

不確実性のモデル化と予測

竹村彰通 合原一幸 駒木文保 鈴木秀幸 清智也

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、外乱の存在など、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。不確実性のモデル化と予測グループでは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して研究している。

1 はじめに

数理モデルにおける確定的な部分と非確定的な部分の切り分け、さらに非確定的な部分の扱い、のバリエーションを考慮すれば、現象のモデル化において様々なアプローチが可能である。このような中で重要な研究目標は、ロバスト性の観点、すなわち与えられた現象へのモデルの安定的な適合と予測の観点から、多くの方法論を統一的に比較しすぐれたモデルを選びだす指針を与えることである。このような指針を与える基礎研究として、個々の手法の一層の深化が必要であることは言うまでもない。

以下では、以上のような目標を念頭におきながら、複雑システムモデリングの理論と応用、個票データベースの安全な利用法、統計的推測・予測の情報幾何、空間統計学におけるモデリングと漸近理論について、この時点での研究成果と今後の研究の展望について述べる。

2 複雑システムモデリング

複雑システムモデリングの理論と応用に関して、カオス写像の結合系 [1] や多状態連想記憶モデル [2] の非線形ダイナミクス、二重回転写像のカントール関数的応答特性 [3]、クラス I* ニューロンの電子回路実装 [4] などの研究を行なった。また、東京大学駒場リサーチキャンパスコンベンションホールにおいて、International Symposium on Complexity Modelling and its Applications 2005 (2005 年 11 月 21 日～23 日)、感染症理論疫学研究大会 2006 (2006 年 1 月 28 日)、マルチニューロン研究会 2006 (2006 年 2 月 20 日～21 日) の 3 つの会議を科学技術振興機構 ERATO 合原複雑数理モデルプロジェクトと共催で開催した。

3 個票データベースの利用と安全管理の手法

官庁統計や社会調査で得られる統計データは、伝統的には統計表の形に整理された上で分析されて来た。最近では、統計調査の際に得られる個々の回答者のデータ (個票データ) は直接デジタルデータとして記録され、また統計パッケージ等の整備により、これらのデータを集計表以前の生の形で解析することが可能になって来た。この際に問題となり得るのは、回答者のプライバシーの問題である。

個票データの安全性の評価には、回答者が特定される確率のモデル化が必要となるが、この目的のために集団遺伝学や計量言語学で発展して来た確率モデルを用いることができる。特に集団遺

伝学での種の分布に関する確率モデルや、計量言語学における語彙の分布に関するモデルが、個票データの問題に直接的な関連を持っている。逆に、個票データの安全評価という観点からこれらのモデルを見なおすことにより、これらのモデルについてより深い理解が得られる。竹村および共同研究者の研究成果については、統計数理研究所の『統計数理』の2003年第51巻第2号において、竹村をオーガナイザーとする「特集 個票開示問題の統計理論」([5])が発行され、この分野におけるわが国の国際的貢献を含んで個票開示問題研究が大いに進展した。竹村自身がここで分野全体のサーベイ([6])を与えている。より最近の研究成果については[7]の報告書に詳しく報告されている。

最近の発展として、個票データを多元の分割表として考えることにより、個票データの安全性の問題が、与えられた周辺和に対応する分割表の存在問題や周辺和を共有する他の分割表の存在問題などの代数的な問題として定式化できることがわかって来た。これにより、個票データの安全性に関する様々な問題が、整数計画法の解の存在問題や、グレブナー基底の構成問題などに帰着される。例えば[8]においては、周辺和を共有する他の分割表の存在問題をグレブナー基底およびマルコフ基底の観点から研究し、すべてのマルコフ基底に現れなければならない分割表の特徴づけを与えている。また[9]においては、与えられた周辺和に対応する分割表の存在問題が、半群とその“saturation”の関係として定式化できることを示している。また、対応する分割表が存在しないような周辺和の集合が有限集合にとどまる場合と無限集合となる場合の違いについて、いくつかの必要十分条件を与えている。これらの結果は、代数学の研究者からも注目されている。

4 統計的推測・予測の情報幾何

性能の良い予測法・推測法を、情報幾何のアプローチを用いて統計モデルとベイズ法に基づいて系統的に構成する手法の研究を行った。

パラメトリックな統計モデルは情報幾何の立場

からは双対接続をもつリーマン多様体と見なせる。これをモデル多様体と呼ぶことにする。モデル多様体の微分幾何学的な性質を調べることにより、適切な事前分布が構成でき、良い予測・推測が可能になることを示した[12]。

応用上重要な具体的なモデルは、それぞれ必然性をともなった美しい幾何学的な構造を持つことが多い。本研究では、応用上重要な具体的なモデルの幾何学的な構造を調べるとともに、性能のよいベイズ予測をあたえる事前分布を具体的に構成した。

全空間での積分が1となる通常の意味での確率測度をパラメータ空間の事前分布として用いるときプロパーな事前分布という。これに対し、全空間での積分が無限大となる事前分布はインプロパーな事前分布と呼ばれる。インプロパーな事前分布は直感的にはプロパーな事前分布の極限として捉えられる。パラメータ空間に対しなんらかの知識を仮定できない場合は、従来理論で構成される自然な事前分布はインプロパーなものとなることが多い。

プロパーな事前密度に基づく予測や推測が、インプロパーな事前密度となる無情報事前分布に基づく予測や推測より、真のパラメータの値によらず、常に良い性能をもつことは素朴な直観からすれば不合理なことにも感じられる。しかし、多変量正規分布の場合には、このような現象が起こることが知られていた。本研究の立場からは、多変量正規分布と、多変量ポアソンモデルはモデル多様体として、類似した構造をもつので、同様の現象が多変量ポアソンモデルの場合にも生じることが予想される。

実際、多変量ポアソンモデルの場合において、プロパーな事前密度に基づく予測密度でジェフリーズ事前分布に基づく予測密度を優越するものがモデルの次元が5次元以上の場合に構成できることを証明し、この現象が正規分布に限らない普遍的なものであることを示した[13]。また、時系列のスペクトル密度の空間での幾何学を調べARモデル、ARMAモデルの情報幾何学的な性質を調べ、具体的な事前分布の形を与えた。

以上のように、個々の具体的な統計モデルの性質を調べて応用するのは、モデルによっては必ずしも容易であるとは限らない。そのようなモデルに対しては漸近理論によるアプローチが必要になる。予測のリスクを漸近的に評価することにより、予測分布の漸近許容性、漸近ミニマックス性などの決定理論を漸近理論の視点から展開した。

また、モデルが複雑で積分が困難な場合にも利用できるブートストラップ予測に関する研究をおこなった [10]。

パラメトリックな統計モデルだけではなく、ノンパラメトリックな手法についても情報幾何、ベイズ理論の視点から研究をおこなった。サポートベクトルマシンが一種のベイズモデルとして捉えられることを示し、情報量規準の構成等に应用した [14]。また、経験尤度法等の性質についても、情報幾何の視点から捉えられることを示した。

また、古典的な確率に基づくモデルから離れ、量子確率論に基づく統計モデルに関して、量子状態の予測密度作用素の研究を行い、古典的な統計学と同様にベイズ法に基づいて構成される予測密度作用素の性能が優れていることを示した [11]。

2005年12月12日から16日まで、東京大学本郷キャンパス山上会館において、第2回「情報幾何とその応用」国際会議を開催した。国内から多くの研究者が参加するとともに、海外から、Dawid, Petz, Pistone など情報幾何関係の著名な研究者や若手研究者が多数参加した。統計的推測理論、ベイズ統計、代数統計学、量子統計モデル、決定理論の幾何学、ブースティング、脳、ニューラルネットワーク、チューブ法、ゲーム論的確率論など、統計学、量子情報、学習理論等の幅広い分野における情報幾何を切り口とした研究に関する発表がなされるとともに、国内外の研究者間で活発で有意義な研究交流がもたれた。この国際会議における発表は、48編の論文からなるプロシーディングス “Proceedings of the second international symposium on information geometry and its applications” にまとめられている。

5 空間統計学におけるモデリングと漸近理論

地球環境データなど、空間的依存性を持つデータの解析は近年ますます重要となっている。本研究では特に、観測領域が有界で、標本の観測地点が密になっていくという設定の下で漸近理論を展開した。このような稠密型漸近理論の設定は現実的である反面、従来のエルゴード性に立脚した漸近理論とは扱い方が本質的に異なるため、十分に解明されていない。

空間データの従う確率分布としてガウス分布を仮定した場合、空間的依存関係は相関関数で記述される。相関関数に強い影響を与えるパラメータとしてフラクタル指数があり、フラクタル指数を精度よく推定することは予測（補間）性能の向上につながる。本研究では擬似最尤法に基づくフラクタル指数の推定量に対し、その分散を稠密型漸近理論の枠組みで決定した。さらに、漸近バイアスの式が、情報幾何学の立場から自然に導かれることを示した。これらの内容を、本プロジェクトと理研脳科学センターの共同主催による国際シンポジウム “The 2nd International Symposium on Information Geometry and its Applications” にて発表した [15]。

一方で、ガウス性の仮定はしばしば現実のデータに適応しないことがある。そのための対処としてガウス分布を変換したモデルが提案されている。従来の変換モデルでは変換のクラスが限定されていた。本研究では、より一般的かつ扱いやすい変換モデルのクラスを構成し、応用・理論の両観点から検討した。結果は現在まとめている段階である。

参考文献

- [1] Gouhei Tanaka, Miguel A. F. Sanjuan, Kazuyuki Aihara. Crisis-Induced Intermittency in Two Coupled Chaotic Maps: Towards Understanding Chaotic Itinerancy, *Physical Review E*, **71**(1), 016219 (2005).

- [2] Gouhei Tanaka, Kazuyuki Aihara. Multi-state Associative Memory with Parametrically Coupled Map Networks, *International Journal of Bifurcation and Chaos*, **15**(4), 1395–1410 (2005).
- [3] Hideyuki Suzuki, Shunji Ito, Kazuyuki Aihara. Double Rotations, *Discrete and Continuous Dynamical Systems*, **13**(2), 515–532 (2005).
- [4] Takashi Takemoto, Takashi Kohno, Kazuyuki Aihara. MOSFET Implementation of Class I* Neurons Coupled by Gap Junctions. to be published in *Journal of Artificial Life and Robotics* (2006).
- [5] 『統計数理』「特集 個票開示問題の統計理論」2003年第51巻第2号, 統計数理研究所, 竹村彰通:オーガナイザー. pp. 181–387.
- [6] 竹村彰通 (2003). 個票開示問題の研究の現状と課題. 「特集 個票開示問題の統計理論」『統計数理』2003年第51巻第2号, pp. 241–260.
- [7] 竹村彰通 (2006). 個票データの秘匿措置と開示データの利用に関する研究. 科学研究費(基盤研究(A)(1) 課題番号14208023) 研究成果報告書. 全386ページ.
- [8] Aoki, S., Takemura, A. and Yoshida, R. (2005). Indispensable monomials of toric ideals and Markov bases. Technical Report METR 05-34, November 2005.
- [9] Takemura, A. and Yoshida, R. (2006). Fundamental holes and saturation points of a commutative semigroup and their applications to contingency tables. Technical Report METR 06-15, March 2006.
- [10] Fushiki, T., Komaki, F., and Aihara, K. (2005). Nonparametric bootstrap prediction, *Bernoulli*, vol. 11, 293–307.
- [11] Tanaka, F., and Komaki, F. (2005). Bayesian predictive density operators for exchangeable quantum statistical models, *Physical Review A*, vol. 71, 052323.
- [12] Komaki, F. (2006). Shrinkage priors for Bayesian prediction, *the Annals of Statistics*, in press.
- [13] Komaki, F. (2006). A class of proper priors for Bayesian simultaneous prediction of independent Poisson observables, *Journal of Multivariate Analysis*, in press.
- [14] Kobayashi, K. and Komaki, F. (2006). Information criteria for kernel machines, *IEEE Trans. on Neural Networks*, in press.
- [15] Sei, T. (2005). Asymptotically efficient estimate of the fractal index of Gaussian processes, Proceedings of the 2nd International Symposium on Information Geometry and its Applications, University of Tokyo, Dec. 12–16, 2005.