Supplement A: Proofs and some detailed examples for "A new prior for discrete DAG models with a restricted set of directions".

Hélène Massam[*], Jacek Wesołowski[†]

April 11, 2016

In this document, references to equation numbers are sometimes to the main file and sometimes to this supplementary file. For readers having access to colour, the difference is clear since references appearing in blue are to this file and those appearing in black are to the main file. However, for the comfort of readers without access to colour, we put a subindex $()_{mf}$ to equation numbers refering to the main file.

2 Preliminaries

2.1 Graph theoretical notions

We first give the proof of Lemma 2.1 .

Proof of Lemma 2.1 (1). • **Existence**: Assume that, on the contrary, $\mathfrak{p}_v \neq \emptyset$ for any $v \in V$. Consider an arbitrary $w_1 \in V$ and for any $k \geq 2$ as w_k choose a vertex from $\mathfrak{p}_{w_{k-1}}$. Then $w_1 \leftarrow w_2 \leftarrow \ldots\ldots \leftarrow w_m \leftarrow \ldots$ Since V is a finite set there is a repetition in the sequence (w_k). Due to acyclicity it is impossible. We got a contradiction.
• **Uniqueness**: Assume that, on contrary, there exists $w_0 \neq v_0$ such that $\mathfrak{p}_{w_0} = \emptyset = \mathfrak{p}_{v_0}$. Obviously, $\{v_0, w_0\} \notin E$. Then consider a minimal path

[*]Department of Mathematics and Statistics, York University, Canada
[†]Politechnika Warszawska, Warszawa, Poland

connecting v_0 and w_0, that is a sequence of vertices $z_0 = v_0, z_1, \ldots, z_m, z_{m+1} = w_0$ such that $\{z_i, z_{i+1}\} \in E,\ i = 0, 1, \ldots, m$ but $\{z_i, z_j\} \notin E, "j \in \{0, \ldots, m+1\} \setminus \{i, i+1\}$. Since v_0 and w_0 are both source vertices there exists $i_0 \in \{1, \ldots, m\}$ such that $\{z_{i_0-1}, z_{i_0+1}\} \subset \mathfrak{p}_{z_{i_0}}$. By morality $\{z_{i_0-1}, z_{i_0+1}\}$ must belong to E but this contradicts the minimality of the path. □

Proof of Lemma 2.1 (2). Denote by K the number of cliques of G. We will define an order of cliques $(C_1, \ldots, C_K)$ which is $\mathfrak{p}$-perfect.

First step. We will now construct C_1. To do so, we will define a sequence of vertices $(v_{1,1}, v_{1,2}, \ldots, v_{1,k_1})$ starting with the source vertex, denoted here by $v_{1,1}$, such that $\mathfrak{q}_{v_{1,i-1}} = \mathfrak{p}_{v_{1,i}}$ and terminates at $k_1 = i$ such that $\mathfrak{q}_{v_{1,i}}$ is a clique. Then we will take $C_1 = \mathfrak{q}_{v_{1,k_1}}$.

(a) We first prove that there exists a vertex $v_{1,2} \neq v_{1,1}$ (not necessarily unique) such that $\mathfrak{p}_{v_{1,2}} = \{v_{1,1}\} = \mathfrak{q}_{v_{1,1}}$. Assume that it is not true. By uniqueness of the source vertex $v_{1,1}$, for any vertex $v \neq v_{1,1}$ we have $\mathfrak{p}_v \setminus \{v_{1,1}\} \neq \emptyset$. Take any vertex $w_1 \neq v_{1,1}$. For any $j \geq 2$ choose as w_j a vertex from $\mathfrak{p}_{w_{j-1}} \setminus \{v_{1,1}\}$. Then $w_1 \leftarrow w_2 \leftarrow \ldots \leftarrow w_m \leftarrow \ldots$. Since the set V is finite there are repetitions in the sequence (w_j), but due to acyclicity this is impossible. So we got a contradiction, therefore there exists a vertex $v_{1,2}$ such that $\mathfrak{p}_{v_{1,2}} = \mathfrak{q}_{v_{1,1}} = \{v_{1,1}\}$.

(b) We will prove now that for any $i \geq 3$ either $\mathfrak{q}_{v_{1,i-1}}$ is a clique or there exists a vertex $v_{1,i} \notin \mathfrak{q}_{v_{1,i-1}}$ such that $\mathfrak{p}_{v_{1,i}} = \mathfrak{q}_{v_{1,i-1}}$. We have proved, in (a) above, that for $j = 2$, there exists $v_{1,j}$ such that $\mathfrak{p}_{v_{1,j}} = \mathfrak{q}_{v_{1,j-1}}$. Let us assume that it is true for $j = 2, \ldots, i-1$ and we are now going to show that it is true for $j = i$.

Assume that $\mathfrak{q}_{v_{1,i-1}}$ is not a clique. Then we will show that there exists a vertex $v_{1,i} \in V \setminus \mathfrak{q}_{v_{1,i-1}}$ such that $\mathfrak{p}_{v_{1,i}} = \mathfrak{q}_{v_{1,i-1}}$. If not, then $\forall v \in V \setminus \mathfrak{q}_{v_{1,i-1}}$, we have $\mathfrak{p}_v \neq \mathfrak{q}_{v_{1,i-1}}$. Since the DAG defined by the parent function $\mathfrak{p}$ is moral, $\mathfrak{q}_v$ for any $v \in V$ is complete and therefore there exists a clique C such that $\mathfrak{q}_{v_{1,i-1}} \subsetneq C$. We know from our induction assumption that $\mathfrak{q}_{v_{1,i-1}} = \{v_{1,1}, v_{1,2}, \ldots, v_{1,i-1}\}$ and therefore for any $z \in \mathfrak{q}_{v_{1,i-1}}$ we have $\mathfrak{p}_z \subset \mathfrak{q}_{v_{1,i-1}}$. As a consequence, for any $w \in C \setminus \mathfrak{q}_{v_{1,i-1}}$ we have $\mathfrak{p}_w \supset \mathfrak{q}_{v_{1,i-1}}$. Indeed, otherwise, there exists $u \in \mathfrak{q}_{v_{1,i-1}},\ u \notin \mathfrak{p}_w$ with $w \to u$ and since all elements in $\mathfrak{q}_{v_{1,i-1}}$ have their parent set in $\mathfrak{q}_{v_{1,i-1}}$, this is impossible.

Moreover, the inclusion $\mathfrak{p}_w \supset \mathfrak{q}_{v_{1,i-1}}$ implies that $v_{1,1} \in \mathfrak{p}_w$ of course and if there was a $u \notin C$ such that $u \to w$, that would create an immorality. Therefore $\mathfrak{p}_w \subset C$. Moreover since $\mathfrak{p}_w \supset \mathfrak{q}_{v_{1,i-1}}$ and by our assumption that $\mathfrak{p}_v \neq \mathfrak{q}_{v_{1,i-1}}$ for any $v \in V \setminus \mathfrak{q}_{v_{1,i-1}}$, we have $\mathfrak{p}_w \setminus \mathfrak{q}_{v_{1,i-1}} \neq \emptyset$. Take an arbitrary

$w_1 \in C \setminus \mathfrak{q}_{v_{1,i-1}}$ and choose $w_k \in \mathfrak{p}_{w_{k-1}} \setminus \mathfrak{q}_{v_{1,i-1}} \subset C \setminus \mathfrak{q}_{v_{1,i-1}}$, $k = 2, 3, \ldots$. Similarly as above we have $w_1 \leftarrow w_2 \leftarrow \ldots \leftarrow w_m \leftarrow \ldots$, and again since C is a finite set there is a repetition in the sequence (w_k) which due to acyclicity is impossible. Thus $C_1 = \mathfrak{q}_{v_{1,k_1}}$ for the first $i = k_1$ such that the set $\mathfrak{q}_{v_{1,i}}$ is a clique.

Second step. Let us assume that we have built $C_1, \ldots, C_{j-1}, j \geq 2$ with the same property that, for some $k_l \geq 2$, $C_l = \mathfrak{q}_{v_{l,k_l}}$, $l = 1 \ldots, j-1$. We define $H_l = \cup_{t=1}^{l} C_t$. We will now construct C_j.

(a) We first show that there exists a vertex $v_{j,1} \in V \setminus H_{j-1}$ such that $\mathfrak{p}_{v_{j,1}} \subset H_{j-1}$. Assume that it is not true. That is for any $v \in V \setminus H_{j-1}$ we have $\mathfrak{p}_v \setminus H_{j-1} \neq \emptyset$. Take an arbitrary $w_1 \in V \setminus H_{j-1}$ and for $k = 2, 3, \ldots$, choose $w_k \in \mathfrak{p}_{w_{k-1}} \setminus H_{j-1}$. With an argument similar to that in (a) we can prove that, due to acyclicity, this is impossible. So there exists $v_{j,1}$ as claimed.

(b) We will prove now that for any $i \geq 2$, $\mathfrak{q}_{v_{j,i-1}}$ is either a clique or there exists a vertex $v_{j,i} \in V \setminus (\mathfrak{q}_{v_{j,i-1}} \cup H_{j-1})$ with the property that $\mathfrak{p}_{v_{j,i}} = \mathfrak{q}_{v_{j,i-1}}$.

Let us prove it first for $i = 2$. Assume that $\mathfrak{q}_{v_{j,1}}$ is not a clique. Then we want to show that there exists a vertex $v_{j,2} \in V \setminus \left(\mathfrak{q}_{v_{j,1}} \cup H_{j-1}\right)$ with the property that $\mathfrak{p}_{v_{j,2}} = \mathfrak{q}_{v_{j,1}}$. Assume the contrary, i.e., $\forall v \in V \setminus \left(\mathfrak{q}_{v_{j,1}} \cup H_{j-1}\right), \mathfrak{p}_v \neq \mathfrak{q}_{v_{j,1}}$. Consider a clique C such that $\mathfrak{q}_{v_{j,1}} \subsetneq C$. We claim that $\left(C \setminus \mathfrak{q}_{v_{j,1}}\right) \cap H_{j-1} = \emptyset$: assume that it is not true, i.e., $\exists w \in \left(C \setminus \mathfrak{q}_{v_{j,1}}\right) \cap H_{j-1}$. Since w is in C but not in $\mathfrak{q}_{v_{j,1}}$, w has to be a child of $v_{j,1}$ and we have $w \leftarrow v_{j,1} \notin H_{j-1}$. But for $w \in H_{j-1}$, by construction of the cliques $C_1, C_2, \ldots, C_{j-1}$, $w \in \mathfrak{q}_{v_{t,k_t}}$ for some $t = 1, \ldots, j-1$ and therefore we have $\mathfrak{p}_w \subset H_{j-1}$. This is a contradiction to the fact that $H_{j-1} \not\ni v_{j,1} \in \mathfrak{p}_w$.

Note that for any $w \in C \setminus \mathfrak{q}_{v_{j,1}}$ we have $\mathfrak{p}_w \supset \mathfrak{q}_{v_{j,1}}$: indeed, otherwise there is $z \in \mathfrak{q}_{v_{j,1}}$ such that $z \leftarrow w$ and this is impossible because either $z = v_{j,1}$ and its parent set is in H_{j-1} or $z \in H_{j-1}$ and the same for its parent set. Moreover, the relation $\mathfrak{p}_w \supset \mathfrak{q}_{v_{j,1}}$, due to morality, implies $\mathfrak{p}_w \subset C$: indeed, we know from above that $\left(C \setminus \mathfrak{q}_{v_{j,1}}\right) \cap H_{j-1} = \emptyset$ and therefore w cannot be in H_{j-1}; so if $\mathfrak{p}_w$ contained a vertex u not in C, we would have the immorality $C \ni v_{j,1} \rightarrow w \leftarrow u$. Since we assumed that $\mathfrak{p}_w \neq \mathfrak{q}_{v_{j,1}}$ and since $\mathfrak{p}_w \supset \mathfrak{q}_{v_{j,1}}$, we have $\mathfrak{p}_w \setminus \mathfrak{q}_{v_{j,1}} \neq \emptyset$. Take $w_1 \in C \setminus \mathfrak{q}_{v_{j,1}}$ and for any $k \geq 2$ choose a vertex $w_k \in \mathfrak{p}_{w_{k-1}} \setminus \mathfrak{q}_{v_{j,1}} \subset C \setminus \mathfrak{q}_{v_{j,1}}$. Therefore we have the path $w_1 \leftarrow w_2 \leftarrow \ldots \leftarrow w_m \leftarrow \ldots$ with $w_m \in C \setminus \mathfrak{q}_{v_{j,1}}$. By finiteness of V we conclude that there is a repetition in the sequence $(w_k)_{k=1,\ldots}$ which produces a cycle and we have a contradiction. We name C_j the clique C.

Let us now assume that there exists $v_{j,k} \in C_j \setminus \mathfrak{q}_{v_{j,k-1}}$ with the property that $\mathfrak{p}_{v_{j,k}} = \mathfrak{q}_{v_{j,k-1}} \subset C_j,\ k = 2, \ldots, i-1$. We are going to prove that either $\mathfrak{q}_{v_{j,i-1}}$ is the clique C_j or we can choose $v_{j,i} \in V \setminus (\mathfrak{q}_{v_{j,i-1}} \cup H_{j-1})$ such that $\mathfrak{p}_{v_{j,i}} = \mathfrak{q}_{v_{j,i-1}}$. Assume that $\mathfrak{q}_{v_{j,i-1}}$ is not a clique. Then we want to show that there exists $v_{j,i} \in V \setminus (\mathfrak{q}_{v_{j,i-1}} \cup H_{j-1})$ such that $\mathfrak{p}_{v_{j,i}} = \mathfrak{q}_{v_{j,i-1}}$. To do so, we can use the same argument by contradiction that we used above for $i = 2$, replacing $v_{j,1}$ by $v_{j,i-1}$, $v_{j,2}$ by $v_{j,i}$ and replacing C by C_j. In this way, we construct the sequence of cliques $C_1, C_2, \ldots, C_K$.

We claim that the ordering $(C_1, \ldots, C_K)$ chosen as above is a $\mathfrak{p}$-perfect ordering of the cliques. To see this we note that $S_j := C_j \cap H_{j-1}$ is equal to $\mathfrak{p}_{v_{j,1}}$: this follows from the fact that $C_j = \mathfrak{p}_{v_{j,1}} \cup \{v_{j,1}, \ldots, v_{j,k_j}\}$. Thus since the DAG is moral, then S_j being a set of parents is complete and moreover since S_j is included in H_{j-1}, we must have $S_j \subset C_i$ for some $i < j$. Therefore $(C_1, \ldots, C_K)$ is a perfect ordering of cliques. Now it is clear that the sets $S_j, j = 2, \ldots, K$ are the separators for the graph G. Since $v_{j,1} \in C_j$ and $\mathfrak{p}_{v_{j,1}} = S_j$, the ordering $(C_1, \ldots, C_K)$ is also $\mathfrak{p}$-perfect.

□

Proof of Lemma 2.1 (3). The numbering in $(2.1)_{mf}$ is now immediately obtained by taking $v_{1,1}$ as defined in the proof of Lemma 2.1 (1) and taking $v_{1,i}, i = 2, \ldots, c_1 = k_1$ as defined in the first step of the proof of Lemma 2.1 (2), Part (a) above. For $j \geq 2$ we take, as v_{j,s_j+1} and as $v_{j,l}, l = s_j + 2, \ldots, c_j$ respectively the vertices $v_{j,1}$ and $v_{j,l}, l = 2, \ldots, k_j$ defined in the proof of Lemma 2.1 (2), Part (b) above, where clearly $k_j = c_j - s_j$. This completes the proof of Lemma 2.1. □

Example 2.1. Let us now illustrate the $\mathfrak{p}$-perfect ordering of cliques and vertices as given in (2.1) and the sets $\mathfrak{P}(V)$ and $\mathfrak{Q}(V)$.

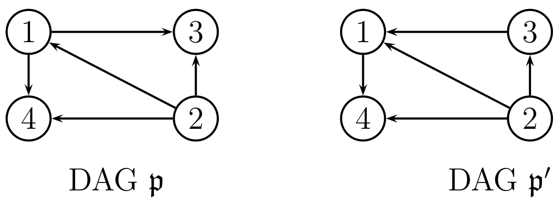


Figure 1: Ordering of the vertices

Consider DAG $\mathfrak{p}$ in Fig. 1. above. Clearly $v_{1,1} = 2$ because $\mathfrak{p}_2 = \emptyset$. Then $v_{1,2} = 1$ because 1 is the only vertex with parent set 1. At this point we can choose between $v_{1,3} = 3$ and $v_{1,3} = 4$ and following the construction in the proof of Lemma 2.1, this determines whether $C_1 = \{1, 2, 3\}$ or $C_1 = \{1, 2, 4\}$ respectively. Both possibilities lead to a $\mathfrak{p}$-perfect ordering of the cliques, that is

$$o_{\mathfrak{p}} = (C_1 = \{1, 2, 3\},\ C_2 = \{1, 2, 4\}) \text{ or } o_{\mathfrak{p}} = (C_1 = \{1, 2, 4\},\ C_2 = \{1, 2, 3\}).$$

Let us now consider DAG $\mathfrak{p}'$ with the same skeleton. Again, $v_{1,1} = 2$ but $v_{1,2}$ is now 3 since this is the only vertex with parent set $\{2\}$ and therefore $C_1 = \{1, 2, 3\}$. We no longer have a choice for the $\mathfrak{p}'$-perfect ordering of cliques: it has to be

$$o_{\mathfrak{p}'} = (C_1 = \{1, 2, 3\}, C_2 = \{1, 2, 4\}).$$

For $\mathcal{P} = \{\mathfrak{p}, \mathfrak{p}'\}$, the sets $\mathfrak{P}(V)$ and $\mathfrak{Q}(V)$ are

$$\mathfrak{P}(V) = \{\emptyset,\ \{2\},\ \{1, 2\}\}, \qquad \mathfrak{Q}(V) = \{\{2\},\ \{1, 2, 3\},\ \{1, 2, 4\}\}$$

and we verify that $\mathfrak{P}(V) \setminus \mathcal{S} = \mathfrak{Q}(V) \setminus \mathcal{C} = \{\{2\}\}$.

4 Moments

4.2 The $\mathcal{P}$-Dirichlet

4.2.1 Example 3.1 continued

Given nonnegative integers $r_{\underline{i}}$, $\underline{i} \in \mathcal{I}$, for any $D \subset V$ we define

$$r^D_{\underline{m}} = \sum_{\underline{i} \in \mathcal{I} :\, \underline{i}_D = \underline{m}} r_{\underline{i}}.$$

For $D=\emptyset$ we write $r^{\emptyset}=r=\sum_{\underline{i}\in\mathcal{I}} r_{\underline{i}}$. The equality of the moments obtained from $(4.1)_{mf}$ (and (4.1) below) for both $\mathfrak{p}$ and $\mathfrak{p}'$ yields

$$\begin{aligned}
\mathbb{E}\prod_{\underline{i}\in\mathcal{I}}\mathbf{p}(\underline{i})^{r_{\underline{i}}} &= \frac{\prod_{m\in\mathcal{I}_2}\left(\alpha^2_m\right)^{r^2_m}}{(\tilde{\alpha})^{r}} \prod_{n\in\mathcal{I}_2}\frac{\prod_{m\in\mathcal{I}_5}\left(\alpha^{5|2}_{m|n}\right)^{r^{25}_{(n,m)}}}{\left(\tilde{\alpha}^2_n\right)^{r^2_n}} \prod_{\underline{n}\in\mathcal{I}_{25}}\frac{\prod_{m\in\mathcal{I}_1}\left(\alpha^{1|25}_{m|\underline{n}}\right)^{r^{125}_{(m,\underline{n})}}}{\left(\tilde{\alpha}^{25,1}_{\underline{n}}\right)^{r^{25}_{\underline{n}}}}\\
&\quad\times\prod_{\underline{n}\in\mathcal{I}_{25}}\frac{\prod_{m\in\mathcal{I}_3}\left(\alpha^{3|25}_{m|\underline{n}}\right)^{r^{325}_{(m,\underline{n})}}}{\left(\tilde{\alpha}^{25,3}_{\underline{n}}\right)^{r^{25}_{\underline{n}}}} \prod_{\underline{n}\in\mathcal{I}_{35}}\frac{\prod_{m\in\mathcal{I}_4}\left(\alpha^{4|35}_{m|\underline{n}}\right)^{r^{435}_{(m,\underline{n})}}}{\left(\tilde{\alpha}^{35}_{\underline{n}}\right)^{r^{35}_{\underline{n}}}}\\
&= \frac{\prod_{m\in\mathcal{I}_3}\left(\beta^3_m\right)^{r^3_m}}{(\tilde{\beta})^{r}} \prod_{n\in\mathcal{I}_3}\frac{\prod_{m\in\mathcal{I}_5}\left(\beta^{5|3}_{m|n}\right)^{r^{35}_{(n,m)}}}{\left(\tilde{\beta}^3_n\right)^{r^3_n}} \prod_{\underline{n}\in\mathcal{I}_{35}}\frac{\prod_{m\in\mathcal{I}_2}\left(\beta^{2|35}_{m|\underline{n}}\right)^{r^{235}_{(m,\underline{n})}}}{\left(\tilde{\beta}^{35,2}_{\underline{n}}\right)^{r^{35}_{\underline{n}}}}\\
&\quad\times\prod_{\underline{n}\in\mathcal{I}_{35}}\frac{\prod_{m\in\mathcal{I}_4}\left(\beta^{4|35}_{m|\underline{n}}\right)^{r^{435}_{(m,\underline{n})}}}{\left(\tilde{\beta}^{35,4}_{\underline{n}}\right)^{r^{35}_{\underline{n}}}} \prod_{\underline{n}\in\mathcal{I}_{25}}\frac{\prod_{m\in\mathcal{I}_1}\left(\beta^{1|25}_{m|\underline{n}}\right)^{r^{125}_{(m,\underline{n})}}}{\left(\tilde{\beta}^{25}_{\underline{n}}\right)^{r^{25}_{\underline{n}}}}.
\end{aligned}$$

Since there are no factorial powers in r^2_m on the right-hand side of the equation above the terms in r^2_m on the left-hand side must cancel out, that is $\alpha^2_m=\tilde{\alpha}^2_m$. Similarly, $\beta^3_m=\tilde{\beta}^3_m$. The factorial power $r^{125}_{\underline{n}}$ on the right- and left-hand side must be the same and therefore $\alpha^{1|25}_{m|\underline{n}}=\beta^{1|25}_{m|\underline{n}}$. Similarly, $\alpha^{3|25}_{m|\underline{n}}=\beta^{2|35}_{m|\underline{n}}$, $\alpha^{4|35}_{m|\underline{n}}=\beta^{4|35}_{m|\underline{n}}$ and also $\tilde{\alpha}=\tilde{\beta}$. For the factorial powers in $r^{25}_{\underline{n}}$ we observe that on the left-hand side there is one power in the numerator and two in the denominator, while on the right-hand side there is only one power in the denominator. Therefore the factorial power in the numerator must cancel with one of the two factorial powers of $\tilde{\alpha}^{25,3}_{\underline{n}}$ or of $\tilde{\alpha}^{25,1}_{\underline{n}}$ in the denominator. This means that $\forall\,\underline{n}\in\mathcal{I}_{25}$

- either we have the cancelation $\alpha^{5|2}_{\underline{n}}=\tilde{\alpha}^{25,3}_{\underline{n}}$ and therefore $\tilde{\beta}^{25}_{\underline{n}}=\tilde{\alpha}^{25,1}_{\underline{n}}$,
- or we have the cancelation $\alpha^{5|2}_{\underline{n}}=\tilde{\alpha}^{25,1}_{\underline{n}}$ and therefore $\tilde{\beta}^{25}_{\underline{n}}=\tilde{\alpha}^{25,3}_{\underline{n}}$.

The first choice means that we associate the separator $\{2,5\}$ with the clique $\{1,2,5\}$ while in the second we associate $\{2,5\}$ with the clique $\{2,3,5\}$. This two choices correspond to two different $\mathfrak{p}$-perfect orders of the cliques:

$$\begin{aligned}
o^{(1)}_{\mathfrak{p}} &= (C_1=\{2,3,5\},\ C_2=\{1,2,5\},\ C_3=\{3,4,5\})\\
o^{(2)}_{\mathfrak{p}} &= (C_1=\{1,2,5\},\ C_2=\{2,3,5\},\ C_3=\{3,4,5\}),
\end{aligned}$$

respectively. Of course, we could also exchange the cliques C_2 and C_3 in both orders. What is important is the pairings $(\{2,5\},\{1,2,5\})$ or $(\{2,5\},\{2,3,5\})$, respectively.

Similarly, for the factorial powers $r^{35}_{\underline{n}}$ on the right-hand side one can choose to cancel the factorial power of $r^{35}_{\underline{n}}$ in the numerator with factorial powers of either $\tilde{\beta}^{35,2}_{\underline{n}}$ or $\tilde{\beta}^{35,4}_{\underline{n}}$. Consequently, $\forall\, \underline{n} \in \mathcal{I}_{35}$

- either we have the cancelation $\beta^{5|3}_{\underline{n}} = \tilde{\beta}^{35,2}_{\underline{n}}$ and therefore $\tilde{\alpha}^{35}_{\underline{n}} = \tilde{\beta}^{35,4}_{\underline{n}}$,
- or we have the cancelation $\beta^{5|3}_{\underline{n}} = \tilde{\beta}^{35,4}_{\underline{n}}$ and therefore $\tilde{\alpha}^{35}_{\underline{n}} = \tilde{\alpha}^{35,2}_{\underline{n}}$,

which corresponds to the two $\mathfrak{p}'$-perfect orders:

$$\begin{aligned} o^{(1)}_{\mathfrak{p}'} &= (C_1 = \{2,3,5\},\ C_2 = \{3,4,5\},\ C_3 = \{1,2,5\}) \\ o^{(2)}_{\mathfrak{p}'} &= (C_1 = \{3,4,5\},\ C_2 = \{2,3,5\},\ C_3 = \{1,2,5\}), \end{aligned}$$

respectively. Again here we could exchange C_2 and C_3 in both cases. What is important are the pairings $(\{3,5\},\{3,4,5\})$ and $(\{3,5\},\{2,3,5\})$, respectively.

From any of these cancelation possibilities we obtain the same formula of moments

$$\mathbb{E}\prod_{\underline{i}\in\mathcal{I}} \mathbf{p}(\underline{i})^{r_{\underline{i}}} = \frac{\prod_{\underline{n}\in\mathcal{I}_{125}}\left(\nu^{125}_{\underline{n}}\right)^{r^{125}_{\underline{n}}} \prod_{\underline{n}\in\mathcal{I}_{235}}\left(\nu^{235}_{\underline{n}}\right)^{r^{235}_{\underline{n}}} \prod_{\underline{n}\in\mathcal{I}_{345}}\left(\nu^{345}_{\underline{n}}\right)^{r^{345}_{\underline{n}}}}{(\mu)^{r} \prod_{\underline{m}\in\mathcal{I}_{25}}\left(\mu^{25}_{\underline{m}}\right)^{r^{25}_{\underline{m}}} \prod_{\underline{m}\in\mathcal{I}_{35}}\left(\mu^{35}_{\underline{m}}\right)^{r^{35}_{\underline{m}}}},$$

but with different constraints for the parameters:

either (I)

$$\mu = \sum_{\underline{m}\in\mathcal{I}_{125}} \nu^{125}_{\underline{m}} = \sum_{\underline{m}\in\mathcal{I}_{345}} \nu^{345}_{\underline{m}} \quad (= \sum_{\underline{m}\in\mathcal{I}_{235}} \nu^{235}_{\underline{m}}),$$

$$\mu^{25}_{\underline{n}} = \sum_{m\in\mathcal{I}_1} \nu^{125}_{(m,\underline{n})} = \sum_{m\in\mathcal{I}_3} \nu^{235}_{(m,\underline{n})},$$

$$\mu^{35}_{\underline{n}} = \sum_{m\in\mathcal{I}_4} \nu^{345}_{(m,\underline{n})} = \sum_{m\in\mathcal{I}_2} \nu^{235}_{(m,\underline{n})},$$

corresponding to the family of orders $O_{\mathcal{P}} = (o^{(2)}_{\mathfrak{p}},\, o^{(2)}_{\mathfrak{p}'})$.

or (II)

$$\mu = \sum_{\underline{m}\in\mathcal{I}_{125}} \nu^{125}_{\underline{m}} = \sum_{\underline{m}\in\mathcal{I}_{235}} \nu^{235}_{\underline{m}},$$

$$\mu_{\underline{n}}^{25} = \sum_{m \in \mathcal{I}_1} \nu_{(m,\underline{n})}^{125} = \sum_{m \in \mathcal{I}_3} \nu_{(m,\underline{n})}^{235},$$

$$\mu_{\underline{n}}^{35} = \sum_{m \in \mathcal{I}_4} \nu_{(m,\underline{n})}^{345},$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathfrak{p}}^{(2)},\, o_{\mathfrak{p}'}^{(1)})$.

or (III)

$$\mu = \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235} = \sum_{\underline{m} \in \mathcal{I}_{345}} \nu_{\underline{m}}^{345},$$

$$\mu_{\underline{n}}^{25} = \sum_{m \in \mathcal{I}_1} \nu_{(m,\underline{n})}^{125},$$

$$\mu_{\underline{n}}^{35} = \sum_{m \in \mathcal{I}_4} \nu_{(m,\underline{n})}^{345} = \sum_{m \in \mathcal{I}_2} \nu_{(m,\underline{n})}^{235},$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathfrak{p}}^{(1)},\, o_{\mathfrak{p}'}^{(2)})$.

or (IV)

$$\mu = \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235},$$

$$\mu_{\underline{n}}^{25} = \sum_{m \in \mathcal{I}_1} \nu_{(m,\underline{n})}^{125},$$

$$\mu_{\underline{n}}^{35} = \sum_{m \in \mathcal{I}_4} \nu_{(m,\underline{n})}^{345},$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathfrak{p}}^{(1)},\, o_{\mathfrak{p}'}^{(1)})$.

We note that we obtained four different families of distributions, that is as many as the number of combinations of pairs (S_l, C_l), where $S_l = \{2,5\}$ or $S_l = \{3,5\}$. Of course, in general, choices will multiply with the number of separators with different possible pairings. In fact, more generally, choices may multiply with the number of elements of $\mathfrak{P}$ with different possible pairings in $\mathfrak{Q}$ and also with the size of $\mathcal{P}$.

In this example, we see that we have the poset $(IV) \to (II, III) \to (I)$ of families of $\mathcal{P}$-Dirichlet distributions, with family (IV) being the maximal family, while the minimal family (I) is just the hyper Dirichlet. Though in our example it is easy to see that (IV) is the maximal family of the poset we are unable to prove that for any given $\mathcal{P}$, there exists such a unique maximal family. On the other hand, the minimal family is always the hyper Dirichlet.

4.2.2 The moment formula

Proof of Theorem 4.1. Assume that $(\mathbf{p}(\underline{i}),\ i \in \mathcal{I})$ is $\mathcal{P}$-Dirichlet distributed. That is representation $(3.1)_{mf}$ holds and for any $\mathfrak{p} \in \mathcal{P}$, the random vectors $(\mathbf{p}^{v|\mathfrak{p}_v}_{m|\underline{k}}, m \in \mathcal{I}_v)$ follow a Dirichlet distribution $\mathrm{Dir}(\alpha^{v|\mathfrak{p}_v}_{m|\underline{k}},\ m \in \mathcal{I}_v)$, $\underline{k} \in \mathcal{I}_{\mathfrak{p}_v}$, $v \in V$ and are independent. We have to identify the parameters $\alpha^{v|\mathfrak{p}_v}_{m|\underline{k}}$, $m \in \mathcal{I}_v$, $\underline{k} \in \mathcal{I}_{\mathfrak{p}_v}$, $v \in V$, such that (see $(4.1)_{mf}$ and $(4.2)_{mf}$)

$$\mathbb{E} \prod_{v\in V} \prod_{\underline{k}\in\mathcal{I}_{\mathfrak{p}_v}} \prod_{m\in\mathcal{I}_v} \left(\mathbf{p}^{v|\mathfrak{p}_v}_{m|\underline{k}}\right)^{r^{\mathfrak{q}_v}_{(\underline{k},m)}} = \prod_{v\in V} \prod_{\underline{k}\in\mathcal{I}_{\mathfrak{p}_v}} \frac{\prod_{m\in\mathcal{I}_v} (\alpha^{v|\mathfrak{p}_v}_{m|\underline{k}})^{r^{\mathfrak{q}_v}_{(\underline{k},m)}}}{(\tilde{\alpha}^{\mathfrak{p}_v}_{\underline{k}})^{r^{\mathfrak{p}_v}_{\underline{k}}}}, \tag{4.1}$$

with

$$\tilde{\alpha}^{\mathfrak{p}_v}_{\underline{k}} = \sum_{m\in\mathcal{I}_v} \alpha^{v|\mathfrak{p}_v}_{m|\underline{k}} \qquad \forall\, \underline{k} \in \mathcal{I}_{\mathfrak{p}_v}, \tag{4.2}$$

is equal to the right-hand side of $(4.3)_{mf}$ with consistency conditions given in $(4.4)_{mf}$.

Note that the result in the opposite direction, that is the fact that the formula $(4.3)_{mf}$ for moments, together with the constraints $(4.4)_{mf}$, implies that the $\mathcal{P}$-Dirichlet distribution for $(\mathbf{p}(\underline{i}),\ i \in \mathcal{I})$, follows immediately from the property that the distribution is uniquely determined by moments.

Consider an arbitrary collection $O_{\mathcal{P}}$ of $\mathfrak{p}$-perfect orders, $\mathfrak{p} \in \mathcal{P}$. Fix an arbitrary $\mathfrak{p} \in \mathcal{P}$ and consider a $\mathfrak{p}$-perfect order $o_{\mathfrak{p}} \in O_{\mathcal{P}}$. We will now relate the sets Q^C_i, $i = 0, 1, \ldots, j_C$, $C \in \mathcal{C}$, to the numbering of vertices imposed by $o_{\mathfrak{p}}$ as given in (2.1). Clearly $C = C_l$ for some $l \in \{1, \ldots, K\}$. For ease of notation, we suppress the subscript l on C_l in the remainder of this proof. We define $j(l,i),\ i \in \{1, \ldots, j_C - 1\}$ to be the index of the vertex which satisfies

$$Q^C_i = \mathfrak{q}_{v_{l,j(l,i)}} = \mathfrak{p}_{v_{l,j(l,i)+1}}\ . \tag{4.3}$$

We also note that

$$C = Q^C_0 = \mathfrak{q}_{v_{l,c_l}} \qquad \text{and} \qquad S = S_l = Q^C_{jC} = \mathfrak{p}_{v_{l,s_l+1}}. \tag{4.4}$$

We will now define the $\alpha^{v|\mathfrak{p}_v}_{m|\underline{k}}$'s in terms of the ν^A's. For any $l \in \{1, \ldots, K\}$, if $v = v_{l,j(l,i)}$, set

$$\alpha^{v|\mathfrak{p}_v}_{m|\underline{k}} := \nu^{Q^C_i}_{(\underline{k},m)} \qquad \forall\, (\underline{k}, m) \in \mathcal{I}_{Q^C_i}. \tag{4.5}$$

For any $l \in \{1, \ldots, K\}$, if $v = v_{l,j}$ and $j \neq j(l,i)$ for any i, we define

$$i_j = \min\{i :\ \mathfrak{q}_v \subset Q^{C_l}_i\} \tag{4.6}$$

and then set

$$\alpha_{m|\underline{k}}^{v|\mathfrak{p}_v} := \sum_{\underline{n} \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathfrak{q}_v} \nu_{(\underline{n},m,k)}^{Q_{i_j}^C} \qquad \forall\, (\underline{k}, m) \in \mathcal{I}_{\mathfrak{q}_v}. \tag{4.7}$$

We will now show that, if $v_{l,j}$ is such that $j \neq j(l,i)$ for any $i \in \{0, \ldots, j_C - 1\}$, then we have

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathfrak{p}_{v_{l,j}}} = \tilde{\alpha}_{(\underline{k},m)}^{\mathfrak{p}_{v_{l,j+1}}} \quad \forall\, (\underline{k}, m) \in \mathcal{I}_{\mathfrak{p}_{v_{l,j+1}}}. \tag{4.8}$$

Consider first the case when $j(l, i_j) = j + 1$. By (4.7) we have

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathfrak{p}_{v_{l,j}}} = \sum_{n \in \mathcal{I}_{v_{l,j+1}}} \nu_{(n,\underline{k},m)}^{Q_{i_j}^C} .$$

Since $Q_{i_j}^C = \mathfrak{q}_{v_{l,j+1}}$ from the above equality and (4.5) we get

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathfrak{p}_{v_{l,j}}} = \sum_{\underline{n} \in \mathcal{I}_{v_{l,j+1}}} \alpha_{n|(\underline{k},m)}^{v_{l,j+1}|\mathfrak{p}_{v_{l,j+1}}}.$$

Thus (4.8) follows from (4.2).

Second, consider the case $j(l, i_j) > j + 1$. By (4.7) we have

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathfrak{p}_{v_{l,j}}} = \sum_{\underline{n} \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathfrak{q}_{v_{l,j}}} \nu_{(\underline{n},\underline{k},m)}^{Q_{i_j}^C} = \sum_{n_1 \in \mathcal{I}_{v_{l,j+1}}} \sum_{\underline{n}_2 \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathfrak{q}_{v_{l,j+1}}} \nu_{(n_1,\underline{n}_2,\underline{k},m)}^{Q_{i_j}^C},$$

where the second equality follows from the fact that $\mathfrak{q}_{v_{l,j+1}} = \mathfrak{q}_{v_{l,j}} \cup \{v_{l,j+1}\}$. Applying (4.7) to the inner sum we obtain

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathfrak{p}_{v_{l,j}}} = \sum_{n_1 \in \mathcal{I}_{v_{l,j+1}}} \alpha_{n_1|(\underline{k},m)}^{v_{l,j+1}|\mathfrak{p}_{v_{l,j+1}}}.$$

Thus (4.8) follows from (4.2).

Due to (4.8) we have cancelations in the right-hand side of (4.1) and the only terms left are:

- in the numerator: $\alpha_{m|\underline{k}}^{v|\mathfrak{p}_v} = \nu_{(\underline{k},m)}^{Q_i^C}$ for $v = v_{l,j(l,i)}$, where $i \in \{0, \ldots, j_C - 1\}$.

- in the denominator: $\tilde{\alpha}_{\underline{k}}^{\mathfrak{p}_v}$ for $v = v_{l,j(l,i)+1}$ where $i \in \{1, \ldots, j_C\}$.

In particular, $j(l, j_C) = s_l$ in general and for $l = 1$, $s_l = 0$ so that, in the denominator, we have parameters indexed by $\mathfrak{p}_{v_{l,s_l+1}} = S_l$ and $\mathfrak{p}_{v_{1,1}} = \emptyset$.

To complete the proof, that is to show that the right-hand side of (4.1) is equal to the right-hand side of $(4.3)_{mf}$, it remains to show that for any $l \in \{1, \ldots, K\}$ and $\underline{i} \in \{1, \ldots, j_C\}$, we have

$$\mu_{\underline{k}}^{Q_i^C} = \tilde{\alpha}_{\underline{k}}^{\mathfrak{p}_{v_{l,j(l,i)+1}}} \quad \forall \underline{k} \in \mathcal{I}_{Q_i^C}. \tag{4.9}$$

Note that

(i) either $j(l, i) + 1 = j(l, i - 1)$, that is $\mathfrak{q}_{v_{l,j(l,i)+1}} = Q_{i-1}^C$,

(ii) or $j(l, i)+1$ is not of the form $j(l, \tilde{i})$ for some $\tilde{i} \in \{i+1, \ldots, j_C\}$ (observe - see (4.6) - that in this case we have $i_{j(l,i)+1} = i - 1$).

Let's consider case (i) first. Using (4.2) and then (4.5) for all $\underline{k} \in \mathcal{I}_{Q_i^C}$ (note that $Q_i^C = \mathfrak{p}_{v_{l,j(l,i)+1}}$) we obtain

$$\tilde{\alpha}_{\underline{k}}^{\mathfrak{p}_{v_{l,j(l,i)+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \alpha_{m|\underline{k}}^{v_{l,j(l,i)+1}|\mathfrak{p}_{v_{l,j+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \nu_{(\underline{k},m)}^{Q_{i-1}^C}.$$

Since $\{v_{l,j(l,i)+1}\} = \mathfrak{q}_{v_{l,j(l,i)+1}} \setminus \mathfrak{q}_{v_{l,j(l,j)}} = Q_{i-1}^C \setminus Q_i^C$, due to $(4.4)_{mf}$ we obtain (4.9).

For case (ii), we use again (4.2) and then (4.7) to arrive at

$$\tilde{\alpha}_{\underline{k}}^{\mathfrak{p}_{v_{l,j(l,i)+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \sum_{\underline{n} \in \mathcal{I}_{Q_{i-1}^C \setminus \mathfrak{q}_{v_{l,j(l,i)+1}}}} \nu_{(m,\underline{n},\underline{k})}^{Q_{i-1}^C} = \sum_{(m,\underline{n}) \in \mathcal{I}_{Q_{i-1}^C \setminus \mathfrak{q}_{v_{l,j(l,i)}}}} \nu_{(m,\underline{n},\underline{k})}^{Q_{i-1}^C},$$

where the last equation follows from the fact that $\mathfrak{q}_{v_{l,j(i,l)+1}} = \mathfrak{q}_{v_{l,j(l,i)}} \cup \{v_{l,j(l,i)+1}\}$. Moreover $\mathfrak{q}_{v_{l,j(l,i)}} = Q_i^C$, therefore (4.9) follows now from assumption (4.2). □

5 The $\mathcal{P}$-Dirichlet as a prior distribution

5.1 Dimension of the $\mathcal{P}$-Dirichlet family

Proof of Theorem 5.1. From $(4.3)_{mf}$ and $(4.4)_{mf}$, we see that the parameters are the $\nu_{\underline{m}}^A$ and we need not count the $\mu_{\underline{n}}^B$ since they are defined by the constraints of the type $(4.4)_{mf}$. Clearly, there are $\sum_{Q\in\mathfrak{Q}}\prod_{v\in Q}|\mathcal{I}_v|$ such parameters $\nu^A_{\underline{m}}$. They are not all free since an element $S\in\mathcal{S}$ can be equal to an element $Q^C_{j_C}$ for several $C\in\mathcal{C}$. More precisely for all $C\in\mathcal{C}$ such that there exists $o\in O_{\mathcal{P}}$ with $S\overset{o}{\to}C$, we would have

$$Q^C_{j_C}\subsetneq Q^C_{j_C-1}$$

and therefore by $(4.4)_{mf}$, we have (N_S-1) equality of the type

$$\mu^S_{\underline{n}}=\mu^{Q^C_{j_C}}_{\underline{n}}=\sum_{\underline{k}\in I_{Q^C_{j_C-1}}\setminus S}\nu^{Q^C_{j_C-1}}_{(\underline{n},\underline{k})}$$

and thus (N_S-1) constraints for a given $\nu^{Q^C_{j_C-1}}_{(\underline{n},\underline{k})}$ and thus a total of $(N_S-1)\prod_{v\in S}|\mathcal{I}_v|$ constraints for each S. We now note that if $B\in\mathfrak{R}$, that is if B is not a separator, the corresponding equation $(4.4)_{mf}$ is not a constraint since then there is only one clique to which B can belong , i.e. only one A such that

$$B=Q^C_i\subsetneq A=Q^C_{i-1}\ \text{ and }\ \mu^B_{\underline{n}}=\sum_{\underline{k}\in I_{A\setminus B}}\nu^A_{(\underline{n},\underline{k})}\ .$$

It follows that $(5.1)_{mf}$ is proved.

In the case of the hyper Dirichlet, a similar argument shows us that the total number of parameters is $\sum_{C\in\mathcal{C}}\prod_{v\in Q}|\mathcal{I}_v|$. The constraints given by equations of the type $(4.4)_{mf}$ are of the form

$$\mu^S_{\underline{n}}=\sum_{\underline{k}\in C\setminus S}\nu^C_{(\underline{n},\underline{k})}$$

for any C containing S. Since considering the hyper Dirichlet is equivalent to taking $\mathcal{P}$ as the set of all DAGs Markov equivalent to G, N_S is nothing but the number of cliques containing S and equation $(5.2)_{mf}$ follows.

To see that $\mathcal{N}_{\mathcal{P}}$ is always strictly greater than $\mathcal{N}_{HP}$, we observe that if $\mathfrak{R}\neq\emptyset$, then $\mathcal{C}\subsetneq\mathfrak{Q}$ and therefore

$$\sum_{Q\in\mathfrak{Q}}\prod_{v\in Q}|\mathcal{I}_v|>\sum_{C\in\mathcal{C}}\prod_{v\in C}|\mathcal{I}_v|$$

Moreover, even if $\mathfrak{R}=\emptyset$ but the $\mathcal{P}$-Dirichlet is not the hyper Dirichlet, then, by Theorem 4.4, $(4.16)_{mf}$ cannot be satisfied and for each S, N_S in the $\mathcal{P}$-Dirichlet is less than or equal to the corresponding N_S in the hyper Dirichlet. Inequality $(5.3)_{mf}$ follows immediately. □

5.2 Conjugacy and directed strong hyper Markov property

Proof of Theorem 5.2. The conditional distribution of $\mathbf{N}$ given $\mathbf{p}$ has the density (with respect to the counting measure) which, up to a multiplicative scalar, is equal to $\prod_{\underline{i}\in\mathcal{I}}[\mathbf{p}(\underline{i})]^{N(\underline{i})}$. Then, by the generalized Bayes formula for any table $(r(\underline{i}),\, \underline{i}\in\mathcal{I})$ of nonnegative integers

$$\mathbb{E}\left(\prod_{\underline{i}\in\mathcal{I}}[\mathbf{p}(\underline{i})]^{r_{\underline{i}}}\,\middle|\,\mathbf{N}=(n(\underline{i}),\,\underline{i}\in\mathcal{I})\right)=\frac{\mathbb{E}\prod_{\underline{i}\in\mathcal{I}}[\mathbf{p}(\underline{i})]^{r_{\underline{i}}+n(\underline{i})}}{\mathbb{E}\prod_{\underline{i}\in\mathcal{I}}[\mathbf{p}(\underline{i})]^{n(\underline{i})}}.$$

Applying $(4.3)_{mf}$ to the numerator and denominator we see that the right hand side above can be written as

$$\frac{\prod_{A\in\mathfrak{Q}}\prod_{\underline{m}\in\mathcal{I}_A}(\nu_{\underline{m}}^A)^{r_{\underline{m}}^A+n_{\underline{m}}^A}}{\prod_{A\in\mathfrak{Q}}\prod_{\underline{m}\in\mathcal{I}_A}(\nu_{\underline{m}}^A)^{n_{\underline{m}}^A}}\,\frac{\prod_{B\in\mathfrak{P}}\prod_{\underline{k}\in\mathcal{I}_B}(\mu_{\underline{k}}^B)^{n_{\underline{k}}^B}}{\prod_{B\in\mathfrak{P}}\prod_{\underline{k}\in\mathcal{I}_B}(\mu_{\underline{k}}^B)^{r_{\underline{k}}^B+n_{\underline{k}}^B}}.$$

Note that

$$\frac{(\nu_{\underline{m}}^A)^{r_{\underline{m}}^A+n_{\underline{m}}^A}}{(\nu_{\underline{m}}^A)^{n_{\underline{m}}^A}}=(\nu_{\underline{m}}^A+n_{\underline{m}}^A)^{r_{\underline{m}}^A}\qquad\text{and}\qquad\frac{(\mu_{\underline{k}}^B)^{r_{\underline{k}}^B+n_{\underline{k}}^B}}{(\mu_{\underline{k}}^B)^{n_{\underline{k}}^B}}=(\mu_{\underline{k}}^B+n_{\underline{k}}^B)^{r_{\underline{k}}^B}.$$

Consequently,

$$\mathbb{E}\left(\prod_{\underline{i}\in\mathcal{I}}[\mathbf{p}(\underline{i})]^{r_{\underline{i}}}\,\middle|\,\mathbf{N}\right)=\frac{\prod_{A\in\mathfrak{Q}}\prod_{\underline{m}\in\mathcal{I}_A}(\nu_{\underline{m}}^A+n_{\underline{m}}^A)^{r_{\underline{m}}^A}}{\prod_{B\in\mathfrak{P}}\prod_{\underline{k}\in\mathcal{I}_B}(\mu_{\underline{k}}^B+n_{\underline{k}}^B)^{r_{\underline{k}}^B}}$$

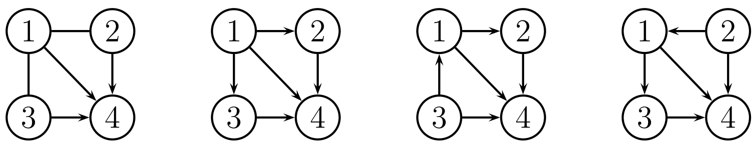
and thus it follows from $(4.3)_{mf}$ and the fact that the distribution is uniquely determined by moments that the posterior distribution of $\mathbf{p}$ given the counts $\mathbf{N}=(n(\underline{i}),\,\underline{i}\in\mathcal{I})$ is $\mathcal{P}$-Dirichlet with parameters updated by counts. We note that the parameters: $\nu_{\underline{m}}^A+n_{\underline{m}}^A$, $A\in\mathfrak{Q}$, and $\mu_{\underline{k}}^B+n_{\underline{k}}^B$, $B\in\mathfrak{P}$, of the posterior distribution of $\mathbf{p}$ satisfy the constraints of the type $(4.4)_{mf}$. Indeed, this is due to the facts that these constraints are linear in the parameters, that the original parameters $\nu_{\underline{m}}^A$ and $\mu_{\underline{k}}^B$ satisfy such constraints by assumption and that such constraints are also trivially satisfied by the counts $n_{\underline{m}}^A$ and $n_{\underline{m}}^B$. This proves that the $\mathcal{P}$-Dirichlet forms a conjugate family of distribution.

The directed strong hyper Markov property of the $\mathcal{P}$-Dirichlet holds true for every $\mathfrak{p}\in\mathcal{P}$ because of the independences (see Def. 3.1 in the main file) given in its construction. □

5.3 Arbitrary DAGs

In the main file, we mentioned the possibility of extending the $\mathcal{P}$-Dirichlet to arbitrary DAGs, not necessarily moral. For such an extension, given an undirected graph G, we take the family $\mathcal{P}$ to be the family of arbitrary DAGs Markov equivalent to a given essential graph. We will now give two examples that suggest what this extended $\mathcal{P}$-Dirichlet might look like.

Let the essential graph of the Markov equivalence class of DAGs be as Figure 2 (a) below (see also Figure 3 of [1]). The equivalence class consists of graphs (b), (c) and (d) in Figure 2.



(a) *(b)* *(c)* *(d)*

Figure 2: Essential graph (a) and its equivalence class {(b), (c), (d)}.

Equating moments as in the proof of Theorem 4.1 we see that $(4.3)_{mf}$ holds with $\mathfrak{P} = \{\emptyset, 1, 123\}$ and $\mathfrak{Q} = \{12, 13, 1234\}$ and the following constraints

$$\sum_{m\in\mathcal{I}_3} \nu^{13}_{(n,m)} = \sum_{m\in\mathcal{I}_2} \nu^{12}_{(n,m)} = \mu^1_n, \quad n \in \mathcal{I}_1,$$
$$\sum_{m\in\mathcal{I}_4} \nu^{1234}_{(\underline{n},m)} = \mu^{123}_{\underline{n}}, \quad \underline{n} \in \mathcal{I}_{123}.$$

In general, essential graphs are chain graphs without flags and with components equal to decomposable graphs. In the example above there is one component, the three-chain 2-1-3. From the moments and constraints we see that the distribution of $(\mathbf{p}^{123}_{\underline{m}}, \underline{m} \in \mathcal{I}_{123})$ is hyper-Dirichlet and the conditional distribution of $(\mathbf{p}^{4|123}_{m|\underline{n}}, m \in \mathcal{I}_4)$ for any $\underline{n} \in \mathcal{I}_{123}$ is another independent Dirichlet. This pattern may extend to arbitrary essential graphs. Since like in the standard hyper Dirichlet, we have used all the DAGs in the equivalence class to derive the distribution of $\mathbf{p}$ we would call this new distribution the G^* hyper Dirichlet. If instead of considering all DAGs in the Markov equivalence class we only use a subset $\mathcal{P}$ of them, e.g. only (b) and (c) in our example, then we would obtain a distribution analogous to the $\mathcal{P}$-Dirichlet but for arbitrary DAGs from the given equivalence class.

As a second example of the family $\mathcal{P}$ being the family of DAGs represented by an essential graph, let us consider the essential graph in Figure 3 (a). The two graphs (b) and (c) in Figure 3 make up the class of Markov equivalent graphs.

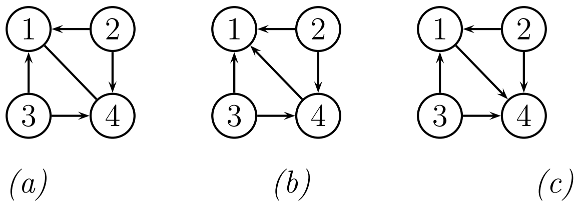


Figure 3: The essential graph (a) and its equivalence class {(b), (c)}

For this family of DAGs, we have

$$\mathfrak{P} = \{\emptyset, \emptyset, 23\}, \quad \mathfrak{Q} = \{2, 3, 1234\}.$$

Our usual moment argument yields the moment formula $(4.3)_{mf}$ with constraints

$$\begin{aligned} \sum_{m \in \mathcal{I}_2} \nu_m^2 &= \mu^{\emptyset,2} \\ \sum_{m \in \mathcal{I}_3} \nu_m^3 &= \mu^{\emptyset,3} \\ \sum_{\underline{m} \in \mathcal{I}_{14}} \nu_{(\underline{m},\underline{n})}^{1234} &= \mu_{\underline{n}}^{23}, \quad \underline{n} \in \mathcal{I}_{23} . \end{aligned}$$

We see that $\mathbf{p}^2$ follows a Dirichlet and so does $\mathbf{p}^3$ and they are independent. The conditional distribution of $\mathbf{p}^{14}$ given $\mathbf{p}^{23}$ is a hyper Dirichlet (on the complete induced subgraph G_{14} and therefore a Dirichlet).

6 Characterization by local and global independence

6.1 The $\mathcal{P}$-Dirichlet and the hyper Dirichlet

Proof of Theorem 6.1. In this proof, to avoid double layers of brackets, we write $r^{\mathfrak{q}_v}_{\underline{k},l}$ rather than $r^{\mathfrak{q}_v}_{(\underline{k},l)}$ as we did in the main file and so far in this file.

Note that according to Def. 3.2 it suffices to show that the formula for moments as given in $(4.1)_{mf}$ and $(4.2)_{mf}$ holds for any $\mathfrak{p} \in \mathcal{P}$.

From $(6.1)_{mf}$ it follows that for any DAG from $\mathcal{P}$ with parent function $\mathfrak{p}$ and any table of non-negative integers $\underline{r} = (r(\underline{i}), \underline{i} \in \mathcal{I})$

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_v}} f^{\mathfrak{p}_v}_{\underline{k}} \left(r^{\mathfrak{q}_v}_{\underline{k},l},\, l \in \mathcal{I}_v \right), \tag{6.1}$$

where for any $v \in V$ and any $\underline{k} \in \mathcal{I}_{\mathfrak{p}_v}$

$$f^{\mathfrak{p}_v}_{\underline{k}}(z_l,\, l \in \mathcal{I}_v) = \mathbb{E} \prod_{l \in \mathcal{I}_v} \left[\mathbf{p}^{v|p_v}_{l|\underline{k}} \right]^{z_l}, \qquad z_l \in \{0, 1, \ldots\},\ l \in \mathcal{I}_v.$$

In order to prove Theorem 6.1 we identify the functions $f^{\mathfrak{p}_v}_{\underline{k}}$, $\underline{k} \in \mathcal{I}_{\mathfrak{p}_v}$, $v \in V$. Our method relies on the identification of the general form of the functions $f^{\mathfrak{p}_v}_{\underline{k}}$, which will appear to be a ratio of products of gamma functions as in the formula for the moments of the $\mathcal{P}$-Dirichlet distribution. Our main tool is equation (8.2) of Lemma 8.1. The proof is divided into two parts, **(a)** and **(b)**. In part **(a)** we transform the moment equation (6.1) into the seemingly cumbersome but useful (6.5). In part **(b)** through a judicious choice of sparse $\underline{r}$'s in (6.5) we will obtain the general form of $f^{\mathfrak{p}_v}_{\underline{k}}$'s using Lemma 8.1.

(a) We first aim for the simplified functional equation (6.5). Fix an arbitrary $\underline{\tau} = (\tau_v \in \mathcal{I}_v,\, v \in V)$ and consider a d-way table $\underline{\epsilon} = (\epsilon_{\underline{i}},\, \underline{i} \in \mathcal{I})$ such that

$$\epsilon_{\underline{i}} = \begin{cases} 1, & \text{if } \underline{i} = \underline{\tau}, \\ 0, & \text{otherwise.} \end{cases}$$

Changing $\underline{r}$ into $\underline{r} + \underline{\epsilon}$ in (6.1) we get

$$\begin{aligned} \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}} + \epsilon_{\underline{i}}} \quad &= \quad \prod_{v \in V} [f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}} \left(r^{\mathfrak{q}_v}_{\tau_{\mathfrak{q}_v}} + 1,\ r^{\mathfrak{q}_v}_{\tau_{\mathfrak{p}_v},l},\ l \in \mathcal{I}_v \setminus \{\tau_v\} \right) \\ &\quad \times \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_v} \setminus \{\tau_{\mathfrak{p}_v}\}} f^{\mathfrak{p}_v}_{\underline{k}} \left(r^{\mathfrak{q}_v}_{\underline{k},l},\ l \in \mathcal{I}_v \right)] \end{aligned} \tag{6.2}$$

We will now obtain an equation of the type (8.2) by equating the right-hand side of (6.2) for different $\mathfrak{p}$'s from $\mathcal{P}$. Fix a DAG in $\mathcal{P}$, that is a $\mathfrak{p} \in \mathcal{P}$, and fix a vertex $v \in V$. Then, by separation property $(6.2)_{mf}$ there exists another DAG in $\mathcal{P}$ with parent function $\mathfrak{p}'$ such that $\mathfrak{p}'_v \neq \mathfrak{p}_v$. For each of $\mathfrak{p}$ and $\mathfrak{p}'$ the right-hand side of (6.2) is split into three parts: the first (first line below) concerns v, the second (second line) $\mathfrak{c}_v$ and the third (third line) the remainder of V. Thus we obtain

$$
\begin{aligned}
& f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}}\left(r^{\mathfrak{q}_v}_{\tau_{\mathfrak{q}_v}}+1,\ r^{\mathfrak{q}_v}_{\tau_{\mathfrak{p}_v},l},\ l \in \mathcal{I}_v \setminus \{\tau_v\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_v} \setminus \{\tau_{\mathfrak{p}_v}\}} f^{\mathfrak{p}_v}_{\underline{k}}\left(r^{\mathfrak{q}_v}_{\underline{k},l},\ l \in \mathcal{I}_v\right) \\
& \prod_{w \in \mathfrak{c}_v} f^{\mathfrak{p}_w}_{\tau_{\mathfrak{p}_w}}\left(r^{\mathfrak{q}_w}_{\tau_{\mathfrak{q}_w}}+1,\ r^{\mathfrak{q}_w}_{\tau_{\mathfrak{p}_w},l},\ l \in \mathcal{I}_w \setminus \{\tau_w\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_w} \setminus \{\tau_{\mathfrak{p}_w}\}} f^{\mathfrak{p}_w}_{\underline{k}}\left(r^{\mathfrak{q}_w}_{\underline{k},l},\ l \in \mathcal{I}_w\right) \\
& \prod_{w \notin \mathfrak{c}_v \cup \{v\}} f^{\mathfrak{p}_w}_{\tau_{\mathfrak{p}_w}}\left(r^{\mathfrak{q}_w}_{\tau_{\mathfrak{q}_w}}+1,\ r^{\mathfrak{q}_w}_{\tau_{\mathfrak{p}_w},l},\ l \in \mathcal{I}_w \setminus \{\tau_w\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_w} \setminus \{\tau_{\mathfrak{p}_w}\}} f^{\mathfrak{p}_w}_{\underline{k}}\left(r^{\mathfrak{q}_w}_{\underline{k},l},\ l \in \mathcal{I}_w\right) \\
& = f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}\left(r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{q}'_v}}+1,\ r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},l},\ l \in \mathcal{I}_v \setminus \{\tau_v\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}'_v} \setminus \{\tau_{\mathfrak{p}'_v}\}} f^{\mathfrak{p}'_v}_{\underline{k}}\left(r^{\mathfrak{q}'_v}_{\underline{k},l},\ l \in \mathcal{I}_v\right) \qquad (6.3) \\
& \prod_{w \in \mathfrak{c}'_v} f^{\mathfrak{p}'_w}_{\tau_{\mathfrak{p}'_w}}\left(r^{\mathfrak{q}'_w}_{\tau_{\mathfrak{q}'_w}}+1,\ r^{\mathfrak{q}'_w}_{\tau_{\mathfrak{p}'_w},l},\ l \in \mathcal{I}_w \setminus \{\tau_w\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}'_w} \setminus \{\tau_{\mathfrak{p}'_w}\}} f^{\mathfrak{p}'_w}_{\underline{k}}\left(r^{\mathfrak{q}'_w}_{\underline{k},l},\ l \in \mathcal{I}_w\right) \\
& \prod_{w \notin \mathfrak{c}'_v \cup \{v\}} f^{\mathfrak{p}'_w}_{\tau_{\mathfrak{p}'_w}}\left(r^{\mathfrak{q}'_w}_{\tau_{\mathfrak{q}'_w}}+1,\ r^{\mathfrak{q}'_w}_{\tau_{\mathfrak{p}'_w},l},\ l \in \mathcal{I}_w \setminus \{\tau_w\}\right) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}'_w} \setminus \{\tau_{\mathfrak{p}'_w}\}} f^{\mathfrak{p}'_w}_{\underline{k}}\left(r^{\mathfrak{q}'_w}_{\underline{k},l},\ l \in \mathcal{I}_w\right)
\end{aligned}
$$

We now write this equation above for two distinct values first for $\tau_v = \rho$ and then for $\tau_v = \sigma$ in $\mathcal{I}_v$, while keeping τ_k the same for all $k \neq v$. We obtain two equations, say E_ρ and E_σ and we then write the identity

$$\frac{\text{lhs}(E_\rho)}{\text{lhs}(E_\sigma)} = \frac{\text{rhs}(E_\rho)}{\text{rhs}(E_\sigma)}. \qquad (6.4)$$

Many simplifications occur (see part 8.2 of Appendix) and we arrive at

$$\frac{f^{\mathfrak{p}v}_{\tau_{\mathfrak{p}v}}\left(r^{\mathfrak{q}v}_{\tau_{\mathfrak{p}v},\rho}+1,\ r^{\mathfrak{q}v}_{\tau_{\mathfrak{p}v},l},\ l\in\mathcal{I}_v\setminus\{\rho\}\right)}{f^{\mathfrak{p}v}_{\tau_{\mathfrak{p}v}}\left(r^{\mathfrak{q}v}_{\tau_{\mathfrak{p}v},\sigma}+1,\ r^{\mathfrak{q}v}_{\tau_{\mathfrak{p}v},l},\ l\in\mathcal{I}_v\setminus\{\sigma\}\right)}$$

$$\prod_{w\in\mathfrak{c}_v}\frac{f^{\mathfrak{p}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\rho)}\left(r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\rho),\tau_w}+1,\ r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\rho),l},\ l\in\mathcal{I}_w\setminus\{\tau_w\}\right)}{f^{\mathfrak{p}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\sigma)}\left(r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\sigma),\tau_w}+1,\ r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\sigma),l},\ l\in\mathcal{I}_w\setminus\{\tau_w\}\right)}\frac{f^{\mathfrak{p}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\sigma)}\left(r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\sigma),l},\ l\in\mathcal{I}_w\right)}{f^{\mathfrak{p}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\rho)}\left(r^{\mathfrak{q}w}_{(\tau_{\mathfrak{p}w\setminus\{v\}},\rho),l},\ l\in\mathcal{I}_w\}\right)}$$

$$=\frac{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}\left(r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},\rho}+1,\ r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},l},\ l\in\mathcal{I}_v\setminus\{\rho\}\right)}{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}\left(r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},\sigma}+1,\ r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},l},\ l\in\mathcal{I}_v\setminus\{\sigma\}\right)} \tag{6.5}$$

$$\prod_{w\in\mathfrak{c}'_v}\frac{f^{\mathfrak{p}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\rho)}\left(r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\rho),\tau_w}+1,\ r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\rho),l},\ l\in\mathcal{I}_w\setminus\{\tau_w\}\right)}{f^{\mathfrak{p}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\sigma)}\left(r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\sigma),\tau_w}+1,\ r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\sigma),l},\ l\in\mathcal{I}_w\setminus\{\tau_w\}\right)}\frac{f^{\mathfrak{p}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\sigma)}\left(r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\sigma),l},\ l\in\mathcal{I}_w\right)}{f^{\mathfrak{p}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\rho)}\left(r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},\rho),l},\ l\in\mathcal{I}_w\}\right)}$$

(b) We now simplify (6.5) further by writing it for properly chosen sparse $\underline{r}$'s. This will lead us to functional equations for functions defined on $\mathcal{I}_v$.

We define

$$\mathfrak{d}_v=\mathfrak{p}'_v\cap\mathfrak{c}_v \qquad \text{and} \qquad \mathfrak{d}'_v=\mathfrak{p}_v\cap\mathfrak{c}'_v.$$

Note that due to the separation property $(6.2)_{mf}$ at least one of them is not empty. Without loss of generality let us assume that $\mathfrak{d}_v\neq\emptyset$. Fix $\xi_{\mathfrak{d}_v}\in\mathcal{I}_{\mathfrak{d}_v}$ such that $\xi_i\neq\tau_i$ for any $i\in\mathfrak{d}_v$. For any $l\in\mathcal{I}_v$ denote by $\underline{i}(l)$ the cell with labels

$$i_v=l,\ i_{\mathfrak{d}_v}=\xi_{\mathfrak{d}_v},\ i_y=\tau_y \text{ for } y\notin\mathfrak{d}_v\cup\{v\}.$$

Define

$$x_l=r_{\underline{i}(l)} \qquad l\in\mathcal{I}_v.$$

Consider any $\underline{r}=(r_{\underline{i}})$ such that $r_{\underline{i}}=0$ for all $\underline{i}\notin\{\underline{i}(l),\ l\in\mathcal{I}_v\}$. Since $\mathfrak{p}_v\cap\mathfrak{d}_v=\emptyset$, by $(6.2)_{mf}$

$$r^{\mathfrak{q}_v}_{\tau_{\mathfrak{p}_v},l}=x_l, \qquad l\in\mathcal{I}_v. \tag{6.6}$$

Again, by $(6.2)_{mf}$ and since $\mathfrak{p}'_v\supset\mathfrak{d}_v\neq\emptyset$ we have

$$r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},l}=0, \qquad l\in\mathcal{I}_v. \tag{6.7}$$

Moreover, for $l\in\mathcal{I}_w$ and $k\in\mathcal{I}_v$ (particularly for $k=\rho$ or $k=\sigma$, which we shall use here)

$$r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w\setminus\{v\}},k),l}=\begin{cases} x_k, & \text{if } \mathfrak{p}_w\cap\mathfrak{d}_v=\emptyset \text{ and either } (w\notin\mathfrak{d}_v, \text{ and } l=\tau_w) \text{ or } (w\in\mathfrak{d}_v \text{ and } l=\xi_w),\\ 0, & \text{otherwise}\end{cases}$$

and

$$r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},k),l} = \begin{cases} x_k, & \text{if } \mathfrak{p}'_w \cap \mathfrak{d}_v = \emptyset \text{ and either } (w \notin \mathfrak{d}_v, \text{ and } l = \tau_w) \text{ or } (w \in \mathfrak{d}_v \text{ and } l = \xi_w), \\ 0, & \text{otherwise.} \end{cases}$$

These last two observations imply that the products $\prod_{w\in\mathfrak{c}_v}$ and $\prod_{w\in\mathfrak{c}'_v}$ on lines 2 and 4 of equation (6.5) factor into a function of x_ρ and a function of x_σ. Therefore their quotient can be written as $a_{v,\rho}(x_\rho)/a_{v,\sigma}(x_\sigma)$. Note, that potentially these functions may depend of $\mathfrak{p}$ and $\mathfrak{p}'$, but it will not impact our final result.

Moreover, by (6.6) and (6.7) it follows that (6.5) assumes the form

$$\frac{f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}}(x_\rho+1,\ x_l,\ l\in\mathcal{I}_v\setminus\{\rho\})}{f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}}(x_\sigma+1,\ x_l,\ l\in\mathcal{I}_v\setminus\{\sigma\})} = K_v\,\frac{a_{v,\rho}(x_\rho)}{a_{v,\sigma}(x_\sigma)},$$

where

$$K_v = \frac{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(1_\rho,\ 0_l,\ l\in\mathcal{I}_v\setminus\{\rho\})}{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(1_\sigma,\ 0_l,\ l\in\mathcal{I}_v\setminus\{\sigma\})}.$$

Since $\tau_{\mathfrak{p}_v}$ was arbitrary in $\mathcal{I}_{\mathfrak{p}_v}$ we conclude from Lemma 8.1 that for any $\underline{k}\in\mathcal{I}_{\mathfrak{p}_v}$ either

$$f^{\mathfrak{p}_v}_{\underline{k}}(z_l,\ l\in\mathcal{I}_v) = \frac{\prod_{l\in\mathcal{I}_v}\left(A^{v|\mathfrak{p}_v}_{l|\underline{k}}\right)^{z_l}}{\left(\tilde{A}^{\mathfrak{p}_v}_{\underline{k}}\right)^{|\underline{z}|}}, \tag{6.8}$$

where $|\underline{z}| = \sum_{l\in\mathcal{I}_v} z_l$ and $\tilde{A}^{\mathfrak{p}_v}_{\underline{k}} = \sum_{l\in\mathcal{I}_v} A^{v|\mathfrak{p}_v}_{l,\underline{k}}$ or it is a product of univariate power functions

$$f^{\mathfrak{p}_v}_{\underline{k}}(z_l,\ l\in\mathcal{I}_v) = \prod_{l\in\mathcal{I}_v}\left[A^{v|\mathfrak{p}_v}_{l|\underline{k}}\right]^{z_l}. \tag{6.9}$$

However the latter case is impossible due to the parameter independence assumption which requires that the distribution of the random vector $(\mathbb{P}_{\mathbf{p}}(X_v = l|\mathbf{X}_{\mathfrak{p}_v}=\underline{k}),\ l\in\mathcal{I}_v)$ is non-degenerate.

We now want to identify the functions $f^{\mathfrak{p}'_v}_{\underline{k}}$, $\underline{k}\in\mathcal{I}_{\mathfrak{p}'_v}$. If $\mathfrak{d}'_v\neq\emptyset$ we can repeat the argument used to derive $f^{\mathfrak{p}_v}_{\underline{k}}$ and obtain an analogue of (6.8) with $\mathfrak{p}$ replaced by $\mathfrak{p}'$. If $\mathfrak{d}'_v=\emptyset$ we need to do some more work. We will use another sparse $\underline{r}$ with new $\underline{i}(l)$'s defined by substituting $\mathfrak{d}'_v$ for $\mathfrak{d}_v$. Note that under this new sparsity pattern for any $k\in\mathcal{I}_v$ (particularly for $k=\rho$ or $k=\sigma$, which we shall use here)

$$r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w\setminus\{v\}},k),l} = r^{\mathfrak{q}'_w}_{(\tau_{\mathfrak{p}'_w\setminus\{v\}},k),l} = \begin{cases} x_k & \text{if } l=\tau_w, \\ 0 & \text{if } l\neq\tau_w, \end{cases}$$

and

$$r^{\mathfrak{q}_v}_{\tau_{\mathfrak{p}_v},l} = r^{\mathfrak{q}'_v}_{\tau_{\mathfrak{p}'_v},l} = x_l, \qquad l \in \mathcal{I}_v.$$

Thus, (6.5) becomes

$$\frac{f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}}(x_\rho + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\rho\})}{f^{\mathfrak{p}_v}_{\tau_{\mathfrak{p}_v}}(x_\sigma + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\sigma\})} = \frac{a_{v,\rho}(x_\rho)}{a_{v,\sigma}(x_\sigma)} \frac{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(x_\rho + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\rho\})}{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(x_\sigma + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\sigma\})}.$$

Plugging (6.8) into the left hand side above we obtain

$$\frac{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(x_\rho + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\rho\})}{f^{\mathfrak{p}'_v}_{\tau_{\mathfrak{p}'_v}}(x_\sigma + 1,\ x_l,\ l \in \mathcal{I}_v \setminus \{\sigma\})} = \frac{A^{v|\mathfrak{p}_v}_{\rho|\tau_{\mathfrak{p}_v}} + x_\rho}{A^{v|\mathfrak{p}_v}_{\sigma|\tau_{\mathfrak{p}_v}} + x_\sigma} \frac{a'_{v,\sigma}(x_\sigma)}{a'_{v,\rho}(x_\rho)} \quad \text{or} \quad \frac{A^{v|\mathfrak{p}_v}_{\rho|\tau_{\mathfrak{p}_v}}}{A^{v|\mathfrak{p}_v}_{\sigma|\tau_{\mathfrak{p}_v}}} \frac{a'_{v,\sigma}(x_\sigma)}{a'_{v,\rho}(x_\rho)},$$

respectively. Again we use Lemma 8.1 in the Appendix below to conclude that one of the representations (6.8) or (6.9) (with $\mathfrak{p}$ changed into $\mathfrak{p}'$) holds also for $f^{\mathfrak{p}'_v}_{\underline{k}}$ for any $\underline{k} \in \mathcal{I}_{\mathfrak{p}'_v}$. Similarly, as above, we conclude that non-degeneracy implies that the representation given in (6.9) is not a valid one.

Given an arbitrary $v \in V$, so far, we have derived the expression of $f^{\mathfrak{p}_v}_{\underline{k}}$ and $f^{\mathfrak{p}'_v}_{\underline{k}}$ in (6.8) for an arbitrary separating pair $\mathfrak{p}, \mathfrak{p}' \in \mathcal{P}$. Clearly, (6.8) is valid for any $\mathfrak{p} \in \mathcal{P}$ and any $v \in V$. Indeed, given $v \in V$ and the separating pair $\mathfrak{p}$, $\mathfrak{p}'$, consider another $\mathfrak{p}''$. Then, either $\mathfrak{p}''(v) \neq \mathfrak{p}(v)$ and then we consider the pair $\mathfrak{p}$, $\mathfrak{p}''$ or $\mathfrak{p}''(v) \neq \mathfrak{p}'(v)$ and then we consider the pair $\mathfrak{p}'$, $\mathfrak{p}''$.

Now, returning to (6.1) we see that for any d-way table $\underline{r} = (r_{\underline{i}},\ \underline{i} \in \mathcal{I})$ of non-negative integers, any $\mathfrak{p} \in \mathcal{P}$, any $v \in V$ and any $(\underline{k}, l) \in \mathcal{I}_{\mathfrak{q}_v}$ there exist numbers $A^{v|\mathfrak{p}_v}_{l|\underline{k}}$ such that

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in I_{\mathfrak{p}_v}} \frac{\prod_{l \in \mathcal{I}_v} \left(A^{v|\mathfrak{p}_v}_{l|\underline{k}}\right)^{r^{\mathfrak{p}_v,v}_{\underline{k},l}}}{\left(\tilde{A}^{\mathfrak{p}_v}_{\underline{k}}\right)^{r^{\mathfrak{p}_v}_{\underline{k}}}}, \tag{6.10}$$

where

$$\tilde{A}^{\mathfrak{p}_v}_{\underline{k}} = \sum_{l \in \mathcal{I}_v} A^{v|\mathfrak{p}_v}_{l|\underline{k}}. \tag{6.11}$$

□

8 Appendix

8.1 An auxiliary result on a functional equation

Lemma 8.1. *Let F be a positive function defined on the n-th cartesian product of the non-negative integers such that $F(\underline{0}) = 1$ and*

$$F(\underline{x}) = \sum_{i=1}^{n} F(\underline{x} + \underline{\epsilon}_i), \tag{8.1}$$

where $\underline{\epsilon}_i$ has all components equal to 0 except for the i-th component which is 1. Assume that for any distinct $p, q \in \{1, \ldots, n\}$

$$\frac{F(\underline{x} + \underline{\epsilon}_p)}{F(\underline{x} + \underline{\epsilon}_q)} = \frac{h_p(x_p)}{h_q(x_q)} \qquad \forall\, \underline{x} = (x_1, \ldots, x_n) \in \{0, 1, \ldots\}^n \tag{8.2}$$

for some functions h_i, $i = 1, 2, \ldots, n$.

Then there exists a vector $\underline{A} = (A_1, \ldots, A_n) \in \mathbb{R}^n$ such that $\forall\, \underline{x} = (x_1, \ldots, x_n) \in \{0, 1, \ldots\}^n$ either

$$F(\underline{x}) = \frac{\prod_{i=1}^{n} (A_i)^{x_i}}{(|\underline{A}|)^{|\underline{x}|}},$$

where $|\underline{u}| = u_1 + \ldots + u_n$ for any vector $\underline{u} = (u_1, \ldots, u_n)$ or

$$F(\underline{x}) = \prod_{i=1}^{n} x_i^{A_i}.$$

Lemma 8.1, as given above, is a special version of Lemma 3.1 from [18] (it simply suffices to take $A = \{1, \ldots, n\}$ in this lemma). This lemma is also closely related to the argument used in the proof of Theorem 2 in [3].

8.2 Proof of (6.5)

1. Let $A_1(\rho)$ and $A_1(\sigma)$ be the values of $A_1 = \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_v} \setminus \{\tau_v\}} f_{\underline{k}}^{\mathfrak{p}_v}(r_{\underline{k},l}^{\mathfrak{q}_v},\, l \in \mathcal{I}_v)$, the second factor in the first line of the left hand side of (6.3), for $\tau_v = \rho$ and

$\tau_v = \sigma$, respectively. Let $A'_1(\rho)$ and $A'_1(\sigma)$ be the analog quantities for the right-hand side of (6.3). Clearly

$$A_1(\rho) = \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_v} \setminus \{\tau_{\mathfrak{p}_v}\}} f_{\underline{k}}^{\mathfrak{p}_v}(r^{\mathfrak{q}_v}_{\underline{k},\rho},\, r^{\mathfrak{q}_v}_{\underline{k},\sigma},\, r^{\mathfrak{q}_v}_{\underline{k},l},\, l \in \mathcal{I}_v \setminus \{\rho,\sigma\}) = A_1(\sigma).$$

Similarly $A'_1(\rho) = A'_1(\sigma)$.

2. Let $A_2(\rho)$ and $A_2(\sigma)$ be the values of the factor for fixed $w \in \mathfrak{c}_v$ in the second line of the left hand side of (6.3), for $\tau_v = \rho$ and $\tau_v = \sigma$, respectively. Let $A'_2(\rho)$ and $A'_2(\sigma)$ be the analog quantities for the right-hand side of (6.3). Clearly

$$\begin{aligned} A_2(\rho) \;=\;& f^{\mathfrak{p}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\rho)}(r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\rho),\tau_w} + 1,\, r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\rho),l}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \\ & f^{\mathfrak{p}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma)}(r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma),l},\, l \in \mathcal{I}_w) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_w} \setminus \{\tau_{\mathfrak{p}_w}, (\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma)\}} f^{\mathfrak{p}_w}_{\underline{k}}(r^{\mathfrak{q}_w}_{\underline{k},l},\, l \in \mathcal{I}_w) \end{aligned}$$

and

$$\begin{aligned} A_2(\sigma) \;=\;& f^{\mathfrak{p}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma)}(r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma),\tau_w} + 1,\, r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\sigma),l}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \\ & f^{\mathfrak{p}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\rho)}(r^{\mathfrak{q}_w}_{(\tau_{\mathfrak{p}_w \setminus \{v\}},\rho),l},\, l \in \mathcal{I}_w) \prod_{\underline{k} \in \mathcal{I}_{\mathfrak{p}_w} \setminus \{\tau_{\mathfrak{p}_w}, (\tau_{\mathfrak{p}_w \setminus \{v\}},\rho)\}} f^{\mathfrak{p}_w}_{\underline{k}}(r^{\mathfrak{q}_w}_{\underline{k},l},\, l \in \mathcal{I}_w). \end{aligned}$$

Note that the two sets appearing in the indices: $\{\tau_{\mathfrak{p}_w}, (\tau_{\mathfrak{p}_w \setminus \{v\}}, \sigma)\}$ in $A_2(\rho)$ and $\{\tau_{\mathfrak{p}_w}, (\tau_{\mathfrak{p}_w \setminus \{v\}}, \rho)\}$ in $A_2(\sigma)$ are identical. Therefore, for each $w \in \mathfrak{c}_v$ the ratio $\frac{A_2(\rho)}{A_2(\sigma)}$ is equal to the factor in the second line of the left hand side of (6.5). Similarly, for each $w \in \mathfrak{c}_v$ the ratio $\frac{A'_2(\rho)}{A'_2(\sigma)}$ is equal to the factor in the second line of the right-hand side of (6.5).

3. Since v appears in the third line of neither the left hand side nor the right-hand side of (6.3), these lines cancel out in the ratios of (6.4).