

# 平均場近似の数理

## — 確率伝搬法という名の物理と情報の交差点 —

東北大学 大学院情報科学研究科 田中 和之<sup>1</sup>

### 概要

ここ10年ほどの間に確率的情報処理あるいは情報統計力学といわれる研究テーマが物理学・統計学・情報科学の境界領域を舞台として様々の分野の研究者により研究され、徐々に認知されるようになってきている。特に、平均場近似は、大規模確率モデルを用いて定式化されることの多い確率的情報処理システムを絵空事ではなく具体化するキーアプリケーションとして注目を集めつつある。また、平均場近似の拡張版であるベータ近似は確率推論における確率伝搬法と呼ばれる計算手法とその数理構造がほぼ同等であることが明らかになり、物理と情報が思わぬ交差点で出くわし、そこで何かが起こりつつある。

本稿では確率をてなずけ、データの統計的性質をてなずけ、ひいてはそこに内在するゆらぎをてなずける手法のひとつとしての平均場近似とその拡張について画像処理および確率推論を例にとり、ベイズ統計と自由エネルギー最小の変分原理という2つの舞台の上で理論的詳細についてできるだけ式の導出の詳細を省略せずにわかりやすく紹介する。

## 1 はじめに

近年、情報処理に関するさまざまな分野で、問題を確率的な枠組みのもとで取り扱おうという方法論が注目を集めている [1]。情報の処理、伝送に関する基礎理論は情報理論であり、情報理論は確率的な枠組みにもとづいて構成されているわけだから、確率的な枠組みで情報処理を扱うのは本来は自然なことである。しかしながら、必要な計算量が大きくなってしまふなどの問題があったため、確率論的なアプローチはこれまではあまり注目されてはいなかった。ところが、近年になってコンピュータの性能が飛躍的に向上し、情報処理の種々の分野で大規模な確率モデルの取り扱いが実際に検討されるようになり、並行してそれらの検討を支える理論的枠組みの必要性も徐々に意識されるようになってきた。このような背景の中で大規模な確率モデルを取り扱う上でのキーアプリケーションの一つとして平均場近似が注目を集めつつある。

歴史的には平均場近似は統計力学において約100年にもわたる長い歴史を持っているが、確率モデルを扱う処方箋として、物理学以外の分野で似たような手法が全く考えられていなかったわけではない。人工知能の分野で確率推論という研究対象があり、ベイジアンネットワークとも呼ばれている [2]。そこでは確率伝搬法という名の下に平均場近似の拡張版のひとつであるベータ近似とほぼ同じ構造をもつアルゴリズムが考案されている [3, 4, 5]。画像処理の分野ではマルコフ確率場という名の下で古典スピン系をベースにした確率モデルと平均場近似を用いた実用的アルゴリズムが数多く提案されている [6, 7, 8, 9, 10, 11]。符号理論の世界ではランダムスピン系とベータ近似をベースにしたアルゴリズムが復号アルゴリズムのチャンピオンデータを出しつつある [12, 13, 14]。

<sup>1</sup>E-mail: kazu@statp.is.tohoku.ac.jp

最近では移動体通信における CDMA 復調方式にランダムスピン系と平均場近似を応用しようという試みも始まりつつある [15, 16]. もともと平均場近似についてはこれまで日本は多くの蓄積を持っている. 特に 1970 年代のスピングラスの研究, 1980 年代から 90 年代にかけてコヒーレント異常法の登場に伴う平均場近似の拡張, 1990 年代におけるニューロコンピューティングにおけるアルゴリズム・平均場近似の情報幾何という見方の提案など過去 30 年以上にわたる蓄積が, 現在, 確率的情報処理の日本の研究が世界的に先行している状況の基盤となっている. 平均場近似の情報処理への応用についての最近の研究成果を集めた解説書と啓蒙書が文献 [17, 18] という形で出版されている. 更に情報統計物理学という更に広い枠組みで基礎的部分から最近の動向まで詳しく書かれた啓蒙書 [19, 20, 21] も出版されている.

このような歴史的背景の中で特に注目されるのが確率推論における確率伝搬法と統計力学におけるベータ近似の同等性である [17, 22]. 両者は歴史的にみて交流はほとんどなかったにも関わらず, 確率モデルという共通の研究対象のなかで, 同じ手法が独立に考案されて来たわけである. そして, 近年, その同等性に着目した様々の研究が行われつつあるのである. 2002 年 12 月には Neural Information Processing Systems という国際会議の中でこの確率伝搬法の数理についての情報工学, 統計科学, 統計力学の研究者による Workshop も開催されている [23]. まさに確率伝搬法という物理と情報の交差点で, 今, 何かが起こりつつある.

本稿では, 主に画像処理と人工知能を例にとり, 平均場近似というものが具体的に情報処理技術にどのように役立てられるかについて具体的な式の導出を可能な限り示しながら解説する. 第 2 節でベイズ統計の枠組みで確率推論と確率モデルの関係について基本的な数理構造を説明する. 第 3 節では厳密な取り扱いの容易ないくつかの確率モデルを紹介する. 第 4 節では 2 つの確率分布間の近さを表す量として情報理論でよく用いられるカルバックライブラー情報量と平均場理論において重量な統計量である自由エネルギーとの関連について解説する. 第 5 節では第 4 節での枠組みをもとに平均場近似についてイジング模型を例にとって解説する. 第 6 節では平均場近似を用いた確率的画像処理アルゴリズムについて紹介する. 第 7 節では第 6 節の枠組みをベータ近似を用いたアルゴリズムへと拡張する. 第 8 節は確率推論についての概説と平均場近似の更なる拡張であるクラスター変分法を用いた具体的アルゴリズムについて概説する. 第 9 節はまとめである.

## 2 確率モデルとベイズ統計

情報処理に統計力学を応用するわけであるから, 当然, 確率モデルを用いた推定へと問題を定式化する必要がある. その際, キーになるのがベイズの公式である. 本節では, ベイズ統計を用いた情報処理の基礎について概説する.

ある事象を考え, その事象として起こりうるすべての場合が合計で  $M$  個として, それに  $1, 2, \dots, M$  などの番号を付けたとする. この番号の中のどれかをとり変数  $A$  を導入し, 「例えば 1 番という番号の付けられた事象が起こったことを “ $A = 1$ ” という数学的記号を用いて表すことにした」とする. この時,  $A$  を確率変数といい, “ $A = 1$ ” で表現された事象の起こる確率を  $\Pr\{A = 1\}$  という記号を用いて表すことにする. 一般に確率変数がある実現値  $1, 2, \dots, M$  のなかの様々な値をとり

うような場合を考えると、 $A = a$  ( $a = 1, 2, \dots, M$ ) により指定された事象の確率は  $\Pr\{A = a\}$  という表現により与えられる。確率  $\Pr\{A = a\}$  は

$$\Pr\{A = a\} \geq 0 \quad (a = 1, 2, \dots, M), \quad \sum_{z=1}^M \Pr\{A = z\} = 1 \quad (1)$$

という条件を満たさなければならない。確率変数  $A$  の確率が

$$\Pr\{A = a\} = P(a) \quad (2)$$

という形で  $a$  の関数  $P(a)$  により与えられたとき、この  $P(a)$  を確率変数  $A$  の確率分布という。式 (1) から確率分布は

$$P(a) \geq 0 \quad (a = 1, 2, \dots, M), \quad \sum_{z=1}^M P(z) = 1 \quad (3)$$

を満たさなければならない。

次に、2 つの事象を考え、その確率変数を  $A_1, A_2$  とし、それぞれの事象として起こりうる場合の総数がそれぞれ  $M_1$  個および  $M_2$  個であるとする。このとき「 $A_1 = a_1$ 」、 $A_2 = a_2$ 」により与えられた事象が両方起きる、すなわち「 $(A_1 = a_1) \cup (A_2 = a_2)$ 」である確率

$$\Pr\{A_1 = a_1, A_2 = a_2\} \quad (a_1 = 1, 2, \dots, M_1; a_2 = 1, 2, \dots, M_2) \quad (4)$$

を確率変数  $A_1$  と  $A_2$  に対する結合確率と呼ぶ。結合確率  $\Pr\{A_1 = a_1, A_2 = a_2\}$  を最初に定義した上で  $A_2$  としてどの事象が起こるかということとは無関係に  $A_1$  が起こる確率を考えた場合、これは

$$\Pr\{A_1 = a_1\} = \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \delta_{a_1, z_1} \Pr\{A_1 = z_1, A_2 = z_2\} = \sum_{z_2=1}^{M_2} \Pr\{A_1 = a_1, A_2 = z_2\} \quad (5)$$

$$\Pr\{A_2 = a_2\} = \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \delta_{a_2, z_2} \Pr\{A_1 = z_1, A_2 = z_2\} = \sum_{z_1=1}^{M_1} \Pr\{A_1 = z_1, A_2 = a_2\} \quad (6)$$

と与えられる。 $\delta_{a,b} \equiv 1$  ( $a = b$ ),  $\delta_{a,b} \equiv 0$  ( $a \neq b$ ) はクロネッカーのデルタである。この時、 $\Pr\{A_k = a_k\}$  ( $k = 1, 2$ ) を  $\Pr\{A_1 = a_1, A_2 = a_2\}$  の確率変数  $A_k$  についての周辺確率と呼ぶ。

話を一気に一般化してみよう。いま  $K$  個の事象を考え、その確率変数を  $A_k$  とし、それぞれの事象として起こりうる場合の総数がそれぞれ  $M_k$  個であるとする。事象「 $(A_1 = a_1) \cup (A_1 = a_1) \cup \dots \cup (A_K = a_K)$ 」が起こる結合確率は次のように与えられる。

$$\Pr\{A_1 = a_1, A_2 = a_2, \dots, A_K = a_K\} \quad (a_k = 1, 2, \dots, M_k, k = 1, 2, \dots, K) \quad (7)$$

以後、確率変数の集合  $\{A_k | k = 1, 2, \dots, K\}$  およびその実現値  $\{a_k | k = 1, 2, \dots, K\}$  をそれぞれ  $\mathbf{A}, \mathbf{a}$  という記号で表すことにすると結合確率は  $\Pr\{\mathbf{A} = \mathbf{a}\}$  と表される。結合確率  $\Pr\{\mathbf{A} = \mathbf{a}\}$  は

$$\Pr\{\mathbf{A} = \mathbf{a}\} \geq 0 \quad (a_k = 1, 2, \dots, M_k, k = 1, 2, \dots, K), \quad \sum_{\mathbf{z}} \Pr\{\mathbf{A} = \mathbf{z}\} = 1 \quad (8)$$

という条件を満たさなければならない。ここで、 $\sum_{\mathbf{z}}$  は  $\mathbf{z} = \{z_k | k = 1, 2, \dots, K\}$  のすべての  $z_k$  に対する和を意味する。

$$\sum_{\mathbf{z}} \equiv \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \cdots \sum_{z_K=1}^{M_K} \quad (9)$$

確率変数  $\mathbf{A}$  の確率が

$$\Pr\{\mathbf{A} = \mathbf{a}\} = P(\mathbf{a}) \quad (10)$$

という形で  $\mathbf{a}$  の関数  $P(\mathbf{a})$  により与えられたとき、この  $P(\mathbf{a})$  を確率変数  $\mathbf{A}$  の結合確率分布という。式 (8) から結合確率分布  $P(\mathbf{a})$  は

$$P(\mathbf{a}) \geq 0, \quad \sum_{\mathbf{z}} P(\mathbf{z}) = 1 \quad (11)$$

を満たさなければならない。結合確率  $\Pr\{\mathbf{A} = \mathbf{a}\}$  に対して

$$\Pr\{A_k = a_k\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (12)$$

$$\Pr\{A_k = a_k, A_{k'} = a_{k'}\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \delta_{a_{k'}, z_{k'}} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (13)$$

$$\Pr\{A_k = a_k, A_{k'} = a_{k'}, A_{k''} = a_{k''}\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \delta_{a_{k'}, z_{k'}} \delta_{a_{k''}, z_{k''}} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (14)$$

という形で様々な周辺確率を定義することができる。

結合確率  $\Pr\{A_1 = a_1, A_2 = a_2\}$  から条件付き確率  $\Pr\{A_1 = a_1 | A_2 = a_2\}$  および  $\Pr\{A_2 = a_2 | A_1 = a_1\}$  は次のように定義される。

$$\begin{aligned} \Pr\{A_1 = a_1 | A_2 = a_2\} &\equiv \frac{\Pr\{A_1 = a_1, A_2 = a_2\}}{\Pr\{A_2 = a_2\}} \\ \Pr\{A_2 = a_2 | A_1 = a_1\} &\equiv \frac{\Pr\{A_1 = a_1, A_2 = a_2\}}{\Pr\{A_1 = a_1\}} \end{aligned} \quad (15)$$

この式から

$$\begin{aligned} \Pr\{A_1 = a_1, A_2 = a_2\} &= \Pr\{A_1 = a_1 | A_2 = a_2\} \Pr\{A_2 = a_2\} \\ &= \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\} \end{aligned} \quad (16)$$

という式が導かれる。両辺を  $\Pr\{A_2 = a_2\}$  で割ることにより次の等式が与えられる。

$$\Pr\{A_1 = a_1 | A_2 = a_2\} = \frac{\Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}}{\Pr\{A_2 = a_2\}} \quad (17)$$

ここで更に周辺確率の定義から

$$\Pr\{A_2 = a_2\} = \sum_{a_1=1}^{M_1} \Pr\{A_1 = a_1, A_2 = a_2\} = \sum_{a_1=1}^{M_1} \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\} \quad (18)$$

であることを考慮すると

$$\Pr\{A_1 = a_1 | A_2 = a_2\} = \frac{\Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}}{\sum_{a_1=1}^{M_1} \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}} \quad (19)$$

が得られる. 式 (17) および式 (19) がベイズの公式である. 式 (19) は  $A_1 = a_1$  という事象が起こる確率  $\Pr\{A_1 = a_1\}$  と  $A_1 = a_1$  という事象が起こったという条件のもとで事象  $A_2 = a_2$  が起こる確率  $\Pr\{A_2 = a_2|A_1 = a_1\}$  から, 事象  $A_2 = a_2$  が起こったという条件のもとで事象  $A_1 = a_1$  が起こっている確率  $\Pr\{A_1 = a_1|A_2 = a_2\}$  が表現できるということを表している. ベイズ統計では  $\Pr\{A_1 = a_1\}$  は事前確率,  $\Pr\{A_1 = a_1|A_2 = a_2\}$  は事後確率と呼ばれている. ベイズ統計の戦略は, 一言で言えば,  $A_1 = a_1$  は原情報,  $A_2 = a_2$  はデータに対応し, 原情報が生成され, それがデータに変換されるという順過程からベイズの公式を用いて逆過程に対する確率, すなわち事後確率を構成し, これをもとにデータから原情報を推定しようというものである.

もう少し複雑な場合として, 3 つの確率変数  $A_1, A_2, A_3$  による場合を考えてみる. 結合確率  $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\}$  は条件付き確率から次のように 2 つの表現で与えられる.

$$\begin{aligned}\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} &= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_1 = a_1, A_2 = a_2\} \\ &= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (20)$$

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \Pr\{A_1 = a_1, A_2 = a_2|A_3 = a_3\}\Pr\{A_3 = a_3\}\quad (21)$$

つまり,

$$\begin{aligned}\Pr\{A_1 = a_1, A_2 = a_2|A_3 = a_3\}\Pr\{A_3 = a_3\} \\ = \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (22)$$

が成り立つわけである. この両辺を  $a_2$  に関して和をとることにより,

$$\begin{aligned}\Pr\{A_1 = a_1|A_3 = a_3\}\Pr\{A_3 = a_3\} \\ = \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (23)$$

すなわち,

$$\begin{aligned}\Pr\{A_1 = a_1|A_3 = a_3\} \\ = \frac{1}{\Pr\{A_3 = a_3\}} \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (24)$$

$$\Pr\{A_3 = a_3\} = \sum_{a_1=1}^{M_1} \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\quad (25)$$

が導かれる. 式 (24) は事象  $A_1 = a_1$  が起こり, 事象  $A_2 = a_2$  が起こり, そしてその結果として事象  $A_3 = a_3$  が起こるとい順過程についての確率が与えられたときに, 逆に事象  $A_3 = a_3$  が起こったという状況のもとで事象  $A_1 = a_1$  が起こっていたかどうかという逆過程に対する条件付き確率を構成する形をとっており, その意味でベイズの公式の拡張版と見なすことができる. 一般にこのような順過程の確率から逆過程を条件付き確率と結合確率の間の関係式をもとに構成してゆく手順を総称してベイズ規則とよんでいる.

3 個の確率変数  $A_1, A_2, A_3$  に対するベイズ規則において結合確率  $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\}$  が

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = P(a_1, a_2, a_3) \quad (26)$$

という形にある関数  $P(a_1, a_2, a_3)$  を用いて与えられれば、逆過程における推論に必要な条件付き確率は定義されることを表しているわけであるが、この結合確率分布は

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \exp\left(-E(a_1, a_2, a_3)\right) \quad (27)$$

$$E(a_1, a_2, a_3) \equiv -\ln P(a_1, a_2, a_3) \quad (28)$$

と書き直すことができる。これは  $z = \exp(\ln(z))$  というよく知られた等式を使っただけであるが、式 (27) をみると  $E(a_1, a_2, a_3)$  をハミルトニアンと見することもできる。式 (27) を形式的に

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \frac{\exp\left(-E(a_1, a_2, a_3)\right)}{\sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \sum_{z_3=1}^{M_3} \exp\left(-E(z_1, z_2, z_3)\right)} \quad (29)$$

と書き換えれば更にわかりやすくなる。つまり、確率モデルは  $P(a_1, a_2, a_3) > 0$  であれば、基本的には統計力学で言うギブス分布 (ボルツマン分布) で表せるのである。あとは  $E(a_1, a_2, a_3)$  がどのような形で与えられるかで、場合によっては統計力学でよく研究されているモデルと対応がつけられることもあるわけである。この状況は確率変数の個数がいくら増えていっても可算個であるかぎり同様である。確率変数の個数、すなわち体系の大きさ (サイズ) が大きくなればそれだけ計算は大変になる。ここで統計力学が更に威力を発揮する。すなわち、統計力学が熱力学的極限という意味での巨大なサイズを持つ体系の巨視的性質を研究してきたわけであるが、その計算技術を適用することができるというわけである。

### 3 厳密なとり扱いの容易な確率モデル

本節では、厳密な取り扱いの容易なくつかの基本的確率モデルについて解説する。

$K$  個の確率変数  $\mathbf{A} = \{A_k | k = 1, 2, \dots, K\}$  からとその実現値  $\mathbf{a} = \{a_k | k = 1, 2, \dots, K\}$  に対する結合確率分布  $\Pr\{\mathbf{A} = \mathbf{a}\} = P(\mathbf{a})$  において平均  $m_k = \sum_{\mathbf{a}} a_k P(\mathbf{a})$ 、分散  $V_k = \sum_{\mathbf{a}} (a_k - m_k)^2 P(\mathbf{a})$  をはじめとする統計量をその定義にもとづいてコンピューターを用いて計算しようとするとき一般には  $K$  重のループ構造からなるプログラムとなり、 $\mathcal{O}(\exp(K))$  のオーダーの計算量が必要となってしまう。例えば画像処理の場合、市販の携帯電話に搭載されたデジタルカメラで 100 万画素程度の画素数を持っているわけで、この場合  $K = 1000000$  と言うことになるわけである。しかし、一部の特殊な例では  $\mathcal{O}(K)$  のオーダーの計算量で計算できてしまう場合がある。本節ではその一部の例を紹介する。

結合確率分布  $\Pr\{\mathbf{A} = \mathbf{a}\} = P(\mathbf{a})$  が

$$P(\mathbf{a}) = \prod_{k=1}^K P_k(a_k) \quad (30)$$

により与えられる場合を考える. ここで  $P_k(a_k)$  は  $P_k(a_k) \geq 0$  ( $a_k = 1, 2, \dots, M_k$ ) および  $\prod_{z_k=1}^{M_k} P_k(z_k) = 1$  を満足するものとする. これはいわゆる確率変数  $A$  が互いに独立な場合であり, 任意の自然数  $n$  に対して以下の等式が成り立つことは容易に確かめられる.

$$\sum_{\mathbf{z}} z_k^n P(\mathbf{z}) = \left( \prod_{l=1}^{k-1} \sum_{z_l=1}^{M_l} P_l(z_l) \right) \left( \sum_{z_k=1}^{M_k} z_k^n P_k(z_k) \right) \left( \prod_{l=k+1}^K \sum_{z_l=1}^{M_l} P_l(z_l) \right) = \sum_{z_k=1}^{M_k} z_k^n P_k(z_k) \quad (31)$$

式 (31) から平均  $m_k$ , 分散  $V_k$  などの重要な統計量が計算させる.

結合確率分布  $\Pr\{A = \mathbf{a}\} = P(\mathbf{a})$  が

$$P(\mathbf{a}) = \frac{\prod_{k=1}^K W_{k,k+1}(a_k, a_{k+1})}{\sum_{\mathbf{z}} \prod_{k=1}^K W_{k,k+1}(z_k, z_{k+1})} \quad (32)$$

により与えられる場合を考える. この確率モデルに対して  $\mathcal{L}_{k-1 \rightarrow k}(a_k)$  および  $\mathcal{R}_{k+1 \rightarrow k}(a_k)$  という 2 つの量を次のように導入する.

$$\mathcal{L}_{k-1 \rightarrow k}(a_k) = \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \cdots \sum_{z_k=1}^{M_k} \delta_{a_k, z_k} \prod_{l=1}^{k-1} W_{l,l+1}(z_l, z_{l+1}) \quad (33)$$

$$\mathcal{R}_{k+1 \rightarrow k}(a_k) = \sum_{z_k=1}^{M_k} \sum_{z_{k+1}=1}^{M_{k+1}} \cdots \sum_{z_K=1}^{M_K} \delta_{a_k, z_k} \prod_{l=k}^{K-1} W_{l,l+1}(z_l, z_{l+1}) \quad (34)$$

これらの量は

$$\begin{aligned} \mathcal{L}_{k-1 \rightarrow k}(a_k) &= \sum_{z_{k-1}=1}^{M_{k-1}} \sum_{z_k=1}^{M_k} \delta_{a_k, z_k} W_{k-1,k}(z_{k-1}, z_k) \left( \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \cdots \sum_{z_{k-2}=1}^{M_{k-2}} \prod_{l=1}^{k-2} W_{l,l+1}(z_l, z_{l+1}) \right) \\ &= \sum_{z_{k-1}=1}^{M_{k-1}} \sum_{z_k=1}^{M_k} \delta_{a_k, z_k} W_{k-1,k}(z_{k-1}, z_k) \left( \sum_{z'_1=1}^{M_1} \sum_{z'_2=1}^{M_2} \cdots \sum_{z'_{k-1}=1}^{M_{k-1}} \delta_{z_{k-1}, z'_{k-1}} \prod_{l=1}^{k-2} W_{l,l+1}(z'_l, z'_{l+1}) \right) \\ &= \sum_{z_{k-1}=1}^{M_{k-1}} \sum_{z_k=1}^{M_k} \delta_{a_k, z_k} W_{k-1,k}(z_{k-1}, z_k) \mathcal{L}_{k-2 \rightarrow k-1}(z_{k-1}) \end{aligned} \quad (35)$$

$$\begin{aligned} \mathcal{R}_{k+1 \rightarrow k}(a_k) &= \sum_{z_k=1}^{M_k} \sum_{z_{k+1}=1}^{M_{k+1}} \delta_{a_k, z_k} W_{k,k+1}(z_k, z_{k+1}) \left( \sum_{z_{k+2}=1}^{M_{k+2}} \sum_{z_{k+3}=1}^{M_{k+3}} \cdots \sum_{z_K=1}^{M_K} \prod_{l=k+1}^K W_{l,l+1}(z_l, z_{l+1}) \right) \\ &= \sum_{z_k=1}^{M_k} \sum_{z_{k+1}=1}^{M_{k+1}} \delta_{a_k, z_k} W_{k,k+1}(z_k, z_{k+1}) \left( \sum_{z'_{k+1}=1}^{M_{k+1}} \sum_{z'_{k+2}=1}^{M_{k+2}} \cdots \sum_{z'_K=1}^{M_K} \delta_{z_{k+1}, z'_{k+1}} \prod_{l=k+1}^K W_{l,l+1}(z'_l, z'_{l+1}) \right) \\ &= \sum_{z_k=1}^{M_k} \sum_{z_{k+1}=1}^{M_{k+1}} \delta_{a_k, z_k} W_{k,k+1}(z_k, z_{k+1}) \mathcal{R}_{k+2 \rightarrow k+1}(z_{k+1}) \end{aligned} \quad (36)$$

という形に書き直される. すなわち  $\mathcal{L}_{k \rightarrow k+1}(a_{k+1})$  および  $\mathcal{R}_{k \rightarrow k-1}(a_{k-1})$  に対する漸化式として次のようにまとめられる.

$$\mathcal{L}_{k \rightarrow k+1}(a_{k+1}) = \sum_{z_k=1}^{M_k} \sum_{z_{k+1}=1}^{M_{k+1}} \delta_{a_{k+1}, z_{k+1}} \mathcal{L}_{k-1 \rightarrow k}(z_k) W_{k,k+1}(z_k, z_{k+1}) \quad (37)$$

$$\mathcal{R}_{k \rightarrow k-1}(a_{k-1}) = \sum_{z_{k-1}=1}^{M_{k-1}} \sum_{z_k=1}^{M_k} \delta_{a_{k-1}, z_{k-1}} \mathcal{L}_{k+1 \rightarrow k}(z_k) W_{k-1,k}(z_{k-1}, z_k) \quad (38)$$

漸化式 (37) および (38) を用いて逐次的に得られた  $\mathcal{L}_{k-1 \rightarrow k}(a_k)$  および  $\mathcal{R}_{k-1 \rightarrow k}(a_k)$  を用いて  $n$  次のモーメントは

$$\sum_{\mathbf{z}} z_k^n P(\mathbf{z}) = \frac{\sum_{z_k=1}^{M_k} z_k \mathcal{L}_{k-1 \rightarrow k}(z_k) \mathcal{R}_{k+1 \rightarrow k}(z_k)}{\sum_{z_k=1}^{M_k} \mathcal{L}_{k-1 \rightarrow k}(z_k) \mathcal{R}_{k+1 \rightarrow k}(z_k)} \quad (39)$$

と与えられ、これにより平均  $m_k$ 、分散  $V_k$  が求められることになる。

これと同様の取り扱いより複雑な構造をもつ確率モデルに一般化される。これをできるだけ明確に表すために、確率分布の構造をグラフ表現として表すことを行う。まず  $A = \{A_1, A_2, \dots, A_K\}$  に対して  $K$  個のノード  $\Omega = \{1, 2, \dots, K\}$  を考える。このノードのうちのいくつかの対を選んでリンクで結ぶ。リンクで結ばれたノード対を最近接ノード対と呼ぶことにし、すべての最近接ノード対の集合を  $B$  により表すことにする。この確率変数  $A$  とその実現値  $a$  および集合  $B$  に対して結合確率分布  $\Pr\{A = a\} = P(a)$  が

$$P(a) = \frac{\prod_{kl \in B} W_{k,l}(a_k, a_l)}{\sum_{\mathbf{z}} \prod_{kl \in B} W_{k,l}(z_k, z_l)} \quad (40)$$

を考える。仮に  $B$  が

$$B = \{kl | k = 1, 2, \dots, K-1, l = k+1\} \quad (41)$$

により与えられる時、式 (40) は式 (32) と等価となる。式 (41) に付随して与えられるグラフ表現を 1 次元鎖と呼ぶことにする。また、 $B$  に属する任意の最近接ノード対  $kl$  に対して  $B$  から  $kl$  を除いた集合  $B \setminus \{kl\}$  を考える。この  $B$  と  $B \setminus \{kl\}$  に対してグラフ表現を考えたとき、 $B$  に付随するグラフが単連結グラフであり、且つ  $B \setminus \{kl\}$  は常に 2 つの単連結グラフに分かれてしまう時、 $B$  に付随するグラフを木構造を持つという。 $B$  に付随するグラフが木構造を持つとき、式 (40) で与えられた確率モデルに対する  $n$  次のモーメントは次のように与えられる。

$$\sum_{\mathbf{z}} z_k^n P(\mathbf{z}) = \frac{\sum_{z_k=1}^{M_k} z_k^n \prod_{l \in \mathbf{c}_k} \Lambda_{l \rightarrow k}(z_k)}{\sum_{z_k=1}^{M_k} \prod_{l \in \mathbf{c}_k} \Lambda_{l \rightarrow k}(z_k)} \quad (42)$$

ここで  $\Lambda_{k' \rightarrow k}(a_k)$  は漸化式

$$\Lambda_{k' \rightarrow k}(a_k) = \frac{\sum_{z_k=1}^{M_k} \sum_{z_{k'}=1}^{M_{k'}} \delta_{a_k, z_k} W_{k,k'}(z_k, z_{k'}) \prod_{l \in \mathbf{c}_{k'} \setminus \{k\}} \Lambda_{l \rightarrow k'}(z_{k'})}{\sum_{z_k=1}^{M_k} \sum_{z_{k'}=1}^{M_{k'}} W_{k,k'}(z_k, z_{k'}) \prod_{l \in \mathbf{c}_{k'} \setminus \{k\}} \Lambda_{l \rightarrow k'}(z_{k'})} \quad (k \in \Omega, k' \in \mathbf{c}_k) \quad (43)$$

から逐次的に計算される。ここで、 $\mathbf{c}_k$  はノード  $k$  のすべての最近接ノードの集合を意味し、 $\mathbf{c}_{k'} \setminus \{k\}$  は  $\mathbf{c}_{k'}$  からノード  $k$  を除いた集合を表す<sup>2</sup>。また従って、この場合も  $\Lambda_{k' \rightarrow k}(a_k)$  から平均  $m_k$ 、分散  $V_k$  が求められることになる。

<sup>2</sup>例えば、式 (41) で  $B$  が与えられた場合、 $\mathbf{c}_k = \{k-1, k+1\}$ 、 $\mathbf{c}_k \setminus \{k \pm 1\} = \{k \mp 1\}$  となる。



ここまでの確率モデルの構造をまとめると

- (a) 式 (30) に与えられるように確率変数が互いに独立なら、個別の確率変数ごとの計算に帰着される。
- (b) 式 (40) に与えられる確率分布において最近接ノード対の集合  $B$  に付随するグラフ構造が一次元鎖または木構造である場合、そのグラフ上で与えられるある局所的な漸化式に従って逐次的に計算する問題へと帰着される。

ということになる。この 2 つの場合は式 (31), (37), (38), (43) において  $\sum_{z_k=1}^{M_k}$  を数値的に計算したとしても、基本的には (数値計算の際に桁落ち等で生じる誤差を除いては) 厳密な取り扱いをしていると見なすことができる。問題はこれらのいずれにも属さない確率モデルをどのように扱えば良いかということになる。(b) はグラフ構造の閉路がないということの意味しているが閉路を有するグラフ構造のある場合は本節の定式化の範囲では取り扱い切れなくなってしまう。確率推論の分野ではこの閉路の問題を、一部の確率モデルに対しては局所的なある種の変換により、木構造をもつグラフ構造に変換することで解決してきた。しかし、一つのグラフ表現のなかに存在する閉路の個数が多くなってくると、このような変換だけでは対応しきれない場合がでてくる。情報処理では厳密な取り扱いの難しい確率モデルに対しては厳密な取り扱いをあきらめて、近似を導入し、統計量を得るためのできるだけ近似精度のよいアルゴリズムを構成するという視点で、その方法論の一つとして平均場近似が採用されている。

## 4 カルバック・ライブラー情報量と自由エネルギー

本節では、平均場近似の定式化を解説する上での基礎として、自由エネルギーの解釈を情報理論におけるカルバック・ライブラー情報量という量にもとづいて説明する。

ある確率変数  $A$  とその実現値  $a$  に対して 2 つの結合確率分布  $P(a)$  と  $Q(a)$  を考えたとき、

$$D[P||Q] \equiv \sum_z Q(z) \ln \left( \frac{Q(z)}{P(z)} \right) \quad (44)$$

という量を導入する。この量はカルバック・ライブラー情報量と呼ばれ、この 2 つの確率分布の分布間近さに対応しており、次の性質を持つ。

- (i)  $D[P||Q] \geq 0$
- (ii)  $P(a) = Q(a) \Rightarrow D[P||Q] = 0$

任意の  $x > 0$  に対して  $\ln(x) \leq x - 1$  が成り立ち、等号は  $x = 1$  のときのみ成り立つ。このことから得られる不等式  $\ln \left( \frac{P(z)}{Q(z)} \right) \leq \frac{P(z)}{Q(z)} - 1$  を  $D[P||Q]$  の表式に適用することにより、

$$D[P||Q] \equiv \sum_z Q(z) \ln \left( \frac{Q(z)}{P(z)} \right) \geq \sum_z Q(z) \left( 1 - \frac{P(z)}{Q(z)} \right) \geq \sum_z P(z) - \sum_z Q(z) = 1 - 1 = 0 \quad (45)$$

という形で  $D[P||Q] \geq 0$  が示される。もちろん、 $D[P||Q] = D[Q||P]$  が常に成り立つわけではないので数学的意味に置いて距離と呼ぶことは言い過ぎであるが、2 つの確率分布間の近さを表す量として情報理論などでよく用いられる。

そこで  $a$  のある関数  $E(a)$  を考え, この  $E(a)$  に対して

$$P(a) = \frac{\exp(-E(a))}{\sum_z \exp(-E(z))} \quad (T > 0) \quad (46)$$

という形で与えられる確率分布を導入する. この確率分布は統計力学ではギブス分布あるいはボルツマン分布と呼ばれて, 関数  $E(a)$  はハミルトニアンに,  $T$  は温度に対応している. このギブス分布を式 (44) に代入すると,

$$\mathcal{D}[P||Q] = \mathcal{F}[Q] + \ln\left(\sum_z \exp(-E(z))\right) \quad (47)$$

$$\mathcal{F}[Q] \equiv \sum_z E(z)Q(z) - \left(-\sum_z Q(z)\ln(Q(z))\right) \quad (48)$$

が得られる.  $\mathcal{F}[Q]$  は自由エネルギーに対応しており, 式 (48) の左辺の第 1 項は内部エネルギーに第 2 項はエントロピーに対応している. 確率分布  $P(a)$  がギブス分布 (46) で与えられたときに,  $\mathcal{D}[P||Q]$  をできるだけ小さくするようにして, 確率分布  $Q(a)$  を  $P(a)$  に近づけるということは, 自由エネルギー  $\mathcal{F}[Q]$  をできるだけ小さくするように  $Q(a)$  を選ぶことに対応していると言える.

規格化条件  $\sum_z Q(z) = 1$  を拘束条件として自由エネルギー  $\mathcal{F}[Q]$  の最小化の変分原理

$$\hat{Q}(z) = \operatorname{argmin}_Q \left\{ \mathcal{F}[Q] \mid \sum_z Q(z) = 1 \right\} \quad (49)$$

を考えることにより, ギブス分布を逆に導くことができる. まず, 規格化条件に対してラグランジュの未定係数  $\lambda$  を導入する.

$$\mathcal{L}[Q] \equiv \mathcal{F}[Q] - \lambda \left( \sum_z Q(z) - 1 \right) \quad (50)$$

汎関数  $\mathcal{L}[Q]$  を  $Q(z)$  について変分をとると次のような極値条件が得られる.

$$\hat{Q}(a) = \exp(-E(a) - 1 + \lambda) \quad (51)$$

式 (51) を規格化条件の式に代入することにより  $\lambda$  が決定され,  $\hat{Q}(z)$  の表式は式 (46) の右辺として与えられる. また, 式 (46) の右辺を式 (48) の  $Q(z)$  に代入することにより

$$\mathcal{F}[\hat{Q}] = \min_Q \mathcal{F}[Q] = -\ln\left(\sum_z \exp(-E(z))\right) \quad (52)$$

という表式として最小化された自由エネルギーの表式が得られる.

## 5 平均場近似の情報論的理解

本節では大規模統計モデルに対する統計力学的近似解析手法としてなじみ深い平均場近似を自由エネルギー最小の変分原理という立場から解説する.

平均場近似という物理を専攻する学生は、学部か大学院の講義で「個々の確率変数に着目してその周りからの確率変数からの影響をある種の期待値で置き換えてしまう近似である。」という形に教わる。例えば、簡単のために正方格子  $\Omega = \{(x, y) | x = 1, 2, \dots, M, y = 1, 2, \dots, N\}$  上のイジング模型を考える。各格子点  $(x, y)$  上の確率変数  $S_{x,y}$  はそれぞれ  $\pm 1$  の値をとり、その確率分布は

$$\Pr\{S = s\} = P(s) \equiv \frac{\exp\left(\sum_{(x,y) \in \Omega} (B_{x,y}s_{x,y} + C s_{x,y}s_{x+1,y} + C s_{x,y}s_{x,y+1})\right)}{\sum_{z} \exp\left(\sum_{(x,y) \in \Omega} (B_{x,y}z_{x,y} + C z_{x,y}z_{x+1,y} + C z_{x,y}z_{x,y+1})\right)} \quad (53)$$

により与えられる。ここで  $\sum_z$  は

$$\sum_z = \prod_{(x,y) \in \Omega} \sum_{z_{x,y} = \pm 1} \quad (54)$$

により定義される。平均場近似では確率変数  $S_{x,y}$  の期待値  $m_{x,y} \equiv \sum_z z_{x,y} \Pr\{S = z\}$  を用いて確率変数  $S$  はその実現値として  $(s_{x,y} - m_{x,y})(s_{x+1,y} - m_{x+1,y}) \simeq 0$  および  $(s_{x,y} - m_{x,y})(s_{x,y+1} - m_{x,y+1}) \simeq 0$  を満たす  $s = \{s_{x,y} | (x,y) \in \Omega\}$  の確率がそれ以外の場合に比べて非常に高いと仮定する。これがいわゆる「ゆらぎを無視する」という意味である。この場合、 $(s_{x,y} - m_{x,y})(s_{x+1,y} - m_{x+1,y}) \simeq 0$  は次のように変形される。

$$s_{x,y}s_{x+1,y} \simeq s_{x,y}m_{x+1,y} + m_{x,y}s_{x+1,y} - m_{x,y}m_{x+1,y} \quad (55)$$

$(s_{x,y} - m_{x,y})(s_{x,y+1} - m_{x,y+1}) \simeq 0$  についても同様である。これを式 (53) に代入し、 $m_{x,y}$  を計算することにより次のような  $\{m_{x,y}\}$  に対する漸化式を導くことができる (図 1)。

$$m_{x,y} = \tanh\left(B_{x,y} + C(m_{x+1,y} + m_{x-1,y} + m_{x,y+1} + m_{x,y-1})\right) \quad (56)$$

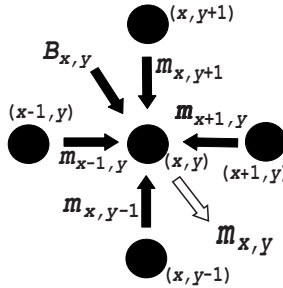


図 1: 平均場方程式 (56) の構造。

同じ方程式は自由エネルギー最小の変分原理からも導出できる。この確率分布に対して各確率変数  $S_{x,y}$  ごとの周辺確率分布

$$P_{x,y}(s_{x,y}) \equiv \sum_z \delta_{s_{x,y}, z_{x,y}} P(s) \quad (57)$$

を計算する必要があるとする。定義に従って計算しようとする $2^{|\Omega|-1}$ 通りの状態についての和を計算しなければならない。そこで、この確率分布  $P(s)$  を近似する確率分布として

$$Q(s) \equiv \prod_{(x,y) \in \Omega} Q_{x,y}(s_{x,y}) \quad (58)$$

$$Q_{x,y}(s_{x,y}) \equiv \sum_z \delta_{s_{x,y}, z_{x,y}} Q(z) \quad (59)$$

を満たす分布  $Q(s)$  を導入し、自由エネルギー (48) に代入する。

$$\begin{aligned} \mathcal{F}[Q] &= \mathcal{F}_{\text{MFA}}[\{Q_{x,y}|(x,y) \in \Omega\}] \\ &\equiv - \sum_{(x,y) \in \Omega} \left( \sum_{\zeta = \pm 1} \zeta Q_{x,y}(\zeta) \right) \left( B_{x,y} + C \sum_{\zeta = \pm 1} \zeta Q_{x+1,y}(\zeta) + C \sum_{\zeta = \pm 1} \zeta Q_{x,y+1}(\zeta) \right) \\ &\quad + \sum_{(x,y) \in \Omega} \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln Q_{x,y}(\zeta) \end{aligned} \quad (60)$$

規格化条件  $\sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1$  を拘束条件として  $\{Q_{x,y}(\zeta)| (x,y) \in \Omega\}$  についての変分をとることにより、 $D[P||Q]$  という尺度で  $Q(s)$  が  $P(s)$  に最も近くなるように決めた  $\{Q_{x,y}(\zeta)| (x,y) \in \Omega\}$  を  $\{\hat{Q}_{x,y}(\zeta)| (x,y) \in \Omega\}$  により表す。

$$\hat{Q}_{x,y}(\zeta) = \operatorname{argmin}_Q \left\{ \mathcal{F}_{\text{MFA}}[\{Q_{x,y}|(x,y) \in \Omega\}] \mid \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1 \ ((x,y) \in \Omega) \right\} \quad (61)$$

カルバック・ライブラー情報量  $D[P||Q]$  を最小にすることと、自由エネルギー  $\mathcal{F}[Q]$  を最小にすることは等価なので、 $\{\hat{Q}_{x,y}(\zeta)| (x,y) \in \Omega\}$  は  $\{P_{x,y}(\zeta)| (x,y) \in \Omega\}$  の良い近似になっているものと見なすことができる。

規格化条件  $\sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1$  に対するラグランジュの未定係数  $\lambda_{x,y}$

$$\mathcal{L}[\{Q_{x,y}|(x,y) \in \Omega\}] \equiv \mathcal{F}_{\text{MFA}}[\{Q_{x,y}|(x,y) \in \Omega\}] - \sum_{(x,y) \in \Omega} \lambda_{x,y} \left( \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) - 1 \right) \quad (62)$$

という形で導入し、 $\{Q_{x,y}(\zeta)| (x,y) \in \Omega\}$  についての変分をとり、その後に規格化条件を満たすように  $\lambda_{x,y}$  を決めることにより、 $\{\hat{Q}_{x,y}(\zeta)| (x,y) \in \Omega\}$  に対する決定方程式は次のように得られる。

$$\hat{Q}_{x,y}(\zeta) = \frac{\exp \left( \left( B_{x,y} + C \sum_{(x',y') \in \mathbf{c}_{x,y}} \left( \sum_{\zeta' = \pm 1} \zeta' \hat{Q}_{x',y'}(\zeta') \right) \right) \zeta \right)}{\sum_{\zeta = \pm 1} \exp \left( \left( B_{x,y} + C \sum_{(x',y') \in \mathbf{c}_{x,y}} \left( \sum_{\zeta' = \pm 1} \zeta' \hat{Q}_{x',y'}(\zeta') \right) \right) \zeta \right)} \quad (63)$$

$$\mathbf{c}_{x,y} \equiv \left\{ (x+1, y), (x-1, y), (x, y+1), (x, y-1) \right\} \quad (64)$$

ここで、 $S_{x,y}$  の期待値の定義を周辺確率分布  $Q_{x,y}(\zeta)$  で書き換えた表式

$$m_{x,y} \simeq \sum_z z_{x,y} Q(z) = \sum_{\zeta = \pm 1} \zeta Q_{x,y}(\zeta) \quad (65)$$

を式 (63) に代入することにより式 (56) に与えられる, 物理においてなじみのある正方格子上的イジング模型に対する平均場方程式が得られる. ひとつの見方ではあるが, 平均場近似とは与えられた確率分布をその周辺確率分布の積の形に与えられた関数系で与えられた確率分布でカルバック・ライブラー情報量という尺度のもとで最も近くなるように決めた近似であると言える.

通常, 物理で教わるイジング模型は  $B_{x,y} = h/T ((x,y) \in \Omega)$ ,  $C = J/T$  と置いたものであり,  $h$  は外場,  $J$  は相互作用,  $T$  は温度と呼ばれ, 格子  $\Omega$  のサイズも  $M \times N$  の有限系ではなく, 無限に大きなサイズを持つ場合が考えられる. この場合, 確率変数  $S_{x,y}$  の期待値  $m_{x,y}$  は格子点  $(x,y)$  の位置によらず一定であるから,  $m_{x,y} = m$  と置くことができる.  $m$  は統計力学では秩序パラメータとよばれ, 強磁性体を想定すれば  $s_{x,y}$  はスピン,  $h$  は磁場であり,  $m$  は磁化に対応する. 従って平均場方程式 (56) は

$$m = \tanh\left(\frac{h}{T} + \frac{4J}{T}m\right) \quad (66)$$

という形に与えられる. 式 (56) をみて, 「こんな式, 覚えがないぞ」と思った物理学科出身の学生さんも式 (66) をみればどこかで見たかと思っただけのものと思う. 式 (66) を  $h$  と  $J$  を固定して様々の  $T$  の値に対して反復法を用いて数値的に解くアルゴリズムを以下に与える.

平均場方程式 (66) を解く反復計算アルゴリズム

Step 1:  $J$  と  $h$  の値を入力する.

Step 2:  $T$  の初期値を  $T = 8$  に設定する.

Step 3: 初期値  $m = 0$  を設定する.

Step 4: 次の操作を  $\varepsilon < 10^{-6}$  となるまで繰り返す.

$$\mu \leftarrow m, \quad m \leftarrow \tanh\left(\frac{h}{T} + \frac{4J}{T}\mu\right), \quad \varepsilon \leftarrow |m - \mu|$$

Step 5:  $T \leftarrow T - 0.10$  として,  $T/J < 0.10$  なら終了し, そうでなければ Step 4 へ戻る.

$h = 0.00010$ ,  $J = 1$  と設定して, 上の反復法を用いて計算した  $m$  の値を図 (2) に与える. 図 2 は統

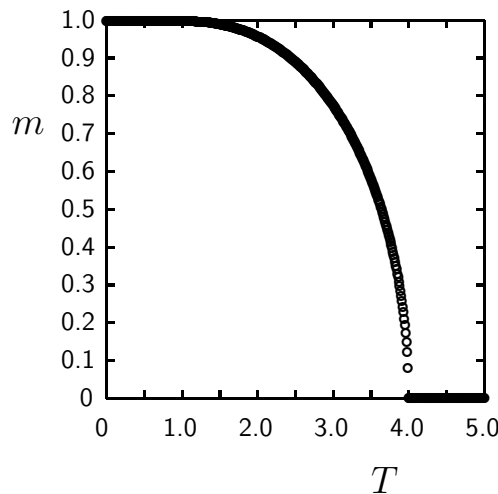


図 2:  $h = 0.00010$ ,  $J = 1$  の場合に平均場方程式 (66) を反復法により解いて得られた  $m$  の  $T$  依存性.

計力学の教科書でよく目にするイジング模型の磁場のないときの磁化の温度依存性の曲線である.<sup>3</sup>

<sup>3</sup>テクニカルなことであるが, 磁場がないと称して, 実際には  $h = 0.00010$  と設定しているのは式 (66) を完全に  $h = 0$

## 6 確率的画像処理のベイズ的アプローチ

ここから、具体的情報処理の問題に入ることにする。本節では 2 値画像の画像修復を例にとり、ごくごく簡単な場合に限定して、ベイズ統計の立場で確率的画像処理の枠組みを解説する。なお、本節および次節で与えられる定式化の詳細は文献 [9, 10, 11, 24, 25, 26] に与えられている。

まず、画素の位置を  $(x, y)$ 、原画像および劣化画像の階調値についての確率変数を  $F_{x,y}$  および  $G_{x,y}$  により表すことにする。画素は正方格子  $\Omega \equiv \{(x, y) | x = 1, 2, \dots, M, y = 1, 2, \dots, N\}$  上に並べられ周期境界条件が課されているものとする。2 次元的なラベル付けをされた確率変数の集合は確率場と呼ばれることがある。原画像の確率場は  $F = \{F_{x,y} | (x, y) \in \Omega\}$ 、劣化画像の確率場は  $g = \{g_{x,y} | (x, y) \in \Omega\}$  である。このとき、原画像  $f = \{f_{x,y} | (x, y) \in \Omega\}$  と劣化画像  $g = \{g_{x,y} | (x, y) \in \Omega\}$  に対してベイズの公式は次のように与えられる

$$\Pr\{F = f | G = g\} = \frac{\Pr\{G = g | F = f\} \Pr\{F = f\}}{\sum_z \Pr\{G = g | F = z\} \Pr\{F = z\}} \quad (67)$$

ここで  $\sum_z$  は

$$\sum_z = \prod_{(x,y) \in \Omega} \sum_{z_{x,y} = \pm 1} \quad (68)$$

により定義される。  $\Pr\{G = g | F = f\}$  は原画像  $f$  から劣化画像  $g$  が生成される確率である。  $\Pr\{F = f\}$  は原画像  $f$  そのものの事前確率である。ベイズの公式の右辺は、事前確率  $\Pr\{F = f\}$  をもとに、まず原画像  $f$  が生成され、その原画像  $f$  から劣化過程  $\Pr\{G = g | F = f\}$  を通して劣化画像  $g$  が生成されるといういわゆる順過程の状況を表している。観測された劣化画像はこの順過程によって生成されたものであると考えれば、ベイズの公式は「順過程を表す式」が「劣化画像  $g$  が与えられたという条件のもとでの原画像  $f$  に対する事後確率  $\Pr\{F = f | G = g\}$ 」に等しいということの意味している (図 3)。

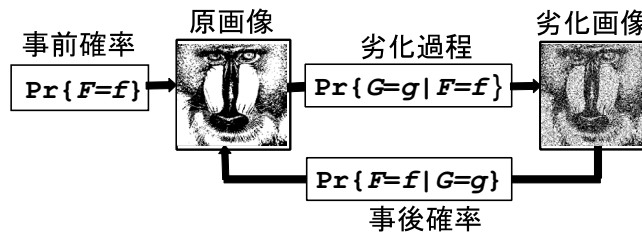


図 3: ベイズ統計における原画像の推定メカニズム。

最も基本的な場合として、各画素が階調値として  $-1$  と  $1$  のみをとる 2 値画像を考え、「劣化過程は原画像から各画素で独立にある確率で劣化されていること」、「原画像の各画素の階調値はその近傍画素の階調値と同じ値をとる確率が高い」という 2 つの仮定が基本的であると考え、これらを明確に式を用いて書き下すと次のようになる。

と設定して、 $m = 0$  を初期値として反復法で解いた場合には  $m = 0$  という解しか得ることができないので、ほんの少しだけ  $h = 0.00010$  という形で磁場をかけて  $m \neq 0$  の解が存在する場合にはそれが得られるようにしている。

**劣化過程** 劣化画像  $g$  は原画像  $f$  の各画素の状態が各画素ごとに独立に確率  $p$  で  $-1$  から  $1$  へ、または  $1$  から  $-1$  へと置き換わることにより生成されるものとする (図 4).

$$\Pr\{G = g | F = f\} \equiv \prod_{(x,y) \in \Omega} \left( (1-p)\delta_{f_{x,y},g_{x,y}} + p(1 - \delta_{f_{x,y},g_{x,y}}) \right) \quad (69)$$

**原画像に対する事前情報** 原画像  $f$  は次の確率分布の高い確率を与える画像のひとつである (図 5).

$$\Pr\{F = f\} = \frac{\left( \prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(f_{x,y} - f_{x+1,y})^2\right) \right) \left( \prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(f_{x,y} - f_{x,y+1})^2\right) \right)}{\sum_z \left( \prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(z_{x,y} - z_{x+1,y})^2\right) \right) \left( \prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(z_{x,y} - z_{x,y+1})^2\right) \right)} \quad (70)$$

確率変数  $f_{x,y}$  の集合すなわち確率場  $f$  は各画素  $(x, y)$  の階調値  $f_{x,y}$  がその近傍画素  $(x \pm 1, y)$ ,  $(x, y \pm 1)$  の階調値にのみ依存する形になっており、この性質を持つ確率場  $f$  は総称してマルコフ確率場と呼ばれる。

2 元対称通信路は情報理論で考えられる最も基本的な劣化過程の一つである。図 4 に示すように各画素で独立に原画像と異なる階調値に置き換えられるということなので、例えば  $p = 0.2$  と設定すると原画像と劣化画像で 20% 程度の画素が劣化されているということになる。式 (70) は最近接

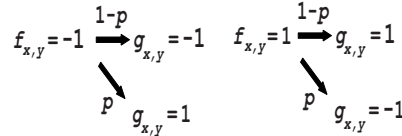


図 4: 2 元対称通信路.

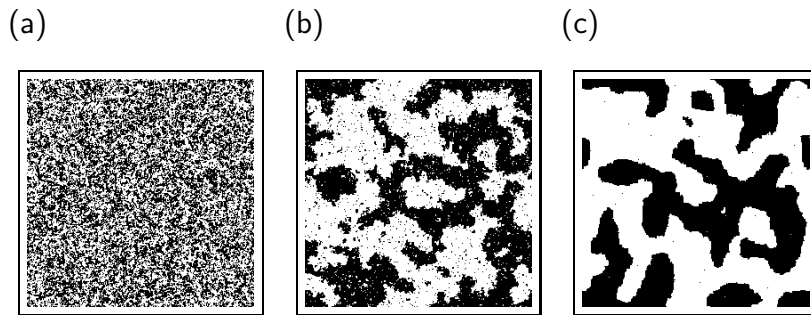


図 5: いくつかの  $\alpha$  の値に対して確率分布 (70) に従うモンテカルロ・シミュレーションのスナップショットとして生成された 2 値画像. (a)  $\alpha = 0.25$ . (b)  $\alpha = 0.5$ . (c)  $\alpha = 1$ .

画素対が同じ階調値をとればとるほど確率が高くなるわけなので、当然、最大確率を与える画像は

真っ白か真っ黒な画像である。そんな画像ばかり扱っていたのでは使えない。当然の話である。しかし、どの程度かなどという堅いことを言わないで比較的高い確率を与える画像はと考えると話は多少変わってくる。いくつかの  $\alpha$  の値に対してモンテカルロシミュレーションのスナップショットとして生成された画像の例が図 5 である。確率モデル (70) のスナップショットを持ってきてこれがわれわれの扱える原画像でございますと言われても現実世界の画像は様々な意味と構造を持っていて、そんな画像で置き換えられるわけではない。しかし、このような仮定の下で画像処理アルゴリズムを構成すると現実世界の画像も含めてうまく処理できてしまうから不思議である。著者がこの説明をするとき、「現実世界の画像の部分部分のパターンをみると図 5 のような画像がみえてくると思うことは無理がないとは考えられませんか?」という苦しい説明をする。このことを深く深く考えてゆくと話がここで終わってしまうので、これについてはとりあえず認めていただくということで話をすすめさせていただきたい。

式 (69) および式 (70) をベイズの公式 (67) に代入することにより、 $\Pr\{F = f|G = g\}$  は次のように与えられる。

$$\Pr\{F = f|G = g\} = \frac{\exp(-E(f|g))}{\sum_z \exp(-E(z|g))} \quad (71)$$

$$E(f|g) \equiv - \sum_{(x,y) \in \Omega} \left( -\frac{1}{2}\beta(g_{x,y} - f_{x,y})^2 - \frac{1}{2}\alpha(f_{x,y} - f_{x+1,y})^2 - \frac{1}{2}\alpha(f_{x,y} - f_{x,y+1})^2 \right) \quad (72)$$

$$\beta \equiv \frac{1}{2} \ln \left( \frac{1-p}{p} \right) \quad (73)$$

修復画像  $\hat{f} = \{\hat{f}_{x,y} | (x,y) \in \Omega\}$  はこの事後確率分布  $\Pr\{F = f|G = g\}$  から

$$\hat{f}_{x,y} = \operatorname{argmax}_{f_{x,y}} \Pr\{F_{x,y} = f_{x,y} | G = g\} \quad (74)$$

$$\Pr\{F_{x,y} = f_{x,y} | G = g\} \equiv \sum_z \delta_{f_{x,y}, z_{x,y}} \Pr\{F = z | G = g\} \quad (75)$$

により、各画素ごとに階調値を推定することにより決定される。 $\Pr\{F_{x,y} = f_{x,y} | G = g\}$  は事後確率分布  $\Pr\{F = f | G = g\}$  の確率変数  $f_{x,y}$  についての周辺確率分布とよばれる。式 (75) の定義にもとづいて  $\Pr\{F_{x,y} = f_{x,y} | G = g\}$  を計算しようとする、 $2^{|\Omega|-1}$  通りの画像についての和を計算する必要があり、膨大な計算量が必要となる。次節ではこれを計算する方法として確率伝搬法について説明する。

## 7 ベーテ近似を用いた確率的画像処理アルゴリズム

本節では、確率モデル (71) の各画素ごとの確率変数  $F_{x,y}$  の周辺確率分布  $\Pr\{F_{x,y} = f_{x,y} | G = g\}$  を高精度に計算する近似として、ベーテ近似について自由エネルギー最小の変分原理にもとづいて概説する [9, 10, 11, 24, 25, 26]。統計力学ではベーテ近似は平均場近似の拡張版として知られている。



まず, 式 (71) で与えられた確率分布は次の形に書き換えることができる.

$$\Pr\{\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g}\} = \frac{1}{Z} \left( \prod_{(x,y) \in \Omega} W_{x,y}(f_{x,y}) \right) \times \left( \prod_{(x,y) \in \Omega} \frac{W_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})}{W_{x,y}(f_{x,y}) W_{x+1,y}(f_{x+1,y})} \right) \left( \prod_{(x,y) \in \Omega} \frac{W_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})}{W_{x,y}(f_{x,y}) W_{x,y+1}(f_{x,y+1})} \right) \quad (76)$$

$$\begin{aligned} E(\mathbf{f} | \mathbf{g}) &= - \sum_{(x,y) \in \Omega} \ln W_{x,y}(f_{x,y}) \\ &\quad - \sum_{(x,y) \in \Omega} \left( \ln W_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y}) - \ln W_{x,y}(f_{x,y}) - \ln W_{x+1,y}(f_{x+1,y}) \right) \\ &\quad - \sum_{(x,y) \in \Omega} \left( \ln W_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1}) - \ln W_{x,y}(f_{x,y}) - \ln W_{x,y+1}(f_{x,y+1}) \right) \end{aligned} \quad (77)$$

ここで  $Z$  は規格化定数であり,  $W_{x,y}(\xi)$  および  $W_{x,y}^{x',y'}(\xi, \xi')$  は次のように定義される.

$$W_{x,y}(\xi) \equiv \exp\left(-\frac{1}{2}\beta(\xi - g_{x,y})^2\right) \quad (78)$$

$$W_{x,y}^{x',y'}(\xi, \xi') \equiv \exp\left(-\frac{1}{2}\beta(\xi - g_{x,y})^2 - \frac{1}{2}\beta(\xi' - g_{x',y'})^2 - \frac{1}{2}\alpha(\xi - \xi')^2\right) \quad (79)$$

第 3 節の説明と同様にして式 (71) で与えられるギブス分布は次のような自由エネルギー-最小の変分原理を満足する.

$$\Pr\{\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g}\} = \min_Q \mathcal{F}[Q] \quad (80)$$

$$\mathcal{F}[Q] \equiv \mathcal{E}[Q] - \mathcal{S}[Q] \quad (81)$$

$$\mathcal{E}[Q] \equiv \sum_{\mathbf{z}} E(\mathbf{z} | \mathbf{g}) Q(\mathbf{z}), \quad \mathcal{S}[Q] \equiv - \sum_{\mathbf{z}} Q(\mathbf{z}) \ln Q(\mathbf{z}) \quad (82)$$

ここで, 試行関数  $Q(\mathbf{f})$  に対して

$$Q_{x,y}(f_{x,y}) \equiv \sum_{\mathbf{z}} \delta_{f_{x,y}, z_{x,y}} Q(\mathbf{z}) \quad (83)$$

$$Q_{x,y}^{x',y'}(f_{x,y}, f_{x',y'}) \equiv \sum_{\mathbf{z}} \delta_{f_{x,y}, z_{x,y}} \delta_{f_{x',y'}, z_{x',y'}} Q(\mathbf{z}) \quad (84)$$

を導入する.  $Q_{x,y}(f_{x,y})$  および  $Q_{x,y}^{x',y'}(f_{x,y}, f_{x',y'})$  は  $f(\mathbf{s}) = P(\mathbf{s} | \mathbf{h})$  であれば  $\Pr\{\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g}\}$  のある画素または画素対の周辺確率分布であることは容易に理解できる. 式 (72), 式 (83) および式 (84) を式 (82) に代入することにより

$$\begin{aligned} \mathcal{E}[Q] &= - \sum_{(x,y) \in \Omega} \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln(W_{x,y}(\zeta)) \\ &\quad - \sum_{(x,y) \in \Omega} \left( \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') \ln(W_{x,y}^{x+1,y}(\zeta, \zeta')) \right. \\ &\quad \quad \left. - \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln(W_{x,y}(\zeta)) - \sum_{\zeta = \pm 1} Q_{x+1,y}(\zeta) \ln(W_{x+1,y}(\zeta)) \right) \\ &\quad - \sum_{(x,y) \in \Omega} \left( \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') \ln(W_{x,y}^{x,y+1}(\zeta, \zeta')) \right. \\ &\quad \quad \left. - \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln(W_{x,y}(\zeta)) - \sum_{\zeta = \pm 1} Q_{x,y+1}(\zeta) \ln(W_{x,y+1}(\zeta)) \right) \end{aligned} \quad (85)$$

という式が導かれる. すなわち, これで汎関数  $\mathcal{E}[Q]$  は周辺確率分布  $Q_{x,y}(f_{x,y}), Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})$  および  $Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})$  のみにより表されたことになる. ここで,  $\mathcal{S}[Q]$  についても同様の書き換えができれば,  $\mathcal{F}[Q]$  も周辺確率分布  $Q_{x,y}(f_{x,y}), Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})$  および  $Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})$  のみにより表されたことになるが, 厳密にそのような書き換えができることは難しい. 式 (76) の形をもとにベータ近似では  $Q(f)$  をその周辺確率分布を用いて

$$Q(f) = \left( \prod_{(x,y) \in \Omega} Q_{x,y}(f_{x,y}) \right) \left( \prod_{(x,y) \in \Omega} \frac{Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})}{Q_{x,y}(f_{x,y})Q_{x+1,y}(f_{x+1,y})} \right) \times \left( \prod_{(x,y) \in \Omega} \frac{Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})}{Q_{x,y}(f_{x,y})Q_{x,y+1}(f_{x,y+1})} \right) \quad (86)$$

という形に表された関数系に制限してその範囲で自由エネルギーを最小にするという尺度をもって試行関数  $Q(f)$  ができるだけ式 (71) で与えられたギブス分布に近くなるように周辺確率分布を決定するという戦略がとられている. 式 (86) を式 (82) に代入することにより,  $\mathcal{S}[Q]$  を次のような近似的表式で置き換えることができる.

$$\mathcal{S}[Q] \simeq \sum_{(x,y) \in \Omega} \mathcal{S}_{x,y} + \sum_{(x,y) \in \Omega} \left( \mathcal{S}_{x,y}^{x+1,y} - \mathcal{S}_{x,y} - \mathcal{S}_{x+1,y} \right) + \sum_{(x,y) \in \Omega} \left( \mathcal{S}_{x,y}^{x,y+1} - \mathcal{S}_{x,y} - \mathcal{S}_{x,y+1} \right) \quad (87)$$

$$\mathcal{S}_{x,y} \equiv - \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln(Q_{x,y}(\zeta)) \quad (88)$$

$$\mathcal{S}_{x,y}^{x',y'} \equiv - \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} Q_{x,y}^{x',y'}(\zeta, \zeta') \ln(Q_{x,y}^{x',y'}(\zeta, \zeta')) \quad (89)$$

この表式はまず  $\mathcal{S}[Q]$  をより小数の画素からなる周辺確率分布を用いて表そうとすると, まず現れるのが各画素ごとの  $\mathcal{S}_{x,y}$  の和であると考えerことは自然であろう. もし  $\alpha = 0$  の場合を考えているのであればこの項のみで終わりである. ところが今考えている確率モデルでは最近接画素対間の相互作用  $\alpha$  が存在している. そこで最近接画素対についての周辺確率分布による項をこれに何らかの形で加える必要がある. 上式の場合, その最近接画素対からの寄与の項を, すべての最近接画素対から画素の寄与を差し引いた量  $\mathcal{S}_{x,y}^{x+1,y} - \mathcal{S}_{x,y} - \mathcal{S}_{x+1,y}$  および  $\mathcal{S}_{x,y}^{x,y+1} - \mathcal{S}_{x,y} - \mathcal{S}_{x,y+1}$  として考慮している. 画素の寄与をわざわざ引いているのは第 1 項で既に足しているから差し引かなければ足しすぎになるという考え方からくるものである. 式 (136) と式 (139) から  $\mathcal{F}[f]$  の近似表現は次のように与えられる.

$$\begin{aligned} \mathcal{F}[Q] &\simeq \mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \\ &\equiv \sum_{(x,y) \in \Omega} \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln\left(\frac{Q_{x,y}(\zeta)}{W_{x,y}(\zeta)}\right) \\ &\quad + \sum_{(x,y) \in \Omega} \left( \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') \ln\left(\frac{Q_{x,y}^{x+1,y}(\zeta, \zeta')}{W_{x,y}^{x+1,y}(\zeta, \zeta')}\right) \right. \\ &\quad \quad \left. - \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln\left(\frac{Q_{x,y}(\zeta)}{W_{x,y}(\zeta)}\right) - \sum_{\zeta = \pm 1} Q_{x+1,y}(\zeta) \ln\left(\frac{Q_{x+1,y}(\zeta)}{W_{x+1,y}(\zeta)}\right) \right) \\ &\quad + \sum_{(x,y) \in \Omega} \left( \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') \ln\left(\frac{Q_{x,y}^{x,y+1}(\zeta, \zeta')}{W_{x,y}^{x,y+1}(\zeta, \zeta')}\right) \right) \end{aligned}$$

$$- \sum_{\zeta=\pm 1} Q_{x,y}(\zeta) \ln\left(\frac{Q_{x,y}(\zeta)}{W_{x,y}(\zeta)}\right) - \sum_{\zeta=\pm 1} Q_{x,y+1}(\zeta) \ln\left(\frac{Q_{x,y+1}(\zeta)}{W_{x,y+1}(\zeta)}\right) \quad (90)$$

周辺確率分布の定義 (83), (84) から次の等式が導かれる.

$$\begin{aligned} Q_{x,y}(\zeta) &= \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') = \sum_{\zeta'=\pm 1} Q_{x-1,y}^{x,y}(\zeta', \zeta) \\ &= \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') = \sum_{\zeta'=\pm 1} Q_{x,y-1}^{x,y}(\zeta', \zeta) \quad (\zeta = \pm 1) \end{aligned} \quad (91)$$

そこで,  $Q_{x,y}(s_{x,y})$ ,  $Q_{x,y}^{x+1,y}(\zeta, \zeta')$ ,  $Q_{x,y}^{x,y+1}(\zeta, \zeta')$  はその規格化条件および等式 (91) を拘束条件として  $\mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}]$  の最小化に対する変分原理をとることにより決定される.

$$\begin{aligned} & \left\{ \widehat{Q}_{x,y}, \widehat{Q}_{x,y}^{x+1,y}, \widehat{Q}_{x,y}^{x,y+1} | (x,y) \in \Omega \right\} \\ &= \arg \min_{\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}} \left\{ \mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \right\} \\ & \left. \begin{aligned} Q_{x,y}(f_{x,y}) &= \sum_{\zeta=\pm 1} Q_{x,y}^{x+1,y}(f_{x,y}, \zeta) = \sum_{\zeta=\pm 1} Q_{x-1,y}^{x,y}(\zeta, f_{x,y}) \\ &= \sum_{\zeta=\pm 1} Q_{x,y}^{x,y+1}(f_{x,y}, \zeta) = \sum_{\zeta=\pm 1} Q_{x,y-1}^{x,y}(\zeta, f_{x,y}), \\ \sum_{\zeta=\pm 1} Q_{x,y}(\zeta) &= \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') = \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') = 1 \end{aligned} \right\} \quad (92) \end{aligned}$$

各拘束条件に対してラグランジュの未定係数を次のように導入する.

$$\begin{aligned} & \mathcal{L}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \\ &= \mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \\ & \quad - \sum_{(x,y) \in \Omega} \Lambda_{x,y}^{x+1,y}(\zeta) \left( Q_{x,y}(\zeta) - \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') \right) \\ & \quad - \sum_{(x,y) \in \Omega} \Lambda_{x,y}^{x-1,y}(\zeta) \left( Q_{x,y}(\zeta) - \sum_{\zeta'=\pm 1} Q_{x-1,y}^{x,y}(\zeta', \zeta) \right) \\ & \quad - \sum_{(x,y) \in \Omega} \Lambda_{x,y}^{x,y+1}(\zeta) \left( Q_{x,y}(\zeta) - \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') \right) \\ & \quad - \sum_{(x,y) \in \Omega} \Lambda_{x,y}^{x,y-1}(\zeta) \left( Q_{x,y}(\zeta) - \sum_{\zeta'=\pm 1} Q_{x,y-1}^{x,y}(\zeta', \zeta) \right) \\ & \quad - \sum_{(x,y) \in \Omega} \nu_{x,y} \left( \sum_{\zeta=\pm 1} Q_{x,y}(\zeta) - 1 \right) \\ & \quad - \sum_{(x,y) \in \Omega} \nu_{x,y}^{x+1,y} \left( \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') - 1 \right) \\ & \quad - \sum_{(x,y) \in \Omega} \nu_{x,y}^{x,y+1} \left( \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') - 1 \right) \end{aligned} \quad (93)$$

$\mathcal{L}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}]$  の極値条件から  $Q_{x,y}(\xi)$ ,  $Q_{x,y}^{x+1,y}(\xi, \xi')$ ,  $Q_{x,y}^{x,y+1}(\xi, \xi')$  に

対する表式が拘束条件に対して導入されたラグランジュの未定係数を用いて表される.

$$Q_{x,y}(\xi) = \frac{W_{x,y}(\xi) \prod_{(x',y') \in \mathbf{c}_{x,y}} \exp\left(\frac{1}{|\mathbf{c}_{x,y}|-1} \Lambda_{x',y'}^{x',y'}(\xi)\right)}{\sum_{\zeta=\pm 1} W_{x,y}(\zeta) \prod_{(x',y') \in \mathbf{c}_{x,y}} \exp\left(\frac{1}{|\mathbf{c}_{x,y}|-1} \Lambda_{x',y'}^{x',y'}(\zeta)\right)}, \quad (94)$$

$$Q_{x,y}^{x',y'}(\xi, \xi') = \frac{W_{x,y}^{x',y'}(\xi, \xi') \exp\left(\Lambda_{x,y}^{x',y'}(\xi)\right) \exp\left(\Lambda_{x',y'}^{x,y}(\xi')\right)}{\sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} W_{x,y}^{x',y'}(\zeta, \zeta') \exp\left(\Lambda_{x,y}^{x',y'}(\zeta)\right) \exp\left(\Lambda_{x',y'}^{x,y}(\zeta')\right)}, \quad (95)$$

ここで  $\{\nu_{x,y}, \nu_{x,y}^{x',y'} | (x,y) \in \Omega, (x',y') \in \mathbf{c}_{x,y}\}$  は規格化定数から決定されている. 残ったラグランジュの未定係数  $\{\Lambda_{x,y}^{x',y'}(\zeta), \Lambda_{x',y'}^{x,y}(\zeta') | (x,y) \in \Omega, (x',y') \in \mathbf{c}_{x,y}, \zeta \in \{\pm 1\}\}$  は周辺確率分布についての規格化条件および等式 (91) を満たすように決定されるという形の非線形方程式に帰着される. このラグランジュの未定係数から変数変換

$$\exp\left(\Lambda_{x,y}^{x',y'}(\xi)\right) = \prod_{(x'',y'') \in \mathbf{c}_{x,y} \setminus \{(x',y')\}} \mathcal{M}_{x,y}^{x'',y''}(\xi) \quad (96)$$

により, 新たな未知変数  $\mathcal{M}_{x,y}^{x',y'}(\xi)$  を導入することにより, 周辺確率分布  $Q_{x,y}(\xi)$ ,  $Q_{x,y}^{x+1,y}(\xi, \xi')$ ,  $Q_{x,y}^{x,y+1}(\xi, \xi')$  は次のようなより見やすい形に整理される.

$$Q_{x,y}(\xi) = \frac{W_{x,y}(\xi) \prod_{(x',y') \in \mathbf{c}_{x,y}} \mathcal{M}_{x,y}^{x',y'}(\xi)}{\sum_{\zeta=\pm 1} W_{x,y}(\zeta) \prod_{(x',y') \in \mathbf{c}_{x,y}} \mathcal{M}_{x,y}^{x',y'}(\zeta)}, \quad (97)$$

$$Q_{x,y}^{x',y'}(\xi, \xi') = \frac{W_{x,y}^{x',y'}(\xi, \xi') \prod_{(x'',y'') \in \mathbf{c}_{x,y} \setminus \{(x',y')\}} \mathcal{M}_{x,y}^{x'',y''}(\xi) \prod_{(x''',y''') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} \mathcal{M}_{x',y'}^{x''',y'''}(\xi')}{\sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} W_{x,y}^{x',y'}(\zeta, \zeta') \prod_{(x'',y'') \in \mathbf{c}_{x,y} \setminus \{(x',y')\}} \mathcal{M}_{x,y}^{x'',y''}(\zeta) \prod_{(x''',y''') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} \mathcal{M}_{x',y'}^{x''',y'''}(\zeta')}, \quad (98)$$

$$\mathcal{M}_{x,y}^{x',y'}(\xi) = \frac{\sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\xi, \zeta)}{W_{x,y}(\xi)}\right) \prod_{(x'',y'') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} \mathcal{M}_{x',y'}^{x'',y''}(\zeta)}{\sum_{\zeta'=\pm 1} \sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\zeta', \zeta)}{W_{x,y}(\zeta')}\right) \prod_{(x'',y'') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} \mathcal{M}_{x',y'}^{x'',y''}(\zeta)}. \quad (99)$$

$$\mathbf{c}_{x,y} \equiv \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\} \quad (100)$$

ここで式 (99) は式 (97) および式 (98) を等式 (91) に代入することにより得られた  $\{\mathcal{M}_{x,y}^{x',y'}(\xi)\}$  の決定方程式である. これらの式は一見複雑に見えるが実はよく見るときれいな構造を持っている. まず式 (97) で与えられた周辺確率分布は画素  $(x, y)$  にその隣接画素からの影響を  $\mathcal{M}_{x,y}^{x',y'}$  ( $(x', y') \in \mathbf{c}_{x,y}$ ) という形で取り込んだ構造をしている. これをグラフ表現にしたのが図 6 である.

式 (98) で与えられた周辺確率分布は例えば  $f_{x,y}^{x+1,y}(\zeta, \zeta')$  の場合, 画素  $(x, y)$  には  $(x+1, y)$  を除く隣接画素からの影響を  $\mathcal{M}_{x',y'}^{x,y} ((x', y') \in c_{x,y} \setminus \{(x+1, y)\})$  という形で取り込んだ構造をしている. また, 画素  $(x+1, y)$  には  $(x, y)$  を除く隣接画素からの影響を  $\mathcal{M}_{x',y'}^{x,y} ((x', y') \in c_{x,y} \setminus \{(x, y)\})$  という形で取り込んだ構造をしている. これをグラフ表現にしたのが図 7 である. そして, 式 (102) は例えば  $(x', y') = (x+1, y)$  と設定した場合, 画素  $(x, y)$  に  $(x+1, y)$  を除く隣接画素から  $\mathcal{M}_{x',y'}^{x,y} ((x', y') \in c_{x,y} \setminus \{(x+1, y)\})$  が入力として入り,  $(x+1, y)$  へと  $\mathcal{M}_{x',y'}^{x,y}$  が出力として伝搬されてゆく構造を持っている. これをグラフ表現にしたのが図 8 である.

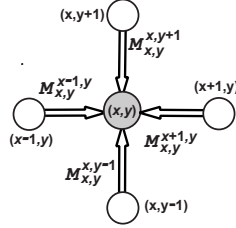


図 6: 式 (97) で与えられた周辺確率分布  $f_{x,y}(\zeta)$  のグラフ表現.

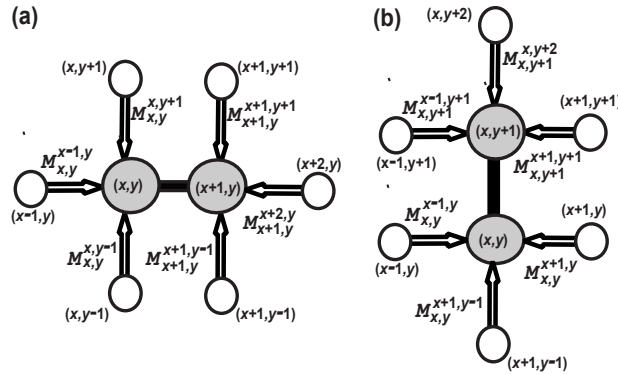


図 7: 式 (98) で与えられた周辺確率分布  $f_{x,y}^{x+1,y}(\zeta, \zeta')$  および  $f_{x,y}^{x,y+1}(\zeta, \zeta')$  のグラフ表現. (a)  $(x', y') = (x+1, y)$ . (b)  $(x', y') = (x, y+1)$ .

式 (83)-(144) において,  $\mathcal{M}_{x',y'}^{x'',y''}(\zeta)$  の変数  $\zeta$  は 0, 1 しかとらないことを考慮すると

$$\mathcal{M}_{x',y'}^{x'',y''}(-1) = \frac{\exp(-\lambda_{x',y'}^{x'',y''})}{2\cosh(\lambda_{x',y'}^{x'',y''})}, \quad \mathcal{M}_{x',y'}^{x'',y''}(1) = \frac{\exp(\lambda_{x',y'}^{x'',y''})}{2\cosh(\lambda_{x',y'}^{x'',y''})} \quad (101)$$

と変数変換することにより, 特に式 (144) は次のように書き換えられる.

$$\lambda_{x',y'}^{x'',y''} = \operatorname{arctanh} \left( \tanh(\alpha) \tanh \left( \beta h_{x,y} + \sum_{(x'',y'') \in c_{x,y} \setminus \{(x',y')\}} \lambda_{x,y}^{x'',y''} \right) \right) \quad (102)$$

本節で紹介したベーテ近似は式 (139) が重要な鍵になっている. そこでは最近接画素対の相関の効果までを重要視し, それより大きな画素の集合の効果は小さいとして無視している.

修復画像  $\hat{f} = \{\hat{f}_{x,y} | (x, y) \in \Omega\}$  は式 (74) により決定される. 周辺確率分布  $\Pr\{F_{x,y} = f_{x,y} | G = g\}$  は式 (97)-(99) を解き, 画素  $(x, y)$  のおける周辺確率分布  $Q_{x,y}(f_{x,y})$  を得ることにより近似的に計

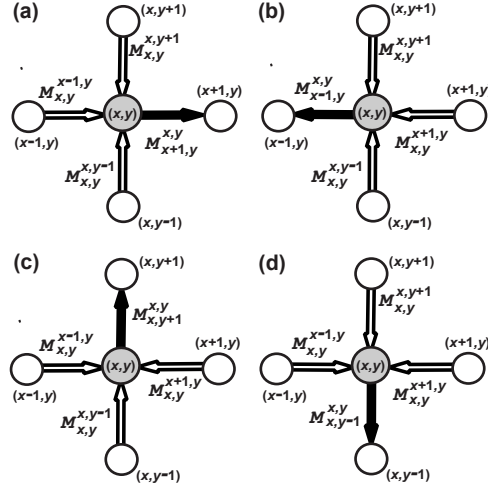


図 8: 式 (102) で与えられた反復方程式のグラフ表現. (a)  $(x', y') = (x + 1, y)$ . (b)  $(x', y') = (x - 1, y)$ . (c)  $(x', y') = (x, y + 1)$ . (d)  $(x', y') = (x, y - 1)$ .

算される.  $\alpha$  と  $p$  の値をひとつ固定したときのベータ近似を用いた画像修復のための反復計算アルゴリズムは以下に与えられる.

**ベータ近似を用いた画像修復アルゴリズム**

**Step 1:** 原画像  $f = \{f_{x,y} | (x,y) \in \Omega\}$ , 劣化画像  $g = \{g_{x,y} | (x,y) \in \Omega\}$  と  $p, \alpha, R, \Delta$  の値を設定する.

**Step 2:** 初期値を  $M_{x,y}^{x',y'}(\xi) = 0$  ( $(x,y) \in \Omega, (x',y') \in \mathbf{c}_{x,y}, \xi \in \{\pm 1\}$ ),  $T \leftarrow 1 + R\Delta$  と設定する.

**Step 3:**  $T \leftarrow T - \Delta, r \leftarrow 0$  と設定する.

**Step 4:**  $r \leftarrow r + 1, a_{x,y}^{x',y'}(\zeta) \leftarrow M_{x,y}^{x',y'}(\zeta)$  ( $(x,y) \in \Omega, (x',y') \in \mathbf{c}_{x,y}, \zeta \in \{\pm 1\}$ ) と設定する.

**Step 5:**  $\{M_{x,y}^{x+1,y}(\xi), M_{x,y}^{x,y+1}(\xi) | (x,y) \in \Omega, \xi \in \{\pm 1\}\}$  の値を次の更新則により更新する.

$$M_{x,y}^{x',y'}(\xi) \leftarrow \frac{\sum_{\zeta=\pm 1} \left( \frac{W_{x,y}^{x',y'}(\xi, \zeta)}{W_{x,y}(\xi)} \right)^{1/T} \prod_{(x'',y'') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} a_{x',y'}^{x'',y''}(\zeta)}{\sum_{\zeta'=\pm 1} \sum_{\zeta=\pm 1} \left( \frac{W_{x,y}^{x',y'}(\zeta', \zeta)}{W_{x,y}(\zeta')} \right)^{1/T} \prod_{(x'',y'') \in \mathbf{c}_{x',y'} \setminus \{(x,y)\}} a_{x',y'}^{x'',y''}(\zeta)} \quad ((x',y') \in \mathbf{c}_{x,y}) \quad (103)$$

**Step 6:**

$$\frac{1}{MN} \sum_{(x,y) \in \Omega} \sum_{(x',y') \in \mathbf{c}_{x,y}} \sum_{\zeta=\pm 1} |a_{x,y}^{x',y'}(\zeta) - M_{x,y}^{x',y'}(\zeta)| < 10^{-6} \quad (104)$$

を満足しなければ Step 4 にもどり, 満足すれば Step 7 に進む.

**Step 7:**  $T = 1$  を満足しなければ Step 3 にもどり, 満足すれば修復画像  $\hat{f} = \{\hat{f}_{x,y} | (x,y) \in \Omega\}$  を

$$\hat{f}_{x,y} \leftarrow \arg \max_{\zeta=\pm 1} \left( W_{x,y}(\xi) \prod_{(x',y') \in \mathbf{c}_{x,y}} M_{x,y}^{x',y'}(\xi) \right) \quad (105)$$

により決定し, 終了する.

数値実験を行った一例を図 9 及び図 10 に与える. 図 9(a) の原画像は  $\alpha = 0.5$  と設定し, 確率分布 (70) に従うモンテカルロ・シミュレーションにより生成された 2 値画像である. すなわち図 9 の数値実験は前節で与えたベイズ的アプローチの枠組みの順過程に忠実に従って生成されたデー

タである劣化画像からの原画像の推定を行った結果である。修復の際には  $\alpha$  は原画像を生成したときの値を用い、 $p$  も劣化させたときに設定した値をそのまま用いている。図 10(a) の原画像は信号処理の分野で標準的に用いられている “Mandrill” と呼ばれる画像に対してある種のしきい値処理を行って作成した画像である。すなわち、図 10 は現実世界の画像をもとに生成された 2 値画像に対して行った数値実験であり、必ずしもデータがベイズ的アプローチの枠組みの順過程に従って生成されているとは限らない場合である。修復の際には  $p$  は劣化させたときに設定した値をそのまま用いているが、 $\alpha$  は、そもそも原画像自体が式 (70) を使って生成されたものではないため、手動で修復画像が原画像に近くなるように調整して決めている。いずれの場合も良好な結果が得られることがわかる。

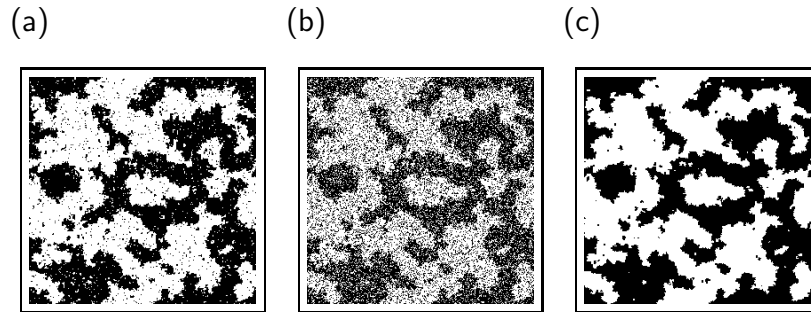


図 9: 2 値画像の画像修復例 (a) 原画像 ( $\alpha = 0.5$  と設定し、確率分布 (70) に従うモンテカルロ・シミュレーションにより生成された 2 値画像). (b) 劣化画像 ( $p = 0.2$  と設定し、劣化過程 (69) を通して、生成された 2 値画像). (c) ベーテ近似による修復画像 ( $p = 0.2, \alpha = 0.5$ )

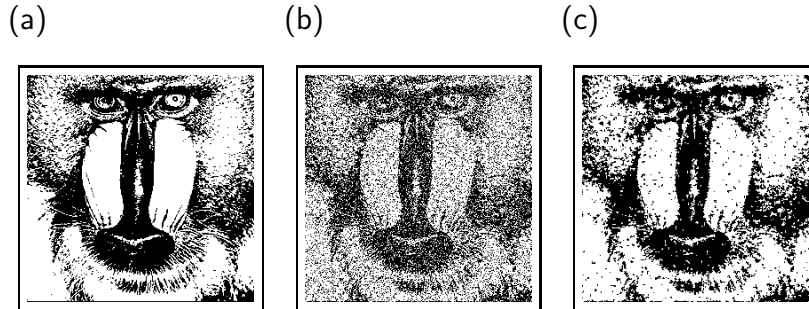


図 10: 2 値画像の画像修復例 (a) 原画像. (b) 劣化画像 ( $p = 0.2$ ). (c) ベーテ近似による修復画像 ( $p = 0.2, \alpha = 0.35$ )

ここまでの話の中で  $p$  と  $\alpha$  を劣化画像から推定することについては述べて来なかったが、実際の画像処理システムにおいてはこの  $p$  および  $\alpha$  を与えられたデータのみから推定しなければならない場合が多い。原画像が式 (70) における  $\alpha$  のどのようなモデルパラメータ  $\alpha$  と  $p$  の推定値  $\hat{\alpha}$ 、 $\hat{p}$  はベイズ統計においてはハイパラメータと呼ばれ、周辺尤度最大化とよばれる最尤推定の立場において次のように決定されることが多い [9, 11].

$$(\hat{\alpha}, \hat{p}) = \max_{(\alpha, p)} \sum_{\mathbf{z}} \Pr\{\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{z}\} \Pr\{\mathbf{F} = \mathbf{z}\} \quad (106)$$

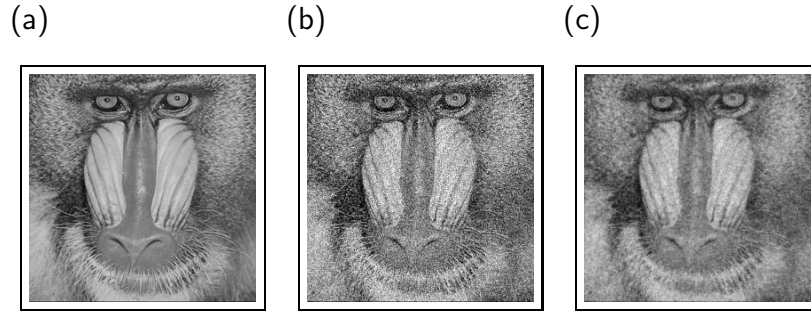


図 11: 濃淡画像の画像修復例 (a) 原画像. (b) 加法的白色ガウス雑音による劣化画像 ( $\mathcal{N}[0, 30^2]$ ). (c) ペーテ近似による修復画像 ( $\sigma = 30, \alpha = 0.0005$ )

右辺の量  $\Pr\{G = g\} \equiv \sum_z \Pr\{G = g|F = z\}\Pr\{F = z\}$  が周辺尤度と呼ばれる量であり, これもペーテ近似を用いて計算することができる.

本節と前節で与えた枠組みは各画素の階調値が 0, 1 ではなく, 任意の実数値をとる場合にも容易に拡張できる. 例えば劣化過程として平均 0, 分散  $\sigma^2 = \frac{1}{2\beta}$  の加法的白色ガウス雑音  $\mathcal{N}[0, \sigma^2]$  を考える. この場合, 劣化過程は

$$\Pr\{G = g|F = f\} \equiv \prod_{(x,y) \in \Omega} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(f_{x,y} - g_{x,y})^2\right) \quad (107)$$

という形に与えられる. 計算の詳細は省略するが  $\sigma = 20$  の場合の画像修復例を図 11 に与える.

本節で紹介した 2 値画像の画像処理アルゴリズムの C 言語を用いた具体的プログラムはホームページ (URL: <http://www.statp.is.tohoku.ac.jp/~kazu/SMAPIP-KazuKazu/>) からダウンロードすることができる.

## 8 確率的推論

本節では統計力学を用いた確率的情報処理のもう一つの例として人工知能における確率的推論をとりあげる. ベイズ統計にもとづいて構成された推論機構は一般にベイジアンネットと呼ばれている. 確率推論の代表的な教科書としては Pearl の教科書等があげられる [3]. また, 最近の発展については文献 [2] 等を参照していただきたい. ここで, 統計力学的手法の一つとして知られるクラスター変分法を用いた確率的推論アルゴリズムについて解説する. クラスター変分法は前節のペーテ近似を更に一般的に拡張した統計力学的手法である.

確率的推論は複数のノードと呼ばれる点とそのノード間の因果関係を表すリンクからグラフ表現の形で与えられる. そこでまず, 例として 4 個のノードを考え, 図 12 に示すようなグラフ表現で与えられる確率的推論機構を用いて, どのようなものかを初心者にわかりやすく解説しよう. この例題は K. Murphy 氏によって作成されたホームページ <http://www.ai.mit.edu/~murphyk/Bayes/bayes.html> にベイジアンネットの解説とともに掲載されているものである. その 4 個のノードに対してそれぞれに [曇りである], [スプリンクラーが作動している], [雨が降る], [芝生がぬれている] という命題を割り当て, そのそれぞれの命題に対応して true (T) と false (F) という 2 つの状態をとる確率変数



をそれぞれ  $A_C, A_S, A_R, A_W$  により表すものとする. 結合確率と条件付き確率の関係式により

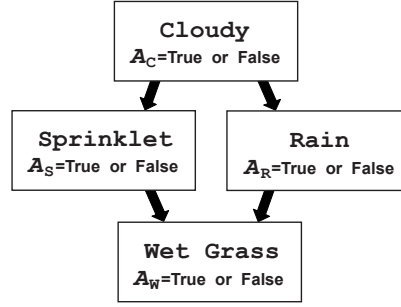


図 12: 4 個のノードからなる確率的推論機構の例.

$$\begin{aligned}
 & \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} \\
 &\quad \times \Pr\{A_C = a_C, A_S = a_S\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} \\
 &\quad \times \Pr\{A_S = a_S | A_C = a_C\} \Pr\{A_C = a_C\} \tag{108}
 \end{aligned}$$

が導かれる. ここで, 各命題に対する因果関係として「曇りである」とすればいずれ「雨が降る」可能性はある. 「曇りでない」とすれば晴れているということで芝生を枯らさないために「スプリンクラーを作動させる」ということになる. そして「雨が降る」かまたは「スプリンクラーを作動させる」ということにより「芝生がぬれている」という状況が発生する.  $\Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\}$  は  $A_C$  すなわち「曇っているかどうか」には直接は依存せず,  $\Pr\{A_R = a_R | A_C = a_C, A_S = a_S\}$  は  $S$  すなわち「スプリンクラーが作動しているかどうか」には直接依存しないことを考慮するとそれぞれ

$$\begin{aligned}
 & \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} = \Pr\{A_W = a_W | A_S = a_S, A_R = a_R\} \\
 & \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} = \Pr\{A_R = a_R | A_C = a_C\} \tag{109}
 \end{aligned}$$

と与えられる. すなわち, 結合確率  $\Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\}$  は次のように与えられる.

$$\begin{aligned}
 & \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \\
 &= \Pr\{A_W = a_A | A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C\} \\
 &\quad \times \Pr\{A_S = a_S | A_C = a_C\} \Pr\{A_C = a_C\} \tag{110}
 \end{aligned}$$

式 (110) における  $\Pr\{A_W = a_A | A_S = a_S, A_R = a_R\}$ ,  $\Pr\{A_R = a_R | A_C = a_C\}$ ,  $\Pr\{A_S = a_S | A_C = a_C\}$ ,  $\Pr\{A_C = a_C\}$  の値は表 1 に与える. さて, ここで「芝生がぬれている」のは一体何が原因な

表 1: 式 (110) における  $\Pr\{A_W = a_A | A_S = a_S, A_R = a_R\}$ ,  $\Pr\{A_R = a_R | A_C = a_C\}$ ,  $\Pr\{A_S = a_S | A_C = a_C\}$ ,  $\Pr\{A_C = a_C\}$  の値. T は true, F は false をそれぞれ表す.

$a_C$	$\Pr\{A_C = a_C\}$
T	0.5
F	0.5

$a_S$	$\Pr\{A_S = a_S   A_C = T\}$	$\Pr\{A_S = a_S   A_C = F\}$
T	0.1	0.5
F	0.9	0.5

$a_R$	$\Pr\{A_R = a_R   A_C = T\}$	$\Pr\{A_R = a_R   A_C = F\}$
T	0.8	0.2
F	0.2	0.8

$a_W$	$\Pr\{A_W = a_W   A_S = T, A_R = T\}$	$\Pr\{A_W = a_W   A_S = T, A_R = F\}$
T	0.99	0.9
F	0.01	0.1
$a_W$	$\Pr\{A_W = a_W   A_S = F, A_R = T\}$	$\Pr\{A_W = a_W   A_S = F, A_R = F\}$
T	0.9	0
F	0.1	1

のかを推定したいとする. この場合,

$$\Pr\{A_R = a_R | A_W = a_W\} = \frac{\Pr\{A_R = a_R, A_W = a_W\}}{\Pr\{A_W = a_W\}}$$

$$\Pr\{A_S = a_S | A_W = a_W\} = \frac{\Pr\{A_S = a_S, A_W = a_W\}}{\Pr\{A_W = a_W\}} \quad (111)$$

を計算して比較することになる. これらの条件付き確率は次の周辺確率を計算することにより得られる.

$$\Pr\{A_R = a_R, A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_S=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (112)$$

$$\Pr\{A_S = a_S, A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_R=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (113)$$

$$\Pr\{A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_S=T, F} \sum_{a_R=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (114)$$

今の場合,  $\Pr\{A_S = \text{true}, A_W = \text{ture}\} = 0.2781$ ,  $\Pr\{A_R = \text{true}, A_W = \text{true}\} = 0.4581$ ,  $\Pr\{A_W = \text{true}\} = 0.6471$  なので

$$\Pr\{A_R = \text{true} | A_W = \text{true}\} = \frac{\Pr\{a_R = \text{true}, A_W = \text{true}\}}{\Pr\{A_W = \text{true}\}} = \frac{0.4581}{0.6471} = 0.7079$$

$$\Pr\{A_S = \text{true} | A_W = \text{true}\} = \frac{\Pr\{A_S = \text{true}, A_W = \text{true}\}}{\Pr\{A_W = \text{true}\}} = \frac{0.2781}{0.6471} = 0.4298 \quad (115)$$

これにより、「芝生がぬれている」のは「雨が降った」ためだったということが推論できるわけである。

もう少し複雑な推論機構として図 13 に示す推論機構を考えよう。この例題はベイジアンネットの標準的な例題を掲載したホームページ <http://www.norsys.com/networklibrary.html> に「Asia」という名前で掲載されているものであり、もともとの出典は文献 [4] とされている。この推論機構は

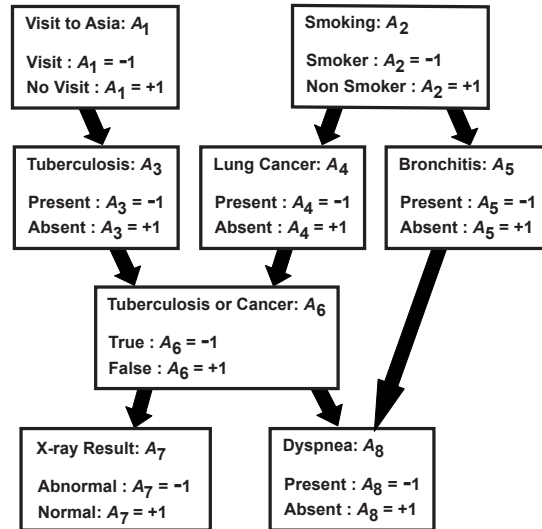


図 13: 確率的推論機構の例.

8 個のノードから構成されている。各ノードに 1 番から 8 番までの通し番号をふる。これにより図 13 の確率的推論機構は図 14 のようなグラフ表現として与えられる。各ノードには前の例と同じよ

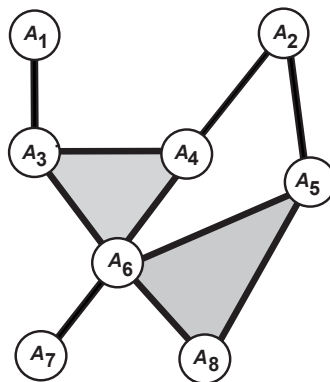


図 14: 図 13 の確率的推論機構のグラフ表現.

うに “+1” と “-1” という 2 つの状態が割り当てられている。この時、結合確率分布はベイズ規則をもとに次のように構成される。

$$\begin{aligned}
 & \Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\} \\
 &= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &\quad \times \Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\}
 \end{aligned}$$

$$\begin{aligned}
&= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
&\quad \times \Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
&\quad \times \Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
&= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
&\quad \times \Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
&\quad \times \Pr\{A_3 = a_3, A_4 = a_4 | A_1 = a_1, A_2 = a_2\} \\
&\quad \times \Pr\{A_1 = a_1, A_2 = a_2\} \tag{116}
\end{aligned}$$

最後の表式に見られる条件付き確率は図 14 では

$$\left\{ \begin{array}{l} \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\ \quad = V_{56 \rightarrow 8}(a_8 | a_5, a_6) V_{6 \rightarrow 7}(a_7 | a_6) \\ \Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\ \quad = V_{2 \rightarrow 5}(a_5 | a_2) V_{34 \rightarrow 6}(a_6 | a_3, a_4) \\ \Pr\{A_3 = a_3, A_4 = a_4 | A_1 = a_1, A_2 = a_2\} = V_{1 \rightarrow 3}(a_3 | a_1) V_{2 \rightarrow 4}(a_4 | a_2) \\ \Pr\{A_1 = a_1, A_2 = a_2\} = V_1(a_1) V_2(a_2) \end{array} \right. \tag{117}$$

という形に与えられる。これにより式 (116) に与えられた結合確率分布は次のような形に書き換えられる。

$$\begin{aligned}
&\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\} \\
&= V_{56 \rightarrow 8}(a_8 | a_5, a_6) V_{6 \rightarrow 7}(a_7 | a_6) V_{34 \rightarrow 6}(a_6 | a_3, a_4) \\
&\quad \times V_{2 \rightarrow 5}(a_5 | a_2) V_{2 \rightarrow 4}(a_4 | a_2) V_{1 \rightarrow 3}(a_3 | a_1) V_2(a_2) V_1(a_1) \tag{118}
\end{aligned}$$

ここで、新たに

$$\left\{ \begin{array}{l} W_{568}(a_5, a_6, a_8) = V_{56 \rightarrow 8}(a_8 | a_5, a_6) \quad W_{67}(a_6, a_7) = V_{6 \rightarrow 7}(a_7 | a_6) \\ W_{346}(a_3, a_4, a_6) = V_{34 \rightarrow 6}(a_6 | a_3, a_4) \quad W_{25}(a_2, a_5) = V_{2 \rightarrow 5}(a_5 | a_2) V_2(a_2) \\ W_{24}(a_2, a_4) = V_{2 \rightarrow 4}(a_4 | a_2) V_2(a_2) \quad W_{13}(a_1, a_3) = V_{1 \rightarrow 3}(a_3 | a_1) V_1(a_1) \\ W_2(a_2) = V_2(a_2) \quad W_3(a_3) = W_4(a_4) = W_5(a_5) = W_6(a_6) = 1 \end{array} \right. \tag{119}$$

という因子を導入し、更に式 (118) は

$$\begin{aligned}
&\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\} \\
&= \frac{1}{Z} \left( \frac{W_{568}(a_5, a_6, a_8) W_{67}(a_6, a_7) W_{346}(a_3, a_4, a_6) W_{25}(a_2, a_5) W_{24}(a_2, a_4) W_{13}(a_1, a_3)}{W_2(a_2)^2 W_3(a_3) W_4(a_4) W_5(a_5) W_6(a_6)} \right) \tag{120}
\end{aligned}$$

という形に書き直される。ここで、 $Z$  は規格化定数である。更に新しい記号として

$$\mathbf{C} \equiv \{568, 67, 346, 25, 24, 13, 2, 3, 4, 5, 6\} \tag{121}$$

$$\left\{ \begin{array}{l} \mu(568) = \mu(67) = \mu(346) = \mu(25) = \mu(25) = \mu(24) = \mu(13) = -1 \\ \mu(2) = 2, \quad \mu(3) = \mu(4) = \mu(5) = \mu(6) = 1 \end{array} \right. , \tag{122}$$

$$\mathbf{A}_\alpha \equiv \{A_i | i \in \alpha\}, \quad \mathbf{a}_\alpha \equiv \{a_i | i \in \alpha\} \quad (\alpha \in \mathbf{C}) \tag{123}$$

を導入する.  $C$  の各要素はそれぞれがノードの集合であり,

$$\begin{aligned} 2 < 24, 2 < 25, 3 < 13, 3 < 346, \\ 4 < 24, 4 < 346, 5 < 25, 5 < 568, \\ 6 < 346, 6 < 67, 6 < 568, \end{aligned} \quad (124)$$

という包含関係にもとづく順序関係が存在し,  $C$  は半順序集合となっている.  $i < \alpha$  はノード  $i$  が 2 個以上のノードからなる集合  $\alpha$  の要素の一つであることを示している. すなわち  $i$  は  $\alpha$  の新部分集合となっているということもできる. ここで定義されたノードの集合  $\alpha$  はクラスターとも呼ばれる. これらの記号を用いると式 (118) は

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \frac{\prod_{\alpha \in C} W_{\alpha}(\mathbf{a}_{\alpha})^{-\mu(\alpha)}}{\sum_{\mathbf{z}} \prod_{\alpha \in C} W_{\alpha}(\mathbf{z}_{\alpha})^{-\mu(\alpha)}} \quad (125)$$

$$\sum_{\mathbf{z}} \equiv \prod_{i=1}^8 \sum_{z_i = \pm 1} \quad (126)$$

というコンパクトな形にまとめられる.

$W_{\alpha}(\mathbf{a}_{\alpha})$  と  $\mu(\alpha)$  から関数  $E(\mathbf{a})$ :

$$E(\mathbf{a}) \equiv - \sum_{\alpha \in C} \mu(\alpha) \ln(W_{\alpha}(\mathbf{a}_{\alpha})) \quad (127)$$

を導入すると,  $\Pr\{\mathbf{A} = \mathbf{a}\}$  はギブス分布

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \frac{\exp(-E(\mathbf{a}))}{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))} \quad (128)$$

として表現されることから,  $E(\mathbf{a})$  はハミルトニアンとなっていることがわかる. 第 3 節の説明と同様にして  $\Pr\{\mathbf{A} = \mathbf{a}\}$  は次のような形の自由エネルギー最小の変分原理を満たすことが確かめられる.

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \min_Q \mathcal{F}[Q] \quad (129)$$

$$\mathcal{F}[Q] \equiv \mathcal{E}[Q] - \mathcal{S}[Q] \quad (130)$$

$$\mathcal{E}[Q] \equiv \sum_{\mathbf{z}} E(\mathbf{z}) Q(\mathbf{z}), \quad \mathcal{S}[Q] \equiv - \sum_{\mathbf{z}} Q(\mathbf{z}) \ln Q(\mathbf{z}) \quad (131)$$

ここで, 試行関数としての確率分布  $Q(\mathbf{f})$  に対して周辺確率分布

$$Q_{\alpha}(\mathbf{a}_{\alpha}) \equiv \sum_{\mathbf{z}} \left( \prod_{\{i | i \in \alpha\}} \delta_{a_i, z_i} \right) Q(\mathbf{z}) \quad (132)$$

を導入すると,

$$\sum_{\mathbf{z}_{\alpha}} Q_{\alpha}(\mathbf{z}_{\alpha}) = 1 \quad (\alpha \in C) \quad (133)$$

$$Q_\alpha(\mathbf{a}_\alpha) = \sum_{\mathbf{z}_\gamma} \left( \prod_{\{i|i \in \alpha\}} \delta_{a_i, z_i} \right) Q_\gamma(\mathbf{z}_\gamma) \quad (\alpha < \gamma \in C) \quad (134)$$

$$\sum_{\mathbf{z}_\alpha} \equiv \prod_{\{i|i \in \alpha\}} \sum_{z_i = \pm 1} \quad (135)$$

が成り立つことは容易に確かめられる。更に、式 (132) を用いると  $\mathcal{E}[Q]$  は周辺確率分布  $Q_\alpha(\mathbf{a}_\alpha)$  ( $\alpha \in C$ ) だけを用いて

$$\mathcal{E}[Q] = \sum_{\alpha \in C} \mu(\alpha) \sum_{\mathbf{z}_\alpha} Q_\alpha(\mathbf{z}_\alpha) \ln(W_\alpha(\mathbf{z}_\alpha)) \quad (136)$$

という形に表される。

クラスター変分法においては  $Q(\mathbf{a})$  を

$$Q(\mathbf{a}) = \frac{Q_{568}(a_5, a_6, a_8) Q_{67}(a_6, a_7) Q_{346}(a_3, a_4, a_6) Q_{25}(a_2, a_5) Q_{24}(a_2, a_4) Q_{13}(a_1, a_3)}{Q_2(a_2)^2 Q_3(a_3) Q_4(a_4) Q_5(a_5) Q_6(a_6)} \quad (137)$$

すなわち

$$Q(\mathbf{a}) = \prod_{\alpha \in C} Q_\alpha(\mathbf{a}_\alpha)^{-\mu(\alpha)} \quad (138)$$

という形に近似的に制限し、式 (133) および式 (134) を拘束条件として自由エネルギー最小の変分原理 (129) のもとで決定された  $Q_\alpha(\mathbf{a}_\alpha) = \hat{Q}_\alpha(\mathbf{a}_\alpha)$  が周辺確率  $\Pr\{A_\alpha = \mathbf{a}_\alpha\}$  の近似として採用される。エントロピー  $S[Q]$  は式 (132) および式 (138) から次のように書き直される。

$$S[Q] = \sum_{\alpha \in C} \mu(\alpha) \sum_{\mathbf{z}_\alpha} Q_\alpha(\mathbf{z}_\alpha) \ln(Q_\alpha(\mathbf{z}_\alpha)) \quad (139)$$

従って自由エネルギー (130) は次のように書き下される。

$$\mathcal{F}[Q] \simeq \mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}] \equiv \sum_{\alpha \in C} \mu(\alpha) \sum_{\mathbf{z}_\alpha} Q_\alpha(\mathbf{z}_\alpha) \ln\left(\frac{Q_\alpha(\mathbf{z}_\alpha)}{Q_\alpha(\mathbf{z}_\alpha)}\right) \quad (140)$$

$\mathcal{F}[Q]$  を最小化する  $\{Q_\alpha | \alpha \in C\}$  は近似自由エネルギー  $\mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}]$  の条件付き変分原理として次のように与えられる。

$$\begin{aligned} \{\hat{Q}_\alpha | \alpha \in C\} &\simeq \arg \min_{\{Q_\alpha | \alpha \in C\}} \left\{ \mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}] \middle| \sum_{\mathbf{z}_\alpha} Q_\alpha(\mathbf{z}_\alpha) = 1 \quad (\alpha \in C), \right. \\ &\quad \left. Q_\alpha(\mathbf{a}_\alpha) = \sum_{\mathbf{z}_\gamma} \left( \prod_{\{i|i \in \alpha\}} \delta_{a_i, z_i} \right) Q_\gamma(\mathbf{z}_\gamma) \quad (\alpha < \gamma \in C) \right\}. \quad (141) \end{aligned}$$

各拘束条件に対してラグランジュの未定係数を導入し、極値の条件を求め、その後、導入したラグランジュの未定係数を拘束条件を満たすように決めるという手順において、 $\{\hat{Q}_\alpha | \alpha \in C\}$  に対する決定方程式は次のように与えられる。

$$\hat{Q}_i(a_i) = \frac{W_i(a_i) \prod_{\{\alpha | \alpha \in \mathbf{c}_i\}} \mathcal{M}_{\alpha \rightarrow i}(a_i)}{\sum_{z_i} W_i(z_i) \prod_{\{\alpha | \alpha \in \mathbf{c}_i\}} \mathcal{M}_{\alpha \rightarrow i}(z_i)}, \quad (142)$$

$$\hat{Q}_\alpha(\mathbf{a}_\alpha) = \frac{W_\alpha(\mathbf{a}_\alpha) \prod_{\{i|i \in \alpha\}} \prod_{\{\gamma|\gamma \in \mathbf{c}_i \setminus \{\alpha\}\}} \mathcal{M}_{\gamma \rightarrow i}(a_i)}{\sum_{\mathbf{z}_\alpha} W_\alpha(\mathbf{z}_\alpha) \prod_{\{i|i \in \alpha\}} \prod_{\{\gamma|\gamma \in \mathbf{c}_i \setminus \{\alpha\}\}} \mathcal{M}_{\gamma \rightarrow i}(z_i)}, \quad (143)$$

$$\mathcal{M}_{\alpha \rightarrow i}(a_i) = \frac{\sum_{\mathbf{z}_\alpha} \delta_{a_i, z_i} \left( \frac{W_\alpha(\mathbf{z}_\alpha)}{W_i(z_i)} \right) \prod_{\{j|j < \alpha\}} \prod_{\{\gamma|\gamma \in \mathbf{c}_j \setminus \{\alpha\}\}} \mathcal{M}_{\gamma \rightarrow j}(z_j)}{\sum_{\mathbf{z}_\alpha} \left( \frac{W_\alpha(\mathbf{z}_\alpha)}{W_i(z_i)} \right) \prod_{\{j|j < \alpha\}} \prod_{\{\gamma|\gamma \in \mathbf{c}_j \setminus \{\alpha\}\}} \mathcal{M}_{\gamma \rightarrow j}(z_j)}. \quad (144)$$

このような取り扱いはどこまで一般化できるのか？ 気になる話である。ある種の条件の下で与えられた任意の推論システムでこの決定方程式が成り立てば、汎用性のある推論ソフトを作ることができる。実はこのからくりの根底には次の性質がある。

- (i)  $C$  はノードの集合を要素とする集合である。
- (ii) すべてのノードからなる集合を  $\Omega$  とすると集合  $C$  から集合  $\Omega$  を除いた集合  $C \setminus \Omega$  の任意の要素  $\gamma$  と  $\gamma'$  に対して  $\gamma < \gamma'$  または  $\gamma > \gamma'$  という関係が成り立つことはない。
- (iii)  $\mu(\gamma)$  は  $\mu(\gamma) = -1$  ( $\gamma \in C \setminus \Omega$ ),  $\mu(i) = -1 - \sum_{\{\gamma|\gamma > i, \gamma \in C\}} \mu(\gamma)$  ( $i \in \Omega$ ) により定義される。
- (iv) 全ノードに対する結合確率分布  $\Pr\{A = \mathbf{a}\}$  が式 (125) の形に与えられる。

具体的に図 13 および図 14 で与えられた確率的推論について  $V_{56 \rightarrow 8}(a_8|a_5, a_6)$ ,  $V_{34 \rightarrow 6}(a_6|a_3, a_4)$ ,  $V_{6 \rightarrow 7}(a_7|a_6)$ ,  $V_{2 \rightarrow 5}(a_5|a_2)$ ,  $V_{2 \rightarrow 4}(a_4|a_2)$ ,  $V_{1 \rightarrow 3}(a_3|a_1)$ ,  $V_1(a_1)$ ,  $V_2(a_2)$  の値を表 2 のように設定し、式 (142)-(144) を解くことにより得られた周辺確率分布 ( $\hat{Q}_i(+1), \hat{Q}_i(-1)$ ) と定義 (132) にもとづいて得られた厳密な周辺確率分布 ( $P_i(+1), P_i(-1)$ ) の値を表 3 に与える。更に周辺確率分布  $\hat{Q}_{25}(a_2, a_5)$ ,  $\hat{Q}_{36}(a_3, a_6)$  および条件付き確率  $\Pr\{A_2 = a_2|A_5 = a_5\} \simeq \hat{Q}_{25}(a_2, a_5)/\hat{Q}_5(a_5)$ ,  $\Pr\{A_3 = a_3|A_6 = a_6\} \simeq \hat{Q}_{36}(a_3, a_6)/\hat{Q}_6(a_6)$  の値を表 4 および表 5 に与える。

ここで紹介したクラスター変分法を用いた確率的推論アルゴリズムの一部は実は人工知能の分野では確率伝搬法という名前で統計力学とはまったく独立に研究されていた。最近になってニューロコンピューティング分野の何人かの研究者によって、自由エネルギーの概念とクラスター変分法の導入が確率伝搬法のアルゴリズムの総括的理解につながるということが指摘されている [22, 27]。本節で紹介した例において、直接、因果関係で結ばれたノード間の推論は計算できることは説明したが、例えば 1 と 8 のような、直接、因果関係で結ばれていないようなノード間の推論の計算については触れなかった。これを可能にするためには、クラスター変分法に加えて、線形応答理論というもう一つの新しい概念を導入する必要がある。詳細は参考文献 [28] を参照していただきたい。

## 9 おわりに

本稿では、平均場近似、ベータ近似そしてクラスター変分法について自由エネルギーと呼ばれる量の最小化に対する変分原理にもとづいて概説した。ベータ近似とクラスター変分法は確率伝搬法あるいは一般化された確率伝搬法と数理的構造がほぼ同じであることが最近の研究で明らか

表 2: 式 (118) における  $V_{56 \rightarrow 8}(a_8|a_5, a_6)$ ,  $V_{34 \rightarrow 6}(a_6|a_3, a_4)$ ,  $V_{6 \rightarrow 7}(a_7|a_6)$ ,  $V_{2 \rightarrow 5}(a_5|a_2)$ ,  $V_{2 \rightarrow 4}(a_4|a_2)$ ,  $V_{1 \rightarrow 3}(a_3|a_1)$ ,  $V_1(a_1)$ ,  $V_2(a_2)$  の値.

$a_1$	$V_1$	$a_2$	$V_2$
+1	0.99	+1	0.50
-1	0.01	-1	0.50

$a_1$	$a_3$	$V_{1 \rightarrow 3}$	$a_2$	$a_4$	$V_{2 \rightarrow 4}$
+1	+1	0.99	+1	+1	0.99
+1	-1	0.01	+1	-1	0.01
-1	+1	0.95	-1	+1	0.90
-1	-1	0.05	-1	-1	0.10

$a_2$	$a_5$	$V_{2 \rightarrow 5}$	$a_6$	$a_7$	$V_{6 \rightarrow 7}$
+1	+1	0.70	+1	+1	0.95
+1	-1	0.30	+1	-1	0.05
-1	+1	0.40	-1	+1	0.02
-1	-1	0.60	-1	-1	0.98

$a_3$	$a_4$	$a_6$	$V_{34 \rightarrow 6}$	$a_5$	$a_6$	$a_8$	$V_{56 \rightarrow 8}$
+1	+1	+1	0.99	+1	+1	+1	0.90
+1	+1	-1	0.01	+1	+1	-1	0.10
+1	-1	+1	0.01	+1	-1	+1	0.30
+1	-1	-1	0.99	+1	-1	-1	0.70
-1	+1	+1	0.01	-1	+1	+1	0.20
-1	+1	-1	0.99	-1	+1	-1	0.80
-1	-1	+1	0.01	-1	-1	+1	0.10
-1	-1	-1	0.99	-1	-1	-1	0.90

になりつつある [22]. 確率を用いた情報処理は従来の決定木上の if-then-else を基本とする推論に比べて、起こる可能性の低い事象も含めて検討した推論を要求される。このことがどうしても計算量の増加につながってしまう要因の一つとなっている。しかしながら、如何に確率モデルとはいえ、あまりにも起こる可能性の低い事象までまじめに取り扱う必要はないわけで、ある程度の妥協を系統的に行っていくことは確率的情報処理システムを実現していく上で不可欠と考えられる。この系統的な妥協を平均場近似が与えてくれているわけである。それとともに平均場近似に始まる統計力学的近似解析手法に対する関心はますます高まりつつある。平均場近似およびその拡張した手法を情報処理に応用する最新の研究成果の一部は科研費特定領域研究「確率的情報処理への統計力学的アプローチ」(平成 14 年度から平成 17 年度までの 4 年間) のホームページ (URL: <http://www.smapip.eei.metro-u.ac.jp/>) から得られるので是非参照していただきたい。



表 3: 式 (142)-(144) を解くことにより得られた周辺確率分布  $(\widehat{Q}_i(+1), \widehat{Q}_i(-1))$  と定義 (132) にもとづいて得られた厳密な周辺確率分布  $(P_i(+1), P_i(-1))$  の値

$i$	$(\widehat{Q}_i(+1), \widehat{Q}_i(-1))$	$(P_i(+1), P_i(-1))$
1	(0.9900, 0.0100)	(0.9900, 0.0100)
2	(0.5000, 0.5000)	(0.5000, 0.5000)
3	(0.9896, 0.0104)	(0.9896, 0.0104)
4	(0.9450, 0.0550)	(0.9450, 0.0550)
5	(0.5500, 0.4500)	(0.5500, 0.4500)
6	(0.9265, 0.0735)	(0.9265, 0.0735)
7	(0.8816, 0.1184)	(0.8816, 0.1184)
8	(0.5574, 0.4426)	(0.5607, 0.4393)

表 4: 式 (142)-(144) を解くことにより得られた周辺確率分布  $\widehat{Q}_{25}(a_2, a_5)$  および条件付き確率  $\Pr\{A_2 = a_2 | A_5 = a_5\} \simeq \widehat{Q}_{25}(a_2, a_5) / \widehat{Q}_5(a_5)$  の値

$a_2$	$a_5$	$\widehat{Q}_{25}(a_2, a_5)$	$\Pr\{A_2 = a_2   A_5 = a_5\}$
+1	+1	0.35	0.70
+1	-1	0.15	0.30
-1	+1	0.20	0.40
-1	-1	0.30	0.60

## 参考文献

- [1] 田中和之, 樺島祥介, “大規模統計モデルにもとづく情報処理 – モノの理とコトの技 –,” 計測と制御, vol.42, no.8, pp.622-625, 2003.
- [2] 本村陽一, “ベイジアンネットワークソフトウェア,” 人工知能学会誌, vol.17, no.5, pp.559-565, 2002.
- [3] J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, (Morgan Kaufmann, 1988).
- [4] S. L. Lauritzen and D. J. Spiegelhalter, “Local computations with probabilities on graphical structures and their application to expert systems,” J. Royal Statistics Society B, vol.50, no.2, pp.157-194, 1988.
- [5] 田中利幸, “人工知能と確率推論,” 数理科学 2004 年 3 月号, No.489, 2004.
- [6] R. Chellappa and A. Jain (eds), *Markov Random Fields: Theory and Applications*, Academic Press, New York, 1993.

表 5: 式 (142)-(144) を解くことにより得られた周辺確率分布  $\widehat{Q}_{36}(a_3, a_6)$  および条件付き確率  $\Pr\{A_3 = a_3 | A_6 = a_6\} \simeq \widehat{Q}_{36}(a_3, a_6) / \widehat{Q}_6(a_6)$  の値

$a_3$	$a_6$	$\widehat{Q}_{36}(a_3, a_6)$	$\Pr\{A_3 = a_3   A_6 = a_6\}$
+1	+1	0.9264	0.9361
+1	-1	0.0632	0.0639
-1	+1	0.0001	0.0120
-1	-1	0.0103	0.9880

- [7] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, 1995.
- [8] 田中和之, “統計力学的手法をもとにした画像修復,” 日本物理学会誌, vol.54, no.1, pp.25-33, 1999.
- [9] K. Tanaka, “Statistical-mechanical approach to image processing,” *Journal of Physics A: Mathematical and General*, vol.35, no.20, pp.R81-R150, 2002.
- [10] 田中和之, “統計力学を用いた確率的画像処理アルゴリズムの基礎— 確率伝搬法と統計力学 —,” 計測と制御, vol.42, no.8, pp.631-636, 2003.
- [11] 田中和之, “確率モデルと画像分析— ゆらぎの生み出す新しい情報処理の世界 —,” 数理科学 2004 年 3 月号, No.489, pp.15-21, 2004.
- [12] 樺島祥介, “コトの物理 — 誤り訂正符号を例として —,” 日本物理学会誌, vol.58, no.4, pp.239-246, 2003.
- [13] Y. Kabashima, T. Murayama and D. Saad, “Cryptographical Properties of Tsing Spin Systems,” *Physical Review Letters*, vol.84, no.9, pp.2030-2033, 2000.
- [14] T. Murayama, “Statistical mechanics of data compression theorem,” *Journal of Physics A: Mathematical and General*, vol.35, no.8, pp.L95-L100, 2002.
- [15] 田中利幸, “移動体通信技術とスピングラスとの意外な関係,” 日本物理学会誌, vol.56, no.9, pp. 660-666, 2001.
- [16] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Transactions on Information Theory*, vol.48, no.11, pp.2888- 2910, 2002.
- [17] M. Opper and D. Saad (eds), “Advanced Mean Field Methods — Theory and Practice —,” MIT Press, 2001.
- [18] 樺島祥介, “岩波講座 物理の世界/学習と情報の平均場理論,” 岩波書店, 2002.
- [19] 特集/知識情報処理の統計力学のアプローチ, 数理科学, 1999 年 12 月号.
- [20] 西森秀稔, “スピングラス理論と情報統計物理学,” 新物理学選書, 岩波書店, 1999.
- [21] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, Oxford, 2001.
- [22] J. S. Yedidia, W. T. Freeman and Y. Weiss: “Generalized belief propagation,” *Advances in Neural Information Processing Systems*, vol.13, pp.689-695, 2001 (Cambridge, MA: MIT Press).
- [23] 池田思朗, 田中利幸, M. Welling, NIPS\*2002 post conference workshop on Propagation Algorithms on Graphs with Cycles: Theory and Applications (December 13th and 14th, 2002, Whistler, Canada), URL: <http://prop.algo.eei.metro-u.ac.jp/>.
- [24] 田中和之, “画像修復に対する周辺ゆう度最大化によるハイパラメータ推定のクラスタ変分法を用いた理論的検討,” 電子情報通信学会論文誌 (A), vol.J83-A, no.10, pp.1148-1160, 2000.
- [25] K. Tanaka, “Maximum Marginal Likelihood Estimation and Constrained Optimization in Image Restoration,” *人工知能学会論文誌*, vol.16, no.2, pp.246-258, 2001.
- [26] K. Tanaka, J. Inoue and D. M. Titterton, “Probabilistic image processing by means of Bethe approximation for  $Q$ -Ising model, *J. Phys. A: Math. & Gen.*, vol.36, no.43, pp.11023-11036, 2003.
- [27] H. J. Kappen and W. Wiegner: “Novel iteration schemes for the cluster variation method,” *Advances in Neural Information Processing System*, vol.14, pp.415-422, 2002 (Cambridge, MA: MIT Press).
- [28] K. Tanaka, “Probabilistic Inference by means of Cluster Variation Method and Linear Response Theory,” *IEICE Transactions on Information and Systems*, vol.E86-D, no.7, pp.1228-1242, 2003.