the contrary, a Bayesian would argue for a probability distribution of adjacency matrices and try to sample that distribution, for example, using MCMC.

The Bayesian approach seems more compelling. Indeed, it seems unclear what would the frequentist inferred network biologically represent. Biological networks are dynamic and adaptive; edges can appear or disappear with time. For instance, some interactions between genes occur only during certain time-frames or in response to certain environmental changes. Furthermore, network inference is an undetermined problem and many solutions explain equally well the data. In Bayesian terms, this means that the posterior distribution of the graph does not favour any specific graph, but instead is diffuse with several high-scoring networks. As Friedman et al. (2000) point out, what we should really be doing is analyzing a set of plausible inferred networks and determining features (*e.g.* edges, motifs) that are common to all of them. Such an analysis can be undertaken using the Bayesian framework by sampling graphs from the posterior distribution using algorithms such as MCMC. The Bayesian method that combines these different models is called BMA. The marginal posterior probability of a feature can then be computed as its probability across the space of all possible graphs (Hill et al., 2012a). Nonetheless, for practical purposes, a threshold can be set in order to obtain one unique network.

From the frequentist framework, a similar analysis can be carried out using the bootstrap procedure. Indeed, it has been shown that a bootstrap distribution is an approximative non-parametric, non-informative posterior distribution for a parameter (Hastie et al., 2009). The frequentist method that combines bootstraped models is called bootstrap aggregating or bagging. This strategy has been applied namely in Schäfer and Strimmer (2005). Nonetheless, this impedes the incorporation of prior biological data. Furthermore, in the frequentist approach, the credibility of a feature is most often accessed using p-values. These p-values represent the probability of the data given a null hypothesis. They are computed using theoretical null distributions or using permutation tests.

The meaning of posterior probability is arguably easier to understand than that of the often misunderstood p-value. Moreover, posterior probabilities provide more powerful inference (Bradley P. Carlin and Thomas A. Louis, 2008). Indeed, p-values, contrary to posterior probabilities, do not take into account an alternative hypothesis. They only provide evidence against the null hypothesis which may lead to many false positives. In a study concerning metabolic networks from Çakır et al. (2009), the authors show that false negatives (*i.e.* non-significant p-values) often correspond to weak interactions. Another objection to p-values is that, unlike posterior probabilities, they are not subject to the likelihood principle. This means that their value will depend on the experimental design. There is currently no standardized experiment for network inference, which may constitute a serious problem for frequentist network inference. Finally, p-values carry the now well know problem of multiple comparisons that the Bayesian posterior probability handles naturally.

4.3 Integration of prior knowledge

The data matrix used for GRN inference contains the expression levels for different genes found in different samples. However, it has been shown that the inclusion of additional data such as perturbation data, ChIP data or knowledge contained in databases such as Gene Ontology improves and speeds up the network-inference process by reducing the search-space (Bonneau, 2008).

Bayesian methods provide a strong theoretical framework to integrate prior biological knowledge. This is done through the prior probability distribution of the network. The Bayesian methodology allows to update the prior probability in light of new data to obtain the posterior distribution of the network. This general strategy can be used for network inference for feature selection problems (Santra, 2014) (Young et al., 2014), in penalized likelihood models such as the LASSO, and for the computation of the posterior probabilities of a network (Mukherjee and Speed, 2008) (Praveen and Fröhlich, 2013). Nonetheless, it is not always clear how what information should be used or how prior probabilities should be constructed and different authors argue for different strategies. Although this may first appear as having the caveat of subjectivism, it is always possible to resort to empirical Bayes as in Schäfer and Strimmer (2005). Nonetheless, in certain cases, it could be argued that the inclusion of prior is not justified. An example would be the inference of a cancer network using priors derived from physiological networks. Nonetheless, there is evidence that mis-specified priors including incorrect information regarding individual edges can still result in an increased accuracy when compared to a flat prior (Mukherjee and Speed, 2008).

Prior information can also be incorporated to frequentist methods through penalized likelihood models such as the LASSO (Studham et al., 2014). This is usually done by setting weights to the different penalty parameters that have to be determined using cross-validation or criteria similar to the Bayesian information criterion (Wang, 2012). Although, this method does not solve the problems describe above, it does have the advantage of being conceptually simpler to understand and to implement efficiently.

4.4 Robustness to noise, small sample size and missing data

While frequentist assume that the underlying parameters are fixed, Bayesians assume that parameters are unknown and describe their uncertainty probabilistically. Because of this conceptual difference, Bayesian methods are said to be less sensitive to noise then frequentist methods (Villaverde and Banga, 2014). Moreover, Bayesian methods are known to be more robust to small sample sizes because they do not rely on asymptotic approximations. These advantages seem very important because, in biological sciences, stochastiscity plays in an important role and small samples are the norm (Wilkinson, 2007). The impact of noise on network inference has been studied in Nagarajan and Scutari (2013). Using algorithms based on pairwise association measures and BNs, the authors found that noise can have a non-trivial impact on results and can affect biological conclusions. Unfortunately, they did not compare the robustness to noise of both models.

Up until now, the studies we have cited are either frequentist or Bayesian. Consequently, we have been making a Manichean distinction between the two approaches. Nonetheless, Bayarri and Berger (2004) argue that there are situations where a joint frequentist-Bayesian approach is required. Although this study concerns inference of rate parameters in a GRN and not inference of its structure, it is still interesting to note that such a joint approach has been applied in Raue et al. (2013). The authors identify that, from the frequentist's perspective, insufficient data leads to lack of identifiability. This leads to non convergent MCMCs, and thus make Bayesian methods non-applicable. As a solution, they use frequentist profile-likelihood and experimental design to identify and resolve non-identifiability. Once this is resolved, they apply their MCMC-based Bayesian method to estimate rate parameters with success.

4.5 Computational time

Computation of posterior probability distribution is often only possible through expensive sampling methods such as MCMC. On the contrary, computation of association measures and regression coefficient do not require numerical integration and are thus much faster. One may thus question the applicability of Bayesian methods to genome-wide network-inference. However, in Young et al. (2014), the authors implement a Dynamic BN using Bayesian Model Averaging and successfully apply it to genome-wide time-series data with speed and performance comparable to the LASSO and association measure based methods.

5 Conclusion

Our comparison of frequentist and Bayesian methods for GRN inference has highlighted advantages and disadvantages of these different approaches. The frequentist approach has the undeniable asset of simplicity. For example, in absence of prior knowledge, they do not require defining a flat prior probability which is non-trivial in high-dimensional settings. Moreover, frequentist procedures are also much easier to implement than Bayesian methods that require expensive sampling algorithms with fine tuning and that may never converge. Nonetheless, Bayesian methods have conceptual but also practical advantages. They naturally allow incorporating prior knowledge by Bayesian updating. They provide more powerful and interpretable inference by yielding summary statistics such as posterior probabilities for the presence of edges. Finally, they do not rely on p-values for assessing uncertainty.

To paraphrase George Box, "statisticians, like artists, have the bad habit of falling in love with their models". Notwithstanding, in the field of GRN inference, no method dominates any other in terms of prediction accuracy. The choice of the methodology should depend on other statistical considerations, on the scientific question, and on the biological context. For example, in absence of prior biological knowledge, no matter how attractive Bayesian methodologies may be, they may ultimately be unjustifiably complicated when compared to frequentist approaches. Finally, it should be emphasized that is it always a good idea to try out different methods and to compare results.

Prior knowledge has been shown to improve prediction accuracy. Nonetheless, no consensus has yet been reached on how prior should be constructed and what information should be used. Future research in the area should concern itself in creating a unifying framework. Additionally, one critic made to the DREAM competition is that it should not become an abstract exercise in absence of all biological context (Stolovitzky et al., 2007). Future DREAM competitions should objectively assess the improvement of accuracy by the incorporation of prior knowledge. Finally, it seems that experimental design has received little attention in the literature. Given the limited amount of data in GRN inference, research to optimize data collection appears pivotal. The theory behind the design of experiments has been most extensively studied with the frequentist framework, namely by Fisher. Future strategies to network inference could thus develop a frequentist-Bayesian method similar to Raue et al. (2013). Careful frequentist experimental design allows optimization of the collection of data, while network inference with a Bayesian method allows easily incorporating prior biological knowledge and provides powerful inference.

References

Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, January 2011. ISSN 1471-0056. doi: 10.1038/nrg2918.

M. J. Bayarri and J. O. Berger. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*, 19(1):58–80, February 2004. ISSN 0883-4237, 2168-8745. doi: 10.1214/088342304000000116.

Richard Bonneau. Learning biological networks: from modules to dynamics. *Nature Chemical Biology*, 4(11):658–664, November 2008. ISSN 1552-4450. doi: 10.1038/nchembio.122.

Richard Bonneau, David J. Reiss, Paul Shannon, Marc Facciotti, Leroy Hood, Nitin S. Baliga, and Vesteinn Thorsson. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biology*, 7(5):1–16, 2006. ISSN 1474-760X. doi: 10.1186/gb-2006-7-5-r36.

Bradley P. Carlin and Thomas A. Louis. *Bayesian Methods for Data Analysis, Third Edition.* Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, June 2008. ISBN 978-1-58488-697-6.

Frank J. Bruggeman and Hans V. Westerhoff. The nature of systems biology. *Trends in Microbiology*, 15(1):45–50, January 2007. ISSN 0966-842X. doi: 10.1016/j.tim.2006.11.003.

A. J. Butte and I. S. Kohane. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 418–429, 2000. ISSN 2335-6936.

Erika Check Hayden. Is the $1,000 genome for real? *Nature*, January 2014. ISSN 1476-4687. doi: 10.1038/nature.2014.14530.

Riet De Smet and Kathleen Marchal. Advantages and limitations of current network inference methods. *Nature Reviews Microbiology*, 8(10):717–729, October 2010. ISSN 1740-1526. doi: 10.1038/nrmicro2419.

Matthias Dehmer, Frank Emmert-Streib, Armin Graber, and Armindo Salvador. *Applied Statistics for Network Biology: Methods in Systems Biology - Matthias Dehmer, Frank Emmert-Streib, Armin Graber, et al.* 2011.

Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Molecular Medicine*, 2:38, 2014. doi: 10.3389/fcell.2014.00038.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, July 2008. ISSN 1465-4644, 1468-4357. doi: 10.1093/biostatistics/kxm045.

Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian Network Classifiers. *Machine Learning*, 29(2-3):131–163, November 1997. ISSN 0885-6125, 1573-0565. doi: 10.1023/A:1007465528199.

Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using Bayesian Networks to Analyze Expression Data. *Journal of Computational Biology*, 7(3-4):601–620, August 2000. ISSN 1066-5277. doi: 10.1089/106652700750050961.

Alex Greenfield, Aviv Madar, Harry Ostrer, and Richard Bonneau. DREAM4: Combining Genetic and Dynamic Information to Identify Biological Networks and Dynamical Models. *PLoS ONE*, 5(10), October 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0013397.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2009. ISBN 978-0-387-84857-0 978-0-387-84858-7.

Steven M. Hill, Yiling Lu, Jennifer Molina, Laura M. Heiser, Paul T. Spellman, Terence P. Speed, Joe W. Gray, Gordon B. Mills, and Sach Mukherjee. Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. *Bioinformatics*, 28(21):2804–2810, November 2012a. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts514.

Steven M Hill, Richard M Neve, Nora Bayani, Wen-Lin Kuo, Safiyyah Ziyad, Paul T Spellman, Joe W Gray, and Sach Mukherjee. Integrating biological knowledge into variable selection: an empirical Bayes approach with an application in cancer biology. *BMC Bioinformatics*, 13:94, May 2012b. ISSN 1471-2105. doi: 10.1186/1471-2105-13-94.

Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, 5(9), September 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0012776.

Judea Pearl and Stuart Russell. Bayesian Networks. In *The Handbook of Brain Theory and Neural Networks*, pages 157–160. MIT Press, 2003. ISBN 978-0-262-01197-6.

Yongsoo Kim, Seungmin Han, Seungjin Choi, and Daehee Hwang. Inference of dynamic networks using time-course data. *Briefings in Bioinformatics*, 15(2):212–228, March 2014. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbt028.

Paola Lecca and Corrado Priami. Biological network inference for drug discovery. *Drug Discovery Today*, 18(5–6):256–264, March 2013. ISSN 1359-6446. doi: 10.1016/j.drudis.2012.11.001.

Yupeng Li and Scott A. Jackson. Gene Network Reconstruction by Integration of Prior Biological Knowledge. *G3: Genes|Genomes|Genetics*, 5(6): 1075–1079, June 2015. ISSN , 2160-1836. doi: 10.1534/g3.115.018127.

Jörg Linde, Sylvie Schulze, Sebastian G. Henkel, and Reinhard Guthke. Data- and knowledge-based modeling of gene regulatory networks: an update. *EXCLI Journal*, 14:346–378, March 2015. ISSN 1611-2156. doi: 10.17179/excli2015-168.

Piyush B. Madhamshettiwar, Stefan R. Maetschke, Melissa J. Davis, Antonio Reverter, and Mark A. Ragan. Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Medicine*, 4:41, 2012. ISSN 1756-994X. doi: 10.1186/gm340.

Stefan R. Maetschke, Piyush B. Madhamshettiwar, Melissa J. Davis, and Mark A. Ragan. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Briefings in Bioinformatics*, 15(2):195–211, March 2014. ISSN 1467-5463. doi: 10.1093/bib/bbt034.

Daniel Marbach, Robert J. Prill, Thomas Schaffter, Claudio Mattiussi, Dario Floreano, and Gustavo Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*, 107(14):6286–6291, April 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0913357107.

Daniel Marbach, James C. Costello, Robert Küffner, Nicole M. Vega, Robert J. Prill, Diogo M. Camacho, Kyle R. Allison, The DREAM5 Consortium, Manolis Kellis, James J. Collins, and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8): 796–804, August 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2016.

Adam A. Margolin and Andrea Califano. Theory and Limitations of Genetic Network Inference from Microarray Data. *Annals of the New York Academy of Sciences*, 1115(1):51–72, December 2007. ISSN 1749-6632. doi: 10.1196/annals.1407.019.

Sach Mukherjee and Terence P. Speed. Network inference using informative priors. *Proceedings of the National Academy of Sciences*, 105 (38):14313–14318, September 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0802272105.

Radhakrishnan Nagarajan and Marco Scutari. Impact of Noise on Molecular Network Inference. *PLoS ONE*, 8(12), December 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0080735.

Christopher A. Penfold and David L. Wild. How to infer gene networks from expression profiles, revisited. *Interface Focus*, 1(6):857–870, December 2011. ISSN 2042-8898. doi: 10.1098/rsfs.2011.0053.

Christopher A. Penfold, Vicky Buchanan-Wollaston, Katherine J. Denby, and David L. Wild. Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics*, 28(12):i233–i241, June 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts222.

Paurush Praveen and Holger Fröhlich. Boosting Probabilistic Graphical Model Inference by Incorporating Prior Knowledge from Multiple Sources. *PLoS ONE*, 8(6), June 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0067410.

Andreas Raue, Clemens Kreutz, Fabian Joachim Theis, and Jens Timmer. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Phil. Trans. R. Soc. A*, 371(1984):20110544, February 2013. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2011.0544.

Ben Readhead and Joel Dudley. Translational Bioinformatics Approaches to Drug Development. *Advances in Wound Care*, 2(9):470–489, November 2013. ISSN 2162-1918. doi: 10.1089/wound.2012.0422.

Tapesh Santra. A Bayesian Framework That Integrates Heterogeneous Data for Inferring Gene Regulatory Networks. *Frontiers in Bioengineering and Biotechnology*, 2, May 2014. ISSN 2296-4185. doi: 10.3389/fbioe.2014.00013.

Juliane Schäfer and Korbinian Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6): 754–764, March 2005. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bti062.

Gustavo Stolovitzky, Don Monroe, and Andrea Califano. Dialogue on Reverse-Engineering Assessment and Methods. *Annals of the New York Academy of Sciences*, 1115(1):1–22, December 2007. ISSN 1749-6632. doi: 10.1196/annals.1407.021.

Matthew E. Studham, Andreas Tjärnberg, Torbjörn E.M. Nordling, Sven Nelander, and Erik L. L. Sonnhammer. Functional association networks as priors for gene regulatory network inference. *Bioinformatics*, 30(12):i130–i138, June 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu285.

Alejandro F. Villaverde and Julio R. Banga. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of The Royal Society Interface*, 11(91):20130505, February 2014. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2013.0505.

Hao Wang. Bayesian Graphical Lasso Models and Efficient Posterior Computation. *Bayesian Analysis*, 7(4):867–886, December 2012. ISSN 1936-0975, 1931-6690. doi: 10.1214/12-BA729.

Adriano V. Werhli, Marco Grzegorczyk, and Dirk Husmeier. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics*, 22(20):2523–2531, October 2006. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btl391.

Darren J. Wilkinson. Bayesian methods in bioinformatics and computational systems biology. *Briefings in Bioinformatics*, 8(2):109–116, March 2007. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbm007.

Tong Tong Wu, Yi Fang Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–721, March 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp041.

William Chad Young, Adrian E. Raftery, and Ka Yee Yeung. Fast Bayesian inference for gene regulatory networks using ScanBMA. *BMC Systems Biology*, 8:47, 2014. ISSN 1752-0509. doi: 10.1186/1752-0509-8-47.

Pietro Zoppoli, Sandro Morganella, and Michele Ceccarelli. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics*, 11:154, 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-154.

Tunahan Çakır, Margriet M. W. B. Hendriks, Johan A. Westerhuis, and Age K. Smilde. Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics*, 5(3):318–329, February 2009. ISSN 1573-3882, 1573-3890. doi: 10.1007/s11306-009-0156-4.