論 文 の 内 容 の 要 旨

論文題目 **On a non-linear risk analysis based on the theory of KM$_2$O-Langevin equations**

（KM$_2$O-ランジュヴァン方程式論に基づく非線形リスク解析について）

氏名　鈴木　健二郎

時系列に潜む異常性の前兆を検出する体系的な方法論は，さまざまな応用分野において大いに役立つと考えられる．たとえば，株価指標のような金融時系列から金融市場の暴落の前兆をつかむことができれば，市場に参加する投資家は何らかの行動を事前に起こし，重大な損失から免れることができるかもしれない．また，予防医学の分野では，疾患の早期発見が治療の成功において重要な鍵を握っているであろうことは間違いない．腎疾患の指標である血清クレアチニン値などの時系列において，そこに潜む異常性を事前につかむことができたならば，人工透析などの施術に関する意思決定において，大きな後ろ盾となるであろう．あるいは，地震学においては誰しもが日々感じているように，異変の前兆を地震データから発見することができれば，状況に的確に対処することで，災害を最小限に抑えることができるかもしれない．

この論文の目的は，さまざまな時系列に潜む異常の前兆を定量的に検出するための，非線形リスク解析を展開し，先行研究である KM$_2$O-ランジュヴァン方程式論に基づく非線形異常解析を補強することにある． 東京大学大学院情報理工学研究科の岡部ら（2002）は，異常性を「時系列の背後にある確率過程における定常性の破れの度合い」と定義し，その前兆を検出するテスト Test(ABN) を提案した．Test(ABN) の結果は異常グラフと定常グラフという 2 つのグラフを通じて解析される．異常グラフでは，実線が実データのチャートを表しており，灰色の網掛けは定常性のテスト Test(S) を通過しなかった箇所を示している．一方定常グラフでは，実線で Test(S) を通過した 2 次元時系列の個数が表されており，定常性の破れの度合いを調べることができる．実際，岡部らは，いくつかの株式市場指標から市場崩壊の前兆を探るという点で前向きな成果を得たが，実験結果には問題点も残されていた．その一つとして，定常グラフにおいて，実際の異常の前兆として検出される箇所のほかにも，定常グラフが急激に下がったり，あるいは 0 になる箇所が現れ，それらの見分けがつかないということである．これは，定常性の破れという観点とは別の尺度により，定量的に異常の前兆を調べる必要性があることを示唆している．

本論文では，この問題を解決するために，異常性の前兆を検出するもうひとつの方法を導入する．理論的背景となるのは，松浦-岡部 (2001) に基づく確率過程に対する非線形予測解析である．まず，非線形予測誤差によって定義される「リスク」の概念を導入する．これは，過去の情報を用いた予測の精度の低下をもって，リスクの拡大と解釈するところにその礎がある．確率過程に対する予測誤差を計算する理論について簡単に説明する．$\mathbf{X} = (X(n); 0 \le n \le N)$ を確率空間 $(\Omega, \mathcal{B}, P)$ 上で定義された 2 次元確率過程とする．非線形予測問題とは，p 期先の非線形予測子 $P_{\mathbf{N}_0^n(\mathbf{X})} X(n+p)$ をいかに計算するか，ということであるが，この非線形予測子は，条件付期待値 $E(X(n+p)|\mathcal{B}_0^n(\mathbf{X}))$ に対応している．この問題を解決するために，松浦-岡部は，非線形情報空間 $\mathbf{N}_0^n(\mathbf{X})$ の生成系を導入した．そこでは，階数有限の非線形変換された多次元確率過程 $\tilde{\mathbf{X}}^{(q)}$ が構成され，これを用いた有限次元の非線形予測子 $P_{\mathbf{M}_0^n(\tilde{\mathbf{X}}^{(q)})} X(n+p)$ の計算アルゴリズムが提案されており，$P_{\mathbf{M}_0^n(\tilde{\mathbf{X}}^{(q)})} X(n+p)$ の q を無限大に飛ばすことにより，理論的には $P_{\mathbf{N}_0^n(\mathbf{X})} X(n+p)$ が得られることになる．さらに岡部-金子 (2001) は，この理論を多次元確率過程に拡張し，気象データを用いて実証分析を行っている．

非線形予測子の精度を測る際，我々は非線形予測誤差 $e(\mathbf{X}|\mathbf{N}_0^n(\mathbf{X}))(n+p) \equiv ||X(n+p) - P_{\mathbf{N}_0^n(\mathbf{X})} X(n+p)||^2$ を計算する．予測誤差は，通常予測値と実現値との乖離を定量的に表したものであり，その値は $0 \le e(\mathbf{X}|\mathbf{N}_0^n(\mathbf{X}))(n+p) \le ||X(n+p)||^2$ を満たす．この非線形予測誤差の近似として，我々は先の階数有限の非線形予測子から階数有限の非線形予測誤差 $e(\mathbf{X}|\mathbf{M}_0^n(\tilde{\mathbf{X}}^{(q)}))(n+p)$ を計算することができる．これを時間経過とともに繰り返し計算した際，その値が以前よりも小さくなった場合，$\mathbf{X}$ の将来の挙動に対する非線形予測子の説明力があることを示唆しており，逆に予測誤差に増加傾向が現れた場合，予測子が十分には将来の挙動を説明できなくなってきていることを示唆する．そこで， 本研究ではこのメカニズムに着目し，時系列の異常解析に積極的に応用することをねらいとした．すなわち, 解析の対象となる時系列の背後にある確率過程において，何らかの異常の前兆が潜在的に存在するとすれば，そこに非線形予測子の説明力が低下し，非線形予測誤差の拡大が現れるのではないかと考えたのである．したがって，時系列に対して Test(ABN) と同様，時間域をシフトしながら一定の階数有限の非線形予測誤差を計算し，その挙動を調べることで，定常性の破れとは別の観点から異常性の前兆を探ることができ，異常性解析で先に挙げた問題点の一解決策となると考えられる．

非線形リスク解析の概要を説明する．前述の自然数 L $(1 \le L \le N)$ と各 $s(L \le s \le N)$ に対し，時系列 $\mathcal{Z} = (\mathcal{Z}(n); 0 \le\le N)$ から時系列 $\mathcal{Z}^{(t)} = (\mathcal{Z}^{(t)}(n+t-L); 0 \le n \le L)$ を抜き出す．さらにその時系列から見本共分散行列関数 $R^{\mathcal{Z}^{(t)}\tilde{\mathcal{X}}_{0j}^{(t)}}$ を計算し，これに基づいてその時間域の階数有限の非線計予測誤差 $e(\mathcal{Z}^{(t)})(L_{0j}) \equiv R_{11}^{\mathcal{Z}^{(t)}\tilde{\mathcal{X}}_{0j}^{(t)}} - C_{N_{0j}}(\mathcal{Z}^{(t)}|\tilde{\mathcal{X}}_{0j}^{(t)})^2$ を計算する．ここで， $R_{11}^{\mathcal{Z}^{(t)}\tilde{\mathcal{X}}_{0j}^{(t)}}$ は，$\mathcal{Z}^{(t)}$ の分散，$C_{N_{0j}}(\mathcal{Z}^{(t)}|\tilde{\mathcal{X}}_{0j}^{(t)})$ は階数有限の非線形予測子の自乗ノルムを表す．さらに，時間域をシフトしながらこの値をプロットすることで，その推移を一つのグラフに描くことができる．これをリスク関数 $RF(\mathcal{Z})$ といい，そのグラフを RFgraph と呼ぶことにする．

非線形リスク解析は，d 次元確率過程 $\mathbf{X}$ の非線形情報解析に基づいている．本論文では，これまでの非線形情報解析で確立された生成系とは別の生成系を構成することで，確率過程の構造に関して新たなアプローチを行っている．まず，以下のような直交分解を考える：

$$X(n) = P_{\mathbf{M}_0^{n-1}(\tilde{\mathbf{X}}^{(q)})} X(n) + \eta_+(\mathbf{X}|\tilde{\mathbf{X}}^{(q)})(n). \tag{1}$$

この分解の第 1 項の具体的な表現により，$\mathbf{X}$ の階数有限の非線形構造を把握することができることになり，先行研究におけるさまざまな非線形解析では，この分解が基礎となっている． 我々はここで (1) を第 1 種直交分解と呼ぶことにする．しかしこれに加えて，$\tilde{\mathbf{X}}^{(q)}$ に付随する前向き KM$_2$O-ランジュヴァン揺動過程 $\nu_+(\tilde{\mathbf{X}}^{(q)}) = (\nu_+(\tilde{\mathbf{X}}^{(q)})(n); 0 \leq n \leq N)$ の非線形構造の解析を通じて，$\mathbf{X}$ についての構造を解析する方法を導入する．$\nu_+(\tilde{\mathbf{X}}^{(q)})$ の典型的な性質は，直交性である．したがって，$\mathbf{M}_0^{n-1}(\nu_+(\tilde{\mathbf{X}}^{(q)}))$ の線形情報からは構造について新たに何かをつかむことはできない．しかし，(1) は，$\nu_+(\tilde{\mathbf{X}}^{(q)})$ の独立性については何も保証するものでなく，したがって非線形な相関構造の存在は否定されていない．そこで，次のようなもう一つの直交分解を考える：

$$X(n) = P_{\mathbf{M}_0^{n-1}(\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')})} X(n) + \eta_+(\mathbf{X}|\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')})(n), \tag{2}$$

ここで，$\mathbf{M}_0^{n-1}(\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')})$ は，$\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')}$ の線形情報空間であり，次の因果関係が成り立つ：

$$\mathbf{M}_0^{n-1}(\tilde{\mathbf{X}}^{(q)}) \subset \mathbf{M}_0^{n-1}(\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')}). \tag{3}$$

この関係式は， (2) に基づいた解析の方が (1) よりも，$\mathbf{X}$ の構造についてのより詳細な情報が得られる可能性を示唆している． (1) を，第 1 種直交分解と呼ぶのに対して，(2) を第 2 種直交分解と呼ぶことにする． これら 2 種類の分解式に基づいて，それぞれから階数有限の非線形予測誤差，$e(\mathbf{X}|\mathbf{M}_0^n(\tilde{\mathbf{X}}^{(q)}))(n+p)$ および $e(\mathbf{X}|\mathbf{M}_0^n(\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')}))(n+p)$ を計算することができる．これらを，それぞれ p 期先の第 1 種および第 2 種非線形予測誤差と呼ぶ.

さらに，これを時系列解析に応用してリスク関数を計算する場合にも，同様に 2 種類のリスク関数を定義することができる．時系列に対して第 1 種の 階数有限の非線形予測誤差から得られるリスク関数を，第 1 種リスク関数，そのグラフを RF graph 1 と呼び，第 2 種の場合も同様，第 2 種リスク関数，そのグラフを RF graph 2 と呼ぶことにする．

本論文の内容を章ごとに追って説明する：

第 2 章 Non-linear analyses for stochastic process （確率過程に対する非線形解析）では，KM$_2$O-ランジュヴァン方程式論の中の理論的基礎となる先行研究と，本研究での新たな理論展開について説明する.

2.1 節 KM$_2$O-Langevin equations associated with stochastic process（確率過

程に付随する KM_42O-ランジュヴァン方程式) ではまず，非退化な場合の d-次元確率過程 $\mathbf{X}$ に対して，線形情報空間 $\mathbf{M}_0^n(\mathbf{X})$ を導入し，$\mathbf{X}$ の線形の時間発展の表現として，これに付随する KM_2O-ランジュヴァン方程式を導く．さらに，当該確率過程が定常性を持つ場合について考察し，散逸散逸定理 (DDT)，揺動散逸定理 (FDT) および偏自己相関関係式 (PAC) について説明する (Okabe(1988), Okabe(1988), Okabe(2000))．つづいて，退化した確率過程に対する理論の展開について言及する (Matsuura-Okabe(2001)).

2.2 節 Non-linear information analysis （非線形情報解析） では，確率過程 $\mathbf{X}$ の非線形情報空間 $\mathbf{N}_0^n(\mathbf{X})$ を導入し，この解析により非線形構造を調べる．まず，岡部靖憲教授と松浦真也助手によって展開された，$\mathbf{N}_0^n(\mathbf{X})$ の生成系 (生成系 (1)) を用いた非線形情報解析の手法について説明する (Matsuura-Okabe(2001))．つづいて，筆者と岡部靖憲教授による新しい研究成果として，確率過程 $\tilde{\mathbf{X}}^{(q)}$ に付随する KM_2O-ランジュヴァン揺動過程 $\nu_+(\tilde{\mathbf{X}}^{(q)})$ を用いた第二の生成系 (生成系 (2)) について説明する．本研究ではこれに基づき，元の確率過程の非線形変換の階数を固定した上で，KM_2O-ランジュヴァン揺動過程に潜む非線形情報を抜き出すことにより，従来の方法では捉えきれない複雑な構造を解析する．

2.3 節 Non-linear prediction analysis （非線形予測解析） では，最初に生成系 (1) を用いた従来のアルゴリズムにより，階数有限の非線形予測子 $(1)P_{\mathbf{N}_0^n(\mathbf{X})}X(n+p)$ を求め，その階数 q に関する極限として，$P_{\mathbf{N}_0^n(\mathbf{X})}X(n+p)$ を求める非線形予測公式 (1) (Okabe-Kaneko(2000), Matsuura-Okabe(2001)) を紹介する．その後，それに基づく非線形予測誤差 (1) を求める．つづいて前節で新たに導入された生成系 (2) に基づき，KM_2O-ランジュヴァン揺動過程を用いて階数有限の非線形予測子 $P_{\mathbf{M}_0^n(\nu_+(\tilde{\mathbf{X}}^{(q)})^{(q')})}X(n+p)$ を求め，q' に関する極限として，非線形予測公式 (2) を導く．さらにそれに基づく非線形予測誤差 (2) を求める．

第 3 章 Non-linear analyses for time series （時系列に対する非線形解析） では，時系列の異常性の前兆を定量的に探るための新たな手法として，リスク関数を導入する．

3.1 節 The first kind of risk function （第 1 種リスク関数） では，1 次元時系列 $\mathcal{Z}$ に対し，第 1 種リスク関数を計算する手順を説明する．まず最初に，cut length の長さだけ切り出したものを $\mathcal{Z}^{(t)}$ とする．これに対して階数 6 の非線形変換を施すことで，19 個の 1 次元時系列 $\mathcal{Z}_i^{(t)}$ を得る．時間域を変更した後，2 次元の時系列 $\tilde{\mathcal{X}}_{0j}^{(t)}$ を構成し，これに対して Test(S) を実行する．Test(S) を通過したものに対して，見本共分散行列関数 $R^{\mathcal{X}_0^{(t)}\tilde{\mathcal{X}}_{0j}^{(t)}}$ を求め，これを用いて第 1 種非線形予測値 $\hat{\mathcal{X}}_0^{(t,i)}(L_j)$ を計算する．その予測誤差として，第 1 種リスク値 $r^{(1)}$ を定義する．最後に時間域をシフトしながら，その変動を記録することによって，第 1 種リスク関数のグラフを得る．

3.2 節 The second kind of risk function （第 2 種リスク関数） では，前節の第 1 種リスク値を計算する過程で得られる KM_42O-ランジュヴァン揺動列 $\tilde{\xi}(\tilde{\mathcal{X}}_{0j}^{(t,L_j-N_{0j})})$ の非線形情報を利用して第 2 種リスク値を求める手順について説明する．まず，$\tilde{\xi}(\tilde{\mathcal{X}}_{0j}^{(t,L_j-N_{0j})})$ の規格化を行い，階数 4 の非線形変換を施すことで，28 個の 1 次元時系列 $\xi(\tilde{\mathcal{X}}_{0j,k}^{(t)})$ を得る．こ

れに対して，第 1 種の場合と同様の手順をを踏み，第 2 種リスク値 $r^{(2)}$ を求める．時間域をシフトしながら，その変動を記録することにより、第 2 種リスク関数のグラフを得る．

第 4 章 Results of empirical analyses （実証解析の結果）　では，前章で導入した時系列解析の手法を各種実データに適用し，非線形リスク解析を行う．

4.1 節 Stock market indexes （株式市場指数）では，アメリカ，イギリスおよび日本各国の代表的株価指数として，Dow，Ftse100 および日経 225 の 1984 年〜1988 年の日次データを取り上げ，各種指数に対して異常性のテスト Test(ABN) を行った上で，第 1 種および第 2 種リスク関数を求めた．この時期には，大きな出来事として，1985 年 9 月 22 日のプラザ合意と 1987 年 10 月 19 日のブラックマンデーが含まれている．実験の結果，特に Dow と日経 225 のリスク関数において，プラザ合意までは減少傾向にあったが，その直後から一転して上昇傾向に変化し，ブラックマンデーの時期まで一貫して増加している現象が観察された．これは，Test(ABN) では区別できなかった異常性の前兆を捉えていると考えることができる．一方従来の市場リスク指標である Value at Risk では，そのような傾向は見られなかった．

4.2 節 Serum creatinine concentration （血清クレアチニン濃度）では，腎機能の指数である血清クレアチニン濃度のデータを取り上げ，前述と同様の実験を行う．その結果，腎機能の悪化を示唆する 2.0 の値を越える前で，リスク関数は一度減少した後に急上昇する傾向が得られる．この場合も Test(ABN) では区別できなかった異常性の前兆を捉えていると考えることができる．また，緊急人工透析の開始および離脱の前後で一貫してリスク関数は上昇しており，これらの施術判断への利用が期待できる．

4.3 節 Seismic waves （地震波）では，通常の地震と深部低周波地震という異なる 2 種類の地震波について，前述と同様の実験を行う．その結果，特に通常の地震において，上下成分では P 波到達の直前で一度減少した後で急上昇するという傾向が現れたが，一方で南北・東西成分ではその傾向が現れず，P 波の前兆を捉えた Test(ABN) の結果を補強するものとなった．また S 波到達の直前では，上下・南北成分では大きな上昇傾向が得られたが東西成分では上昇傾向が現れなかった．この結果は Test(ABN) からは捉えられなかった変化であるが，東京大学地震研究所の武尾実教授により指摘された，震源に対する観測地点の位置関係に起因する地理的条件を示唆している可能性がある．