

■ 8.3.3 - 8.3.4 概要

8.3節では、マルコフ確率場（無向グラフィカルモデル）における条件付き独立性と分解特性について学んだ。分解特性については、ポテンシャル関数と分配関数を定義することによって同時確率分布を因数分解できることを確かめた。今回はマルコフ確率場の応用例と有向グラフとの関係性を学ぶ。

8.3.3項では、マルコフ確率場の分解特性によって画像中のノイズを除去する例を学ぶ。画像中の画素をノードとして表現し、観測された画像は本来の画像にノイズが加わったものとする。このときイジングモデルを仮定すると、エネルギーが小さくなるよう変数を更新することでノイズが取り除かれることを確認する。

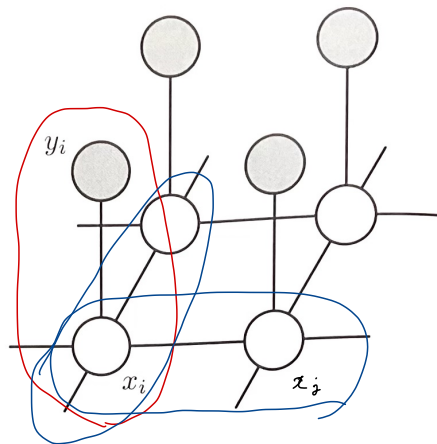
8.3.4項では、有向グラフと無向グラフの関係を学ぶ。無向グラフは有向グラフから構築することができる（ただし逆はあまりやられない）。また、有向グラフと無向グラフでは一般に表現できる分布の集合が異なり、どちらか一方に属する分布、どちらにも属する分布、どちらにも属さない分布が存在する。

■ 8.3.3 有向グラフとの関係

（細かい前提は読み合わせとします）

右図中の、ノード x_i を本来の画像の画素とし、 y_i をノイズが加わった画像の画素とする。これはイジングモデルと呼ばれ、本来の画像の隣接する画素同士が相関し、また、本来の画素とノイズありの画素も相関している様子を表現している。

本来の画像 $\{x_i\}$ を推定ために、同時分布（尤度）をポテンシャル関数に因数分解し、これを最大にする $\{x_i\}$ を求める方針をとる。



このグラフにおける極大クリークは、同一画素の x_i , y_i 同士（赤枠）と、ノイズなし画像における隣接するノード x_i , x_j （青枠）の2種類のみである。

これらの極大クリークにおけるポテンシャル関数を考える代わりにエネルギー関数を考える。
エネルギーは低い方が安定していると考えられるので、同符号（つまり同じ値を取る）のとき小さい値を取るよう設定するのが肝である。

(感想)

エネルギー関数の設定の仕方がだいぶ適当だったのが気になった。そもそもエネルギーと呼ぶ理由もよくわからないし、はじめからポテンシャル関数のみで扱う方が個人的には分かりやすく感じる。

$$(x_i, y_i) \text{ のエネルギー} : -\eta x_i y_i$$

$$(x_i, x_j) \text{ のエネルギー} : -\theta x_i x_j$$

また、本節の例では、ノイズなし画像の x_i の関数 $-h x_i$ をエネルギーに追加しても良い。(←ここは謎でした)

このモデルの全エネルギーは次のようになる。

$$E(x, y) = -h \sum_i x_i - \theta \sum_{\{i, j\}} x_i x_j - \eta \sum_i x_i y_i$$

これは、各クリークのポテンシャル関数を次のように設定することからも確かめられる。

$$(x_i, y_i) \text{ のポテンシャル} : e^{\eta x_i y_i}$$

$$(x_i, x_j) \text{ のポテンシャル} : e^{\theta x_i x_j}$$

$$(x_i) \text{ のポテンシャル} : e^{-h x_i} \quad (\text{極大クリークではないが...})$$

$$p(x, y) = \frac{1}{Z} \left\{ \prod_i e^{-h x_i} \cdot \prod_{\{i, j\}} e^{\theta x_i x_j} \cdot \prod_i e^{\eta x_i y_i} \right\}$$

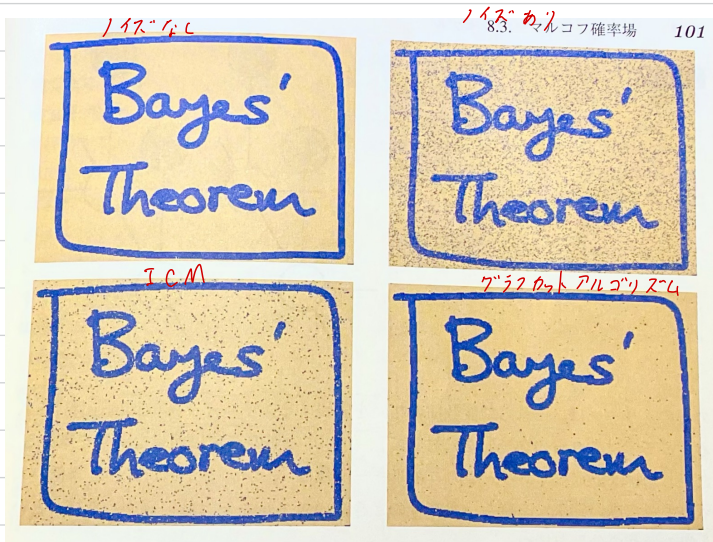
$$= \frac{1}{Z} \exp \left\{ - \left(\lambda \sum_i x_i - \theta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i \right) \right\}$$

$$= \frac{1}{Z} \exp \left\{ - E(x, y) \right\}$$

エネルギー関数としてとらえる。

同時分布（尤度）の最大化は、エネルギー $E(x, y)$ の最小化に等しく、**反復条件付きモード (ICM)** と呼ばれる手続きで行われる。まず $\{x_i\}$ を $x_i = y_i$ などで初期化する。次にノード x_i を1つずつ選んでその他のノード変数の値を固定したまま $x_i \in \{1, -1\}$ の2つの可能な状態におけるエネルギーを評価し、エネルギーが小さくなる方に値を更新する。これを適切な停止基準が満たされるまで実施する。

この手続きによって得られた画像は左下の画像であり、右上のノイズありの図からいくらかノイズが取り除かれていることがわかる。



多くの場合、グラフカットアルゴリズム（右下図）はICMよりも良い解を与える。

グラフカットアルゴリズムについてはこちらのpdfが参考になった。

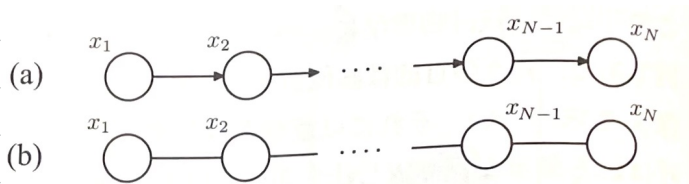
<http://www.f.waseda.jp/hfs/SSIITutorial.pdf>

8.3.4 有向グラフとの関係

8.3節の最後はこれまでの話題から外れて、性質の紹介的な内容となる。無向グラフから有向グラフを構築する方法を紹介し、有向グラフで表現できる分布と無向グラフで表現できる分布の関係を示す。

まず、無向グラフから有向グラフを構築するには、ノードの親同士をリンクさせることで得られる。これは**モラル化**と呼ばれ、得られた無向グラフは**モラルグラフ**と呼ばれる。

下図の例ではグラフ(a) → (b) に変換することでモラル化を行なっている。



このとき、有向グラフにおける条件付き確率の因数分解は次のように表現される。

$$p(x) = p(x_1) p(x_2 | x_1) p(x_3 | x_2) \cdots p(x_N | x_{N-1})$$

これに対し、ポテンシャル関数を次のように対応づけると、無向グラフにおける因数分解の形も得られる。

$$\left\{ \begin{array}{l} \psi_{1,2}(x_1, x_2) = p(x_1) p(x_2 | x_1) \\ \psi_{2,3}(x_2, x_3) = p(x_3 | x_2) \\ \vdots \\ \psi_{N-1,N}(x_{N-1}, x_N) = p(x_N | x_{N-1}) \end{array} \right. \quad \text{これは 2 項}$$

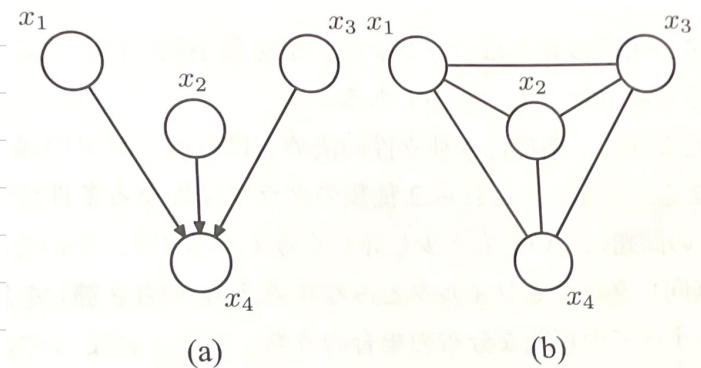
とすると、

$$p(x) = \frac{1}{Z} \prod_{i=1}^{N-1} \psi_{i,i+1}(x_{i+1} | x_i)$$

この例では有向グラフとモラルグラフでどちらも同じ条件付き独立性を有している。

すなわち、どちらのグラフにおいても $x_i \perp\!\!\!\perp x_j \mid \langle x_i \text{ と } x_j \text{ の 間の ノード} \rangle$ を表現している。

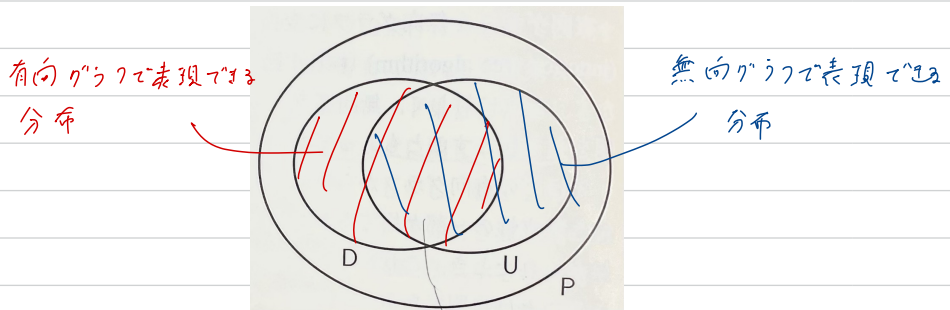
次の例はモラルグラフの方が元のグラフより少ない条件付き独立を有する例である。



例えば、(a) では $x_1 \perp\!\!\!\perp x_2 \mid \phi$ や $x_1 \perp\!\!\!\perp x_3 \mid \phi$ が成り立つが、(b) ではいかなる条件付き独立も成立しない。このように、モラルグラフを形成することは、有向グラフで表現できていた条件付き独立性を捨ててしまうことがある。

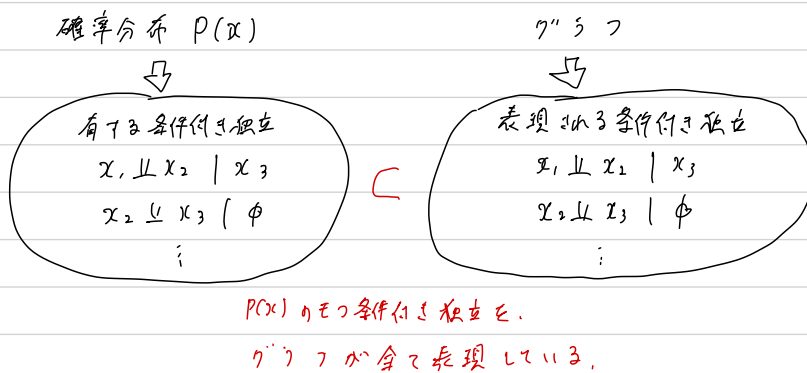
逆に無向グラフから有向グラフを形成する方法についても議論したくなるが、規格化制約が問題を起すこともあって、あまり行われていないらしい (よくわからん)

最後に、有向グラフが表現可能な条件付き独立を持つ分布と、無向グラフが表現可能な条件付き独立をもつ分布を比較する。結論から述べると両者に包含関係はなく、下記ベン図のような関係になる。

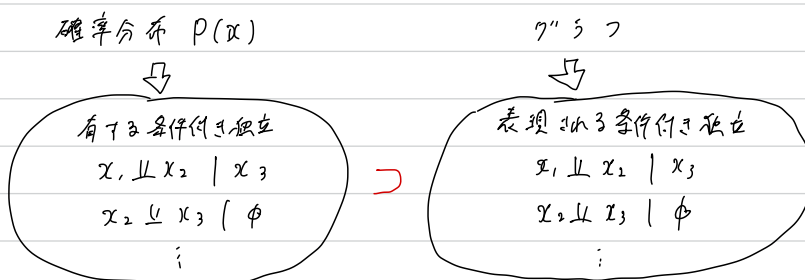


これを示すために、**依存性マップ (D-map)** や**独立性マップ (I-map)**、**完全マップ**と呼ばれる概念を導入する。これらは、ある確率分布にとってのそのグラフとの関係を示す。

まず、**依存性マップ**の定義は「ある分布の条件付き独立をその分布が全て有する」ことである。図にすると下のイメージ



一方で、**依存性マップ**は「グラフの持つ条件付き独立をある分布が全て有する」ことを言う。図にすると下のイメージ

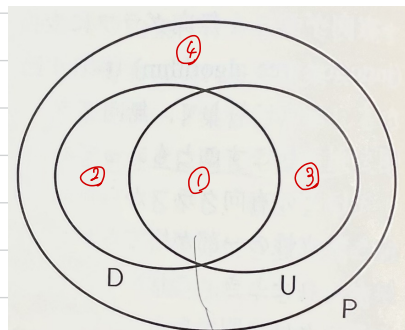


完全マップはこれらを同時に満たすグラフのことを指す。

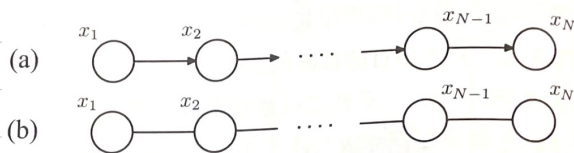
(完全マップにおける私の個人的なイメージは、「その分布における確率的関係を「完全に」表現しているグラフィカルモデルである」ということ)

最後に、完全マップが存在する分布の集合について考察する。全ての分布は次のいずれかのケースに属する。

- ① 有向グラフの完全マップも無向グラフの完全マップも存在する。
- ② 有向グラフの完全マップが存在するが、無向グラフの完全マップは存在しない
- ③ 無向グラフの完全マップが存在するが、有向グラフの完全マップは存在しない
- ④ 完全マップが存在しない

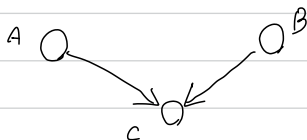


① の例



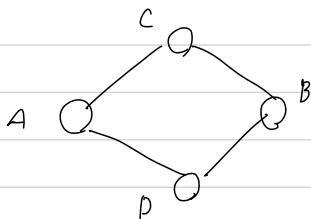
(a) と (b) はどちらも x_i と x_j の間のノードが観測されたときに $x_i \perp x_j$ が成立するような分佈を表現している。

② 例



これは $\{ A \perp B \mid \emptyset \}$ を満たす分佈の完全ママルクだが、無向グラフの完全ママルクは存在しない。

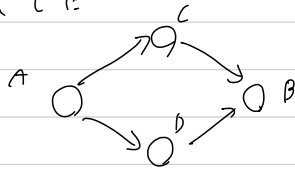
③ 例



これは $\left\{ \begin{array}{l} A \perp B \mid C, D \\ C \perp D \mid A, B \end{array} \right\}$ を満たす分佈の完全ママルクだが、

無向グラフの完全ママルクは存在しない。

式に



を考えると,

$$\left\{ \begin{array}{c|c} A \perp\!\!\!\perp B & C, D \\ C \perp\!\!\!\perp D & A \end{array} \right\} \quad \text{となる.}$$

④ 例

(思いつかなかった . もしありそうなら一稿に書きたい)

次回からはグラフィカルモデルを用いた推論の話に入る。グラフィカルモデルのうちいくつかのノードが観測されたときに、未知のノードを効率よく推論する方法について学んでいく。