2001 年 ベイジアンネットチュートリアル
Tutorial on Bayesian Networks (BN2001)
Tokyo, Japan, July 29, 2001.

ベイジアンネットワークにおける確率伝播について

鈴木 譲*
大阪大学大学院理学研究科数学専攻

Abstract: 最初に、ベイジアンネットワークの本来の定義と階層化手順との関係を述べる。次に、単一結合の場合の確率伝播を解説する。最後に、ファクターグラフを用いた場合の確率伝播がオリジナルの方法とどのように対応するのかを述べる。

1 まえがき

ベイジアンネットワークの研究を始めたが、適切な解説がなく、特に確率伝播の動作を理解するのに苦しんでいるという声を聞く。わかりやすく書いてある書物や記事がほとんどないといっても過言ではなく、そのことが、日本でベイジアンネットワークを用いた推論・学習がなかなか広まっていかない理由のひとつになっている。しかしながら、最近はユーザモデルやターボ符号の復号などの応用分野が進展してきて、ベイジアンネットワークの利用のニーズが極めて高くなってきた。

あの黒くて厚い本に悩まされた読者は多いと思う。ベイジアンネットワークの定義などはその本に詳しく解説されているし、かなり勉強しているごく少数の人には感謝されていると思う。ただ、証明が不適切であったり、記法の説明が十分でなかったり、評判がよくないというのも事実である (本稿では、記法に関しては神経質すぎるぐらいに注意を払っている)。

本稿は、オリジナルな結果を含む論文ではない。しかしながら、ベイジアンネットワークを用いた確率伝播と無向ファクターグラフを用いた確率伝播との関係など、解釈の上で著者なりの見方を入れている。また、思い切って枝葉は捨てて、ベイジアンネットワークの定義と単一結合の場合の確率伝播の動作の 2 項目にしぼることにした。例えば、ターボ符号の復号であれば、ベイジアンネットワークのトポロジーが単一結合ではなく、また連続値をとる確率変数を扱わなければならない。したがって、本稿に書かれていることだけでは不十分かもしれない。しかし、本稿で解説している基本事項の理解なくして、その上に積み上げられる応用に関して正確な理解が得られないと思われる。

以下、$\bar{X} - Y$ で 2 集合 X, Y の差集合 $X \cap \bar{Y}$ を表す。

*〒 305-8568 豊中市待兼山町 1-1, tel: 06-6850-5315, e-mail: suzuki@math.sci.osaka-u.ac.jp, URL: http://www.math.sci.osaka-u.ac.jp/ suzuki/suzuki-home.html

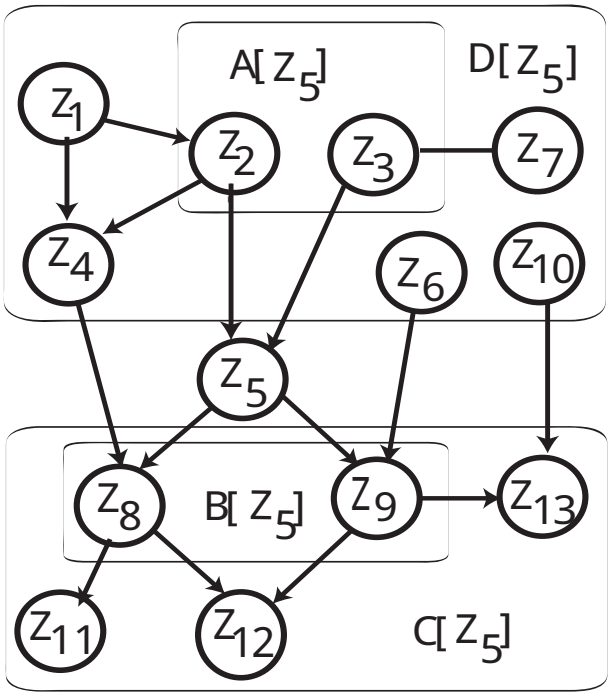

図 1: 親集合 $A[z]$、子集合 $B[z]$、子孫集合 $c[z]$、非子孫集合 $D[z]$

2 ベイジアンネットワーク

2.1 有向非巡回グラフ

定義 **1** V を有限集合、$\vec{E}$ を $\{(v, v') | v, v' \in V, v \neq v'\}$ の部分集合とする。対 $\vec{G} = (V, \vec{E})$ を頂点集合 V と有向辺集合 $\vec{E}$ からなる有向グラフ (directed graph) という。そして、$(v_{i-1}, v_i) \in \vec{E},\ i = 1, \cdots, m$ のとき、$((v_0, v_1), (v_1, v_2), \cdots (v_{m-1}, v_m))$ を v_0 から v_m までの有向経路と呼び、$v_0 \to v_1 \to \cdots \to v_m$ で記す。また、$C[v] := \{v_m \in V | v = v_0 \to v_1 \to \cdots \to v_m\}$ を v の子孫集合とよぶ。

定義 **2** V を有限集合、E を $\{\{v, v'\} | v, v' \in V, v \neq v'\}$

の部分集合とする。対 $G=(V,E)$ を頂点集合 V と無向辺集合 E からなる無向グラフ (undirected graph) という。特に、有向グラフ $\vec{G}=(V,\vec{E})$ について、$E=\{\{v,v'\}|(v,v')\in\vec{E}\}$ を $\vec{E}$ の無向辺集合、$G=(V,E)$ を $\vec{G}$ の無向グラフという。$\{v_{i-1},v_i\}\in E, i=1,\cdots,m$ のとき、$(\{v_0,v_1\},\{v_1,v_2\},\cdots\{v_{m-1},v_m\})$ を v_0 から v_m までの無向経路と呼び、$v_0-v_1-\cdots-v_m$ で記す。また、$S[v]:=\{v_m\in V|v=v_0-v_1-\cdots-v_m\}$ を v の連結とよぶ。

以下、有向グラフ $\vec{G}=(V,\vec{E})$ と任意の $v\in V$ について、$A[v]:=\{v'\in V|(v',v)\in\vec{E}\}$($v$ の親集合)、$B[v]:=\{v'\in V|(v,v')\in\vec{E}\}$($v$ の子集合)、$D[v]:=V-\{v\}-C[v]$(図 1, v の非子孫集合) とおく。

定義 3　1. $v\in C[v]$ なる $v\in V$ が存在しないとき、有向グラフ $\vec{G}=(V,\vec{E})$ は非巡回 (acyclic) であるという。

2. $v\in S[v]$ なる $v\in V$ が存在しないとき、無向グラフ $G=(V,E)$ は単一結合 (singly connected) であるという。また、有向グラフ $\vec{G}$ の無向グラフ G が単一結合の場合、$\vec{G}=(V,\vec{E})$ は単一結合であるという。

例 1　$\vec{G}_0=\{\{z_1,z_2,z_3,z_4,z_5,z_6,z_7,z_8,z_9,z_{10},z_{11},z_{12},z_{13}\},\{(z_1,z_2),(z_1,z_4),(z_2,z_4),(z_2,z_5),(z_3,z_5),(z_3,z_7),(z_4,z_8),(z_5,z_8),(z_5,z_9),(z_6,z_9),(z_8,z_{11}),(z_8,z_{12}),(z_9,z_{12}),(z_9,z_{13}),(z_{10},z_{13})\}\}$ において (図 1)、z_5 の親集合 $A[z_5]=\{z_2,z_3\}$、子集合 $B[z_5]=\{z_8,z_9\}$、子孫集合 $C[z_5]=\{z_8,z_9,z_{11},z_{12},z_{13}\}$、非子孫集合 $D[z_5]=\{z_1,z_2,z_3,z_4,z_6,z_7,z_{10}\}$ をそれぞれ示した。

命題 1　$\vec{G}$ が単一結合ならば、$\vec{G}$ は非巡回である (逆は成立しない)。

例 2　$\vec{G}_1=(\{z_1,z_2,z_3\},\{(z_1,z_2),(z_1,z_3)\})$ (図 2(a))、$\vec{G}_2=(\{z_1,z_2,z_3\},\{(z_2,z_1),(z_1,z_3)\})$ (図 2(b))、および $\vec{G}_3=(\{z_1,z_2,z_3\},\{(z_1,z_2),(z_2,z_3),(z_1,z_3)\})$ (図 2(c)) は有向非巡回グラフであるが、$\vec{G}_4=(\{z_1,z_2,z_3\},\{(z_1,z_2),(z_2,z_3),(z_3,z_1)\})$ (図 2(d)) は有向非巡回グラフではない。また、$\vec{G}_1,\vec{G}_2$ は単一結合だが、$\vec{G}_3$ は単一結合ではない。

命題 2　$\vec{G}=(V,\vec{E})$ が有向非巡回グラフであれば、$v\in V$ に関して

$$A[v]\subseteq D[v] \tag{1}$$

が成立する。

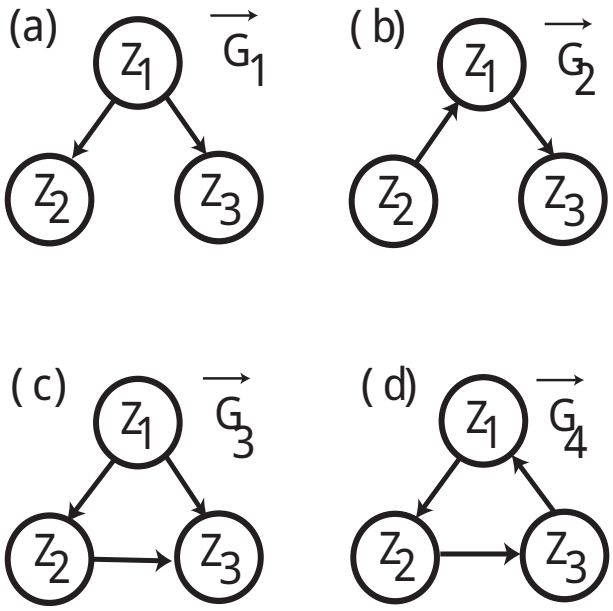

図 2: 有向グラフ $\vec{G}_1,\vec{G}_2,\vec{G}_3,\vec{G}_4$

命題 2 の証明: $A[v]$ の定義から $v\notin A[v]$。したがって、$A[v]\cap C[v]=\{\}$ をいえば十分である。しかし、$u\in A[v]\cap C[v]$ なる $u\in V$ が存在すれば、$u\to v\to v_1\to\cdots\to v_{m-1}\to u$ $(m\geq 1, v_0=v)$ なる $v_1,\cdots,v_{m-1}\in V$ が存在するので、$\vec{G}$ が非巡回であることに矛盾する。(証明終)

2.2　依存分離性

定義 4　有向非巡回グラフ $\vec{G}=(V,\vec{E})$ の互いに共通部分をもたない V の部分集合 X,Y,U の任意の $x\in X,\ y\in Y,\ x-v_1-v_2-\cdots-v_{m-1}-v_m$ について、ある $k(1\leq k\leq m)$ が存在し、

条件 1　$v_{k-1}\to v_k\to v_{k+1}$ または $v_{k+1}\to v_k\to v_{k-1}$ が存在し、$v_k\in U$ (図 3(a))

条件 2　$v_k\to v_{k-1}, v_k\to v_{k+1}$ が存在し、$v_k\in U$ (図 3(b))

条件 3　$v_{k-1}\to v_k, v_{k+1}\to v_k$ が存在し、$(C[v_k]\cup\{v_k\})\cap U=\{\}$ (図 3(c))

のいずれかが成り立つ (無向経路 $x-v_1-v_2,\cdots,v_{k-1}-v_k-v_{k+1},\cdots,v_{m-1}-v_m-y$ が v_k によってブロックされる) とき、U のもとで X は Y から依存分離的 (dependency-separate) である、もしくは U は X を Y から分離 (separate) するといい、$<X|U|Y>_{\vec{G}}$ と書く。

例 3　$\vec{G}_0$(図 1) に関して、

1. $U=\{\}$ のもとで、無向経路 $z_6-z_9-z_{13}-z_{10}$ が z_{13} によってブロックされるので (条件 3)、$X=\{z_6\}$ は $Y=\{z_{10}\}$ から依存分離的である。

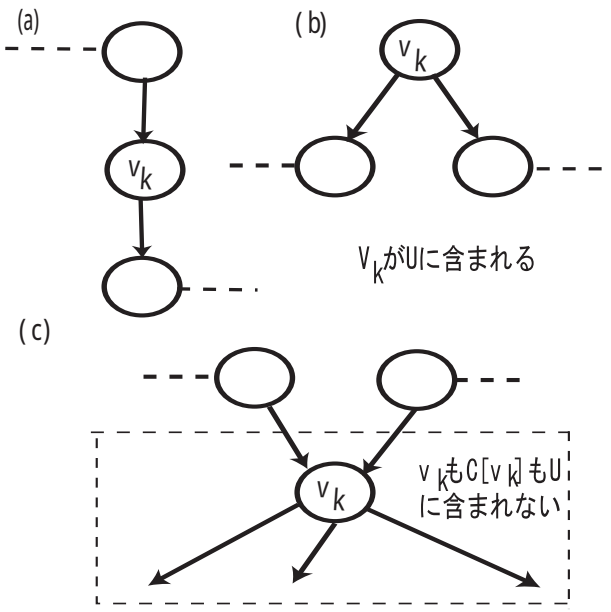

図 3: 依存分離性: 条件 1, 条件 2, 条件 3

2. $U = \{z_9\}$ のもとで、無向経路 $z_6 - z_9 - z_{13} - z_{10}$) がブロックされるので (条件 1)、$X = \{z_6\}$ は $Y = \{z_{10}\}$ から依存分離的である。
3. $U = \{z_{13}\}$ のもとで、無向経路 $z_6 - z_9 - z_{13} - z_{10}$ が 3 条件のいずれも満足しないので、$X = \{z_6\}$ は $Y = \{z_{10}\}$ は依存分離的ではない。
4. 以下の考察から、$U = \{z_5\}$ のもとで、$X = \{z_2\}$ は $Y = \{z_9\}$ から依存分離的である。
   (a) 無向経路 $z_2 - z_5, z_5 - z_9$ は z_5 によってブロックされる (条件 1)。
   (b) 無向経路 $z_2 - z_4 - z_8 - z_{12} - z_9$ および $z_2 - z_1 - z_4 - z_8 - z_{12} - z_9$ は、z_{12} によってブロックされる (条件 3)。
   (c) 無向経路 $z_2 - z_4 - z_8 - z_5 - z_9$ および $z_2 - z_1 - z_4 - z_8 - z_5 - z_9$ は、z_5 によってブロックされ (条件 2)、z_8 によってブロックされる (条件 3)。
   (d) 無向経路 $z_2 - z_5 - z_8 - z_{12} - z_9$ は、z_5 によってブロックされ (条件 1)、z_{12} によってブロックされる (条件 3)。
5. $U = \{\}$ のもとで、無向経路 $z_2 - z_5 - z_3$ および $z_2 - z_5 - z_3 - z_7$ は z_5 によってブロックされる (条件 3) ので、$X = \{z_2\}$ は $Y = \{z_3, z_7\}$ から依存分離的である。
6. $U = \{z_{12}\}$ のもとで、$z_{12} \in C[z_5]$ であり、無向経路 $z_2 - z_5 - z_3$ は 3 条件のいずれも満足しないので、$X = \{z_2\}$ は $Y = \{z_3, z_7\}$ から依存分離的ではない。

以下では、成立する $< X|U|Y >_{\vec{G}}$ の集合を $D(\vec{G})$、また $D(\vec{G})_0 = \{< X|U|Y >_{\vec{G}} \in D(\vec{G}) : X = \{\}$ または $Y = \{\}\}$ とおく。

例 **4**
1. $D(\vec{G_1}) = D_0(\vec{G_1}) \cup \{< \{z_2\}|\{z_1\}|\{z_3\} >, < \{z_3\}|\{z_1\}|\{v_2\} >\}$
2. $D(\vec{G_2}) = D_0(\vec{G_2}) \cup \{< \{z_2\}|\{z_1\}|\{z_3\} >, < \{z_3\}|\{z_1\}|\{v_2\} >\} = D(\vec{G_1})$
3. $D(\vec{G_3}) = D_0(\vec{G_3})$

命題 **3** 有向非巡回グラフ $\vec{G} = (Z, \vec{W})$ の任意の $z \in Z$ に関して、$< \{z\}|A[z]|D[z] - A[z] >_{\vec{G}}$

命題 3 の証明: z から $A[z]$ を経る無向経路は、定義 4 の条件 1,2 よりブロックされる。z から $A[z]$ の頂点を経ないで $D[z] - A[z]$ の頂点に至る経路では、$C[z]$ の頂点を経て、ある $z' \in C[z]$ の親 $z'' \in A[z']$ に至ることになる。しかし、$z'' \in A[z]$ とすれば、$\vec{G}$ が非巡回であることに矛盾し、$z'' \notin U = A[z]$ なので、定義 4 の条件 3 より、この経路はブロックされる。(証明終)

2.3 条件付独立

定義 **5** $Z := \{z_1, \cdots, z_N\}$ を確率変数の集合、$P(z_1, \cdots, z_N)$ その同時確率とする。互いに共通部分をもたない Z の部分集合 X, Y, U に関して、

$$X' \subseteq X, Y' \subseteq Y, P(U, Y') > 0 \Rightarrow P(X'|U, Y') = P(X'|U) \tag{2}$$

が成立するとき、U が与えられたとき X と Y は互いに条件付独立 (conditionally independent) であるといい、$I(X, U, Y)$ が成立するという。

例 **5** $Z = \{z_1, z_2, z_3\}$, $P(z_1, z_2, z_3) > 0$ について、
1. $P(z_2|z_1) = P(z_2) \Rightarrow I(\{z_2\}, \{\}, \{z_1\}))$
2. $P(z_3|z_1, z_2) = P(z_3) \Rightarrow I(\{z_3\}, \{\}, \{z_1, z_2\})$
3. $P(z_3|z_1, z_2) = P(z_3|z_1) \Rightarrow I(\{z_3\}, \{z_1\}, \{z_2\})$

命題 **4** 互いに共通部分をもたない任意の Z の部分集合 X, U, Y, W について以下がいえる。

$$I(X, U, \{\}) \tag{3}$$

$$I(X, U, Y) \Longleftrightarrow I(Y, U, X) \tag{4}$$

$$I(X, U, Y \cup W) \Longrightarrow I(X, U \cup W, Y) \tag{5}$$

$$I(X, U, Y \cup W) \Longrightarrow I(X, U, Y) \wedge I(X, U, W) \tag{6}$$

$$I(X, U, Y) \wedge I(X, U \cup Y, W) \Longrightarrow I(X, U, Y \cup W) \tag{7}$$

Z に含まれる確率変数の確率分布が既知であれば、確率分布 $P(z_1, \cdots, z_N)$ から Z の互いに共通部分をもたない任意の部分集合 X, Y, U に関しての $I(X, U, Y)$ の真

偽はすべて得られる。しかしながら、実際の知識情報処理では、$P(z_1,\cdots,z_N)$ が既知という仮定は一般的ではない。仮定する $I(X,U,Y)$ の集合を Z の依存モデル (dependency model) という。

以下では、Z の依存モデル M が任意に与えられたときに、それを有向非巡回グラフで表現することを考える。以下、依存モデル M によって $I(X,U,Y)$ が真になることを $I(X,U,Y)_M$ のように添え字 "M"をつけ、確率変数 $Z=\{z_1,\cdots,z_N\}$ の確率分布 $P(z_1,\cdots,z_N)$ が既知であり、その情報に基づいて $I(X,U,Y)$ が真になる場合は $I(X,U,Y)_Z$ のように添え字 Z をつけて表すものとする。また、

$$M_0=\{I(X,U,Y)_M : X=\{\} \text{ または } Y=\{\}\}$$

と表すものとする。

例 **6** $I(\{z_2\},\{z_1\},\{z_3\})$ と (3)(4)(5)(6)(7) から演繹できる $I(\cdot,\cdot,\cdot)$ の集合 M は、

$$M \quad = \quad M_0 \cup \{I(\{z_2\},\{z_1\},\{z_3\}), I(\{z_3\},\{z_1\},\{z_2\})\}$$

となる。

M の依存モデルが Z の依存モデルと一致するか否かはどうかは議論しない。何らかの有向非巡回グラフ $\vec{G}$ で依存モデル M が表現できるか否かを問題とする。

2.4 ベイジアンネットワーク

定義 **6** Z を有限個の確率変数の集合とし、M をその依存モデルとする。有向非巡回グラフ $\vec{G}=(Z,\vec{W})$ に関して、互いに共通部分をもたない Z の任意の部分集合 X,U,Y について

$$< X|U|Y >_{\vec{G}} \Rightarrow I(X,U,Y)_M \tag{8}$$

が成立し、$\vec{W}$ を任意の真部分集合 $\vec{W'}\subset\vec{W}$ に置き換えたときに *(8)* が成立しなくなるとき、$\vec{G}$ は M のベイジアンネットワークであるという。また、M が Z の確率分布に基づいて正しく与えられ、

$$< X|U|Y >_{\vec{G}} \Rightarrow I(X,U,Y)_Z \tag{9}$$

が成立するとき、$\vec{G}$ は Z のベイジアンネットワークであるという。

すなわち、ベイジアンネットワークは、確率変数間の依存性モデルを頂点間の依存分離性を用いて表現した有向非巡回グラフであるということができる。

例 **7** 例 6 の M に関して、例 4 より

1. 有向非巡回グラフ $\vec{G_1}$ は、M のベイジアンネットワークである。
2. 有向非巡回グラフ $\vec{G_2}$ は、M のベイジアンネットワークである。
3. 有向非巡回グラフ $\vec{G_3}$ は、M のベイジアンネットワークではない。実際、$\vec{G_1}$ の有向辺集合 $\{(z_1,z_2),(z_1,z_3)\}$ は $\vec{G_3}$ の有向辺集合 $\{(z_1,z_2),(z_2,z_3),(z_1,z_3)\}$ の真部分集合である。

定義 **7** M を確率変数の有限集合 $Z=\{z_1,\cdots,z_N\}$ の依存モデルとし、任意に置換 $\pi:\{1,\cdots,N\}\to\{1,\cdots,N\}$ を固定し、また、$Z^\pi[z]:=\{z_k|\pi(k)<\pi(j),z=z_j\}$ とおく。各 $z\in Z$ について、$A=A^{(\pi)}[z]$ に関して

$$I(\{z\},A,Z^{(\pi)}[z]-A)_M \tag{10}$$

が成立するように $A^{(\pi)}[z]\subseteq Z^{(\pi)}[z]$ を各 $z\in Z$ でさだめると、$\vec{W}=\{(z',z)|z\in Z,z'\in A^{(\pi)}[z]\}$ として有向非巡回グラフ $\vec{G}=(Z,\vec{W})$ が生成される。このとき、対 $L_\pi=(\pi,A^{(\pi)})$ を M の階層化手順 (stratified protocol) とよぶ。また、M が Z の確率分布に基づいて正しく与えられるとき、同様の対 $L_\pi=(\pi,A^{(\pi)})$ を Z の階層化手順 (stratified protocol) という。

命題 **5** 任意のベイジアンネットワーク $\vec{G}=(Z,\vec{W})$ には、$z'\in D[z]\Rightarrow z<z'$ を満足する Z 間の半順序関係 $<$ が存在する。すなわち、$Z=\{z_1,\cdots,z_N\}$ として、$z_j\in D[z_i]\in\vec{W}\Rightarrow\pi(i)<\pi(j)$ を満足する置換 $\pi:\{1,\cdots,N\}\to\{1,\cdots,N\}$ が存在する。

命題 5 の証明: $\vec{G}$ は有向非巡回グラフであるので、任意の $z_i,z_j\in Z,\ i\neq j$ に関して、$z_i\in C[z_j]$ と $z_j\in C[z_i]$ は同時に成立しない。したがって、グラフの制約条件と矛盾しない置換 π が存在する。(証明終)

命題 5 の条件を満足する置換 π を $\vec{G}$ の許容可能 (admissible) な置換とよぶ。

例 **8** 例 7 の M に関して、

1. $\pi(1)=1,\pi(2)=2,\pi(3)=3$, $A^{(\pi)}[z_1]=\{\}$, $A^{(\pi)}[z_2]=\{z_1\},A^{(\pi)}[z_3]=\{z_1\}$ であれば $\vec{G_1}$ (図 3(a))、
2. $\pi(1)=2,\pi(2)=1,\pi(3)=3$, $A^{(\pi)}[z_1]=\{z_2\}$, $A^{(\pi)}[z_2]=\{\},A^{(\pi)}[z_3]=\{z_1\}$ であれば $\vec{G_2}$ (図 3(b))、
3. $\pi(1)=1,\pi(2)=2,\pi(3)=3$, $A^{(\pi)}[z_1]=\{\}$, $A^{(\pi)}[z_2]=\{z_1\},A^{(\pi)}[z_3]=\{z_1,z_2\}$ であれば $\vec{G_3}$ (図 3(c))、

になるので、$\vec{G_1}, \vec{G_2}, \vec{G_3}$ は、M の階層化手順によって生成される。

定理 1 (Verma) $Z = \{z_1, \cdots, z_N\}$ を確率変数の有限集合、M を (3)(4)(5)(6)(7) を満足する Z の依存モデルとする。M の階層化手順によって生成された有向非巡回グラフ $\vec{G}$ に関して、(8) が成立する。また、各 $A^{(\pi)}[z], z \in Z$ として最小のもの (各 π に対して一意) を選ぶとき、またそのときに限り、$\vec{G}$ はベイジアンネットワークになり、さらに

$$< X|U|Y >_{\vec{G}} \Leftarrow I(X, U, Y)_M \tag{11}$$

が成立する。

また、M が確率分布 $P(z_1, \cdots, z_N)$ に基づいて与えれたとき、命題 4 より、(3)(4)(5)(6)(7) が満足されるので、以下が成立する。

系 1 Zの階層化手順によって生成された有向非巡回グラフ $\vec{G}$ に関して、(9) が成立する。また、各 $A^{(\pi)}[z], z \in Z$ として最小のもの (各 π に関して一意) を選ぶとき、またそのときに限り、$\vec{G}$ はベイジアンネットワークになり、さらに

$$< X|U|Y >_{\vec{G}} \Leftarrow I(X, U, Y)_Z \tag{12}$$

が成立する。

系 2 Z のベイジアンネットワーク $\vec{G}$ で、許容可能な置換 $\pi : \{1, \cdots, N\} \to \{1, \cdots, N\}$ を任意に固定する。また、$z \in Z$ について、

$$I(\{z\}, A^{(\pi)}[z], Z^{(\pi)}[z] - A^{(\pi)}[z])_M \tag{13}$$

が成立する最小の $Z^{(\pi)}[z]$ の部分集合を $A^{(\pi)}[z]$ とする。このとき、π によらず、すべての $z \in Z$ で、$A^{(\pi)}[z] = A[z]$ である。

例 9

1. 図 3(a) は、$\pi(1) = 1, \pi(2) = 2, \pi(3) = 3$ とみなせば、$A^{(\pi)}[z_1] = \{\}, A^{(\pi)}[z_2] = \{z_1\}, A(\pi)[z_3] = \{z_1\}$、$\pi(1) = 1, \pi(2) = 3, \pi(3) = 2$ とみなせば、$A^{(\pi)}[z_1] = \{\}, A^{(\pi)}[z_2] = \{z_1\}, A^{(\pi)}[z_3] = \{z_1\}$ となる。
2. 図 3(b) は、$\pi(1) = 2, \pi(2) = 1, \pi(3) = 3$ とみなせば、$A^{(\pi)}[z_2] = \{\}, A^{(\pi)}[z_1] = \{z_2\}, A[z_3]^{(\pi)} = \{z_2\}$、$\pi(1) = 2, \pi(2) = 3, \pi(3) = 1$ とみなせば、$A^{(\pi)}[z_2] = \{\}, A^{(\pi)}[z_3] = \{z_2\}, A[z_1]^{(\pi)} = \{z_2\}$、となる。

系 3 $Z = \{z_1, \cdots, z_N\}$ を確率変数の有限集合、M を (3)(4)(5)(6)(7) を満足する Z の依存モデルとする。有向非巡回グラフ $\vec{G} = (Z, \vec{W})$ が M のベイジアンネットワークであるための必要十分条件は、各 $z \in Z$ について、$A = A[z]$ に関して

$$I(\{z\}, A, D[z] - A)_M \tag{14}$$

が成立し、$A \subset A[z]$ に関して (14) が成立しないことである。

系 3 の証明: 以下の 4 ステップからなる。

1. $\vec{G} = (Z, \vec{W})$ が M のベイジアンネットワークであるとする。定義 7 と命題 3 より、$A = A[z]$ は (14) を満足しなければならない。
2. ある $z \in Z$ と $A \subset A[z]$ に関して (14) が成立すると仮定する。$\vec{G}$ は有向非巡回グラフであるから、命題 5 の条件を満足する置換 π が存在する。しかし、そのような π および $A^{(\pi)}[z] = A[z] \subseteq Z^{(\pi)}[z]$(系 2) に関して、(6) と $Z^{(\pi)}[z] \subseteq D[z]$ から、
$$I(\{z\}, A, D[z] - A)_M \Rightarrow I(\{z\}, A, Z^{(\pi)}[z] - A)_M \tag{15}$$
が成立し、$A^{(\pi)}[z]$ は、(10) を満足する $Z^{(\pi)}[z]$ の最小の部分集合でなかったことになる。これは、$\vec{G} = (Z, \vec{W})$ が M のベイジアンネットワークであることと矛盾する。
3. 逆に、系 3 の条件が満足されれば、$\vec{G}$ は有向非巡回グラフであるから、命題 5 の条件を満足する置換 π が存在する。そのような π および $A^{(\pi)}[z] = A[z] \subseteq Z^{(\pi)}, z \in Z$ に関して、(15) より、$A = A^{(\pi)}[z]$ は、(10) を満足する。
4. ある $z \in Z$ と $A \subset A[z]$ に関して (10) が成立すると仮定する。$X = \{z\}, U = A, Y = A[z] - A, W = D[z] - A[z]$ とおくと、(7) と $A[z] \subseteq D[z]$ より、$I(\{z\}, A, D[z] - A)_M$ が成立し、系 3 と矛盾。

(証明終)

例 10 例 6 の M に関して、

1. $\vec{G_1}$ (図 2(a)) は、M のベイジアンネットワークである。実際、
   (a) $C[z_1] = \{z_2, z_3\}$ なので、$D[z_1] = Z - \{z_1\} - C[z_1] = \{\}$。また、$A[z_1] = \{\}$。したがって、$A[z_1]$ は $I(\{z_1\}, A[z_1], D[z_1] - A[z_1])_M$ を満足する $D[z_1]$ の最小の部分集合になっている。
   (b) $C[z_2] = \{\}$ なので、$D[z_2] = Z - \{z_2\} - C[z_2] = \{z_1, z_3\}$。また、$A[z_2] = \{z_1\}$。一方、$I(\{z_2\}, \{z_1\}, \{z_3\})_M$ は成立するが、$I(\{z_2\}, \{\}, \{z_1, z_3\})_M$ は成立しない。したがって、$A[z_2]$ は $I(\{z_2\}, A[z_2],$

$D[z_2] - A[z_2])_M$ を満足する $D[z_2]$ の最小の部分集合になっている。

(c) $C[z_3] = \{\}$ なので、$D[z_3] = Z - \{z_3\} - C[z_3] = \{z_1, z_2\}$。また、$A[z_3] = \{z_1\}$。一方、$I(\{z_3\}, \{z_1\}, \{z_2\})_M$ は成立するが、$I(\{z_3\}, \{\}, \{z_1, z_2\})_M$ は成立しない。しかし、1.(c) より、$A[z_3]$ は $I(\{z_3\}, A[z_3], D[z_3] - A[z_3])_M$ を満足する $D[z_3]$ の最小の部分集合になっている。

2. 1. と同様の考察から、$\vec{G_2}$ (図 2(b)) は、M のベイジアンネットワークであることがわかる。

3. $\vec{G_3}$ (図 2(c)) は、M のベイジアンネットワークではない。実際、

(a) $C[z_3] = \{\}$ なので、$D[z_3] = Z - \{z_3\} - C[z_3] = \{z_1, z_2\}$。また、$A[z_3] = \{z_1, z_2\}$。一方、$I(\{z_3\}, \{z_1\}, \{z_2\})_M$ が成立する。したがって、$A[z_3]$ は $I(\{z_3\}, A[z_3],$
$D[z_3] - A[z_3])_M$ を満足する $D[z_3]$ の最小の部分集合ではない。

したがって、一般的な依存モデル M ではなく、Z の確率分布から出発する場合には、次のようになる。$Z = \{z_1, \cdots, z_N\}$ の確率分布 $P(z_1, \cdots, z_N)$ と Z のベイジアンネットワークが与えられれば、$P(z_1, \cdots, z_N)$ が

$$P(z_1, \cdots, z_N) = \prod_{j=1}^{N} P(z_j | A[z_j]))$$

のように一意に因数分解できる。また、置換 $\pi : \{1, \cdots, N\} \to \{1, \cdots, N\}$ が与えられれば、

$$P(z_1, \cdots, z_N) = \prod_{j=1}^{N} P(z_j | Z_k^{(\pi)}) = \prod_{j=1}^{N} P(z_j | A^{(\pi)}[z_j])) \tag{16}$$

に対応したベイジアンネットワーク $\vec{G} = (Z, \vec{W}), \vec{W} = \{(z', z) | z \in Z, z' \in A^{(\pi)}[z]\}$ が一意に決まる。ただし、$A^{(\pi)}[z_j])$ は (16) を満足する $Z_k^{(\pi)}$ の最小の部分集合とした。

3 単一結合ベイジアンネットワークの確率推論

$\vec{G} = (Z, \vec{W})$ を $Z = \{z_1, \cdots, z_N\}$ のベイジアンネットワークとする。確率変数 Z の集合のみが与えられても、ベイジアンネットワークは一意ではない。Z の要素間の順序をあらわす置換 π の情報が必要である。置換 π は、実際上は、Z の要素間の因果関係によって定められる。すなわち、$z_{\pi(1)}$ から $z_{\pi(N)}$ にむかって、事象が過去から現在に、あるいは原因から結果に推移するように π が定められる。

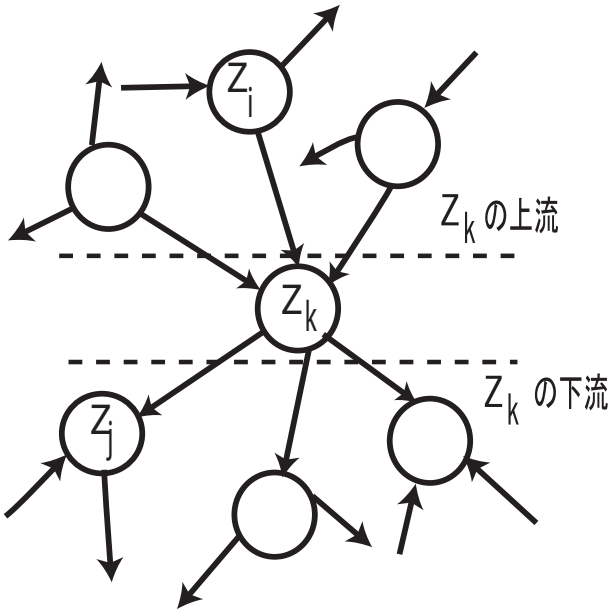

図 4: z_k の上流、下流

一般に確率的推論という場合、$P(Y|X)$ の推定や $P(Y|X)$ に基づく決定 (X, Y は Z の互いに共通部分を持たない部分集合) などを意味することもあるが、ベイジアンネットワークによる推論という場合は、$P(Y|X)$ の計算を意味することが多い。すなわち、$X := \{x_1, \cdots, x_L\}, Y = \{y_1, \cdots, y_M\}$ e_i を確率変数 x_i の実現値として、$x_1 = e_1, \cdots, x_L = e_L$ の観測 $e = (e_1, \cdots, e_L)$ が得られたときの y_j, $j = 1, \cdots, M$ の確率分布 $P(Y|X = e)$ の計算である。そのためには、$U := Z - X - Y$ とおいたときに、

$$P(Y|X = e) = \frac{\sum_U P(X = e, Y, U)}{\sum_{Y,U} P(X = e, Y, U)} \tag{17}$$

を求めれば十分である。しかしながら、各確率変数が有限個の値をとる場合、(17) を求めるための計算量が N とともに指数的に増大し、連続値をとる場合にはさらに困難を極めることになる。その計算の効率化が、以下の議論の中心になる。

本節では、各確率変数が有限個のいずれかの値をとり、ベイジアンネットワーク $\vec{G}$ が単一結合 (定義 2) になっていることを仮定する。

3.1 ベイジアンネットワークによる確率伝播

ベイジアンネットワーク $\vec{G} = (Z, \vec{W})$ の任意の頂点 $z \in Z$ を固定する。各 $z' \in A[z]$ について、$\vec{G}$ から (z', z) を除くと、2 個の有向グラフができる。このうち、z' が含まれている方を $\vec{G}[z']$、他方を $\vec{G} - \vec{G}[z']$ で表記する。また、$\vec{G}[z']$ の頂点を z' の側にある頂点とよぶ。また、$\cup_{z' \in A[z]} \vec{G}[z']$ の頂点を z の上流にある頂点とよぶ。同様

に、各 $z'' \in B[z]$ について、$\vec{G}$ から (z, z'') を除くと、2 個の有向グラフができる。このうち、z'' が含まれている方を $\vec{G}[z'']$、他方を $\vec{G} - \vec{G}[z'']$ で表記する。また、$\vec{G}[z'']$ の頂点を z'' の側にある頂点とよぶ。また、$\cup_{z'' \in B[z]} \vec{G}[z'']$ の頂点を z の下流にある頂点とよぶ。z の上流にあっても $A[z]$ 以外の要素は z の先祖であるとは限らないし、z の下流にあっても $B[z]$ 以外の要素は z の子孫であるとは限らない (図 4)。

BEL の計算

z より上流のいくつかの確率変数で観測 e^+ が得られ、下流のいくつかの確率変数で観測 e^- が得られた (e^+, e^- は、それぞれ z の上流・下流で観測された確率変数の数だけの大きさをもつベクトルとみなす) とする。

$\vec{G} = (Z, \vec{W})$ はベイジアンネットワークであることから、系 3 より、$I(A[z], \{z\}, C[z])_Z$ が成立する。したがって、$P(e^-|z, e^+) = P(e^-|z)$ が成立し、$\lambda(z) = P(e^-|z)$, $\pi(z) = P(z|e^+)$ として、

$$BEL(z) := P(z|e^+, e^-) = \frac{P(z|e^+)P(e^-|z)}{P(e^-)} = \alpha\lambda(z)\pi(z) \tag{18}$$

が成立する。ただし、$\alpha = [P(e^-)]^{-1}$。$\lambda(z)$ は、観測 e^- のもとでの z の尤度とよばれ、$\sum_z \lambda(z) = 1$ になる必要がなく、確率変数 z の各実現値の比のみが意味を持つ。$\pi(z)$ は、観測 e^+ のもとでの z の事後確率とよばれる。

以下では、確率変数の集合 $Z = \{z_1, \cdots, z_N\}$ に対して、$\pi_{k \to l}(z_k)$ で、$z_k \to z_l$ の側の観測のもとでの z_k の事後確率をあらわし、$\pi_{k+}(z_k)$ で、z_k より上流の観測のもとでの z_k の事後確率をあらわす。また、$\lambda_{l \to k}(z_k)$ で、$z_k \to z_l$ の側の観測のもとでの z_k の尤度、$\lambda_{k-}(z_k)$ で、z_k より下流の観測のもとでの z_k の尤度をあらわす。また、各 $k \in \{1, \cdots, N\}$ について、$z_i \in A[z_k] \Leftrightarrow i \in a[k]$ および $z_i \in B[z_k] \Leftrightarrow i \in b[k]$ となるように $a[k], b[k] \subseteq \{1, \cdots, N\}$ を定義する。

確率伝播は、以下の 3 種類の頂点 z_k から始まる。

1. z_k の上流の頂点がなければ、$\pi_{k+}(z_k)$ には z の各実現値の事前確率 $P(z_k)$ をおく。
2. z_k の下流の頂点がなければ、$\lambda_{k-}(z_k)$ にはすべて等しい値 (通常は 1) をおく。

$$\lambda_{k-}(z_k) = P(e^-|z_k) = 1 \Rightarrow P(z_k = e_k|e^-)$$
$$= \frac{P(z_k = e_k)\lambda_{k-}(z_k = e_k)}{\sum_{z_k} P(z_k)\lambda_{k-}(z_k)} = P(z_k = e_k)$$

より、このようにおくことは、下流での観測 e^- のもとでの事後確率を事前確率のままにしておくことと等価である。

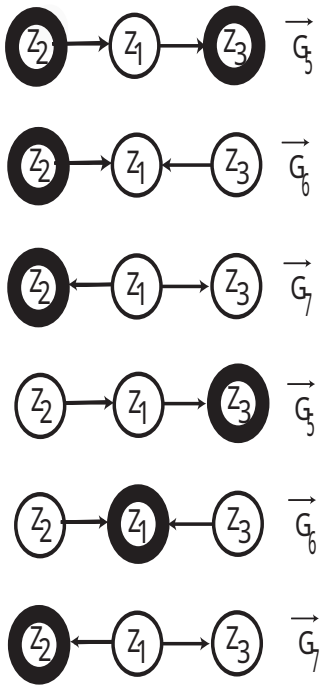

図 5: 有向グラフ $\vec{G}_5, \vec{G}_6, \vec{G}_7$

3. z_k の値自身が観測された場合は、その観測値 $z_k = e_k$ および $i \in b[k]$ に関して $\lambda_{i \to k}(e_k) = 1$ で、$z_k \neq e_k$ に関しては $\lambda_{i \to k}(z_k) = 0$ とおく。このようにおくことは、$P(z_k = e_k|e^-) = 1$ とおくことと等価である。

例 11 $Z = \{z_1, z_2\}$ のベイジアンネットワーク $\vec{G}_5 = (\{z_2, z_3\}, \{(z_2, z_1), (z_1, z_3)\})$ で、z_2 に $P(z_2)$ の各値が、z_1 に $P(z_1|z_2)$ の各値が、z_3 に $P(z_3|z_1)$ の各値が格納される。$z_2 = e_2, z_3 = e_3$ が観測された場合 (図 5(a))、$z_2 = e_2$ の観測のもとでの z_2 の事後確率は

$$\pi_{1+}(z_1) = P(z_1|z_2 = e_2)$$

$z_3 = e_3$ の観測のもとでの z_1 の尤度は

$$\lambda_{1-}(z_1) = P(z_3 = e_3|z_1)$$

したがって、

$$BEL(z_1) \propto \lambda_{1-}(z_1)\pi_{1+}(z_1) = P(z_3 = e_3|z_1)P(z_1|z_2 = e_2)$$

z_k の上流での観測 e^+ が、各 $i \in a[k]$ で、$z_i \to z_k$ の側でなされ e_i とすれば、$\vec{G}$ が単一結合なので、$< \{z_i, e_i\}|\{\}|\{z_j, e_j\} >_G$ と定義 4 の条件 3 および定義 7 から、$i \neq j, i, j \in a[k]$ で $I(\{z_i, e_i\}, \{\}, \{z_j, e_j\})_M$ が成立する。したがって、

$$\pi_{k+}(z_k) = P(z_k|e^+) = P(z|e_i, i \in a[k])$$
$$= \sum_{z_i, i \in a[k]} P(z_k|z_i, i \in a[k])P(z_i, i \in a[k]|e_i, i \in a[k])$$

$$= \sum_{z_i, i\in a[k]} P(z_k|z_i, i \in a[k]) \prod_{j\in a[k]} P(z_j|e_j)$$
$$= \sum_{z_i, i\in a[k]} P(z_k|z_i, i \in a[k]) \prod_{j\in a[k]} \pi_{j\to k}(z_j) \quad (19)$$

例 12 $Z = \{z_1, z_2, z_3\}$ のベイジアンネットワーク $\vec{G}_6 = (\{z_1, z_2, z_3\}, \{(z_2, z_1), (z_3, z_1)\})$ で、z_2 に $P(z_2)$ の各値が、z_3 に $P(z_3)$ の各値、z_1 に $P(z_1|z_2, z_3)$ の各値が格納されている。$z_2 = e_2$ が観測された場合 (図 5(b))、

$$\pi_{2+}(z_2) = \begin{cases} 1, & (z_1 = e_1) \\ 0, & (z_1 \neq e_1) \end{cases}, \ \pi_{3+}(z_3) = P(z_3)$$

$\pi_{2\to 1}(z_2) = \pi_{2+}(z_2)$、$\pi_{3\to 1}(z_3) = \pi_{3+}(z_3)$、および (19) より、

$$\pi_{1+}(z_1) = \sum_{z_2, z_3} P(z_1|z_2, z_3)\pi_{2\to 1}(z_2)\pi_{3\to 1}(z_3)$$
$$= \sum_{z_3} P(z_1|z_2 = e_2, z_3)P(z_3)$$
$$\lambda_{1-}(z_1) = 1$$

(18) より、

$$BEL(z_1) \propto \pi_{1+}(z_1)\lambda_{1-}(z_1) = \sum_{z_3} P(z_1|z_2 = e_2, z_3)P(z_3)$$

となる。

z_k の下流での観測 e^- が、各 $i \in b[k]$ で、$z_k \to z_i$ の側でなされ e_i とすれば、$\vec{G}$ が単一結合なので、$i, j \in b[k], i \neq j$ として $(z_k, z_i), (z_k, z_j) \in \vec{W}$ と定義 4 の条件 2 および定義 7 から、$i \neq j, i, j \in b[k]$ で $I(e_i, \{z_k\}, e_j)_M$ が成立する。したがって、

$$\lambda_{k-}(z_k) = P(e^-|z_k) = P(z_i = e_i, i \in b[k]|z_k)$$
$$= \prod_{i\in b[k]} P(z_i = e_i|z_k) = \prod_{i\in b[k]} \lambda_{i\to k}(z_k) \quad (20)$$

例 13 $Z = \{z_1, z_2, z_3\}$ のベイジアンネットワーク $\vec{G}_7 = (\{z_1, z_2, z_3\}, \{(z_1, z_2), (z_1, z_3)\})$ で、z_1 に $P(z_1)$ の各値が、z_2 に $P(z_2|Z_1)$ の各値、z_3 に $P(z_3|z_1)$ の各値が格納されている。$z_2 = e_2$ が観測された場合 (図 5(c))、

$$\lambda_{2\to 1}(z_1) = P(z_2 = e_2|z_1), \ \lambda_{3\to 1}(z_1) = 1$$

(19) より、

$$\lambda_{1-}(z_1) = \lambda_{2\to 1}(z_1)\lambda_{3\to 1}(z_1) = P(z_2 = e_2|z_1)$$

$$\pi_{1+}(z_1) = P(z_1)$$

(18) より、

$$BEL(z_1) \propto \pi_{1+}(z_1)\lambda_{1-}(z_1) = P(z_1)P(z_2 = e_2|z_1)$$

となる。

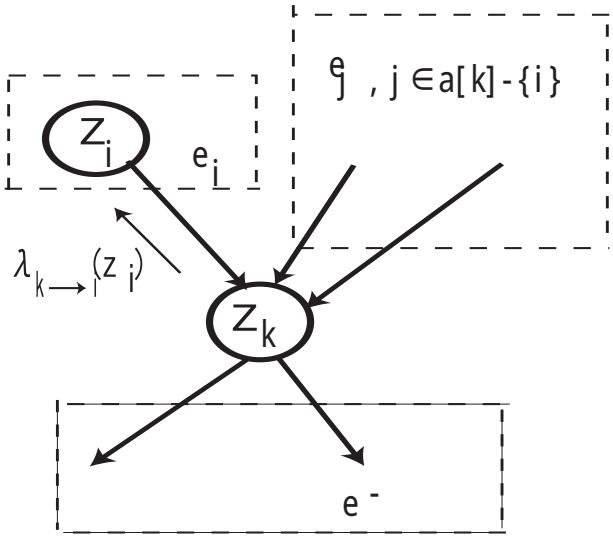

図 6: λ の更新

λ の更新

$k \in \{1, \cdots, N\}$ として、$\lambda_{k+}(z_k)$ と観測 $e^-, e_j, \ j \in a[k] - \{i\}$ から、各 $i \in a[k]$ について、$\lambda_{k\to i}(z_i)$ を計算する (図 6)。

命題 6 $\beta_i = P(e_j, j \in a[k] - \{i\})$ として

$$\lambda_{k\to i}(z_i) = \beta_i \sum_{z_k} \lambda_{k-}(z_k) \sum_{z_j, j\in a[k]-\{i\}} P(z_k|z_j, j \in a[k]) \cdot \prod_{l\in a[k]-\{i\}} \pi_{l\to k}(z_l) \quad (21)$$

命題 6 の証明: まず、

$$\lambda_{k\to i}(z_i)$$
$$= P(e^-, e_j, j \in a[k] - \{i\}|z_i)$$
$$= \sum_{z_j, j\in a[k]\cup\{k\}-\{i\}} P(e^-, e_j, j \in a[k] - \{i\}|z_j, j \in a[k] \cup \{k\} - \{i\})P(z_j, j \in a[k] \cup \{k\} - \{i\}|z_i)$$
$$= \sum_{z_j, j\in a[k]\cup\{k\}-\{i\}} P(e^-|z_k) \cdot P(e_j, j \in a[k] - \{i\}|z_j, j \in a[k] - \{i\}) \cdot P(z_j, j \in a[k] \cup \{k\} - \{i\}|z_i) \quad (22)$$

がいえる。ただし、最後の等号は $I(\{e_j, j \in a[k]-\{i\}\}, \{z_k\}, e^-)_M$ および $I(\{e_j, j \in a[k] - \{i\}\}, \{z_j, j \in a[k] - \{i\}\}, \{z_i\})_M$ によった。

$$P(e^-|z_k) = \lambda_{k+}(z_k) \quad (23)$$

$$P(e_j, j \in a[k] - \{i\}|z_j, j \in a[k] - \{i\})$$
$$= P(e_j, j \in a[k] - \{i\}) \cdot \frac{P(z_j, j \in a[k] - \{i\}|e_j, j \in a[k] - \{i\})}{P(z_j, j \in a[k] - \{i\})}$$

$$= \beta_i \cdot \frac{\prod_{j\in a[k]-\{i\}} P(z_j|e_j)}{P(z_j, j \in a[k]-\{i\})}$$
$$= \beta_i \cdot \frac{\prod_{j\in a[k]-\{i\}} \pi_{j\to k}(z_j)}{P(z_j, j \in a[k]-\{i\})} \quad (24)$$

さらに、$I(\{z_i\},\{\},\{e_j^+, j \in a[k]-\{i\}\})_M$ より、

$$P(z_j, j \in a[k]\cup\{k\}-\{i\}|z_i)$$
$$= P(z_k|z_j, j \in a[k])P(z_j, j \in a[k]-\{i\}|z_i)$$
$$= P(z_k|z_j, j \in a[k])P(z_j, j \in a[k]-\{i\}) \quad (25)$$

(22)(23)(24)(25) より、(21) が成立する。(証明終)

(21) を各 $i \in a[k]$ で求めて (20) に代入する。この情報は、ベイジアンネットワークの矢印の逆方向 (下流から上流) に流れる。

例 14 $\vec{G_5}$ で、$z_3 = e_3$ が観測された場合 (図 5(d))、

$$\lambda_{1-}(z_1) = P(z_3 = e_3|z_1)$$

(21) より、

$$\lambda_{2-}(z_2) \propto \sum_{z_1} \lambda_{1-}(z_1)P(z_1|z_2)$$
$$\pi_{2+}(z_2) = P(z_2)$$

(18) より、

$$BEL(z_2) \propto \pi_{2+}(z_2)\lambda_{2-}(z_2)$$
$$= P(z_2)\sum_{z_1} P(z_3 = e_3|z_1)P(z_1|z_2)$$

が求まる。

例 15 $\vec{G_6}$ で、$z_1 = e_1$ が観測された場合 (図 5(e))、

$$\lambda_{1-}(z_1) = \begin{cases} 1 & z_1 = e_1 \\ 0 & z_1 \neq e_1 \end{cases}$$

$\pi_{3\to1}(z_3) = \pi_{3+}(z_3) = P(z_3)$ および (21) より、

$$\lambda_{1\to2}(z_2) \propto \sum_{z_3} \lambda_{1-}(z_1)P(z_1|z_2, z_3)\pi_{3\to1}(z_3)$$
$$= \sum_{z_3} P(z_1 = e_1|z_2, z_3)P(z_3)$$
$$\pi_{2+}(z_2) = P(z_2)$$

$\lambda_{2-}(z_2) = \lambda_{1\to2}(z_2)$ および (18) より、

$$BEL(z_2) \propto \pi_{2+}(z_2)\lambda_{2-}(z_2)$$
$$= P(z_2)\sum_{z_3} P(z_1 = e_1|z_2, z_3)P(z_3)$$

が求まる。対称性から

$$BEL(z_3) \propto P(z_3)\sum_{z_2} P(z_1 = e_1|z_2, z_3)P(z_2)$$

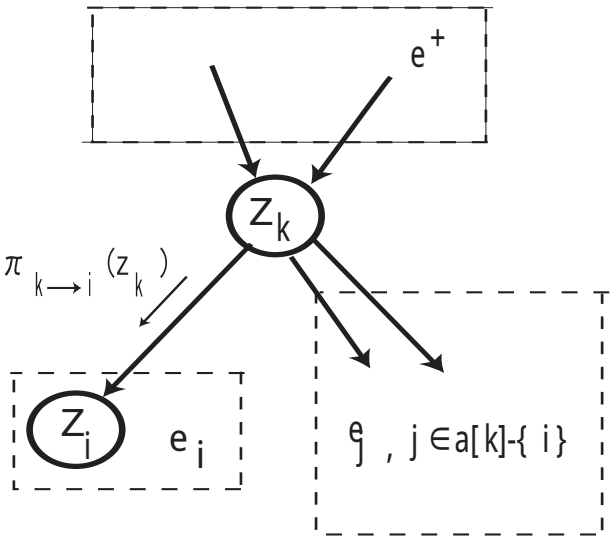

図 7: π の更新

π の更新

$k \in \{1,\cdots,N\}$ として、$\pi_{k+}(z_k)$ と観測 $e^+, e_j,\ j \in b[k]-\{i\}$ から、各 $i \in b[k]$ について、$\pi_{k\to i}(z_k)$ を計算する (図 7)。

命題 7 $\gamma_i = [P(e_j, j \in b[k]-\{i\})]^{-1}$ として、

$$\pi_{k\to i}(z_k) = \gamma_i \pi_{k+}(z_k) \prod_{j\in b[k]-\{i\}} \lambda_{j\to k}(z_k) \quad (26)$$

命題 7 の証明:

$$\pi_{k\to i}(z_k) = P(z_k|e^+, e_j^-, j \in b[k]-\{i\})$$
$$= P(z_k|e^+)P(\{e_j, j \in b[k]-\{i\}|e^+, z)$$
$$= \gamma_i P(z_k|e^+)\prod_{j\in b[k]-\{i\}} P(e_j|z_k)$$
$$= \gamma_i \pi_{k+}(z_k)\prod_{j\in b[k]-\{i\}} \lambda_{j\to k}(z_k)$$

となる。(証明終)

(26) を各 $i \in b[k]$ でもとめて (19) に適用する。この情報は、ベイジアンネットワークの矢印と同じ方向 (上流から下流) に流れる。

(18)(20)(26) より、

$$BEL(z_k) \propto \lambda_{k-}(z_k)\pi_{k+}(z_k) \propto \lambda_{i\to k}(z_k)\pi_{k\to i}(z_k)$$

(図 8)。

例 16 $\vec{G_7}$ で、$z_2 = e_2$ が観測された場合 (図 5(f))、(26) より、

$$\pi_{1\to3}(z_1) \propto \pi_{1+}(z_1)\lambda_{2\to1}(z_1) = P(z_2 = e_2|z_1)P(z_1)$$
$$\pi_{3+}(z_3) \propto \sum_{z_1} P(z_3|z_1)\pi_{1\to3}(z_1)$$

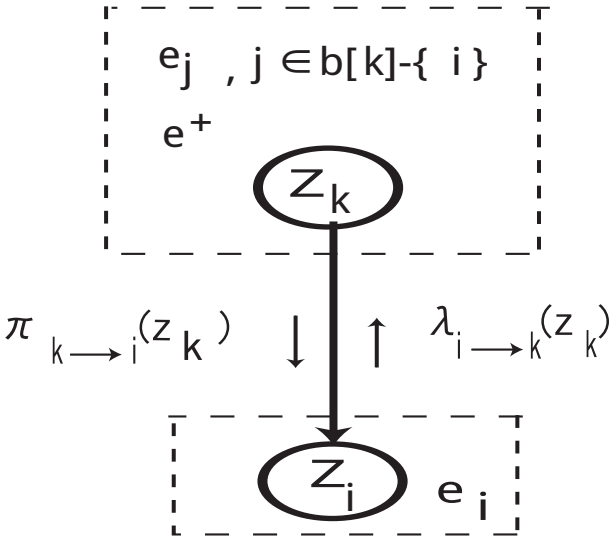

図 8: $BEL(z_k)$ の計算

$$= \sum_{z_1} P(z_3|z_1)P(z_2 = e_2|z_1)P(z_1)$$

(18) および $\lambda_{3-}(z_3) = 1$ より、

$$BEL(z_3) \propto \pi_{3+}(z_3)\lambda_{3-}(z_3)$$

$$= \sum_{z_1} P(z_3|z_1)P(z_2 = e_2|z_1)P(z_1)$$

となる。

確率伝播

上流に頂点がない場合の $\pi_{k+}(z_k)$、下流に頂点がない場合の $\lambda_{k-}(z_k)$、z_k で値が観測された場合の $\pi_{k+}(z_k)$ および $\lambda_{k-}(z_k)$ 以外は、値が計算されていない。以下の説明のため、

準備状態 頂点 z_k について、高々1 個の $i \in a[k] \cup b[k]$ が存在し、$j \in a[k] - \{i\}$ に対して$\pi_{j\to k}(z_j)$ の値が、$j \in b[k] - \{i\}$ に対して$\lambda_{j\to k}(z_k)$ の値が計算されている。

完了状態 頂点 z_k について、$j \in a[k]$ に対して$\pi_{j\to k}(z_j)$ の値が、$j \in b[k]$ に対して$\lambda_{j\to k}(z_k)$ の値が計算されている。

の 2 状態を定義する。準備状態になったときに、

1. $k \to i$ に矢印があれば、(19) によって $\pi_{k+}(z_k)$ を求めてから、(26) によって $\pi_{k\to i}(z_k)$ を計算する。
2. $i \to k$ に矢印があれば、(20) によって $\lambda_{k-}(z_k)$ を求めてから、(21) によって $\lambda_{k\to i}(z_i)$ を計算する。

ベイジアンネットワークが単一結合であることを仮定しているので、有向非巡回グラフ $\vec{G}$ は有向木であり、方向を無視すれば、無向木ができる (任意の頂点を根とみなしてよい)。したがって、この無向木の葉から頂点に向けて準備状態がすべての頂点に伝播する。また、その無向木の根であれば、必然的に完了状態になるが、どの頂点もその無向木の根になっているので、どの頂点も最後は完了状態になる。z_k が完了状態になれば、(20) によって $\lambda_{k-}(z_k)$ を、(19) によって $\pi_{k+}(z_k)$ を求めてから、(18) によって $BEL(z_k)$ が計算される。

例 **17** $G_8 = (\{z_1, z_2, z_3, z_4, z_5\}, \{(z_1, z_4), (z_2, z_3), (z_3, z_4), (z_3, z_5)\})$ (図 9) で、z_1 に $P(z_1)$ の値が、z_2 に $P(z_2)$ の値が、z_3 に $P(z_3|z_2)$ の値が、z_4 に $P(z_4|z_1, z_3)$ の値が、z_5 に $P(z_5|z_3)$ の値が格納されている。$z_2 = e_2, z_5 = e_5$ が観測されたとき、

1. $\pi_{1\to 4}(z_1) = P(z_1)$
2. $\pi_{2\to 3}(z_2) = \pi_{2+}(z_2) = \begin{cases} 1, & z_2 = e_2 \\ 0, & z_2 \neq e_2 \end{cases}$
3. z_4 はすでに準備状態にある。$\lambda_{4-}(z_4) = 1$ および (21) より、$\lambda_{4\to 3}(z_3) = 1$
4. z_3 はすでに準備状態である。(19) によって、$\pi_{3+}(z_3) = \sum_{z_2} P(z_3|z_2)\pi_{2\to 3}(z_2) = P(z_3|z_2 = e_2)$。(26) より、

$$\pi_{3\to 4}(z_3) = \pi_{3+}(z_3)\lambda_{5\to 3}(z_3)$$
$$= P(z_3|z_2 = e_2)P(z_5 = e_5|z_3)$$

5. z_4 はすでに完了状態にある。
   (a) $\pi_{4+}(z_4) = \sum_{z_1,z_3} P(z_4|z_1, z_3)\pi_{1\to 4}(z_1)\pi_{3\to 4}(z_3) = \sum_{z_3} P(z_4|z_1, z_3)P(z_1)P(z_3|z_2 = e_2)P(z_5 = e_5|z_3)$
   (b) $\lambda_{4-}(z_4) = 1$

$$BEL(z_4) = \sum_{z_1}\sum_{z_3} P(z_1)P(z_4|z_1, z_3) \cdot P(z_3|z_2 = e_2)P(z_5 = e_5|z_3)$$

6. z_3 はすでに完了状態にある。
   (a) $\pi_{3+}(z_3) = P(z_3|z_2 = e_2)$
   (b) $\lambda_{3-}(z_3) = \lambda_{4\to 3}(z_3)\lambda_{5\to 3}(z_3) = P(z_5 = e_5|z_3)$

$$BEL(z_3) = P(z_3|z_2 = e_2)P(z_5 = e_5|z_3)$$

例 **18** 泥棒と警報の問題

泥棒と警報の問題 (the bulglar alarm problem) を用いて、前節の方法を具体的に適用してみよう。

$$\vec{G} = (\{a, b, e\}, \{(b, a), (e, a)\})$$

(図 10) は、泥棒だけでなく地震にも敏感な警報の依存モデル M をベイジアンネットワークで表している。$Z = \{a, b, e\}$ の確率変数は、$a = 1/a = 0$ が警告 (alarm) あり/なし、$b = 1/b = 0$ が泥棒 (burglary) あり/なし、

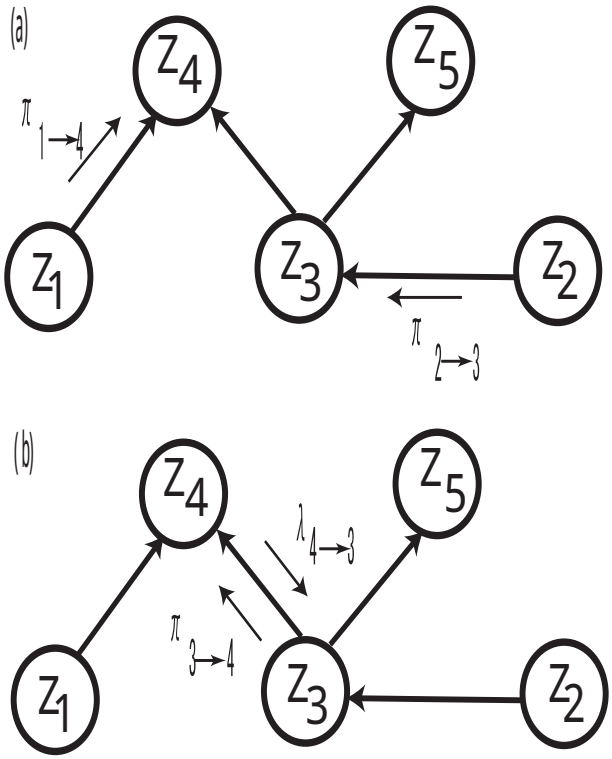

図 9: 例 17 におけ確率伝播

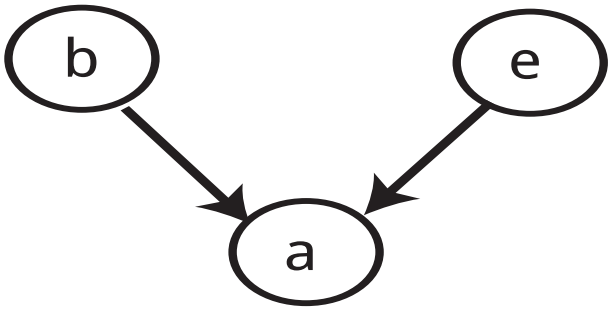

図 10: 泥棒と警報の問題

$e = 1/e = 0$ が地震 (earthequake) あり/なしを表している。以下のような確率分布が与えられているとする

$$P(b = 1) = 0.1, P(e = 1) = 0.1,$$

$$P(a = 1|b = 0, e = 0) = 0.001,$$

$$P(a = 1|b = 1, e = 0) = 0.368,$$

$$P(a = 1|b = 0, e = 1) = 0.135,$$

$$P(a = 1|b = 1, e = 1) = 0.607$$

そして、学会に参加していて携帯電話から警報のみを受け取った ($a = 1$ の観測) ときに、どのような確率推論をすべきであろうか。これを前節の確率伝播によって求めてみよう。

実は、例 15 に数値を代入すれば、直ちに解が得られる。$z_1 = a, z_2 = b, z_3 = e$ として、

$$\begin{pmatrix} BEL(z_2 = 0) \\ BEL(z_2 = 1) \end{pmatrix}$$
$$= \begin{pmatrix} P(z_2 = 0) \sum_{z_3} P(z_1 = 1|z_2 = 0, z_3)P(z_3) \\ P(z_2 = 1) \sum_{z_3} P(z_1 = 1|z_2 = 1, z_3)P(z_3) \end{pmatrix}$$
$$= \begin{pmatrix} 0.249 \\ 0.751 \end{pmatrix}$$

$$\begin{pmatrix} BEL(z_3 = 0) \\ BEL(z_3 = 1) \end{pmatrix}$$
$$= \begin{pmatrix} P(z_3 = 0) \sum_{z_2} P(z_1 = 1|z_3 = 0, z_2)P(z_2) \\ P(z_3 = 1) \sum_{z_2} P(z_1 = 1|z_3 = 1, z_2)P(z_2) \end{pmatrix}$$
$$= \begin{pmatrix} 0.651 \\ 0.349 \end{pmatrix}$$

3.2 ファクターネットワークによる確率伝播

Brendan J. Frey らが考案したファクターネットワークによる確率伝播を紹介し、オリジナルのベイジアンネットワークを用いる方法との関連を示す。

$X = \{x_1, \cdots, x_N\}$ を変数の集合、$X_j, j = 1, \cdots, J$ を X の空でない部分集合とする。また、f_j を X_j に含まれる変数の関数とする。このとき、変数 $x_1, \cdots, x_N$ の関数 $f(x_1, \cdots, x_N)$ は、各 f_j の定数倍を除いて以下のように一意にかける。

$$f(x_1, \cdots, x_N) \propto \prod_{j=1}^{J} f_j(x_j) \tag{27}$$

各 f_j は f の局所関数とよばれる。N 個の変数 $x_1, \cdots, x_N$ および J 個の局所関数 $f_1, \cdots, _J$ を要素にもつ有限集合 $V = \{x_1, \cdots, x_N, f_1, \cdots, f_N\}$ と

$$\{x_i, f_j\} \in E \Leftrightarrow x_i \in X_j$$

なる有限集合 E との対 (無向グラフ)$G = (V, E)$ を考えることができる。これを $f(x_1, \cdots, x_N)$ のファクターグラフという。

以下では、X を確率変数の集合 $Z = \{z_1, \cdots, z_N\}$ の確率分布 $P(z_1, \cdots, z_N)$ のファクターグラフを考える。

例 19 確率変数 $Z = \{z_1, \cdots, z_4\}$ が

$$P(z_1, z_2, z_3, z_4) = P_A(z_1)P_B(z_2|z_1)P_C(z_3|z_1, z_2)P_D(z_4|z_3)$$

の分布にしたがうとする。

$$f(z_1, z_2, z_3, z_4) = P(z_1, z_2, z_3, z_4),$$

$$f_A(z_1) = P_A(z_1), f_B(z_1, z_2) = P_B(z_2|z_1),$$

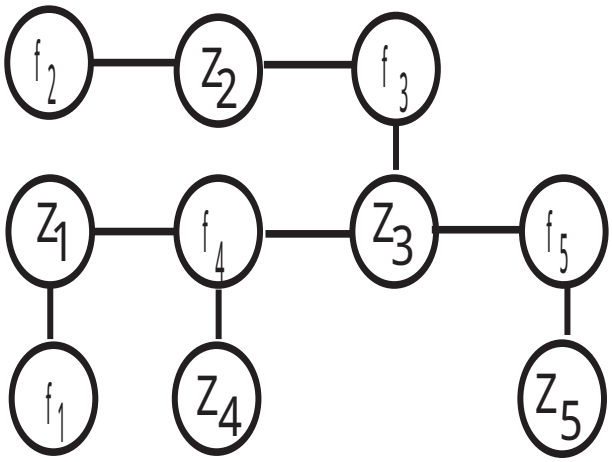

図 11: 例 19

$f_C(z_1,z_2,z_3)=f_C(z_3|z_1,z_2), f_D(z_3,z_4)=P_D(z_4|z_3)$

とおけば、$f(z_1,z_2,z_3,z_4)$ のファクターグラフは、図 11 のように与えられる。

そもそも、$\{z_j=e_j\}$ から z_k の確率分布を求める計算は、ベイジアンネットの置換 π が与えられいなくても、確率分布 $P(z_1,\cdots,z_N)$ だけから計算することができる。

例 **20** 例 19 で、$u=e$ が観測されたときの、z の確率分布は、

$$\sum_{z_1}\sum_{z_2} f_A(z_1)f_B(z_2)f_C(z_1,z_2,z_3)f_D(z_3,z_4=e_4)$$

を計算すればよい。

単一結合の場合、この計算が容易に行われる。

1. $f_A(z_1), f_B(z_2)$ はそれぞれ z_1, z_2 からなるの唯一の関数なので $\sum_{z_1}\sum_{z_2} f_A(z_1)f_B(z_2)f_C(z_1,z_2,z_3)$ から z のみの関数を得る。
2. $\sum_{z_1}\sum_{z_2} f_A(z_1)f_B(z_2)f_C(z_1,z_2,z_3)$ および $f_D(z_3,z_4=e_4)$ は z_3 のみの関数であり、両者の積が最終的な解である。

これを一般化すると、

ルール **1:** 1 変数 z_2 のみからなる関数 $\mu_{2\to A}(z_2)$、1 変数 z_3 のみからなる関数 $\mu_{3\to A}(z_3)$、および関数 $f_A(z_1,z_2,z_3)$ から、

$$\lambda_{A\to 1}(z_1)=\sum_{z_2}\sum_{z_3} f_A(z_1,z_2,z_3)\pi_{2\to A}(z_2)\pi_{3\to A}(z_3) \tag{28}$$

を計算する (変数 z_2, z_3 は任意個 (0 個でも可) の変数でよい)。

ルール **2:** 変数 z_1 のみからなる 2 関数 $\pi_{B\to z_1}(z_1)=f_B(z_1)$ と $\pi_{C\to z_1}(z_1)=f_C(z_1)$ から、

$$\pi_{1\to A}(z_1)=\lambda_{B\to 1}(z_1)\lambda_{C\to 1}(z_1) \tag{29}$$

を計算する (関数 f_B, f_C は任意個の関数でよい)。

単一結合ベイジアンネットワークであれば、辺の方向を無視して無向ファクターグラフを求め、その局所関数とルール 1,2 と以下のルール 3 を加えて、$\{z_j=e_j\}$ の観測のもとでの z_k の確率分布を計算することができる。

ルール **3:** (29) で z_1 の値 (e_1 とする) が観測されれば、

$$\pi_{1\to A}(z_1)=\begin{cases}1, & z_1=e_1\\ 0, & z_1\neq e_1\end{cases} \tag{30}$$

例 **21** 例 17 のベイジアンネットワークをのファクターグラフを用いて計算する。

$$f_A(z_1)=P_A(z_1), f_B(z_2)=P_B(z_2), f_C(z_2,z_3)=P_C(z_3|z_2),$$

$$f_D(z_1,z_3,z_4)=P_D(z_4|z_1,z_3), f_E(z_3,z_5)=P_E(z_5|z_3)$$

1. ルール 2 より、$\pi_{1\to D}(z_1)=\lambda_{A\to 1}(z_1)=P(z_1)$
2. ルール 3 より、$\pi_{2\to C}(z_2)=\begin{cases}1, & z_2=e_2\\ 0, & z_2\neq e_2\end{cases}$
3. ルール 2 より、$\pi_{4\to D}(z_1)=1$。ルール 1 より、$\lambda_{D\to 3}(z_3)=\sum_{z_1}\sum_{z_4} f_D(z_1,z_3,z_4)\pi_{1\to D}(z_1)\pi_{4\to D}(z_1)=1$。
4. ルール 1 より、$\lambda_{C\to 3}(z_3)=\sum_{z_2} f_C(z_2,z_3)\pi_{2\to C}(z_2)=P(z_3|z_2=e_2)$。ルール 2 より、$\pi_{3\to D}(z_3)=\lambda_{C\to 3}(z_3)$ $\lambda_{E\to 3}(z_3)=P(z_3|z_2=e_2)P(z_5=e_5|z_3)$

5(a) ルール 1 より、$\lambda_{D\to 4}(z_4)=\sum_{z_1}\sum_{z_3} f_D(z_1,z_3,z_4)$ $\pi_{1\to D}(z_1)\pi_{3\to D}(z_3)=\sum_{z_1}\sum_{z_3} P(z_4|z_1,z_3)$ $P(z_1)P(z_3|z_2=e_2)P(z_5=e_5|z_3)$

(b) ルール 2 より、$\pi_{4\to D}(z_4)=1$。
ルール 1 より、$BEL(z_4)=\lambda_{D\to 4}(z_4)\pi_{4\to D}(z_4)=\sum_{z_1}\sum_{z_3} P(z_4|z_1,z_3)P(z_1)P(z_3|z_2=e_2)P(z_5=e_5|z_3)$

6(a) 4. より、$\lambda_{C\to 3}(z_3)=P(z_3|z_2=e_2)$

(b) 3. より、$\lambda_{D\to 3}=1$。ルール 3 より、$\pi_{5\to E}(z_5)=\begin{cases}1, & z_5=e_5\\ 0, & z_5\neq e_5\end{cases}$。ルール 1 より、
$\lambda_{E\to 3}(z_3)=\sum_{z_2} f_E(z_3,z_5)\pi_{5\to E}(z_5)=P(z_5=e_5|z_3)$。ルール 2 より、$\pi_{3\to C}(z_3)=\lambda_{D\to 3}(z_3)$ $\lambda_{E\to 3}(z_3)=P(z_5=e_5|z_3)$
ルール 2 より、$BEL(z_3)=\lambda_{D\to 3}(z_3)\pi_{3\to D}(z_3)=P(z_3|z_2=e_2)P(z_5=e_5|z_3)$

ルール 3 がなぜ必要かという理由は直感的に理解されるかもしれないが、ここではオリジナルのベイジアンネットワークによる確率伝播と対比させることにより、意味を明確にしておこう。

そのために、オリジナルの方法で $\pi_{k+}(z_k)$ および $\lambda_{k-}(z_k)$ の記法をやめ、それぞれ $\lambda_{k\to k}(z_k)$ および $\pi_{k\to k}(z_k)$ とい

うように、また $a^*[k] = a[k] \cup \{k\}$ および $b^*[k] = b[k] \cup \{k\}$ と書くものとする。このとき、(19)(20)(21)(26) は、

$$\lambda_{i\to k}(z_i) \propto \sum_{z_k} \sum_{z_j, j\in a^*[k]-\{i\}} P(z_k|z_j, j \in a^*[k]) \cdot \prod_{l\in a^*[k]-\{i\}} \pi_{l\to k}(z_l) \tag{31}$$

$$\pi_{k\to i}(z_k) \propto \prod_{j\in b^*[k]-\{i\}} \lambda_{k\to j}(z_k) \tag{32}$$

と等価である。特に、ファクターグラフの方で、関数名の添え字 $A, B, \cdots$ をやめて、$f_A = P(z_1|\cdots), f_B = P(z_2|\cdots), \cdots$ のときに、$1, 2, \cdots$ と書けば、(31)(32) とルール 1,2 は一致する。

したがって、ベイジアンネットワークで z_k の値が観測されたときに、$i \in b[k]$ として、

$$\pi_{k\to i}(z_k) = \begin{cases} 1, & z_k = e_k \\ 0, & z_k \neq e_k \end{cases}$$

とおくことと、ルール 3 は等価である。

参考文献

[1] Judea Pearl, *Probablistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, 1988.

[2] Brendan J. Frey, *Graphical Models for Machine Learning and Digital Communication*, MIT Press, 1998.

[3] Thomas Verma, "Causal Networks: Semantics and Expressiveness", Technical Report, R-65-I (Cognitive Systems Laboratory, UCLA), June 1987.