

第 58 回物性若手夏の学校

“人工知能”のための統計力学

樺島 祥介¹東京工業大学大学院総合理工学研究科
知能システム科学専攻

概要

気がつけば“人工知能”がそばにいる世の中になっていた。HAL9000 は暴走していないし、ネコ型ロボットが街を闊歩しているわけでもない。しかし、スマホは次に入力するフレーズを予測し、デジカメはヒトの顔を枠で囲む。Google は検索結果を瞬時に返してくるし、Amazon で買い物をすると好みの品物をワンサカ薦めてくる。活気のある立ち会いが好景気の象徴だった株の売り買いも、いまではミリ秒単位の自動売買である。確かに、SF の世界で描かれる、ヒトのように自律的に考え判断する機械はまだ現れていない。けれども、それほど複雑ではないけれどコツや経験が必要な、ひと昔前ならヒトでなければできなかった知的作業を機械が代行する。そういった社会はすでに到来しているのである。こうした“人工知能”の共通点として、手元にあるデータから有益な情報を抽出する「推論」を実現していることが挙げられる。これを可能にした要因として真っ先に思い浮かぶのは計算機の高性能化と普及である。でもそれだけでは十分ではない。データに含まれる情報を最大限に引き出すためには適切なモデル化と、それを現実的制約下で計算機上に実装する技術が必要になる。データの生成過程を適切な確率モデルで表現し、ベイズの公式にしたがって推論を行う枠組みは特に「確率推論」とよばれる。本講義では、スピングラスに端を発する不規則系の統計力学と確率推論との構造的類似性にもとづいて統計力学の理論を“人工知能”の技術として活かす研究について紹介する。

1 はじめに

携帯電話やインターネットなどの通信インフラ、データセンターの整備、また、さまざまな計測機器の発達を背景として、大規模なデータからの情報抽出への関心が高まっている。適当な確率モデルを仮定しベイズの公式にもとづいて観測データから未知の確率変数を推定する確率推論は、こうした情報抽出を行うための有力な手段である。ところが、確率推論に必要な計算量は一般に確率変数の次元に関して指数関数的に増大し、実行が困難になる。この問題を解決する実際的手段として、近年、人工知能研究をはじめとした情報／計算機科学では統計力学への関心が高まっている。

本稿では、そうした物理学の枠を越えた統計力学の利用法について紹介する。

本稿の構成は以下の通りである。次節では、統計力学と確率推論それぞれの特徴を概観し、それらの構造的な類似点と相違点について述べる。本稿で主に述べるのは、確率推論を実行する近似アルゴリズムとしての平均場近似である。その準備として、第3節では強磁性体のイジング模型を例題として統計力学でもちいられる平均場近似について説明する。物質の一様性や規則性を取り入れながら定式化されている物理学の平均場近似は、多様な確率分布があらわれる確率推論にはそのままの形では使い難い。そこで、一般的な分布への平均場近似の拡張について第4節で述べる。第5節はまとめにあてられる。

なお、講義時間の都合上、性能評価問題への統計力学の利用法については今回は省略した。文献 [1, 2] などをご覧頂きたい。

2 統計力学、確率推論、不規則系の統計力学

本節では統計力学と確率推論との関係について述べる。

¹E-mail: kaba@dis.titech.ac.jp

2.1 統計力学とは

2.1.1 分配関数, 自由エネルギー, 内部エネルギー, エントロピー

古典系のみ限定して話を進める. 微視的な状態変数の組 \mathbf{S} に対し, ハミルトニアン $H(\mathbf{S})$ が定義されているシステムを考える. エッセンスだけ抜き出すと, このシステムの絶対温度 T での熱平衡状態の性質が**カノニカル分布 (canonical distribution)**

$$P(\mathbf{S}) = \frac{e^{-\beta H(\mathbf{S})}}{Z(\beta)} \quad (1)$$

に関する期待値 (平均) で評価できる, というのが (平衡) 統計力学の主張である. ただし, $\beta = (k_B T)^{-1}$ であり (k_B はボルツマン定数), 規格化定数 $Z(\beta) = \text{Tr}_{\mathbf{S}} e^{-\beta H(\mathbf{S})}$ はしばしば**分配関数 (partition function)** とよばれる².

カノニカル分布に関する期待値を具体的に評価する際には分配関数, あるいはその対数で与えられる**自由エネルギー (free energy)** $F = -\beta^{-1} \ln Z(\beta)$ が大きな役割を果たす. たとえば, カノニカル分布に関するハミルトニアンの期待値

$$U = \text{Tr}_{\mathbf{S}} H(\mathbf{S}) P(\mathbf{S}) = \text{Tr}_{\mathbf{S}} H(\mathbf{S}) \frac{e^{-\beta H(\mathbf{S})}}{Z(\beta)} \quad (2)$$

を考えよう. 合成関数の微分に関する公式を使うと, これは

$$U = -\frac{\partial}{\partial \beta} \ln Z(\beta) = \frac{\partial}{\partial \beta} (\beta F) \quad (3)$$

によって評価できることがわかる. U を熱力学における内部エネルギー, F を熱力学における自由エネルギーと対応づけると, 熱力学の関係式からエントロピーも

$$S = \frac{1}{T} (U - F) = -k_B \beta \frac{\partial}{\partial \beta} \ln Z(\beta) + k_B \ln Z(\beta) \quad (4)$$

のように分配関数を通じて評価することができる.

2.1.2 例1) 古典単原子分子理想気体

体積 V の容器内にある 1 分子あたりの質量が m の単原子分子古典理想気体を考える. 分子数を N , 分子 $i (= 1, 2, \dots, N)$ の位置と運動量をそれぞれ $\mathbf{q}_i, \mathbf{p}_i$ とすると, $\mathbf{S} = \{(\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_2, \mathbf{p}_2), \dots, (\mathbf{q}_N, \mathbf{p}_N)\}$ であり, ハミルトニアンは

$$H(\mathbf{S}) = \sum_{i=1}^N \left(\frac{|\mathbf{p}_i|^2}{2m} + \Phi(\mathbf{q}_i) \right) \quad (5)$$

で与えられる. ただし, $\Phi(\mathbf{q})$ は \mathbf{q} が容器の内部にあるとき 0, それ以外では無限大になるポテンシャルエネルギーである. 不確定性原理および気体分子は互いに区別できないことから, プランク定数を h として $\text{Tr}_{\mathbf{S}} \rightarrow h^{-3N} (N!)^{-1} \int \prod_{i=1}^N (d\mathbf{q}_i d\mathbf{p}_i)$ となる. ここで, ハミルトニアンが各粒子に関する和の形に書けていることから $\text{Tr}_{\mathbf{S}}$ は各粒子添字に関して独立な積分となり

$$Z(\beta) = \frac{(2\pi m k_B T)^{3N/2} V^N}{h^{3N} N!} \quad (6)$$

が得られる. これから直ちに内部エネルギーが

$$U = -\frac{\partial}{\partial \beta} \ln Z(\beta) = \frac{3N}{2\beta} = \frac{3N k_B T}{2} \quad (7)$$

² $\text{Tr}_{\mathbf{S}}$ はすべての可能な状態について和あるいは積分を取るという意味

で与えられること、よって、定積熱容量が

$$C = \frac{\partial U}{\partial T} = k_B \beta^2 \frac{\partial^2 \ln Z(\beta)}{\partial \beta^2} = \frac{3Nk_B}{2} \quad (8)$$

となることが導かれる。また、圧力 p と自由エネルギーを関係づける式 $p = -(\partial/\partial V)F$ に $F = -\beta^{-1} \ln Z(\beta)$ を代入することで状態方程式

$$pV = Nk_B T \quad (9)$$

が得られる。これらは古典的な扱いが正当化される高温低密度の状況で実験的に観測される気体の振る舞いをよく記述する。こうした結果を通じて統計力学が有用な理論であることが確かめられる。

2.1.3 例2) 強磁性体のイジング模型

理想気体は統計力学が“うまくいく”ことを簡単に示すよい教材である。ただし、ここまで簡単に事が進む例は他にほとんどない。計算が途中で手に負えなくなり、先に進めなくなるからである。強磁性体の**イジング模型 (Ising model)** はそうした“計算できない”例の典型である。

強磁性体 (= 磁石) の振る舞いを数理的に考察するために導入されたこの模型では、与えられた格子上の格子点 i に磁性の根源であるスピンを意味する2値変数 $S_i \in \{+1, -1\}$ を定義する。強磁性体では隣接するスピン変数は相互作用の結果、互いに同じ値を取ろうとする傾向があると考えられる。この性質を簡潔に表現したハミルトニアンとして

$$H(S) = -J \sum_{(ij)} S_i S_j - h \sum_{i=1}^N S_i \quad (10)$$

を仮定する。ここで、 $J > 0$ は相互作用の強さ、 h は外部磁場、 (ij) は隣接する格子点の対をそれぞれあらわすものとする。 $J > 0$ と仮定したことで、カノニカル分布 (1) では隣接するスピン対が同符号になる方が確率が大きくなる。その結果、強磁性体の性質 (隣接スピン対が同符号になりたがる、という性質) が簡潔に反映されている。

強磁性体では、ある臨界的な温度 T_c を境に磁化の強さ (= 磁石の強度) M の振る舞いが外部磁場 h をゼロとした極限で

$$\lim_{h \rightarrow 0} M \begin{cases} > 0 & T < T_c \\ = 0 & T > T_c \end{cases} \quad (11)$$

のように定性的に変化する**相転移 (phase transtion)** が生じることが知られている。一般に相転移は自由エネルギーの解析性が破れる現象として特徴づけられる。外場ゼロの強磁性転移では自由エネルギーの温度に関する2階微分と対応する比熱に不連続性が生じることが実験的に知られている。

大胆な単純化を導入しているイジング模型は相転移を記述できるだろうか? 分配関数と内部エネルギーを結びつける関係式 (3) はモデルやハミルトニアンの詳細には依存しない。イジング模型に対しても分配関数さえ評価できれば理想気体に関する (7), (8) と同様の計算によって比熱が求まり、したがって、相転移の有無を吟味することができる。

ところが、この分配関数の計算が難しい。その理由はスピン間相互作用をあらわす (10) の右辺第1項の存在にある。これにより、理想気体の場合とは異なり、 Tr_S を評価する際には各スピン添字ごとに独立に和を計算し、その結果をあとでまとめるという評価ができなくなる。もちろん、並進対称性や相互作用、外場の一様性などを利用して部分的な和を“手で”評価し、計算の効率化をはかることは可能かもしれない。実際、1次元格子 (鎖) や $h = 0$, $N \rightarrow \infty$ の2次元格子では広い意味でこの方針が成功し、厳密解が得られる [3, 4]。けれども、これらは“例外”なのであり、外場のある2次元格子、現実との対応の観点からもっとも興味深い3次元格子についての厳密解は未だ得られていない。

分配関数を評価するためには、要素数 N について指数関数的に増える全状態数と同程度の項数に関する和を愚直に計算する以外に一般的な方法は知られていない。そのため、物理学の研究とはいえ、統計力学の研究の実際では、こうした爆発する計算量への対処法の開発が非常に大きなウエートを占めている。

2.2 確率推論とは

次に確率推論について述べる。

2.2.1 例3) 情報通信

通信路を介して N ビットの信号 $\mathbf{x} \in \{0,1\}^N$ を送信する状況を考える。通信路では何らかのプロセスが作用し、受信側では \mathbf{x} とは異なる信号 \mathbf{y} が受信されるとする。 \mathbf{y} は離散ベクトルでも連続ベクトルでもよく、また、一般に \mathbf{x} とは異なる次元を持つ。送信信号 \mathbf{x} は事前分布 $P(\mathbf{x})$ にしたがって生成され、また、 \mathbf{x} が送信されたとき条件付き確率 $P(\mathbf{y}|\mathbf{x})$ にしたがって受信信号 \mathbf{y} が受信されるとする。

$P(\mathbf{x})$, $P(\mathbf{y}|\mathbf{x})$ の関数形は受信側で既知であるとしよう。情報通信では \mathbf{y} を受信した際に送信した信号 \mathbf{x} が何であったかを推定 (復号) することが必要になる。どのような復号を行えばよいのだろうか? 意外なことに、 $P(\mathbf{x})$, $P(\mathbf{y}|\mathbf{x})$ の具体的関数形を与えないままでも、この問題の答えは形式的に求まる [5]。

[最適な復号法] \mathbf{y} を得た際の復号結果を一般的に $\hat{\mathbf{x}}(\mathbf{y}) = (\hat{x}_i(\mathbf{y}))$ と記す。復号結果の良し悪しをビット誤り率

$$\text{BER}(\hat{\mathbf{x}}(\cdot)) = \frac{1}{N} \sum_{i=1}^N \Pr(\hat{x}_i(\mathbf{y}) \neq x_i) \quad (12)$$

によって評価することにする。このとき、以下が成立する。

$$\text{BER}(\hat{\mathbf{x}}(\cdot)) \geq \frac{1}{N} \sum_{i=1}^N \text{Tr}_{\mathbf{y}} \left(\min_{x_i \in \{0,1\}} \{P_i(x_i|\mathbf{y})\} P(\mathbf{y}) \right) \quad (13)$$

ただし、ベイズの公式から

$$P(\mathbf{y}) = \text{Tr}_{\mathbf{x}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \quad (14)$$

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \quad (15)$$

であり、 $P_i(x_i|\mathbf{y}) = \text{Tr}_{\mathbf{x} \setminus x_i} P(\mathbf{x}|\mathbf{y})$ は事後分布 $P(\mathbf{x}|\mathbf{y})$ を x_i を残したまま他の成分 $x_{j \neq i}$ について周辺化した (和を取った) **周辺事後分布 (marginal posterior distribution)** をあらわしている。また、本稿では以降一般に $A \setminus a$ はベクトルや集合をあらわす A から成分や要素をあらわす a を除外する操作を意味するものとする。等号は

$$\hat{x}_i^{\text{opt}}(\mathbf{y}) = \underset{x_i \in \{0,1\}}{\text{argmax}} \{P_i(x_i|\mathbf{y})\} \quad (16)$$

の場合に成立する。ただし、 $\text{argmax}_x \{f(x)\}$ は一般に関数 $f(x)$ を最大にする x の値をあらわすとす。このことは、(16) がビット誤り率を最小にする意味で、すべての復号法の中で最良の方法であることを意味している。

以下に証明を与える。

[証明] \mathbf{x} を送信し、 \mathbf{y} が受信される同時確率がベイズの公式をもちいることで $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y})$ のように表現できることに着目する。 \mathbf{x} に対して、復号結果 $\hat{\mathbf{x}}(\mathbf{y})$ が含む誤りビット数は $\sum_{i=1}^N \delta(\hat{x}_i(\mathbf{y}) \neq x_i)$ で与えられる。ただし、 $\delta(\dots)$ は \dots が真の場合 1, 偽の場合 0 となるインジケータ関数であるとする。ビット誤り率はこれを同時確率 $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y})$ で平均し、総ビット数 N で割ったものである。よって、

$$\text{BER}(\hat{\mathbf{x}}(\cdot)) = \frac{1}{N} \sum_{i=1}^N \Pr(\hat{x}_i(\mathbf{y}) \neq x_i)$$

$$\begin{aligned}
&= \frac{1}{N} \operatorname{Tr}_{\mathbf{y}} \operatorname{Tr}_{\mathbf{x}} \sum_{i=1}^N \delta(\hat{x}_i(\mathbf{y}) \neq x_i) P(\mathbf{x}|\mathbf{y}) P(\mathbf{y}) \\
&= \frac{1}{N} \operatorname{Tr}_{\mathbf{y}} \sum_{i=1}^N \sum_{x_i \in \{0,1\}} (1 - \delta(\hat{x}_i(\mathbf{y}) = x_i)) P_i(x_i|\mathbf{y}) P(\mathbf{y}) \\
&= 1 - \frac{1}{N} \sum_{i=1}^N \operatorname{Tr}_{\mathbf{y}} P_i(\hat{x}_i(\mathbf{y})|\mathbf{y}) P(\mathbf{y}) \\
&\geq 1 - \frac{1}{N} \sum_{i=1}^N \operatorname{Tr}_{\mathbf{y}} \left(\max_{x_i \in \{0,1\}} \{P_i(x_i|\mathbf{y})\} P(\mathbf{y}) \right) \\
&= \frac{1}{N} \sum_{i=1}^N \operatorname{Tr}_{\mathbf{y}} \left(\min_{x_i \in \{0,1\}} \{P_i(x_i|\mathbf{y})\} P(\mathbf{y}) \right) \tag{17}
\end{aligned}$$

がなりたつ。不等号は各 \mathbf{y} に対し、一般に

$$P_i(\hat{x}_i(\mathbf{y})|\mathbf{y}) \leq \max_{x_i \in \{0,1\}} \{P_i(x_i|\mathbf{y})\} \tag{18}$$

がなりたつことにもとづいている。復号法が(16)の場合にはあきらかに等号が成立する。よって、(16)はビット誤り率を最小にする復号法であり、(13)の右辺はこの通信路に対して原理的に達成可能なビット誤り率の最小値をあらわしている。

2.2.2 確率推論

知っている事柄にもとづいて未知の事柄を知ろうとする営みは一般に「推論」とよばれる。過去のデータにもとづいて今後の天気を予測する天気予報、検査結果や症状にもとづいて病気の原因を探る医療診断、購買履歴にもとづいて次に買ったような商品を薦めるリコメンデーション、これらはすべて推論である。数学の証明のように仮定から結論を曖昧さなく導く決定的な推論もあるが、実社会の多くの状況では不確実な環境の下で推論を行わなければならない。

先に示した情報通信の問題はこうした不確実環境下における推論の雛形になっている。この例では、環境の不確実性を $P(\mathbf{y}|\mathbf{x})$, $P(\mathbf{x})$ といった確率形式で表現した**確率モデル (probabilistic model)** にもとづいて最適な推論の方法を導くことができた。ところで、情報通信の文脈で説明したが、最適性の証明では具体的な通信の方式やノイズの発生原因などの詳細は何も使っていなかったことを思い出そう。そのため、最適化すべき目的関数は問題ごとに変わるものの、 \mathbf{y} を過去の天気/検査結果・症状/購買履歴、 \mathbf{x} を今後の天気/病気の原因/次に買いたい商品、に変え、それぞれの場合に適当な確率モデルを導入すれば、天気予報/医療診断/リコメンデーションに最適な推論方法をまったく同様の方針で導くことができる。こうした、確率モデルにもとづいた推論方法を、以下、一般に**確率推論 (probabilistic inference)** とよぶ。

2.3 統計力学と確率推論：類似点と相違点

情報通信の例が示すように確率推論を行うためには、事後分布 $P(\mathbf{x}|\mathbf{y})$ について周辺化や最適化を行うことが必要になる。 $\mathbf{x} \in \{0,1\}^N$ である場合、これに必要な計算量は特別な場合を除き、一般に N について指数関数的に増大する。これは強磁性体のイジング模型の分配関数評価と同様の計算量的困難を生み出す。そのため、確率推論の研究でもこうした計算爆発への対処法が大きなウェイトを占めている。

統計力学と確率推論の間には相違点もある。統計力学で通常扱われる問題のカノニカル分布は高次元の確率モデルである。しかしながら、物質の多様性を反映してさまざまなモデルが考えられるものの、それらの多くは格子状のグラフの隣接頂点に定義された変数対の同一関数で記述される2体相互作用 (= 確率変数間の依存関係) を想定している。こうした視点に立つと統計力学であらわれる問題は一般の確率モデルからすると特殊なクラスに属していると考えざるを得ない。より象徴的には、統計力学がカノニカル分布という「分布」に関する問題であるのに対し、確率推論は事後分布という「条件付き分布」に対する問題である、ということもできる。条件付き分布は観測値 \mathbf{y} の実現値ごとに \mathbf{x} に異なる分布を与える。そのため、具体的な推論アルゴリズムを

開発する際にも、単一の分布に対する有効性ではなく、観測値 \mathbf{y} のさまざまな可能性に対する有効性を考慮しなければならない。

計算量的困難という類似した問題を抱えつつも近年になるまで2つの分野間に活発な交流が生まれなかった主な原因の一つは、分野間で主に考察している問題に関するこうした構造的相違点にある、と考えられる。

2.4 不規則系の統計力学

統計力学と確率推論との距離は不規則系の統計力学の登場によって縮まることになる。

2.4.1 例4) スピングラス模型

非磁性の金属(金, 銀, 銅など)に, 磁性体(鉄, マンガンなど)を低濃度(0.1~10%程度)で混ぜ高温から低温に急冷(クエンチ)する。こうして低濃度で混ぜられた磁性体は磁性不純物とよばれるが, クエンチされた結果, 磁性を担うスピン間相互作用には強磁性と反強磁性の相互作用が実効的に入り交じったランダム性が生じることになる。スピングラス模型とは, こうした状況を簡潔に記述するために導入された数理模型であり, たとえば, イジングスピン $\mathbf{S} \in \{+1, -1\}^N$ に対してハミルトニアン

$$H(\mathbf{S}|\mathbf{J}) = - \sum_{(ij)} J_{ij} S_i S_j - h \sum_{i=1}^N S_i \quad (19)$$

を仮定することで定義する。(19)は強磁性体のハミルトニアン(10)とほぼ同じ形をしているが, ランダム性を反映して, スピン間相互作用係数 J_{ij} が相互作用対 (ij) ごとにある確率分布 $P(J_{ij})$ から選んだサンプル値として定義されているところに特徴がある。このようにハミルトニアンを定めるパラメータが乱数で与えられるシステムに関する統計力学は特に**不規則系の統計力学 (statistical mechanics of disordered systems)** とよばれている。

$P(J_{ij})$ としてはガウス分布 $(2\pi\bar{J}^2)^{-1/2} \exp(-(J_{ij} - J_0)^2/(2\bar{J}^2))$ やベルヌーイ分布 $p\delta(J_{ij} - 1) + (1-p)\delta(J_{ij} + 1)$ ($0 < p < 1$) など正負両方の値を取り得る分布を仮定することが多い。スピングラスの最も際立った特徴は正負の相互作用係数がランダムに選ばれた結果, 複数の相互作用対間で安定なスピン配位が矛盾する**フラストレーション (frustration)** が生じる点にある。1970年代の提案以降, 多くの理論研究者がシンプルでありながら質的に新しい物性を記述するスピングラス模型の解析に取り組んできた。

2.4.2 不規則系の統計力学に求められること

ハミルトニアン(19)の詳細は $\mathbf{J} = (J_{ij})$ の実現値ごとに異なる。そのため, 不規則系のカノニカル分布はパラメータ \mathbf{J} で条件付けされた条件付き分布

$$P(\mathbf{S}|\mathbf{J}) = \frac{1}{Z(\mathbf{J}, \beta)} \exp(-\beta H(\mathbf{S}|\mathbf{J})) \quad (20)$$

となる。ただし, $Z(\mathbf{J}, \beta) = \text{Tr}_{\mathbf{S}} \exp(-\beta H(\mathbf{S}|\mathbf{J}))$ は \mathbf{J} に対して定まる分配関数である。このことはスピンの熱平均値 $\langle S_i \rangle = \text{Tr}_{\mathbf{S}} S_i P(\mathbf{S}|\mathbf{J})$ は \mathbf{J} に依存した確率変数になることを意味する。ただし, $\langle \dots \rangle$ はカノニカル分布に関する平均をあらわすものとする。こうした相互作用係数のランダム性に依存して揺らぐ熱平均の統計的な性質を解明する, 具体的には, 熱平均に関するモーメント

$$[\langle S_i \rangle^m] = \text{Tr}_{\mathbf{J}} \left(\text{Tr}_{\mathbf{S}} S_i P(\mathbf{S}|\mathbf{J}) \right)^m P(\mathbf{J}) \quad (21)$$

($i = 1, 2, \dots, N; m = 1, 2, \dots$) を評価することが不規則系の統計力学の目標となる。ただし, 熱平均と区別するため配位平均とよばれる \mathbf{J} に関する平均を $[\dots]$ とあらわした。

2.4.3 不規則系の統計力学における代表的な解析法

J の1つのサンプルについて S に関する平均を計算するのでさえ困難であるのに、その結果をさらに J について平均しなければならない不規則系の統計力学は通常の統計力学と比較して格段に難しい。そのため、不規則系の統計力学の理論は平均評価が例外的に容易になる**無限レンジ模型 (infinite range model)** や解析的評価が例外的に可能になる条件を中心に発展してきた。その結果得られた代表的な解析法の要点を以下に記す [2, 6, 7].

- **レプリカ法 (replica method)** : (21) を形式的に

$$[\langle S_i \rangle^m] = \lim_{n \rightarrow 0} \text{Tr}_{\mathbf{J}} \left(\text{Tr}_{\mathbf{S}} S_i P(\mathbf{S}|\mathbf{J}) \right)^m Z^n(\mathbf{J}, \beta) P(\mathbf{J}) \quad (22)$$

と書き換える。 $P(\mathbf{S}|\mathbf{J}) = \exp(-\beta H(\mathbf{S}|\mathbf{J})) / Z(\mathbf{J}, \beta)$ であること、また、 $Z(\mathbf{J}, \beta) = \text{Tr}_{\mathbf{S}} \exp(-\beta H(\mathbf{S}|\mathbf{J}))$ であることをもちいると $n \geq m = 1, 2, \dots \in \mathbb{N}$ について

$$\begin{aligned} & \text{Tr}_{\mathbf{J}} \left(\text{Tr}_{\mathbf{S}} S_i P(\mathbf{S}|\mathbf{J}) \right)^m Z^n(\mathbf{J}, \beta) P(\mathbf{J}) \\ &= \text{Tr}_{\mathbf{J}} \text{Tr}_{S^1, S^2, \dots, S^m} S_i^1 S_i^2 \dots S_i^m \exp \left(-\beta \sum_{a=1}^n H(\mathbf{S}^a|\mathbf{J}) \right) P(\mathbf{J}) \end{aligned} \quad (23)$$

という表現が得られる。 S^1, S^2, \dots, S^n は同じ J を共有する元のシステムの複製と解釈できるので**レプリカ (replica)** とよばれる。ここで $P(\mathbf{J})$ が相互作用ごとに独立な分布である場合には J に関する平均は比較的容易に評価することができる。その結果にもとづいて

$$\tilde{H}(S^1, S^2, \dots, S^n) = -\frac{1}{\beta} \ln \left(\text{Tr}_{\mathbf{J}} \exp \left(-\beta \sum_{a=1}^n H(\mathbf{S}^a|\mathbf{J}) \right) P(\mathbf{J}) \right) \quad (24)$$

を定義すると、これは n 個のレプリカ系全体に対するランダム性を含まない有効ハミルトニアンであると解釈できる。この有効ハミルトニアンが無限レンジ模型を意味している場合には (23) を n の関数として厳密に評価することができる。その結果の数学的正当性は $n \in \mathbb{N}$ に対してしか保証されないが、得られた表現が実数 $n \in \mathbb{R}$ についても成立すると仮定して (22) の右辺にもちいる。無限レンジ模型でない場合も、平均場近似により (23) を近似的に評価することで同様に対処できる。こうした一連の評価法はレプリカ法とよばれる。

- **キャビティ法 (cavity method)** : J の1つのサンプルが与えられたとする。通常の統計力学の無限レンジ模型では、後に説明する分子場近似が $N \rightarrow \infty$ で正しい結果を導く。ただし、不規則系の場合には、たとえ無限レンジ模型であるとしても $P(\mathbf{J})$ で特徴づけられる J の統計的性質に合わせた平均場近似を行う必要がある。広く考察されているある模型のクラスでは、変数間の依存関係が局所的にツリー (ループがない) であると近似する平均場近似の方法が正しい結果を導く。こうした近似法は変数間の依存関係をグラフで表現した際、変数や相互作用をあらわすノードを取り除き／加えながら自己無撞着な関係式を構成するアルゴリズムに帰着される。ノードの除去がグラフに“穴 (キャビティ: cavity)”をあけることに対応するのでしばしばキャビティ法とよばれる。
- **ゲージ理論 (gauge theory)** : ゲージ対称性のあるモデルでは、ゲージ変数を導入し状態和を取る自由度を増やすことにより、相互作用パラメータの生成分布 $P(\mathbf{J})$ によって定まる特別な温度 (**西森温度 (Nishimori temperature)**) に対して、分配関数を評価することなしに内部エネルギーや比熱の上限が任意の次元で厳密に求まる場合がある。また、その結果、相図に関していくつかの重要な知見を得ることができる。レプリカ法やキャビティ法は実質的に無限レンジ模型でしか正確な結果を得ることができない。そのため、ゲージ理論によるこうした分析は有限次元模型を解析する際の数少ない有力な手段となっている。

2.5 不規則系の統計力学と確率推論：構造的類似性

磁性不純物の混ざった非磁性金属の物性を解明するために提案された物理の模型であるスピンガラス模型は、高次元の条件付き分布として表現されている、という点において確率推論の問題と構造的に類似していることに注意しよう。通常の統計力学の計算法は均質性や規則性といった物

理モデルの特殊性にもとづいているため、確率推論の問題にそのまま適用することは難しい。ところが、不規則系の統計力学で発展した上述の解析法や考え方なら、構造的類似性にもとづいて確率推論の問題に直接的にもちいることができる。

主に情報科学で検討されてきた確率推論の研究では数学的厳密性や一般性を重視する傾向が強い。物理学をはじめ自然科学の教育を受けた人には意外かもしれないが、例外的に計算が容易になる模型や条件を深く分析し、近似や摂動的な取り扱いをもちいることで適用範囲を広める、といった物理学での常套手段は情報科学においてはまったく常識ではないのである。たとえば、1.2.1節で示した確率推論の最適性に関し、最適な推論を行った結果 (13) の右辺がどのくらいの値になるのか、といった定量的な評価は一般に難しく、情報科学では適当な不等式による上下界評価以外の系統的な評価法はほとんど知られていない。ところが、情報科学の文脈で眺めてみると、レプリカ法やゲージ理論は、例外的な模型や条件といった特別な状況ではあるが、こうした**性能評価 (performance evaluation)** の問題に対して直接的かつ系統的な評価法を与えている。また、キャピティ法をはじめとした平均場近似の方法は計算量爆発を実際的に解決する**近似的推論アルゴリズム (approximate inference algorithm)** と位置づけることができる。これら物理の方法を適用すれば、情報科学の問題に対して新規的な結果を系統的に得ることができるかもしれない。このようなモノ (コト?) の見方が、1990年代初頭にパーセプトロンとよばれるパターン認識の基本モデルの解析に不規則系の統計力学が応用されて [8] 以降徐々に広まっていった。

3 統計力学における平均場近似

計算が困難になる原因は要素間に相互作用が存在するからである。それならば、理想気体のような相互作用のないシステムで近似してやればよい。このような素朴な発想にもとづいた計算量的困難への近似的解決法は一般に**平均場近似 (mean field approximation)** とよばれる。

本節では、確率推論でもちいられる平均場近似の準備として統計力学における平均場近似の方法を強磁性体のイジング模型を例として紹介する。

3.1 分子場近似

最も基本的な平均場近似である**分子場近似 (molecular field approximation)** は以下のように構成される [9]。簡単のために外部磁場 (外場) のない $h = 0$ の場合に限定して話を進める。隣接格子点の数が z の正方格子 (図 1(a)) を想定し、格子点 i のスピン S_i に着目する。 S_i は i に関する隣接格子点の集合 (近傍) ∂i に属する格子点 $j \in \partial i$ のスピン S_j と相互作用するが、熱平衡状態における S_j からの影響はその熱平均値 $\langle S_j \rangle$ に置き換えて評価してもさほど問題はないだろう (図 1(b))。また、格子の並進対称性からすべてのスピン S_k ($k = 1, 2, \dots, N$) は同じ熱平均値 $\langle S_k \rangle = m$ を持つと考えてよい。これらのことから、熱平衡状態における S_i の 1 体分布は有効ハミルトニアン

$$H_i^{\text{eff}}(S_i) = -JS_i \sum_{j \in \partial i} \langle S_j \rangle = -zJmS_i \quad (25)$$

に対するカノニカル分布で近似できると期待される。(25) は相互作用のないスピンに zJm の外場が印加されている系のハミルトニアンと解釈できる。 zJm は相互作用する相手 (分子) を代替する外場であるため**分子場 (molecular field)** あるいは**平均場 (mean field)** とよばれる。これが分子場近似、平均場近似の名前の由来である。

ところで、 S_i は勝手に着目したスピンなので、その熱平均値は他のスピンと同じはずである。そこで (25) に対するカノニカル分布を使って S_i の熱平均値を評価し、それを m と等値すると

$$\langle S_i \rangle = \sum_{S_i \in \{+1, -1\}} S_i \frac{e^{\beta z J m S_i}}{2 \cosh(\beta z J m)} = \tanh(\beta z J m) = m \quad (26)$$

という、 m に関する自己無撞着方程式が得られる。 $\beta_c = (zJ)^{-1}$ とおくと、この方程式は

$$\beta \begin{cases} > \beta_c & \rightarrow \pm m^* \ (m^* > 0) \text{ の解を持つ} \\ \leq \beta_c & \rightarrow m = 0 \text{ しか解を持たない} \end{cases} \quad (27)$$

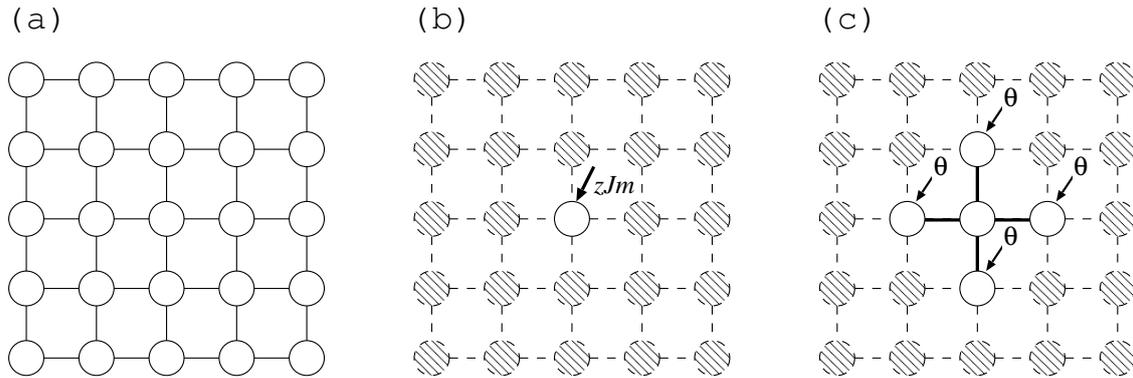


図 1: (a): 正方格子上的強磁性体のイジングスピン模型. \bigcirc がイジングスピン $S_i \in \{+1, -1\}$ をあらわしている. (b): 分子場近似. 1つのスピンの着目し, 他のスピンからの影響は分子場 zJm で代替されると考える. (c): ベーテ近似. 1つのスピンとそれと最隣接しているスピンの組に着目. それら以外のスピンからの影響は最隣接スピンの印加されたキャビティ場 θ で代替されると考える. (b), (c) ではそれぞれ m, θ を自己無撞着に定める.

という振る舞いを示す. $\beta = \beta_c$ を境とした熱平均値の定性的変化は強磁性相転移を記述しており, $\beta > \beta_c$ における $m \neq 0$ の解は自発磁化の出現に対応していると解釈される.

上記の 1 体近似をもちいると, 1 スピンあたりの内部エネルギー密度 u も分配関数を経由することなしに

$$u = \frac{1}{N} \langle H(\mathbf{S}) \rangle = -\frac{J}{N} \sum_{\langle ij \rangle} \langle S_i S_j \rangle \simeq -\frac{J}{N} \sum_{\langle ij \rangle} \langle S_i \rangle \langle S_j \rangle = -\frac{zJ}{2} m^2 \quad (28)$$

と評価することができる. これから 1 スピンあたりの比熱 c は

$$c = \frac{\partial u}{\partial T} = -k_B \beta^2 \frac{\partial u}{\partial \beta} = k_B z J \beta^2 m \frac{\partial m}{\partial \beta} \quad (29)$$

となる. (26) を代入すると臨界温度 $T = T_c = (k_B \beta_c)^{-1} = zJ/k_B$ において

$$c = \begin{cases} 3k_B/2, & T \rightarrow T_c - 0 \\ 0, & T \rightarrow T_c + 0 \end{cases} \quad (30)$$

となることが示される.

相互作用の相手を平均値で置き換える, という大胆な近似を行ったにも関わらず相転移における比熱の不連続性といった定性的な特徴はうまくとらえられている. 素朴な発想にもとづいているため, 適用範囲も広い. こうした理由から, 分子場近似は相互作用下にある多体問題のもっとも基本的な解析法となっている.

3.2 ベーテ近似

分子場近似により, 計算量的困難を回避しながら相転移の特徴を定性的に再現することができた. しかしながら, 臨界温度 T_c の値など定量的な正確さという観点では改善の余地がある, たとえば, 外場のない 2 次元強磁性イジングスピン模型については厳密解から $k_B T_c / J \simeq 2.27$ であることがわかっている. 一方, 分子場近似での評価値は $k_B T_c / J = 4$ であり, かなり過大評価されている. 加えて, 厳密解においては比熱は $T = T_c$ で対数発散するが, 分子場近似では不連続であるものの発散はしないという違いもある. そこで, 分子場近似を改良する方法が多数検討されてきた. **ベーテ近似 (Bethe approximation)** はその代表例である [10].

ベーテ近似の基本的な発想は, 着目したスピンに対し, 最隣接相互作用は厳密に評価してそれより遠くに位置するスピンの影響は 1 体分布で近似する, というものである. 隣り合うスピンの

の相互作用は正しく考慮されているので分子場近似によるものよりも近似の精度は改善されると期待される。分子場近似の場合と同様、外場のない $h = 0$ の状況を想定しよう。着目する格子点を i としてハミルトニアンを

$$\begin{aligned} H(\mathbf{S}) &= -J \sum_{\langle ij \rangle} S_i S_j = -J S_i \sum_{j \in \partial i} S_j - J \sum_{\langle kl \rangle | k \neq i, l \neq i} S_k S_l \\ &= -J S_i \sum_{j \in \partial i} S_j + H_{\setminus i}(\mathbf{S} \setminus S_i) \end{aligned} \quad (31)$$

のように変形する。ただし、 $H_{\setminus i}(\mathbf{S} \setminus S_i)$ は $H(\mathbf{S})$ から S_i に関係する項をすべて除いた関数、すなわち、元のシステムから S_i を取り除いて得られるシステムのハミルトニアンを表現している。

ペーテ近似では、 $H_{\setminus i}(\mathbf{S} \setminus S_i)$ が $j \in \partial i$ のスピンに与える影響を補助的な外場 θ ($= 1$ 体分布のパラメータ) で代替する (図 1(c))。 S_i が不在システムを代替する外場なので、 θ はしばしば**キャビティ (空孔) 場 (cavity field)** とよばれる。これは $H(\mathbf{S})$ が i とその近傍の格子点 $j \in \partial i$ のスピンに関して、有効ハミルトニアン

$$H^{\text{eff}}(S_i, \{S_{j \in \partial i}\}) = -J S_i \sum_{j \in \partial i} S_j - \theta \sum_{j \in \partial i} S_j \quad (32)$$

で近似できることを意味している。 S_i と $\{S_{j \in \partial i}\}$ の結合分布を (32) に関するカノニカル分布によって $P(S_i, \{S_{j \in \partial i}\}) \propto \exp(-\beta H^{\text{eff}}(S_i, \{S_{j \in \partial i}\}))$ と表現する。 $x \in \{+1, -1\}$, $\forall A \in \mathbb{R}$ に対して

$$e^{Ax} = 2 \cosh(A) \times \frac{e^{Ax}}{2 \cosh(A)} = 2 \cosh(A) \times \frac{1 + x \tanh(A)}{2} \quad (33)$$

$$\sum_{x \in \{+1, -1\}} x \times \frac{1 + x \tanh(A)}{2} = \tanh(A) \quad (34)$$

がなりたつことを使い結合分布 $P(S_i, \{S_{j \in \partial i}\})$ を周辺化すると、 S_i の一体分布が

$$P(S_i) = \sum_{\{S_{j \in \partial i}\} \in \{+1, -1\}^z} P(S_i, \{S_{j \in \partial i}\}) = \frac{1 + S_i \tanh(\beta z \hat{\theta})}{2} \quad (35)$$

と評価されることがわかる。ただし、

$$\hat{\theta} = \frac{1}{\beta} \tanh^{-1}(\tanh(\beta J) \tanh(\beta \theta)) \quad (36)$$

とおいた。 $\hat{\theta}$ はしばしば**有効場 (effective field)** あるいは**キャビティバイアス (cavity bias)** とよばれる。これより、 S_i の熱平均値 $\langle S_i \rangle = m$ と θ を結ぶ関係式

$$m = \sum_{S_i \in \{+1, -1\}} S_i P(S_i) = \tanh(\beta z \hat{\theta}) \quad (37)$$

が得られる。

キャビティ場 θ の決め方をまだ述べていない。これには、 θ が $H_{\setminus i}(\mathbf{S} \setminus S_i)$ を代替するための外場である、ということをもちいる。元のシステムから S_i を取り除き、 i の近傍にある格子点 $j \in \partial i$ のスピン S_j に着目する。 S_i が取り除かれているため j の近傍にあるスピン数は $z - 1$ 個になるが、 S_j とその近傍にあるスピン $\{S_{k \in \partial j \setminus i}\}$ の組について、上と同様の議論を繰り返す。ただし、 S_i が取り除かれたシステムに関する評価なので、 S_j の熱平均値は外場 θ のみで定まる熱平均値と一致しなければならない。具体的には (37) において z を $z - 1$ に置き換えたものと $\tanh(\beta \theta)$ が一致しなければならない。このことから θ に関する自己無撞着方程式

$$\theta = (z - 1) \hat{\theta} = \frac{1}{\beta} (z - 1) \tanh^{-1}(\tanh(\beta J) \tanh(\beta \theta)) \quad (38)$$

が得られる。実際に計算を行う際には、(38)にもとづいてキャビティ場 θ を求めたのち、スピンの熱平均を(37)から求めるという手順になる。

2次元強磁性体イジング模型に対して、ベータ近似をもちいると、臨界温度は $k_B T_c / J \simeq 2.886\dots$ となり、分子場近似と比較して厳密解の評価値へ大幅に近づいている。最隣接相互作用を正しく評価しただけで、これほど近似精度が改良するのであれば正しく評価する相互作用の範囲を増やせば評価値はさらに改善されると期待される。こうしたアイデアは**クラスタ変分法 (cluster variation method)** として定式化されている [11, 12].

3.3 平均場近似が厳密な結果を与える例

平均場近似により得られる評価値は一般に厳密解からずれる。しかしながら、特別な性質を満たす対象では平均場近似の結果が適当な条件下で厳密解と一致する。こうした特別な例は適切な近似法を選択するための指針として重要である。

3.3.1 分子場近似：強磁性体の無限レンジ模型

正方格子における隣接格子点数 z と空間の次元 d の間には $z = 2d$ の関係がある。つまり、次元数を増やすほど相互作用する隣接スピンの数は増加する。この極限として各スピンの他のすべての $N - 1$ 個のスピンの相互作用するモデルは**無限レンジ (次元) 模型 (infinite range model)** とよばれる。強磁性体の無限レンジ模型のハミルトニアンは

$$H(\mathbf{S}) = -\frac{J}{N} \sum_{i>j} S_i S_j - h \sum_{i=1}^N S_i \quad (39)$$

で与えられる。相互作用係数が N^{-1} 倍されているのは、相互作用によって生じる典型的なエネルギーの値を有限次元の模型と同様に $O(N)$ とするためである。

強磁性体の無限レンジ模型では分子場近似が厳密解を与える。 $m = N^{-1} \sum_{i=1}^N S_i$ とすると(39)が

$$H(\mathbf{S}) = -N \left(\frac{Jm^2}{2} + hm + \frac{1}{N} \right) = -N \left(\frac{Jm^2}{2} + hm \right) + O(1) \quad (40)$$

と表現できることに着目する。一方、スターリングの公式 $\ln N! \simeq N \ln N - N$ をもちいると与えられた $m \in [-1, 1]$ の値に対応する状態 \mathbf{S} の数は

$$\begin{aligned} & \frac{N!}{(N(1+m)/2)! (N(1-m)/2)!} \\ &= \exp \left(-N \left(\frac{1+m}{2} \ln \left(\frac{1+m}{2} \right) + \frac{1-m}{2} \ln \left(\frac{1-m}{2} \right) \right) + O(1) \right) \end{aligned} \quad (41)$$

と評価される。これらを組み合わせると、分配関数について

$$Z(\beta) = \text{Tr}_{\mathbf{S}} \exp(-\beta H(\mathbf{S})) = \int_{-1}^{+1} dm \exp(-N\beta\phi(m; \beta) + O(1)) \quad (42)$$

という表現が得られる。ただし

$$\phi(m; \beta) = -\frac{Jm^2}{2} - hm + \frac{1}{\beta} \left(\frac{1+m}{2} \ln \left(\frac{1+m}{2} \right) + \frac{1-m}{2} \ln \left(\frac{1-m}{2} \right) \right) \quad (43)$$

とした。

(42) は $N \rightarrow \infty$ では、1自由度あたりの自由エネルギーが

$$f = -\frac{1}{N\beta} \ln Z(\beta) = \min_{m \in [-1, +1]} \{\phi(m; \beta)\} \quad (44)$$

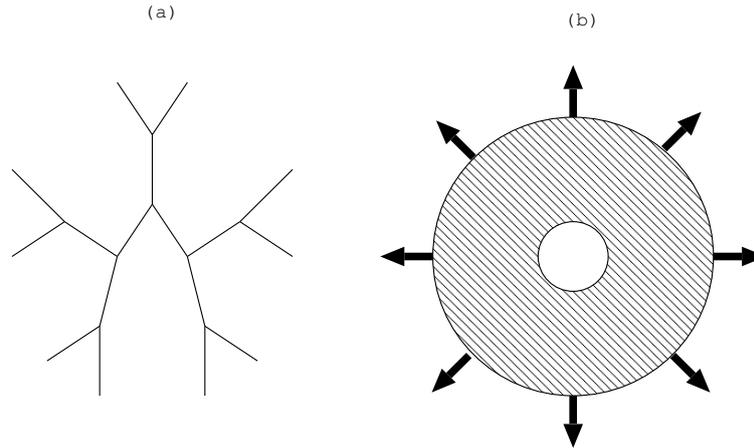


図 2: (a) : $z = k + 1 = 3$ のケーリーツリー. (b) : 十分大きなケーリーツリー (斜線の部分) の中心付近 (白い部分) がベーテ格子.

によって厳密に評価できることを意味している. 極値条件 $\partial\phi(m; \beta)/\partial m = 0$ から

$$m = \tanh(\beta(Jm + h)) \quad (45)$$

が得られる³. 一方, 外場のある無限レンジ模型に対する分子場近似の自己無撞着方程式は

$$m = \tanh\left(\beta\left(\frac{(N-1)J}{N}m + h\right)\right) \quad (46)$$

であり, $N \rightarrow \infty$ の極限で (45) に帰着する. (45) は厳密解を与える方程式なので, このことは強磁性体の無限レンジ模型については, $N \rightarrow \infty$ で分子場近似が厳密な結果を導くことを意味している.

3.3.2 ベーテ近似 (1) : ベーテ格子

中心点から $z = k + 1$ 本の辺を伸ばし, それぞれの辺の空いている方の端点のそれぞれからさらに k 本の辺を伸ばすことを適当な回数繰り返す. このようにして出来たツリー状のグラフは一般に**ケーリーツリー (Cayley tree)** とよばれる (図 2(a)). また, 十分大きなケーリーツリーの中心付近は特に**ベーテ格子 (Bethe lattice)** とよばれる (図 2(b)). ケーリーツリーの上で強磁性体のイジング模型を定義する. ベーテ格子に対し, ベーテ近似はこの模型の厳密解を与える.

ケーリーツリーが中心点から G 回辺を伸ばすことで作られているとし, 中心点からツリーの外に向けて頂点を同心円状に第 0 世代, 第 1 世代, ..., 第 G 世代と分類する. 任意の世代 g の頂点を 1 つ選びそこから G 世代まで世代を増やしてできる部分木に限定したスピン系を考える. 対称性から, この系の分配関数は世代 g のみに依存し選んだ頂点の位置の詳細には依存しない. よって, これを Z_g と記す. また, 選んだ頂点のスピンを残し残りのスピンについて周辺化することで得られる 1 体分布は, 適当な有効外場 θ_g をもちいて必ず $P(S) = \exp(\beta\theta_g S)/(2 \cosh(\beta\theta_g))$ と表現できることに注意する.

$g-1$ 世代の, ある頂点を 0, そこから伸びている k 本の辺の端点にあたる g 世代の頂点を $1, 2, \dots, k$ と記す. S_0 と S_1, \dots, S_k との間の相互作用を考察することで, $g \geq 2$ に対し, Z_g と Z_{g-1} および θ_g と θ_{g-1} に関する以下の漸化式が得られる.

$$Z_{g-1} = Z_g^k \times \sum_{S_0, S_1, S_2, \dots, S_k} \exp\left(\beta\left(h + J \sum_{j=1}^k S_j\right) S_0\right) \prod_{j=1}^k \frac{\exp(\beta\theta_g S_j)}{2 \cosh(\beta\theta_g)}$$

³(45) の解は必ずしも, スピンの熱平均値 $N^{-1} \sum_{i=1}^N \langle S_i \rangle$ (並進対称性から各格子点上のスピンの熱平均値と一致) を意味しない. $h \sim O(1)$ の場合, $\phi(m; \beta)$ を最小にする m の値は 1 つであり, その値は $N^{-1} \sum_{i=1}^N \langle S_i \rangle$ と一致する. 一方, $h \rightarrow 0$ の場合, 強磁性相を意味する $\beta > J^{-1}$ では $m = \pm m^*$ ($m^* > 0$) という 2 つの解が $\phi(m; \beta)$ を最小にする. この場合, $N^{-1} \sum_{i=1}^N \langle S_i \rangle$ は $h \rightarrow 0$ の近づけ方の詳細に応じて区間 $[-m^*, +m^*]$ に含まれる任意の値を取り得る.

$$\begin{aligned}
&= (2 \cosh(\beta J))^k Z_g^k \\
&\quad \times \sum_{S_0 \in \{+1, -1\}} \exp(\beta h S_0) \left(\frac{1 + S_0 \tanh(\beta J) \tanh(\beta \theta_g)}{2} \right)^k \\
&= \left(\frac{2 \cosh(\beta J) Z_g}{2 \cosh(\beta \hat{\theta}_g)} \right)^k \times 2 \cosh(\beta(h + k \hat{\theta}_g))
\end{aligned} \tag{47}$$

$$\theta_{g-1} = h + k \hat{\theta}_g \tag{48}$$

ただし,

$$\hat{\theta}_g = \frac{1}{\beta} \tanh^{-1} (\tanh(\beta J) \tanh(\beta \theta_g)) \tag{49}$$

とした.

一方, 中心点のスピンの評価となる $g = 1$ については $k \rightarrow k + 1$ とした

$$\begin{aligned}
Z_0 &= Z_1^{k+1} \times \sum_{S_0, S_1, S_2, \dots, S_{k+1}} \exp \left(\beta \left(h + J \sum_{j=1}^{k+1} S_j \right) S_0 \right) \prod_{j=1}^{k+1} \frac{\exp(\beta \theta_1 S_j)}{2 \cosh(\beta \theta_1)} \\
&= (2 \cosh(\beta J))^{k+1} Z_1^{k+1} \\
&\quad \times \sum_{S_0 \in \{+1, -1\}} \exp(\beta h S_0) \left(\frac{1 + S_0 \tanh(\beta J) \tanh(\beta \theta_1)}{2} \right)^{k+1} \\
&= \left(\frac{2 \cosh(\beta J) Z_1}{2 \cosh(\beta \hat{\theta}_1)} \right)^{k+1} \times 2 \cosh(\beta(h + (k + 1) \hat{\theta}_1))
\end{aligned} \tag{50}$$

$$\theta_0 = h + (k + 1) \hat{\theta}_1 \tag{51}$$

という漸化式になる. Z_0 は全系 (ケーリーツリー上で定義されたスピン系) の分配関数, $\tanh(\beta \theta_0)$ は中心点におけるスピンの熱平均値を意味している.

十分大きな G に対し, ケーリーツリーの境界 (G 世代目のスピン) から漸化式 (47), (48), (49) を更新しながら中心点まで世代を遡る. このとき, 中心付近, すなわち, ベーテ格子に対応する部分では漸化式が十分沢山更新されている. よって, $\theta_g, \hat{\theta}_g$ は固定点の値に収束していると考えられる. それは非線形連立方程式

$$\hat{\theta} = \frac{1}{\beta} \tanh^{-1} (\tanh(\beta J) \tanh(\beta \theta)) \tag{52}$$

$$\theta = h + k \hat{\theta} \tag{53}$$

の解である. また, 中心点のスピン S_c の熱平均値はこの解にもとづいて

$$\langle S_c \rangle = \tanh(\beta(h + (k + 1) \hat{\theta})) \tag{54}$$

と評価される. $z = k + 1$ を代入すると, (52), (53) は外場のない系に対するベーテ近似の自己無撞着方程式 (38), (54) は有効場から磁化を求める式 (37) に他ならない. 上述の漸化式による評価はケーリーツリー全体で厳密になりたっている⁴. このことはベーテ格子ではベーテ近似が厳密な結果を導くことを意味している.

⁴ケーリーツリーでは, 境界にあるスピン数がバルクにあるスピン数よりも圧倒的に多いことに注意. そのため, 漸化式にもとづく解析は厳密解を与えるが, それは境界条件と世代に強く依存する. 通常の物質で観察される一様な性質はベーテ格子に対応する中心付近でしか得られない.

3.3.3 ベーテ近似 (2) : スパースランダムグラフ

ベーテ格子上の強磁性体のイジング模型に対して、ベーテ近似が厳密な結果を与えることを示した。ただし、十分大きなケーリーツリーの中心付近といった複雑な定義であるベーテ格子には、ベーテ近似を厳密に成立させるために人工的に考案されたモデルシステムであるといった印象が否めない。ところが、各頂点に対し $z \sim O(1)$ 個の近傍点をランダムに定めてできる無限に大きなスパースランダムグラフとベーテ格子との等価性が、近年、徐々に認識されるようになった。スパースランダムグラフは、複雑ネットワーク、情報通信、理論計算機科学などの問題などで広く知られる“自然な”数理モデルである。このことから、現在、ベーテ近似は有限次元の格子系に対する近似手法としての位置づけに加えて、こうした非物理系の数理モデルの厳密な解析手法として利用される機会も増えている [13, 14]。

スパースランダムグラフとベーテ格子との等価性はグラフ内に存在する循環経路 (ループ) の長さを見積もるとわかりやすい。グラフ内の各頂点から出発し、辺を伝いながら頂点を經由し移動していく。逆戻りは許さず、また、同じ辺は2度と通らないとしたとき、何ステップで出発した頂点に戻ることができるかでループの長さの特徴づける。ツリーはすべての頂点に対してループ長が無限大のグラフである。 $d \geq 2$ 次元以上の正方格子であれば次元によらずすべての頂点に対してループ長は4である。一方、 $z \sim O(1)$ である限り、ランダムに構成されたグラフではループ長は典型的な頂点に対して $O(\ln N)$ 程度となり $N \rightarrow \infty$ で発散する。つまり、十分大きなスパースランダムグラフの局所的な構造はツリーに漸近する。

このことは、着目した頂点から順に $z \sim O(1)$ 個の近傍点をランダムに選んでいくと、適当な世代 L までは、それまでに選ばれた頂点が再度選ばれる確率が $N \rightarrow \infty$ でゼロに漸近する、という性質にもとづいている。ただし、 L 世代までに含まれる頂点の数が N と同程度になると、それまでに選ばれた頂点が再度近傍に登場するようになり、ループが形成される。このときの L が典型的なループ長の特徴づける。臨界条件を見積もると

$$\begin{aligned} & 1 + z + z(z-1) + z(z-1)^2 + \dots + z(z-1)^{L-1} \\ &= 1 + \frac{z(z-1)^L - z}{z-2} \sim N \end{aligned} \quad (55)$$

となり、したがって

$$L \sim \frac{\ln N}{\ln(z-1)} \rightarrow \infty \quad (N \rightarrow \infty) \quad (56)$$

が結論づけられる。

4 確率推論のための平均場近似

強磁性体のイジング模型を例として、統計力学における計算量的困難の問題が平均場近似によりどのように解決されるか、について述べてきた。しかしながら、相互作用係数や外部磁場が格子上で一様である強磁性体のモデルは確率モデルとしては特殊すぎるため、それに特化して表現された平均場近似はそのままでは一般のシステムには使えない。本節では、分子場近似、ベーテ近似にもとづいて一般の確率推論に使いやすい平均場近似の表現を求める。

4.1 ナイーブ平均場近似

頂点の集合 V と辺の集合 E で特徴づけられる一般的なグラフ $G = (V, E)$ 上で定義されたイジングスピン系を考える。ハミルトニアンは

$$H(\mathbf{S}) = - \sum_{(ij) \in E} J_{ij} S_i S_j - \sum_{i \in V} h_i S_i \quad (57)$$

で定義されるとする。強磁性体のモデルとは異なり、 J_{ij} , h_i は辺 (ij) および頂点 i に依存して異なる値を取るとする。

この模型に対して、分子場近似の発想を適用しよう。分子場近似は着目した頂点 i の近傍 ∂i のスピンをすべて熱平均値に置き換えることで S_i に関する有効ハミルトニアンを構成し、 S_i の熱平均の評価を 1 体問題に帰着させることにもとづいていた。これを適用し

$$H_i^{\text{eff}}(S_i) = -S_i \left(h_i + \sum_{j \in \partial i} J_{ij} \langle S_j \rangle \right) = -S_i \left(h_i + \sum_{j \in \partial i} J_{ij} m_j \right) \quad (58)$$

によって S_i に関する有効ハミルトニアンを定義する。(57)には並進対称性がないため、スピンの熱平均値は頂点ごとに異なる。そのため、頂点に関する依存性を持たせ $\langle S_j \rangle = m_j$ とおいた。

(58) から S_i に関する 1 体分布は $P(S_i) = \exp(-\beta H_i^{\text{eff}}(S_i)) / (\sum_{S_i \in \{+1, -1\}} \exp(-\beta H_i^{\text{eff}}(S_i)))$ と定義される。これをもちいて、 S_i の熱平均値 $\langle S_i \rangle = m_i$ を評価すると

$$m_i = \sum_{S_i \in \{+1, -1\}} S_i \frac{e^{\beta(h_i + \sum_{j \in \partial i} J_{ij} m_j) S_i}}{2 \cosh \beta(h_i + \sum_{j \in \partial i} J_{ij} m_j)} = \tanh \left(\beta \left(h_i + \sum_{j \in \partial i} J_{ij} m_j \right) \right) \quad (59)$$

が $i = 1, 2, \dots, N$ のそれぞれに対して得られる。

強磁性体モデルの場合と異なり、(59)は 1 変数で閉じた方程式ではない。しかしながら、 N 個の未知変数 m_1, m_2, \dots, m_N を決定するのに十分な N 個の条件式を与えている。この連立非線形方程式を解析的に解くことは一般に難しいが反復代入法などにより数値解を求めることは今日の標準的な計算機をもちいれば容易である。1 度の反復に必要な計算量は 1 スピンあたり近傍サイズ程度である。収束までに十分な回数反復することが必要であるが、それを考慮しても熱平均値を厳密に評価するために必要な計算量 $O(2^N)$ と比較すると圧倒的に少ない計算コストで実行することができる。分子場近似を一般化したこのような近似計算法はしばしば**ナイーブ平均場近似 (naive mean field approximation)** とよばれる。

4.1.1 例5) 画像ノイズ除去

ナイーブ平均場近似の応用例として確率推論にもとづいたデジタル画像のノイズ除去を紹介しよう [15]。デジタルカメラなどで画像を撮影した場合には、手ぶれや熱雑音などによって画像が劣化することがある。画像ノイズ除去とはそうしたノイズの影響を何らかの指針にしたがって除去する問題である。

デジタル画像は 2 次元正方格子の各格子点にピクセル値とよばれる画像情報をあらわす数値を割り当てることで構成されている。数値の表現にはいくつかの方法があるが、白黒画像の場合にはイジングスピンのように 2 値のピクセル値 $x_i \in \{+1, -1\}$ をもちいてあらわすこともできる。さて、自然画像では隣り合うピクセル値は同じ値を取る傾向がある。この傾向は強磁性体のイジングスピン模型と同じである。そこで自然画像の生成に対応する確率モデルとして事前分布

$$P(\mathbf{x}) = \frac{\exp \left(\beta \sum_{(ij)} x_i x_j \right)}{Z} \quad (60)$$

($\beta > 0$) を仮定しよう。ただし、 $Z = \text{Tr}_{\mathbf{x}} \exp \left(\beta \sum_{(ij)} x_i x_j \right)$ であり、 (ij) は 2 次元正方格子の最隣接格子対をあらわすものとする。

劣化過程にはさまざまなタイプのものが考えられるが、ここでは、本来のピクセル値が格子点の位置によらず同一の確率 $p (< 1)$ で反転する 2 元対称通信路を考えることにする。 $\alpha = (1/2) \ln((1-p)/p)$ とすると原画像 \mathbf{x} に対して、劣化画像 \mathbf{y} が得られる確率は条件付き確率

$$P(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^N \frac{\exp(\alpha y_i x_i)}{2 \cosh(\alpha)} \quad (61)$$

で与えられる。(60), (61) より、劣化画像 \mathbf{y} を観測した際の \mathbf{x} の事後分布は、ベイズの公式から

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \propto \exp \left(\beta \sum_{(ij)} x_i x_j + \alpha \sum_{i=1}^N y_i x_i \right) \quad (62)$$

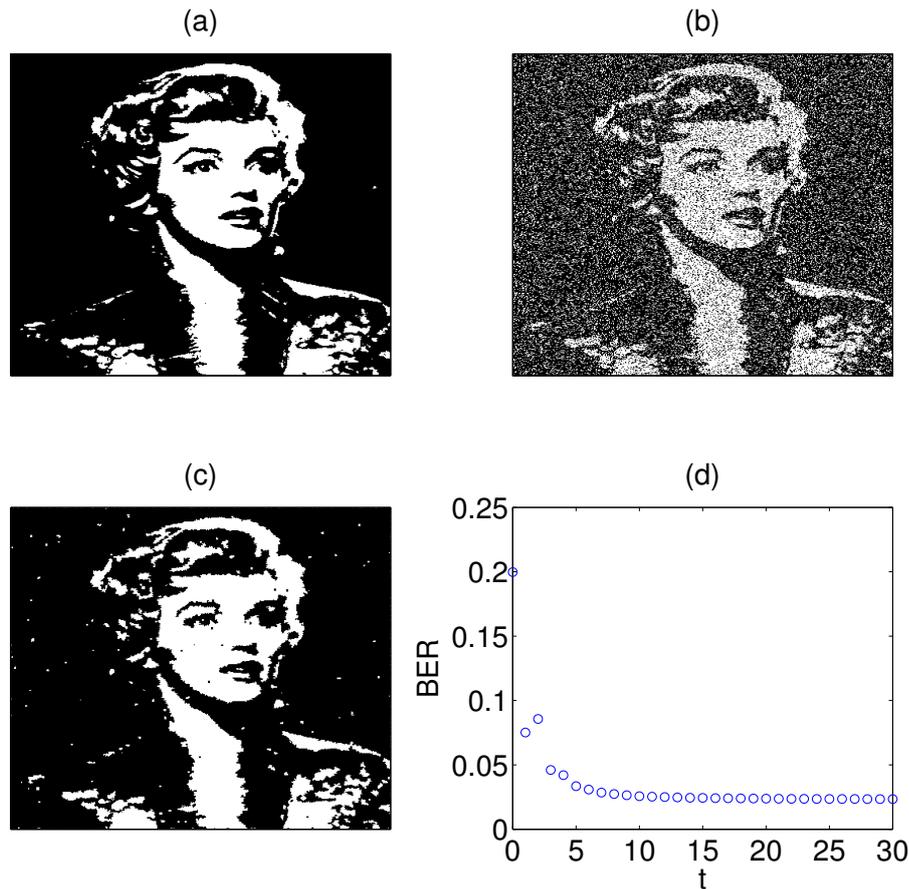


図 3: ナイーブ平均場近似をもちいた画像ノイズ除去. 2次元強磁性体イジング模型において強磁性相転移が生じる臨界逆温度に対応させて $\beta \simeq 1/2.27$ とした. また, 反転確率 p は既知であると し $\alpha = (1/2) \ln((1-p)/p)$ とした. (a): 原画像. (b): 20%の画素値が反転した劣化画像. (c): 復元された画像. (d): (63) を反復代入した回数と復元画像の現画像に対するビット誤り率との関係.

となる.

2.2.1 節と同様の考察により, 上記の確率モデルが正しいとした仮定の下では各格子点において (62) の周辺分布を最大にするようにピクセル値を与える方法がビット誤り率を最小にする意味で最適である. イジングスピン表現 $x_i \in \{+1, -1\}$ をもちいた場合, これは各スピンの事後分布による平均値の符号を推定結果とする方法とみなすことができる. ナイーブ平均場近似をもちいると事後分布による平均値は平均場方程式

$$m_i = \tanh \left(\alpha y_i + \beta \sum_{j \in \partial i} m_j \right) \quad (63)$$

から近似的に求めることができる. この方程式の解にもとづいて $\hat{x}_i = \text{sign}(m_i) = m_i/|m_i|$ により, ノイズ除去後の画像が得られる (図 3).

ここに述べた方法を実行するためには, 事前分布と条件付き確率の詳細を定めるハイパーパラメータ α, β を与える必要がある. 現実的な状況では, α や β はあらかじめわかっていない場合が多い. そのような場合には, 劣化画像 \mathbf{y} の尤度 $P(\mathbf{y}) = \sum_{\mathbf{x}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$ を最大化するように α, β を推定する方法が考えられる [16]. ただし, 尤度の評価も計算量的に困難であるため, その解決のためにも適宜ナイーブ平均場近似がもちいられる.

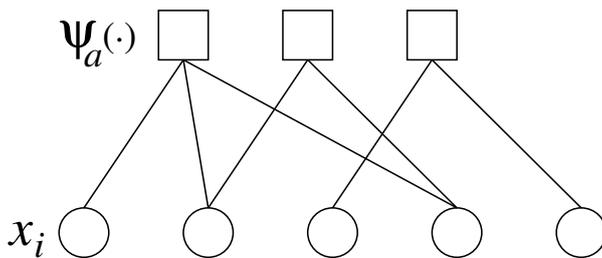


図 4: $P(\mathbf{x}) = (1/Z)\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2)\psi_3(\mathbf{x}_3)$ ($\mathbf{x}_1 = (x_1, x_2, x_4), \mathbf{x}_2 = (x_2, x_4), \mathbf{x}_3 = (x_3, x_5)$) に関するファクターグラフ. ○ノードは左から順に x_1, x_2, \dots, x_5 , □ノードは左から順に $\psi_1(\mathbf{x}_1), \psi_2(\mathbf{x}_2), \psi_3(\mathbf{x}_3)$ をあらわしている.

4.2 キャビティ法

ここまでは、変数間に依存関係がある場合として2体相互作用を含むハミルトニアンで表現されるモデルを考えてきた。しかしながら、一般の確率モデルでは、相互作用が3つ以上の変数を含む関数で表現される場合やそもそも確率モデルがハミルトニアンの形式で表現されていない場合も多い。**キャビティ法 (cavity method)** はそういった一般的な確率モデルに対してベータ近似を拡張した方法であると位置づけられる [6, 7].

4.2.1 ファクターグラフ

N 次元の確率変数 $\mathbf{x} = (x_i)$ に関する結合分布が

$$P(\mathbf{x}) = \frac{1}{Z} \prod_a \psi_a(\mathbf{x}_a) \tag{64}$$

のように表現されている状況を想定する⁵。 $\psi_a(\mathbf{x}_a) (\geq 0)$ はポテンシャル関数とよばれ \mathbf{x} に含まれる一部の成分の組 \mathbf{x}_a の関数である。たとえば、強磁性体のイジング模型の場合、隣接格子点対 (ij) が a , $\exp(\beta JS_i S_j)$ がポテンシャル関数, (S_i, S_j) が \mathbf{x}_a にそれぞれ対応する。 Z は分配関数に対応する。

x_i が関係しているポテンシャル関数に関する添字の集合を ∂i , \mathbf{x}_a に含まれる成分の添字の集合を ∂a とあらわす。確率変数の各成分を○, 各ポテンシャル関数を□といった2種類のノードで表現し, すべての a に対し, 対応する□ノードと ∂a に含まれる添字に対応するすべての○ノードを辺でつなぐ。すると, 結合分布 (64) の変数間の依存関係は**ファクターグラフ (factor graph)** とよばれる向きのない2部グラフで表現される (図4)。

変数間の依存関係をグラフ的に表現する方法には, これ以外にも (64) を条件付き確率の積で表現しなおし向き付きグラフで表現する方法や1種類のノードのみで表現する方法などいくつかの種類がある。それらはしばしば**グラフィカルモデル (graphical model)** と総称される。ファクターグラフによる表現では

確率変数に対応する○ノードが2つの互いに交わりのない異なる部分グラフに含まれる \implies 2つの確率変数は統計的に独立である

という命題が一般になりつつ, 理想気体と強磁性イジング模型との比較が示すように, 確率変数の統計的独立性は確率推論における計算コストに大きく影響する。このことは確率モデルのグラフ表現が確率推論で必要となる計算コストと密接に関係していることを意味している。

⁵変数が1つだけ含まれているポテンシャル関数を特別視する表現もある。そうした表現はクラスタ変分法など, より進んだ近似法へ拡張する場合に都合がよい。ただし, 本稿のレベルでは特に利点がないので, ここでは表現の簡潔性を優先しすべてのポテンシャル関数を同等に扱う。

4.2.2 周辺分布とキャビティ分布との間になりたつ関係式

2.2.1 節に示したように確率推論では結合分布 (64) からの周辺分布 $P_i(x_i) = \text{Tr}_{\mathbf{x} \setminus x_i} P(\mathbf{x})$ の評価を求められることが多い。これに関連して、以下の等式が任意のシステムに対して厳密に成立することに着目する。

$$P_i(x_i) = \frac{\langle \prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \rangle_{\setminus x_i}}{\text{Tr}_{x_i} \langle \prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \rangle_{\setminus x_i}} \quad (65)$$

ただし、 $\langle \dots \rangle_{\setminus x_i}$ は x_i に直接関係しないすべてのポテンシャル関数によって定義される x_i を除いた確率変数の組 $\mathbf{x} \setminus x_i$ に対する結合分布（しばしば**キャビティ分布 (cavity distribution)** とよばれる）

$$P_{\setminus i}(\mathbf{x} \setminus x_i) = \frac{\prod_{b \notin \partial i} \psi_b(\mathbf{x}_b)}{\text{Tr}_{\mathbf{x} \setminus x_i} \prod_{b \notin \partial i} \psi_b(\mathbf{x}_b)} \quad (66)$$

に関する平均をあらわしている。(65) は、 x_i に直接関わるすべてのポテンシャル関数の積をキャビティ分布で平均して得られる有効ポテンシャル関数 $\psi_i^{\text{eff}}(x_i) = \langle \prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \rangle_{\setminus x_i}$ を規格化したものが x_i の周辺分布に一致する、ということの意味している。

[証明] ポテンシャル関数の積を x_i を含むものと含まないものに分けて

$$P(\mathbf{x}) = \frac{\prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \prod_{b \notin \partial i} \psi_b(\mathbf{x}_b)}{\text{Tr}_{\mathbf{x}} \prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \prod_{b \notin \partial i} \psi_b(\mathbf{x}_b)} \quad (67)$$

と表現する。さらに右辺の分母と分子をともに定数 $\text{Tr}_{\mathbf{x} \setminus x_i} \prod_{b \notin \partial i} \psi_b(\mathbf{x}_b)$ で割ると (66) から

$$P(\mathbf{x}) = \frac{(\prod_{a \in \partial i} \psi_a(\mathbf{x}_a)) P_{\setminus i}(\mathbf{x} \setminus x_i)}{\text{Tr}_{\mathbf{x}} (\prod_{a \in \partial i} \psi_a(\mathbf{x}_a)) P_{\setminus i}(\mathbf{x} \setminus x_i)} \quad (68)$$

が得られる。周辺化 Tr_{x_i} を行い、さらに分母に対して $\text{Tr}_{\mathbf{x}}(\dots) = \text{Tr}_{x_i} \{ \text{Tr}_{\mathbf{x} \setminus x_i}(\dots) \}$ であることをもちいると (65) が得られる。

4.2.3 ビリーフプロパゲーション

(65) は有効ポテンシャル $\psi_i^{\text{eff}}(x_i) = \langle \prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \rangle_{\setminus x_i}$ の評価に要する計算量と周辺分布 $P_i(x_i)$ を評価するために必要な計算量が同程度であることを意味している。残念ながら一般にはキャビティ分布 $P_{\setminus i}(\mathbf{x} \setminus x_i)$ にも変数間に複雑な依存関係があり、この分布に関する平均が必要な $\psi_i^{\text{eff}}(x_i)$ の評価は計算量的に困難である。しかしながら、 $P(\mathbf{x})$ を表現したファクターグラフがループを含まない**ハイパーツリー (hypertree)** (図5(a)) となる場合には、効率的に $\psi_i^{\text{eff}}(x_i)$ や $P_i(x_i)$ を評価することが可能になる。

各 $a \in \partial i$ に対し、 $\psi_a(\mathbf{x}_a)$ をのぞいた系 (a キャビティ系とよぶ) に関する \mathbf{x}_a に含まれる成分 x_j の周辺 (1 体) 分布関数が評価できているとして $m_{j \rightarrow a}(x_j)$ と書く。ファクターグラフがハイパーツリーの場合には

- $a, b, c, \dots \in \partial i$ について $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \dots$ に共通して含まれる変数は x_i のみである
- x_i をあらわす○ノードをとりのぞくと $a, b, c, \dots \in \partial i$ に対応する□ノードを含む部分グラフは、それぞれ互いに交わりを持つことなく分離される

という性質がすべての○ノードについてなりたつ (図5(b))。前述のとおり、交わりを持たない異なる部分グラフに含まれる確率変数は統計的に独立である。つまり、ハイパーツリーで表現され

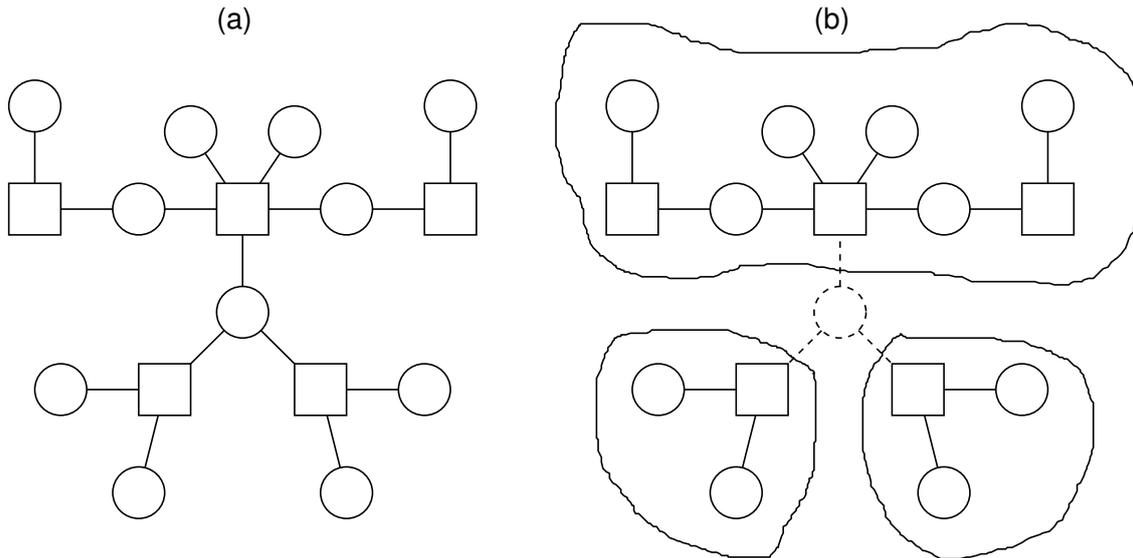


図 5: (a) : ハイパーツリー=ループを含まないファクターグラフの例. (b) : ハイパーツリーではすべての○ノードについて、そのノードを除去するとグラフの残りの部分が必ず互いに交わりを持たない部分グラフに分離される.

るシステムのキャビティ分布に関しては、統計的独立性にもとづき部分グラフ間で独立して平均評価をおこなうことが可能になり

$$\begin{aligned}\psi_i^{\text{eff}}(x_i) &= \text{Tr}_{\mathbf{x} \setminus x_i} \left(\prod_{a \in \partial i} \psi_a(\mathbf{x}_a) \right) P_{\setminus i}(\mathbf{x} \setminus x_i) \\ &= \prod_{a \in \partial i} \left(\text{Tr}_{\mathbf{x}_a \setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j \in \partial a \setminus i} m_{j \rightarrow a}(x_j) \right)\end{aligned}\quad (69)$$

が厳密になりつつ、 ∂i および、すべての $a \in \partial i$ に対する ∂a 、の大きさがともに $O(1)$ である限り、(69) の右辺は $O(1)$ 程度の計算コストで評価できる。すなわち、計算量的困難は生じない。同様に a キャビティ系の有効ポテンシャル $\psi_{i \rightarrow a}^{\text{eff}}(x_i)$ に関する厳密な表現

$$\psi_{i \rightarrow a}^{\text{eff}}(x_i) = \prod_{b \in \partial i \setminus a} \left(\text{Tr}_{\mathbf{x}_b \setminus x_i} \psi_b(\mathbf{x}_b) \prod_{j \in \partial b \setminus i} m_{j \rightarrow b}(x_j) \right)\quad (70)$$

が得られるが、この右辺の評価にも計算量的困難は生じない。ところで、定義から $m_{i \rightarrow a}(x_i)$ とは $\psi_{i \rightarrow a}^{\text{eff}}(x_i)$ を規格化したものに他ならない。このことから

$$m_{i \rightarrow a}(x_i) = \frac{\prod_{b \in \partial i \setminus a} \left(\text{Tr}_{\mathbf{x}_b \setminus x_i} \psi_b(\mathbf{x}_b) \prod_{j \in \partial b \setminus i} m_{j \rightarrow b}(x_j) \right)}{\text{Tr}_{x_i} \prod_{b \in \partial i \setminus a} \left(\text{Tr}_{\mathbf{x}_b \setminus x_i} \psi_b(\mathbf{x}_b) \prod_{j \in \partial b \setminus i} m_{j \rightarrow b}(x_j) \right)}\quad (71)$$

あるいは2種類の分布関数にわけた表現

$$m_{a \rightarrow i}(x_i) = \alpha_{a \rightarrow i} \text{Tr}_{\mathbf{x}_a \setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j \in \partial a \setminus i} m_{j \rightarrow a}(x_j)\quad (72)$$

$$m_{i \rightarrow a}(x_i) = \alpha_{i \rightarrow a} \prod_{b \in \partial i \setminus a} m_{b \rightarrow i}(x_i)\quad (73)$$

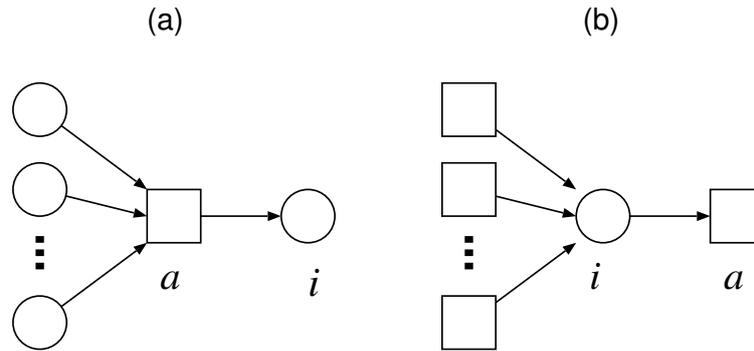


図 6: (a): 更新式 (72) によるメッセージ伝搬の様子. (b): 更新式 (73) によるメッセージ伝搬の様子. ビリーフプロパゲーションはファクターグラフを構成する各ノードで (a), (b) のようなメッセージ ($m_{a \rightarrow i}(x_i), m_{i \rightarrow a}(x_i)$) のやり取りを行うアルゴリズムとみなすことができる.

が得られる. ただし $\alpha_{a \rightarrow i}, \alpha_{i \rightarrow a}$ はそれぞれ $m_{a \rightarrow i}(x_i), m_{i \rightarrow a}(x_i)$ が分布関数になるように導入された規格化定数である. $m_{a \rightarrow i}(x_i)$ をもちいると, x_i の周辺分布は

$$P_i(x_i) = \alpha_i \prod_{a \in \partial i} m_{a \rightarrow i}(x_i) \quad (74)$$

(α_i は規格化定数) により厳密に評価される.

以上は a キャビティ系での $j \in \partial a$ に関する 1 体分布関数 $m_{j \rightarrow a}(x_j)$ が得られていることを前提とした議論であったが, (72), (73) はファクターグラフの各辺に対して定義された 2 つの 1 体分布関数 $m_{a \rightarrow i}(x_i), m_{i \rightarrow a}(x_i)$ のすべてを逐次的に定めるのに十分な条件を与えている. すべての添字 a, i について, $\partial a, \partial i$ に含まれる要素の数がそれぞれ $O(1)$ である場合, すなわち, ハイパーツリーがスパースな場合には (72) および (73) の右辺の評価に必要な計算コストは高々 $O(1)$ である. $a \rightarrow i$ および $i \rightarrow a$ はグラフの各辺について向きを付けることに対応している. つまり, すべての確率変数の成分に対して (72), (73) を評価したとしても計算コストは辺の数の 2 倍に比例する程度である. その結果を (74) に代入すれば, 全体として確率変数の数に比例する程度の計算コストですべての確率変数について周辺分布が求まる. (72), (73) および (74) にもとづいたハイパーツリーに関する効率的な周辺分布の評価法は **ビリーフプロパゲーション (belief propagation)** あるいは **確率伝搬法 (probability propagation)** とよばれる. (72), (73) は一見複雑であるが, ファクターグラフ内の a と i を結ぶ辺について

- (72): その辺以外から \square ノード (a) に入ってくるメッセージ ($m_{j \rightarrow a}(x_j)$) を集めて和 ($\text{Tr}_{\mathbf{x}_a \setminus x_i}$) を評価し, \circ ノード (i) にメッセージ ($m_{a \rightarrow i}(x_i)$) を送る
- (73): その辺以外から \circ ノード (i) に入ってくるメッセージ ($m_{b \rightarrow i}(x_i)$) を集めて積 ($\prod_{b \in \partial i \setminus a}$) を評価し, \square ノード (a) にメッセージ ($m_{i \rightarrow a}(x_i)$) を送る

アルゴリズムとしてグラフィカルに表現するとわかりやすい (図 6). そのため分野によっては **和積アルゴリズム (sum-product algorithm)** とよばれることもある.

2 体相互作用を \square ノードで表現すれば, 1 次元鎖やケーリーツリーもハイパーツリーとして表現できる. 1 次元鎖について知られている転送行列法, 3.3.2 節で示したケーリーツリー上の漸化式はビリーフプロパゲーションの特別な場合と位置づけることができる.

4.2.4 ルーピービリーフプロパゲーション

ファクターグラフがハイパーツリーになる場合にはビリーフプロパゲーションによって周辺分布を効率的に評価できることがわかった. 残念ながら, 一般の確率モデルに対応するファクターグラフはハイパーツリーにはならない. ところが (72), (73), (74) を見直してみると, それぞれの右辺にある評価式は, グラフの局所的な情報しかもちいられておらず, ファクターグラフにループが含まれる/含まれないといった大域的な性質とは関係なく実行可能であることがわかる. そのため,

ループを含む一般のグラフに対しても、ビリーフプロパゲーションを近似的に適用する方が考えられる。このような近似的な評価を目的として一般のグラフに対し適用されるビリーフプロパゲーションはしばしば**ルーピービリーフプロパゲーション (loopy belief propagation)** とよばれる。ただし、ハイパーツリーでない場合、(72), (73) をグラフ上で一度伝搬させても $m_{a \rightarrow i}(x_i)$, $m_{i \rightarrow a}(x_i)$ は確定しない。そのため、これらの関数が収束したとみなされるまで適当な回数更新を反復する必要がある。

強磁性体のイジング模型に関して、 $\exp(\beta J_{ij} S_i S_j)$ および $\exp(\beta h S_i)$ をポテンシャル関数とみなし、有効場 $\hat{\theta}$, 空孔場 θ をもちいて $m_{a \rightarrow i}(S_i) = e^{\beta \hat{\theta} S_i} / (2 \cosh(\beta \hat{\theta}))$ ($\psi_a(\mathbf{x}_a)$ が $\exp(\beta J_{ij} S_i S_j)$ に対応するとき), $e^{\beta h S_i} / (2 \cosh(\beta h))$ ($\psi_a(\mathbf{x}_a)$ が $\exp(\beta h S_i)$ に対応するとき), $m_{i \rightarrow a}(S_i) = e^{\beta \theta S_i} / (2 \cosh(\beta \theta))$ と表現すれば、3.2 節に示したベータ近似はルーピービリーフプロパゲーションとして解釈することができる。

4.2.5 例6) 低密度パリティ検査符号の復号

(ルーピー) ビリーフプロパゲーションの応用例として低密度パリティ検査符号の復号を紹介する。情報を送信したり、保存したりする際にはノイズによる劣化に備えて情報表現に誤り訂正機能を持たせる必要がある。誤り訂正を目的とした符号化方式は一般に**誤り訂正符号 (error correcting code)** とよばれる。**低密度パリティ検査符号 (low-density parity-check code)** とは現在最も高い誤り訂正能力を誇る実用的な符号である [17]。

誤り訂正機能を生み出す鍵は情報を冗長に表現することにある。 K ビットのメッセージ $\mathbf{m} \in \{0, 1\}^K$ を $N (> K)$ ビットの符号語 $\mathbf{x} \in \{0, 1\}^N$ に 1 対 1 で対応させて表現を冗長化する。誤り訂正符号の主要なクラスである線形符号ではこのために検査行列とよばれる $(N - K) \times N$ の 0 または 1 を成分とする行列 H を用意し

$$H\mathbf{x} = \mathbf{0} \pmod{2} \quad (75)$$

を満たすベクトル \mathbf{x} のみを符号語として利用する。mod 2 は 2 を法にする演算を意味する。 \mathbf{m} から \mathbf{x} への変換には、 H に対し、掃き出し法などをもちいて $HG^T = \mathbf{0} \pmod{2}$ を満たす $N \times K$ 行列 G^T (生成行列とよばれる) をあらかじめ用意しておく

$$\mathbf{x} = G^T \mathbf{m} \pmod{2} \quad (76)$$

と変換すればよい。こうすれば必ず $H\mathbf{x} = HG^T \mathbf{m} = \mathbf{0} \mathbf{m} = \mathbf{0} \pmod{2}$ が保証される。また、 $N > K$ であるため \mathbf{x} から \mathbf{m} への逆写像も一意に定まる。低密度パリティ検査符号とは、この中で特に各行/各列あたりの 1 の割合が $N, K \rightarrow \infty$ とする際にゼロに漸近する検査行列 H で特徴づけられるものを指す。以下、 H はこの性質を満たすランク落ちのない行列であるとす。

H の μ 行目 ($\mu = 1, 2, \dots, N - K$) で成分が 1 である列添字 i の集合を ∂_μ とし、 i 列目 ($i = 1, 2, \dots, N$) で成分が 1 である行添字 μ の集合を ∂_i と記す。また、表現を簡潔にするためにバイナリ・バイポーラ変換とよばれる変換

$$S_i = (-1)^{x_i} \iff x_i = \frac{1 - S_i}{2} \quad (77)$$

($i = 1, 2, \dots, N$) により、符号語 $\mathbf{x} \in \{0, 1\}^N$ をイジングスピン $\mathbf{S} \in \{+1, -1\}^N$ によって表現する。これにより検査方程式 (75) の μ 行目の条件式はポテンシャル関数

$$\psi_\mu(\mathbf{S}_\mu) = \frac{1 + \prod_{i \in \partial_\mu} S_i}{2} = \begin{cases} 1 & (\text{条件を満たしている}) \\ 0 & (\text{条件を満たしていない}) \end{cases} \quad (78)$$

によって簡潔にあらわすことができる。メッセージ $\mathbf{m} \in \{0, 1\}^K$ が一様分布にしたがって生成されるとしよう。すると、 \mathbf{S} のしたがう事前分布は

$$P(\mathbf{S}) = \frac{1}{Z_H} \prod_{\mu=1}^{N-K} \psi_\mu(\mathbf{S}_\mu) \quad (79)$$

(ただし、 $Z_H = \text{Tr}_{\mathbf{S}} \prod_{\mu=1}^{N-K} \psi_\mu(\mathbf{S}_\mu)$) によって与えられる。 \mathbf{S} の劣化過程としては、各成分独立に同じ確率 $0 < p < 1/2$ で符号が反転する 2 元対称通信路を仮定しよう。劣化したベクトル

ルを $\xi \in \{+1, -1\}^N$ と表現すると真のベクトル \mathbf{S} から劣化によって ξ が得られる過程は $\beta_p = (1/2) \ln((1-p)/p)$ として、条件付き確率

$$P(\xi|\mathbf{S}) = \prod_{i=1}^N \frac{\exp(\beta_p \xi_i S_i)}{2 \cosh(\beta_p)} = \frac{1}{(2 \cosh(\beta_p))^N} \prod_{i=1}^N \psi_i(S_i) \quad (80)$$

で与えられる。ただし、 $\psi_i(S_i) = \exp(\beta_p \xi_i S_i)$ とした、ベイズの公式をもちいると (79) と (80) から ξ を観測した際の \mathbf{S} の事後分布

$$P(\mathbf{S}|\xi) = \frac{P(\xi|\mathbf{S})P(\mathbf{S})}{P(\xi)} = \frac{1}{Z} \prod_{\mu=1}^{N-K} \psi_{\mu}(\mathbf{S}_{\mu}) \prod_{i=1}^N \psi_i(S_i) \quad (81)$$

が得られる。誤り訂正は劣化したベクトル ξ から正しいベクトル \mathbf{S} を復号することでなされる。2.2.1 節の議論から (81) にもとづいた確率推論がビット誤り率を最小にする意味で最適な復号であることが導かれる。

この確率推論にもとづく復号に (ルーピー) ビリーフプロパゲーションが使われる。そのために、 $\psi_{\mu}(\mathbf{S}_{\mu})$ および $\psi_i(S_i)$ をポテンシャル関数とみなし (81) をファクターグラフで表現する (図 7)。 $\mu = 1, 2, \dots, N-K$ および $i = 1, 2, \dots, N$ の両方がポテンシャル関数の添字 a に対応する。イジングスピンに対しては必ず $m_{a \rightarrow i}(S_i) = \exp(\theta_{a \rightarrow i} S_i) / (2 \cosh(\theta_{a \rightarrow i}))$, $m_{i \rightarrow a}(S_i) = \exp(\theta_{i \rightarrow a} S_i) / (2 \cosh(\theta_{i \rightarrow a}))$ という表現が可能であることをもちいると (72) に対応して

$$\theta_{\mu \rightarrow i} = \tanh^{-1} \left(\prod_{j \in \partial \mu \setminus i} \tanh(\theta_{j \rightarrow \mu}) \right) \quad (\psi_{\mu}(\mathbf{S}_{\mu}) \text{ に関して}) \quad (82)$$

$$\theta_{i \rightarrow i} = \beta_p \xi_i \quad (\psi_i(S_i) \text{ に関して}) \quad (83)$$

が、(73) に対応して

$$\theta_{i \rightarrow \mu} = \beta_p \xi_i + \sum_{\nu \in \partial i \setminus \mu} \theta_{\nu \rightarrow i} \quad (84)$$

が、それぞれ得られる。これらから決まる変数をもちいて、復号結果は

$$\hat{S}_i = \begin{cases} +1 & \text{if } \beta_p \xi_i + \sum_{\mu \in \partial i} \theta_{\mu \rightarrow i} > 0 \\ -1 & \text{if } \beta_p \xi_i + \sum_{\mu \in \partial i} \theta_{\mu \rightarrow i} < 0 \end{cases} \quad (85)$$

により求まる。

各行/各列あたりの 1 の個数を $O(1)$ に固定した条件下で H をランダムに生成した場合、対応するファクターグラフはスパースなランダム 2 部グラフとなる。3.3.3 節と同様の議論からスパースなランダム 2 部グラフの典型的なループ長は $O(\ln N)$ 程度で増大する。つまり、典型的な H のサンプルに対するファクターグラフの局所的な構造は十分大きな N に対してハイパーツリーに漸近する。このことは、適切な初期条件を与えることができれば、十分大きな N に対して、ビリーフプロパゲーションにもとづく復号の性能が厳密な復号の性能に漸近することを意味している。

5 おわりに

ベイズの公式にもとづく確率推論の問題と不規則系の統計力学との構造的な類似性にもとづいて統計力学の研究で開発された平均場近似の方法が、現実的な計算コストで推論を実施するための近似アルゴリズムとして有用であることを述べた。本稿では特に、ナイーブ平均場近似とキャビティ法 (およびそれを具体的に実行するアルゴリズムとしてビリーフプロパゲーション) を取り上げ、応用例として画像ノイズ除去、低密度パリティ検査符号の復号を紹介した。

最近「ビッグデータ」という言葉をよく耳にするが、大量の高次元データからの情報抽出に関する需要は産業応用のみならず学術面でも今後さらに増えていくものと予想される。多くの場合、高次元データの処理には計算量的困難がつきまとう。従来、物理の応用といえば、半導体や材料、機械などモノのイメージが強かったが、推論や計算といったコトの分野でも物理の理論が大活躍する。物理学のこうした新しい使われ方が広く定着する日もそう遠くはないかもしれない。

最後に本稿で詳しく紹介できなかった最近の話題について簡単に触れておく。

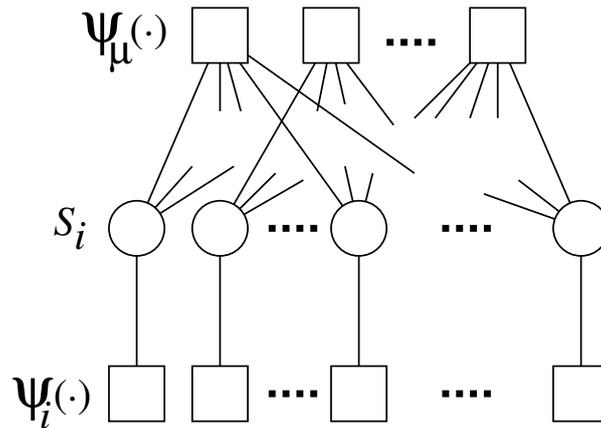


図 7: 低密度パリティ検査符号の復号問題に対するファクターグラフ。

- 逆イジング問題 (inverse Ising problem)**: 近年, 電気生理学において複数神経細胞の活動状態に関する同時計測ができるようになった. 神経細胞は発火および非発火の2つの特徴的な状態を取ることが知られている. i 番目の細胞に対し, この2状態をイジングスピン $S_i \in \{+1, -1\}$ に対応させる. 多細胞の同時計測データから細胞間の相互作用の様子を推定したい. イジングスピンによる記述にしたがうとすれば, これはデータから結合行列 J_{ij} や外場 h_i を推定する問題に他ならない. 与えられた J_{ij} や h_i と観測される統計量とを結びつけることは計算量的に困難である. そこで, J_{ij} や h_i の推定精度がなるべくよい平均場近似の方法を開発し問題解決をはかる [18, 19, 20].
- 圧縮センシング (compressed sensing)**: 線形変換として表現できる信号計測の結果, M 個のデータを得た. これらにもとづいて N 次元の信号を復元したい. 計測をあらゆる線形変換はランク落ちていないとしよう. 任意の信号を確実に復元するためには, $\alpha = M/N \geq 1$ が必要十分であることが線形代数の一般論から示される. しかしながら, 対象となる信号がスパースである (ゼロになる成分が多い) ことが統計的に期待される場合には, $\alpha < 1$ であっても信号復元が可能になるかもしれない. このようなアイデアにもとづく信号計測/復元の枠組みは圧縮センシングとよばれ, 現在盛んに研究されている [21]. 圧縮センシングの信号復元はスパースな信号をあらゆる事前分布を導入することで確率推論の問題として定式化できる. ここに平均場近似を導入することによりできるだけ小さな α での復元が保証され実際の計算コストで実施できる信号復元アルゴリズムを開発する [22, 23].
- 潜在変数モデリング (latent variable modeling)**: M 個の N 次元データの組 $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ が観測されているとする. これらのデータの規則性を分析する際, しばしば, 各データ \mathbf{x}_μ が観測されない潜在変数 \mathbf{z}_μ とデータに共通の行列 A にもとづいて $\mathbf{x}_\mu = A\mathbf{z}_\mu + \mathbf{n}_\mu$ というプロセス (\mathbf{n}_μ はノイズ) で生成されると仮定し, X から A と \mathbf{z}_μ の両方を求めることがなされる. こうしたモデル化の方法を潜在変数モデリングとよぶことがある. $Z = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M)$, $N = (\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_M)$ と表記すれば, これは Z や N に関する適当な仮定 (たとえば, スパースである, 統計的に独立である, など) の下で X を $X = AZ + N$ の形に分解する問題として定式化できる. このような問題は一般に計算量的に困難である. そこで, 平均場近似によって問題解決をはかる [24, 25, 26].

参考文献

- [1] 樺島祥介, 学習と情報の平均場理論, 岩波書店 (2002)
- [2] 西森秀稔, スピングラス理論と情報統計力学, 岩波書店 (1999)
- [3] E. Ising, Z. Phys. **31**, 253 (1925)

- [4] L. Onsager, *Phys. Rev. (2)* **65**, 117 (1944)
- [5] Y. Iba, *J. Phys. A* **32**, 3875 (1999)
- [6] M. Mézard, G. Parisi and M. A. Virasoro, *Spin glass theory and beyond*, World Scientific (1987)
- [7] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press (2009)
- [8] T. H. L. Watkin, A. Rau and M. Biehl, *Rev. Mod. Phys.* **65**, 499 (1993)
- [9] P. Weiss, *J. Phys. Theor. Appl.* **6**, 661 (1907)
- [10] H. A. Bethe, *Proc. Roy. Soc. London A* **150**, 552 (1935)
- [11] R. Kikuchi, *Phys. Rev.* **81**, 988 (1951)
- [12] T. Morita, M. Suzuki, K. Wada and M. Kaburagi (eds), *Foundations and Applications of Cluster Variation Method and Path Probability Method (Prog. Theor. Phys. Suppl. 115)* (1994)
- [13] M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001)
- [14] K. Nakagawa and H. Yamaguchi, *IEICE Trans. Inf. Syst.* **E96-D**, 433 (2013)
- [15] 田中和之, *確率モデルによる画像処理技術入門*, 森北出版 (2006)
- [16] R. Molina, A. K. Katsaggelos and J. Mateos, *IEEE Trans. Image Processing* **8**, 231 (1999)
- [17] 和田山正, *低密度パリティ検査符号とその復号法*, トリケップス (2002)
- [18] E. Schneidman, M. J. Berry, R. Segev and W. Bialek, *Nature* **440**, 1007 (2006)
- [19] S. Cocco and R. Monasson, *Phys. Rev. Lett* **106**, 090601 (2011)
- [20] H. Huang and Y. Kabashima, *Phys. Rev. E* **87**, 062129 (2013)
- [21] D. L. Donoho, *IEEE Trans. on Inf. Theor.* **52**, 1289 (2006)
- [22] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun and L. Zdeborová, *Phys. Rev. X* **2**, 021005 (2012)
- [23] M. C. Angelini, F. Ricci-Tersenghi and Y. Kabashima, in *Proc. 50th Annual Allerton Conference*, pp. 808–814 (2012); arXiv:1207.2853
- [24] A. Sakata and Y. Kabashima, *EPL (Europhysics Letters)* **103**, 28008 (2013)
- [25] A. Sakata and Y. Kabashima, in *Proc. 2013 IEEE International Symposium on Information Theory (Istanbul, Turkey, July 7–12, 2013)*, pp. 669–673 (2013); arXiv:1301.6199
- [26] F. Krzakala, M. Mézard and L. Zdeborová, in *Proc. 2013 IEEE International Symposium on Information Theory (Istanbul, Turkey, July 7–12, 2013)*, pp. 659–663 (2013); arXiv:1301.5898