数理解析研究所講究録
第 418 巻 1981 年 2-8

点群の空間配置図から相互作用ポテンシャル関数を尤度法によって推定する。

統数研　尾形良彦
種村正美

1　序.

多次元空間（特に平面上）に於ける点群の配置に関する統計的解析の研究は長い歴史を有する。それは、まず配置データが完全にランダムな点群（ポアソン配置）であるか否か、ということの検定から始まった。与えられた配置から計算される様様の検定量は、次には非ポアソン配置の特徴を示す指標として広く生態学などで使用されてきた（例えば[1, 2, 3]参照）。時系列解析の発展に伴って、多次元点過程の２次モーメント情報にもとづいた解析が問題にされてきた。それはスペクトル解析（例えば[4]）だったり、統計物理で使われている 動径分布関数の類似品だったりする。このようにして最近の主流は、配置データ毎に、考えうる構造の理論モデルをたてて、シミュレーションを行なったりしながら上記の諸統計量の変化を示し、実データのそれらと比較するノンパラメト

リックで ad hoc な手法の寄せ集めによる検討である([5,6,7]など)。これらの手法は、配置データの生成にかんする科学的理論(もし明確に存在するなら)の検証などに有効である。

われわれの目的とするところは、全く統計的な立場だけから、尤度法に依拠したパラメトリックな推定方式の可能性と将来性を示唆することである。この方式は情報量基準(AIC)による統計モデルの比較をすることによって更に威力を発揮できる。

2. 尤度

有界領域 V 内に配置しているN個の点が与えられたとき、われわれはこれがしかるべき相互作用ポテンシャルのもとでの平衡状態と考える。すなわち点の配置法則はポテンシャル関数 $\phi_\theta(r)$ によって特徴づけられるものとする。ここに r は2点間の距離であり、ポテンシャル関数は θ によって適当にパラメータ化されている。

このとき熱力学的平衡分布である Gibbs 分布によって尤度を与えることができる。すなわち点の位置 $\mathbb{X}=\{x_1,\cdots,x_N\}$ に対して

$$L(\theta;\mathbb{X}) = \exp\{-U(\mathbb{X};\theta)\}/Z_N(\theta)$$

ただし温度は U に組み入れられて

$$U(\mathbb{X};\theta)=\sum_{i<j}\sum\phi_\theta(r_{ij})\quad,\quad r_{ij}=|x_i-x_j|$$

そして

$$Z_N(\theta)=\int_{V\times\cdots\times V}\exp\{-U(\mathbb{X};\theta)\}\,dx_1\cdots dx_N$$

は、それぞれ全相互作用エネルギーおよび配位分配関数と呼ばれるものである。

3．希薄気体のもとでの近似尤度

分配関数 $Z_N(\theta)$ は尤度の解析的計算を極めて困難にしている。しかし密度 $N/|V|$ が十分小であって、点が互いに近接にない限り相互作用が弱くて、二体衝突しか考えなくても良い状況では

$$Z_N(\theta)\approx\{1-a(\theta)/|V|\}^{N(N-1)/2}$$

の近似が許される。$a(\theta)$ は第2クラスター積分と呼ばれ

$$a(\theta)=\int_0^\infty\{1-e^{-\phi_\theta(r)}\}\,2\pi r\,dr$$

である。こうして、われわれは近似対数尤度として

$$\log L(\theta;\mathbb{X})=-\sum_{i<j}\sum\phi_\theta(r_{ij})-\frac{N(N-1)}{2}\log\left\{1-\frac{a(\theta)}{|V|}\right\}$$

を得る。最尤推定量は、これを最大にするパラメータをさがせば良い。

4. ポテンシャル関数のパラメータ化

第2クラスター積分 $a(\theta)$ が解析的に計算可能なポテンシャルとして次のような3つのモデルを考察する。

1) $\phi_\theta(r) = -\log\{1+(\alpha r-1)e^{-\beta r^2}\}$, $\theta=(\alpha,\beta), \alpha \geqq 0, \beta > 0$

2) $\phi_\theta(r) = -\log\{1+(\alpha-1)e^{-\beta r^2}\}$, $\theta=(\alpha,\beta), \alpha \geqq 0, \beta > 0$

3) $\phi_\theta(r) = \beta(\sigma/r)^n - \alpha(\sigma/r)^m$, $\theta=(\alpha,\beta,\sigma)$, $n>m, \beta \geqq 0$

モデル 3)は Lennard-Jones ポテンシャルと呼ばれている。これらはパラメータ値を適当に変えることによって、引力、斥力、引力斥力混合型の一定の相互作用を表現する。もちろんいずれのモデルもポアソンモデル $\phi_\theta(r) \equiv 0$ を含む。上記のモデルに対応する第2クラスター積分は

1) $a(\alpha,\beta) = (\pi/\beta)\left(1-\frac{\alpha}{2}\sqrt{\pi/\beta}\right)$

2) $a(\alpha,\beta) = \pi(1-\alpha)/\beta$

3) $a(\alpha,\beta,\sigma) = -\frac{\pi}{m}\beta^{1/m}\sigma^2\sum_{k=0}^{\infty}\frac{1}{k!}\Gamma\left(\frac{mk-2}{2m}\right)\alpha^k\beta^{-k/2}$,
$n=2m,\ m>2$

である。

5. モンテカルロ・シミュレーションと一般の尤度計算

指定されたパラメータ値 θ_0 をもつポテンシャル $\phi_{\theta_0}(r)$ のもとでの平衡状態にある点配置を計算機シミュレーションで

求める方法は[8]等で与えられている。これはGibbs分布を定常周辺分布にするようなマルコフ連鎖の遷移確率を構成し、これを格子上の点群のランダムウォークによって実現するものである。

これによれば $\mathbb{X}=\{x_1,\cdots,x_N\}$ の関数 $g(\mathbb{X})$ のGibbs分布による期待値 $\langle g\rangle$ は点群のランダムウォークの各ステップ t で実現される配置 $\mathbb{X}_t=\{x_1(t),\cdots,x_N(t)\}$ に対する関数値 $g(\mathbb{X}_t)$ の時間平均によって推定される。すなわち

$$\langle g\rangle \approx \frac{1}{T}\sum_{t=1}^{T} g(\mathbb{X}_t) 。$$

いま g として $\exp\{U(\mathbb{X};\theta)\}$ をとれば $\langle g\rangle=|V|^N/Z_N(\theta)$ であるから、これにもとづいて分配関数の値を推定できる。この値と3節に於ける近似値を比較することができる。

3節に於けるような近似が許されない場合でも固定したパラメータ θ の推定尤度値として

$$\hat{L}(\theta;\mathbb{X})=\frac{1}{T}\sum_{t=1}^{T}\exp\{U(\mathbb{X}_t;\theta)-U(\mathbb{X};\theta)-N\log|V|\}$$

が考えられる。ただし $\mathbb{X}$ は配置データであるが $\{\mathbb{X}_t, t=1,2,\cdots T\}$ はポテンシャル $\phi_\theta(r)$ についてモンテカルロ・シミュレーションで得られた点群のランダムウォーク（配置の時系列）である。したがって原理的には、これによって最尤推定を得たいところだが、以下のベイズ的な推定法の方が、その性格上、安定

で現実的であろう。すなわちパラメータ θ について適当な事前分布 $\pi(\theta)$ を想定すると

$$\hat{\phi}(r;\mathbb{X}) = \int \phi_\theta(r) L(\theta;\mathbb{X})\pi(\theta)d\theta \Big/ \int L(\theta;\mathbb{X})\pi(\theta)d\theta$$

はポテンシャル関数 $\phi(r)$ のベイズ推定である。しかし上記の積分は極めて困難なので、分布 $\pi(\theta)$ にもとづいて θ_i，$i=1,2,\cdots,M$ を生成することによって

$$\tilde{\phi}(r;\mathbb{X}) = \sum_{i=1}^{M} \phi_{\theta_i}(r)\tilde{L}(\theta_i;\mathbb{X}) \Big/ \sum_{i=1}^{M} \tilde{L}(\theta_i;\mathbb{X})$$

をポテンシャル関数の推定とする。

6．結び

3節に於ける近似尤度のもとでの、最尤法や最小 AIC 法によるモデル選択や、ポテンシャルの推定の有効性は[9]に於いて詳しく議論されている。このときのわれわれの議論は、従来からのポアソン検定の議論を含む。なぜならポアソン配置は理想気体だから。

ポテンシャルのパラメータ化は、目的に応じてなされるべきであり、モデル構成の立場からみて重要である。

参考文献

[1] Morisita, M. (1959). Measuring of the dispersion of individuals and analysis of the distributional patterns. Mem. Fac. Sci. Kyushu Univ. Ser. E (Biol.) 2, 215-235.

[2] Hopkins, B. and Skellam, J. G. (1954). A new method of determining the type of distribution of plant individuals. Ann. Bot. Lond. N.S. 18, 213-227.

[3] 種村正美, 尾形良彦 (1981). 点の配置パターンを測る —なわばりの生態学—. 数理科学 3月号, サイエンス社

[4] Bartlett, M. S. (1964). The spectral analysis of two-dimensional point processes. Biometrika 51, 299-311.

[5] Ripley, B. D. (1977). Modelling Spatial Patterns (with discussion). J. Roy. Statist. Soc. B39, 172-212.

[6] Diggle, P. J. (1979). On Parameter Estimation and Goodness-of-fit Testing for Spatial Point Patterns. Biometrics 35, 87-101.

[7] 種村正美, 長谷川政美 (1980). カモメ類の巣配置パターンのモデル化. 統数研彙報 27.

[8] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087-1092.

[9] Ogata, Y. and Tanemura, M. (1980). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. Res. Memo. No. 191, Inst. Statist. Math.