

不確実性のモデル化と予測

竹村彰通 合原一幸 駒木文保 青木敏 下川英敏 鈴木秀幸
情報理工学系研究科数理情報学専攻

概要

現実の現象をモデル化するには、不確実性のモデル化が避けられない。しかしながら現象のどの部分を確定的に扱い、どの部分を統計モデルなどにより不確実性として扱うかの切り分けはあきらかではない。ロバスト性という観点からモデル化の方法論を確立することをめざす。

1 研究の成果と今後の展望

本研究では、確率場の理論と応用、統計的予測理論、マルコフ連鎖・モンテカルロ法による離散データ解析、複雑システムモデリング、において本年度も多くの研究成果が得られており、それぞれロバスト計算原理の観点から重要な成果であると考えられる。ここでは今後の展望をふまえながら、これらの成果をまとめる。

1.1 確率場の理論と応用

ガウス確率場の理論においては、最近になって「オイラー標数法」や「チューブ法」の方法が急速に発展している。[3]では、ガウス確率場が有限な直交関数展開を有する場合について、モースの定理を拡張することにより、「オイラー標数法」と「チューブ法」の同等性を証明した。この結果は[2]において非常に一般的な設定に拡張され、永年の課題であったオイラー標数法の一般的な正当化が与えられた。また臨界半径が0となる非正則な場合においては、チューブ法の近似が部分的にしか保証されないが、その場合の詳しい誤差評価が[1]で与えられた。その後も、竹村と統計数理研究

所の栗木哲氏、スタンフォード大学の Jonathan Taylor 氏、イスラエル Technion の Robert Adler 氏との共同研究が進行中であり、ラプラス法を用いてオイラー標数法の誤差についてさらに詳しい評価が得られつつある。

1.2 統計的予測理論

統計科学におけるさまざまな重要な問題を、予測問題として定式化し、予測分布の性能を評価することにより、理論的な解決をあたえる研究を行った。

統計的予測では、ベイズ法が有力な方法のひとつである。ベイズ統計学では、いわゆる無情報事前分布としてどのような分布を採用すべきかについて数多くの研究がなされてきた。現在では、どのような目的にも使える万能な無情報事前分布を、伝統的な主観確率の考え方に基づいて構成することは困難であることが共通の認識となっている。本研究では、ベイズ法を統計解析の目的に応じた予測を構成するための手段としてとらえ、予測分布の性能の評価を評価することにより、事前分布の構成やモデルの選択を行うことを提案した。

本研究におけるアプローチと密接に関連するものとして、情報理論的な方法により無情報事前分布を構成する研究がある。Bernardo は、パラメータと観測量の相互情報量を最大にする事前分布を利用することを提案し、ヒューリスティックな議論により Jeffreys 事前分布を導出した。Barron は、パラメータ空間をコンパクトな部分集合に制限した上でパラメータと観測量との相互情報量の漸近的な評価を厳密に行い、相互情報量を漸的に最大にする分布として Jeffreys 分布を導出して

いる。しかし、このようなコンパクト性に関する人為的な仮定の下での議論では、Jeffreys 事前分布が必ずしも良いベイズ推測を与えない多くの例を説明できない。本研究では、観測量を与えたもとでの、パラメータと予測すべき量との条件付相互情報量が予測のリスクであることから、条件付相互情報量を評価することによりこの困難が自然に回避できることを示した。これは、ベイズ予測の理論的な基礎となるものである。

さらに、統計的予測法に関する以下の研究を行った。

1) ブートストラップ法を利用した予測の性能の漸近的な評価を与え、この手法が推定量を直接利用する方法より優れていることを示した。さらにブートストラップ法とベイズ法との関係を明らかにした [4]。

2) 局所漸近混合正規性をもつモデルについてベイズ予測の性能を評価し、さらに Kullback-Leibler ダイバージェンスに基づく予測のリスクの不偏推定量を求めた。これは、局所漸近混合正規性をもつモデルの情報量規準となる量である。従来のベイズ法やモデル選択に関連する漸近理論は、ほとんど局所漸近正規性を仮定したものに限られていた。

3) AR モデル, MA モデルのモデル多様体の断面曲率の符号を統一的に評価した。特に AR(2) モデル多様体が負曲率多様体であることを示し、性質の良い事前分布の構成を行った [5]。

2003 年 12 月に香港で開催された Bernoulli 協会の国際会議, 12 月にインドで開催された確率と統計に関する国際会議で、これらの研究成果についての発表を行い高い評価を得た。

1.3 マルコフ連鎖・モンテカルロ法による離散データ解析

集団を、性別や年齢、あるいは疾患の有無や生活習慣などの要因で多重分類し、それぞれの人数を表の形で表したものは分割表と呼ばれる。分割表に要約されたデータから、要因間のさまざまな関連を調べるための解析手法は、その、医学、疫

学、工学、自然科学などのさまざまな分野における応用上の重要性もあり、発展してきたが、特に近年、計算機の進歩や、インターネットを利用した大規模なデータを取り扱う可能性などを背景に、サンプリングベースの統計量の計算手法が注目されている。類似の例では、遺伝性疾患の関連遺伝子の同定において、候補となる数十～数百の部位を同時に解析しなければならない、というような状況がある。そのような場合に、研究上興味のない多くの不確実性（統計学的には局外母数に対応する）を、興味の対象となる不確実性と同列に扱うことは全くナンセンスであるため、局外母数の値によらない推定方式が必要となる。このように、より具体的には、われわれが遭遇する多くの統計的推測の問題は、集約的には、ある統計量の条件付期待値の推定問題として定式化することができ、マルコフ連鎖・モンテカルロ法は、その数値的評価のためのひとつのアルゴリズムである。とくにそれは、単純なモンテカルロ積分が実行できない（直接的なサンプリングが不可能）というようなケースを想定しており、Importance Sampling 法とはその目的を共有するものである。

離散データ解析におけるマルコフ連鎖・モンテカルロ法は、1998 年に Diaconis and Sturmfels によって提案され、注目を集めた。この方法は、マルコフ連鎖を構成するための基底を、代数アルゴリズムを用いて算出するものであり、理論上はこの方法で、任意の離散の条件付分布からのサンプリングが可能になった、という点で興味深いものである。しかし一方で、この代数アルゴリズムを用いた基底の算出方法には問題点も多く、中でも、計算時間の問題と、得られる基底が極小でないという問題は重大である。計算時間の問題は、代数アルゴリズムの理論的な計算量が、変数の数の二重指数オーダーであることに起因しており、比較的小さなサイズの問題に対しても、実際の計算はすぐに破綻してしまう。また、基底が極小でないという問題は、代数アルゴリズムが変数間に項順序を与えて計算するものであるため、変数間の対称性が崩れる、ということに起因している。

これに対し、青木、竹村は、変数間の対称性に

注目した手法により、代数アルゴリズムを用いずに直接極小な基底を算出する方法を提案した。提案する手法により、比較的小さなサイズの問題に対しては、代数アルゴリズムを用いるよりもはるかに効率的に極小基底を求めることが可能となる。特に、3元分割表の無三因子交互作用モデルに対しては、[6]において、 $3 \times 3 \times K$ の問題に対する極小基底を導出した。その後、より大きなサイズ($3 \times 4 \times K, 4 \times 4 \times 4$)の問題に対する解も得られ、現在投稿中である。さらに、[7]において、極小基底が一意に存在するための必要十分条件を導出し、さまざまな問題に対する極小基底とその一意性を確認した。これらの結果は、2003年12月にカリフォルニアで行なわれたARCCワークショップ(Computational Algebraic Statistics)において発表され、好評を得た。その後の研究では、分割表の、水準の入れ替えに関する不変性をもつ基底に注目し、不変極小基底の性質とその一意性に関する理論的な結果も得られ、現在投稿中である。この研究で得られた諸結果は、理論的な重要性のみならず、応用上、非常に大きな意味を持つものであり、今後も研究を継続し、国内外で発表していく予定である。

1.4 複雑システムモデリングの理論と応用

複雑なシステムのモデル化とその予測等への応用に関して、以下の研究を行なった。

(1) 時系列データの解析はシステムを理解し、予測や制御をする上で重要である。しかし、経済や気象、生体など実世界の複雑な現象に対しては既存の手法は必ずしも有効ではない。本研究ではマルチフラクタルの観点から新しい時系列解析の手法を開発した。時系列を線分上の確率分布として捉えることで時系列のスペクトルを求め、時系列の変動との間に相関があるような特徴パラメータをスペクトルから抽出した。円ドル為替レートの時系列を解析した結果、価格変動は長い時間スケールにおいて厳密にランダムではなく、この手法が予測に有効である可能性が示された。また、確率過程と仮想市場の観点から短かい時間スケ-

ルでの価格変動の性質についても調べ、価格変動のメカニズムの理解を深めた [8]

(2) 大規模な非線形ネットワークにおいて普遍的に起こることが知られている擬アトラクタ間の不規則な遍歴現象の発生要因を調べたため、同様の現象を低次元力学系を通じて解析した。扱った2次元結合写像では、2つの離れたカオスアトラクタのいずれかに収束していた軌道が、パラメータ変化とともに突然それら2つの間を遷移し始める現象が見られる。この発生要因は、2つのカオスアトラクタとそれらのフラクタルベイスン境界が接触する大域的分岐現象であることを明らかにした。さらに、このフラクタルベイスンが発生、変形、崩壊する様子は、境界のフラクタル次元の変化でとらえることができることを示した [9]。

(3) 従来のいわゆるニューラルネットワーク理論は、入力を全て線形に加算し、その結果に対して一回の非線形変換を施すユニットを多数つなげた際の計算能力についての理論である。近年の実験技術の進歩によって、生物の脳の神経細胞は、こうした人工ニューラルネットの素子とは違って、そこでの入力の加算も本質的に非線形であることが分かってきた。こうした知見は、脳機能の解明にとって重要であるのみならず、そうした実際の神経細胞に似た特性を持つユニットを組み合わせ、従来の人工ニューラルネットとは異なる計算能力を持ったデバイスを構成する可能性をももたらすものである。本研究では、神経科学分野でここ数年の間にコンセンサスとなってきた知見を元にして、各神経細胞において、正の入力と負の入力の差し引きが入力源ごとに局所的に行われるような神経回路の数理モデルを構築し、それが従来の同様のニューラルネットとは異なる計算特性を持つことを見出した。さらに、このモデルをアナログVLSI回路として実装する可能性について検討した [10]。

(4) 一般の1次元離散時間力学系における相空間の分割について理論的な解析を行った。離散時間力学系では、生成的な分割と呼ばれる方法で相空間を分割すると、もとの力学系と分割後の離散状態の力学系が、位相同型になることが知られて

いる。本研究では、相空間の非生成的な分割による離散状態力学系の挙動を調べた。特に、テント写像について、分割の位置と、離散状態力学系の位相エントロピーが非単調な「悪魔の階段」状の構造をもつこと(いたるところ傾き0であること)を示した。また、この結果の応用として、実験等で観測した時系列から、現象を支配する力学系の不安定周期軌道を求める方法を提案した。これは、時系列を、一旦、生成的な分割によって離散状態時系列に変換し、その上での周期記号列を求める方法である [11]。

参考文献

- [1] Takemura, A. and Kuriki, S. (2003). Tail probability via the tube formula when the critical radius is zero. *Bernoulli*, Vol.9, No.3, pp.535–558.
- [2] Taylor, J., Takemura, A. and Adler, R. (2003). Validity of the expected Euler characteristic heuristic. Technical Report METR 03-26, August 2003, tentatively accepted for publication in *Annals of Probability*.
- [3] Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Annals of Applied Probability*, **12**, 768–796.
- [4] Fushiki, T., Komaki, F. and Aihara, K. (2004). Nonparametric bootstrap prediction. *Scandinavian Journal of Statistics*, to appear.
- [5] Tanaka, F. and Komaki, F. (2003). The sectional curvature of AR model manifolds. *Tensor*, to appear.
- [6] Aoki, S. and Takemura, A. (2003). Minimal basis for connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals. *Australian and New Zealand Journal of Statistics*, **45**, pp. 229–249.
- [7] Takemura, A. and Aoki, S. (2004). Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. *Annals of Institute of Statistical Mathematics*, to appear.
- [8] T. Ohnishi, K. Aihara, M. Takayasu, and H. Takayasu (2003). Statistical properties of the moving average price in dollar-yen exchange rates. *Proc. of Applications of Physics in Financial Analysis*, **4**, Vol.27C, p.A86.
- [9] G. Tanaka and K. Aihara (2003). An explosion of chaotic attractors in two coupled logistic maps. *Proc. of the 11th workshop on Nonlinear Dynamics of Electronic Systems*, pp.275–278.
- [10] K. Morita and K. Aihara (2003). A pre-integration lateral inhibition model in continuous time. *Proc. of Shanghai International Symposium on Nonlinear Science and Applications*, Shanghai, China.
- [11] K. Fukuda and K. Aihara (2004). Symbolic dynamics of a chaotic neuron model. *Artificial Life and Robotics*. (in press).