

確率伝搬法の情報幾何

池田思朗 (九工大 & 科技団)

田中利幸 (都立大)

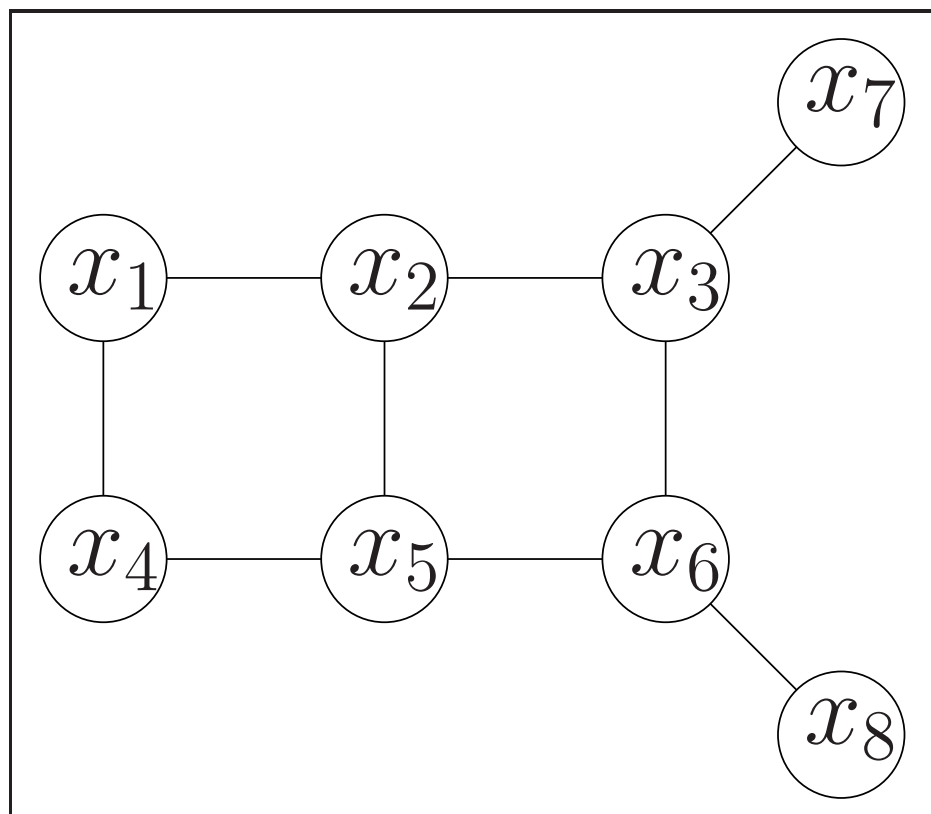
甘利俊一 (理研)

2002 年 9 月 2 日

確率伝搬法

- 各ノードの周辺分布を計算するための手法
- 木構造のグラフに対しては正確な値を与える．ループのあるグラフに対しては近似解．
- ノードの数が多いときには現実的な方法
- アルゴリズムの収束性，安定性など基本的な性質が完全には解明されていない

ベイジアンネットの例



ノード

$$x_i = \{x_1, \dots, x_8\}$$

リンク

$$\mathcal{L}' = \{(1, 2), \dots, (6, 8)\}$$

確率伝搬法の関連分野

- 統計物理
 - グラフ構造は与えられる．全ノードの平均，分散に興味がある．グラフの規模は大きい ($m \rightarrow \infty$) ．
- 情報理論
 - 低密度パリティ検査符号，誤り訂正符号ターボ符号．グラフは符号の設計と通信路から与えられる．各ビットの期待値に興味がある．グラフの規模は大きい．

確率伝搬法の関連分野

- 人工知能
 - － 場合によってはグラフ構造も推定する．各ノードの分布に興味がある．グラフの大きさは必ずしも大きくない．
- 情報幾何
 - － 本研究の目的: 解析のための数理的枠組を与える

ベイジアンネットの表現

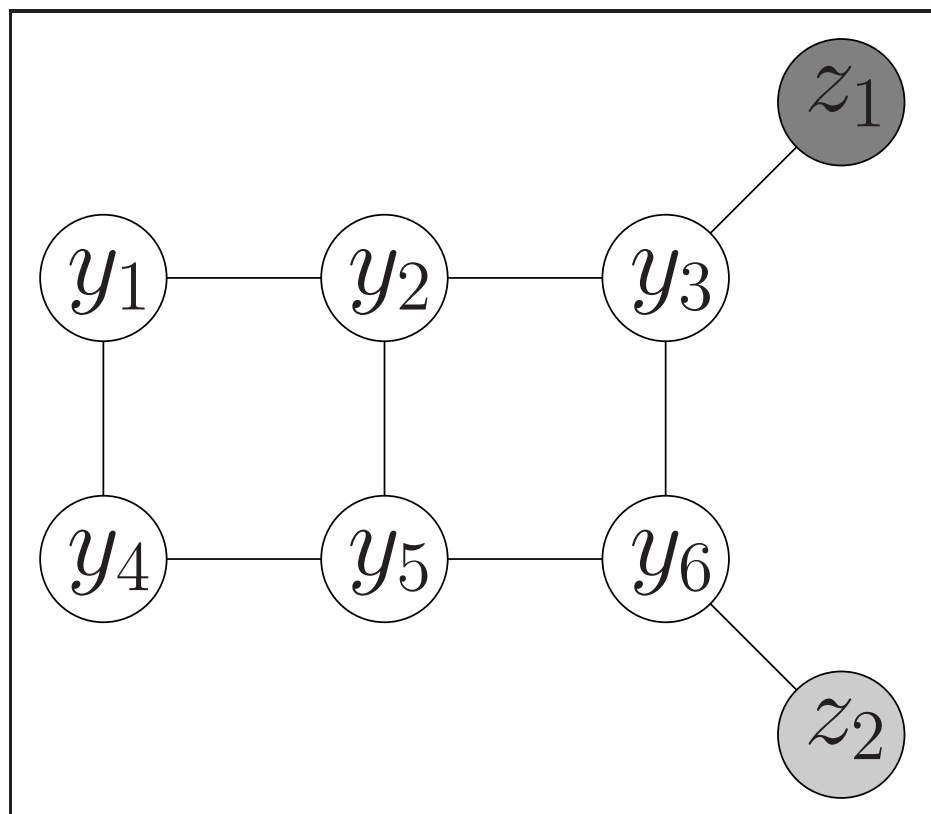
確率変数 $\boldsymbol{x} = (x_1, \dots, x_n)^T$ に対する同時確率密度関数は

$$q(\boldsymbol{x}) = \frac{1}{Z} \prod_{(i,j) \in \mathcal{L}} \psi_{ij}(x_i, x_j)$$

$\psi_{ij}(x_i, x_j) > 0$, Z は規格化定数

$$Z = \sum_{\boldsymbol{x}} \left(\prod_{(i,j) \in \mathcal{L}} \psi_{ij}(x_i, x_j) \right)$$

いくつかのノードが観測された場合



非観測ノード

$$y_i = \{y_1, \dots, y_6\}$$

観測ノード

$$z_i = \{z_1, z_2\}$$

リンク

$$\mathcal{L} = \{(1, 2), \dots, (5, 6)\}$$

いくつかのノードが観測された場合

確率変数 $\mathbf{y} = (y_1, \dots, y_m)^T$ に対する同時確率密度関数は

$$q(\mathbf{y}) \stackrel{\text{def}}{=} q(\mathbf{y}|\mathbf{z}) = \frac{1}{Z} \underbrace{\prod_{i=1}^m \varpi_i(y_i)}_{\text{観測ノード}} \underbrace{\prod_{(i,j) \in \mathcal{L}} \psi_{ij}(y_i, y_j)}_{\text{非観測ノード}}$$

$$\varpi_i(x_i) \stackrel{\text{def}}{=} \prod_{(i,j) \in \mathcal{L}'} \psi_{ij}(y_i, z_{j-m})$$

ビリーフ

- 真のビリーフとは $q(y_i)$ のことである .

$$q(y_i) = \sum_{y_j: j \neq i} q(\mathbf{y})$$

特に $y_i \in \{-1, +1\}$ の2値である場合 , 真のビリーフは y の期待値を求めることと等しい . η と定義すると

$$\eta = \sum_{\mathbf{y}} \mathbf{y} q(\mathbf{y})$$

- 推定されたビリーフは $b(y_i)$ と書くことにする .

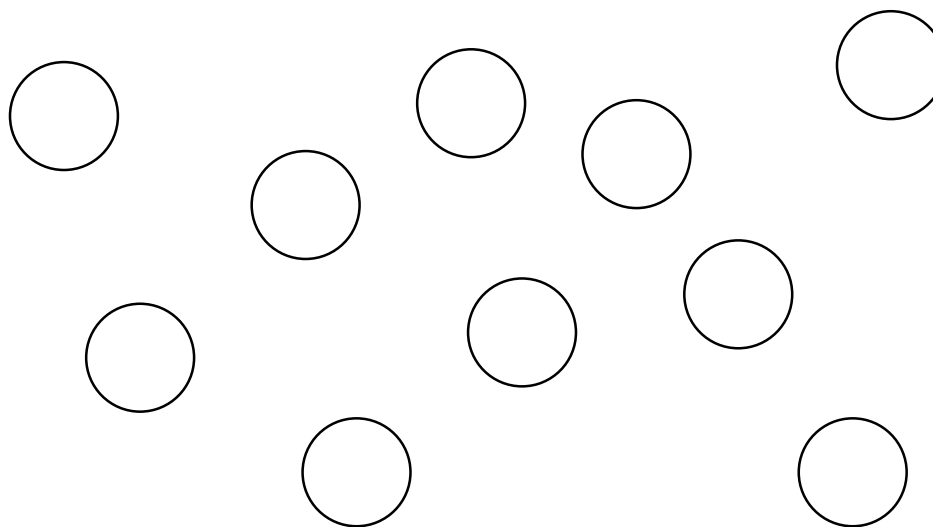
ビリーフ

ビリーフの計算は実用上重要である．一部のデータが観測されたという条件付きで，他のノードの確率分布が分かれば，その結果を用いて応用ができる．

- ユーザーモデル
- ロボット
- 誤り訂正符号

真のビリーフを計算するための計算量は **グラフの構造** によって異なる．

ビリーフの計算が容易なグラフ

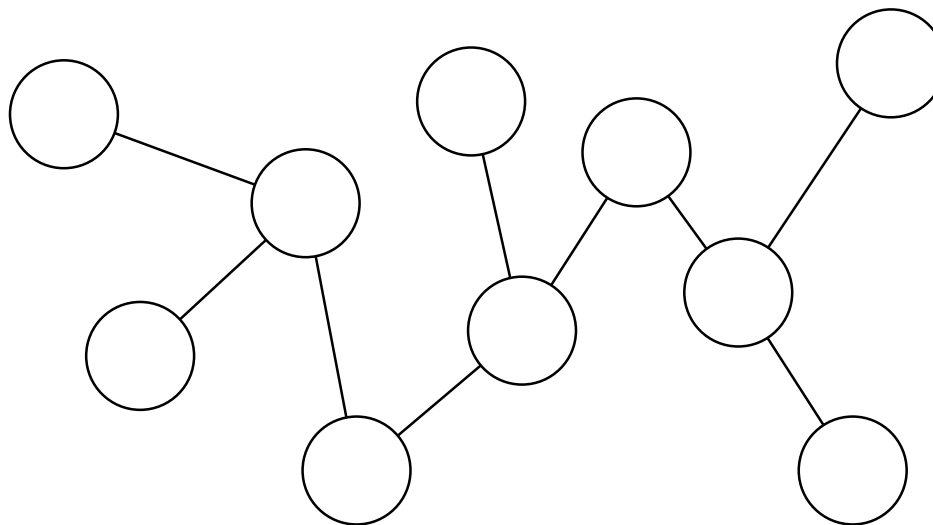


この場合は同時分布 $q(\mathbf{y})$ の形が

$$q(\mathbf{y}) = q(y_1) \cdot q(y_2) \cdots q(y_m)$$

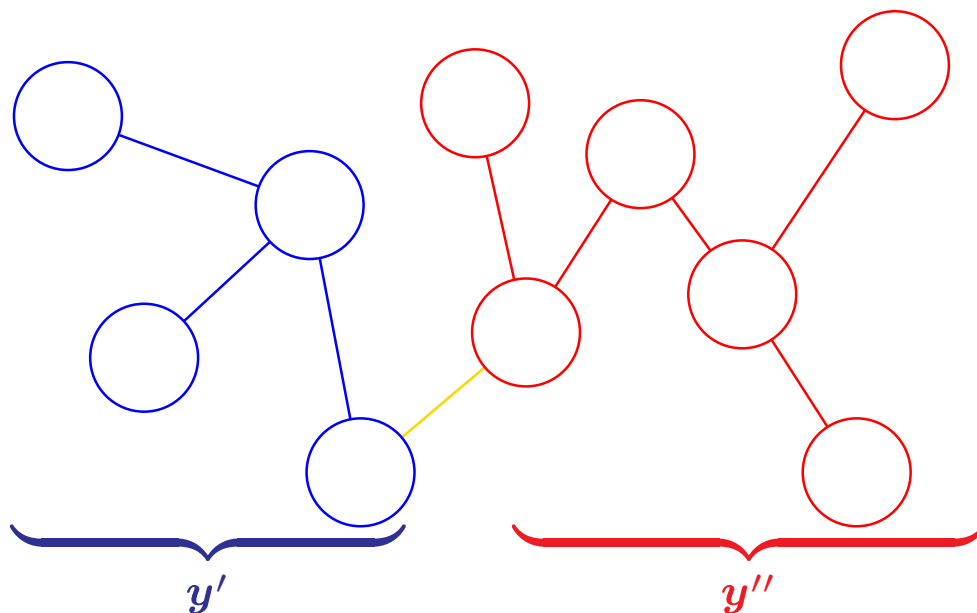
となるため、各ノードの計算のみでよい。

ビリーフの計算が比較的容易なグラフ



このようにグラフが“木”の構造をしている場合には効率の良い計算方法が存在する．これもやはり $q(y)$ の形から理解することができる．

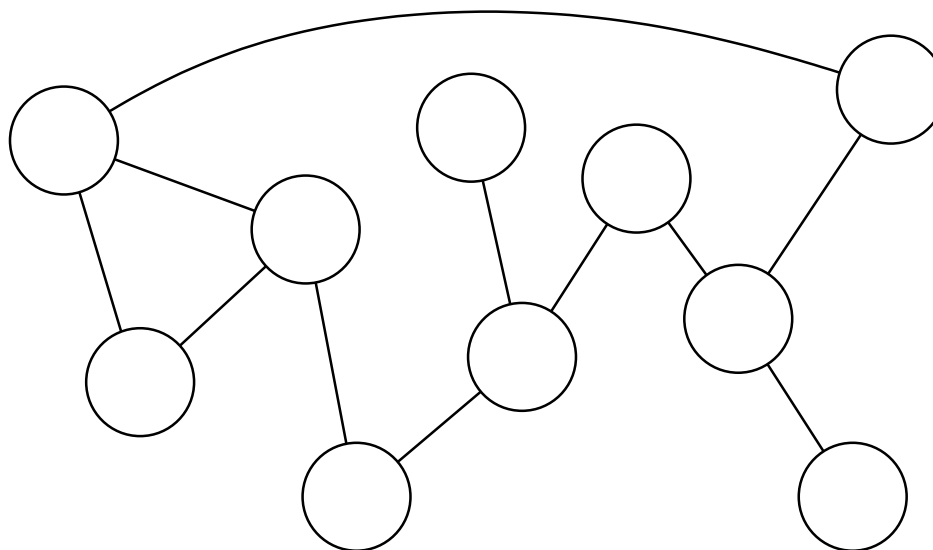
ビリーフの計算が比較的容易なグラフ



同時分布は以下のようなになるため，端から順に計算していけば，比較的容易に計算できる．

$$q(\mathbf{y}) = \frac{1}{Z} q(\mathbf{y}') \psi_{ij}(y'_i, y''_j) q(\mathbf{y}'')$$

ビリーフの計算が困難なグラフ



このようにループがあると，端から計算していても計算量を減らすことができない．

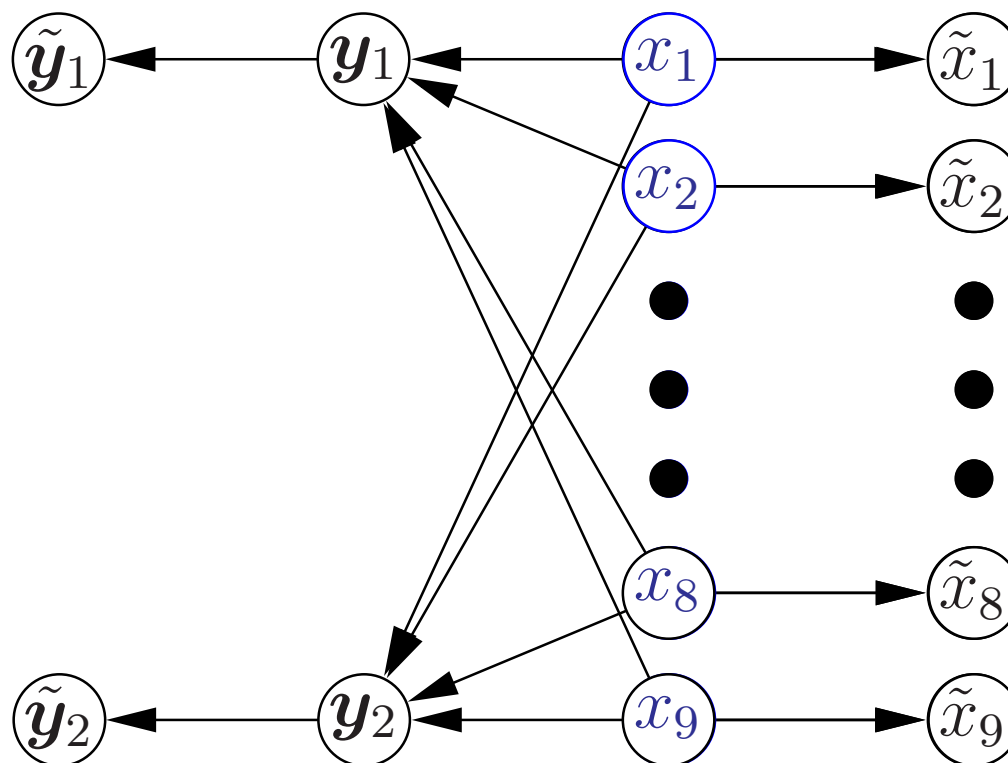
ビリーフの計算

NP-hard

仮りに $y_i \in \{-1, +1\}$ とすると η を求めるためには, 2^m の項の和を計算する必要がある. m の増加とともに計算量が指数的に増加し, 現実的には計算できなくなる.

ノードの数が 10 個程度なら 2^m 項の和も可能だが, 例えば誤り訂正符号などの場合には非観測ノードの数は数百である.

ベイジアンネットによるターボ符号の表現

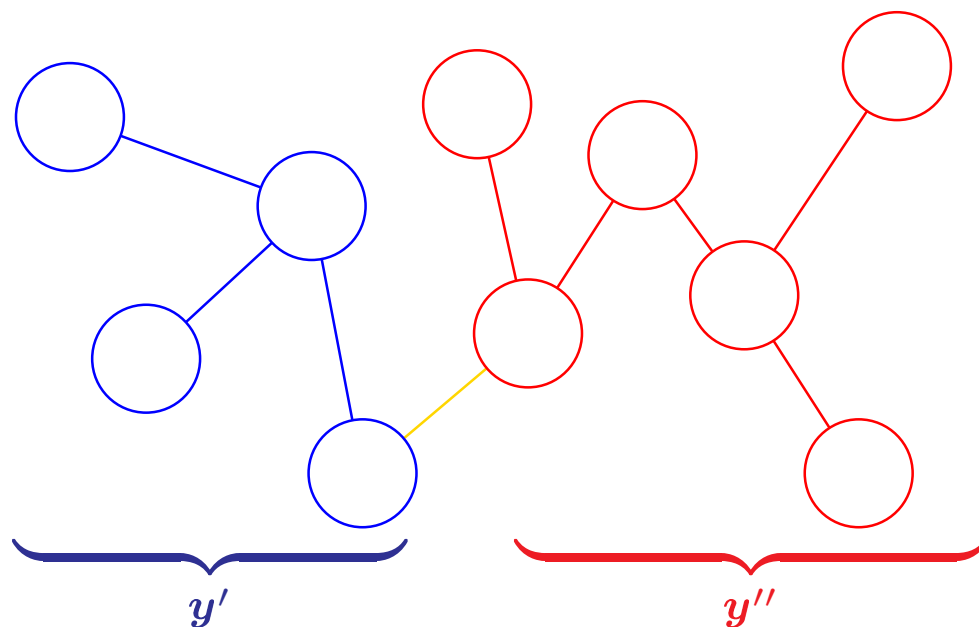


確率伝搬法の登場

確率伝搬法 とはビリーフを求める計算を効率行なうための手法である．木の構造をしているネットワークに対しては正しいビリーフが計算できる．

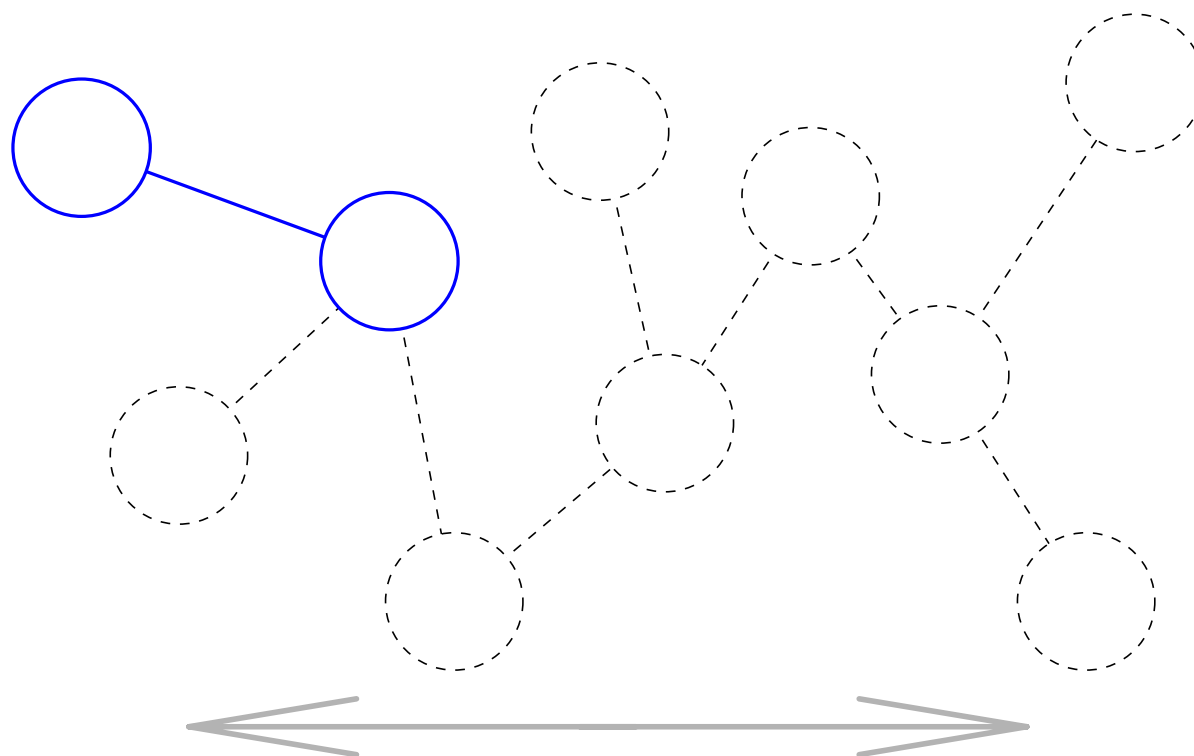
Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo. CA: Morgan Kaufmann, 1988.

確率伝搬法

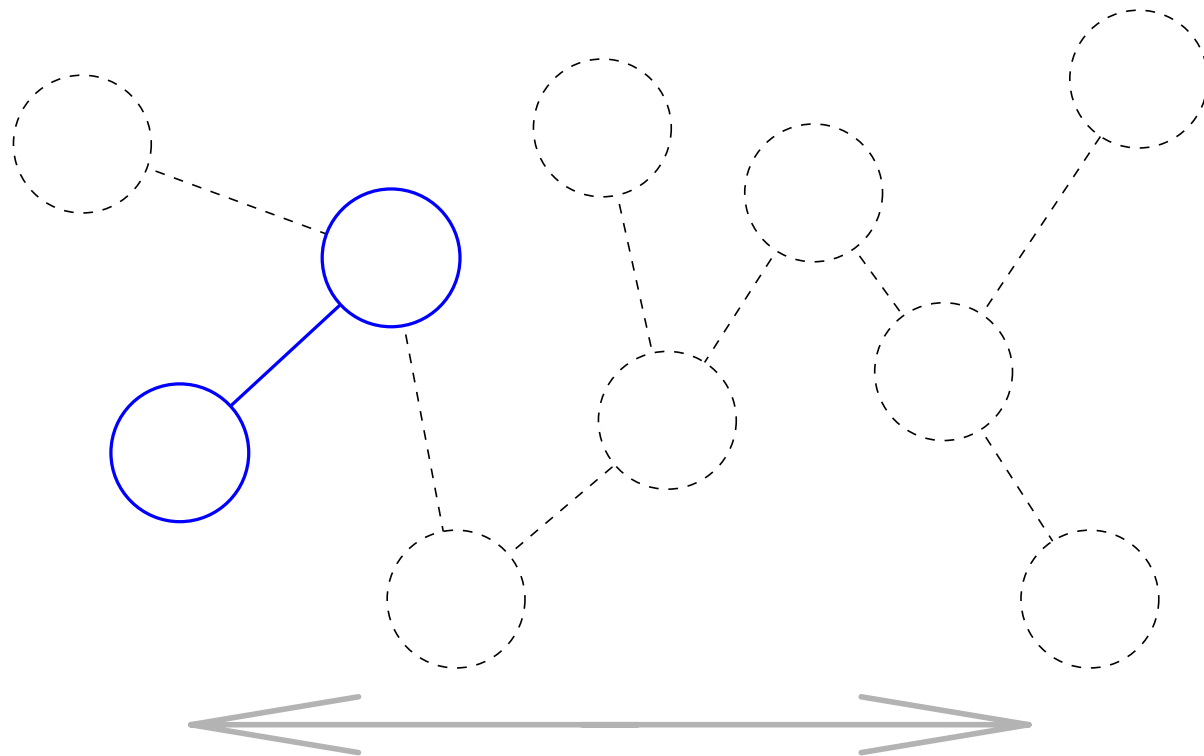


木のグラフでは端から順に計算していけば，比較的容易に周辺分布が計算できる．確率伝搬法はこの考え方をシステムティックに表現したもの

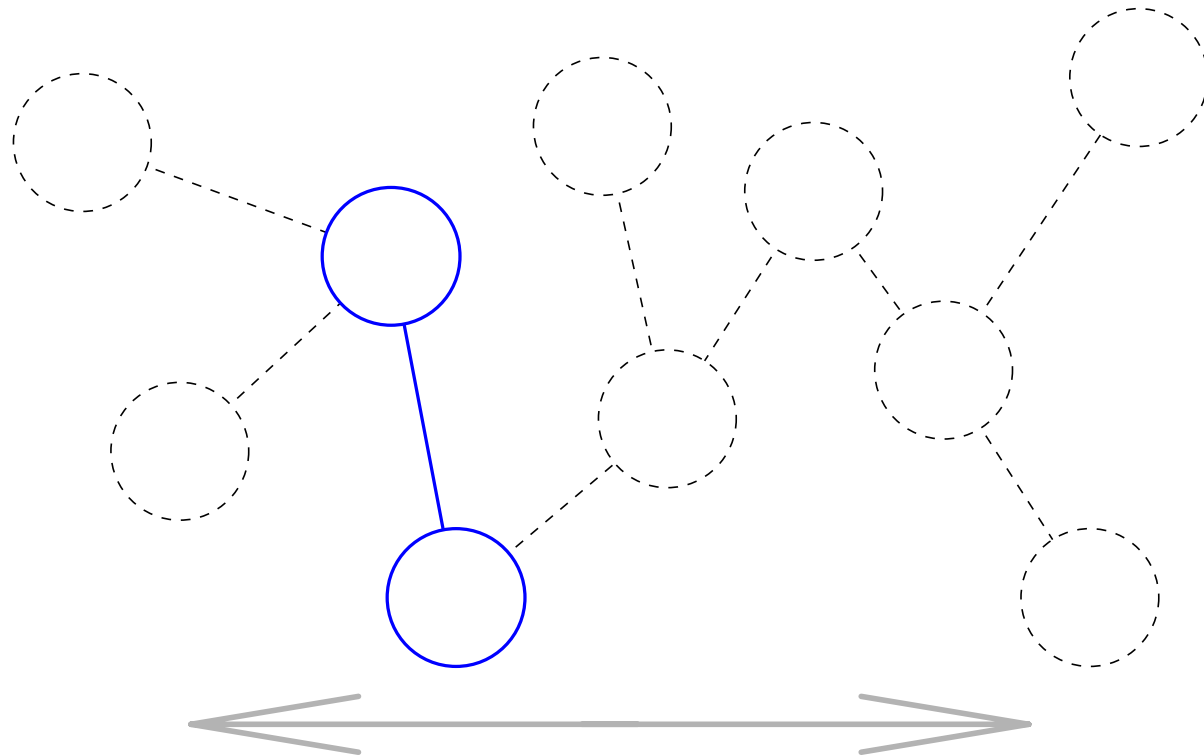
確率伝搬法



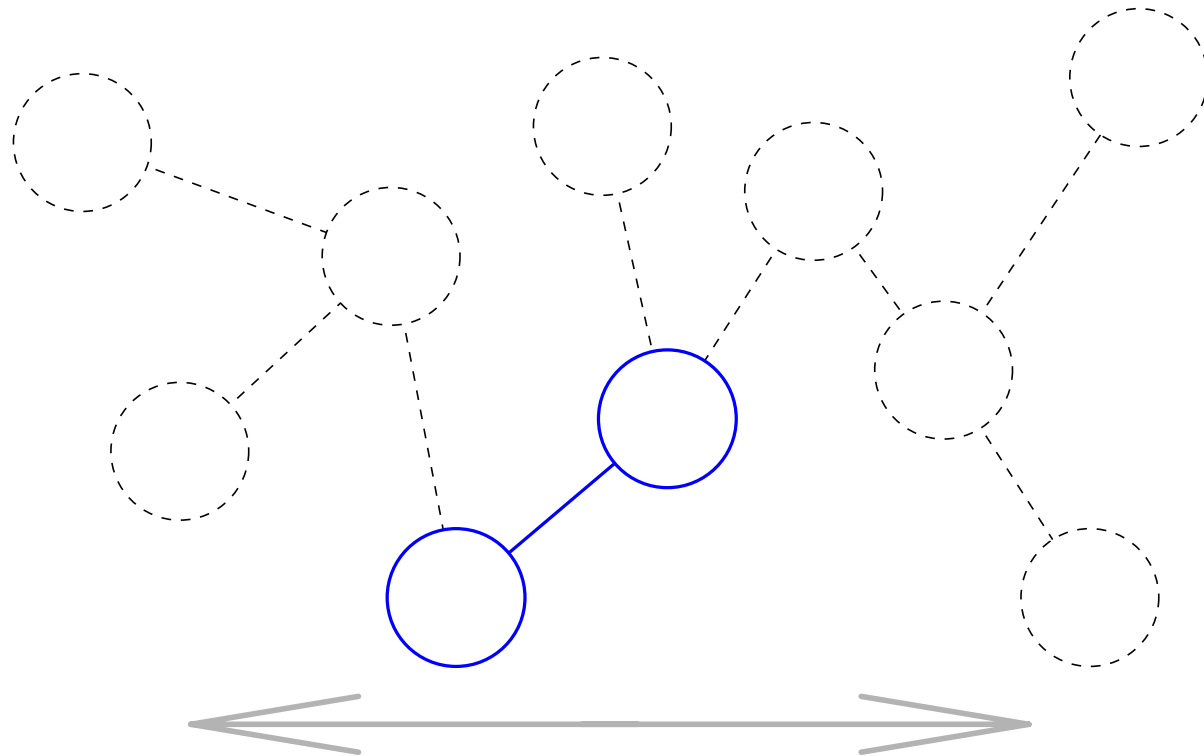
確率伝搬法



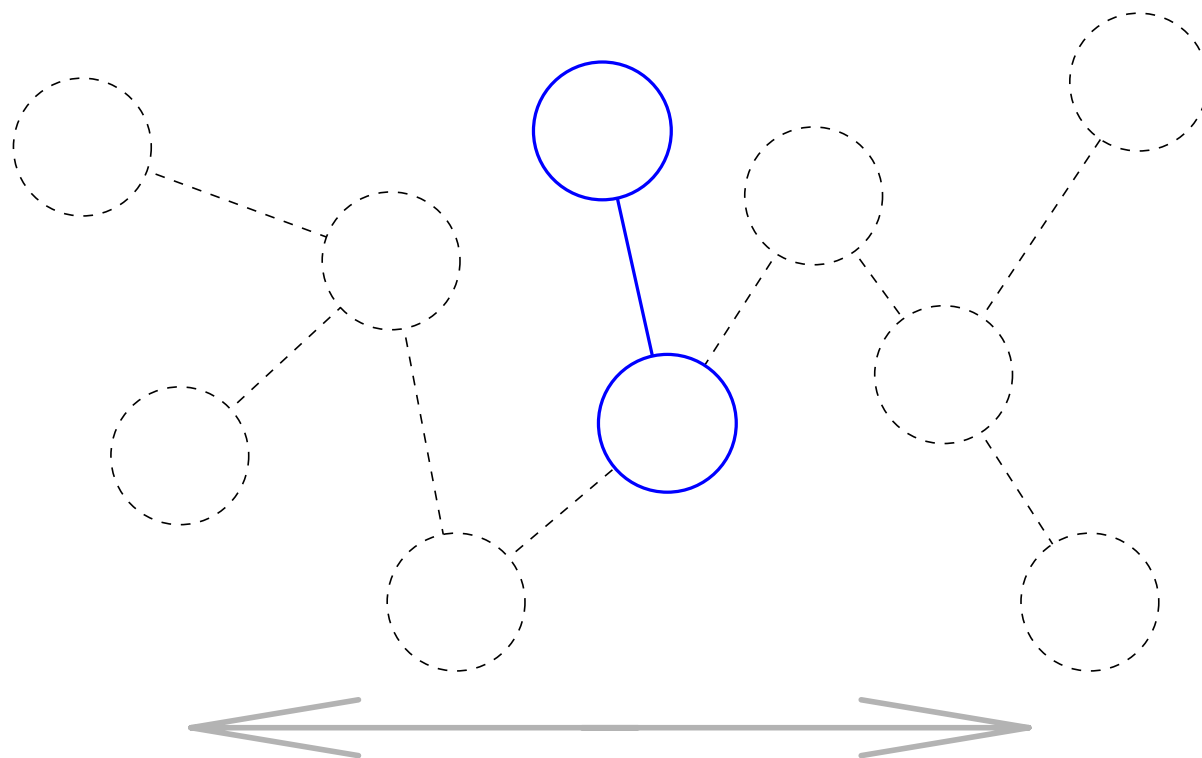
確率伝搬法



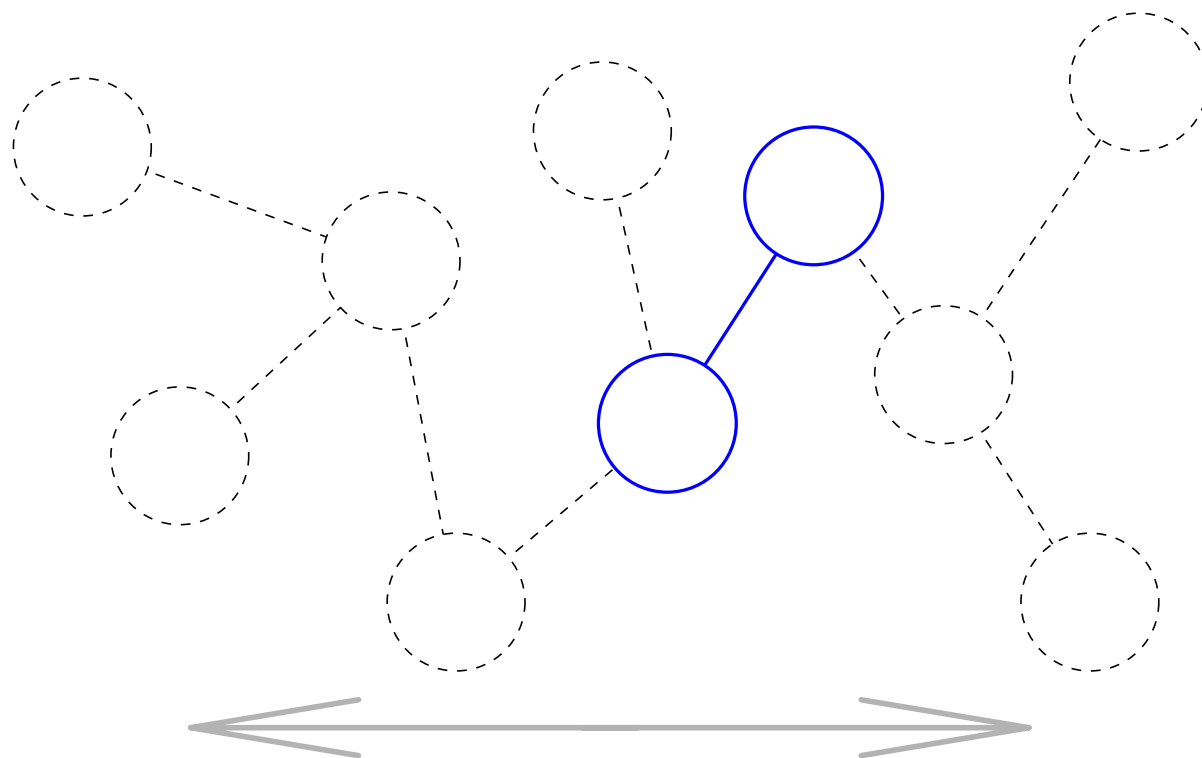
確率伝搬法



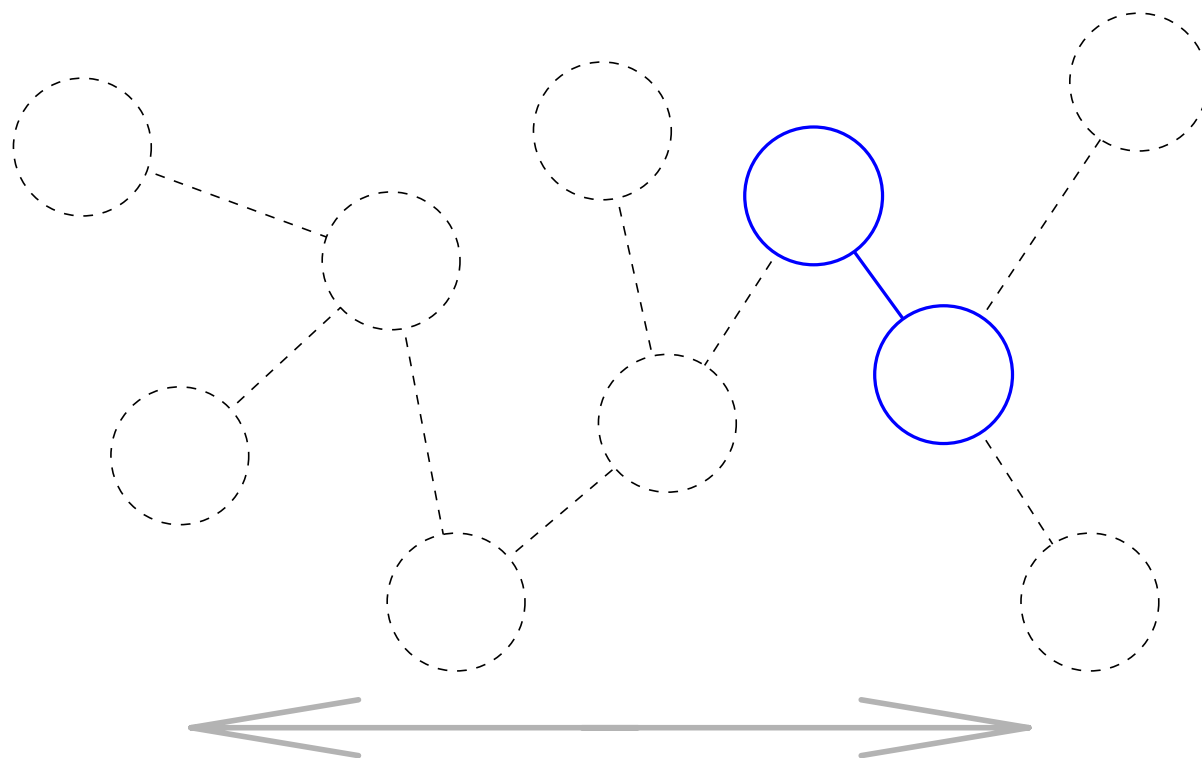
確率伝搬法



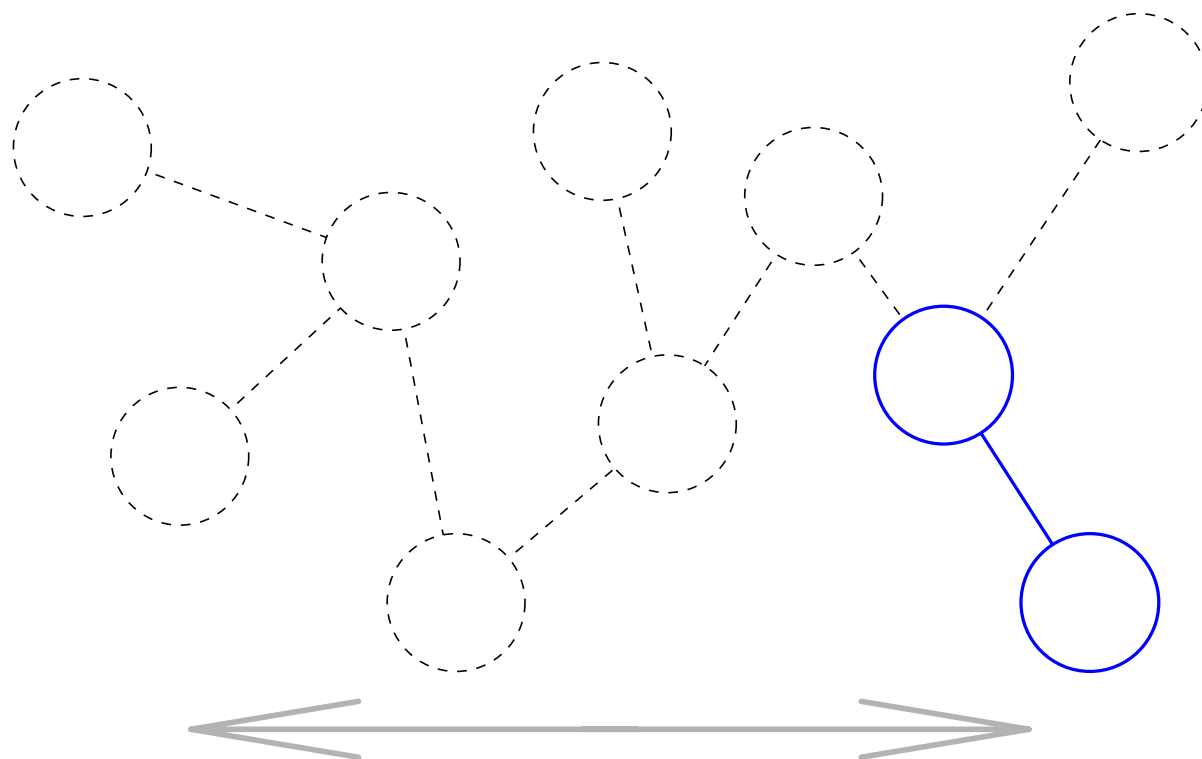
確率伝搬法



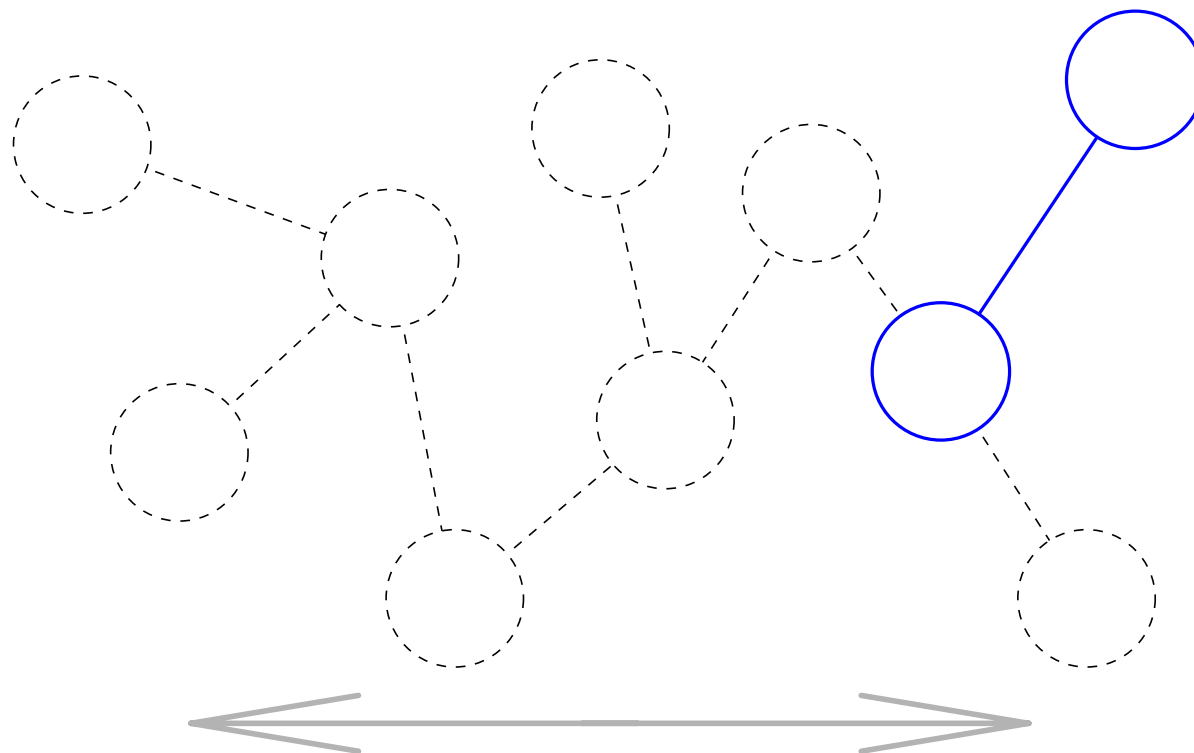
確率伝搬法



確率伝搬法



確率伝搬法

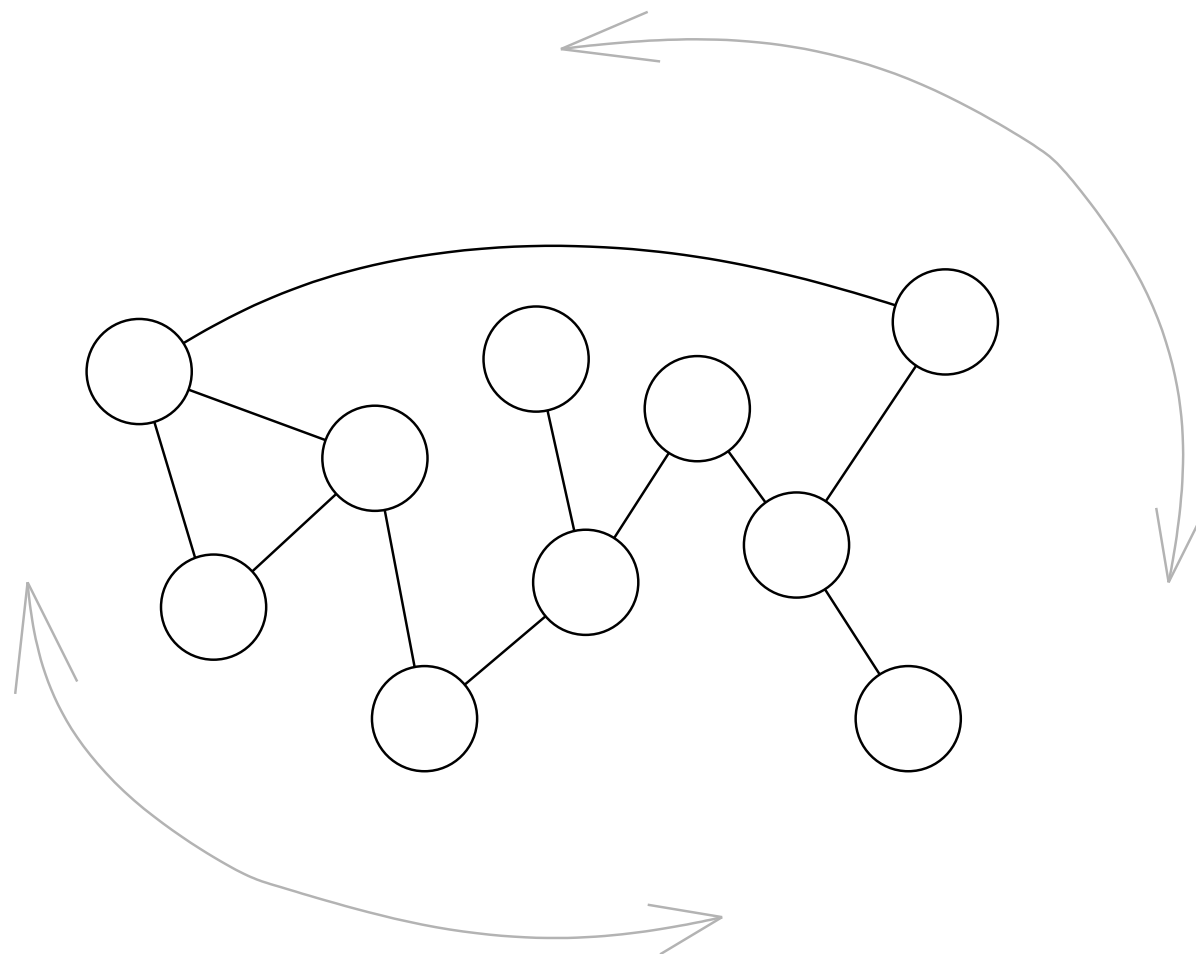


確率伝搬法

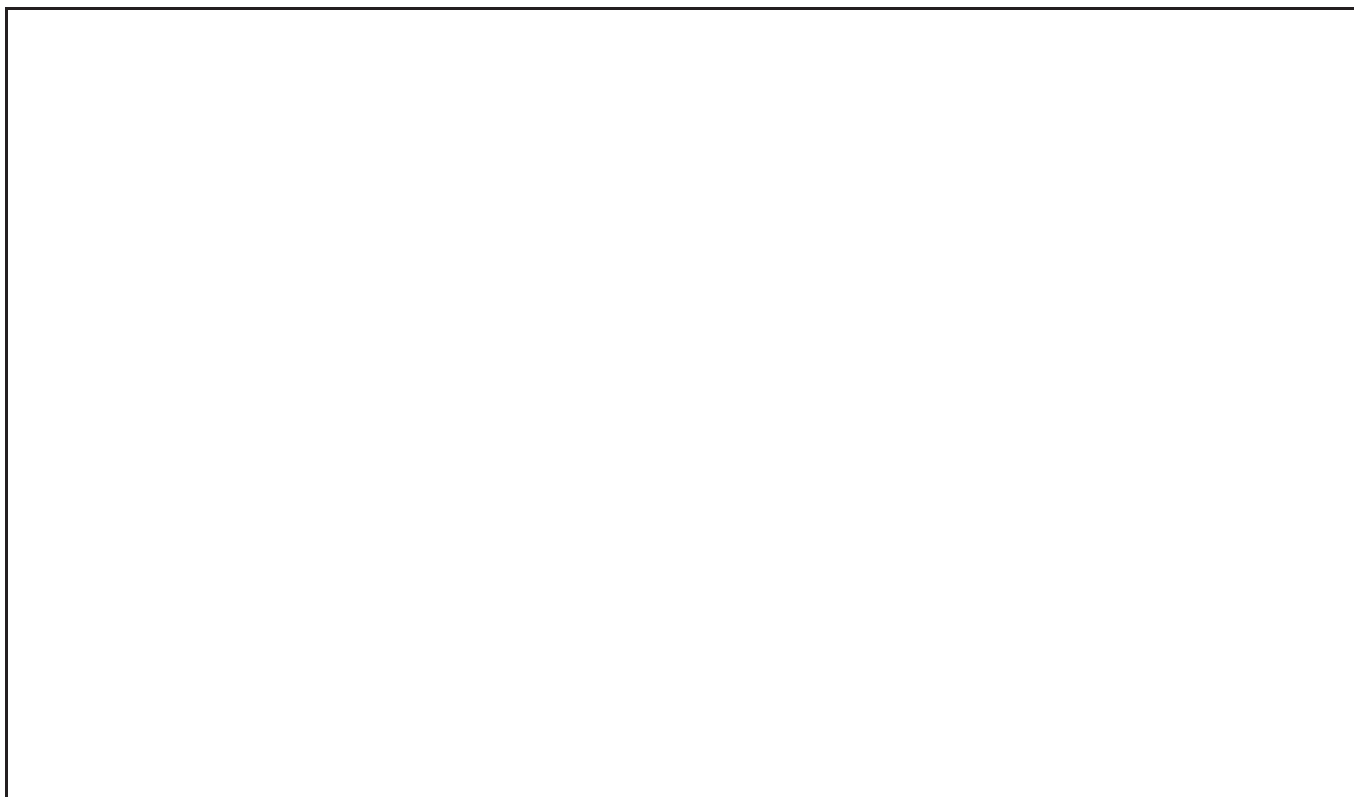
確率伝搬法はループのあるグラフに対しても用いられる．しかし，

- 繰り返し演算となる
- 近似解が得られる

確率伝搬法



確率伝搬法



確率伝搬法

しかし、それでも 現実的には有効

- 低密度パリティ検査符号，ターボ符号では高性能な誤り訂正能力を発揮

問題点

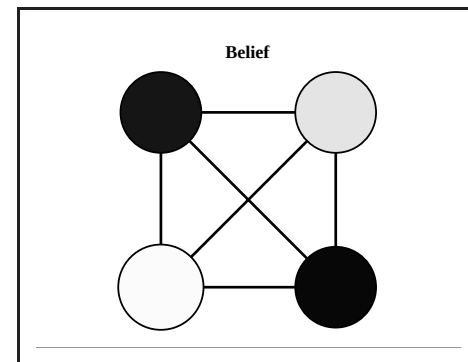
- 収束性
- 近似精度

数理的な解析は困難

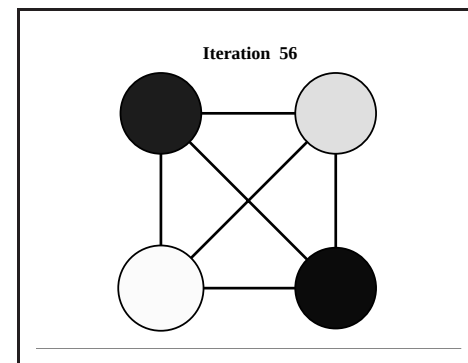
確率伝搬法の例:4 ビットモデル



真のビリーフ

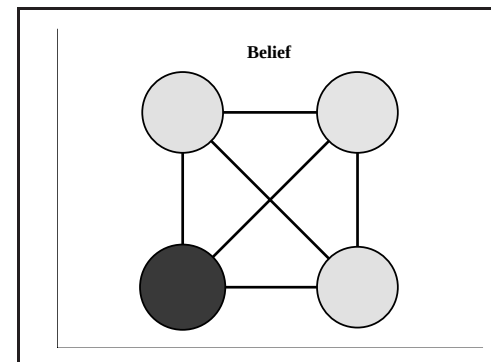


確率伝搬法の結果



確率伝搬法の例:4 ビットモデル

真のビリーフ



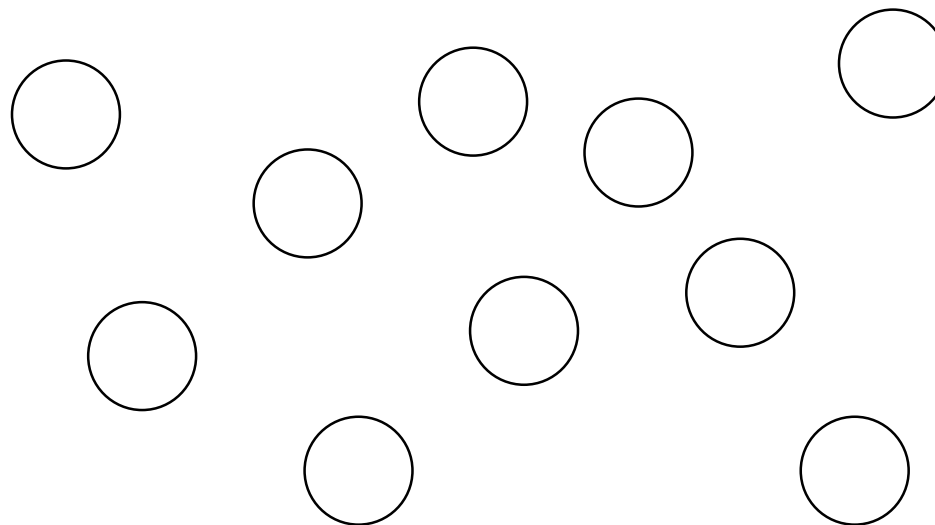
真の分布の表現

$q(\mathbf{y})$ の表現を再考する .

$$\begin{aligned} q(\mathbf{y}) &= \frac{1}{Z} \prod_{i=1}^m \varpi_i(y_i) \prod_{(i,j) \in \mathcal{L}} \psi_{ij}(y_i, y_j) \\ &= \exp \left(k_0(\mathbf{y}) + \sum_{(i,j) \in \mathcal{L}} \ln \psi_{ij}(y_i, y_j) - C \right) \\ &= \exp \left(k_0(\mathbf{y}) + \sum_{r=1}^L c_r(\mathbf{y}) - C \right) \end{aligned}$$

$c_r(\mathbf{y})$ は $\ln \psi_{ij}(y_i, y_j)$ に対応している . また , $k_0(\mathbf{y})$ は y_i に対して線形な関数である .

ビリーフの表現



一方ビリーフの表現は次のようになる．これは上の図のグラフで表現される分布である．

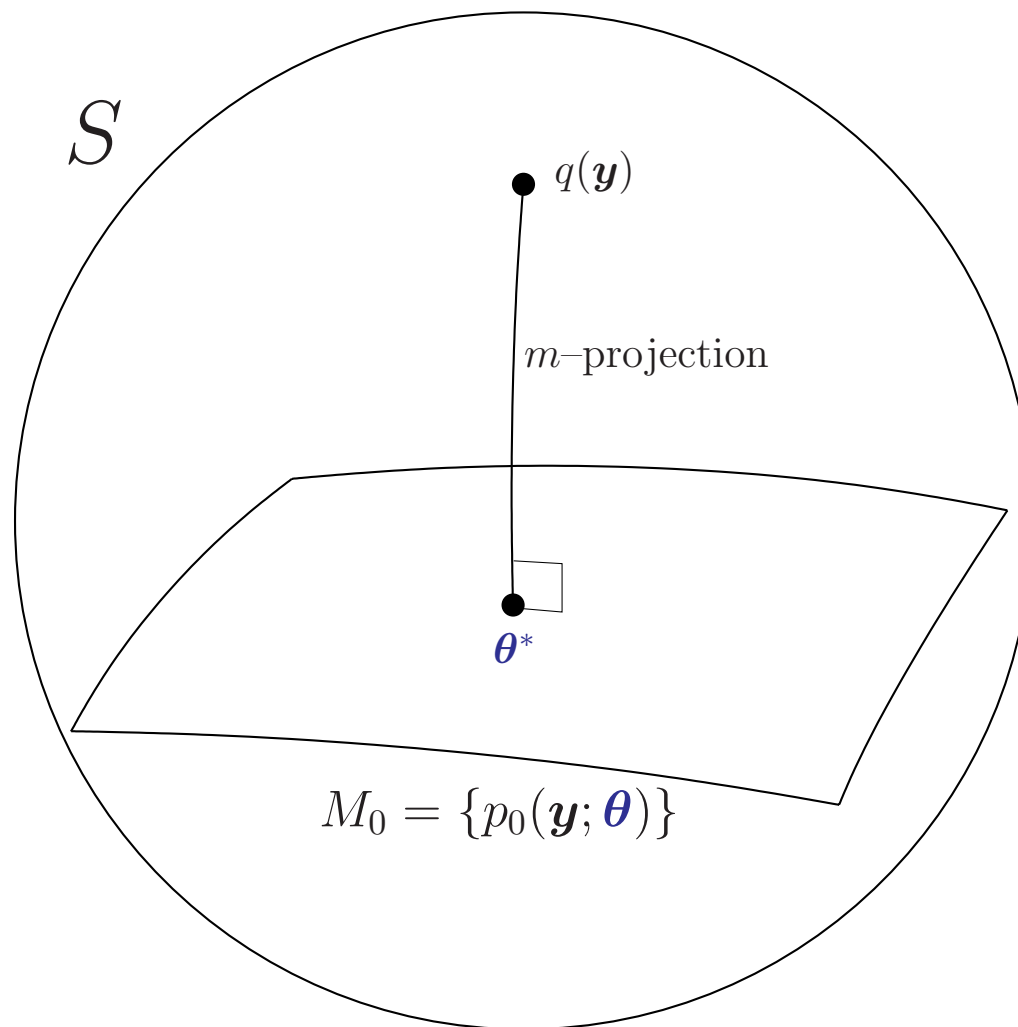
$$p_0(\mathbf{y}; \boldsymbol{\theta}) = \exp\left(k_0(\mathbf{y}) + \boldsymbol{\theta} \cdot \mathbf{y} - \varphi(\boldsymbol{\theta})\right)$$

真のビリーフ

真のビリーフは次の式で定義される θ^* を用いて, $p_0(\mathbf{y}; \theta^*)$ とかける .

$$\begin{aligned}\theta^* &\stackrel{\text{def}}{=} \operatorname{argmin}_{\theta} KL(q(\mathbf{y}); p_0(\mathbf{y}; \theta)) \\ &= \operatorname{argmin}_{\theta} \sum_{\mathbf{y}} q(\mathbf{y}) \log \frac{q(\mathbf{y})}{p_0(\mathbf{y}; \theta)}\end{aligned}$$

真のビリーフ:情報幾何的な表現



確率伝搬法の近似

確率伝搬法では $q(\mathbf{y})$ を $p_0(\mathbf{x}; \boldsymbol{\theta})$ によって次のように近似する .

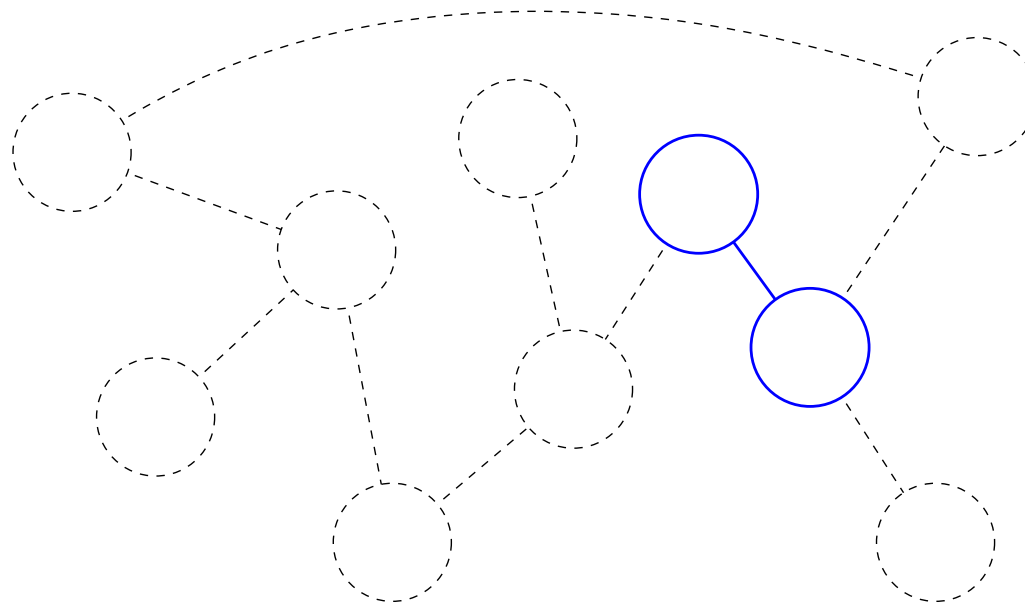
$$q(\mathbf{y}) = \exp(k_0(\mathbf{y}) + c_1(\mathbf{y}) + c_2(\mathbf{y}) + \cdots + c_L(\mathbf{y}) - C)$$

$$p_0(\mathbf{x}; \boldsymbol{\theta}) = \exp(k_0(\mathbf{y}) + \boldsymbol{\xi}_1 \cdot \mathbf{y} + \boldsymbol{\xi}_2 \cdot \mathbf{y} + \cdots + \boldsymbol{\xi}_L \cdot \mathbf{y} - \varphi_0(\boldsymbol{\theta}))$$

$$\text{ただし } \boldsymbol{\theta} = \boldsymbol{\xi}_1 + \cdots + \boldsymbol{\xi}_L.$$

$c_r(\mathbf{y}) \rightarrow \boldsymbol{\xi}_r \cdot \mathbf{y}$ と表現する . 実際には $c_r(\mathbf{y})$ には y_i と y_j し
か含まれていないことから , $c_r(y_i, y_j) \rightarrow \boldsymbol{\xi}_r^i y_i + \boldsymbol{\xi}_r^j y_j$ と表現
する .

確率伝搬法の近似



グラフィカルモデルでは, $q(\mathbf{y})$ の周辺分布を直接求めることはできない. しかし, 任意の $c_r(\mathbf{y})$ を一つ含むモデルの周辺分布は求めることができる.

リンクをひとつ含む確率分布

リンクをひとつ含む確率分布は

$$p_r(\boldsymbol{x}; \zeta_r) = \exp(k_0(\boldsymbol{y}) + c_r(\boldsymbol{y}) + \zeta_r \cdot \boldsymbol{y} - \varphi_r(\zeta_r))$$

となる．確率伝搬法では ζ_r , ξ_r , θ の間に次の関係を仮定する．

$$\theta = \zeta_r + \xi_r, \quad \theta = \frac{1}{L-1} \sum_r \zeta_r, \quad \theta = \sum_r \xi_r$$

重要な確率分布

$$q(\mathbf{y}) = \exp(k_0(\mathbf{y}) + c_1(\mathbf{y}) + c_2(\mathbf{y}) + \cdots + c_L(\mathbf{y}) - C)$$

$$p_0(\mathbf{x}; \boldsymbol{\theta}) = \exp(k_0(\mathbf{y}) + \boldsymbol{\xi}_1 \cdot \mathbf{y} + \boldsymbol{\xi}_2 \cdot \mathbf{y} + \cdots + \boldsymbol{\xi}_L \cdot \mathbf{y} - \varphi_0(\boldsymbol{\theta}))$$

$$p_1(\mathbf{x}; \boldsymbol{\zeta}_1) = \exp(k_0(\mathbf{y}) + c_1(\mathbf{y}) + \boldsymbol{\xi}_2 \cdot \mathbf{y} + \cdots + \boldsymbol{\xi}_L \cdot \mathbf{y} - \varphi_1(\boldsymbol{\zeta}_1))$$

$$p_2(\mathbf{x}; \boldsymbol{\zeta}_2) = \exp(k_0(\mathbf{y}) + \boldsymbol{\xi}_1 \cdot \mathbf{y} + c_2(\mathbf{y}) + \cdots + \boldsymbol{\xi}_L \cdot \mathbf{y} - \varphi_2(\boldsymbol{\zeta}_2))$$

$$\vdots$$

$$p_L(\mathbf{x}; \boldsymbol{\zeta}_L) = \exp(k_0(\mathbf{y}) + \boldsymbol{\xi}_1 \cdot \mathbf{y} + \boldsymbol{\xi}_2 \cdot \mathbf{y} + \cdots + c_L(\mathbf{y}) - \varphi_L(\boldsymbol{\zeta}_L))$$

確率伝搬法の情報幾何的表現

まず, 任意の確率分布 $r(\mathbf{y})$ を周辺化する手続きを表わす記号を定義する

$$\pi \circ r(\mathbf{y}) \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\operatorname{argmin}} D[r(\mathbf{y}); p_0(\mathbf{y}; \boldsymbol{\theta})].$$

定義より $\boldsymbol{\theta}^* = \pi \circ q(\mathbf{y})$ であるが, これは計算できない. 一方 $\pi \circ p_r(\mathbf{y}; \boldsymbol{\zeta}_r)$ は計算可能である.

確率伝搬法の情報幾何的表現

1. $t = 0$ において, $\theta^t = \zeta_r^t = \mathbf{o}$, $r = 1, \dots, L$ とする .
2. $t = 1, 2, 3, \dots$ に対して, 以下の手続きを行なう .
 - (a) ξ_r^{t+1} , $r = 1, \dots, L$ を次のように更新する .

$$\xi_r^{t+1} = \pi \circ p_r(\mathbf{y}; \zeta_r^t) - \zeta_r^t.$$

- (b) θ , ζ_r , $r = 1, \dots, L$ を次のように更新する .

$$\theta^{t+1} = \sum_{r=1}^L \xi_r^{t+1}, \quad \zeta_r^{t+1} = \theta^{t+1} - \xi_r^{t+1}.$$

3. ξ_r^{t+1} が収束したならば終了する .

収束点の満たす条件

収束点では ξ_r^* , ζ_r^* , θ^* の間に次の関係がある .

$$1) \quad \pi \circ p_r(\mathbf{y}; \zeta_r^*) = \theta^*, \quad r = 1, \dots, L$$

$$2) \quad \theta^* = \sum_r \xi_r^* = \frac{1}{L-1} \sum_r \zeta_r^*.$$

1) は次のように書き直せる

$$\sum_{\mathbf{y}} \mathbf{y} p_r(\mathbf{y}; \zeta_r^*) = \sum_{\mathbf{y}} \mathbf{y} p_0(\mathbf{y}; \theta^*), \quad r = 1, \dots, L$$

$$\eta_r(\zeta_r^*) = \eta_0(\theta^*), \quad r = 1, \dots, L$$

多様体の定義

M_0 は $p_0(\mathbf{y}; \boldsymbol{\theta})$ の集合 , M_r は $p_r(\mathbf{y}; \boldsymbol{\zeta}_r)$ の集合であり , 次のように定義する .

$$M_0 = \left\{ p_0(\mathbf{y}; \boldsymbol{\theta}) = \exp(k_0(\mathbf{y}) + \boldsymbol{\theta} \cdot \mathbf{y} - \varphi_0(\boldsymbol{\theta})) \mid \boldsymbol{\theta} \in \mathcal{R}^m \right\}$$

$$M_r = \left\{ p_r(\mathbf{y}; \boldsymbol{\zeta}_r) = \exp(k_0(\mathbf{y}) + c_r(\mathbf{y}) + \boldsymbol{\zeta}_r \cdot \mathbf{y} - \varphi_r(\boldsymbol{\zeta}_r)) \mid \boldsymbol{\zeta}_r \in \mathcal{R}^m \right\} .$$

収束点の満たす条件: 等平均多様体 $M(\boldsymbol{\theta})$

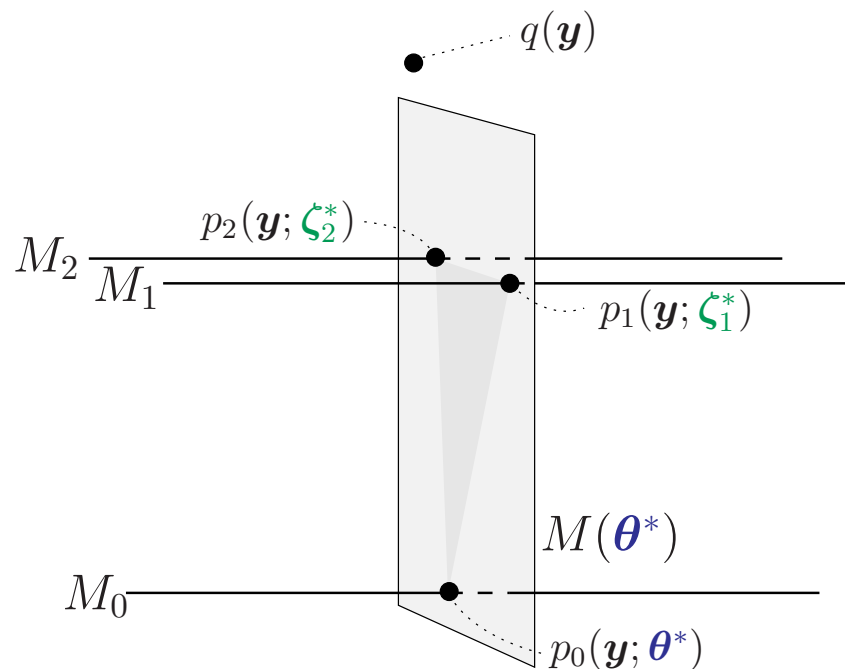


Fig 1: $M(\boldsymbol{\theta}) = \left\{ p(\mathbf{y}) \mid \sum_{\mathbf{y}} \mathbf{y} p(\mathbf{y}) = \sum_{\mathbf{y}} \mathbf{y} p_0(\mathbf{y}; \boldsymbol{\theta}) \right\}$

収束点の満たす条件: 対数線形多様体 $E(\boldsymbol{\theta}^*)$

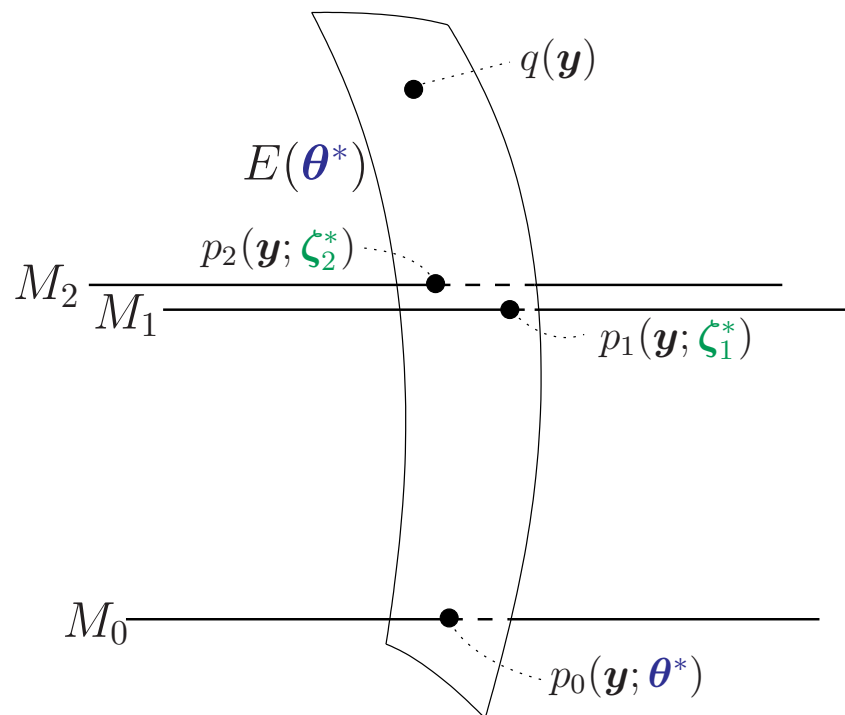
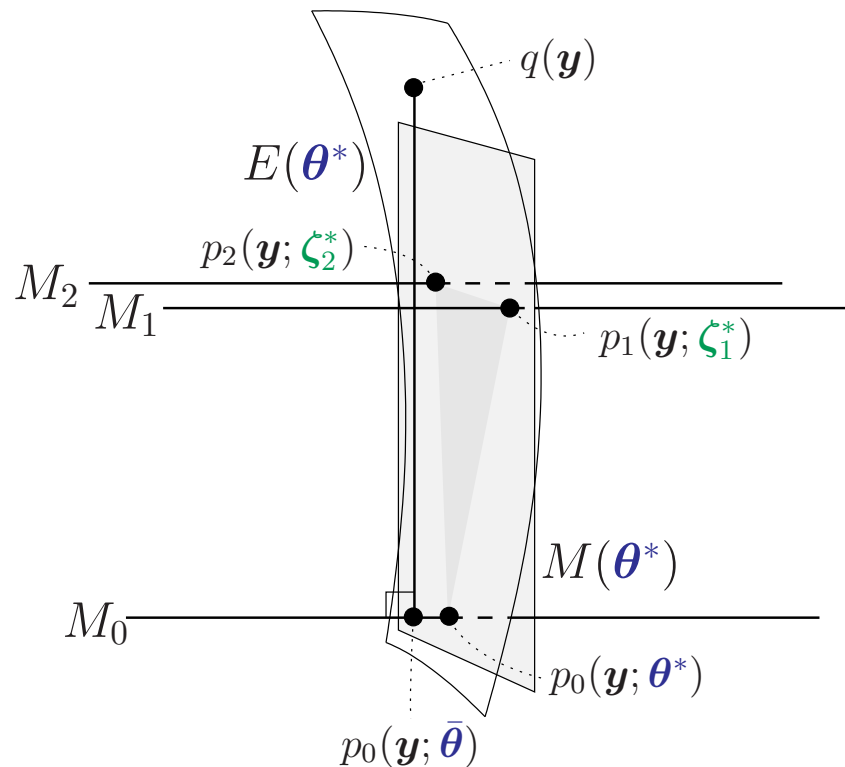


Fig 2: $E(\boldsymbol{\theta}) = \left\{ p = C p_0^{t_0} p_1^{t_1} p_2^{t_2} \mid \sum_{r=0}^2 t_r = 1 \right\}$

収束点の持つ性質



$$p_0(x; \theta^*), p_r(y; \zeta_r^*) \in M(\theta^*)$$

$$p_0(y; \theta^*), p_r(y; \zeta_r^*), q(y) \in E(\theta^*)$$

真の解からのずれ

真の分布 $q(\mathbf{y})$ による平均 $\bar{\eta}$ と確率伝搬法による解 $\eta(\boldsymbol{\theta}^*)$ の差は以下のように近似できる .

$$\bar{\eta} \simeq \eta(\boldsymbol{\theta}^*) + \frac{1}{2} \sum_{r \neq s} B_{rs} \eta(\boldsymbol{\theta}^*).$$

ただし ,

$$B_{rs} \stackrel{\text{def}}{=} \left(\partial_r - \sum_k I^{kk} A_{kr} \partial_k \right) \left(\partial_s - \sum_j I^{jj} A_{js} \partial_j \right)$$

$$A_{ir} \stackrel{\text{def}}{=} \partial_r \eta_i(\boldsymbol{\theta}^*, \mathbf{o}) = \text{Cov}[y_i, c_r(\mathbf{y})].$$

収束性を向上する

$$1) \quad \eta_r(\zeta_r^*) = \eta_0(\theta^*),$$

$$2) \quad \theta^* = \sum_r \xi_r^* = \frac{1}{L-1} \sum_r \zeta_r^*.$$

を満たすパラメータを求めるアルゴリズムを考える．例えば，

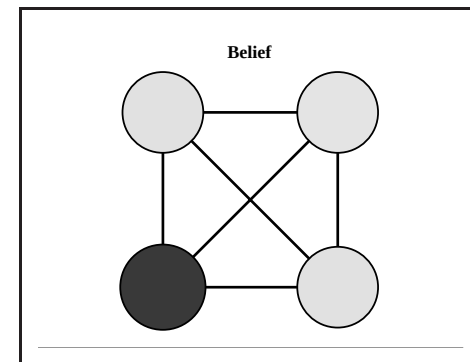
$$E = \sum_r |\eta_r(\zeta_r) - \eta_0(\theta)|^2$$

を 0 とするように最急降下法を考えることができる．一回の勾配を求めるために通常の変換法を 2 回用いれば良い．

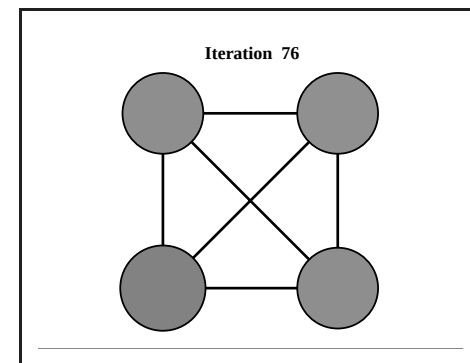
確率伝搬法の例:4 ビットモデル



真のビリーフ



確率伝搬法の結果



まとめ

- 情報幾何によって確率伝搬法を表現した．
- その表現に基づき，アルゴリズムの収束性，誤差についての解析が可能となる．我々も初歩的な結果をここに示した．