
Dmitry Rusakov, Dan Geiger
{rusakov,dang}@cs.technion.ac.il
Computer Science Department
Technion, Israel.

Abstract

We investigate parameter priors for discrete DAG models. It was shown in previous works that a Dirichlet prior on the parameters of a discrete DAG model is inevitable assuming global and local parameter independence for all possible complete DAG structures. A similar result for Gaussian DAG models hinted that the assumption of local independence may be redundant.

Herein, we prove that the local independence assumption is necessary in order to dictate a Dirichlet prior on the parameters of a discrete DAG model. We explicate the minimal set of assumptions needed to dictate a Dirichlet prior, and we derive the functional form of prior distributions that arise under the global independence assumption alone.

1 Introduction

A directed graphical model is a representation of a family of joint probability distributions for a collection of random variables via a Directed Acyclic Graph (DAG). In particular, each node in the DAG corresponds to a random variable, and the lack of an edge between two nodes represents a conditional independence assumption. A specific joint probability distribution can be represented by a given directed graphical model by specifying the values for the set of associated parameters. The DAG along with such a distribution is called a Bayesian network. Graphical models and Bayesian networks have been extensively studied in AI, Statistics, Machine learning, and in many application areas [3-9,12].

Bayesian networks encode a probability distribution with a manageable number of parameters (due to the factorization introduced by underlying graph), thus reducing the complexity of the representation and reducing the complexity of decision making based on this distribution. Bayesian networks are also useful when constructed directly from expert knowledge because they introduce cause-effect relationships that are intuitive to human experts. These features made Bayesian networks a premier tool for representing probabilistic knowledge and reasoning under uncertainty.

In this paper we focus on learning—the process of updating both the parameters and the structure of a Bayesian network based on data. To compute goodness-of-fit of data to a network structure in a closed form, researchers have made a number of assumptions. Among them, global and local parameter independence for all network structures, Dirichlet distribution on network parameters, and some other assumptions [3]. It was later shown that the assumption of global and local parameter independence for all nodes in every complete network structure dictates that the only possible prior parameter distribution for discrete DAG models is a Dirichlet prior [7, 9].

In contrast, in a subsequent work, it was shown that for Gaussian DAG models, which consist of a recursive set of linear regression models, global independence alone dictates that the only feasible parameter prior is the Normal-Wishart distribution, assuming models with at least three nodes [6]. It was thus natural to hypothesize that the proofs for discrete and continuous case can be unified and, as a result, the assumption of local independence will turn out to be redundant also in the characterization of the Dirichlet distribution.

This work gives a negative answer to the question of similarity between the discrete and continuous cases. In other words, it shows that, while global independence dictates a Normal-Wishart prior for Gaussian DAG models with more than 3 nodes, global independence alone does not dictate a Dirichlet prior for discrete DAG models with more than 3 nodes. We provide a minimal set of assumptions needed to dictate a Dirichlet prior
and, in addition, we specify the class of discrete probability distributions, which is larger than the Dirichlet family, that arise under global independence assumption alone via a solution of a new set of functional equations.

2 DAG Models

A Directed Acyclic Graphical (DAG) model \( m \triangleq m(s, F_s) \) for a set of variables \( X = \{X_1, \ldots, X_n\} \) each associated with a set of possible values \( D_i = \{d_{i1}, \ldots, d_{ni}\} \), respectively, is a set of joint probability distributions for \( D = D_1 \times \ldots \times D_n \) specified via two components: a structure \( s \) and a set of local distribution families \( F_s \). The structure \( s \) for \( X \) is a directed graph with no directed cycles (i.e., a DAG) having for every variable \( X_i \) in \( X \) a node labeled \( X_i \). We denote the parents of \( X_i \) by \( Pa_i \). The structure \( s \) represents the set of conditional independence assertions, and only these conditional independence assertions, which are implied by a factorization of a joint distribution for \( X \) given by \( p(x) = \prod_{i=1}^n p(x_i|Pa_i^s) \), where \( x \) is a value for \( X \) (an n-tuple), \( x_i \) is a value for \( X_i \) and \( Pa_i^s \) is a value for \( Pa_i \). When \( x_i \) has no incoming arcs in \( s \) (no parents), \( p(x_i|Pa_i^s) \) stands for \( p(x_i) \).

The local distributions are the \( n \) conditional and marginal distributions that constitute the factorization of \( p(x) \). Each such distribution belongs to the specified family of allowable probability distributions \( F_s \), which depends on a finite set of numerical parameters \( \theta_m \in \Theta_m \subseteq \mathbb{R}^k \) (a parametric family). The parameters \( \theta_m \) for a local distribution is a set of real numbers that completely determine the functional form of \( p(x_i|pa_i^s) \).

A DAG model is often called a Bayesian network, although the later name sometimes refers to a specific joint probability distributions that factorizes according to a DAG, and not, as we mean herein, a set of joint probability distributions each factorizing according to the same DAG. A DAG model is complete if it has no missing arcs. Note that any two complete DAG models for \( X \) encode the same set of conditional independence assertion, namely none.

2.1 Parameter Priors for DAG Models

In the Bayesian approach for learning of Bayesian networks from data, the parameters \( \theta_m \) and the model hypothesis \( m \) are uncertain but the parametric families are known. To compute the posterior probability of a model hypothesis \( m \) given data \( d = \{x_1, \ldots, x_m\} \) in closed form one must access an a priori probability distribution for the parametric family \( \Theta_m \). We now present a set of assumptions that simplify the assessment of parameter priors for DAG models. These assumptions follow the assumptions presented in [6] and are formulated in spirit of calculus on manifolds (see Chapter 2 in [11]).

Let \( P_m \) denote the set of all joint distributions on \( X = \{X_1, \ldots, X_n\} \) which can be represented by model \( m \) and let \( p_m \) denote a function that defines a specific joint probability distribution according to parameters \( \theta_m \in \Theta_m \), i.e., \( p_m : \Theta_m \rightarrow P_m \).

Assumption 1 (Unique representation) For any DAG model \( m \) the function \( p_m \) is one-to-one.

The corollary of this assumption is the existence of one-to-one inverse function \( \theta_m \triangleq p_m^{-1} : P_m \rightarrow \Theta_m \), which can be viewed as a coordinate function for \( P_m \). (We allow the common abuse of notation here, denoting by \( \theta_m \) both the coordinate function and the coordinates themselves).

Assumption 2 (Complete model equivalence) If \( m_1 = m_1(s_1, F_{s_1}) \) and \( m_2 = m_2(s_2, F_{s_2}) \) are complete DAG models for \( X \) then \( P_{m_1} = P_{m_2} \).

The corollary of the Assumptions 1 and 2 is that for every two complete DAG models \( m_1 \) and \( m_2 \) for \( X \) there exists a one-to-one mapping \( f_{12} \triangleq p_{m_2}^{-1} \circ p_{m_1} \) between the parameters \( \Theta_{m_1} \) of \( m_1 \) and the parameters \( \Theta_{m_2} \) of \( m_2 \) such that \( p_{m_1}(\theta_{m_1}) = p_{m_2}(\theta_{m_2}) \) where \( \theta_{m_2} = f_{12}(\theta_{m_1}) \).

Assumption 3 (Regularity) For any two complete DAG models \( m_1 \) and \( m_2 \) the Jacobian \( |\partial \theta_{m_1} / \partial \theta_{m_2}| \) exists and is non-zero for all values of \( \Theta_{m_1} \) (for one-to-one mapping \( \theta_{m_2} = f_{12}(\theta_{m_1}) \)).
Assumptions 1-3 imply that for any \( m_1 \) and \( m_2 \) and for any \( p_1 \) pdf on \( \Theta_{m_1} \), the corresponding pdf on \( \Theta_{m_2} \) is:

\[
p_2(\theta_{m_2}) = \left( \frac{\partial \theta_{m_2}}{\partial \theta_{m_1}} \right)^{-1}(\theta_{m_2})p_1(\theta_{m_1}) \text{ where } \theta_{m_1} = f_{12}^{-1}(\theta_{m_2}).
\]

From now on, we will restrict our discussion for the discrete DAG models, where local distributions \( p(x_i | \text{pa}_i) \) are specified by multinomial parameters \( \theta_i = \{\theta_i | \text{pa}_i | x_i \in \{d_i^1, \ldots, d_i^{D_i-1}\}, \text{pa}_i \in D \} \). We assume that \( \theta_i \)'s do not overlap, and \( \theta_m \) is a concatenation of \( \theta_i \)'s, i.e. \( \theta_m = (\theta_i^1, \theta_i^2, \ldots, \theta_i^n) \). Note, that this assumption in no sense restricts the class of models under discussion, since if two local distributions depend on the same parameter it can be represented by two different parameters that are always equal (by specifying a proper \( \theta_m \)). Assumptions 1-3 naturally hold for discrete DAG models as defined above and the parametric families for all complete models that represent all possible distributions for \( X \) are the same, namely \( \Theta = [0,1]^{2D-1} \), where \( D = \prod_{i=1}^n |D_i| \).

In Bayesian framework, we suppose that there is some a priori distribution on natural parameters \( p(\Theta_X) \) that induces the distributions of local parameters for each model \( p(\theta_m | m) \) via change of parameters formula. We explicitly assume the regularity of parameter distributions.

Assumption 4 (Parameter Priors Regularity) The probability distribution functions (p.d.f.) on natural parameters and corresponding p.d.f.'s on model parameters are everywhere positive and twice differentiable.

So far, we have not made specific assumptions on the form of a parameter distribution \( p(\theta_m | m) \). The specific form of prior distribution can be either assumed directly, e.g. Dirichlet, or indirectly, by requiring some properties from prior distribution. Two common properties that are requested from prior distributions are global and local parameter independence.

Definition Parameters \( \theta_m \) of a DAG model \( m \) are said to be globally independent if \( \{\theta_i^m\}_{i=1}^n \) are mutually independent, i.e. \( p(\theta_i^m | m) = \prod_{i=1}^n p(\theta_i^m | m) \).

Definition Parameters \( \theta_{m_i} \) of node \( X_i \) of a DAG model \( m(s, F_s) \) are said to be locally independent if subsets \( \theta_{X_i | \text{pa}_i} = \{\theta_{x_i | \text{pa}_i} | x_i \in \{d_j^1, \ldots, d_j^{D_j-1}\} \} \) of \( \theta_i \) are mutually independent, i.e. \( p(\theta_i^m | m) = \prod_{i=1}^n p(\theta_i^m | m) \).

Heckerman, Geiger and Chickering [9] prove the following theorem that links between parameter independence and Dirichlet distribution:

Theorem 1 (Heckerman, Geiger & Chickering, 1995) Let \( m_1 \) and \( m_2 \) be two complete discrete DAG models for \( X \) with variable orderings \( (X_1, \ldots, X_n) \) and \( (X_n, X_1, \ldots, X_{n-1}) \), respectively. If parameters of \( m_1 \) and \( m_2 \) satisfy global independence and local independence (for all nodes) then \( p(\theta_X) \) is Dirichlet, where \( \theta_X \) are natural multinomial parameters for \( X \).

It was speculated (in [7] and [6]) that Dirichlet distribution may arise even under a weaker set of conditions, probably under the assumption of global independence for all complete networks, similarly to the result for Gaussian DAG models [6].

In this paper we investigate prior distributions for discrete DAG models that satisfy global independence requirement, and show that this class is strictly larger than class of Dirichlet distributions for all discreet models. In addition, we specify the minimal set of global and local independence assumptions that is needed to ensure that prior distributions are indeed Dirichlet.

3 Two Node Networks

We commence by analyzing in detail prior distribution for parameters of complete two node network. The results and techniques developed in this section will be the basis for advanced derivations in multiple node networks.

Consider the following complete two node network, as shown on Figure 1, with variables \( X, Y \) having \( n \) and \( k \) states accordingly. Since this network is complete it is capable of describing any multinomial distribution of two random variables. Any multinomial distribution which is naturally described by a set of parameters \( \{\theta_{x_i, y_j}\}_{i=1}^{n} \rightleftharpoons_{j=1}^{k} \rightleftharpoons_{1}^{n} \) (to be denoted shortly by \( \{\theta_{ij}\} \)) that sum to unity can be described by this
network by specifying $\theta_{X,i} = \sum_{j=1}^{k} \theta_{ij}$ and $\theta_{Y,j|x_i} = \theta_{ij}/\theta_{X,i}$ (commonly denoted as $\theta_i$ and $\theta_{jk}$) for $1 \leq i \leq n$ and $1 \leq j \leq k$.

We are interested in finding a functional form of a prior distributions $p(\theta_{ij})$ that satisfy a global independence assumption for all complete network for $\{X, Y\}$, namely $X \rightarrow Y$ (shown above) and $X \leftarrow Y$. Thus, such distributions should satisfy the following two functional equations, that describe global independence assumption:

$$
\begin{align*}
\frac{p(\theta_{ij})}{g(\theta_{ij})} &= J_1^{-1} f_1(\theta_i) \quad & \text{for } 1 \leq i \leq n \\
\frac{p(\theta_{ij})}{g(\theta_{ij})} &= J_2^{-1} f_2(\theta_j) \quad & \text{for } 1 \leq j \leq k
\end{align*}
$$

(1)

where $J_1, J_2$ are appropriate Jacobians and $\theta_{ij}, \theta_{jk}$ are defined similarly to $\theta_i$ and $\theta_{jk}$.

**Theorem 2** Any distribution of $\{\theta_{ij}\}$ that satisfies global independence assumption for all complete DAG models for $\{X, Y\}$, i.e. satisfies Equation 1, is of the form

$$
p(\theta_{ij}) = C \prod_{i=1}^{n} \prod_{j=1}^{k} \theta_{ij}^{\alpha_{ij}} H \left( \left\{ \frac{\theta_{ij}}{\theta_{i+1,j}} \theta_{i+1,j+1} \right\} \right) \quad \left( \frac{\theta_{i+1,j}}{\theta_{i,j+1}} \theta_{i,j+1} \right)
$$

(2)

where $\alpha_{ij}$ are arbitrary constants, $H()$ is an arbitrary twice-differentiable function of $(n-1)(k-1)$ variables, and $C$ is a normalization constant.

The proof of this theorem is based on the direct solution of functional equations 1. The general approach is given in the following subsection.

Theorem 2 implies that for two node discrete DAG models global independence assumption alone does not guarantee the Dirichlet distribution of priors. In Section 4 we will prove the similar result for all discrete DAG models.

### 3.1 The Functional Equation

In this subsection we present the functional equation which implies from Equation 1 and is the basis for the proof of Theorem 2.

Consider the following two sets of variables $\{y_{i}\} \quad 1 \leq i \leq n - 1 \}$ and $\{z_{ji}\} \quad 1 \leq i \leq n, 1 \leq j \leq k - 1 \}$. The domain of each of these variables is $(0, 1)$. These sets correspond to the sets $\{\theta_i\}$ and $\{\theta_{jk}\}$ of multinomial parameters discussed above but we shall not make a formal use of this correspondence.

We define,

$$
\begin{align*}
y_i &= 1 - \sum_{i=1}^{n-1} y_i \\
z_{ji} &= 1 - \sum_{j=1}^{k-1} z_{ji}, \quad 1 \leq i \leq n \\
x_j &= \sum_{i=1}^{n} z_{ji} y_i, \quad 1 \leq j \leq k \\
w_{ji} &= \sum_{i=1}^{n} z_{ji} y_i, \quad 1 \leq j \leq k, \quad 1 \leq i \leq n.
\end{align*}
$$

(3)

Note that $x_k = 1 - \sum_{j=1}^{k-1} x_j$ and $w_{jn} = 1 - \sum_{i=1}^{n-1} w_{ji}$ (for $1 \leq j \leq k$). Here, $\{x_j\}$ corresponds to $\{\theta_{j}\}$ and $\{w_{ji}\}$ corresponds to $\{\theta_{ij}\}$.

Finally, we let,

$$
\begin{align*}
Y &= (y_1, \ldots, y_{n-1}), & Z_t &= (z_{1,i}, \ldots, z_{k-1,i}), \\
X &= (x_1, \ldots, x_{k-1}), & W_j &= (w_{j,1}, \ldots, w_{j,n-1}) \\
Z &= (Z_1, \ldots, Z_n), & W &= (W_1, \ldots, W_k)
\end{align*}
$$

(4)

Thus, for example,

$$g(Z) = g(z_{1,1}, \ldots, z_{k-1,1}, \ldots, z_{1,i}, \ldots, z_{k-1,i}, \ldots, z_{1,n}, \ldots, z_{k-1,n})$$
Now let $Z_i = \left( \frac{w_{1,i}, \ldots, w_{n,i}}{y_i}, \ldots, \frac{w_{n-1,i}}{y_i}, y_i \right)$, $1 \leq i \leq n$. Consequently, we obtain the equation

$$F(Y)g(Z) = G(X)f(W)$$

which is identical to Equation 5 except that now $X$ and $W$ are free variables rather than $Y$ and $Z$. This dual representation of Equation 5 will be used in the derivation of its solution.

The solution of Equation 5, which is presented in Appendix A.2, uses Assumption 4 and is based on the technique of reduction of functional equation to partial differential equations ([1], page 324).

4 Multiple Node Networks: Globally Independent Parameters

Consider a complete DAG models of $n$ discrete nodes: $X = X_1, \ldots, X_n$, each having $|D_1|, \ldots, |D_n|$ values respectively. Such models provide a various ways (according factorization given by network structure) to describe a set of natural multinomial parameters $\Theta_X = \{\theta_{x_1,\ldots,x_n} | (x_1, \ldots, x_n) \in D \}$. The Bayesian networks corresponding to some DAG model $m(s,F_s)$ can be specified by a set of multinomial parameters $\theta_m = \{\theta_{[\text{pa}_i]} | i \}$ which describe the conditional distribution of $X_i$ given its parents $\text{pa}_i$ in the network structure $s$. In this section we would like to investigate prior the prior distributions on $\{\theta_{x_1,\ldots,x_n}\}$ (and thus on $\{\theta_{[\text{pa}_i]}\}$) that satisfy the global independence assumption for complete network structures, i.e.

**Assumption 5 (Global Parameter Independence)** For every complete DAG model $m(s,F_s)$ for $X$, $p(\theta_m | m) = \prod_{i=1}^{n} p(\theta_{m_i} | m)$ where $\theta_{m_i} = \{\theta_{[\text{pa}_i]}\}$

We say that $\Theta_X$ are globally independent if global parameters independence holds for all complete DAG models for $X$.

The Assumption 5 means that $p(\Theta_X)$ must satisfy the following $n!$ functional equations:

$$p(\Theta_X) = J_l^{-1} \prod_{j=1}^{n} f_{l,j}(\{\theta_{x_{i_1},\ldots,x_{i_{l-1}}}\}), \text{ s.t. } l = \langle i_1, \ldots, i_n \rangle \text{ is a permutation on } \langle 1, \ldots, n \rangle$$

where $f_{l,j}(\cdot)$ are measurable functions, $\{\theta_{x_{i_1},\ldots,x_{i_{l-1}}}\} \triangleq \{\theta_{x_{i_1}} | x_{i_1}, \ldots, x_{i_{l-1}} \}$ are expressed in terms of $\Theta_X$ and $J_l$ denotes the Jacobian of transformation from natural parameters to the...
parameters of the complete Bayesian network with topological order of nodes specified by $I$. These Jacobians can be 'hidden' inside $f_{I,j}$, since they are a function of $\{\theta_{x_{i_{j_1}}x_{i_{j_2}} \ldots x_{i_{j_{m_{-1}}}}}\}$, namely (using the result from [9]):

$$J_I = \prod_{j=1}^{n-1} \prod_{x_{i_{j_1}}, \ldots, x_{i_{j_{m_{-1}}}}} [\theta_{x_{i_{j_1}}x_{i_{j_2}} \ldots x_{i_{j_{m_{-1}}}}}] \left[\prod_{r \not\in I} |D_r|^{-1}\right]^{-1} \tag{9}$$

In order to solve Equation 8 we make use of the solution for two node network as given by Theorem 2. Consider a two discrete random variables $Y_i = \{Y_i, Y\}$, where $Y_i = X_i$ and $Y = X_1 \times \ldots \times X_{i-1} \times X_{i+1} \times \ldots \times X_n$. We claim the following lemma:

**Lemma 3** If $\Theta_X$ are globally independent then $\Theta_{Y_i}$ are also globally independent for all $i = 1, \ldots, n$.

**Proof:** The proof is immediate after noting the correspondence between $\Theta_X$ and $\Theta_{Y_i}$ and the fact that global independence holds for the networks with node ordering: $X_i, X_{i+1}, \ldots, X_{i-1}, X_{i+1}, X_{i+2}, \ldots, X_n$ and $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n, X_i$. In fact, the assumption that $\Theta_X$ is globally independent for all network structures is redundant, it is enough to assume global independence for only two networks: with $X_i$ being first and last node correspondingly. ■

Lemma 3 allows to apply Theorem 2 for $Y_i$ and conclude that any $p(\Theta_X)$ that satisfies Equation 8 should satisfy the following $n$ equations (for $i = 1, \ldots, n$):

$$p(\Theta_X) = C_i \prod_{r \not\in D} \theta_{x_{i_{r_1}}, r, \ldots, r, r'} H_i \left( \left\{ \theta_{x_{i_{r_1}}, r, \ldots, r, r'}, \theta_{r, r'} \in D \text{ s.t. } \begin{align*} [r']_{x_{i_{r_1}}} &= [r']_{x_{i_{r_2}}} + 1 = [r']_{x_{i_{r_3}}} + 1 \\ [r'']_{x_{i_{r_1}}} &= [r'']_{x_{i_{r_2}}} + 1 = [r'']_{x_{i_{r_3}}} + 1 \\ \text{where } [r']_{x_{i_{r_1}}} &\triangleq r_{x_{i_{r_1}}} \end{align*} \right\} \right) \tag{10}$$

where $[r']_{x_{i_{r_1}}}$ and $[r']_{x_{i_{r_2}}}$ denote the values of vector $r$ on $X_i$ and $X \setminus X_i$ respectively. The $' + 1'$ operation denotes taking the next value of the corresponding vector in any fixed order. (We assume here the alphabetical order for the convenience, but the sets of parameters of $H$ are equivalent under all orders.)

Now by analyzing Equation 10 we can see that Lemma 3 actually holds in two directions:

**Lemma 4** If $\Theta_{Y_i}$ are globally independent for all $i = 1, \ldots, n$ then $\Theta_X$ are also globally independent.

**Proof:** Since the distributions of $\Theta_X$ that are globally independent for all $\Theta_{Y_i}$, $i = 1, \ldots, n$, are described by Equation 10 we must prove that any solution for Equation 10 is a solution for Equation 8.

Consider $p(\Theta_X) = f(\Theta_X)$ a solution for Equation 10. We would like to show that $f(\Theta_X)$ satisfies Equation 8 for an arbitrary $l = \{i_1, \ldots, i_k\}$. Indeed, since $f$ is a solution for Equation 10 it can be represented as

$$f(\Theta_X) = C_i \prod_{r \not\in D} \theta_{x_{i_{r_1}}, r} H_i \left( \left\{ \theta_{x_{i_{r_1}}, r, r', r''} \in D \text{ s.t. } \begin{align*} [r']_{x_{i_{r_1}}} &= [r']_{x_{i_{r_2}}} + 1 = [r']_{x_{i_{r_3}}} + 1 \\ [r'']_{x_{i_{r_1}}} &= [r'']_{x_{i_{r_2}}} + 1 = [r'']_{x_{i_{r_3}}} + 1 \\ \text{where } [r']_{x_{i_{r_1}}} &\triangleq r_{x_{i_{r_1}}} \end{align*} \right\} \right) \tag{11}$$

Changing the variables to $\theta_{x_{i_{j}}} |\text{variables of complete network with topological order of nodes specified by } I|$ we see that $f(\Theta_X)$ factorizes according to $I$. We have:

$$f(\Theta_X) = C_i \prod_{j=1}^{n} \prod_{r \in D_{i_{j}}, r \in D_{i_{j+1}}, \ldots, D_{i_{j_{m-1}}}} \theta_{x_{i_{j_{1}}}, \ldots, x_{i_{j_{m-1}}}, x_{i_{j_{m}}}} H \left( \left\{ \theta_{x_{i_{j}}, r, r+1} H_i \in Di_{j} \text{ s.t. } \begin{align*} r, r+1 \in D_{i_{j}} \times \ldots \times D_{i_{j_{m-1}}} \\ \hat{r} \hat{r+1} \in \hat{D}_{i_{j}} \end{align*} \right\} \right) \tag{12}$$

where $\theta_{x_{i_{j_{1}}}, \ldots, x_{i_{j_{m}}}} = \sum_{r \in D, r+1 \in D} \theta_x \theta_{x_{i_{j_{1}}}, \ldots, x_{i_{j_{m}}}}$. These concludes the proof, since we showed that $f(\Theta_X)$ factorizes according to arbitrary complete network order $I$, therefore distribution $f(\Theta_X)$ is globally independent. ■

The solution of the Equation 10 could be almost an impossible task without the help of following lemma:
Lemma 5 Consider the following system of $m$ functional equations:

\[
\begin{align*}
f(x_1, \ldots, x_n) &= \sum_{i=1}^{n} a_{i1}x_i + h_1(\tilde{b}_{11}x, \tilde{b}_{12}x, \ldots, \tilde{b}_{1k_1}x) \\
f(x_1, \ldots, x_n) &= \sum_{j=1}^{n} a_{j2}x_j + h_2(\tilde{b}_{21}x, \tilde{b}_{22}x, \ldots, \tilde{b}_{2k_2}x) \\
&\vdots \\
f(x_1, \ldots, x_n) &= \sum_{l=1}^{n} a_{m1}x_l + h_m(\tilde{b}_{m1}x, \tilde{b}_{m2}x, \ldots, \tilde{b}_{mk_m}x)
\end{align*}
\]

(13)

where $f, h_1, \ldots, h_m$ are unknown functions, $\alpha_{ij}$ unknown constants and $\tilde{b}_{ij}$ arbitrary (given) $n$-dimensional vectors. Equations 13 express the fact that function $f$ can be represented as a linear sum of its variables plus arbitrary function of any given form: $h_1, \ldots, h_m$. For applications in this paper, $\tilde{b}_{ij} \in \{-1,0,1\}^n$ and $k_1 = k_2 = \ldots = k_m$.

The general solution for $f$ in Equation 13 is:

\[
f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i + h(\tilde{b}_1x, \ldots, \tilde{b}_kx)
\]

(14)

where $h$ is an arbitrary function, $\{a_i\}$ are arbitrary constants and $\tilde{b}_1, \ldots, \tilde{b}_k$ is the basis of the linear space $\bigcap_{l=1}^{m} B_l$, where $B_l$ is a linear space spanned by $\tilde{b}_{l1}, \ldots, \tilde{b}_{lk_l}$.

Since Equations 10 can be transformed to the form of Equation 13 by taking a logarithm of both sides of each equation and changing the variables to $\ln \theta$, Lemma 5 provides a powerful tool for solving Equation 10. The proof of this lemma is given in Appendix A.3.

In the following subsections we give an exact functional form of globally independent prior distributions for binary valued network and show the class of globally independent distributions for general networks. Our results demonstrate that global independence assumption alone is not enough to ensure Dirichlet prior.

4.1 Binary-Valued Networks

Consider $n$ binary-valued random variables: $X = X_1, \ldots, X_n$, that can be described by $2^n$ variables, $\{\theta_x\}_{x \in \{0,1\}^n}$ (of them $2^n - 1$ independent). The following theorem specifies the functional form of any distribution on $\Theta_X$ that satisfies global independence assumption.

Theorem 6 Any distribution of $\Theta_X$, where $X = X_1, \ldots, X_n$ are binary random variables, that satisfies global independence assumption for all complete DAG models for $X$, i.e. satisfies Equation 8, is of the form

\[
p(\Theta_X) = C \prod_{x \in \{0,1\}^n} \alpha_x \theta_x \left( \prod_{x \in \{0,1\}^n} \theta_x \right)^h \left( \prod_{x \in \{0,1\}^n, x^y \neq x^z} \theta_x \right)^{\theta_{xz}}
\]

(15)

where $h$ is an arbitrary function, $\alpha_x$ are arbitrary constants and $C$ is a normalization constant.

The proof is given in Appendix A.4. For example, the prior for 4-node binary network is:

\[
p(\{\theta_x\}_{x \in \{0,1\}^4}) = C \prod_{x \in \{0,1\}^4} \alpha_x \theta_x \left( \theta_{0000}\theta_{0101}\theta_{1010}\theta_{1100}\theta_{0110}\theta_{1001}\theta_{1110}\theta_{1111} \right)^h
\]

(16)

4.2 The General Network

The functional equations that describe admissible distributions for arbitrary $n$-node network get extremely complex, as the number of nodes and their values grow. We have not succeeded to give explicit expression for the general solution of Equations 8 in the most general case. For any specific network, however, the globally independent prior distribution can be calculated using Equation 10 and Lemma 5.

The following theorem is a direct generalization of the general solution for a binary valued networks. We believe, that general solution of Equation 8 is given by the theorem below, but we are able to prove one-way inclusion only.
Theorem 7 Consider an $n$ discrete random variables $X = X_1, \ldots, X_n$ with $X_i$ receiving values from $D_i$. The joint multinomial distribution of $X$, i.e. the distribution of $n$-tuples from $D = D_1 \times \ldots \times D_n$ is described by natural parameters $\Theta_X = \{\theta_x | x \in D\}$.

Let $a_n$ denote the arguments of the arbitrary functions in solution for binary networks (Equation 15):,

$$a_n\left((t_x)_{x \in [0,1]}\right) = \prod_{x \in [0,1]} t_x^{a_{x,x}} = \prod_{x \in [0,1]} t_x^{a_{x,x}}$$ (17)

then any $f(\Theta_X)$ of the form given below gives a globally independent prior on $\Theta_X$:

$$f(\Theta_X) = \prod_{x \in D} \theta_x^{a_{x,x}} h \left(\{a_{x_1,x_2} | x_1, x_2 \in D, \forall i, 1 \leq i \leq n, x_i \neq \bar{x}_i | \bar{y}_i\} \right)$$ (18)

where $h$ is an arbitrary function with parameters $a_{x_1,x_2} = a_n(t_{y_i})$ where $t_{y_i} = \theta_{x,y_i}$ for $x_{ji} = y_{ji}$.

Note that not all of $a_{x_1,x_2}$ are independent.

The proof is immediate, since Equation 18 satisfies all Equations 8. An example of parameter distribution of this form is the distribution for two node network, as specified by Equation 2.

The Equation 18 provided a loosey form (with inter-dependencies among parameters of $h$) of the admissible distributions for a general network under global independence assumption. For any specific set of $X$ that can have values from $D$ we can specify the admissible distribution in symbol form by solving the appropriate Equation 10 using Lemma 5. For example, for three ternary variables $X_1, X_2, X_3$ global independence on all complete network structures dictates a parameter prior of the form:

$$f(\{\theta_x | x \in [1,3]^3\}) = C \prod_{x \in [1,3]^3} \theta_x^{a_{x,x}} h \left(\{a_{x_1,x_2} | x_1, x_2 \in D, \forall i, 1 \leq i \leq n, x_i \neq \bar{x}_i | \bar{y}_i\} \right)$$ (19)

Note that in terms of Equation 18, the arguments of $h$ are $a_{11,12,22}, a_{12,13,23}, a_{11,21,32}, a_{12,21,32}, a_{11,22,32}, a_{12,22,32}$, and $a_{12,11,32}, a_{12,21,32}, a_{11,12,22}, a_{11,21,32}$. The matlab code for computing the symbolic form of such distributions for an arbitrary $X$ is shown in Appendix B.

5 Dirichlet Priors: The Minimal Set of Assumptions

We have shown in the previous sections that global independence alone is not enough to ensure Dirichlet prior on the network parameters. The natural question is: “What is the minimal set of independence requirements that can ensure Dirichlet prior?” In this section we give an answer to this question. We start by providing some additional results that link between global independence in various networks.

Definition Parameters $\theta_m = \{\theta_x | m\}$ of node $X_i$ in a DAG model $m(s, F_s)$ are said to be globally independent if $p(\theta_m | m) = p(\theta_m | m)p(\theta_m | m)$.

Lemma 8 Let $m_1$ be an arbitrary complete $n$-node network with topological order of nodes $X_1, \ldots, X_n$, \{i_1, \ldots, i_n\} = \{1, \ldots, n\}$ and let $m_2$ be another complete network, with order $X_{j_1}, \ldots, X_{j_n}$ \{j_1, \ldots, j_n\} = \{1, \ldots, n\}. Then given $i_k = j_k$ and $\{i_1, \ldots, i_{k-1}\} = \{j_1, \ldots, j_{k-1}\}$: $\theta_{m_1}$, globally independent if $\theta_{m_2}$ globally independent.

Proof: Since $m_1$, $m_2$ are arbitrary, only one direction of “iff” must be proved. Consider $\theta_{m_1}$ globally independent, then:

$$p(\theta_{m_1} | m_2) = J^{-1}_{\theta_{m_1} \rightarrow \theta_{m_2}} p(\theta_{m_1} | m_1) = J^{-1}_{\theta_{m_1} \rightarrow \theta_{m_2}} p(\theta_{m_1} | m) p(\theta_{m_1} | \theta_{m_1} | m_1)$$ (20)

where $\theta_{m_1} = f(\theta_{m_2})$ is an appropriate transformation, and $J_{\theta_{m_1} \rightarrow \theta_{m_2}}$ the Jacobian of this transformation. Since $X_{j_1}$ have the same parents in both networks, $J_{\theta_{m_1} \rightarrow \theta_{m_2}} (\theta_{m_2}) \equiv \theta_{m_2}^{\theta_{m_1}}$, and $f_{\theta_{m_1} \rightarrow \theta_{m_2}}$ depends only on
the parameters $\theta_{m_1}\setminus \theta_{m_2}^{i_1}$ and the Jacobian $J_{\theta_{m_1}\setminus \theta_{m_2}^{i_1}}$ depends only on $\theta_{m_2}\setminus \theta_{m_2}^{i_1}$. We have:

$$p(\theta_{m_2}|m_2) = f_1(\theta_{m_2}^{i_1}) f_2(\theta_{m_2}\setminus \theta_{m_2}^{i_1})$$

where $f_1(\theta_{m_2}^{i_1}) = p(\theta_{m_2}^{i_1}|m_1)$ and $f_2(\theta_{m_2}\setminus \theta_{m_2}^{i_1}) = J_{\theta_{m_1}\setminus \theta_{m_1}^{i_1}} f(\theta_{m_2}\setminus \theta_{m_2}^{i_1}|m_1)$. Thus $\theta_{m_2}$ are globally independent. 

We can now present the main result of this section:

**Theorem 9** Let $X = X_1, \ldots, X_n$ be a random variables over $D_1, \ldots, D_n$. Let $m_1(s_1, F_{s_1})$ be an arbitrary, complete DAG model for $X$ with topological order of nodes $X_{i_1}, \ldots, X_{i_n}$, $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$ and let $m_2(s_2, F_{s_2})$ be another complete DAG model for $X$, with order $X_{j_1}, \ldots, X_{j_n}$ $\{j_1, \ldots, j_n\} = \{1, \ldots, n\}$, s.t. $j_n = i_1$.

- Parameters of $X_{i_1}$ in $m_1$ are globally independent, i.e.

$$p(\theta_{m_1}|m_1) = p(\theta_{m_1}^{i_1}|m_1) p(\theta_{m_2}\setminus \theta_{m_2}^{i_1}|m_1)$$

- Parameter of $X_{j_n}$ in $m_2$ are globally and locally independent, i.e.

$$p(\theta_{m_2}|m_2) = p(\theta_{m_2}\setminus \theta_{m_2}^{j_1}) \prod_{\text{pa}_{j_n}^{i_j} \in D_j \text{pa}_{j_n}^{i_j}} p(\theta_{X_{j_n}\setminus \text{pa}_{j_n}^{i_j}}|m_2)$$

where $\theta_{X_i\setminus \text{pa}_i} = \{\theta_{x_i, \text{pa}_i}|x_i \in \{d_1, \ldots, d_{|D_i|-1}\} \}$

then $p(\Theta_X)$ is Dirichlet and this set of conditions is minimal in sense that the elimination of any one of these two conditions extends the class of admissible priors beyond Dirichlet.

The theorem states that among all the set of global and local independence assumptions used by previous authors, one actually need only two assumptions: global independence for the network parameters for the first node in some network, and global and local independence for the same node in some network where this node is the last node.

**Proof:** The minimality of this two assumptions is straightforward, since eliminating any one of them will allow any distribution of form 22 or 23. So Lemma 8 holds, we can assume without the loss of generality that two DAG models under consideration are models with node orders $X_n, X_1, \ldots, X_{n-1}$ and $X_1, \ldots, X_n$ respectively. By treating nodes $X_1, \ldots, X_{n-1}$ as a one super node for a random variable $Y = X_1 \times X_2 \times \ldots \times X_{n-1}$ the problem reduces to determining prior distributions for two node networks with global independence for all configurations and local independence for one last node in one network, that turn out to be Dirichlet (Appendix A.5). 

### 6 Discussion

In this paper we investigated the prior parameter distributions for discrete DAG models that arise under different independence conditions. We have shown that global independence alone can not ensure the Dirichlet prior. We explicitly defined the class of prior distributions that satisfy global independence assumption for binary valued networks and provided an algorithmic way to compute the functional form of admissible distributions for any discrete network.

In addition, the minimal set of global and concurrent (on the same network) local independence assumptions that result in Dirichlet prior was shown. As a byproduct, a number of important functional equations was solved (e.g. Lemma 5). Among the open topics, the question of explicit functional form of general solution to Equation 8 remains unanswered. This, however, may be of little practical importance, since application of Equation 10 and Lemma 5 can provide a solution for any specific set of random variables $X$.

All the results presented in this paper were achieved under the assumption of local parameter distributions being twice differentiable and everywhere positive. One may hope to derive the properties of twice
Proof: Let $v$ ariables/, the differential equation becomes
\[ \frac{dz}{dt} = \frac{z}{z + 1} \]
where $g$, $f$, $G$ and $f$ are unknown functions and $z_0$, $z_1$, $y$ are variables from $(0, 1)$. The question “Is there any measurable solution that is not of the Dirichlet form?” remains open.

A Proofs and Derivations

A.1 Useful Lemmas

In this section we prove several lemmas that are useful for solving equations presented in this paper.

Lemma 10 The general solution of the following partial differential equation for $f(x_1, \ldots, x_n)$,

\[ f_{x_i} - f_{x_j} = \frac{\alpha}{x_i} + \frac{\beta}{x_j} \]

is given by

\[ f(x_1, \ldots, x_n) = a \log x_i - b \log x_j + h(x_i + x_j, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_j, x_{i+1}, \ldots, x_n) \]

where $h$ is an arbitrary differentiable function having $n - 1$ arguments.

Proof: Let $s = x_i + x_j$ and $t = x_i - x_j$. Thus, $f_{x_i} = f_s + f_t$, $f_{x_j} = f_s - f_t$. Hence, after a change of variables, the differential equation becomes

\[ f_t = \frac{\alpha}{s + t} + \frac{\beta}{s - t} \]

Integrating wrt $t$ and changing back to the original variables yields the desired solution.

Lemma 11 Let $f(x_1, \ldots, x_n)$ be a twice-differentiable function from $(0, 1)^n$ to $(0, 1)$. If for all $1 \leq i < j \leq n$,

\[ f(x_1, \ldots, x_n) = a_i \log x_i + a_j \log x_j + f_{ij}(x_i + x_j, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_j, x_{i+1}, \ldots, x_n) \]

where $f_{ij}$ are arbitrary twice differentiable functions from $(0, 1)^{n-1}$ to $(0, 1)$, then

\[ f(x_1, \ldots, x_n) = g\left(\sum_{i=1}^{n} a_i \log x_i \right) \]

where $g$ is an arbitrary twice-differentiable function from $(0, 1)$ to $(0, 1)$.

Proof: We shall prove the following stronger claim. For every $2 \leq l \leq n$, and for every permutation of the indices of $x_1, \ldots, x_n$,

\[ f(x_1, \ldots, x_n) = h_l\left(\sum_{i=1}^{l} x_i, x_{l+1}, \ldots, x_n\right) + \sum_{i=1}^{l} a_i \log x_i \]

where $h_l$ is an arbitrary twice differentiable function from $(0, 1)^{n-l+1}$ to $(0, 1)$. The function $h_l$ depends on the permutation, although this fact is not reflected in our notation. The base case $l = 2$ is assumed by the lemma and the case $l = n$ is needed to be proved.
By the induction hypothesis we assume Eq 28 and for the permutation

\[(1, \ldots, n) \rightarrow (l, 1, \ldots, l-1, l+1, \ldots, n)\]

we also assume (by the induction hypothesis),

\[f(x_1, \ldots, x_n) = g_l(x_l, x_{l+1} + \frac{1}{2} x_l + \frac{l-1}{2} x_l + \cdots + \frac{l-1}{2} x_l + b_l \log x_l + b_{l+1} \log x_{l+1}) \tag{29}\]

Let \(x = \sum_{i=1}^{l-1} x_i, c_i = b_i - a_i\) and \(x = (x_{l+1}, \ldots, x_n)\). From Eqs 28 and 29 we get,

\[h_l(x_l + x, x_{l+1}, \mathbf{x}) = g_l(x_l, x_{l+1} + x, \mathbf{x}) + \sum_{i=1}^{l-1} c_i \log x_i - a_l \log x_l + b_{l+1} \log x_{l+1} \tag{30}\]

Set \(x_i = \frac{1}{x_i - 1}, i = 1, \ldots, l - 1\). Thus, \(x = \frac{1}{x_l}\) and Eq 30 yields,

\[h_l(x_l + \frac{1}{2}, x_{l+1}, \mathbf{x}) = g_l(x_l, x_{l+1} + \frac{1}{2}, \mathbf{x}) + \sum_{i=1}^{l-1} c_i \log \left(\frac{1}{2(l - 1)}\right) - a_l \log x_l + b_{l+1} \log x_{l+1} \tag{31}\]

Plugging Eq 31 into Eq 30 and letting

\[\hat{g}_l(x_l, x_{l+1} + x, \mathbf{x}) = g_l(x_l, x_{l+1} + x, \mathbf{x}) - a_l \log x_l \tag{32}\]

yields,

\[\hat{g}_l(x_l + x - \frac{1}{2}, x_{l+1} + \frac{1}{2}, \mathbf{x}) = \hat{g}_l(x_l, x_{l+1} + x, \mathbf{x}) + \sum_{i=1}^{l-1} c_i \log x_i - \sum_{i=1}^{l-1} c_i \log(2l - 1)) \tag{33}\]

By taking a derivative wrt \(x_j, 1 \leq j \leq l - 1\) of Eq 33 we get,

\[\hat{g}_l(x_l + x - \frac{1}{2}, x_{l+1} + \frac{1}{2}, \mathbf{x})_1 = \hat{g}_l(x_l, x_{l+1} + x, \mathbf{x})_2 + c_j / x_j \tag{34}\]

Similarly by taking the derivatives wrt \(x_i\) we get,

\[\hat{g}_l(x_l + x - \frac{1}{2}, x_{l+1} + \frac{1}{2}, \mathbf{x})_1 = \hat{g}_l(x_i, x_{l+1} + x, \mathbf{x})_1 \tag{35}\]

Consequently,

\[\hat{g}_l(x_l, x_{l+1} + x, \mathbf{x})_1 - \hat{g}_l(x_i, x_{l+1} + x, \mathbf{x})_2 = c_j / x_j \tag{36}\]

for \(j = 1, \ldots, l - 1\).

We now show that \(c_j = 0\). If \(l > 2\), then set \(j = j_1\) and \(j = j_2, 1 \leq j_1 < j_2 \leq l - 1\), in Eq 36 and subtract the two equations. Consequently, \(c_{j_1} / x_{j_1} = c_{j_2} / x_{j_2}\) and therefore \(c_{j_1} = c_{j_2} = 0\). If \(l = 2\), then, \(x = x_1\) and Eq. 36 becomes

\[h'_1(x_2, x_3 + x_1, \mathbf{x}) - h'_2(x_2, x_3 + x_1, \mathbf{x}) = c_1 / x_1 \tag{37}\]

Let \(u = x_1 + x_3, w = x_1 - x_3\) and rewrite the last equation,

\[h'_1(x_2, u, \mathbf{x}) - h'_2(x_2, u, \mathbf{x}) = \frac{2c_1}{u + w} \tag{38}\]

Since the left hand side is not a function of \(w\) we have \(c_1 = 0\).
Now let \( s = x_t + (x + x_{t+1}) \) and \( t = x_t - (x + x_{t+1}) \) and rewrite the differential equation (Eq 36) by changing variables to \( s, t \) and \( \overline{x} \) (using \( c_j = 0 \)). We get,

\[
\frac{\partial \overline{x}(s, t, \overline{x})}{\partial t} = 0 \tag{39}
\]

Thus, \( \dot{g}(s, t, \overline{x}) = i(s, \overline{x}) \) where \( i \) is a function of just \( s \) and \( \overline{x} \). Consequently, by switching back to the original variables, we get,

\[
f(x_1, \ldots, x_n) = \sum_{i=1}^{t+1} a_i \log x_i + \mathcal{I}(\sum_{i=1}^{t+1} x_i, x_{t+2}, \ldots, x_n) \tag{40}
\]

Since the above equation can be derived for any permutation of the indices of \( x_i \) using the arguments above, the induction hypothesis has been proved.

**Lemma 12** Let \( t = (t_1, \ldots, t_n) \) be an \( n \)-dimensional vector of variables (or values) and let \( \hat{t} \) denote the first \( n-1 \) elements of \( t \), i.e. \( \hat{t} = (t_1, \ldots, t_{n-1}) \). Let \( \Delta_m = \{ (a_1, \ldots, a_m) | a_i > 0, \sum_{i=1}^{m} a_i < 1 \} \) denote the \( m \)-dimensional unit simplex.

Consider the following general type of functional equation:

\[
f(t_1, \ldots, t_n) = h_i(t, f_i(g_i(t))), \quad i = 1, \ldots, k \tag{41}
\]

where \( f, f_1, \ldots, f_k \) are unknown functions and \( t = (t_1, \ldots, t_n) \in D \subseteq \mathbb{R}^n \) are independent variables. Suppose \( D \) is such that \( \{ t | t > 0, \sum_{j=1}^{n} t_j = 1 \} \subset D \).

Consider the same equation, with \( t_1, \ldots, t_n \) being dependent, with \( t_n = 1 - \sum_{j=1}^{n-1} t_j \):

\[
\dot{f}(\hat{t}) = h_i(q(\hat{t}), f_i(g_i(q(\hat{t})))), \quad i = 1, \ldots, k \tag{42}
\]

where \( q(\hat{t}) = (t_1, \ldots, t_{n-1}, 1 - \sum_{j=1}^{n-1} t_j) \) and \( \hat{t} \in \Delta_{n-1} \).

The relationship between the solutions of Equations 41 and 42 is:

1. For any solution \( f \) of Equation 41, \( \dot{f}(\hat{t}) \triangleq f(q(\hat{t})) \) is a solution for Equation 42.

2. If \( h_i \) and \( g_i \) are scale-independent on \( D \), i.e. \( h_i(at, b) = h_i(t, b) \) and \( g_i(at) = g_i(t) \) for all \( i = 1, \ldots, k, \ a, b \in \mathbb{R} \) and \( t, at \in D \). Then any solution \( \dot{f} \) of Equation 42 is of the form \( \dot{f}(\hat{t}) \triangleq f(q(\hat{t})) \), where \( f \) is a solution for Equation 41.

**Proof:** The first part: Since \( f \) is a solution for Equation 41 there exist functions \( f_1, \ldots, f_k \) such that Equation 41 holds for all \( t \in D \). We have (for \( 1 \leq i \leq k \)):

\[
\dot{f}(\hat{t}) = f(q(\hat{t})) = h_i(q(\hat{t}), f_i(g_i(q(\hat{t})))) \tag{43}
\]

for all \( \hat{t} \in \Delta_{n-1} \). Thus \( \dot{f} \) is a solution of Equation 42.

The second part: We first show that for any \( \dot{f} \) a solution for Equation 42 \( \dot{f}(\hat{t}) \triangleq \dot{f}(\hat{t})/\sigma \) is a solution for Equation 41, where \( \sigma = \sum_{j=1}^{n-1} t_j \). Indeed, since \( \dot{f} \) is a solution for Equation 42 there exist functions \( f_1, \ldots, f_k \) such that Equation 42 holds for all \( \hat{t} \in \Delta_{n-1} \) (and for all \( t \in D, \dot{f}(\hat{t})/\sigma \in \Delta_{n-1} \)). We have (for \( 1 \leq i \leq k \)):

\[
f(t) = \dot{f}(\hat{t})/\sigma = h_i(q(\hat{t})/\sigma), f_i(g_i(q(\hat{t})/\sigma))) = h_i(t/\sigma, f_i(g_i(t/\sigma))) = h_i(t, f_i(g_i(t))) \tag{44}
\]

for all \( t \in D \).

Now it is easy to check that \( \dot{f}(\hat{t}) = f(q(\hat{t})) \). We have (for \( 1 \leq i \leq k \)):

\[
f(q(\hat{t})) = \dot{f}(\hat{t})/\sigma \Rightarrow f(q(\hat{t})) = \dot{f}(\hat{t})/1 = \dot{f}(\hat{t}) \tag{45}
\]

Thus any solution for Equation 42 can be found by using a solution of Equation 41. \( \square \)
A.2 The Solution of Functional Equation for Two Node Networks

We now solve Equation 5 for any $n, k$. We use the following notations. Let $\hat{h}(x)$ denote $\ln h(x)$ for any positive function $h$. Also let

$$\hat{F}_i(Y) = \frac{\partial \hat{F}(Y)}{\partial y_i} \quad 1 \leq i \leq n - 1$$

$$\hat{F}_{i,j}(Y) = \frac{\partial^2 \hat{F}(Y)}{\partial y_i \partial y_j} \quad 1 \leq i, j \leq n - 1$$

$$\hat{g}_{j,i}(Z) = \frac{\partial \hat{g}(Z)}{\partial x_j} \quad 1 \leq i, j \leq n$$

$$\hat{g}_{j,i}(j_{j,i})(Z) = \frac{\partial^2 \hat{g}(Z)}{\partial z_{j,i} \partial z_{j,i}} \quad 1 \leq i, j \leq n, 1 \leq j, j_1, j_2 \leq k - 1$$

and similarly for $G$ and $f$.

Taking partial derivatives of $X$ and $W$ gives

$$\frac{\partial x_j}{\partial y_i} = z_{j,i} - z_{j,\nu} \quad 1 \leq j \leq k, \quad 1 \leq i \leq n - 1$$

$$\frac{\partial x_j}{\partial z_{j,i}} = \begin{cases} y_k & j = m \\ 0 & j \neq m \end{cases}, \quad \frac{\partial z_{j,i}}{\partial x_j} = -y_k \quad 1 \leq m, j \leq k - 1, \quad 1 \leq i \leq n$$

$$\frac{\partial y_{i,j}}{\partial y_i} = \begin{cases} \frac{z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu})}{x_j^2} & i \neq l \\ -\frac{z_{y_{i}y_{j}}^2}{x_j^3} + \frac{z_{y_{i}y_{j}}}{x_j} & i = l \end{cases} \quad 1 \leq j \leq k, \quad 1 \leq i, l \leq n - 1$$

$$\frac{\partial y_{i,j}}{\partial x_{i,j}} = \begin{cases} \frac{z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu})}{x_j^2} & i \neq l, m = j \\ -\frac{z_{y_{i}y_{j}}^2}{x_j^3} + \frac{z_{y_{i}y_{j}}}{x_j} & i = l, m = j \end{cases} \quad 1 \leq j, m \leq k - 1, \quad 1 \leq i, l \leq n, \quad i \neq n$$

$$\frac{\partial z_{i,j}}{\partial x_{i,j}} = \begin{cases} \frac{z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu})}{x_j^2} - \frac{z_{y_{i}y_{j}}}{x_j} & i \neq l \\ \frac{z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu})}{x_j^2} & i = l \end{cases} \quad 1 \leq m \leq k - 1, \quad 1 \leq i, l \leq n, \quad i \neq n$$

(Note, that there is no error in line 3; $\frac{\partial y_{i,j}}{\partial x_{i,j}}$ indeed can be written in the given form). Additional calculations give:

$$\frac{\partial \hat{G}}{\partial y_i} = \sum_{j=1}^{k} [z_{j,i} - z_{j,\nu}] \hat{G}_j \quad 1 \leq i \leq n - 1$$

$$\frac{\partial \hat{G}}{\partial z_{j,i}} = y_k \hat{G}_j \quad 1 \leq i \leq n, 1 \leq j \leq k - 1$$

$$\frac{\partial \hat{f}}{\partial y_i} = \sum_{j=1}^{k} [z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu}) \hat{f}_j + \hat{z}_{y_{i}y_{j}} \hat{f}_j \hat{f}_j] \quad 1 \leq i \leq n - 1$$

$$\frac{\partial \hat{f}}{\partial x_{i,j}} = \sum_{j=1}^{k} [z_{y_{i}y_{j}}(z_{i,j}-z_{i,\nu}) \hat{f}_j + \hat{z}_{y_{i}y_{j}} \hat{f}_j \hat{f}_j + \hat{z}_{y_{i}y_{j}} \hat{f}_j \hat{f}_j - \frac{\hat{z}_{y_{i}y_{j}}}{x_j} \hat{f}_{j,i}] \quad 1 \leq i \leq n - 1, 1 \leq j \leq k - 1$$

Let $\hat{C}_j = \sum_{j=1}^{k} [z_{y_{i}y_{j}} \hat{f}_j]$. By taking the logarithm and then a derivative wrt $y_k$ ($1 \leq i \leq n - 1$) of Equation 5, we get,

$$\hat{F}_i(Y) = \sum_{j=1}^{k} [z_{j,i} - z_{j,\nu}] \hat{G}_j(X) + \sum_{j=1}^{k} [z_{j,i} - z_{j,\nu}] \hat{C}_j(x_j, W) + \frac{\hat{z}_{y_{i}y_{j}}}{x_j} \hat{f}_j(W)$$

(49)

By taking the logarithm and then a derivative wrt $z_{j,i}$ ($1 \leq i \leq n - 1, 1 \leq j \leq k - 1$) of Equation 5, we get,

$$\hat{g}_{j,i}(Z) = y_k \hat{G}_j(X) + y_k \hat{C}_j(x_j, W) - y_k \hat{C}_k(x_k, W) + \frac{\hat{W}}{x_j} \hat{f}_j(W) - \frac{\hat{W}}{x_k} \hat{f}_{k,i}(W)$$

(50)

and by the same operation wrt $z_{j,\nu}$ ($1 \leq j \leq k - 1$):

$$\hat{g}_{j,\nu}(Z) = y_k \hat{G}_j(X) + y_k \hat{C}_j(x_j, W) - y_k \hat{C}_k(x_k, W)$$

(51)

From Equations 50 and 51 we get: ($1 \leq i \leq n - 1, 1 \leq j \leq k - 1$)

$$\frac{1}{y_k} \hat{g}_{j,i}(Z) - \frac{1}{y_k} \hat{g}_{j,\nu}(Z) = \frac{1}{x_j} \hat{f}_{j,i}(W) - \frac{1}{x_k} \hat{f}_{k,i}(W)$$

(52)
Solving Equation 52 for $\frac{1}{x_j}\hat{f}_{ji}(W)$ and substitution into Equation 51 gives:

$$\sum_{l=1}^{n} \frac{\hat{z}_{jl}}{x_j} \hat{g}_{jl}(Z) = \hat{G}_j(X) - \sum_{l=1}^{n-1} (w_{jl} - w_{kl}) \frac{1}{x_k} \hat{f}_{ki}(W)$$  \hspace{1cm} (53)

Simplifying Equation 49 by using Equations 51, 52 and recalling $\hat{z}_{ki} = 1 - \sum_{j=1}^{k-1} \hat{z}_{ji}$ we get (for $1 \leq i \leq n-1$):

$$\hat{F}_i(Y) = \sum_{j=1}^{k-1} \left( \frac{\hat{z}_{ji}}{y_i} \hat{g}_{ji}(Z) - \frac{\hat{z}_{jn}}{y_i} \hat{g}_{jn}(Z) \right) + \frac{1}{x_i} \hat{f}_{ki}(W)$$  \hspace{1cm} (54)

Multiplying Equations 54 by $(w_{ji} - w_{ki})$, taking the sum of the resulting equations and substitution $\sum_{l=1}^{n-1} (w_{jl} - w_{kl}) \frac{1}{x_l} \hat{f}_{kl}(W)$ from Equation 53 gives (for $1 \leq j \leq k - 1$):

$$\sum_{l=1}^{n-1} (w_{jl} - w_{kl}) \hat{F}_l(Y) = \sum_{l=1}^{n-1} \left( (w_{jl} - w_{kl}) \frac{1}{x_j} \hat{f}_{jl}(Y) + \frac{1}{y_i} \hat{G}_j(X) - \sum_{l=1}^{n-1} \frac{z_{jl}}{x_j} \hat{g}_{jl}(Z) \right)$$  \hspace{1cm} (55)

After some simplifications (recall that $w_{ji} = \frac{z_{ji} y_i}{z_{ij}}$) we get:

$$\sum_{l=1}^{n-1} (w_{jl} - w_{kl}) \hat{F}_l(Y) = \sum_{l=1}^{n-1} \left[ \left( \frac{z_{jl}}{x_j} - \frac{z_{kl}}{x_k} \right) \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) \right] + \hat{G}_j(X) - \sum_{l=1}^{n-1} \frac{z_{jl}}{x_j} \hat{g}_{jl}(Z)$$  \hspace{1cm} (56)

Taking a derivative by $z_{ji}$ we get:

$$\sum_{l=1}^{n-1} \left( \frac{z_{jl}}{x_j} \frac{y_i}{x_i} - \frac{z_{kl}}{x_k} \frac{y_i}{x_i} \right) \hat{F}_l(Y) + \frac{y_i}{x_i} \hat{F}_i(Y) + \frac{y_i}{x_i} \hat{F}_i(Y) =$$

$$\sum_{l=1}^{n} \left[ \left( \frac{z_{jl}}{x_j} - \frac{z_{kl}}{x_k} \right) \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) \right] + \frac{y_i}{x_i} \hat{G}_j(X) - \sum_{l=1}^{n-1} \frac{z_{jl}}{x_j} \hat{g}_{jl}(Z)$$  \hspace{1cm} (57)

Substituting $z_{ji} = \frac{1}{k}$ (and thus $x_j = \frac{1}{k}$, $w_{ji} = y_i$) for $1 \leq i \leq n$, $1 \leq j \leq k$ we get (1 $\leq$ $i$ $\leq$ $n$ $-1$):

$$- \sum_{l=1}^{n-1} \frac{y_i}{x_i} \hat{F}_l(Y) + \hat{F}_i(Y) = \frac{1}{y_i} C_i + A$$  \hspace{1cm} (58)

where

$$C_i = \frac{1}{k} \sum_{m=1}^{k-1} \hat{g}_{ml}(\frac{1}{k}) - \frac{1}{2k} \sum_{l=1}^{n} \hat{g}_{jl}(\frac{1}{k})(\frac{1}{k}) - \frac{1}{2} \hat{g}_{ji}(\frac{1}{k})$$  \hspace{1cm} (59)

$$A = - \sum_{l=1}^{n-1} \frac{1}{k} \hat{g}_{ml}(\frac{1}{k}) + \frac{1}{2k} \hat{G}_{jj}(\frac{1}{k}) + \frac{1}{2} \sum_{l=1}^{n} \hat{g}_{jl}(\frac{1}{k})$$

where $C_i$'s and $A$ are computed for some value of $j$.

In case $n = 2$, ($Y = \{y\}$, $\hat{F}_1(Y) = \hat{F}'(y)$) we have only one equation

$$-y \hat{F}'(y) + \hat{F}'(y) = \frac{1}{y} C_1 + A$$  \hspace{1cm} (60)
Thus \( F(y) = C y^C_1 (1 - y)^{-(A + C_1)} \) (i.e. \( F \) is of Dirichlet form).

In case \( n \geq 3 \) by subtracting two Equations 58 for \( 1 \leq i_1, i_2 \leq n - 1, i_1 \neq i_2 \), we get:

\[
\hat{F}_{i_1}(Y) - \hat{F}_{i_2}(Y) = \frac{C_{i_1}}{y_{i_1}} - \frac{C_{i_2}}{y_{i_2}}
\]  

(61)

The general solution for Equations 61 is (by Lemmas 10 and 11):

\[
\hat{F}(Y) = h \left( \sum_{i=1}^{n-1} y_i \right) + \sum_{i=1}^{n-1} C_i \ln y_i
\]

(62)

Substituting the solution into Equation 58, and solving for \( h \), we get:

\[
F(Y) = C \prod_{i=1}^{n} y_i^{C_i}
\]

(63)

where \( C \) is an arbitrary constant, \( y_i = 1 - \sum_{i=1}^{n-1} y_k \) and \( C_n = -A - \sum_{i=1}^{n-1} C_i \).

Similarly, \( G(X) = B \prod_{j=1}^{k} x_j^{B_j} \), and by substitution \( y_k = \frac{1}{n} (x_j = \frac{1}{n} z_j, w_{ji} = \frac{z_{ij}}{z_j}, \) where \( z_j = \sum_{i=1}^{n} z_{ij} \)

Equation 56 transforms to:

\[
\sum_{i=1}^{n-1} \left( \frac{z_{ij}}{z_j} - \frac{z_{kl}}{z_k} \right) (C_i - C_n) - \left( \frac{B_j}{z_j} - \frac{B_k}{z_k} \right) = \sum_{i=1}^{n} \left[ \left( \frac{z_{ij}}{z_j} - \frac{z_{kl}}{z_k} \right) \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) \right] - \sum_{i=1}^{n} \frac{z_{ij}}{z_j} \hat{g}_{jl}(Z)
\]

(64)

By substitution we can see that \( \hat{g}_p(Z) = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ji} \ln z_{ji} \) is a particular solution of Equation 64. From Equations 59 and 64, the coefficients \( \{a_{ji}\} \), \( \{C_i\} \) and \( \{B_j\} \) satisfy: \( C_i = \sum_{m=1}^{k} a_{mi}, \ B_j = \sum_{j=1}^{k} a_{ji}, \ B_k = \frac{1}{k} \sum_{m=1}^{k} B_m \) and, by symmetry, \( C_n = \frac{1}{n} \sum_{i=1}^{n} C_i \) (see Section A.2.1 for derivations)

We must now solve the homogeneous first-order partial differential equation:

\[
\sum_{i=1}^{n} \left[ \left( \frac{z_{ij}}{z_j} - \frac{z_{kl}}{z_k} \right) \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) \right] - \sum_{i=1}^{n} \frac{z_{ij}}{z_j} \hat{g}_{jl}(Z) = 0
\]

(65)

Multiplying each equation by \( z_j \), taking the sum by \( j, 1 \leq j \leq k - 1 \) and simplifying we get:

\[
\sum_{i=1}^{n} \frac{z_{ij}}{z_j} \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) = 0
\]

(66)

so Equations 65 transforms to (for \( 1 \leq j \leq k - 1 \)):

\[
\sum_{i=1}^{n} \frac{z_{ij}}{z_j} \sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) = \sum_{i=1}^{n} \frac{z_{ij}}{z_j} \hat{g}_{jl}(Z)
\]

(67)

Substitution \( \hat{g}(Z) = \hat{g}^{(j)}(Z) \), where \( z_{ij} = \frac{z_{ij}}{z_j} + \frac{z_{ij}}{z_j} \) for \( 1 \leq j \leq k - 1, 1 \leq i \leq n \) gives:

\[
z_{ij} \hat{g}_{ij}(Z) = \begin{cases} 
\frac{z_{ij}}{z_j} \hat{g}_{ij}(Z) + \frac{z_{ij}}{z_j} \hat{g}_{k-1,i}(Z) & j = 1 \\
-\frac{z_{ij}}{z_j} \hat{g}_{j-1,i}(Z) + \frac{z_{ij}}{z_j} \hat{g}_{ij}(Z) + \frac{z_{ij}}{z_j} \hat{g}_{k-1,i}(Z) & 2 \leq j \leq k - 1
\end{cases}
\]

(68)

and

\[
\sum_{m=1}^{k-1} z_{ml} \hat{g}_{ml}(Z) = \frac{1}{z_{kl}} \hat{g}_{kl-1,l}(Z)
\]

(69)

Plugging Equations 68 and 69 into Equations 67 gives

\[
\sum_{i=1}^{n} \frac{z_{ij}}{z_j} \hat{g}_{ij}(Z) = 0 \quad 2 \leq j \leq k - 1
\]

(70)
Thus for all \( j, (1 \leq j \leq k - 1) \):

\[
\sum_{i=1}^{n} x_{ji} \tilde{g}_{ji}(\tilde{Z}) = 0
\]  

(71)

Additional substitution \( \tilde{g}(\tilde{Z}) = \tilde{g}(\tilde{Z}) \), where

\[
x_{ji} = \frac{z_{ji}}{z_{j,i+1}} \quad 1 \leq i \leq n - 1, \quad 1 \leq j \leq k - 1
\]

\[
x_{jn} = \frac{z_{jn}}{z_{j,n+1}} \quad 1 \leq j \leq k - 1
\]

(72)

gives (for \( 1 \leq j \leq k - 1 \)):

\[
x_{jn} \tilde{g}_{jn}(\tilde{Z}) = 0
\]

(73)

The general solution to Equations 73 is

\[
\tilde{g}(\tilde{Z}) = h(\{z_{ji}|1 \leq i \leq n - 1, \quad 1 \leq j \leq k - 1\}) + C \delta
\]

(74)

where \( h \) is an arbitrary differentiable function of \((n - 1) \times (k - 1)\) variables and \( C \delta \) is an arbitrary constant.

Thus the general solution for \( g \) of the functional equation 5 is:

\[
g(Z) = C \left[ \prod_{i=1}^{n} \prod_{j=1}^{k} \left( a_{ji}^{j}\right) \right] H \left( \left\{ \left( \frac{z_{ji}z_{j+1,i+1}}{z_{j+1,i}z_{j,i+1}} \right) \mid 1 \leq i \leq n - 1, \quad 1 \leq j \leq k - 1 \right\} \right)
\]

(75)

where \( H \) is an arbitrary differentiable function of \((n - 1) \times (k - 1)\) variables and \( C \) is an arbitrary constant.

Using \( z_{ji} = \theta_{ij} = \frac{\theta_i}{\theta_j}, y_k = \theta_i \) (for \( 1 \leq i \leq n, \quad 1 \leq j \leq k \)), as well as properties of the particular solution of Equation 64 \( C_i = \sum_{m=1}^{n} a_{mi} \) we get from Equations 63,75:

\[
p(\{\theta_{ij}\}) = C \left[ \prod_{i=1}^{n} \prod_{j=1}^{k} \left( a_{ji}^{j}\right) \right] H \left( \left\{ \left( \frac{\theta_{ij}\theta_{i+1,j+1}}{\theta_{i+1,j}\theta_{ij+1}} \right) \mid 1 \leq i \leq n - 1, \quad 1 \leq j \leq k - 1 \right\} \right)
\]

(76)

This solves Equation 1 and completes the proof of Theorem 2. \( \square \)

### A.2.1 Some technical details

In this appendix we want to show that \( \tilde{g}_b(Z) = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ji} \ln z_{ji} \) is a particular solution of Eq 64, where the coefficients \( \{a_{ji}\} \) should satisfy some constrains imposed by Eq 64 and Eq 59. Note that \( \frac{\partial \tilde{g}_b}{\partial z_{ji}}(Z) = \frac{a_{ji}}{z_{ji}} - \frac{a_{kj}}{z_{kj}} \).

Substituting \( \tilde{g}_b(Z) \) into right hand side of Eq 64, we get:

\[
\sum_{i=1}^{n} \left[ \left( \frac{z_{ji}}{z_{j,i+1}} - \frac{z_{kji}}{z_{k,j+1}} \right) \sum_{m=1}^{k} \sum_{l=1}^{k} a_{ml} \left( \frac{a_{ml}}{z_{ml}} - \frac{a_{kl}}{z_{kl}} \right) \right] = \sum_{i=1}^{n} \frac{z_{ji}}{z_{j,i+1}} \left( \frac{a_{ji}}{z_{ji}} - \frac{a_{kji}}{z_{kji}} \right) - \sum_{i=1}^{n} \frac{z_{kji}}{z_{k,j+1}} \left( \frac{a_{kji}}{z_{kji}} - \frac{a_{kkl}}{z_{kkl}} \right)
\]

(77)

Thus, from Eq 77 and Eq 64, we get:

\[
C_i = C_i = \sum_{m=1}^{k} a_{ml} - \sum_{m=1}^{k} a_{mn} \quad 1 \leq i \leq n - 1
\]

\[
B_j = \sum_{i=1}^{n} a_{ji} \quad 1 \leq j \leq k
\]

(78)
If \( \hat{g}_p(Z) \) is indeed a particular solution to Eq 64 then \( \{a_{ji}\} \) and \( \{C_i\} \) should satisfy Eq 59. Substituting \( \hat{g}_p(Z) \) into Eq 59 and using Eq 78, we get (1 \( \leq i \leq n - 1 \)):

\[
\begin{align*}
C_i &= \sum_{m=1}^{n-1} (a_{mi} - a_{ki}) + k (a_{ji} + a_{ki}) - k (a_{ji} - a_{ki}) = \sum_{m=1}^{n} a_{mi} \\
A &= -k \sum_{m=1}^{n} (a_{mi} - a_{ki}) - k (B_j + B_k) + \frac{1}{n} \sum_{m=1}^{n} (a_{ji} - a_{ki}) 
\end{align*}
\]

(79)

Thus, \( C_n = \sum_{m=1}^{n} a_{mn} \) (from Eqs 78,79) and using the fact that \( A = -\sum_{i=1}^{n} C_i \) (from the definition of \( C_n \), after Eq 63), we get:

\[
k \sum_{m=1}^{n} (B_m - B_k) + k B_k = \sum_{i=1}^{n} C_i 
\]

(80)

Thus \( B_k = \frac{1}{k} \sum_{m=1}^{n} B_m \) and, by symmetry, \( C_n = \frac{1}{n} \sum_{i=1}^{n} C_i \).

### A.3 The Proof of Lemma 5

The proof is given for \( m = 2 \), but can be easily extended for a general case by induction. Suppose that we have a following system of functional equations:

\[
\begin{align*}
f(x_1, \ldots, x_n) &= \sum_{i=1}^{n} a_{i} x_{i} + h(\tilde{a}_1^T \tilde{x}, \ldots, \tilde{a}_k^T \tilde{x}) \\
f(x_1, \ldots, x_n) &= \sum_{i=1}^{n} b_{i} x_{i} + g(\tilde{b}_1^T \tilde{x}, \ldots, \tilde{b}_k^T \tilde{x})
\end{align*}
\]

(81)

By comparing the two equations we have:

\[
h(\tilde{a}_1^T \tilde{x}, \ldots, \tilde{a}_k^T \tilde{x}) = \sum_{i=1}^{n} (\beta_i - a_i) x_i + g(\tilde{b}_1^T \tilde{x}, \ldots, \tilde{b}_k^T \tilde{x})
\]

(82)

It is easy to see that \( V = \{ \tilde{a}^T \tilde{x} \mid \tilde{a} \in \mathbb{R}^n, \tilde{x} = (x_1, \ldots, x_n)^T \} \) is a vector space and that \( V_h = \text{span}(\tilde{a}_1^T \tilde{x}, \ldots, \tilde{a}_k^T \tilde{x}) \) and \( V_g = \text{span}(\tilde{b}_1^T \tilde{x}, \ldots, \tilde{b}_k^T \tilde{x}) \) are subspaces of \( V \). We can now change the variables \( x_1, \ldots, x_n \) to \( n \)-independent variables \( y_1, \ldots, y_k \) where \( y_i = \tilde{g}_i^T \tilde{x} \) such that:

\[
\begin{align*}
V_h \cap V_g &= \text{span}(y_1, \ldots, y_{k^*}) \\
V_h &= \text{span}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_k) \\
V_g &= \text{span}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_{k+k'-k^*}) \\
V &= \text{span}(y_1, \ldots, y_k)
\end{align*}
\]

(83)

where \( k^* = \dim(V_h \cap V_g) \). Let \( \Gamma_h \) be the \( n \times k \) matrix with columns \( \tilde{g}_1, \ldots, \tilde{g}_k \), let \( \Gamma_g \) be the \( n \times k' \) matrix with columns \( \tilde{g}_1, \ldots, \tilde{g}_{k'}, \tilde{g}_{k+1}, \ldots, \tilde{g}_{k+k'-k^*} \) and let \( \Gamma \) denote the \( n \times n \) matrix with columns \( \tilde{g}_1, \ldots, \tilde{g}_n \).

Equation 82 now transforms to:

\[
\hat{h}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_k) = \sum_{i=1}^{n} \delta_i y_k + \hat{g}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_{k+k'-k^*})
\]

(84)

where

\[
\begin{align*}
\hat{h}(t_1, \ldots, t_k) &= h((\Gamma_h^T \tilde{a}_1)^T t_1, \ldots, (\Gamma_h^T \tilde{a}_k)^T t_k) \\
\hat{g}(t_1, \ldots, t_k) &= g((\Gamma_g^T \tilde{b}_1)^T t_1, \ldots, (\Gamma_g^T \tilde{b}_k)^T t_k) \\
\delta &= \Gamma^{-1}(\beta - \tilde{a})
\end{align*}
\]

(85)

and \( \Gamma_h^T \tilde{a}_i \) (or \( \Gamma_g^T \tilde{b}_i \)) are coordinates of \( \tilde{a}_i^T \tilde{x} \) (or \( \tilde{b}_i^T \tilde{x} \)) in the corresponding basis of \( y \)'s. For any matrix \( B \), \( B^+ \) denotes pseudo-inverse, \( B^+ = (B^T B)^{-1} B^T \), which is well defined since \( \Gamma_h, \Gamma_g \) are of full rank and \( n > k, k' \) (since \( \text{rank}(B^T B) = \text{rank}(B) \), see [2] page 20).

Equation 84 can be further transformed to:

\[
\hat{h}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_k) = \sum_{i=k+k'-k^*+1}^{n} \delta_i y_k + \hat{g}(y_1, \ldots, y_{k^*}, y_{k^*+1}, \ldots, y_{k+k'-k^*})
\]

(86)
where
\[
\tilde{h}(t_1, \ldots, t_k) = \tilde{h}(t_1, \ldots, t_k) - \sum_{i=1}^{k} \delta_i t_i
\]
\[
\tilde{g}(t_1, \ldots, t_k) = \tilde{g}(t_1, \ldots, t_k) + \sum_{i=1}^{k} \delta_i + \delta_{k+1}
\]
(87)

Now it is clear from Equation 86 that \(\delta_i = 0\) for \(i = k + k' - k^* + 1, \ldots, n\) and functions \(\tilde{h}\) and \(\tilde{g}\) depend only on variables \(y_1, \ldots, y_k\). Going back to original variables \(x_1, \ldots, x_n\) we can obtain the solution to the original system 81, which is:
\[
f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i x_i + \tilde{h}(\gamma_1^T \bar{x}, \ldots, \gamma_k^T \bar{x})
\]
(88)

where \(\{\lambda_i\}_{i=1}^{n}\) are arbitrary constants, \(\tilde{h}\) is an arbitrary function of \(k^*\) variables and \(\gamma_1^T \bar{x}, \ldots, \gamma_k^T \bar{x}\) is the basis for \(V_h \cap V_b\).

**A.4 The Proof of Theorem 6**

Lemmas 3 and 4 state that in order to find all distributions that are dictated by global independence assumption it is enough to find a general solution for Equation 10. Equation 10 satisfies the scale-independence conditions of Lemma 12, thus its general solution can be found by treating \(\theta_{\bar{x}}\) as independent variables.

Equations 10 for the binary \(X\) can be written as (normalization constant is inside \(H_i\)):
\[
p(\theta_{\bar{x}} | \bar{x} \in \{0,1\}^n) = \prod_{\bar{x} \in \{0,1\}^n} \theta_{\bar{x}}^{\bar{x} \cdot \bar{x}} H_i \left( \left\{ \theta_{\bar{x}} | \bar{x} \in \{0,1\}^n, \left| \bar{y} \right| \neq 1 \right\} \right)
\]
(89)

Applying logarithm to both sides of Equations 10 and to arguments of \(H_i\) and changing the variables to \(\ln \theta_{\bar{x}} | \bar{x} \in \{0,1\}^n\) we get Equations 13 with:
\[
\tilde{b}_{ij}[m] = \begin{cases} 
1 & \bar{x} \cdot x_i = \bar{y}, x_i = 0 \text{ or } \bar{x} \cdot x_i = \bar{y} + 1, x_i = 1 \\
-1 & \bar{x} \cdot x_i = \bar{y}, x_i = 1 \text{ or } \bar{x} \cdot x_i = \bar{y} + 1, x_i = 0 \\
0 & \text{otherwise}
\end{cases}
\]
(90)

where \(1 \leq i \leq n, 0 \leq j \leq 2^{n-1} - 2, 0 \leq m \leq 2^n - 1\) and \(\bar{y}, \bar{x}\) denote the binary representations of \(j\) and \(m\) respectively. The application of Lemma 5 and the following lemma (for \(k = n\)) concludes the proof.

**Lemma 13** Given \(\tilde{b}_{ij}\) as specified by Equation 90,
\[
\bigcap_{i=1}^{k} B_i = A_k, \quad \text{for } k = 2, \ldots, n
\]
(91)

where \(B_i = \text{span}(\tilde{b}_{i,0}, \ldots, \tilde{b}_{i,2^{n-1}-2})\) and \(A_k = \text{span}(\tilde{a}_{k,0}, \ldots, \tilde{a}_{k,2^{n-1}-1})\), s.t.
\[
\tilde{a}_{k,\bar{y}}[\bar{x}] = \begin{cases} 
1 & \bar{x} \cdot [k+1, \ldots, n] = \bar{y}, \bar{x} \cdot [1, \ldots, k] \equiv 0 \pmod{2} = 0 \\
-1 & \bar{x} \cdot [k+1, \ldots, n] = \bar{y}, \bar{x} \cdot [1, \ldots, k] \equiv 1 \pmod{2} = 1 \\
0 & \text{otherwise.}
\end{cases}
\]
(92)

We use \(\bar{y}\) and \(\bar{x}\) to denote binary representation of indexes: \(0 \leq \bar{y} \leq 2^{n-k} - 1\) and \(0 \leq \bar{x} \leq 2^n - 1\).

**Proof:** By induction on \(k\).

**Induction Basis:** Consider \(M_1, M_2\) matrices with rows \(\{\tilde{b}_{1,j}\}_{j=0}^{2^n-1-2}\) and \(\{\tilde{b}_{2,j}\}_{j=0}^{2^n-1-2}\) respectively:
\[
M_1 = \begin{bmatrix} 
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & & & & & & & & & & & 
\end{bmatrix}
\]
(93)
and

\[ M_1 = \begin{bmatrix}
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & \ldots & 0 \\
\vdots & & & & & & & & & & & & \\
\end{bmatrix} \]  \quad (94)

Note that the same pattern as in the first two rows, repeats itself with 4-column shift for all \(2^n-1 - 1\) rows of \(M_1\) and \(M_2\) except the very last row. Considering the matrix \(M = [M_1^T M_2^T]^T\), with row space equal to \(B_1 + B_2\), we can see that it is exactly of rank \(3 \cdot 2^{n-3} - 2\), since rows \(\{b_{1,j}\}_{j=1,3,\ldots,2^n-3}\) and \(\{b_{2,j}\}_{j=1,3,\ldots,2^n-3}\) are linearly independent and \(\tilde{b}_{1,j} = \tilde{b}_{2,j}\) for \(j = 0, 2, \ldots, 2^n-2\). Thus, \(\text{dim}(B_1 \cap B_2) = \text{dim}(B_1) + \text{dim}(B_2) - \text{dim}(B_1 + B_2) = (2^n - 1) + (2^n - 1) - (3 \cdot 2^{n-2} - 2) = 2^{n-2}\) and the basis for \(B_1 \cap B_2\) is specified by \(\tilde{a}_{2,3}, \tilde{a}_{2,1,2}\), for \(i = 0, \ldots, 2^{n-3} - 1\).

**Induction Step:** We will demonstrate the correctness of the induction step for \(k = 3\). The proof is easily extended for arbitrary \(k\).

Consider matrices \(M_{A_3} M_{B_3}\) with rows \(\{\tilde{a}_{3,j}\}_{j=0}^{2^n-1}\) and \(\{\tilde{b}_{3,j}\}_{j=0}^{2^n-1}\) respectively, i.e.

\[ M_A = \begin{bmatrix}
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & & & & & & & & & & & & \\
\end{bmatrix} \]  \quad (95)

and

\[ M_B = \begin{bmatrix}
1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & & & & & & & & & & & & \\
\end{bmatrix} \]  \quad (96)

The first 4-row pattern of \(M_{B_3}\) repeats itself with 8-column shift for all \(2^n-1 - 1\) rows of \(M_{B_3}\) (the last pattern is truncated to 3 rows). For a general \(M_{A_k}\) matrix - the pattern of the first \(2^k-1\) rows will repeat itself with shift of \(2^k\) columns, which is twice the length of non-zero segment in \(\{\tilde{a}_{k-1,j}\}\).

It is clear now, that for the matrix \(M = [M_{A_3}^T M_{B_3}^T]^T\) the rows \(\{\tilde{b}_{3,j}\}_{j=0}^{2^n-1}\) and \(\{\tilde{a}_{3,j}\}_{j=1,3,\ldots,2^n-1}\) are linearly independent and \(\tilde{a}_{2,j} = \tilde{b}_{3,j} = \tilde{b}_{3,j+2} + \tilde{a}_{3,j+1}\), for \(j = 0, 2, \ldots, 2^{n-2} - 2\). Thus, \(\text{dim}(A_2 \cap B_3)\) is equal to number of dependent rows in \(M\), i.e. \(2^n-3\), and the basis for \(A_3\) is \(\tilde{a}_{5,j} = \tilde{a}_{2,2,j} - \tilde{a}_{2,2,j+1} = \tilde{b}_{3,2,j} - \tilde{b}_{3,2,j+2}\), for \(j = 0, \ldots, 2^{n-3} - 1\).

For the general \(k = 3, \ldots, n\), the rows \(\{\tilde{b}_{k,j}\}_{j=0}^{2^n-1}\) and \(\{\tilde{a}_{k,j}\}_{j=1,3,\ldots,2^n-(n+1)}\) are linearly independent and \(\tilde{a}_{k,j} = \tilde{a}_{k-1,j} - \tilde{a}_{k-1,j+1} = \sum_{j=0}^{2^n-3} (-1)^j \tilde{b}_{k,2,j+3j}\) for \(j = 0, \ldots, 2^{n-k} - 1\). \(\blacksquare\)

### A.5 The Proof of Theorem 9 for Two Node Networks

In this section we would like to complete the proof of Theorem 9 by solving the following system of functional equations that correspond to assumptions of Theorem 9 for arbitrary two node network (similar to Equation 1):

\begin{align*}
p(\{\theta_{ij}\}) = f_1(\{\theta_{ij}\}_{i=1,\ldots,n-1})g_1(\{\theta_{ijk}\}_{j=1,\ldots,k-1}) \\
p(\{\theta_{ij}\}) = f_2(\{\theta_{ij}\}_{j=1,\ldots,k-1}) \prod_{j=1}^{k} h_j(\{\theta_{ij}\}_{i=1,\ldots,n-1}) \quad (97)
\end{align*}

(Jacobians are “hided” inside \(f_1, f_2, g_1, h_1, \ldots, h_k\)).

Any solution \(p\) that satisfies Equation 97 satisfies also Equation 1 and thus can be written in form given by Theorem 2 (Equation 2), so we have (from Equations 97 and 2):

\[ C \left( \prod_{i=1}^{n} \prod_{j=1}^{k} \theta_{ij}^{a_{ij}} \right) H \left( \left\{ \frac{\theta_{ij+1,j+1} \ldots \theta_{ij+k-1,j+k-1}}{\theta_{i+1,j+1} \ldots \theta_{i+k-1,j+k-1}} \right\} \right) = f_1(\{\theta_{ij}\}_{i=1,\ldots,n-1}) \prod_{j=1}^{k} h_j(\{\theta_{ij}\}_{i=1,\ldots,n-1}) \quad (98) \]
Changing the variables on the left hand side to $\{\theta \}_{j}$ and $\{\theta \}_{ij}$, we have:

$$
\sum_{j=1}^{k} \prod_{i=1}^{n} \left( \sum_{j=1}^{\omega_i} \alpha_i \right) \left[ w_{ji} \right]^{\alpha_i} \right] H \left( \left( \frac{w_{ji+1}}{w_{ji+1}} \right)^{\omega_i} \right) = f_{2} \left( \left\{ x_{j} \right\}_{j=1}^{n}, \left\{ \omega_{ji} \right\}_{j=1}^{n} \right) \prod_{j=1}^{k} \ h_{j} \left( \left\{ w_{ji} \right\}_{i=1}^{n-1} \right) \quad (99)
$$

where $x_{j} \triangleq \theta_{j}$ and $w_{ji} \triangleq \theta_{ij}$ (similar to Section 3.1). Clearly $f_{2} \left( \left\{ x_{j} \right\}_{j=1}^{n}, \left\{ \omega_{ji} \right\}_{j=1}^{n} \right) = \prod_{j=1}^{k} x_{j}^{\omega_{ji}}$. Eliminating $\{x_{i}\}$, hiding $C$ inside $H$, $\{\omega_{ji}\}$ inside $h_{j}$ and taking the logarithm, we get:

$$
\hat{H} \left( \left( \frac{w_{ji+1}}{w_{ji+1}} \right)^{\omega_i} \right) = \sum_{j=1}^{k} \ h_{j} \left( \left\{ w_{ji} \right\}_{i=1}^{n-1} \right) \quad (100)
$$

where

$$
\hat{H}(t_{1,1}, \ldots , t_{(k-1),(n-1)}) = \ln \left[ C \cdot H(t_{1,1}, \ldots , t_{(n-1),(k-1)}) \right] \\
\hat{h}_{j}(t_{j,1}, \ldots , t_{j,n-1}) = \ln \left[ \prod_{i=1}^{\omega_i} \ h_{j} \left( t_{j,1}, \ldots , t_{j,n-1} \right) \right] \quad 1 \leq j \leq k.
$$

(101)

Changing the variables to $\tilde{w}_{ji} = \ln(\frac{w_{ji}}{w_{ji+1}})$ (this mapping is one-to-one, since $w_{jn} = 1 - \sum_{i=1}^{n-1} w_{ji}$), we have:

$$
\hat{H} \left( \left\{ \tilde{w}_{ji} \right\}_{j=1}^{n}, \left\{ \tilde{w}_{ji+1} \right\}_{j=1}^{n-1} \right) = \sum_{j=1}^{k} \ h_{j} \left( \left\{ \tilde{w}_{ji} \right\}_{i=1}^{n-1} \right) \quad (102)
$$

where $\hat{H}$ and $\hat{h}$ are such that:

$$
\hat{H}(t_{1,1}, \ldots , t_{(k-1),(n-1)}) = \hat{H}(\ln t_{1,1}, \ldots , \ln t_{(k-1),(n-1)}) \\
\hat{h}_{j}(t_{j,1}, \ldots , t_{j,n-1}) = \hat{h}_{j}(\ln t_{j,1}, \ldots , \ln t_{j,n-1}) \ln \left( \frac{t_{j,n-1}}{1 - \sum_{i=1}^{n-1} t_{ji}} \right)
$$

(103)

Taking derivatives wrt $\{\tilde{w}_{ji}\}$ we get (for $1 \leq i \leq n$):

$$
\frac{\partial \hat{H}}{\partial \tilde{w}_{ji}} = \frac{\partial \hat{h}_{j}}{\partial \tilde{w}_{ji}}, \quad j = 1 \\
\frac{\partial \hat{H}}{\partial \tilde{w}_{ji}} = \frac{\partial \hat{h}_{j}}{\partial \tilde{w}_{ji}}, \quad 2 \leq j \leq k - 1 \\
- \frac{\partial \hat{H}}{\partial \tilde{w}_{j-1,i}} = \frac{\partial \hat{h}_{j}}{\partial \tilde{w}_{ji}}, \quad j = k 
$$

(104)

where $t_{ji} = \frac{\tilde{w}_{ji}}{\tilde{w}_{ji+1}}$ (for $1 \leq i \leq n - 1, 1 \leq j \leq k - 1$) denote the arguments of $\hat{H}$. Summing Equations 104 we get (for all $1 \leq i \leq n$):

$$
\sum_{j=1}^{k} \frac{\partial \hat{h}_{j}}{\partial \tilde{w}_{ji}} (\tilde{w}_{j1}, \ldots , \tilde{w}_{jn-1}) = 0 \quad (105)
$$

Since arguments of all $h_{j}$’s are distinct, we get:

$$
\frac{\partial \hat{h}_{j}}{\partial \tilde{w}_{ji}} (\tilde{w}_{j1}, \ldots , \tilde{w}_{jn-1}) = c_{ji}, \quad 1 \leq j \leq k, \quad 1 \leq i \leq n - 1
$$

(106)

where $c_{ji}$ are arbitrary constants s.t. $\sum_{j=1}^{k} c_{ji} = 0$. From Equation 106 we get:

$$
\hat{h}_{j}(\tilde{w}_{j1}, \ldots , \tilde{w}_{jn-1}) = c_{j} + \sum_{i=1}^{n-1} c_{ji} \tilde{w}_{ji}, \quad 1 \leq j \leq k.
$$

(107)

and changing back to the original variables:

$$
h_{j}(w_{j1}, \ldots , w_{jn-1}) = \beta_{i} \prod_{i=1}^{n} w_{ji}^{\alpha_{i}}, \quad 1 \leq j \leq k
$$

(108)
where \( b_j, b_{j1}, \ldots, b_{jn} \) are appropriate constants. Combining the results of Equation 108, Equation 99 and Equation 97, we get:

\[
p(\theta_{ij}) = C' \prod_{i=1}^{n} \prod_{j=1}^{k} \theta_{ij}^{\theta_{ij}}
\]

(109)

i.e. \( \{b_{ij}\} \) are distributed Dirichlet.

**B Matlab Code**

This appendix presents a matlab code for determining the symbolic form of an arbitrary globally independent distribution (Section 4.2).

```matlab
function [Mjoint] = joint_pairwise_space(nodes);
% Input: 'nodes' - list of number of states of each variable.
% Output: 'Mjoint' - each column corresponds to one argument of
% the arbitrary function.
% Mjoint[i,j] describes the role of \( \theta_i \) in argument #j.
% 1 - \( \theta_i \) is in the nominator of argument #j
% -1 - \( \theta_i \) is in the denominator of argument #j
% 0 - \( \theta_i \) doesn't participate in argument #j

% Description agreement:
% States are counted from 0.
% State A_1(n1,n2,...,n_n) will be described in place
% \sum_{i=1}^{n} n_i = tailprod(i)/nodes[i]
% where tailprod(i) = \prod_{j=i+1}^{n} nodes[j]

n = size(nodes,2);
tailprod = fliplr(cumprod(fliplr(nodes)))./fliplr(nodes);
Mjoint = [];
for p = 1:n
    % construct descriptions of pairwise solutions:
    % node #p vs. all other nodes.
    M = [];
    scount = prod(nodes)/nodes(p);
    for i = 1:nodes(p)-1;
        state = zeros(size(nodes));
        state(p) = i-1;
        for j = 1:scount-1;
            % building states
            nstate = next_state(nodes,state,[p]);
            state = state;
            state(p) = i;
            nstate1 = nstate;
            nstate1(p) = i;
            % building description.
            desc = zeros(1,prod(nodes));
            desc(state*tailprod + 1) = 1;
            desc(nstate1*tailprod + 1) = 1;
            desc(state1*tailprod + 1) = -1;
            desc(nstate1*tailprod + 1) = -1;
```
M = [M, desc'];

% moving to next state.
state = nstate;
end
if isempty(Mjoint)
Mjoint = M;
else
m = null([Mjoint, M], 'r');
if isempty(m)
disp('no free function');
return
end
m = m(1:size(Mjoint,2),:);
Mjoint = Mjoint*m;
end
return

function [out] = next_state(nodes, state, exclude_list);
out = state;
for i = size(nodes,2):-1:1;
    if isempty(exclude_list) | sum(exclude_list == i) == 0
        out(i) = out(i) + 1;
        if out(i) >= nodes(i);
            out(i) = 0;
        else
            break;
        end
    end
end
return

References


