MORALITY and IMMORALITY in DISCRETE GRAPHICAL MODELS

Jacek Wesołowski
(Politechnika Warszawska)

Mathematical Methods in Modern Statistics
LUMINY, July 10-14, 2017

with
Helene Massam (York Univ., Toronto),
John Noble (Uniw. Warszawski, Warszawa)

Plan

1. Markov properties in graphical language
2. Morality: decomposable models
3. Bayesian perspective: hyper-Dirichlet and beyond
4. Global and local parameter independence: characterizations
5. Immoralities: from DAGs to essential graphs through CCC
6. Literature

Discrete model: $\underline{X} = (X_v,\ v \in V) \in \mathcal{I} = \times_{v \in V} \mathcal{I}_v$

$p_{\underline{i}} := \mathbb{P}(\underline{X} = \underline{i}) > 0$, $\underline{i} \in \mathcal{I}$ and $\#\mathcal{I} < \infty$

- *marginal probability*: for $A \subset V$ denote $\underline{X}_A = (X_v,\ v \in A)$,

$$p^A_{\underline{i}_A} := \mathbb{P}(\underline{X}_A = \underline{i}_A) = \sum_{\underline{i}_{A^c} \in \mathcal{I}_{A^c}} p(\underline{i}), \quad \underline{i}_A \in \mathcal{I}_A := \times_{v \in A} \mathcal{I}_v,$$

- *conditional probability*: for $A, B \subset V$ disjoint

$$p^{A|B}_{\underline{k}|\underline{m}} := \frac{p^{A \cup B}_{(\underline{k},\underline{m})}}{p^B_{\underline{m}}}, \quad (\underline{k}, \underline{m}) \in \mathcal{I}_{A \cup B}.$$

- for $A, B, S \subset V$ disjoint $\underline{X}_A \perp\!\!\!\perp \underline{X}_B \mid \underline{X}_S$ if

$$p^{A \cup B|S}_{(\underline{k},\underline{m})|\underline{n}} = p^{A|S}_{\underline{k}|\underline{n}}\, p^{B|S}_{\underline{m}|\underline{n}}, \quad (\underline{k}, \underline{m}, \underline{n}) \in \mathcal{I}_{A \cup B \cup S}.$$

Notation: $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{S}$

Markov property wrt an undirected graph $G = (V, E)$

Distribution $p = (p_{\underline{i}})_{\underline{i} \in \mathcal{I}}$, is Markov wrt G if anyone of 2 conditions holds (**equivalent** since $p_{\underline{i}} > 0 \ \forall \ \underline{i} \in \mathcal{I}$)

- $v \perp\!\!\!\perp w \mid V \setminus \{v, w\}$ if only $v \sim w \notin E$,
- Hammersley-Clifford factorization

$$p_{\underline{i}} = \prod_{A \subset V:\, G_A \text{ is complete}} \psi_A(\underline{i}_A) \qquad \forall \ \underline{i} \in \mathcal{I}$$

for some functions ψ_A.

G_A denotes the subgraph induced in G by $A \subset V$.

Markov property wrt a DAG

A DAG $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$ with skeleton $G = (V, E)$ is defined by parent function $\mathfrak{pa} : V \to 2^V$

$$\mathfrak{pa}(v) = \{w \in V : w \to v\}, \quad v \in V.$$

A distribution $p = (p_{\underline{i}})_{\underline{i} \in \mathcal{I}}$ is Markov wrt $\mathcal{G}$ if any of 2 **equivalent** conditions holds:

- $\forall v \in V$

$$v \perp\!\!\!\perp \mathfrak{nd}(v) \setminus \mathfrak{pa}(v) \mid \mathfrak{pa}(v),$$

  where $\mathfrak{nd}(v) = \{w \in V : \neg(v \to u_1 \to \ldots u_k \to w)\}$;
- recursive factorization:

$$p_{\underline{i}} = \prod_{v \in V} p^{v|\mathfrak{pa}(v)}_{i_v|\underline{i}_{\mathfrak{pa}(v)}}, \quad \underline{i} \in \mathcal{I}.$$

Who is moral?

Let $G = (V, E)$ be a decomposable **(chordal)** undirected graph, i.e. any loop of size ≥ 4 has a chord.

A DAG $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$ is **moral** if

$$\forall\, v \in V \qquad G_{\mathfrak{pa}(v)} \text{ is complete.}$$

Example:

- $$a \to b \to c, \qquad a \leftarrow b \leftarrow c, \qquad a \leftarrow b \to c$$

  are moral DAGs.

- $$a \to b \leftarrow c$$

  is not.

Markov factorizations for chordal $G = (V, E)$

$\mathcal{C}$ - set of cliques (maximal complete subgraphs);
$\mathcal{S}$- set of separators (minimal complete subgraphs removal of which makes the rest of G disconnected).

A distribution $p = (p_{\underline{i}}, \underline{i} \in \mathcal{I})$

- is Markov wrt undirected chordal graph $G = (V, E)$, i.e.

$$p_{\underline{i}} = \frac{\prod_{C \in \mathcal{C}} p^{C}_{\underline{i}_C}}{\prod_{S \in \mathcal{S}} p^{S}_{\underline{i}_S}}, \qquad \underline{i} \in \mathcal{I}.$$

  iff

- it is Markov wrt a moral DAG $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$, i.e.

$$p_{\underline{i}} = \prod_{v \in V} p^{v|\mathfrak{pa}(v)}_{i_v|\underline{i}_{\mathfrak{pa}(v)}}, \quad \underline{i} \in \mathcal{I};$$

  iff

- it is Markov wrt any moral DAG $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$.

Multinomial law for cell counts

Let $\underline{X}_1, \ldots, \underline{X}_n$ be iid with distribution $p = (p_{\underline{i}})_{\underline{i} \in \mathcal{I}}$. Let

$$M_{\underline{i}} = \sum_{j=1}^{n} \mathbb{I}(\underline{X}_j = \underline{i}), \quad \underline{i} \in \mathcal{I}.$$

Then $M = (M_{\underline{i}},\ \underline{i} \in \mathcal{I})$ has a multinomial distribution, $\mathrm{mn}_{\mathcal{I}}(n,\ p)$, i.e.

$$\mathbb{P}(M = m) = \binom{n}{m} \prod_{\underline{i} \in \mathcal{I}} p_{\underline{i}}^{m_{\underline{i}}},$$

$$m = (m_{\underline{i}}, \underline{i} \in \mathcal{I}) \in \mathbb{N}^{\#\mathcal{I}}, \qquad \sum_{\underline{i} \in \mathcal{I}} m_{\underline{i}} = n.$$

In the Bayesian approach the parameter, $p = (p_{\underline{i}},\ \underline{i} \in \mathcal{I})$, becomes a random vector, $\mathbf{p} = (\mathbf{p}_{\underline{i}},\ \underline{i} \in \mathcal{I})$.

Markov property wrt complete G means nothing!

The only restrictions on $\mathbf{p}$ are:

$$\mathbf{p}_{\underline{i}} > 0, \quad \underline{i} \in \mathcal{I}, \quad \text{and} \quad \sum_{\underline{i} \in \mathcal{I}} \mathbf{p}_{\underline{i}} = 1.$$

A standard prior law is **Dirichlet** $\mathrm{D}_{\mathcal{I}}(\alpha_{\underline{i}}, \underline{i} \in \mathcal{I})$ defined e.g. by its moments

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} \mathbf{p}_{\underline{i}}^{r_{\underline{i}}} = \frac{\prod_{\underline{i} \in \mathcal{I}} (\alpha_{\underline{i}})^{r_{\underline{i}}}}{(|\alpha|)^{|r|}},$$

where $(r_{\underline{i}})_{\underline{i} \in \mathcal{I}} \in \mathbb{N}^{\#\mathcal{I}}$, $|\mathbf{c}| = \sum_{\underline{\mathbf{i}} \in \mathcal{I}} \mathbf{c}_{\underline{\mathbf{i}}}$ and $(\mathbf{c})^{\mathbf{k}} = \frac{\Gamma(\mathbf{c}+\mathbf{k})}{\Gamma(\mathbf{c})}$.

Conjugacy: If $M|\mathbf{p}$ is multinomial $\mathrm{mn}_{\mathcal{I}}(n, \mathbf{p})$ and $\mathbf{p} \sim \mathrm{D}_{\mathcal{I}}(\alpha_{\underline{i}}, \underline{i} \in \mathcal{I})$, then

$$\mathbf{p}|M \sim \mathrm{D}_{\mathcal{I}}(\alpha_{\underline{i}} + M_{\underline{i}}, \underline{i} \in \mathcal{I}).$$

The easiest non-trivial case

Let $G = \mathbf{1} \sim \mathbf{2} \sim \mathbf{3}$, $\mathcal{I} = \{0,1\}^3$ and let

$$p = (p_{ijk} = \mathbb{P}(X_1 = i, X_2 = j, X_3 = k), \quad i,j,k \in \{0,1\}).$$

p is Markov wrt $\mathbf{1} \sim \mathbf{2} \sim \mathbf{3}$:

$$p_{ijk} = \frac{\mathbb{P}(X_1=i, X_2=j)\,\mathbb{P}(X_2=j, X_3=k)}{\mathbb{P}(X_2=j)}$$

iff it is Markov wrt $\mathbf{1} \to \mathbf{2} \to \mathbf{3}$:

$$p_{ijk} = \mathbb{P}(X_1 = i)\,\mathbb{P}(X_2 = j | X_1 = i)\,\mathbb{P}(X_3 = k | X_2 = j)$$

iff it is Markov wrt $\mathbf{1} \leftarrow \mathbf{2} \to \mathbf{3}$:

$$p_{ijk} = \mathbb{P}(X_1 = i | X_2 = j)\,\mathbb{P}(X_2 = j)\,\mathbb{P}(X_3 = k | X_2 = j).$$

5-dimensional manifold in 8-dimensional space

By calculation it follows that $p = (p_{ijk})$ is Markov iff

$$p_{101} = \frac{p_{100}p_{001}}{p_{000}} \quad \text{and} \quad p_{111} = \frac{p_{110}p_{011}}{p_{010}}.$$

One needs a probability measure on 5-dimensional manifold in 8-dimensional space defined by the conditions:

$$x_i > 0, \quad i = 0, \ldots, 7,$$

$$\sum_{i=0}^{7} x_i = 1,$$

$$x_5 = \frac{x_4 x_1}{x_0}, \qquad x_7 = \frac{x_6 x_3}{x_2}.$$

Dawid & Lauritzen (AS'93) to the rescue!

Let $G = (V, E)$ be a chordal graph with cliques $\mathcal{C}$ and separators $\mathcal{S}$. Then **p** has a hyper-Dirichlet distribution $\mathrm{HD}_G(\nu^C_{\underline{i}_C}, \underline{i}_C \in \mathcal{I}_C, C \in \mathcal{C})$ if for any $r = (r_{\underline{i}}, \underline{i} \in \mathcal{I}) \in \mathbb{N}^{\#\mathcal{I}}$,

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} \mathbf{p}_{\underline{i}}^{r_{\underline{i}}} = \frac{\prod_{C \in \mathcal{C}} \prod_{\underline{i}_C \in \mathcal{I}_C} \left(\nu^C_{\underline{i}_C}\right)^{r^C_{\underline{i}_C}}}{\prod_{S \in \mathcal{S}} \left(\mu^S_{\underline{i}_S}\right)^{r^S_{\underline{i}_S}}},$$

where for any $S \in \mathcal{S}$ and $\underline{m} \in \mathcal{I}_S$

$$\mu^S_{\underline{m}} = \sum_{\underline{n} \in \mathcal{I}_{C \setminus S}} \nu^C_{(\underline{m}, \underline{n})} \quad \text{if only } S \subset C \in \mathcal{C}.$$

Here we assume that $\emptyset \in \mathcal{S}$, $\mathcal{I}_\emptyset = \{0\}$ and thus

$$\mu^\emptyset_0 = \sum_{\underline{m} \in \mathcal{I}_C} \nu^C_{\underline{m}} \quad \forall\, C \in \mathcal{C} \quad \text{and} \quad r^\emptyset_0 = \sum_{\underline{i} \in \mathcal{I}} r_{\underline{i}} =: |r|.$$

HD_G is conjugate in multinomial model

- If $\mathbb{P}(\underline{X} = \underline{i}|\mathbf{p}) = \mathbf{p}_{\underline{i}}$, $\underline{i} \in \mathcal{I}$, and $\mathbf{p} \sim \mathrm{HD}_G$ then $\mathbf{p}$, the conditional distribution $\underline{X}|\mathbf{p}$, is Markov wrt to G.
- If G is a complete then $\mathcal{C} = \{G\}$ and $\mathcal{S} = \{\emptyset\}$, Thus moments formula imply: $\mathrm{HD}_G = \mathrm{D}_{\mathcal{I}}$.
- Let $M|\mathbf{p} \sim \mathrm{mn}_{\mathcal{I}}(n, \mathbf{p})$ and $\mathbf{p} \sim \mathrm{HD}_G(\nu^C_{\underline{k}}, \underline{k} \in \mathcal{I}_C, C \in \mathcal{C})$. By the generalized Bayes rule

$$\mathbb{E}\left(\prod_{\underline{i}\in\mathcal{I}} \mathbf{p}_{\underline{i}}^{r_{\underline{i}}} \middle| M = m\right) = \frac{\mathbb{E}\prod_{\underline{i}\in\mathcal{I}} \mathbf{p}_{\underline{i}}^{m_{\underline{i}}+r_{\underline{i}}}}{\mathbb{E}\prod_{\underline{i}\in\mathcal{I}} \mathbf{p}_{\underline{i}}^{m_{\underline{i}}}} = \frac{\prod_{C\in\mathcal{C}}\prod_{\underline{i}_C\in\mathcal{I}_C}\left(\nu^C_{\underline{i}_C}+m^C_{\underline{i}_C}\right)^{r^C_{\underline{i}_C}}}{\prod_{S\in\mathcal{S}}\left(\mu^S_{\underline{i}_S}+m^S_{\underline{i}_S}\right)^{r^S_{\underline{i}_S}}}$$

since $(\mathbf{a})^{\mathbf{i}+\mathbf{j}} = (\mathbf{a})^{\mathbf{i}}(\mathbf{a}+\mathbf{i})^{\mathbf{j}}$. Thus

$$\mathbf{p}|M \sim \mathrm{HD}_G(\nu^C_{\underline{i}_C} + M^C_{\underline{i}_C}, \underline{i}_C \in \mathcal{I}_C, C \in \mathcal{C}),$$

where $M^C_{\underline{i}_C}$ are marginal counts.

Directional properties of **p** $\sim \mathrm{HD}_G$ for any moral DAG

- **Parameters Independence (PI)**: Random conditional probabilities

$$\mathbf{p}^{v|\mathfrak{pa}(v)}_{\underline{i}_{\mathfrak{pa}(v)}} := \left(\mathbf{p}^{v|\mathfrak{pa}(v)}_{i_v|\underline{i}_{\mathfrak{pa}(v)}},\ i_v \in \mathcal{I}_v\right), \quad \underline{i}_{\mathfrak{pa}(v)} \in \mathcal{I}_{\mathfrak{pa}(v)},\ v \in V$$

  are independent. (global and local independence of parameters - **two in one!**)

- **Dirichlet conditionals (DC)**: All random vectors

$$\mathbf{p}^{v|\mathfrak{pa}(v)}_{\underline{i}_{\mathfrak{pa}(v)}}, \quad \underline{i}_{\mathfrak{pa}(v)} \in \mathcal{I}_{\mathfrak{pa}(v)},\ v \in V$$

  have classical Dirichlet laws.

If **p** is Markov wrt to G and satisfies PI and DC for a given DAG $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$ we say that its law is **$\mathcal{G}$-Dirichlet**.

Can one determine HD_G through PI and DC?

Let $G = (V, E)$ be chordal. Let $\mathbf{p}$, Markov wrt G, has a $\mathcal{G}$-Dirichlet law for **any** moral $\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$. Then **PI** implies

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} \mathbf{p}_{\underline{i}}^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{pa}(v)}} \mathbb{E} \prod_{m \in \mathcal{I}_v} \left[\mathbf{p}_{m|\underline{k}}^{v|\mathfrak{pa}(v)} \right]^{r^{\mathfrak{q}(v)}_{(\underline{k},m)}},$$

where $\mathfrak{q}(v) = \mathfrak{pa}(v) \cup \{v\}$.

Since by **DC**

$$\mathbf{p}_{\underline{k}}^{v|\mathfrak{pa}(v)} \sim \mathrm{D}_{\mathcal{I}_v}(\alpha_{m|\underline{k}}^{v|\mathfrak{pa}(v)}, m \in \mathcal{I}_v), \quad \underline{k} \in \mathcal{I}_{\mathfrak{pa}(v)}, \quad v \in V,$$

it follows that

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} \mathbf{p}_{\underline{i}}^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{pa}(v)}} \frac{\prod_{m \in \mathcal{I}_v} \left(\alpha_{m|\underline{k}}^{v|\mathfrak{pa}(v)} \right)_{r^{\mathfrak{q}(v)}_{(\underline{k},m)}}}{\left(\left| \alpha_{\underline{k}}^{v|\mathfrak{pa}(v)} \right| \right)_{r^{\mathfrak{pa}(v)}_{\underline{k}}}}. \tag{1}$$

$\mathcal{P}$-Dirchlet distribution

Let $\mathcal{P}$ be a family of moral DAGs with a chordal skeleton G.

If the law of $\mathbf{p}$ is $\mathcal{G}$-Dirichlet for any $\mathcal{G} \in \mathcal{P}$ we call it **$\mathcal{P}$-Dirichlet distribution**.

To describe its properties we need to

- define several **new objects**;
- prove several **new results**!

No time **!** See H. Massam & JW, *AS*'16.

Here we concentrate on HD_G!

When $\mathcal{P}$-Dirichlet is a hyper-Dirichlet?

Proposition

Let $\mathcal{P}$ be a family of moral DAGs, with a chordal skeleton $G = (V, E)$ with cliques $\mathcal{C}$ and separators $\mathcal{S}$.

Let

- $$\bigcap_{\mathcal{G}\in\mathcal{P}} \mathfrak{pa}(V) = \mathcal{S};$$
- $\mathcal{P}$ *be a* **pairing family**, *i.e.*

$$\forall\, S \in \mathcal{S},\ C \in \mathcal{C} \quad \text{such that} \quad S \subset C$$

$$(\exists\, \mathcal{G} \in \mathcal{P},\ \exists\ v \in C \setminus S) : \quad S = \mathfrak{pa}(v).$$

Then any $\mathcal{P}$-Dirichlet distribution is a HD_G distribution.

1. Markov properties in graphical language
2. Morality: decomposable models
3. Bayesian perspective: hyper-Dirichlet and beyond
4. Global and local parameter independence: characterizations
5. Immoralities: from DAGs to essential graphs through CCC
6. Literature

Heckerman, Geiger, Chickering, *ML*'95, Geiger, Heckerman, *AS*'97

$G = (V, E)$ - complete graph, $V = \{1, \dots, d\}$.
$\mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$ corresponds to permutation $\sigma \in S_d$:

$$\mathfrak{pa}(\sigma_1) = \emptyset, \quad \mathfrak{pa}(\sigma_k) = \{\sigma_1, \dots, \sigma_{k-1}\}, \quad k = 2, \dots, d.$$

Theorem

G - complete, $\mathbf{p}$ has
a "smooth" density, which is strictly positive on the unit simplex.
If $\mathbf{p}$ satisfies the PI condition for 2 DAGs:

$$\mathcal{G} \equiv \sigma = (1, 2, \dots, d) \qquad \text{and} \qquad \mathcal{G}' \equiv \sigma' = (d, 1, 2 \dots, d-1),$$

then its distribution is a classical Dirichlet.

Why "smooth" densities? Why densities at all? Why such two DAGs $\mathcal{G}$ and $\mathcal{G}'$? Why complete G?

When do PI conditions yield HD_G distribution?

$\mathcal{P}$ - family of DAGs with skeleton $G = (V, E)$

Separating family if $\forall\, v \in V$

$$\exists\, \mathcal{G}, \mathcal{G}' \in \mathcal{P} : \quad \mathfrak{pa}(v) \neq \mathfrak{pa}'(v).$$

Theorem

Let $\mathcal{P}$ be a family of moral DAGs, with a chordal skeleton $G = (V, E)$. Assume that $\mathcal{P}$ is **pairing**, **separating** *and*

$$\bigcap_{\mathcal{G} \in \mathcal{P}} \mathfrak{pa}(V) = \mathcal{S} \qquad \textit{(set of separtors).}$$

If $\forall\, \mathcal{G} \in \mathcal{P}$ PI holds for **p** *then* **p** *has a HD_G law.*

How to extend GHC theorem for complete graphs?

Theorem

Let $\mathbf{p}$ *be a vector of random probabilities. Let G be a complete graph with vertices* $\{1, \ldots, d\}$*. Consider 2 DAGs* $\mathcal{G} \equiv \sigma \in S_d$ *and* $\mathcal{G}' \equiv \sigma' \in S_d$*:*

$$\sigma(\{1, \ldots, j\}) \neq \sigma'(\{1, \ldots, j\}), \qquad j = 1, \ldots, d-1.$$

If $\mathbf{p}$ *satisfies PI conditions wrt to* $\mathcal{G}$ *and* $\mathcal{G}'$ *then its law is classical Dirichlet.*

The case of GHC'95 and GH'97:

$$\sigma(\{1, \ldots, j\}) = \{1, \ldots, j\} \quad \text{and} \quad \sigma'(\{1, \ldots, j\}) = \{d, 1, \ldots, j-1\}.$$

Moral DAGs on T-ree and PI characterization of HD_T

$T = (V, E)$ - tree; L set of leaves of T.

A moral DAG $\mathcal{G} = \mathcal{G}(T, \mathfrak{pa})$ is determined by its source vertex $v_0 \in V$ - we write $\mathcal{G} = \mathcal{G}_{v_0}$.

Theorem

Let $\mathbf{p}$ *be a vector of random probabilities, Markov wrt a tree* T*.*

If $\mathbf{p}$ *satisfies the PI condition wrt to* $\mathcal{G}_{v_0}$ *for all* $v_0 \in L$*, then its law is a hyper-Dirichlet* HD_T*-distiribution.*

Example: If $T1 \sim \ldots \sim d$ is a chain, then PI wrt 2 DAGs:

$$\mathcal{G}_1 = 1 \to \ldots \to d \quad \text{and} \quad \mathcal{G}_d = 1 \leftarrow \ldots \leftarrow d$$

characterizes HD_T law.

Who is immoral?

$G = (V, E)$ - undirected graph

Γ - family of all DAGs with skeleton G

- Triplet of vertices $(a; b, c)$ - **immorality** in $\mathcal{G} \in \Gamma$ if

$$\mathfrak{pa}(a) \supset \{b, c\} \notin E.$$

- $\mathcal{G}, \mathcal{G}' \in \Gamma$ are **graphically** (morally?) equivalent, $\mathcal{G} \sim \mathcal{G}'$, iff they have identical sets of immoralities.

  $\mathrm{ES}(G) = \Gamma / \sim$ - equivalence classes of $\sim$.

Essential graph as Markov equivalence class

Any $[\mathcal{G}] \in \mathrm{ES}$ with skeleton $G = (V, E)$ can be represented by a mixed graph $\mathcal{E}$ with vertices V and edges:

- **directed a $\to$ b**, if $a \to b \in E'$ for any $\mathcal{G}' = (V, E') \in [\mathcal{G}]$.
- **undirected a $\sim$ b**, otherwise, if only $a \sim b \in E$

$\mathcal{E} \equiv [\mathcal{G}]$ is called an **essential graph**.

Proposition (Markov equivalence classes)

Let $\mathcal{E} \equiv [\mathcal{G}]$ be an essential graph.
A probability p is Markov wrt to $\mathcal{G}$ iff it is Markov wrt to any DAG in $\mathcal{E} \equiv [\mathcal{G}]$ (i.e. to any DAG sharing immoralities with $\mathcal{G}$).

Essential graphs are special **chain graphs**.

Chain graph - $G = (V, E, \mathfrak{pa}, \mathcal{T}, \mathcal{D}, \mathfrak{pa}_{\mathcal{D}})$:

Chain graph is a mixed graph with no (partially) directed cycles.

- $\mathfrak{pa}$ - parent function:

$$\mathfrak{pa}(b) = \{a \in V : a \to b \in E\}, \quad b \in V;$$

- $v \equiv w$ if there exists an undirected path in G connecting $v,\ w \in V$;
- $(V/\equiv) =: \mathcal{T} \ni \tau$ - **chain component** (CC);
- $\mathcal{D}$ - **DAG of chain components** with parent function $\mathfrak{pa}_{\mathcal{D}}$:

$$\mathfrak{pa}_{\mathcal{D}}(\tau) = \{\sigma \in \mathcal{T} :\ \sigma \cap \mathfrak{pa}(\tau) \neq \emptyset\} \quad \left(\mathfrak{pa}(\tau) = \bigcup_{\mathbf{v} \in \tau} \mathfrak{pa}(\mathbf{v}) \right).$$

Andersson, Madigan, Perlman *AS*'97

G - mixed graph

- $a \to b \sim c$ - **flag** in G if $G_{abc} = a \to b \sim c$;
- $a \to b$ - **strongly protected** in G if one of
  - $a \to b \leftarrow c$, (immorality)
  - $c \to a \to b$, (compelled arrow)
  - $a \to c \to b$ and $a \to b$, (compelled arrow)
  - $a \sim c_i \to b$, $i = 1, 2$, and $a \to b$ (compelled arrow)

  is an induced subgraph in G.

Theorem (AMP)

A mixed graph G is an essential graph iff G is a chain graph with chordal CCs;

- *G has no flags;*
- *all arrows of G are strongly protected.*

A new characterization of essential graphs

Theorem

A mixed graph G is an essential graph iff it is a chain graph $G = (V, E, \mathfrak{pa}, \mathcal{T}, \mathcal{D}, \mathfrak{pa}_{\mathcal{D}})$ with chordal CCs;
for any $\tau \in \mathcal{T}$

(I) $\mathfrak{pa}(v) = \mathfrak{pa}(\tau), \quad v \in \tau;$

(II) $\forall \sigma \in \mathfrak{pa}_{\mathcal{D}}(\tau)$

$$\mathfrak{pa}(\sigma) = \mathfrak{pa}(\tau) \setminus \sigma \quad \Rightarrow \quad G_{\sigma \cap \mathfrak{pa}(\tau)} \text{ is not complete.}$$

Following Frydenberg'90

$\mathcal{P}$ - family of DAGs with skeleton $G = (V, E)$
$G^{\mathcal{P}} = (V, E_{\mathcal{P}})$ - chain graph with $E_{\mathcal{P}}$ defined in two steps:

F1
- if $a \to b$ in any $\mathcal{G} \in \mathcal{P}$ then
$$a \to b \in \tilde{E}_{\mathcal{P}};$$
- otherwise, if $a \sim b \in E$ then
$$a \sim b \in \tilde{E}_{\mathcal{P}}.$$

$(V, \tilde{E}_{\mathcal{P}})$ **may not be a chain graph**

F2 $E_{\mathcal{P}}$ inherits edges from $\tilde{E}_{\mathcal{P}}$ except:

if $a \to b \in \tilde{E}_{\mathcal{P}}$ is in a (partially) directed cycle, then
$$a \sim b \in E_{\mathcal{P}}.$$

q(uasi)-essential graphs

Proposition

Let $\mathcal{P} \subset [\mathcal{G}] \equiv \mathcal{E}$. Then for $G^{\mathcal{P}} = (V, E, \mathfrak{pa}, \mathcal{T})$

1. *all its CCs are chordal;*
2. $\mathfrak{pa}(v) = \mathfrak{pa}(\tau), \quad v \in \tau \in \mathcal{T}$. **i.e. (I) or "no flags"**

Definition

A chain graph G satisfying (1) and (2) is called **quasi-essential**.

Proposition

If G is a q-essential graph then there exists a family $\mathcal{P}$ of Markov equivalent DAGs (with skeletons as G) such that

$$G = G^{\mathcal{P}}.$$

Then quasi-essential G is Markov equivalent to $\mathcal{E} \equiv [\mathcal{G}]$, $\mathcal{G} \in \mathcal{P}$.

What is condition (II) responsible for ?

$G = (V, E, \mathfrak{pa}, \mathcal{T}, \mathcal{D}, \mathfrak{pa}_{\mathcal{D}})$ - a chain graph

If (II) does not hold for $\sigma, \tau \in \mathcal{T}$:

$$\begin{cases} \sigma \in \mathfrak{pa}_{\mathcal{D}}(\tau), \\ \mathfrak{pa}(\sigma) = \mathfrak{pa}(\tau) \setminus \sigma, \\ G_{\sigma \cap \mathfrak{pa}(\tau)} \quad \text{is complete,} \end{cases} \tag{2}$$

then

$$\psi_{\sigma,\tau}(G) = G' = (V, E')$$

is a mixed graph with E' defined as follows:

- $\forall\, w \in \sigma$ and $\forall\, v \in \tau$

$$\psi_{\sigma,\tau}(w \to v) = w \sim v;$$

- other edges of E' are inherited from E.

$\psi_{\sigma,\tau}(G)$ for q-essential G

Proposition

Let G as above be a q-essential graph and $G' = \psi_{\sigma,\tau}(G)$ for $\sigma, \tau \in \mathcal{T}$ satisfying (2).

Then $G' = (V, E, \mathfrak{pa}', \mathcal{T}', \mathcal{D}', \mathfrak{pa}_{\mathcal{D}'})$ is a q-essential graph with

$$\mathfrak{pa}'(v) = \mathfrak{pa}'(\rho) = \begin{cases} \mathfrak{pa}(\rho), & v \in \rho \neq \tau, \\ \mathfrak{pa}(\tau) \setminus \sigma, & v \in \rho = \tau, \end{cases}$$

$\mathcal{T}' = (\mathcal{T} \setminus \{\sigma, \tau\}) \cup \{\sigma \cup \tau\}$ *and*

$$\mathfrak{pa}_{\mathcal{D}'}(\rho) = \begin{cases} \mathfrak{pa}_{\mathcal{D}}(\sigma), & \text{if } \rho = \sigma \cup \tau, \\ (\mathfrak{pa}_{\mathcal{D}}(\rho) \setminus \{\tau\}) \cup \{\sigma \cup \tau\}, & \text{if } \rho \in \mathcal{T} \text{ and } \tau \in \mathfrak{pa}_{\mathcal{D}}(\rho), \\ \mathfrak{pa}_{\mathcal{D}}(\rho), & \text{otherwise.} \end{cases}$$

Moreover, G' and G are in the same Markov equivalence class.

CCC algorithm $\sim O(n^3)$ - alternative to AMP'97 $\sim O(n^6)$

CCC algorithm: from DAG $\mathcal{G}$ to $\mathcal{E} \equiv [\mathcal{G}]$.

1. $G_0 = \mathcal{G} = \mathcal{G}(G, \mathfrak{pa})$;

   for $k = 0, 1, \dots$
2. G_k - q-essential graph with CCs set $\mathcal{T}_k$;

   choose $\sigma, \tau \in \mathcal{T}_k$ satisfying (2) and set

$$G_{k+1} = \psi_{\sigma,\tau}(G_k);$$

3. if

$$k^* = \min\{k : \ \text{no chain components } \sigma, \ \tau \in \mathcal{T}_k \text{ satisfy (2)}\}.$$

   then $G_{k*} = \mathcal{E} \equiv [\mathcal{G}]$ and the algorithm stops.