

の更新式を用いて統一的に表されることが知られている。いまあるクラスター C_i と C_j が結合されたとき、この結合によって得られたクラスター $(C_i \cup C_j)$ と、これ以外のクラスター $C_k (k \neq i, j)$ との更新距離は4つのパラメータ $(\alpha_i, \alpha_j, \beta, \gamma)$ とクラスター間の距離により、つぎのように与えられ、パラメータの値を変えることにより、複数の手法が表されることから組み合わせ的算法という ([2], [4])。

$$d(C_i \cup C_j, C_k) \equiv \alpha_i d(C_i, C_k) + \alpha_j d(C_j, C_k) + \beta d(C_i, C_j) + \gamma |d(C_i, C_k) - d(C_j, C_k)|$$

$$(i \neq j, k \neq i, j)$$

(ここで、 $d(C_i, C_j) < d(C_i, C_k) < d(C_j, C_k)$)。

この算法でクラスターを作るとき、更新距離は明らかにもとの距離を保存せず、これを“距離のひずみ”といい、手法によってその挙動(ひずみの大きさや鎖状効果など)が異なることや階層構造の単調性(新しい結合距離はそれより前に作られたそれよりも大きいという距離空間の単調性)が崩れること(結合距離の逆転現象)などが知られていた([2], [3])。これらはいずれも実験的に観察されていたが、これを数理的に示すために、クラスター間の距離についてのある順序関係を用意し、これにより距離空間のひずみを、保存・拡大・縮小の3つに区分し、各手法がとりうる状態が、これらのいずれに相当するかを示した。これらは、DuBien らの結果の拡張でもある([1])。たとえば、群平均法は常に空間保存、最近隣法は縮小、最遠隣法は拡大であること、ワード法は従来言われていたこととは異なり、保存・拡大の混在、可変法は拡大・保存・縮小がパラメータのとりうる範囲によって異なること、などがすべて明らかになった。さらに、パラメータ $(\alpha_i, \alpha_j, \beta)$ の作る空間を考えると、各手法の相対的な位置関係が説明されるとともに、より一般的に保存・拡大・縮小となるための各パラメータの許容領域が示される。これらの結果を用いて、可変法が最近隣法、最遠隣法をも含むこと(したがってパラメータ γ は不要であること)、距離空間を保存の状態において、可変法と群平均法の性質を併せ持つような折衷型の新しい算法(調整型可変法)が考えられること、などを示した。

参 考 文 献

- [1] DuBien, J.L. and Warde, W.D. (1979). A mathematical comparison of the members of an infinite family of agglomerative clustering algorithms, *Canad. J. Statist.*, **7**, 29-38.
- [2] Lance, G.N. and Williams, W.T. (1967). A general theory of classificatory sorting strategies, I. Hierarchical systems, *Comput. J.*, **9**, 373-380.
- [3] Milligan, G.W. (1979). Ultrametric hierarchical clustering algorithms, *Psychometrika*, **44**, 343-346.
- [4] Williams, W.T. and Lance, G.N. (1977). Hierarchical classificatory methods, (eds. K. Enslein, A. Ralston and H.S. Wilf), *Statistical Methods for Digital Computers*, Vol. 3, Wiley, New York, 269-295.

分子動力学法における統計的諸問題

種 村 正 美

物質を形作る原子・分子のミクロなレベルから出発して、その物質の示す物理的性質を導き出すために計算機実験が行なわれる。その方法を大別するとモンテカルロ法と分子動力学法がある。分子動力学法は Newton の運動方程式を解いていくもので、モンテカルロ法と比較して確率・統計とは一見無縁のようであるが実は種々の統計的問題が潜んでいる。

われわれは分子動力学法による実験の信頼性に関係して現われるいくつかの問題点を整理し、実際に実験を行なうことによって検討した。その一つは計算精度の問題であるが、計算に用いる精度が1ワード32, 36および64ビットの場合について、それぞれ同一の初期条件から出発して、ある時間ステップ進んだ時点で時間の進行方向を反転させて、どれくらい以前のミクロな状態が再現されるかを調べた。もしも数値誤差が存在しなければ、いくらでも元の径路を遡ることが出来るはずである。検討の結果、少

なくとも倍精度での計算が必要なことが明らかになった。

実験を行なうと各時刻ごとに変動するデータが得られる。しかるに、従来この分野では、実験で得られた熱力学量の時系列を近代的な統計手法で処理をすることがあまり行なわれていない。物理量の時間変動の大きな様子を捉えたいことが往々にして生じる。定常時系列の場合には、それを時間平均によって求めることが出来るが、非定常時系列の場合には時間平均を使えない。通常は、生データを眺めて定性的に捉えることで終わってしまうことが多い。そこで、データをトレンド成分と不規則成分に分解できれば、非定常の場合も取扱いが簡単になる。そのために、我々は TIMSAC-84 にある石黒氏のソフトウェアによっていくつかのデータを解析したところ、決定論的に得られた変動データであるにも拘らず、分解が良好に行なわれるという結果を得た。今後はこのような手法の適用が望ましい。

最近、系を一定の温度や圧力に保つ種々の分子動力学法が開発されているが、それらを適用する現象によっては問題が生じる可能性が現われる。われわれは時間的な変化を詳細に追うダイナミカルな現象において、通常一定エネルギーの分子動力学法と他の分子動力学法を比較して問題点の所在を明らかにした。

本研究は上田 顯 (京大)・荻田直史 (理研)・小川 泰 (筑波大) の各氏との統計数理研究所 共同研究 (課題番号 62-共研-36) に基づくものである。

連結ベクトルの分布とグラフ解析

馬 場 康 維

k 個の状態で表現される系を考え、 k 個の状態を L_1, L_2, \dots, L_k とする。ここでいう状態とは離散型確率変数のとり得る値でも良いし、カテゴリーでも良い、あるいは連続無限個の状態を考えれば連続量でも良い。

状態 L_j をベクトルの方向で表現することにし、それを θ_j とする。状態 L_j に属する標本をベクトル

$$x_j = (w_j \cos \theta_j, w_j \sin \theta_j) \quad (0 \leq \theta_j \leq \pi)$$

で表現することになると、このベクトルを連結することによってデータの構造が表現できる。

各状態に対応するベクトルの合成ベクトルを

$$y = \sum_{j=1}^k x_j = (c, s)$$

とすると合成ベクトルの終点の座標 (c, s) に関して以下のことが成り立つ。

- (1) (c, s) を極座標 (R, ϕ) に変換すると (R, ϕ) は連結ベクトルによって表現される標本分布の集中度と平均に対応する。
- (2) (c, s) の漸近分布は 2 次元正規分布になる。

これらの結果から (1) 順位データのグラフ解析、(2) グラフを用いた分布の選択などが可能になる。

参 考 文 献

- [1] 馬場康維 (1987). 連結ベクトルの分布とその応用, 第 55 回日本統計学会予稿集, 55-56.
- [2] Baba, Y. (1987). Distribution of linked vectors and its application to graphical analysis, 46th Session of ISI, 25-26.

社会組織の防災力に関する問題

水 野 欽 司

自然災害対策では、常になんらかの「防災力」を暗黙に想定している。しかし「防災力」の社会的概念や計量化を具体的に問題として捉えた研究は少ない。それは、災害の様相の事前予想が難しく、被災