Spiegler (2017, REStud), Data Monkeys, Proposition 2

Kyohei Okumura*

July 17, 2018

- The equivalence class of a DAG R is denoted as $[R]$
- The v-structure of a DAG R is denoted as $v(R) := \{(i,j,k) \mid i \to j, j \to k, i \nrightarrow j, j \nrightarrow i\}$.

Prop. 0.1 (Spiegler(2017, REStud) Prop.2, Spiegler(2016, QJE) Prop.7)**.** *Let R be a DAG and let $C \subseteq N$.*

$$[\forall p \in \Delta(X) \forall x; p_R(x_C) = p(x_C)] \iff [\exists Q \in [R]; C \text{ is an ancestral clique in } Q].$$

[2018/07/16: $\Leftarrow$ is correct; $\Rightarrow$ is not sure.]

Proof. First, note that for any DAG R, the following holds:

$$\begin{aligned} p_R(x_C) &= \sum_{x'_{N-C}} p_R(x_C, x'_{N-C}) \\ &= \sum_{x_{N-C}} \prod_{i \in C} p(x_i \mid x_{R(i)\cap C}, x'_{R(i)-C}) \prod_{i \notin C} p(x'_i \mid x_{R(i)\cap C}, x'_{R(i)-C}) \end{aligned} \tag{1}$$

$\Leftarrow$) Fix C such that C is an ancestral clique in some $Q \in [R]$. Note that $R(i) - C = \emptyset$ for all $i \in C$. Then,

$$\prod_{i \in C} p(x_i \mid x_{R(i)\cap C}, x'_{R(i)-C}) = \prod_{i \in C} p(x_i \mid x_{R(i)\cap C}) = p(x_C) \;(\because \text{topological sort})$$

Hence, by (1),

$$p_R(x_C) = p_Q(x_C) = p(x_C) \underbrace{\sum_{x_{N-C}} \prod_{i \notin C} p(x'_i \mid x_{R(i)\cap C}, x'_{R(i)-C})}_{1} = p(x_C).$$

e.g. 0.1. *For example, consider the following DAG:*

$$\begin{tikzcd} 1 \arrow[r] & 2 \arrow[dr] & 4 \\ & & 3 \end{tikzcd}$$

Let $C := \{1, 2\}$. Then,

$$p_R(x_1, x_2) = \sum_{x'_3, x'_4} p_R(x_1, x_2, x'_3, x'_4) = p(x_1, x_2) \sum_{x'_3, x'_4} p(x'_4) p(x'_3 \mid x_2) = p(x_1, x_2)$$

$\Rightarrow$) [We need to make some fix in this direction.]

We show contrapositive: we show the following:

$$[\forall Q \in [R];\ C \text{ is not an ancestral clique in } Q] \implies [\exists p \exists x;\ p_R(x_C) = p(x_C)]$$

Fix any $Q \in [R]$ that C is not an ancestral clique in Q. We divide the proof into two cases:

*E-mail: kyohei.okumura@gmail.com

Case (i): In case C is not a clique in Q. OK.

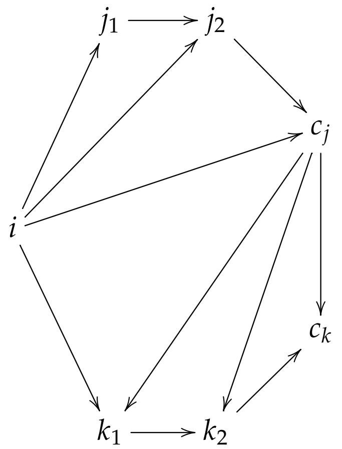
Case (ii): In case C is a clique, but not an ancestral clique. In the original proof in Spiegler(2017), there is a lemma like the following (pp.1838, Case 2, the first paragraph), but the lemma is wrong:

Lem. 0.1. *Let R be a DAG. Assume the following two:*

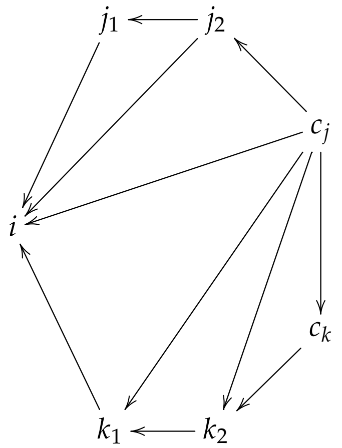
1. *$\forall j \in C$; j has no unmarried parents in R.*
2. *$\forall i \notin C$; if there is a directed path from i to some node $j \in C$ in R, then i has no unmarried parents in R.*

Transform R into another DAG R' by inverting every link along every such path; R and R' has the same v-structure.

e.g. 0.2 (Counter example for Lem.0.1)**.** *Let R be the graph below:*



Let $C := \{c_j, c_k\}$. Note that for all $k \in N \setminus C$ such that k has a path to some $c \in C$, k has no unmarried parents. R' is as follows:



Though $v(R) = \emptyset$, we have $v(R') = \{(j_1, i, k_1), (j_1, i, c_j), (j_2, i, k_1)\}$. Therefore, Lem.0.1 does not hold.

$\square$