

# グラフィカルモデルへの imset を用いた代数的アプローチ

数理情報学専攻 48-096209 柏村拓哉

指導教員 竹村彰通 教授

## 1 導入

条件付き独立性は、確率変数間の相関関係や絡みなどの質的関係を与えるものであり、統計学や人工知能の分野で非常によく用いられている概念である。

条件付き独立性の構造は、伝統的にグラフや特別なリストで表せられるが、これらでは完全に構造を表すことはできない。そこで、imset と呼ばれる整数ベクトルを用いた手法が、M. Studený [4] により提案された。この手法を用いることにより、離散確率測度による条件付き独立性の構造を完全に記述することができる。

この imset を用いた手法では、複数の imset が同じ条件付き独立性の構造を記述することがある。したがって、この imset の同値類の中で適切な代表元を求めることが問題となる。Studený は、その問題の一つとして、古典的なグラフィカルモデルに対する imset を求めることを考えた。特に、非巡回的有向グラフや分解可能グラフに対し、standard imset と呼ばれる imset を用いた表現法を与えている [4]。連鎖グラフに対しても、特別な場合においては imset による表現法を導いている [5]。

本論文では、これらの先行研究の結果を拡張し、無向グラフおよび一般の連鎖グラフに対する imset による表現法を導出した。ここで得られた imset は、先行研究の一般化となっている。

## 2 Imset

有限な変数集合を  $N$  とする。条件付き独立性を考える際には、通常 3 つの互いに排反な部分集合  $A, B, C \subseteq N$  を考える。ここでは、これらの 3 つ組を  $\langle A, B | C \rangle$  と書く。この 3 つ組すべてからなる集合を  $\mathcal{T}(N)$  で表す。

Imset は  $\mathbb{Z}^{2^N}$  上のベクトルである。集合  $A \subseteq N$  に対し、

$$\delta_A(B) = \begin{cases} 1 & (B = A), \\ 0 & (B \neq A, B \subseteq N) \end{cases}$$

を basic imset と呼ぶ。この basic imset は  $\mathbb{R}^{2^N}$  の標準基底をなす。また、特別な imset として、

$$u_{\langle A, B | C \rangle} = \delta_{ABC} + \delta_C - \delta_{AC} - \delta_{BC}, \langle A, B | C \rangle \in \mathcal{T}(N)$$

を semi-elementary imset という。これは、離散測度  $P$

における条件付き独立性  $A \perp\!\!\!\perp B | C [P]$  の定義

$$P^{ABC}(x_A, x_B, x_C) \cdot P^C(x_C) = P^{AC}(x_A, x_C) \cdot P^{BC}(x_B, x_C)$$

において、両辺の対数を取ったものに対応している。特に、 $A = \{a\}, B = \{b\}$  のとき、 $u_{\langle \{a\}, \{b\} | C \rangle}$  を elementary imset と呼ぶ。さらに、elementary imset の張る凸錐内の格子点を structural imset と呼び、その集合を  $\mathcal{S}(N)$  で表す。定義より、structural imset は、elementary imset の非負有理数結合で表せることがわかる。

Structural imset  $u \in \mathcal{S}(N)$  と 3 つ組  $\langle A, B | C \rangle \in \mathcal{T}(N)$  に対して、 $A \perp\!\!\!\perp B | C [u]$  を、 $\exists k \in \mathbb{N}, k \cdot u - u_{\langle A, B | C \rangle} \in \mathcal{S}(N)$  が成り立つときと定義する。また、 $\mathcal{M}_u = \{\langle A, B | C \rangle \in \mathcal{T}(N) : A \perp\!\!\!\perp B | C [u]\}$  とする。任意の離散確率測度  $P$  に対して、ある structural imset  $u \in \mathcal{S}(N)$  が存在し、

$$A \perp\!\!\!\perp B | C [u] \iff A \perp\!\!\!\perp B | C [P], \quad \forall \langle A, B | C \rangle \in \mathcal{T}(N)$$

が成り立つことが知られている [4]。

## 3 分解可能グラフに対する imset (先行研究)

分解可能グラフに対する imset は Studený [4] により導出されている。

単純な無向グラフ  $G$  に対し、 $C \subset N$  を取り除いたグラフ  $G_{N \setminus C}$  において、排反な部分集合  $A, B \subseteq N \setminus C$  が連結でないとき、 $A \perp\!\!\!\perp B | C [G]$  と表す。また、 $\mathcal{M}_G = \{\langle A, B | C \rangle \in \mathcal{T}(N) : A \perp\!\!\!\perp B | C [G]\}$  とする。 $G$  の極大クリークの集合を  $\mathcal{K}(G)$  としたとき、あるこの要素の順序  $K_1, \dots, K_m$  が存在し、RIP

$$2 \leq \forall i \leq m, 1 \leq \exists k < i, S_i \equiv K_i \cap \bigcup_{j < i} K_j \subseteq K_k$$

を満たすとき、 $G$  は分解可能であるという [2]。また、 $\mathcal{S}(G) = \{S_2, \dots, S_m\}$  とする。分解可能グラフ  $G$  の standard imset は、次のように定義される [4]:

$$u_G = \delta_N - \sum_{K \in \mathcal{K}(G)} \delta_K + \sum_{S \in \mathcal{S}(G)} \nu_G(S) \cdot \delta_S. \quad (1)$$

この imset は、グラフ  $G$  が規定する条件付き独立性を完全に表している。つまり、 $\mathcal{M}_{u_G} = \mathcal{M}_G$  が成り立つ。

## 4 無向グラフに対する imset の導出

前節の分解可能の場合の結果を利用するために、グラフに辺を付け加えて分解可能にすることを考える。この操作でできたグラフを三角化 (triangulation [1]) と呼ぶ。このうち、辺の包含関係に対し極小なものを極小三角化と呼ぶ。

一般の無向グラフ  $G$  に対しては、すべての極小三角化  $MT(G)$  を考えることにより、この無向グラフを表わす imset を構成することができる。

**定理 4.1.** 無向グラフ  $G$  に対し、 $v_G = \sum_{G' \in MT(G)} u_{G'}$  (ただし、 $u_{G'}$  は式 (1) で与えられる) とする。このとき、 $M_G = M_{v_G}$  が成り立つ。

誘導部分グラフ  $G_V$ 、 $V \subset N$  がクリークセパレータを持たない部分グラフの中で極大であるとき、これを mp-subgraph であるという [3]。グラフの極小三角化は mp-subgraph と呼ばれる誘導部分グラフに分解して、各グラフごとで極小三角化を考えれば十分であることがわかる。よって、無向グラフに対する imset を次で定義する：

$$u_G = \delta_N - \sum_{V \in \mathcal{V}(G)} \delta_V + \sum_{S \in \mathcal{S}(G)} \nu_G(S) \cdot \delta_S + \sum_{V \in \mathcal{V}(G)} \sum_{H \in MT(G_V)} u_H. \quad (2)$$

ただし、 $\mathcal{V}(G)$  を  $G$  の mp-subgraph の頂点集合の族であり、各  $H \in MT(G_V)$  ( $V \in \mathcal{V}(G)$ ) に対し、 $u_H$  は式 (1) で定義される standard imset である。この imset も  $G$  の表す条件付き独立性を完全に記述している。

**定理 4.2.** 無向グラフ  $G$  に対し、 $u_G$  を式 (2) で定義する。このとき、 $M_{u_G} = M_G$  が成り立つ。

## 5 連鎖グラフに対する imset の導出

連鎖グラフは、有向辺と無向辺を持ち、有向サイクルを持たないグラフとして定義される。これは、非巡回的有向グラフや無向グラフの一般化となっている。連鎖グラフ  $G$  には、moralization と呼ばれる分離基準があり [2]、ここから規定される条件付き独立性の集合を  $M_G$  で表す。連鎖グラフ  $G$  に対しては、その standard imset を

$$u_G = \delta_N - \delta_\emptyset + \sum_{C \in \mathcal{C}(G)} \left\{ \delta_{\text{pa}_G(C)} - \delta_{\text{CUpa}_G(C)} + u_{\bar{G}(C)} \right\} \quad (3)$$

で定義する。ただし、 $\mathcal{C}(G)$  は、グラフ  $G$  の連結成分の集合であり、任意の  $C \in \mathcal{C}(G)$  について、 $u_{\bar{G}(C)}$  は閉包グラフと呼ばれる無向グラフ  $\bar{G}(C)$  に対する standard imset (式 (2)) である。連鎖グラフのうち、ある非巡回的有向グラフと同じ条件付き独立性を規定する場合には、式 (3) で  $u_{\bar{G}(C)}$  を分解可能グラフに対する imset (式 (1)) としたもので与えられている [5]。よって、式 (3) は先行研究の一般化となっている。この imset は、連鎖グラフに対する imset となっている。

**定理 5.1.** 連鎖グラフ  $G$  に対し、 $u_G$  を式 (3) で定義する。このとき、 $M_{u_G} = M_G$  が成り立つ。

これは、無向グラフと同様の議論を行うことにより得られる。そのために、本論文では連鎖グラフに対し、三角化を一般化した概念を定義した。

## 6 まとめと今後の課題

本論文では、無向グラフと連鎖グラフに対して imset の導出を行った。この過程で重要な概念は極小三角化である。また他の成果として、同じ条件付き独立性を表す連鎖グラフに対し、与えた imset が一意に定まることを示した。これは、[5] の結果の一般化となっている。

今後の課題としては、imset を elementary imset の和で表したとき、その足されている elementary imset の個数 (次数) が少ないものを求めることが考えられる。これは、[4] で代表元の基準として考えられているものの一つである。また、本論文では、無向グラフに対する imset を、性質の良いグラフ (分解可能グラフ) の場合に帰着することにより求めた。よって、一般の条件付き独立モデルについて、何らかの意味で三角化の概念を一般化し、グラフィカルモデルの場合に帰着できれば、これらの imset を求めることができると考えられる。

## 参考文献

- [1] P. Heggenes: Minimal triangulations of graphs: A survey. *Discrete Mathematics*, **306**, pp. 297–317, 2006.
- [2] S. L. Lauritzen: *Graphical Models*. Oxford University Press, Oxford, 1996.
- [3] H. G. Leimer: Optimal decomposition by clique separators. *Discrete Mathematics*, **113**, pp. 99–123, 1993.
- [4] M. Studený: *Probabilistic Conditional Independence Structures*. Springer-Verlag, London, 2005.
- [5] M. Studený, A. Roverato, and Š. Štěpánová: Two operations of merging and splitting components in a chain graph. *Kybernetika*, **45**, no. 2, pp. 208–248, 2009.