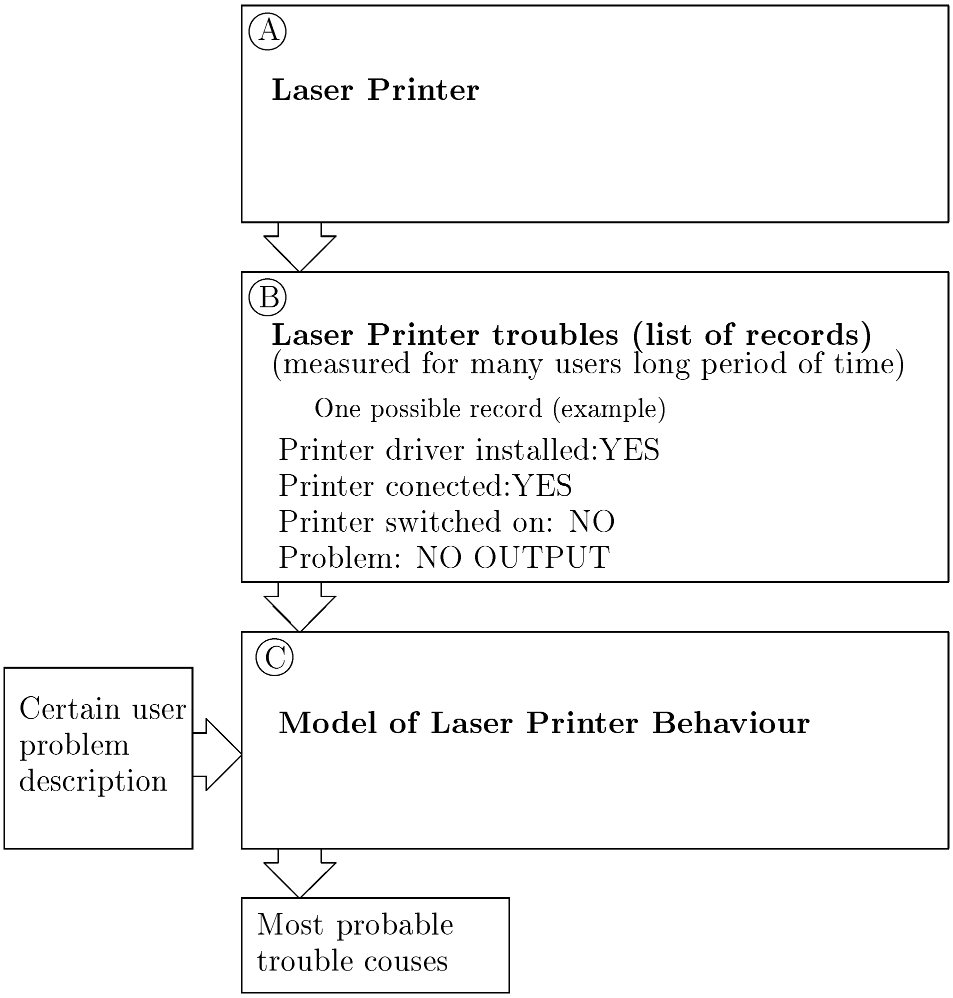
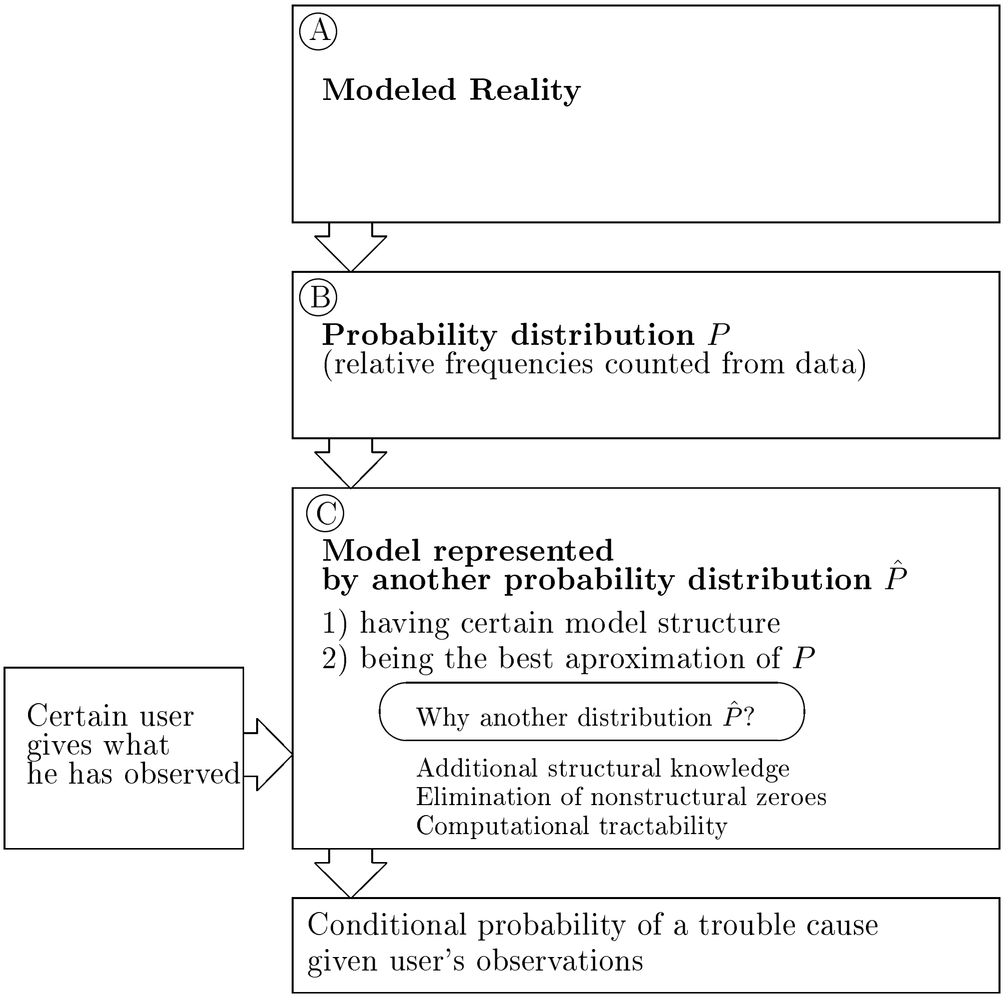
- Motivation. From an example to the theory. Basic definitions.
- Description and history of Iterative Proportional Fitting Procedure.
- Hypergraphs, diferent types of acyclicity in hypergraphs, conformal hypergraphs, chordal and strongly chordal graphs.
- IPFP for decomposable and β-acyclic generating classes.
- IPFP in inconsistent case.


A
Laser Printer
B
Laser Printer troubles (list of records)
(measured for many users long period of time)
One possible record (example)
Printer driver installed:YES
Printer conected:YES
Printer switched on: NO
Problem: NO OUTPUT
C
Model of Laser Printer Behaviour
Certain user problem description
Most probable trouble couses

PROBLEM FORMULATION USING PROBABILITY THEORY



Ⓐ **Reality Description**

Set of n discrete random variables

Ⓑ **Data represented by probability distribution P**

Relative frequencies counted from data
joint distribution, oligodistributions

Ⓒ **Model represented by prob. distribution P***

1)
Hypergraph encoding conditional independence relations between variables

Probability distribution from class of hierrachical log-linear model induced by hypergraph

2)
Minimizing I-divergence between P and P*

Having marginals equivalent to oligodistributions

Users set certain values

Conditional probability

A) Reality Description

Index set $V = \{1, 2, \ldots, n\}$
Set of discrete random variables $\{X_1, X_2, \ldots, X_n\}$,
X_i has values from $\mathbb{X}_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,i_n}\}$,
$\mathbb{X} = \times_{i \in V} \mathbb{X}_i$

B) Probability distribution P (relative frequencies)

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
abreviated $P(X_1, X_2, \ldots, X_n)$, $P(x)$, or just P
Oligodistributions $P_i = P(X_j)_{j \in E_i}$

C) Model represented by probability P^*

1a) **Hypergraph** $H = (V, \mathcal{E})$
vertices(nodes) $V = \{1, 2, \ldots, n\}$ (variables' indices)
set of edges (generating class) $\mathcal{E} = \{E_1, E_2, \ldots E_K\}$,
$\forall i = 1, 2, \ldots, K : E_i \subseteq V$,
$\forall i, j = 1, 2, \ldots K, i \neq j : E_i \nsubseteq E_j$ (reduced hypergraph)

1b) **Hierachical log-linear model** induced by hypergraph
$P^* \in \Pi_{\mathcal{E}} = \{Q(x) : \log Q(x) = \sum_{J \subseteq V, J \subseteq E_i, E_i \in \mathcal{E}} \lambda^J_{x^J}\}$

2a) I-divergence $I(P \parallel P^*) = \sum_{x \in \mathbb{X},\, P(x) > 0} P(x) \log \frac{P(x)}{P^*(x)}$

2b) $P^* \in \Pi(\mathcal{E}) = \{Q(x) : Q^{E_i} = P_i, i = 1, 2, \ldots, K\}$,
where $Q^{E_i} = Q^{E_i}(X_j)_{j \in E_i}$ is marginal of Q

Theorem 1 *Suppose P is strictly positive probability distribution. Then probability distribution P^* that fulfill conditions 1b), 2a), and 2b) is unique and can be found by the iterative proportional fitting procedure -* **IPFP***.*

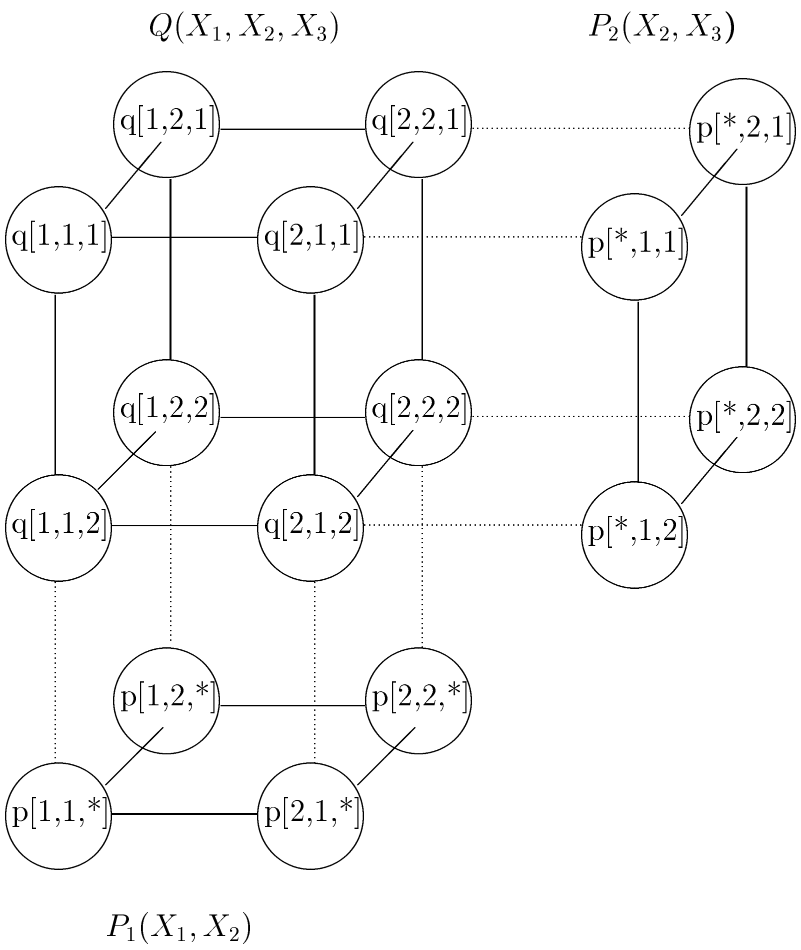
Remark. If P is not strictly positive probability distribution then there exist P^* such that P^* are results of IPFP, but $P^* \notin \Pi_{\mathcal{E}}$. In order to allow distributions having zeroes, instead of $\Pi_{\mathcal{E}}$ its closure $\overline{\Pi}_{\mathcal{E}}$ should be used:

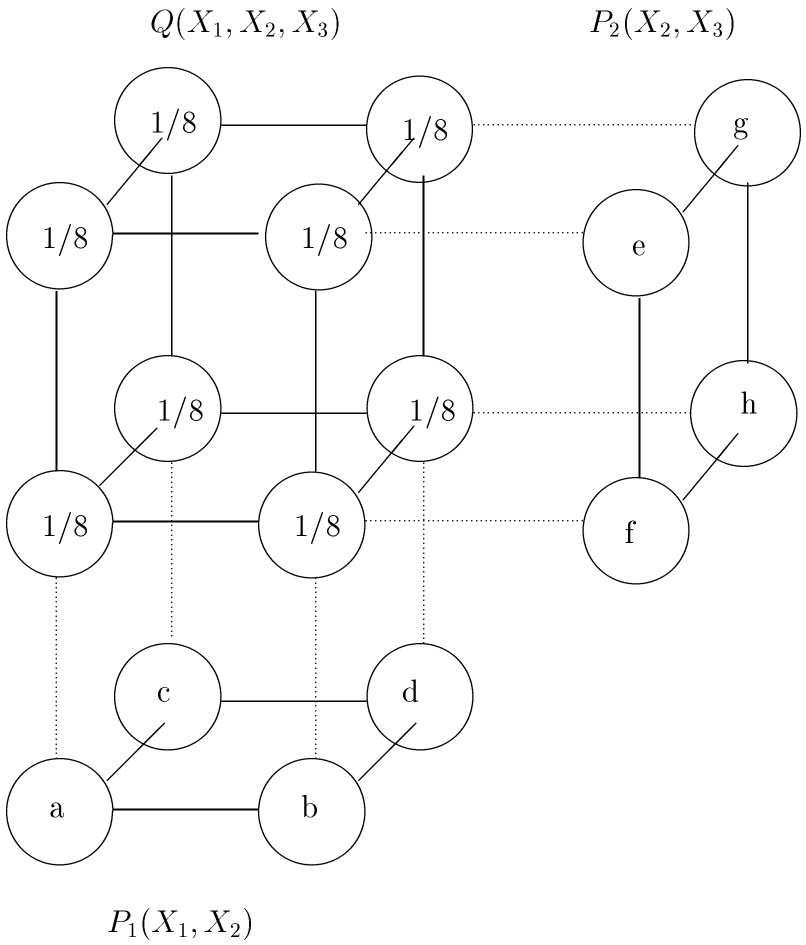
$$\overline{\Pi}_{\mathcal{E}} = \{Q(X_1, \ldots, X_n) : \exists P_n \in \Pi(\mathcal{E}), n = 1, 2, \ldots : \lim_{n \to \infty} P_n = Q\}$$

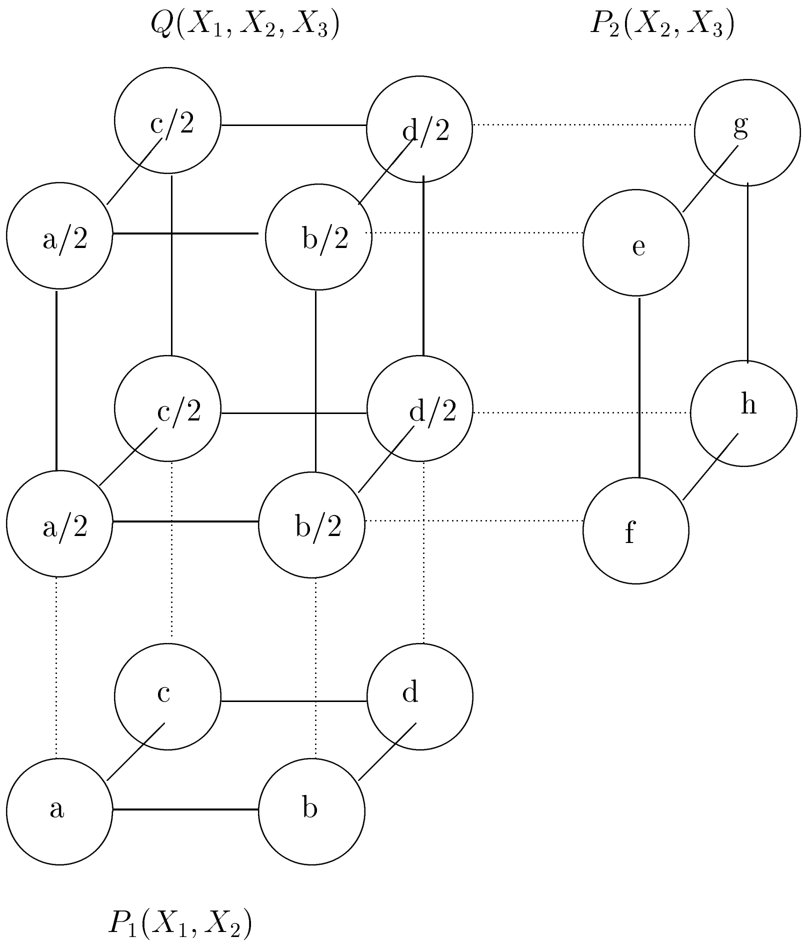
Definition 1 (IPFP) Let Q_0 be the joint uniform probability distribution defined for all the variables $(X_i)_{i \in V}$. The computational process defined by the formula

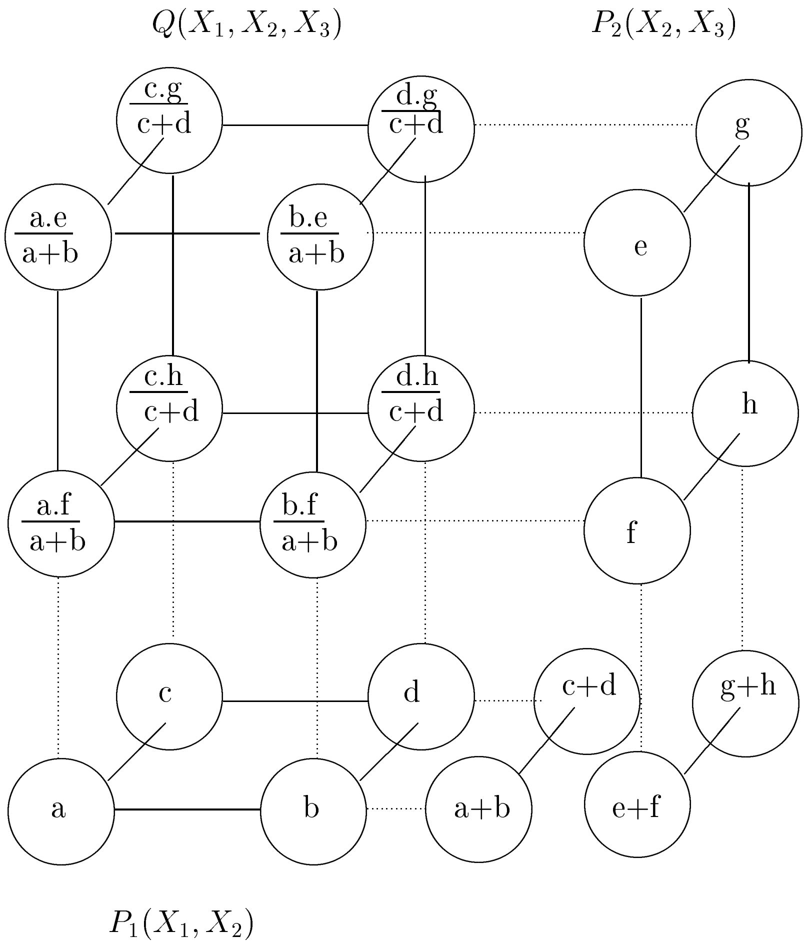
$$Q_q = Q_{q-1} \cdot \frac{P_j}{Q_{q-1}^{E_j}} \qquad \text{for} \quad j = ((q-1) \bmod K) + 1$$

is called *iterative proportional fitting procedure.*


$Q(X_1, X_2, X_3)$
$P_2(X_2, X_3)$
q[1,2,1]
q[2,2,1]
p[*,2,1]
q[1,1,1]
q[2,1,1]
p[*,1,1]
q[1,2,2]
q[2,2,2]
p[*,2,2]
q[1,1,2]
q[2,1,2]
p[*,1,2]
p[1,2,*]
p[2,2,*]
p[1,1,*]
p[2,1,*]
$P_1(X_1, X_2)$


$Q(X_1, X_2, X_3)$
$P_2(X_2, X_3)$
1/8
1/8
1/8
1/8
1/8
1/8
1/8
1/8
g
e
h
f
c
d
a
b
$P_1(X_1, X_2)$


$Q(X_1, X_2, X_3)$
$P_2(X_2, X_3)$
c/2
d/2
g
a/2
b/2
e
c/2
d/2
h
a/2
b/2
f
c
d
a
b
$P_1(X_1, X_2)$



P_1 and P_2 are consistent iff $P_1(X_1, X_2)^{\{2\}} = P_2(X_2, X_3)^{\{2\}}$

1940 W. E. Deming and F. F. Stephan proposed IPFP for adjustment of frequencies in contingency tables. They conjectured the procedure minimized square distance

$$\sum_{x\in\mathbb{X},P(x)>0} \frac{(P(x)-P^*(x))^2}{P(x)}.$$

1959 D. T. Brown gave a proof of IPFP convergence based on the fact, that the likelihood of the iterated solution is a monotonic decreasing function.

1968 C. T. Ireland and S. Kullback's proof of IPFP convergence contained an error.

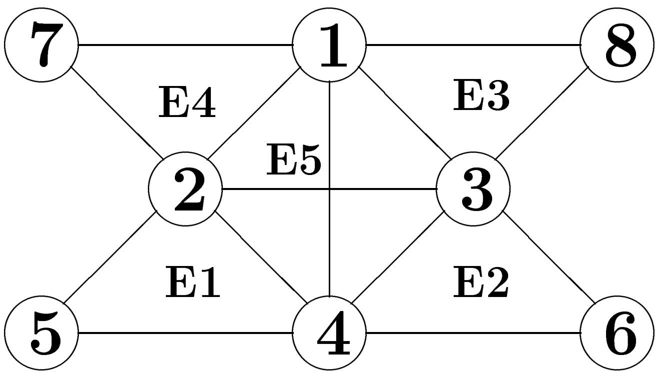
1970 S. Fienberg showed that IPFP does not minimize square distance and proved IPFP convergence for strictly positive distribution.

1974 S. J. Haberman showed that IPFP is a special case of the cyclic ascent method of function maximization. She proved that for strongly decomposable generating classes procedure for every order of oligodistributions converges within K steps. She conjectured that for decomposable classes IPFP converges within $2 * K$ steps.

1975 I. Csiszar proved IPFP convergence using I-divergence geometry without assumption of strict positivity.

Definition 2 (Decomposability of hypergraphs) Hypergraph $H = (V, \mathcal{E})$ is decoposble if either has one edge or is the union of two disjoint decomposable hypergraphs $H_1 = (V_1, \mathcal{E}_1)$ and $H_2 = (V_2, \mathcal{E}_2)$, i.e. $V_1 \subseteq V, V_2 \subseteq V$, $V = V_1 \cup V_2$, $\mathcal{E}_1 \cup \mathcal{E}_2 = \mathcal{E}$, and $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$, such that there exist edges $E_1 \in \mathcal{E}_1$ and $E_2 \in \mathcal{E}_2$ that:

$$((\bigcup_{E' \in \mathcal{E}_1} E') \bigcap (\bigcup_{E'' \in \mathcal{E}_2} E'')) = E_1 \bigcap E_2.$$



Sequence of intersections that can decompose hypergraph is

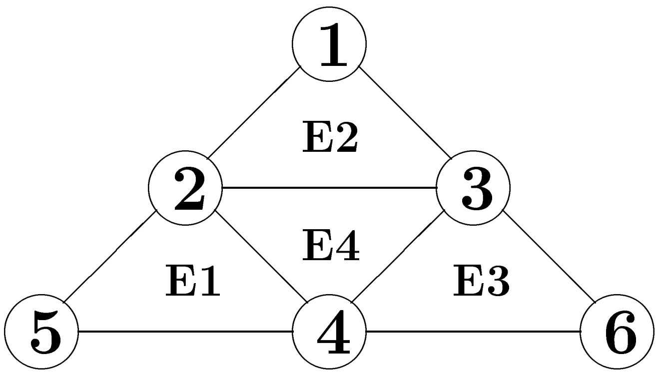
$$E_1 \cap E_5, \quad E_2 \cap E_5, \quad E_3 \cap E_5, \quad E_4 \cap E_5$$

In 1983 R. Fagin defined four basic degrees of acyclicity for relational database schemes:

Berge-acyclicity $\Rightarrow$ γ-acyclicity $\Rightarrow$ β-acyclicity $\Rightarrow$ α-acyclicity

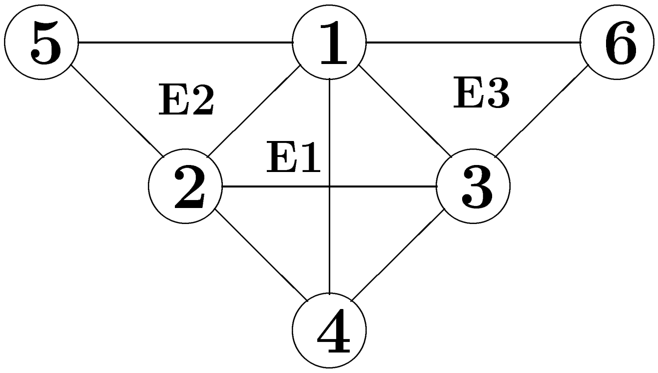
Set of edges is nontrivial if it has more than one edge. Let $\mathcal{M}$ be set of hypergraph nodes. The set of partial edges $\mathcal{F}$ generated by this set is $\{E \cap \mathcal{M} : E \in \mathcal{E}\} - \{\emptyset\}$ after its reduction. Let $F, G \in \mathcal{F}$ be two diferent partial edges and $Q = F \cap G$. Q is an articulation set of $\mathcal{F}$ if $\{E - Q : E \in \mathcal{F}\} - \{\emptyset\}$ is not a connected set.

Definition 3 (α-acyclicity) Hypergraph H is *α-acyclic* iff all nontrivial, conected, node generated sets of partial edges have an articulation set.



It is α-acyclic hypergraph. For example an articulation set for $\{E_1, E_4\}$ is the set $\{2, 4\}$. Subset $\{E_1, E_2, E_3\}$ has no articulation set, but it is not node-generated.

Definition 4 (β-cyclicity) Hypergraph H is *β-cyclic* iff some nontrivial, connected reduced set of edges has no articulation set.



This is an example of β-acyclic hypergraph.

Graph $G = (V, E)$ consits of list of vertices $v \in V$ and list of edges $(v_i, v_j) \in E$.

Definition 5 (Clique of a graph) *Clique* is maximal complete (all nodes are adjacent) subgraph of a graph.

Definition 6 (Cycle in a graph)
Given a graph G, a *cycle* c in G is sequence $(v_1, v_2, \ldots, v_q)$, $q \geq 3$ of distinct graph nodes such that for every $1 \leq i < q$, (v_i, v_{i+1}) are graph edges and (v_1, v_q) is a graph edge too.

Definition 7 (Chord)
Chord in a cycle $c = (v_1, v_2, \ldots, v_q)$ is an edge (v_i, v_j), such that $1 < |i - j| < q - 1$.

Definition 8 (Chordal graph)
A *chordal graph* (triangulated graph) is a graph in which every cycle with at least four distinc nodes has a chord.

Definition 9 (Strong chord)
Strong chord in a even cycle $c = (v_1, v_2, \ldots, v_q)$ is an edge (v_i, v_j), such that $1 < |i-j| < q-1$ and $|i-j|$ is odd ($|i-j| = 3, 5, 7, \ldots$).

Definition 10 (Strongly chordal graph)
A *strongly chordal graph* is a graph that in every even cycle with at least six nodes contains a strong chord.

Theorem 2 *Hypergraph H is α-acyclic iff it is decomposable.*

Definition 11 (Graph of a hypergraph)
A graph $G(H)$ is *graph of hypergraph* $H = (V, \mathcal{E})$ iff it has the same vertex set V and two vertices are joined by an edge if they occur together in some edge $E \in \mathcal{E}$.

Definition 12 (Conformal hypergraph) Hypergraph H is *conformal* iff cliques of its graph $G(H)$ are exactly edges of hypergraph H.

Consider a bijection that associates to each conformal hypergraph H its graph $G(H)$ and to a graph G hypergraph of its maximal cliques.

Theorem 3 *Hypergraph H is α-acyclic iff its graph $G(H)$ is chordal.*

Theorem 4 *Hypergraph H is β-acyclic iff its graph $G(H)$ is strongly chordal.*

Definition 13 (Strong consistency) The system of oligodistribution $\{P_1, \ldots, P_K\}$ is *strongly consistent* iff

$$\Pi(\mathcal{E}) = \{Q(x) : Q^{E_i} = P_i, i = 1, 2, \ldots, K\} \neq \emptyset.$$

Definition 14 (Weak consistency) The system of oligodistribution $\{P_1, \ldots, P_K\}$ is *weakly consistent* iff for any couple of different indices $k, \ell \in \{1, \ldots, K\}$

$$P_k\big((X_i)_{i\in(E_k\cap E_\ell)}\big) = P_\ell\big((X_i)_{i\in(E_k\cap E_\ell)}\big)$$

or, equivalently

$$P_k^{E_k\cap E_\ell} = P_\ell^{E_k\cap E_\ell}.$$

Elements of generating class $\mathcal{E} = \{E_i : i = 1, 2, \ldots, K\}$ are edges of hypergraph H therefore properties of hypergraph H will be given to generating class $\mathcal{E}$ directly.

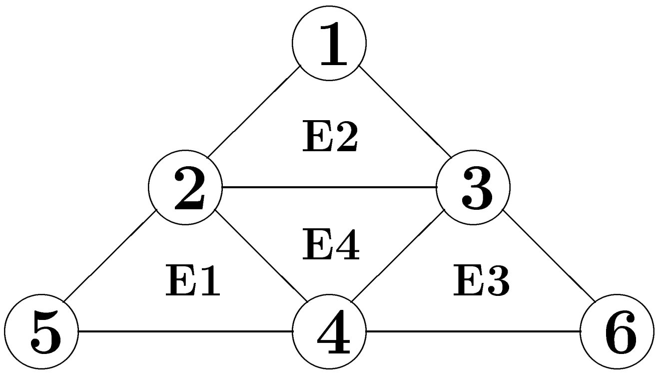
Theorem 5 *Suppose generating class $\mathcal{E}$ is decomposable. If oligodistributions $P_1, P_2, \ldots, P_K$ are weakly consistent, than they are strongly consistent too.*

Theorem 6 *Suppose generating class $\mathcal{E}$ is decomposable and oligodistributions $P_1, P_2, \ldots, P_K$ are weakly consistent. Then there exist an ordering $\mathcal{E} = \{E_i : i = 1, 2, \ldots, K\}$ such that IPFP converges within K steps (one cycle).*

One of possible orderings that guarantee convergence in one cycle are orderings that meet Running Intersection Property (RIP).

Definition 15 (Running Intersection Property) Let E_{j_k} for $k = 1, \ldots, K$ be an ordering of sets $E_1, E_2, \ldots, E_K \in \mathcal{E}$. This ordering meets the *Running Intersection Property (RIP)* if

$$\forall l = 2, \ldots, K \;\; \exists k \;\; (1 \leq k < l) \;\; ((E_l \cap \bigcup_{m=1}^{l-1} E_m) \subset E_k).$$



An example of ordering that meets RIP is $\{E_1, E_4, E_2, E_3\}$:

$$\begin{gathered} l = 2 : E_4 \cap E_1 \subset E_1, E_4 \\ l = 3 : E_2 \cap (E_1 \cup E_4) \subset E_4 \\ l = 4 : E_3 \cap (E_1 \cup E_4 \cup E_2) \subset E_4 \end{gathered}$$

Theorem 7 [Sufficient but not necessary condition] *Suppose generating class $\mathcal{E} = \{E_i : i = 1, 2, \ldots, K\}$ is β-acyclic and oligodistributions $P_1, P_2, \ldots, P_K$ are weakly consistent. Then IPFP converges within K steps (one cycle) no matter how $\mathcal{E}$ is ordered.*

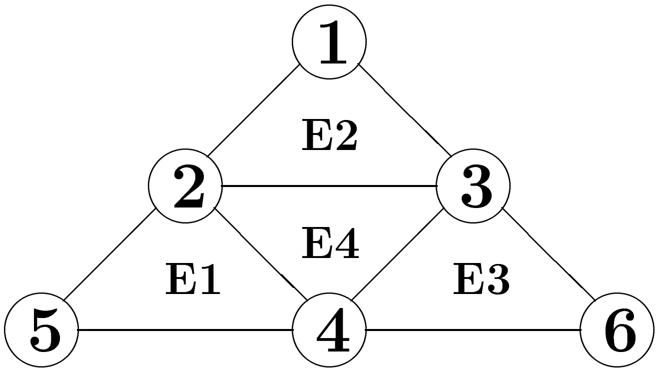


Figure 1: Generating class for which IPFP converge in one cycle no matter how $\mathcal{E}$ is ordered

Theorem 8 [Sufficient but not necessary condition] *Suppose in generating class $\mathcal{E} = \{E_i : i = 1, 2, \ldots, K\}$ is a set E^* such that $E_i \cap E_j \subset E^*, \forall i, j = 1, 2, \ldots K, i \neq j$ and oligodistributions $P_1, P_2, \ldots, P_K$ are weakly consistent. Then IPFP converges within K steps (one cycle) no matter how $\mathcal{E}$ is ordered.*

Conjecture 1 [Haberman, 1974] *Suppose generating class* $\mathcal{E} = \{E_i : i = 1, 2, \ldots, K\}$ *is decomposable and* $P_1, P_2, \ldots, P_K$ *are weakly consistent. Then IPFP converges within* $2 * K$ *steps (two cycles) no matter how* $\mathcal{E}$ *is ordered.*

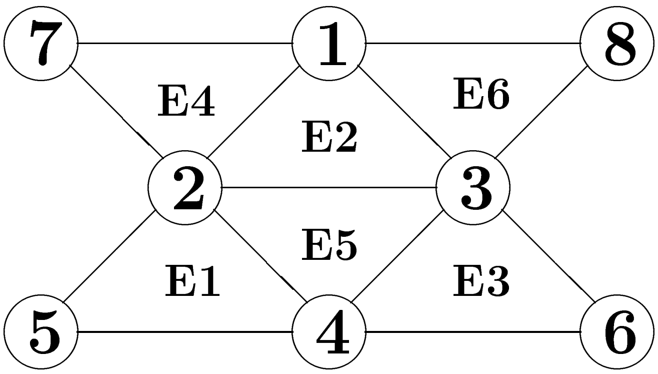


Figure 2: The smallest decomposable generating class for which IPFP need not to converge in two cycles for the ordering $\{E_1, E_2, E_3, E_4, E_5, E_6\}$

In the case of incomplete data, each oligodistribution can be counted from data vectors, that are complete for variables on which is given oligodistribution defined. In this case consistency of system of oligodistributions can not be guaranteed and IPFP need not converge.

F. Matúš proposed an IPFP modification, that is in fact a special case of EM-algorithm.

$$Q_q \;=\; \sum_{j=1}^{K} w_j \cdot Q_{q-1} \cdot \frac{P_j}{Q_{q-1}^{E_j}}, \tag{1}$$

where $w_j = \frac{N_j}{N}$ are weights allocated to every distribution P_j. N is number of all data vectors, N_j is number of vectors used for calculation of P_j.

F. Matúš proved that the sequence of iterations defined by the previous formula converges and minimizes:

$$\psi(Q|P_1, P_2, \ldots, P_K) = \sum_{l=1}^{L} w_l \cdot I(P_l||Q^{E_l}).$$

CONCLUSIONS

- In comparison with Newton-Raphson gradient method IPFP is slower in convergence rate, but it does not require complicated calculations (as matrix inversion in Newton-Raphson).
- For decomposable classes IPFP converge in one cycle, but for some classes proper ordering must be found.
- IPFP modifications are able to find good solution even if data are not complete.