# はじめに

筆者は長年、統計科学の基礎理論とその諸 分野への応用に従事してきた。この流れで、数 理科学と産学連携の雑感を述べる。

時と共に変動する偶然量の観測値の系列を 時系列という。数学的にはこの系列を1つの確 率過程の実現したものとみなす。確率過程の 統計解析を時系列解析という。通常の統計学 の議論は、主に独立標本に対する議論である が、時系列解析は過去、現在、未来の系列が 互いに影響しあっている状況での統計解析で、 自然科学、環境、工学、金融・経済、生物・医学、 発達心理等、極めて広汎な分野のデータに適 用可能である。つまり、時系列解析は産学連携 の基礎を担うものといえる。

## 最適推測論

初期の時系列解析では、特殊なモデルに対して正規性等の都合のよい条件を課し、遂行された統計推測も極めて原始的なもので、最適性への理論的裏付けも極めて脆弱なものであった。二十数年ほど前から、独立標本でLeCam教授によって開発されたLAN性に基づくアプローチが、ベルギーのHallin教授や筆者によって広汎な確率過程モデルに対して展開され、確率過程に対する最適統計推測論がシステマチックに構築されてきた([1],[2])。確率過程の長さnの観測系列の確率分布をPnで表し、これ

が、未知母数ベクトル $\theta$ で規定されているとし、 $P_n(\theta)$ と表す。近接する対立仮説の列: $\theta_n = \theta + h/\sqrt{n}$ , (hは定数ベクトル)、に対して対数尤度 比 $\Lambda = \log\{P_n(\theta_n)/P_n(\theta)\}$ がnを大きくした場合の 確率展開

 $\Lambda=h'\triangle(\theta)-0.5h'F(\theta)h+o_p(1)$ を持つとき確率分布族 $\{P_n(\theta)\}$ はLAN性を持つという。ここに $\Delta=\triangle(\theta)$ は確率変数ベクトルでcentral sequenceと呼ばれ、 $F=F(\theta)$ はノンランダムな行列でFisher 情報行列と呼ばれる。LAN性の下、最適な推定量、検定統計量、判別統計量等、の(漸近)最適性は、 $\Delta$ とFの言葉で記述され極めて統一的な最適統計手法が与えられる。この議論の流れで推定に関しては、広汎な確率過程モデルに対して最尤推定量が最適であることが示されている。

# 生体解析への応用

文献[2]は多変量金融データの解析で、次式で定義されるm次元非正規、非線形時系列モデル (CHARN Model):

 $X_t = F_{\theta}(X_{t-1}, \cdots, X_{t-p}) + H_{\theta}(X_{t-1}, \cdots, X_{t-q}) U_t$ を提案した。ここに $\{U_t\}$ は独立で同分布に従う m次元確率変数列、 $F_{\theta}$ と $H_{\theta}$ は、それぞれm次元 ベクトル値、mxm 行列値可測関数、 $\theta$ は未知母 数ベクトルである。このCHARNモデルは極めて 一般的な時系列モデルで、文献 [4] は、これに 対して LAN 性に基づいて最適推測論を展開 し、最尤推定量の最適性を示した。さらに、この 理論結果を、ある疾病患者の脳波と筋電波の解析に適用して(図1を参照)、大脳皮質一次運動野からの手先運動を示す表面筋電位への意味ある情報伝達機構を明らかにした。

CHARNモデルは、元々金融データの解析で提案されたものであるが、筆者等の論文で生体波のデータ解析にも有効であることを見た。これは、極めて汎用性のあるモデルであるので、最適に推測された CHARNモデルは、金融や医学・生体解析だけでなく、自然科学、工学、環境等、膨大な分野に応用可能と思われる。



# 金融解析への応用

時系列解析の重要な応用先として金融統計解析がある。実証分析の立場からは、金融収益率データは、非正規、非定常、従属過程の兆候を持つ。これを数理モデルで記述するとすれば、

 $X_{t,T} = \int_{-\pi}^{\pi} e^{-it\lambda} A_{t,T}(\lambda) d\xi(\lambda),$ 

が典型的なものの1つであろう。ここに、tは時刻を表し、 $t=1,2,\cdots,T$ 、とする。 $A_{t,T}(\lambda)$ は応答関数で基準化された滑らかな関数  $A(u,\lambda)$ 、(u=t/T)、でTを大きくする漸近展開で近似されるとする。また $\xi(\lambda)$ は直交増分過程とする。この場合  $\{X_{t,T}\}$  の時変スペクトル密度関数  $f(u,\lambda)$  は  $A(u,\lambda)$ のゲインとして定義される。

現実の場面で、 $\{X_{t,T}\}$ が次の時変スペクトル密度行列に関する2つのカテゴリー

 $\Pi_1$ :  $f(u,\lambda)$ ,  $\Pi_2$ :  $g(u,\lambda)$ 

# 数理科学と産学連携の

理工学術院 基幹理工学部 応用数理学科 教授 谷口正信



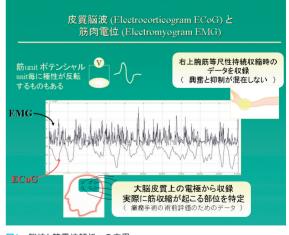


図1 脳波と筋電波解析への応用

のどちらかに属することが分かっているが、その どちらかが不明のとき、{X<sub>+</sub> <sub>+</sub>}が∏<sub>+</sub>に属するか、 Ⅱ。に属するかを観測系列から判別したいとい う事は諸分野でよく見受けられる。文献[6]で は、観測データが2つの時変スペクトル f=f(11 \lambda) とg=g(u, λ)の、どちらに近いかを測るある種の 距離 D(f:g)を導入して、D(f:g)>0ならば、データ は∏1に属し、D(f:g)≤0なら、∏2に属するという 判別方式を導入した。この判別方式の誤判別 確率がfとgの言葉で評価され、最適判別も議 論された。判別カテゴリーを2群から一般のk 群、すなわち  $\Pi_i$ :  $f_i(u,\lambda)$ ,  $j=1,\cdots,k$ , にする 拡張も同様に可能である。また実際の問題では カテゴリーを記述する時変スペクトル密度関数 も未知である。このような場合、各カテゴリーに 属することが分かっている観測系列が得られる とき、これらから、 $f_i$ の非母数的な推定量 $\hat{f}_i$ を構 成できる。文献[3]は日本の13社の日次収益率 データに対して $\hat{f}_i$ を求め、 $D(\hat{f}_i:\hat{f}_i)$  を用いた階 層的分類(クラスター)分析を行い結果を樹形図 (図2)で表し、13社が業種ごとに明快に分類さ れることを示した。

近年、金融工学においては企業の財務デー タに基づく格付けの研究が盛んになってきてい るが、その多くは、原始的な独立標本に基づく 判別・クラスター解析法を用いており、今後は上 記のような従属構造や非定常構造を本質的に 用いた解析が期待できよう。また大学と金融企 業、医学・製薬企業との連携も、双方にとって高 度な研究発展の中で可能になろう。



# 年金分野への応用

我が国の120兆円におよぶ巨大な年金積立 金は5つの金融資産に分散投資され増益を得 るべく運用されている。金融収益率は、非正規、 非定常過程で記述されるのが自然であるの で、前述の多変量非定常過程X<sub>+T</sub>で表せるとし て、投資比率を並べたベクトルαで線形結合(ポ ートフォリオ) a'X+Tを作り、これを最大化するaを 求めることがもくろまれている。図3は変動の大 きい資産1と変動の小さい資産2とその算術平 均で定義されるポートフォリオの動きをプロットし たものである。これにより、ポートフォリオが資産 の変動を小さくして、変動が小さいが収益の低 い資産を改善することが容易に見えよう。

ポートフォリオは確率変数なので、ポートフォリ オの平均 $\mu(a)$ と分散V(a)を考慮した最適ポート フォリオが提案されている。最も代表的なものは

 $a_{opt} = arg \max_{\alpha} \{ \mu(\alpha) - \beta V(\alpha) \}$ で定義される。ここにβは解析者が与える正定 数である。従って、α<sub>ont</sub>は、収益率過程の平均μ と分散 V の関数で表される。 すなわち  $\alpha_{ont}$ = g(u, V) の形である。従来は、金融収益率過程 が独立であることを想定してµ、Vをそれぞれ標 本平均ûと標本分散Ŷで置き換えたポートフォリ オ推定量  $\hat{a}_{opt} = g(\hat{\mu}, \hat{V})$  を使った。しかしなが ら、金融収益率が従属性を持つ場合は、ântは 一般に最適推定量にならないことが示され、文

献[5]は前述の非正規、非定常収益率過程に 対して最適ポートフォリオ推定量を構成した。こ れは従来の年金・金融解析への重要な警告 と、年金積立金の運用ポートフォリオ構成にも新 しい視点を与えている。



# ( ) 数理科学と産学連携

前節で見たように数理科学の部分を構成す る統計科学の文脈だけでも、生体・医学、金融、 年金等に広汎な応用をもつ。当然、これらに関 係する企業、産業との連携は可能で、筆者は既 に年金積立金運用の法人と協業している。実 はここで述べた応用、協業はさらに広がり、工 学、自然科学、環境、保険、遺伝子、薬学、経済 などデータが出る分野すべてに応用可能で、 関連する産業と連携が可能である。前述の遺 伝子データは、ノンランダムなものであるが、既 に遺伝子データに時系列手法を適用して、 DNA塩基のある種の周期性や特性検出など も試みられており、これに基づいた病理判別な ども可能であろう。このことは、時系列手法に基 づいて、数論分野で素数の現れ方、間隔、その 独立性の検定などが議論できる可能性を示唆 している。つまり、大学での研究結果を産業分 野に応用するだけにとどまらず、逆に数学や数 理基礎理論への新しいブレークスルーが生ま れるものと思われる。

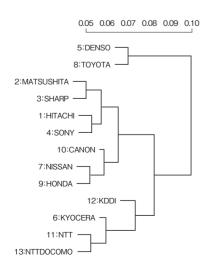


図2 株価収益率による企業の分類

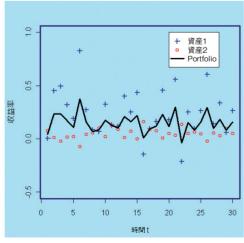


図3 ポートフォリオの動き

# 参考文献

- [1] Hallin, M., Ingenbleek, J.F. and Puri, M.L. (1985). Linear serial rank tests for randomness against ARMA alternatives. Ann. Statist. 13, 1156-1181.
- [2] Härdle, W., Tsybakov, A. and Yang, L. (1998). Nonparametric vector autoregression. J. Statist. Plann, Inference, 68, 221-245.
- [3] Hirukawa, J. (2006). Cluster analysis for non-Gaussian locally stationary processes. Int. J. Theor. Appl. Finance, 9, 113-132.
- [4] Kato, H., Taniguchi, M. and Honda, M. (2006). Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. IEEE Trans. Signal Process., 54, 3414-3425.
- [5]Shiraishi, H. and Taniguchi, M. (2007). Statistical estimation of optimal portfolios for locally stationary returns of assets. Int. J. Theor. Appl. Finance, 10, 129-154.
- [6] Taniguchi, M., Hirukawa, J. and Tamaki, K. (2008). Optimal Statistical Inference in Financial Engineering. Chapman & Hall/CRC, New York.
- [7] Taniguchi, M. and Kakizawa, Y. (2000). Asymptotic Theory of Statistical Inference for Time Series. Springer-Verlag, New York.