「モンテカルロ法の新展開3|

Tsallis 統計による拡張アンサンブル法の展開

産業技術総合研究所,生物情報解析研究センター 福田育夫

1. はじめに

Tsallis 統計 [1] は、従来のBoltzmann-Gibbs (BG) 統計を非加法的に拡張した熱・統計力学の枠 組みであり、熱・統計力学の基礎、数理に関するものから、乱流、粉粒体などの物理系あるいは異常 拡散現象等との関連、さらにネットワークフローや経済現象等におけるデータ解析など非常に広範囲 の様々な分野への応用まで、議論が展開されてきている [2]。

我々は、Tsallis 分布を生成する分子動力学 (MD) の手法であるTsallis dynamics (TD)を導き、こ れを用いて物理系のsimulation を可能にすることを示した[3] [我々が主に対象とするようなfull-atomic 記述の一般に複数分子から成るような系のsimulation では、モンテカルロ (MC) 法よりも分子動力学 (MD) 法が扱いやすいという観点から、ここでは主にMD法を中心とした議論を行わせて頂く]。さらに、 従来のsimulation 手法では、その状態サンプリング性能に限界があり、BG分布が生成できない、とい った取扱いが困難な系に対しても、Tsallis分布のエネルギーに関する緩やかな減少性等の利用によ り、効率良く正確な結果を与えること、またペプチド系においてエネルギー空間の幅広い探索が可能 なことを示した [4]。

しかしながら、本方程式を状態サンプリング手法として用いる場合、最も有効な結果を与えるTsallis 分布のパラメタ値 [5] をどのようにして求めるか、が問題になる。これに対し、ヒューリスティックな方法 等も可能であるが、対象とする系が大きくなると、一般に計算コストの増大が避けられない。従って、よ りスムースな応用を目指すためには、システマティックな方法が必要になってくる。

今回、エネルギー分布のデザインが可能になるTsallis パラメタ値についての新規の導出法を提案 した [6]。さらにまた、別アプローチとして、適切なパラメタ値を限定するのではなく、各パラメタ値を伴 った多数の分布を同時に生成するMDの一般的な手法も開発した [7] ので報告する。

2. Tsallis パラメタ値の導出

Tsallis 分布の密度関数

$$\rho_{\text{Tsallis}}(\omega) = \left[1 - (1 - q)\beta(E(\omega) + \varepsilon)\right]^{q/(1 - q)}$$

は、3つのパラメタをもつ (エスコート分布の形を用いた): $\beta = 1/k_{\rm B}T$ は、(逆)温度パラメタ、qはTsallis 指数パラメタ [$q \rightarrow 1$ の極限でBG密度関数 exp[$-\beta E(\omega)$]になる], ε はエネルギーのシフトパラメタ

研究会報告

である [8,9]。我々はまず、[8]の議論を精密化し再考することにより、この3つのパラメタの非独立性を 明確な形で示し、パラメタ値探索の際の指針を与えた。最初に、qを固定して、(ε ,T) への依存性を考 えるとき、我々は分布を不変にするパラメタ(ε ,T) 間のある変換に着目した。即ち、パラメタ値として全 ての(ε ,T) 値を探索する必要はなく、この変換で互いに写る同値なものの中から、代表を一つずつ探 索して行けばよい。ところで、この同値類は(ε ,T) の空間における傾き非零の半無限直線になることか ら、例えば、Tを任意の正値に固定して ε を探索するだけで十分となる。この結果は任意のq値に対し 成立するので、結局、Tを任意値に固定しqと ε の2つのパラメタ値の探索で十分なのである。

2つのパラメタ $q, \varepsilon \varepsilon$ 、Tsallisエネルギー分布 P_{r} に対する独立な2つの条件にて決める。我々は、これら2つのパラメタ値を、ある与えた2エネルギー値 E_{1}, E_{2} での $\ln P_{r}$ の傾きが各々 d_{1}, d_{2} になるという2つの条件を満たすように一意的に定められることを見出し、次の厳密な表式をもつことを示した [6]:

$$q = 1 + \frac{1}{a - 1},\tag{2-1}$$

$$\varepsilon = \frac{1-a}{\beta} - b, \tag{2-2}$$

ここで、

$$a = (E_1 - E_2) \frac{\kappa_1 \kappa_2}{\kappa_2 - \kappa_1},$$

$$b = \frac{\kappa_1 E_1 - \kappa_2 E_2}{\kappa_1 - \kappa_2},$$
(2-3)

と置いた。ただし、

 $\kappa_i = D \ln P_i(E_i) - D \ln \rho_i(E_i) - d_i, \qquad (2-4)$

であり、 ρ_i は、BG密度 $\rho_i(E) = \exp(-E/k_B T_i)$ 、のように明確な関数形で表現される分布密度であり、 $P_i \propto \rho_i \Omega$ は ρ_i とエネルギー状態密度 Ω から決まるエネルギー分布関数である。従って、例えば [3] で提案されているような方法を用いた適当な事前シミュレーションで得られる統計量を基に P_i を算 出できる。こうして、(2-1)、(2-2)式に従い、Tsallisのパラメタ*q*と ε の値を定めると、前述の

$$d_i = D \ln P_{\rm T}(E_i), \quad i = 1,2$$
 (2-5)

なる関係を得る。

この手法により決定したパラメタ値を用いたTD法 [3]をアラニンペプチドAc-Ala-Ala-NMe系に適用 したsimulationを行った。T = 200 K, $\rho_i \epsilon BG密度$, とし、 E_i は温度 T_i のBGエネルギー分布の最大 エネルギー点とした。常温付近でエネルギー分布が平坦になるように、 $T_1 = 300$ K, $d_1 = 0$ とし、さら に高温領域での分布を設計するため、 $T_2 = 700$ Kとし、 E_2 での分布の傾き d_2 にいくつかの値を与 えた。図2-1に示したsimulation結果は、設定した傾き d_1 , d_2 が良く再現されたことを示すものである。 このような例を通し、我々は(2-5)式の成立を検証し、本手法の妥当性を確認した。なお、ここではMD 法を用いたが、本手法は同様にMC法にも適用できることを指摘しておく。



図2-1: BGエネルギー分布(実線)及びTD法により生成した各種 Tsallis エネルギー分布

3. 多数の分布を同時に生成する MD 手法

前節の取り組みは、分布のパラメタを与えてそのエネルギー分布をデザインする方法であったが、 取り扱う分布を単一の分布に限定するのではなく、多数の分布を生成する方が、より自由度が高く簡 便な場合もある。このような考えの下で開発されたものとしてレプリカ交換MD (REMD) [10] があり、 様々な研究によりその有効性が示されてきた [11]。REMDでは、いくつかのBG分布を、予め定めた 適当なタイミングで、適当な確率をもって交換、といった操作が必要である。この交換の確率の問題は、 詳細釣合いの原理等に従うことで厳密な取扱いが可能であるが、交換のタイミング値の設定について は、充分な解析がなされておらず、実際この値の違いによるシミュレーション結果への依存性といった 問題も指摘されている [12]。我々は、このような不定のパラメタなしで多数の分布を生成する方法の 試みとして、REMDと異なり、決定論的手段により多数の分布を同時に生成する手法、を開発した[7]。

対象となる分布が*M* 個あるとし、その密度関数を、 ρ^a : $\mathbf{R}^N \to \mathbf{R}$, a = 1, ..., M, とする。物理系 を運動エネルギー $K(p) = \frac{1}{2} \|p\|^2$ とポテンシャルエネルギーU(x)で規定し、 ρ^a をこれらを通した次 のような形で与えることにする。

$$\rho^{a}(\omega) \equiv \rho^{a}(x, p, \zeta) \equiv \rho_{P}^{a}(U(x), K(p))\rho_{z}(\zeta), \qquad (3-1)$$

ここで ρ, は新たに導入された実変数 ζ の適当な実数値関数である。 従って、目的は、

$$\rho(\omega) \equiv \sum_{a=1}^{M} \rho^{a}(x, p, \zeta) = \sum_{a=1}^{M} \rho_{P}^{a}(U(x), K(p)) \rho_{z}(\zeta)$$

の形の分布密度を実現することである。このために、Nosé-Hooverの手法 [13,14] に基づき次のような

研究会報告

常微分方程式を導いた:

$$\dot{x}_i = \tau_2(x, p) p_i, \quad i = 1, ..., n,$$
 (3-2)

$$\dot{p}_{i} = -\tau_{1}(x, p)D_{i}U(x) - \tau_{3}(\zeta)p_{i}, \quad i = 1,...,n,$$
(3-3)

$$\dot{\zeta} = \tau_2(x, p) \|p\|^2 - nT,$$
 (3-4)

ここで、

$$\tau_{\alpha}(x,p) \equiv -T \frac{\sum_{a=1}^{M} D_{\alpha} \rho_{p}^{a}(U(x), K(p))}{\sum_{a=1}^{M} \rho_{p}^{a}(U(x), K(p))}, \quad \alpha = 1, 2,$$
(3-5)

$$\tau_3(\zeta) \equiv -T D \ln \rho_z(\zeta) \tag{3-6}$$

である。各 ρ^a が $\rho \equiv \sum_{a=1}^{M} \overline{\rho}^a$, ここで $\overline{\rho}^a \equiv \rho^a/Z^a$, $Z^a \equiv \int_{\Gamma} \rho^a(\omega) d\omega$, の形で規格化されている 場合を考えると、 Z^a を求めるのは通常困難なので、次のような対策をとる:

(a) $\rho \equiv \sum_{a=1}^{M} \overline{\rho}^{a}$ の代わりに $\tilde{\rho} \equiv \sum_{a=1}^{M} \rho^{a}/Y^{a}$, ここで $Y^{a} \equiv Z^{a}/Z^{c}$, cは $\{1,...,M\}$ 内の固定のひとつ、を用いる。

(b) Y^aは Z^aを計算するより容易であり、一般的な方法ではfree-energy perturbation 法や thermodynamic integration 法などが知られている。しかし、精密な Y^aを算出することはあくまでよいサ ンプリングを得るという目的を達成するための十分条件であり、実際は例えば以下の様な簡便な方法:

 $(P^{j-1}(E)/P^{j}(E))((\rho^{j}(E)/\rho^{j-1}(E)) = Z^{j}/Z^{j-1},$

を用いることができる。ここで、P^eは p^e から得られたエネルギー分布密度である。

(c) p から p への変更により、式 (3-2)-(3-6)で変更を受けるのは、式 (3-5)のみであり次の様になる:

$$\tau_{\alpha}(x,p) = -T \frac{\sum_{a=1}^{M} D_{\alpha} \rho_{p}^{a}(U(x), K(p)) / Y^{a}}{\sum_{a=1}^{M} \rho_{p}^{a}(U(x), K(p)) / Y^{a}}, \quad \alpha = 1, 2.$$
(3-7)

このとき、もし系が不変測度 $\tilde{\rho}d\omega$ に関してergodic ならば、 $\int_{U} |f \tilde{\rho}| d\omega < +\infty$ な関数 f に対し、

$$\lim_{\tau \to \infty} \frac{1}{\tau} \int_{0}^{\tau} f(\omega(t)) dt = \int_{\Gamma} f \tilde{\rho} d\omega / \int_{\Gamma} \tilde{\rho} d\omega$$
$$= \int_{\Gamma} f(\omega) \rho(\omega) d\omega / \int_{\Gamma} \rho(\omega) d\omega$$

となる。つまり f の長時間平均は、'multi'分布 $\rho = \sum_{a=1}^{M} \rho^{a}/Z^{a}$ における期待値に等しくなる。これは、 式 (3-2)-(3-4), (3-6), (3-7)がこのmulti分布を実現することを表す。同様に、本手法で得るポテンシャル エネルギー分布密度 P_{U} については、各aのそれ、 $P_{U}^{a}(u)$ 、の平均となることが証明される:

$$P_U(u) = \frac{1}{M} \sum_{a=1}^{M} P_U^a(u).$$
(3-8)

Y^aを求める際には、分布間の重なりが重要になってくる。Tsallis 分布では、通常対象となるBG分 布に比べて分布の広がりが大きいので、これを複数個利用すれば分布間の充分な重なりをより容易 に実現することができるので有利である。そこで、分布 p^aの具体形としてTsallis 分布を用いたものを 考察する。この場合、(3-7)式は次のようになる。

$$\tau_{1}(x,p) = \tau_{2}(x,p) = -T \frac{\sum_{a=1}^{M} \alpha_{a} b_{a} [1 + \alpha_{a} (E(x,p) + \varepsilon_{a})]^{b_{a}-1} / Y^{a}}{\sum_{a=1}^{M} [1 + \alpha_{a} (E(x,p) + \varepsilon_{a})]^{b_{a}} / Y^{a}}.$$
(3-9)

ここで、 $\alpha_a = (q_a - 1)/T_a$, $b_a = 1/(1 - q_a)$ [もしくは $q_a/(1 - q_a)$]. ところで、式 (3-2)-(3-4), (3-6), (3-9)で、 M = 1 とおくことは、単一の Tsallis 分布を生成することになるが、そのときこれらの式[で $\alpha_1 = (q - 1)/T$, $b_1 = q/(1 - q)$ としたもの]は、TDの方程式 [3] に(全体の*T*因子を除いて)同じにな るので、本手法はTD法の一つの拡張になっている。

本手法により、アラニンペプチドの系において、2つのTsallis 分布を合成させた場合のポテンシャ ルエネルギー分布密度を図3-1に示した。単一のTsallis 分布における結果の平均値とよく一致したこ とから、(3-8)式の成立が示され、また実際に単一Tsallis 分布を考えていただけでは容易にカバーさ れない非常に広い領域をカバーできることが分かった。図3-2にポテンシャルエネルギーの軌跡を示し た。本手法で得られたポテンシャルエネルギー軌跡は、単一のTsallis 分布を生成させた場合の軌跡 と本質的に異なるものであり、各々の単一分布の中心的エネルギー領域間で自然に起こっている遷 移が観察される。具体的にどの様な組合せの分布の合成が有効かについては [7] を参照されたい。



図3-1: パラメタの異なる (T=15 K 及び T=60 K) 2つのTsallis ポテンシャルエネ ルギー分布(破線及び一点鎖線)、それらの平均 (点線)、及び、本手法にて2つの Tsallis 分布を合成させたときのポテンシャルエネルギー分布 (実線)。







図3-2: 図3-1に示した以下の分布を生成したポテ ンシャルエネルギーの軌跡。(a) T=15 KのTsallis分 布; (b) T=60 KのTsallis分布; (c) 本手法にて上記2 つのTsallis分布を合成させた分布。

謝辞

本研究は、産業技術総合研究所,jbircの堀江将氏、大阪大学蛋白研の中村春木先生との共同研究 に基づくものであり、NEDOの補助を受けたものである。

参考文献

- [1] C. Tsallis, J. Stat. Phys. 52, 479 (1988).
- [2] http://tsallis.cat.cbpf.br/biblio.htm
- [3] I. Fukuda and H. Nakamura, Phys. Rev. E 65, 026105 (2002).
- [4] I. Fukuda and H. Nakamura, Chem. Phys. Lett. 382, 367 (2003); J. Phys. Chem. B 108, 4162 (2004).
- [5] I. Fukuda and H. Nakamura, AIP Conf. Proc. 708, 356 (2004).
- [6] I. Fukuda, M. Horie, and H. Nakamura, Chem. Phys. Lett. 405, 364 (2005).
- [7] I. Fukuda and H. Nakamura, Phys. Rev. E 71, 046708 (2005).
- [8] A. R. Plastino and A. Plastino, Phys. Lett. A 193, 140 (1994).
- [9] C. Tsallis, R. S. Mendes, and A. R. Plastino, Physica A 261, 534 (1998).
- [10] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. 314, 141 (1999).
- [11] A. Mitsutake and Y. Okamoto, J. Chem. Phys. 121, 2491 (2004).
- [12] R. Yamamoto and W. Kob, Phys. Rev. E 61, 5473 (2000).
- [13] S. Nosé, J. Chem. Phys. 81, 511 (1984).
- [14] W. G. Hoover, Phys. Rev. A 31, 1695 (1985).