

講義ノート

コトの物理学*

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

樺島 祥介

講義ノート作成：大阪大学サイバーメディアセンター 杉浦正康

(2008 年 6 月 30 日受理)

1 コトの物理学

モノ（物質）の科学とコト（情報）の科学はエントロピーの概念を共有しているが、長らく別々に発展してきた。しかし、この 10 年、両者が急接近している。これは物理学においてスピングラスの理論（統計力学）が発展したことと、情報分野において多体問題への関心が高まったことで、機が熟したのだといえる。さらにその背景には計算機、情報インフラの飛躍的な進歩がある。

この講義は、具体的な事例を紹介しながらコトの研究にモノの理論を生かす際の勘所を伝えることを目的としている。レプリカ法が情報分野に現れる様々な問題へアプローチするための強力な方法であることを示し、また、統計力学と情報論の間で発展している近似法について説明していく。

1.1 ランダム K-SAT 問題

1.1.1 ランダム K-SAT 問題とは

情報の世界には、その複雑さのため、解き明かすことが困難な様々な問題が発生する。そういった問題に対して、統計力学の手法を用いてどのようにアプローチできるのかということを探っていきたい。例として、まずはランダム K-SAT 問題を議論する。

まず、真偽に対応した N 個の論理変数（リテラル）の組

$$\boldsymbol{x} = (x_i) = \{0, 1\}^N \quad (1)$$

があるとして、 $\boldsymbol{x}$ から選んだ K 個の変数に対して、それ自身またはその否定を or でつないだ論理式の基本要素を“ K -節（クローズ）”と呼ぶことにする。たとえば 3-節は次のようになる。

$$\begin{aligned} C_1(\boldsymbol{x}) &= x_1 \vee \bar{x}_3 \vee x_7 \\ C_2(\boldsymbol{x}) &= x_4 \vee x_5 \vee \bar{x}_9 \\ &\dots \end{aligned} \quad (2)$$

K-SAT 問題は M 個の K -節を and でつなげてできる論理式 (K-CNF)

$$F(\boldsymbol{x}|\boldsymbol{C}) = C_1(\boldsymbol{x}) \wedge C_2(\boldsymbol{x}) \wedge \dots \wedge C_M(\boldsymbol{x}) \quad (3)$$

に対して、その値を 1（真）にする解（完全充足解＝SAT 解）が存在するかどうかを判定する決定問題である。

なぜここで K-SAT 問題を取り上げるかというと、これが NP 完全問題の標準形となっているからである。NP 完全問題とは問題サイズに関する多項式程度の時間で互いに書き換えることのできる組み合わせ問題の、あるクラスの集まりであり、現時点では任意の NP 完全問題を効率的に（問題サイズが多

*これは 2006 年 11 月に大阪大学大学院理学研究科で行われた集中講義の講義ノートである。

項式時間で) 解く方法は見つかっていない. けれども, 互いに多項式程度の時間で書き換えられるということは, NP 完全に属するある問題クラスの任意の問題を多項式時間で解く方法が発見できれば, NP 完全に属するすべての問題が多項式時間で解けることを意味する. K-SAT 問題はこういった文脈において組み合わせ問題の標準的な表現として広く取り上げられているのである.

K-SAT 問題の構造を調べるために, 「節数 M に対して, SAT 解の存在する問題の割合はどのように変化していくか」, ということについて調べる. そのためにランダム K-SAT 問題について考察していく. これはランダムに K-SAT 問題を生成し, 生成した問題のうち SAT 解をもつ問題がどれくらいの割合であるのを調べていくというものである.

まず直感的に考えると, K が 3 以上のときは, 節に含まれる変数のうちどれかひとつが真であれば, その節は真となるので, 各節を真にするのは比較的簡単である. 節数 M が増えると, 全部の節を満足するのは難しくなるが, M が十分多ければ次に示すように節の間での矛盾がみえやすくなり, 判定は易しくなる.

節間で矛盾が生じる例: x_1, x_2, x_3 に関するすべての 3-節を同時に真とすることはできない.

$$\begin{array}{cccc} x_1 \vee x_2 \vee x_3 & \bar{x}_1 \vee x_2 \vee x_3 & x_1 \vee x_2 \vee \bar{x}_3 & \bar{x}_1 \vee x_2 \vee \bar{x}_3 \\ x_1 \vee \bar{x}_2 \vee x_3 & \bar{x}_1 \vee \bar{x}_2 \vee x_3 & x_1 \vee \bar{x}_2 \vee \bar{x}_3 & \bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3 \end{array}$$

それでは $\alpha = M/N$ に対して SAT 解をもつ問題の割合がどのように変化するかをみてみる. $M, N \gg 1$, $\alpha = O(1)$ の状況を考えると次のようになることがわかる.

- $K = 1$: α によらず, UNSAT 解の割合が圧倒的に多い.
- $K = 2$: $\alpha < 1$ のとき SAT 解が圧倒的に多く, $\alpha > 1$ のとき UNSAT 解が圧倒的に多い.
- $K = 3$ 以上: $\alpha < \alpha_c$ のとき SAT 解が圧倒的で, $\alpha > \alpha_c$ のとき UNSAT 解が圧倒的となる α_c が存在する.

ただし, $K = 1, 2$ については数学的に厳密な結果, $K = 3$ 以上については数値解に基づく予想である. ここで, $\alpha = \alpha_c$ での振舞いを理論的に解明し, できれば $\alpha_c(K)$ の値を理論的に評価したいのだが, ここに困難が生じる. というのは, この問題をナイーブに解こうとすると, まず SAT 問題を発生し, それに対して SAT 解があるかどうかを調べることになるが, その手間が指数関数的に増大してしまうのである. さらに SAT 解の割合

$$P_{SAT} = \sum_C P(C) \delta(\exists \mathbf{x} \text{ s.t. } F(\mathbf{x}|C) = 1) \quad (4)$$

を求めたいのだが, これが与えられた節数の下での SAT 問題のすべての可能性に関して平均を必要とする「2重平均」の構造になっていることで問題がもう一段複雑になっている. この「2重平均」を評価する有力な手段として物理の方法であるレプリカ法が登場する.

1.1.2 レプリカ法の概略

レプリカ法は統計力学において, 不規則系の分配関数の一般化モーメントを求める方法であるが, 次のようにレプリカトリックを用いて, 対数の平均値を求める目的で使われることが多い.

$$\begin{aligned} [\ln Z(J)]_J &= \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \ln[Z^n(J)]_J \\ &= \lim_{n \rightarrow 0} \frac{[Z^n(J)]_J - 1}{n} \end{aligned} \quad (5)$$

n が自然数のとき、次のような展開ができることがポイントである。

$$\begin{aligned}
 [Z^n(J)]_J &= \sum_J P(J) \left(\sum_S e^{-\beta H(S|J)} \right)^n \\
 &= \sum_{S^1, S^2, \dots, S^n} \sum_J P(J) e^{-\beta \sum_{k=1}^n H(S^k|J)} \\
 &= \sum_{S^1, S^2, \dots, S^n} e^{-\beta \sum_{k=1}^n H_{\text{eff}}(S^1, S^2, \dots, S^n; \beta)}
 \end{aligned} \tag{6}$$

n を自然数として展開を行ったが、レプリカトリックにおいては $n \rightarrow 0$ という操作が必要である。レプリカ法では、この操作を自然数に対して得られた (6) 式に関する解析的表現を実数に解析接続することにより実行する。自然数から実数への解析接続という飛躍を伴う計算法であるため、レプリカ法で得られた結果が正当性をもつかどうかには十分注意する必要がある。

(6) 式の展開において $H(S|J)$ が現れた。これはエネルギー関数であるが、K-SAT 問題についてはエネルギー関数を次のように導入することでレプリカ法による解析を行うことができる。

$$\begin{aligned}
 H(x|C) &= \sum_{\mu=1}^M (1 - C_{\mu}(x)) = \text{「非充足節の個数」} \\
 (1 - C_{\mu}(x)) &= \begin{cases} 0 & \text{if } C_{\mu}(x) \text{ satisfied} \\ 1 & \text{if } C_{\mu}(x) \text{ unsatisfied} \end{cases}
 \end{aligned} \tag{7}$$

このようにエネルギー関数を書けば、K-SAT 問題は「エネルギー関数の最小値がゼロかどうか」を判定する問題と言い換えることができる。数式で表現すれば次のようなことになる。

x が SAT 解である条件

$$F(x|C) = 1 \leftrightarrow H(x|C) = 0$$

SAT 解が存在する条件

$$F(\exists x|C) = 1 \leftrightarrow \min_x \{H(x|C)\} = 0$$

要は最小エネルギーがゼロかどうかを判定できればよい (図 1)。

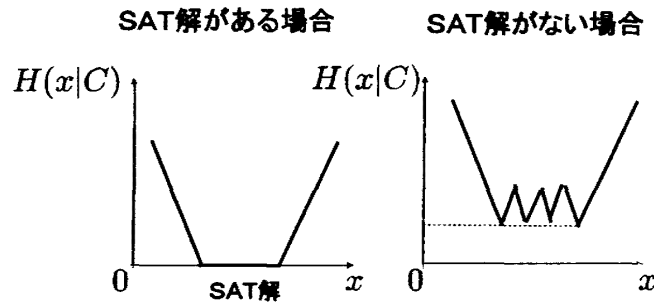


図 1: SAT 問題のエネルギー描像

つまり、SAT 解を持つ問題の割合 P_{SAT} は

$$\begin{aligned}
 P_{\text{SAT}} &= \sum_C P(C) \delta(\exists x \text{ s.t. } F(x|C) = 1) \\
 &= \sum_C P(C) \delta(\min_x H(x|C) = 0)
 \end{aligned} \tag{8}$$

となり、「最小エネルギーがゼロとなる」問題の割合ということになる。しかし、ランダムに与えた問題 C の最小エネルギーは C に複雑に依存するランダム変数となるため、その統計的振舞いを解析するのは、やはり一筋縄ではいかない。

その苦境を救ってくれるのが、最小エネルギーの下界を表す分配関数である。分配関数 $Z(\beta|C)$ を導入することで、K-SAT 問題を次のように統計力学的に定式化することができる。

$$\begin{aligned} e^{-\beta \min_{\mathbf{x}} \{H(\mathbf{S}|\mathbf{J})\}} &\leq \sum_{\mathbf{x}} e^{-\beta H(\mathbf{S}|\mathbf{J})} = Z(\beta|C) \\ \min_{\mathbf{x}} \{H(\mathbf{x}|C)\} &\geq -\frac{1}{\beta} \ln Z(\beta|C) \\ \min_{\mathbf{x}} \{H(\mathbf{x}|C)\} &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \ln Z(\beta|C) \end{aligned} \quad (9)$$

(9) 式は情報理論の通信路符号化定理に関連して用いられる不等式であり、しばしば Arimoto 不等式と呼ばれる。つまり、任意の逆温度 $\beta > 0$ の分配関数（の対数）の統計的振舞いがわかれば、 $\beta \rightarrow \infty$ の極限で、最小エネルギーの振舞い、つまり、SAT 問題の振舞いがわかる。ここで分配関数から自由エネルギー f が

$$Nf = -\frac{1}{\beta} \ln Z(\beta|C) \quad (10)$$

と得られるが、この自由エネルギー f が (9) 式からわかるように評価すべき対象となる。 f は C に依存するが、 f がその典型的な値をとり得る確率は、大偏差統計に従って N の値に関して指数的に 1 に近づき、相転移現象が観測される $N, M \gg 1$ の状況では以下のようなスケーリングが期待される。

$$\begin{aligned} Z(\beta|C) &\sim e^{-N\beta f} \\ P[f = -\frac{1}{N\beta} \ln Z(\beta|C)] &\sim e^{-Nc(f;\beta)} \\ (f &\sim O(1)) \end{aligned} \quad (11)$$

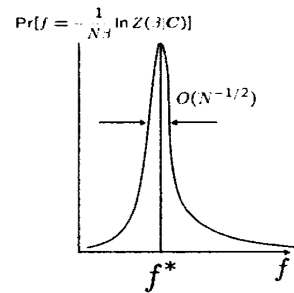


図 2: 大偏差統計により、典型値 f^* が得られる確率はほぼ 1

大偏差性に着目すると、SAT-UNSAT 転移について図 3 のような描像が得られる。大偏差統計での典型的な値とは

$$f^* = \begin{cases} -\frac{1}{N\beta} \ln[Z(\beta|C)]_C & \times \\ -\frac{1}{N\beta} [\ln Z(\beta|C)]_C & \circ \end{cases}$$

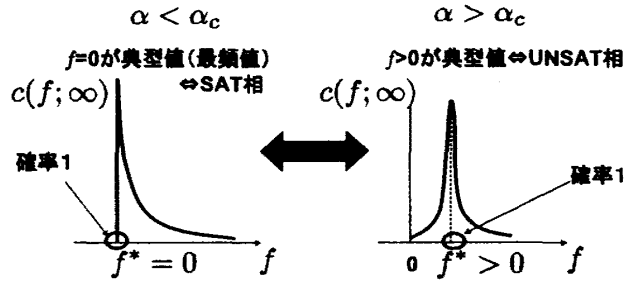


図 3: SAT-UNSAT 相転移

である。なぜならば

$$\begin{aligned} [Z(\beta|C)]_C &= \int df \exp[-N(\beta f + c(f; \beta))] \\ &\simeq \exp\left[-N(\min_f \{\beta f + c(f; \beta)\})\right] \end{aligned} \quad (12)$$

となり $[Z(\beta|C)]_C$ は大偏差統計にひきずられて典型値（最頻値）からずれてしまうからである。

分配関数の対数の形で与えられた典型値の評価は、(5) 式のレプリカトリックにより可能になる（有限の n に関する評価を行えば、非典型値の評価も可能）。

$$\begin{aligned} f^* &= -\frac{1}{N\beta} [\ln Z(\beta|C)]_C \\ &= -\lim_{n \rightarrow 0} \frac{\partial}{\partial n} \frac{1}{N\beta} \ln [Z(\beta|C)]_C \end{aligned} \quad (13)$$

あとは、 $\beta \rightarrow \infty$ の極限をとり、最小エネルギーの典型値が評価できる。

$$\begin{aligned} f^* &= -\lim_{\beta \rightarrow \infty, n \rightarrow 0} \frac{\partial}{\partial n} \frac{1}{N\beta} \ln [Z(\beta|C)]_C \\ &\begin{cases} f^* = 0 \rightarrow \text{SAT 相} \\ f^* > 0 \rightarrow \text{UNSAT 相} \end{cases} \end{aligned} \quad (14)$$

ここで詳しい計算については説明しないが、レプリカ解析から得られる描像は図 4 のようになる。

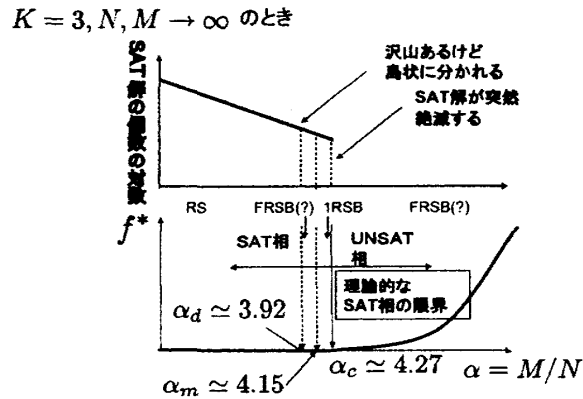


図 4: SAT 問題についてレプリカ解析から得られる描像

ランダム K-SAT 問題にレプリカ法を適用した流れをまとめると次のようになる。

- ランダムに発生した K-SAT 問題が SAT 解を持つ割合を調べる際に，2 重平均の構造が現れる。
- レプリカ法は 2 重平均の困難を解決する方法で，特に分配関数の対数の平均を求めるのによく使われる。
- 最小エネルギー値は分配関数と結びつく。
-最大化，最小化は平均計算の特別な場合。
- 大偏差統計では，典型値（最頻値）＝平均値
-対数の平均が必要になる→‘レプリカ法’

いくつかの数学的な問題点も指摘しておく。まず 1 つ目の問題点として，(13) 式をみると， $N \rightarrow \infty$ の解析なので， $O(1)$ の誤りが無視されていて，大偏差的解析では解けたことになっていない。しかし，それでも数値実験とは良く合うことがわかっている。

2 つ目の問題点として，解析接続がよいのかどうか，そもそも離散値から連続値への接続がユニークなのかという問題がある。これは Carlson の定理により有限系であれば今の問題では大丈夫である。

3 つ目の問題点は熱力学的極限に関するものであり，次の極限の入れ替えの正当性についてである。

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \longrightarrow \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} \quad (15)$$

これに起因して一種の相転移（レプリカ対称性の破れ）が現れるという問題があることはわかっている。しかし，必ずよいかはわからないが，経験的にこれに対する対処法はわかっている。

1.2 CDMA マルチユーザ復調問題

1.2.1 CDMA について

次に携帯電話の方式の一つである CDMA（Code Division Multiple Access）について議論する。携帯電話サービスでは，各エリアに基地局が置かれており，あるエリアで端末から発信された情報はそのエリアの基地局に送られる（図 5）。各エリアのことをセル（cell）と呼び，携帯電話を英語で cellular phone というのはそのためである。

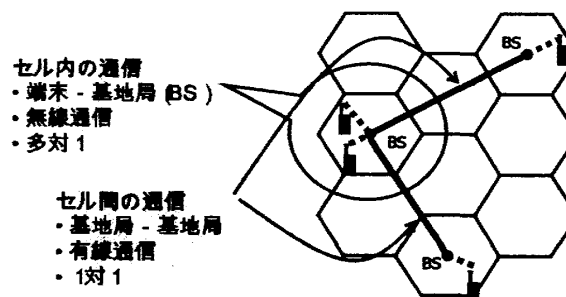


図 5: 携帯電話の通信方式

ユーザのデジタル情報は図 6 に示すように，矩形波の連なりで表現される。ここで，仮定としてユーザ間での波の位相は同期しているとする。このような同期は実用レベルで実現可能である。

問題となるのは，一つの基地局が複数のユーザを抱えているため，ユーザ間干渉が起きてしまうことである。図 7 のように同一セル内に Alice と Bob がいるとする。この二人が同時に携帯電話を使う状況を考える。Alice と Bob の音声信号はデジタル化されて基地局に送られる。デジタル信号は通常 0,1 で表されるが，ここでは便宜上 +1, -1 の二値で表すことにする。Alice と Bob が，もし同じ周波数の信号

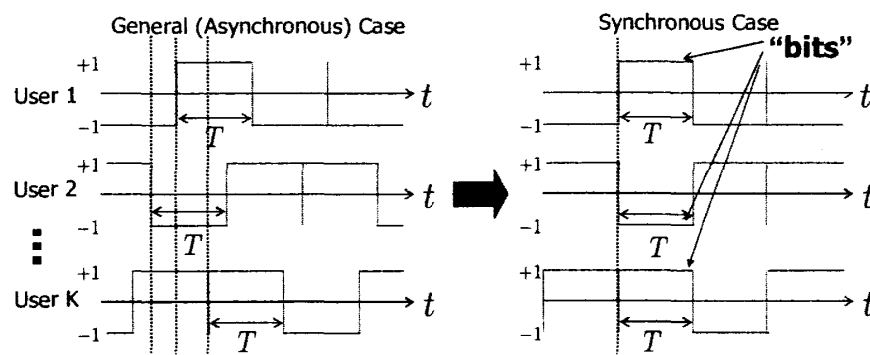


図 6: デジタル信号の同期

を利用すると、+1 と -1 を同時に送るとき、それらの信号は重ねあわせで 0 になり通信がうまくいかない。そこで、同一セル内で同時刻に通信を行うためには何らかの工夫が必要となる。

そのような多元接続の方法はいくつか検討され実用化されている。FDMA(Frequency Division Multiple Access) は、周波数帯域全体をいくつかの帯域に分割し、各帯域に個別のチャンネルを割り当て、複数のユーザが同時に通信を行なう方式である。また TDMA(Time Division Multiple Access) は、複数のユーザが使用する無線チャンネルに対し、1つの無線周波数を使用する幾つかのタイムスロットに分割する。各ユーザは異なるタイムスロットを使用し、1つの周波数を短時間ずつ交代で複数のユーザで共有する。

FDMA と TDMA は一つの基地局で複数のユーザの通信を同時に処理することができるが、あらかじめ処理できるユーザの数は限定される。

あらかじめ処理できるユーザ数が決まらない方法の一つが CDMA である。CDMA では、図 7 に示すとおり、Alice の -1 信号と Bob の +1 信号に各々別の拡散系列を掛けて変調している。拡散系列は、送信信号 1 ビットが送られる時間内にその送信信号よりも細かく分けられたランダムパターンである。

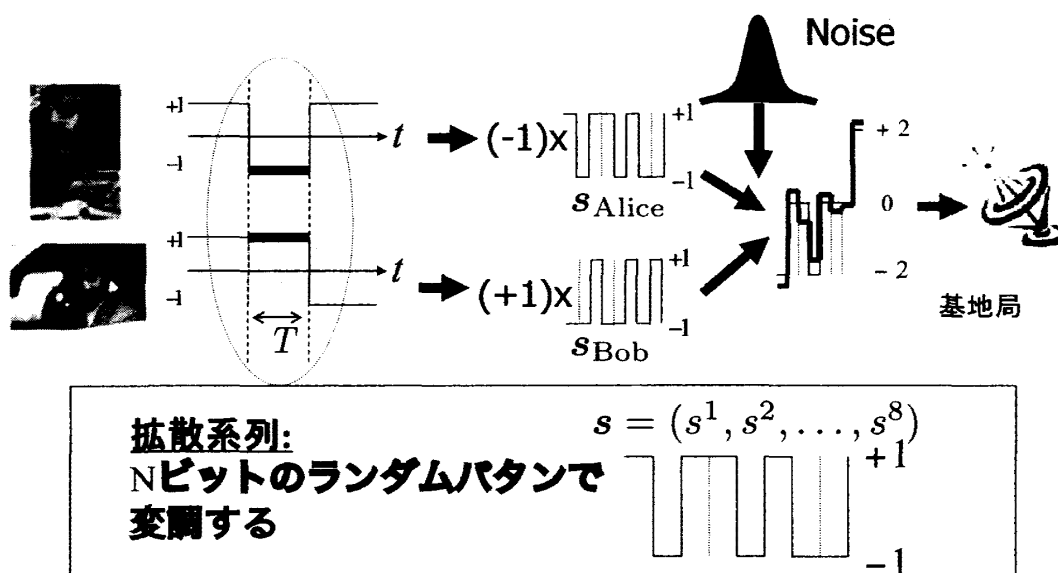


図 7: CDMA の概念図

同一セル内で K 人のユーザが同時に携帯電話を使う場合に拡張して考えよう。 k 番目のユーザのあるビットを b_k とする。 K 人のユーザの送るビット $b_1, b_2, \dots, b_K$ が同期しており、伝送中に重ねあわされる。拡散系列の細かく分割されたそれぞれをチップと呼び、送信信号 1 ビットに対する拡散系列は N 個

のチップから成るものとする． k 番目のユーザの送信信号 b_k に $\mathbf{s}_k = \{s_k^1, s_k^2, \dots, s_k^N\}$ の拡散系列を掛けて送信する．さらに，伝送中に μ 番目のチップの部分に偶然発生した Gaussian noise $\{\sigma_0 n^\mu\}$ が加わる．こうして，基地局で受信される μ 番目のチップの部分の信号は

$$y^\mu = \frac{1}{\sqrt{N}} \sum_{k=1}^N s_k^\mu b_k + \sigma_0 n^\mu \quad (16)$$

となる．

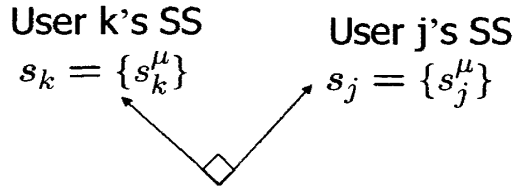
復調問題は拡散系列 $\{s_k^\mu\}$ を既知であるとして， $\{y^\mu\}$ から $\{b_k\}$ を推定する問題である．実用のために求められることは，すばやく復調できること，そしてエラーが少ないことである．(16) 式は線型方程式であるが，離散変数のため解くことが計算量的に困難である．そこで，拡散系列 $\{s_k^\mu\}$ の性質を利用してうまく復調する近似法を考える．拡散係数の各チップは

$$P(s) = \frac{1}{2}\delta(s+1) + \frac{1}{2}\delta(s-1) \quad (17)$$

でランダムに与えられるので，独立変数として扱うことができるため

$$\frac{1}{N} \mathbf{s}_k \cdot \mathbf{s}_j = \frac{1}{N} \sum_{\mu=1}^N s_k^\mu s_j^\mu = \begin{cases} 1 & k=j \\ O(N^{-1/2}) & k \neq j \end{cases} \quad (18)$$

となって，統計的にほぼ直交している．



この直交性を利用する簡便な復調方式として実用化されているシングルユーザ復調について説明する．受信信号 $\{y^\mu\}$ に拡散系列 s_k を作用させると

$$h_k = \frac{1}{\sqrt{N}} \sum_{\mu=1}^N s_k^\mu y^\mu = b_k + \sum_{j \neq k} \frac{1}{N} \sum_{\mu=1}^N s_k^\mu s_j^\mu b_j + \frac{1}{\sqrt{N}} \sum_{\mu=1}^N s_k^\mu \sigma_0 n^\mu \quad (19)$$

となる．第2項は平均0，分散 K/N のガウス分布に従う．よって，ユーザ数 K より，チップ数 N が十分多ければ，基地局が複数ユーザを担当しているにもかかわらず，信号の劣化は実質的にノイズの影響しか受けないことになる．シングルユーザ推定ではこの原理を直接用いて， h_k から送信されたビットを

$$\hat{b}_k^{SUD} = \text{sgn}(h_k) \quad (20)$$

によって推定する．この方法は計算コストがそれほどかからないという利点がある．欠点は，ユーザ数が増えるとビット誤り率が大きくなることである．

1.2.2 復調問題への統計力学的アプローチ

ここからはさらに通信効率の向上をよくするための方法として，マルチユーザ復調について説明する．マルチユーザ復調ではベイズ決定理論を使う．まず例として，表が確率 p で出ることがわかっているコインをトスするとき，出る目を予測する方法としてどのようなものが最適か，ということを考える．確率 q で表が出ると予想すると，予想が的中する確率は

$$\begin{aligned} R(q|p) &= p \times q + (1-p)(1-q) \\ &= (2p-1)q + 1-p \end{aligned} \quad (21)$$

となる。よって最適な戦略は $R(q|p)$ を q に関して最大化することにより

$$\begin{cases} p > 1/2 & \rightarrow & q = 1 & (\text{表}) \\ p < 1/2 & \rightarrow & q = 0 & (\text{裏}) \end{cases} \quad (22)$$

となる。つまり問題に確率的要素があっても、ベイズ決定理論による最適な戦略は決定論的になる。

CDMA に対して同様にベイズ決定理論による議論をする。

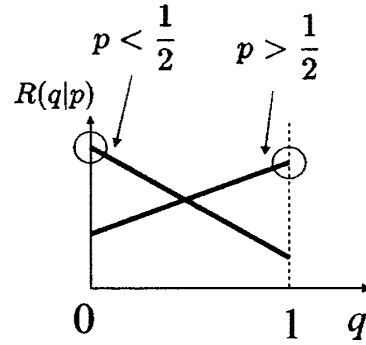


図 8: ベイズ決定によるコイン投げ問題に対する最適戦略

$\mathbf{b} = \{b_1, \dots, b_K\}$ のあるパターンが得られる確率 $P(\mathbf{b})$ は、それぞれのユーザの送る信号が +1 か -1 の値をとる確率はそれぞれ 1/2 と仮定するのが自然なので、

$$P(\mathbf{b}) = \frac{1}{2^K} \quad (23)$$

となる。Gaussian noise の影響を受けて、(16) 式より y^μ は平均 $1/\sqrt{N} \sum_{k=1}^K s_k^\mu b_k$ 、分散 σ_0 のガウス分布に従い、

$$P(y^\mu | \mathbf{b}, (s_{k=1,2,\dots,K}^\mu)) = \frac{1}{\sqrt{2\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} \left(y^\mu - \frac{1}{\sqrt{N}} \sum_{k=1}^K s_k^\mu b_k \right)^2 \right] \quad (24)$$

という確率になる。ベイズ決定理論に従えば、 $\mathbf{y} = y^\mu$ を受信した後の $\mathbf{b}$ の事後確率は

$$P(\mathbf{b} | \mathbf{y}, s_k) \propto P(\mathbf{b}) \prod_{\mu=1}^N P(y^\mu | \mathbf{b}, (s_{k=1,2,\dots,K}^\mu)) \propto \exp \left[-\frac{1}{2\sigma^2} \sum_{\mu=1}^N \left(y^\mu - \frac{1}{\sqrt{N}} \sum_{k=1}^K s_k^\mu b_k \right)^2 \right] \quad (25)$$

となり、コイン投げの例で言えば、「表が出る真の確率」に対応している。送信ビットの推定は事後確率が最大になるものを選ぶことによっておこなう。その際、複数ユーザの送ったビット $\mathbf{b}$ をまとめて推定しておこなう復調を MAP 復調といい、推定ビット $\hat{\mathbf{b}}$ が、真の送信ビット $\mathbf{b}^0$ と異なる確率 $\text{Prob}(\hat{\mathbf{b}} \neq \mathbf{b}^0)$ を最小にする。MAP 復調では、

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\text{argmax}} \{ P(\mathbf{b} | \mathbf{y}, \{s_k\}) \} \quad (26)$$

により、ユーザの送ったビットを推定する。

各ユーザの送ったビットをそれぞれ別々に推定しておこなう復調は MPM 復調という。これは、 $\text{Prob}(\hat{b}_k \neq b_k^0)$ を最小にするようにする。MPM 復調での送信ビットの推定は

$$\begin{aligned} \hat{b}_k &= \underset{b_k}{\text{argmax}} \{ P(b_k | \mathbf{y}, \{s_k\}) \} \\ &= \underset{b_k}{\text{argmax}} \left\{ \sum_{\mathbf{b} \setminus b_k} P(\mathbf{b} | \mathbf{y}, \{s_k\}) \right\} \end{aligned} \quad (27)$$

によっておこなう。

MAP 復調や MPM 復調をするには、すべてのユーザの系列が必要になる。そのため、これらの復調は「マルチユーザ復調 (MUD)」と呼ばれる。あるユーザの送ったビットを推定するのに、他のユーザの情報も使うことで、シングルユーザに比べてよい復調になることが期待できる。

ベイズ決定理論により事後確率を最大にするようにビットを推定することでマルチユーザ復調ができることがわかった。しかし、厳密計算は計算量的に困難である。この問題の解決のために高性能な近似的なアルゴリズムの開発がおこなわれているが、このことについては後で簡単に触れる。

マルチユーザ復調のほかの問題として、性能評価の問題がある。最適性は保障されるものの、ベイズ決定理論だけではシングルユーザ復調と比べてどのくらいよいのかはいえない。そこで、統計力学を使って定量的に性能評価をおこなう。まずはスピン系との類似性を指摘しておこう。 $\mathbf{b} = \{+1, -1\}^K$ として、送信されるビットを Ising Spins と対応させ、 σ^2 を温度に対応させることで、次のようなハミルトニアンとボルツマン分布を構成して、統計力学の技法を使った性能評価を行うことができる。

$$H(\mathbf{b}|\mathbf{y}, \{\mathbf{s}_k\}) = \frac{1}{2} \sum_{\mu=1}^N (y^\mu - \frac{1}{\sqrt{N}} \sum_{k=1}^K s_k^\mu b_k) ^2 \quad (28)$$

$$P(\mathbf{b}|\mathbf{y}, \{\mathbf{s}_k\}) \propto \exp \left[-\frac{1}{\sigma^2} H(\mathbf{b}|\mathbf{y}, \{\mathbf{s}_k\}) \right] \quad (29)$$

$N, K \rightarrow \infty$, $\beta = \frac{N}{K} \sim O(1)$ の極限における、さまざまな $\mathbf{y}$ と $\{\mathbf{s}_k\}$ に対する典型的な性能評価を行うことは、2重平均の構造になっている。そのために、ここでもレプリカ法を使って解析する。

レプリカ法の手順はこれまでも説明したように、分配関数 Z を導入して、自然数 n に対して、配位平均

$$[Z^n]_{\mathbf{y}, \{\mathbf{s}_k\}} = \sum_{\mathbf{b}^1, \mathbf{b}^2, \dots, \mathbf{b}^n} \left[\exp \left\{ -\frac{1}{\sigma^2} \sum_{\sigma=1}^n H(\mathbf{b}^\sigma | \mathbf{y}, \{\mathbf{s}_k\}) \right\} \right]_{\mathbf{y}, \{\mathbf{s}_k\}} \quad (30)$$

を計算して、

$$\frac{1}{K} [\ln Z]_{\mathbf{y}, \{\mathbf{s}_k\}} = \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \frac{1}{K} \ln [Z^n]_{\mathbf{y}, \{\mathbf{s}_k\}} \quad (31)$$

を求める。

ここでは、詳しい計算をせず、レプリカ法で評価された結果をいくつか紹介することにする。まず、レプリカ対称 (RS: Replica Symmetric) 自由エネルギーについて考える。レプリカ対称仮定

$$\begin{aligned} \frac{1}{K} \sum_{k=1}^K [b_k^0 \langle b_k^a \rangle]_{\mathbf{y}, \{\mathbf{s}_k\}} &= m^a = m \quad (\forall a = 1, 2, \dots, n) \\ \frac{1}{K} \sum_{k=1}^K [\langle b_k^a \rangle \langle b_k^b \rangle]_{\mathbf{y}, \{\mathbf{s}_k\}} &= q^{ab} = q \quad (\forall a > b) \end{aligned} \quad (32)$$

において、RS 自由エネルギーは

$$\begin{aligned} \frac{1}{K} [\ln Z]_{\mathbf{y}, \{\mathbf{s}_k\}} &= \text{Extr}_{m, q, \hat{m}, \hat{q}} \left\{ -\frac{1}{2\beta} \ln(1 + \sigma^{-2} \beta(1 - q)) - \frac{1 - 2m + q + \beta^{-1} \sigma_0^2}{2\sigma^2(1 + \sigma^{-2} \beta(1 - q))} \right. \\ &\quad \left. - \hat{m}m - \frac{\hat{q}(1 - q)}{2} + \int \frac{dz e^{-z^2/2}}{\sqrt{2\pi}} \ln 2 \cosh(\sqrt{\hat{q}}z + \hat{m}) \right\} \quad (\beta = K/N) \end{aligned} \quad (33)$$

となる。この極値問題を解くことで、 m, q の典型的な値が決まる。

また、性能評価の問題としてビットエラー率についてもレプリカ法による評価が可能である。ビットエラー率は真の送信ビットと推定ビットが異なる確率であり、

$$\text{Prob}(\hat{b}_k \neq b_k^0) = \int_{\hat{m}/\sqrt{\hat{q}}}^{\infty} \frac{dz e^{-z^2/2}}{\sqrt{2\pi}} = \text{erfc} \left(\frac{\hat{m}}{\sqrt{\hat{q}}} \right) \quad (34)$$

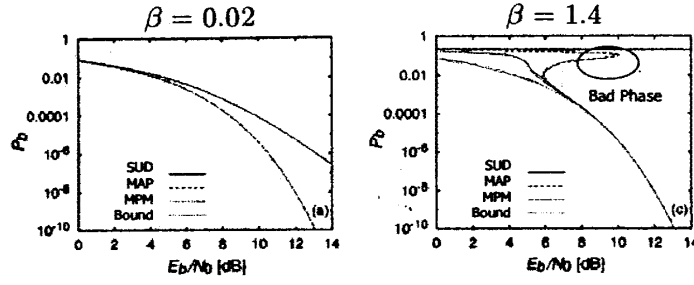


図 9: ビットエラー率 (Tanaka, 2002[2] より改変). 横軸は SN 比 (信号とノイズの比) であり, ここでは情報 1 ビット当たりの電力 E_0 を通信路ノイズの片側電力スペクトル密度 N_0 で割ったものである. この値が大きいほど良好な状態といえる.

となる. 図9は異なる負荷率 $\beta = K/N$ でのビットエラー率を表している. 性能は, $\text{MPM} > \text{MAP} > \text{SUD}$ (シングルユーザ復調), となる. 大きな負荷のときは“ダメな相”が出現している. これは実験的に観測される“Waterfalling”現象への回答になっている. グラフに見られる S 字の部分では, 同一の SN 比に対して 3 つの解が存在しているが, 真ん中の解は不安定であり, 上下の 2 つの安定な解のどちらかが正しい解を与える. このうち正しくない解は準安定解と呼ばれる. レプリカ解析結果では, S 字の「右肩」に相当する解が準安定解である. 準安定解が正しくないから重要でないかということ, そうではない. 図 10 にモンテカルロ法による実験結果との比較を示すが, 実験結果は準安定解である「右肩」に沿うような振舞いをみせている. このことは実際の推定を行う際の推定誤り率を評価する場合には, 準安定解が重要であることを示している.

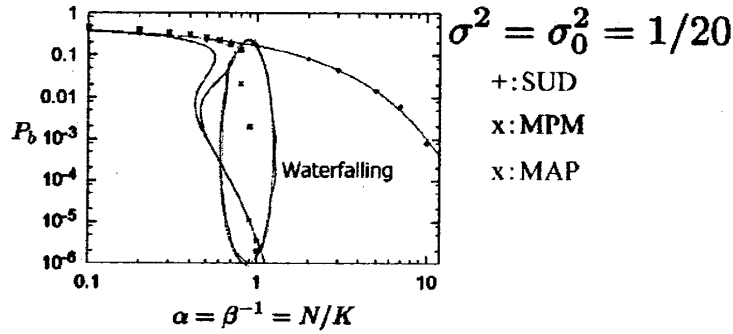


図 10: ビットエラー率の実験結果 (宇田新介氏, 田中利幸氏からの提供)

ビット推定において, ここではレプリカ対称性を仮定した結果を示したが, その妥当性はどうだろうか. RS 解は以下の条件が成立しないと熱力学的に不安定であることがわかっている.

AT 条件

$$1 - \frac{\beta}{(\sigma^2 + \beta(1-q))^2} \int \frac{dz e^{-z^2/2}}{\sqrt{2\pi}} (1 - \tanh^2(\sqrt{q}\hat{z} + \hat{m}))^2 > 0 \quad (35)$$

これは σ^2/σ_0^2 が十分小さいと不安定化する. これは次のことを意味する.

σ^2 は (29) 式などを見るとわかるように, ボルツマン因子における温度に対応している. MAP 復調と

MPM 復調は

$$\begin{aligned} \text{MAP: } \hat{\mathbf{b}} &= \underset{\mathbf{b}}{\operatorname{argmax}} \{P(\mathbf{b}|\mathbf{y}, \{s_k\})\} \Leftrightarrow \begin{matrix} \sigma^2/\sigma_0^2 \rightarrow 0 \\ (\text{zero-temp.}) \end{matrix} \\ \text{MPM: } \hat{b}_k &= \underset{b_k}{\operatorname{argmax}} \left\{ \sum_{\mathbf{b} \setminus b_k} P(\mathbf{b}|\mathbf{y}, \{s_k\}) \right\} \Leftrightarrow \begin{matrix} \sigma^2/\sigma_0^2 = 1 \\ (\text{finite-temp.}) \end{matrix} \end{aligned}$$

といった対応があり、温度の違いとして理解可能である (Rujan (1993), Nishimori (1993), Sourlas (1994)). よって、AT 条件によれば MAP 解は不安定ということになる。また、RS 解は $\sigma^2/\sigma_0^2 = 1$ が成り立つといつでも安定であるという Nishimori 条件と呼ばれるものが示されている。これらのことは $\sigma^2 = \sigma_0^2$ の場合の MPM 解に関する性能評価は RS 解で十分であり、MAP 解の詳細な解析にはレプリカ対称解の破れを考慮する必要があることを示している。

2 コトを解き明かす

レプリカ法は2重平均の数理構造をもつ問題を評価することができ、情報分野の研究に対して性能評価に利用できることを説明した。しかし、その具体的な計算は見せなかった。計算を見せなかった理由は、ほとんどの問題でレプリカ法の計算は、添え字がたくさん出てきたり、一部数値計算が必要になるなど、煩雑で面倒だからである。また、レプリカ法はその手順において、いくつか数学的に「怪しい」部分があり、そこには十分に注意する必要がある。しかし、レプリカ法を知るためには具体的な計算をやってみる必要もあるだろう。ここでは、例外的に手計算で扱うことのできるモデルを取り上げ、レプリカ法の勘所を伝える。

2.1 ランダムエネルギーモデル

2.1.1 ランダムエネルギーモデルの説明

レプリカ法は分配関数 Z のモーメント $[Z^n(J)]$ をまず自然数 n に対して計算して、それから実数の n に解析接続することで求める方法である。具体的な例としてランダムエネルギーモデル (REM: Derrida (哲学者 Jack Derrida の甥), 1981) に関してレプリカ法の計算をみていこう。

2^N 個の状態の $\tau = \{+1, -1\}^N$ の各々に対するエネルギー値 $E(\tau)$ を

$$P(E) = \frac{1}{\sqrt{N\pi}} \exp \left[-\frac{E^2}{N} \right] \quad (36)$$

から独立同分布抽出で定めたシステムを考える (図 11)。

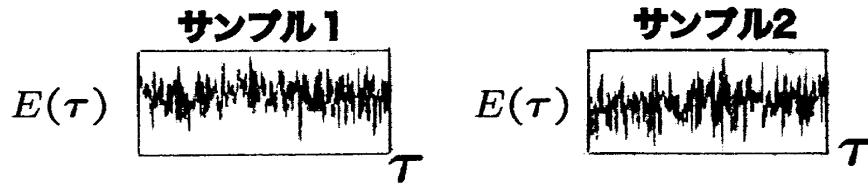


図 11: ランダムエネルギーランドスケープ. (36) 式の確率分布から得られた REM のランドスケープの例 ($N = 8$)

このシステムはスピングラス、高分子の研究等で用いられ、情報理論のランダム符号化とも密接に関係している。状態 $S = \{+1, -1\}^N$ に対してエネルギー関数は

$$H(S|E) = \sum_{\tau} E(\tau) \delta(S, \tau) \quad \left(\delta(S, \tau) = \begin{cases} 1, & (S = \tau) \\ 0, & (S \neq \tau) \end{cases} \right) \quad (37)$$

となる． $\mathbf{E} = \{E(\tau)\}$ である．エネルギーランドスケープへの依存性を明示した表現にするため， $E(\mathbf{S})$ と書かず， $E(\tau)$ としている．絶対温度 $T = \beta^{-1}$ のときの分配関数は

$$Z(\beta|\mathbf{E}) = \sum_{\mathbf{S}} e^{-\beta H(\mathbf{S}|\mathbf{E})} \quad (38)$$

となり，ここから自由エネルギー（密度）の平均値

$$f(\beta) = -\frac{1}{N\beta} [\ln Z(\beta|\mathbf{E})]_{\mathbf{E}} \quad (39)$$

を評価する． $[\dots]_{\mathbf{E}}$ は $\mathbf{E}$ に関して平均（配位平均）をとることを意味する．

2.1.2 レプリカ法を用いない解法

REM を例にした理由は，自由エネルギーがレプリカ法を用いなくても求まるので，レプリカ法の検証に使えるからである．そこで，まず正解をレプリカ法を使わずに求めて，それからレプリカ法を用いて正解を導くにはどのような工夫が必要かを示す（それによりレプリカ法のいかがわしさもわかる）．

1 自由度当たりの状態密度

$$E \in \{Ne, N(e + \delta e)\} \sim \mathcal{N}(e|\mathbf{E}) \quad (40)$$

に対して，平均と分散は

$$\begin{aligned} [\mathcal{N}(e|\mathbf{E})]_{\mathbf{E}} &= 2^N \times P(E = Ne)(N\delta e) \sim \exp[N(\ln 2 - e^2)] \\ \text{Var}\{\mathcal{N}(e|\mathbf{E})\} &= 2^N \times P(Ne)(N\delta e)(1 - P(Ne)(N\delta e)) \\ &\sim \exp[N(\ln 2 - e^2)] \end{aligned} \quad (41)$$

となる．これより N が大きいときを考えると状態密度の形は図 12 に示すとおり

- $|e| < \sqrt{\ln 2}$ 指数的．揺らぎ/平均は無視できる
- $|e| > \sqrt{\ln 2}$ ゼロとみなしてよい

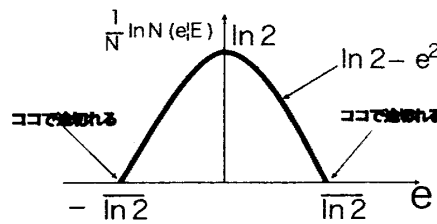


図 12: REM における状態密度の形

となるので，実質的にランダム性を考えなくてもよいことになり，ほとんど全ての実現値に対して典型時（1 に漸近する確率で）の評価をすることができる．

鞍点法を用いると，ほとんど全てのエネルギーランドスケープの実現値に対して

$$\begin{aligned} \frac{1}{N} \ln Z(\beta|\mathbf{E}) &= \frac{1}{N} \ln \int de e^{-\beta Ne} \mathcal{N}(e|\mathbf{E}) \sim \max_e \left\{ -\beta e + \frac{1}{N} \ln \mathcal{N}(e|\mathbf{E}) \right\} \\ &= \begin{cases} \frac{\beta^2}{4} + \ln 2, & \beta < \beta_c = 2\sqrt{\ln 2} \\ \beta\sqrt{\ln 2}, & \beta > \beta_c = 2\sqrt{\ln 2} \end{cases} \end{aligned} \quad (42)$$

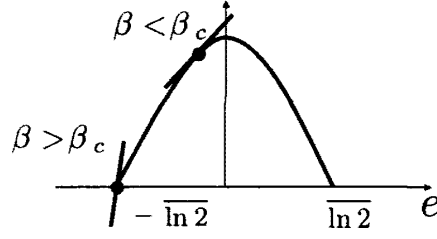


図 13: 鞍点法による評価

となり, これより自由エネルギー $f(\beta)$, エネルギー $u(\beta)$, エントロピー $s(\beta)$ は次のように相転移を起こすことが示される.

$$\begin{aligned} f(\beta) &= \begin{cases} -\frac{\beta}{4} - \frac{\ln 2}{\beta}, & \beta < \beta_c = 2\sqrt{\ln 2} \\ -\sqrt{\ln 2}, & \beta = \beta_c = 2\sqrt{\ln 2} \end{cases} \\ u(\beta) &= \begin{cases} -\frac{\beta}{2}, & \beta < \beta_c = 2\sqrt{\ln 2} \\ -\sqrt{\ln 2}, & \beta = \beta_c = 2\sqrt{\ln 2} \end{cases} \\ s(\beta) &= \beta(u(\beta) - f(\beta)) = \begin{cases} -\frac{\beta^2}{4} + \ln 2, & \beta < \beta_c = 2\sqrt{\ln 2} \\ 0, & \beta = \beta_c = 2\sqrt{\ln 2} \end{cases} \end{aligned} \quad (43)$$

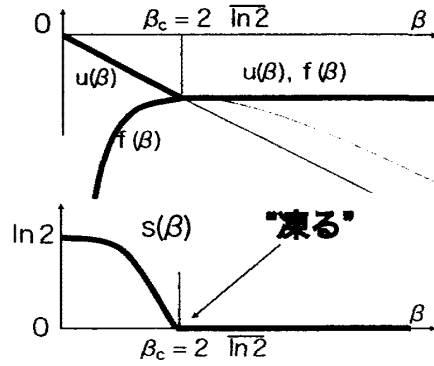


図 14: REM の相転移

2.1.3 レプリカ法による評価

次に REM のレプリカ法による評価を行う. レプリカ法ではレプリカトリックを用いて自由エネルギーを

$$\begin{aligned} f(\beta) &= -\frac{1}{N\beta} [\ln Z(\beta|\mathbf{E})]_{\mathbf{E}} \\ &= \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \lim_{N \rightarrow \infty} \frac{1}{N\beta} \ln [Z^n(\beta|\mathbf{E})]_{\mathbf{E}} \end{aligned} \quad (44)$$

というように計算する。計算の方針はこれまでも述べたように $n = 1, 2, \dots$ に対して $[Z^n(\beta|E)]_E$ を評価して、 n を実数（あるいは複素数）に解析接続するというものである。REM の分配関数 (38) 式から、

$$[Z^n(S|E)]_E = \left[\left(\sum_S e^{-\beta H(S|E)} \right)^n \right]_E = \sum_{S^1, S^2, \dots, S^n} \left[e^{-\beta \sum_{a=1}^n H(S^a|E)} \right]_E \quad (45)$$

によって $[Z^n(S|E)]_E$ を計算することになるが、(37) 式を使って指数の肩を書き換えると

$$[Z^n(S|E)]_E = \left[\left(\sum_S e^{-\beta \sum_{\tau} E(\tau) \delta(S, \tau)} \right)^n \right]_E = \sum_{S^1, S^2, \dots, S^n} \left[e^{-\beta \sum_{\tau} E(\tau) \times \sum_{a=1}^n \delta(S^a, \tau)} \right]_E \quad (46)$$

となる。各状態 τ のエネルギーが (36) 式の確率からランダムに発生することより、各状態 τ における平均の計算は

$$\begin{aligned} \left[e^{-\beta E(\tau) \times \sum_{a=1}^n \delta(S^a, \tau)} \right]_E &= \int \frac{dE(\tau)}{\sqrt{\pi N}} e^{-\frac{E^2(\tau)}{N} - \beta E(\tau) \times (\sum_{a=1}^n \delta(S^a, \tau))} \\ &= \exp \left[\frac{N\beta^2}{4} \left(\sum_{a=1}^n \delta(S^a, \tau) \right)^2 \right] \end{aligned} \quad (47)$$

となる。よって

$$\begin{aligned} [Z^n(S|E)]_E &= \sum_{S^1, S^2, \dots, S^n} \prod_{\tau} \exp \left[\frac{N\beta^2}{4} \left(\sum_{a=1}^n \delta(S^a, \tau) \right)^2 \right] \\ &= \sum_{S^1, S^2, \dots, S^n} \exp \left[\frac{N\beta^2}{4} \sum_{\tau} \left(\sum_{a=1}^n \delta(S^a, \tau) \right)^2 \right] \\ &= \sum_{S^1, S^2, \dots, S^n} \exp \left[\frac{N\beta^2}{4} \sum_{\tau} \sum_{a=1}^n \sum_{b=1}^n \delta(S^a, \tau) \delta(S^b, \tau) \right] \\ &= \sum_{S^1, S^2, \dots, S^n} \exp \left[\frac{N\beta^2}{4} \sum_{a,b} \delta(S^a, S^b) \right] \end{aligned} \quad (48)$$

となる。 $\delta(S^a S^b)$ は $\delta(S^a S^a) = 1$ で、 a, b に関して対称なので、独立な変数の数は $\frac{n(n-1)}{2}$ である。次の恒等式が成り立つことに注意する。

$$\sum_{q_{ab}=0,1} \prod_{a>b} \delta[\delta(S^a, S^b) - q_{ab}] = 1 \quad (49)$$

ここで q_{ab} はこの系のオーダーパラメータになっている。この式を (48) 式に代入すると

$$\begin{aligned} [Z^n(\beta|E)] &= \sum_{q_{ab}=0,1} \sum_{S^1, \dots, S^n} \exp \left[\frac{N\beta^2}{4} \sum_{a,b} \delta(S^a, S^b)^2 \right] \prod_{a>b} \delta[\delta(S^a, S^b) - q_{ab}] \\ &= \sum_{q_{ab}=0,1} \exp \left[\frac{N\beta^2}{4} \sum_{a,b} q_{ab} \right] \times \underbrace{\sum_{S^1, \dots, S^n} \prod_{a>b} \delta[\delta(S^a, S^b) - q_{ab}]}_{\exp[N S(\{q_{ab}\})]} \end{aligned} \quad (50)$$

$$= \sum_{q_{ab}=0,1} \exp \left[N \left\{ \frac{\beta^2}{4} \sum_{a,b} q_{ab} + S(\{q_{ab}\}) \right\} \right] \quad (51)$$

を得る。最後の式に現れた

$$S(\{q_{ab}\}) = \frac{1}{N} \ln \left[\sum_{S^1, \dots, S^n} \prod_{a>b} \delta[\delta(S^a, S^b) - q_{ab}] \right] \quad (52)$$

はあるオーダーパラメータ $\{q_{ab}\}$ に対するレプリカスピンの微視的状态数の対数をシステムサイズ N で割ったものになっており, つまり $\{q_{ab}\}$ によって特徴付けられる巨視的状态に対する (1 自由度当たりの) エントロピーになっている.

有限の N に対して, (51) 式では $\frac{n(n-1)}{2}$ 個の q_{ab} の和をとることになる. さらに各 q_{ab} のそれぞれに対して $S^1, S^2, \dots, S^n$ に関する 2^{nN} 回の和をとらなければならない. しかし, 熱力学極限 $N \rightarrow \infty$ では, 各項間の大きさの優劣がその指数ではっきり現れてくるので, 鞍点法により (51) 式の最後の式の指数の肩はその値がもっとも大きくなるときの $\{q_{ab}\}$ を代入した値

$$\max_{\{q_{ab}\}} N \left\{ \frac{\beta^2}{4} \sum_{a,b} q_{ab} + S(\{q_{ab}\}) \right\} \quad (53)$$

で代表して構わない. これにより, $n = 1, 2, \dots$, に対しては $[Z^n(\beta|E)]$ を

$$[Z^n(\beta|E)] = \sum_{q_{ab}=0,1} \exp \left[\max_{\{q_{ab}\}} N \left\{ \frac{\beta^2}{4} \sum_{a,b} q_{ab} + S(\{q_{ab}\}) \right\} \right] \quad (54)$$

で評価することができる.

ここまでで, (51) 式の指数の肩を最大にする $\{q_{ab}\}$ を選べば, $[Z^n(\beta|E)]$ を評価できることがわかった. しかし, その評価を行うためには $\frac{n(n-1)}{2}$ 個の q_{ab} を選ぶ離散最大化問題を解かなくてはならない. これはナイーブにやろうとすると骨の折れる問題である.

そこで, レプリカ系のもつ特徴を使って問題をうまく簡単化する方法を考える. まず, レプリカ系 $S^1, S^2, \dots, S^n$ は同じエネルギーランドスケープを共有している同等なものである. このことから (50) 式の指数の肩は添え字 $1, 2, \dots, n$ の入れ替えに関して不変となる対称性を有する. よって目的の指数の肩を最大にする点もレプリカ変数に関して対称性があり, 任意の $a > b$ に対して

$$q_{ab} = q \quad (55)$$

となることが期待できる. この仮定をレプリカ対称 (replica symmetric: RS) 仮定と呼び, その仮定の下で得られる解をレプリカ対称 (RS) 解という. q は 0 か 1 のどちらかの値しか取りえないので, RS 解もそれに応じて 2 つ存在しか存在し得ない.

まず RS1 として, $q = 0$ のとき, つまり

$$(q_{ab}) = \left(\begin{array}{ccc} \overbrace{1 \quad \dots \quad 1}^n & & 0 \\ & \ddots & \\ 0 & & \overbrace{1 \quad \dots \quad 1}^n \end{array} \right) \Bigg\}_n \quad (56)$$

における解を求める. このとき

$$\begin{aligned} \sum q_{ab} &= n \\ e^{NS(\{q_{ab}\})} &= 2^N \times (2^N - 1) \times (2^N - 2) \cdots (2^N - n) \\ &\simeq 2^{nN} = e^{Nn \ln 2} \end{aligned} \quad (57)$$

が得られて

$$[Z^n(\beta|E)]_{RS1} = \exp \left[Nn \left(\frac{\beta^2}{4} + \ln 2 \right) \right] \quad (58)$$

となる.

次に RS2 として, $q = 1$ のとき, つまり

$$(q_{ab}) = \left(\overbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}^n \right) \Bigg\}_n \quad (59)$$

のときの解を求める. このときは

$$\begin{aligned} \sum q_{ab} &= n^2 \\ e^{NS(\{q_{ab}\})} &= 2^N \times 1 \cdots 1 \\ &= e^{N \ln 2} \end{aligned} \quad (60)$$

が得られて

$$[Z^n(\beta|E)]_{RS2} = \exp \left[N \left(\frac{n^2 \beta^2}{4} + \ln 2 \right) \right] \quad (61)$$

となる.

2つのレプリカ対称解が得られたが, それらの妥当性を検討しなければならない. $\frac{1}{N} [\ln Z(\beta|E)]$ は RS1 では

$$\begin{aligned} \frac{1}{N} [\ln Z^n(\beta|E)]_{RS1} &= \lim_{n \rightarrow 0} \frac{1}{n} n \left(\frac{\beta^2}{4} + \ln 2 \right) \\ &= \frac{\beta^2}{4} + \ln 2 \end{aligned} \quad (62)$$

となり, RS2 では

$$\begin{aligned} \frac{1}{N} [\ln Z^n(\beta|E)]_{RS1} &= \lim_{n \rightarrow 0} \frac{1}{n} \left(\frac{n^2 \beta^2}{4} + \ln 2 \right) \\ &= \infty \end{aligned} \quad (63)$$

となる.

よってレプリカ法を使わないで求めた解と合う解は RS1 ということになり, RS2 ではおかしい結果を導いてしまっている. RS 解を求める際に鞍点法を用いた. 鞍点法によれば得られた解のうち大きい方を選択することになるはずなのだが, 今, RS2 は正しい解を与えない. (58) と (61) 式の解は $n \in \mathbf{N}$ については妥当性が数学的に保証されているのであるが, これを実数 $n \in \mathbf{R}$ に解析接続すると主要解の選択が自明ではなくなるのである. このことについては後でもう一度検討することにする.

$n \in \mathbf{N}$ に対しレプリカ対称性を課して鞍点解を求めると, 自然に実数への解析接続が可能な形で RS 解が与えられることがわかった. しかし, (42) 式と比較すると, 低温相で RS 解は正しい結果を導かないことが分かる. これまでの研究から, この問題を解決するためにはレプリカ変数の対称性を破った RSB(replica symmetric breaking) 解を構成する必要があることが分かっている.

具体的には, $n \times n$ の q_{ab} 行列をすべての要素が 1 の $m \times m (m < n)$ のブロック行列を対角成分に並べ

て、それ以外はゼロに分けた次のようなものを考える。

$$(q_{ab}) = \begin{pmatrix} \boxed{\begin{matrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{matrix}} & 0 & 0 \\ 0 & \boxed{\begin{matrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{matrix}} & 0 \\ 0 & 0 & \boxed{\begin{matrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{matrix}} \end{pmatrix} \quad (64)$$

これを1ステップRSB(1RSB)とよぶ。問題によっては、 q_{ab} が0,1以外の様々な値をとることもある。そういった場合は、 $m \times m$ 行列に対して1RSBの構造を入れ子で課していくことで2ステップRSB, 3ステップRSBの解を構成していくことも可能である。REMでは1RSBで十分である。1RSBでは、

$$\begin{aligned} \sum q_{ab} &= m^2 \times \frac{n}{m} = mn \\ e^{NS(\{q_{ab}\})} &= 2^N \times (2^N - 1) \times \dots \times \left(2^N - \frac{n}{m} + 1\right) \times {}_nC_m \times {}_{n-m}C_m \times \dots \times {}_mC_m \\ &\simeq \exp \left[N \frac{n}{m} \ln 2 \right] \end{aligned} \quad (65)$$

となり、よって

$$[Z^n(\beta|\mathbf{E})]_{1RSB} = \exp \left[N \frac{n}{m} \left(\frac{m^2 \beta^2}{4} + \ln 2 \right) \right] \quad (66)$$

が得られる。ここでブロックの大きさ m をどのように決めるかが問題になるが、ここでは少々天下り的に極値条件により決めることにする。すると、そのような m は

$$m = \frac{2\sqrt{\ln 2}}{\beta} \quad (67)$$

であり、1RSB解

$$\begin{aligned} \frac{1}{N} [\ln Z^n(\beta|\mathbf{E})]_{RS1} &= \lim_{n \rightarrow 0} \frac{1}{n} n \beta \sqrt{\ln 2} \\ &= \beta \sqrt{\ln 2} \end{aligned} \quad (68)$$

が得られる。これは(42)式が示す低温相での正解と一致している。

ここでもう一度どのような解を選択するべきなのか、ということを考えてみよう。これまでにRS1, RS2, 1RSBと三つの解が得られた。そして、 $\beta > \beta_c$ の高温側でRS1の $\frac{1}{N} [Z^n(\beta|\mathbf{E})] = \frac{\beta^2}{4} + \ln 2$, $\beta < \beta_c$ の低温側で1RSBの $\frac{1}{N} [Z^n(\beta|\mathbf{E})] = \beta \sqrt{\ln 2}$ という“正解”を導いた。 $n \in \mathbf{N}$ に対する高温相と低温相での解の振舞いをまとめると図15のようになる。

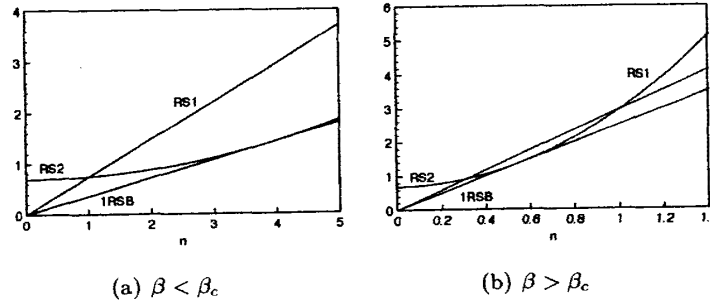


図 15: (a) $\beta < \frac{1}{2} < 2\sqrt{\ln 2} = \beta_c$ と (b) $\beta = 3 > \beta_c$ に対する RS1, RS2, 1RSB の大小関係. (a), (b) ともに RS1 と RS2 が $n = 1$ で交差する. また, RS2 と 1RSB の接点 $n_{1RSB} = \frac{\beta_c}{\beta}$ は (a) では $n_{1RSB} > 1$, (b) では $0 < n_{1RSB} < 1$ となる.

図の通り $n \in \mathbb{N}$ に対して正しい結果を導く「値の大きな鞍点解を選択する」という選択方法を $n \in \mathbb{R}$ にそのままあてはめると、いずれの場合も $n \rightarrow 0$ で RS2 が選ばれるべきである。しかしながら、 $\frac{1}{N} [Z^n(\beta) | \mathbf{E}]_{RS2} = \ln 2$ となり、これは間違った結果を導いている。なぜなら $Z(\beta | \mathbf{E})$ は正定値なので $\lim_{n \rightarrow 0} [Z^n(\beta | \mathbf{E})] = 1$ となるはずだからである。また、ここでは詳しく説明しないが、エントロピーが負になってしまうという問題も生じる。つまり、 $n \in \mathbb{N}$ での鞍点解の選択基準は自然数以外の n に対して一般には通用しないのである。

それでは、どのような解を選択するべきなのか。結論から述べると、数学的に正当性が保証されている一般的な処方箋は今のところ存在しない。しかしながら、ここで説明した REM など、経験的に選択基準が確立している問題もあり、これまでの研究においては、負のエントロピーを避けるなど物理的によく合う解を試行錯誤して選ぶことで矛盾を解消する処方箋が経験的に作られてきた。この経験則をまとめると以下ようになる。

i) $n = 1$ のすぐ右側で優勢な RS 解を選択する。

ii) i) で選択した RS 解に $n < 1$ のどこかで不都合が生じた場合には適宜 RSB 解を構成する。

こういった数学的ないかがわしさはある程度許容しながら、物理的にうまく解釈していくことが現時点においてレプリカ法を使いこなすには大切なことなのである。

2.2 ランダム行列の固有値分布

2.2.1 ランダム行列とは

レプリカ法を使う次の例題としてランダム行列の固有値分布を考える。ランダム行列は乱数の要素を持つ行列で、応用範囲が広く解析数論、組み合わせ論、原子核物理学、素粒子物理学、固体物理学、生態学、金融工学など、多くの分野で使われている。ここではランダム対称行列に関する固有値の漸近分布がレプリカ法により求まることを示す。

行列 C は

$$C \sim P(C) \quad (69)$$

のように確率 $P(C)$ で決まるとする。 $\lambda_1, \lambda_2, \dots, \lambda_N$ を C の固有値とすると、その分布は

$$\rho_N(\lambda) = \frac{1}{N} \sum_{\mu=1}^N \delta(\lambda - \lambda_\mu) \quad (70)$$

となる行列のアンサンブル $P(C)$ に対する固有値の漸近分布は

$$\rho(\lambda) = \lim_{N \rightarrow \infty} [\rho_N(\lambda)]_C = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\mu=1}^N [\delta(\lambda - \lambda_\mu)]_C \quad (71)$$

と定義される. $[\dots]_C$ は C に関する平均である. これも 2 重平均の構造であり, レプリカ法による評価が有効であると考えられる. ただし, この評価では孤立固有値は無視されており, また, 有限サイズ依存性が見えなくなっていることは注意しておこう.

ここからの計算では Kramers-Kronig の分散公式

$$\begin{aligned} \delta(\lambda - \lambda_\mu) &= \lim_{\epsilon \rightarrow +0} \frac{1}{\pi} \text{Im} \frac{1}{\lambda - \lambda_\mu - i\epsilon} = \frac{1}{\pi} \text{Im} \frac{1}{\lambda - \lambda_\mu} \\ &= \frac{1}{\pi} \text{Im} \left\{ \frac{\partial}{\partial \lambda} \ln(\lambda - \lambda_\mu) \right\} \end{aligned} \quad (72)$$

が鍵となる. これにより, 固有値分布は

$$\begin{aligned} \rho_N(\lambda) &= \frac{1}{N} \sum_{\mu=1}^N \delta(\lambda - \lambda_\mu) \\ &= \frac{1}{\pi N} \text{Im} \left\{ \frac{\partial}{\partial \lambda} \ln(\lambda - \lambda_\mu) \right\} = \frac{1}{\pi N} \frac{\partial}{\partial \lambda} \text{Im} \left\{ \ln \prod_{\mu=1}^N (\lambda - \lambda_\mu) \right\} \\ &= \frac{1}{\pi N} \text{Im} \left\{ \frac{\partial}{\partial \lambda} \ln \det(\lambda - C) \right\} \end{aligned} \quad (73)$$

となる. 対称行列を仮定すると, ここで現れた行列式はガウス積分の公式によって

$$\begin{aligned} \det(\lambda - C)^{-\frac{1}{2}} &= (2\pi)^{-\frac{N}{2}} \int d\mathbf{x} e^{-\frac{1}{2} \mathbf{x}^T (\lambda - C) \mathbf{x}} \\ &= (2\pi)^{-\frac{N}{2}} \int d\mathbf{x} e^{-\frac{\lambda}{2} |\mathbf{x}|^2} \times \exp \left[\frac{1}{2} \mathbf{x}^T C \mathbf{x} \right] \\ &= Z(\lambda|C) \end{aligned} \quad (74)$$

と分配関数の形に表現することが可能であり, 固有値漸近分布は

$$\rho(\lambda) = \frac{-2}{\pi} \text{Im} \left\{ \lim_{N \rightarrow \infty} \frac{1}{N} \frac{\partial}{\partial \lambda} [\ln Z(\lambda|C)]_C \right\} \quad (75)$$

となる.

ここでレプリカ法を使う. 今までと同じ手続きで自然数 n に対して

$$\begin{aligned} [Z^n(\lambda|C)]_C &= \int d\mathbf{x}^1 d\mathbf{x}^2 \dots d\mathbf{x}^n e^{-\frac{\lambda}{2} \sum_{a=1}^n |\mathbf{x}^a|^2} \times \left[e^{\frac{1}{2} \sum_{a=1}^n (\mathbf{x}^a)^T C \mathbf{x}^a} \right]_C \\ &= \int \prod_{a=1}^n N dQ_a \prod_{a>b} N dq_{ab} \underbrace{\int \prod_{a=1}^n d\mathbf{x}^a \prod_{a=1}^n \delta(|\mathbf{x}^a|^2 - N Q_a) \prod_{a>b} \delta(\mathbf{x}^a \cdot \mathbf{x}^b - N q_{ab}) e^{-\frac{\lambda}{2} \sum_{a=1}^n |\mathbf{x}^a|^2}}_{(B)} \\ &\quad \times \underbrace{\left[e^{\frac{1}{2} \sum_{a=1}^n (\mathbf{x}^a)^T C \mathbf{x}^a} \right]_C}_{(A)} \end{aligned} \quad (76)$$

を計算する. 行列 C は問題により様々な要素を持ったものになるが, ここでは次のようにガウス型アンサンブルをとるもの考える.

$$P(C) = \left(\frac{N}{2\pi} \right)^{-N(N-1)/2} \exp \left[-\frac{N}{2} \sum_{i>j} c_{ij}^2 \right] \quad \left(\begin{array}{l} c_{ij} = c_{ji} \\ c_{ii} = 0 \end{array} \right) \quad (77)$$

この結果はよく知られたものになるが、それをレプリカ法で確認してみよう。(76) 式の (A) と (B) をそれぞれ別々に計算する。

(A) 各 c_{ij} に注目

$$\begin{aligned} \int dc_{ij} \sqrt{\frac{N}{2\pi}} e^{-\frac{N}{2} c_{ij}^2} \exp \left[c_{ij} \sum_{a=1}^n x_i^a x_j^a \right] &= \exp \left[\frac{1}{2N} \left(\sum_{a=1}^n x_i^a x_j^a \right)^2 \right] \\ \therefore \left[e^{\frac{1}{2} \sum_{a=1}^n (\mathbf{x}^a)^T \mathbf{C} \mathbf{x}^a} \right]_C &= \exp \left[\frac{1}{2N} \sum_{i>j} \left(\sum_{a=1}^n x_i^a x_j^a \right)^2 \right] \\ \sum_{i>j} \left(\sum_{a=1}^n x_i^a x_j^a \right)^2 &\simeq N^2 \times \left(\frac{1}{2} \sum_{a=1}^n Q_a^2 + \sum_{a>b} q_{ab}^2 \right) \quad \begin{cases} Q_a \equiv \frac{1}{N} \sum_{i=1}^N (x_i^a)^2 \\ q_{ab} \equiv \frac{1}{N} \sum_{i=1}^N x_i^a x_i^b \end{cases} \\ \therefore \left[e^{\frac{1}{2} \sum_{a=1}^n (\mathbf{x}^a)^T \mathbf{C} \mathbf{x}^a} \right]_C &= \exp \left[N \left(\frac{1}{4} \sum_{a=1}^n Q_a^2 + \frac{1}{2} \sum_{a>b} q_{ab}^2 \right) \right] \end{aligned}$$

(B)

$$\begin{aligned} \int \prod_{a=1}^n d\mathbf{x}^a \prod_{a=1}^n dQ_a \delta(|\mathbf{x}_a|^2 - NQ_a) \prod_{a>b} dq_{ab} \delta(\mathbf{x}^a \cdot \mathbf{x}^b - Nq_{ab}) e^{-\frac{\lambda}{2} \sum_{a=1}^n |\mathbf{x}^a|^2} \\ \begin{cases} \delta(|\mathbf{x}_a|^2 - NQ_a) = \int d\hat{Q}_a e^{\frac{\hat{Q}_a}{2} (|\mathbf{x}_a|^2 - NQ_a)} \\ \delta(\mathbf{x}^a \cdot \mathbf{x}^b - Nq_{ab}) = \int d\hat{q}_{ab} e^{\hat{q}_{ab} (\mathbf{x}^a \cdot \mathbf{x}^b - Nq_{ab})} \end{cases} \\ \propto (\det(\lambda - \hat{\mathbf{Q}}))^{-\frac{N}{2}} \exp \left[-\frac{N}{2} \text{Tr} \hat{\mathbf{Q}} \mathbf{Q} \right] \\ \mathbf{Q} = \left(\begin{array}{ccc} \overbrace{Q_1}^n & & q_{ab} \\ & \ddots & \\ q_{ab} & & Q_n \end{array} \right) \Bigg\}_n \quad \hat{\mathbf{Q}} = \left(\begin{array}{ccc} \overbrace{\hat{Q}_1}^n & & \hat{q}_{ab} \\ & \ddots & \\ \hat{q}_{ab} & & \hat{Q}_n \end{array} \right) \Bigg\}_n \end{aligned} \quad (78)$$

ここで、レプリカ対称 (RS) 仮定として、 $Q_a = Q, q_{ab} = q, \hat{Q}_a = \hat{Q}, \hat{q}_{ab} = \hat{q}$ とすると、

$$\begin{aligned} \frac{1}{N} [\ln Z(\lambda | C)]_C &= \lim_{n \rightarrow 0} \frac{1}{n} \frac{1}{N} \ln [Z^n(\lambda | C)]_C \\ &= \text{Extr}_{Q, q, \hat{Q}, \hat{q}} \left\{ \frac{1}{4} (Q^2 - q^2) - \frac{1}{2} (Q\hat{Q} - q\hat{q}) \right. \\ &\quad \left. - \frac{1}{2} \ln(\lambda - \hat{Q} + \hat{q}) + \frac{\hat{q}}{2[\lambda - \hat{Q} + \hat{q}]} + \text{const} \right\} \end{aligned} \quad (79)$$

$$\begin{cases} \frac{Q}{2} - \frac{\hat{Q}}{2} = 0 \\ -\frac{q}{2} + \frac{\hat{q}}{2} = 0 \end{cases} \quad \begin{cases} -\frac{1}{2} Q + \frac{1}{2(\lambda - \hat{Q} + \hat{q})} + \frac{\hat{q}}{2[\lambda - \hat{Q} + \hat{q}]^2} = 0 \\ \frac{1}{2} q - \frac{1}{2(\lambda - \hat{Q} + \hat{q})} + \frac{1}{2(\lambda - \hat{Q} + \hat{q})} - \frac{\hat{q}}{2[\lambda - \hat{Q} + \hat{q}]^2} = 0 \end{cases} \quad (80)$$

となり、まとめると、

$$q = 0$$

$$-Q + \frac{1}{\lambda - Q} = 0 \Rightarrow Q^2 - \lambda Q + 1 = 0 \Rightarrow Q = \frac{\lambda \pm \sqrt{\lambda^2 - 4}}{2} \quad (81)$$

が得られて、(75) 式により固有値の漸近分布は

$$\begin{aligned} \rho(\lambda) &= \frac{-2}{\pi} \text{Im} \left\{ \lim_{N \rightarrow \infty} \frac{1}{N} \frac{\partial}{\partial \lambda} [\ln Z(\lambda|C)]_C \right\} \\ &= \frac{1}{\pi} \text{Im} \left\{ \frac{1}{\lambda - \bar{Q}} \right\} = \frac{1}{\pi} \text{Im}(Q) = \frac{\sqrt{4 - \lambda^2}}{2\pi} \end{aligned} \quad (82)$$

となる (図 16). この結果は Wigner 半円則として知られている.

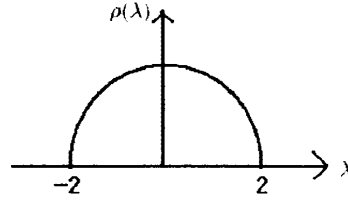


図 16: Wigner 分布

分散公式とレプリカ法に基づいてランダム行列の固有値に関する漸近分布の評価を行った. この手法はガウス型のアンサンブルに限らず広い行列アンサンブルのクラスに利用することができる. しかしながら, ここで示した評価方法は万能ではない. まず, こぼれ (孤立固有値) が評価できない. それに N に対する有限サイズスケーリングも不明である. さらに順位統計, 最大固有値の分布なども不明である. それでもなお, 漸近分布を評価する際には極めて強力で汎用性の高い方法である.

3 コトを手なずける

ここまで, 統計力学におけるスピングラス理論で発展してきたレプリカ法を用いて情報論や様々な問題について説明できることを示してきた.

大自由度系では, 次元に関して計算量が指数的に爆発してしまう. 平衡統計力学の歴史はこの問題との戦いの歴史である, と言っても過言ではない. そのため, 統計力学は大自由度系における物理量を評価する様々な方法を発展させてきた. その一つである平均場近似の理論が最近情報科学との境界で発展している. ここではそれを紹介していく.

3.1 グラフ表現とベータ近似 (信念伝搬)

3.1.1 ベータ近似

多次元の確率分布に関する平均量の計算は

$$\langle O(S) \rangle = \sum_S O(S) P(S) \quad (83)$$

によってなされる. ここで $O(S)$ は系がとり得る N 個の要素を持つ状態変数 S における物理量, $P(S)$ は状態 S が現れる確率である. この厳密な値を求めるのに必要な計算量は一般に次元数の指数で増加す

る。ここでは、正しい分布を計算の容易な別の分布に近似することで問題解決を図ることにする。そのための方法が平均場近似である。

平均場近似にもいろいろな近似の仕方があるが、ベータ近似というものについて説明していく。簡単のため

$$O(S) = S \quad (84)$$

の場合について述べることにするが、それ以外の場合への拡張も可能である。(84)式に対して、(83)式の計算をするのも要素数 N に関して指数関数的に計算量が増大し、現実時間での厳密計算の実行は困難になる。

ベータ近似の着眼点は分布のグラフ構造と計算量の関係を用いることである。図17のようなグラフ表現(2部グラフ表現)について考えよう。

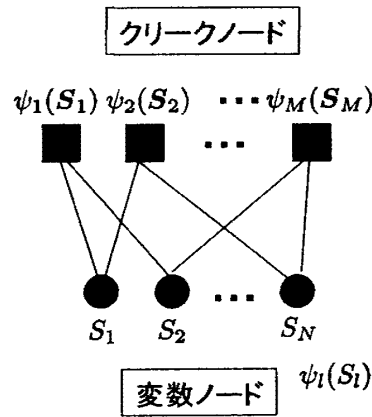


図 17: 分布の2部グラフ表現

2部グラフは状態 S となる確率の計算

$$P(S) = \frac{\prod_{\mu=1}^M \psi_{\mu}(S_{\mu}) \prod_{l=1}^N \psi_l(S_l)}{Z} \quad (85)$$

をグラフ表現したものである。 $\psi_{\mu}(S)$ はグラフの構成要素 μ が確率に与える影響を表す関数である。グラフの構成要素 μ をクリークと呼ぶ。また、 $\psi_l(S)$ は確率変数ノードが確率に与える影響を表す関数である。

一般統計モデルについて変数とクリークの繋がりをグラフ表現した際にグラフにループ(循環構造)が存在しなければ(図18), 平均量を少ない計算量で評価できることがわかっている。ループのないグラフ

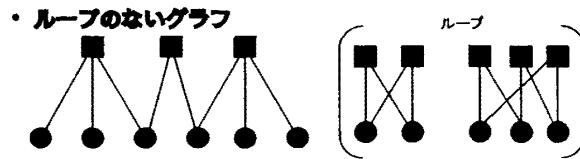


図 18: ループのないグラフとループのあるグラフ

では周辺化表現

$$P(S) = \frac{\prod_{\mu=1}^M P_{\mu}(S_{\mu})}{\prod_{l=1}^N P_l^{C_l-1}(S_l)} \quad \left(\begin{array}{l} P_{\mu}(S_{\mu}) = \sum_{S \setminus S_{\mu}} P(S) \\ P_l(S_l) = \sum_{S \setminus S_l} P(S) \end{array} \right) \quad (86)$$

が成立する．ここで， $\mathbf{S} \setminus \mathbf{S}_\mu$ は $\mathbf{S}$ からクリーク $\mathbf{S}_\mu$ の要素を除くことを意味する． $\mathbf{S} \setminus S_l$ なども同様の表現である．また C_l は変数 S_l が含まれるクリークの数である．

一般に近似を行うには，元の分布と近似した分布がどれ程近いかという尺度を導入すると便利である．そこで，2つの分布 $P(\mathbf{S})$ と $Q(\mathbf{S})$ の間の近さを測る量として Kullback-Leibler (以下，KL) ダイバージェンス

$$KL(Q, P) = \sum_{\mathbf{S}} Q(\mathbf{S}) \ln \frac{Q(\mathbf{S})}{P(\mathbf{S})} \geq 0 \quad (87)$$

を導入する．KL ダイバージェンスは常に 0 以上という条件があり，分布間の近さを表すものであるが $KL(Q, P) \neq KL(P, Q)$ など距離の公理を満たさないので普通の意味での距離ではない．KL ダイバージェンスに

$$\begin{cases} P(\mathbf{S}) = \frac{\prod_{\mu=1}^M \psi_{\mu}(\mathbf{S}_{\mu}) \prod_{l=1}^N \psi_l(S_l)}{Z} \\ Q(\mathbf{S}) = \frac{\prod_{\mu=1}^M b_{\mu}(\mathbf{S}_{\mu})}{\prod_{l=1}^N b_l^{C_l-1}(S_l)} \end{cases} \quad (88)$$

を代入して，拘束条件

$$\sum_{\mathbf{S}_{\mu}} b_{\mu}(\mathbf{S}_{\mu}) = 1 \quad \sum_{S_l} b_l(S_l) = 1 \quad \sum_{\mathbf{S}_{\mu} \setminus S_l} b_{\mu}(\mathbf{S}_{\mu}) = b_l(S_l) \quad (89)$$

のもとで最小化するように $b_l(S_l)$ を決める．これはベータ自由エネルギー

$$F_{\text{Bethe}}(\{b_{\mu}, b_l\}) = -\ln Z + KL(\{b_{\mu}, b_l\} | P) \geq -\ln Z \quad (90)$$

$$\begin{aligned} F_{\text{Bethe}}(\{b_{\mu}, b_l\}) &= \sum_{\mu=1}^M \sum_{\mathbf{S}_{\mu}} b_{\mu}(\mathbf{S}_{\mu}) \ln \frac{b_{\mu}(\mathbf{S}_{\mu})}{\psi_{\mu}(\mathbf{S}_{\mu}) \prod_{l \in \mathcal{L}(\mu)} \psi_l(S_l)} \\ &\quad + \sum_{l=1}^N (1 - C_l) \sum_{S_l} b_l(S_l) \ln \frac{b_l(S_l)}{\psi_l(S_l)} \end{aligned} \quad (91)$$

を最小にすることと同じである．この最小化は Belief Propagation (BP: 信念伝搬) を使うことで高速に実現できる．

3.1.2 Belief Propagation

KL ダイバージェンスおよびベータ自由エネルギーの最小化問題は，試験分布に関する拘束条件

$$\sum_{\mathbf{S}_{\mu} \setminus S_l} b_{\mu}(\mathbf{S}_{\mu}) = b_l(S_l) \quad \Rightarrow \quad + \lambda_{\mu l}(S_l) \left(\sum_{\mathbf{S}_{\mu} \setminus S_l} b_{\mu}(\mathbf{S}_{\mu}) - b_l(S_l) \right) \quad (92)$$

$\lambda_{\mu l}(S_l)$: ラグランジュ未定乗数

を用いた拘束条件付最適化問題に帰着する．この極値条件を求めると，周辺化分布は未定乗数を用いて

$$\begin{aligned} b_{\mu}(\mathbf{S}_{\mu}) &= \frac{\psi_{\mu}(\mathbf{S}_{\mu}) \prod_{l \in \mathcal{L}(\mu)} (\psi_l(S_l) e^{-\lambda_{\mu l}(S_l)})}{\sum_{\mathbf{S}_{\mu}} \psi_{\mu}(\mathbf{S}_{\mu}) \prod_{l \in \mathcal{L}(\mu)} (\psi_l(S_l) e^{-\lambda_{\mu l}(S_l)})} \\ b_l(S_l) &= \frac{\psi_l(S_l) \prod_{\mu \in \mathcal{M}(l)} e^{-\lambda_{\mu l}(S_l)/(C_l-1)}}{\sum_{S_l} \psi_l(S_l) \prod_{\mu \in \mathcal{M}(l)} e^{-\lambda_{\mu l}(S_l)/(C_l-1)}} \end{aligned} \quad (93)$$

と表現される．未定乗数を用いて (92) 式を整理すると

$$e^{-\hat{\lambda}_{\mu l}(S_l)} \propto \sum_{S_\mu \setminus S_l} \psi_\mu(S_\mu) \prod_{j \in \mathcal{L}(\mu) \setminus l} (\psi_j(S_j) e^{-\lambda_{\mu j}(S_j)}) \quad (94)$$

$$\psi_l(S_l) e^{-\hat{\lambda}_{\mu l}(S_l)} \propto \psi_l(S_l) \prod_{\nu \in \mathcal{M}(l) \setminus \mu} e^{-\hat{\lambda}_{\nu l}(S_l)} \quad (95)$$

$$\left(\hat{\lambda}_{\mu l}(S_l) = \frac{1}{C_l - 1} \sum_{\nu \in \mathcal{M}(l)} \lambda_{\mu l}(S_l) - \lambda_{\mu l}(S_l) \right)$$

が得られる．ここから最適な乗数を決めればよい．ループのないグラフでは要素，グラフに適当な順序関係を定義することができ，その順序関係に沿ってグラフを一通り伝搬することで解くことができる．このような方法を Belief Propagation(以下，BP) という．

この種の非線形方程式は同期的な反復法を使って解くのが一般的である．ここでは変数を $\hat{m}_{\mu \rightarrow l}(S_l) \propto e^{-\hat{\lambda}_{\mu l}(S_l)}$ ， $m_{l \rightarrow \mu}(S_l) \propto \psi_l(S_l) e^{-\lambda_{\mu l}(S_l)}$ のように書き直し前向き反復

$$\begin{aligned} \hat{m}_{\mu \rightarrow l}^{t+1}(S_l) &= \alpha_{\mu \rightarrow l}^{t+1} \sum_{S_\mu \setminus S_l} \psi_\mu(S_\mu) \prod_{j \in \mathcal{L}(\mu) \setminus l} m_{j \rightarrow \mu}^t(S_j) \\ m_{l \rightarrow \mu}^t(S_l) &= \alpha_{l \rightarrow \mu}^t \psi_l(S_l) \prod_{\nu \in \mathcal{M}(l) \setminus \mu} \hat{m}_{\nu \rightarrow l}^t(S_l) \end{aligned} \quad (96)$$

を用いて解く．BP では図 19 のように，グラフ上でメッセージの受け渡しをして伝搬する．計算量はクリークサイズについて指数的になるが，計算はグラフ上を一通り伝搬するだけでよい．

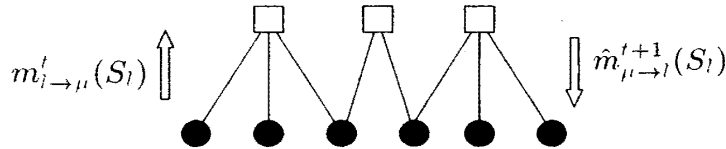


図 19: BP におけるグラフの伝播

試験分布の周辺化表現は，(96) で反復法により得られた解を (93) 式に代入することにより

$$b_l(S_l) = \frac{\psi_l(S_l) \prod_{\mu \in \mathcal{M}(l)} \hat{m}_{\mu \rightarrow l}(S_l)}{\sum_{S_l} \psi_l(S_l) \prod_{\mu \in \mathcal{M}(l)} \hat{m}_{\mu \rightarrow l}(S_l)} \quad (97)$$

となる．

BP は 1 次元鎖の場合，転送行列法のことである．ここでそのことを示していくが，以下の議論はループのない一般的なグラフでも同じである．

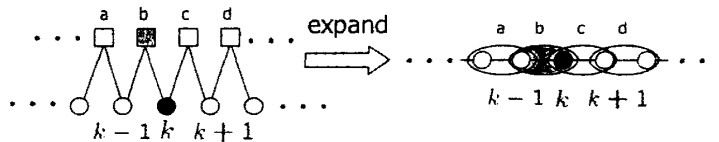


図 20: ループなし 2 部グラフと 1 次元鎖の対応関係

次元鎖のある要素 k に着目すると分布は

$$P(S) = \frac{\prod_k \psi_{k,k+1}(S_k, S_{k+1})}{\sum_S \prod_k \psi_{k,k+1}(S_k, S_{k+1})} \quad (98)$$

となる。ここで、この結合分布から周辺化された分布が

$$P(S_l) = \sum_{S \setminus S_l} P(S) \propto \underbrace{\left(\sum_{S_j < S_l} \prod_{j < l} \psi_{j,j+1}(S_j, S_{j+1}) \right)}_{\hat{m}_{k-1 \rightarrow k}(S_k)} \underbrace{\left(\sum_{S_i > S_l} \prod_{i > l} \psi_{i,i+1}(S_i, S_{i+1}) \right)}_{\hat{m}_{k+1 \rightarrow k}(S_k)} \quad (99)$$

と表現できることに着目する (図 21)。

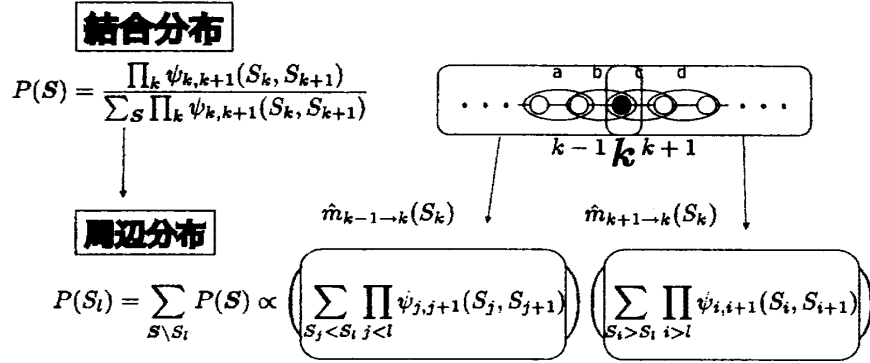


図 21: 次元鎖の分布の周辺化

$\hat{m}_{k-1 \rightarrow k}(S_k)$, $\hat{m}_{k+1 \rightarrow k}(S_k)$ を直接評価するのに必要な計算は添え字 k の位置に依存し、それぞれ S_k の左、右に位置する要素の個数の指数に比例して増大する。ところが、これらに関して漸化式

$$\begin{aligned} \hat{m}_{k \rightarrow k+1}(S_{k+1}) &= \sum_{S_j < k+1} \prod_{j < k+1} \psi_{j,j+1}(S_j, S_{j+1}) \\ &= \sum_{S_k} \psi_{k,k+1}(S_k, S_{k+1}) \left(\sum_{S_j < k} \prod_{j < k} \psi_{j,j+1}(S_j, S_{j+1}) \right) \\ &= \sum_{S_k} \psi_{k,k+1}(S_k, S_{k+1}) \hat{m}_{k-1 \rightarrow k}(S_k) \end{aligned} \quad (100)$$

が成立することに注意しよう。この漸化式を利用すると、現時点での $\hat{m}_{k-1 \rightarrow k}(S_k)$ あるいは $\hat{m}_{k+1 \rightarrow k}(S_k)$ を利用し、1 ステップ当たり $O(1)$ の計算量で隣の要素に関する $\hat{m}_{k-1 \rightarrow k}(S_k)$, $\hat{m}_{k+1 \rightarrow k}(S_k)$ を計算することが可能になる。すべての要素に関して、これを実行するのはグラフ上を一通り伝播するのに必要な高々 $O(N)$ の計算量で済む。これで、次元鎖分布の計算量が $O(2^N)$ から $O(N)$ に削減できることがわかった。

ループがあると KL の関係が成り立たない。

$$F_{\text{Bethe}}(\{b_\mu, b_l\}) \neq -\ln Z + KL(\{b_\mu, b_l\} | P) \geq -\ln Z \quad (101)$$

しかし、ベーテ自由エネルギーの最小化は“近似手法 (ベーテ近似)”として利用できる。この際、近似解の探索アルゴリズムとして BP も利用できる (この場合には特に loopy BP と呼ばれることがある) が、残念ながら、収束の保証はない。

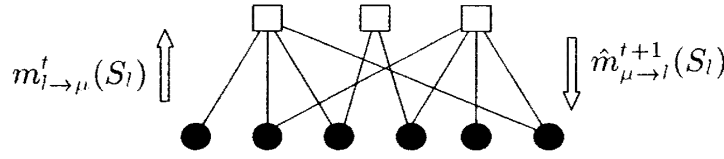


図 22: Loopy BP

3.1.3 ベーテ近似の応用例：誤り訂正符号

ベーテ近似の応用例として誤り訂正符号問題の復号問題がある。誤り訂正符号とは送信情報に冗長性を持たせることでノイズなどにより生じる情報の誤りを訂正するための技術である。\$K\$ ビットの 2 元 \$(0,1)\$ ベクトル \$\mathbf{x} = (x_1, x_2, \dots, x_K)\$ を原情報として、これを線形変換で \$N(> K)\$ ビットの 2 元ベクトル (符号語) \$\mathbf{y}^0 = (y_1^0, y_2^0, \dots, y_N^0)\$ へ符号化する線形符号を考える (図 23)。受信された情報の復号をする

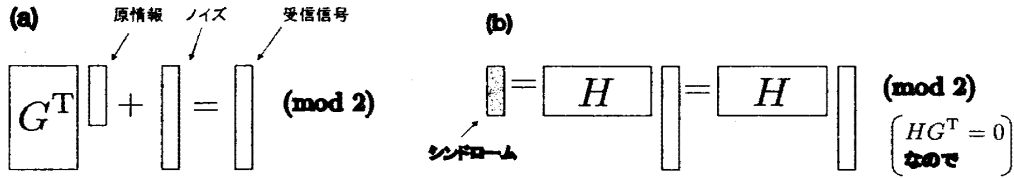


図 23: 線形符号の (a) 符号化と (b) 復号

ための、0 または 1 を成分とする \$(N - K \times N)\$ のパリティ検査行列 \$H\$ を定義する。\$H\$ が与えられると

$$HG^T = 0 \pmod{2} \quad (102)$$

を満たす、0 または 1 を成分とする \$N \times K\$ の生成行列 \$G^T\$ を構成することができる。線形符号化は \$G^T\$ を用いて

$$\mathbf{y}^0 = G^T \mathbf{x} \pmod{2} \quad (103)$$

により行われる。受信される信号 \$\mathbf{y}\$ はこれに \$N\$ ビットのノイズ \$\mathbf{n}\$ が加わった

$$\mathbf{y} = \underbrace{G^T \mathbf{x}}_{\mathbf{y}^0} + \mathbf{n} \pmod{2} \quad (104)$$

となる。簡単のため、ノイズとしては各ビットの値が、独立に確率 \$0 < p < 1/2\$ で反転するようなものを考える。このような通信路を 2 元対称通信路という。復号は符号語 \$\mathbf{y}\$ にパリティ検査行列 \$H\$ を左から掛けてシンδροーム \$\mathbf{z} = H\mathbf{y} \pmod{2}\$ を計算し、パリティ検査方程式

$$\mathbf{z} = H\mathbf{y} = H(G^T \mathbf{x} + \mathbf{n}) = H\mathbf{n} \pmod{2} \quad (105)$$

からノイズベクトル \$\mathbf{n}\$ を推定することにより行われる。例えば具体的には

$$\underbrace{\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}}_{\text{シンδροーム}} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \end{pmatrix}}_{\text{ノイズ}} \quad (106)$$

のような方程式になる．原情報 $\mathbf{x}$ は，ノイズベクトルの推定値 $\hat{\mathbf{n}}$ が得られた後，

$$G^T \mathbf{x} = \mathbf{y} - \hat{\mathbf{n}} \pmod{2} \quad (107)$$

から求められる．誤り訂正符号ではパリティ検査方程式により原情報に推定が行えることがわかった．しかし，パリティ検査方程式からのノイズ $\mathbf{n}$ の推定は一般に計算量的に困難な問題である．

そこで，ノイズの推定の計算に工夫が必要となるが，ここでベータ近似を使う．ノイズの成分とシンδροームを 2 部グラフにおける要素とクリークとみなすと

$$\begin{aligned} \psi_1(\mathbf{n}) &= \delta(z_1; n_1 + n_2 + n_3 + n_5) \\ \psi_2(\mathbf{n}) &= \delta(z_2; n_2 + n_3 + n_4 + n_6) \\ \psi_3(\mathbf{n}) &= \delta(z_3; n_1 + n_3 + n_4 + n_7) \\ \psi_i(n_i) &= p^{n_i} (1-p)^{1-n_i} \end{aligned} \quad (108)$$

を構成することができて，(85) 式に対応する分布が得られる．2 部グラフで表現することができたことで，BP で解くことが可能となる．

近年，低密度パリティ検査 (LDPC: low-density parity-check) 符号とよばれる符号族に関しては，数万程度の長いビット長に対してもベータ近似により実際の時間での復号が可能であることが認識されてきた．そして，LDPC 符号は世界最高水準の性能が得られることが実験により明らかになっている．ここでは，代表的な LDPC 符号であるギャラガー (Gallager) 符号に対して，ベータ近似による復号アルゴリズムを導出する．

ギャラガー符号は 0,1 の要素からなる $(N-K) \times K$ 疎行列 C_1 と $(N-K) \times (N-K)$ の可逆な疎行列 C_2 を横に接続し，パリティ検査行列が $H = (C_1 C_2)$ で与えられる (図 24)．疎行列とは値が 1 となる要素数の割合が少ない行列という意味である．

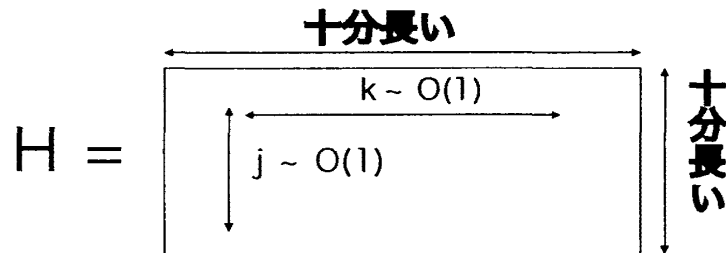


図 24: ギャラガー符号におけるパリティ検査行列

以下では， C_1, C_2 はそれぞれ各列あたり， $j(=1, 2, \dots)$ この要素だけが 1 となり，各行あたりでは 1 の個数なるべくランダムに構成された行列であるとし，ビット長 K, N は十分大きな状況を考える．以上のように構成されたパリティ検査行列 H に対しては， $K \times K$ 単位行列 I_K と $(N-K) \times K$ 行列 $C_2^{-1} C_1$ を縦に接続した $G^T = \begin{pmatrix} I_K \\ C_2^{-1} C_1 \end{pmatrix}$ によって生成行列が構成される．ただし，逆行列は $(\text{mod } 2)$ の演算に対して定義する．このように生成行列を構成すると $HG^T = 0$ が成り立つことに注意．

それでは，ギャラガー符号に対して， $\mathbf{n}$ の推定を考えよう．ここまでは， $(0,1)$ で 2 値情報を表現してきたが，便宜上 $(+1, -1)$ の 2 値で表すように変更する．同型変換 $\{0, 1, +(\text{mod } 2)\} \leftrightarrow \{+1, -1, \times\}$ を用いてすべての変数を $(+1, -1)$ によって表現しなおす．すると， $(0,1)$ 表示で各要素 $n_i = 0, 1$ となる確率が独立にそれぞれ $1-p, p$ という事前確率は， $(+1, -1)$ 表示では $F = (1/2) \ln[(1-p)/p]$ を用いて

$$P(\mathbf{n}) = \frac{\exp \left[F \sum_{l=1}^N n_l \right]}{(2 \cosh F)^N} \quad (109)$$

となる。また、ノイズ $\mathbf{n}$ に対してシンドローム $\mathbf{z}$ は確定的に与えられるのでパリティ検査行列 H に対する条件付き確率 $P(\mathbf{z}|\mathbf{n})$ は (0,1) 表示では $H\mathbf{n} = \mathbf{z} \pmod{2}$ ならば 1, それ以外では 0 となるデルタ関数 $\delta(H\mathbf{n} = \mathbf{z})$ となるが, (+1, -1) 表示では

$$P(\mathbf{z}|\mathbf{n}) = \prod_{\mu=1}^{N-K} \delta\left(z_{\mu}; \prod_{l \in \mathcal{L}(\mu)} n_l\right) = \prod_{\mu=1}^{N-K} \frac{1 + z_{\mu} \prod_{l \in \mathcal{L}(\mu)} n_l}{2} \quad (110)$$

のように表現できる。ここで、 μ はパリティ検査符号 H の行番号を、 $\mathcal{L}(\mu)$ は H の μ 番目の行に含まれる要素が 1 の列添字が作る集合をそれぞれ表す。

ベイズの公式を用いると、事後確率は

$$P(\mathbf{n}|\mathbf{z}) = \frac{P(\mathbf{z}|\mathbf{n})P(\mathbf{n})}{\sum_{\mathbf{n}} P(\mathbf{z}|\mathbf{n})P(\mathbf{n})} \quad (111)$$

となる。事後確率から様々な目的関数について最適復号戦略が決まる。ここでは、真のノイズとその推定値がビットごとに一致する確率を最大にする MPM 復号を考えることにしよう。この復号では事後分布に関する各ノイズ成分の平均値

$$\langle n_l \rangle = \sum_{\mathbf{n}} n_l P(\mathbf{n}|\mathbf{z}) \quad (112)$$

を計算し

$$\hat{n}_l = \text{sign}(\langle n_l \rangle) \quad (113)$$

を推定結果とする。

ノイズの平均値 (112) の計算は困難であるので、ベータ近似を用いて近似的に評価する。そのために、 μ 番目のパリティを構成する要素の組をクリークに、 l 番目のノイズ成分をそれぞれ対応させて、ラグランジュ未定乗数を (94),(95) 式に代入すると、

$$\psi_l(S_l) e^{-\lambda_{\mu l}(S_l)} \propto \frac{1 + m_{\mu l} n_l}{2} \quad (114)$$

$$e^{-\hat{\lambda}_{\mu l}(S_l)} \propto \frac{1 + \hat{m}_{\mu l} n_l}{2} \quad (115)$$

を導入する。今回のグラフにはループが存在するが、(114),(115) 式を (94),(95) 式に代入すると、

$$\begin{aligned} \hat{m}_{\mu \rightarrow l}^{t+1} &= z_{\mu} \prod_{j \in \mathcal{L}(\mu) \setminus l} m_{j \rightarrow \mu}^t \\ m_{l \rightarrow \mu}^t &= \tanh \left(\frac{1}{2} \ln \frac{1-p}{p} + \sum_{\nu \in \mathcal{M}(l) \setminus \mu} \tanh^{-1}(\hat{m}_{\nu \rightarrow l}^t) \right) \end{aligned} \quad (116)$$

が得られる。この反復式からベータ近似によるノイズの推定値

$$\hat{n}_l = \text{sign} \left(\frac{1}{2} \ln \frac{1-p}{p} + \sum_{\mu \in \mathcal{M}(l)} \tanh^{-1}(\hat{m}_{\mu \rightarrow l}^t) \right) \quad (117)$$

が求まる。この結果からギャラガー符号は、(loopy)BP により、高々、ビット長に比例する程度の計算量で復号できる。

ベータ近似はループのないグラフの性質を利用した近似法あり、BP が効率的解法として利用できる。直感的には疎なランダムグラフで良い近似になる。それについてはここで LDPC 符号についての例を示した。しかし、密なグラフに対しても良い近似になることがある。例えば、CDMA の復調 (Kabashima, 2003) についての議論がある。またレプリカ化による要約伝搬 (Survey Prop:SP) という方法が K-SAT 問題の効率的解探索法 (Mezard et al, 2002) で世界記録を樹立したが、この方法とベータ近似とのつながりを示す議論もある (Kabashima, 2005)。

3.2 ギブス自由エネルギー形式と適応 TAP 法

3.2.1 自由エネルギー形式

確率分布 $P(S)$ について物理量 $O(S)$ の期待値を求めるのに自由エネルギーがよく使われる。自由エネルギーにはヘルムホルツ自由エネルギー

$$\Psi(\theta) = -\ln \left[\sum_S P(S) e^{\theta \cdot O(S)} \right] \quad (118)$$

とギブス自由エネルギー

$$\begin{aligned} \Phi(m) &= \min_Q \{ KL(Q, P) | \langle O(S) \rangle_Q = m \} \\ &= \max_{\theta} \{ \theta \cdot m + \Psi(\theta) \} \end{aligned} \quad (119)$$

があり、求めたい量、あるいは目的によって使い分けられる..

自由エネルギーの 1 階微分、または自由エネルギー最小原理により、物理量を

$$\langle O(S) \rangle = - \frac{\partial \Psi(\theta)}{\partial \theta} \bigg|_{\theta \rightarrow 0} = \underset{m}{\operatorname{argmin}} \{ \Phi(m) \} \quad (120)$$

から求めることができる。

また、2 階微分から分散・共分散 ((情報論では) フィッシャー情報) 行列

$$\langle O_i(S) O_j(S) \rangle - \langle O_i(S) \rangle \langle O_j(S) \rangle = - \left(\frac{\partial^2 \Psi(\theta)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta} \right) = \left(\frac{\partial^2 \Psi(m)}{\partial m_i \partial m_j} \right)^{-1} \quad (121)$$

が得られる。以上のように物理学、情報論において、自由エネルギーから様々な量を得ることが期待できる。しかし、自由エネルギーを使うからといって計算量が減るわけではない。一般には自由エネルギーの評価でも計算は爆発する。しかし近似計算においては自由エネルギーを用いるメリットがある。それは、系統的な近似法の構成が容易だからである。先ほどは触れなかったが、ベータ近似で現れたベータ自由エネルギーもギブス自由エネルギーの近似式と解釈することができる。

ここでは、特に以下の場合を考察する。

$$P(S) \propto \exp \left[\sum_{l>k} \underbrace{J_{lk} S_l S_k}_{\text{全結合系}} \right] \prod_{l=1}^N \underbrace{(\delta(S_l - 1) + \delta(S_l + 1))}_{\text{イジングスピン}} \quad (122)$$

$$O(S) = S \quad (123)$$

相互作用がなければギブス自由エネルギーは計算が容易である。しかしながら、相互作用がある場合は何らかの工夫が必要になる。そこで、相互作用係数に展開パラメータ α を付けて、 $\alpha = 0$ の周りでテイラー展開し、 $\alpha = 0$ を代入するという、一種の高温展開を行う。

$$\Phi(m; \alpha) = \max_{\theta} \left\{ \theta \cdot m - \ln \left[\sum_S e^{\sum_{l>k} \alpha J_{lk} S_l S_k + \theta \cdot m} \right] \right\} \quad (124)$$

$$\begin{aligned} \Phi(m; \alpha = 0) &= \sum_{l=1}^N \sum_{S_l = \pm 1} \frac{1 + m_l S_l}{2} \ln \frac{1 + m_l S_l}{2} \\ \frac{\partial}{\partial \alpha} \Phi(m; \alpha = 0) &= - \sum_{l>k} J_{lk} m_l m_k \\ \frac{\partial^2}{\partial \alpha^2} \Phi(m; \alpha = 0) &= - \sum_{l>k} J_{lk}^2 (1 - m_l^2) (1 - m_k^2) \end{aligned}$$

これをブレフカ展開という。相互作用が $J_{lk} \sim N(J_0/N, J^2/N)$ の正規分布に従う場合、2 次の展開式は Thouless-Anderson-Palmar (TAP) がベータ近似に対して与えた自由エネルギー (TAP 自由エネルギー)

$$\Phi(\mathbf{m}; \alpha = 0) \simeq \sum_{l=1}^N \sum_{S_l = \pm 1} \frac{1 + m_l S_l}{2} \ln \frac{1 + m_l S_l}{2} - \sum_{l > k} J_{lk} m_l m_k - \frac{1}{2} \sum_{l > k} J_{lk}^2 (1 - m_l^2)(1 - m_k^2) \quad (125)$$

に一致する。ただし、以下の条件が満たされない限り高次項の影響は無視できずこの評価式は正しくないことが知られている。

ブレフカ条件 (収束半径の条件)

$$1 - \frac{J^2}{N} \sum_{l=1}^N (1 - m_l^2)^2 > 0 \quad (126)$$

3.2.2 適応 TAP 法

ブレフカ展開には問題もある。まず、相互作用行列 $\mathbf{J}$ は統計性は一般にはわからない。それに、一般に展開は無限級数になる。この手間は指数的に増大し、高々数次のオーダーで手計算では手に負えなくなる。なので、できればギブス自由エネルギーを非摂動的に評価する方法があるとよい。

そこで適応 TAP 法 (Oppen and Winther, 2001) というものを考える。この方法の着眼点はガウス分布を利用するということにある。ガウス分布は多次元でも計算が容易である。また、今考えている全結合系では中心極限定理から他の影響はガウス分布でよく近似できると期待される。

適応 TAP 自由エネルギーはイジングスピンをガウススピんに換えてもよく近似されるという仮定の下で、拡張された自由エネルギーに関する 1 階微分の関係式

$$\frac{\partial}{\partial \alpha} \Phi(\mathbf{m}; \alpha) = - \left\langle \sum_{l > k} J_{lk} m_l m_k \right\rangle_{\alpha} \simeq \frac{\partial}{\partial \alpha} \Phi_{\text{Gauss}}(\mathbf{m}; \alpha) \quad (127)$$

を 0 から 1 まで積分することで、次のように構成される。

$$\Phi_{\text{ada}}(\mathbf{m}) \simeq \Phi_{\text{Gauss}}(\mathbf{m}; \alpha = 1) - \Phi_{\text{Gauss}}(\mathbf{m}; \alpha = 0) + \Phi(\mathbf{m}; \alpha = 0) \quad (128)$$

ただし、 Φ_{Gauss} の計算ではイジングスピンを表す重み $[\delta(S-1) + \delta(S+1)]$ の代わりにガウス重み

$$P_0(S_l) \propto \exp \left\{ -\frac{\Lambda_l}{2} S_l^2 \right\} \quad \Lambda_l : \quad \text{s.t.} \quad \langle S_l^2 \rangle_{\alpha} = 1 \quad (129)$$

を用いて分配関数の計算を行う。ガウススピンに対する自由エネルギーは

$$\begin{aligned} \Phi_{\text{Gauss}}(\mathbf{m}; \alpha) &= \max_{\boldsymbol{\theta}, \boldsymbol{\Lambda}} \left\{ -\sum_{l=1}^N \frac{\Lambda_l}{2} + \boldsymbol{\theta} \cdot \mathbf{m} \right. \\ &\quad \left. - \ln \left(\int d\mathbf{S} \exp \left[-\sum_{l=1}^N \frac{\Lambda_l S_l^2}{2} + \alpha \sum_{l > k} J_{lk} S_l S_k + \sum_{l=1}^N \theta_l S_l \right] \right) \right\} \\ &= \max_{\boldsymbol{\Lambda}} \left\{ -\sum_{l=1}^N \frac{\Lambda_l}{2} + \frac{1}{2} \mathbf{m}^T (\boldsymbol{\Lambda} - \alpha \mathbf{J}) \mathbf{m} + \frac{1}{2} \ln (\boldsymbol{\Lambda} - \alpha \mathbf{J}) \right\} \end{aligned} \quad (130)$$

となり、 $\boldsymbol{\Lambda}$ を決める条件は

$$1 = (\boldsymbol{\Lambda} - \alpha \mathbf{J})_{ll}^{-1} + m_l^2 \quad (\langle S_l^2 \rangle_{\alpha} = 1) \quad (131)$$

である。よって適応 TAP 自由エネルギーは

$$\begin{aligned}\Phi_{\text{ada}}(m) = & \overbrace{\max_{\Lambda} \left\{ -\sum_{l=1}^N \frac{\Lambda_l}{2} + \frac{1}{2} m^T (\Lambda - J) m + \frac{1}{2} \ln (\Lambda - J) \right\}}^{\Phi_{\text{Gauss}}(\alpha=1)} \\ & + \sum_{l=1}^N \frac{1}{2} \ln(1 - m_l^2) \quad \Leftarrow \quad -\Phi_{\text{Gauss}}(\alpha=0) \\ & + \underbrace{\sum_{l=1}^N \sum_{S_l \pm 1} \frac{1 + m_l S_l}{2} \ln \frac{1 + m_l S_l}{2}}_{\Phi_{\text{Ising}}(\alpha=0)}\end{aligned}\quad (132)$$

となる。極値条件より適応 TAP 方程式が導かれ、オーダーパラメータが

$$\begin{cases} m_l = \tanh \left[\sum_{k \neq l} J_{lk} m_k - \underbrace{\left(\Lambda_l - \frac{1}{1 - m_l^2} \right)}_{\text{オンサーガ反跳場項}} m_l \right] \\ \underbrace{(\Lambda - J)_{ll}^{-1} = 1 - m_l^2}_{\text{オンサーガ反跳場項を適応的に決める条件}} \end{cases} \quad (133)$$

から得られる。

ここで適応 TAP 法の性質について、いくつか述べておく。適応 TAP 法を用いる際には相互作用行列 J の統計性に関する情報は必要ない。この性質のため、生命情報などのデータ解析への応用に大きな期待を持たれている。

J の統計性を与える場合は、例えば次のように既存結果を再現することができる。

$$\Lambda_l - \frac{1}{1 - m_l^2} = \begin{cases} \beta^2 J^2 (1 - Q) & (\text{SK model}) \\ \frac{\alpha \beta^2 (1 - Q)}{1 - \beta(1 - Q)} & (\text{Hopfield model}) \end{cases} \quad \left(Q = \frac{1}{N} \sum_{l=1}^N m_l^2 \right) \quad (134)$$

ただ、計算コストは $O(N^3)$ となり、効率的な解法をみつけることは未解決の問題である。

また、次のようにパーセプトロン型の問題にも応用可能であり、判別分析への導入も容易である。

$$\begin{aligned}P(S) & \propto \prod_{\mu=1}^M f(\mathbf{x}_{\mu} \cdot \mathbf{S}) = \int d\mathbf{u} d\hat{\mathbf{u}} \prod_{\mu=1}^M f(u_{\mu}) e^{\hat{\mathbf{u}}_{\mu} (\mathbf{x}_{\mu} \cdot \mathbf{S} - u_{\mu})} \\ & = \int d\hat{\mathbf{u}} e^{\overbrace{\hat{\mathbf{u}}^T \mathbf{x} \mathbf{S}}^{2 \text{ 体相互作用}}} \underbrace{\prod_{\mu=1}^M \int du_{\mu} f(u_{\mu}) e^{-\hat{\mathbf{u}}_{\mu} u_{\mu}}}_{\hat{\mathbf{u}} \text{ の事前分布}}\end{aligned}\quad (135)$$

この性質を利用して、適応 TAP 法をソナー信号処理で爆弾か石かを判別する問題 (ニューラルネットの有名なベンチマーク問題でもある) に応用する例もある。(Gorman, R.P. and Sejnowski, T.J., 1988).

4 まとめ

これまでの 10 年間でモノ (物質) の理論をコト (情報) の研究に生かすことが盛んに行われてきた。元々、物理学と情報論はエントロピーの概念を共有しており、最近の研究でモノとコトの類似性を強調してきたとも言える。これからは、モノとコトの既存理論の相違点を意識し、双方の得意・不得意を認識することが重要になる。

また、これまで、統計力学で研究されてきたトイモデルを用い、無限系で対称性の高い場合のモノの理論を第ゼロ近似としてコトの研究に応用することで、テリトリを確保してきた。これからの1つの方向性として、有限次元で対称性の低い、より実践的な多変量データ解析の理論の研究が挙げられる。また、量子情報における多体問題の検討など、さらに新しい領域での研究も進むだろうと思われる。

参考文献

- [1] R. Monasson, et. al., “Determining computational complexity from characteristic ‘phase transition’”, *Nature*, vol.400, pp.133-137, 1999.
- [2] T.Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors”, *IEEE, trans. Inform. Theory*, vol.48, no.11, pp.2888-2910, 2002.
- [3] 汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功:「統計科学のフロンティア / 計算統計 I – 確率計算の新しい手法」, 岩波書店 (2003).
- [4] 田中和之・編:「確率的情報処理と統計力学—様々なアプローチとそのチュートリアル」, 臨時別冊・数理科学, サイエンス社 (2006)
- [5] 西森秀稔:「スピングラス理論と情報統計力学」, 新物理学選書, 岩波書店 (1999).
- [6] B. Derrida, “Random-energy model: An exactly solvable model of disordered systems”, *Phys. Rev. B*, vol.24, no.5, pp.2613-2626, 1981.
- [7] R.G. Gallager, “Low density parity check codes”, *IRE Trans. Inform. Theory*, vol.IT-8, pp.21-28, 1962.
- [8] Y. Kabashima, “A CDMA multiuser detection algorithm on the basis of belief propagation”, *J. Phys. A: Math. Gen.*, vol.36, no.43, pp.11111-11121, 2003.
- [9] M. Mezard, et. al., “Analytical and Algorithmic Solution of Random Satisfiability Problems”, *Science*, vol.297, no.5582, pp.812-815, 2002.
- [10] Y. Kabashima, “Replicated Bethe Free Energy: A Variational Principle behind Survey Propagation”, *J. Phys. Soc. Jpn.*, vol.74, pp.2133-2136.
- [11] D.J. Thouless, P.W. Anderson, R.G. Palmer, “Solution of ‘Solvable model of a spin glass’”, *Phil. Mag.*, vol.35, pp.593-601, 1977.
- [12] T. Plefka, “Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model”, *J. Phys. A: Math. Gen.*, vol.15, pp.1971-1978, 1982.
- [13] M. Oppen and O. Winther, “Tractable Approximation for Probabilistic Models; The Adaptive Thouless-Anderson-Palmer Mean Field Approach”, *Phys. Rev. Lett.*, vol.86, pp.3695-3699, 2001.
- [14] R.P. Gorman and T.J. Sejnowski, “Learned classification of sonar targets using a massively parallel network”, *Neural Networks*, vol.1, pp.75-89, 1988.