Some open problems from the research kitchen 'Learning Conditional Indepecence Models'

1. Inclusion problem

Let K, L are acyclic directed graphs with a set of nodes V and $\mathcal{I}(K)$ denotes the (conditional independence) model over V induced by K through d-separation criterion. Characterize the inclusion $\mathcal{I}(K) \subseteq \mathcal{I}(L)$ with help of graphical conditions on relation K and L.

Meek's conjecture (about transformational characterization of inclusion)
$\mathcal{I}(K) \subseteq \mathcal{I}(L)$ iff there exists a sequence of acyclic directed graphs $K = H_1, \ldots, H_n = L$, $n \geq 1$ over V such that H_{i+1} is obtained from H_i for $i = 1, \ldots n-1$ either by removal of an arrow or by reversal of an arrow $a \to b$ such that $pa(b) = pa(a) \cup a$.
Here, $pa(a) \equiv \{c \in N\,;\, c \to a\,\}$.

2. Reconstruction of decomposable models

Definition 1 Let $P((X_i)_{i \in V})$ be a $|V|$-dimensional ($|V| < \infty$) probability distribution of finite-valued variables and a system $K_1, \ldots, K_n$ be a cover of V. $K_1, \ldots, K_n$ is called a *compact directed informational support* of P if there exists a Bayesian network with an acyclic directed graph $G = (V, E)$ representing distribution P such that

1. $\forall i \in V \exists k \in \{1, \ldots, n\}$, for which $\{i\} \cup pa(i) \subseteq K_k$;
2. if there is a directed path in G from i to j, for which $i, j \in K_k$ for some $k \in \{1, \ldots, n\}$, then all the nodes from this path must belong also to K_k.

Remark The condition 1. guarantees possibility to reconstruct the Bayesian network just from the marginals $P((X_i)_{i \in K_1}), \ldots, P((X_i)_{i \in K_n})$. The *compactness* condition 2. expresses the fact that the information from these marginals is in a sense local.

Proposition A cover of V is a compact directed informational support for P *iff* it can be ordered $K_1, \ldots, K_n$ in such a way that for all $k = 2, \ldots, n$

$$(X_i)_{i \in K_k \setminus (K_1 \cup \ldots \cup K_{k-1})} \perp\!\!\!\perp (X_i)_{i \in (K_1 \cup \ldots \cup K_{k-1}) \setminus K_k} | (X_i)_{i \in K_k \cap (K_1 \cup \ldots \cup K_{k-1})} \, [P] \,.$$

Definition 2 Let $P((X_i)_{i \in V}$ be a $|V|$-dimensional ($|V| < \infty$) probability distribution of finite-valued variables and a system $K_1, \ldots, K_n$ be a cover of V. $K_1, \ldots, K_n$ is called a *compact decomposable informational support* of P if there exists a triangulated graph $G = (V, E)$ (cliques of G are denoted $C_1, \ldots, C_m$ in the sequel), with respect to which distribution P factorizes, such that

1. $\forall j \in \{1, \ldots, m\} \exists k \in \{1, \ldots, n\}$, for which $C_j \subseteq K_k$;
2. if there is K_k containing two different cliques C_j, C_ℓ of G then it has to contain also all the cliques which are on the path from C_j to C_ℓ in any junction tree coresponding to G.

The problem What are the necessary and sufficient condition for $K_1, \ldots, K_n$ to be a compact decomposable informational support of P ?

3. Hidden variable

Problem What is the least natural number m such that for every set Y of cardinality m and every probability distribution P on $X_1 \times \ldots \times X_n$ where $|X_i| = 2$ there exists a probability distribution Q on $Y \times X_1 \times \ldots \times X_n$ such that

- $Q(y, x_1, \ldots, x_n) = Q(y) \cdot \prod_{i=1}^{n} Q(x_i|y)$ for $x_i \in X_i$, $y \in Y$,
- marginal of Q on $X_1 \times \ldots \times X_n$ is P.

4. Likelihood projection

Definition Let X be a non-empty finite set and $\mathcal{P}^+$ the class of densities p of probability distributions on X (with respect to the counting measure) which are strictly positive. Let $\mathcal{F} \subseteq \mathcal{P}^+$ be an *exponential family* of distributions, i.e. a class satisfying

$$\forall p, q \in \mathcal{F}\ \forall \alpha \in \Re\ \ \exists \text{ a constant } c(\alpha, p, q) \text{ such that } c(\alpha, p, q) \cdot p \cdot q \in \mathcal{F}\,.$$

Remark An example of exponential family is the class of distributions satisfying prescribed saturated conditional independence statements.

Proposition Then for every $p \in \mathcal{P}^+$ unique $q \in \mathcal{F}$ minimizing the Kulback-Leibler divergence $I(p\|q)$ exists. It is called *likelihood projection* of p and denoted by $\pi_{\mathcal{F}}(p)$.

The problem Suppose that $\emptyset \neq \mathcal{F} = \mathcal{F}_1 \cap \mathcal{F}_2$ is the intersection of two exponential families and $p \in \mathcal{P}^+$. Construct a sequence of densities as follows:

$$p_0 = p, \quad p_i = \pi_{\mathcal{F}_1}(p_{i-1}) \text{ for } i \text{ odd}, \quad p_i = \pi_{\mathcal{F}_2}(p_{i-1}) \text{ for } i \text{ even}\,.$$

Under which conditions the constructed sequence converges to $\pi_{\mathcal{F}}(p)$?

5. Information projection

Definition Let X be a non-empty finite set and $\mathcal{P}$ the class of densities p of probability distributions on X (with respect to the counting measure). Let $\mathcal{F} \subseteq \mathcal{P}$ be a *linear family* of distributions, i.e. a class satisfying

$$\forall p, q \in \mathcal{A}\ \forall \alpha, \beta \in \Re\ \text{ whenewer } \alpha \cdot p + \beta \cdot q \in \mathcal{P} \text{ then } \alpha \cdot p + \beta \cdot q \in \mathcal{A}\,.$$

By *information projection* of $p \in \mathcal{P}$ is understood unique $q \in \mathcal{A}$ minimizing the Kulback-Leibler divergence $I(q\|p)$.

Remark An example of linear family is the class of densities with prescribed marginals.

The problem Suppose that $\emptyset \neq \mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ is the intersection of two linear families and $p \in \mathcal{P}$. Construct a sequence of densities as follows:

$$p_0 = p, \quad p_i = \pi_{\mathcal{A}_1}(p_{i-1}) \text{ for } i \text{ odd}, \quad p_i = \pi_{\mathcal{A}_2}(p_{i-1}) \text{ for } i \text{ even}\,.$$

Does it converge to the information projection $\pi_{\mathcal{A}}(p)$?

6. Axiomatization of intersection of UG-models

Let $\mathcal{G} = \{G_i; i \in I\}$ be a (finite) collection of undirected graphs with a set of nodes V. Let $\mathcal{I}(\mathcal{G})$ denote the intersection $\bigcap_{i \in I} \mathcal{I}(G_i)$ where $\mathcal{I}(G_i)$ denotes the (conditional independence) model induced by G_i through separation criterion.

Task Characterize models $\mathcal{I}(\mathcal{G})$ in an axiomatic way using properties of semigraphoid type.

7. Semigraphoid closure

Definition Let $\mathcal{E}(V)$ denotes the class of 'elementary independencies' $\langle \{i,j\} | K \rangle$, where $K \subset V$, $i,j \in V \setminus K$, $i \neq j$. Supposing $\mathcal{L} \subseteq \mathcal{E}(V)$ let $cl(\mathcal{L})$ denote the closure of $\mathcal{L}$ with respect to semi-graphoid inference rules (reformulated for the case of elementary independencies).
Supposing G is an acyclic directed graph over V and $\mathcal{I}(G)$ the (conditional independence) model induced by G though the d-separation criterion let $com(G)$ denote the *elementary semigraphoid complexity* of $\mathcal{I}(G)$, that is the least cardinality of $\mathcal{L} \subseteq \mathcal{E}(N)$ such that $cl(\mathcal{L}) = \mathcal{I}(G) \cap \mathcal{E}(N)$.

Task Compute for each natural number n the number c_n which is defined as the maximum of $com(G)$ through all acyclic directed graphs G having n nodes.

8. Finite multiinformation

Definition Let P be a probability measure on $\prod_{i \in N} X_i$ where X_i are general measurable spaces. Provided that P is absolutely continuous with respect to the product of its one-dimensional marginals $\prod_{i \in N} P^{\{i\}}$ its *multiinformation* is defined as Kulback-Leibler divergence $I(P \| \prod_{i \in N} P^{\{i\}})$.

Question Let P, Q are probability measures on $\prod_{i \in N} X_i$ with finite multiinformation and $\alpha \in [0,1]$. Has their convex combination $\alpha \cdot P + (1-\alpha) \cdot Q$ finite multiinformation as well?

9. Gaussian conditional independence models

Question Let P, Q are non-degenerate Gaussian measures on $\Re^N$. Is there a non-degenerate Gaussian measure R on $\Re^N$ such that the conditional independence model induced by R is the intersection of conditional indendence models induced by P and Q?