Structure Learning of Bayesian Network

Kyoto Univerity
Ryota Miyagi

参考文献

- 特に書かない場合は以下の参考文献から

  - 植野真臣．**ベイジアンネットワーク**．コロナ社(2013)
  - 鈴木 譲，植野 真臣，黒木 学，清水 昌平，湊 真一，石畠 正和，樺島 祥介，田中 和之，本村 陽一，玉田 嘉紀．**確率的グラフィカルモデル**．共立出版(2016)
  - 黒木 学．**統計的因果モデルの基礎**．共立出版(2017)
  - Mykel J. Kochenderfer, Tim A. Wheeler and Kyle H. Wray**. Algorithms for Decision Making**. MIT Press(2022)
  - Marco Scutari, Jean-Baptiste Denis. **Bayesian Networks with Examples in R Second Edition**. CRC Press(2022)

記法について

- 確率変数は大文字（e.g. X）
- 確率変数の集合は大文字太字（e.g. $\boldsymbol{V}$）
- 確率変数のとる値は小文字（e.g. x, $\mathrm{P}(X = x)$）

ここからの流れとモチベーション

- 対象の互いに素な部分集合 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ に対して，“knowing $\boldsymbol{Z}$ renders $\boldsymbol{X}$ and $\boldsymbol{Y}$ independent”という関係を考える
  - e.g. 条件付き独立性 $P(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) = P(\boldsymbol{X} \mid \boldsymbol{Z})\, P(\boldsymbol{Y} \mid \boldsymbol{Z})$
  - e.g. ある知識体系$\boldsymbol{Z}$が与えられ，命題集合$\boldsymbol{X}$の真偽を確かめたい場合， $\boldsymbol{Z}$に含まれない別の命題集合$\boldsymbol{Y}$を参照する価値があるかどうか
- この関係$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$をどのように表現するか？
  - 関係$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$を列挙する
    - △互いに素な部分集合$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$の組み合わせは指数的
  - 関係$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y})$をグラフで表現する
    - ○時間的にも空間的にも効率的
    - △互いに素な部分集合$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$のすべての組み合わせを表現することはできない
- 一つの解決策がベイジアンネットワーク

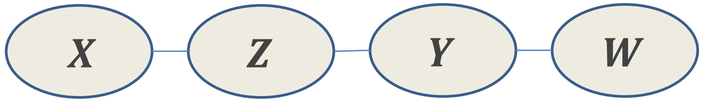
条件付き独立性とその無向グラフ表現

- 同時確率分布Pで特徴づけられる離散確率変数集合$\boldsymbol{V}$があるとする
- $\boldsymbol{Z}$を所与として$\boldsymbol{X}, \boldsymbol{Y}$が条件付き独立を以下で定義し，$I(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{Y})$と書く
  - $P(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) = P(\boldsymbol{X} \mid \boldsymbol{Z})P(\boldsymbol{Y} \mid \boldsymbol{Z}),\ P(\boldsymbol{Z}) > 0$
- $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})$を「ノード集合$\boldsymbol{Z}$がノード集合$\boldsymbol{X}, \boldsymbol{Y}$を分離する」で符号化すると，以下の条件付き独立性の満たすべき性質は保存される

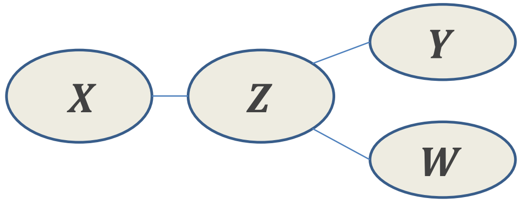
Judea Pearl and Azaria Paz. GRAPHOIDS: A Graph-based Logic for Reasoning about Relevance Relations, In Proceedings, ECAI-86, Brighton, United Kingdom, June 1986

条件付き独立性とその無向グラフ表現

- 対称性(symmetry):
  - $I(\boldsymbol{X},\boldsymbol{Y} \mid \boldsymbol{Z}) \Leftrightarrow I(\boldsymbol{Y},\boldsymbol{X} \mid \boldsymbol{Z})$
- 縮約性(contraction):
  - $I(\boldsymbol{X},\boldsymbol{W} \mid \boldsymbol{Z} \cup \boldsymbol{Y})$ かつ $I(\boldsymbol{X},\boldsymbol{Y} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X},\boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z})$



- 分解性(decomposition):
  - $I(\boldsymbol{X},\boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X},\boldsymbol{Y} \mid \boldsymbol{Z})$ かつ $I(\boldsymbol{X},\boldsymbol{W} \mid \boldsymbol{Z})$
- 弱結合性(week union):
  - $I(\boldsymbol{X},\boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X},\boldsymbol{W} \mid \boldsymbol{Z} \cup \boldsymbol{Y})$ かつ$I(\boldsymbol{X},\boldsymbol{Y} \mid \boldsymbol{Z} \cup \boldsymbol{W})$
- 交差性(interaction):
  - $I(\boldsymbol{X},\boldsymbol{W} \mid \boldsymbol{Z} \cup \boldsymbol{Y})$ かつ $I(\boldsymbol{X},\boldsymbol{Y} \mid \boldsymbol{Z} \cup \boldsymbol{W}) \Rightarrow I(\boldsymbol{X},\boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z})$



Judea Pearl and Azaria Paz. GRAPHOIDS: A Graph-based Logic for Reasoning about Relevance Relations, In Proceedings, ECAI-86, Brighton, United Kingdom, June 1986

無向グラフによる$I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})$の表現

- $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_P$: 同時確率分布$P$における真の条件付き独立性
- $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_G$: 無向グラフ$G$でノード集合$\boldsymbol{Z}$がノード集合$\boldsymbol{X}, \boldsymbol{Y}$を分離
- 定義: Perfect map (*P-map*)
  - $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_G \Leftrightarrow I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_P$のとき，$G$は$P$のP-map
- 定義: Independent map (*I-map*)
  - $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_G \Rightarrow I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_P$のとき，$G$は$P$のI-map
  - 完全グラフは$I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_G$なる$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$は存在しないので常にI-map
- 定義: 最小 I-map
  - I-map G が辺を取り除くとI-map でなくなるとき， GはPの最小 I-map
- i.e., 限界はあるが，無向グラフで条件付き独立性が符号化できる

Judea Pearl and Azaria Paz. GRAPHOIDS: A Graph-based Logic for Reasoning about Relevance Relations, In Proceedings, ECAI-86, Brighton, United Kingdom, June 1986

無向グラフによる$I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})$の表現

- 定理: すべてのPは一意の最小 I-map G_0を持ち，以下で構成できる
  - 完全グラフから$I(X, Y \mid \boldsymbol{V} - X - Y)_P$なる$(X, Y)$をすべて取り除く
  - 後述する制約ベース構造学習につながる
- 定義: マルコフ境界(Markov boundary)
  - 変数αのマルコフ境界B_α^Pはαを他のすべての変数と独立とする極小変数集合
- 定理: 各変数αのマルコフ境界B_α^PはG_0におけるαとの隣接頂点集合$B_\alpha^{G_0}$と一致する.
- i.e., 最小 I-map G_0における変数αの隣接頂点集合$B_\alpha^{G_0}$を知ることで，変数αは他のすべての変数と独立となる

Judea Pearl and Azaria Paz. GRAPHOIDS: A Graph-based Logic for Reasoning about Relevance Relations, In Proceedings, ECAI-86, Brighton, United Kingdom, June 1986

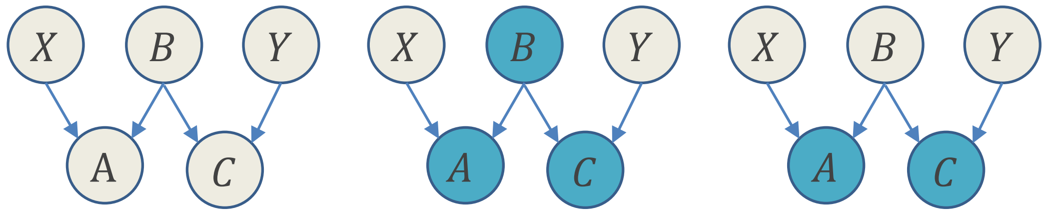
寄り道：グラフォイド (graphoid)

- 条件付き独立性を無向グラフで表現することを考えてきた
- 他の「$\boldsymbol{Z}$を知ると$\boldsymbol{X}$と$\boldsymbol{Y}$は独立」という関係を表現するための拡張
- 互いに素な対象の部分集合 $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ に対し以下の公理を満たす3項関係の集合を *graphoid*と呼ぶ
  - 対称性(symmetry): $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) \Leftrightarrow I(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{Z})$
  - 分解性(decomposition): $I(\boldsymbol{X}, \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})$ かつ $I(\boldsymbol{X}, \boldsymbol{W} \mid \boldsymbol{Z})$
  - 弱結合性(week union): $I(\boldsymbol{X}, \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X}, \boldsymbol{W} \mid \boldsymbol{Z} \cup \boldsymbol{Y})$ かつ$I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z} \cup \boldsymbol{W})$
  - 縮約性(contraction): $I(\boldsymbol{X}, \boldsymbol{W} \mid \boldsymbol{Z} \cup \boldsymbol{Y})$ かつ $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) \Rightarrow I(\boldsymbol{X}, \boldsymbol{Y} \cup \boldsymbol{W} \mid \boldsymbol{Z})$
- DBの独立関係(Embedded Multi-valued Dependence, EMVD)は graphoid らしい

Judea Pearl and Azaria Paz. GRAPHOIDS: A Graph-based Logic for Reasoning about Relevance Relations, In Proceedings, ECAI-86, Brighton, United Kingdom, June 1986

有向グラフによる$I(X, Z, Y)$の表現

- 無向グラフでの分離に対応する有向グラフ版の分離が必要
- 定義: 有向分離(d-separation)
  - 有向グラフGにおいてノード集合$\boldsymbol{X}$の任意の要素Xと$\boldsymbol{Y}$の任意の要素Yを結ぶすべての道のそれぞれについて，$\boldsymbol{X} \cup \boldsymbol{Y}$ と背反なノード集合$\boldsymbol{Z}$が次の条件のいづれかを満たすとき，$\boldsymbol{Z}$は$\boldsymbol{X}$と$\boldsymbol{Y}$を有向分離すると定義し，$\mathrm{I}(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_d$と書く
    - (1) XとYを結ぶ道上の合流点で，その合流点とその子孫が$\boldsymbol{Z}$に含まれないものがある
    - (2) XとYを結ぶ道上の非合流点で，$\boldsymbol{Z}$に含まれるものがある
  - 具体例
    - $\emptyset$は$\{X\}, \{Y\}$を有向分離する
    - $\{\mathrm{A}, B, C\}$は$\{X\}, \{Y\}$を有向分離する
    - $\{\mathrm{A}, C\}$は$\{X\}, \{Y\}$を有向分離しない（！）



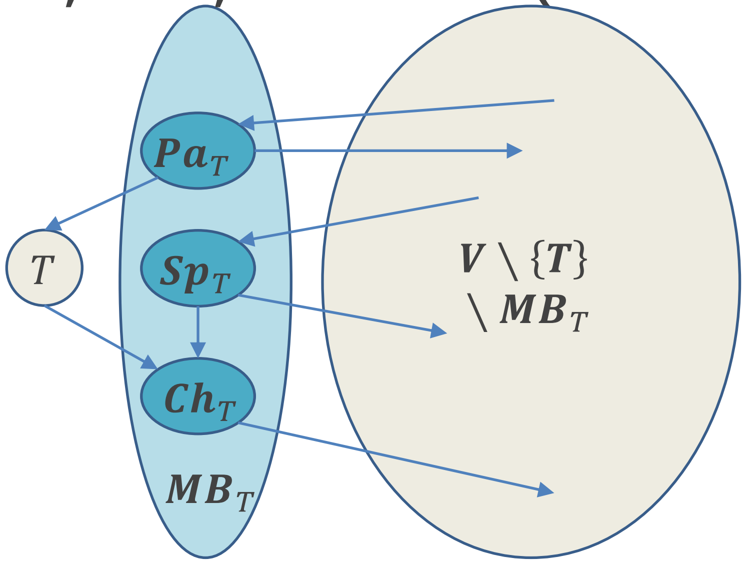
- ちょっと複雑なので，視覚的な理解を次に示す

有向分離の視覚的理解

- 有向グラフ上の3ノードA, B, Cの関係は以下のどれか
  - Note: AとCの間に辺がある場合，そもそも有向分離されないので
- 逐次結合(serial, chain) $A \to B \to C$
  - Bは非合流点なので$B \in \boldsymbol{Z}$のとき，有向分離する
- 分岐結合(diverging, fork) $A \leftarrow B \to C$
  - Bは非合流点なので$B \in \boldsymbol{Z}$のとき，有向分離する
- 合流結合(coverging, inverted fork) $A \to B \leftarrow C$
  - Bは合流点なので$B \notin \boldsymbol{Z}$のとき，有向分離する
  - 特別なので，V構造(v-structure)と名前がついている

マルコフブランケット(Markov Blanket, MB)

- T の親集合，子集合，配偶者集合をそれぞれ $\boldsymbol{Pa}_T, \boldsymbol{Ch}_T, \boldsymbol{Sp}_T$ と書く
- 定義: マルコフブランケット(Markov Blanket, MB)
  - 変数Tのマルコフブランケット$\boldsymbol{MB}_T$はTを他のすべての変数から条件付き独立とする極小変数集合
  - i.e., $\forall X \in \boldsymbol{V} \setminus \{T\} \setminus \boldsymbol{MB}_T,\ \boldsymbol{I}(T, X \mid \boldsymbol{MB}_T)_{\boldsymbol{P}}$
- 定理: ノード T のマルコフブランケットは T の親，子， 配偶者(T と子を共有するノードの和集合
  - i.e., $\boldsymbol{MB}_T = \boldsymbol{Pa}_T \sqcup \boldsymbol{Ch}_T \sqcup \boldsymbol{Sp}_T$
- i.e., 変数Tの$\boldsymbol{MB}_T$を知ることで
- 変数Tは他のすべての変数と独立となる

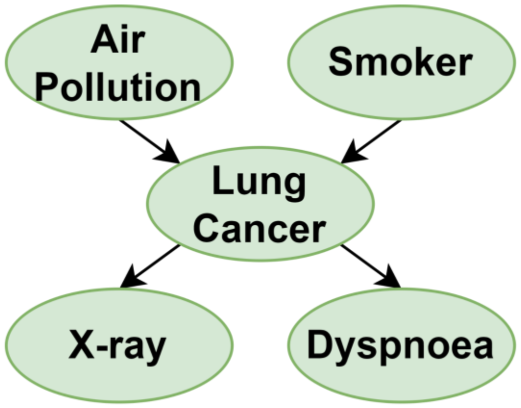


ベイジアンネットワーク(Bayesian Network, BN)

- 離散確率変数集合 $\boldsymbol{V} = \{\mathrm{X}_1, X_2, \dots, X_n\}$のベイジアンネットワークは以下の組で定義される$(G, \boldsymbol{\Theta})$
  - Gは$\boldsymbol{V}$の各変数に対応するノード集合によって構成されるDAG
  - $\boldsymbol{\Theta} = \{\, \mathrm{p}(\mathrm{X} \mid \boldsymbol{Pa}_{\mathrm{X}}) \,\}$は条件付き確率パラメータ集合
- 有向分離と非巡回性の仮定のみから導かれる最も自然なモデル

Air Pollution	
H	L
0.9	0.1

Smoker	
T	F
0.3	0.7



Lung Cancer			
Pollut	Smk	T	F
H	T	0.05	0.95
H	F	0.02	0.98
L	T	0.03	0.97
L	F	0.01	0.99

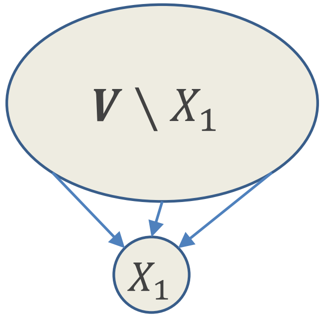
X-ray		
Cnc	T	F
T	0.9	0.1
F	0.20	0.80

Dyspnoea		
Cnc	T	F
T	0.65	0.35
F	0.30	0.70

Pearl, Judea. "Bayesian networks: A model of self-activated memory for evidential reasoning." *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*. 1985

ベイジアンネットワークのうれしさ

- 同時確率分布をコンパクトな条件付き確率の積に因数分解する
  - $X_1, X_2, \ldots, X_n$はトポロジカル降順になっているとすると
    - $\mathrm{p}(\boldsymbol{V})$
    - $= \mathrm{p}(\mathrm{X}_1 \mid \boldsymbol{V} \setminus \{X_1\})\, p(\boldsymbol{V} \setminus \{X_1\}) \quad \because Bayes'\ theorem$
    - $= \mathrm{p}\left(\mathrm{X}_1 \mid \boldsymbol{Pa}_{\mathrm{X}_1}\right) \quad p(\boldsymbol{V} \setminus \{X_1\}) \quad \because \forall X \in \boldsymbol{V} \setminus \{X_1\} \setminus \boldsymbol{Pa}_{\mathrm{X}_1},\ I(X, X_1 \mid \boldsymbol{Pa}_{\mathrm{X}_1})$
    - $= \cdots$
    - $= \Pi_i^n \mathrm{p}\left(\mathrm{X}_i \mid \boldsymbol{Pa}_{\mathrm{X}_1}\right)$
  - 巡回がある場合，このような分解はできないので，非巡回性の仮定が必要

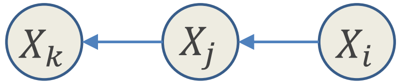


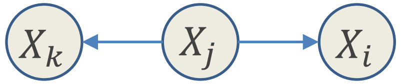
- パラメータ数の組み合わせ爆発を抑える
  - もし$X_1, X_2, \ldots, X_n$がすべて2値離散変数の場合，$\mathrm{p}(\boldsymbol{V})$を直接推定するのに必要なパラメータの数は $2^n \Rightarrow \sum_i^n 2^{\left|\boldsymbol{Pa}_{X_i}\right|+1}$ ，妥当なパラメータ推定に必要なデータ数も減る
- i.e., BNは条件付き独立性を利用した離散変数の同時確率分布の圧縮表現

Pearl, Judea. "Bayesian networks: A model of self-activated memory for evidential reasoning." *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*. 1985

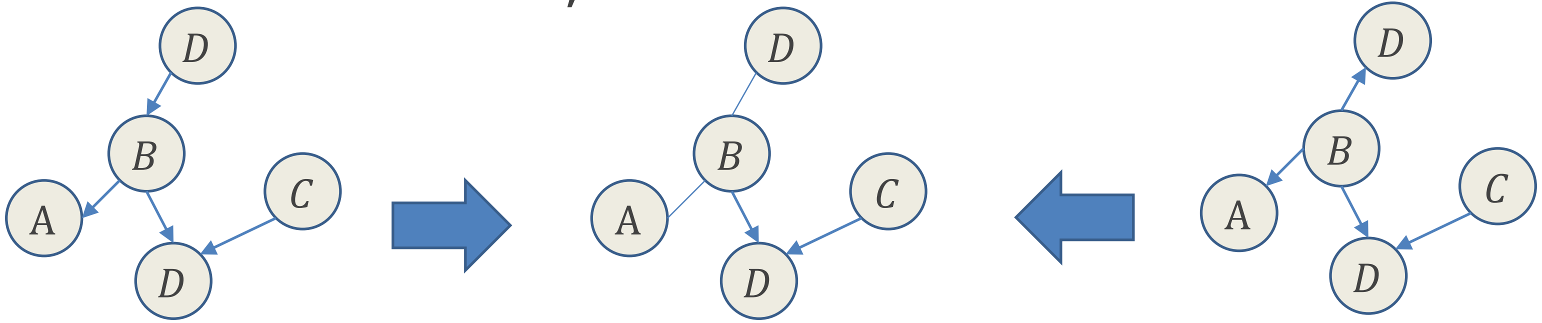
ベイジアンネットワークの同値類(Equivalence classes)

- $p(X_k \mid X_j)p(X_j \mid X_i)p(X_i)$ ←逐次結合
- $= p(X_k \mid X_j)p(X_i, X_j)$
- $= p(X_k \mid X_j)p(X_j)p(X_i \mid X_j)$ ←分岐結合



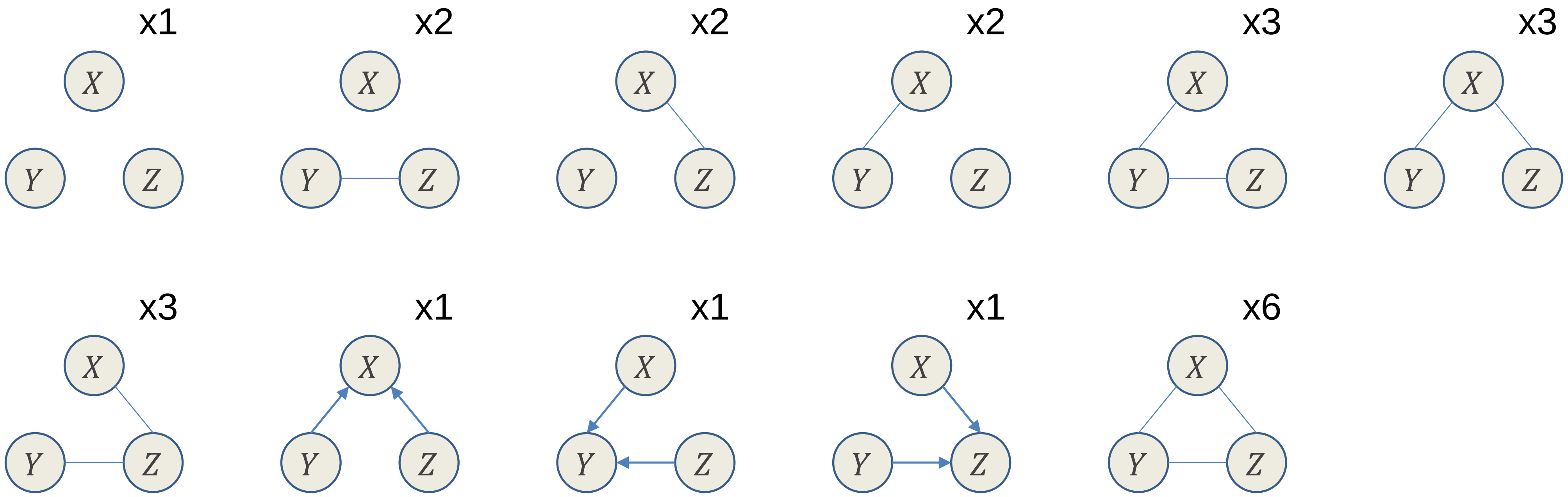


- 同値類(equivalence classes)
  - 二つのDAGが無向グラフにしたときに同一でかつ同一のV構造を持つとき，両者は同値（符合化する条件付き独立性は同一）
- completed partially directed acyclic graph, CPDAG
  - V構造以外を無向化したもの，つまり同値類の代表表現

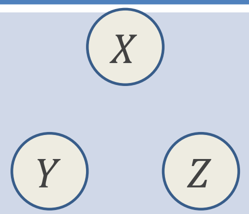
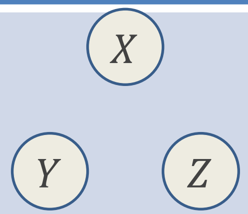
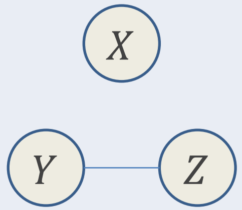
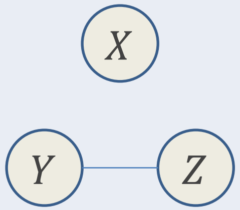
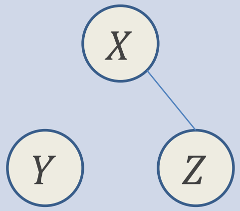
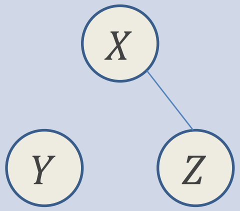
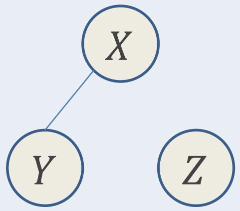
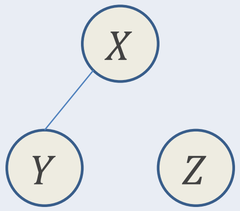


ベイジアンネットワークの同値類(Equivalence classes)

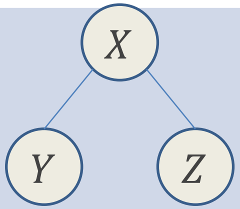
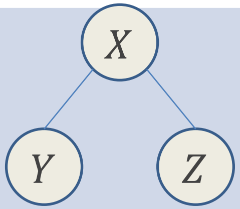
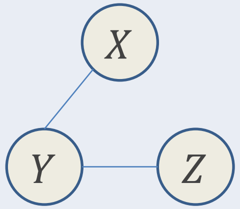
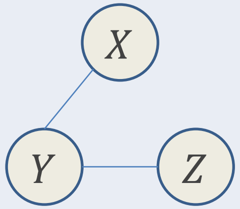
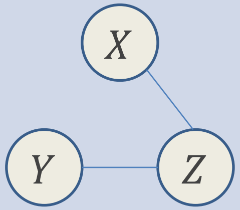
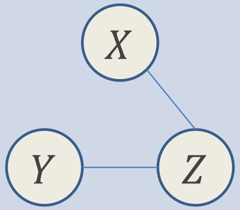
- 3変数X, Y, ZのDAGとしては3^3-2 = 25個だがCPDAGは11個）



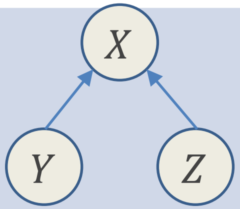
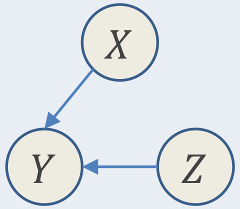
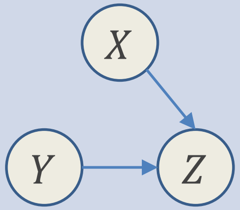
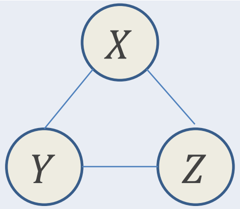
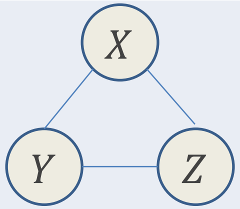
有向グラフ表現と無向グラフ表現の限界

符号化する条件付き独立性	有向グラフ表現(CPDAG)	無向グラフ表現
$I(X,Y \mid \emptyset), I(Y,Z \mid \emptyset), I(Z,X \mid \emptyset),$ $I(\{XY\},Z \mid \emptyset), I(\{YZ\},X \mid \emptyset), I(\{ZX\},Y \mid \emptyset)$		
$I(X,Y \mid \emptyset)$, $I(X,Z \mid \emptyset)$, $I(X,YZ \mid \emptyset)$		
$I(X,Y \mid \emptyset)$, $I(Y,Z \mid \emptyset)$, $I(Y,XZ \mid \emptyset)$		
$I(X,Z \mid \emptyset)$, $I(Y,Z \mid \emptyset)$, $I(Z,XY \mid \emptyset)$		

有向グラフ表現と無向グラフ表現の限界

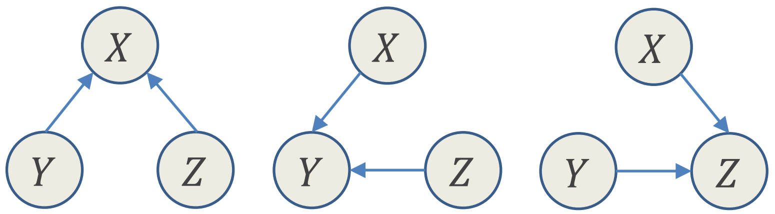
符号化する条件付き独立性	有向グラフ表現(CPDAG)	無向グラフ表現
$I(Y,Z \mid X)$		
$I(X,Z \mid Y)$		
$I(X,Y \mid Z)$		

有向グラフ表現と無向グラフ表現の限界

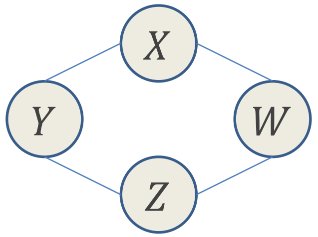
符号化する条件付き独立性	有向グラフ表現(CPDAG)	無向グラフ表現
$I(Y, Z \mid \emptyset)$		ナシ
$I(X, Z \mid \emptyset)$		ナシ
$I(X, Y \mid \emptyset)$		ナシ
ナシ		

有向グラフ表現と無向グラフ表現の限界

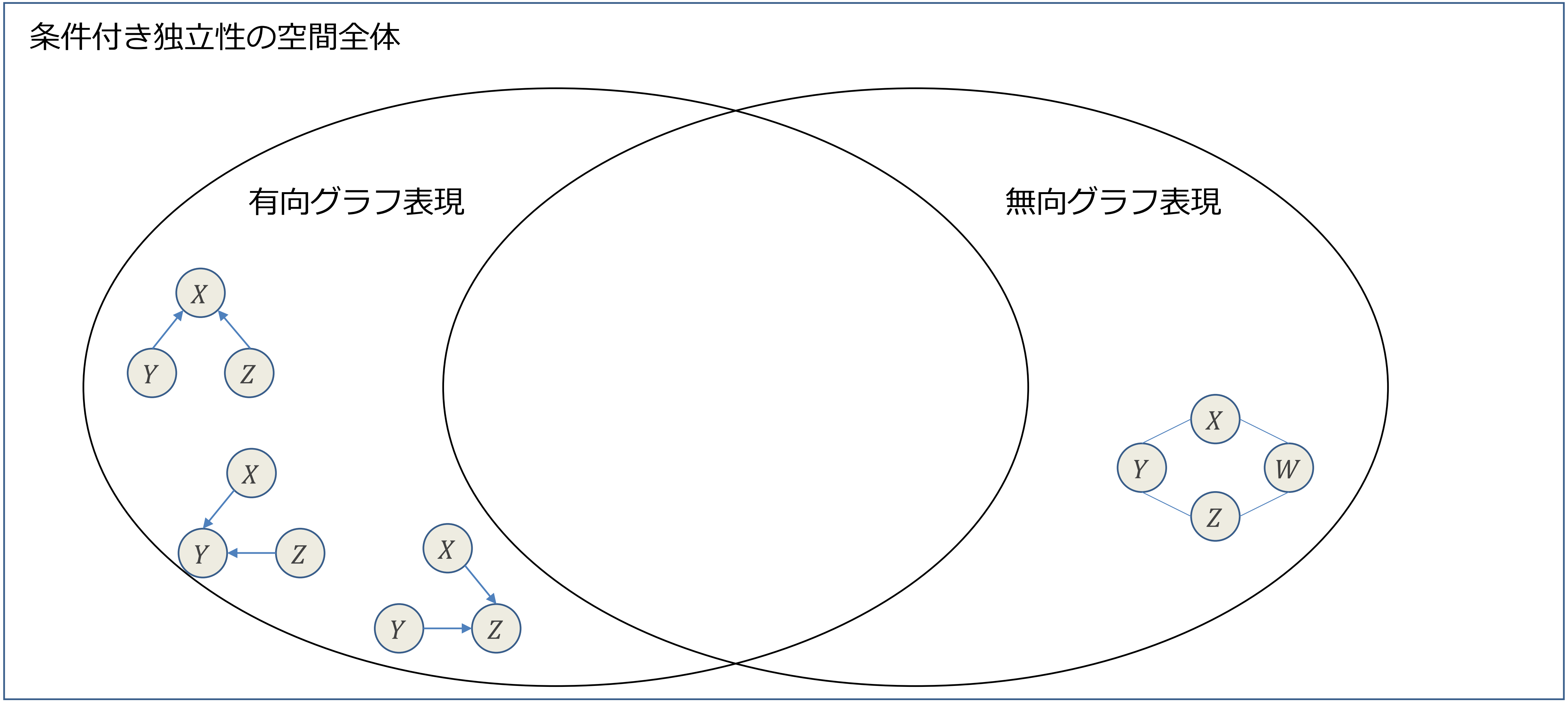
- 3変数では無向グラフで表現できない条件付き独立性が存在した.
  - $I(Y,Z \mid \emptyset), I(X,Z \mid \emptyset), I(X,Y \mid \emptyset)$
  - 無向グラフ表現はV-Structureを表現できない



- 実は4変数では有向グラフで表現できない条件付き独立性が存在する.
  - $I(X,Z \mid Y,W)$かつ$I(Y,W \mid X,Z)$



有向グラフ表現と無向グラフ表現の限界



統計的因果推論(causal inference in statistics)

- 統計的因果推論とは？
  - データから事象の因果効果を統計的に推定する
  - 大きく2つの流派
    - Rubin派: 潜在反応モデル (Rubin causal model, RCM)
    - Pearl派: 構造的因果モデル (Structural Causal Models, SCM)
      - DAGによって因果関係を表現する
- ベイジアンネットワーク自体は因果関係は表現していない

因果ダイアグラム(causal diagram)

- 定義: 因果ダイアグラム
  - 確率変数集合 $\mathbf{V} = \{X_1, X_2, \dots, X_n\}$に対し，DAG Gが確率変数の関係を
$$X_i = g_i(\boldsymbol{Pa}_{X_i}, \epsilon_i)$$
  - なる形に規定し，確率変数がこの関数関係に従って生成されるとき， Gを因果ダイアグラムという．ここで誤差変数ϵ_iは互いに独立である．
- 物理的なデータの生成過程を表している
- このとき以下が成り立つ
$$\mathrm{p}(\boldsymbol{V}) = \Pi_i^n \mathrm{p}\left(\mathrm{X}_i \mid \boldsymbol{Pa}_{\boldsymbol{X}_i}\right)$$

Pearl, Judea. "Causal diagrams for empirical research." *Biometrika* 82.4 (1995): 669-688.

因果ダイアグラムとの関係

> ベイジアンネットワーク

- 条件付き独立性の表現
- 非巡回性の仮定からDAGになる
- 同じ条件付き独立性の符号化でも，矢印が逆になる場合もある

> 因果ダイアグラム

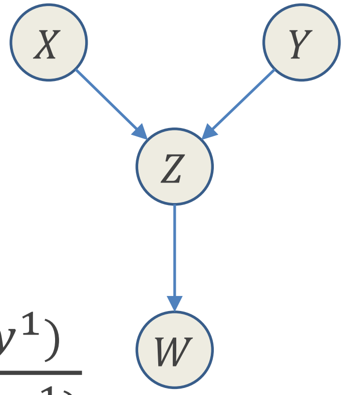
- データの因果関係を表現
- 必ずDAGになる
- 矢印が逆にはならない

いったんまとめ

- 条件付独立性は無向/有向グラフにおける分離/有向分離で表現できる
  - ただし，すべての条件付き独立性を表現できるわけではない
- さらに非巡回性を仮定してベイジアンネットワークが定義される．
  - 同時確率分布をコンパクトな条件付き確率の積に因数分解する
  - パラメータ数の組み合わせ爆発を抑える
  - i.e., BNは条件付き独立性を利用した同時確率分布の圧縮表現
- ベイジアンネットワーク自体は因果関係は表現していない
  - ただし，因果関係の副残物として条件付き独立性が成立してる…？

ベイジアンネットワークによる推論

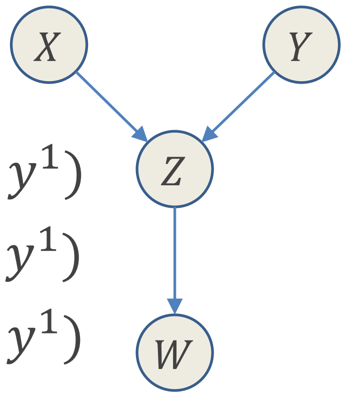
- BN$(G, \boldsymbol{\Theta})$と変数の具体的な値(証拠)が与えられて事後分布を推定
  - 例えば， 4変数のBNで$X = x^1$と$Y = y^1$からWがw^1をとる確率を求めるとか
  - $p(W = w^1 \mid X = x^1, Y = y^1 \})$
  - $= \dfrac{p(X=x^1, Y=y^1, W=w^1)}{p(X=x^1, Y=y^1)}$
  - $= \dfrac{\sum_z p(X=x^1, Y=y^1, Z=z, W=w^1)}{\sum_z \sum_w p(X=x^1, Y=y^1, Z=z, W=w)}$
  - $= \dfrac{\sum_z p(W=w^1 | Z=z)\, p(Z=z | X=x^1, Y=y^1)\, p(X=x^1)\, p(Y=y^1)}{\sum_z \sum_w p(W=w | Z=z)\, p(Z=z | X=x^1, Y=y^1)\, p(X=x^1)\, p(Y=y^1)}$



- 厳密推論(exact inference)
  - NP困難で数百変数が限度
- 近似推論(approximate inference)
  - BN$(G, \boldsymbol{\Theta})$からサンプリング

Compiling Bayesian Networks

- BN($G, \mathbf{\Theta}$)をある種の離散構造にコンパイルして推論を高速化する
- 確率分布を展開してMulti-Linear Function (MLF)を作る
  - 変数がとる値を表現する変数を導入する
    - $= \lambda_{x^1}\lambda_{y^1}\lambda_{z^1}\lambda_{w^1}\, p\big(W = w^1 \mid Z = z^1\big)p\big(Z = z^1 \mid X = x^1, Y = y^1\big)p(X = x^1)p(Y = y^1)$
    - $+ \lambda_{x^1}\lambda_{y^1}\lambda_{z^1}\lambda_{w^2}\, p\big(W = w^2 \mid Z = z^1\big)p\big(Z = z^1 \mid X = x^1, Y = y^1\big)p(X = x^1)p(Y = y^1)$
    - $+ \lambda_{x^1}\lambda_{y^1}\lambda_{z^2}\lambda_{w^1}\, p\big(W = w^1 \mid Z = z^2\big)p\big(Z = z^2 \mid X = x^1, Y = y^1\big)p(X = x^1)p(Y = y^1)$
    - …
    - $+ \lambda_{x^2}\lambda_{y^2}\lambda_{z^2}\lambda_{w^2}\, p\big(W = w^2 \mid Z = z^2\big)p\big(Z = z^2 \mid X = x^2, Y = y^2\big)p\big(X = x^2\big)p\big(Y = y^2\big)$



- このMLFをうまく因数分解してコンパクトに表現したい
    - CNFとして表現しd-DNNFというグラフ構造に変換する
    - BDD/ZDDに変換する

M. Chavira and A. Darwiche, "Compiling Bayesian networks with local structure," IJCAI(2005)

Shin-ichi Minato, Ken Satoh, Taisuke Sato, "Compiling Bayesian Networks by Symbolic Probability Calculation Based on Zero-suppressed BDDs". IJCAI(2007)

ベイジアンネットワークの学習

- データ$\boldsymbol{D}$から$(G, \boldsymbol{\Theta})$を推論（学習）する
- $G, \boldsymbol{\Theta}$を同時に最適化するのは難しいので，一般には 2stepに分ける
    - $p(G, \boldsymbol{\Theta} \mid \boldsymbol{D})$
    - $= p(\boldsymbol{\Theta} \mid G, \boldsymbol{D}) \quad p(G \mid \boldsymbol{D})$
    - $= p(\boldsymbol{\Theta} \mid G, \boldsymbol{D}) \quad p(G)\, p(\boldsymbol{D} \mid G) \,/\, p(\boldsymbol{D})$
    - $\propto p(\boldsymbol{\Theta} \mid G, \boldsymbol{D}) \quad p(G)\, p(\boldsymbol{D} \mid G)$
    - $= p(\boldsymbol{\Theta} \mid G, \boldsymbol{D}) \quad p(G) \int p(\boldsymbol{D} \mid G, \boldsymbol{\Theta})\, p(\boldsymbol{\Theta} \mid G) d\boldsymbol{\Theta}$

1. 構造学習(structure learning)
    - DAG構造に関する事前の信念$p(G)$と$\boldsymbol{D}$から$p(G)p(\boldsymbol{D} \mid G)$を最大化する$G$を求める
2. パラメータ学習(parameter learning)
    - $\boldsymbol{D}$と構造学習で得られたGから$p(\boldsymbol{\Theta} \mid G, \boldsymbol{D})$を最大化する$\boldsymbol{\Theta}$を求める（簡単なので割愛）

- Note: $p(G)p(\boldsymbol{D} \mid G)$は最大化するが，$p(G, \boldsymbol{\Theta} \mid \boldsymbol{D})$を最大化しない$(G, \boldsymbol{\Theta})$も存在しうるので，妥協案

構造学習(structure learning)

- 既知のデータ$\boldsymbol{D}$からDAG G を求める
  - 基本的にDAG空間の探索，変数の数が増えると非現実的
- 膨大な数の手法が提案されている
  - 制約ベース(Constraint-based)
    - 条件付き独立性検定（CI test）による
  - スコアベース(Score-based)
    - スコア関数を最大化するDAGを探索する組み合わせ最適化問題として定式化
  - Local-to-global
    - 局所構造を見つけてできるだけ整合性を保ちつつ統合

制約ベース構造学習

- 忠実性の仮定が置かれる
  - DAG G が同時確率分布 Pと忠実: $I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_{\boldsymbol{d}} \Leftrightarrow I(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z})_P$
  - つまり，データの従う分布PがDAGで符号化できること
- 完全無向グラフから，条件付き独立性(CI)テストによって不要はエッジを取り除き，最後に有向化する

スコアベース構造学習

- DAGGからデータ$\boldsymbol{D}$が得られる尤度$p(\boldsymbol{D} \mid G)$を最大化する最適化問題
  - ここで，$p(\boldsymbol{D} \mid G)$は以下のように分解できる
$$p(\boldsymbol{D} \mid G) = \prod_{X \in \boldsymbol{V}} p(X \mid \boldsymbol{Pa}_X^G)$$
  - つまり，同時確率分布の因数分解の尤度を最大化する組み合わせ最適化
- 十分なデータの下で，離散変数のスコアベース構造学習はNP-hard
- DP, SA,
- 整数計画 (Integer Programming, IP)
  - 親変数の大きさの制約があれば，数十変数の最適DAG構造を見つけられる.
- 二次制約なし二値最適化(Quadratic Unconstrained Binary Optimization, QUBO)

local-to-global 構造学習

- 変数の数が増えると構造学習は非現実的
- Local-to-Global
  - 1) 少ない変数で局所構造学習（通常は各変数TのMB_T）を学習する
    - スコアベース or 制約ベース
    - e.g. S^2TMB
      - 忠実性の仮定の下でTの真のMB_Tを見つけられる
  - 2) 局所構造を最大限の整合性で組み合わせ，コンフリクトを解決
    - e.g. GGSL
      - 無限のデータと忠実性の仮定の下で真のDAG（と等価なDAG）を回復できる
- 各局所構造学習は小さくかつ並列に実行できる

Gao, Tian, and Qiang Ji. "Efficient score-based Markov Blanket discovery." International Journal of Approximate Reasoning 80 (2017): 277-293.

Gao, Tian, Kshitij Fadnis, and Murray Campbell. "Local-to-global Bayesian network structure learning." International Conference on Machine Learning. PMLR, 2017.

単一のDAGを決めない推論

- 単一のDAGを決めず既知のデータセット$\boldsymbol{D}$から未知のデータdを推定
- ベイジアンモデル平均(Bayesian model averaging)
  - モデル選択による不確実性を減らしたい
    - $p(d \mid \boldsymbol{D}) = \sum_{\text{あるモデル}}$(あるモデルの尤度)(あるモデルによる推定値)
  - ベイジアンネットワークにおいては以下のようになる
    - $p(d \mid \boldsymbol{D})$
    - $= \sum_{\text{ある}\mathrm{DAG}}$(ある*DAG*の尤度)(ある*DAG*による推定値)
    - $= \sum_{G} p(G \mid \boldsymbol{D}) \int p(d \mid G, \boldsymbol{\Theta}_G)\, p(\boldsymbol{\Theta}_G \mid G, \boldsymbol{D})$
  - △適当なDAGをランダムにとってくる
    - ほとんどのDAGの尤度は小さいので無駄が多い
  - ○マルコフ連鎖モンテカルロによるDAG空間の重点サンプリング
    - 尤もらしいDAGだけとってきたい（当たる宝くじだけ買いたい）

マルコフ連鎖モンテカルロ

- マルコフ連鎖: 離散的かつ直前の状態にだけ依存する確率過程
- モンテカルロ法: 乱数を用いた試行を繰り返して数値的に近似値を求める

- 以下を満たすように遷移核$p(X_{n+1} \mid X_n)$を定めサンプル列$X_0, X_1, \cdots, X_N$を生成するとサンプル列の確率分布が確率密度関数 $f(X)$に収束する.
  - マルコフ性: $p(X_{n+1} \mid X_n, \cdots, X_0) = p(X_{n+1} \mid X_n)$
  - 既約性: $\forall i, j \in \boldsymbol{C}, \exists n > 0 \;\; s.t. \; p(X_n = i \mid X_0 = j) > 0$
  - 非周期性: $\forall i \in \boldsymbol{C}, \; \gcd\{n \mid p(X_n = i \mid X_0 = i) > 0\} = 1$
  - 詳細平衡条件:

$$\forall i, j \in \boldsymbol{C}, \quad p(X_{n+1} = j \mid X_n = i) f(X_n = i) = p(X_{n+1} = i \mid X_n = j) f(X_n = j)$$

- $f(X)$の詳細な形がわからなくても，履歴に依存せず重点サンプリング

単一のDAGを決めない推論

- サンプリングの空間はいろいろ考えれらる
  - Structure-MCMC
    - エッジの追加・削除・反転で直接次のDAGを生成，DAG空間をサンプリング
  - Order-MCMC
    - トポロジカル順序の空間
  - Partition-MCMC
    - 変数集合の順序付き分割の空間
  - Layering-MCMC
    - 変数集合の順序付き分割の順序付き分割の空間

Kuipers, Jack, and Giusi Moffa. "Partition MCMC for inference on acyclic digraphs." *Journal of the American Statistical Association* 112.517 (2017): 282-299.

Viinikka, Jussi, and Mikko Koivisto. "Layering-MCMC for Structure Learning in Bayesian Networks." *Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020.