

PRML輪読ゼミ

第8章：グラフィカルモデル

M2 芝原

2

グラフィカルモデルの便利な特徴

- 確率モデルの構造を視覚化する簡単な方法を提供し、新しいモデルの設計方針を決めるのに役立つ.
 - グラフの構造を調べることで、条件付き独立性などのモデルの性質に関する知見が得られる.
 - 精巧なモデルにおいて推論や学習を実行するためには複雑な計算が必要となるが、これを（数学的な表現を暗に伴う）グラフ上の操作として表現できる.
-

3

グラフについて

- ここで扱うグラフ：
 - ノード＝確率変数
 - リンク＝変数間の確率的関係を表現
- 有向グラフィカルモデル：ベイジアンネットワーク
 - 確率変数間の因果関係を表現
- 無向グラフィカルモデル：マルコフ確率場
 - 確率変数間の緩い束縛関係を表現

4

大まかなもくじ

■ 8章：グラフィカルモデル

- 8.1 ベイジアンネットワーク（有向グラフ）
 - 8.2 条件付き独立性
 - 8.3 マルコフ確率場（無向グラフ）
 - 8.4 グラフィカルモデルにおける推論
-

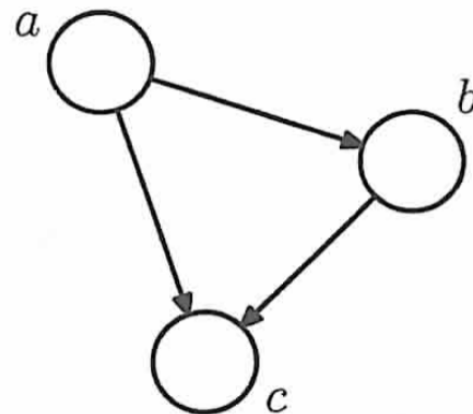
5

ベイジアンネットワーク（有向グラフ）

- 確率の乗法定理：同時分布を条件付き分布に分解

$$p(a,b,c) = p(c|a,b)p(a,b)$$

$$p(a,b,c) = p(c|a,b)p(b|a)p(a)$$



- ベイジアンネットワーク（有向グラフ）

- 上式の右辺をグラフ表現する
- 条件付き分布のノードに向かって、条件付けられた変数ノードから有向リンクをつなぐ.
- aはbの**親ノード**， bはaの**子ノード**
- 左辺は対称的だが， 右辺は非対称であることに注意！
- 暗に順番を選んでいた． 順序が違えば異なるグラフ表現になる．

- 全結合：全ノード組にリンクが貼られている状態のこと．

$$p(x_2, \dots, x_K) = p(x_K | x_1, \dots, x_{K-1}) \cdots p(x_2 | x_1) p(x_1)$$

6

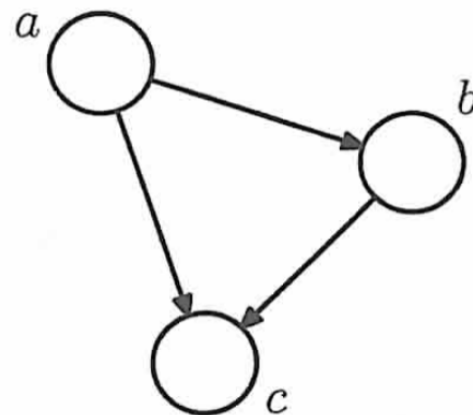
ベイジアンネットワーク（有向グラフ）

- 一般に K 個のノードを持つグラフに対応する同時分布は,

$$p(X) = \prod_{k=1}^K p(x_k \mid \underline{pa_k})$$

x_k の親ノード集合

$$X = \{x_1, \dots, x_K\}$$



- 分解特性：上の式のノードに変数集合や、ベクトル値変数に対応させることも可能.
- これ以降考える有向グラフは、**有向閉路**（あるノードから出発して、また初めのノードに戻ってくるような閉じた経路）を持たない**有向非循環グラフ**のみ考える.

7

グラフの記法

■ (例) ベイズ多項式回帰モデル

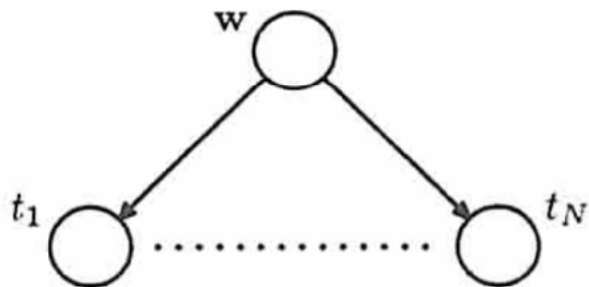
ー 確率変数

- 多項式係数ベクトル w
- 観測データ $t = (t_1, \dots, t_N)^T$

$$\underline{p(t, w)} = \underline{p(w)} \prod_{n=1}^N \underline{p(t_n | w)}$$

同時分布 事前分布 $n=1$ N個の条件付き分布

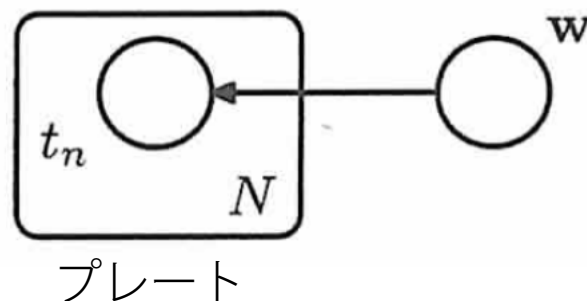
普通の有向グラフ表現



煩雑なので.



プレートを用いた有向グラフ表現



8

グラフの記法

■ (例) ベイズ多項式回帰モデルのパラメータ

- 確率変数

- 多項式係数ベクトル w

- 観測データ $t = (t_1, \dots, t_N)^T$

- 入力データ

 $X = (x_1, \dots, x_N)^T$

- ノイズの分散

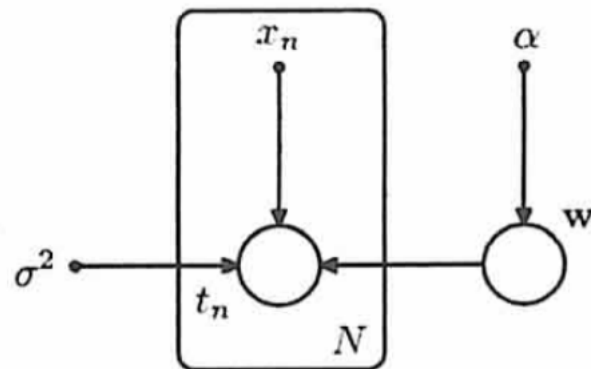
 σ^2 - 超パラメータ α (w 上のガウス事前分布の精度を表す)

パラメータを陽に書いた場合

$$p(t, w | x, \alpha, \sigma^2) = \underbrace{p(w | \alpha)}_{\text{事前分布}} \prod_{n=1}^N \underbrace{p(t_n | w, x_n, \sigma^2)}_{N\text{個の条件付き分布}}$$

グラフの記法

- 確率変数：○
- 決定的パラメータ：●



9

グラフの記法

■ (例) ベイズ多項式回帰モデルのパラメータ

- 確率変数

- 多項式係数ベクトル w

- 観測データ $t = (t_1, \dots, t_N)^T$

- 入力データ

$$X = (x_1, \dots, x_N)^T$$

- ノイズの分散

$$\sigma^2$$

- 超パラメータ α (w 上のガウス事前分布の精度を表す)

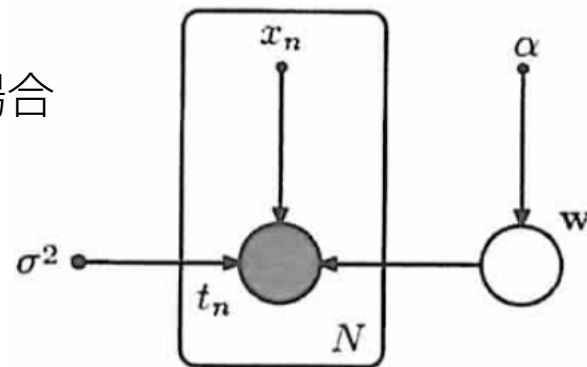
$$t = (t_1, \dots, t_N)^T$$

w

が観測されていて,
が観測されていない (隠れ変数) の場合

グラフの記法

- | | |
|----------------|--------------|
| ・ 確率変数 : ○ | ・ 観測変数 : 影あり |
| ・ 決定的パラメータ : ● | ・ 潜在変数 : 影なし |



10

生成モデル

- 伝承サンプリング：番号の小さい順にサンプリングする

$$p(X) = \prod_{k=1}^K p(x_k \mid \underline{pa_k})$$

x_k の親ノード集合

※子ノードは親ノードより大きい番号を振られているとする.

1. $p(x_1)$ に従うサンプルを発生させる. → これを $\hat{x}_1$ とする.

2. 順々に各ノードで $p(x_n \mid pa_n)$ に従うサンプルを発生させる.
親変数 pa_n の値にはサンプリングされた値を用いる.

3. 最後の変数 x_K のサンプルが得られたら終了.
全変数上の同時分布に従うサンプルを1つずつ発生させたことになる.

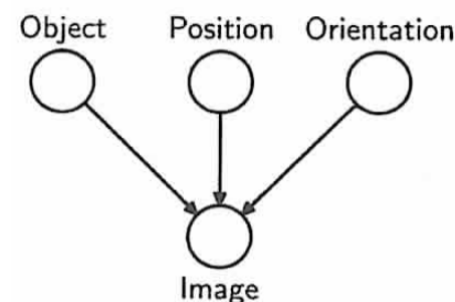
11

生成モデル

- 生成モデル：観測データが生成される因果過程を表現するモデル.

- (例) 物体認識問題：像が各観測データ点に対応. この観測データから物体の種類を推定する.

- 観測データ：物体の像
- 隠れ変数：物体の種類, 位置, 向き (相互に独立)



- 多項式回帰モデルは, 入力変数 x に関する確率分布があたえられておらず, 人工的なデータ点が生成できないため, 生成モデルではない.
- 伝承サンプリングを適用すると, 観測データの生成過程の模倣により, 観測データと同じ確率分布に従う架空データを発生させられる.

12

離散変数

- K 個の状態を取りうる離散変数 X 上の確率分布は、以下で与えられ、 N 個のパラメータで支配される。

$$p(X|\mu) = \prod_{k=1}^K \mu_k^{x_k}$$

規格化の制約
(全確率=1) $\sum_k \mu_k = 1$

パラメータ $\mu = (\mu_1, \dots, \mu_N)^T$

規格化の制約条件を使えば、
 $K-1$ 個の μ_k を指定すれば分布が決まる。

- 2つの K 状態離散変数 X_1 と X_2 の同時分布のモデル化

$$p(x_1, x_2 | \mu) = \prod_{k=1}^K \prod_{l=1}^K \mu_{kl}^{x_{1k} x_{2l}}$$

$$\sum_k \sum_l \mu_{kl} = 1$$

パラメータ μ_{kl} $x_{1k} = 1$ かつ $x_{2l} = 1$ という観測結果が得られる確率
 X_1 の第 k 成分 X_2 の第 l 成分

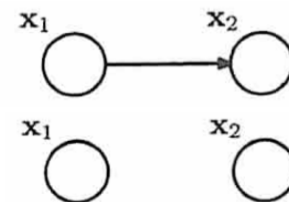
K^2-1 個の μ_{kl} を指定すれば分布が決まる。

13

離散変数

■ パラメータ数

1. 条件付き分布： K^2-1 個 $p(x_1, x_2) = p(x_2 | x_1) p(x_1)$
2. 互いに独立： $2(K-1)$ 個 $p(x_1, x_2) = p(x_2) p(x_1)$



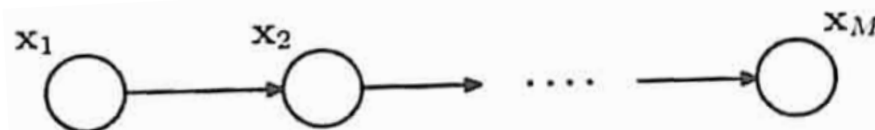
→リンクを切るとパラメータ数が減る.

■ 連鎖のパラメータ数

$$\underline{K-1} + \underline{(M-1)K(K-1)}$$

$p(x_1)$ のパラメータ数

$p(x_i | x_{i-1})$ の1個あたりパラメータ数 (x_{i-1} のK個の可能な状態に対して, x_i はK-1個のパラメータが必要)



– K に関して2次, 連鎖の長さ M に関して線的に増加.

14

線形ガウスモデル

前節＝離散変数集合上の同時確率分布の構成方法

ここ＝線形ガウスモデルに対応するグラフで，多変量ガウス分布を表現する．

- D個の変数上の任意の有向非循環グラフを考える．
 - － 線形ガウスモデルでは，分布の平均はノード i の親ノード pa_i 状態の線形結合で表せる．

$$p(x_i | pa_i) = N\left(x_i \mid \sum_{j \in pa_i} \underline{w_{ij}} x_j + \underline{b_i}, \underline{v_i}\right)$$

$N(x | m, \sigma^2)$ 変数 x を平均 m ，分散 σ^2 で表せるガウス分布（正規分布）

平均を支配するパラメータ

条件付き分布の分散

15

線形ガウスモデル

- 同時分布の対数は，グラフ内の全ノード上の条件付き分布の積の対数だから，

$$\ln p(x) = \sum_{i=1}^D \ln p(x_i | pa_i)$$

$$= - \sum_{i=1}^D \frac{1}{2v_i} \left(x_i - \sum_{j \in pa_i} w_{ij} x_j - b_i \right)^2 + \text{const.} \quad \rightarrow X \text{の成分に関する2次関数}$$

$$X = (x_1, \dots, x_D)^T$$

= 多変量ガウス分布

- この同時分布の平均と共分散は再帰的に求められる。

変数： $x_i = \sum_{j \in pa_i} w_{ij} x_j + b_i + \sqrt{v_i} \varepsilon_i$ ← 右の式より. $p(x_i | pa_i) = N\left(x_i \mid \sum_{j \in pa_i} w_{ij} x_j + b_i, v_i\right)$

平均0, 分散1のガウス確率変数

$$E[\varepsilon_i] = 0, \quad E[\varepsilon_i \varepsilon_j] = I_{ij}$$

平均： $E[x_i] = \sum_{j \in pa_i} w_{ij} E[x_j] + b_i$ ← 番号の小さいノードから順番に計算すれば，
 $E[X] = (E[x_1], \dots, E[x_D])^T$ の各成分が得られる。

16

線形ガウスモデル

- $p(x)$ の共分散行列の i, j 成分は, 共分散も再帰的に求められる.

$$\begin{aligned}\text{cov}[x_i, x_j] &= E\left[(x_i - E[x_i])(x_j - E[x_j])\right] \\ &= E\left[(x_i - E[x_i])\left\{\sum_{k \in pa_j} w_{jk} (x_k - E[x_k]) + \sqrt{v_j} \varepsilon_j\right\}\right] \\ &= \sum_{k \in pa_j} w_{jk} \text{cov}[x_i, x_k] + I_{ij} v_j\end{aligned}$$

17

線形ガウスモデル

■ 2つの極端な場合を考える.

– D個の孤立ノードの集合

- w_{ij} が存在しない.
- 平均: b_i , 共分散行列: v_i で決定.
- 同時分布は2D個のパラメータを持つ.
- 独立な1変量ガウス分布がD個集まったものと表現される.

– 全結合グラフ

- 各ノードが自分より小さい番号の親を持つ.
- つまり, w_{ij} は第 i 行が $i-1$ 個の成分を持つ下三角行列.
- w_{ij} のパラメータ数は $D(D-1)/2$ 個.
- v_i がD個あるので, 合計パラメータ数は $D(D+1)/2$
- これは一般の対象共分散行列の自由度に相当する.

– 中間的な複雑さのグラフは, 共分散行列が部分的に制約された同時ガウス分布に対応する.

$$x_i = \sum_{j \in pa_i} w_{ij} x_j + b_i + \sqrt{v_i} \varepsilon_i$$

$$E[x_i] = \sum_{j \in pa_i} w_{ij} E[x_j] + b_i$$

$$\text{cov}[x_i, x_j] = \sum_{k \in pa_j} w_{jk} \text{cov}[x_i, x_k] + I_{ij} v_j$$

18

条件付き独立性

$$p(a|b,c) = p(a|c) \quad \leftarrow \text{これは, } a \text{ の条件付き分布が } b \text{ の値に依存しないことを指す.}$$

$$\begin{aligned} p(a,b|c) &= p(a|b,c)p(b|c) \\ &= p(a|c)p(b|c) \end{aligned} \quad \leftarrow c \text{ で条件づけられた } a \text{ の周辺分布と } b \text{ の周辺分布の積で表される.}$$

- 条件付き独立：変数 c が与えられた時, a が b に対して独立であること（同時分布が周辺分布の積で書けること）

$$a \perp\!\!\!\perp b \mid c$$

- グラフィカルモデルは条件付き独立性が直接, グラフから読み取れる

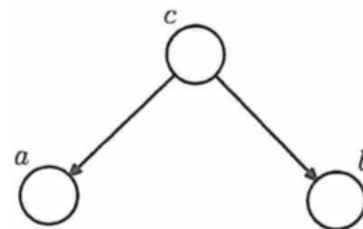
19

3つのグラフの例

条件付き独立を示す，代表的な3つのグラフを示す。

■ 右のグラフの同時分布

$$p(a,b,c) = p(a|c)p(b|c)p(c)$$



どの変数も観測されていないとする。
aとbが独立かどうかは，両辺をcに関して周辺化すれば良い。

$$p(a,b) = \sum_c p(a|c)p(b|c)p(c)$$

条件付き独立でない

→ $a \not\perp b$ $|\emptyset$ 空集合

積の形に分解できないので，

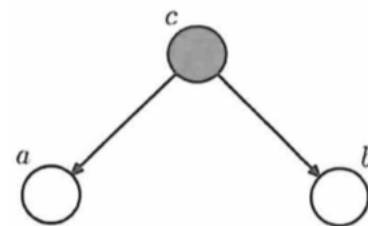
20

3つのグラフの例

■ 右のグラフの同時分布

c が与えられて、2つの変数を条件付けるとすると、

$$\begin{aligned} p(a, b | c) &= \frac{p(a, b, c)}{p(c)} \\ &= p(a | c) p(b | c) \end{aligned}$$



$$\rightarrow a \perp\!\!\!\perp b \mid c$$

ノード a から b への経路を考えると、ノード c は **tail-to-tail** である。

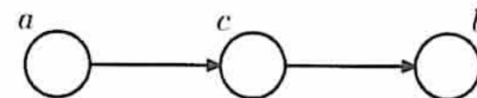
この条件付きノードは、 a から b への経路を **遮断** して条件付き独立にする。

21

3つのグラフの例

■ head-to-tail

同時分布 $p(a, b, c) = p(a)p(c|a)p(b|c)$



c に関して $p(a, b) = p(a) \sum_c p(c|a)p(b|c)$
周辺化する

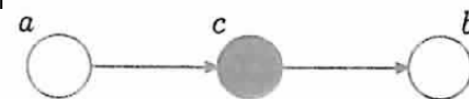
$$= p(a)p(b|a)$$

$$\rightarrow a \not\perp b \mid \emptyset$$

c で条件付け： c を観測すると， a から b への経路が遮断

$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)}$$

$$= p(a|c)p(b|c)$$



$$\rightarrow a \perp b \mid c$$

22

3つのグラフの例

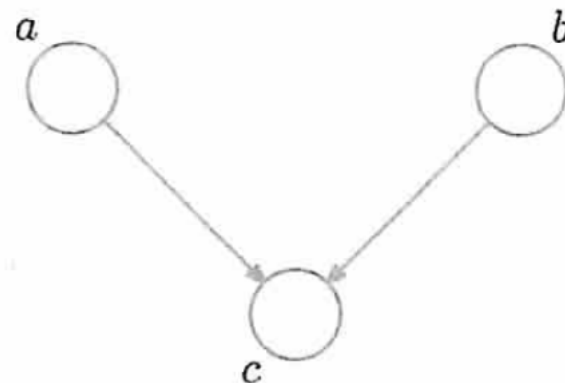
■ head-to-head:

同時分布 $p(a,b,c) = p(a)p(b)p(c|a,b)$

cに関して
周辺化する $p(a,b) = p(a)p(b)$

ノード c が観測されていないとき,
このノードが a から b への経路を「遮断」

※どの変数も観測されていないとき,
a と b が独立.



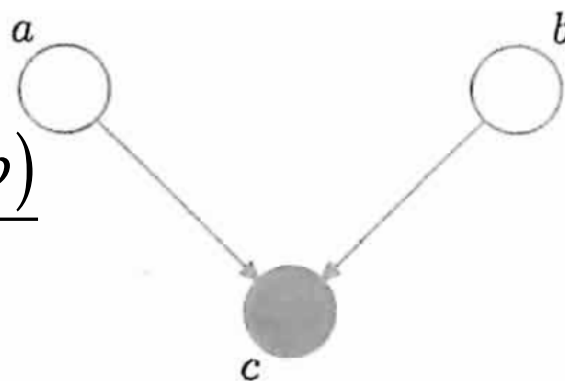
$$\rightarrow a \perp\!\!\!\perp b \mid \emptyset$$

c で条件付け：c を観測すると，遮断が解かれる

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(b)p(c|a,b)}{p(c)}$$

子孫：head-to-headノードか，その子孫

(矢印に従った経路のあるノード) のうちのいずれかが
観測されれば，その経路の遮断は解かれる． (証明略)



$$\rightarrow a \not\perp\!\!\!\perp b \mid c$$

23

有向分離

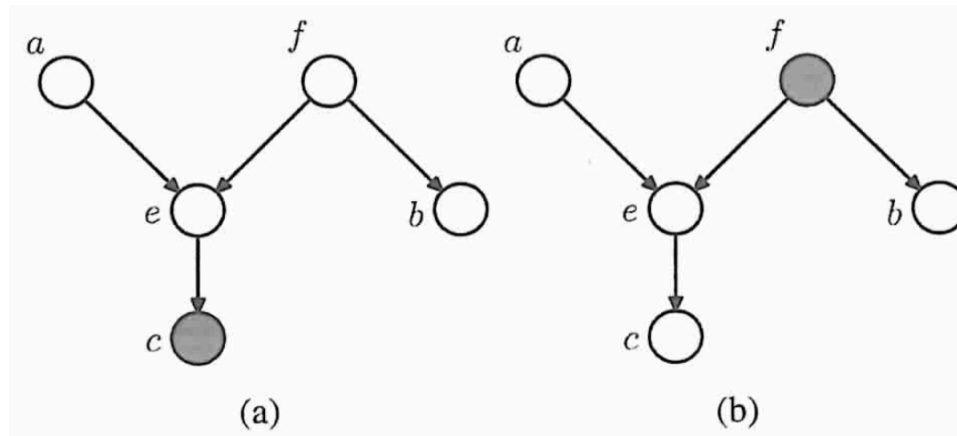
- 集合 A, B, C について, $A \perp\!\!\!\perp B \mid C$ を調べたいとき, さっきの3つのグラフの特性を用いて調べる.
- 「遮断されている」の定義
 - 以下の条件のいずれかを満たすノードを含む経路は遮断されている.
 - 1. 集合 C に含まれるノード (観測値) であって, 経路に含まれる矢印がそこでhead-to-tailか, tail-to-tailである.
 - 2. 経路に含まれる矢印がそのノードでhead-to-headで, 自身かそのすべての子孫がどれも集合 C に含まれない.
(観測されていない)

(例) 有向分離

■ 「遮断されている」の定義

1. 集合 C に含まれるノード（観測値）であって、経路に含まれる矢印がそこでhead-to-tailか、tail-to-tailである。
2. 経路に含まれる矢印がそのノードでhead-to-headで、自身かそのすべての子孫がどれも集合 C に含まれない。（観測されていない）

(例)



(a)

- $a \rightarrow b$ は、ノード f で遮断されない。 : tail-to-tail
- ノード e でも遮断されない。 : head-to-headだが、子孫が観測値

(b)

- $a \rightarrow b$ は、ノード f で遮断される。 : tail-to-tailな観測値ノード
- ノード e でも遮断される。 : head-to-headで、子孫に観測値がない。

25

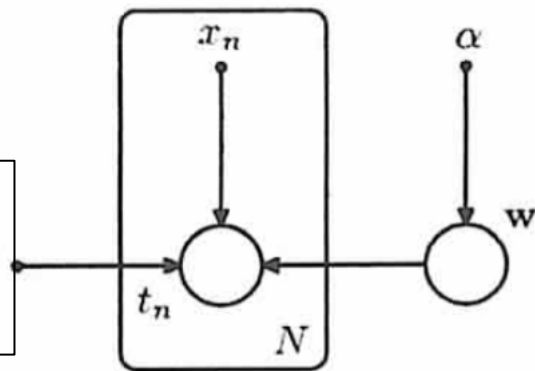
(例) 有向分離

- (例) ベイズ多項式回帰モデルのパラメータノード
 - 確率変数
 - 多項式係数ベクトル w
 - 観測データ $t = (t_1, \dots, t_N)^T$
 - 入力データ $X = (x_1, \dots, x_N)^T$
 - ノイズの分散 σ^2
 - 超パラメータ α (w 上のガウス事前分布の精度を表す)

グラフの記法

- 確率変数：○
- 決定的パラメータ：•

- パラメータノードは、観測済みノードと考える。
- が、周辺分布がなく、親ノードを持つことがないので、 σ^2 必ずtail-to-tailになる。→「遮断されている」と考える。



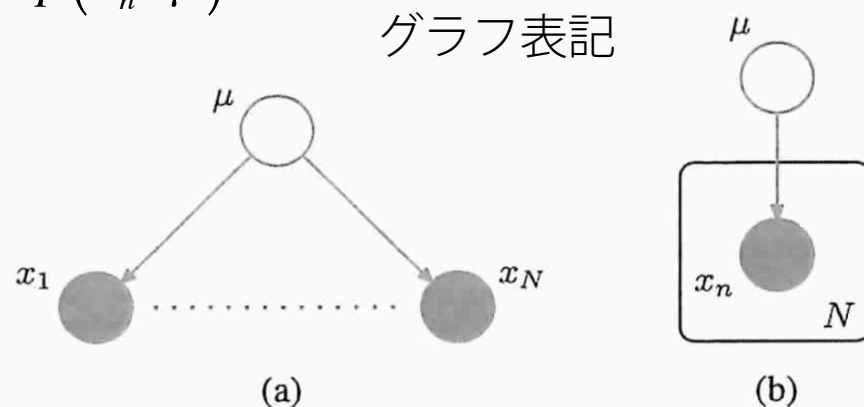
26 有向分離

- 1変量ガウス分布の平均の事後分布を得たい。

事前分布 $p(\mu)$ 条件付き分布の集合 $p(x_n | \mu)$
で同時分布が決まる。

この問題の目的は,

$D = \{x_1, \dots, x_N\}$ が観測された下で,
 μ を推論すること。



全経路は, 変数ノード μ がtail-to-tailなので遮断される。

→ μ が与えられた下で観測値 $D = \{x_1, \dots, x_N\}$ は独立。

$$p(D | \mu) = \prod_{n=1}^N p(x_n | \mu)$$

μ を積分消去した場合には
観測値は一般に独立でない。

$$p(D) = \int_{-\infty}^{\infty} p(D | \mu) p(u) du \neq \prod_{n=1}^N p(x_n)$$

- D個のノードを持つ有向グラフで表される同時分布は,

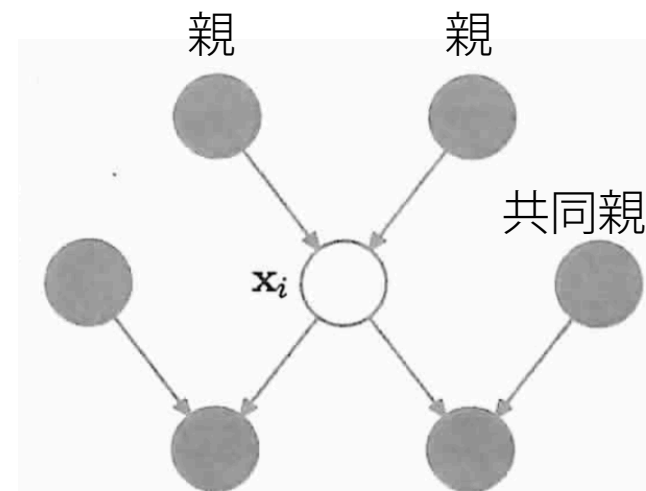
$$p(x_i | x_{\{j \neq i\}}) = \frac{p(x_1, \dots, x_D)}{\int p(x_1, \dots, x_D) dx_i} = \frac{\prod_k p(x_k | pa_k)}{\int \prod_k p(x_k | pa_k) dx_i}$$

x_i に関係ない項は分子と分母でキャンセルされる。

残るのは,

- ・ ノード x_i 自身の条件付き分布 $p(x_i | pa_i)$
- ・ x_i を親に持つノードの条件付き分布 $p(x_k | pa_k)$

- マルコフブランケット：ノードを孤立させるための最小のノード集合のこと．すべての共同親ノードも観測する必要がある．（head-to-headノードの子孫を考慮）



28

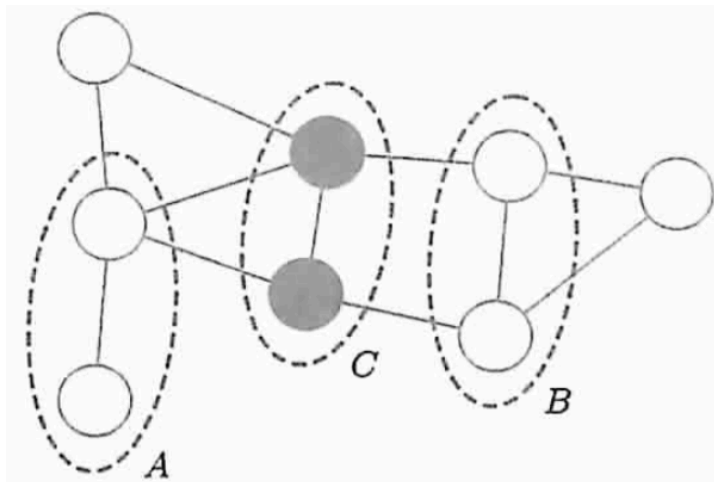
マルコフ確率場／条件付き独立性

■ 無向グフィカルモデル＝マルコフ確率場

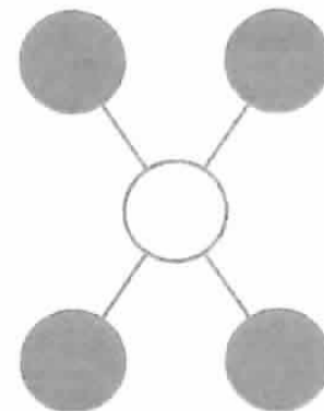
- 親ノード・子ノードの関係性がない.
- head-to-headノードのようなわかりにくい現象が起こらない.

■ 条件付き独立性

- 集合 A に含まれるノード→集合 B に含まれるノードの任意の経路で, 集合 C を通らないこと.



マルコフ確率場の
マルコフブランケット



29

分解特性

- グラフ全体の同時分布が，局所的な変数集合上の関数の積として，どう表現されるか考える．「局所性」とは？？

1つのリンクで直接接続されない x_i と x_j の2つのノードを考える．

グラフ上の他のすべてのノードの値が与えられた下で，条件付き独立である．

$$p(x_i, x_j \mid x_{\setminus \{i,j\}}) = p(x_i \mid x_{\setminus \{i,j\}}) p(x_j \mid x_{\setminus \{i,j\}})$$

すべての変数集合から， x_i と x_j を取り除いたもの

→ 局所的なクリーク概念を導入すると考えやすそう．

30

分解特性

- クリーク：全ノード組にリンクが貼られた，グラフ中の部分集合のこと．クリークは全結合．
- 極大クリーク：もうひとつノードを加えるとクリークで無くなってしまいうようなクリーク．

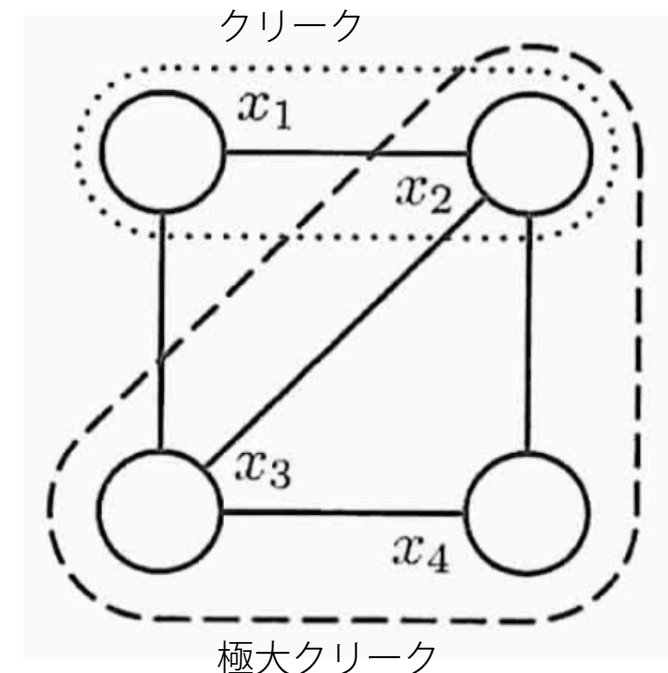
(例)

クリーク：

$$(\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_2\}, \{x_1, x_3\})$$

極大クリーク：

$$(\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\})$$



31

分解特性

- 同時分布を因数分解したときの各因子を，クリークが含む変数集合の関数にすれば，局所→全体が表現できる.
- クリークは極大クリークの部分集合→クリーク上で任意の関数を定義したなら，クリークの関数は極大クリークの関数に含まれる.
- つまり，同時分布は，グラフの極大クリーク上のポテンシャル関数の積の形で書ける.

同時分布

$$p(X) = \frac{1}{Z} \prod_C \psi_C(X_C)$$

C 極大クリーク
 X_C 極大クリーク内の変数集合
 ψ_C ポテンシャル関数

規格化定数

$$Z = \sum_X \prod_C \psi_C(X_C)$$

ボルツマン分布
で表すと便利.

$$\psi_C(x_C) = \exp\{-E(X_C)\}$$

エネルギー関数

※ポテンシャル関数は確率的なものに限定されない。
 ので，正しく規格化される必要がある。

32

(例) 画像のノイズ除去

観測画像2値ピクセル値 $y_i \in \{-1, 1\}$ の二次元配列

(ノイズのない2値画像 $x_i \in \{-1, 1\}$ からランダムに反転)

2種類のクリークを考える. それぞれのエネルギー関数は,

1) ノイズレベルが低いので, ノイズ付加前後で相関

$-\eta x_i y_i$ 同符号のとき低いエネルギー (高い確率)

2) 隣接するピクセル間の強い相関

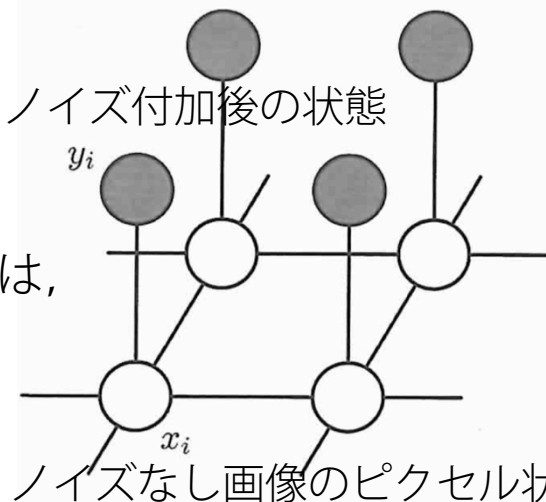
$-\beta x_i x_j$

ノイズなしピクセル値の関数として, hx_i

まとめると, このモデルの全エネルギーは,

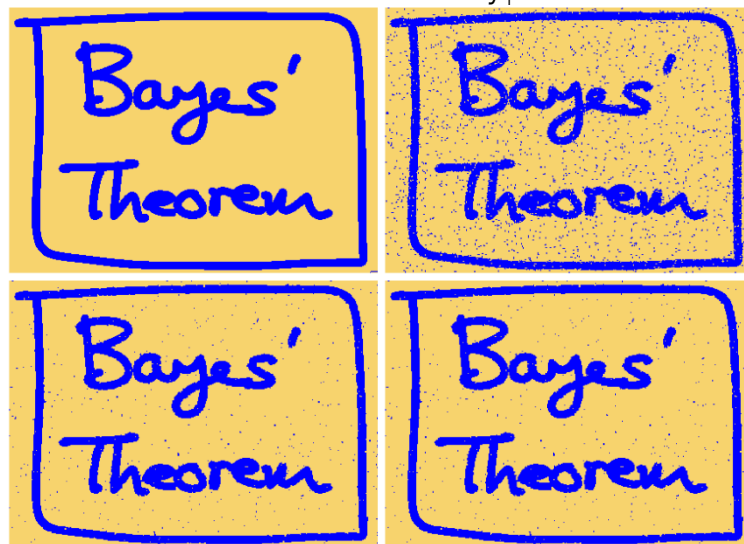
$$E(x, y) = h \sum_i x_i - \beta \sum_{\{i, j\}} x_i x_j - \eta \sum_i x_i y_i$$

同時確率: $p(x, y) = \frac{1}{Z} \exp\{-E(x, y)\}$



ノイズなし

y_i (10%反転)



ICM法 (96%一致)

グラフカットアルゴリズム
(99%一致)

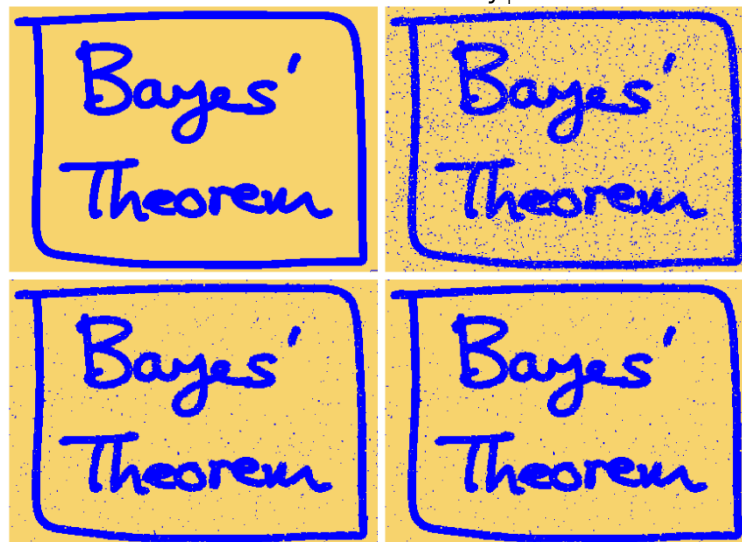
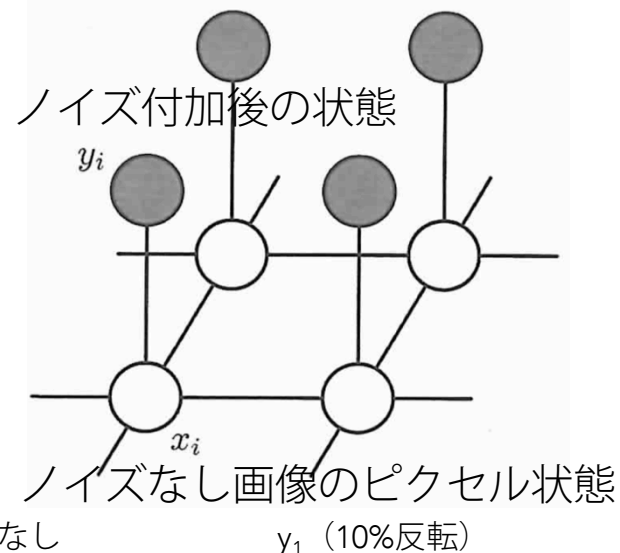
33

(例) 画像のノイズ除去

ICM法（反復条件付きモード法）

1. 変数 $\{x_i\}$ を初期化する. 例えば, $x_i = y_i$
2. ノード x_i を1つ選んで, そのノードの状態だけを変えたときの全エネルギーを評価.
3. ノード x_i の状態を, 全エネルギーが小さくなる状態に設定.
4. 全ノードを選び評価し終わった時点で終了.

※大域的最適解に至るとは限らない.



ICM法（96%一致）

グラフカットアルゴリズム
(99%一致)

34

有向グラフとの関係

■ 連鎖の有向グラフを無向グラフへ変換

(a) を条件付き確率で記述

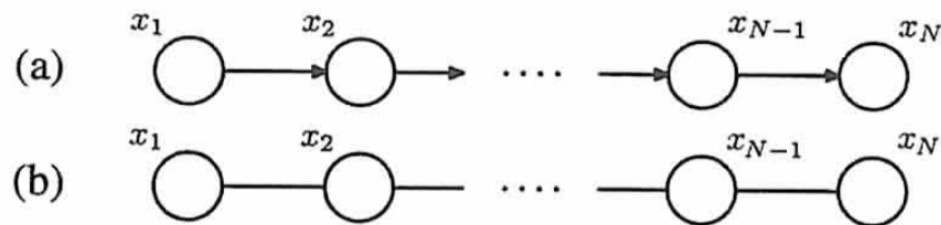
$$p(X) = p(x_1)p(x_2 | x_1)p(x_3 | x_2) \cdots p(x_N | x_{N-1})$$

(b) を極大クリークのポテンシャル関数で記述

$$p(X) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$



(a) と (b) の対応付けから,



$$\psi_{1,2}(x_1, x_2) = p(x_1)p(x_2 | x_1)$$

$$\psi_{2,3}(x_2, x_3) = p(x_3 | x_2)$$

⋮

$$\psi_{N-1,N}(x_{N-1}, x_N) = p(x_N | x_{N-1}) \quad Z = 1$$

矢印をとるだけで
条件付き分布がクリークポテンシャルに吸収される。

35 モラル化

■ 一般的な有向グラフから無向グラフへの変換

(a) を条件付き確率で記述

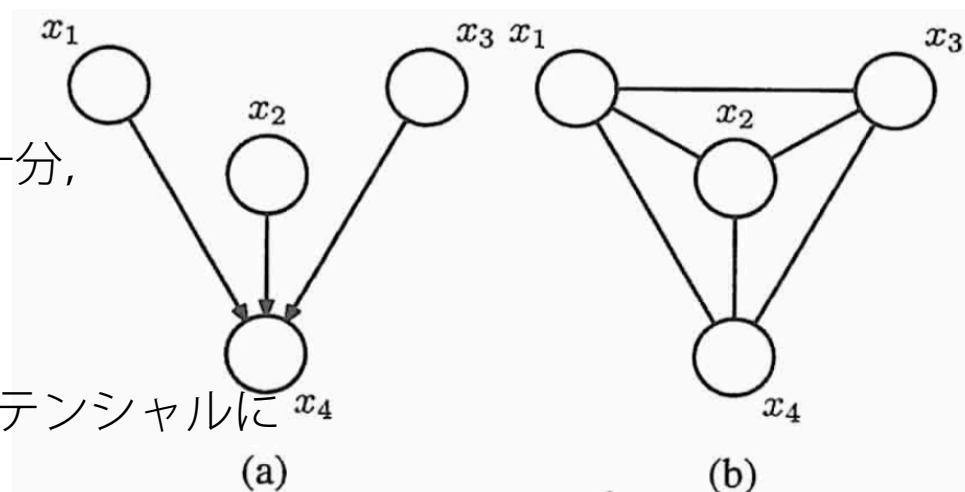
$$p(x) = p(x_1)p(x_2)p(x_3)p(x_4 | x_1, x_2, x_3)$$

親を2つ以上持つ有向グラフのノード
に関しては、矢印をとるだけでは不十分、

→ head-to-head経路ノードが厄介.

この条件付き分布を1つのクリークポテンシャルに
吸収させたい.

→ リンクをつなぐ必要がある.



■ モラル化：親同士の結婚を行うこと?? (言葉が, , ,) モラル グラフが作成される.

- グラフの幾つかのノードが観測値に固定されたとき、残ったノードに関する事後分布を計算したい.
- グラフの構造を利用することで、アルゴリズムの構造の見通しをよくできる.
- 多くのアルゴリズムは、局所的なメッセージのグラフ全体にわたる伝播として表現できる.

37

グラフィカルモデルにおける推論

■ ベイズの定理の復習

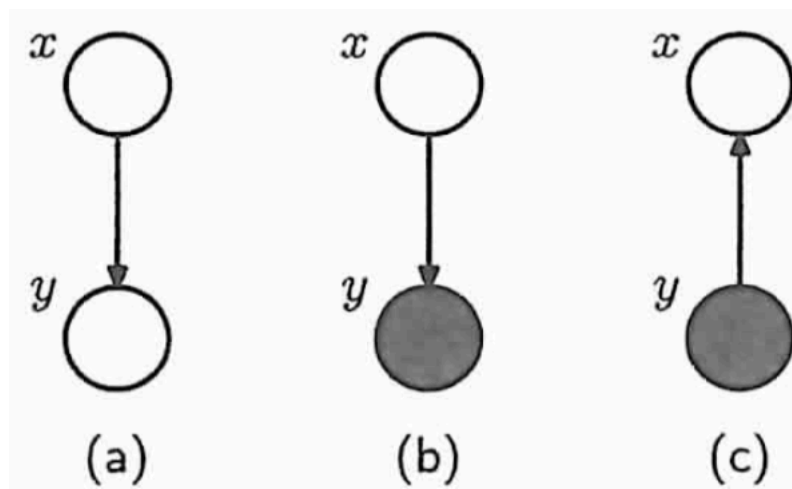
$$(a) \quad p(x, y) = p(x)p(y|x)$$

- (b) 変数 y が観測されたとき $\rightarrow$ 潜在変数 x の周辺分布 $p(x)$ は事前分布
 目的は、この事前分布に対する x 上の事後分布を推論すること。
 確率の加法定理と乗法定理より、

$$p(y) = \sum_{x'} p(y|x')p(x')$$

ベイズの定理から、

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



$\rightarrow$ 同時分布が矢印の向きが反転したグラフ (c) で表現出来る！

38

連鎖における推論

有向連鎖 (a) と無向連鎖 (b) とは全く同じ条件付き独立性を表現する. このグラフの同時分布は,

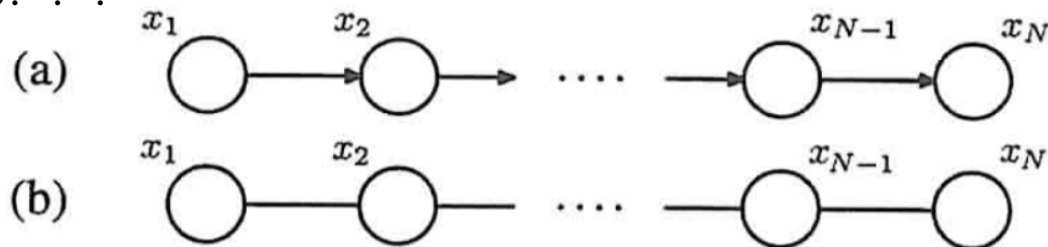
$$p(x) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

- 各ポテンシャル関数は $K \times K$ の表で与えられる. ので, 同時分布は $(N-1)K^2$ 個のパラメータを持つ. 連鎖の途中のノード x_n の周辺分布は,

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(x)$$

- 計算量は, ,

- K 状態変数が N 個あるので, 取りうる状態数は K^N 個. よって, 同時分布を計算してそれらを保存し, さらに周辺化して $p(x_n)$ を求めるには, 連鎖の長さ N に関して指数オーダーの計算量になる. . .



39

連鎖における推論

- ここで、条件付き独立性を用いる。代入し足し算と掛け算の順序を入れ替えて、連鎖の端から順に計算すれば良い。

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} \frac{1}{Z} \psi_{1,2}(x_1, x_2) \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

- グラフ上では、和演算1回あたり1つずつノードが除去されていく。
 - 計算量は $O(NK^2)$ で、連鎖の長さ N について線形。指数的なコスト増加を緩和。

$$p(x_n) = \frac{1}{Z}$$

周辺分布が2つの因子の積に規格化定数をかけたもので表される。

$\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \right] \right]$	$\longrightarrow$	$\mu_\alpha(x_n)$
$\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \right]$	$\longrightarrow$	$\mu_\beta(x_n)$

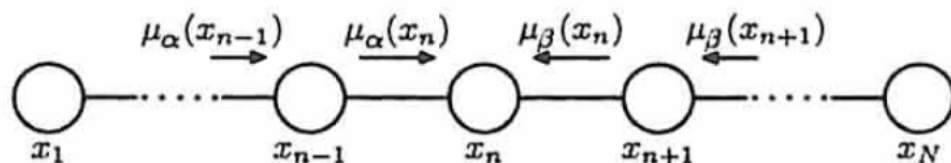
40

連鎖における推論

周辺分布を，局所的なメッセージの伝播の集積と考える．

マルコフ連鎖

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$



$\mu_\alpha(x_n)$: ノード x_{n-1} からノード x_n へ前向きに伝わるメッセージ

$\mu_\beta(x_n)$: ノード x_n からノード x_{n+1} へ後向きに伝わるメッセージ

$$\mu_\alpha(x_n) = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \left[\sum_{x_{n-2}} \cdots \right] = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_\alpha(x_{n-1})$$

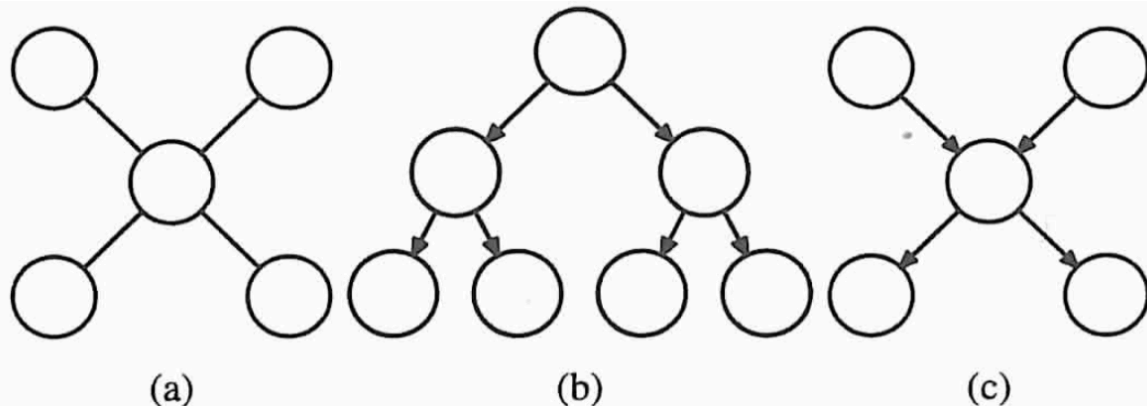
初期値： $\mu_\alpha(x_2) = \sum_{x_1} \psi_{1,2}(x_1, x_2)$ を計算したら，あとは再帰的に得られる。
後ろ向きも同様。

全ノードの周辺分布を計算する前に，2つの向きであらかじめ計算しておけば良い。

さらに一般的に考えるため、木 (tree) を考える.

■ 木

- 無向グラフ (a) : 任意のノードの組の間に唯一の経路が存在するグラフ.
- 有向グラフ (b) : 根ノードと, 他の全てのノードはそれぞれ親を1つずつ持つグラフとして定義される.
- モラル化してもリンクを付け加える必要がない.
- 有向多重木 (c) : 2つ以上の親を持つノードが存在するが, 任意の2ノード間の経路が1つしかない有向グラフ. モラル化すると閉路を持つ.

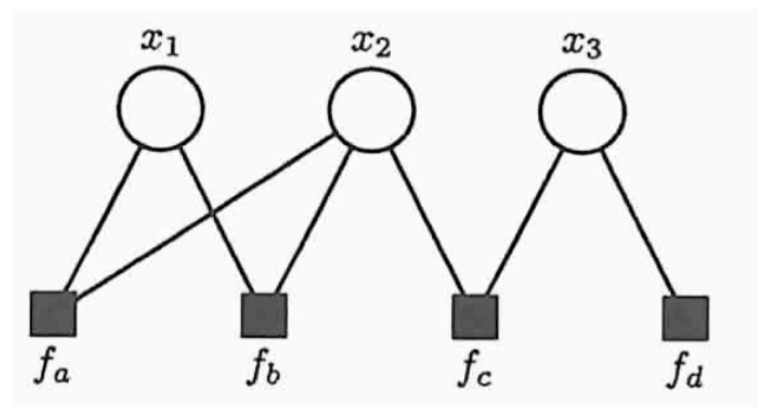


42 因子グラフ

- 因子グラフ：変数を表現するノードに因子そのものに対応するノードを付け加えたグラフ。
 - リンクは異種のノード間のみをつなぐと再定義.
 - 同一変数集合上で定義される2つの因子があることに注意. 内在する因数分解に関してより詳細な情報を表現.

一般的に, $p(X) = \prod_s f_s(X_s)$ とかく. $f_s(X_s)$: 因子 (変数集合の関数)

(例) $p(X) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$



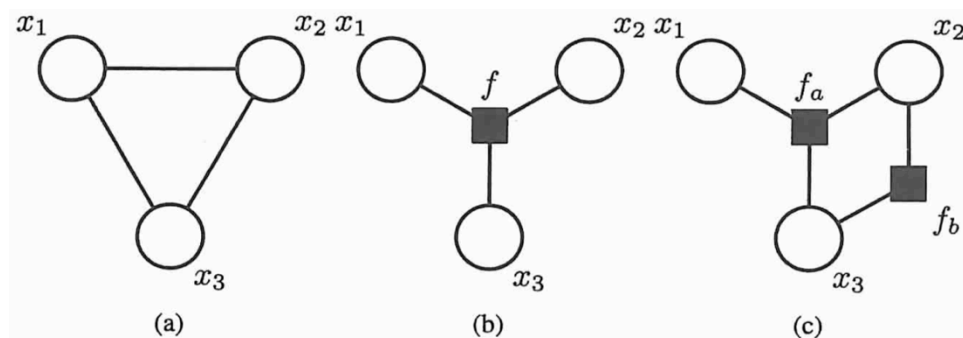
43 因子グラフ

■ 無向グラフからの変換：クリークポテンシャルを変換

(a) $\psi(x_1, x_2, x_3)$

(b) $f(x_1, x_2, x_3) = \psi(x_1, x_2, x_3)$

(c) $f_a(x_1, x_2, x_3) f_b(x_2, x_3) = \psi(x_1, x_2, x_3)$

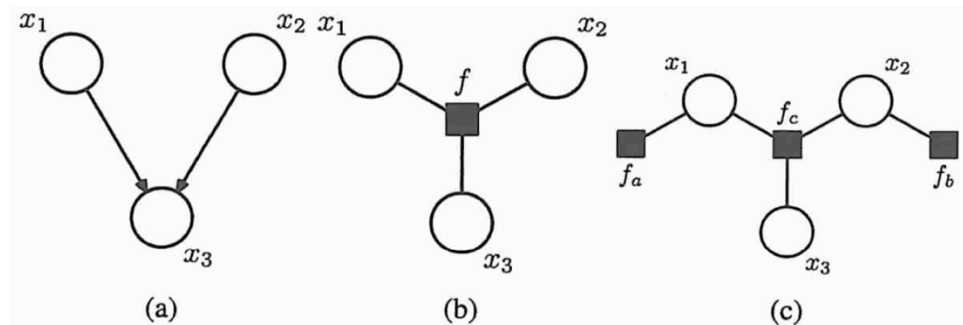


■ 有向グラフからの変換

(a) $p(x_1)p(x_2)p(x_3 | x_1, x_2)$

(b) $f(x_1, x_2, x_3) p(x_1) p(x_2) p(x_3 | x_1, x_2)$

(c) $f_a(x_1) = p(x_1), f_b(x_2) = p(x_2),$
 $f_c(x_1, x_2, x_3) = p(x_3 | x_1, x_2)$



どちらも、複数の因子グラフが元のグラフに対応する。

44

積和アルゴリズム

- もともとのグラフは無向木, 有向木, 多重木のいずれかと仮定.
- ここでの目的は,
 1. 周辺分布を求めるための効率良い厳密推論アルゴリズムを得ること.
 2. 複数の周辺分布を計算したいとき, 計算の重複をなくして効率化すること.

周辺分布の定義 (x を除くすべての変数に関する同時分布の和) から,

$$p(x) = \sum_{X \setminus x} p(X) \longleftarrow p(X) = \prod_s f_s(X_s)$$

に因子グラフ表現を用い, 和演算と積演算を交換して効率的なアルゴリズムを得る.

45

積和アルゴリズム

- 下図のような, グラフの1部分について考える.

$$p(x) = \sum_{X \setminus x} p(X)$$

- 木構造なので, 同時分布の因子を変数ノード x に隣接する各因子ノードごとにグルーピングできる.

$$p(X) = \prod_{s \in ne(x)} F_s(x, X_s)$$

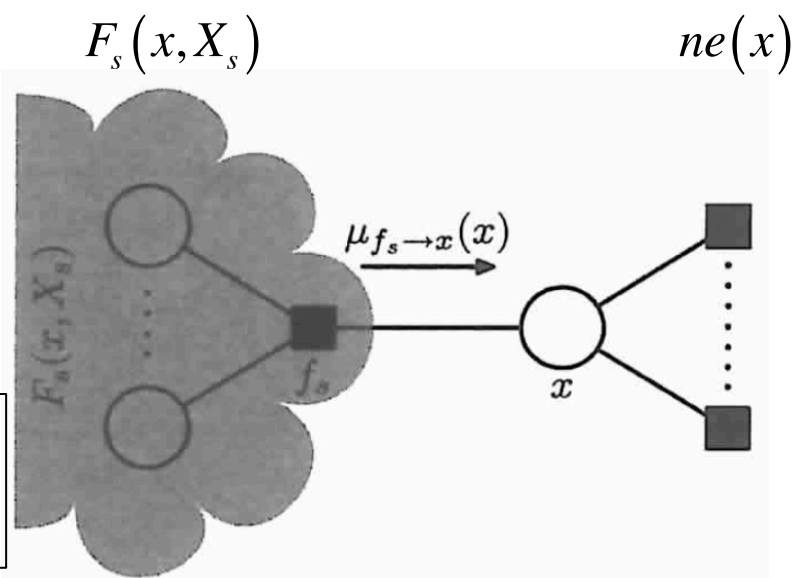
$ne(x)$: x に隣接する因子ノードの集合

X_s : 因子ノード f_s を通して
変数ノード x に接続される
部分木に属するすべての変数

$$p(x) = \prod_{s \in ne(x)} \left[\sum_{X_s} F_s(x, X_s) \right]$$

$$= \prod_{s \in ne(x)} \mu_{f_s \rightarrow x}(x)$$

因子ノード f_s から
変数ノード x への
メッセージ $\mu_{f_s \rightarrow x}(x) \equiv \sum_{X_s} F_s(x, X_s)$



46

積和アルゴリズム

さらに、メッセージの各因子も因数分解可能だから、（下図）

$$F_s(x, X_s) = f_s(x, x_1, \dots, x_M) G_1(x_1, X_{s1}) \dots G_M(x_M, X_{sM})$$

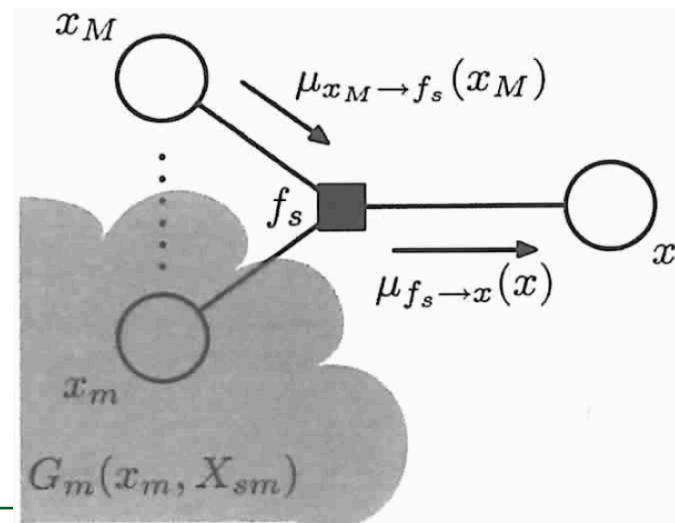
メッセージの式に代入

$$\begin{aligned} \mu_{f_s \rightarrow x}(x) &= \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in ne(f_s) \setminus x} \left[\sum_{X_{sm}} G_m(x_m, X_{sm}) \right] \\ &= \sum_{x_1} \dots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in ne(f_s) \setminus x} \mu_{x_m \rightarrow f_s}(x_m) \end{aligned}$$

$ne(f_s)$: 因子ノード f_s に隣接する変数ノードの集合

変数ノードから因子ノードへのメッセージ

$$\mu_{x_m \rightarrow f_s}(x_m) \equiv \sum_{X_{sm}} G_m(x_m, X_{sm})$$



47

積和アルゴリズム

■ 因子→変数のメッセージの計算方法は,

1. 因子ノードに接続された他のすべてのリンクを通して入ってくるメッセージの積を計算
2. このノードに対応する因子をかける.
3. 入ってきたメッセージに関連するすべての変数について周辺化する.

$$\mu_{f_s \rightarrow x}(x) = \underbrace{\sum_{x_1} \cdots \sum_{x_M}}_3 \underbrace{f_s(x, x_1, \dots, x_M)}_2 \underbrace{\prod_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \rightarrow f_s}(x_m)}_1$$

48

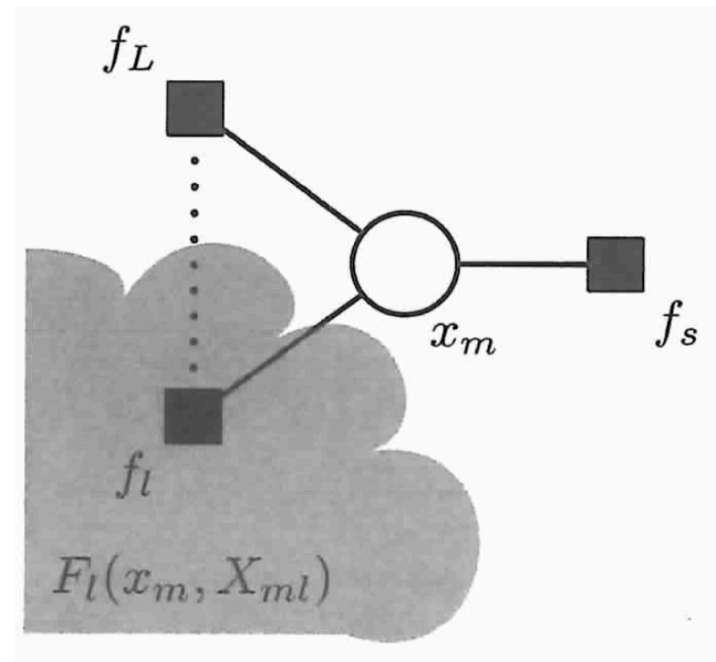
積和アルゴリズム

最後にもう一度，変数ノードから因子ノードへ伝わるメッセージを導出する．

$$G_m(x_m, X_{sm}) = \prod_{l \in ne(x_m) / f_s} F_l(x_m, X_{ml})$$

$$\mu_{x_m \rightarrow f_s}(x_m) = \prod_{l \in ne(x_m) / f_s} \left[\sum_{X_{ml}} F_l(x_m, X_{ml}) \right]$$

$$= \prod_{l \in ne(x_m) / f_s} \mu_{f_l \rightarrow x_m}(x_m)$$



つまり，ある変数ノードからそれに隣接するある因子ノードに，それらの間のリンクを伝わって送られるメッセージを計算するには，残りのリンクから伝わるメッセージの積を計算すれば良い．

49

積和アルゴリズム

- ノード x を木の根ノードであるとみなし, すべての葉ノードからのメッセージ伝播を考えれば良い.

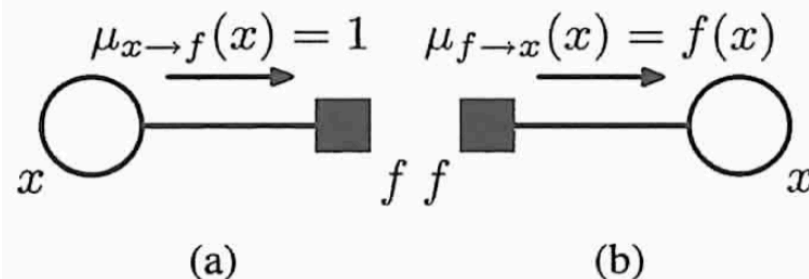
$$\mu_{x_m \rightarrow f_s}(x_m) = \prod_{l \in \text{ne}(x_m) \setminus f_s} \mu_{f_l \rightarrow x_m}(x_m)$$

$$\mu_{f_s \rightarrow x}(x) = \sum_{x_1} \cdots \sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m \in \text{ne}(f_s) \setminus x} \mu_{x_m \rightarrow f_s}(x_m)$$

- 上の定義から,

葉ノードが変数ノードなら, $\mu_{x \rightarrow f}(x) = 1$

葉ノードが因子ノードなら, $\mu_{f \rightarrow x}(x) = f(x)$



50

積和アルゴリズムのまとめ

1. 変数ノード x を因子グラフの根ノードとみなし, 式からグラフの葉ノードにおけるメッセージを初期化する.
2. メッセージパッシングの式を再帰的に適用する. すべてのリンクを伝播し, 根ノードがすべての隣接ノードからのメッセージを受け取るまで続ける.
3. 根ノードがすべての隣接ノードからメッセージを受け取り, 求める周辺分布を下の式から計算する.

$$p(x) = \prod_{s \in ne(x)} \mu_{f_s \rightarrow x}(x)$$

51

全ノードについて計算したい

1. 任意の（変数か因子）ノードを選んでそれを根ノードとする.
2. 全葉ノードから根ノードへメッセージを伝播させる.
3. 根ノードからすべての隣接ノードにメッセージを送る.
4. 次に, 根ノードの隣接ノードから, 根ノードから遠ざかる方向にメッセージを送る.
5. メッセージはグラフ中のすべてのリンクにわたって双方向に伝達される. 全変数ノードが全変数ノードからメッセージを受け取ったことになり, グラフ上の任意の変数に対する周辺分布が計算できる.