University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Damian Skrzypiec
Student no.: 320335

Structure Learning Algorithms for Chain Graphs

Master's thesis
in MATHEMATICS

Supervisor:
John Noble, PhD.
Institute of Applied Mathematics and Mechanics

April 2017

Supervisor's statement

Hereby I confirm that the present thesis was prepared under my supervision and that it fulfils the requirements for the degree of Master of Mathematics.

Date Supervisor's signature

Author's statement

Hereby I declare that the present thesis was prepared by me and none of its contents was obtained by means that are against the law. The thesis has never before been a subject of any procedure of obtaining an academic degree. Moreover, I declare that the present version of the thesis is identical to the attached electronic version.

Date Author's signature

Abstract

In this place will be abstract of this project.

Keywords

graphical model, chain graph, structure learning

Thesis domain (Socrates-Erasmus subject area codes)

11.2 Statistics

Subject classification

62 Statistics
62C10 Bayesian Problems

Title of the thesis in Polish

Contents

Chapter 1

Introduction

The purpose of this project is to present algorithms for learning conditional independence structure of joint probability distributions represented by chain graphs. This is a special case of learning probabilistic graphical models which provides convenient representation of factorisation probability distribution using graphs. Two most common classes of probabilistic graphical models (PGMs) are Bayesian Networks where PGM is represented by directed acyclic graph and Markov Fields where PGM is represented by undirected graph. Chain graphs is a class of graphs that does not contains cycles (formal definition in 2.1.9). It contains both directed and undirected edges in graph representation hence it is natural generalization of Bayesian Networks and Markov Fields. Such a generalization was needed because of limitation of Markov Fields and Bayesian Networks. An edge in a Markov Field model represent that there is a correlation between two random variables but it does not specify what type of correlation it is. On the other hand Bayesian Network models contains only directed edges which represents only cause-effect relationships without possibility of existence of mutual correlation between two random variables. [TO BE CHANGED] In this paper we present one algorithm for learning chain graphs and one algorithm for learning undirected graphical models. Both algorithms are based on idea of graph decomposition which suppose to decrease complexity of algorithms. [/TO BE CHANGED]

Chapter 2

Preliminaries

2.1. Graph Theory Terminology

This section provides definitions of graph theory objects required for completeness of further sections. In this section, when is not mention different, V is default notation for set of graph's vertices and E is default notation for set of graph's edges.

Definition 2.1.1. ***(Undirected edge)***
For vertices $u, v \in V$ we say that there is an undirected edge between vertices u and v if $(u, v) \in E$ and $(v, u) \in E$. Undirected edge between u and v is marked as $u - v$.

Definition 2.1.2. ***(Directed edge)***
For vertices $u, v \in V$ we say that there is a directed edge from vertex u to vertex v if $(u, v) \in E$ and $(v, u) \notin E$. Directed edge from u to v is marked as $u \to v$.

Definition 2.1.3. ***(Skeleton)***
Skeleton of graph $G = (V, E)$ is a graph $G' = (V', E')$ where $V = V'$ and the set of edges E' is obtained by replacing directed edges of set E by undirected edges.

Definition 2.1.4. ***(Route)***
A route in graph $G = (V, E)$ is a sequence of vertices $(v_0, \ldots, v_k)$, $k \geq 0$, such that

$$(v_{i-1}, v_i) \in E \quad or \quad (v_i, v_{i-1}) \in E$$

for $i = 1, \ldots, k$. The vertices v_0 and v_k are called terminals. A route is called descending if $(v_{i-1}, v_i) \in E$ for $i = 1, \ldots, k$. Descending route from u to v is marked as $u \mapsto v$.

Definition 2.1.5. ***(Path)***
A route $r = (v_0, v_1, \ldots, v_k)$ in graph $G = (V, E)$ is called a path if all vertices in r are distinct.

Definition 2.1.6. ***(Complex)***
A path $\pi = (v_1, v_2, \ldots, v_k)$ in graph $G = (V, E)$ is called complex if

1. $v_1 \to v_2$
2. $\forall_{i \in \{2,3,\ldots k-2\}}\ v_i - v_{i+1}$
3. $v_{k-1} \leftarrow v_k$
4. *There is not additional edges in graph G for vertices in path π.*

Vertices v_1 and v_k are called parents of the complex, set of vertices $\{v_2, v_3, \ldots, v_{k-1}\}$ is called region of the complex and number $k-2$ is the degree of the complex.

Definition 2.1.7. ***(Moral Graph)***
Let $G = (V, E)$ be a graph. A moral graph $G^m = (V, E^m)$ of graph G is a graph obtained by firstly join parents of complexes in graph G and then replace all edges by undirected edges.

Definition 2.1.8. ***(Cycle)***
A route $r = (v_0, v_1, \ldots, v_k)$ in graph $G = (V, E)$ is called a pseudocycle if $v_0 = v_k$ and a cycles if further route is a path and $k \geq 3$.

A graph with only directed edges is called an *undirected graph.* A graph without directed cycles and with only directed edges is called a *directed acyclic graph* (DAG).

Definition 2.1.9. ***(Chain graph)***
A graph $G = (V, E)$ is called a chain graph if it does not have directed (pseudo) cycles.

Definition 2.1.10. ***(Section)***
A subroute $\sigma = (v_i, \ldots, v_j)$ of route $\rho = (v_0, \ldots, v_k)$ in graph G is called section if σ is the maximal undirected subroute of route ρ. That means $v_i - \cdots - v_j$ for $0 \leq i \leq j \leq k$. Vertices v_i and v_j are called terminals of section σ. Further vertex v_i is called a head-terminal if $i > 0$ and $v_{i-1} \to v_i$ in graph G. Analogically vertex v_j is called a head-terminal if $j < k$ and $v_j \leftarrow v_{j+1}$ in graph G.

A section with two head-terminals is called *head-to-head* section. Otherwise the section is called *non head-to-head.* For a given set of vertices $S \subset V$ in graph G and section $\sigma = (v_i, \ldots, v_j)$ we say that section is hit by S if $\{v_i, \ldots, v_j\} \cap S \neq \emptyset$. Otherwise we say that section σ is outside set S.

Definition 2.1.11. ***(Intervention)***
A route ρ in graph $G = (V, E)$ is blocked by a subset $S \subset V$ of vertices if and only if there exists a section σ of route ρ such that one of the following conditions is satisfied.

1. *Section σ is head-to-head with respect to ρ and σ is outside of S.*
2. *Section σ is non head-to-head with respect to ρ and σ is hit by S.*

Example 2.1.1. *Based on the following two graphs we present examples of above defined definitions. Let graph presented in figure 2.1 be denoted as G. In graph G as example of descending route is (A, B, C, D) and example of non-descending route is (D, E, F, G). Graph G contains two complexes. Complex (A, B, C, D, E) is of degree equal to 3 and the other one (F, G, H, I) is of degree equal to 2. Graph G contains one cycle (I, J, K, I). Route (F, G, H, I) in graph contains section (G, H) which is head-to-head section.*

Graph presented in figure 2.2 is moral graph of graph G. Additional edges $[A, E]$ and $[F, I]$ came from connecting parents of two complexes in original graph G.

2.2. Graphical Model Terminology

Our main goal is to find an conditional independence structure of given joint probability distribution, hence we start from recalling definition of conditional independence.

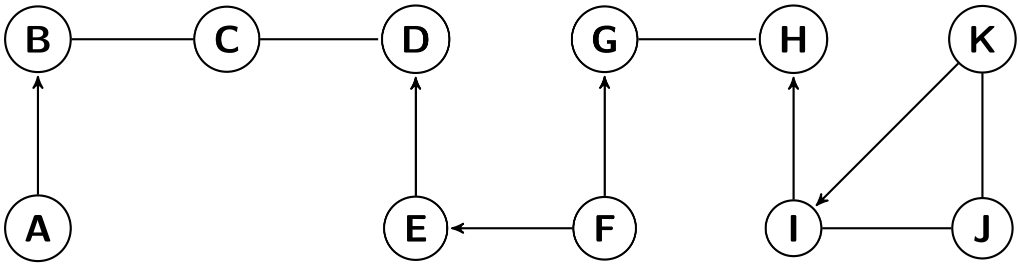

Figure 2.1: Example graph

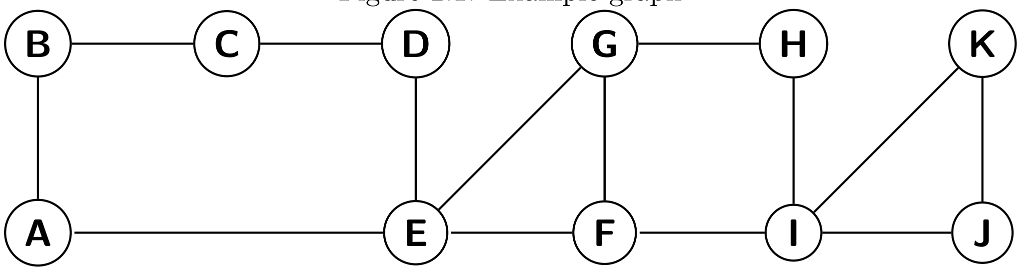

Figure 2.2: Moral graph of graph in figure 2.1

Definition 2.2.1. ***Conditional Independence***
Let $(X_1, X_2, \ldots, X_n)$ be a random vector over probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that random vectors $X_A = \{X_a \mid a \in A\}$ and $X_B = \{X_b \mid b \in B\}$ are conditional independent given $X_S = \{X_s \mid s \in S\}$ when for all $A_1, A_2, A_3 \in \mathcal{F}$

$$\mathbb{P}(X_A \in A_1, X_B \in A_2 \mid X_S \in A_3) = \mathbb{P}(X_A \in A_1 \mid X_S \in A_3)\mathbb{P}(X_B \in A_2 \mid X_S \in A_3) \tag{2.1}$$

where $A, B, S \subset 1, 2, \ldots, n$. Conditional independence of X_A and X_B given X_S is denoted as $X_A \perp\!\!\!\perp X_B \mid X_S$.

The following definition of c-separation is an analogical version of d-separation, used in Bayesian Networks, for chain graphs. This definition was introduced by Studeny and Bouckaert in [4]. The notation c-separation is short of "chain separation" and it is written in this form to present analogy to definition of d-separation.

Definition 2.2.2. ***(c-separation)***
Let $G = (V, E)$ be a chain graph. Let A, B, S be three disjoint subsets of the vertex set V, such that A and B are nonempty. We say that A and B are c-separated by S on G if every route within one of its terminals in A and the other in B is blocked by S. We call S a c-separator for A and B and mark as $\langle A, B \mid S\rangle_{\mathcal{G}}^{sep}$.

Definition 2.2.3. ***(faithfulness)***
Let $G = (V, E)$ be a chain graph with random variables X_v associated with vertex $v \in V$. Let note domain of random variable X_v as $\mathcal{X}_v$. A probability measure $\mathbb{P}$ defined on $\prod_{v \in V} \mathcal{X}_v$ is faithful with respect to G if for any triple (A, B, S) of disjoint subsets of V where A and B are non-empty we have

$$\langle A, B \mid S\rangle_{\mathcal{G}}^{sep} \iff X_A \perp\!\!\!\perp X_B \mid X_S \tag{2.2}$$

In the same setup a probability measure $\mathbb{P}$ is called Markovian with respect to G if

$$\langle A, B \mid S\rangle_{\mathcal{G}}^{sep} \implies X_A \perp\!\!\!\perp X_B \mid X_S \tag{2.3}$$

The following theorem from Frydenberg's paper [1] provides convenient tool for testing if two given chain graphs are the same in respect to Markov equivalent class.

Proposition 2.2.1. ***(Markov equivalence of chain graphs)*** *[Theorem 5.6 from [1]]* *Two chain graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ have the same Markov properties if and only if they same the same skeleton and the same complexes.*

Chapter 3

Structural Learning of Chain Graphs

3.1. Algorithm

Chapter 4

Undirected Graphical Model Selection

4.1. Algorithm

Bibliography

[1] M. Frydenberg, *The Chain Graph Markov Property*, Scandinavian Journal of Statistics 17 (1990) 333-353

[2] R. D. Nowak and D. Vats, *A Junction Tree Framework for Undirected Graphical Model Selection*, Journal of Machine Learning Research 15 (2014) 147-191

[3] Z. Ma, X. Xie and Z. Geng, *Structural Learning Of Chain Graphs via Decomposition*, Journal of Machine Learning Research 9 (2008) 2847-2880

[4] M. Studeny and R.R. Bouckaert, *On chain graph models for description of conditional independence structures*, Annals Of Statistics 26 (1998) 1434-1495