グラフィカルモデルの基礎

統計数理研究所　数理・推論研究系

坂田 綾香

ayaka@ism.ac.jp

2021年度「計算推論基礎」講義資料

目次

参考文献

1. 「確率的グラフィカルモデル」共立出版　１章・２章　⏩
    - ✦ グラフィカルモデルに関連した分野の話題を概観できる
2. 「パターン認識と機械学習(下)」Springer Japan, 8章
    - ✦ 機械学習の一手法として説明している点が特徴的
3. Pearl, “Probabilistic Reasoning in Intelligent Systems” (1988) 3章
    - ✦ AI研究に必要な知識の一つとして扱っている
    - ✦ 現在では一部否定されている話も載っているので要注意
4. Koller & Friedman, “Probabilistic Graphical Models” (2009) 3章・4章
    - ✦ グラフィカルモデルに関する推定や学習を詳しく扱っている

☆ 1, 3, 4は統数研図書室に注文中です ☆

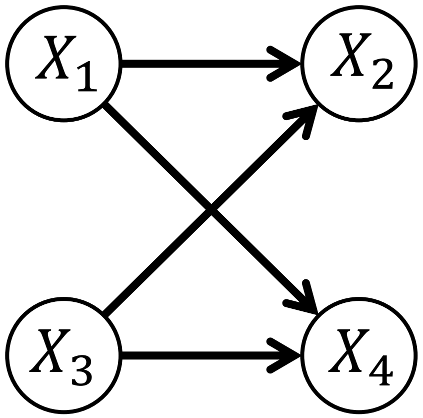
1. はじめに

グラフィカルモデルとは

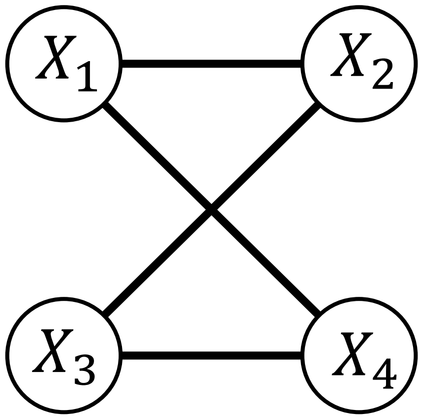
変数間の関係をグラフとして表したもの

- ✦ 4つの確率変数X_1, X_2, X_3, X_4に対するグラフィカルモデルの例

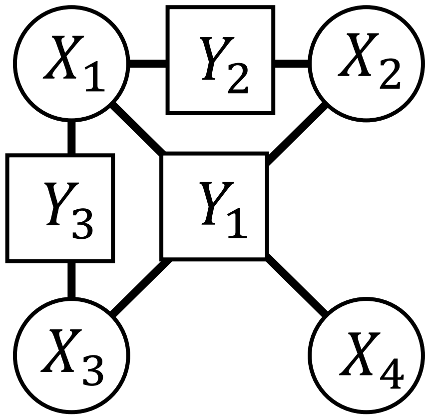
有向グラフ



無向グラフ



因子グラフ
（ファクターグラフ）



※ 今回の講義では因子グラフは扱いませんが，
Mezard & Montanari (2009)などを参照してください．

グラフィカルモデルを使う目的

1. 変数間の依存関係を視覚的に捉えられる
   - 独立性，条件付き独立性
   - モデル(確率分布)の設計に役立つ
2. グラフ上での推論を行うことができる
   - 構造学習
   - 因果推論
3. グラフの構造を活かした近似計算ができる
   - 確率伝搬法など

この講義の流れと目標

- 1回目と2回目：グラフィカルモデルの基礎
- 3回目：グラフィカルモデルにおける変数消去
- 4回目：グラフィカルモデルにおけるアルゴリズム (確率伝搬法)

■ 条件付き確率の推定，構造学習，因果推論などの話題は扱いませんが，
この講義を経て「確率的グラフィカルモデル」の後ろの章や
“Probabilistic Graphical Models”の他の章を
自分で読めるようになっていることが目標

1.1 条件付き確率の復習

確率分布による条件付き独立性の表現

定義：独立性

- 確率変数X, Yの同時確率分布を$p(X, Y)$とする
- 周辺化分布を$p(X) = \sum_Y p(X, Y)$, $p(Y) = \sum_X p(X, Y)$とする
- 「確率変数XとYが独立である」ということを$X \perp\!\!\!\perp Y$と表記する

確率変数の独立性は，次のように表現される

$$X \perp\!\!\!\perp Y \iff p(X, Y) = p(X)p(Y)$$

※ $\perp\!\!\!\perp$はDawid's notationと呼ばれますが
powerpointの記号に入っていないので，
'¥parallel'に下線をひいて表現しています．

定義：条件付き独立性

- 条件付き確率を次のように定義する

$$p(X|Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X,Y)}{\sum_X p(X,Y)}, \qquad p(Y|X) = \frac{p(X,Y)}{p(X)} = \frac{p(X,Y)}{\sum_Y p(X,Y)}$$

- 「Zが所与のもとでXとYが条件付き独立である」を$X \perp\!\!\!\perp Y|Z$と表す

$$X \perp\!\!\!\perp Y|Z \Leftrightarrow p(X,Y|Z) = p(X|Z)p(Y|Z) \qquad \forall z \text{ with } p(z) > 0$$

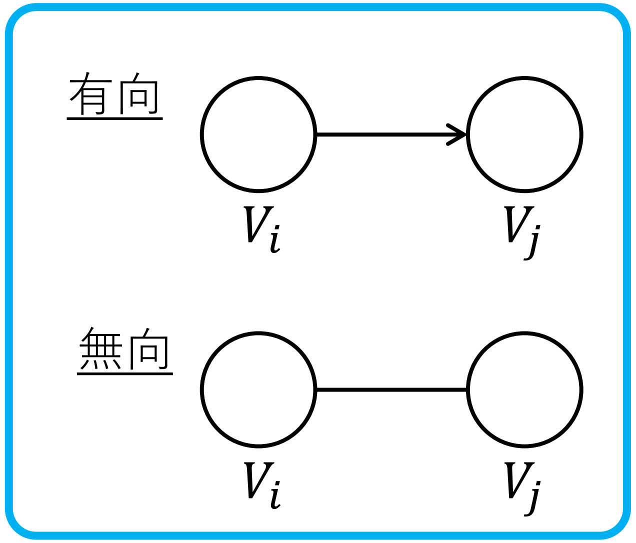
$$X \perp\!\!\!\perp Y|Z \Leftrightarrow p(X|Y,Z) = p(X|Z) \qquad \forall (y,z) \text{ with } p(y,z) > 0$$

- このStatementを$I(X,Y|Z)_p$と表記する.

1.2 グラフィカルモデルとは

グラフィカルモデルとは

- グラフを$G = (\mathcal{V}, \mathcal{E})$と表記する
  - $\mathcal{V}$: node(頂点)集合…確率変数を表す
  - $\mathcal{E}$: edge(辺)集合…確率変数の依存関係を表す
    - V_iとV_jの間にedgeが存在すればE_{ij}と書くことにする
    - 有向edge: $E_{ij} \in \mathcal{E},\ \ E_{ji} \notin \mathcal{E}$
    - 無向edge: $E_{ij} \in \mathcal{E},\ E_{ji} \in \mathcal{E}$



- 全てのedgeが有向edgeのグラフを有向グラフ(directed graph)と呼ぶ

  …有向グラフと確率分布のセットを**ベイジアンネットワーク**(BN)モデルという

- 全てのedgeが無向edgeのグラフを無向グラフ(undirected graph)と呼ぶ

  …無向グラフと確率分布のセットを**マルコフネットワーク**(MN)モデルという

2. ベイジアンネットワーク

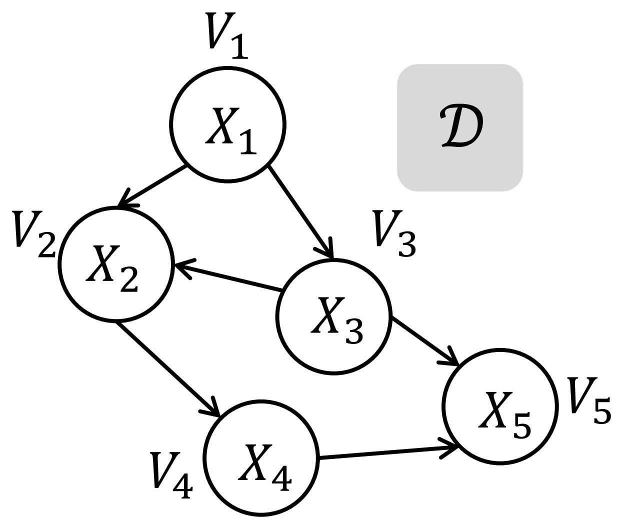
ベイジアンネットワークとは

N個の変数集合$\boldsymbol{X} = \{X_1, \dots, X_N\}$をもつベイジアンネットワークモデルは$(\mathcal{D}, p_{\mathcal{B}})$で表現される

- ✦ $\mathcal{D}$: ノード集合と有向エッジにより構成される**DAG**
  - Bayesian network structureとよぶ
- ✦ $p_{\mathcal{B}}(\boldsymbol{X}; \mathcal{D}, \Theta)$: Underlying distribution
  - DAG $\mathcal{D}$により指定されるルールで条件付き確率の集合Θを掛け合わせたもの

※ 今日の内容ではΘについてはほとんど言及しないので省略

- 全ての可能な確率分布を$\mathcal{P}$とすると$p_{\mathcal{B}} \subset \mathcal{P}$
  - ✦ BNは一部の確率分布しか表現できない

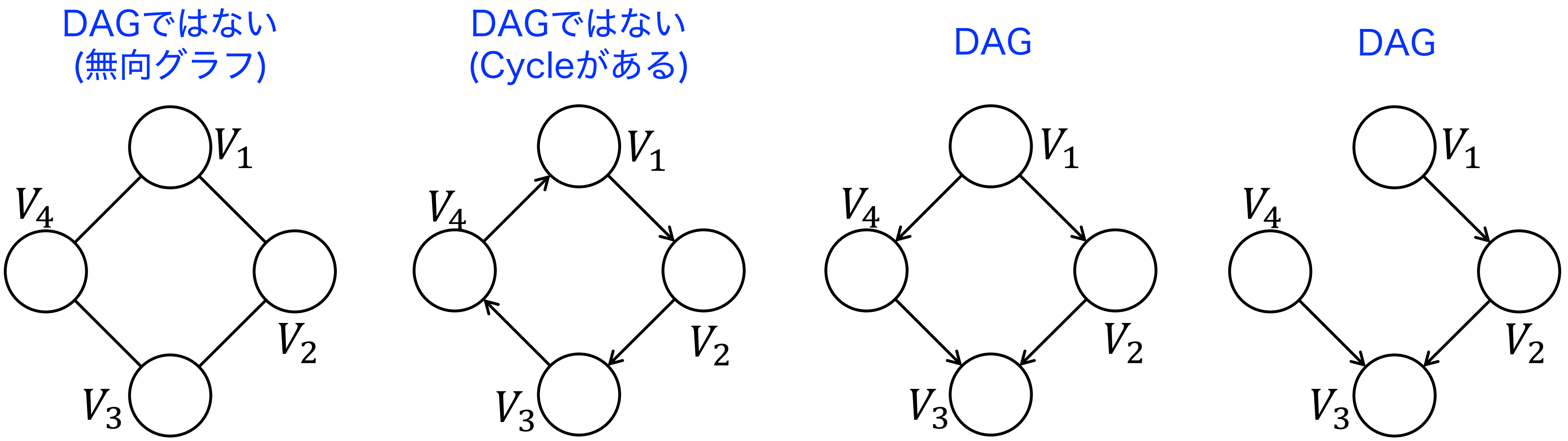


$$p_{\mathcal{B}}(\boldsymbol{X}; \mathcal{D}, \Theta) = p\left(\boldsymbol{X}_i | \boldsymbol{X}_{\pi(i)}\right)$$

- $\pi(i)$: iの**親ノード**集合（$\mathcal{D}$に依存する）

DAG

閉路のない有向グラフのことを
Directed Acyclic Graph (有向非巡回グラフ，DAG)とよぶ.



DAGはDirected Local Markov Independence
と呼ばれる独立性を表現する

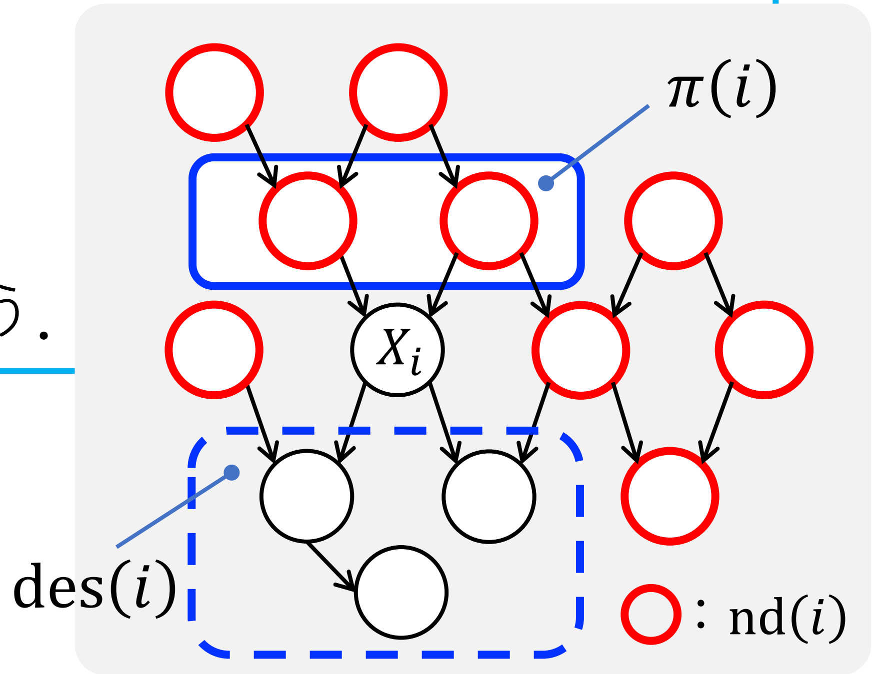
定義：Directed Local Markov Independence

- ノードV_iから矢印を辿って到達できるノード全てを**子孫ノード**とよび，$\mathrm{des}(i)$と表記する．
- ノードV_iの**非子孫ノード**を$\mathrm{nd}(i) = V \setminus \big(i \cup \mathrm{des}(i)\big)$と定義する．

Bayesian network structure $\mathcal{D}$は，$i \in V$について

独立性の仮定$i \perp\!\!\!\perp \mathrm{nd}(i) | \pi(i)$を表現している．

これをDirected Local Markov Independenceという．



✦ DAG $\mathcal{D}$において$i \perp\!\!\!\perp \mathrm{nd}(i)|\pi(i)$が表現されることを $I\big(i \perp\!\!\!\perp \mathrm{nd}(i)|\pi(i)\big)_{\mathcal{D}}$と表記する．

定義：Factorized form

ベイジアンネットワークにおけるUnderlying distributionは
次のように表現される

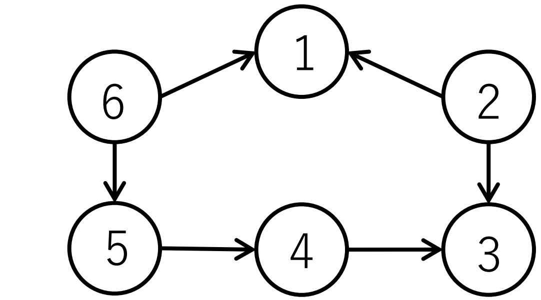
$$p_{\mathcal{B}}(\boldsymbol{X}; \mathcal{D}, \Theta) = \prod_{i=1}^{N} p\left(X_i | \boldsymbol{X}_{\pi(i)}\right)$$

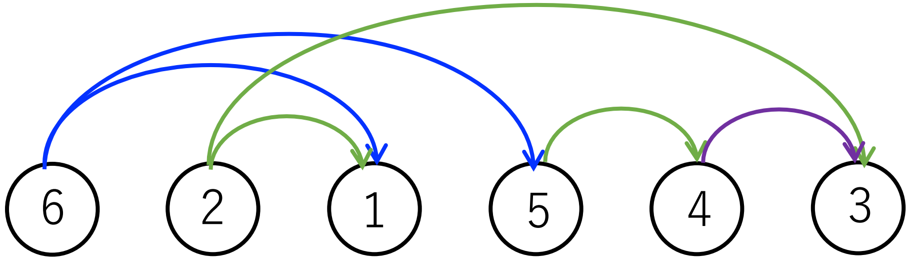
- 確率分布のChain ruleに
Directed Local Markov Independenceを適用したものと解釈できる

※ ここで**Topological Sorting**を導入する

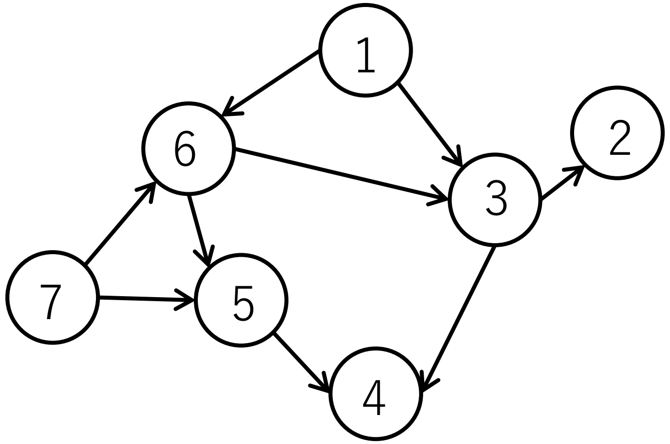
準備：Topological sorting

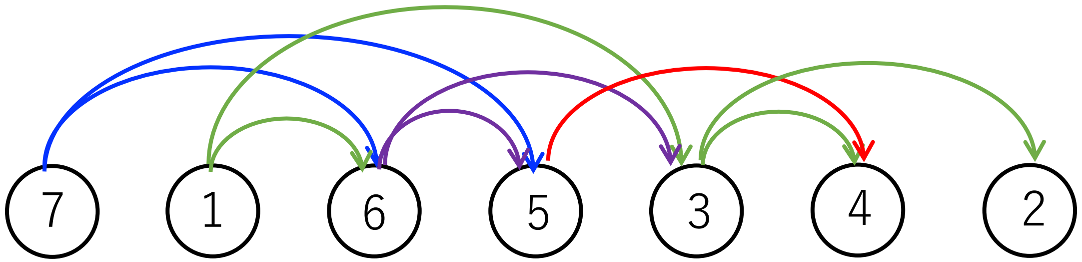
- DAGに対して定義されるノードの並べ方
  - ✦ u, vの順に並んでいるとき，uは常にvより上流になければならない





※ 2, 6, 5, 4, 3, 1でもよい





※ 1, 7, 6, 3, 5, 2, 4などでもよい

定義：Factorized form

ベイジアンネットワークにおけるUnderlying distributionは
次のように表現される

$$p_{\mathcal{B}}(\boldsymbol{X}) = \prod_{i=1}^{N} p\left(X_i | \boldsymbol{X}_{\pi(i)}\right)$$

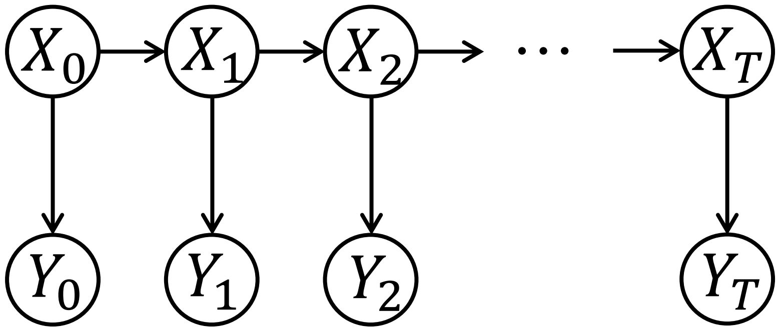
- 確率分布のChain ruleに
  Directed Local Markov Independenceを適用したものと解釈できる
  - ✦ 変数がTopological sortingにより並んでいるとき，Chain ruleより

$$P(\boldsymbol{X}) = p(X_N | X_1, \dots, X_{N-1}) p(X_{N-1} | X_1, \dots, X_{N-2}) p(X_1, \dots, X_{N-2})$$

$$= \prod_{i=1}^{N} p(X_i | X_1, \dots, X_{i-1}) \xrightarrow[\substack{i \perp\!\!\!\perp \mathrm{nd}(i) | \pi(i) \\ \pi(i) \subseteq \{X_1, \dots, X_{i-1}\}}]{} p_{\mathcal{B}} = \prod_{i=1}^{N} p\left(X_i | \boldsymbol{X}_{\pi(i)}\right)$$

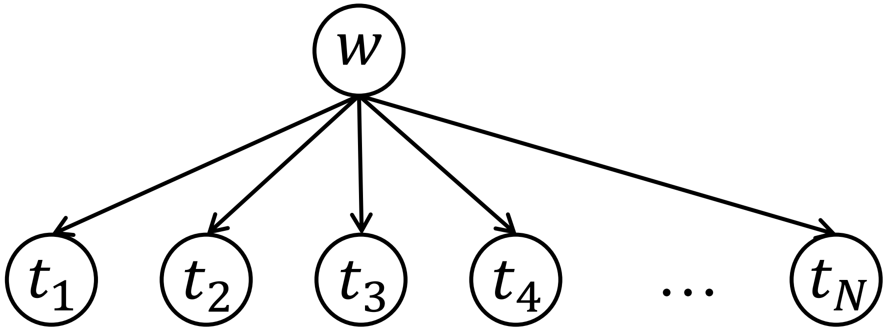
ベイジアンネットワークモデルの例

隠れマルコフモデル (Hidden Markov Model)



$$P(\boldsymbol{X}, \boldsymbol{Y}) = p(X_0) p(Y_0|X_0) \prod_{t=1}^{T} p(X_t|X_{t-1}) p(Y_t|X_t)$$

ベイズ多項式回帰モデル



$$p(t, w) = \prod_{i=1}^{N} p(t_i|w)$$

d分離性

Local Markov Independence以外に，グラフから読み取れる独立性はあるか？

- **d分離性**：DAG $\mathcal{D}$において独立性を定義する
  - DAG $\mathcal{D}$のもとで，Local Markov Independenceに基づいて Factorizeされた確率分布$p_{\mathcal{B}}$が，d分離性により与えられる条件付き独立性を示すことを見ていく

2.1 d分離性

グラフの構造から条件付き独立性を読み取る

d分離 (有向分離, d-separation)

- グラフ構造から条件付き独立性を定義するための概念
  - ✦ dは”directional”の略
- d分離は，有向グラフにおいて3つの結合の中で定義される

(1) **逐次結合** (serial connections)

(2) **分岐結合** (diverging connections)

(3) **合流結合** (converging connections)

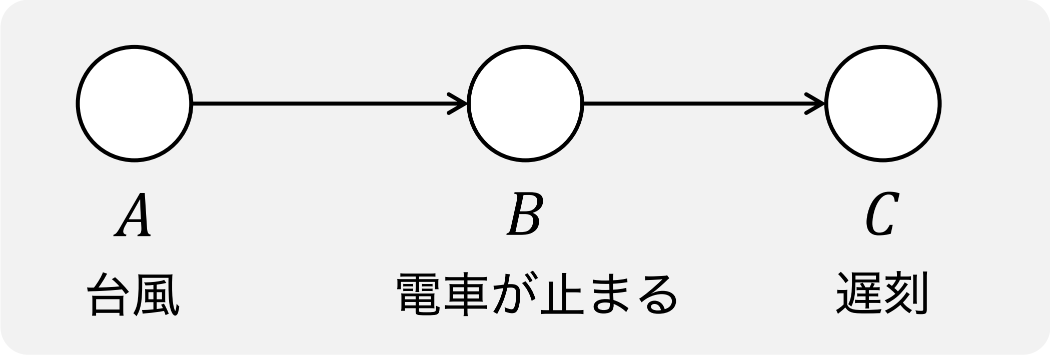
(1) 逐次結合

「ノードAとCは，Bがインスタンス化されたときd分離である」
と言い，これを$I(A,C|B)_D$と表記する．

✦変数の状態がわかることを「インスタンス化された」という

- 確率変数の実現値が得られた，ということと同じ

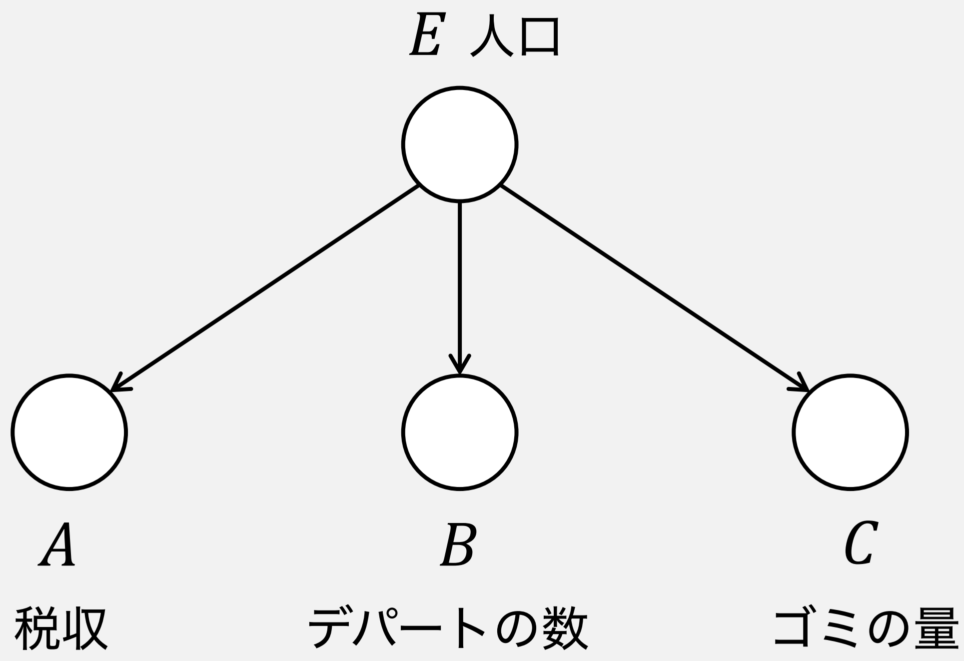
✦ノードBの状態がわかっているのであれば，ノードCの状態を知る上でノードAの情報は必要ない



(2) 分岐結合

「ノードA, B, Cは，ノードEがインスタンス化されたとき
d分離である」と言い，これを$I(A, B, C|E)_{\mathcal{D}}$と表記する．

- Eが分かれば、A, B, Cをそれぞれ知る上で他の変数の情報は必要ない



- ノードA, B, Cを
ノードEの**子ノード**とよぶ

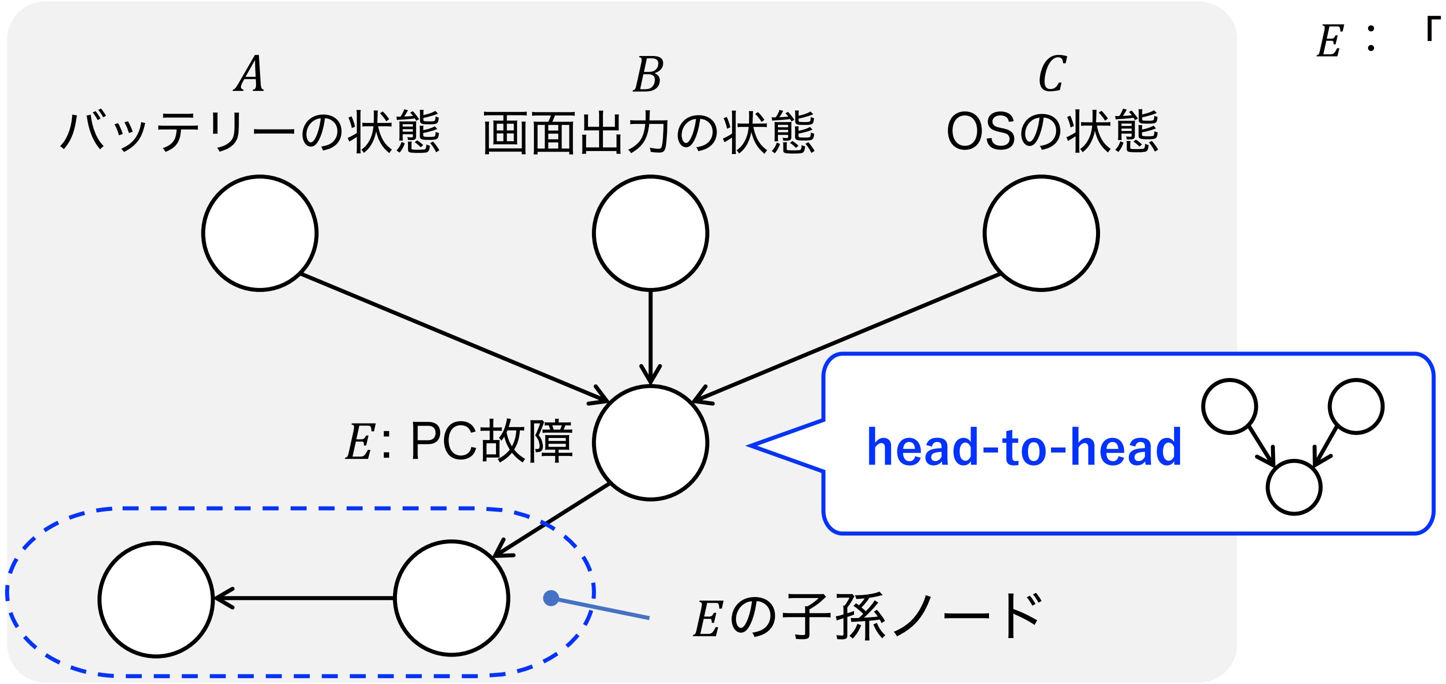
※ 子ノード⊆子孫ノード

- ノードEは
ノードA, B, Cの親ノード

(3) 合流結合

「ノードA,B,Cは，ノードEまたはEの子孫ノードが
インスタンス化されなければ，ノードEを介してd分離である」という．

- 合流結合，head-to-head nodeはDAGの特徴



E：「PC故障」についてYesという情報を得る

⬇

A,B,Cの何処かに原因がある

⬇

Aは「異常なし」

⬇

B,Cに異常がある可能性が高まる

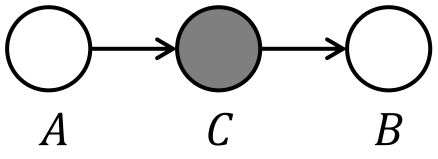
d分離性まとめ

- A, B, CをDAG $\mathcal{D}$におけるノードのdisjoint subsetとする
- DAG $\mathcal{D}$ の有向Edgeを無向Edgeにしたグラフを**Skelton**と呼び$S[\mathcal{D}]$と表す

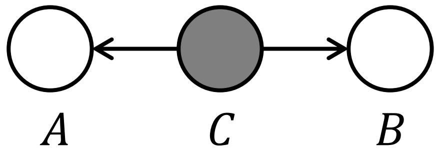
$S[\mathcal{D}]$ 上での$A-B$間の全経路が次の条件を満たすとき，
CはAとBをd分離するといい，$I(A,B|C)_{\mathcal{D}}$と表記する．

(1) 経路上のhead-to-head以外のnodeがCに含まれる

逐次結合

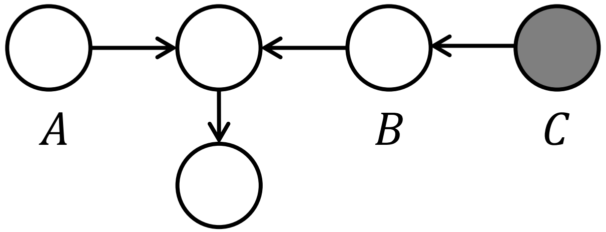


分岐結合



(2) 経路上のhead-to-head nodeとその子孫ノードがCに含まれない

合流結合



経路をブロックする

d分離の言い換え

次の条件が成立するとき，ある経路は**ブロック**されているという

(a) 経路上にhead-to-head nodeがなく，
経路上の変数がインスタンス化されている

(b) 経路上にhead-to-head nodeがあり，
そのnodeおよび子孫ノードがインスタンス化されていない

- あるノードからあるノードへ情報が伝わらない，
  という意味で「ブロック」と呼ぶ

補足：エビデンス

変数がインスタンス化されたときの値を**エビデンス**という

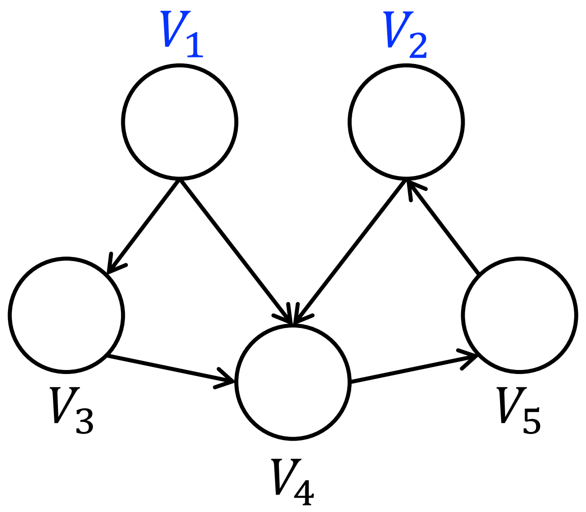
- 値がわかっている場合をハードエビデンス,
  値を取る確率がわかっている場合をソフトエビデンスと呼ぶ
  - ✦ d分離性の議論では，ハードエビデンスを考える
    - 経路をブロックするにはハードエビデンスが必要

d分離の例

- SkeltonにおけるV_1からV_2へ全ての経路を考える

✦ 経路：

- $V_1 \rightarrow V_4 \leftarrow V_5$: 合流結合
- $V_1 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5 \rightarrow V_2$: 逐次結合



✦ V_1とV_2がd分離である条件：

- V_4がインスタンス化されていない
  - ➢ $I(V_1, V_2|C)_D \ \ (C \subset V \setminus \{V_1, V_2\},\ V_4 \notin C)$
- V_3, V_5のいずれかがインスタンス化されている
  - ➢ $I(V_1, V_2|V_3)_D,\ I(V_1, V_2|V_5)_D$

※ $A \setminus B$は，集合Aから集合Bの要素を除いた要素の集合，という意味

d分離の例

- SkeltonにおけるV_1からV_2へ全ての経路を考える

✦ 経路：

- $V_1 \rightarrow V_3 \rightarrow V_7 \leftarrow V_4 \rightarrow V_8 \leftarrow V_6 \leftarrow V_2$
- $V_1 \rightarrow V_3 \rightarrow V_7 \leftarrow V_4 \leftarrow V_5 \rightarrow V_6 \leftarrow V_2$

✦ V_1とV_2がd分離である条件：

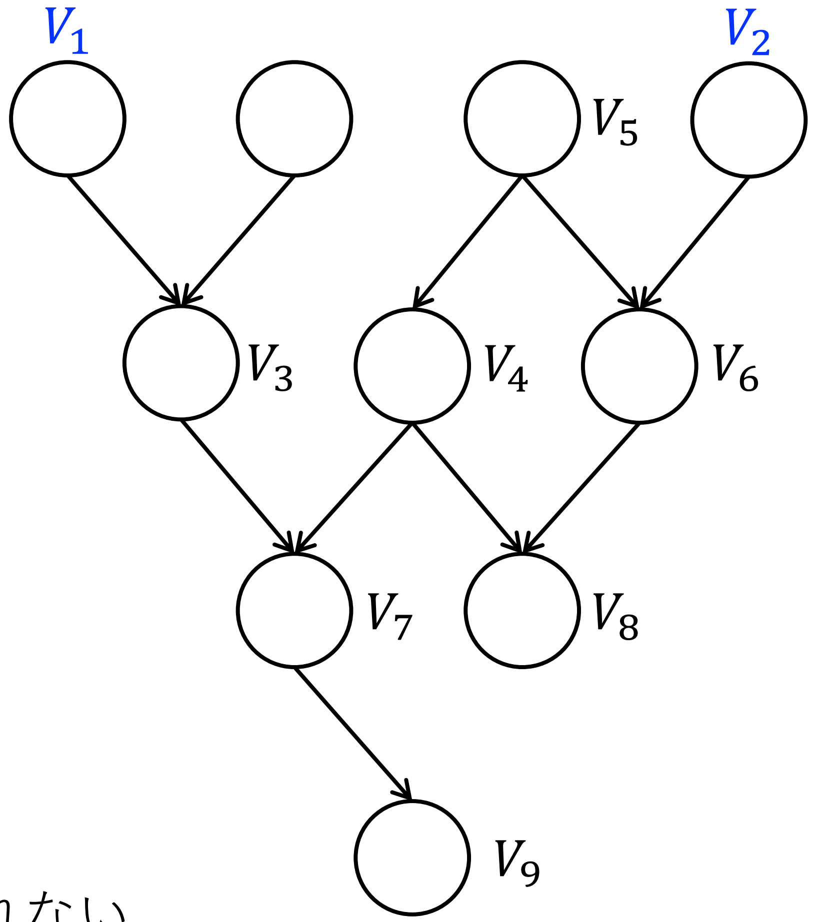
- V_7, V_9がインスタンス化されない
  - ➢ $I(V_1, V_2|C)\ \ C \subset V, \{V_7, V_9\} \notin C$

または

- V_4がインスタンス化される
  - ➢ $I(V_1, V_2|V_4)_D$

または

- V_2がインスタンス化され，V_7がインスタンス化されない
  - ➢ $I(V_1, V_2|C)\ \ C \subset V, V_2 \in C, V_7 \notin C$



定義：Directed Global Markov Independence

DAG $\mathcal{D}$において，disjointなノードのsubset A, B, Cがあったとする．

A, BがCによりd分離されるとき，

Bayesian network structure $\mathcal{D}$は，$i \in V$について

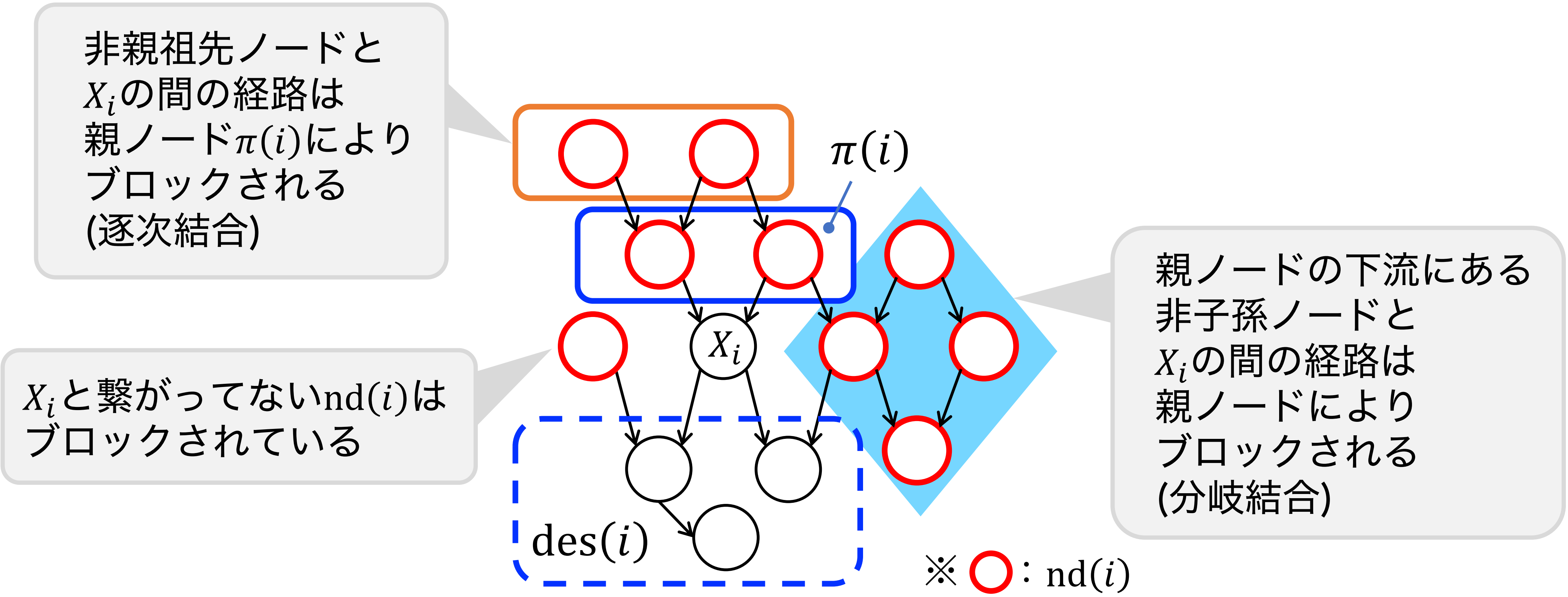
独立性の仮定$A \perp\!\!\!\perp B | C$を表現している．

この独立性をDirected Global Markov Independenceという．

✦ DAG $\mathcal{D}$において$A \perp\!\!\!\perp B | C$が表現されることを $I(A, B | C)_{\mathcal{D}}$と表記する．

Directed Global / Directed Local Independence

- Directed Local IndependenceはDirected Global Independenceの一例
  - $\mathrm{nd}(i) = V \setminus \left(i \cup \mathrm{des}(i)\right)$について$i \perp\!\!\!\perp \mathrm{nd}(i) | \pi(i)$という性質は d分離性から示せる



ここまでのまとめ

- ベイジアンネットワークモデルは$\{\mathcal{D}, p_{\mathcal{B}}\}$により与えられる
  - $\mathcal{D}$：DAG
  - $p_{\mathcal{B}}$：$\mathcal{D}$のunderlying distribution
    - DAGのLocal Markov Independenceに基づいてFactorizeされる

- その他の独立性として，d分離性に基づく
  Global Markov Independenceを導入した.

！ グラフから読み取る条件付き独立性と
確率分布の条件付き独立性の関連は非自明 → これから議論していきます

2.2 独立性の示す性質

Semi-graphoid, Grapoid

独立性のまとめ

- DAGから読み取る独立性：d分離性
  - ✦ CがAとBをd分離しているとき，C所与のもとAとBが独立とする：$I(A,B|C)_{\mathcal{D}}$
  - ✦ DAG $\mathcal{D}$が仮定する条件付き独立性の集合を$\mathcal{I}(\mathcal{D})$と表記する
    - $\mathcal{I}(\mathcal{D}) = \{I(A,B|C)_{\mathcal{D}} | C$はAと$B$をd分離する$\}$
- 確率分布の条件付き独立性
  - ✦ C所与のもとAとBが独立なことを$I(A,B|C)_p$と表記
  - ✦ 確率分布の条件付き独立性の集合を$\mathcal{I}(p)$とおく。
    - $\mathcal{I}(p) = \left\{I(A,B|C)_p | p(A,B|C) = p(A|C)p(B|C)\right\}$

$I(A,B|C)_{\mathcal{D}}$と$I(A,B|C)_p$ の満たす性質は異なる

Semi-graphoid/Graphoid

<u>条件付き独立性に対する5つの性質</u>

a. **Symmetry**：$I(A,B|C) \Leftrightarrow I(B,A|C)$

b. **Decomposition**：$I(A,B \cup D|C) \Rightarrow I(A,B|C)$ and $I(A,D|C)$

c. **Weak Union**：$I(A,B \cup D|C) \Rightarrow I(A,B|C \cup D)$

d. **Contraction**：$I(A,B|C)$ and $I(A,D|C \cup B) \Rightarrow I(A,B \cup D|C)$

e. **Intersection**：$I(A,B \cup D|C)$ and $I(A,B \cup C|D) \Rightarrow I(A,B \cup D|C)$

[Pearl & Paz (1986)]

- a~dを満たす独立性のモデルを**Semi-graphoid**とよぶ
- a~eを満たす独立性のモデルを**Graphoid**とよぶ

Graphical interpretation

※ は"conditioned"を意味する

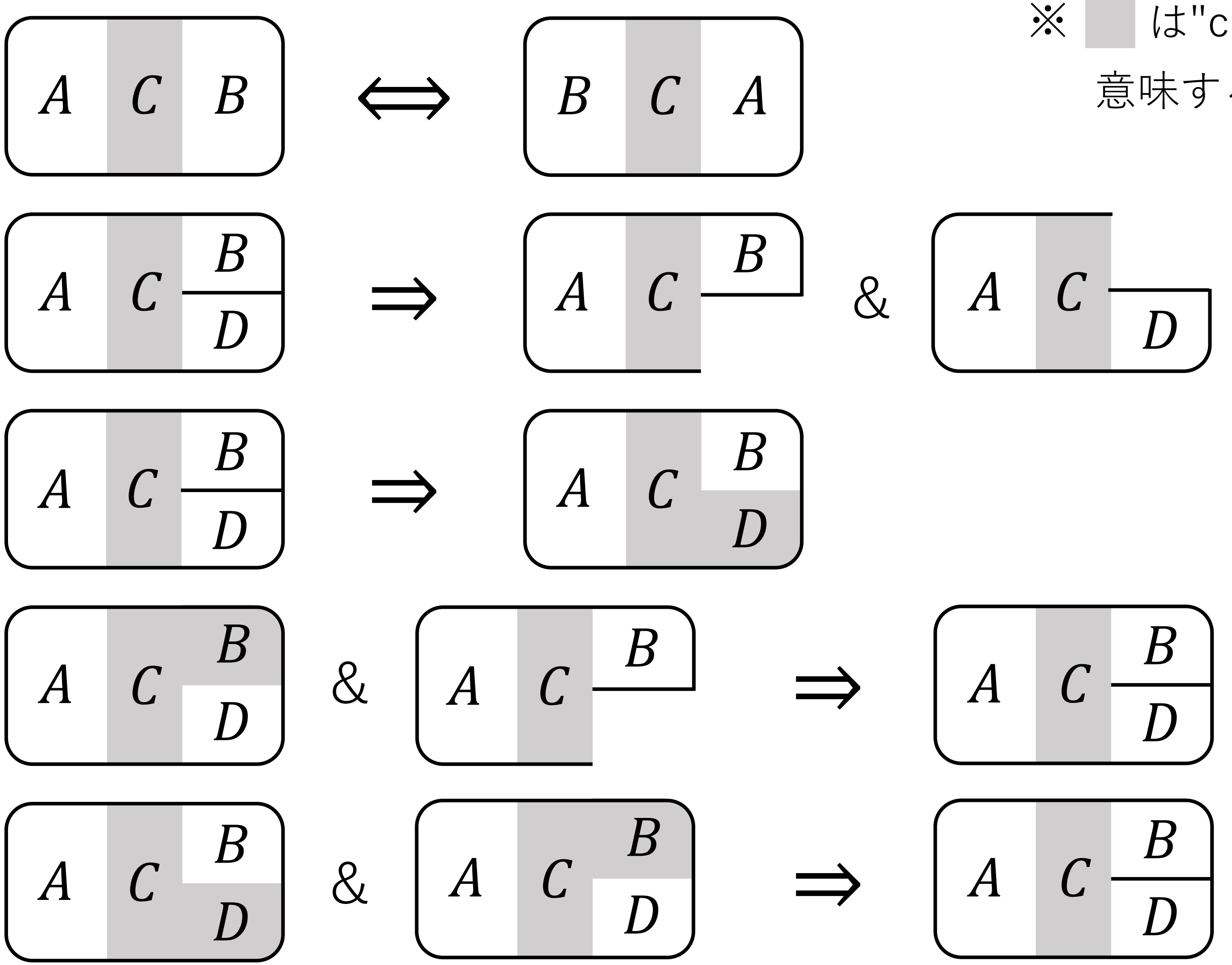
a. Symmetry

b. Decomposition

c. Weak Union

d. Contraction

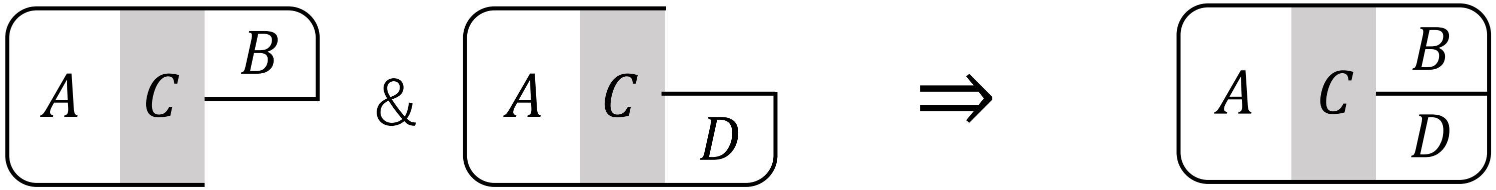
e. Intersection



Graph-induced graphoids

- DAG $\mathcal{D}$においてd分離性から定義された独立性を$I(A,B|C)_{\mathcal{D}}$とする.
- $I(A,B|C)_{\mathcal{D}}$ はGraphoid axiomとComposition propertyを満たすので「Compositional Graphoid」とも呼ばれる

f. Composition ：$I(A,B|C)$ and $I(A,D|C) \Rightarrow I(A,B\cup D|C)$



- Decompositionと合わせると$I(A,B|C)_{\mathcal{D}}$ and $I(A,D|C)_{\mathcal{D}} \Leftrightarrow I(A,B\cup D|C)_{\mathcal{D}}$
- これは次の表現と等価：$I(A,B|C)_{\mathcal{D}} \Leftrightarrow I(a,b|C)_{\mathcal{D}}\ \ \forall a\in A, b\in B$

※ あとで導入する 無向グラフにおける独立性も同じ性質をみたす.

Probabilistic graphoids

- 確率分布による条件付き独立性を$I(A,B|C)_p$と表記する.

$$I(A,B|C)_p \Leftrightarrow p(A,B|C) = p(A|C)p(B|C)$$

- 常にsemi-graphoidであり，strictly positive $(p(\boldsymbol{X}) > 0)$のときはgraphoidである.（証明は付録⏩）

- 一般にはComposition propertyを満たさない
  - ✦ Composition propertyを満たす例として多変量ガウス分布が知られる (付録⏩)

その他の独立性の例：Geometric Orthogonality

ユークリッド空間あるいはヒルベルト空間における

ベクトル$\boldsymbol{L}, \boldsymbol{M}, \boldsymbol{N}$に関する独立性

$$I(\boldsymbol{L}, \boldsymbol{M} | \boldsymbol{N})_O \Leftrightarrow (\boldsymbol{L} \ominus \boldsymbol{N}) \perp (\boldsymbol{M} \ominus \boldsymbol{N})$$

- ここで$\boldsymbol{L} \ominus \boldsymbol{N} = \boldsymbol{L} \cap \boldsymbol{N}^{\perp}$
- $I(\boldsymbol{L}, \boldsymbol{M} | \boldsymbol{N})_O$のとき，$\boldsymbol{L}$と$\boldsymbol{M}$は”meet orthogonally in $\boldsymbol{N}$” という。

- Semigraphoidである
- Intersection propertyは成立しない
- Composition propertyは満たす

補足：完備な公理系による条件付き独立性の表現への試み

$I(A,B|C)$に対する公理系の**完備性(Completeness)**とは：
公理に従えば，全ての可能な条件付き独立性を記述できること

[Pearl (1986)]
- Semi-graphoid/Graphoid axiomsにより全ての確率分布の条件付き独立性を表現できる

[Gelger (1987)]
- $I(A,B|C)_{\mathcal{D}}$に対する完備な公理系が存在しないことを証明

[Studeny (1989)]：Pearl(1986)の否定
- Graphoid/Semi-graphoidで表現できない$I(A,B|C)_p$の性質を発見

[Studeny (1992)]
- $I(A,B|C)_p$に対する完備な公理系が存在しないことを証明

現在では，
公理系としてではなく，
$I(A,B|C)_{\mathcal{D}}, I(A,B|C)_p$が
満たす便利な性質として
(a)-(e)が用いられている．

2.3 DAGと確率分布の条件付き独立性の関係

$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(p_{\mathcal{B}})$の証明

Factorized form再掲：$\mathbb{DF}$

- 一般の同時確率分布で成立するChain rule

$$P(\boldsymbol{X}) = p(X_N|X_1,\dots,X_{N-1})p(X_1,\dots,X_{N-1})$$
$$= p(X_N|X_1,\dots,X_{N-1})p(X_{N-1}|X_1,\dots,X_{N-2})p(X_1,\dots,X_{N-2})$$
$$= \prod_{i=1}^{N} p(X_i|X_1,\dots,X_{i-1})$$

- ベイジアンネットワークモデル

$$p_{\mathcal{B}}(\boldsymbol{X}) = \prod_{i=1}^{N} p\left(X_i|\boldsymbol{X}_{\pi(i)}\right)$$

確率分布$\boldsymbol{p}$が，DAG $\mathcal{D}$に規定される Local Markov Independenceのもと Factorized formで表現されることを $\mathbb{DF}$と表記する．

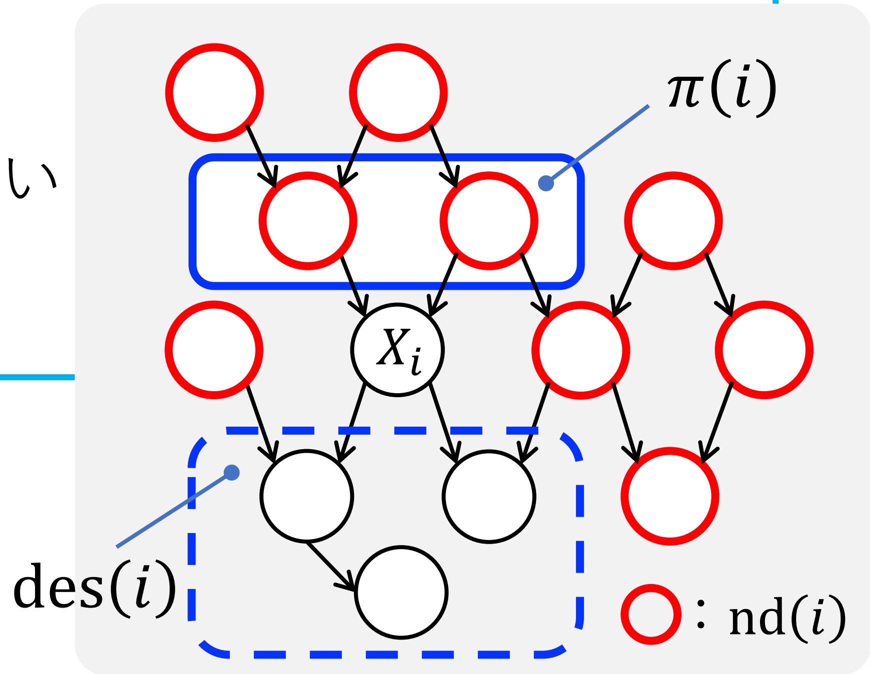
定義：Directed Local Markov Property $\mathbb{DL}$

- ノードV_iから矢印を辿って到達できるノード全てを子孫ノードとよび，$\mathrm{des}(i)$と表記する．
- ノードV_iの非子孫ノードを$\mathrm{nd}(i) = V \setminus \left(i \cup \mathrm{des}(i)\right)$と定義する．

確率分布pが$i \perp\!\!\!\perp \mathrm{nd}(i) | \pi(i)\ \forall i \in V$を満たすとき，

pは**Directed local Markov property**を満たすといい

$\mathbb{DL}$と表記する．



✦ 確率分布が$i \perp\!\!\!\perp \mathrm{nd}(i) | \pi(i)$ を満たすことを $I\left(i, \mathrm{nd}(i) | \pi(i)\right)_p$ と表記する．

$\mathbb{DF} \Rightarrow \mathbb{DL}$は（今の段階では）言えない

- Topological sortingにより確率変数を並べる：$\mathrm{des}(i) \not\subset \{X_1, \dots, X_{i-1}\}$

- 変数の集合を定義する

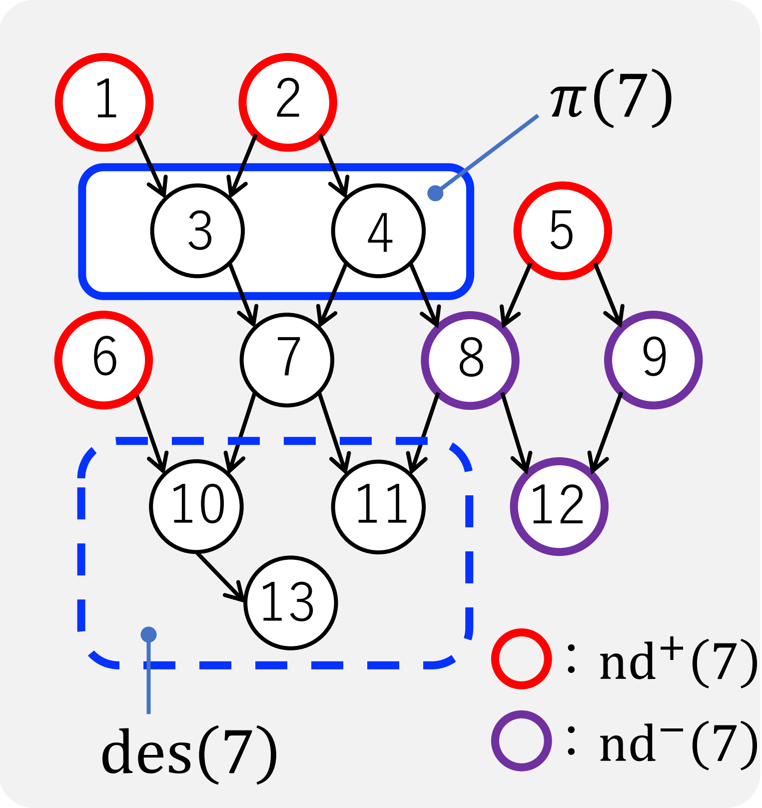
  ✦ $\mathrm{nd}^+(i) = \mathrm{nd}(i) \cap \{X_1, \dots, X_{i-1}\} \setminus \pi(i)$
  … non-descendantのうち$\pi(i)$を除いて
  上流($j < i, j \in V$)にあるもの

  ✦ $\mathrm{nd}^-(i) = \mathrm{nd}(i) \setminus \mathrm{nd}^+(i) \cup \pi(i)$
  … non-descendantのうち下流にあるもの

  - よって$\{X_1, \dots, X_{i-1}\} = \mathrm{nd}^+(i) \cup \pi(i)$

- $\mathbb{DF}$を適用すると$p(X_i | X_1, \dots, X_{i-1}) = p(X_i | \boldsymbol{X}_{\pi(i)})$,
  つまり$I\left(X_i, \{X_1, \dots, X_{i-1}\} \setminus X_{\pi(i)} | X_{\pi(i)}\right)_p = I\left(X_i, X_{\mathrm{nd}^+(i)} | X_{\pi(i)}\right)$

示したいのは$I\left(X_i, X_{\mathrm{nd}(i)\setminus\pi(i)} | X_{\pi(i)}\right)$なので，これだけでは不十分.



定理：$\mathbb{DL} \Rightarrow \mathbb{DF}$

DAG $\mathcal{D}$において確率分布pが$\forall\alpha \in V$について$I\left(X_\alpha, \boldsymbol{X}_{\mathrm{nd}(\alpha)} | \boldsymbol{X}_{\pi(\alpha)}\right)_p$であるとき，
pは$\mathcal{D}$が規定する親ノードに関する条件付き確率により因子分解される.

▶ 証明の概要

- $k(< N)$個のノード$\mathcal{V}_k$からなるサブグラフ$\mathcal{D}_{\mathcal{V}_k}$で$\mathbb{DF}$が成立しているとする.
  - ✦ $k = 1$はあきらかに成立する
- $\mathbb{DL}$が成立しているとき，$\mathcal{D}_{\mathcal{V}_{k+1}}$でも$\mathbb{DF}$が成立することを示す.

◆ 証明 ◆

- $\mathcal{D}_{\mathcal{V}_k}$で$\mathbb{DF}$が成立しているとき，$\mathcal{D}_{\mathcal{V}_{k+1}}$での同時確率は次のように与えられる

$$p(X_1, X_2, \dots, X_k, X_{k+1}) = \prod_{i=1}^{k} p\left(X_i | \boldsymbol{X}_{\pi(i)}\right) p(X_{k+1} | X_1, X_2, \dots, X_k)$$

Decomposition Propertyを用いる

- $\mathbb{DL}$：$I\left(X_\alpha, \boldsymbol{X}_{\mathrm{nd}(\alpha)\setminus\pi(\alpha)} | \boldsymbol{X}_{\pi(\alpha)}\right)_p \Rightarrow I\left(X_\alpha, \boldsymbol{X}_{\mathrm{nd}^+(\alpha)} | \boldsymbol{X}_{\pi(\alpha)}\right)_p$ and $I\left(X_\alpha, \boldsymbol{X}_{\mathrm{nd}^-(\alpha)} | \boldsymbol{X}_{\pi(\alpha)}\right)_p$なので
$p(X_{k+1} | X_1, X_2, \dots, X_k) = p\left(X_{k+1} | \boldsymbol{X}_{\pi(k+1)}\right)$となり，$\mathcal{D}_{\mathcal{V}_{k+1}}$でも$\mathbb{DF}$が成立する.

定義：$\mathbb{DG}$

確率分布pが，DAG $\mathcal{D}$のもと

d分離性から定義される条件付き独立性を満たす性質を

Global Markov Propertyと呼び，$\mathbb{DG}$と表記する．

$$\mathbb{DG}：I(A,B|C)_{\mathcal{D}} \Rightarrow I(A,B|C)_p$$

定理：$\mathbb{DF} \Rightarrow \mathbb{DG}$

確率分布$p_\mathcal{B}$がDAG $\mathcal{D}$においてfactorizedされているとき,

CがAとBをd-分離するdisjointなノード集合A, B, Cについて

$p_\mathcal{B}$は$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C$を満たす.

▶ 証明の概要（詳細は付録⏩）

- 帰納法を用いる：その際にTopological sortingを行う
  - $\mathcal{D}$から末端ノードωを取り除いたDAG $\mathcal{D}_{\backslash\omega}$において, $I(A', B'|C')_{\mathcal{D}_{\backslash\omega}} \Rightarrow I(X_{A'}, X_{B'}|X_{C'})_p$が成立しているとする
    - $\mathcal{D}_{\backslash\omega}$のノード数が1の場合は自明に$\mathbb{DF} \Rightarrow \mathbb{DG}$が成立
  - ωに対する$\mathbb{DF}$と$I(A', B'|C')_{\mathcal{D}_{\backslash\omega}} \Rightarrow I(X_{A'}, X_{B'}|X_{C'})_p$から, $I(A, B|C)_\mathcal{D} \Rightarrow I(X_A, X_B|X_C)_p$を示す.
    - つまりノード数$k-1$で$\mathbb{DF} \Rightarrow \mathbb{DG}$が成立していれば$k$でも成立する

Contraction Propertyと Weak Union Propertyを用いる

定理：$\mathbb{DG} \Rightarrow \mathbb{DL}$

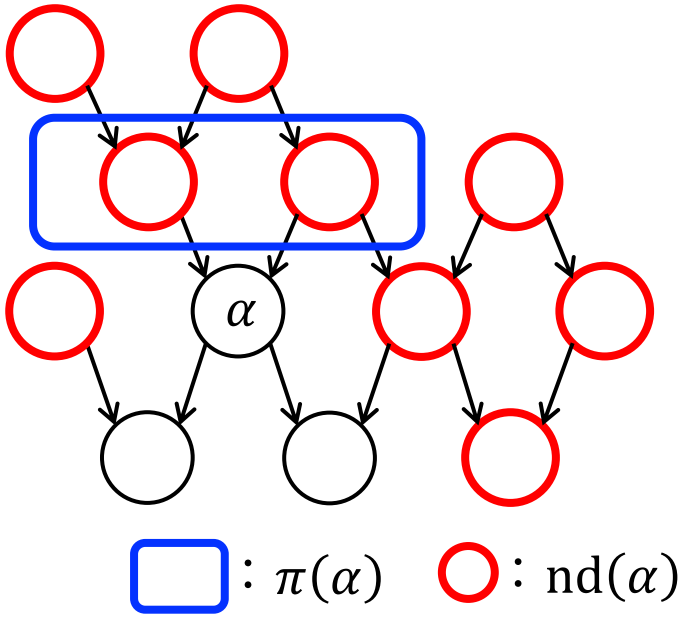
ある確率分布pがDAG $\mathcal{D}$において

d分離性から示される条件付き独立性を満たすとき,

pは$\forall \alpha \in V$について$\alpha \perp\!\!\!\perp \mathrm{nd}(\alpha) | \pi(\alpha)$を満たす.

▶ 証明の概要 (詳細⏩)

- あるノードαに注目し，DAG $\mathcal{D}$のSkelton $S[\mathcal{D}]$において，αから非子孫ノード$\mathrm{nd}(\alpha)$への可能な経路を考える
- 確率分布pがd分離性から示される条件付き独立性を満たすとき，親ノード$\pi(\alpha)$により条件づけられるとαから$\mathrm{nd}(\alpha)$への可能な経路はすべてブロックされることを示す（つまり$\mathbb{DL}$: $\alpha \perp\!\!\!\perp \mathrm{nd}(\alpha) | \pi(\alpha)$が成立）



ここまでのまとめ

- $\mathbb{DF} \Rightarrow \mathbb{DG} \Rightarrow \mathbb{DL} \Rightarrow \mathbb{DF}$が示された．よって$\mathbb{DF} \Leftrightarrow \mathbb{DG} \Leftrightarrow \mathbb{DL}$
  - ✦ Factorizeされた確率分布はd分離性から示される条件付き独立性を満たす.
  - ✦ 確率分布がd分離性から示される条件付き独立性を満たすのなら，Factorized formで表現できる
  - ✦ つまり$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(p_{\mathcal{B}})$

    ※ $I(A,B|C)_{\mathcal{D}} \in \mathcal{I}(\mathcal{D}) \Rightarrow I(A,B|C)_{p_{\mathcal{B}}} \in \mathcal{I}(p_{\mathcal{B}})$の意味.

    あとで登場するI-mapという概念と対応する.

- 証明において用いた仮定：Symmetry, Decomposition, Weak Union, Contraction
  - ✦ 確率分布の独立性に限らず，Semi-graphoidな独立性であれば同じ結論
- $\mathbb{DG} \Rightarrow \mathbb{DF}$の直接の証明も可能

定理：$\mathbb{DG} \Rightarrow \mathbb{DF}$

DAG $\mathcal{D}$におけるd分離性から与えられる条件付き独立性を

確率分布$p_{\mathcal{B}}$が満たすとき，確率分布$p_{\mathcal{B}}$はFactorizeされる．

▶ 証明の概要

- 証明のためには，いくつかの概念が必要

  - ✦ 無向グラフにおける**Markov Property**
  - ✦ **Ancestral set**
  - ✦ **Moral graph**

  無向グラフにおける条件付き独立性や，BNとMNの対応関係から説明します．

→ この資料の最後で$\mathbb{DG} \Rightarrow \mathbb{DF}$の証明を扱います．
（詳しくは「5. ベイジアンネットワーク再訪」⏩）

$\mathbb{DF} \Leftrightarrow \mathbb{DG}$を認めると

- 条件付き独立性を仮定したChain ruleから，対応するDAGを得ることが可能
  - ✦ 一般のChain rule：$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)$

- さまざまなConditional independenceを仮定

(1) $p(X_1)p(X_2)p(X_3)$

(2) $p(X_1)p(X_2|X_1)p(X_3)$, $p(X_1)p(X_2)p(X_3|X_1)$, $p(X_1)p(X_2)p(X_3|X_2)$

(3) $p(X_1)p(X_2)p(X_3|X_1, X_2)$

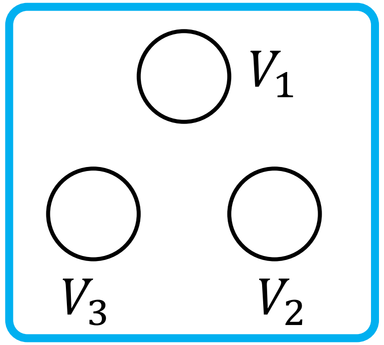
(4) $p(X_1)p(X_2|X_1)p(X_3|X_2)$

(5) $p(X_1)p(X_2|X_1)p(X_3|X_1)$

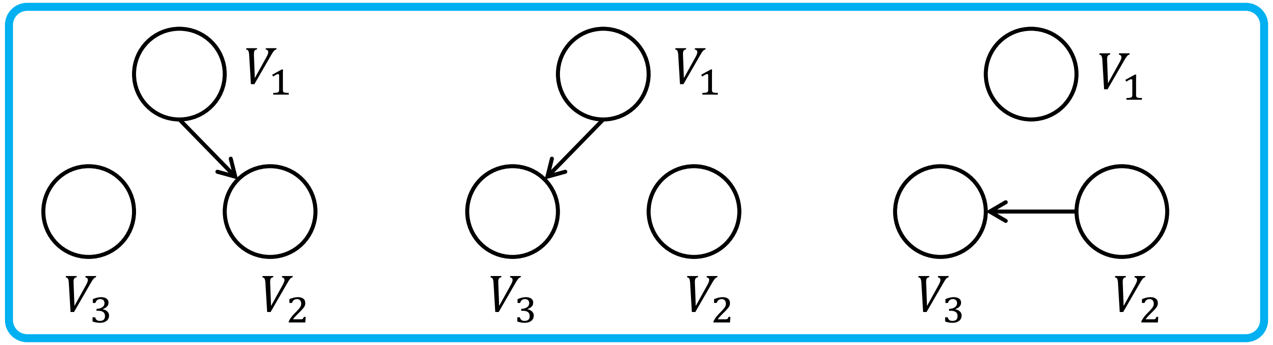
(6) $p(X_1)p(X_2|X_1)p(X_3|X_1, X_2)$

DAGによる条件付き独立性の表現

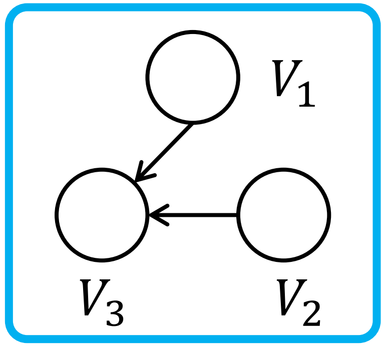
(1) $p(X_1)p(X_2)p(X_3)$



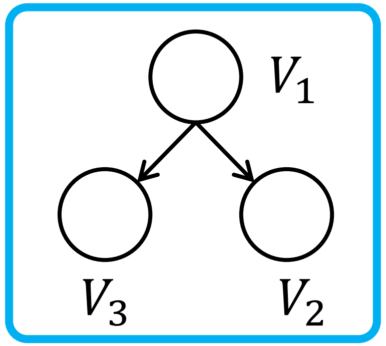
(2)$p(X_1)p(X_2|X_1)p(X_3)$など



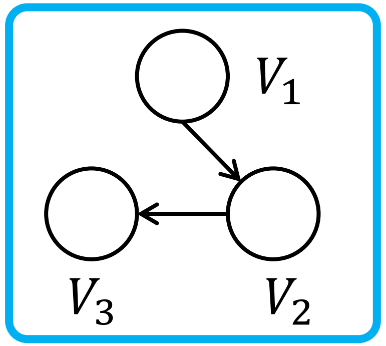
(3)$p(X_1)p(X_2)p(X_3|X_1,X_2)$



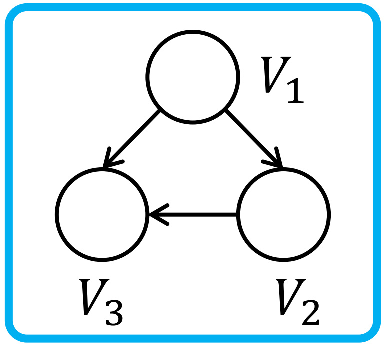
(4) $p(X_1)p(X_2|X_1)p(X_3|X_1)$



(5) $p(X_1)p(X_2|X_1)p(X_3|X_2)$

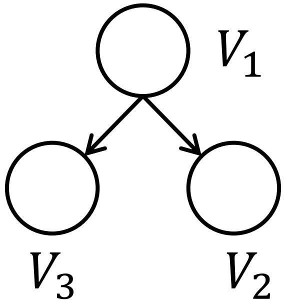


(6) $p(X_1)p(X_2|X_1)p(X_3|X_1,X_2)$

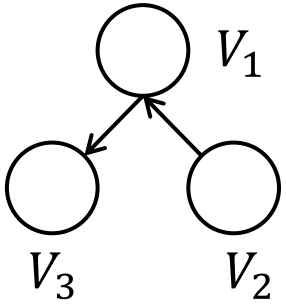


逐次結合と分岐結合の等価性

$p(X_1)p(X_2|X_1)p(X_3|X_1)$



$p(X_2)p(X_1|X_2)p(X_3|X_1)$



同時分布で表現すると

（左） $P(X_1)P(X_2|X_1)P(X_3|X_1) = \dfrac{P(X_1, X_2)P(X_1, X_3)}{P(X_1)}$

（右） $P(X_2)P(X_1|X_2)P(X_3|X_1) = \dfrac{P(X_1, X_2)P(X_1, X_3)}{P(X_1)}$

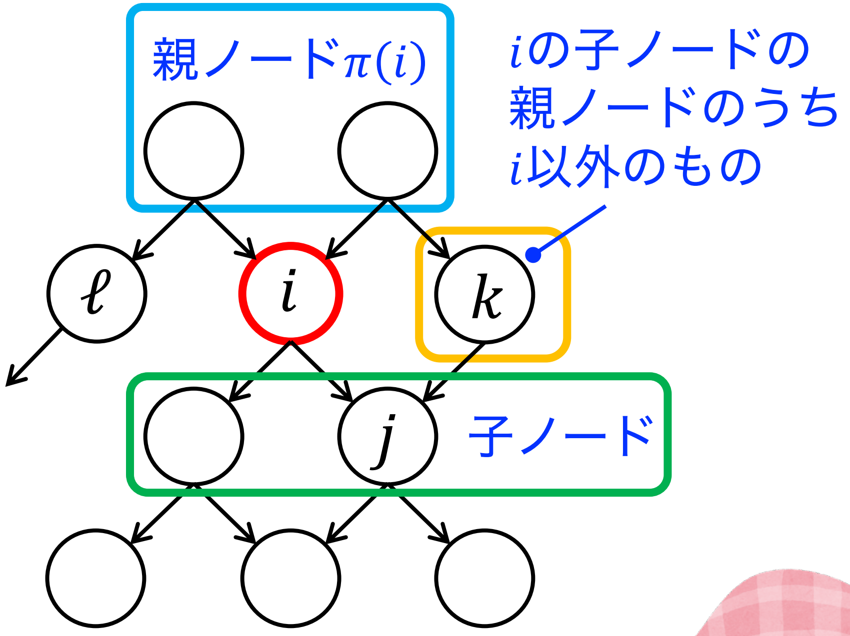
同じ分布を表現している.

これを**I-equivalence**という

Markov blanket

ノードV_iの親ノード, 子ノードと，その子ノードの

ノードV_i以外の親ノードの集合を**Markov blanket**と呼び$V_{M_{\mathcal{D}}(i)}$と表記する.

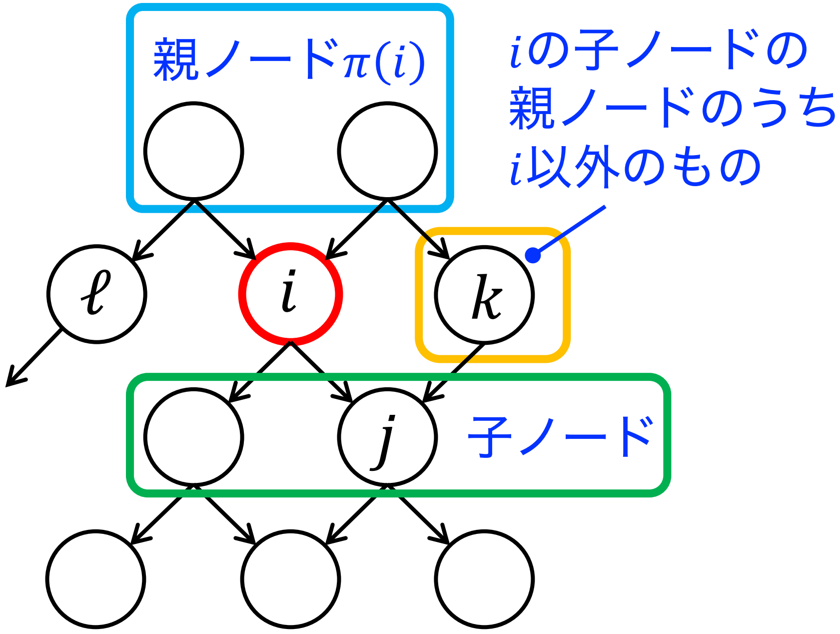


- $I(i, V \setminus M_{\mathcal{D}}(i) | M_{\mathcal{D}}(i))_{\mathcal{D}}$である
  - ✦ iの親ノード$\pi(i)$がインスタンス化されると $\pi(i)$の子孫ノード(たとえばℓ)とiは独立. またさらに先の祖先ノードとも独立.
  - ✦ 子ノードjがインスタンス化されるとiとkは依存. ただしjから先についてはブロックされる

おふとんにd分離されています

Markov blanket (条件付き確率からの説明)

- ベイジアンネットワークのunderlying distributionについて，
  $\boldsymbol{X}_{\setminus i}$で条件づけられた$X_i$の周辺化分布を考える



$$p_{\mathcal{B}}(\boldsymbol{X}) = \prod_{j=1}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right)$$

$$p_{\mathcal{B}}\left(X_i \middle| \boldsymbol{X}_{\setminus i}\right) \equiv \frac{\prod_{j=1}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right)}{\int \prod_{j=1}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right) d\boldsymbol{X}_i}$$

$$= \frac{p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right) \prod_{j\in\gamma(i)}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right) \cancel{\prod_{j\notin\gamma(i)}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right)}}{\int p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right) \prod_{j\in\gamma(i)}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right) d\boldsymbol{X}_i \cancel{\prod_{j\notin\gamma(i)}^{N} p\left(X_j \middle| \boldsymbol{X}_{\pi(j)}\right)}}$$

※ $\gamma(i)$: iの子ノード集合

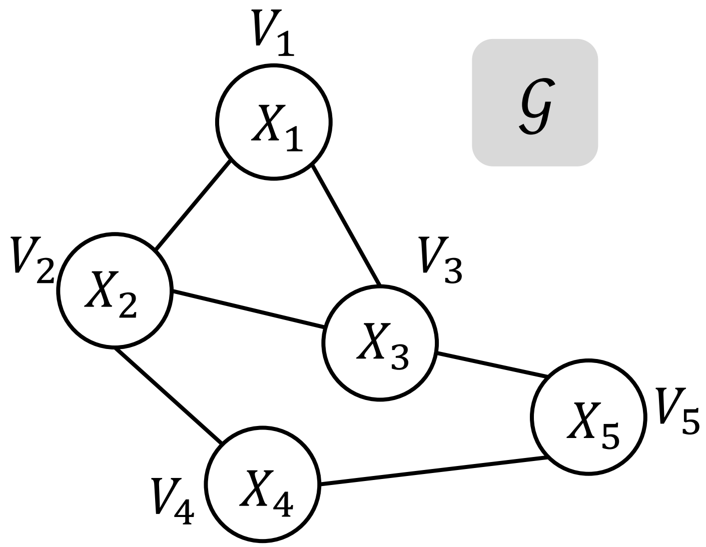
Markov blanket nodeの変数のみに依存する

3. マルコフネットワーク

マルコフネットワークとは

N個の変数集合$\boldsymbol{X} = \{X_1, \dots, X_N\}$をもつマルコフネットワークモデルは$(\mathcal{G}, p_{\mathcal{M}})$で表現される

- ✦ $\mathcal{G}$: ノード集合により構成される無向グラフ
  - Markov network structureとよぶ
- ✦ $p_{\mathcal{M}}$: Underlying distribution
  - $\mathcal{G}$により指定されるルールで **ファクターϕ**を掛け合わせたもの



- 全ての可能な確率分布をPとすると$p_{\mathcal{M}} \subset P$
  - ✦ MNは一部の確率分布しか表現できない

$$p_{\mathcal{M}}(\boldsymbol{X}) = \frac{1}{Z} \prod_a \phi_a$$

- Z：分配関数(規格化定数)

定義：Global Markov Independence

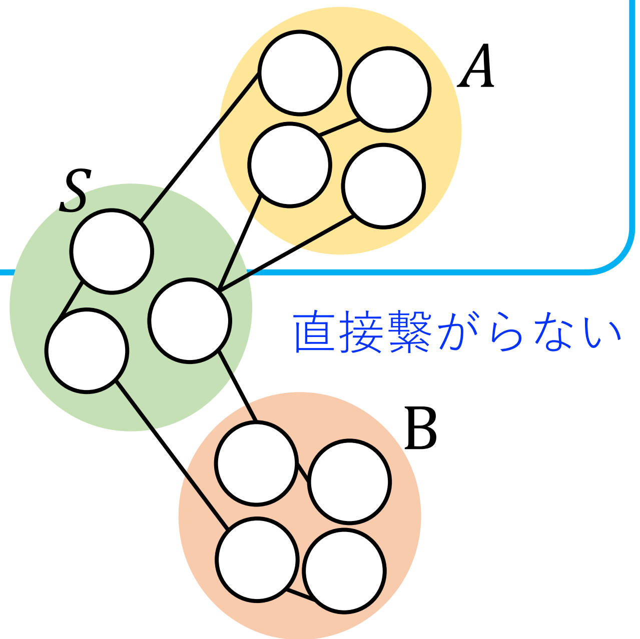
無向グラフ $\mathcal{G}$において，disjointなノードのsubset A, B, Cがあったとする.

A, B内のノード間の全ての経路が Cの少なくとも1つのノードを含むとき，

Markov network structure $\mathcal{G}$は， $i \in V$について

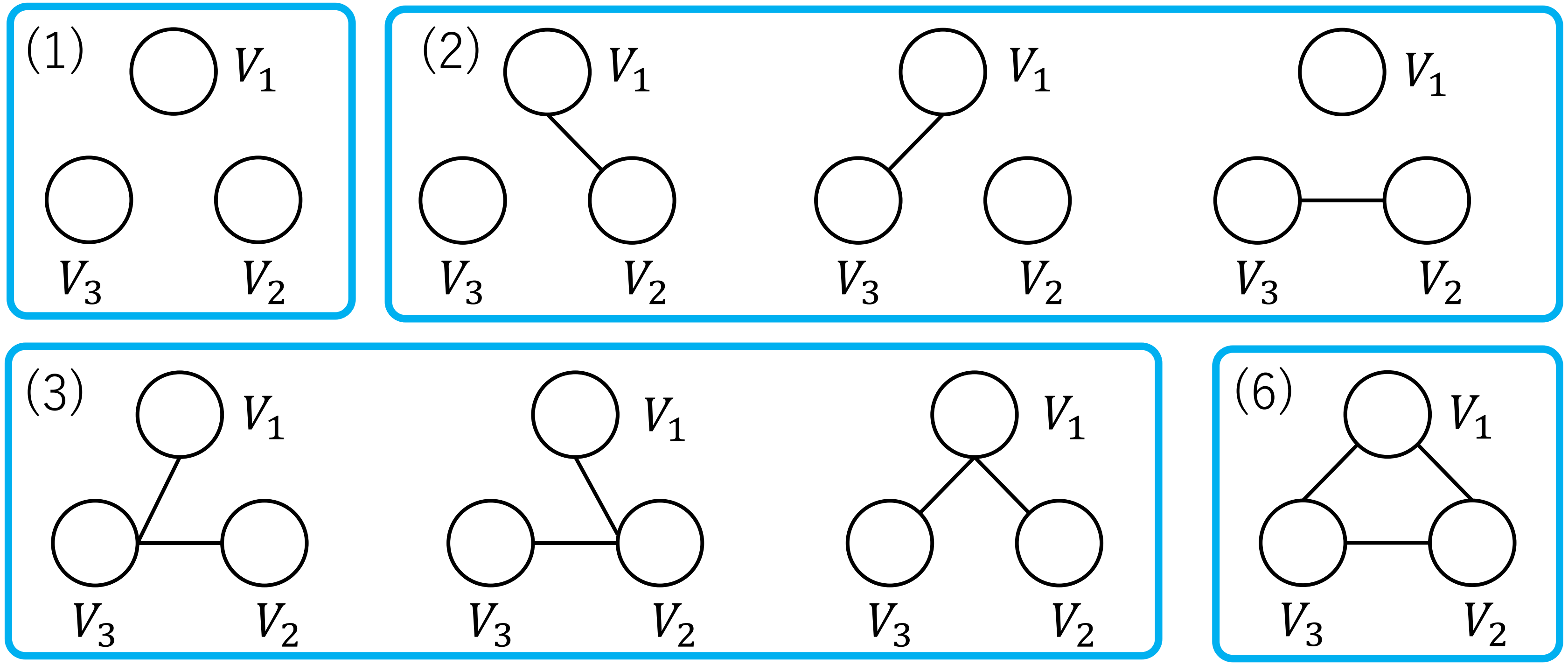
独立性の仮定$A \perp\!\!\!\perp B | C$を表現している.

この独立性をGlobal Markov Independenceという.



- 無向グラフ$\mathcal{G}$において$A \perp\!\!\!\perp B | C$が表現されることを $I(A, B | C)_{\mathcal{G}}$と表記する.

$N = 3$でのMN



(1) $X_1 \perp\!\!\!\perp X_2, X_1 \perp\!\!\!\perp X_3, X_2 \perp\!\!\!\perp X_3, X_1 \perp\!\!\!\perp \{X_2, X_3\}, X_2 \perp\!\!\!\perp \{X_3, X_1\}, X_3 \perp\!\!\!\perp \{X_1, X_2\}$

(2の左端)$X_1 \perp\!\!\!\perp X_3, X_2 \perp\!\!\!\perp X_3, X_3 \perp\!\!\!\perp \{X_1, X_2\}$

(3の左端)$X_1 \perp\!\!\!\perp X_2 | X_3$

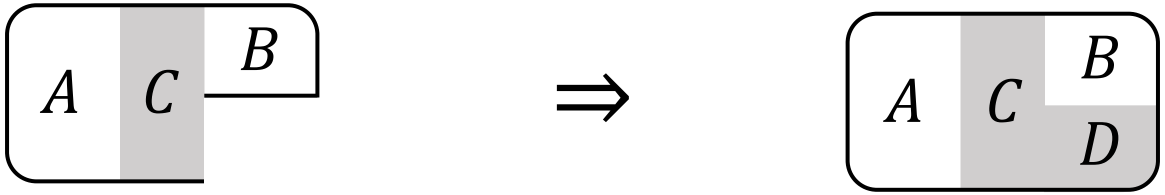
(6) 独立でない

合流結合と分岐結合が区別できないため
$X_1 \perp\!\!\!\perp X_2$ $(X_1 \not\!\perp\!\!\!\perp X_2 | X_3)$などが表現できない

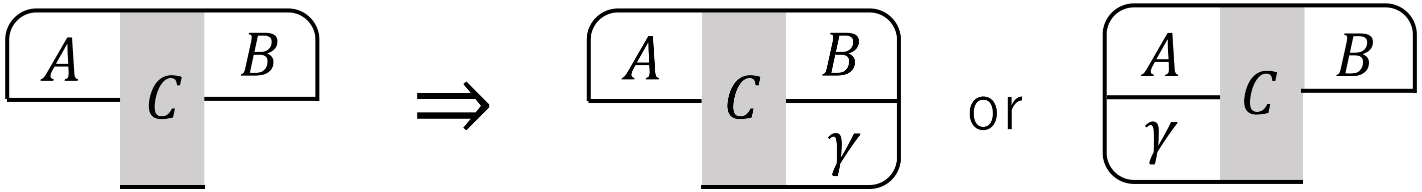
$I(A,B|C)_{\mathcal{G}}$の性質

- Graphoidである
- Symmetry, Decomposition, Intersection, Strong union, Transitivityは $\mathcal{I}(\mathcal{G}) = \{I(A,B|C)_{\mathcal{G}}\ |C$は$A$と$B$を分離する$\}$の完備な公理系である

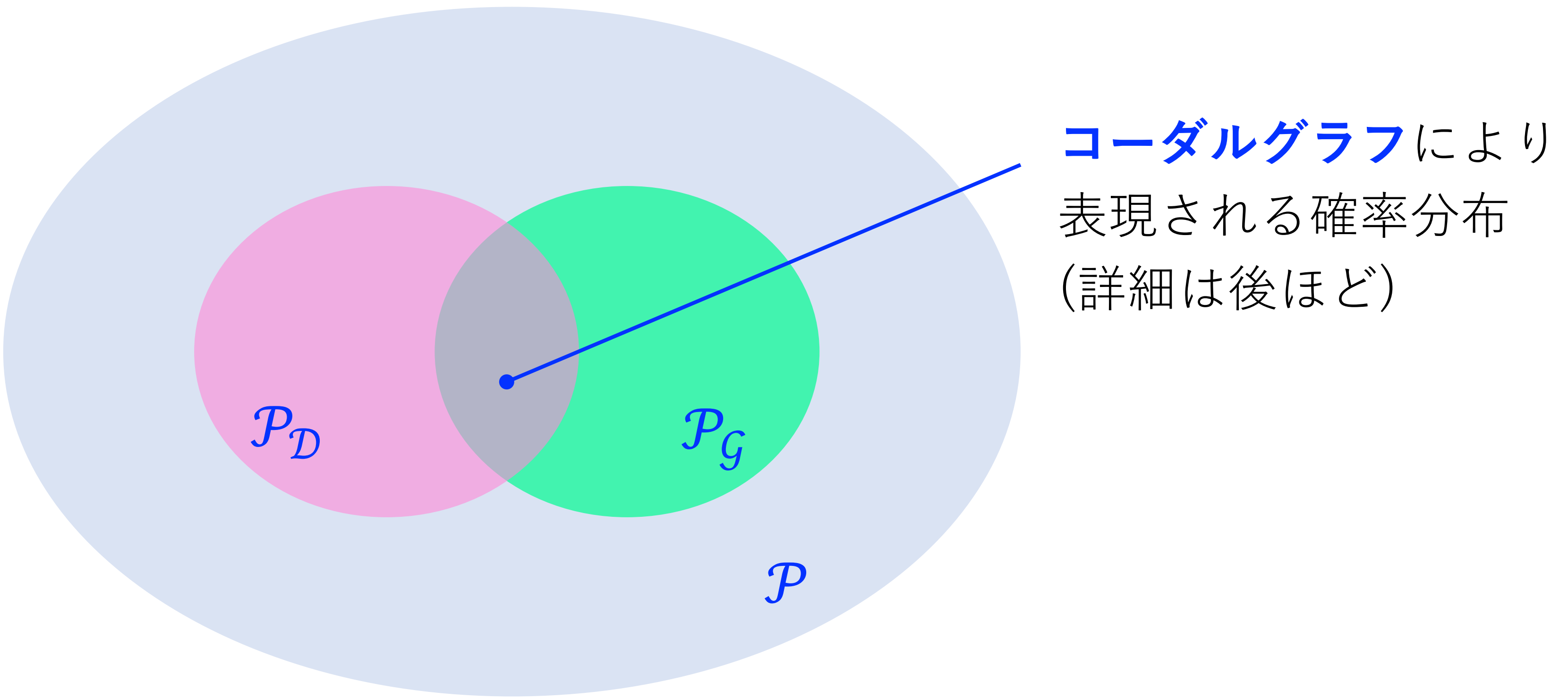
g. Strong union ：$I(A,B|C) \Rightarrow I(A,B|C \cup D)$



h. Transitivity ：$I(A,B|C) \Rightarrow I(A,\gamma|C)$ or $I(\gamma,B|C),\ \gamma \notin (A \cup B \cup C)$



BNとMNの比較



- $\mathcal{P}$: 全ての可能な確率分布の条件付き独立性
- $\mathcal{P}_\mathcal{D}$: BNによって表現可能な条件付き独立性
- $\mathcal{P}_\mathcal{G}$: MNによって表現可能な条件付き独立性

3.1 クリークポテンシャルとGibbs分布

マルコフネットワークにおけるUnderlying distribution

Factorization Property: $\mathbb{F}$ (completenessによる定義)

確率分布pが次のように表現されるとき,

無向グラフ$\mathcal{G}$においてpはfactorizeされているという.

$$p(\boldsymbol{X}) = \frac{1}{Z} \prod_{a:a \text{ is complete}} \phi_a(\boldsymbol{X}_a)$$

このような分布は**Gibbs分布**ともよばれる.

ここで$Z = \sum_{\boldsymbol{X}} \prod_{a:a \text{ is complete}} \phi_a(\boldsymbol{X}_a)$ は**分配関数**(規格化定数)である.

また関数$\phi_a(X_a)$を**ファクター**とよぶ.

※ ベイジアンネットワークでは，条件付き確率により同時分布をFactorizeしているため，
規格化定数が必要ない.

Complete graph

ノード集合が全てエッジで繋がっているとき，

完全グラフ(complete graph)であるという

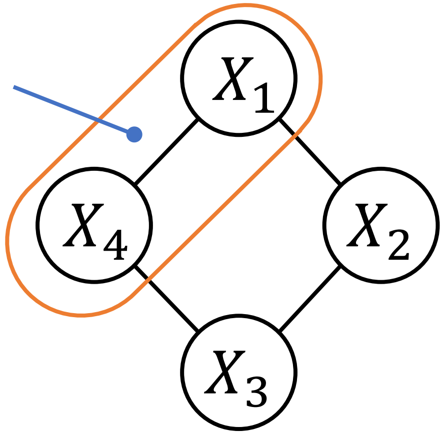
- Complete：

  $\{X_1, X_2\}, \{X_2, X_3\}, \{X_3, X_4\}, \{X_4, X_1\}$

- Completeでない：

  $\{X_1, X_2, X_3\}, \{X_1, X_2, X_4\}, \{X_1, X_3, X_4\},$
  $\{X_2, X_3, X_4\}, \{X_1, X_2, X_3, X_4\}$

Completeな例



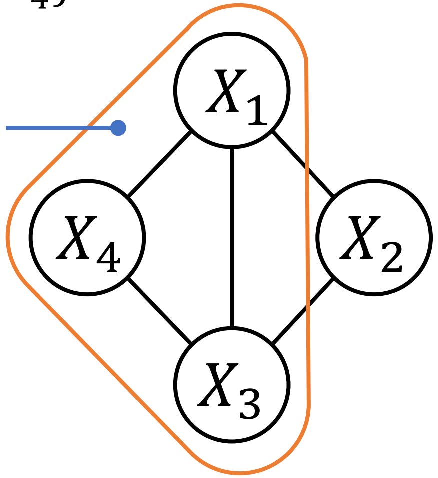
- Complete：

  $\{X_1, X_2\}, \{X_2, X_3\}, \{X_3, X_4\}, \{X_4, X_1\}$
  $\{X_1, X_2, X_3\}, \{X_1, X_3, X_4\}$

- Completeでない：

  $\{X_1, X_2, X_3, X_4\}$

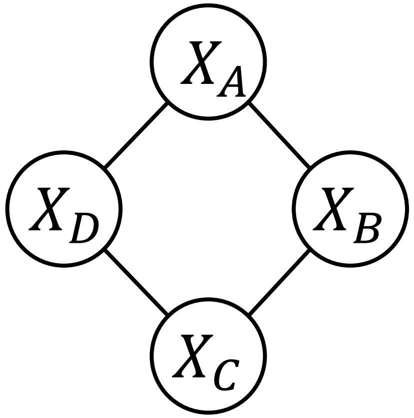
Completeな例



具体例

- Completeなノード集合は $\{X_A, X_B\}, \{X_B, X_C\}, \{X_C, X_D\}, \{X_D, X_A\}$
  - Xは± 1の値をとる
- ファクター集合は$\Phi = \{\phi_1(X_A, X_B), \phi_2(X_B, X_C), \phi_3(X_C, X_D), \phi_4(X_D, X_A)\}$
  - 次の関数を仮定する：$\phi_i(a, b) = \exp(J_i ab)$
- 確率分布

$$p_{\mathcal{M}}(X_A, X_B, X_C, X_D) \propto \exp(J_1 X_A X_B + J_2 X_B X_C + J_3 X_C X_D + J_4 X_D X_A)$$



- 分配関数

$$Z = \sum_{X_A} \sum_{X_B} \sum_{X_C} \sum_{X_D} \exp(J_1 X_A X_B + J_2 X_B X_C + J_3 X_C X_D + J_4 X_D X_A)$$

$$= 4 \sum_{X_B} \sum_{X_D} \cosh(J_1 X_B + J_4 X_D) \cosh(J_2 X_B + J_3 X_D)$$

$$= 4\{\cosh(J_1) \cosh(J_4) \cosh(J_2) \cosh(J_3) + \sinh(J_1) \sinh(J_4) \sinh(J_2) \sinh(J_3)\}$$

Factorization Property: $\mathbb{F}$ (Cliqueによる定義)

確率分布pが次のように表現されるとき,
無向グラフ$\mathcal{G}$においてpはfactorizeされているという.

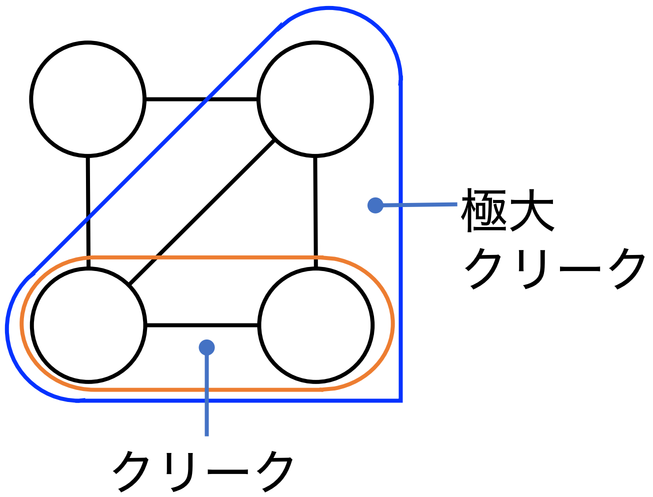
$$p(\boldsymbol{X}) = \frac{1}{Z}\prod_{\gamma \in C} \psi_\gamma(\boldsymbol{C}_\gamma)$$

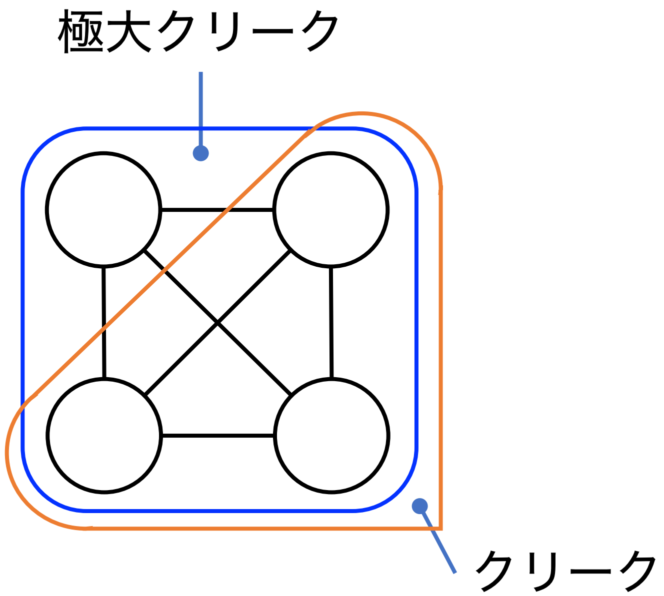
Cは**極大クリーク**の集合で, $\boldsymbol{C}_\gamma$はγ番目のクリークに属す変数の集合である.
また関数$\psi_\gamma(\boldsymbol{X}_\gamma)$は**クリークポテンシャル**とよばれる.

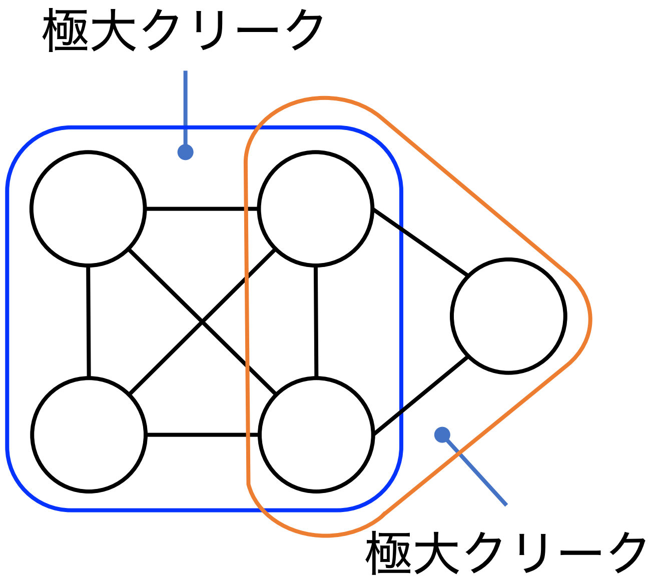
分配関数は次のように与えられる. $Z = \sum_{\boldsymbol{X}} \prod_{\gamma \in C} \psi_\gamma(\boldsymbol{X}_\gamma)$

クリーク

- グラフ$\mathcal{G}$の部分グラフが完全グラフであるとき,
  その部分グラフを**クリーク**と呼ぶ
- クリークを部分グラフとして含む他のクリークが存在しないとき,
  そのクリークを**極大クリーク**と呼ぶ





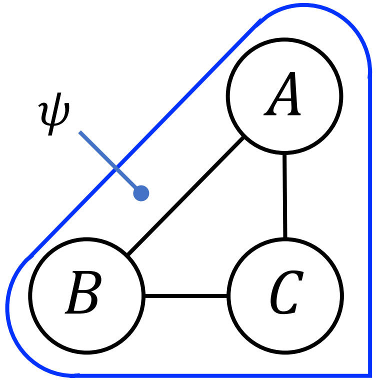


ファクター→クリークポテンシャル

ファクターϕに含まれる変数集合を**スコープ**とよび，Scope[ϕ]と表記する．

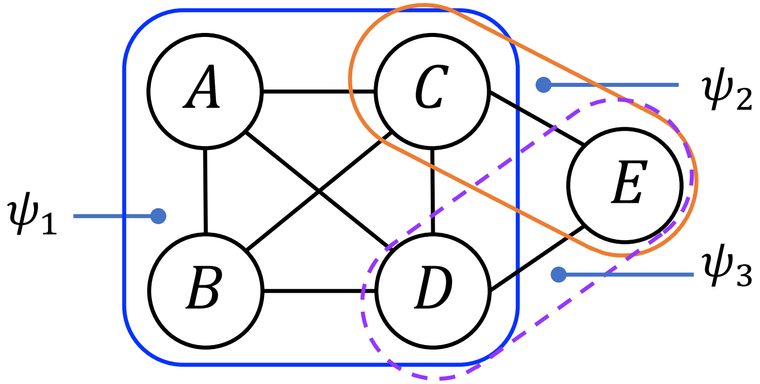
ファクター集合Φのもと，クリークポテンシャルは次のように与えられる

$$\psi_\gamma(\boldsymbol{C}_\gamma) = \prod_{\phi \in \Phi : \mathrm{Scope}[\phi] \subseteq \boldsymbol{C}_\gamma} \phi$$



$\Phi = \{\phi(A,B), \phi(B,C), \phi(C,A)\}$

$\psi(A,B,C) = \phi(A,B)\phi(B,C)\phi(C,A)$



$\Phi = \{\phi(A,B,C), \phi(B,C), \phi(C,D), \phi(C,E), \phi(E,D)\}$

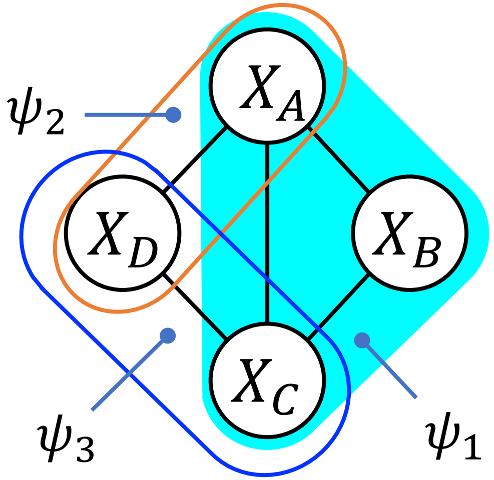
$\psi_1 = \phi(A,B,C)\phi(B,C)\phi(B,D)$

$\psi_2 = \phi(C,E),\qquad \psi_3 = \phi(E,D)$

具体例：Completeグラフによる表現

- Completeなノード集合は $\{X_A, X_B\}, \{X_A, X_C\}, \{X_B, X_C\}, \{X_C, X_D\}, \{X_D, X_A\}$
  - Xは± 1の値をとる
- ファクター集合：$\Phi = \{\phi_1(X_A, X_B), \phi_2(X_B, X_C), \phi_3(X_C, X_D), \phi_4(X_D, X_A), \phi_5(X_A, X_C)\}$
  - 次の関数を仮定する：$\phi_i(a, b) = \exp(J_i ab)$
- 確率分布

$$p_{\mathcal{M}}(X_A, X_B, X_C, X_D) \propto \exp(J_1 X_A X_B + J_2 X_B X_C + J_3 X_C X_D + J_4 X_D X_A + J_5 X_A X_C)$$



- 分配関数

$$Z = \sum_{X_A} \sum_{X_B} \sum_{X_C} \sum_{X_D} \exp(J_1 X_A X_B + J_2 X_B X_C + J_3 X_C X_D + J_4 X_D X_A + J_5 X_A X_C)$$

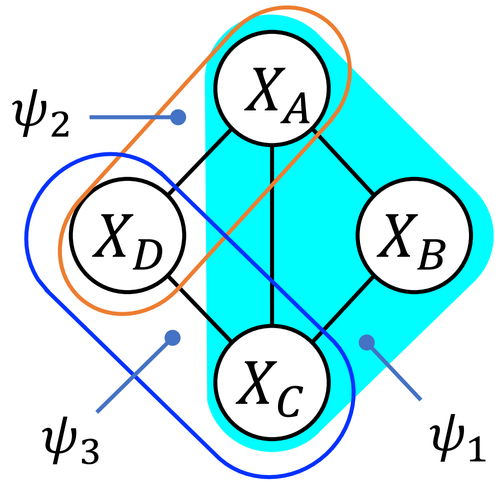
$$= 4 \sum_{X_A} \sum_{X_C} \{\cosh(J_1) \cosh(J_2) + X_A X_C \sinh(J_1) \sinh(J_2)\}$$

$$\times \{\cosh(J_3) \cosh(J_4) + X_A X_C \sinh(J_3) \sinh(J_4)\} \{\cosh(J_5) + X_A X_C \sinh(J_5)\}$$

$= \cdots$（略）

具体例：クリークによる表現

- クリークは$C_1 = \{X_A, X_B, X_C\}, C_2 = \{X_A, X_C, X_D\}$
  - Xは± 1の値をとる
- ファクター集合：$\Phi = \{\phi_1(X_A, X_B), \phi_2(X_B, X_C), \phi_3(X_C, X_D), \phi_4(X_D, X_A), \phi_5(X_A, X_C)\}$
  - 次の関数を仮定する：$\phi_i(a, b) = \exp(J_i ab)$
  - クリークポテンシャル：
    - ✦ $\psi_1 = \phi_1(X_A, X_B)\phi_2(X_B, X_C)\phi_5(X_A, X_C)$
    - ✦ $\psi_2 = \phi_4(X_D, X_A)$
    - ✦ $\psi_3 = \phi_3(X_C, X_D)$



- 確率分布

$$p_{\mathcal{M}}(X_A, X_B, X_C, X_D) \propto \psi_1\psi_2\psi_3$$

3.2 分離性とGibbs分布の関係

無向グラフにおける独立性と$\boldsymbol{p}_{\mathcal{M}}$の独立性

無向グラフ$\mathcal{G}$から規定される条件付き独立性：$I(A,B|S)_{\mathcal{G}}$

Disjointなノード集合A,B,Sについて，SがA,Bを分離するとき条件付き独立

$I(A,B|S)_{\mathcal{G}}$と$I(A,B|S)_{p_{\mathcal{M}}}$の関係を議論する．

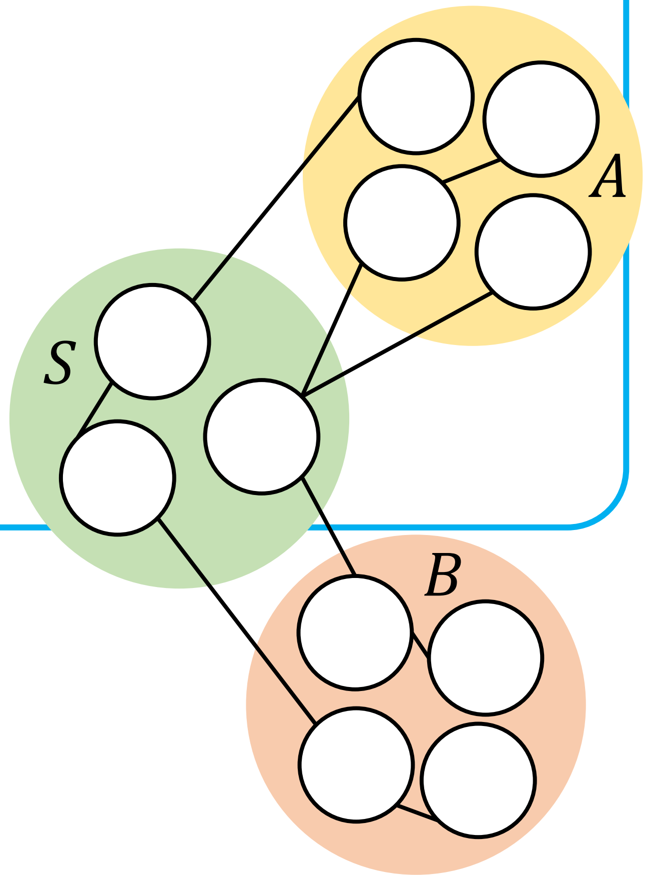
無向グラフ$\mathcal{G}$のUnderlying distribution $p_{\mathcal{M}}$

クリークポテンシャルによりFactorizeされる（Gibbs分布）

定義：Global Markov Property $\mathbb{G}$

グラフ$\mathcal{G}$におけるdisjointなノード集合A, B, Sについて
SがAとBを分離しているとする.

確率分布pにおいて$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_S$が成立するとき
pは**Global Markov Property**を満たすという.



✦ $\mathbb{F} \Leftrightarrow \mathbb{G}$が示せれば$\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(p_{\mathcal{M}})$が言える

※ $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(p_{\mathcal{M}})$は$I(A, B|S)_{\mathcal{G}} \in \mathcal{I}(\mathcal{G}) \Rightarrow I(A, B|S)_{p_{\mathcal{M}}} \in \mathcal{I}(p_{\mathcal{M}})$の意

定理：$\mathbb{F} \Rightarrow \mathbb{G}$

確率分布pが無向グラフ$\mathcal{G}$において

Factorized formで表現されるとき,

確率分布pはGlobal Markov Propertyを満たす.

▶ 証明の概要 (詳細は⏩)

- AとBがSにより分離されているとき,
  AとBにまたがるクリークが存在しない性質を示す
- この性質を利用してFactorization formを適用する

Decomposition Propertyを用いる

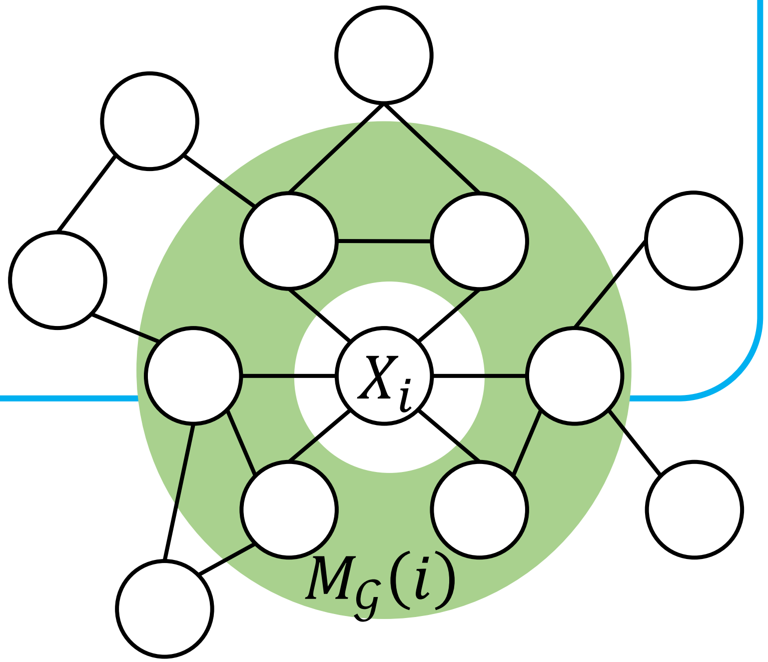
定義：Local Markov property $\mathbb{L}$

確率分布pが無向グラフ$\mathcal{G}$について次の性質を持つとき,

Local Markov Propertyを満たすという

$$p\left(X_i, \boldsymbol{X}_{\backslash i\cup M_{\mathcal{G}}(i)} | \boldsymbol{X}_{M_{\mathcal{G}}(i)}\right) = p\left(X_i | \boldsymbol{X}_{M_{\mathcal{G}}(i)}\right) p\left(\boldsymbol{X}_{\backslash i\cup M_{\mathcal{G}}(i)} | \boldsymbol{X}_{M_{\mathcal{G}}(i)}\right) \quad \forall i \in V$$

$M_{\mathcal{G}}(i)$はiの周辺ノード集合であり,
無向グラフにおける**Markov Blanket**と
よばれる.



定理：$\mathbb{G} \Rightarrow \mathbb{L}$

確率分布pが無向グラフ$\mathcal{G}$において

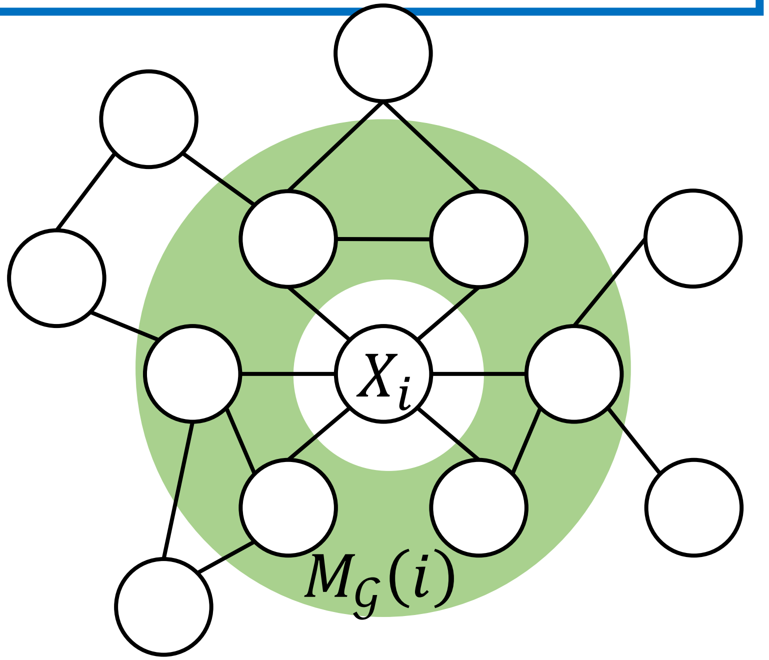
Global Markov Property $\left(I(A,B|S)_{\mathcal{G}} \Rightarrow I(A,B|S)_p\right)$を満たすとき，

Local Markov Propertyも満たす.

◆ 証明 ◆

iは$M_{\mathcal{G}}(i)$により$\boldsymbol{X}_{\setminus i \cup M_{\mathcal{G}}(i)}$から分離される.

したがって$\mathbb{G} \Rightarrow \mathbb{L}$が直ちに示される.



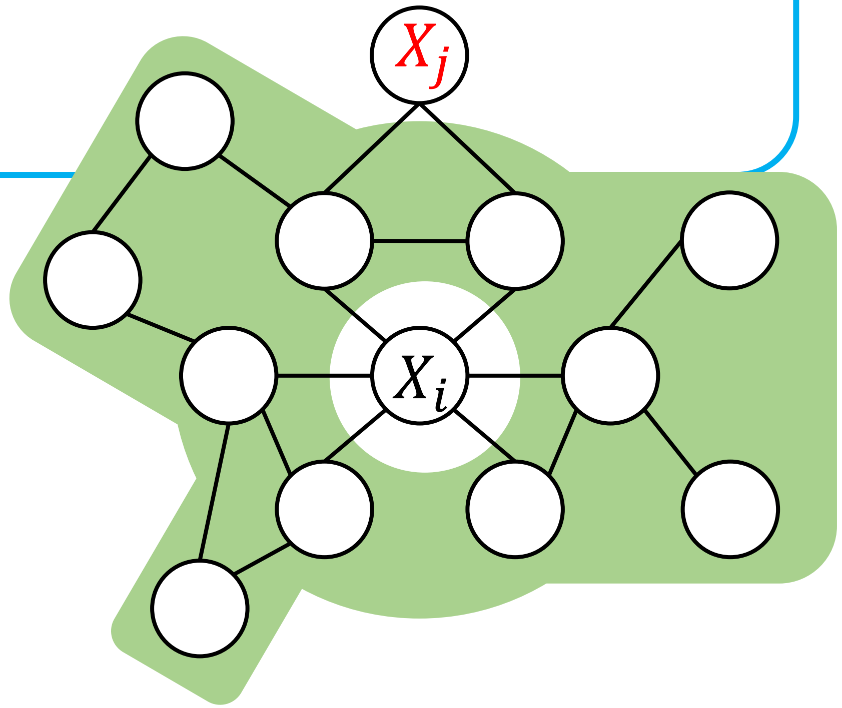
定義：Pairwise Markov Property $\mathbb{P}$

全ノードペア$(i,j) \notin \mathcal{E}$について次の関係が成立するとき,

$$p(X_i, X_j | \boldsymbol{X}_{\setminus i,j}) = p(X_i | \boldsymbol{X}_{\setminus i,j}) p(X_j | \boldsymbol{X}_{\setminus i,j})$$

確率分布pはグラフ$\mathcal{G}$について

Pairwise Markov Propertyを持つという



定理：$\mathbb{L} \Rightarrow \mathbb{P}$

確率分布pが無向グラフ$\mathcal{G}$において

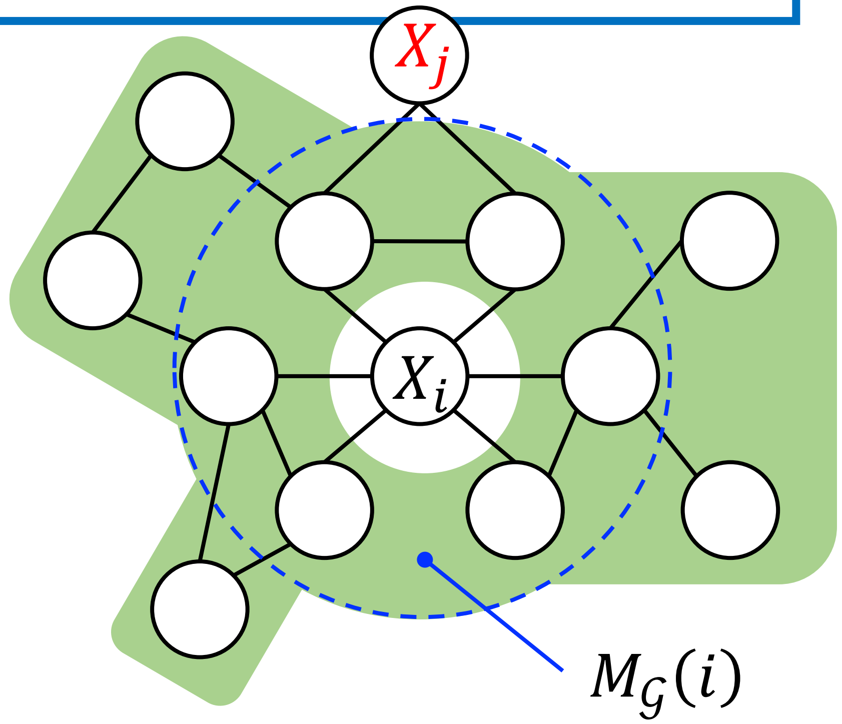
Local Markov Property $X_i \perp\!\!\!\perp \left(X_j \cup \boldsymbol{X}_{\backslash i,j}\right) | \boldsymbol{X}_{M_{\mathcal{G}}(i)}$ $\forall i \in V$を満たすとき,

Pairwise Markov Propertyも満たす.

◆ 証明 ◆

- $X_i \perp\!\!\!\perp \left(X_j \cup \boldsymbol{X}_{\backslash i,j}\right) | \partial_i$にWeak Union propertyを適用すると$X_i \perp\!\!\!\perp X_j | \left(\partial_i \cup \boldsymbol{X}_{\backslash i,j}\right)$
- $\boldsymbol{X}_{M_{\mathcal{G}}(i)} \subset \boldsymbol{X}_{\backslash i,j}$なので$X_i \perp\!\!\!\perp X_j | \boldsymbol{X}_{\backslash i,j}$

Weak Union Propertyを用いる



ここまでのまとめ

$$\mathbb{F} \Rightarrow \mathbb{G} \Rightarrow \mathbb{L} \Rightarrow \mathbb{P}$$

- 一般に逆は成立しない
- 以下のページでは，次のことを述べる
  - ✦ $p(\boldsymbol{X}) > 0$では$\mathbb{P} \Rightarrow \mathbb{G}$が成立するので$\mathbb{G} \Leftrightarrow \mathbb{L} \Leftrightarrow \mathbb{P}$
  - ✦ Hammersley-Cliffordの定理から， $p(\boldsymbol{X}) > 0$で$\mathbb{P} \Rightarrow \mathbb{F}$が成立

定理：$\mathbb{P} \Rightarrow \mathbb{G}$

確率分布pがintersection propertyを満たし，

また無向グラフ$\mathcal{G}$においてpが

Pairwise Markov Propertyを満たすならば

Global Markov Propertyも満たす

▶ 証明の概要 (詳細は⏩)

- 3つのdisjointなノード集合A, B, Cに対して帰納法を用いる（CはAとBを分離している）

  (i) $|C| = N - 2$の場合(すなわち$|A| = |B| = 1$), $\mathbb{P} \Rightarrow \mathbb{G}$は直ちに成立する

  (ii) $|C| > n$で$\mathbb{P} \Rightarrow \mathbb{G}$が成立しているとして，$|C| = n$ で$\mathbb{P} \Rightarrow \mathbb{G}$が成立するかを考える

定理：$\mathbb{P} \Rightarrow \mathbb{F}$

Hammersley-Cliffordの定理

無向グラフ$\mathcal{G}$において確率分布$p(\boldsymbol{X}) > 0$が

Pairwise Markov propertyを満たすとき,

pは$\mathcal{G}$のもとGibbs分布で与えられる.

▶ 証明の概要 (詳細は⏩)

- あるノード集合bと$V \setminus b$に対して，$H_b(\boldsymbol{X}_b) = \log p\left(\boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b}\right)$として関数$H_b$を定義
  - $\boldsymbol{X}^*_{V\setminus b}$は実現値
- H_bに対してMöbius反転公式を適用することで，確率分布がノード集合に対してFactorizeされることを示す
- さらに，completeでないノード集合からの寄与が無いことを示す

まとめ

- 一般に,

$$\mathbb{F} \Rightarrow \mathbb{G} \Rightarrow \mathbb{L} \Rightarrow \mathbb{P}$$

使った性質
- Decomposition
- Weak Union

- 確率分布が$p(\boldsymbol{X}) > 0$である場合は

$$\mathbb{F} \Leftrightarrow \mathbb{G} \Leftrightarrow \mathbb{L} \Leftrightarrow \mathbb{P}$$

- Intersection
- Decomposition

✦ Gibbs分布は分離性から示される条件付き独立性を満たす.

✦ 分離性から示される条件付き独立性を満たす確率分布は，Gibbs分布で表現できる.

✦ つまり$\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(p_{\mathcal{M}})$ （$I(A, B|C)_{\mathcal{G}} \in \mathcal{I}(\mathcal{G}) \Rightarrow I(A, B|C)_{p_{\mathcal{M}}} \in \mathcal{I}(p_{\mathcal{M}})$の意味）

  - あとで登場するI-mapという概念と対応する.

※ Contraction propertyを満たさない独立性にも適用できる

4. 確率分布とBN, MNの関係性

Minimal I-mapとBN→MN, MN→BN変形

用語の整理

注）グラフの性質なのか，分布の性質なのかを把握しないと混乱する

	グラフから 条件付き独立性を 読み取る際に 参照する性質	グラフの性質 (確率分布の独立性に 対して定義される)	グラフ上での 確率分布の形	確率分布の 独立性の性質 (グラフの独立性に 対して定義される)
BN	d分離性	• D-map • I-map • Perfect map	Factorized Form (親ノードによる 条件付け)	Directed Markov Property
MN	分離性	• D-map • I-map • Perfect map	Factorized Form (クリークポテンシャル による分解, Gibbs分布)	Markov Property

あとで
説明

これから説明します．

”map”とは，真の確率分布の
条件付き独立性をマップしているという意味．

4.1 確率分布に対するグラフの特徴づけ

定義：D-map

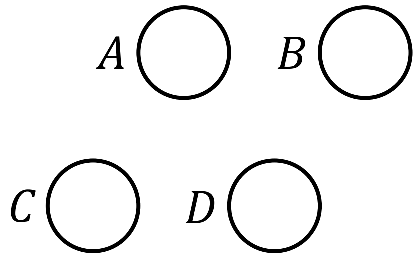
確率分布pが持つ条件付き独立性がグラフGにより全て表現されるとき，
つまり任意のdisjointなノード集合A, B, Sについて

$$I(\boldsymbol{X}_A, \boldsymbol{X}_B | \boldsymbol{X}_S)_p \Rightarrow I(A, B|S)_G$$

が成立するとき，グラフGを
確率分布pに対する**依存性マップ**(Dependency map, D-map) とよぶ.

- GはDAGでも無向グラフでもよい。
  DAGなら$I(A, B|S)_G \leftarrow I(A, B|S)_{\mathcal{D}}$,
  無向グラフなら$I(A, B|S)_{\mathcal{G}}$とする

- 完全に分離されたグラフは,
  全ての分布に対する自明なD-map



$A \perp\!\!\!\perp B, A \perp\!\!\!\perp B | C, A \perp\!\!\!\perp B | \{C, D\}, \ldots$
あらゆる独立性を表現できる

定義：Maximal D-map

D-mapのうち最大のグラフGを**最大依存性マップ**(Maximal D-map)とよぶ.

D-mapであるGに1つでもedgeを追加すると依存性マップでなくなるとき,
Gは最大依存性マップである.

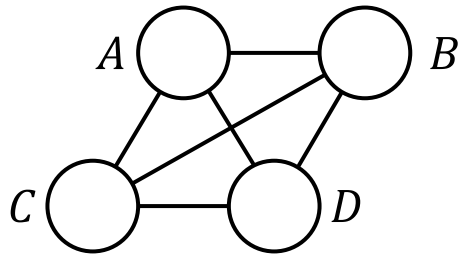
定義：I-map

グラフGによって規定される全ての条件付き独立性が
確率分布pにより満たされるとき，すなわち任意のノード集合A, B, Sについて

$$I(A, B|S)_G \Rightarrow I(A, B|S)_p$$

が成立するとき，グラフGを
確率分布pに対する**独立性マップ**(Independence map, I-map)と呼ぶ。

- 完全グラフは
  あらゆる条件付き独立性を表現できないので，
  全ての分布に対する自明なI-mapである
- DAG $\mathcal{D}$は$p_{\mathcal{B}}$のI-mapであり，
  無向グラフ$\mathcal{G}$は$p_{\mathcal{M}}$のI-mapであることは既に見た



あらゆる独立性を表現できない

定理：$\mathcal{D}$はpのI-map⇒ $\mathbb{DF}$

DAG $\mathcal{D}$が確率分布pのI-mapであるとき，pはFactorizeされる

◆ 証明 ◆

- 確率変数$\boldsymbol{X}$のラベルをTopological sortingにより決めておく
- 確率分布のChain ruleより$p(\boldsymbol{X}) = \prod_{j=1}^{N} p\left(X_j \middle| X_1, \dots, X_{j-1}\right)$
- $I\left(i, \mathrm{nd}(i) \middle| \pi(i)\right)_{\mathcal{D}} \in \mathcal{I}(\mathcal{D})$, $\forall i \in V$である.
- $\mathcal{D}$はpのI-mapなので$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(p)$であり，$I\left(X_i, \boldsymbol{X}_{\mathrm{nd}(i)} \middle| \boldsymbol{X}_{\pi(i)}\right)_p \in \mathcal{I}(p)$, $\forall i \in V$である.
- $\boldsymbol{X}_{\mathrm{nd}(i)} \subset \left\{X_1, \dots, X_{j-1}\right\}$であるので,
  $p(X_i | X_1, \dots, X_{i-1}) = p\left(X_i \middle| X_{\pi(i)}\right)$ $\forall i \in V$となり，Factorized formが得られる.

※ Local Markov Independenceの仮定からFactorized formを得る手続きと同じ.

定理： $\mathbb{DF} \Rightarrow \mathcal{D}$は$p$のI-map

確率分布pがDAG $\mathcal{D}$のもとでFactorizeされるとき，

DAG $\mathcal{D}$はpのI-mapである．

◆ 証明 ◆

- $\mathbb{DF} \Leftrightarrow \mathbb{DG}$なので，確率分布$p$は
  グラフで規定される全ての条件付き独立性を見たす．
- よってDAG $\mathcal{D}$はpのI-mapである．

定義：Minimal I-map

I-mapのうち最小のグラフGを**最小独立性マップ**(Minimal I-map)とよぶ.

I-mapであるGから1つでもedgeを取り除くと独立性マップでなくなるとき,
Gは最小独立性マップである.

BNではMinimal I-mapは一意ではない

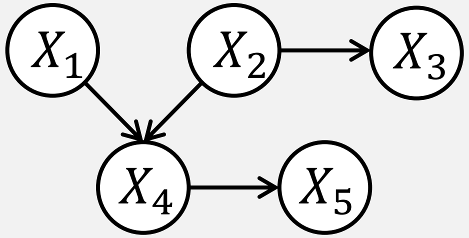
Minimal I-mapの生成アルゴリズム

1. 変数を適当に並べ，ラベル$i = 1, \ldots, N$を与える
2. $X_i \perp\!\!\!\perp \left(\{X_1, \ldots, X_{i-1}\} \setminus \boldsymbol{X}_{\pi(i)}\right) | \boldsymbol{X}_{\pi(i)}$となる最小の$\pi(i) \in \{1, \ldots, i-1\}$を選ぶ

変数の並べ方に依存して，異なるMinimal I-mapが生成される

$\mathcal{I}(p_{\mathcal{B}}) = \mathcal{I}(\mathcal{D})$

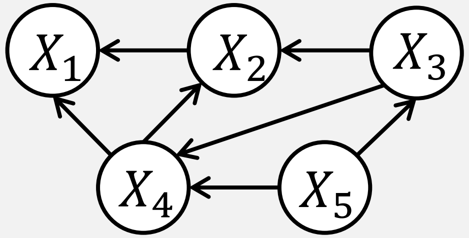
$\mathcal{D}$



$5 \to 3 \to 4 \to 2 \to 1$
とした場合の$p_{\mathcal{B}}$のMinimal I-map

$\mathcal{I}(\mathcal{D}_1) = \{I(X_3, \emptyset | X_5), I(X_4, X_3 | X_5),$
$I(X_2, X_5 | X_3 \cup X_4), I(X_1, X_3 \cup X_5 | X_2, X_4)\}$

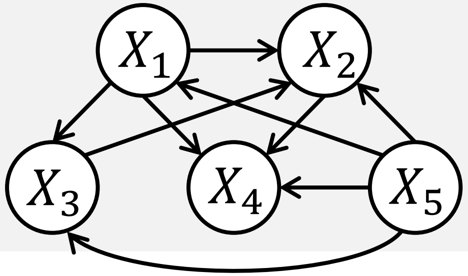
$\mathcal{D}_1$



$5 \to 1 \to 3 \to 2 \to 4$
とした場合の$p_{\mathcal{B}}$のMinimal I-map

$\mathcal{I}(\mathcal{D}_2) = \{I(X_1, \emptyset | X_5), I(X_3, \emptyset | X_1, X_5),$
$I(X_2, X_3 | X_1 \cup X_4 \cup X_5), I(X_4, X_3 \cup X_5 | X_2, X_4)\}$

$\mathcal{D}_2$



MNではMinimal I-mapは一意である

Strictly positiveな確率分布pにおいて，

$p(X_i, X_j | \boldsymbol{X}_{\setminus i,j}) \neq p(X_i | \boldsymbol{X}_{\setminus i,j}) p(X_j | \boldsymbol{X}_{\setminus i,j})$を満たす$(i,j)$の集合$\Omega_p$を構成する．

このとき，$E_{ij} \in \mathcal{E} \left(\forall (i,j) \in \Omega_p\right)$とするedge集合$\mathcal{E}$からなる無向グラフ$\mathcal{G}$は

確率分布pのuniqueな最小I-mapである．

◆ 証明 ◆

- グラフの構成方法から，確率分布pはグラフ$\mathcal{G}$のもとで$\mathbb{P}$である．
  - ✦ $E_{ij} \notin \mathcal{E}$について$p(X_i, X_j | \boldsymbol{X}_{\setminus i,j}) = p(X_i | \boldsymbol{X}_{\setminus i,j}) p(X_j | \boldsymbol{X}_{\setminus i,j})$が成立している
- pはstrictly positiveなので，$\mathbb{P} \Leftrightarrow \mathbb{L} \Leftrightarrow \mathbb{G}$となり，$\mathcal{G}$は$p$のI-mapである。
- 最小性：$\mathcal{G}$のedgeを1つでも抜くと，pにない性質を記述してしまう．
- 一意性：他のI-map $\mathcal{G}'$があったとすると， $\mathcal{G} \subset \mathcal{G}'$である．

定義：Perfect map

グラフGにより確率分布pが持つ全ての条件付き独立性を表現でき,

かつグラフGが規定する全ての条件付き独立性を確率分布pが持つとき,

すなわち次の関係が成立するとき

$$I(X,Y|Z)_p \Leftrightarrow I(X,Y|Z)_G$$

グラフGは確率分布pの**完全マップ**(Perfect map)であるという.

- つまり完全マップはD-mapかつI-mapである

■ 全ての確率分布が完全マップを持つわけではない

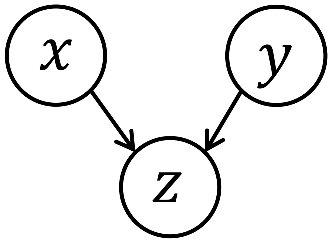
BNで表現できない例 (1)

- XOR関数で与えられる確率分布

$$p(x,y,z)=\begin{cases}1/12 & x\oplus y\oplus z=0\\ 1/6 & x\oplus y\oplus z=1\end{cases}$$

  - $x,y,z\in\{0,1\}$
  - $0\oplus 0=0, 1\oplus 0=1, 1\oplus 1=0$

✦ 右表より$I(x,y|\emptyset)_p, I(x,z|\emptyset)_p\in\mathcal{J}(p)$



$I(x,y|\emptyset)_{\mathcal{D}}$のみ成立

$p_{\mathcal{B}}=p(x)p(y)p(z|x,y)$

x	y	z	$p(x,y,z)$
0	0	0	1/12
0	0	1	1/6
0	1	0	1/6
0	1	1	1/12
1	0	0	1/6
1	0	1	1/12
1	1	0	1/12
1	1	1	1/6

$$p(x)=\begin{cases}1/2 & x=0\\ 1/2 & x=1\end{cases}$$

※ y,zも同様

x	y	$p(x,y)$
0	0	1/4
0	1	1/4
1	0	1/4
1	1	1/4

$p(x,y)=p(x)p(y)$

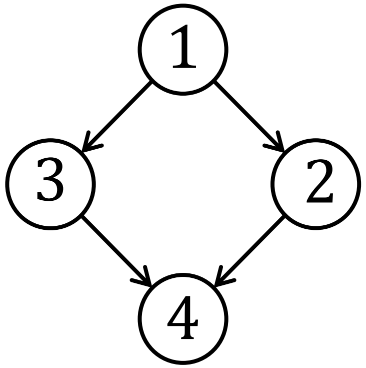
x	z	$p(x,z)$
0	0	1/4
0	1	1/4
1	0	1/4
1	1	1/4

$p(x,z)=p(x)p(z)$

BNで表現できない例 (2)

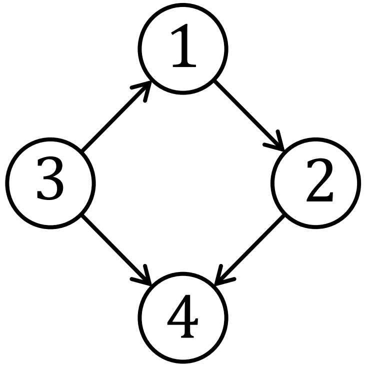
- 「$I(V_1, V_4 | \{V_2, V_3\})$かつ$I(V_2, V_3 | \{V_1, V_4\})$」という独立性

※ head-to-head nodeがV_4の場合のみ列挙



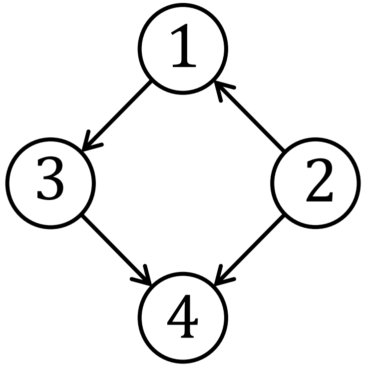
$V_1 \perp\!\!\!\perp V_4 | \{V_2, V_3\}$

$V_2 \not\perp\!\!\!\perp V_3 | \{V_1, V_4\}$



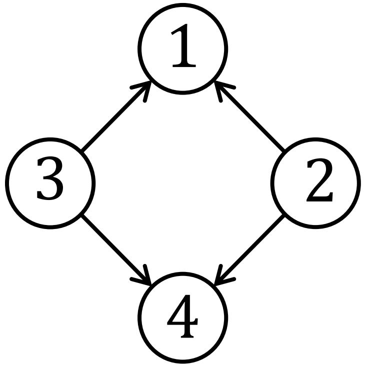
$V_1 \perp\!\!\!\perp V_4 | \{V_2, V_3\}$

$V_2 \not\perp\!\!\!\perp V_3 | \{V_1, V_4\}$



$V_1 \perp\!\!\!\perp V_4 | \{V_2, V_3\}$

$V_2 \not\perp\!\!\!\perp V_3 | \{V_1, V_4\}$



$V_1 \perp\!\!\!\perp V_4 | \{V_2, V_3\}$

$V_2 \not\perp\!\!\!\perp V_3 | \{V_1, V_4\}$

（マルコフネットワークであれば表現できる）

定義：Faithful

$I(X,Y|Z)_p \in \mathcal{J}(p)$であるとき常に$I(X,Y|Z)_{\mathcal{D}} \in \mathcal{J}(\mathcal{D})$であり，

また逆も成立するとき，

確率分布pはDAG $\mathcal{D}$に対してFaithfulであるという．

- 「$\mathcal{D}$はpのperfect mapである」の言い換え

測度論による議論から知られていること

- 多変量ガウス分布の集合を$\mathcal{P}_{\mathcal{N}}$とする.
- $\mathcal{P}_{\mathcal{N}}$ のI-mapとなるDAG $\mathcal{D}$上で
  $\mathcal{P}_{\mathcal{N}}$の表現に必要なパラメータ集合(条件付き確率の集合)を$\pi_{\mathcal{D}}^{\mathcal{N}}$とする.

[**Existence**]

あらゆるDAG $\mathcal{D}$は，完全マップとなる$p \in \mathcal{P}_{\mathcal{N}}$をもつ

[**Measure zero**]

$\pi_{\mathcal{D}}^{\mathcal{N}}$上のLebesgue測度について，Unfaithfulな分布は測度ゼロである.

✦ 多項分布の集合$\mathcal{P}_{\mathcal{M}}$およびパラメータ集合$\pi_{\mathcal{D}}^{\mathcal{M}}$についても同じ性質が成立

- 正規分布について：Geiger and Pearl (1998), Spirtes et al (1993)
- 多項分布について：Meek (1995)

直観的な理解

- 同じFactorized formでも，条件付き分布が異なれば分布は異なる
  - ✦ $p_{\mathcal{B}}(X;\mathcal{D},\Theta),\Theta$は条件付き確率の集合

（例）A, Bという2状態を取る確率変数を考える．親ノードの状態は$2^{|\pi(i)|}$通り．
親ノードの状態がσのとき，X_iがAを取る確率を$p(X_i = A|X_{\pi(i)} = \sigma) = \theta_{i\sigma}$とする．
このとき$\Theta = \left\{\theta_{i\sigma}\middle|i = 1,\dots,N,\sigma = 1,\dots,2^{|\pi(i)|}\right\} \in \Omega_\Theta \equiv [0,1]^{\sum_{i=1}^{N} 2^{|\pi(i)|}}$となる．

- Θの要素内に独立性があると，グラフ構造だけでは独立性を記述できない．
- Ω_Θの中で，そのようなΘの測度はゼロである．
  - ✦ Ω_Θの大きさに比べて非常に小さい．

$\mathcal{I}(p_{\mathcal{B}}) \not\subseteq \mathcal{I}(\mathcal{D})$という例は作ることができる

$\mathcal{D}$

$U \to Y \to Z \to W$

$p_{\mathcal{B}} = p(U)p(Y|U)p(Z|Y)p(W|Z)$

CPT

$p(u_1) = a$	$p(y_1\|u_1) = 1-(b+c)$	$p(z_1\|y_1) = e$	$p(w_1\|z_1) = g$
$p(u_2) = 1-a$	$p(y_2\|u_1) = c$	$p(z_2\|y_1) = 1-e$	$p(w_2\|z_1) = 1-g$
	$p(y_3\|u_1) = b$	$p(z_1\|y_2) = e$	$p(w_1\|z_2) = h$
	$p(y_1\|u_2) = 1-(b+d)$	$p(z_2\|y_2) = 1-e$	$p(w_2\|z_2) = 1-h$
	$p(y_2\|u_2) = d$	$p(z_1\|y_3) = f$	
	$p(y_3\|u_2) = b$	$p(z_2\|y_3) = 1-f$	

$$p(z_1|u_1) = \sum_Y p(z_1|Y)p(Y|u_1) = e - eb + fb$$
$$p(z_1|u_2) = \sum_Y p(z_1|Y)p(Y|u_2) = e - eb + fb$$

ベクトルの内積と見做せる.

また$p(z_2|u_1) = p(z_2|u_2)$なので$U \perp\!\!\!\perp Z$である.
この独立性は$\mathcal{I}(\mathcal{D})$に含まれない.

$\boldsymbol{\alpha} = \begin{pmatrix} p(z_1|y_1) \\ p(z_1|y_2) \\ p(z_1|y_3) \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} p(y_1|u_1) \\ p(y_2|u_2) \\ p(y_3|u_3) \end{pmatrix}$, $\gamma = \sum_Y p(z_1|Y)p(Y|u_1)$ としたとき, $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\beta} + \gamma = 0$を満たす$\boldsymbol{\alpha}, \boldsymbol{\beta}$は測度ゼロ

$\mathcal{I}(p_{\mathcal{B}}) \not\subseteq \mathcal{I}(\mathcal{D})$となるCPTはほとんど実現しない.

I-map/P-mapまとめ

グラフ構造のもとFactorizeされた確率分布と
d分離性/分離性の関係は次の通り

ベイジアンネットワーク

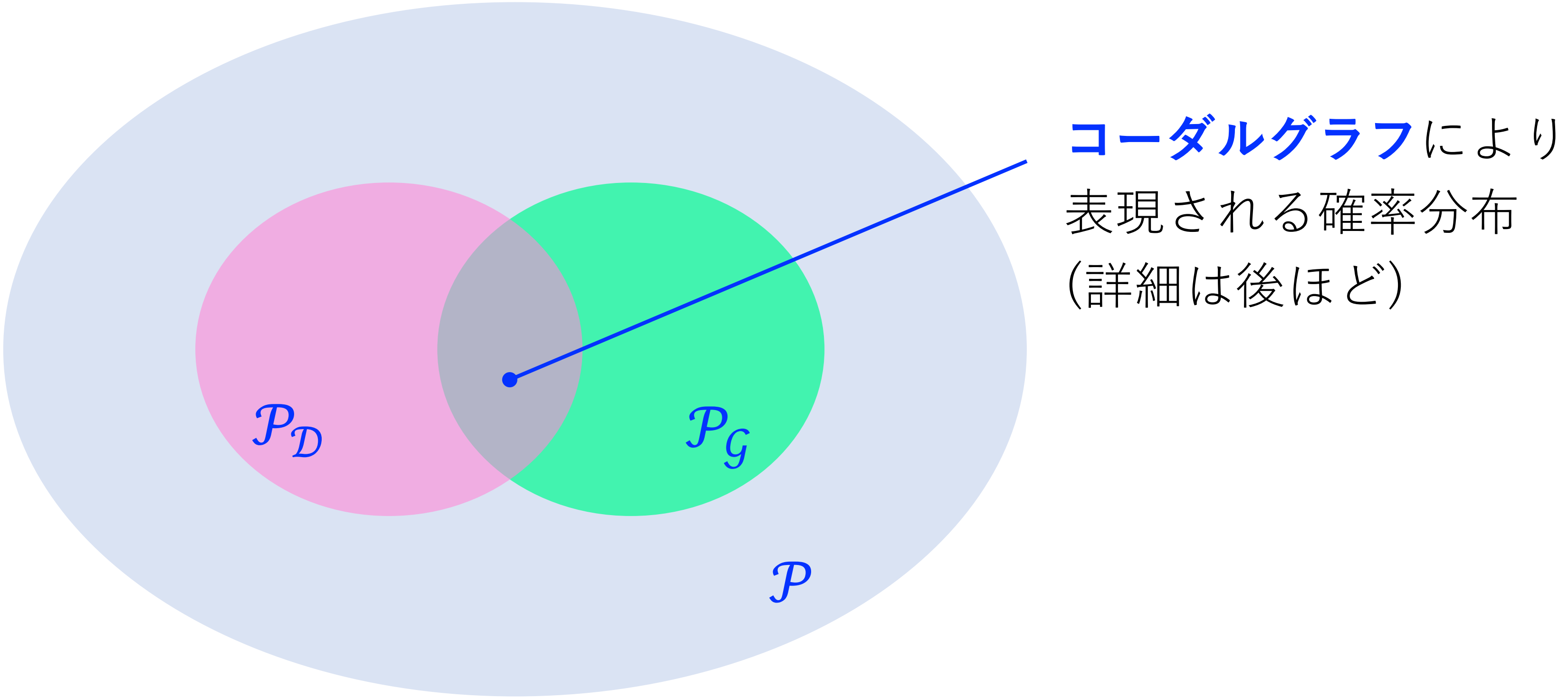
- ✦ $\mathbb{DF} \Rightarrow \mathbb{DG}$が成立する⏩ので$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(p_{\mathcal{B}})$
  - Factorizeされているのであれば，
    d分離性から規定される
    全ての条件付き独立性を満たす
- ✦ 一部の分布では，
  $\mathcal{I}(\mathcal{D}) \neq \mathcal{I}(p_{\mathcal{B}})$となる$p_{\mathcal{B}} \in \mathcal{P}_{\mathcal{D}}$が測度ゼロ．
  つまりほとんどの$p_{\mathcal{B}} \in \mathcal{P}_{\mathcal{D}}$で$\mathcal{I}(\mathcal{D}) = \mathcal{I}(p_{\mathcal{B}})$.

マルコフネットワーク

- ✦ $\mathbb{F} \Rightarrow \mathbb{G}$が成立するので$\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(p_{\mathcal{M}})$
  - Factorizeされているのであれば，
    分離性から規定される
    全ての条件付き独立性を満たす
- ✦ 一部の分布では，
  $\mathcal{I}(\mathcal{G}) \neq \mathcal{I}(p_{\mathcal{M}})$となる$p_{\mathcal{M}} \in \mathcal{P}_{\mathcal{G}}$が測度ゼロ．
  つまりほとんどの$p_{\mathcal{M}} \in \mathcal{P}_{\mathcal{G}}$で$\mathcal{I}(\mathcal{G}) = \mathcal{I}(p_{\mathcal{M}})$.

4.2 BNとMNの関係

BNとMNの比較



- $\mathcal{P}$: 全ての可能な確率分布の条件付き独立性
- $\mathcal{P}_\mathcal{D}$: BNによって表現可能な条件付き独立性
- $\mathcal{P}_\mathcal{G}$: MNによって表現可能な条件付き独立性

定義：モラルグラフ

DAG $\mathcal{D}$の有向エッジを無向エッジとし,

また親ノードを全て結合させたグラフを

モラルグラフと呼び$\mathcal{M}[\mathcal{D}]$と表記する.

「モラルグラフ」という呼び方はどうかと思いますが
慣習的に使われているのでここでもそう呼びます.

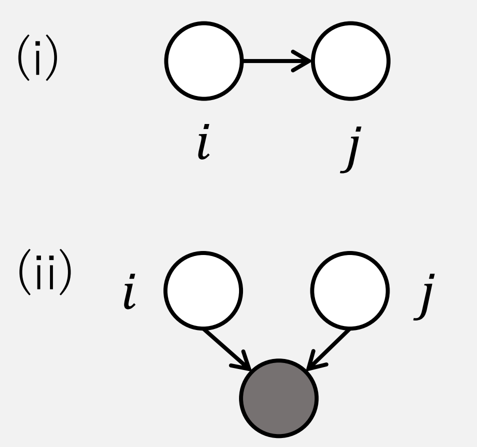
定理：$\mathcal{D}$と$\mathcal{M}[\mathcal{D}]$ (1)

有向グラフ$\mathcal{D}$をモラル化したグラフ$\mathcal{M}[\mathcal{D}]$は，$\mathcal{D}$のMinimal I-mapである

◆ 証明 ◆　… 無向グラフにおけるMinimal I-mapの作り方⏩にしたがう．

- $I(A,B|S)_{\mathcal{G}} \Rightarrow I(A,B|S)_{\mathcal{D}}$となるMarkov Network $\mathcal{G}$を構成することを考える．
- $I\left(X_i,X_j|\boldsymbol{X}_{\backslash i\cup j}\right)_{\mathcal{D}}$が成立しないペア$(i,j)$の集合を$\Omega_{\mathcal{D}}$とする．
  (i) 隣り合う(i,j)は$\Omega_{\mathcal{D}}$に含まれる．
  (ii) 同じ子ノードに繋がっている(i,j)は$\Omega_{\mathcal{D}}$に含まれる．
- エッジ集合$E_{ij} \in \mathcal{E}\ \forall(i,j) \in \Omega_{\mathcal{D}}$となる無向グラフ$\mathcal{G}$は$\mathcal{D}$のMinimal I-mapである．
  - (i)に由来するエッジは$\mathcal{G}$と$\mathcal{D}$で共通
  - (ii)はモラル化していることに対応
- よって$\mathcal{D}$のMinimal I-mapである無向グラフは$\mathcal{M}[\mathcal{D}]$である．

$\Omega_{\mathcal{D}}$に含まれる(i,j)の例



定理：$\mathcal{D}$と$\mathcal{M}[\mathcal{D}]$ (2)

有向グラフ$\mathcal{D}$において親ノードが全結合である場合,

モラル化されたグラフ$\mathcal{M}[\mathcal{D}]$は$\mathcal{D}$のperfect mapである

▶ 証明の概要　(詳細は⏩)

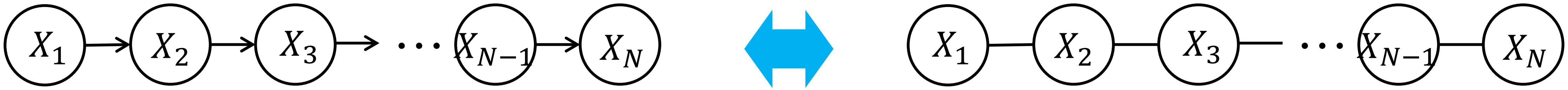
…前ページで一般の$\mathcal{D}$について$\mathcal{I}(\mathcal{M}[\mathcal{D}]) \subseteq \mathcal{I}(\mathcal{D})$を示したので,
　$\mathcal{D}$において親ノードが全結合である場合に$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{M}[\mathcal{D}])$を示せばよい.

- 背理法により証明する
  - $I(X,Y|Z) \in \mathcal{I}(\mathcal{D}), I(X,Y|Z) \notin \mathcal{I}(\mathcal{M}[\mathcal{D}])$という独立性の存在を仮定する.
  - $\mathcal{D}$の親ノードが全結合であるとき$I(X,Y|Z) \in \mathcal{I}(\mathcal{D}), I(X,Y|Z) \notin \mathcal{I}(\mathcal{M}[\mathcal{D}])$は否定され,
    結果$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{M}[\mathcal{D}])$が言える.

BNからMNへの変形例(1)

$\mathcal{D}$ $\qquad\qquad$ $\mathcal{M}[\mathcal{D}]$

$\mathcal{I}(\mathcal{D}) = \mathcal{I}(\mathcal{M}[\mathcal{D}])$



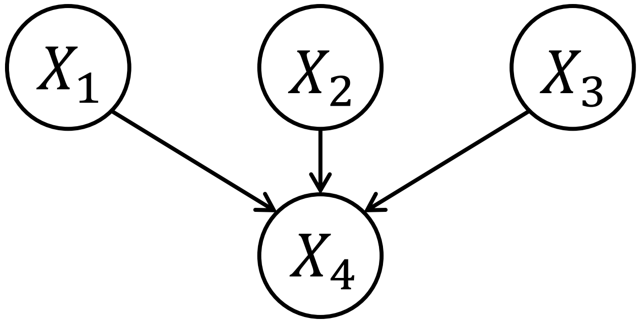
$$p_{\mathcal{B}}(\boldsymbol{X}) = p(X_1)p(X_2|X_1) \times p(X_3|X_2)\ldots p(X_N|X_{N-1})$$

$$p_{\mathcal{M}}(\boldsymbol{X}) = \frac{1}{Z}\psi_{12}(X_1, X_2) \times \psi_{23}(X_2, X_3)\ldots\psi_{N-1,N}(X_{N-1}, X_N)$$

- クリークポテンシャルと条件付き確率は次のように対応する．
  - ✦ $\psi_{12}(X_1, X_2) = p(X_1)p(X_2|X_1)$
  - ✦ $\psi_{i,i+1}(X_i, X_{i+1}) = p(X_{i+1}|X_i)\ (i > 1)$
  - ✦ $Z = 1$
- Parent nodeが1つだけなので，モラル化は有向edgeを無向edgeに変えることに対応する．

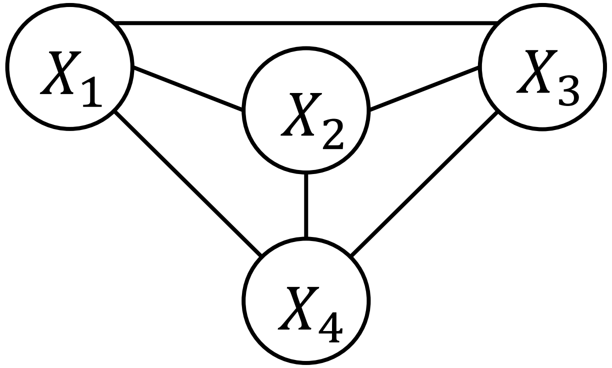
BNからMNへの変形例(2)

$\mathcal{D}$

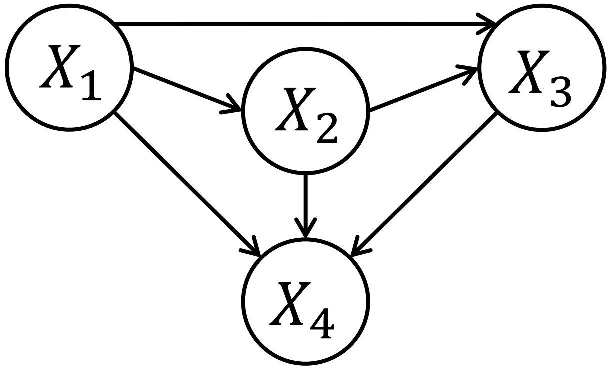


$\mathcal{I}(\mathcal{M}[\mathcal{D}]) \subseteq \mathcal{I}(\mathcal{D})$

$\mathcal{M}[\mathcal{D}]$

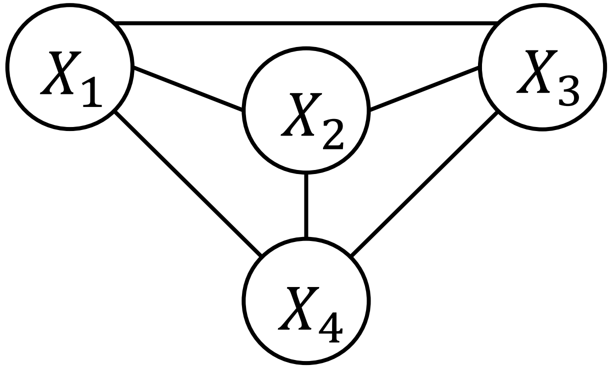


$\mathcal{D}$



$\mathcal{I}(\mathcal{M}[\mathcal{D}]) = \mathcal{I}(\mathcal{D})$

$\mathcal{M}[\mathcal{D}]$



定理：MNからBNへの変形

無向グラフ$\mathcal{G}$の最小I-mapである有向グラフを$\mathcal{D}$とする．

このとき$\mathcal{D}$の親ノードは全結合である．

◆ 証明 ◆

1. I-mapであることから：$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{G})$
2. 最小I-mapであることから：
   $\mathcal{D}$からエッジを１本でも抜いたときに生じる独立性は，$\mathcal{I}(\mathcal{G})$に含まれない

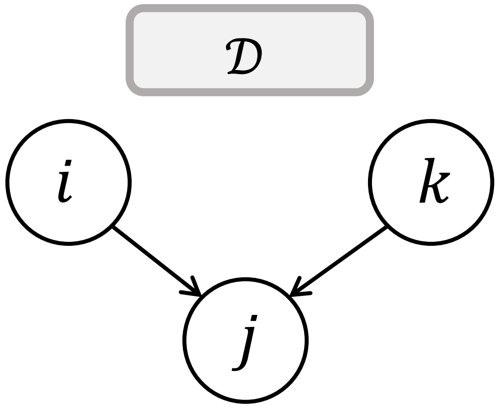
- $\mathcal{D}$に1つでも結合していない親ノードがあるとき，1と2を満たすことができない

言い換え

無向グラフ$\mathcal{G}$の最小I-mapである有向グラフを$\mathcal{D}$とする．

このとき$\mathcal{D}$は**コーダルグラフ**である．

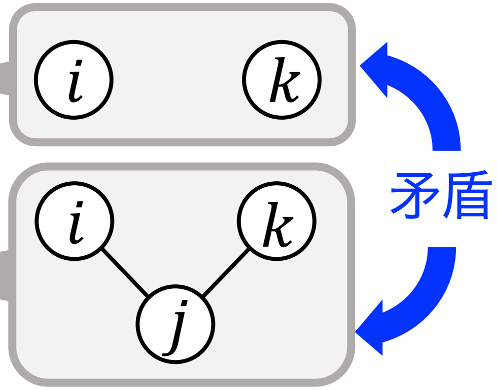
結合していない親ノードがあるとき



- 親ノードが繋がっていないDAG $\mathcal{D}$がMinimal I-mapだったとする.
  1. $\mathcal{D}$のもつ独立性は$I(i,k|\emptyset)$なので$I(i,k|\emptyset)_{\mathcal{G}} \in \mathcal{I}(\mathcal{G})$
- Minimal I-mapなので，エッジを消去したときに生じる独立性は$\mathcal{I}(\mathcal{G})$ に含まれない.
  2. $I(i,j|k)_{\mathcal{G}}, I(i,j|\emptyset)_{\mathcal{G}}, I(j,k|i)_{\mathcal{G}}, I(j,k|\emptyset)_{\mathcal{G}} \notin \mathcal{I}(\mathcal{G})$
- 1, 2を満たす無向グラフ$\mathcal{G}$を作ってみる.

1.の$I(i,k|\emptyset) \in \mathcal{I}(\mathcal{G})$より，$\mathcal{G}$においてi$i$と$k$は分離していなくてはならない

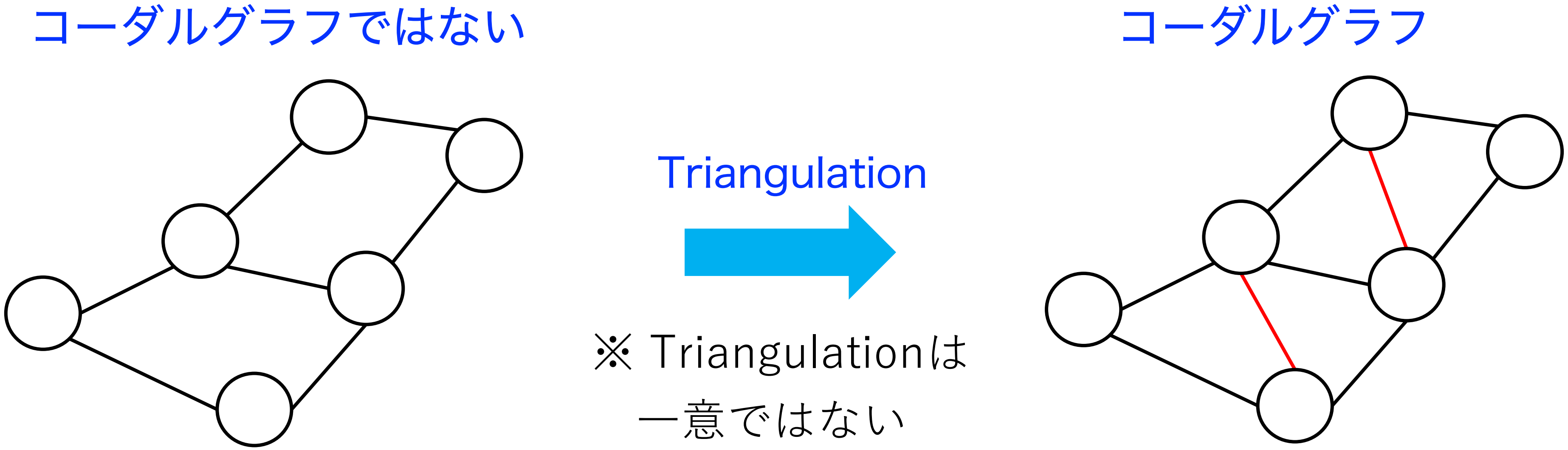
2.の$I(i,j|\emptyset), I(j,k|\emptyset) \notin \mathcal{I}(\mathcal{G})$より，$i$と$j$，$j$と$k$間にはエッジがある



定義：コーダルグラフ(Chordal graph)

長さが４以上の全てのサイクルに

少なくとも１つedgeがあるグラフを**コーダルグラフ**とよぶ



※ Triangulationは一意ではない

- $N=3$のループしかないグラフ，$N=2$でループがないグラフもコーダルグラフとする

コーダルグラフの完全マップ

コーダルグラフ$\mathcal{H}$のもとでのマルコフネットワークモデルについて,

完全マップとなるベイジアンネットワークモデルが存在する

- まず**クリークツリー**を導入する.

定義：Clique Tree

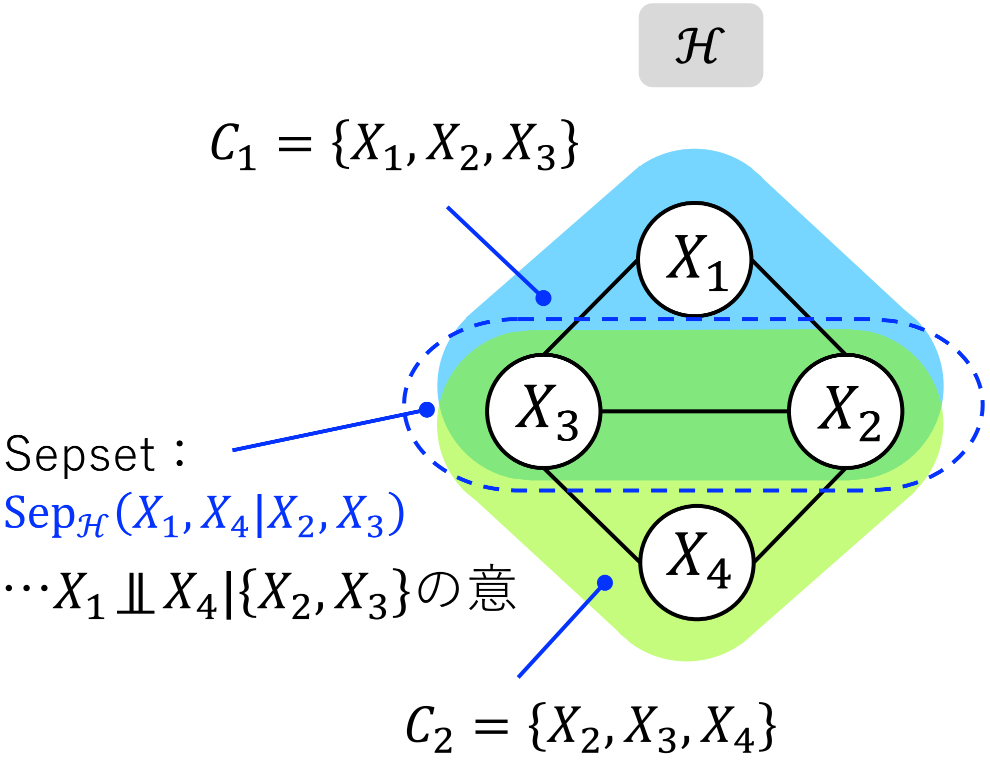
- 無向グラフ$\mathcal{G}$において， i番目のクリークに属す変数集合をC_iとする.
- C_iとC_jのintersectionをSepsetと定義して$\mathbf{Sep}_{ij}$と表記する.

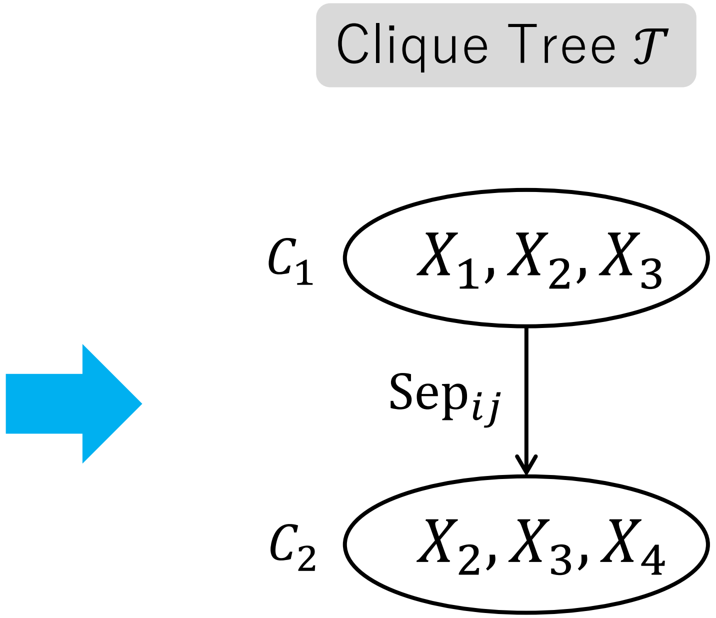
ノード$\{\boldsymbol{C}_i\}$とエッジ$\{\mathbf{Sep}_{ij}\}$を持つツリー$\mathcal{T}$が次の性質を満たすとき,
無向グラフ$\mathcal{G}$に対するクリークツリーとよぶ.

- $\mathcal{G}$の最大クリークは$\mathcal{T}$のノードである.
- $\mathcal{G}$において$\mathbf{Sep}_{ij}$が$\boldsymbol{C}_i$より手前のクリークと$\boldsymbol{C}_j$より後のクリークを分ける.

Clique Treeの例

無向コーダルグラフ$\mathcal{H}$は，対応するクリークツリー$\mathcal{T}$を持つ．

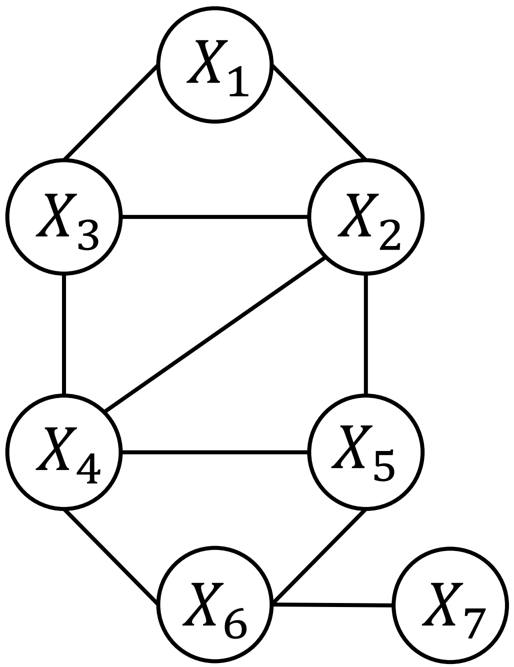




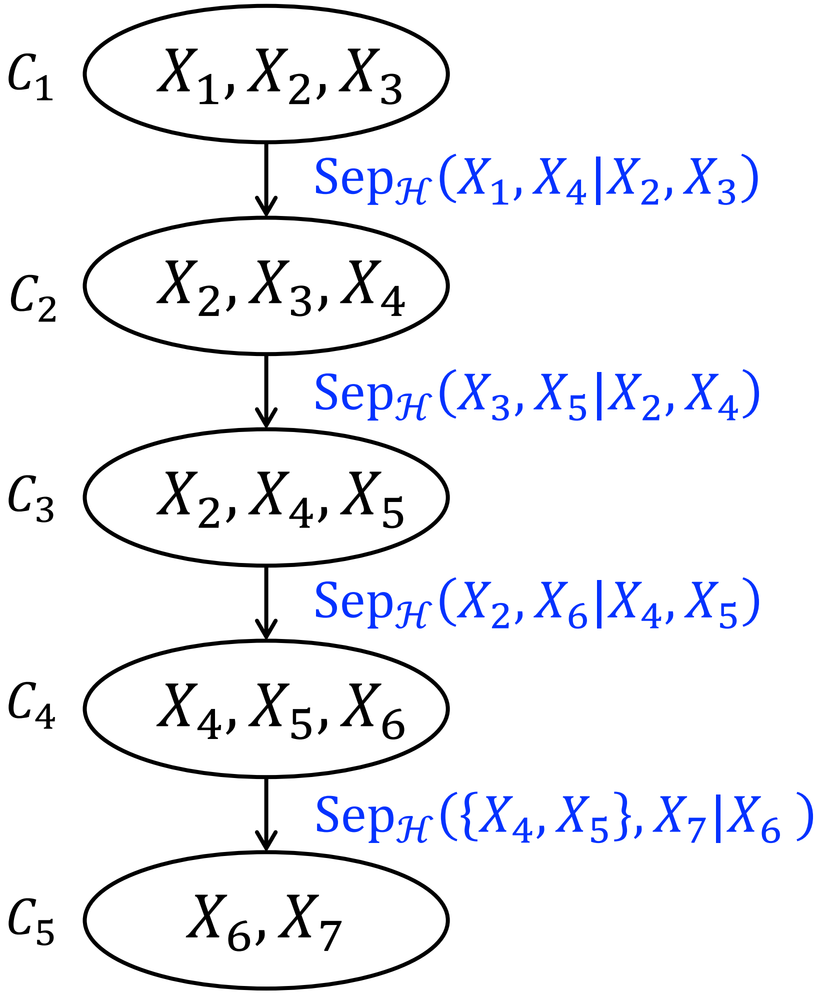
- $\mathrm{Sep}_{ij} = \{X_2, X_3\}$

コーダルグラフ→Clique Tree→BN

コーダルグラフ$\mathcal{H}$



Clique Tree $\mathcal{T}$



対応する
BNを
構成できる

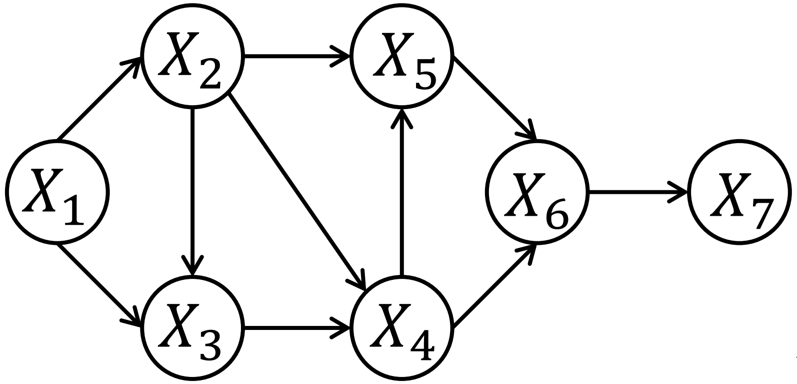
Algorithm: Clique Tree to Bayesian Network

- X_iが最初に登場するクリークを$C_{k(i)}$とする.
- Parent node $\pi(i)$を次のように決める.

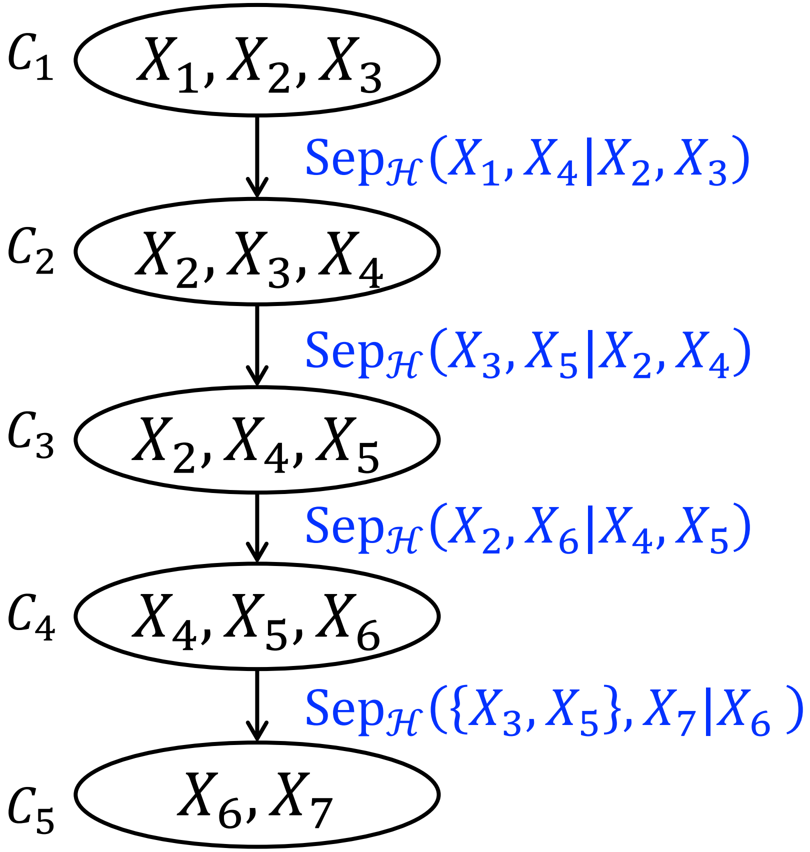
$$\pi(i) = \left(C_{k(i)} \setminus X_i\right) \cap \{X_1, \ldots, X_{i-1}\}$$

✦ $i = 1 : \pi(1) = (C_1 \setminus X_1) \cap \emptyset \rightarrow \pi(1) = \emptyset$
✦ $i = 2 : \pi(2) = (C_1 \setminus X_2) \cap X_1 \rightarrow \pi(2) = X_1$
✦ $i = 3 : \pi(3) = (C_2 \setminus X_3) \cap \{X_1, X_2\} \rightarrow \pi(3) = \{X_1, X_2\}$
…

結果として得られるBN



Clique Tree $\mathcal{T}$



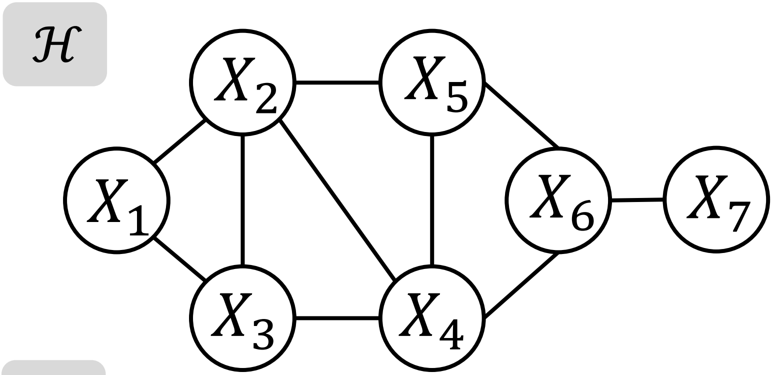
※ C_1に入っている変数から$X_1, X_2, \ldots$とラベルを振る必要がある

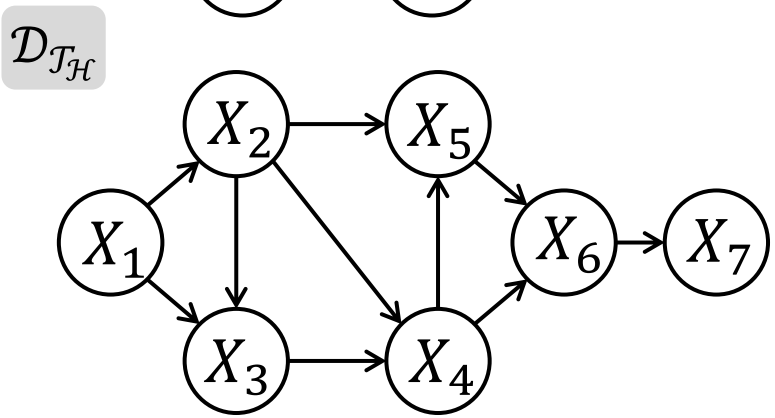
コーダルグラフの完全マップ：まとめ

コーダルグラフ$\mathcal{H}$のもとでのマルコフネットワークモデルについて，

完全マップとなるベイジアンネットワークモデルが存在する

- $\mathcal{H}$のクリークツリー$\mathcal{T}_{\mathcal{H}}$から，
  $\mathcal{H} = \mathcal{M}\left[\mathcal{D}_{\mathcal{T}_{\mathcal{H}}}\right]$となるDAG $\mathcal{D}_{\mathcal{T}_{\mathcal{H}}}$を構成できる．
- $\mathcal{D}_{\mathcal{T}_{\mathcal{H}}}$において，親ノードは全結合である．
- 「有向グラフ$\mathcal{D}$において親ノードが全結合である場合，
  モラル化されたグラフ$\mathcal{M}[\mathcal{D}]$は$\mathcal{D}$のperfect mapである 」
  という定理⏩から，
  $\mathcal{I}(\mathcal{H}) = \mathcal{I}\left(\mathcal{D}_{\mathcal{T}_{\mathcal{H}}}\right)$が言える．





BN→MNとMN→BNのまとめ

- BNからMNへの変形
  - 一般に，$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{M}[\mathcal{D}])$
  - $\mathcal{D}$がコーダルグラフ(親ノードが全結合)である場合は$\mathcal{I}(\mathcal{D}) = \mathcal{I}(\mathcal{M}[\mathcal{D}])$

- MNからBNへの変形
  - 一般に，$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{G})$であるとき，$\mathcal{D}$はコーダルグラフ
  - $\mathcal{G}$がコーダルグラフの場合，$\mathcal{I}(\mathcal{G}) = \mathcal{I}(\mathcal{D})$となる$\mathcal{D}$を作ることができる．

5. ベイジアンネットワーク再訪

$\mathbb{DF} \Rightarrow \mathbb{DG}$の証明とDirected Pairwise Markov Property

復習(再掲)

- $\mathbb{DF} \Rightarrow \mathbb{DG} \Rightarrow \mathbb{DL} \Rightarrow \mathbb{DF}$が示された．よって$\mathbb{DF} \Leftrightarrow \mathbb{DG} \Leftrightarrow \mathbb{DL}$
  - ✦ Factorizeされた確率分布はd分離性から示される条件付き独立性を満たす.
  - ✦ 確率分布がd分離性から示される条件付き独立性を満たすのなら, Factorized formで表現できる
  - ✦ つまり$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(p_{\mathcal{B}})$: $\mathcal{D}$は$p_{\mathcal{B}}$のI-map

    ※ $I(A, B|C)_{\mathcal{D}} \in \mathcal{I}(\mathcal{D}) \Rightarrow I(A, B|C)_{p_{\mathcal{B}}} \in \mathcal{I}(p_{\mathcal{B}})$の意味.

- 証明において用いた仮定：Symmetry, Decomposition, Weak Union, Contraction

- $\mathbb{DG} \Rightarrow \mathbb{DF}$の直接の証明も可能 ⇐ この部分を説明します
  - まずDirected Markov Propertyの別の定義を導入し, d分離性と等価であることを示す.

定義：Ancestral sets

ノードαの祖先ノードを$an(\alpha)$と定義する．

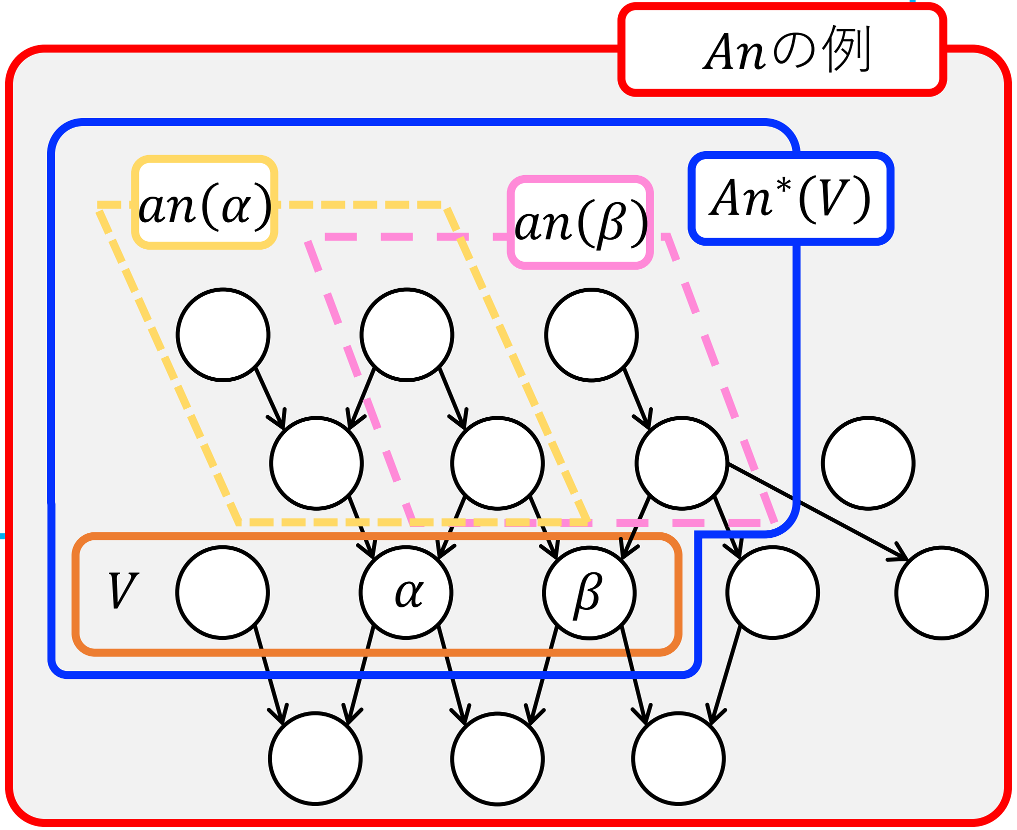
ある集合Anが，すべての$W \in An$について$\pi(W) \in An$となるとき，

AnをAncestral set とよぶ．

またVを含むAncestral set のうち

最小のものを最小Ancestral setとよび，

$An^*(V)$と表記する．



- 定義より$An^*(V) = V \cup_{\alpha \in V} an(\alpha)$，

またAn$^*(\alpha) = \alpha \cup an(\alpha)$である．

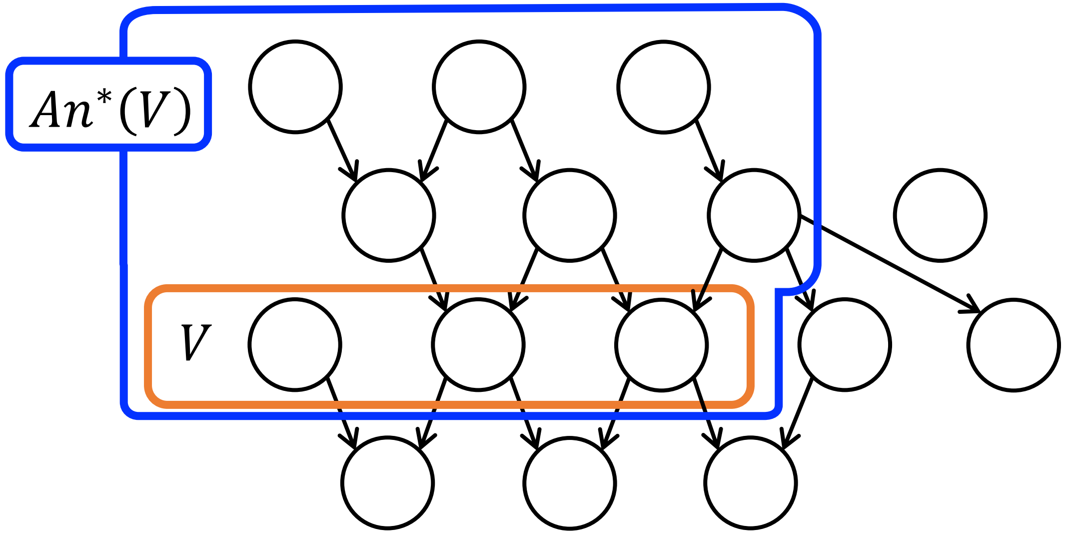
定義：Directed Global Markov Property $\mathbb{DG}'$

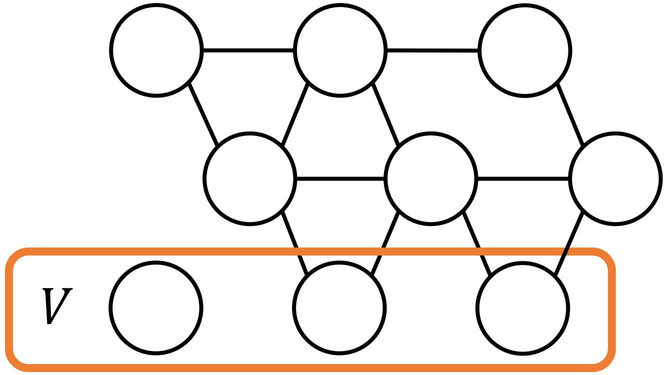
DAG $\mathcal{D}$におけるノード集合A, B, Cについて，

最小ancestral set $An^*(A \cup B \cup C)$からなるサブグラフを$\mathcal{D}_{An^*(A \cup B \cup C)}$とする．

Moral graph $\mathcal{M}\left[\mathcal{D}_{An^*(A \cup B \cup C)}\right]$においてて$I(A, B|C)_{\mathcal{G}} \Rightarrow I(A, B|C)_p$となる性質を

DAG Gに対する**Directed Global Markov Property**とよび$\mathbb{DG}'$と書く．





最小Ancestral set に対する$\mathcal{M}\left[G_{An^*(V)}\right]$

補題：$\mathbb{DF} \Rightarrow \mathcal{M}[\mathcal{D}]$において$\mathbb{G}$

DAG $\mathcal{D}$において$\mathbb{DF}$が成立しているとき，$\mathcal{M}[\mathcal{D}]$において$\mathbb{G}$が成立する

◆ 証明 ◆

「$\mathbb{DF} \Rightarrow \mathcal{M}[\mathcal{D}]$において$\mathbb{F}$が成立する $\Rightarrow \mathcal{M}[\mathcal{D}]$において$\mathbb{G}$が成立する」

という流れで証明する

- $\mathcal{M}[\mathcal{D}]$では，parent nodesは完全グラフを構成する
- $\mathbb{DF}$ が成立するとき，確率分布はparent nodesに条件づけられた分布の積となる
- この分布は$\mathcal{M}[\mathcal{D}]$においてcomplete setごとに分解された分布と解釈できる
- したがって，$\mathcal{M}[\mathcal{D}]$において$\mathbb{F}$が成立しており，すでに示された定理から$\mathbb{G}$が成立する

定理：$\mathbb{DF} \Rightarrow \mathbb{DG}'$

$\mathbb{DF}$が成立しているとする．DAG G上の

disjointなノード集合A, B, Cの最小Ancestral set $An^*(A \cup B \cup C)$に対応する

Moral graph $M[G_{An(A \cup B \cup C)}]$において$\mathbb{G}$が成立する．

◆ 証明 ◆

$$\mathbb{DF}:\ p(\boldsymbol{X}) = \prod_{i \in V} p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right) = \left(\prod_{i \in A_n} p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right)\right)\left(\prod_{i \in V \setminus A_n} p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right)\right)$$

Anはある最小Ancestral set

- Ancestral setの定義から$i \in A_n$ならば$\pi(i) \in A_n$なので

$$p(\boldsymbol{X}_{A_n}) = \sum_{\boldsymbol{X}_{V \setminus A_n}} \prod_{i \in V} p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right) = \prod_{i \in A_n} p\left(X_i \middle| \boldsymbol{X}_{\pi(i)}\right)$$

つまりAncestral set内で$\mathbb{DF}$が成立

- Anによるサブグラフに対しても「DAG $\mathcal{D}$において$\mathbb{DF} \Rightarrow \mathcal{M}[\mathcal{D}]$で$\mathbb{G}$が成立」が適用できる

命題：$\mathbb{DG}'$と$\mathbb{DG}$の等価性

A,B,CをDAG $\mathcal{D}$におけるdisjointなノード集合とする.

このとき$\mathcal{M}\left[\mathcal{D}_{An^*(A\cup B\cup C)}\right]$において$I(A,B|C)_{\mathcal{G}}$あれば,

$\mathcal{D}$において$I(A,B|C)_{\mathcal{D}}$である.

また$\mathcal{D}$において$I(A,B|C)_{\mathcal{D}}$あれば，$\mathcal{M}\left[\mathcal{D}_{An^*(A\cup B\cup C)}\right]$において$I(A,B|C)_{\mathcal{G}}$である.

▶ 証明の概要 (詳細⏩)

- DAG $\mathcal{D}$において： CはAとBをd分離しない $\Leftrightarrow$ AーB間に，Cによりblockされない経路がある

$\Updownarrow$ **ここを示す**

- $\mathcal{M}\left[\mathcal{D}_{An^*(A,B,C)}\right]$において：$C$は$A$と$B$を分離しない $\Leftrightarrow$ AーB間にCを経由しない経路がある

ここまでのまとめ

- マルコフネットワークでは$\mathbb{F} \Leftrightarrow \mathbb{G}$を示すために
  Pairwise Markov Propertyを用いた.

- ベイジアンネットワークの場合には,
  Directed Pairwise Markov Propertyを用いずに
  $\mathbb{DF} \Leftrightarrow \mathbb{DG}$を示すことができる

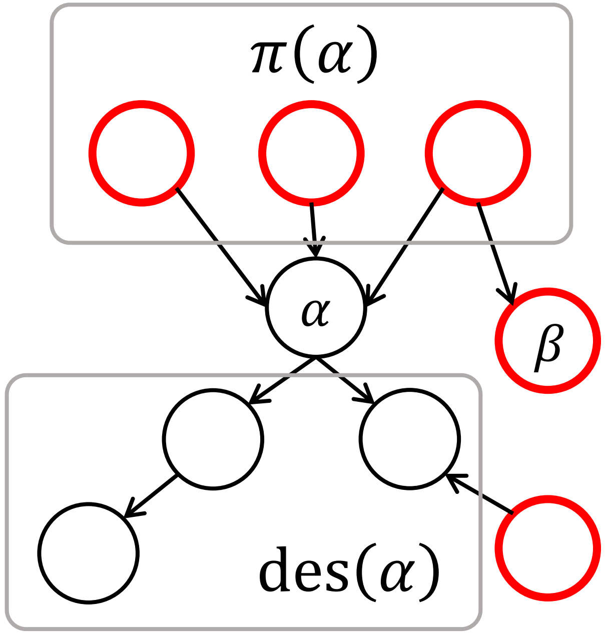
- ここまでIntersection property $(p_{\mathcal{B}}(\boldsymbol{X}) > 0)$を使っていない！
  - Directed Pairwise Markov Propertyが$\mathbb{DF}, \mathbb{DG}$と等価になるにはIntersection propertyが必要

定義：$\mathbb{DP}$

DAG Gのノードαと$\beta \in \mathrm{nd}(\alpha)$について，確率分布$p$が

$\alpha \perp\!\!\!\perp \beta | \mathrm{nd}(\alpha) \setminus \beta$を満たすとき，

確率分布pは**Directed Pairwise Markov property**を持つという．



○：$\mathrm{nd}(\alpha)$

定理：$\mathbb{DL} \Rightarrow \mathbb{DP}$

確率分布pがDAG Gに対して$\mathbb{DL}$を満たすとき，$\mathbb{DP}$も満たす.

◆ 証明 ◆

- $\mathbb{DL}$が成立するとき$\alpha \perp\!\!\!\perp \mathrm{nd}(\alpha)|\pi(\alpha)$
- $\beta \in \mathrm{nd}(\alpha)$に注目すると$\alpha \perp\!\!\!\perp (\mathrm{nd}(\alpha) \setminus \beta \cup \beta)|\pi(\alpha)$
- Weak Union Propertyから

$$\alpha \perp\!\!\!\perp (\mathrm{nd}(\alpha) \setminus \beta \cup \beta)|\pi(\alpha) \Rightarrow \alpha \perp\!\!\!\perp \mathrm{nd}(\alpha) \setminus \beta|\pi(\alpha) \cup \beta \text{ and } \alpha \perp\!\!\!\perp \beta|\pi(\alpha) \cup \mathrm{nd}(\alpha) \setminus \beta$$

- $\pi(\alpha) \in \mathrm{nd}(\alpha)$なので$\alpha \perp\!\!\!\perp \beta|\pi(\alpha) \cup \mathrm{nd}(\alpha) \setminus \beta = \alpha \perp\!\!\!\perp \beta|\mathrm{nd}(\alpha) \setminus \beta$

定理：$\mathbb{DP} \Rightarrow \mathbb{DL}$

確率分布pが$\mathbb{DP}$を満たし，

またintersection propertyを満たすとき※，pは$\mathbb{DL}$も満たす

※ pがstrictly positiveであることと同じ．

◆ 証明 ◆

- $\mathrm{nd}(\alpha) \setminus \pi(\alpha) = \{\beta_1, \beta_2, \dots, \beta_n\}$とする．
- $m \in \{1, \dots, n\}$について$\alpha \perp\!\!\!\perp \{\beta_1 \cup \cdots \cup \beta_m\} | \pi(\alpha) \cup \{\beta_{m+1} \cup \cdots \cup \beta_N\}$が成立するかを検討する
  - ✦ $m = n$は$\mathbb{DL}$に対応

- $\mathbb{DP}$より$\alpha \perp\!\!\!\perp \beta_1 | \pi(\alpha) \cup \{\beta_2 \cup \cdots \cup \beta_N\}$, $\alpha \perp\!\!\!\perp \beta_2 | \pi(\alpha) \cup \{\beta_1 \cup \beta_3 \cup \cdots \cup \beta_N\}$が成立
- Intersection propertyより$A \perp\!\!\!\perp B | (C \cup D)$ and $A \perp\!\!\!\perp C | (B \cup D) \Rightarrow A \perp\!\!\!\perp (B \cup C) | D$なので
- $\alpha \perp\!\!\!\perp (\beta_1 \cup \beta_2) | \pi(\alpha) \cup \{\beta_3 \cup \cdots \cup \beta_N\}$
- この操作を繰り返すと$\mathbb{DL}$が得られる．

まとめ

- 一般のDAG $\mathcal{D}$, 確率分布$p_{\mathcal{B}}$について以下が成立

$$\mathbb{DF} \Leftrightarrow \mathbb{DL} \Leftrightarrow \mathbb{DG} \Leftrightarrow \mathbb{DG}' \Rightarrow \mathbb{DP}$$

- intersection propertyを満たすとき$(p_{\mathcal{B}}(\boldsymbol{X}) > 0)$

$$\mathbb{DF} \Leftrightarrow \mathbb{DL} \Leftrightarrow \mathbb{DG} \Leftrightarrow \mathbb{DG}' \Leftrightarrow \mathbb{DP}$$

6. グラフィカルモデルの展開について

グラフィカルモデルに関連する話題(1)

推定：グラフ構造を活用した効率的推定方法

- 変数消去 (3回目の講義で扱います)
  - 「確率的グラフィカルモデル」1 章, “Probabilistic Graphical models” 9章
- 確率伝搬法 (4回目の講義で扱います)
  - “Probabilistic Graphical models” 10章, 11章

グラフィカルモデルに関連する話題(2)

学習：条件付き確率とグラフ構造をデータから学習する

- パラメータ推定
  - 「確率的グラフィカルモデル」1章, “Probabilistic Graphical models” 17章
- グラフの構造学習
  - 「確率的グラフィカルモデル」2章, “Probabilistic Graphical models” 18章

因果推論：因果関係をDAGにより表す

- 構造方程式モデル
  - 「確率的グラフィカルモデル」3章, 4章, “Probabilistic Graphical models” 21章
  - 黒木学「構造的因果モデルの基礎」共立出版 ⏩

付録 (1)

確率分布と条件付き独立性の性質

⏪ 本文に戻る

定理：確率分布はDecomposition Propertyを満たす

$$\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C \Longrightarrow \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C \text{ and } \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_D | \boldsymbol{X}_C$$

◆ 証明 ◆

- $\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C$なので$p(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) = p(\boldsymbol{X}_A | \boldsymbol{X}_C) p(\boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) \cdots ☆$

1. $\sum_{\boldsymbol{X}_D} p(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) = p(\boldsymbol{X}_A, \boldsymbol{X}_B | \boldsymbol{X}_C)$
2. ☆を用いると$\sum_{\boldsymbol{X}_D} p(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) = p(\boldsymbol{X}_A | \boldsymbol{X}_C) p(\boldsymbol{X}_B | \boldsymbol{X}_C)$

1.2 の比較から$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C$が言える

3. $\sum_{\boldsymbol{X}_B} p(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) = p(\boldsymbol{X}_A, \boldsymbol{X}_D | \boldsymbol{X}_C)$
4. ☆を用いると$\sum_{\boldsymbol{X}_B} p(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_D | \boldsymbol{X}_C) = p(\boldsymbol{X}_A | \boldsymbol{X}_C) p(\boldsymbol{X}_D | \boldsymbol{X}_C)$

3.4 の比較から$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_D | \boldsymbol{X}_C$が言える

定理：確率分布はWeak Union Propertyを満たす

$$\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C \Longrightarrow \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | (\boldsymbol{X}_C \cup \boldsymbol{X}_D) \text{ and } \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_D | (\boldsymbol{X}_C \cup \boldsymbol{X}_B)$$

◆ 証明 ◆

- $\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C$から$p(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_C, \boldsymbol{X}_D) = p(\boldsymbol{X}_A | \boldsymbol{X}_C)$
- またDecomposition Propertyを用いると
  $\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C$ から$p(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D) = p(\boldsymbol{X}_A | \boldsymbol{X}_C)$が言える
- よって$p(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_C, \boldsymbol{X}_D) = p(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D)$となり$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | (\boldsymbol{X}_C \cup \boldsymbol{X}_D)$.
- またDecomposition Propertyを用いると
  $\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_D) | \boldsymbol{X}_C$ から$p(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_B) = p(\boldsymbol{X}_A | \boldsymbol{X}_C)$が言える
- よって$p(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_C, \boldsymbol{X}_D) = p(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_C)$となり$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_D | (\boldsymbol{X}_B \cup \boldsymbol{X}_C)$.

定理：確率分布$\boldsymbol{p(X)} > \mathbf{0}$はIntersection propertyを満たす

Strictly positiveな確率分布について,

$$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | (\boldsymbol{X}_C \cup \boldsymbol{X}_D) \text{ and } \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_C | (\boldsymbol{X}_B \cup \boldsymbol{X}_D) \Longrightarrow \boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_C) | \boldsymbol{X}_D$$

◆ 証明 ◆

- $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | (\boldsymbol{X}_C \cup \boldsymbol{X}_D)$なので
  $P(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_C | \boldsymbol{X}_D) = P(\boldsymbol{X}_A, \boldsymbol{X}_B | \boldsymbol{X}_C, \boldsymbol{X}_D) P(\boldsymbol{X}_C | \boldsymbol{X}_D) \to P(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D) P(\boldsymbol{X}_B | \boldsymbol{X}_C, \boldsymbol{X}_D) P(\boldsymbol{X}_C | \boldsymbol{X}_D)$
- $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_C | (\boldsymbol{X}_B \cup \boldsymbol{X}_D)$なので
  $P(\boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_C | \boldsymbol{X}_D) = P(\boldsymbol{X}_A, \boldsymbol{X}_C | \boldsymbol{X}_B, \boldsymbol{X}_D) P(\boldsymbol{X}_B | \boldsymbol{X}_D) \to P(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_D) P(\boldsymbol{X}_C | \boldsymbol{X}_B, \boldsymbol{X}_D) P(\boldsymbol{X}_B | \boldsymbol{X}_D)$

$P(\boldsymbol{X}_B | \boldsymbol{X}_C, \boldsymbol{X}_D) \neq 0$等を仮定すると$P(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D) = P(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_D)$. ここから次のことが言える

✦ $P(\boldsymbol{X}_A | \boldsymbol{X}_D) = \sum_{\boldsymbol{X}_B} P(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_D) P(\boldsymbol{X}_B | \boldsymbol{X}_D) = \sum_{\boldsymbol{X}_B} P(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D) P(\boldsymbol{X}_B | \boldsymbol{X}_D) = P(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D)$

✦ $P(\boldsymbol{X}_A | \boldsymbol{X}_D) = \sum_{\boldsymbol{X}_C} P(\boldsymbol{X}_A | \boldsymbol{X}_C, \boldsymbol{X}_D) P(\boldsymbol{X}_C | \boldsymbol{X}_D) = \sum_{\boldsymbol{X}_C} P(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_D) P(\boldsymbol{X}_C | \boldsymbol{X}_D) = P(\boldsymbol{X}_A | \boldsymbol{X}_B, \boldsymbol{X}_D)$

よって$\boldsymbol{X}_A \perp\!\!\!\perp (\boldsymbol{X}_B \cup \boldsymbol{X}_C) | \boldsymbol{X}_D$

付録 (2)

多変量正規分布がCompositional Graphoidであることを示す

⏪ 本文に戻る

準備：多変量ガウス変数の線形変換

多変量ガウス分布に従う変数を線形変換した変数も 多変量ガウス分布に従う．
具体的に，$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\xi}, \Sigma)$に対して$\boldsymbol{Y} = L\boldsymbol{X} + \boldsymbol{b}$は$\mathcal{N}\left(L\boldsymbol{\xi} + \boldsymbol{b}, L\Sigma L^{\mathrm{T}}\right)$に従う

- $L \in \mathbb{R}^{r \times d}, \boldsymbol{b} \in \mathbb{R}^{r}$とする．$L$は横長・縦長行列でも良い．

▶ 証明の概要

- 「変数aとbのモーメント母関数が一致するとき，
  変数aとbは同じ確率分布に従う」という性質を使う

◆ 証明 ◆

- $\boldsymbol{X}$のモーメント母関数：$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left[\exp\left(\boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}\right)\right] = \exp\left(\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\Sigma\boldsymbol{t} - \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{t}\right)$

- $\boldsymbol{Y}$のモーメント母関数$M_{\boldsymbol{Y}}(\boldsymbol{t}) = \exp\left(\boldsymbol{t}^{\mathrm{T}}\boldsymbol{b}\right) E[\exp(\boldsymbol{t}^{\mathrm{T}}L\boldsymbol{X})]$は次のように与えられる

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = \exp\left(\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}L\Sigma L^{\mathrm{T}}\boldsymbol{t} - (L\boldsymbol{\xi} + \boldsymbol{b})^{\mathrm{T}}\boldsymbol{t}\right) \quad \cdots \mathcal{N}\left(L\boldsymbol{\xi} + \boldsymbol{b}, L\Sigma L^{\mathrm{T}}\right)\text{のモーメント母関数と一致}$$

定理：多変量ガウス分布はCompositional Graphoid

- 多変量ガウス分布（変数の和をNとする）

$$p(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\xi})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\xi})\right)$$

  - 精度行列$K = \Sigma^{-1}$を用いた場合は

$$p(\boldsymbol{X}) = \sqrt{\frac{|K|}{(2\pi)^N}} \exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\xi})^{\mathrm{T}}K(\boldsymbol{X}-\boldsymbol{\xi})\right)$$

▶ 証明の概要

- 確率変数をDisjointなsubset A, B ($A \cup B = V$)に分け，次の表現を導入する

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_A \\ \boldsymbol{\xi}_B \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}, \qquad K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}$$

(1) 「$\boldsymbol{X}$が多変量正規分布に従うとき，$\boldsymbol{Y} = L\boldsymbol{X} + \boldsymbol{b}$も多変量正規分布に従う」を利用して周辺化分布を求める

- 確率変数をDisjointなsubset A, B $(A \cup B = V)$に分け，次の表現を導入する

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_A \\ \boldsymbol{\xi}_B \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}$$

  - $X \in \mathbb{R}^N, X_A \in \mathbb{R}^{N_A}, X_B \in \mathbb{R}^{N_B}, N_A + N_B = N$とする
  - $\boldsymbol{\xi}_A \in \mathbb{R}^N$は$\boldsymbol{X}_A = \{X_i | i \in A\}$の平均，$\boldsymbol{\xi}_B \in \mathbb{R}^N$は$\boldsymbol{X}_B \in \{X_i | i \in B\}$の平均
  - $\Sigma_{AA} \in \mathbb{R}^N$は$\boldsymbol{X}_A$内の分散共分散行列，$\Sigma_{AB} \in \mathbb{R}^N$は$\boldsymbol{X}_A, \boldsymbol{X}_B$間の分散共分散行列，

- $L = \begin{array}{cc} \begin{pmatrix} 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} & N_A \\ \quad\underbrace{\hphantom{0\ \cdots\ 0}}_{N_A}\ \underbrace{\hphantom{0\ \cdots\ 0}}_{N_B} & \end{array}$とすると$\boldsymbol{X}_B = L\boldsymbol{X}$なので周辺化分布を得ることができる．

$$p(\boldsymbol{X}_B) = \frac{1}{\sqrt{2\pi|\Sigma_{BB}|}} \exp\left(-\frac{1}{2} (\boldsymbol{X}_B - \boldsymbol{\xi}_B)^T \Sigma_{BB}^{-1} (\boldsymbol{X}_B - \boldsymbol{\xi}_B) \right)$$

(2) 条件付き確率の導出

$p(\boldsymbol{X}_A|\boldsymbol{X}_B) = \frac{p(\boldsymbol{X}_A, \boldsymbol{X}_B)}{p(\boldsymbol{X}_B)}$なので

$$p(\boldsymbol{X}_A|\boldsymbol{X}_B) \propto \exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\xi})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\xi}) + \frac{1}{2}(\boldsymbol{X}_B-\boldsymbol{\xi}_B)^T\Sigma_{BB}^{-1}(\boldsymbol{X}_B-\boldsymbol{\xi}_B)\right)$$

- Concentration matrix $K = \Sigma^{-1}$を定義し，分散教分散行列と同様に $K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}$とする．

$$p(\boldsymbol{X}_A|\boldsymbol{X}_B = \boldsymbol{x}_B) \propto \exp\left(-\frac{1}{2}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)^{\mathrm{T}}K_{AA}(\boldsymbol{X}_A-\boldsymbol{\xi}_A) + (\boldsymbol{x}_B-\boldsymbol{\xi}_B)^{\mathrm{T}}K_{AB}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)\right)$$

$$\propto \exp\left(-\frac{1}{2}\boldsymbol{X}_A^T K_{AA}\boldsymbol{X}_A + \boldsymbol{\xi}_A^{\mathrm{T}}K_{AA}\boldsymbol{X}_A + (\boldsymbol{x}_B-\boldsymbol{\xi}_B)^{\mathrm{T}}K_{AB}\boldsymbol{X}_A\right)$$

規格化すると

$$p(\boldsymbol{X}_A|\boldsymbol{X}_B = \boldsymbol{x}_B) = \frac{1}{\sqrt{2\pi\left|K_{A|B}\right|}}\exp\left(-\frac{1}{2}\left(\boldsymbol{X}_A-\boldsymbol{\xi}_{A|B}\right)^{\mathrm{T}}K_{A|B}^{-1}\left(\boldsymbol{X}_A-\boldsymbol{\xi}_{A|B}\right)\right)$$

$\xi_{A|B} = \xi_A - K_{AA}^{-1}K_{AB}(x_B-\xi_B), K_{A|B} = K_{AA}^{-1}$となる

$\Sigma_{AA}, \Sigma_{AB}, \Sigma_{BB}$とも正則とする

$$K_{AA}^{-1} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \qquad K_{AA}^{-1}K_{AB} = -\Sigma_{AB}\Sigma_{BB}^{-1}$$

(3) 条件付き独立性

- $p(\boldsymbol{X}_A|\boldsymbol{X}_B=\boldsymbol{x}_B)=p(\boldsymbol{X}_A)$のとき$\boldsymbol{X}_A$は独立

$$p(\boldsymbol{X}_A|\boldsymbol{X}_B=\boldsymbol{x}_B)=\frac{1}{\sqrt{2\pi|K_{AA}|}}\exp\left(-\frac{1}{2}\left(\boldsymbol{X}_A-\boldsymbol{\xi}_{A|B}\right)^{\mathrm{T}}K_{A|B}^{-1}\left(\boldsymbol{X}_A-\boldsymbol{\xi}_{A|B}\right)\right)$$

いずれにしろガウス分布

$$\xi_{A|B}=\xi_A+\Sigma_{AB}\Sigma_{BB}^{-1}(x_B-\xi_B),\qquad K_{A|B}=\Sigma_{AA}-\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}$$

$$p(\boldsymbol{X}_A)=\frac{1}{\sqrt{2\pi|\Sigma_{AA}|}}\exp\left(-\frac{1}{2}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)^T\Sigma_{AA}^{-1}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)\right)$$

二つの分布の比較から$\Sigma_{AB}=0\Rightarrow X_A\perp\!\!\!\perp X_B$

- 一方で$X_A\perp\!\!\!\perp X_B$であるなら

$$p(\boldsymbol{X}_A,X_B)\propto\exp\left(-\frac{1}{2}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)^T\Sigma_{AA}^{-1}(\boldsymbol{X}_A-\boldsymbol{\xi}_A)-\frac{1}{2}(\boldsymbol{X}_B-\boldsymbol{\xi}_B)^T\Sigma_{BB}^{-1}(\boldsymbol{X}_B-\boldsymbol{\xi}_B)\right)$$

$K=\begin{pmatrix}\Sigma_{AA}^{-1} & 0\\ 0 & \Sigma_{BB}^{-1}\end{pmatrix}$このとき$\Sigma=K^{-1}=\begin{pmatrix}\Sigma_{AA} & 0\\ 0 & \Sigma_{BB}\end{pmatrix}$　　$X_A\perp\!\!\!\perp X_B\Rightarrow\Sigma_{AB}=0$

(3) 一般化

- Conditionされた多変量正規分布は 多変量正規分布であるため
- X_Cについてconditionされた分布を次のように置く

$$p(\boldsymbol{X}_{\backslash C} | \boldsymbol{X}_C = \boldsymbol{x}_C) = \frac{1}{\sqrt{2\pi|\Sigma^{\backslash C}|}} \exp\left(-\frac{1}{2}(\boldsymbol{X}_{\backslash C} - \boldsymbol{\xi}^{\backslash C})^{\mathrm{T}} {\Sigma^{\backslash C}}^{-1} (\boldsymbol{X}_{\backslash C} - \boldsymbol{\xi}^{\backslash C})\right)$$

$$a = \{A, B\}$$

$$b = \{A, D\}$$

$$p(X_A, \boldsymbol{X}_B) = \frac{1}{\sqrt{2\pi|\Sigma_a|}} \exp\left(-\frac{1}{2}(\boldsymbol{X}_a - \boldsymbol{\xi}_a)^T \Sigma_{aa}^{-1} (\boldsymbol{X}_a - \boldsymbol{\xi}_a)\right)$$

$$p(\boldsymbol{X}_A | X_B = x_B, \boldsymbol{X}_C = \boldsymbol{x}_C) = \frac{1}{\sqrt{2\pi\left|K_{AA}^{\backslash C}\right|}} \exp\left(-\frac{1}{2}\left(\boldsymbol{X}_A - \boldsymbol{\xi}_{A|B}^{\backslash C}\right)^{\mathrm{T}} {K_{A|B}^{\backslash C}}^{-1} \left(\boldsymbol{X}_A - \boldsymbol{\xi}_{A|B}^{\backslash C}\right)\right)$$

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow \Sigma_{AB}^{\backslash C} = 0$$

$$\Sigma_{AB}^{\backslash C} = 0 \text{ and } \Sigma_{AD}^{\backslash C} = 0 \Rightarrow X_A \perp\!\!\!\perp X_B \cup X_D | X_C$$

$$p(\boldsymbol{X}_a, X_b | \boldsymbol{X}_B) \propto \exp\left(-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\xi})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\xi}) + \frac{1}{2}(\boldsymbol{X}_B - \boldsymbol{\xi}_B)^T \Sigma_{BB}^{-1} (\boldsymbol{X}_B - \boldsymbol{\xi}_B)\right)$$

$d = B \cup D$とすれば$\Sigma_A^{\backslash C} = 0$

付録 (3)

$\mathbb{DF} \Rightarrow \mathbb{DG}$の証明

⏪ 本文に戻る

目標　DAG $\mathcal{D}$において，disjoint set A, B, Cがあり$I(A, B|C)_{\mathcal{D}}$とする．
このとき$\mathbb{DF}$を満たすpについて，$I(X_A, X_B|X_C)_p$が成立することを示す．

- 方針：
  - Topological Sortingによりノードを並べる
    - ✦ 末端ノードには子孫がいない
  - $\mathcal{D}$から末端ノードωを取り除いたDAG $\mathcal{D}_{\backslash\omega}$において，
    $I(A', B'|C')_{\mathcal{D}_{\backslash\omega}} \Rightarrow I(X_{A'}, X_{B'}|X_{C'})_p$が成立しているとする
    - ✦ $\mathcal{D}_{\backslash\omega}$のノード数が1の場合は自明に$\mathbb{DF} \Rightarrow \mathbb{DG}$が成立
  - ωに対する$\mathbb{DF}$と$I(A', B'|C')_{\mathcal{D}_{\backslash\omega}} \Rightarrow I(X_{A'}, X_{B'}|X_{C'})_p$から，$I(A, B|C)_{\mathcal{D}} \Rightarrow I(X_A, X_B|X_C)_p$を示す．
    - ✦ つまりノード数$k-1$で$\mathbb{DF} \Rightarrow \mathbb{DG}$が成立していれば$k$でも成立する

- 前提をまとめると　(i) $I(A, B\ |C)_{\mathcal{D}}$である
  (ii) $I(A', B'|C')_{\mathcal{D}_{\backslash\omega}} \Leftrightarrow I(X_{A'}, X_{B'}|X_{C'})_p$が成立している

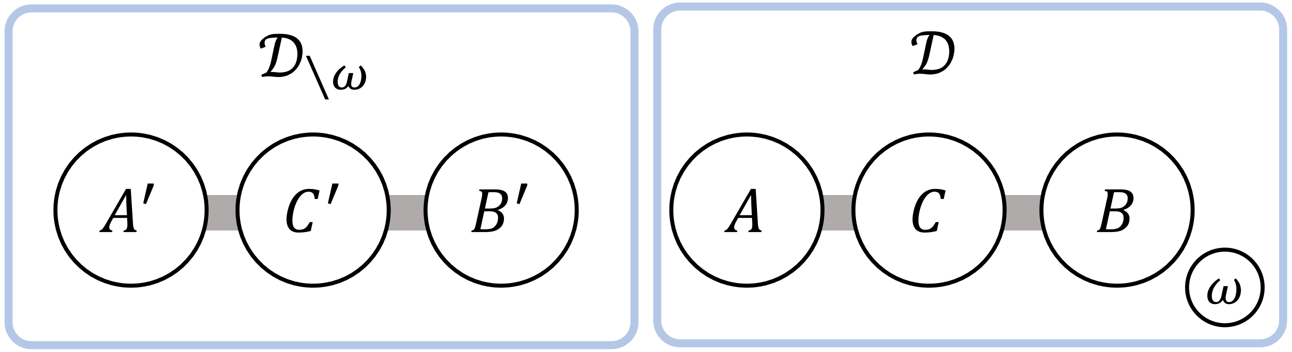
- ωとA, B, Cの関係により次の3通りを考える
  (1) ωがA, B, Cに含まれない場合
  (2) $\omega \in A$の場合 ($\omega \in B$の場合も同様)
  (3) ωがCに含まれる場合

(1) ωがA, B, Cに含まれない場合

$A = A', B = B', C = C'$なので，直ちに

$I(A, B|C)_{\mathcal{D}} \Leftrightarrow I(X_A, X_B|X_C)_p$

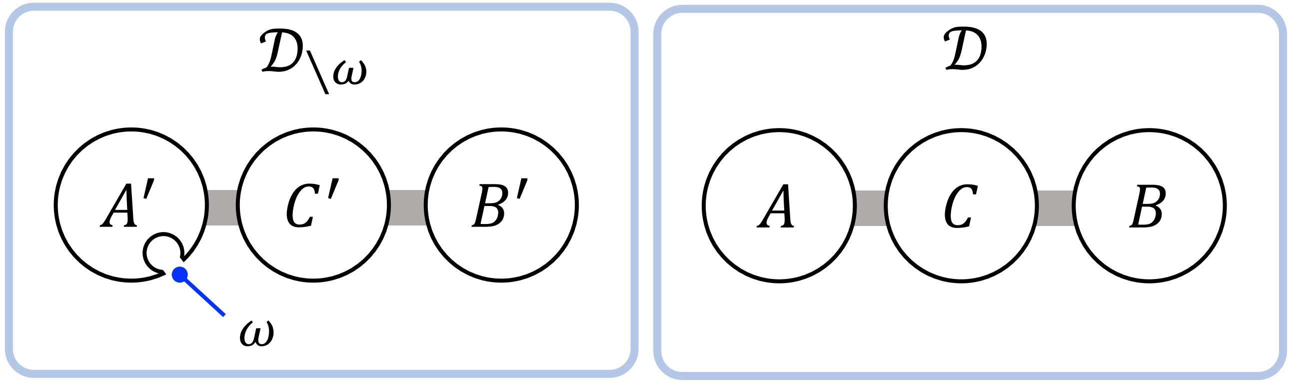
(1)の場合の$\mathcal{D}_{\backslash\omega}$と$\mathcal{D}$



(2) $\omega \in A$の場合（$\omega \in B$の場合も同様）

- $A = A' \cup \omega$
- $I(A, B|C)_{\mathcal{D}} \Rightarrow I(\omega, B|C)_{\mathcal{D}}, \pi(\omega) \notin B$

(2)の場合の$\mathcal{D}_{\backslash\omega}$と$\mathcal{D}$



Chain ruleより

$$\begin{aligned} p(A, B, C) &= p(\omega, A', B, C) \\ &= p(\omega|A', B, C)p(A', B, C) \\ &= p(\omega|A', C)p(A', B, C) \quad \cdots \pi(\omega) \notin B \text{なので}\mathbb{DF}\text{を適用} \end{aligned}$$

よって$I(X_\omega, X_B|X_{A'} \cup X_C)_p$.

$I(X_{A'}, X_{B'}|X_{C'})_p$と合わせると，Contractionより$I(X_{A'} \cup X_\omega, X_B'|X_C')_p = I(X_A, X_B|X_C)_p$

(3) ωがCに含まれる場合

- $C = C' \cup \omega$
- $I(A,B|C' \cup \omega)_{\mathcal{D}}, I(A,B|C')_p, I(A,B|C')_{\mathcal{D}_{\backslash\omega}}$

✦ $I(A,B|C')_{\mathcal{D}_{\backslash\omega}}$ と$I(A,B|C' \cup \omega)_{\mathcal{D}}$ の両立は，$A-B$間経路で，ωにブロックされるものはないことを意味する

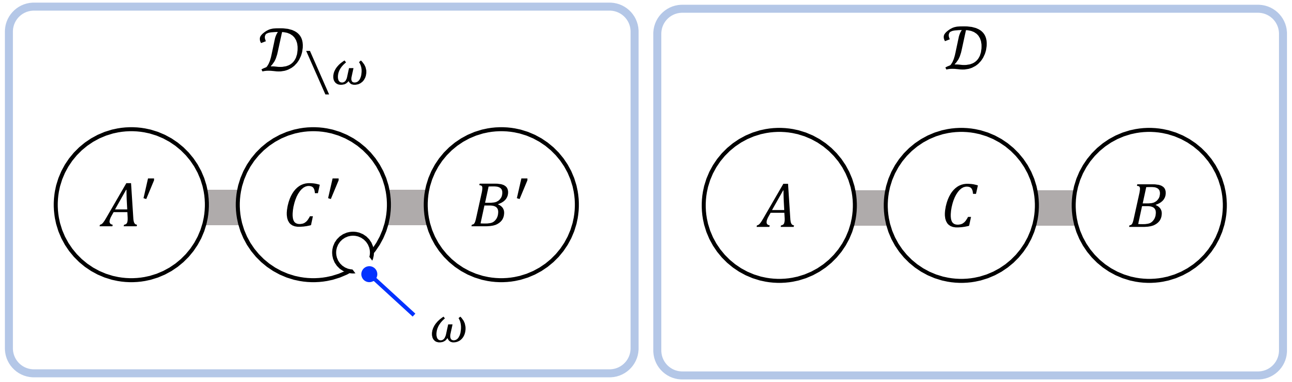
✦ ωはC'によってAまたはBとd分離される必要がある(右図(b,c,d)参照)

✦ Chain ruleより

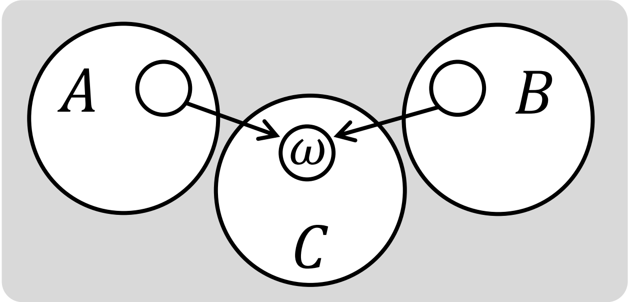
$$p(A,B,C) = p(\omega|A,B,C')p(A,B|C')p(C')$$

- (b)の場合 $p(\omega|A,B,C') = p(\omega|C')$
  - ☐ Weak Unionより
    $I(\omega, A \cup B|C')_p \Rightarrow I(\omega, B|A \cup C')_p$
  - ☐ Contractionより
    $I(A,\omega|C')_p$ and $I(A,B|C' \cup \omega)_p$
    $\Rightarrow I(A \cup \omega, B|C')_p \Rightarrow I(A,B|C' \cup \omega)_p$
    $= I(A,B|C)_p$
- (c),(d)の場合も同様に示すことができる．

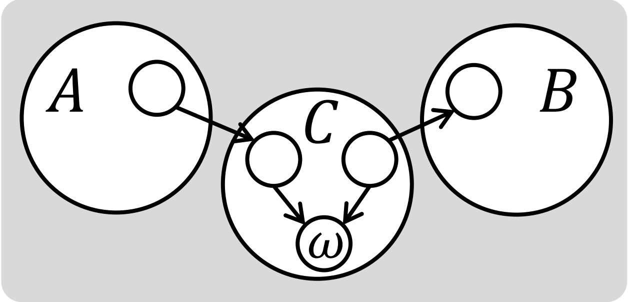
(3)の場合の$\mathcal{D}_{\backslash\omega}$と$\mathcal{D}$



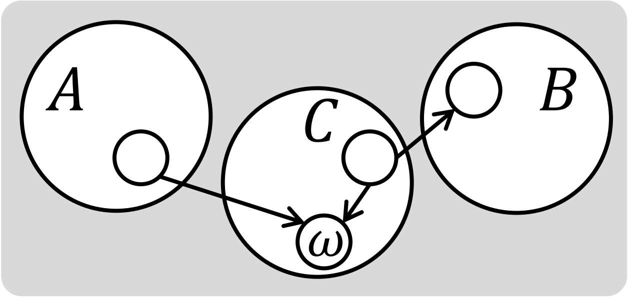
(a) $I(A,B|C)_{\mathcal{D}}$と矛盾



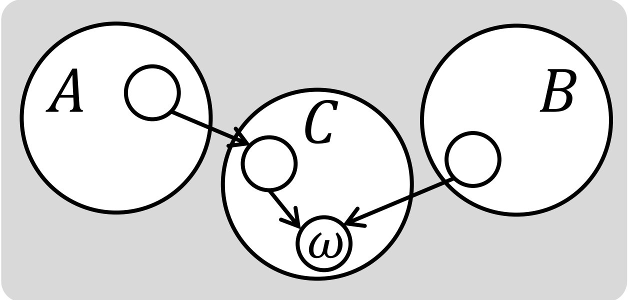
(b) $I(A,B|C)_{\mathcal{D}}$と矛盾しない
：$I(A \cup B, \omega|C')_{\mathcal{D}}$



(c) $I(A,B|C)_{\mathcal{D}}$と矛盾しない
：$I(B,\omega|C')_{\mathcal{D}}$



(d) $I(A,B|C)_{\mathcal{D}}$と矛盾しない
：$I(A,\omega|C')_{\mathcal{D}}$



付録 (4)

$\mathbb{F} \Rightarrow \mathbb{G}$の証明

⏪ 本文に戻る

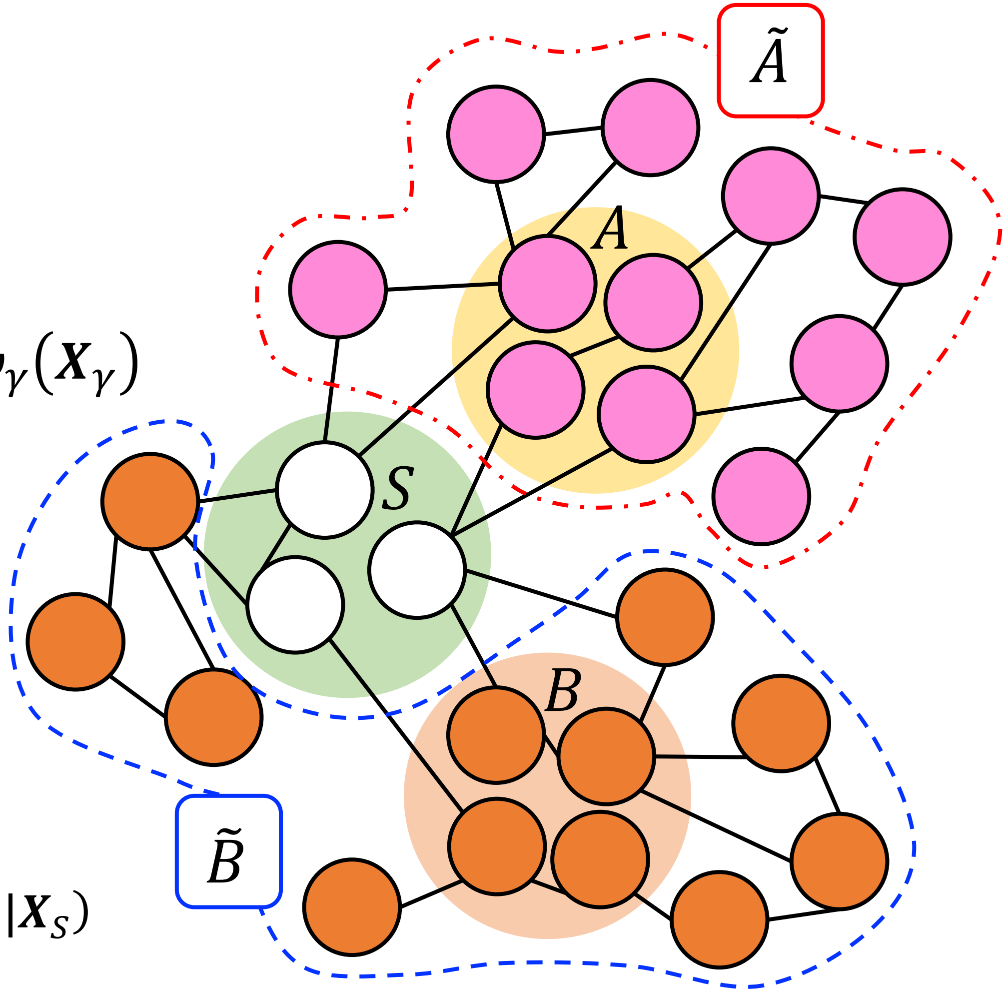
- disjointなノード集合A,B,Sを考える．
- $V \setminus S$のうち，Aを含むconnected componentを$\tilde{A}$とする $(B \not\subset \tilde{A})$
- $\tilde{B} = V \setminus (\tilde{A} \cup S)$とする．$B \subset \tilde{B}$となる．また$V = \tilde{A} \cup \tilde{B} \cup S$.
- Sは$\tilde{A}$と$\tilde{B}$を分離するので，$\mathcal{G}$上の全クリークは
  $\tilde{A} \cup S$または$\tilde{B} \cup S$のどちらか一方のノードを含む
- $\mathbb{F}$より $P(\boldsymbol{X}) = P(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S) = \frac{1}{Z} \prod_{C \in \gamma(\tilde{A} \cup S)} \psi_\gamma(\boldsymbol{X}_\gamma) \prod_{C \in \gamma(\tilde{B} \cup S)} \psi_\gamma(\boldsymbol{X}_\gamma)$

$$\equiv f(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_S) g(\boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S)$$

- Factorizeされた形を利用すると

$$P(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_{\tilde{B}} | \boldsymbol{X}_S) = \frac{P(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S)}{\sum_{\boldsymbol{X}_{\tilde{A}}} \sum_{\boldsymbol{X}_{\tilde{B}}} P(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S)}$$

$$= \frac{f(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_S) g(\boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S)}{\sum_{\boldsymbol{X}_{\tilde{A}}} f(\boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_S) \sum_{\boldsymbol{X}_{\tilde{B}}} g(\boldsymbol{X}_{\tilde{B}}, \boldsymbol{X}_S)} \equiv p(\boldsymbol{X}_{\tilde{A}} | \boldsymbol{X}_S) p(\boldsymbol{X}_{\tilde{B}} | \boldsymbol{X}_S)$$



- よって$\boldsymbol{X}_{\tilde{A}} \perp\!\!\!\perp \boldsymbol{X}_{\tilde{B}} | \boldsymbol{X}_S$．さらに$\boldsymbol{X}_A \subset \boldsymbol{X}_{\tilde{A}}, \boldsymbol{X}_B \subset \boldsymbol{X}_{\tilde{B}}$なので$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_S$．（Decomposition Property）

付録 (5)

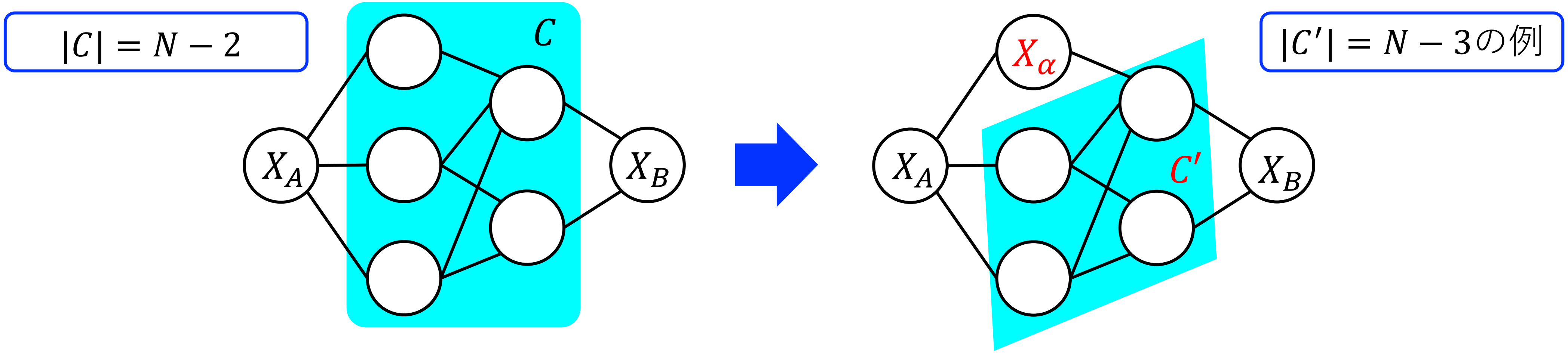
Intersection Propertyが成立するとき$\mathbb{P} \Rightarrow \mathbb{G}$の証明

⏪ 本文に戻る

- $I(A,B|C)_G$を満たすdisjointなノード集合A,B,Cを考える

▶ 証明の概要

- $|C| = N-2$の場合($|A| = |B| = 1$), $\mathbb{P} \Rightarrow \mathbb{G}$は直ちに成立。
- $|C| > n$で$\mathbb{P} \Rightarrow \mathbb{G}$が成立しているとして、 $|C| = n$ で$\mathbb{P} \Rightarrow \mathbb{G}$が成立するかどうかを考える
- これを繰り返すことで,
  任意のdisjointなノード集合に対して$I(A,B|C)_G$が成立することを示す



$\mathbb{P} \Rightarrow \mathbb{G}$が成立

$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C$

$\mathbb{P} \Rightarrow \mathbb{G}$が成立する？

$\boldsymbol{X}_{A'} \perp\!\!\!\perp \boldsymbol{X}_{B'} | \boldsymbol{X}_{C'}$?

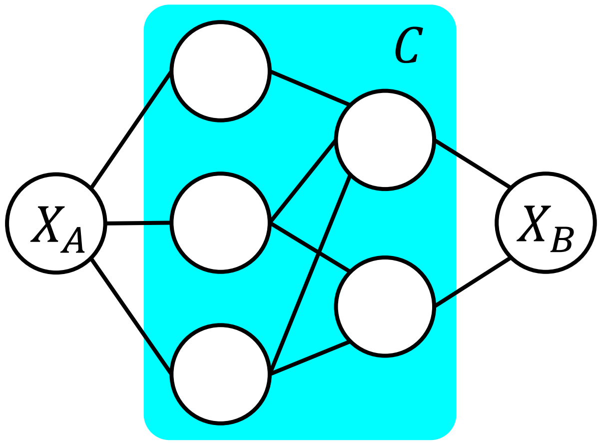
- 新しいノード集合をA', B', C'とする

2通りの可能性が考えられる

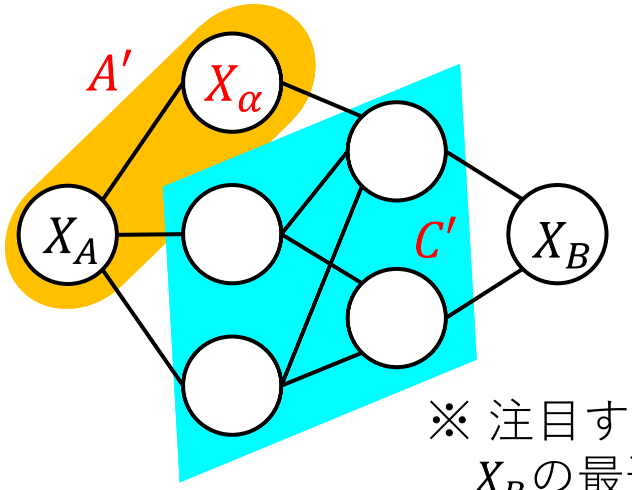
(1) $A' \cup B \cup C' = V$ (または$A \cup B' \cup C' = V$)

- Cの一部をAまたはBに加える

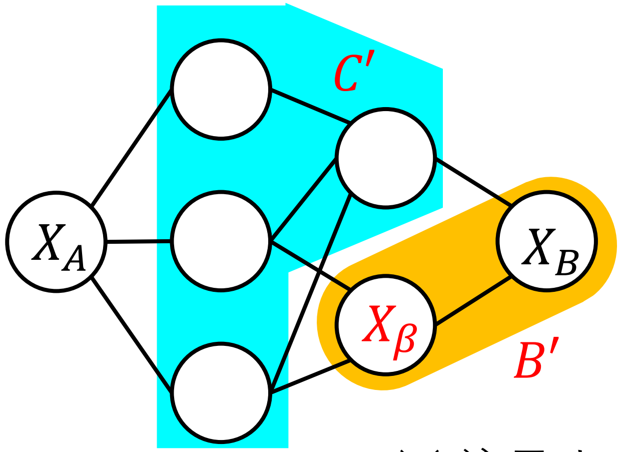
$|C| = N - 2$



$|C'| = N - 3$の例



※ 注目するαが
X_Bの最近傍でない場合



※ 注目するβが
X_Aの最近傍でない場合

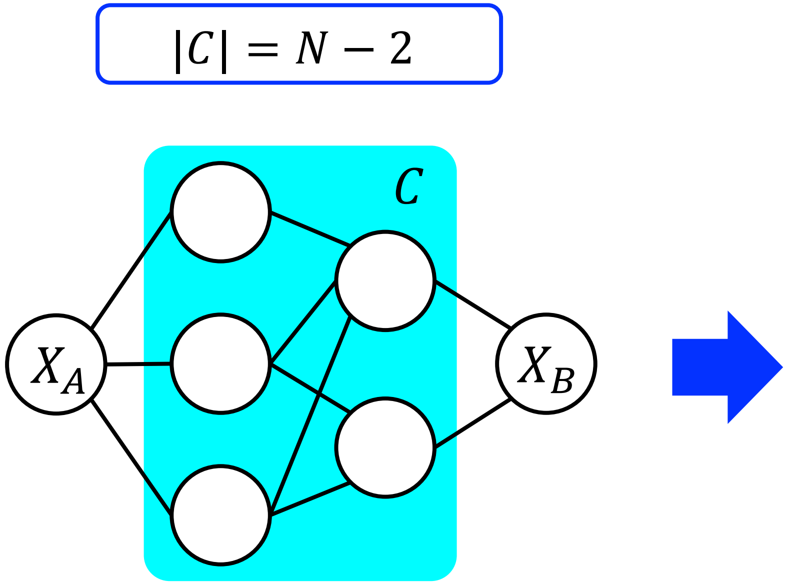
(右上の，A', B, C'とした例のみ示す)

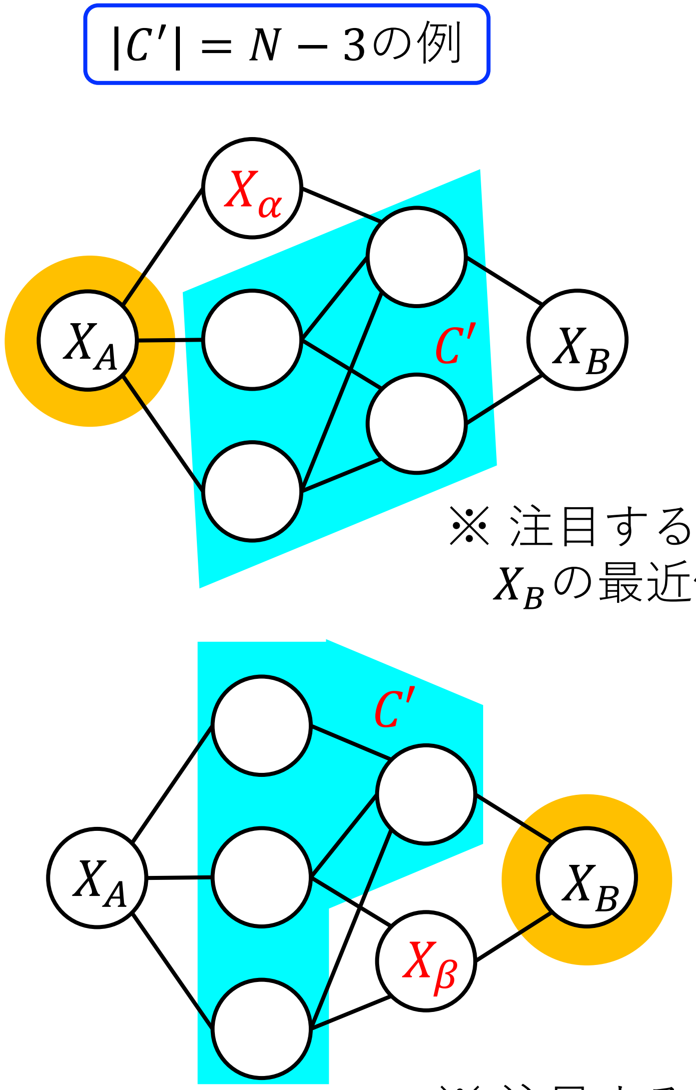
- $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C \Rightarrow \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_{C'} \cup \boldsymbol{X}_\alpha$
- また$X_\alpha \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_{C'} \cup \boldsymbol{X}_A$なので，
  Intersection Propertyより$\boldsymbol{X}_B \perp\!\!\!\perp \boldsymbol{X}_{A'} | \boldsymbol{X}_{C'}$

2通りの可能性が考えられる

(2) $(A \cup B \cup C') \subset V$

- Cの一部を取り除き，AまたはBに加えない





※ 注目するαが X_Bの最近傍でない場合

※ 注目するβが X_Aの最近傍でない場合

(右上の例のみ示す)

- $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_C \Rightarrow \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_{C'} \cup \boldsymbol{X}_\alpha$
- $X_\alpha \perp\!\!\!\perp \boldsymbol{X}_B | \boldsymbol{X}_{C'} \cup \boldsymbol{X}_A$なので，
  Intersection Propertyより$\boldsymbol{X}_B \perp\!\!\!\perp (\boldsymbol{X}_A \cup X_\alpha) | \boldsymbol{X}_{C'}$，
  またDecomposition Propertyより$\boldsymbol{X}_B \perp\!\!\!\perp \boldsymbol{X}_A | \boldsymbol{X}_{C'}$

付録 (6)

$\mathbb{P} \Rightarrow \mathbb{F}$: Hammersley-Cliffordの定理の証明

⏪ 本文に戻る

準備：Möbius inversion lemma

- $a \subseteq V$とする
- Vの可能なsubsetに対して定義される関数Φ, Ψを用意する
- このとき次の二つの表現は等価である

$$\Psi(a) = \sum_{b:b \subseteq a} \Phi(b) \qquad \cdots ☆$$

$$\Phi(a) = \sum_{b:b \subseteq a} (-1)^{|a \setminus b|} \Psi(b) \qquad \cdots ★$$

Möbius inversion lemmaの証明

- ★→☆を示す
  - ★の両辺について和をとる

$$
\begin{aligned}
\sum_{b:b\subseteq a} \Phi(b) &= \sum_{b:b\subseteq a} \sum_{c:c\subseteq b} (-1)^{|b\setminus c|} \Psi(c) \\
&= \sum_{c:c\subseteq a} \Psi(c) \left\{ \sum_{b:c\subseteq b\subseteq a} (-1)^{|b\setminus c|} \right\} \\
&= \sum_{c:c\subseteq a} \Psi(c) \left\{ \sum_{h:h\subseteq a\setminus c} (-1)^{|h|} \right\} \\
&= \Psi(a)
\end{aligned}
$$

※ non-empty setについて,
要素数が偶数の集合と奇数の集合は同数存在する

Möbius inversion lemmaの証明(つづき)

- ☆→ ★を示す
  - ☆の両辺について和をとる

$$\sum_{b:b\subseteq a}(-1)^{|a\backslash b|}\Psi(b)=\sum_{b:b\subseteq a}(-1)^{|a\backslash b|}\sum_{c:c\subseteq b}\Phi(c)$$

$$=\sum_{c:c\subseteq a}\Phi(c)\left\{\sum_{b:c\subseteq b\subseteq a}(-1)^{|a\backslash b|}\right\}$$

$$=\sum_{c:c\subseteq a}\Phi(c)\left\{\sum_{h:h\subseteq a\backslash c}(-1)^{|h|}\right\}$$

$$=\Phi(a)$$

※ non-empty setについて,
要素数が偶数の集合と奇数の集合は同数存在する

$\mathbb{P} \Rightarrow \mathbb{F}$: Hammersley-Cliffordの定理の証明

▶ 証明の概要

- あるノード集合bと$V \setminus b$に対して，$H_b(\boldsymbol{X}_b) = \log p\left(\boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b}\right)$とする．$\boldsymbol{X}^*_{V\setminus b}$は実現値.

(1) H_bに対してMöbius 反転公式を適用

$$\Phi_a(\boldsymbol{X}_a) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|} H_b(\boldsymbol{X}_b)$$

$$H_V(\boldsymbol{X}) \equiv \sum_{a:a\subseteq V} \Phi_a(\boldsymbol{X}_a)$$

→ここから，$p(\boldsymbol{X}) = \prod_{a\subseteq V} \exp(\Phi_a(\boldsymbol{X}_a))$ という形になることがわかる

(2) さらに，ノード集合aがcompleteでない場合$\Phi_a(\boldsymbol{X}_a) = 0$であることを示す

→ここから，$\Phi_a(\boldsymbol{X}_a)$がポテンシャルとみなせることを示す

◆ 証明 ◆

- $\alpha, \beta \in a\ (E_{\alpha\beta} \notin E)$という二つのノードを考える．また$c = a \setminus \{\alpha, \beta\}$とする

$$\Phi_a(\boldsymbol{X}_a) = \sum_{b:b\subseteq a} (-1)^{|a\setminus b|} H_b(\boldsymbol{X}_b)$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} H_b(\boldsymbol{X}_b) + \sum_{b:b\subseteq c} (-1)^{|a\setminus b\cup\{\alpha,\beta\}|} H_b(\boldsymbol{X}_b)$$

$$+ \sum_{b:b\subseteq c} (-1)^{|a\setminus b\cup\{\alpha\}|} H_b(\boldsymbol{X}_b) + \sum_{b:b\subseteq c} (-1)^{|a\setminus b\cup\{\beta\}|} H_b(\boldsymbol{X}_b)$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} \left\{H_b(\boldsymbol{X}_b) + (-1)^{-2} H_{b\cup\{\alpha,\beta\}}\left(\boldsymbol{X}_{b\cup\{\alpha,\beta\}}\right) + (-1)^{-1} H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha}) + (-1)^{-1} H_{b\cup\beta}\left(\boldsymbol{X}_{b\cup\beta}\right)\right\}$$

$$= \sum_{b:b\subseteq c} (-1)^{|a\setminus b|} \left\{H_b(\boldsymbol{X}_b) + H_{b\cup\{\alpha,\beta\}}\left(\boldsymbol{X}_{b\cup\{\alpha,\beta\}}\right) - H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha}) - H_{b\cup\beta}\left(\boldsymbol{X}_{b\cup\beta}\right)\right\}$$

- $H_b(\boldsymbol{X}_b)$, $H_{b\cup\{\alpha,\beta\}}(\boldsymbol{X}_{b\cup\{\alpha,\beta\}})$, $H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha})$, $H_{b\cup\beta}(\boldsymbol{X}_{b\cup\beta})$についてそれぞれ考える.

✦ $$H_{b\cup\{\alpha,\beta\}}(\boldsymbol{X}_{b\cup\{\alpha,\beta\}}) = \log p\left(X_\alpha, X_\beta, \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$
$$= \log p\left(X_\alpha, X_\beta | \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right) + \log p\left(\boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$
$$= \log p\left(X_\alpha | \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right) + \log p\left(X_\beta | \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right) + \log p\left(\boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$

… Pairwise Markov Propertyを用いた.

✦ $$H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha}) = \log p\left(X_\alpha, \boldsymbol{X}_b, X^*_\beta, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$
$$= \log p\left(X_\alpha | \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right) + \log p\left(X^*_\beta, \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$

… Pairwise Markov Propertyを用いた.

よって

$$H_{b\cup\{\alpha,\beta\}}(\boldsymbol{X}_{b\cup\{\alpha,\beta\}}) - H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha}) = \log p\left(X_\beta, \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right) - \log p\left(X^*_\beta, \boldsymbol{X}_b, \boldsymbol{X}^*_{V\setminus b\cup\{\alpha,\beta\}}\right)$$

✦ $H_{b\cup\beta}(\boldsymbol{X}_{b\cup\beta}) = \log p\left(X_\beta, \boldsymbol{X}_b, X_\alpha^*, \boldsymbol{X}_{V\backslash b\cup\{\alpha,\beta\}}^*\right) = \log p\left(X_\beta, \boldsymbol{X}_b, \boldsymbol{X}_{V\backslash b\cup\{\alpha,\beta\}}^*\right)$

✦ $H_b(\boldsymbol{X}_b) = \log p\left(\boldsymbol{X}_b, \boldsymbol{X}_{V\backslash b}^*\right) = \log p\left(\boldsymbol{X}_b, X_\alpha^*, X_\beta^*, \boldsymbol{X}_{V\backslash b\cup\{\alpha,\beta\}}^*\right) = \log p\left(\boldsymbol{X}_b, X_\beta^*, \boldsymbol{X}_{V\backslash b\cup\{\alpha,\beta\}}^*\right)$

よって，$H_b(\boldsymbol{X}_b) + H_{b\cup\{\alpha,\beta\}}\left(\boldsymbol{X}_{b\cup\{\alpha,\beta\}}\right) - H_{b\cup\alpha}(\boldsymbol{X}_{b\cup\alpha}) - H_{b\cup\beta}\left(\boldsymbol{X}_{b\cup\beta}\right) = 0$なので,

ノード集合がdisjointなノードを含むとき常に$\Phi_a(\boldsymbol{X}_a) = 0$である.

このことから,

$$H_V(\boldsymbol{X}) = \sum_{a:a\subseteq V} \Phi_a(\boldsymbol{X}_a) = \sum_{a:a\text{ is complete}} \Phi_a(\boldsymbol{X}_a)$$

となり，よって,

$$p(\boldsymbol{X}) = \prod_{a:a\text{ is complete}} \exp\left(\Phi_a(\boldsymbol{X}_a)\right)$$

… これはFactorized formである.

付録 (7)

MoralなDAG $\mathcal{D}$の完全マップが$\mathcal{M}[\mathcal{D}]$であることの証明

- $I(X,Y|Z) \in \mathcal{I}(\mathcal{D}), I(X,Y|Z) \notin \mathcal{I}(\mathcal{M}[\mathcal{D}])$という独立性の存在を仮定する.
  - ✦ $\mathcal{M}[\mathcal{D}]$においてZを経由しない$X-Y$経路がある.
  - ✦ 同じ経路は$\mathcal{D}$にも存在しなければならない ($\mathcal{D} \to \mathcal{M}[\mathcal{D}]$により追加される経路はないため).
- 仮定から，$X-Y$経路は$\mathcal{D}$においてZによりブロックされていなければならない
  - $X-Y$経路上にhead-to-head nodeがある必要がある。
- しかし，$\mathcal{D}$では親ノードが全結合なので，head-to-head nodeの
  親ノードは繋がっているため,
  head-to-head nodeでブロックされる経路とは別の経路が生じてしまう
- よって，$\mathcal{D}$の親ノードが繋がるとき$I(X,Y|Z) \in \mathcal{I}(\mathcal{D}), I(X,Y|Z) \notin \mathcal{I}(\mathcal{M}[\mathcal{D}])$は否定され,
  以下のいずれかが成立する.
  - ✦ $I(X,Y|Z) \in \mathcal{I}(\mathcal{D}), I(X,Y|Z) \in \mathcal{I}(\mathcal{M}[\mathcal{D}])$
  - ✦ $I(X,Y|Z) \notin \mathcal{I}(\mathcal{D}), I(X,Y|Z) \notin \mathcal{I}(\mathcal{M}[\mathcal{D}])$
  - ✦ $I(X,Y|Z) \notin \mathcal{I}(\mathcal{D}), I(X,Y|Z) \in \mathcal{I}(\mathcal{M}[\mathcal{D}])$
- これは$\mathcal{I}(\mathcal{D}) \subseteq \mathcal{I}(\mathcal{M}[\mathcal{D}])$を意味する.

付録 (8)

$\mathbb{DG} \Leftrightarrow \mathbb{DG}'$の証明

⏪ 本文に戻る

【DAG $\mathcal{D}$においてAーB間にCによりblockされていない経路がある

$\Rightarrow \mathcal{M}\left[\mathcal{D}_{An^*(A \cup B \cup C)}\right]$において$A$ー$B$間に$C$を経由しない経路がある】

- $A - B$経路上の全てのノードが$An^*(A \cup B \cup C)$にある場合を考えれば良い

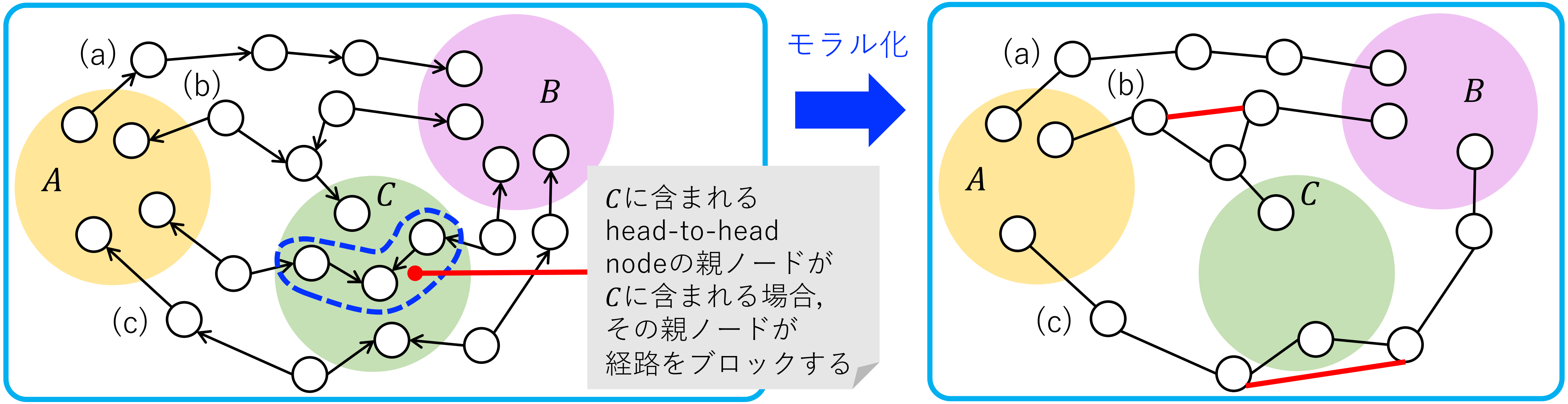
DAG $\mathcal{D}$においてAーB間がblockされていないということは，以下いずれかの経路がある．

(a) Cを通らない$A - B$間経路にhead-to-head nodeがない

(b) Cを通らない$A - B$間経路にhead-to-head nodeがあり，その子孫がCに含まれる

(c) $A - C - B$間で，Cに含まれるhead-to-head nodeがある．
ただしそのhead-to-headの親ノードのうち少なくとも2つはCに含まれない．

いずれの経路も，モラル化によりCを経由しない経路を生じさせる．

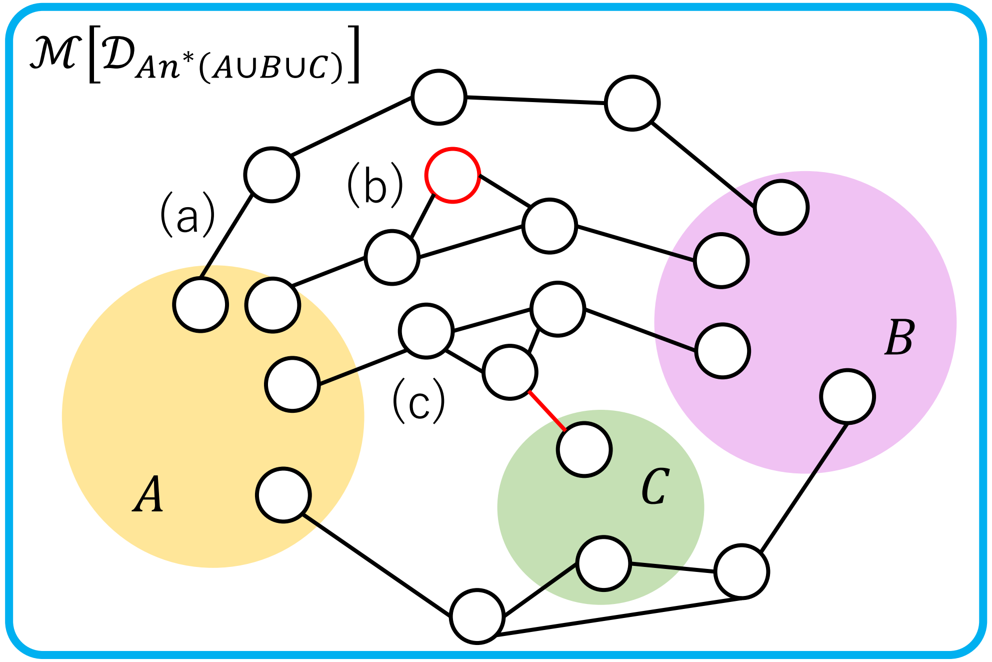


【 $\mathcal{M}\left[\mathcal{D}_{An^*(A\cup B\cup C)}\right]$においてA－B間にCを経由しない経路がある
⇒ DAG $\mathcal{D}$においてA－B間にCによりblockされていない経路がある】

- $\mathcal{M}\left[\mathcal{D}_{An^*(A\cup B\cup C)}\right]$で，$C$を経由しない$A-B$経路のパターンを全て考える(下図参照).

(a) $N=3$のループがない
- $\mathcal{D}$においてこの経路にはhead-to-head node がなく，ブロックされない

(b) Cのノードを含まない$N=3$のループがある
- $An^*(A\cup B\cup C)$を考えているので，$\mathcal{D}$において頂点(赤丸)はhead-to-headではない
  ※ $\mathcal{D}$においてhead-to-head nodeであれば$An^*(A\cup B\cup C)$に含まれない)
- よってこの経路はブロックされない



(c) Cに含まれるノードにつながる$N=3$のループがある
- $\mathcal{D}$において，ループとつながるedge(赤)はCに入る向きになる.
- このとき，ループの矢印の向きがCに分離されない

(d) Cを含む$N=3$のループがある
- Cに含まれるノードはhead-to-head nodeであり，Cにより分離されない

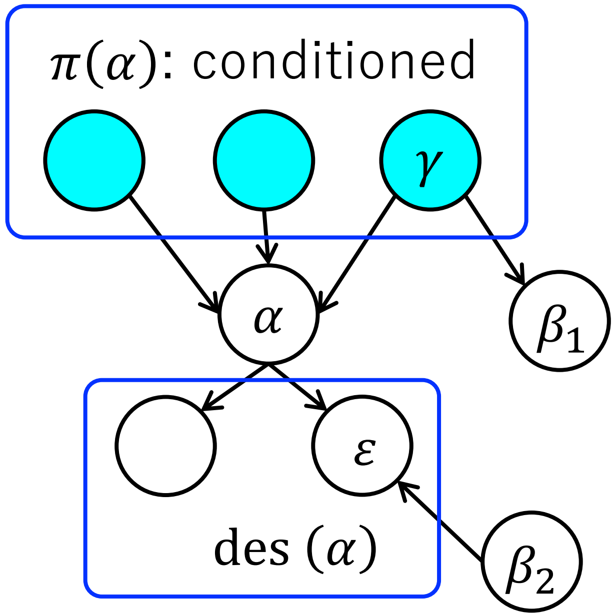
付録 (9)

$\mathbb{DG} \Rightarrow \mathbb{DL}$の証明

⏪ 本文に戻る

- ノードαの親ノード集合を$\pi(\alpha)$，子孫ノード集合を$\mathbf{des}(\alpha)$とする．

- DAG $\mathcal{D}$のSkelton $S[\mathcal{D}]$において，
  αからnon-descendants $\mathbf{nd}(\alpha) = V \setminus \left(\alpha \cup \mathbf{des}(\alpha)\right)$への
  可能な経路を考える（$\pi(\alpha) \subset \mathbf{nd}(\alpha)$である）



- 3通りの経路が考えられる

(i) αから$\pi(\alpha)$への経路：$\pi(\alpha)$によりブロックされる

(ii) $\gamma \in \pi(\alpha)$を通り$\beta_1 \in \mathbf{nd}(\alpha) \setminus \pi(\alpha)$に至る経路

- γはhead-to-head nodeではない
- $\pi(\alpha)$は条件づけられているので，$\alpha - \beta_1$の経路はblockされている

(iii) $\varepsilon \in \mathbf{des}(\alpha)$を通り$\beta_2 \in \mathbf{nd}(\alpha) \setminus \pi(\alpha)$に至る経路

- 経路上にhead-to-head nodeが存在する必要があり，これにより経路はblockされる

(i),(ii),(iii)を合わせると，d分離性より$\alpha \perp\!\!\!\perp \mathbf{nd}(\alpha) | \pi(\alpha)$