

論 文

α -分布族に対する平均場近似と変分法

豊泉 太郎[†] 合原 一幸[†]

Mean-field and Variational Methods for α -families

Taro TOYOIZUMI[†] and Kazuyuki AIHARA[†]

あらし

近年 Tsallis の非加法統計, 発達した乱流, 経済などの分野でベキ型の確率分布が注目されている. 本研究では真の分布と近似する分布間の α -ダイバージェンスに基づく "距離" を極小化することによって, ベキ型の分布の統計量を近似的に求める手法を示す. さらに, Tsallis の非加法統計力学の枠組から導出される自由エネルギーの上界, 下界が α -ダイバージェンスを用いて評価できることを示す.

キーワード α -ダイバージェンス, Tsallis 統計, 情報幾何, 平均場近似, Bogolyubov 不等式

1. ま え が き

確率的に振舞う素子からなる系において, それぞれの素子が互いに複雑に相互作用している場合, その平均的な振舞いを厳密に求めることは難しい. そこで他の素子からの相互作用を平均的な場によって置き換えるナイーブ平均場近似, さらにその場の揺らぎも考慮に入れた TAP 近似などが解析に用いられる [2]. これらの手法は物理学のみならず学習理論 [3], 誤り訂正符号 [4] 等にも広く応用されている手法である.

近年, 相互作用項の小さい条件でこれらの平均場近似が確率分布間の "距離" の一つである KL-ダイバージェンスを最小化することによっても導出されることが示され [5, 7], 平均場近似の幾何的な描像が明らかになった. また甘利らは, より一般的な確率分布間の "距離" として α -ダイバージェンスを導入することで, ナイーブ平均場近似は 1-ダイバージェンスの, TAP 近似は相互作用項の小さい場合の近似として -1 -ダイバージェンスの最小化に対応していることを示し, これらの手法の違いが確率分布間の "距離" の定義の違いに起因するものとして解釈できることを明らかにするとともに, 一般の α -ダイバージェンスの最小化に対応する平均場近似の可能性を指摘した [1]. しかし現在の

ところ指数型分布族を考える限り, 一般の α -ダイバージェンスの最小化は KL ダイバージェンスの最小化同様, 近似的な摂動計算に頼らなくてはならない.

一方で, 近年物理学の分野においては従来の Boltzmann-Gibbs 統計の枠組みを拡張しようという動きがある. 即ち, 従来の Boltzmann-Gibbs による加法的なエントロピーに代えて, より一般的な非加法的エントロピーに基づく統計力学を再構成しようという試みが盛んに行われている [10, 13–15, 17, 18]. そしてその体系において, 従来の指数型分布に代わる重要な役割を担っているのがベキ型の分布である.

本研究では, 指数型分布族を形式的に含む形で拡張したより一般的なベキ型の分布族に対して, その分布の指数の値に対応する特定の α -ダイバージェンスの最小化が容易に実行できることを説明するとともに, それを用いてベキ型の分布の統計量を近似的に求める手法を示す. そしてこの手法によって得られる解を数値計算と比較することで結果の妥当性を議論する. また α -ダイバージェンスの非負性より Tsallis 型の自由エネルギーの上界または下界が評価できることを示した. この手法は Tsallis 統計に拡張した場合の Bogolyubov の不等式 [14] の一般化であることがわかる.

2. では上に述べた Tsallis の非加法統計力学の枠組を概観する. 3. では α -ダイバージェンスのもつ性質と情報幾何の枠組を使ってそのダイバージェンスから導出される幾何 [8] について触れるとともに, α -射影を用いた平均場近似 [1] について説明する. 4. では α -分

[†] 東京大学大学院新領域創成科学研究科複雑理工学専攻, 東京都 Department of Complexity Science and Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

布のある部分多様体への α -射影を求めることで、もとの α -分布の統計量を近似的に計算する手法を示す。5. では α -ダイバージェンスの非負性にもとづいて導出される自由エネルギーの上界を評価する。6. では本研究の結果をまとめる。

2. Tsallis の非加法統計力学と α -分布族

近年、長距離相互作用をもつハミルトン系において、系が準定常状態に留まる時間が系の微視的自由度に比例して長くなる場合があるということが報告されている [12]。通常、平衡統計力学の枠組では微視的自由度 $N \rightarrow \infty$ 、時間 $t \rightarrow \infty$ の極限を考えることが多いが、上記の例からわかるように、この二つの極限操作は一般に可換でない。つまり上記の例において $N \rightarrow \infty$ の系で考えるならばエルゴードの仮定は成り立っていないということになる。このような系の性質を記述する有力な候補として、現在、Tsallis の非加法統計力学が注目を集めている [11]。

Tsallis は、マルチフラクタル系で基本となるようなスケールされた量 (確率分布のベキ乗) を含む形でエントロピーの定義を

$$S_q = k \frac{\sum_{\mathbf{x}} p^q(\mathbf{x}) - 1}{1 - q} \quad (1)$$

と変更した [10, 13]。以降では $k = 1$ とし、 \mathbf{x} は状態変数である。ここでパラメーター q を用いて指数関数と対数関数を一般化しておくくと便利である。即ち、

$$e_q^x \equiv [1 + (1 - q)x]^{1/(1-q)} \xrightarrow{q \rightarrow 1} e^x, \quad (2)$$

$$\log_q(x) \equiv \frac{x^{1-q} - 1}{1 - q} \xrightarrow{q \rightarrow 1} \log(x), \quad (3)$$

である。これを用いて (1) を書き直すと $S_q = -\sum_{\mathbf{x}} p^q(\mathbf{x}) \log_q p(\mathbf{x})$ となり、 $q \rightarrow 1$ の極限でこれが Gibbs-Shannon エントロピーと一致するのは明らかである。

$\langle \mathcal{H} \rangle_q \equiv \sum_{\mathbf{x}} \mathcal{H}(\mathbf{x}) p^q(\mathbf{x})$ が一定のもとでの Tsallis エントロピーの最大化から、カノニカル分布が、

$$p(\mathbf{x}) = \frac{e_q^{-\beta \mathcal{H}(\mathbf{x})}}{Z} \quad (4)$$

と求まる。ここで Z は規格化定数、また $\beta \equiv 1/T$ であり $[1 - (1 - q)\beta \mathcal{H}(\mathbf{x})] \leq 0$ の場合は $p(\mathbf{x}) = 0$ とする。Tsallis 統計における自由エネルギーは、

$$F \equiv \langle \mathcal{H} \rangle_q - TS_q = -\frac{1}{\beta} \log_q Z \quad (5)$$

と書ける。

いま \mathcal{H} がパラメーター θ に関して線形であるとするれば (4) は情報幾何でいう α 分布族 [8]、

$$p(\mathbf{x}) = \left[1 + \frac{1 - \alpha}{2} \theta^\mu F_\mu(\mathbf{x}) \right]^{\frac{2}{1 - \alpha}} e^{-\psi(\theta)} \quad (6)$$

と同等である。ここで $\psi(\theta) = \log Z$ であり、 $\theta^\mu F_\mu = \sum_{\mu=0}^M \theta^\mu F_\mu$ と Einstein の記法を採用している。実際に流体の異常拡散、発達した乱流、経済システムなどにおいてはこのような形の分布が表れることが分かっている [6, 13]。

3. α -ダイバージェンスと情報幾何

統計学および情報理論では確率分布間の擬距離としてダイバージェンスを用いる。中でも KL-ダイバージェンスはよく用いられるが、ここではより一般的な Csiszár の f -ダイバージェンス [19] について説明する。ここで $f(u)$ は $u > 0$ 上で凹な連続関数とする。また $f(1) = 0$ とする。ここで任意の確率分布 p_1, p_2 間の f -ダイバージェンスは、

$$D_f(p_1 || p_2) \equiv \sum_{\mathbf{x}} p_1(\mathbf{x}) f\left(\frac{p_2(\mathbf{x})}{p_1(\mathbf{x})}\right) \quad (7)$$

と定義される。このとき関数 f の凹性から $D_f(p_1 || p_2) \geq 0$ であり、等号が $p_1 = p_2$ の場合のみ成り立つことが確かめられる。情報幾何において重要となる α -ダイバージェンスは (7) において $f(u) = \frac{4}{1 - \alpha^2} (1 - u^{\frac{1 + \alpha}{2}})$ としたものに他ならない。ここで $\lim_{\alpha \rightarrow -1} D_\alpha(p_1 || p_2) = KL(p_1 || p_2)$ となって、 α -ダイバージェンスは $\alpha \rightarrow -1$ の極限で KL-ダイバージェンスと一致する。

いま確率分布全体の成す集合 \mathcal{S} を多様体と考え、その中の部分多様体を $\mathcal{M} (\subset \mathcal{S})$ とする。真の確率分布 $p \in \mathcal{S}$ が与えられた時、ナイーブな平均場近似とは情報幾何的な観点から見ると、 \mathcal{M} として各状態変数について因数分解可能な確率分布族からなる e -自己平行な多様体を考え、 $D_1(p || p_0)$ を最小化する $p_0 \in \mathcal{M}$ の期待値パラメーターをもって p の期待値パラメーターの近似値とする手法と理解できる [1]。即ち p から部分多様体 \mathcal{M} への 1-射影がナイーブ平均場近似に対応する。このとき 1-射影は計算量的には容易に求まるもののその幾何構造は一般に複雑であり、1-射影の解 ($D_1(p || p_0)$ の極小解) は一般には複数存在する。一方で p から \mathcal{M} への -1 -射影 ($D_{-1}(p || p_0)$ の極小解) は

唯一に定まり幾何構造は単純であるが計算量的には困難であるというトレードオフがある。また p の -1 -射影の期待値パラメータは p の期待値パラメータと一致する [1]。

4. α -ダイバージェンスと平均場近似

さて、ここでは (6) の形の分布 p を期待値の計算が容易な分布族のなす部分多様体 \mathcal{M} への α' -射影、即ち、

$$\arg \min_{p_0 \in \mathcal{M}} D_{\alpha'}(p||p_0) \quad (8)$$

で近似する手法を示す。問題設定は以下の通りである。 $p \in \mathcal{S}$ として、

$$p(\mathbf{x}; \boldsymbol{\theta}) = e^{-\psi(\boldsymbol{\theta})} \rho(\mathbf{x}; \boldsymbol{\theta}), \quad (9)$$

$$\rho(\mathbf{x}; \boldsymbol{\theta}) = \left[1 + \frac{1-\alpha}{2} \theta^\mu F_\mu(\mathbf{x}) \right]^{\frac{2}{1-\alpha}}, \quad (10)$$

というベキ型の分布を考える。いま、この分布 p に関する期待値パラメータ

$$\eta_i = \sum_{\mathbf{x}} p(\mathbf{x}) F_i(\mathbf{x}), \quad (i = 0, 1, \dots, m) \quad (11)$$

の値を求めたい。また、特に Tsallis 統計の表記に従って q -期待値、 $\langle \langle F_i \rangle \rangle_q \equiv \frac{\sum_{\mathbf{x}} F_i(\mathbf{x}) p^q(\mathbf{x})}{\sum_{\mathbf{x}} p^q(\mathbf{x})}$ が知りたい場合は以降の式において p を $\frac{p^q(\mathbf{x})}{\sum_{\mathbf{x}} p^q(\mathbf{x})}$ と置き換えればよい^(注1)。例えば $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ のそれぞれの変数 x_i が k 個の離散状態をとる場合、期待値パラメータ η_i の値を求めるのに必要な計算量は一般に $O(k^N)$ となって N の増加に伴って指数的に増える。 x_i が連続値をとる場合についても事情は同様である。この為 N が大きい場合の期待値パラメータの厳密計算は一般には困難となる。

このような問題点を解決する為に (10) を $p_0 \in \mathcal{M}$ で近似することを考える。ここで $\mathcal{M} \subset \mathcal{S}$ 上では期待値パラメータ $\eta_{0,\mu} \equiv \sum_{\mathbf{x}} p_0(\mathbf{x}) F_\mu(\mathbf{x})$, $\mu = 0, 1, \dots, M (> m)$, の値は容易に計算できるものと仮定する。逆に言えば、 p_0 についての期待値パラメータが容易に計算できるように、うまく部分多様体 \mathcal{M} を選んで、そこに p を射影することを考える。

さて、ここで確率分布 p と p_0 の間の α' -ダイバー

ジェンスは以下のように書ける。

$$D_{\alpha'}(p||p_0) = \frac{4}{1-\alpha'^2} \left[1 - \sum_{\mathbf{x}} p^{\frac{1-\alpha'}{2}} p_0^{\frac{1+\alpha'}{2}} \right]. \quad (12)$$

ここで (12) を $G_{\alpha'} \equiv -\frac{4}{1-\alpha'^2} \log \sum_{\mathbf{x}} p^{\frac{1-\alpha'}{2}} p_0^{\frac{1+\alpha'}{2}}$ を用いて書き換えると、

$$D_{\alpha'}(p||p_0) = \frac{4}{1-\alpha'^2} \left[1 - e^{-\frac{1-\alpha'}{2} (\psi + \frac{1+\alpha'}{2} G_{\alpha'})} \right] \quad (13)$$

となる。ここで定義した $G_{\alpha'}(p||p_0)$ は、Rényi の G -ダイバージェンス [9] と定数差、定数倍を除いて等価である。このとき (13) から

$$\arg \min_{p_0 \in \mathcal{M}} D_{\alpha'}(p||p_0) = \arg \min_{p_0 \in \mathcal{M}} G_{\alpha'}(p||p_0) \quad (14)$$

となることがわかる。(14) より $G_{\alpha'}(p||p_0)$ を極小化することで p の \mathcal{M} への α' -射影としての p_0 が得られる。

真の確率分布 p を \mathcal{S} の中のどのような部分多様体に射影するかについて、ここでは特に \mathcal{M} として e -自己平行な部分多様体を選ぶことにする。したがって、

$$p_0(\mathbf{x}; \mathbf{h}) = \exp(h^i F_i(\mathbf{x}) - \psi_0(\mathbf{h})), \quad (15)$$

であり、 $\psi_0(\mathbf{h}) = \log \sum_{\mathbf{x}} \exp(h^i F_i(\mathbf{x}))$ とする。このとき (13) の $G_{\alpha'}$ は、

$$G_{\alpha'} = -\frac{4}{1-\alpha'^2} \left[-\frac{1+\alpha'}{2} \psi_0(\mathbf{h}) + \psi_0\left(\frac{1+\alpha'}{2} \mathbf{h}\right) + \log \langle \rho^{\frac{1-\alpha'}{2}} \rangle^{(\alpha')} \right], \quad (16)$$

のように書き直すことができる。ただし、

$$\begin{aligned} \langle \cdot \rangle^{(\alpha')} &\equiv \frac{\sum_{\mathbf{x}} \cdot p_0^{\frac{1+\alpha'}{2}}(\mathbf{x}; \mathbf{h})}{\sum_{\mathbf{x}} p_0^{\frac{1+\alpha'}{2}}(\mathbf{x}; \mathbf{h})} \\ &= \sum_{\mathbf{x}} \cdot p_0\left(\mathbf{x}; \frac{1+\alpha'}{2} \mathbf{h}\right), \end{aligned} \quad (17)$$

である。一般に任意の実数 α' について $G_{\alpha'}$ を極小化する α' -射影を考えることができるが、特に $\alpha' = \alpha$ と選んだ場合は $G_{\alpha'}$ の極小化が容易に実行できる。なぜならこの場合、(16) の第3項において、 $\langle \rho^{\frac{1-\alpha}{2}} \rangle^{(\alpha)} = 1 + \frac{1-\alpha}{2} \theta^\mu \langle F_\mu \rangle^{(\alpha)}$ となり、 $\{F_\mu(\mathbf{x})\}$ の非線型項の平均がでてこないからである。

(注1) : α' -射影を実際に計算するとき、 $\alpha' = \alpha$ の場合は $D_{\alpha'}(p||p_0)$ の極小化が容易に実行できる。 q -期待値を近似的に求めたい場合、即ち $D_{\alpha'}\left(\frac{p^q(\mathbf{x})}{\sum_{\mathbf{x}} p^q(\mathbf{x})} || p_0\right)$ を極小化するには $\alpha' = 1 - \frac{1-\alpha}{q}$ と選ぶと容易に計算ができる。

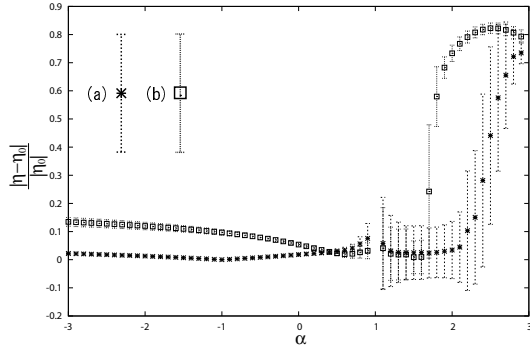


図 1 α -ダイバージェンスの極小化および Callen の等式に対するナイーブ平均場近似により求めた期待値パラメータの誤差。

Fig. 1 Errors of the expectation parameters derived from (a) the minimization of the α -divergence and (b) the naive mean-field approximation to the Callen identity.

例えば自由エネルギーの極小値を探索するアルゴリズムとして最急降下法を用いれば、 p_0 のパラメータ \mathbf{h} の更新規則は、

$$\begin{aligned} \Delta h^i &= -\delta \frac{dG_\alpha}{dh^i} \\ &= \frac{2\delta}{1-\alpha} \left[-\eta_{0i} + \eta_{0i}^{(\alpha)} + g_{0i\mu}^{(\alpha)} \frac{\partial \log \rho^{\frac{1-\alpha}{2}}(\boldsymbol{\eta}_0^{(\alpha)})}{\partial \eta_{0\mu}^{(\alpha)}} \right], \end{aligned} \quad (18)$$

となる。ただし、

$$\eta_{0\mu}^{(\alpha)} \equiv \langle F_\mu \rangle^{(\alpha)}, \quad (19)$$

$$g_{0i\mu}^{(\alpha)} \equiv \frac{2}{1+\alpha} \frac{\partial \eta_{0\mu}^{(\alpha)}}{\partial h^i}, \quad (20)$$

と定義した。 δ は最適化のパラメータである。ここで $p_0 \in \mathcal{M}$ に関する仮定より $\eta_{0\mu}^{(\alpha)} = \eta_{0\mu}^{(\alpha)}(\mathbf{h})$ は容易に求まるから、 \mathbf{h} が与えられれば $\eta_{0\mu}^{(\alpha)}, g_{0i\mu}^{(\alpha)}$ が計算でき、(18) に従って \mathbf{h} を更新してゆくことで $D_\alpha(p||p_0)$ の極小解が求まる。

4.1 2 値モデルへの適用例

ここでは (18) の適用例として N 個の 2 値素子からなる系を考える。系の状態を $\mathbf{x} = \{x_i = \pm 1 | i = 1, 2, \dots, N\}$ で表し、 $\sum_{\mathbf{x}}$ はそれらの取り得る状態 2^N 個全てについての和とする。また $F_{ij} = x_i x_j, F_i = x_i, F_0 = 1$ とすると、(10) において、

$$\theta^\mu F_\mu(\mathbf{x}) = \theta^0 + \theta^i x_i + \theta^{ij} x_i x_j, \quad (21)$$

となる。簡単な為、 $\min_{\mathbf{x}} [1 + \frac{1-\alpha}{2} \theta^\mu F_\mu(\mathbf{x})] = c (> 0)$

となるよう θ^0 を定めるものとする。

また $p_0 \in \mathcal{M}$ として各変数について因数分解可能な分布族を考える。即ち、

$$p_0(\mathbf{x}) = \prod_{i=1}^N p_{0i}(x_i) \quad (22)$$

である。いま各 x_i の取り得る状態は ± 1 の 2 値であるから (22) の分布は一般性を失う事なく指数型分布の形で表現できて、

$$p_0(\mathbf{x}) = \exp(h^i x_i - \psi_0(\mathbf{h})) \quad (23)$$

となる。ここで (18) にこの p, p_0 を代入すると、

$$\begin{aligned} \Delta h^i &= \frac{2\delta}{1-\alpha} \left[-\eta_{0i} + \eta_{0i}^{(\alpha)} \right. \\ &\quad \left. + g_{0ii}^{(\alpha)} \frac{\frac{1-\alpha}{2} (\theta^i + \theta^{ij} \eta_{0j}^{(\alpha)})}{1 + \frac{1-\alpha}{2} (\theta^0 + \theta^i \eta_{0i}^{(\alpha)} + \theta^{ij} \eta_{0i}^{(\alpha)} \eta_{0j}^{(\alpha)})} \right], \end{aligned} \quad (24)$$

$$\eta_{0i}^{(\alpha)} = \tanh\left(\frac{1+\alpha}{2} h^i\right), \quad (25)$$

$$g_{0ii}^{(\alpha)} = 1 - (\eta_{0i}^{(\alpha)})^2, \quad (26)$$

と求まり、(24) の反復計算によって求めたい $\boldsymbol{\eta}_0$ が得られる。また (24) において停留点の条件 $\Delta h^i = 0$ からは $\alpha \rightarrow 1$ の極限で、良く知られたナイーブ平均場近似の反復計算の式が導出される。

図 1 に N が小さく $\boldsymbol{\eta}$ の厳密計算が可能な場合において (24) から得られた近似値 $\boldsymbol{\eta}_0$ と $\boldsymbol{\eta}$ を比較した。横軸に (10) の α の値を、縦軸に $\frac{|\boldsymbol{\eta} - \boldsymbol{\eta}_0|}{|\boldsymbol{\eta}|}$ をとった。ここではノルムとして L_1 ノルムを用いた。ここで調べるパラメータ領域として、 $\{\theta^{ij}\}$ と $\{\theta^i\}$ はそれぞれ $\theta^{ij} \sim \mathcal{N}(0, 1)$ と $\theta^i \sim \mathcal{N}(0, 10)$ の正規分布に従って発生させた。ただし $c = 0.1$ とする。また、各 α の値ごとに 100 個のサンプルの平均値と標準偏差を示した。ここで、 $\boldsymbol{\eta}_0$ の初期値は $[-0.05, 0.05]$ の一様乱数とした。(a) に示すのは (24) のアルゴリズムによる結果である。ただし、各 α に対し最適化のパラメータを $\delta = \{1, 2, 3, 4, 5\}$ と変化させ、それぞれ 30 回の反復計算を行い、そのうちで最終的に最も (16) の α -ダイバージェンスの値の小さかった結果を採用した。(b) は Tsallis 統計に拡張した場合の Callen の等式をナイーブ平均場近似 [18] したもので、

$$\eta_i \approx \frac{p(x_i = 1, \boldsymbol{\eta}_{\setminus i}) - p(x_i = -1, \boldsymbol{\eta}_{\setminus i})}{p(x_i = 1, \boldsymbol{\eta}_{\setminus i}) + p(x_i = -1, \boldsymbol{\eta}_{\setminus i})}, \quad (27)$$

与えられる反復計算を 150 回繰り返した後の解である。ここで、 $\boldsymbol{\eta}_i = \{\eta_1, \dots, \eta_{i-1}, \eta_{i+1}, \dots, \eta_N\}$ である。両手法ともに一回の反復計算に必要な時間は $O(N^2)$ と評価できる。ただし $\alpha \rightarrow 1$ の極限における両アルゴリズムの定常解は、ナイーブ平均場近似のそれと同一であり、両手法の差は探索法による差のみである^(注2)。グラフに示すように $a \approx 1$ では両手法による結果に有意な差は認められない。一方で $\alpha \leq 0.2, 1.8 \leq \alpha \leq 2.5$ においては、(24) のアルゴリズムによる誤差が (27) による誤差よりも、今回調べたサンプル間の標準偏差分以上、小さくなっていることが読み取れる。今回特に、 \mathcal{M} として e-自己平行な多様体をとっているため、 $\alpha = -1$ において (24) の定常解は唯一に定まり、そこでは $\boldsymbol{\eta}$ と $\boldsymbol{\eta}_0$ は完全に一致する。つまり今回のように、 α -ダイバージェンスの最小化を期待値パラメータの近似値を求める目的で行った場合、よい近似を得られるための十分条件は $|\alpha + 1|$ が十分小さい場合だと言える。ただし α -ダイバージェンス D_α をもとに位相が定義できるのは $|\alpha| < 1$ の場合のみであり [19]、それ以外に関しては、 p または p_0 が小さな状態に“引っ張られ”ロバストでなくなることが知られている [20]。

5. α -ダイバージェンスと変分法

4. では α -ダイバージェンスの最小化に基く平均場近似について説明したが、 α -ダイバージェンスは Tsallis 統計における変分法とも深く関係している。Plastino らによる一般化 Bogolyubov の不等式 [14] は $D_{3-2q}(p||p_0) \geq 0$ の条件に対応していると解釈できる。ここでは Tsallis 統計に拡張した場合の Bogolyubov の不等式と α -ダイバージェンスとの関係、また α -ダイバージェンスが非負であるという条件から導かれる変分法について議論する。

以降では Tsallis 統計の枠組で考えるので p は (4) のように表されているとする。このとき (13) を、

$$\psi = -\frac{1+\alpha}{2}G_\alpha - \frac{2}{1-\alpha}\log\left[1 - \frac{1-\alpha^2}{4}D_\alpha\right], \quad (28)$$

のように書き換えることができる。また $D_\alpha(p||p_0)$ の非負性より任意の確率分布 $p \in \mathcal{S}$, $p_0 \in \mathcal{S}$ に対し (5) の F は

(注2) : (24) の左辺を 0 とおいた反復アルゴリズムを考えることはできるが、この場合 $\alpha > 1$ の領域ではアルゴリズムが収束しない場合が多く不安定である。

$$F \leq -\frac{1}{\beta}\log_q e^{-\frac{1+\alpha}{2}G_\alpha(p||p_0)}, \quad (\alpha \geq -1) \quad (29)$$

の不等式を満たす。即ち、 p に対する自由エネルギー F を直接計算することが困難な場合に、 $\mathcal{M}(C \mathcal{S})$ について

$$\min_{p_0 \in \mathcal{M}} G_\alpha(p||p_0) \quad (30)$$

の値を求めることで、(29) より α の値に応じて F の上界(または下界)を評価することができる。

(29) において α の選び方には任意性があり、特に $\alpha = 3 - 2q$ とした場合には一般化された Bogolyubov 不等式 [14] が再現される。即ち $p = e_q^{-\beta\mathcal{H}}e^{-\psi_0}$, $p_0 = e_q^{-\beta\mathcal{H}_0}e^{-\psi_0}$ と表したとすれば、任意の $\mathcal{H}, \mathcal{H}_0$ について

$$F \leq \frac{F_0}{H} + \left(1 - \frac{1}{H}\right) \frac{1}{(1-q)\beta}, \quad (q \leq 2) \quad (31)$$

の不等式が成り立つ。ただし上式において $H \equiv \sum_{\mathbf{x}} p_0 \left[\frac{1-(1-q)\beta\mathcal{H}_0}{1-(1-q)\beta\mathcal{H}} \right]$, $F_0 = -\frac{1}{\beta}\log_q e^{-\psi_0}$ である。従って (29) は (31) をパラメータ α によって拡張したものだと考えることができる。そして 4. 同様、 p のパラメータ q に応じて α の値をうまく定めることで $G_\alpha(p||p_0)$ の値を容易に評価することができる。

また、 F の値は \mathcal{H} を \mathcal{H}_0 のまわりで摂動展開することによっても評価することができる。このとき $q > 0$ について、

$$F \leq F_0 + \sum_{\mathbf{x}} p_0^q (\mathcal{H} - \mathcal{H}_0), \quad (32)$$

が成り立つ [15]。

5.1 調和振動子に対する F の上界の評価

ここではこのような手法が有用である例として、 F の厳密計算が可能な調和振動子からなる系を考える。また簡単の為、 $N = 1$ とする。このとき $p = e_q^{-\beta\mathcal{H}}e^{-\psi}$,

$$\mathcal{H}(\mathbf{x}) = \frac{x_1^2}{2m} + \frac{m\omega^2 x_2^2}{2} \quad (33)$$

である。いま x_1, x_2 はそれぞれ $(-\infty, \infty)$ の連続値を取るので状態和の計算は位相空間での積分 $\frac{1}{2\pi\hbar} \iint_{p(\mathbf{x}) > 0} dx_1 dx_2$ に置き換えられる。 \hbar はプランク定数を 2π で割ったものである。 $(2-q)\beta > 0$ の場合 F は厳密に計算できて

$$F_{\text{exact}} = -\frac{1}{\beta}\log_q \frac{1}{(2-q)\beta\hbar\omega} \quad (34)$$

となる。一方 \mathcal{M} として β_0 をパラメータとした 1 次

元の多様体を考え、その元を

$$p_0(\mathbf{x}; \beta_0) = e^{-\beta_0 \mathcal{H}(\mathbf{x}) - \psi_0} \quad (35)$$

のように書く。このとき 4. と同様の理由で G_{2q-1} の値は容易に求めることができ、

$$\frac{1}{1-q} \left\{ -q\psi_0 + \psi_0^{(q)} + \log \left[1 + \frac{(1-q)\beta}{q} \frac{\partial \psi_0^{(q)}}{\partial \beta_0} \right] \right\}, \quad (36)$$

となる。ここで、

$$\begin{aligned} \psi_0^{(q)} &\equiv \log \frac{1}{2\pi\hbar} \iint_{p(\mathbf{x})>0} dx_1 dx_2 e^{-q\beta_0 \mathcal{H}} \\ &= \begin{cases} \log \frac{1}{q\beta_0 \hbar \omega}, & (1-q)\beta < 0 \\ \log \frac{1}{q\beta_0 \hbar \omega} \left(1 - e^{-\frac{q}{1-q} \frac{\beta_0}{\beta}} \right), & (1-q)\beta > 0 \end{cases} \end{aligned} \quad (37)$$

である。いま $\beta > 0, 1 < q < 2$ の場合を考えると、(29), (30) の変分法によって F の上界を求めることができ、 $G_{2q-1}(p||p_0)$ は $\beta_0 = \frac{2-q}{q}\beta$ のとき最小となることがわかる。従って F を、

$$F < -\frac{\frac{(r_1 \beta \hbar \omega)^q}{r_2 \beta \hbar \omega} - 1}{\beta(1-q)} \quad (38)$$

と評価できる。ただし $r_1 = (2-q)/q, r_2 = (2-q)^2$ である。

図 2 では横軸を $T = 1/k\beta$ として、 $p_0 = e_q^{-\beta \mathcal{H}_0} e^{-\psi_0}$, $\mathcal{H}_0 = \frac{x_2^2}{2m} + V_0(x_2)$ とした場合に (32) の変分法から求めた F の上界^(注3) と (34) の F_{exact} との差: ΔF_1 (実線), と (38) による F と F_{exact} との差: ΔF_2 (破線), を比較した。ただし $m = \omega = k = 2\pi\hbar = 1$ とした。

6. むすび

本論文では (6) の α 分布族に対し α -ダイバージェンスの最小化という観点から期待値パラメータの近似アルゴリズムを定式化した。数値計算との比較によって相互作用項が外部入力による項に比べて十分小さい場合、本手法による近似が、 α の比較的広い領域で、相互作用項をその平均値で置き換えるという意味でのナイーブ平均場近似に比べて良い結果を与えることが確認できた。 p を射影する分布族 \mathcal{M} として今回は

(注3) : $V_0(x_2)$ として $|x_2| \leq L/2$ では $V_0 = 0, |x_2| > L/2$ では $V_0 = \infty$ という井戸型のポテンシャルを考えた場合。(32) 右辺の最小化より $L = \frac{4}{3-q} \left(\frac{3}{m\omega^2\beta} \right)^{\frac{1}{2}}$ が求まる [15].

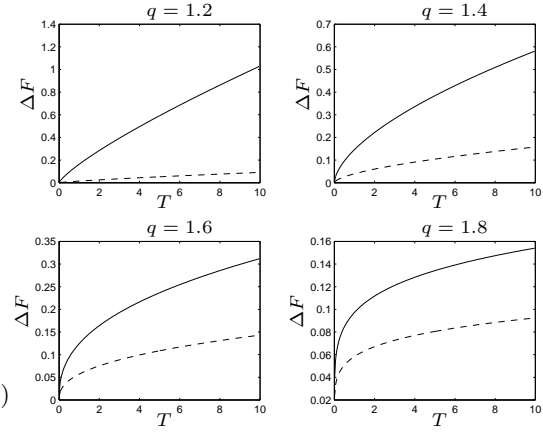


図 2 Tsallis の自由エネルギーの誤差の差と温度
Fig. 2 The two errors of the Tsallis free energy: ΔF_1 (solid line) and ΔF_2 (broken line) vs the temperature.

各変数について因数分解可能な、相互作用のない分布族をとったが、構造を持った分布を用いて近似することで近似精度を上げられる可能性がある。また、 α -ダイバージェンスは Tsallis の非加法統計力学 [13] とその変分法などに関連して深い関わりを持つが、 α -ダイバージェンスの非負性に基いて導かれる変分法によって Tsallis 型の自由エネルギーの上界または下界を押しやえることが可能であることを示した。

発達した乱流の速度差分布等は実際にベキ型の分布で特徴づけられることが知られている。本研究における近似手法をそのような実際の系に適用した場合、本研究における近似がどの程度有効であるかを調べることは今後の課題である。また本手法では p と p_0 の間の“距離”が近い場合に α -射影を -1 -射影の近似として用いたが、この α -射影が物理学的に何に対応しているのかということは興味深い問題である。

謝辞 本研究を進めるにあたり貴重な御助言を頂いた下川英敏氏、田中利幸氏、樺島祥介氏に感謝致します。

文 献

- [1] S. Amari, S. Ikeda, and H. Shimokawa, “Information geometry of α -projection in mean field approximation,” in *Advanced Mean Field Methods*, eds. M. Opper and D. Saad, pp.241-257, The MIT Press, Cambridge, MA, 2001.
- [2] M. Opper and O. Winther, “From naive mean field theory to the TAP equation,” in *Advanced Mean Field Methods*, eds. M. Opper and D. Saad, pp.7-20, The MIT Press, Cambridge, MA, 2001.
- [3] M. Opper and O. Winther, “Gaussian processes for

- classification: Mean field algorithms,” *Neural Computation*, vol.12, no.11, pp.2655-2684, 2000.
- [4] Y. Kabashima and D. Saad, “Belief propagation vs. TAP for decoding corrupted messages,” *Europhysics Letters*, vol.44, no.5, pp.668-674, 1998.
- [5] T. Tanaka, “Information geometry of mean-field approximation,” *Neural Computation*, vol.12, no.8, pp.1951-1968, 2000.
- [6] C. Beck, “Dynamical foundations of nonextensive statistical mechanics,” *Physical Review Letters*, vol.87, 180601, 2001.
- [7] C. Bhattacharyya and S.S. Keerthi, “Information geometry and Plefka’s mean-field theory,” *J. Phys. A*, vol.33 no.7, pp.1307-1312, 2000.
- [8] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS & Oxford University Press, 2000.
- [9] C. Arndt, *Information Measures*, Springer-Verlag, Berlin, 2001.
- [10] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol.52, issue 1-2, pp.479-487, 1988.
- [11] C. Tsallis, A. Rapisarda, V. Latora, and F. Baldovin, “Nonextensivity: from low-dimensional maps to Hamiltonian systems,” [cond-mat/0209168], 2002.
- [12] V. Latora, A. Rapisarda, and C. Tsallis, “Non-Gaussian equilibrium in a long-range Hamiltonian system,” *Physical Review E*, vol.64, 056134, 2001.
- [13] S. Abe and Y. Okamoto, eds., *Nonextensive Statistical Mechanics and Its Application*, Springer-Verlag, Berlin, 2001.
- [14] A. Plastno and C. Tsallis, “Variational method in generalized statistical mechanics,” *J. Phys. A*, vol.26, no.18, pp.L893-L896, 1993.
- [15] E.K. Lenzi, L.C. Malacarne, and R.S. Mendes, “Perturbation and variational methods in nonextensive Tsallis statistics,” *Physical Review Letters*, vol.80, issue 2, pp.218-221, 1998.
- [16] E.K. Lenzi, R.S. Mendes, L.R. da Silva, and L.C. Malacarne, “Remarks on $(1 - q)$ expansion and factorization approximation in the Tsallis nonextensive statistical mechanics,” *Physics Letters A*, vol.289, issue 1-2, pp.44-50, 2001.
- [17] F. Büyükkılıç, D. Demirhan, and U. Tırnaklı, “Generalization of the mean-field Ising model within Tsallis thermostatics,” *Physica A*, vol.238, issue 1-4, pp.285-294, 1997.
- [18] E.F. Sarmiento, “Generalization of the single-site Callen identity within Tsallis statistics,” *Physica A*, vol.218, issue 3-4, pp.482-486, 1995.
- [19] I. Csiszár, “On topological properties of f -divergences,” *Studia Scientiarum Mathematicarum Hungaria*, vol.2, pp.329-339, 1967.
- [20] 長岡 浩司, “情報幾何学の基礎的研究とロバスト推定への応用,” 東京大学工学部計数工学科修士論文, 1981.

(平成 x 年 xx 月 xx 日受付)

豊泉 太郎

平 13 東工大・理・物理卒, 同年東大大学院・新・複雑理工修士課程入学。現在, 最適化問題, 情報幾何の研究に従事

合原 一幸 (正員)

昭 52 東大・工・電気卒, 昭 57 同大大学院工学系研究科電子工学専門課程博士課程了。工博。現在, 東京大学大学院新領域創成科学研究科教授、放送大学客員教授。カオス工学および脳の数理モデリングの研究に従事。IEEE、日本ME学会、生物物理学会、電気学会、電設学会、神経回路学会各会員。