Research Article

Local Structure Recovery of Chain Graphs after Marginalization

Qiang Zhao,^{1,2} Panpan Luo,¹ and Qing Xi¹

¹*School of Mathematical Sciences, Shandong Normal University, Jinan 250014, China*

²*School of Mathematical Sciences, Peking University, Beijing 100871, China*

Correspondence should be addressed to Qiang Zhao; zhaoqmath@gmail.com

Received 15 October 2014; Accepted 6 March 2015

Academic Editor: Bo Yang

Copyright © 2015 Qiang Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper discusses local structure recovery of chain graphs (CGs) when there exist unobserved or latent variables or after marginalization over observed variables. Under a condition presented in this paper, it is explained which edges and directions of edges in local structure can be recovered validly and which cannot after marginalization.

1. Introduction

Graphical models, also known as Markov networks and Bayesian networks, including independence graphs, directed acyclic graphs (DAGs), and chain graphs (CGs) have been applied widely to many fields, such as stochastic systems, data mining, pattern recognition, artificial intelligence, and causal discovery. Chain graphs (CGs) are widely used to represent independence, conditional independence, and causal relationships among the random variables [1–5]. Structure recovery of CGs has been discussed by many authors [4–6]. A statistical conclusion may be reversed after marginalization over some variables, which is called Yule-Simpson paradox [7, 8]. For sampling design, prior knowledge or assumptions on models are necessary for valid structural learning of CGs and parameter estimates since some variables may be unobserved, such as the faithfulness assumption [4, 5] and collapsibility [9, 10]. On the other hand, for data analysis, conditions are necessary for marginalizing over some observed variables. Various conditions have been presented for avoiding a statistical conclusion reversion about association and parameters of linear models [10–14]. Collapsibility of parameter estimates for undirected graphical models over some variables has been discussed in [9, 15–17], and that of DAGs has been discussed in [18, 19].

In this paper, we discuss local structure recovery for a CG when there exist unobserved or latent variables or after marginalization over observed variables. Suppose that there is an unknown true CG with a large number of variables but

we may be interested in construction of a local structure of the CG from a marginal distribution of a subset of variables. We present a condition for this localized recovery and explain which edges and directions of edges in local structure can be recovered validly from the marginal distribution and which edges may be spurious. We say that an edge or a direction is recovered validly from the marginal distribution if it is the same as that recovered from the joint distribution. The condition is useful for both sampling design and data analysis. This localization of structure recovery is related to identification and collapsibility.

Section 2 gives notation and definitions. In Section 3, we present theoretical results on local structure recovery. Finally a discussion is given in Section 4.

2. Notation and Definitions

In this section, we briefly introduce terminologies and notations on graph theory. Readers can refer to [3, 20] for more details.

Let $\mathcal{G}_V = (V, E_V)$ be a graph with a vertex set V and an edge set E_V . We say that there is an undirected edge, or a line, between vertex v and vertex u (denoted as $v - u$), if $\langle v, u \rangle \in E_V$ and $\langle u, v \rangle \in E_V$; and there is a direct edge, or an arrow, from vertex v to u (denoted as $v \rightarrow u$), if $\langle v, u \rangle \in E_V$ and $\langle u, v \rangle \notin E_V$. Chain components of $\mathcal{G}_V$ are obtained by removing all arrows in $\mathcal{G}_V$ and taking the connectivity components of the remaining undirected graph. A chain

graph with only undirected edges is known as an undirected graph (UG). A chain graph with only directed edges and without any directed cycles is known as a directed acyclic graph (DAG).

If $v \rightarrow u$, then we call v a parent of u and u a child of v . In general, we use $\text{pa}_{\mathcal{G}}(u)$ and $\text{ch}_{\mathcal{G}}(u)$ to represent the collection of parents and children of u , respectively. If $v - u$, then we call v a neighbour of u . We use $\text{ne}_{\mathcal{G}}(u)$ to represent the set of neighbours of vertex u in $\mathcal{G}_V$. If there is an edge between the vertices v and u , then we say that v and u are joined or adjacent. The family of u contains u and its parents, which is denoted as $\text{fa}_{\mathcal{G}}(u)$. For a set $A \subseteq V$, we can define similarly $\text{pa}_{\mathcal{G}}(A) = \bigcup_{u \in A} \text{pa}(u) \setminus A$, $\text{ch}_{\mathcal{G}}(A) = \bigcup_{u \in A} \text{ch}(u) \setminus A$, and $\text{fa}_{\mathcal{G}}(A) = A \cup \text{pa}_{\mathcal{G}}(A)$. The boundary of u is the set of neighbours and parents of u , which is denoted as $\text{bd}_{\mathcal{G}}(u)$. What is more, the boundary of a set $A \subseteq V$ is defined as $\text{bd}_{\mathcal{G}}(A) = (\text{pa}_{\mathcal{G}}(A) \cup \text{ne}_{\mathcal{G}}(A)) \setminus A$. In case that the underlying CG is clearly specified, the subscript $\mathcal{G}$ is often ignored to simplify the notations.

If there is a sequence of the vertices $u_0, u_1, \dots, u_{k-1}, u_k$, $k \geq 0$, with $\langle u_{i-1}, u_i \rangle \in E_V$ or $\langle u_i, u_{i-1} \rangle \in E_V$, for all $1 \leq i \leq k$, then we call it a route and u_0 and u_k the ends of the route. Furthermore if for all $1 \leq i \leq k$ there is $\langle u_{i-1}, u_i \rangle \in E_V$, we say the route is descending and use $v \mapsto u$ to denote a descending route from vertex v to u . If the vertices in a route are all distinct from each other, then we call the route a path. If there is a descending path from vertex v to u , then we call v an ancestor of vertex u and u a descendant of vertex v , which are denoted as $\text{an}(u)$ and $\text{de}(v)$, respectively. Similarly, the ancestral set of a set $B \subseteq V$ is defined as $\text{an}(B) = \bigcup_{u \in B} \text{an}(u) \setminus B$, and the descendant set is defined as $\text{de}(B) = \bigcup_{u \in B} \text{de}(u) \setminus B$. Furthermore, we define $\text{An}(b) = \text{an}(b) \cup \{b\}$. A vertex v without any children in $\mathcal{G}_V$ is called a terminal. Besides, if $\text{de}_{\mathcal{G}}(A) = \emptyset$, we call A a terminal set. We call a route a pseudocycle if it satisfies $u_0 = u_k$. Besides, it is called a cycle if it also satisfies $k \geq 3$ and $u_0, u_1, \dots, u_{k-1}$ are distinct vertices. A cycle or pseudocycle is directed if it is descending and has $\langle u_j, u_{j-1} \rangle \notin E_V$ for some $j \in \{1, \dots, k-1, k\}$. Finally, if graph $\mathcal{G}_V$ does not contain any directed cycles or directed pseudocycles, we will call it a CG (chain graph).

A section σ of a route $\rho = (u_0, u_1, \dots, u_{k-1}, u_k)$ in CG $\mathcal{G}_V$ means any maximum undirected subroute σ in ρ , for example, $u_j - \dots - u_i$ with $0 \leq j \leq i \leq k$. We call u_j and u_i the two ends of the section. Besides, if there is $u_{j-1} \rightarrow u_j$ in ρ for some $j > 0$ (or $u_i \leftarrow u_{i+1}$ in ρ for some $i < k$), then we call u_j (or u_i) a head-terminal, or else we call it a tail-terminal. The head-to-head section in reference to route ρ is the section which has two head-terminals, and a non-head-to-head section if it has one head-terminal at most. We use $\{u_j, \dots, u_i\}$ to denote the set of vertices of a section σ . If $\{u_j, \dots, u_i\} \cap S = \emptyset$ with $S \subseteq V$, then we consider that section σ is outside of the set S . If $\{u_j, \dots, u_i\} \cap S \neq \emptyset$, then we consider that σ is hit by set S .

Another important concept in CG is complex κ . A complex κ in CG $\mathcal{G}_V$ is an especial path $(u_0, u_1, \dots, u_{k-1}, u_k)$, $k \geq 2$, with $u_0 \rightarrow u_1$, $u_j - u_{j+1}$ ($1 \leq j \leq k-2$), and $u_{k-1} \leftarrow u_k$ in $\mathcal{G}_V$, where there is no extra edge among vertices on the path. The vertices u_0 and u_k are called the parents of complex

κ , denoted as $\text{par}(\kappa)$, and $\{u_1, u_2, \dots, u_{k-2}, u_{k-1}\}$ is called the region of κ . What is more, the complex with vertex x as one of its parents is denoted by $\kappa_{x \in \text{par}(\kappa)}$. At last, the Markov Blanket of vertex x is defined as $\text{MB}(x) = x \cup \text{ch}(x) \cup \text{ne}(x) \cup \text{pa}(x) \cup \text{par}(\kappa_{x \in \text{par}(\kappa)})$.

Next, we give the definition of c -separation in CG $\mathcal{G}_V$. If a section σ_1 of route ρ satisfies the following two presentations, (1) if σ_1 is a head-to-head section in reference to ρ and σ_1 is outside of set S , or (2) σ_1 is not a head-to-head section in reference to ρ and σ_1 is hit by set S , then we say that ρ is c -separated by the set $S \subseteq V$ in $\mathcal{G}_V$. What is more, we say that the disjoint set $A, B \subseteq V$ in $\mathcal{G}_V$ is c -separated by the disjoint set $S \subseteq V$ if every route ρ between A and B is c -separated by S , and A, B, S form a c -separation which is denoted by $\langle A, B \mid S \rangle_{\mathcal{G}_V}^{\text{sep}}$.

If two CGs share the same c -separation patterns over the same set of vertices, then we say that they are *Markov equivalent*. All chain graphs, which are Markov equivalent with each other, form equivalent class, which is known as *Markov equivalent class*. It is well known that two CGs are *Markov equivalent* if and only if they share the identical global skeleton and complexes [20].

If a probability distribution P over $X_1, X_2, \dots, X_{n-1}, X_n$ permits the following factorization [3],

$$P(x) = \prod_{\tau \in T} P(x_{\tau} \mid x_{\text{pa}(\tau)}), \quad (1)$$

where T is the collection of chain components of CG $\mathcal{G}_V$ and after given $x_{\text{pa}(\tau)}$, the conditional probability distribution of x_{τ} is denoted as $P(x_{\tau} \mid x_{\text{pa}(\tau)})$; then we say that P is a compatible probability distribution in reference to $\mathcal{G}_V$. It is easy to check that if P is a compatible probability distribution in reference to CG $\mathcal{G}_V$, we have

$$\langle A, B \mid S \rangle_{\mathcal{G}_V}^{\text{sep}} \implies X_A \perp\!\!\!\perp X_B \mid X_S, \quad (2)$$

where $X_A \perp\!\!\!\perp X_B \mid X_S$ represents the conditional independence between X_A and X_B given X_S in P . If the condition is strengthened to

$$\langle A, B \mid S \rangle_{\mathcal{G}_V}^{\text{sep}} \iff X_A \perp\!\!\!\perp X_B \mid X_S, \quad (3)$$

then we say that P is faithful to $\mathcal{G}_V$. Let $M(\mathcal{G}_V)$ be the family of compatible probability distributions in reference to CG $\mathcal{G}_V$. A chain graph model is defined as $\mathcal{T} = (\mathcal{G}_V, \mathcal{P})$, where $\mathcal{P}$ is a family of probability distributions that are compatible with $\mathcal{G}_V$.

Example 1. Consider a CG $\mathcal{G}_V = (V, E_V)$ in Figure 1, where $V = \{a, \dots, j\}$ and $E_V = \{\langle a, b \rangle, \langle a, c \rangle, \langle a, d \rangle, \langle b, e \rangle, \langle c, e \rangle, \langle c, f \rangle, \langle d, f \rangle, \langle e, g \rangle, \langle f, h \rangle, \langle g, e \rangle, \langle g, h \rangle, \langle h, j \rangle, \langle i, g \rangle, \langle j, h \rangle\}$. We have $\text{pa}(g) = \{i\}$, $\text{ch}(g) = \{h\}$, $\text{ne}(g) = \{e\}$, $\text{fa}(g) = \{g, i\}$, $\text{bd}(g) = \{i, e\}$, $\text{an}(g) = \{a, b, c, e, i\}$, $\text{de}(g) = \{h, j\}$, and $\text{MB}(g) = \{e, i, h, f\}$. $\kappa = (b, e, g, i)$ and $\kappa' = (g, h, f)$ are complexes, where $\text{par}(\kappa) = \{b, i\}$ and $\text{par}(\kappa') = \{g, f\}$, respectively. The path $l = (a, c, f, h)$ between a and h is c -separated by $\{c\}$ or $\{f\}$, while the path $l' = (a, b, e, g, i)$ between a and g is c -separated by the empty set. The sets $A = \{a\}$ and $B = \{e, f, g, h, i, j\}$ are c -separated by the set $C = \{b, c, d\}$, and A, B, S form a c -separation $\langle A, B \mid S \rangle_{\mathcal{G}_V}^{\text{sep}}$.

FIGURE 1: A chain graph $\mathcal{G}_V$.

3. Local Structure Recovery after Marginalization

Let A , B , and C be a partition of all variables in V . In this section, we assume that $A \perp\!\!\!\perp B \mid C$ and suppose that variables in B are omitted or unobservable. The assumption of conditional independence has been discussed as one of conditions for collapsibility of parameter estimates over unobserved variables [9–12, 14, 18, 19]. Collapsibility of parameter estimates further requires another condition that the separator C is a complete subgraph in the moral graph. A decomposition approach of structural learning proposed in [21] also requires the condition of a complete separator. In this section, we show under the assumption of conditional independence but without the condition of a complete separator that the local structure of CG $\mathcal{G}_V$ over $A \cup C$ can be partially recovered from the marginal distribution of observed variables in $A \cup C$. In many practical applications, the conditional independence $A \perp\!\!\!\perp B \mid C$ can be judged with domain or prior knowledge, such as Markov chain, chain graphical models, and dynamic or temporal models [1]. When all variables in the full set V are observed, we can first construct an undirected independence graph $\bar{G}$ over V from the observed data; then we find a set C which separates A and B in the undirected graph $\bar{G}$, and thus we have that $A \perp\!\!\!\perp B \mid C$ holds [2, 3].

Two CGs have the same Markov property if and only if they have the same skeleton (i.e., an undirected version of $\mathcal{G}_V$) and the same complexes [20]. Thus for recovering the structure of a CG, we can only learn the skeleton and complexes from a distribution of observed variables. A marginal distribution obtained from the distribution P with the Markov property of $\mathcal{G}_V$ may not obey the Markov property of any CG. Although the class of CGs is not closed under marginalization in this sense, we show that local structure of a CG may be partially recovered under some conditions. First we discuss which edges of local structure over $A \cup C$ can be recovered validly and which cannot from a marginal distribution of $A \cup C$.

Theorem 2. *Two vertices in A are c -separated by a subset of V if and only if they are c -separated by a subset of $A \cup C$.*

Proof. The sufficiency is obvious since $V \supseteq A \cup C$. For the necessity, let a and a' be two vertices in A that are c -separated by $S(\subseteq V)$. Thus there is no edge connecting a and a' in

CG $\mathcal{G}_V$. Since a and a' are contained in A , $bd(a)$ and $bd(a')$ are contained in $A \cup C$, otherwise $A \not\perp\!\!\!\perp B \mid C$. Without loss of generality, suppose that a is not an ancestor of a' . Thus we have that $bd(a) \subseteq A \cup C$ c -separates a and a' . $\square$

From Theorem 2, we can see that the existence of edges falling into A can be determined validly from the marginal distribution of $A \cup C$.

Theorem 3. *Let a and c be two vertices in A and C , respectively. Then a and c are c -separated by a subset of V if and only if they are c -separated by a subset of $A \cup C$.*

Proof. The sufficiency is trivial. For the necessity, we show below that $S = [(An(a) \cup An(c)) \cap (A \cup C)] \setminus \{a, c\}$ can c -separate every route connecting a and c . We consider the following two cases separately:

- (1) a route is contained in $An(a) \cup An(c)$,
- (2) a route is not completely contained in $An(a) \cup An(c)$.

We first discuss case (1). Let $\rho = (a, u, v, \dots) \subseteq An(a) \cup An(c)$. Note that u must not be the vertex c since there is no edge between a and c , but v may be c . Further we consider two subcases:

- (1.1) u is not contained in any head-to-head section,
- (1.2) u is contained in some head-to-head section.

For subcase (1.1), we have $u \in A \cup C$ for $a \in A$ and thus $u \in S$ for $\rho \subseteq An(a) \cup An(c)$. So the path ρ is c -separated by node u .

For subcase (1.2), let us suppose the head-to-head section on route ρ is $a \rightarrow u - v - \dots - b_1 \leftarrow b$, where b is the other parent of this head-to-head section. Since ρ is contained in $An(a) \cup An(c)$, then $b_1 \in An(a) \cup An(c)$; thus we have the fact that the node b must not be c . Otherwise, $b_1 \in ch(c)$ and $b_1 \in de(a)$, which is contradictory to $b_1 \in An(a) \cup An(c)$. Because $A \perp\!\!\!\perp B \mid C$ and $a \in A$, we have that $b \notin B$. Thus $b \in A \cup C$ and then $b \in S = [(An(a) \cup An(c)) \cap (A \cup C)] \setminus \{a, c\}$. Thus we have that the route ρ is c -separated by node b .

Now we consider case (2) and show first that such a route contains a head-to-head section. Let $\rho = (a, \dots, m, u, \dots, v, n, \dots, c)$ such that m and n are contained in $An(a) \cup An(c)$ but no vertices from u to v are contained in $An(a) \cup An(c)$ since ρ is not completely contained in $An(a) \cup An(c)$. We know that the arrows must be oriented as $m \rightarrow u$ and $n \rightarrow v$ since $m, n \in An(a) \cup An(c)$ and $u, v \notin An(a) \cup An(c)$. Then there must be a head-to-head section between m and n on ρ , and none of its descendants is in $An(a) \cup An(c)$. Thus vertices on the region of this head-to-head section and its descendants are not in $S = [(An(a) \cup An(c)) \cap (A \cup C)] \setminus \{a, c\}$, and we obtain that ρ is c -separated by S . We proved this theorem. $\square$

According to Theorem 3, the existence of edges crossing A and C can also be determined validly from the marginal distribution of $A \cup C$.

Example 1 (continued). Consider again the CG in Figure 1. Let $A = \{e, f, g, h, i, j\}$, $B = \{a\}$, and $C = \{b, c, d\}$. We

FIGURE 2: A local skeleton recovered from the marginal distribution of $\{b, \dots, j\}$.

FIGURE 3: A local skeleton recovered from the marginal distribution of $\{a, b, c, d\}$.

have $A \perp\!\!\!\perp B \mid C$. Thus according to Theorems 2 and 3, we can obtain a local skeleton from the marginal distribution of $A \cup C$, which may have spurious edges, as shown in Figure 2. Similarly, suppose that variables in $\{a, b, c, d\}$ are observed but those in $\{e, \dots, j\}$ are not observed. Then we obtain the local skeleton from the marginal distribution of $\{a, b, c, d\}$, as shown in Figure 3. Note that the edges (b, c) , (b, d) , and (c, d) falling into C are spurious in Figure 2, but they are absent in Figure 3.

Next we discuss recovery of complexes from a marginal distribution of $A \cup C$. We say that a complex $(a, a_1, \dots, a_k, c)$ ($k \geq 1$) can be determined validly from the marginal distribution of $A \cup C$ if the marginal distribution has the following two conditions: (1) $a \perp\!\!\!\perp c \mid S$ for some $S \subseteq (A \cup C)$ and (2) $a \not\perp\!\!\!\perp c \mid S \cup \{a_1\}$ and $a \not\perp\!\!\!\perp c \mid S \cup \{a_k\}$. Below we discuss which complexes of local structure can be determined validly from the marginal distribution and which cannot. The following two theorems show conditions for determining complexes and for validity of determined complexes, respectively.

Theorem 4. *If at most one vertex of a complex is not contained in A , then the complex can be determined from the marginal distribution of $A \cup C$.*

Proof. Let $(a, a_1, \dots, a_k, c)$ ($k \geq 1$) be a complex where a and c are the parents of this complex and at most one of these vertices is not contained in A . It follows from $A \perp\!\!\!\perp B \mid C$ that $\{a, a_1, \dots, a_k, c\} \subseteq A \cup C$. From Theorems 2 and 3, we can

validly determine the presence of edges $(a, a_1), (a_1, a_2), \dots, (a_k, c)$ and find a subset S of $A \cup C$ that c -separates a and c and does not contain the vertices from a_1 to a_k . Thus this complex $(a, a_1, \dots, a_k, c)$ can be determined from the marginal distribution of (A, C) . $\square$

Theorem 5. *If two parents of a complex are contained in A , then the directions of the complex can be determined from the marginal distribution of $A \cup C$.*

Proof. Let $(a, a_1, a_2, \dots, c_2, c_1, c)$ be a complex where a and c are the parents of this complex and $\{a, c\} \subseteq A$. It follows from $A \perp\!\!\!\perp B \mid C$ that $\{a_1, c_1\} \subseteq A \cup C$ (a_1 and c_1 may be the same vertex). Rewrite this complex as $(a, a_1, \dots, a_i, \dots, c_j, \dots, c_1, c)$, where $(a, a_1, \dots, a_i) \cup (c_j, \dots, c_1, c) \subseteq A \cup C$ and $(a_{i+1}, c_{j+1}) \subseteq B$. Then we can determine the presence of edge (a_i, c_j) from the marginal distribution of (A, C) . From Theorems 2 and 3, we can validly determine the presence of edges $(a, a_1), (a_1, a_2), \dots, (a_{i-1}, a_i), (c_j, c_{j-1}), \dots, (c_2, c_1), (c_1, c)$ as well as finding a subset S of $A \cup C$ that c -separates a and c and does not contain the vertex a_1 nor c_1 . Thus the directions of this complex, $a \rightarrow a_1$ and $c \rightarrow c_1$, can be determined from the marginal distribution of (A, C) . $\square$

According to Theorem 4, from the marginal distribution of $A \cup C$, we can determine all complexes in the local structure which have at most one vertex in C . When two parents of a complex are contained in C , it may not be determined since the two parents may not be c -separated by any subset of $A \cup C$. From Theorem 5, however, if a complex is determined from the marginal distribution of $A \cup C$ whose two parents are contained in A , the directions of the complex must be valid. A complex who has more than one vertex fall into C may be a spurious complex since the edges falling into C may be spurious.

From Theorems 2 to 5, we can get an approach for local structure recovery, in which edges can be recovered according to Theorems 2 to 3, and then directions can be determined according to Theorems 4 and 5. Suppose that we are interested in the local structure recovery of a CG over a set A of variables. Then we must find a set C to be observed based on the domain or prior knowledge such that C is large enough to c -separate A from the set B of unobserved variables.

Example 1 (continued). Now we search complexes in Figures 2 and 3. For Figure 2, suppose that variable $\{a\}$ is not observed. According to Theorem 4, three complexes (b, e, g, i) , (c, e, g, i) , and (f, h, g) can be found from the marginal distribution of $A \cup C$, as shown in Figure 4. The complexes (c, f, d) and (b, e, c) cannot be found since there exist spurious edges between their parents, although the directions of $\langle b, e \rangle$ and $\langle c, e \rangle$ are oriented with other complexes. For Figure 3, there is no complex in this local structure because variables b, c , and d are conditional independent given variable a .

If variable $\{a\}$ is omitted or unobservable, we only obtain the local structure in Figure 4, in which all directions and all edges not falling into C are valid, but edges falling into C (here (b, c) , (b, d) , and (c, d)) may be spurious.

FIGURE 4: A partially oriented local structure over $\{b, \dots, j\}$.

4. Discussion

We showed that the conditional independence $A \perp\!\!\!\perp B \mid C$ is a sufficient condition for local structure recovery when there exist unobserved or latent variables or after marginalization over observed variables. The conditional independence is an important prior knowledge for local structure recovery. This prior may hold in many cases, such as Markov chain, chain graphical models, and dynamic or temporal models [1]. Suppose that we are interested in the local structure of a CG over a set A of variables. We must find a set C which is large enough to separate A from the set B of unobserved variables. Based on the conditional independence $A \perp\!\!\!\perp B \mid C$, we explained which edges and directions of edges in local structure can be recovered validly and which cannot after marginalization.

Based on the theoretical results presented in this paper, we can efficiently recover local structures of a CG. Domain or prior knowledge of conditional independencies can be utilized to facilitate the structural recovery. The theoretical results can also be used for an observational study design and a split questionnaire survey sampling [22, 23].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by NSFC (11001155), Doctoral Fund of Shandong Province (BS2010SW030), and A Project of Shandong Province Higher Educational Science and Technology Program (J11LA08).

References

- [1] D. R. Cox and N. Wermuth, *Multivariate Dependencies: Models, Analysis, and Interpretation*, Chapman and Hall, London, UK, 1996.
- [2] D. Edwards, *Introduction to Graphical Modelling*, Springer Texts in Statistics, Springer, New York, NY, USA, 2nd edition, 2000.
- [3] S. L. Lauritzen, *Graphical Models*, vol. 17 of *Oxford Statistical Science Series*, Oxford University Press, 1996.
- [4] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, UK, 2000.
- [5] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction and Search*, vol. 81 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1993.
- [6] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer, New York, NY, USA, 1999.
- [7] E. H. Simpson, "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society, Series B: Methodological*, vol. 13, pp. 238–241, 1951.
- [8] G. U. Yule, "Notes on the theory of association of attributes in statistics," *Biometrika*, vol. 2, no. 2, pp. 121–134, 1903.
- [9] D. Madigan and K. Mosurski, "An extension of the results of Asmussen and Edwards on collapsibility in contingency tables," *Biometrika*, vol. 77, no. 2, pp. 315–319, 1990.
- [10] N. Wermuth, "Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 49, no. 3, pp. 353–364, 1987.
- [11] D. R. Cox and N. Wermuth, "A general condition for avoiding effect reversal after marginalization," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 65, no. 4, pp. 937–941, 2003.
- [12] Z. Geng, "Collapsibility of relative risk in contingency tables with a response variable," *Journal of the Royal Statistical Society Series B: Methodological*, vol. 54, no. 2, pp. 585–593, 1992.
- [13] N. Wermuth, "Moderating effects of subgroups in linear models," *Biometrika*, vol. 76, no. 1, pp. 81–92, 1989.
- [14] A. S. Whittemore, "Collapsibility of multidimensional contingency tables," *Journal of the Royal Statistical Society, Series B: Methodological*, vol. 40, no. 3, pp. 328–340, 1978.
- [15] S. Asmussen and D. Edwards, "Collapsibility and response variables in contingency tables," *Biometrika*, vol. 70, no. 3, pp. 567–578, 1983.
- [16] M. Frydenberg, "Marginalization and collapsibility in graphical interaction models," *The Annals of Statistics*, vol. 18, no. 2, pp. 790–805, 1990.
- [17] Z. Geng, K. Wan, and T. Feng, "Mixed graphical models with missing data and the partial imputation EM algorithm," *Scandinavian Journal of Statistics: Theory and Applications*, vol. 27, no. 3, pp. 433–444, 2000.
- [18] S.-H. Kim and S.-H. Kim, "A note on collapsibility in DAG models of contingency tables," *Scandinavian Journal of Statistics. Theory and Applications*, vol. 33, no. 3, pp. 575–590, 2006.
- [19] X. Xie and Z. Geng, "Collapsibility for directed acyclic graphs," *Scandinavian Journal of Statistics*, vol. 36, no. 2, pp. 185–203, 2009.
- [20] M. Studeny, "A recovery algorithm for chain graphs," *International Journal of Approximate Reasoning*, vol. 17, no. 2-3, pp. 265–293, 1997.
- [21] Z. Geng, C. Wang, and Q. Zhao, "Decomposition of search for v -structures in DAGs," *Journal of Multivariate Analysis*, vol. 96, no. 2, pp. 282–294, 2005.
- [22] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, New York, NY, USA, 2nd edition, 2002.
- [23] S. Rässler, *Statistical Matching*, vol. 168 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 2002.