構造の観測における物理学と数学

渡辺澄夫
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

統計的推測の問題を考えるときに現れる物理学と数学について紹介する。統計的推測の問題は、ランダム・ハミルトニアンによって定義されるカノニカル分布のもとでマクロな変量の挙動を考える問題と等価である。特に、観測データの背後にある構造を推測する場合には、基底状態が特異点を持つ解析的集合であるカノニカル分布を考察する問題になり、マクロな変量が従う法則を導出する際に代数幾何などの数学的手段が有効である。統計力学と類似する理論の構築により、統計的推測においても普遍性を持つ公式が存在すること、および、その公式は双有理不変量を与えていることが明らかになる。

1　はじめに

この文章は2011年の第56回物性若手夏の学校のテキストです。物理学教室で学んでいる学生のかたが理解しやすいように書かれています。幾つかの数学的な構造について説明していますが、数学的な厳密性についてはこだわらないことにします。

統計力学において、ハミルトニアン $H(w)$ が与えられたとき、カノニカル分布

$$e^{-\beta H(w)}$$

の挙動を考えることは、$H(w)$ によって定まる平衡状態を考えることに相当しています。この文章では、統計的推測の問題が、

$$H(w)=\sum_{i=1}^{n} f(X_i,w) \tag{1}$$

である場合の平衡状態を調べることに相当することを説明し、その場合のカノニカル分布の持つ性質を調べます。ここで $X_1,X_2,...,X_n$ は $\mathbb{R}^N$ に値をとる確率変数で、同じ確率分布 $q(x)$ に独立に従うものとします。また $f(x,w)$ は、$(x,w)\in\mathbb{R}^N\times\mathbb{R}^d$ から実数への与えられた関数です。確率変数の個数 n が多くなっていく極限でカノニカル分布 $\exp(-\beta H(w))$ はどのような挙動を持つでしょうか。また、そこでは、どのような物理学的な法則が見つかるでしょうか。その法則の発見の基礎には、どのような数学があるでしょうか。

注意. $X_1,X_2,...,X_n$ が確率変数なので、$\exp(-\beta H(w))$ は確率過程です。このように確率的に変動するハミルトニアンは、ランダム・ハミルトニアンと呼ばれています。不純物を含んでいる金属の物性を考察するとき、不純物がランダムに混入している様子をモデル化すると、ランダム・ハミルトニアンが現れてくることが知られています。

2　なぜ統計的推測にカノニカル分布が現れるか

まず、なぜ、このような問題を考えるのかについて例を用いて説明します。

2.1 統計的推測とは

$\mathbb{R}^N$ 上に定義された確率分布 $q(x)$ に独立に従う n 個の確率変数の集合を

$$X^n = \{X_1, X_2, ..., X_n\}$$

と表記することにします。「確率変数 X^n から $q(x)$ を推測すること」を統計的推測と言います。推測される $q(x)$ のことを **真の分布** と呼びます。実用においては、真の分布は不明であって、これをデータ X^n から推測します。理論を考えるときには、とにかく、ある $q(x)$ が存在しているものと考えます。これを情報源と呼ぶ場合もあります。

注意. X が確率分布 $q(x)$ に従うとき、集合 $A \subset \mathbb{R}^N$ について X が A の中に値を取る確率 $\mathrm{Prob}(X \in A)$ は

$$\mathrm{Prob}(X \in A) = \int_A q(x)dx$$

になります。ここで $q(x)$ は確率密度関数ですが、物理学の通例に従って、この文章でも確率密度関数のことを確率分布と呼ぶことにしました。

真の確率分布を推測するために確率モデルを使う場合を考えましょう。パラメータ $w \in W \subset \mathbb{R}^d$ によって定まる $x \in \mathbb{R}^N$ の確率分布 $p(x|w)$ が与えられているとします。この $p(x|w)$ のことを**確率モデル** と言います。実問題では、真の分布 $q(x)$ は不明であることが多いので、確率モデル $p(x|w)$ を用いてこれを推測します。

確率モデルは、あくまでもモデルにすぎません。それは真の分布に対して適切である場合もあれば、そうでない場合もあります。どのくらい適切であるかを考えるための基礎になる理論を作ることがこの文章の目的です。「モデル→理論→理論と実験の比較→モデルの評価」のサイクルが可能になるようにすることが目標です。

真の分布 $q(x)$ と 確率モデル $p(x|w)$ の関係を二つ定義します。
(1) あるパラメータ $w_0 \in W$ が存在して、$q(x) = p(x|w_0)$ とできるとき、$q(x)$ は $p(x|w)$ により**実現可能**であるといいます。
(2) $q(x)$ から $p(x|w)$ までの相対エントロピー[1]あるいはカルバック・ライブラ距離

$$S(w) = \int q(x) \log \frac{q(x)}{p(x|w)} dx \tag{2}$$

を最小にするパラメータ $w_0 \in W$ がひとつだけ存在して、ヘッセ行列 $(\nabla\nabla S)(w_0)$ が正則行列（固有値が全て正値の行列）であるとき、$q(x)$ は $p(x|w)$ に対して**正則**であるといいます。

例. 具体的な例を紹介します。
(例 1) コインを振って表が出たとき $X = 1$, 裏が出たとき $X = 0$ とします。確率モデルを $W = \{0 < a < 1\}$ として

$$p(x|a) = a^x(1-a)^{1-x}$$

と定めます。この問題では、コインに特別な仕掛けがない限り、$q(x)$ は $p(x|a)$ で実現可能かつ正則であると考えて良いものと思われます（ただし、表または裏が確率 1 で出る場合を除いています)。
(例 2) 二つの物理量 $X = (X_1, X_2)$ について観測を行って得た結果に最小二乗法で直線を当てはめるという操作を、$W = \{(a, b) \in \mathbb{R}^2\}$ として

$$X_2 = aX_1 + b + \text{正規雑音}$$

[1] この概念の発見者はボルツマンです。常に $S(w) \geq 0$ です。$S(w) = 0$ と $q(x) = p(x|w)$ は等価です。

という確率モデルを用いていると考えることにします。この場合、X_1 と X_2 とが、この確率分布に従っているとは思えないことが多いので、$q(x)$ は $p(x|w)$ に対して正則ですが実現可能ではありません。

(例3) 空間の幾つかの場所 $\mathbf{r}$ で電位 $D(\mathbf{r})$ を測定して、その結果から、複数の点電荷の大きさ a_h と位置 $\mathbf{r}_h$ を推定するという問題を考えます。確率モデルとして $W=\{(a_h,\mathbf{r}_h);h=1,2,...,H\}$ として

$$D(\mathbf{r})=\sum_{h=1}^{H}\frac{a_h}{|\mathbf{r}-\mathbf{r}_h|}+\text{正規雑音}$$

を考えたとします。もしも確率モデルの電荷の個数が現実に配置されている個数よりも多い場合には、$q(x)$ は $p(x|w)$ により実現可能であるが正則ではありません。

(例4) 例えば、生徒100人に量子力学のテストを行うと、その成績は正規分布のように山がひとつではなく、二つあるいは三つの山があるように見える場合があります。そこで、平均が b_h で分散が σ_h^2 の正規分布 H 個を和にしたもの、すなわち $W=\{(a_h,b_h);h=1,2,...,H,\sum a_h=1\}$ で

$$p(x|w)=\sum_{h=1}^{H}\frac{a_h}{\sqrt{2\pi\sigma_h^2}}\exp\Big(-\frac{1}{2\sigma_h^h}(x-b_h)^2\Big)$$

を考えましょう。このような問題では、もはや、真の分布は確率モデルによって実現可能か正則か、わかりません。

通常の確率統計の教科書に書かれているのは、上記の例1に相当するケース、すなわち実現可能かつ正則な場合です。物理学を学んだ人が統計学を現実の問題に適用するとき、「なんだか怪しい感じがする」のは、実現可能かつ正則な場合の理論が、そうでない場合にも使われているからではないでしょうか。現実の問題では、例2、例3、例4の場合だと考えた方が妥当であることが多く、そのような場合の理論は従来の確率統計学では作られておりませんでした。特に、例3、例4のように「データから真の分布の構造を推測する」という問題では、普通の統計学の教科書に書かれていることは成り立ちません。例3、例4のような場合に、統計学の教科書に書かれている検定や推定の方法を使うと、誤った結論が導かれることがあるので注意が必要です。

2.2 統計的推測におけるカノニカル分布

確率モデルを用いて、真の分布を推測する方法を説明します。パラメータの集合 $W\subset\mathbb{R}^d$ 上に確率分布 $\varphi(w)$ を用意します。これを **事前分布** といいます。事前分布の選び方は後で述べますので、最初は、とりあえず、ある確率分布が与えられたと考えてください。

次に **事後分布** を

$$p(w|X^n)=\frac{1}{Z}\varphi(w)\prod_{i=1}^{n}p(X_i|w)^\beta$$

と定義します。定数 Z は $p(w|X^n)$ の w についての積分が1になるための定数であり **分配関数** です。$\beta=1$ のとき **ベイズ推測** といいます。$\beta=\infty$ のとき **最尤推測** といいます。真の分布が確率モデルに対して正則でない場合には、n が大きくなっても事後分布は正規分布に近づきません。この場合には、最尤推測は非常に精度が悪くなることがわかっています[2]。真の分布が確率モデルで実現可能であってもなくても、正則であってもなくても、推測ができる方法を考えたい場合には、最尤推測は適していません。そこで以下ではベイズ推測の場合、つまり $\beta=1$ の場合を考えます。（この文章

[2]このとき、最尤推定量は、一致性も、漸近正規性も、漸近有効性も持ちません。存在しないこともあります。

で述べることは一般の β でも同様の導出ができます)。パラメータ上の平均操作を

$$\mathbb{E}_w[\ ] = \frac{\displaystyle\int (\ )\varphi(w)\prod_{i=1}^{n} p(X_i|w)dw}{\displaystyle\int \varphi(w)\prod_{i=1}^{n} p(X_i|w)dw} \tag{3}$$

と定義します。$\mathbb{E}_w[\ ]$ は平均操作を表していますが、X^n が確率変数なので、こうして計算される平均値は確率変数であることに注意してください。

この平均操作を用いて確率モデルを平均したもの

$$\hat{p}(x) \equiv \mathbb{E}_w[p(x|w)]$$

のことを **予測分布** といいます。これは、サンプル X^n をもとにして、真の分布 $q(x)$ を推測したものです。予測分布 $\hat{p}(x)$ は、真の分布 $q(x)$ の良好な近似になっていることが望まれますが、その保証は、どこにもありません。そこで、統計的推測においては次のことが研究の目標になります。
(1) 予測分布は、どれくらい真の分布に近いでしょうか。
(2) 真の分布が不明な場合でも、予測分布がどのくらい適切か知ることはできるでしょうか。
(3) 予測分布ができるだけ真の分布をよく近似するためには、どのような確率モデルと事前分布を用いたらよいでしょうか。

さて、式 (2) で定義される相対エントロピー $S(w)$ の最小値を与えるパラメータのひとつを w_0 とします。関数 $f(x,w)$ を

$$f(x,w) = \log\frac{p(x|w_0)}{p(x|w)} \tag{4}$$

とおくと、

$$S(w) = S(w_0) + \int q(x)f(x,w)dx$$

より

$$(\forall w \in W) \quad K(w) \equiv \int q(x)f(x,w)dx \geq 0 \tag{5}$$

が成り立ちます。また定義式 (4) から

$$p(x|w) = p(x|w_0)\exp(-f(x,w))$$

と書けるので、$\mathbb{E}_w[\ ]$ を書き換えることができて

$$\mathbb{E}_w[\ ] = \frac{\displaystyle\int (\ )\varphi(w)\exp\Big(-\sum_{i=1}^{n} f(X_i,w)\Big)dw}{\displaystyle\int \varphi(w)\exp\Big(-\sum_{i=1}^{n} f(X_i,w)\Big)dw} \tag{6}$$

になります。これは、式 (1) で表されるランダム・ハミルトニアン $H(w)$ によって定まるカノニカル分布による平均に他なりません。統計的推測の問題は、このカノニカル分布に関する統計力学を考えることに相当していることがわかりました。

さて、サンプル X^n に関する平均を $\mathbb{E}[\ ]$ で表すことにすると、式 (5) より

$$(\forall w \in W) \quad \mathbb{E}[H(w)] = nK(w) \geq 0$$

が成り立ちます。$H(w)$ はランダムな関数ですから、$H(w)$ の値は確率的には負になることもありますが、平均の関数 $nK(w)$ は非負です。また平均の関数の基底状態に相当する集合

$$W_0 \equiv \{w \in W \, ; \, K(w) = 0\}$$

は、真の分布が確率モデルに対して正則でない場合には、一般には一点ではなく、広がりを持った集合になっています。

2.3 マクロな変量

これで平衡状態の特徴づけができましたので、次に、マクロな変量について説明します。

汎化損失 G_n と **経験損失** T_n を次のように定義します。

$$\begin{aligned} G_n &= -\int q(x) \log \mathbb{E}_w[p(x|w)]dx, \\ T_n &= -\frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}_w[p(X_i|w)]. \end{aligned}$$

これらの変量は、カノニカル分布 $\mathbb{E}_w[\ ]$ による平均を用いて定義されています。すなわち、これらは統計力学でいうところのマクロな変量です。また、これらは、サンプル X^n の関数ですから確率変数です。これらの変量を考える理由について説明しましょう。

汎化損失の意味. 予測分布は $\hat{p}(x) = \mathbb{E}_w[p(x|w)]$ ですから、汎化損失は、

$$G_n = -\int q(x) \log q(x) dx + \int q(x) \log \frac{q(x)}{\hat{p}(x)} dx$$

と書くことができます。右辺の第一項は、真の分布のエントロピーですから、確率モデルや事前分布には依存していません。第2項は、真の分布 $q(x)$ から予測分布 $\hat{p}(x)$ までの相対エントロピーです。もしも、運よく $q(x) = \hat{p}(x)$ であれば第2項は0になります。そうでないときには正の値になります。そこで、G_n が小さい値であればあるほど、予測分布 $\hat{p}(x)$ は良い推測であると考えることができます。統計的推測あるいは情報理論の目標は G_n が小さくなるように確率モデルや事前分布を定めることであると考えて良いと思います。なお、G_n は真の分布のエントロピーよりも小さな値になることはありません。

経験損失の意味. 現実の問題で汎化損失が計算できれば良いのですが、汎化損失は人間にとっては未知のものである真の分布を用いた平均なので、この値を直接に求めることはできません。そこで $\int (\ ) q(x) dx$ をサンプル X^n の和で置き換えたものが経験損失です。これは、真の分布を知らなくても人間が計算することができます。しかしながら、X^n は $\mathbb{E}_w[\ ]$ を定義する場合にも使われているので、経験損失の値は汎化誤差と同じではありません。すなわち

$$\begin{aligned} G_n &= \text{「真の分布を評価に使った場合の値」} \\ T_n &= \text{「データを評価に使った場合の値」} \end{aligned}$$

になっています。データだけから計算できる経験損失 T_n から、汎化損失 G_n を計算する方法を作ることが、この文章の目的です。

3 カノニカル分布からマクロな変量を求める公式

統計的推測の問題をまとめると、次のようになります。

目標. ランダム・ハミルトニアン $H(w)=\sum f(X_i,w)$ から定まるカノニカル分布による平均 $\mathbb{E}_w[\ ]$ の性質を解明し、マクロな変量 G_n と T_n の間に成り立つ普遍的な関係式を導出せよ。

問題が定式化されましたので、後は数学の問題になります。ランダム・ハミルトニアンで定義された統計力学の問題では、しばしばレプリカ法が有効であることが知られています。しかしながら、真の分布が確率モデルに対して正則でないときには、レプリカ法を用いても解析を行うことはできません。また従来の統計学や情報理論の方法でも、この問題を考えることはできませんでした。一般に、物理学の根源にある問題は数学と深い結びつきを有していることが多いのですが、この問題もまた数学と関係を持っているのでしょうか。

3.1 相対的に有限な分散

まず、目標を考えるための条件について述べます。

定義. 相対エントロピー $S(w)$ の最小値を与えるパラメータの部分集合を W_0 とします。これは $K(w)$ の零点全体の集合です。

$$W_0=\{w\in W\ ;\ K(w)=0\}.$$

あるパラメータ $w_0\in W_0$ が存在して、条件

$$(\exists c_1>0,\forall w\in W)\quad \int q(x)f(x,w)dx\geq c_1\int q(x)f(x,w)^2dx \tag{7}$$

が成り立つとき、ランダム・ハミルトニアン $H(w)$ は**相対的に有限な分散を持つ** といいます。

次のことは定義からすぐに示すことができます。
(1) $q(x)$ が $p(x|w)$ により実現可能であるとき、$H(w)$ は相対的に有限な分散を持ちます。
(2) $q(x)$ が $p(x|w)$ に対して正則であるとき、$H(w)$ は相対的に有限な分散を持ちます。
(3) $H(w)$ が相対的に有限な分散を持つとき、W_0 の任意の元 w に対して $p(x|w)$ は同じ確率分布を与えます。

以下、この文章では、ハミルトニアンが相対的に有限な分散を持つ場合を考察します。この条件が成立しないときには、統計的推測の構造が大きく変化することがあります。

3.2 ミクロとマクロをつなぐ公式

カノニカル分布 $\mathbb{E}_w[\ ]$ から、マクロな変量 G_n,T_n を計算する公式を作っておきましょう。

ハミルトニアンが相対的に有限な分散を持つとします。w_0 を W_0 に含まれるパラメータとして

$$\begin{aligned}L_0&=-\int q(x)\log p(x|w_0)dx,\\ L_n&=-\frac{1}{n}\sum_{i=1}^{n}\log p(X_i|w_0),\end{aligned}$$

と表記します。$\mathbb{E}[L_n] = L_0$ が成り立ちます。関数 $f(x,w)$ を用いると、汎化損失、経験損失は次のようになります。

$$\begin{aligned} G_n &= L_0 - \int q(x) \log \mathbb{E}_w[e^{-f(x,w)}]dx, \\ T_n &= L_n - \frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}_w[e^{-f(X_i,w)}]. \end{aligned}$$

これらの計算のために、二つの母関数を定義します（キュムラント母関数）。

$$\begin{aligned} \mathcal{G}_n(\alpha) &= -\int q(x) \log \mathbb{E}_w[e^{-\alpha f(x,w)}]dx, \\ \mathcal{T}_n(\alpha) &= -\frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}_w[e^{-\alpha f(X_i,w)}], \end{aligned}$$

すると $\mathcal{G}_n(0) = \mathcal{T}_n(0) = 0$ であり、三次以上の項は無視できる [8] ので

$$G_n = L_0 + \mathcal{G}_n(1) \cong L_0 + \mathcal{G}_n'(0) + \frac{1}{2}\mathcal{G}_n''(0), \tag{8}$$

$$T_n = L_n + \mathcal{T}_n(1) \cong L_n + \mathcal{T}_n'(0) + \frac{1}{2}\mathcal{T}_n''(0), \tag{9}$$

が成り立ちます。統計力学における自由エネルギーの計算と同様にして

$$\mathcal{G}_n'(0) = \int q(x)\mathbb{E}_w[f(x,w)]dx = \mathbb{E}_w[K(w)], \tag{10}$$

$$\mathcal{G}_n''(0) = -\int q(x)\Big\{\mathbb{E}_w[f(x,w)^2] - \mathbb{E}_w[f(x,w)]^2\Big\}dx, \tag{11}$$

$$\mathcal{T}_n'(0) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_w[f(X_i,w)] = \frac{1}{n}\mathbb{E}_w\Big[H(w)\Big], \tag{12}$$

$$\mathcal{T}_n''(0) = -\frac{1}{n}\sum_{i=1}^{n}\Big\{\mathbb{E}_w[f(X_i,w)^2] - \mathbb{E}_w[f(X_i,w)]^2\Big\}, \tag{13}$$

が導出されます。以上のことから次のことがわかりました。カノニカル分布による平均 $\mathbb{E}_w[\ ]$ を用いて、式 (10),(11),(12),(13) で表される 4 つの量の値を求めた後、式 (8),(9) に代入すれば、目標の変量 G_n, T_n の挙動がわかります。

注意. カノニカル分布の式 (6) と $\mathcal{T}_n(\alpha), \mathcal{G}_n(\alpha)$ の定義式をよく見ると、次の等式が成り立つことがわかります。

$$\mathbb{E}[\mathcal{G}_{n-1}(1)] = -\mathbb{E}[\mathcal{T}_n(-1)].$$

これから

$$\mathbb{E}[\mathcal{G}_{n-1}'(0) + \frac{1}{2}\mathcal{G}_{n-1}''(0)] \cong \mathbb{E}[\mathcal{T}_n'(0) - \frac{1}{2}\mathcal{T}_n''(0)] \tag{14}$$

が成り立ちます。

4 数学的基礎

ここでは、必要になる数学的基礎を説明します。しばらく、もとの問題から離れます。

4.1 二つの困難な点

まず、何が数学的な困難であるかを説明します。

関数 $K(w)$ が解析関数であるとします。集合 $W_0=\{w;K(w)=0\}$ は、解析関数の零点全体の集合です。解析関数の零点全体で表される集合は **解析的集合** と呼ばれ、一般には複雑な特異点を含んでいます。特に $K(w)$ が多項式の場合には、この集合は**代数的集合** あるいは **代数多様体** と呼ばれています。基底状態の縮退が、このような集合であるようなハミルトニアンを考察するときには二つの問題があります。

ハミルトニアン $H(w)$ をその平均の関数 $nK(w)$ とゆらぎを表す関数で書くと

$$\begin{aligned} H(w) &= \sum_{i=1}^{n} f(X_i,w) \\ &= nK(w)-\sqrt{nK(w)}\times\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\Big(\frac{K(w)-f(X_i,w)}{\sqrt{K(w)}}\Big)\Big] \end{aligned} \tag{15}$$

となります。

(1) 一般に、解析関数 $K(w)$ は特異点の付近で極めて多様な形状を持つので、$\exp(-nK(w))$ を一般的に取扱うのは容易ではありません。

(2) 上記の式 (15) で [ ] で囲まれた部分は、ゆらぎを表す関数ですが、これは ill-defined です。この関数は、独立な確率変数の n 個の和を $1/\sqrt{n}$ 倍しているので、中心極限定理を適用することができますが、w が W_0 の上にあるとき定義できない関数になります。

4.2 特異点解消定理

上記の問題に基本的な解決を与える数学的基礎として特異点解消定理を説明します。

集合 $W\subset\mathbb{R}^d$ 上に非負値の解析関数 $K(w)$ が定義されているとします。うまく、d 次元多様体 $\mathcal{M}$、および $\mathcal{M}$ から W への解析写像 $w=g(u)$ とを見つけることにより、合成関数 $K(g(u))$ とヤコビアン $|g'(u)|$ を扱い易い形にしたいと考えます。

注意. 一変数の関数 $f(x)$ が解析関数であるとします。もしも

$$\frac{f(x)}{x-a}$$

が $x=a$ の近傍で有界なら、実は $x=a$ の近傍で解析関数です。(つまり、$x=a$ は除去可能な点です)。これは、一変数では因数定理「$f(a)=0$ なら $f(x)$ は $(x-a)$ で割り切れる」が成り立つからです。しかしながら、多変数の多項式は、そのような性質はありません。例えば

$$\frac{(x+y)^2}{x^2+y^2}$$

は原点の近傍で有界ですが、この関数は原点の近傍で解析関数ではありません。原点は除去可能ではありません。しかしながら、多変数関数であっても、特別な形をしたもの

$$K(x,y)=x^3y^5\times(0\text{ にならない関数}) \tag{16}$$

であるときには、

$$\frac{f(x,y)}{K(x,y)}$$

がある開集合の中で有界ならその中で解析関数になる、という性質を持っています。すなわち、関数が式 (16) のように、変数毎の積で書けているときには、多変数関数であっても、一変数と同じで因

数定理が成り立ちます。これは非常に扱い易い性質です。式 (16) のように、多変数の関数が変数毎の積で書けるとき、**正規交差** であるといいます。変数変換によって、与えられた関数を正規交差にする方法を次の例で考えてみましょう。

例. $\mathbb{R}^2$ 上で定義された $K(x,y)=x^2+y^2$ の場合を考えます。$K(x,y)=0$ で定義される集合は原点だけです。しかしながら、関数

$$\psi=\frac{(x+y)^2}{K(x,y)}$$

は、原点の近傍で有界ですが well-defined ではありません。そこで二つの座標 $U_1=\{(x_1,y_1)\}$ と $U_2=\{(x_2,y_2)\}$ を用意して、それぞれの座標から $\mathbb{R}^2$ へ

$$x=x_1=x_2y_2$$
$$y=x_1y_1=y_2$$

という写像 g を考えましょう。和集合 $U_1\cup U_2$ を考えて「写像 g により行き先が同じになる」という同値関係 $\sim$ を導入し、商集合

$$\mathcal{M}=U_1\cup U_2/_{\sim}$$

を考えると $\mathcal{M}$ は多様体になります。二つの座標 U_1 と U_2 は $\mathcal{M}$ の局所座標になります。同値関係 $\sim$ は、行き先が同じになることで定義されていましたから、写像 g は $\mathcal{M}$ から $\mathbb{R}^2$ への写像として well-defined になります。この操作を「座標を貼りあわせる」といいます。それぞれの座標の上で

$$K(g(x_1,y_1))=x_1^2(1+y_1^2),$$
$$K(g(x_2,y_2))=y_2^2(x_2^2+1),$$

となります。これは正規交差な関数です。このとき、関数 ψ は

$$\psi=\frac{(1+y_1)^2}{1+y_1^2}=\frac{(x_2+1)^2}{x_2^2+1}$$

ですから、$\mathcal{M}$ 上の解析関数です。なお、ヤコビアン $|g'(u)|$ は

$$|g'(u)|=|x_1|=|y_2|$$

になります。

例. もうひとつ例をあげます。

$$K(a,b,c)=(ab+c)^2+a^2b^4$$

を考えましょう。このとき、4つの座標

$$U_i=\{(a_i,b_i,c_i)\}\quad (i=1,2,3,4)$$

を考えて

$$\begin{aligned}
a&=a_1c_1 & b&=b_1 & c&=c_1,\\
a&=a_2 & b&=b_2c_2 & c&=a_2(1-b_2)c_2,\\
a&=a_3 & b&=b_3 & c&=a_3b_3(b_3c_3-1),\\
a&=a_4 & b&=b_4c_4 & c&=a_4b_4c_4(c_4-1).
\end{aligned}$$

という写像を作ります。4つの座標は、行き先が同じになる元を同一視することで貼りあわせることができて多様体になります。またそれぞれの座標の上で

$$\begin{aligned} K &= c_1^2\{(a_1b_1+1)^2+a_1^2b_1^4\} \\ &= a_2^2c_2^2(1+b_2^2c_2^2) \\ &= a_3^2b_3^4(c_3^2+1) \\ &= a_4^2b_4^2c_4^4(1+b_4^2) \end{aligned}$$

となります。つまり、K は正規交差です。ヤコビアンは

$$\begin{aligned} |g'(u)| &= |c_1| \\ &= |a_2c_2| \\ &= |a_3b_3^2| \\ &= |a_4b_4c_4|^2 \end{aligned}$$

になっています。

上の二つは特殊な例のように見えるかもしれませんが、実は、どんな解析関数でも、これと同じことができるという定理があります。

特異点解消定理 (広中の定理)([5][2]). $K(w)$ を開集合 $W \subset \mathbb{R}^d$ 上の非負値の解析関数とします。このとき、ある d 次元多様体 $\mathcal{M}$ と解析写像 $g:\mathcal{M}\to W$ が存在して、局所座標ごとに

$$K(g(u)) = u_1^{2k_1}u_2^{2k_2}\cdots u_d^{2k_d}, \tag{17}$$

$$|g'(u)| = b(u)|u_1^{h_1}u_2^{h_2}\cdots u_d^{h_d}|, \tag{18}$$

が成り立ちます。ここで $b(u)>0$ は零にならない解析関数です。また

$$\begin{aligned} k &= (k_1,k_2,...,k_d), \\ h &= (h_1,h_2,...,h_d), \end{aligned}$$

は、非負の整数の集合です（一般に 0 になるものも含みます)。このような k, h のことを多重指数といいます。

この定理は広中平祐氏によって 1964 年に証明された代数幾何の基本定理です。関数 $K(w)$ に対して $w=g(u)$ を見つける方法も作られています。式 (17)(18) を、それぞれ

$$\begin{aligned} K(g(u)) &= u^{2k}, \\ |g'(u))| &= b(u)|u^h|, \end{aligned}$$

と略記します。

さて、この定理を適用すると

$$K(g(u)) = u^{2k}$$

とできることがわかります。ハミルトニアンが相対的に有限な分散を持つことを仮定しているので、式 (7) から

$$(u^k)^2 \geq c_1 \int q(x)f(x,g(u))^2dx$$

が成り立ちます。これより、$f(x,g(u))$ が u^k で割り切れることがわかります。すなわちある well-defined な関数 $a(x,u)$ が存在して

$$f(x,g(u)) = u^k a(x,u)$$

となります。定義である $K(g(u)) = \int f(x,g(u))q(x)dx$ を用いると $\int a(x,u)q(x)dx = u^k$ であることがわかります。関数 $\xi_n(u)$ を

$$\xi_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{u^k - a(X_i,u)\} \tag{19}$$

とおくと、ハミルトニアンは

$$H(w) = nu^{2k} - \sqrt{n}u^k\xi_n(u) \tag{20}$$

と書けることがわかりました。これをランダム・ハミルトニアンの **標準形** と呼びます。また

$$\varphi(w)dw = |u^h|b(u)du$$

となります。

4.3 超関数の漸近展開

次に、$u \in \mathbb{R}^d$ についての超関数でパラメータ $t \in \mathbb{R}$ を持つもの、すなわち、**状態密度関数**

$$\Delta(t,u) = \delta(t-u^{2k})u^h$$

の $t \to 0$ における挙動を考えます。超関数を考えるときには、u の取りうる値の範囲について局所的な集合の上で考えることができます。(超関数が作用する関数を多様体上の「1の分割」を用いて局所的なものの和で書くことができるからです)。そこで、特に $u \in [0,1]^d$ の場合を考えても一般性を失いません。

メリン変換を

$$(Mf)(z) \;=\; \int_0^\infty t^z f(t)dt \quad (z \in \mathbb{C})$$

と定義します。メリン変換はフーリエ変換やラプラス変換と同じように逆変換を持ちますので、対応 $f \leftrightarrow Mf$ は、実質的に一対一と考えることができます。

例. 関数

$$f(t) = \begin{cases} t^{\lambda-1}(-\log t)^{m-1} & (0<t<1) \\ 0 & (\text{上記以外}) \end{cases}$$

のメリン変換は具体的に計算することで

$$(Mf)(z) = \frac{(m-1)!}{(z+\lambda)^m}$$

であることがわかります。

超関数 $\Delta(t,u)$ を t についてメリン変換すると

$$(M\Delta)(z,u) = u^{2kz+h}$$

なります。この関数の超関数としての性質を調べるために任意の関数 $\Phi(u)$ に作用させたもの (**ゼータ関数**)

$$\zeta(z)=\int_{[0,1]^d} u^{2kz+h}\Phi(u)du$$

を考えてみましょう。$\Phi(u)$ の原点の周りでのテーラー展開を好きなオーダーまで考えると

$$\zeta(z)=\frac{\Phi(0)}{\prod_{j=1}^{d}(k_j z+h_j+1)}+\cdots+ \tag{21}$$

となることがわかります。この展開は、Φ のテーラー展開の次数をあげることで、どこまでも行うことができます。つまりゼータ関数は、有理型関数として複素平面全体に一意に解析接続できます。特に d 個の値

$$\frac{h_j+1}{k_j} \quad (j=1,2,..,d)$$

の最小値を λ として、最小値を取る j の回数を m とします。すると、式 (21) の極はすべて負の有理数ですが、最も原点に近いのは $z=-\lambda$ で、その位数は m になります。また座標 $u=(u_1,u_2,...,u_d)$ の中で最小値を取る j に対応するもの $u_a\in\mathbb{R}^m$ とそうでないもの $u_b\in\mathbb{R}^{d-m}$ を使って $u=(u_a,u_b)$ と書くことにすると、

$$u^{2kz+h}\propto\frac{\delta(u_a)u_b^{\mu}}{(z+\lambda)^m}+\cdots+$$

となることがわかります。ここで $\mu=\{\mu_j\}$ は $d-m$ 次元の多重指数で、$\mu_j>-1$ を満たしているので有限な積分値を持ちます。この逆メリン変換を考えることにより

$$\Delta(t,u)=(c_2\delta(u_a)u_b^{\mu})\,t^{\lambda-1}(-\log t)^{m-1}+\cdots+$$

であることがわかりました。ここで $c_2>0$ は定数であり、「$\cdots$」 は $t\to 0$ で主要項よりも早く 0 に収束します（漸近展開)。最後に積分要素を

$$du^*\equiv c_2\delta(u_a)u_b^{\mu}b(u)du$$

とおくと、$t\to 0$ での挙動は

$$\Delta(t,u)du\cong t^{\lambda-1}(-\log t)^{m-1}du^* \tag{22}$$

となります。ゼータ関数の極に関する情報から状態密度関数の挙動が得られました。

注意. 定数 λ は、特異点解消によって得られる多重指数 k と h から定まる定数（有理数）です。一般に与えられた関数 $K(w)$ に対して、その特異点解消は無限にあります。ユニークではありません。多重指数 k,h についてもユニークではありません。しかしながら、λ はどのような特異点解消を考えても同じ値になります。特異点解消に依存しない値のことを **双有理不変量** といいます。この定数 λ は **実対数閾値** と呼ばれる量です。このような超関数の漸近展開の方法は I.M.Gelfand によります [4]。特異点解消定理を用いて任意の解析関数で Gelfand の方法が使えることを示したのは M.F.Atiyah です [2]。特異点解消定理を用いずに w における超関数の展開を考えることもできます [3]。この問題は **代数解析学** を発展させるひとつの動機となりました [6]。統計的推測の問題において λ の値は特異点の解消を計算することで求めることができます [8]。

4.4 経験過程

以上でランダム・ハミルトニアンに関する積分計算のための数学的準備ができました。次にハミルトニアンのゆらぎを扱うための数学的基礎を説明します。

式 (20) の標準形において現れる関数 $\xi_n(u)$ の挙動について考えます。式 (19) において $s(x,u) = u^k - a(x,u)$ と置くと

$$\xi_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, u) \tag{23}$$

です。これは u の関数ですが、まず u を止めて、各 u について考えて見ましょう。一般に、$\{Y_i\}$ が平均 0、分散 σ^2 を持つ同じ確率分布 $q(y)$ に独立に従うとき、

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$$

の確率分布は、平均が 0 で分散が σ^2 の正規分布に収束することが知られています。ここで「確率分布が収束する」ということは、実数上に定義された任意の有界連続関数 $F(\ )$ について

$$\lim_{n\to\infty}\int F\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} y_i\Big)\prod_{i=1}^{n} q(y_i)dy_i = \int F(y)\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{y^2}{2\sigma^2})dy$$

が成り立つということです。これを **中心極限定理** といいます。従って、$\xi_n(u)$ の確率分布は各点 u ごとに正規分布に近づいて行きます。

それでは、$\xi_n(u)$ は、関数としても、上記と同じような性質を持つでしょうか。つまり、ある関数の空間から実数への有界な連続汎関数 $F(\ )$ が与えられたとき、$\xi_n(u)$ と同じ分散共分散関数を持つ正規確率過程 $\xi(u)$ が存在して

$$\lim_{n\to\infty}\int F\Big(\xi_n\Big)\prod_{i=1}^{n} q(x_i)dx_i = \mathbb{E}_{\xi}[F(\xi)]$$

が成り立つでしょうか。ここで $\mathbb{E}_{\xi}[\ ]$ は、正規確率過程 ξ に関する平均を表しています。

この問題に関する数学的な基礎に関心があるかたは、例えば [7] をご覧ください。具体的には、u の集合がコンパクトで、$s(x,u)$ が u 上の解析関数であり、$\int s(x,u)^{2+\delta}q(x)dx < \infty$ $(\delta > 0)$ が成り立つときには、関数空間として sup ノルムで定義されるバナッハ空間を採用することで、収束先の確率過程 ξ はユニークであり、上記の収束が成り立つことを示すことができます。さらに

$$\sup_u\Big|\frac{1}{n}\sum_{i=1}^{n} a(X_i,u) - \int a(x,u)q(x)dx\Big| \to 0, \tag{24}$$

$$\mathbb{E}[\sup_u|\xi_n(u)|^2] \to \mathbb{E}_{\xi}[\sup_u|\xi(u)|^2], \tag{25}$$

などが成り立ちます。このような性質を持つ $\xi_n(u)$ のことを**経験過程** といいます。これは、関数空間上の中心極限定理に相当します。なお、$F(\ )$ として有界関数でなくても、「漸近一様可積分」という性質を持つものに対しては上記の収束を保証することができます。

5 統計的推測の理論

以上で数学的な準備ができました。当初の目標であったマクロな変量の挙動を導出しましょう。

5.1 分配関数

式 (3) で定義されたカノニカル分布を考察するために、ボルツマン因子に関する微小積分

$$\Omega(w)dw \equiv \exp(-H(w))\varphi(w)dw$$

の挙動を考えましょう。これは分配関数の計算を行うことと実質的に同じです。まず、$K(w)$ の特異点を正規交差にする写像 $w=g(u)$ を用いると、ヤコビアンを $|g'(u)|$ として

$$\Omega(w)dw = \exp(-H(g(u)))\varphi(u)|g'(u)|du$$

となります。次に、ハミルトニアンの標準形 (式 (20)) を用いると

$$\Omega(w)dw = \exp(-nu^{2k} + \sqrt{n}u^k\xi_n(u))\ u^h b(u)\ du$$

です。デルタ関数の性質から

$$\Omega(w)dw = \int_0^\infty dt\ \delta(t-u^{2k})u^h\ \exp(-nt+\sqrt{nt}\ \xi_n(u))\ b(u)\ du$$

と書き直すことができます。変数の変換 $t := t/n$ を行ってスケーリングを変えると

$$\Omega(w)dw = \int_0^\infty \frac{dt}{n}\ \delta\Big(\frac{t}{n}-u^{2k}\Big)u^h\ \exp(-t+\sqrt{t}\ \xi_n(u))\ b(u)\ du$$

ここで式 (22) を用いると、主要項は

$$\Omega(w)dw \cong \frac{(\log n)^{m-1}}{n^\lambda}\int_0^\infty dt\ t^{\lambda-1}\ \exp(-t+\sqrt{t}\ \xi_n(u))\ du^* \tag{26}$$

になります。ここで積分要素 du^* のサポートは集合 $g^{-1}(W_0)=\{u;K(g(u))=0\}$ に含まれることに注意してください。これで分配関数の計算ができました。

5.2 マクロな変量の挙動

$n\to\infty$ の極限状態での平均操作 $\langle\ \rangle$ をある関数 $F(t,u)$ について

$$\langle F(t,u)\rangle = \frac{\int du^* \int_0^\infty dt\ F(t,u)\ t^{\lambda-1}\ \exp(-t+\sqrt{t}\ \xi_n(u))}{\int du^* \int_0^\infty dt\ t^{\lambda-1}\ \exp(-t+\sqrt{t}\ \xi_n(u))} \tag{27}$$

と定義します。ここで $\int du^*$ は各局所座標の積分の和です。式 (26) と、スケーリングの関係

$$f(x,g(u)) = u^k a(x,u) = \sqrt{\frac{t}{n}}\ a(x,u)$$

を用いて $\mathbb{E}_w[\ ]$ による平均の極限を $\langle\ \rangle$ の平均で書きなおすことができます。これより、$n\to\infty$ において、非負の整数 j について

$$\mathbb{E}_w[f(x,u)^j] \cong \frac{1}{n^{j/2}}\langle\ (\sqrt{t}\ a(x,u))^j\ \rangle$$

が成立します。また式 (27) で $F(t,u)=t$ のとき、dt についての部分積分を行うことにより、関係

$$\langle t\rangle = \lambda + \frac{1}{2}\langle\sqrt{t}\xi_n(u)\rangle$$

が得られます。以上の二つのことを使うと、$\mathcal{G}_n'(0)$, $\mathcal{G}_n''(0)$, $\mathcal{T}_n'(0)$, $\mathcal{T}_n''(0)$ を計算することができます。

$$\begin{aligned}
\mathcal{G}_n'(0) &\cong \frac{1}{n}\Big(\lambda+\frac{1}{2}\langle\sqrt{t}\xi_n(u)\rangle\Big),\\
\mathcal{G}_n''(0) &\cong \frac{1}{n}\int\{\langle ta(x,u)^2\rangle - \langle\sqrt{t}a(x,u)\rangle^2\}q(x)dx,\\
\mathcal{T}_n'(0) &\cong \frac{1}{n}\Big(\lambda-\langle\frac{1}{2}\sqrt{t}\xi_n(u)\rangle\Big),\\
\mathcal{T}_n''(0) &\cong \frac{1}{n^2}\Big\{\sum_{i=1}^n\langle ta(X_i,u)^2\rangle - \langle\sqrt{t}a(X_i,u)\rangle^2\Big\}.
\end{aligned}$$

ここで、式 (24) において $s(x,u)=a(x,u)^2$ の場合を考えれば、$n\mathcal{G}_n''(0)-n\mathcal{T}_n''(0)$ は $n\to\infty$ のとき、0 に収束することが示せます。そこで $n\mathcal{G}_n''(0)$ を $V(\xi_n)$ と書くと、

$$\begin{aligned} G_n &\cong L_0+\frac{1}{n}\Big(\lambda+\frac{1}{2}\langle\sqrt{t}\xi_n(u)\rangle-\frac{1}{2}V(\xi_n)\Big),\\ T_n &\cong L_n+\frac{1}{n}\Big(\lambda-\frac{1}{2}\langle\sqrt{t}\xi_n(u)\rangle-\frac{1}{2}V(\xi_n)\Big).\end{aligned}$$

一方、式 (14) から

$$\mathbb{E}_\xi[\langle\sqrt{t}\xi(u)\rangle]=\mathbb{E}_\xi[V(\xi)]$$

が成り立つことが得られます。この定数を 2ν とおいて **特異ゆらぎ** と呼ぶことにします。以上のことから、マクロな変量の平均値は、実対数閾値 λ と 特異ゆらぎ ν を用いて

$$\begin{aligned}\mathbb{E}[G_n] &\cong L_0+\frac{\lambda}{n},\\ \mathbb{E}[T_n] &\cong L_0+\frac{\lambda-2\nu}{n},\end{aligned}$$

と表されることがわかりました。さらに

$$V_n=\sum_{i=1}^{n}\Big\{\mathbb{E}_w[(\log p(X_i|w)^2]-\mathbb{E}_w[\log p(X_i|w)]^2\Big\}$$

とおくと $\mathbb{E}[V_n]\to 2\nu$ が成り立ちますので、統計的推測の **状態方程式**

$$\mathbb{E}[G_n]=\mathbb{E}\Big[T_n+\frac{V_n}{n}\Big]+o(\frac{1}{n})$$

が成り立つことがわかりました。変量 WAIC を

$$\mathrm{WAIC}=T_n+\frac{V_n}{n}$$

と定義すると、$\mathbb{E}[G_n]=\mathbb{E}[\mathrm{WAIC}]+o(1/n)$ です。ここで $o(1/n)$ は $1/n$ よりも早く 0 に収束する項です[3]。変量 WAIC は、確率モデルとデータだけから計算することができますので、状態方程式によって、真の分布を知らなくても汎化損失を求めることができます。これは統計的推測において、確率モデルや事前分布の良さが評価できることを意味しています。状態方程式は平均値について成り立つ関係ですが、確率変数としては、

$$\Big(G_n-L_0\Big)+\Big(\mathrm{WAIC}-L_n\Big)=\frac{2\lambda}{n}+o(\frac{1}{n})$$

という性質を持っています。つまり、(G_n-L_0) と $(\mathrm{WAIC}-L_n)$ の分散は同じです。

注意. 特異ゆらぎ ν も特異点の解消に依存しないので双有理不変量です。この値は数学でも調べられていないようです。真の分布が確率モデルによって実現可能であり、かつ、正則であるときには、$\lambda=\nu=d/2$ が成り立ちます。ここで d はパラメータ空間の次元です。このとき、WAIC は、統計学における赤池情報量規準 [1] と一致します。サンプルから構造を推測する問題では、真の分布は、確率モデルで実現可能であるとは限らず、また正則とも限らないのですが、そのような場合でも成立するように一般化することができました。これを広く使える情報量規準 (Widely Applicable Information Criterion) と呼んでいます。

[3]実は、$\mathbb{E}[G_n]$ と $\mathbb{E}[\mathrm{WAIC}]$ の差は $1/n^2$ のオーダーです。

注意. 真の分布がわからないとき、確率モデルと事前分布をどのように決めたらよいのでしょうか。何らかの事情によって確率モデルと事前分布が分かっている場合には、それを使うことになります。そうでない場合には、それはモデルの設計の問題になります。統計学においては **自由エネルギー**

$$F = -\log \int \prod_{i=1}^{n} p(X_i|w)\varphi(w)dw$$

を最小にするという方法と、平均汎化誤差を最小にするという方法がよく使われるものです。自由エネルギーの漸近挙動は、式 (26) からすぐに導出することができて

$$F = nL_n + \lambda \log n - (m-1)\log\log n + (\text{定数オーダーの確率変数})$$

となります。自由エネルギーは、実質的な第一項である $\lambda \log n$ が確率的には揺れていないので、自由エネルギーを規準にするとデータのゆらぎに影響されにくい設計を行うことができます。しかしながら、自由エネルギーの最小化は汎化誤差の最小化と等価ではありません。平均汎化誤差を最小にする規準は汎化誤差そのものを見る方法ですが、実質的な第一項が確率的に揺れているので、データの揺らぎに依存する設計を行うことになります。統計学として、どちらを使うべきかという問題は、正則な場合に限定しても多くの論議があるところです。現実の問題では、F と WAIC の両方とも計算してみて、真の分布に対する確率モデルや事前分布の適切さを両者の角度から検討してみるというのが良いのではないかと思われます。

注意. ここで考えている問題では、カノニカル分布は有限次元空間の上の確率分布ですが、データの個数が無限大になって行く極限を考えているため、ハミルトニアンや事前分布に制御変数が入っているときには **相転移** が生じることがあります。相転移とは、ある点の前後でカノニカル分布の実質的なサポートが大きく変わり、その結果として自由エネルギーや汎化誤差の連続性や微分可能性などが失われることを言います。無限次元空間でないと相転移が起きないということはありません。相転移点は理論的に解明することができますが、それは計算してみて初めてわかる点であり、直観的に自明な点ではありません。相転移点の性質を調べることは、確率モデルや事前分布を設計する場合に、重要な知見を与えてくれます。また、現実の問題に確率モデルや事前分布を用いるとき、相転移点の付近では挙動が不安定になることを知っておくべきだと思います。

注意. 自然科学においては、カノニカル分布の実現は自然が行ってくれています。統計的推測の問題では、カノニカル分布は計算機で実現する他ありません。多くの場合に **マルコフ連鎖モンテカルロ法** (MCMC 法) が利用されています。20 世紀に作られた全てのアルゴリズムの中で、MCMC 法が最も重要なものであると認識されています。メトロポリス法、ハミルトニアン・モンテカルロ法、交換モンテカルロ法などの MCMC 法の主要なアルゴリズムを発明したのは、すべて物理学者でした。これらのアルゴリズムが作られた動機が「もっと正確に物理現象を見たかったから」であり「応用に役立つから」ではなかったことは大変に重要なことです。また、平均場近似によってカノニカル分布を近似する方法についてもよく研究されています。平均場が満たすべき拘束条件を再帰的代入として用いることにより数値的に平均場近似を求めることができます。しかしながら、カノニカル分布を平均場近似したとき、汎化損失や経験損失がどのようになるかについて、また、状態方程式がどのようになるかについては、まだ解明されておりません。相転移がある問題では、その位置がずれることが知られています。

注意. この文章は数学的な厳密性にこだわらないという方針で書かれています。証明が必要なかたは [8] をご覧ください。また、この文章では、ハミルトニアンが相対的に有限な分散を持つ場合を考えてきました。真の分布が確率モデルで実現可能でなく、かつ、正則でもない場合の中には、相対的に有限な分散を持たない場合も起こりえます。このとき、$(G_n - L_0)$, $(T_n - L_n)$ の漸近挙動が $1/n$ のオーダーではなく、別のオーダー、例えば $1/n^{2/3}$ など、に変わります。ただし、その場合でも状態方程式は成立しています [9]。

6 終わりに

統計的推測において、「観測から構造を知る」という問題がランダム・ハミルトニアンによって定まるカノニカル分布の挙動を考える問題になっていることを説明しました。その理論的な解析のために必要な数学的手段を紹介し、統計的推測においても普遍性を持つ関係式が導出できることを示しました。統計的推測の問題が物理学と似ている構造を持つのは、単なる偶然なのでしょうか。それとも、観測から真の分布を推測する、ということは物理学の一部なのでしょうか。

謝辞. 2011年第56回物性若手夏の学校の事務局の皆さまに感謝いたします。特に、この発表についての運営をご担当頂いた京都大学数理解析研究所の岡村和弥さんに感謝いたします。

参考文献

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, Vol.19, pp.716-723, 1974.

[2] M.F. Atiyah. Resolution of singularities and division of distributions. *Communications of Pure and Applied Mathematics*, Vol.13, pp.145-150. 1970.

[3] I.N. Bernstein. The analytic continuation of generalized functions with respect to a parameter. *Functional Analysis and Applications*, Vol.6, pp.26-40, 1972.

[4] I.M. Gelfand and G.E. Shilov. *Generalized Functions*. Academic Press, San Diego, 1964.

[5] H. Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. *Annals of Mathematics*, Vol.79, pp.109-326, 1964.

[6] M. Kashiwara. B-functions and holonomic systems. *Inventiones Mathematicae*, Vol. 38, pp.33-53, 1976.

[7] A. W. van der Vaart, J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer,1996.

[8] S. Watanabe. Algebraic geometry and statistical learning theory. Cambridge University Press, Cambridge, UK, 2009.

[9] S. Watanabe. Asymptotic learning curve and renormalizable condition in statistical learning theory. *Journal of Physics Coneference Series*, 2010, July. Vol.233, No.1, 012014. doi: 10.1088/1742-6596/233/1/012014.